

ΤΕΙ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ / ΜΕΣΟΛΟΓΓΙ

Πτυχιακή εργασία

ΠΡΟΒΛΕΨΗ ΜΕΛΛΟΝΤΙΚΩΝ ΠΩΛΗΣΕΩΝ
ΕΠΙΧΕΙΡΗΣΗΣ ΜΕ ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ
ΔΕΔΟΜΕΝΩΝ

Ελένη Μαρία Λεπεσιώτη

Μεσολόγγι 2017

ΤΕΙ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ / ΜΕΣΟΛΟΓΓΙ

Πτυχιακή εργασία

ΠΡΟΒΛΕΨΗ ΜΕΛΛΟΝΤΙΚΩΝ ΠΩΛΗΣΕΩΝ
ΕΠΙΧΕΙΡΗΣΗΣ ΜΕ ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ
ΔΕΔΟΜΕΝΩΝ

Ελένη Μαρία Λεπεσιώτη

Επιβλέπων καθηγητής:
Νικόλαος Καρούσος

Μεσολόγγι 2017

Η έγκριση της πτυχιακής εργασίας από το Τμήμα Διοίκησης Επιχειρήσεων/Μεσολογγίου του ΤΕΙ Δυτικής Ελλάδας δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

ΠΕΡΙΛΗΨΗ

Η πρόβλεψη έχει πολύ μεγάλη σημασία στη ζωή του ανθρώπου και παράλληλα στις επιχειρήσεις. Οι προβλέψεις χρησιμοποιούνται ως υποστήριξη αποφάσεων, επιχειρηματικών ή μη. Πλέον οι πληροφορίες θεωρούνται ανεκτίμητες για την οικονομική εξέλιξη και την ανταγωνιστική δύναμη μιας επιχείρησης. Οι πληροφορίες απορρέουν από δεδομένα, για τα οποία οι επιχειρήσεις καταναλώνουν κεφάλαιο και πόρους για να τα αποκτήσουν, να τα αποθηκεύσουν και να τα διαχειριστούν. Από την επεξεργασία των δεδομένων διεξάγονται πληροφορίες, οι οποίες επίσης βοηθούν στην ανάπτυξη προβλέψεων οι οποίες ουσιαστικά χρησιμοποιούνται για βελτιστοποίηση των επιχειρηματικών αποφάσεων σχετικά με διάφορους σκοπούς και διεργασίες. Στο παρόν έγγραφο θα γίνει περιγραφή μεθόδων προβλέψεων που βασίζονται στην ύπαρξη ιστορικών δεδομένων. Κατά τη διαδικασία ανάπτυξης μιας πρόβλεψης θα πρέπει να ακολουθηθούν μερικά βασικά βήματα. Εφόσον ο σκοπός για τον οποίο γίνεται η πρόβλεψη γίνει κατανοητός, τα μοντέλα που θα αναπτυχθούν θα πρέπει να εφαρμοστούν στα δεδομένα και να πληρούν συγκεκριμένες προϋποθέσεις. Τα αποτελέσματα θα πρέπει να αξιολογούνται σχετικά με την απόδοση τους με χρήση διάφορων κριτηρίων και ελέγχων ακρίβειας. Τα κριτήρια αυτά συνήθως περιγράφουν τα αποτελέσματα μαθηματικών εξισώσεων που, σαν εισόδους, δέχονται τα σφάλματα των προβλέψεων, δηλαδή τις αποκλίσεις των προβλεπόμενων δεδομένων από τα πραγματικά. Για την τελική επιλογή του κατάλληλου μοντέλου πρόβλεψης, η ανάπτυξη πολλαπλών μοντέλων από διαφορετικές μεθόδους και η σύγκριση αυτών έχει θεμελιώδης σημασία.

Καταρχάς, τα ιστορικά δεδομένα που θα συμπεριληφθούν στα μοντέλα πρόβλεψης θα πρέπει να είναι απολύτως κατανοητά. Επειδή όμως τα δεδομένα αυτά μπορεί να υπάρχουν διαθέσιμα σε μεγάλο όγκο και με διάφορα σφάλματα, κατά τα αρχικά βήματα μιας διαδικασίας πρόβλεψης συνηθίζεται η χρήση εξερευνητικής ανάλυσης με μεθόδους περιγραφικής στατιστικής. Η εξερευνητική στατιστική βοηθά στον σχηματισμό της εικόνας που αναμένεται να έχουν τα αποτελέσματα μιας πρόβλεψης, ώστε να μπορεί να γίνει προσδιορισμός προβλημάτων, εφόσον προκύψουν. Στη μελέτη της περιγραφικής στατιστικής υπάγονται τα γραφήματα, τα οποία βοηθούν ιδιαίτερα στην κατανόηση των δεδομένων, απεικονίζοντας τη γενική εικόνα καθώς και προσφέροντας ενόραση για κρυφά μηνύματα που μπορεί να περιλαμβάνονται. Υπάρχουν πολλά είδη γραφημάτων για διαφορετικής φύσεως δεδομένα, ενώ η μορφή που θα έχει η απεικόνιση των δεδομένων εξαρτάται από τον τύπο

του γραφήματος. Στην περιγραφική στατιστική επίσης υπάγονται τα ιδιαίτερα δημοφιλή μέτρα μεταβλητών όπως είναι ο αριθμητικός μέσος, η διάμεσος, η επικρατούσα τιμή, τα τεταρτημόρια, η τυπική απόκλιση και ο συντελεστής συσχέτισης.

Με την ανάλυση των δεδομένων, αυτά μπορούν να προετοιμαστούν ώστε να υπάρξουν στην κατάλληλη δομή και διάσταση για να εισαχθούν σε μοντέλα πρόβλεψης. Μπορεί να κριθεί αναγκαίο να πραγματοποιηθούν περαιτέρω τροποποιήσεις στα δεδομένα, όπως ημερολογιακές μετατροπές ή λογαριθμικές μεταμορφώσεις, ώστε να ενισχυθεί η στατιστική τους σημασία. Εφόσον έχει πραγματοποιηθεί ανάκτηση των δεδομένων, η επεξεργασία, η περίληψη και η κατανόηση τους, θεωρητικά έχουν χτιστεί τα θεμέλια για την ανάπτυξη μιας πρόβλεψης. Υπάρχουν διάφορες μέθοδοι πρόβλεψης, η υλοποίηση αυτών όμως εξαρτάται από τη φύση των δεδομένων και τον σκοπό της πρόβλεψης. Οι μέθοδοι πρόβλεψης που αναλύονται σε αυτό το έγγραφο είναι η πρόβλεψη του ενός βήματος ή εξομάλυνση των κινούμενων μέσων, η εκθετική εξομάλυνση, η γραμμική παλινδρόμηση και τα στοχαστικά μοντέλα χρονολογικών σειρών ARMA και ARIMA.

Οι παραπάνω μέθοδοι πρόβλεψης εφαρμόζονται σε πραγματικά δεδομένα επιχείρησης, στο πλαίσιο μελέτης περίπτωσης για την οποία προσεγγίστηκε μία επιχείρηση ηλεκτρονικού εμπορίου στην Αγγλία. Η επιχείρηση που εμπορεύεται στη βιομηχανία παιχνιδιών, πουλάει τα προϊόντα της απευθείας σε πελάτες μέσω ηλεκτρονικών καταστημάτων. Το πρόβλημα που αντιμετωπίζει αφορά κυρίως στη διαχείριση των εμπορευμάτων της και συγκεκριμένα τον προσδιορισμό της βέλτιστης χρονολογικά τοποθέτησης των παραγγελιών της στα εργοστάσια με τα οποία συνεργάζεται. Η αδυναμία της επιχείρησης να ανταπεξέλθει στη ζήτηση οδηγεί σε εξαντλήσεις αποθεμάτων και συνεπώς σε ζημία των εσόδων της και των κατατάξεων των σελίδων των προϊόντων της σε σχέση με τους ανταγωνιστές της. Αυτό που χρειάζεται η συγκεκριμένη επιχείρηση είναι ένα σύστημα υποστήριξης αποφάσεων που να πραγματοποιεί πρόβλεψη των μελλοντικών πωλήσεων της επιχείρησης και τον πλήρες προσδιορισμό της ζήτησης των προϊόντων της.

Αρχικά, περιγράφεται η διαδικασία συλλογής των δεδομένων και η οργάνωση αυτών. Στη συνέχεια πραγματοποιείται προκαταρκτική στατιστική ανάλυση και υποδεικνύονται διάφορες τεχνικές που μπορούν να εφαρμοστούν εύκολα ώστε να συνεισφέρουν στην κατανόηση και την προετοιμασία των δεδομένων. Αναπτύσσονται σχετικά γραφήματα με σκοπό να υποστηρίξουν τα αποτελέσματα περίληψης στατιστικών μέτρων και να παρουσιάσουν τη μορφή των δεδομένων της επιχείρησης που τελικά χαρακτηρίζονται ως

χρονολογικές σειρές με υψηλή εποχικότητα. Η φύση των δεδομένων καθιστά τη μέθοδο της εξομάλυνσης κινούμενων μέσων ανίκανη να υλοποιήσει τις προβλέψεις που απαιτούνται από τη φύση του προβλήματος. Ωστόσο, παρουσιάζεται το πώς εφαρμόστηκε παραλλαγή της μεθόδου ώστε να πραγματοποιηθεί ανάκτηση των «χαμένων» πωλήσεων που σημειώθηκαν σε περιόδους εξάντλησης αποθεμάτων. Έπειτα εφαρμόζεται επιτυχώς η μέθοδος της τριπλής εκθετικής εξομάλυνσης του Excel, συνοδευόμενη από πρακτικές όπως το τρεχούμενο άθροισμα και συντελεστές εποχικότητας, οι οποίες επανεμφανίζονται κατά την εφαρμογή των επόμενων δύο μεθόδων πρόβλεψης. Περιγράφεται η εφαρμογή της μεθόδου της πολλαπλής γραμμικής παλινδρόμησης, ενώ εξηγείται γιατί τα δεδομένα της επιχείρησης δεν είναι κατάλληλα για τη συγκεκριμένη μέθοδο. Τα στοχαστικά μοντέλα τελικά αποτελούν την ιδανική μέθοδο πρόβλεψης για τη μελέτη περίπτωσης, αφού η φύση των δεδομένων ταιριάζει με τη μονομεταβλητή φύση της μεθόδου.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ.....	iii
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	vi
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	x
ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ – ΣΤΙΓΜΙΟΤΥΠΩΝ.....	xi
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ.....	xiv
1 Εισαγωγή.....	1
1.1 Η πρόβλεψη	1
1.2 Η πληροφορία.....	2
1.3 Ο ρόλος της πρόβλεψης στον επιχειρηματικό κόσμο	3
1.4 Χρονολογικές σειρές	5
1.5 Στόχοι και μελέτη περίπτωσης	9
2 Είδη πρόβλεψης και βασικά εργαλεία	11
2.1 Βασικά βήματα για τη διαδικασία της πρόβλεψης.....	11
2.2 Είδη τεχνικών πρόβλεψης	13
2.3 Θεμελιώδης αρχές πρόβλεψης.....	14
2.4 Αξιολόγηση ακρίβειας μοντέλων και μεθόδων πρόβλεψης.....	16
2.4.1 Σφάλματα εξαρτώμενα από την κλίμακα	16
2.4.2 Ποσοστιαία σφάλματα.....	17
2.4.3 Κλιμακωτά σφάλματα	18
2.5 Σύγκριση μεθόδων αξιολόγησης ακρίβειας	19
3 Κατανόηση μεταβλητών και στατιστική ανάλυση	23
3.1 Απεικόνιση δεδομένων.....	23
3.2 Εξερευνητική ανάλυση δεδομένων	27
3.2.1 Ο αριθμητικός μέσος	29

3.2.2	Η διάμεσος.....	30
3.2.3	Η επικρατούσα τιμή.....	30
3.2.4	Τεταρτημόρια και εκατοστημόρια.....	31
3.2.5	Η τυπική απόκλιση	31
3.2.6	Συσχέτιση	32
3.3	Μεταμορφώσεις.....	35
3.4	Δομημένη ανάλυση δεδομένων	36
4	Μέθοδοι πρόβλεψης.....	39
4.1	Κινούμενοι μέσοι.....	39
4.1.1	Ακόλουθος κινούμενος μέσος	41
4.1.2	Κεντρικός κινούμενος μέσος.....	41
4.1.3	Σταθμισμένος κινούμενος μέσος	42
4.2	Εκθετική εξομάλυνση.....	42
4.3	Γραμμική παλινδρόμηση	43
4.3.1	Απλή και πολλαπλή γραμμική παλινδρόμηση	44
4.3.2	Στατιστική σημασία και μηδενική υπόθεση.....	45
4.3.3	Τυπικό σφάλμα εκτίμησης και συντελεστής προσδιορισμού	47
4.3.4	Προϋποθέσεις	48
4.4	Ανάλυση χρονολογικών σειρών	49
4.4.1	Η μείξη της αυτοπαλινδρόμησης με τους κινούμενους μέσους (ARMA)	50
4.4.2	Μεταβαλλόμενες και αμετάβλητες χρονολογικές σειρές.....	52
4.4.3	Ενσωμάτωση μεταβαλλόμενων διαδικασιών	53
4.4.4	Η διαδικασία ARIMA των Box και Jenkins.....	54
4.4.5	Επιλογή παραμέτρων και αξιολόγηση.....	55
5	Μελέτη περίπτωσης – ανάλυση	57
5.1	Επισκόπηση	57

5.1.1	Η επιχείρηση.....	57
5.1.2	Ίδρυση	58
5.1.3	Η Αγορά	58
5.1.4	Προμήθεια και αποστολές.....	59
5.1.5	Εποχικότητα	60
5.1.6	Διαχείριση αποθεμάτων.....	61
5.2	Στόχοι και προσδιορισμός του προβλήματος.....	63
5.2.1	Η γενική προσέγγιση	63
5.2.2	Απαιτήσεις χρηστών και ορισμός προδιαγραφών.....	64
6	Σχεδιασμός επίλυσης.....	66
6.1	Κατανόηση του προβλήματος	66
6.2	Συλλογή πληροφοριών	67
6.3	Προκαταρκτική – εξερευνητική ανάλυση.....	67
6.4	Επιλογή και προσαρμογή μοντέλων.....	69
6.5	Εφαρμογή και αξιολόγηση μοντέλων	69
7	Υλοποίηση – προετοιμασία και κατανόηση των δεδομένων.....	71
7.1	Συλλογή των δεδομένων.....	71
7.2	Οργάνωση των δεδομένων	73
7.3	Φιλτράρισμα των δεδομένων και περιγραφική στατιστική.....	81
7.3.1	Περιληπτικός πίνακας προϊόντος - Σύνταξη	81
7.3.2	Περιληπτικός πίνακας προϊόντος – Στατιστική Σύνοψη.....	84
7.3.3	Περιληπτικός πίνακας προϊόντος – Μορφοποίηση υπό όρους	87
7.3.4	Περιληπτικός πίνακας κατηγορίας.....	89
7.4	Απεικόνιση δεδομένων – γραφήματα.....	91
7.5	Απεικόνιση δεδομένων και αυτοματισμός – Δυναμικό Ταμπλό.....	97
7.6	Πολυμεταβλητή Ανάλυση	101

8	Υλοποίηση – Ανάπτυξη μοντέλων πρόβλεψης.....	106
8.1	Ανάκτηση «χαμένων» πωλήσεων	106
8.2	Εφαρμογή μεθόδου κινούμενων μέσων	108
8.3	Εφαρμογή τριπλής εκθετικής εξομάλυνσης.....	114
8.4	Εφαρμογή πολλαπλής γραμμικής παλινδρόμησης.....	119
8.5	Εφαρμογή στοχαστικών μοντέλων ARIMA	124
	ΕΠΙΛΟΓΟΣ.....	129
	ΠΑΡΑΡΤΗΜΑΤΑ.....	132
I.	Τεχνολογίες πληροφορικής	132
II.	Κώδικας R: Τροποποίηση δεδομένων για ανάκτηση «χαμένων» πωλήσεων	133
III.	Κώδικας R: ARIMA.....	136
IV.	Κώδικας R: Αξιολόγηση μοντέλου και γραφήματα.....	138
	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	144

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1:	Σύγκριση των αποδόσεων ακριβείας κοινών μεθόδων προβλέψεων	21
Πίνακας 2:	Περιεχόμενα της Περιγραφικής Στατιστικής	28
Πίνακας 3:	Τύποι δομημένης ανάλυσης δεδομένων	37

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ – ΣΤΙΓΜΙΟΤΥΠΩΝ

Εικόνα 1:	Αποσύνθεση χρονολογικών σειρών – εξόρυξη της τάσης	7
Εικόνα 2:	Αποσύνθεση χρονολογικών σειρών – εξόρυξη της εποχικότητας	8
Εικόνα 3:	Αποσύνθεση χρονολογικών σειρών – εξόρυξη του απροσδιόριστου	9
Εικόνα 4:	Σύγκριση κοινών μεθόδων πρόβλεψης πάνω σε πραγματικά δεδομένα επιχείρησης – Ανάπτυξη γραφήματος με χρήση της κονσόλας R	20
Εικόνα 5:	Γράφημα που συνδυάζει οκτώ σειρές δεδομένων από δύο διαφορετικές μεταβλητές με διπλούς κάθετους άξονες και πίνακα περιεχομένων	25
Εικόνα 6:	Γράφημα σωρευμένων στηλών για πολλαπλές σειρές (πάνω), χρονολογικό γράφημα με ενσωματωμένες προβλέψεις (κάτω), διάγραμμα πίτας (δεξιά).....	26
Εικόνα 7:	Η καμπύλη της κανονικής κατανομής.....	32
Εικόνα 8:	Παραδείγματα συσχέτισης	34
Εικόνα 9:	Επιχειρηματικές αναφορές με λεπτομέρειες πωλήσεων και κίνησης με φίλτρο το ASIN, όπως εμφανίζονται στην Amazon Seller Central.	73
Εικόνα 10:	Οργάνωση των αποθηκευμένων επιχειρησιακών αναφορών	74
Εικόνα 11:	Η αρχική δομή των πηγαίων δεδομένων	74
Εικόνα 12:	Παράδειγμα προβλήματος πολλαπλών SKU	77
Εικόνα 13:	Επίλυση του προβλήματος των πολλαπλών SKU.....	78
Εικόνα 14:	Εμφάνιση των αποτελεσμάτων της αναζήτησης του NetSuite περί κατηγορίας προϊόντων.....	79
Εικόνα 15:	Χρήση περίπλοκης αναζήτησης για τη λύση του προβλήματος μη ακεραιότητας των δεδομένων.....	80
Εικόνα 16:	Συμπληρωματικές στήλες.....	80
Εικόνα 17:	Εμφάνιση ενός προϊόντος ανά τις χρονολογικές εκδοχές του.....	82
Εικόνα 18:	Επιλογή προϊόντων για την εμφάνιση τους στον πίνακα προϊόντος	83
Εικόνα 19:	Απόκρυψη εγγραφών – χρήση της στήλης Exclude.....	84

Εικόνα 20:	Πίνακας τιμών με στατιστικές συναρτήσεις.....	86
Εικόνα 21:	Μορφοποίηση υπό όρους – Εφαρμογή	87
Εικόνα 22:	Μορφοποίηση υπό όρους – Αποτελέσματα	89
Εικόνα 23:	Εμφάνιση μιας κατηγορίας ανά τις χρονολογικές εκδοχές της.....	90
Εικόνα 24:	Πωλήσεις τεμαχίων για ένα προϊόν για τα έτη 2014 έως 2017 ανά εβδομάδα .	91
Εικόνα 25:	Γραμμικά γραφήματα (<i>line charts</i>) προϊόντος (πάνω) και κατηγορίας (κάτω) + πίνακας δεδομένων.....	93
Εικόνα 26:	Γραφήματα με μπάρες (<i>bar charts</i>) προϊόντος (πάνω) και κατηγορίας (κάτω)	95
Εικόνα 27:	Ιστόγραμμα κατανομής πωλήσεων	97
Εικόνα 28:	Ταμπλό και απεικόνιση δεδομένων – Δυναμικοί πίνακες.....	99
Εικόνα 29:	Ταμπλό και απεικόνιση δεδομένων – Δυναμικοί πίνακες – Αυτοματοποίηση	100
Εικόνα 30:	Ταμπλό και απεικόνιση δεδομένων – Γραφήματα.....	101
Εικόνα 31:	Πίνακας συσχετίσεων.....	102
Εικόνα 32:	Διάγραμμα διασποράς (συνεδρίες και πωλήσεις)	103
Εικόνα 33:	Πίνακας τιμών και γράφημα πρόβλεψης κυλιόμενου αθροίσματος (52 εβδομάδων) πωλήσεων ενός προϊόντος για περιόδους του παρελθόντος όπου η προσφορά αδυνατούσε να καλύψει τη ζήτηση	107
Εικόνα 34:	Πίνακας τιμών δείγματος μιας σειράς παρατηρήσεων από πωλήσεις τεμαχίων ενός προϊόντος και η εφαρμογή του εργαλείου κινούμενων μέσων	109
Εικόνα 35:	Τροποποίηση δεδομένων με χρήση παραλλαγμένης μεθόδου των κεντρικών κινούμενων μέσων και εισαγωγή του τυχαίου	111
Εικόνα 36:	Τροποποιημένες παρατηρήσεις με χρήση κώδικα R.....	113
Εικόνα 37:	Παραμετροποίηση πρόβλεψης τριπλής εκθετικής εξομάλυνσης – Excel.....	114
Εικόνα 38:	Πρόβλεψη τριπλής εκθετικής εξομάλυνσης – Excel – πίνακας αποτελεσμάτων	115
Εικόνα 39:	Πρόβλεψη τριπλής εκθετικής εξομάλυνσης – Excel – γράφημα	115

Εικόνα 40: Πρόβλεψη τριπλής εκθετικής εξομάλυνσης – Excel – αναγνώριση εποχικότητας	116
Εικόνα 41: Τρεχούμενο άθροισμα – απεικόνιση τάσης.....	117
Εικόνα 42: Ανάπτυξη πρόβλεψης εκθετικής εξομάλυνσης πάνω σε δεδομένα τρεχούμενων αθροισμάτων και πολλαπλασιασμός αποτελεσμάτων με εποχικούς δείκτες – γράφημα.....	118
Εικόνα 43: Πίνακας συντελεστών συσχέτισης	120
Εικόνα 44: Έξοδος πολλαπλής γραμμικής παλινδρόμησης.....	120
Εικόνα 45: Εφαρμογή συνάρτησης πολλαπλής γραμμικής παλινδρόμησης	121
Εικόνα 46: Εφαρμογή συντελεστών παλινδρόμησης σε ιστορικά δεδομένα και σύγκριση των προβλεπόμενων τιμών (διακεκομμένη γραμμή) με τις πραγματικές (+ γραμμική τάση)	122
Εικόνα 47: Γραφήματα πρόβλεψης πολλαπλής γραμμικής παλινδρόμησης σε λογαριθμικά μεταμορφωμένα δεδομένα (πάνω) και πρόβλεψη τριπλής εκθετικής εξομάλυνσης στα ίδια λογαριθμικά δεδομένα (κάτω).....	123
Εικόνα 48: Συντελεστές εποχικότητας.....	124
Εικόνα 49: Ταμπλό επιλογής δεδομένων για πρόβλεψη – είσοδος R.....	125
Εικόνα 50: Κονσόλα R και υπολογισμός <code>auto.arima()</code>	126
Εικόνα 51: Αποσύνθεση δεδομένων και πρόβλεψη ARIMA – έξοδος R	127
Εικόνα 52: Πίνακας αποτελεσμάτων πρόβλεψης με διαστήματα εμπιστοσύνης – έξοδος R..	127
Εικόνα 53: Γραφήματα προβλέψεων ARIMA – έξοδος R	128
Εικόνα 54: Συντελεστές συναρτήσεων μοντέλων και κριτήρια αξιολόγησης – έξοδος R	128
Εικόνα 55: Πρόβλεψη μελλοντικών πωλήσεων ενός προϊόντος	129
Εικόνα 56: Πρόβλεψη επόμενης εξάντλησης αποθεμάτων	130

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

API	Application Programming Interface
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ASC	Amazon Seller Central
ASIN	Amazon Standard Identification Number
B2B	Business-to-business
B2C	Business-to-consumer
CEO	Chief Executive Officer
EDA	Exploratory Data Analysis
ERP	Enterprise Resource Planning
FBA	Fulfilled by Amazon
FBM	Fulfilled by Merchant
FG	Fairy Glam
GUI	Graphic User Interface
SKU	Stock Keeping Unit
VBA	Visual Basic for Applications

1 Εισαγωγή

1.1 Η πρόβλεψη

Ανέκαθεν ο άνθρωπος έχει περιέργεια. Αναρωτιέται για το οτιδήποτε του είναι άγνωστο και από την εμπειρία του πλάθει λύσεις για προβλήματα που αναμένει ότι θα έχουν μια συγκεκριμένη εξέλιξη. «Ο ουρανός έχει σύννεφα, άρα θα πάρω ομπρέλα». Ο συγκεκριμένος συλλογισμός δεν αποτελεί ένστικτο, αλλά είναι αποτέλεσμα εμπειρικής γνώσης από επαναλαμβανόμενες καταστάσεις ίδιας φύσεως που κάποιος αντιμετωπίζει από μικρό παιδί. Παρατηρώντας τους γύρω του και συνδέοντας την ομπρέλα με τη βροχή καθώς και την έλλειψη της ομπρέλας με την έλλειψη προστασίας από τη βροχή, το παιδί καταλήγει στο συμπέρασμα ότι χρειάζεται την ομπρέλα, ώστε να μη βραχεί. Πριν ακόμα από τη γέννηση, ο εγκέφαλος του μωρού λαμβάνει σήματα, πληροφορίες και εκπαιδεύεται μεθοδικά. Η υφή, η ακοή, η γεύση, είναι αισθήσεις που ο εγκέφαλος αντιλαμβάνεται μέσα ακόμα και από τη μήτρα. Με την είσοδο του μωρού στον κόσμο, ο όγκος των πληροφοριών – όπως πρόσωπα, σχήματα, χρώματα και λοιπά εξωτερικά ερεθίσματα, που είναι διαθέσιμος για τον άπειρο, μικροσκοπικό εγκέφαλο είναι τεράστιος και η εκπαίδευση του ανεκτίμητου αυτού οργάνου που η επιστήμη ακόμα δυσκολεύεται να κατανοήσει, ξεκινά μία ισόβια διαδικασία παρατήρησης, μάθησης και κατανόησης. Μέσα στο επόμενο χρονικό διάστημα, το παιδί θα αρχίσει να κινείται. Θα θελήσει να φτάσει από το σημείο Α στο σημείο Β. Δεν έχει σημασία αν ανάμεσα στα δύο αυτά σημεία βρεθεί η άκρη του κρεβατιού, ή κάποιο σκαλοπάτι, διότι ο εγκέφαλος του παιδιού δεν μπορεί να *προβλέψει* την πτώση ούτε τη συνέπια αυτής, δηλαδή την έκπληξη, την κρούση και τον πόνο. Το ρητό «αν δεν πέσει δε θα μάθει» έχει πολύ ακριβής σημασία. Μετά την πτώση ή ίσως μετά από πολλαπλές πτώσεις, ο εγκέφαλος θα εκπαιδευτεί να αναγνωρίζει άκρες, να αντιλαμβάνεται αποστάσεις και στη συνέχεια, να εφαρμόζει λύσεις για την αποφυγή της πτώσης, είτε κάνοντας παύση σε ασφαλές σημείο, είτε αλλάζοντας πορεία ώστε να προσεγγίσει το σημείο Β με κάποια εναλλακτική λύση.

Η πρόβλεψη αποτελεί πολύ σημαντικό κομμάτι για την ύπαρξη του ανθρώπου, συνειδητά και υποσυνείδητα. Μέσα στους αιώνες, ο άνθρωπος ξεκίνησε την καταγραφή εμπειριών και γνώσεων, ώστε να μπορέσουν να διαιωνιστούν και να φανούν χρήσιμες σε άλλους ανθρώπους. Μέσω της δημοσιοποίησης εμπειριών και ιστορικών δεδομένων, θεωρίες και έργα ζωής διατηρούνται, διαδέχονται και ολοκληρώνονται από τις επόμενες γενιές. Κάπως

έτσι, μετά από αμέτρητες γενιές και άπειρα γεγονότα, από θεωρήματα και επιστήμες της αρχαιότητας καταλήγουμε στα σύγχρονα τεχνολογικά επιτεύγματα. Ωστόσο, ένα είναι σίγουρο: Χωρίς τις απαραίτητες πληροφορίες και την ανθρώπινη περιέργεια, ο σημερινός κόσμος θα ήταν πολύ διαφορετικός. Ακόμα και από τα τέλη του 20^{ου} αιώνα, η πρόσβαση σε πληροφορίες, για τους περισσότερους ανθρώπους, είναι πιο εύκολη από ποτέ. Παλιά, η γνώση και η πρόσβαση σε γραπτές πληροφορίες αποκτιόνταν από εκπαιδευτικά εγχειρίδια και εγκυκλοπαίδειες. Με την είσοδο του ηλεκτρονικού υπολογιστή και με την εξέλιξη του από όγκο ολόκληρου δωματίου σε κάτι που χειριζόμαστε από την παλάμη μας, η ανάγκη για εκτυπωμένες πληροφορίες έχει περιοριστεί, αφού η διαθεσιμότητα των πληροφοριών είναι συντριπτική. Δεν έχει σημασία αν το ερώτημα είναι «Πώς να τηγανίσω ένα αυγό;» ή «Πώς να κάνω ανάλυση χρονολογικών σειρών σε δεδομένα με σφοδρή εποχικότητα», οι απαντήσεις υπάρχουν και είναι αχανής – χωρίς αυτό να σημαίνει ότι είναι πάντα χρήσιμες.

1.2 Η πληροφορία

Στην εποχή της ελεύθερης πληροφορίας, είναι σαφές ότι ο ρόλος αυτών για την επιχειρηματικότητα είναι ιδιαίτερα ουσιώδεις. Ο λόγος που οι επιχειρήσεις στρέφονται προς τις πληροφορίες είναι για να μπορέσουν να μελετήσουν συμπεριφορές καθώς και τις συνέπειες πράξεων του παρελθόντος των ιδίων ή τρίτων οντοτήτων, ώστε να συγκρίνουν και να διαρρυθμίσουν τις επιχειρηματικές τους αποφάσεις κατάλληλα, για την απώτερη μεγιστοποίηση του κέρδους και την μείωση της αβεβαιότητας. (Παπής, 2006) Τι είναι όμως οι πληροφορίες; *Πληροφορία* είναι οτιδήποτε μπορεί να επιφέρει κάποια χρησιμότητα ή ενημέρωση στον αποδέκτη και απορρέει από την επεξεργασία *δεδομένων*. Οι πληροφορίες δεν είναι πάντοτε εύκολα ανιχνεύσιμες. Ακόμα και αν υπάρχουν διαθέσιμες, πολλές φορές απαιτείται το κατάλληλο μέσο για να τις εξορύξει. Για την παραγωγή της πληροφορίας χρειάζεται κάποιο μέσο επεξεργασίας. Αυτό μπορεί να είναι κάποιο λογισμικό που κάνει υπολογισμούς και παράγει κάποιο αποτέλεσμα (πληροφορία) ή ακόμα και ο ανθρώπινος εγκέφαλος. Κάποιος μπορεί να αντιληφθεί διάφορες πληροφορίες από το σύνολο των δεδομένων με το γυμνό μάτι, αν και συνήθως για τη διευκόλυνση της διαδικασίας αυτής προτιμάται κάποια περίληψη στατιστικών υπολογισμών και η οπτική απεικόνιση των δεδομένων μέσω γραφημάτων. Με αυτό τον τρόπο, η παρουσία των πληροφοριών γίνεται ευκολότερα αισθητή.

Τα *δεδομένα* μπορεί να είναι κάτι εντελώς απλό ή περισσότερο σύνθετο, αλλά συνήθως εμφανίζονται σε μετρήσιμη μορφή. Τα δεδομένα συνήθως αποθηκεύονται σε μία βάση δεδομένων, η οποία μπορεί απλούστατα να χαρακτηριστεί ως ένας πολυδιάστατος πίνακας, και μπορούν να κατηγοριοποιηθούν με διάφορους τρόπους ανάλογα με την επιστημονική οπτική γωνία με την οποία θα γίνει η επεξεργασία τους. Όσο αναφορά τους τύπους δεδομένων, είναι ασφαλές να ταξινομηθούν ως εξής: (Keith Ord, 2013)

- *Διατομεακά* δεδομένα είναι μετρήσεις που καταγράφηκαν σε μία συγκεκριμένη χρονική περίοδο. Π.χ. αριθμός ελαττωματικού εμπορεύματος για την πιο πρόσφατη παραλαβή.
- Μία *χρονολογική σειρά* αποτυπώνει την επανειλημμένη καταγραφή σχετικών μετρήσεων για μία συγκεκριμένη μεταβλητή, ανά πολλαπλές χρονικές περιόδους. Π.χ. ο αριθμός επιστροφών ελαττωματικών προϊόντων από πελάτες *μηνιαία* για τα δύο τελευταία χρόνια.
- *Διαστρωματικά* δεδομένα είναι διατομεακά δεδομένα που καταγράφονται τακτικά και επανειλημμένα. Π.χ. εβδομαδιαία καταγραφή των δύο τελευταίων ετών σχετικά με τον αριθμό επιστροφών για πέντε διαφορετικά προϊόντα, με πρόσθετες λεπτομέρειες όπως το υποκατάστημα από όπου πουλήθηκαν, τις ημέρες μεταξύ αγοράς και επιστροφής και τον λόγο επιστροφής τους.

1.3 Ο ρόλος της πρόβλεψης στον επιχειρηματικό κόσμο

Συνοπτικά, με την επιμονή για τη συλλογή δεδομένων σε μεγάλες βάσεις, τη διάθεση μεγάλου χρηματικού κεφαλαίου για λογισμικά πακέτα και εξειδικευμένο προσωπικό, καθώς και χρόνου για εξονυχιστικές αναλύσεις, είναι σαφές ότι οι επιχειρήσεις στοχεύουν όλο και περισσότερο στην απόκτηση της πληροφορίας, της οποίας η αξία για πολλούς είναι ανεκτίμητη. Οι πληροφορίες χρησιμοποιούνται για τον *κατάλληλο σχεδιασμό, προγραμματισμό και τη βελτίωση των επιχειρηματικών αποφάσεων*. Οι επιχειρηματικές αποφάσεις μπορεί να είναι: (Εμίρης, 2012)

- *Βραχυπρόθεσμες* – πχ. προγραμματισμός παραγωγής, διαχείριση διανομών προϊόντων ή υπηρεσιών σε πελάτες

- *Μεσοπρόθεσμες* – πχ. διαχείριση αποθεμάτων και εμπορεύματος, κατανομή πόρων διαδικτυακής διαφήμισης σε διάφορα δίκτυα
- *Μακροπρόθεσμες* – πχ. εισαγωγή σε νέα αγορά, εξαγορά ανταγωνιστικής εταιρίας

Αναπόσπαστο κομμάτι του επιχειρηματικού σχεδιασμού είναι η *πρόβλεψη*. Όλες οι επιχειρήσεις λειτουργούν με σκοπό την επίτευξη των στόχων τους, οι οποίοι κατά κύριο λόγο είναι οικονομικής φύσεως. Ακόμα και οι στόχοι αυτοί αποτελούν προβλέψεις, οι οποίες πηγάζουν από τις επιθυμίες και τις υποχρεώσεις των ανώτατων στελεχών. Λαμβάνοντας υπόψη τις σχετικές μεθόδους που θα χρειαστεί να ακολουθήσουν για να φτάσουν τους στόχους τους, μπορούν να υπολογίσουν κατά πόσο οι στόχοι τους είναι εφικτοί και τα εμπόδια που τυχόν θα προκύψουν. Έτσι, εφόσον έχουν προβλέψει τα εμπόδια, μπορούν να λάβουν μέτρα για να τα αποφύγουν. Επιπλέον, ανάλογα με την αυτοπεποίθησή τους, μπορούν να υπολογίσουν τις πιθανότητες κατάκτησης συγκεκριμένων στόχων. Για παράδειγμα, με τη διάθεση μεγαλύτερου χρηματικού κεφαλαίου για διαφήμιση και την είσοδο ενός νέου προϊόντος στην αγορά, είναι αρκετά εφικτό για μια επιχείρηση να ελπίζει σε μία αύξηση των κερδών κατά 20% στο επόμενο εξάμηνο, αλλά λιγότερο εφικτό αν προσδοκεί τον δεκαπλασιασμό του ετήσιου προϋπολογισμού της για το επόμενο έτος.

Έχοντας τις επιπλέον γνώσεις από τις διαθέσιμες πληροφορίες, η επιχείρηση μπορεί να βελτιστοποιήσει τις επιχειρησιακές διεργασίες προς όλους τους τομείς από τους οποίους αποτελείται, από το να αλλάξει τη συχνότητα των ελέγχων των μηχανημάτων του εργοστασίου της για να μειώσει τη συχνότητα εμφάνισης ελαττωματικών προϊόντων, μέχρι το να αναθεωρήσει τους χρηματικούς πόρους που θα διαθέσει για διαφήμιση. Ακόμα και μία σχετικά μικρή επιχείρηση μπορεί να επηρεάζεται από αμέτρητες μεταβλητές και η λήψη των αποφάσεων γίνεται από διάφορα άτομα. Η διαθεσιμότητα των πληροφοριών λοιπόν, είναι πολύ σημαντική και συμβάλλουν στη γενική διοικητική οργάνωση, όχι μόνο για να συντονίσουν τους αρμόδιους που παίρνουν τις αποφάσεις, αλλά και για να τους βοηθήσουν να λάβουν υπόψη μεταβλητές που μπορεί να τους είχαν «ξεφύγει».

Υπάρχουν πολλοί διαφορετικοί τρόποι για τη διαδικασία ανάπτυξης μιας πρόβλεψης (*forecast*). Οι μέθοδοι προβλέψεων μπορούν γενικά να κατηγοριοποιηθούν στις εξής τρεις κατηγορίες: (Gilchrist, 1976)

- ***Ενστικτώδης*** είναι η κλασική και αρχαιότερη μέθοδος πρόβλεψης. Οι προβλέψεις αυτής της μεθόδου πηγάζουν από τη διαίσθηση του ατόμου και μπορεί να είναι σχετικές με τις προσωπικές του εμπειρίες, αλλά αυτό δεν είναι απαραίτητο.

Ουσιαστικά αποτελούν την έμπειρη γνώμη των στελεχών ή άλλων ειδικών, ομαδικά ή ατομικά, αν και η ομαδική διαμόρφωση αυτών των προβλέψεων συνήθως είναι πιο αξιόπιστη. Ακόμα, ως ενστικτώδης μέθοδος μπορεί να χαρακτηριστεί το αποτέλεσμα μιας έρευνας αγοράς και μιας δημοσκόπησης από δείγμα της αγοράς στόχου.

- **Αιτιοκρατική** είναι μια μέθοδος πρόβλεψης που βασίζεται στη γνώση για τις αιτίες που προκαλούν μια κατάσταση και έτσι προσπαθούν να τοποθετήσουν την επανεμφάνιση της στο μέλλον. Η παρατήρηση μακροοικονομικών εξωτερικών παραγόντων μπορεί να εξηγήσει την εμφάνιση συγκεκριμένων φαινομένων. Επιπλέον, η παρατήρηση των αποδόσεων άλλων οργανισμών με παρόμοια επιχειρηματική δραστηριότητα μπορεί να εμφανίσει πληροφορίες συσχέτισης συνεπειών με συγκεκριμένες πράξεις.
- Η μέθοδος **προεκβολής** περιλαμβάνει τη χρήση ιστορικών δεδομένων για την ανάπτυξη ενός μοντέλου πρόβλεψης, η οποία γίνεται με διάφορες στατιστικές μεθόδους και μεθόδους εξόρυξης δεδομένων. Συγκεκριμένα από τη μέθοδο προεκβολής συνεπάγονται: (Εμίρης, 2012)
 - Μέθοδοι βασισμένες στον μέσο όρο των ιστορικών δεδομένων όπως *κινούμενοι μέσοι (moving averages)* και *εκθετικής εξομάλυνσης (exponential smoothing)*
 - Μοντέλα *παλινδρόμησης (regression)* για προεκβολή *τάσης (trend extrapolation)*
 - Μέθοδοι ανάλυσης χρονολογικών σειρών με ανάπτυξη στοχαστικών μοντέλων.

1.4 Χρονολογικές σειρές

Οι χρονολογικές σειρές αποτελούνται από δεδομένα που περιέχουν τη διάσταση του χρόνου. Τα προβλήματα χρονολογικών σειρών συνήθως αναλύονται ανά αντικείμενο, προϊόν ή οποιασδήποτε φύσης είναι το χαρακτηριστικό ενδιαφέροντος. Η ανάλυση επίσης μπορεί να γίνει αθροιστικά στην περίπτωση που υπάρχουν πολλαπλά αντικείμενα, αρκεί οι υποστηρικτικές μεταβλητές να εμφανίζονται με βάση τη διάσταση του χρόνου. Συνήθως, όταν γίνεται ανάλυση χρονολογικών σειρών, θεωρείται ότι ο χρόνος είναι κρίσιμος και προμηνύει τη συμπεριφορά των δεδομένων. Τα αποτελέσματα των προβλέψεων από την

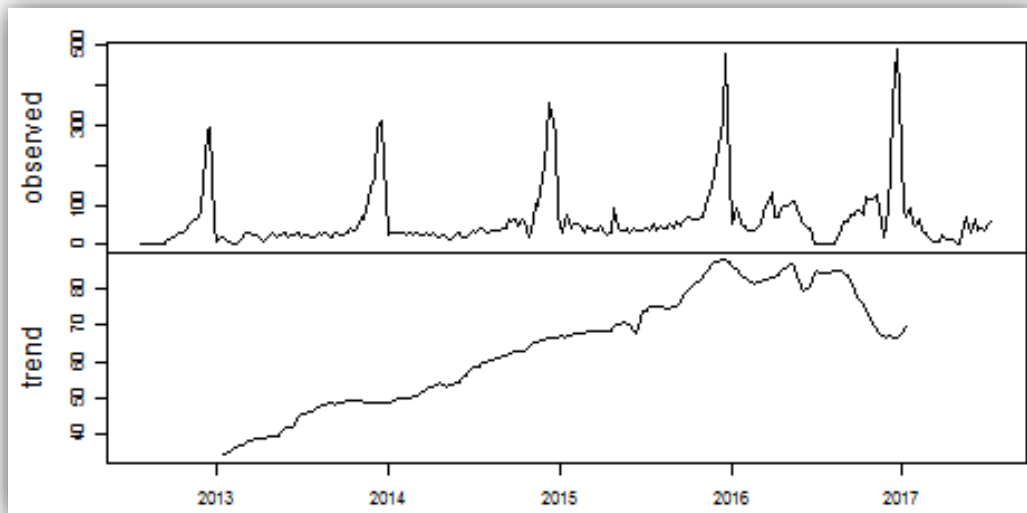
ανάλυση χρονολογικών σειρών είναι κατά κύριο λόγο αριθμητικά δεδομένα, όπως είναι τα τεμάχια που θα πωληθούν από μία κατηγορία προϊόντος μέσα στο επόμενο τρίμηνο, οι συναλλαγματικές αξίες, κλπ.

Η ανάλυση δεδομένων που υπάγονται σε χρονολογικές σειρές αποτελείται κυρίως από στατιστικές μεθόδους ή εργαλεία εξόρυξης πληροφορίας μη στατιστικής φύσεως. Η διάσταση των προκείμενων προβλέψεων περιορίζεται στις μελλοντικές τιμές μιας μεταβλητής. Η παραγωγή προβλέψεων με υψηλή ακρίβεια εξαρτάται πρωτίστως από τη διαθεσιμότητα των σχετικών δεδομένων, αλλά η φύση του προβλήματος παίζει επίσης σημαντικό ρόλο στο αν το πρόβλημα είναι αντιμετωπίσιμο ή όχι. Καταρχάς, κατά τη διαδικασία ανάπτυξης προβλέψεων πάνω σε δεδομένα χρονολογικών σειρών θα πρέπει να εξετάζεται η φύση των δεδομένων και η γραμμικότητά τους. Αν τα δεδομένα δεν είναι χρήσιμα για κάποιο γραμμικό μοντέλο και φέρουν μη ικανοποιητικά αποτελέσματα, τότε ενδείκνυται η μεταμόρφωση δεδομένων ή η χρήση μη γραμμικών μοντέλων. Η αξιολόγηση των αποτελεσμάτων ενός μοντέλου πρόβλεψης δεδομένων χρονολογικών σειρών πρέπει να γίνεται προσεκτικά και ιδιαίτερα αν έχει προηγηθεί πρόβλεψη για δεδομένα μιας μεταβλητής, τα αποτελέσματα της οποίας χρησιμοποιήθηκαν για τη διεκπεραίωση της καταλυτικής πρόβλεψης. Τέλος, η αξιολόγηση του μοντέλου πρόβλεψης θα πρέπει να επαναλαμβάνεται τακτικά ώστε να διασφαλίζεται η σωστή απόδοση του συγκριτικά με τα νέα δεδομένα. (Richard J. Roiger, 2008)

Για την επίλυση ενός προβλήματος που απαιτεί ανάπτυξη πρόβλεψης με μεθόδους προεκβολής, η παρουσία ιστορικών δεδομένων είναι θεμελιώδους σημασίας. Είτε τα δεδομένα υπάγονται σε χρονολογικές σειρές είτε όχι, για να αναπτυχθεί μια πρόβλεψη, πρώτα γίνεται μελέτη των ιστορικών δεδομένων. Η πρόβλεψη με μεθόδους προεκβολής προϋποθέτει τη χρήση των ιστορικών δεδομένων για την παραγωγή των αποτελεσμάτων. Οι μέθοδοι προεκβολής βασίζονται στην ύπαρξη κάποιας *συσχέτισης μεταξύ των μεταβλητών*, που να μπορέσει να προσδιορίσει μελλοντικές τιμές. Στην περίπτωση που η μέθοδος πρόβλεψης θα γίνει με βάση τη διάσταση του χρόνου, η συσχέτιση πολλές φορές δεν αναζητείται στο σύνολο των μεταβλητών, αλλά στις ιστορικές εκδόσεις της ίδιας της μεταβλητής. Αναμένεται λοιπόν ότι θα βρεθεί *αυτοσυσχέτιση* της ίδιας μεταβλητής με τις προηγούμενες εκδόσεις της. (Rob J Hyndman, 2017)

Καταρχάς θα πρέπει να μελετηθεί και να γίνει κατανοητή η εξέλιξη των δεδομένων κατά το παρελθόν. Είναι αναμενόμενο ότι από την παρατήρηση του παρελθόντος θα εμφανιστούν

μοτίβα από τα οποία θα προσδιοριστεί η *τάση* των δεδομένων (η «μακροπρόθεσμη» κίνηση μιας χρονολογικής σειράς χωρίς τις σχετικές ημερολογιακές και ακανόνιστες επιδράσεις, π.χ. ανοδική, καθοδική, στάσιμη), ο *χρονικός κύκλος* ή *περίοδος* που καθορίζει τη συμπεριφορά των δεδομένων και τυχόν παρατηρούμενων φαινομένων *εποχικότητας*. Η εποχικότητα εμφανίζεται πάντα κατά γνωστές, προκαθορισμένες περιόδους (π.χ. ανά μήνα, ανά χρόνο) και αποτελείται από εποχικά μοτίβα. Ένα εποχικό μοτίβο παρουσιάζεται όταν μια χρονολογική σειρά είναι επηρεασμένη από εποχικούς παράγοντες. Τα εποχικά μοτίβα χρονικού κύκλου περιλαμβάνουν αυξομειώσεις των δεδομένων που δεν αντιστοιχούν σε συγκεκριμένη περίοδο. Ένας εποχικός κύκλος συνήθως χαρακτηρίζεται από μοτίβα που παρουσιάζονται για περιόδους τουλάχιστον δύο ετών. (Hyndman, 2011) Ο προσδιορισμός των παραπάνω χαρακτηριστικών μιας χρονολογικής σειράς γίνεται με τη χρήση μεθόδων αποσύνθεσης (στατιστική μέθοδος που διασπάει τα μέρη που συνθέτουν μια χρονολογική σειρά, τα οποία αποτελούν θεμελιώδης κατηγορίες μοτίβων).

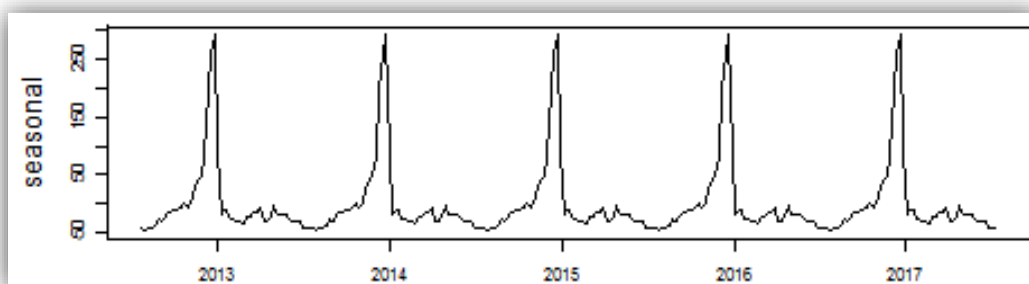


Εικόνα 1: Αποσύνθεση χρονολογικών σειρών – εξόρυξη της τάσης

Όπως φαίνεται στην *εικόνα 1*, στο πάνω πλαίσιο παρουσιάζεται μια πραγματική χρονολογική σειρά, όπως οι τιμές τις παρατηρήθηκαν κατά την περίοδο πέντε ετών. Με μία πρώτη ματιά ίσως να μη γίνεται αισθητή η *ανοδική τάση* των τιμών της χρονολογικής σειράς, αλλά η άνοδος φαίνεται καλύτερα στο κάτω πλαίσιο της εικόνας, όπου η *τάση (trend)* έχει φιλτραριστεί από τα αρχικά δεδομένα με μέθοδο αποσύνθεσης. Το γράφημα τάσης που παρουσιάζεται ανήκει όντως στην εικονιζόμενη χρονολογική σειρά. Αν η αρχική μορφή της

χρονολογικής σειράς είχε εικόνα παρόμοια με αυτή της τάσης, τότε θα ήταν ξεκάθαρο ότι υπάρχει ραγδαία ανοδική πορεία στα δεδομένα. Ωστόσο, δεν είναι εύκολος ο προσδιορισμός φαινομένων εποχικότητας ή των χρονικών κύκλων από μία σειρά που παρουσιάζεται όπως στο κάτω πλαίσιο της εικόνας 1. Αντίθετα, το γράφημα των αρχικών παρατηρήσεων δείχνει ξεκάθαρα ότι οι χρονικοί κύκλοι αντιπροσωπεύουν ένα έτος και ότι προς το τέλος του κάθε κύκλου παρατηρείται ακραία εποχικότητα. Το γράφημα της τάσης μπορεί να αναπτυχθεί αν οι τιμές του παραχθούν από τις αρχικές παρατηρήσεις με τη χρήση κυλιόμενου αθροίσματος σταθερής περιόδου. Το κυλιόμενο άθροισμα ορίζεται ως η πρόσθεση μιας σειράς αριθμών που ενημερώνεται κάθε φορά που προσθέεται μία νέα τιμή. Στην περίπτωση της σταθερής περιόδου, κάθε φορά που προσθέεται μια νέα τιμή αφαιρείται η πιο παλιά. Συνεπώς, για την τάση ισχύουν τα παρακάτω χαρακτηριστικά: (Matteson, 2017)

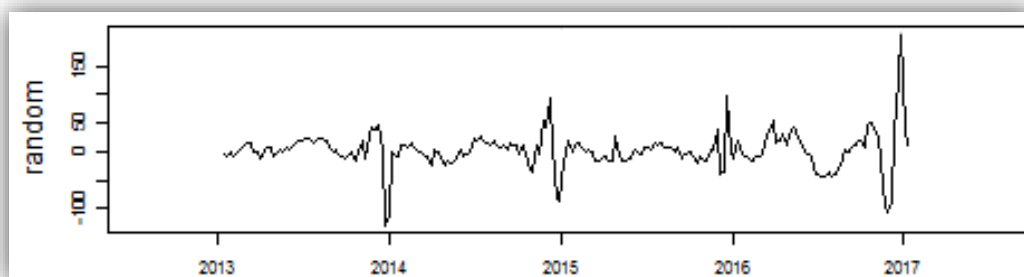
- i. Κάποιες χρονολογικές σειρές δεν παρουσιάζουν ξεκάθαρες χρονικές τάσεις.
- ii. Οι τάσεις μπορεί να είναι *θετικές* (ανοδικές) ή *αρνητικές* (καθοδικές).
- iii. Οι τάσεις μπορεί να εμφανίζουν *γραμμικότητα* ή μη γραμμικότητα (η γραμμική τάση ορίζεται από σημεία δεδομένων που σχηματίζουν μια ευθεία).
- iv. Συνήθως, *μη γραμμικές* τάσεις παρουσιάζουν *ραγδαία* πορεία.
- v. Οι τάσεις μπορεί να είναι *περιοδικές* και να συμπεριλαμβάνουν έναν ή περισσότερους *χρονικούς κύκλους* με κάποια χαρακτηριστικά *εποχικότητας*.



Εικόνα 2: Αποσύνθεση χρονολογικών σειρών – εξόρυξη της εποχικότητας

Ως συνέχεια της αποσύνθεσης της αρχικής χρονολογικής σειράς, στην εικόνα 2 παρουσιάζεται το εποχικό περιεχόμενο των δεδομένων. Εξάγοντας την τάση και την εποχικότητα, τα αρχικά δεδομένα έχουν ξεμείνει με το απροσδιόριστο, τυχαίο χαρακτηριστικό, όπως φαίνεται στην εικόνα 3. Το τυχαίο χαρακτηριστικό (*random walk*) φέρει απροσδιόριστο μέσο ή διακύμανση και παρουσιάζει ισχυρή χρονική εξάρτηση. Οι αλλαγές και οι αυξομειώσεις του τυχαίου σημειώνονται... τυχαία, χωρίς να μπορεί να

προσδιοριστεί η αιτία για τις μεταβολές των σημείων της χρονολογικής σειράς. Οι αυξομειώσεις αυτές επίσης ονομάζονται *λευκός θόρυβος (white noise)*. Μια χρονολογική σειρά λευκού θορύβου είναι το πιο βασικό παράδειγμα μιας *αμετάβλητης* διαδικασίας (*stationary*). Τα χαρακτηριστικά που διέπουν μια διαδικασία λευκού θορύβου είναι ένας σταθερός, αμετάβλητος μέσος, μία σταθερή και αμετάβλητη διακύμανση καθώς και ανύπαρκτη χρονική εξάρτηση, δηλαδή μηδενική αυτοσυσχέτιση. (Matteson, 2017) Εφόσον ο λευκός θόρυβος είναι το κύριο χαρακτηριστικό μιας αμετάβλητης διαδικασίας και το τυχαίο χαρακτηριστικό είναι αυτό που απομένει σε μία χρονολογική σειρά μετά την εξόρυξη της τάσης και της εποχικότητας, συμπερασματικά, οποιαδήποτε διαδικασία περιέχει τάση και εποχικότητα θεωρείται ως *μεταβλητή (non stationary)*.



Εικόνα 3: Αποσύνθεση χρονολογικών σειρών – εξόρυξη του απροσδιόριστου

1.5 Στόχοι και μελέτη περίπτωσης

Στο παρόν έγγραφο καταγράφεται και παρουσιάζεται η διαδικασία ανάπτυξης μοντέλων πρόβλεψης με μεθόδους *προεκβολής*, κατά κύριο λόγο. Στα επόμενα κεφάλαια και στο πρώτο μέρος της παρούσας εργασίας, γίνεται βιβλιογραφική αναφορά σχετικά με τα διαθέσιμα εργαλεία και τις στατιστικές μεθόδους που συνήθως προτιμώνται για προβλήματα της φύσεως των επιχειρηματικών διεργασιών που εξαρτώνται από τον κατάλληλο χρονικό προγραμματισμό. Στα πλαίσια του έργου αυτού πραγματοποιήθηκε μελέτη περίπτωσης για μία επιχείρηση διαδικτυακού εμπορίου του Ηνωμένου Βασιλείου, η οποία μελέτη καταγράφεται στο δεύτερο μέρος αυτής της εργασίας. Τα δεδομένα της επιχείρησης που μελετήθηκε, χρησιμοποιήθηκαν για την ανάλυση και ανάπτυξη των μεθόδων πρόβλεψης που αναφέρονται στο πρώτο μέρος της εργασίας. Τα μοντέλα μεθόδων προεκβολής που αναπτύχθηκαν θα εφαρμοστούν πάνω στα ιστορικά δεδομένα της επιχείρησης, και στη συνέχεια θα αξιολογηθεί η χρησιμότητά τους σύμφωνα με τις ανάγκες της. Το μοντέλο που

τελικά θα έχει την καλύτερη απόδοση, θα αποτελέσει μέρος ενός συστήματος υποστήριξης, στο οποίο τα δεδομένα είναι οργανωμένα, έχουν υποστεί στατιστική ανάλυση και παρουσιάζονται με διάφορα γραφήματα. Στόχος της μελέτης περίπτωσης είναι η περαιτέρω κατανόηση των μεθόδων προβλέψεων και η παρατήρηση της απόδοσης τους εφόσον εφαρμοστούν κατάλληλα και κληθούν να επιλύσουν προβλήματα του πραγματικού κόσμου. Συγκεκριμένα, στο δεύτερο μέρος του εγγράφου αυτού αποτυπώνονται τα εξής:

- Λεπτομέρειες σχετικά με την επιχείρηση και καταγραφή των προβλημάτων που απαιτούν επίλυση.
- Η διαδικασία συλλογής και οργάνωσης των δεδομένων.
- Ο τρόπος αποκόμισης πληροφοριών από τα σύνολα δεδομένων εφαρμόζοντας απλή στατιστική ανάλυση.
- Η παρουσίαση των δεδομένων οπτικά ώστε ένας παρατηρητής να μπορέσει να ερμηνεύσει σχετική πληροφορία από τα γραφήματα.
- Η ανάπτυξη στατιστικών μοντέλων πρόβλεψης πάνω στο Microsoft Excel και η ανάλυση των αποτελεσμάτων τους.
- Η ανάπτυξη και εφαρμογή στοχαστικών μοντέλων πρόβλεψης με χρήση της γλώσσας προγραμματισμού R.

2 Είδη πρόβλεψης και βασικά εργαλεία

2.1 Βασικά βήματα για τη διαδικασία της πρόβλεψης

Για τη δημιουργία ενός μοντέλου πρόβλεψης για το οποίο υπάρχουν μετρήσιμα δεδομένα, πρέπει να ακολουθούνται τα παρακάτω πέντε βήματα: (Spyros Makridakis S. C., 1998)

➤ Προσδιορισμός του προβλήματος

Το αποτέλεσμα μιας πρόβλεψης είναι γενικά αναμενόμενο. Τα φαινόμενα που συμβαίνουν στην τροπόσφαιρα και που αποτελούν αντικείμενο έρευνας και σπουδής της μετεωρολογίας έχουν συγκεκριμένο αριθμό. Τα προγνωστικά για τον καιρό θα ανήκουν πάντα στο σύνολο των καιρικών φαινομένων που υπάρχουν, αν τελικά η πρόβλεψη είναι σωστή. Ακόμα και αν η πρόβλεψη είναι λάθος, το καιρικό φαινόμενο που τελικά θα συμβεί, θα ανήκει επίσης στο σύνολο των συγκεκριμένων καιρικών φαινομένων και πιθανόν να μην είναι πολύ αποκλίνον από την αρχική πρόγνωση. Είναι γεγονός ότι η χιονοθύελλα ανήκει στο σύνολο των καιρικών φαινομένων και, ενώ υπάρχει κάποια πιθανότητα εμφάνισης της, δεν είναι αναμενόμενο να συμβεί στην Ελλάδα τον μήνα Ιούλιο, την ίδια ημέρα για την οποία η Εθνική Μετεωρολογική Υπηρεσία πρόβλεψε ότι η ανώτερη θερμοκρασία της ημέρας θα ήταν 34°C... Ωστόσο, η μετεωρολογία και ο προσδιορισμός του συνόλου των καιρικών φαινομένων αποτελούν ολόκληρη επιστήμη. Συνεπώς, ο προσδιορισμός του προβλήματος που απαιτεί μια πρόβλεψη μπορεί να αποβεί το δυσκολότερο κομμάτι ενός αναλυτή που ασχολείται με την πρόβλεψη. Χρειάζεται χρόνος – ίσως περισσότερος χρόνος απ' ό τι θα ήταν αναμενόμενο, και είναι απαραίτητο ότι ο αναλυτής θα αποκτήσει βαθιά γνώση σχετικά με τις διεργασίες που αφορούν το πρόβλημα και τους ανθρώπους που το διαχειρίζονται.

➤ Συλλογή δεδομένων

Για τη διαδικασία οποιουδήποτε είδους πρόβλεψης, χρειάζονται κάποιες σχετικές πληροφορίες. Οι πληροφορίες βρίσκονται στη μορφή συνήθως μετρήσιμων δεδομένων του παρελθόντος ή της εμπειρικής γνώσης και κρίσης των κατάλληλων ειδικών. Για να αναπτυχθεί ένα χρήσιμο μοντέλο πρόβλεψης θα πρέπει να υπάρχουν διαθέσιμες πληροφορίες και των δύο ειδών. Αφού ληφθεί πρόσβαση στις σωστές πηγές και γίνει αποθήκευση των σχετικών ακατέργαστων δεδομένων, θα χρειαστεί να υποστούν επεξεργασία και την κατάλληλη οργάνωση ώστε να ξεκινήσει η διαδικασία της ανάλυσης και εξόρυξης

πληροφοριών. Αν όντως πραγματοποιηθεί επεξεργασία των αρχικών δεδομένων, θα πρέπει να γίνει πολύ προσεκτικά και με μεθοδικό τρόπο, ώστε να μη διαστρεβλωθεί η στατιστική τους σημασία.

➤ *Προκαταρκτική - εξερευνητική ανάλυση*

Η ανάλυση αυτή εξυπηρετεί δύο σκοπούς: τον καθαρισμό ή την επεξεργασία των δεδομένων και την κατανόηση των δεδομένων. (Keith Ord, 2013) Παραπάνω αναφέρθηκε ότι τα ακατέργαστα δεδομένα μπορεί να χρειαστούν επεξεργασία ώστε να φανούν χρήσιμα σε μια διαδικασία πρόβλεψης. Αυτό είναι πολύ σημαντικό, αλλά δεν είναι δυνατό να συμβεί αν πρώτα δεν υπάρχει κατανόηση περί του σκοπού της πρόβλεψης αλλά και των ίδιων των δεδομένων που υπάρχουν διαθέσιμα. Με την προκαταρκτική ανάλυση ο ερευνητής θα χρησιμοποιήσει στατιστικές μεθόδους περίληψης και γραφήματα ώστε να αποκομίσει όσο γίνεται περισσότερες πληροφορίες. Παράλληλα με τη διεξαγωγή συμπερασμάτων για τα δεδομένα μπορεί να υποβάλει *καθαρισμό*, όπως στην περίπτωση αφαίρεσης μη χρήσιμων μεταβλητών, συμπλήρωσης κενών πεδίων, εφαρμογής λογαριθμικών μεταμορφώσεων ή μετατροπής ακραίων τιμών που επηρεάζουν τα αποτελέσματα σε υπερβολικό βαθμό ενώ δεν αντικατοπτρίζουν την πραγματικότητα.

➤ *Επιλογή και προσαρμογή μοντέλων*

Εφόσον τα δεδομένα είναι έτοιμα, μπορούν να χρησιμοποιηθούν ως είσοδος για στατιστικές μεθόδους ή μοντέλα πρόβλεψης. Πολλές φορές, για την προσαρμογή διαφορετικών μεθόδων θα χρειαστεί περαιτέρω επεξεργασία και εξερευνητική ανάλυση των δεδομένων για την επιτυχή εφαρμογή των μοντέλων. Σε αυτό το σημείο αξιωματικώς σημειωθεί η διαφορά μεταξύ μιας μεθόδου και ενός μοντέλου πρόβλεψης: Τα *μοντέλα* πρόβλεψης αναπτύσσονται από την αλλαγή παραμέτρων μιας μεθόδου πρόβλεψης. *Μέθοδος* θεωρείται η προσέγγιση με την οποία θα γίνει η προσπάθεια ανάπτυξης της πρόβλεψης, π.χ. γραμμική παλινδρόμηση, εκθετική εξομάλυνση. Από μία μέθοδο πρόβλεψης μπορεί να παραχθούν πολλά μοντέλα πρόβλεψης με διαφορετικές παραμέτρους και με διακυμάνσεις στην ακρίβεια τους. Συνεπώς, η σύγκριση προβλέψεων γίνεται όσο αναφορά τα προκείμενα μοντέλα και όχι απαραίτητα τις μεθόδους στις οποίες τα μοντέλα ανήκουν. Η σύγκριση των μοντέλων πρόβλεψης δεν προϋποθέτει ότι τα μοντέλα θα πρέπει να ανήκουν στην ίδια μέθοδο πρόβλεψης, ωστόσο θα χρειαστεί να γίνει έλεγχος των παραμέτρων τους και των κλιμάκων των δεδομένων που χρησιμοποιήθηκαν, ώστε τα αποτελέσματα των συγκρίσεων να είναι ερμηνεύσιμα.

➤ *Εφαρμογή και αξιολόγηση μοντέλων*

Η επιτυχία στην εφαρμογή μοντέλων πάνω στα δεδομένα και η διεξαγωγή αποτελεσμάτων είναι πολύ σημαντική, αλλά εξίσου κρίσιμη είναι και η αξιολόγηση των μοντέλων ως προς τα αποτελέσματα τους σε σχέση με τα πραγματικά δεδομένα. Το μοντέλο που αποδίδει καλύτερα θα θεωρηθεί ως το πιο αξιόπιστο και τα υπόλοιπα θα απορριφθούν. Για το χρονικό διάστημα που η επιλεγμένη μέθοδος χρησιμοποιείται για την παραγωγή προβλέψεων, θα πρέπει να ελέγχεται συχνά ως προς την ακρίβεια της και να αξιολογείται κατάλληλα.

2.2 Είδη τεχνικών πρόβλεψης

Παραπάνω αναφέρθηκαν τα πέντε βασικά βήματα που πρέπει να ακολουθηθούν κατά τη διαδικασία μιας πρόβλεψης. Η βασική προϋπόθεση για την παραπάνω προσέγγιση είναι ότι υπάρχουν διαθέσιμα δεδομένα με στατιστική σημασία ώστε η ανάλυση τους να είναι εφικτή. Υπάρχουν διάφορες τεχνικές πρόβλεψης και η κάθε μία χρησιμοποιείται για διαφορετικό σκοπό. Σχετικά με τη διαθεσιμότητα των δεδομένων, οι τεχνικές πρόβλεψης χωρίζονται στις εξής κατηγορίες: (Spyros Makridakis S. C., 1998)

➤ *Ποσοτικές τεχνικές*

Οι μέθοδοι πρόβλεψης που υπάγονται σε αυτήν την κατηγορία εξαρτώνται από τη διαθεσιμότητα των δεδομένων, που για να χρησιμοποιηθούν με ποσοτικές τεχνικές πρόβλεψης θα πρέπει να υπάρχουν σε αφθονία. Π.χ. για την εφαρμογή γραμμικής παλινδρόμησης είναι απαραίτητη η ύπαρξη τουλάχιστον δύο διαφορετικών μεταβλητών, μίας εξαρτημένης και τουλάχιστον μίας ανεξάρτητης, ή αλλιώς ελεγχόμενης ή επεξηγηματικής μεταβλητής (*predictor*). Για να γίνει ανάλυση χρονολογικών σειρών, κυρίαρχη προϋπόθεση είναι η ύπαρξη δεδομένων *πολλαπλών* χρονολογικών κύκλων.

➤ *Ποιοτικές τεχνικές*

Σε αυτήν τη περίπτωση τα δεδομένα είναι ελάχιστα, δεν υπάρχουν καθόλου ή δεν υπάρχει η πρόσβαση σε αυτά. Συνεπώς, οι ποιοτικές τεχνικές δε βασίζονται στα δεδομένα, εφόσον δεν είναι διαθέσιμα, αλλά εξαρτώνται από την επάρκεια της έμπειρης γνώσης από εξειδικευμένους ερευνητές ή επαγγελματίες. Σε αυτήν την κατηγορία υπάγονται και οι ενστικτώδης μέθοδοι πρόβλεψης.

➤ Απρόβλεπτη

Στην περίπτωση που τα δεδομένα είναι ελάχιστα, μη στατιστικής σημασίας ή ανύπαρκτα, καθώς και είναι αδύνατον να γίνει εφαρμογή ποιοτικών τεχνικών, τότε το προκειμένο πρόβλημα θεωρείται απρόβλεπτο.

2.3 Θεμελιώδης αρχές πρόβλεψης

Κατά τη διαδικασία ανάπτυξης ενός μοντέλου πρόβλεψης, εφόσον το πρόβλημα είναι κατανοητό, τα εφαρμοσμένα μοντέλα θα είναι αυτά που επιφέρουν αποτελέσματα και τη λύση του προβλήματος. Συνήθως η ανάπτυξη, η κατανόηση και η αξιολόγηση των μοντέλων είναι περίπλοκη και χρονοβόρα, οπότε είναι λογικό να πέφτει ιδιαίτερη βαρύτητα σε αυτό το κομμάτι της ανάπτυξης μιας πρόβλεψης. Ωστόσο, όσο σωστή και ποιοτική είναι η δουλειά της ανάπτυξης και επιλογής του μοντέλου πρόβλεψης, αν τα θεμέλια δεν είναι σωστά δομημένα από την αρχή, τότε τα αποτελέσματα θα υστερούν. Τα θεμέλια για ένα μοντέλο πρόβλεψης είναι τα δεδομένα του και είναι σημαντικό να δοθεί ιδιαίτερη σημασία σε κάποιες βασικές αρχές κατά τη διαδικασία της προκαταρκτικής – εξερευνητικής ανάλυσης των δεδομένων. (Keith Ord, 2013)

➤ Επιβεβαίωση ότι έχουν επιλεγεί τα κατάλληλα δεδομένα.

Εφόσον το πρόβλημα και το αναμενόμενο αποτέλεσμα μιας πρόβλεψης είναι απολύτως κατανοητό, μπορεί τότε να αναγνωριστεί το σύνολο δεδομένων που ταιριάζει καλύτερα. Ωστόσο, το καλύτερο αυτό σύνολο δεδομένων μπορεί να μην είναι διαθέσιμο για διάφορους λόγους. Π.χ. Η δημοσίευση μακροοικονομικών δεδομένων πραγματοποιείται με κάποια καθυστέρηση – πολλές φορές αρκετών μηνών, και ακόμα και τότε μπορεί να παρουσιάζεται κατά προσέγγιση. Ο ερευνητής θα πρέπει να ελέγξει τη διαθεσιμότητα των δεδομένων λαμβάνοντας υπόψη την τελική χρήση των προβλέψεων.

➤ Καθαρισμός των δεδομένων

Τα δεδομένα μπορεί να είναι ελλιπή, λανθασμένα, ή επηρεασμένα από ορισμούς που αλλάζουν κατά τη διάρκεια καταγραφής των παρατηρήσεων, αποτελώντας έτσι σύγχυση. Σε αυτήν την περίπτωση είναι συνετή η πρακτική αφαίρεσης μη χρήσιμων μεταβλητών, συμπλήρωσης κενών πεδίων κ.λπ. Οι τροποποιήσεις πρέπει να γίνονται μόνο όπου είναι

απαραίτητο και να διατηρείται σχετικό έγγραφο όπου αναφέρονται οι τροποποιήσεις για μελλοντική χρήση του ίδιου του ερευνητή ή τρίτων ατόμων.

➤ Χρήση μεταμορφώσεων εφόσον απαιτείται από τη φύση των δεδομένων

Οι μεταμορφώσεις γίνονται ως απάντηση σε ένα πρόβλημα όπου τα διαθέσιμα δεδομένα δεν είναι χρήσιμα για κάποια τεχνική πρόβλεψη. Π.χ. η μη-γραμμικότητα των δεδομένων καθιστά προβληματική τη χρήση της γραμμικής παλινδρόμησης, αλλά αυτό είναι αντιμετωπίσιμο με την εφαρμογή της μεθόδου σε δεδομένα που έχουν υποστεί λογαριθμική μεταμόρφωση.

➤ Απεικόνιση δεδομένων

Η χρήση διαγραμμάτων και γραφημάτων είναι ιδιαίτερα σημαντική για την απόκτηση περαιτέρω πληροφορίας. Με αυτόν τον τρόπο μπορεί να προσδιοριστούν δεδομένα που θα μπορούσαν να επωφεληθούν από εναλλακτική προσέγγιση. Τα γραφήματα δεν πρέπει να λείπουν από καμία διαδικασία ανάλυσης δεδομένων με σκοπό την πρόβλεψη.

➤ Τροποποιήσεις ακραίων δεδομένων για απρογραμμάτιστα γεγονότα

Τα δεδομένα μπορεί να είναι επηρεασμένα από ενδογενής ή εξωγενής παράγοντες που δεν αντικατοπτρίζουν τη σύνηθε συμπεριφορά του θέματος που παρατηρείται. Απρόβλεπτοι παράγοντες μπορεί να είναι ακραία καιρικά φαινόμενα, απεργίες εργοστασίων, εξάντληση αποθεμάτων κ.λπ. Σε αυτήν την περίπτωση είναι συνετή η τροποποίηση των επηρεασμένων τιμών ώστε να παρουσιάζουν μια αναμενόμενη εικόνα και να αυξήσουν την ακρίβεια των αποτελεσμάτων της πρόβλεψης.

➤ Περαιτέρω τροποποιήσεις για αναμενόμενα ακραία δεδομένα

Κάποιες φορές οι ακραίες τιμές δεν είναι απρογραμμάτιστες και μπορεί να επαναλαμβάνονται τακτικά. Π.χ. περίοδοι υψηλής εποχικότητας. Αυτές οι επιρροές θα πρέπει να λαμβάνονται υπόψη ώστε τα αποτελέσματα της πρόβλεψης να συμπεριλαμβάνουν τις ακραίες τιμές, εφόσον είναι αναμενόμενες. Ένας τρόπος για να εισαχθούν εποχικά μοτίβα πάνω στα αποτελέσματα μιας πρόβλεψης είναι η χρήση συντελεστών εποχικότητας. Π.χ. χρήση απλών τεχνικών κινούμενων μέσων ή εκθετικής εξομάλυνσης για την πρόβλεψη των μελλοντικών τιμών και πολλαπλασιασμός των παραγόμενων αποτελεσμάτων με τους συντελεστές εποχικότητας ώστε να ρυθμιστεί η βαρύτητα της κάθε τιμής ανάλογα με το πως παρουσιάστηκε στο παρελθόν.

2.4 Αξιολόγηση ακρίβειας μοντέλων και μεθόδων πρόβλεψης

Μία κρίσιμη απόφαση που θα πρέπει να λαμβάνεται υπόψη κατά τη διαδικασία ανάπτυξης ενός μοντέλου πρόβλεψης είναι ο τρόπος με τον οποίο θα γίνεται η μέτρηση της απόδοσης του. Η μέτρηση της απόδοσης ενός μοντέλου είναι πολύ σημαντική ώστε να μπορεί να γίνει επιτυχημένα η σύγκριση μεταξύ διαφορετικών μεθόδων πρόβλεψης και να πραγματοποιηθεί η τελική επιλογή του μοντέλου που αποδίδει καλύτερα. Ύστερα από την τελική επιλογή, είναι εξίσου σημαντική η χρήση των καταμετρήσεων απόδοσης του μοντέλου, ώστε να διαβεβαιωθεί ότι αυτό συνεχίζει να αποδίδει με παρόμοια επίπεδα ακρίβειας. Στην περίπτωση που μετά από κάποιο χρονικό διάστημα το επιλεγμένο μοντέλο δεν αποδίδει σύμφωνα με τις προσδοκίες για αυτό, τότε θα πρέπει να γίνει κάποια παρέμβαση στον τρόπο λειτουργίας του μοντέλου ή ακόμα και να μεσολαβήσει μια επανάληψη της σύγκρισης του μοντέλου με άλλα μοντέλα. (Keith Ord, 2013)

Μία σύνηθες πρακτική για τη διαδικασία αξιολόγησης της ακρίβειας είναι η παρατήρηση των σφαλμάτων, δηλαδή η διαφορά μεταξύ των εκτιμώμενων αποτελεσμάτων με τα πραγματικά δεδομένα. Υπάρχουν διάφοροι μέθοδοι μέτρησης του σφάλματος αυτού αλλά εννοείται ότι η σύγκριση των αποτελεσμάτων θα πρέπει να γίνεται για κάθε μέθοδο μέτρησης ξεχωριστά, καθώς και να επαναλαμβάνονται οι μέθοδοι αυτοί ανά τακτά χρονικά διαστήματα και λαμβάνοντας υπόψη τις κατάλληλες παραμέτρους. (Rob J Hyndman, 2017)

2.4.1 Σφάλματα εξαρτώμενα από την κλίμακα

Το σφάλμα της πρόβλεψης είναι απλώς $e_i = y_i - \hat{y}_i$, δηλαδή η απόκλιση της πραγματικής τιμής από την προβλεπόμενη. Τα σφάλματα αυτά ανήκουν στην ίδια κλίμακα όπως τα δεδομένα, δηλαδή για ποσοστιαία δεδομένα τότε το σφάλμα θα είναι ποσοστιαίας κλίμακας, ενώ για δεδομένα συναλλαγματικής αξίας, τα σφάλματα θα είναι αντίστοιχα (και όχι ποσοστιαίας φύσεως).

Οι δύο συχνότερες μετρήσεις σφαλμάτων εξαρτώμενα από την κλίμακα, βασίζονται στις απόλυτες τιμές και στις ριζικές τετραγωνισμένες τιμές: μέσο απόλυτο σφάλμα (*Mean Absolute Error – MAE*) και ριζικό μέσο τετραγωνισμένο σφάλμα (*Root Mean Squared Error – RMSE*). Όπου e_i εννοείται η τιμή του σφάλματος.

$$MAE = \text{μέσος}(|e_i|)$$

$$RMSE = \sqrt{\text{μέσος}(e_i^2)}$$

Επιπλέον, η απλή διαίρεση του αθροίσματος των σφαλμάτων προς τον μέσο των παρατηρήσεων είναι ένας χρήσιμος τρόπος προσδιορισμού προκατάληψης στην πρόβλεψη. (Keith Ord, 2013). Συγκεκριμένα, αν το πηλίκο του μέσου σφάλματος (*Mean Error – ME*) είναι μεγάλο και θετικό σημαίνει ότι η πραγματική τιμή είναι μεγαλύτερη από την πρόβλεψη. Αντίθετα, αν το πηλίκο είναι αρνητικό τότε η τιμή της πρόβλεψης είναι μεγαλύτερη από την πραγματική τιμή. Όπου $\bar{\mu}$ ή «μέσος» εννοείται ο αριθμητικός μέσος.

$$ME = \sum_{i=1}^{\bar{\mu}} \frac{e_{t+i}}{\bar{\mu}}$$

2.4.2 Ποσοστιαία σφάλματα

Το ποσοστιαίο σφάλμα της πρόβλεψης είναι του τύπου $p_i = \frac{e_i}{y_i} * 100$. Τα ποσοστιαία σφάλματα έχουν το πλεονέκτημα ότι δεν είναι εξαρτώμενα από την κλίμακα του δείγματος δεδομένων που αντιπροσωπεύουν. Εφόσον οι τιμές των σφαλμάτων αυτών εμφανίζονται πάντα με συγκεκριμένη μορφή, είναι ιδιαίτερα χρήσιμα για την περίπτωση που χρειάζεται να γίνει σύγκριση μεταξύ της απόδοσης των προβλέψεων (Y) διαφορετικών μεταβλητών του συνόλου δεδομένων. Επιπρόσθετα, τα ποσοστιαία σφάλματα μπορεί να φανούν χρήσιμα αν οι μεταβλητές που παρατηρούνται είναι αυστηρά θετικές, όπως ο αριθμός των εργαζομένων μιας επιχείρησης ή τα έσοδα από πωλήσεις. Οι συχνότερες μετρήσεις ποσοστιαίων σφαλμάτων είναι το κεντρικό ποσοστιαίο σφάλμα (*Mean Percentage Error – MPE*) και το κεντρικό απόλυτο ποσοστιαίο σφάλμα (*Mean Absolute Percentage Error – MAPE*).

$$MPE = \frac{100}{\bar{x}} \sum_{i=1}^{\bar{\mu}} \frac{e_{t+i}}{Y_{t+i}}$$

$$MAPE = \text{μέσος}(|p_i|)$$

2.4.3 Κλιμακωτά σφάλματα

Οι μετρήσεις που βασίζονται σε ποσοστιαία σφάλματα φέρουν το μειονέκτημα ότι, στην περίπτωση που υπάρχουν μηδενικές τιμές στο δείγμα δεδομένων που εξετάζεται ή τιμές που είναι πολύ κοντά στο μηδέν, τότε τα σφάλματα μπορεί να πλησιάζουν το άπειρο ή κάποια απροσδιόριστη τιμή. Π.χ. το ME είναι χρήσιμο για δεδομένα θερμοκρασίας, αλλά το ποσοστιαίο MPE δεν είναι, διότι οι τιμές μπορούν να πέσουν κάτω από το μηδέν. Τα κλιμακωτά σφάλματα μπορούν να χρησιμοποιηθούν ως εναλλακτική λύση, ώστε να γίνεται σύγκριση των αποδόσεων των μοντέλων προβλέψεων διαφορετικής κλίμακας. Η διαδικασία απαιτεί την «εκπαίδευση» των μέσων απόλυτων σφαλμάτων μέσω μιας απλής μεθόδου πρόβλεψης. Για μη-εποχικές χρονολογικές σειρές, τα κλιμακωτά σφάλματα συμπεριλαμβάνουν μεθόδους *Naïve* (Rob J Hyndman, 2017), που μπορούν να οριστούν ως:

$$q_j = \frac{e_j}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|}$$

Επειδή ο αριθμητής και ο παρονομαστής του παραπάνω κλάσματος εμπεριέχουν τιμές κλίμακας των αρχικών δεδομένων, το q_j είναι ανεξάρτητο από την κλίμακα αυτή. Το κλιμακωτό σφάλμα είναι μικρότερο της μονάδας αν πηγάζει από κάποιο καλύτερο μοντέλο πρόβλεψης από το κοινό αποτέλεσμα της πρόβλεψης *Naïve*. Παρομοίως, αν το κλιμακωτό σφάλμα είναι μεγαλύτερο της μονάδας, αυτό σημαίνει ότι το εξεταζόμενο μοντέλο πρόβλεψης αποδίδει χειρότερα από τη μέθοδο *Naïve*. Η αξιολόγηση των μοντέλων πρόβλεψης που χρησιμοποιούν διατομεακά δεδομένα πραγματοποιείται με χρήση του μέσου. Για διατομεακά δεδομένα, το κλιμακωτό σφάλμα μπορεί να οριστεί ως:

$$q_j = \frac{e_j}{\frac{1}{N} \sum_{i=1}^T |y_t - \bar{y}|}$$

Ο λόγος που η παραπάνω συνάρτηση δε θα παράγει τα βέλτιστα αποτελέσματα για δεδομένα εποχικών χρονολογικών σειρών είναι ότι ο μέσος θα επηρεάσει τυχόν εμφανιζόμενες τάσεις ή εποχικά μοτίβα. Στην περίπτωση των εποχικών χρονολογικών σειρών προτείνεται η χρήση της μεθόδου *Naïve* ως μέτρο σύγκρισης, και ο ορισμός του κλιμακωτού σφάλματος είναι ο παρακάτω:

$$q_j = \frac{e_j}{\frac{1}{T-m} \sum_{t=m+1}^T |y_t - y_{t-m}|}$$

Το μέσο απόλυτο σφάλμα (*Mean Absolute Scaled Error – MASE*) και το ριζικό μέσο τετραγωνισμένο σφάλμα (*Mean Squared Scaled Error – MSSE*) για τα κλιμακωτά σφάλματα ορίζονται ως εξής:

$$MASE = \text{μέσος}(|q_j|)$$

$$MSSE = \sqrt{\text{μέσος}(q_j^2)}$$

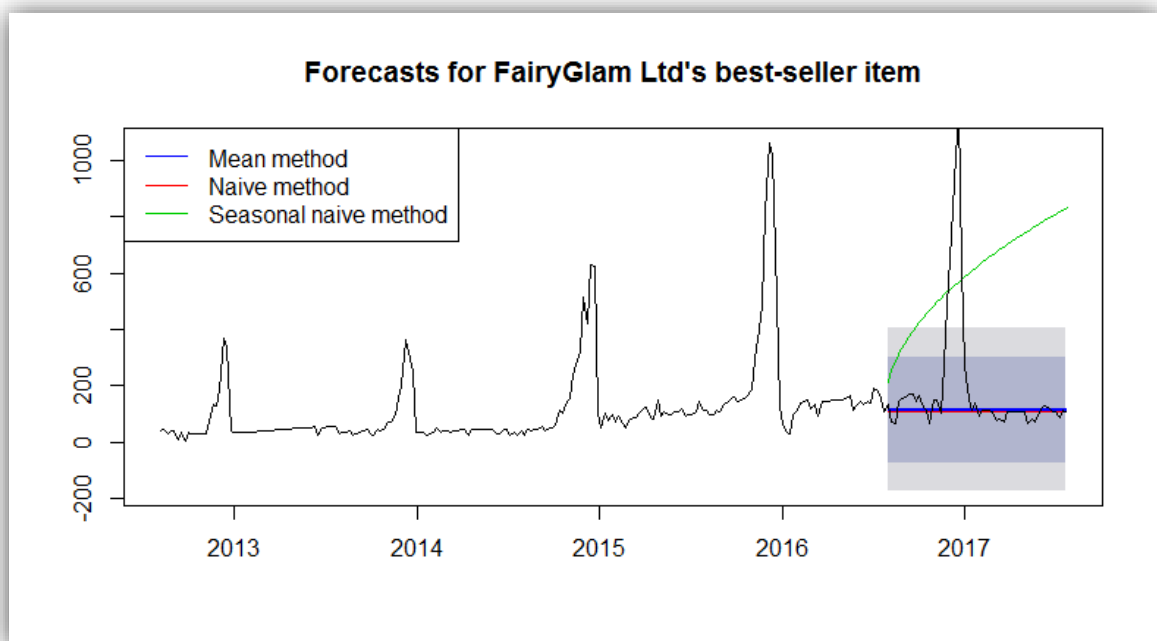
Το ριζικό μέσο τετραγωνισμένο σφάλμα είναι καθιερωμένο και χρησιμοποιείται συχνά για ασκήσεις οικονομετρίας· η δομή του είναι παρόμοια με αυτή του *MASE*, αλλά αντί για απόλυτες τιμές γίνεται χρήση των τετραγώνων. Το *MSSE* είναι επίσης γνωστό και ως το κριτήριο *U* του *Theil*.

2.5 Σύγκριση μεθόδων αξιολόγησης ακρίβειας

Οι μετρήσεις αξιολόγησης ακρίβειας που αναφέρθηκαν παραπάνω μπορούν να φανούν χρήσιμες σε διαφορετικές περιπτώσεις και για διαφορετικά σύνολα δεδομένων. Συγκεκριμένα: (Keith Ord, 2013)

- i. Το *MAPE* πρέπει να χρησιμοποιείται μόνο όταν οι τιμές του δείγματος είναι θετικές. Το *MASE* και *U* δε φέρουν τέτοιο περιορισμό.
- ii. Το *RMSE* χρησιμοποιείται επειδή το *MSE* συμπεριλαμβάνει τετραγωνισμένα σφάλματα και συνεπώς οι μετρήσεις του *MSE* παρουσιάζονται ως τετραγωνισμένες τιμές. Η χρήση του *RMSE* πάνω στις τιμές του *MSE* αντιστρέφει τη μεταμόρφωση των τιμών και διευκολύνει την ερμηνεία των αποτελεσμάτων.
- iii. Το *RMSE* δίνει περισσότερη βαρύτητα σε μεγάλα απόλυτα σφάλματα. Γι' αυτό το λόγο, μπορεί να παρατηρηθεί το φαινόμενο όπου το *RMSE* εμφανίζεται μεγαλύτερο από το *MAE*.
- iv. Οι μετρήσεις που χρησιμοποιούν απόλυτες τιμές θα είναι πάντα ίσες ή μεγαλύτερες από την απόλυτη τιμή μετρήσεων που βασίζονται πάνω στις αρχικές τιμές των σφαλμάτων. Συνεπώς, $MAE \geq ME$ και $MAPE \geq MPE$.
- v. Τα *MAPE*, *MASE* και *U* είναι ανεξάρτητα από την κλίμακα των μεταβλητών, οπότε μπορούν να χρησιμοποιηθούν για να γίνει σύγκριση πάνω σε πολλαπλές σειρές ή μεταβλητές.

Ανεξάρτητα με τη χρήση των παραπάνω μετρήσεων, είναι σημαντικό να σημειωθεί ότι η αξιολόγηση της ακρίβειας μιας πρόβλεψης πρέπει να γίνεται με τη χρήση αυθεντικών προβλέψεων. Η πρόβλεψη δε θεωρείται έγκυρη αν εφαρμοστεί πάνω στα ιστορικά δεδομένα, εφόσον έγινε χρήση αυτών για την ανάπτυξη των παραμέτρων του μοντέλου πρόβλεψης. Η ακρίβεια μιας πρόβλεψης μπορεί να υπολογιστεί με τον έλεγχο της απόδοσης της πάνω σε καινούρια δεδομένα, ανεξάρτητα αν είναι ιστορικά ή μελλοντικά, αρκεί να μην έχουν χρησιμοποιηθεί για την ανάπτυξη του ίδιου του μοντέλου πρόβλεψης. Κατά τη διαδικασία επιλογής του κατάλληλου μοντέλου ή μεθόδου πρόβλεψης, συνήθως ο έλεγχος γίνεται πάνω σε ένα σύνολο δεδομένων, από το οποίο ένα μεγάλο κομμάτι χρησιμοποιείται για την ανάπτυξη της πρόβλεψης και τα υπόλοιπα δεδομένα συγκρίνονται με τα αποτελέσματα της πρόβλεψης. Για τον παραπάνω καταμερισμό συχνά χρησιμοποιείται ο κανόνας του 80-20, αλλά αυτό δεν είναι απαραίτητο. Είναι σαφές ότι όσο περισσότερα και στατιστικά έγκυρα ιστορικά δεδομένα υπάρχουν διαθέσιμα, τόσο πιο επιτυχημένη θα είναι η πρόβλεψη, χωρίς αυτό να σημαίνει ότι η πρόβλεψη δεν μπορεί να αναπτυχθεί με μικρότερο όγκο ιστορικών δεδομένων. Ωστόσο, κατά γενικό κανόνα, το μέγεθος των ιστορικών δεδομένων που θα χρησιμοποιηθούν για την ανάπτυξη της πρόβλεψης θα πρέπει να είναι τουλάχιστον τόσο μεγάλο όσο και το μέγεθος της επιθυμητής περιόδου πρόβλεψης.



Εικόνα 4: Σύγκριση κοινών μεθόδων πρόβλεψης πάνω σε πραγματικά δεδομένα επιχείρησης – Ανάπτυξη γραφήματος με χρήση της κονσόλας R

Για το παράδειγμα που εμφανίζεται στην *εικόνα 4*, χρησιμοποιήθηκαν πραγματικά δεδομένα επιχείρησης. Η σύγκριση των μεθόδων ακρίβειας βασίζεται πάνω στα πραγματικά δεδομένα του τελευταίου εποχικού κύκλου των παρατηρήσεων και στην απόδοση κοινών μεθόδων πρόβλεψης για την ίδια χρονική περίοδο, όπως τη μέθοδο των μέσων και την απλή και εποχική μέθοδο *Naïve*. Η εποχικότητα είναι ένα από τα κυρίαρχα χαρακτηριστικά των δεδομένων που απεικονίζονται καθώς και η ανοδική τάση των πωλήσεων. Όπως μπορεί να παρατηρηθεί αρχικά, η μέθοδος των μέσων και η απλή μέθοδος *Naïve* δεν αντιπροσωπεύουν καθόλου την εποχική συμπεριφορά των πωλήσεων, και αυτό γιατί οι μέθοδοι αυτοί δεν είναι ικανοί να προβλέψουν εποχικότητα. Ωστόσο ανεξάρτητα από την κορυφή που παρατηρείται κατά το τέλος του χρόνου, οι δύο αυτοί μέθοδοι προβλέψεων που μοιάζουν πολύ μεταξύ τους, αντικατοπτρίζουν αρκετά έγκυρα τις αναμενόμενες τιμές για όλο το υπόλοιπο έτος. Από τις τρεις μεθόδους πρόβλεψης που χρησιμοποιήθηκαν, η μόνη που εμφανίζει «επιθυμητά» αποτελέσματα όσο αναφορά την ανοδική τάση είναι η εποχική μέθοδος *Naïve*, αν και αποτυχαίνει να παρουσιάσει την αναμενόμενη εποχική συμπεριφορά.

Πίνακας 1: Σύγκριση των αποδόσεων ακριβείας κοινών μεθόδων προβλέψεων

Μέθοδος		ME	RMSE	MAE	MAPE	MASE
Μέσων	Εντός δείγματος	0	145,80	82,68	145,66	1,50
	Εκτός δείγματος	70,32	226,37	90,78	29,61	1,64
Απλή Naïve	Εντός δείγματος	0,35	78,14	31,41	39,27	0,57
	Εκτός δείγματος	74,94	227,85	90,87	28,50	1,64
Εποχική Naïve	Εντός δείγματος	48,08	98,02	55,30	39,85	1,00
	Εκτός δείγματος	-18,86	69,17	54,00	41,84	0,98

Ο *πίνακας 1* εξηγεί την απόδοση των μεθόδων πρόβλεψης που απεικονίζονται στο γράφημα. Για την κάθε μέθοδο, η παραγόμενες προβλέψεις προσαρμόζονται στα ιστορικά δεδομένα και σε σύγκριση με αυτά, υπολογίζονται οι αποκλίσεις της πρόβλεψης από τα πραγματικά

δεδομένα. Από τα ιστορικά δεδομένα, τα πρώτα τέσσερα χρόνια χρησιμοποιήθηκαν για την εκμάθηση του μοντέλου πρόβλεψης και από αυτά πηγάζει η παραγόμενη πρόβλεψη. Οι παρατηρήσεις που εμφανίζονται σε αυτήν την περίοδο θεωρούνται ότι είναι «εντός δείγματος» ή επίσης ονομάζονται «δείγμα εκπαίδευσης». Το τελευταίο έτος του συνόλου των ιστορικών δεδομένων αποκολλήθηκε από το δείγμα και δε χρησιμοποιήθηκε για την ανάπτυξη των προβλέψεων. Αυτές οι παρατηρήσεις θεωρούνται ότι είναι «εκτός δείγματος» ή αλλιώς «δείγμα ελέγχου», και γίνεται χρήση αυτών για τη σύγκριση των προβλέψεων με την πραγματικότητα. Αν δεν είχε πραγματοποιηθεί ο καταμερισμός του συνόλου των ιστορικών δεδομένων, τότε στον παραπάνω πίνακα δε θα εμφανίζονταν μετρήσεις για την κατηγορία *εκτός δείγματος*.

Υπάρχουν διαφορετικοί τρόποι για να συγκριθεί η απόδοση διαφορετικών μοντέλων και μεθόδων πρόβλεψης. Οι παραπάνω μετρήσεις βοηθούν στην απόφαση του καλύτερου μοντέλου όταν αυτά είναι παρόμοια μεταξύ τους. Στο παράδειγμα που εξετάστηκε είναι προφανές ότι οι τρεις μέθοδοι που χρησιμοποιήθηκαν δεν αποδίδουν με ακρίβεια, οπότε πρέπει να ερευνηθεί η χρήση πιο περίπλοκων μεθόδων που να μπορούν να αναγνωρίσουν την περιοδικότητα των δεδομένων. Ωστόσο, είναι σημαντικό να σημειωθεί ότι τα παραπάνω κριτήρια και οι μετρήσεις απόδοσης θα πρέπει να χρησιμοποιούνται ως υποστήριξη και όχι ως κανόνα: ένα μοντέλο που προσαρμόζεται στα δεδομένα *εντός δείγματος* με μεγάλη ακρίβεια δε σημαίνει απαραίτητα ότι θα παράγει καλή πρόβλεψη. Η υλοποίηση πολλαπλών μεθόδων και μοντέλων πρόβλεψης και η παρατήρηση των παραγόμενων γραφημάτων είναι καθοριστικός παράγοντας για την τελική απόφαση της βέλτιστης μεθόδου και μοντέλου πρόβλεψης.

3 Κατανόηση μεταβλητών και στατιστική ανάλυση

Η διαδικασία ανάπτυξης μιας πρόβλεψης περιλαμβάνει ανάλυση και κατανόηση των σχετικών δεδομένων. Τα δεδομένα αυτά μπορεί να προκύψουν σε μεγάλο όγκο και σε ποικίλες μορφές με πολλαπλές μεταβλητές και κλίμακες. Αν παράλληλα τα δεδομένα αυτά δεν είναι οργανωμένα σε μία βάση, τότε η περιήγηση μέσα σε υπολογιστικά φύλλα υπερφορτωμένα με δεδομένα δεν είναι ευχάριστη υπόθεση και καθόλου βέλτιστη όσο αναφορά την εξόρυξη πληροφοριών από τα συμπυκνωμένα αυτά νούμερα. Κατά συνέπεια, σύνηθες πρακτικές στα πρώτα βήματα της ανάλυσης δεδομένων είναι η ανάπτυξη γραφημάτων και η εξερευνητική ανάλυση με μεθόδους περιγραφικής στατιστικής.

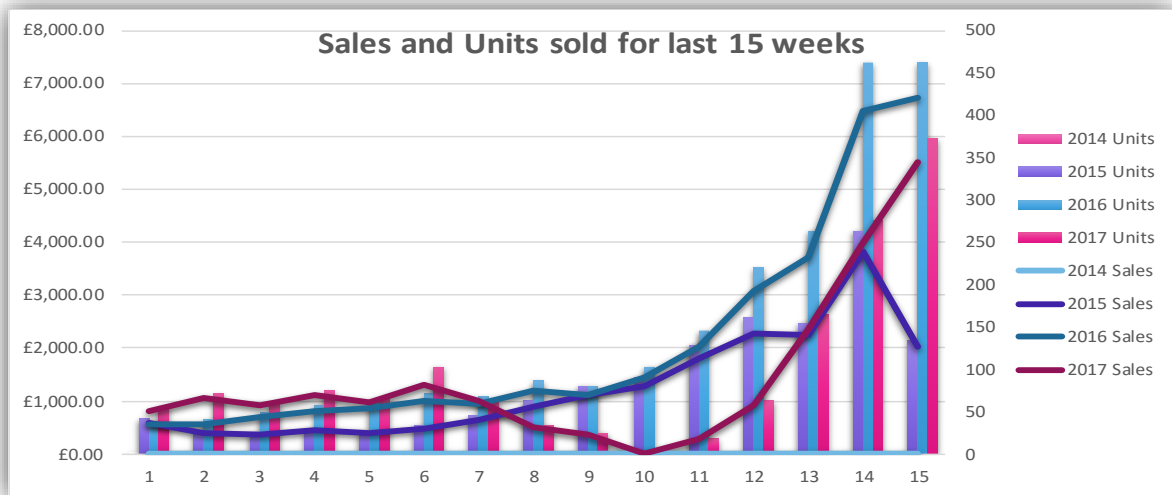
3.1 Απεικόνιση δεδομένων

Ως *γράφημα* θεωρείται η οπτική απεικόνιση αριθμητικών τιμών. Τα γραφήματα αξιοποιούνται ως ενσωματωμένες δυνατότητες υπολογιστικών φύλλων εδώ και πολλά χρόνια, από τότε που το λογισμικό Lotus 1-2-3 ξεκίνησε να χρησιμοποιείται από επιχειρήσεις για την αποθήκευση και διαχείριση των δεδομένων τους. (Walkenbach, 2007) Τα γραφήματα που υπήρχαν διαθέσιμα τότε ήταν αρκετά οπτικά δυσάρεστα σύμφωνα με τα σημερινά δεδομένα, αλλά με τον καιρό, γραφήματα που διατίθενται από διάφορα λογισμικά έχουν βελτιωθεί σημαντικά σχετικά με την ποιότητα και την προσαρμοστικότητα τους. Η απεικόνιση δεδομένων σε ένα καλά δομημένο γράφημα συνεισφέρει στην κατανόηση των δεδομένων και τυχών κρυφών μηνυμάτων που μπορεί να προκύψουν από αυτά. Τα γραφήματα παρουσιάζουν μία εικόνα, οπότε είναι ιδιαίτερα χρήσιμα για την περίληψη αριθμητικών σειρών και των σχέσεων που τις απαρτίζουν. Με την ανάπτυξη ενός γραφήματος, καθίσταται δυνατός ο προσδιορισμός τάσεων και μοτίβων που κατά τα άλλα δε θα είχαν γίνει αντιληπτά.

Προτού αναπτυχθεί ένα γράφημα, πρέπει να υπάρχουν διαθέσιμα δεδομένα που να μπορούν να αποτυπωθούν στο γράφημα. Τα δεδομένα θα πρέπει να αποτελούνται από αριθμούς και αναγνωρίσιμες μεταβλητές. Συνήθως, τα δεδομένα από τα οποία πηγάζουν τα γραφήματα βρίσκονται σε ένα υπολογιστικό φύλλο, αλλά αυτό δεν είναι απαραίτητο. Τα γραφήματα μπορεί να υπάρχουν στα ίδια φύλλα με τα δεδομένα τους ή σε ξεχωριστά. Το κάθε γράφημα

θεωρείται ως ξεχωριστό αντικείμενο και αποτελείται από μία ή περισσότερες σειρές δεδομένων που απεικονίζονται οπτικά. Η μορφή που θα έχει η απεικόνιση των δεδομένων εξαρτάται από τον τύπο του γραφήματος. Για παράδειγμα, η δημιουργία ενός γραμμικού γραφήματος που περιέχει δύο σειρές δεδομένων, θα παρουσιάσει μία εικόνα όπου η κάθε σειρά θα αποτυπώνεται ως μία διαφορετικά μορφοποιημένη γραμμή. Τα σχετικά δεδομένα θα πρέπει να έχουν αποθηκευτεί κατάλληλα, με τη σωστή σειρά σε ξεχωριστές στήλες ή γραμμές. Οι τιμές που παρουσιάζονται στο γράφημα αποτυπώνουν τη σύνδεση των ξεχωριστών κελιών που περιέχουν τα δεδομένα. Ένα πολύ σημαντικό χαρακτηριστικό που προσφέρεται από διάφορα λογισμικά υπολογιστικών φύλλων είναι ότι τα γραφήματα μπορούν να είναι δυναμικά. Αυτό σημαίνει ότι η ανάπτυξη και μορφοποίηση ενός γραφήματος μπορεί να γίνει μόνο μια φορά και από το σημείο αυτό και μετά, όποια τροποποίηση των σχετικών δεδομένων προκύψει, σημαίνει και την αυτόματη ενημέρωση του γραφήματος. Φυσικά το γράφημα μπορεί να τροποποιηθεί όσο αναφορά τα δεδομένα του και τη μορφή του οποιαδήποτε στιγμή, αλλά η δυναμικότητα συνεισφέρει στην ελαχιστοποίηση του χρόνου που απαιτείται για τη συντήρηση του γραφήματος.

Ένα γράφημα αποτελείται από πολλαπλές παραμέτρους. Δεν είναι απαραίτητη η χρήση όλων των παραμέτρων αλλά οι δυνατότητες υπάρχουν. Γίνεται να αναπτυχθούν απλά γραμμικά γραφήματα που να αποτυπώνουν μία σειρά δεδομένων, ή πιο πολύπλοκα γραφήματα με πολλαπλές μεταβλητές και άξονες. Όλα τα γραφήματα φέρουν έναν κάθετο (y) άξονα και έναν οριζόντιο (x). Συνήθως ο άξονας x αποτυπώνει τη χρονική πορεία των δεδομένων και ο άξονας y προσδιορίζει τις τιμές που τα δεδομένα αντιπροσωπεύουν. Η παραπάνω μορφοποίηση θεωρείται ως η συνηθέστερη αλλά το ακριβώς αντίθετο μπορεί επίσης να παρουσιάσει τα ίδια μηνύματα, απλώς προς διαφορετική κατεύθυνση. Επιπλέον, υπάρχει η δυνατότητα να γίνει συνδυασμός δύο γραφημάτων σε ένα και να παρουσιάζονται διαφορετικές πληροφορίες εντός της ίδιας εικόνας αλλά να αντιπροσωπεύονται από τρίτο άξονα y' , ο οποίος θα φέρει διαφορετικές τιμές ή κλίμακα από τον αρχικό y . Στην περίπτωση που απεικονίζονται πολλαπλές σειρές δεδομένων από τις ίδιες ή διαφορετικές μεταβλητές, προτείνεται η χρήση πίνακα περιεχομένων εντός του γραφήματος, ώστε οι προκείμενες πληροφορίες να είναι ερμηνεύσιμες.

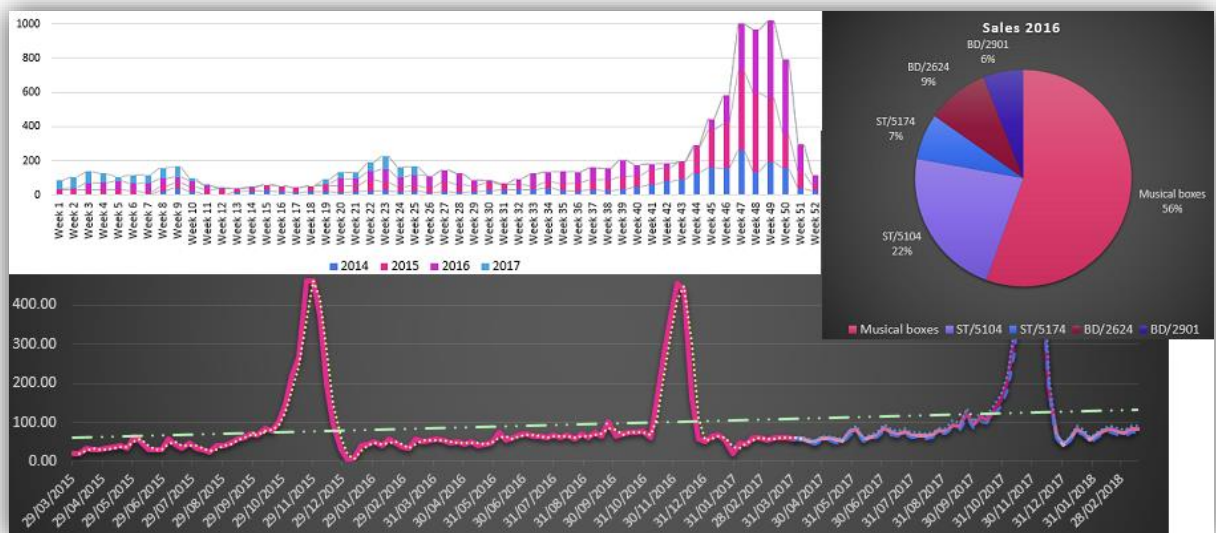


Εικόνα 5: Γράφημα που συνδυάζει οκτώ σειρές δεδομένων από δύο διαφορετικές μεταβλητές με διπλούς κάθετους άξονες και πίνακα περιεχομένων.

Αξίζει να σημειωθεί ότι η πολυπλοκότητα δεν είναι το χαρακτηριστικό που σηματοδοτεί ένα καλό γράφημα, αφού η υπερφόρτωση πληροφοριών καθιστά την κατανόηση τους δύσκολη και τη χρήση τους ανούσια. Η σημασία ενός γραφήματος αποτυπώνεται στον τρόπο που παρουσιάζει τις υπάρχουσες πληροφορίες. Ένας αναλυτής που αναπτύσσει γραφήματα θα πρέπει να μπορεί να παρουσιάσει έγκυρα τα μηνύματα που θα πρέπει να πηγάζουν από τα γραφήματα αυτά. Συνήθως το κυρίαρχο μήνυμα επισημαίνεται στον τίτλο του γραφήματος ή γίνεται αντιληπτό από τον πίνακα περιεχομένων που υποστηρίζει το γράφημα. Το ίδιο το γράφημα προσφέρει οπτική υποστήριξη για το σημαντικότερο του μήνυμα. Η επιλογή του κατάλληλου τύπου γραφήματος είναι επίσης σημαντική για την εικόνα του μηνύματος, οπότε αξίζει να αφιερωθεί παραπάνω προσοχή και χρόνος για την τελική επιλογή του τύπου γραφήματος που θα χρησιμοποιηθεί. Ο πιο κοινός πρωτεύον στόχος για τη χρήση γραφημάτων είναι η κάποιου είδους σύγκριση δεδομένων. Μερικά γενικά και συχνά είδη συγκρίσεων αποτυπώνονται παρακάτω: (Walkenbach, 2007)

- i. Σύγκριση αντικειμένων – Για παράδειγμα, σε ένα γράφημα μπορεί να αποτυπώνονται οι πωλήσεις διαφορετικών προϊόντων μιας επιχείρησης για το ίδιο χρονικό διάστημα.
- ii. Χρονική σύγκριση δεδομένων – Παρατήρηση των πωλήσεων ενός προϊόντος όπως αυτό συμπεριφέρεται για κάποιο χρονικό διάστημα. Από αυτά τα γραφήματα συνήθως προσδιορίζονται οι τάσεις, περιοδικοί κύκλοι, εποχικότητα κλπ.

- iii. Σύγκριση διαφορετικών τύπων γραφημάτων – Γραφήματα πίτας συνήθως χρησιμοποιούνται για εναλλακτική απεικόνιση των αρχικών γραφημάτων. Είναι απλώς ένας εναλλακτικός τρόπος απεικόνισης των πληροφοριών.
- iv. Σύγκριση συχνότητας – Οι συχνότητες εμφάνισης συγκεκριμένων τιμών ή συστάδων συχνά απεικονίζονται ως ιστογράμματα.
- v. Σύγκριση μεταβλητών – Τα διαγράμματα διασποράς χρησιμοποιούνται για την παρατήρηση της συσχέτισης ενός ζεύγους τιμών. Ένα διάγραμμα διασποράς που υποδεικνύει υψηλή γραμμική συσχέτιση σημαίνει ότι η σύγκριση των δύο προκειμένων σειρών δεδομένων θα εμφανίσει αρκετά χαρακτηριστικά πως το ένα επηρεάζει το άλλο, με κάποιο τρόπο.
- vi. Προσδιορισμός σφαλμάτων (καταλοίπων) ή ασυνήθιστων καταστάσεων – Για παράδειγμα, αν ο σκοπός της ανάλυσης είναι ο προσδιορισμός της ζήτησης, μία εξάντληση αποθεμάτων θα μηδενίσει τις πωλήσεις για το χρονικό διάστημα που παρατηρείται αλλά αυτό δε σημαίνει ότι δεν υπήρχε ζήτηση για το συγκεκριμένο προϊόν και για τη χρονική περίοδο που μελετάται.



Εικόνα 6: Γράφημα σωρευμένων στηλών για πολλαπλές σειρές (πάνω), χρονολογικό γράφημα με ενσωματωμένες προβλέψεις (κάτω), διάγραμμα πίτας (δεξιά)

Οι οπτικές περιλήψεις των δεδομένων με χρήση γραφημάτων προσφέρουν ανεκτίμητη διορατικότητα. Τα χρονολογικά γραφήματα και τα διαγράμματα διασποράς αποτελούν

πολύτιμο εργαλείο για τη διαδικασία ανάλυσης δεδομένων και θα πρέπει να χρησιμοποιούνται απαραίτητως στα αρχικά στάδια μιας διαδικασίας ανάπτυξης προβλέψεων για να διευκολύνουν τον αναλυτή να κατανοήσει τα δεδομένα του. Εκτός το κομμάτι της κατανόησης, αυτά τα εργαλεία συνήθως φαίνονται ιδιαίτερα χρήσιμα για την αναγνώριση σφαλμάτων των παραγόμενων προβλέψεων και τη βελτιστοποίηση αυτών. Ακόμα και στην περίπτωση που ο όγκος των προβλέψεων που καλείται κάποιος να αναπτύξει είναι πολύ μεγάλος, τα διαγράμματα αυτά προσφέρουν ιδιαίτερη καθοδήγηση και ενόραση. Ωστόσο, θα πρέπει να αναγνωριστεί ότι συχνά καθίστανται απαραίτητη η χρήση κάποιου είδους αριθμητικής περίληψης, όπως ο μέσος όρος των διαθέσιμων τιμών, οι τυπικές τους αποκλίσεις κλπ. Όπως παρουσιάζεται σε παραπάνω κεφάλαιο, αυτές οι μετρήσεις είναι επίσης χρήσιμες για τη διαδικασία αξιολόγησης απόδοσης των μοντέλων πρόβλεψης και προσδιορισμού τις συνέπειας των σφαλμάτων τους. Στο επόμενο κεφάλαιο θα παρουσιαστούν μέθοδοι της περιγραφικής στατιστικής που μπορούν να χρησιμοποιηθούν για την περίληψη των δεδομένων και την απόκτηση πληροφοριών.

3.2 Εξερευνητική ανάλυση δεδομένων

Κατά την ανάλυση δεδομένων, η χρήση περιγραφικής στατιστικής είναι πολύ σημαντική για την κατανόηση των δεδομένων που εξετάζονται. Η εξερευνητική ανάλυση δεδομένων (*Exploratory Data Analysis – EDA*) αποτελείται από μεθόδους περιγραφικής στατιστικής. Η χρήση της για την εξέταση των διαθέσιμων μεταβλητών, πριν την ανάπτυξη οποιουδήποτε μοντέλου πρόβλεψης είναι ιδιαίτερα σημαντική. Με τη χρήση τεχνικών της *EDA* ένας ερευνητής θα εμβαθύνει την κατανόηση του για τα δεδομένα που καλείται να αναλύσει, εφόσον με τα αποτελέσματα της *EDA* θα μπορέσει να βγάλει συμπεράσματα για να αναγνωρίσει λάθη, να παρατηρήσει μοτίβα, να παράγει υποθέσεις και να αντιληφθεί τυχόν παραβιάσεις των στατιστικών υποθέσεων. Με την *EDA* μπορεί να διαμορφωθεί η εικόνα που θα πρέπει να έχουν τα αποτελέσματα μιας πρόβλεψης ώστε να γίνουν αναγνωρίσιμα τυχόν προβλήματα που μπορεί να εμφανιστούν.

Τα γραφήματα που αναφέρθηκαν παραπάνω υπάγονται επίσης στη μελέτη της περιγραφικής στατιστικής, όπως φαίνεται παραπάνω στον πίνακα 2. Σε αυτό το κεφάλαιο θα αναπτυχθούν μερικά από τα μέτρα και τις μεθόδους της *EDA*, τα οποία θα χρησιμοποιηθούν στο β' μέρος της παρούσας εργασίας κατά την ανάπτυξη μοντέλων προβλέψεων για πραγματικά δεδομένα

επιχείρησης. Ο μέσος όρος είναι μία τιμή που θεωρείται τυπική ή αντιπροσωπευτική για ένα σύνολο δεδομένων. Εφόσον οι τυπικές τιμές συνήθως κυμαίνονται κεντρικά μέσα σε ένα σύνολο μετρήσεων σύμφωνα με το μέγεθος του, οι μέσοι όροι επίσης ονομάζονται και μέτρα κεντρικής τάσης. Υπάρχουν διάφορα αριθμητικά μέτρα θέσης που μπορούν να χρησιμοποιηθούν κατά την περίληψη των δεδομένων για τη συλλογή των αρχικών πληροφοριών από τις παρατηρήσεις. (Murray R. Spiegel, 2014)

Πίνακας 2: Περιεχόμενα της Περιγραφικής Στατιστικής

Περιγραφική στατιστική	
Συνεχής Μεταβλητές	Θέσης Μέσος (Αριθμητικός, Αρμονικός, Γεωμετρικός), Διάμεσος, Επικρατούσα Τιμή, Τεταρτημόρια
	Μεταβλητότητας Εύρος Κατανομής, Διασπορά, Τυπική Απόκλιση, Συντελεστής Μεταβλητότητας
	Μορφής Ασσυμετρία, Κύρτωση
Μετρήσιμες Μεταβλητές	Συντελεστής Διασποράς
Περιληπτικοί Πίνακες	Καταμερισμός Κλάσεων, Σχετική Συχνότητα, Πίνακας Κατανομής Συχνοτήτων
Εξάρτηση	Συσχέτιση (Pearson, Spearman, Kendall), Μερική Συσχέτιση, Διαγράμματα Διασποράς
Γραφήματα	Γράφημα – στηλών, γραμμών, φυσαλίδων, σημείων, ακτινικό, περιοχής Διάγραμμα – συσχέτισεων, πίτας, διασποράς Ιστόγραμμα

(Wikipedia: Descriptive Statistics, 2017)

3.2.1 Ο αριθμητικός μέσος

Ο αριθμητικός μέσος (*mean*) \bar{X} προτιμάται ως μέτρο κεντρικής τάσης και χρησιμοποιείται συνήθως για περαιτέρω στατιστική ανάλυση επειδή υπολογίζεται εύκολα, συμπεριλαμβάνει όλες τις παρατηρήσεις του συνόλου και παράγει μόνο μία τιμή που είναι αντιπροσωπευτική για τα δεδομένα που μελετώνται, αρκεί τα δεδομένα αυτά να ακολουθούν την κανονική κατανομή. Βέβαια, αυτό σημαίνει ότι, εφόσον στο σύνολο συμπεριλαμβάνονται ακραίες τιμές ο μέσος μπορεί να μην είναι απόλυτα αντιπροσωπευτικός για την πλειοψηφία του δείγματος. Στην περίπτωση που το δείγμα δεν ακολουθεί την κανονική κατανομή τότε ο αριθμητικός μέσος θεωρείται ακατάλληλος ως μέτρο κεντρικής τάσης.

Ο αριθμητικός μέσος, ή αλλιώς ο μέσος ενός συνόλου N αριθμών $X_1, X_2, X_3, \dots, X_n$ ορίζεται ως:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N} = \frac{\sum_{j=1}^N X_j}{N} = \frac{\sum X}{N}$$

Στην περίπτωση που οι αριθμοί $X_1, X_2, X_3, \dots, X_n$ εμφανίζονται $f_1, f_2, f_3, \dots, f_n$ φορές, ο αριθμητικός μέσος είναι:

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + f_3 X_3 + \dots + f_N X_N}{f_1 + f_2 + \dots + f_K} = \frac{\sum_{j=1}^K f_j X_j}{\sum_{j=1}^K f_j} = \frac{\sum fX}{N}$$

Ιδιότητες του μέσου: (Murray R. Spiegel, 2014)

- i. Το αλγεβρικό άθροισμα των αποκλίσεων ενός συνόλου αριθμών από τον αριθμητικό μέσο είναι ίσο με το μηδέν.
- ii. Το άθροισμα των τετραγώνων των αποκλίσεων ενός συνόλου αριθμών X_j από οποιοδήποτε αριθμό α είναι ελάχιστο αν και μόνο αν $\alpha = \bar{X}$.
- iii. Αν f_1 αριθμοί (είναι γνωστή η συχνότητα τους αλλά όχι οι αρχικές τιμές) έχουν μέσο \bar{X}_1 , f_2 αριθμοί έχουν μέσο \bar{X}_2, \dots, f_K αριθμοί έχουν μέσο \bar{X}_K , τότε ο μέσος του συνόλου είναι:

$$\bar{X} = \frac{f_1 \bar{X}_1 + f_2 \bar{X}_2 + \dots + f_K \bar{X}_K}{f_1 + f_2 + \dots + f_K}$$

- iv. Αν ένας οποιοσδήποτε αριθμός A θεωρείται ως ο αριθμητικός μέσος και αν $d_j = X_j - A$ είναι οι αποκλίσεις του X_j από τον μέσο A , τότε ισχύει:

$$\bar{X} = \frac{\sum_{j=1}^N X_j}{N} = A + \frac{\sum_{j=1}^N d_j}{N} = A + \frac{\sum d}{N}$$

$$\bar{X} = \frac{\sum_{j=1}^K f_j X_j}{\sum_{j=1}^K f_j} = A + \frac{\sum_{j=1}^K f_j d_j}{\sum_{j=1}^K f_j} = A + \frac{\sum f d}{N}$$

Όπου $N = \sum_{j=1}^K f_j = \sum f$.

3.2.2 Η διάμεσος

Η διάμεσος (*median*) M ενός συνόλου αριθμών είναι η τιμή της μεσαίας παρατήρησης αν το πλήθος των παρατηρήσεων είναι περιττός αριθμός, ή του αριθμητικού μέσου των δύο μεσαίων τιμών αν το πλήθος είναι ζυγός αριθμός, από ένα σύνολο παρατηρήσεων διατεταγμένων σε αύξουσα ή φθίνουσα σειρά. Το πλεονέκτημα της διαμέσου σε σύγκριση με τον αριθμητικό μέσο είναι ότι η διάμεσος δεν επηρεάζεται από ακραίες τιμές, αφού αποτελείται από μία τιμή πραγματικής παρατήρησης. Σε ένα μεγάλο δείγμα και με την προϋπόθεση ότι δεν υπάρχουν πολλές επαναλαμβανόμενες τιμές, περίπου το 50% του δείγματος θα είναι μικρότερο από τη διάμεσο και το υπόλοιπο 50% θα είναι μεγαλύτερο.

3.2.3 Η επικρατούσα τιμή

Η επικρατούσα τιμή (*mode*) ενός συνόλου δεδομένων είναι η τιμή που εμφανίζεται με τη μεγαλύτερη συχνότητα, δηλαδή είναι η συχνότερα παρατηρούμενη τιμή. Δεν είναι απαραίτητο ότι σε όλα τα παρατηρούμενα σύνολα θα υπάρχει κάποια επικρατούσα τιμή, όπως επίσης δε σημαίνει ότι θα είναι μοναδική. Επειδή η παρουσία της και η μοναδικότητα της δεν είναι εγγυημένη, η επικρατούσα τιμή συνήθως δε θεωρείται ιδιαίτερα αξιόπιστη για περαιτέρω στατιστική ανάλυση. Στην περίπτωση που έχει σχεδιαστεί καμπύλη συχνοτήτων για ομαδοποιημένα δεδομένα, η επικρατούσα τιμή είναι εκείνη που εμφανίζεται ως η ψηλότερη κορυφή της καμπύλης. Αν το δείγμα ακολουθεί την κανονική κατανομή, η επικρατούσα τιμή είναι επίσης αμερόληπτη εκτίμηση για τον αριθμητικό μέσο και τη διάμεσο.

3.2.4 Τεταρτημόρια και εκατοστημόρια

Για ένα σύνολο δεδομένων που είναι διατεταγμένο σε αύξουσα σειρά, η μεσαία τιμή ή ο μέσος όρος των δύο μεσαίων τιμών είναι η διάμεσος. Η θέση της διάμεσου βρίσκεται στο 50% των τιμών, χωρίζοντας έτσι το δείγμα σε δύο. Με τον ίδιο τρόπο υπολογίζονται και τα τεταρτημόρια (*quartiles*). Το δεύτερο τεταρτημόριο Q_2 είναι ίσο με τη διάμεσο και βρίσκεται ακριβώς στη μέση των τιμών, ενώ το πρώτο Q_1 και τρίτο Q_3 τεταρτημόρια χωρίζουν τα αρχικά δύο μέρη του δείγματος που είχαν διασπαστεί από το δεύτερο τεταρτημόριο, το καθένα σε άλλα δύο μέρη. Έτσι το δείγμα καταμερίζεται σε τέσσερις ομάδες και πέντε κορυφές, με αρχή τη μικρότερη τιμή (*min*) και τέλος τη μέγιστη (*max*). Ενδιάμεσα κυμαίνονται τα τρία τεταρτημόρια, το πρώτο βρίσκεται στο 25% του συνόλου και το τρίτο στο 75%. Εφόσον οι θέσεις των τεταρτημόριων μπορούν να διευκρινιστούν από το ποσοστό του συνόλου στο οποίο βρίσκονται, παρομοίως απορρέουν τα εκατοστημόρια (*percentiles*), που έχουν τον ίδιο σκοπό όπως τα τεταρτημόρια αλλά περισσότερη ευελιξία. Ισχύει ότι $Q_1 = 25^\circ$ εκατοστημόριο, $Q_2 = M = 50^\circ$ εκατοστημόριο και $Q_3 = 75^\circ$ εκατοστημόριο. (Murray R. Spiegel, 2014)

3.2.5 Η τυπική απόκλιση

Ο βαθμός με τον οποίο αριθμητικά δεδομένα αποκλίνουν από τη μέση τιμή λέγεται *μεταβλητότητα*, ή *διασπορά* του δείγματος. Υπάρχουν διάφορες διαθέσιμες μετρήσεις μεταβλητότητας όπως το εύρος, η μέση απόκλιση, η *τυπική απόκλιση* κλπ.

Η τυπική απόκλιση (σ ή s) ενός συνόλου N αριθμών X_1, X_2, \dots, X_n ορίζεται ως:

$$\sigma = \sqrt{\frac{\sum_{j=1}^N (X_j - \bar{X})^2}{N}} = \sqrt{\frac{\sum x^2}{N}} = \sqrt{(X - \bar{X})^2}$$

Όπου το x αντιπροσωπεύει τις αποκλίσεις του κάθε ένα από τους αριθμούς X_j από τον μέσο \bar{X} . Συνεπώς, η τυπική απόκλιση σ είναι το ριζικό μέσο τετράγωνο (*RMS*) των αποκλίσεων από τον μέσο. Η τυπική απόκλιση επίσης ονομάζεται ριζική μέση τετραγωνική απόκλιση.

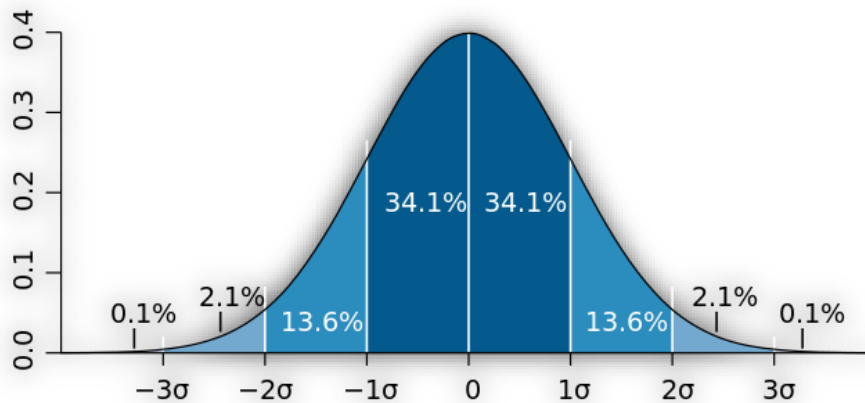
Ιδιότητες της τυπικής απόκλισης: (Murray R. Spiegel, 2014)

- i. Η τυπική απόκλιση μπορεί επίσης να οριστεί ως:

$$\sigma = \sqrt{\frac{\sum_{j=1}^N (X_j - \alpha)^2}{N}}$$

Όπου ο α είναι κάποιος μέσος εκτός του αριθμητικού. Για όλες αυτές τις τυπικές αποκλίσεις το ελάχιστο μπορεί να είναι $\alpha = \bar{X}$.

- ii. Για κανονικές κατανομές, όπως εμφανίζεται στην *εικόνα 7*, ισχύουν τα παρακάτω:
- Το 68,27% των παρατηρήσεων εμφανίζονται μεταξύ των αποστάσεων μίας τυπικής απόκλισης από την κάθε μεριά του μέσου· $\bar{X} - \sigma$ και $\bar{X} + \sigma$
 - Το 95,45% των παρατηρήσεων εμφανίζονται μεταξύ των αποστάσεων δύο τυπικών αποκλίσεων από την κάθε μεριά του μέσου· $\bar{X} - 2\sigma$ και $\bar{X} + 2\sigma$
 - Το 99,73% των παρατηρήσεων εμφανίζονται μεταξύ των αποστάσεων τριών τυπικών αποκλίσεων από την κάθε μεριά του μέσου· $\bar{X} - 3\sigma$ και $\bar{X} + 3\sigma$



(Wikimedia File:Standard deviation diagram.svg, 2015)

Εικόνα 7: Η καμπύλη της κανονικής κατανομής

3.2.6 Συσχέτιση

Η *ανάλυση της εξάρτησης* των μεταβλητών αποτελείται από μία ομάδα από τεχνικές που μετρούν τη *συσχέτιση* μεταξύ δύο μεταβλητών. (Douglas A. Lind, 2015) Εκτός από τις τεχνικές περίληψης και αντιπροσωπευτικών μέτρων των δεδομένων, στην *περιγραφική στατιστική* υπάγεται και η *ανάλυση της εξάρτησης*. Υπάρχουν διάφορες μετρήσεις συσχέτισης μεταβλητών, όπως φαίνεται στον *πίνακα 2*, και μία από αυτές είναι ο συντελεστής γραμμικής συσχέτισης του Pearson (*Pearson's correlation coefficient*). Η γραμμικότητα των δεδομένων

είναι μία από τις προϋποθέσεις που θα πρέπει να πληρείται για την ανάπτυξη μοντέλων γραμμικής παλινδρόμησης, κάτι που θα παρουσιαστεί σε επόμενα κεφάλαια. Συνεπώς, ο συντελεστής συσχέτισης του Pearson και τα διαγράμματα διασποράς έχουν ιδιαίτερη σημασία για τη διαδικασία ανάπτυξης μοντέλων πρόβλεψης, τουλάχιστον για μερικές από τις μεθόδους προβλέψεων που παρουσιάζονται στο παρόν έγγραφο. Γιατί είναι όμως σημαντική η ανάλυση εξάρτησης μεταβλητών και πως αλλιώς μπορεί να φανεί χρήσιμη;

Σε όλους τους επιχειρηματικούς τομείς, από τον προσδιορισμό και τη μελέτη των σχέσεων μεταξύ των μεταβλητών μπορεί να προκύψει γνώση σχετικά με τρόπους μεγιστοποίησης εσόδων, ελαχιστοποίησης κόστους, ή προσδιορισμό μεταβλητών ικανών να προβλέψουν τη ζήτηση. Μία συνήθης επιχειρηματική τεχνική είναι η μείωση των τιμών ενός προϊόντος μέσω προσφορών ή εκπτώτικών κουπονιών με σκοπό την αύξηση του όγκου των πωλήσεων. Η προκείμενη αύξηση του όγκου πωλήσεων ενός συγκεκριμένου προϊόντος μπορεί κατά συνέπεια να αυξήσει τη μελλοντική ζήτηση του προϊόντος, να προκαλέσει την αύξηση πωλήσεων συμπληρωματικών προϊόντων καθώς και να έχει άλλες συνέπειες, επισημαίνοντας έτσι τη συνολική αύξηση των εσόδων. Για να συμβούν όμως τα παραπάνω, θα πρέπει να μελετηθεί η συσχέτιση της μείωσης τιμής με τις πωλήσεις του συγκεκριμένου προϊόντος. Η ανάλυση εξάρτησης των δύο μεταβλητών, με την προϋπόθεση ότι θα υπάρχουν διαθέσιμα δεδομένα για να γίνει οποιαδήποτε ανάλυση, θα πρέπει να αποτελέσει σε υποδείξεις μιας ισχυρής συσχέτισης μεταξύ της τιμής και των πωλήσεων.

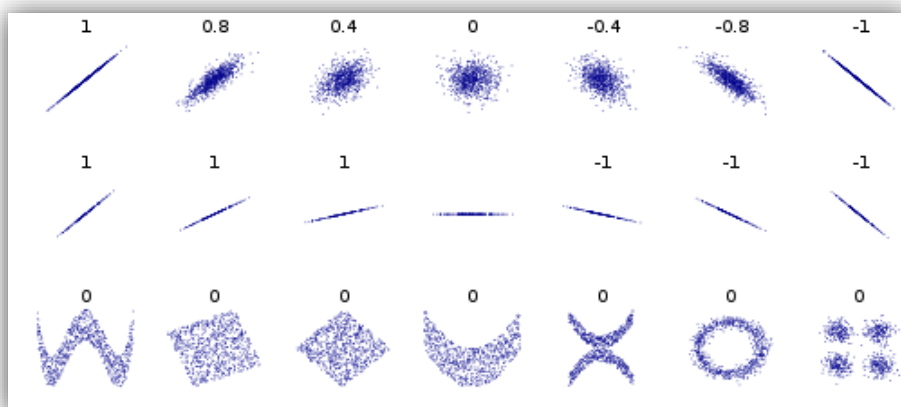
Καταρχάς, η μέτρηση της συσχέτισης θα πρέπει να μελετάται με τη χρήση διαγραμμάτων διασποράς, στα οποία γίνεται σύγκριση των δύο μεταβλητών και η εικόνα που προκύπτει θα υποδεικνύει την ύπαρξη γραμμικής συσχέτισης μεταξύ των μεταβλητών και το είδος αυτής, όπως φαίνεται στην εικόνα 8. Εφόσον το διάγραμμα διασποράς εμφανίζει καλή πιθανότητα για την ύπαρξη γραμμικής συσχέτισης, το επόμενο βήμα είναι ο υπολογισμός του συντελεστή συσχέτισης, ο οποίος αποτελεί ένα μετρήσιμο μέγεθος της σχέσης μεταξύ των μεταβλητών. Ο συντελεστής συσχέτισης (r) του Pearson, περιγράφει τη δύναμη της σχέσης μεταξύ δύο ομάδων μεταβλητών κλίμακας διαστήματος ή αναλογίας. Τα χαρακτηριστικά του συντελεστή συσχέτισης του Pearson είναι τα εξής: (Douglas A. Lind, 2015)

- i. Ο συντελεστής συσχέτισης εμφανίζει την κατεύθυνση και τη βαρύτητα της γραμμικής συσχέτισης δύο μεταβλητών.
- ii. Τα άκρα των πιθανών τιμών του r ορίζονται ως το ± 1 .

- iii. Αν η τιμή του r είναι μηδέν ή πολύ κοντά στο μηδέν, σημαίνει ότι η γραμμική συσχέτιση των μεταβλητών είναι *ελάχιστη* ή *δεν υπάρχει καθόλου*.
- iv. Αν η τιμή του r είναι κοντά στο $+1$, σημαίνει ότι η γραμμική συσχέτιση των μεταβλητών είναι *τέλεια και θετική*. Στο διάγραμμα διασποράς η συσχέτιση με συντελεστή $+1$ έχει *ανοδική κλίση*.
- v. Αν η τιμή του r είναι κοντά στο -1 , σημαίνει ότι η γραμμική συσχέτιση των μεταβλητών είναι *τέλεια και αρνητική*. Στο διάγραμμα διασποράς η συσχέτιση με συντελεστή -1 έχει *καθοδική κλίση*.
- vi. Ο συντελεστής συσχέτισης r μπορεί να οριστεί ως:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}}$$

Όπου ως x ορίζεται η πρώτη μεταβλητή και η δεύτερη ως y . (Rob J Hyndman, 2017) Ο συντελεστής συσχέτισης του Pearson μετράει μόνο την παρουσία μιας *γραμμικής* συσχέτισης: είναι δυνατόν δύο μεταβλητές x και y να έχουν ισχυρή μη-γραμμική συσχέτιση αλλά χαμηλό συντελεστή r . Το πρόβλημα της μη-γραμμικότητας μπορεί προφανώς να αντιμετωπιστεί με χρήση μεθόδων προβλέψεων που δεν προϋποθέτουν γραμμικότητα, αλλιώς προτείνεται η χρήση μεταμορφώσεων των δεδομένων ώστε να διορθωθεί η μη-γραμμικότητα και τα δεδομένα να μπορέσουν κατά συνέπεια να ενταχθούν σε γραμμικά μοντέλα.



(Wikimedia File:Correlation examples.png, 2010)

Εικόνα 8: Παραδείγματα συσχέτισης

3.3 Μεταμορφώσεις

Πολλές φορές, ένας αναλυτής μπορεί να ανησυχεί και μόνο στην ιδέα μιας μεταμόρφωσης των δεδομένων. Τα πηγαία δεδομένα στην αυθεντική τους μορφή γενικώς θεωρούνται ότι έχουν τη μεγαλύτερη στατιστική σημασία, οπότε οποιαδήποτε τροποποίηση συμβεί πάνω στα δεδομένα θα μπορούσε να αλλοιώσει τη στατιστική τους σημασία. Τι συμβαίνει όμως αν μερικές μετατροπές των ιστορικών δεδομένων σηματοδοτήσουν την επιτηδειότητα τους για χρήση σε ένα πιο απλό μοντέλο πρόβλεψης; Αν η ανατροπή της μεταμόρφωσης πάνω στα αποτελέσματα γίνεται εύκολα; Ο στόχος της πρακτικής αυτής είναι η απλοποίηση των μοτίβων στα ιστορικά δεδομένα αφαιρώντας γνωστές πηγές επιρροών ή μετατρέποντας τη ροή των δεδομένων σε πιο σταθερή και λιγότερο μεταβαλλόμενη. Συνήθως, απλούστερες σειρές δεδομένων παραπέμπουν σε προβλέψεις με μεγαλύτερη ακρίβεια. (Rob J. Hyndman, 2017) Υπάρχουν πολλαπλοί τύποι μεταμορφώσεων που χρησιμοποιούνται τακτικά σε διαφόρων ειδών αναλύσεις δεδομένων. Σε αυτό το έγγραφο θα γίνουν δύο ειδών μεταμορφώσεις: μαθηματικές μεταμορφώσεις και ημερολογιακές μετατροπές.

Από τις διάφορες μαθηματικές μεταμορφώσεις που μπορούν να χρησιμοποιηθούν, οι λογάριθμοι συνήθως προτιμώνται διότι είναι ερμηνεύσιμοι: όποια αλλαγή συμβεί σε μία λογαριθμικά μεταμορφωμένη τιμή είναι σχετική και ανάλογη με την αρχική κλίμακα των δεδομένων. Άμα χρησιμοποιηθούν λογάριθμοι με τη βάση του 10, τότε η αύξηση κατά μία μονάδα στη λογαριθμική κλίμακα ισοδυναμεί με τον πολλαπλασιασμό της αρχικής τιμής επί 10. Η λογαριθμική μεταμόρφωση μπορεί να αποτελέσει τη λύση στο πρόβλημα της μη γραμμικότητας των δεδομένων. Με τη μετατροπή των δεδομένων σε γραμμικά, τα διαγράμματα διασποράς και η συσχέτιση του Pearson θα εμφανίσουν ακριβέστερες εξαρτήσεις μεταξύ των μεταβλητών. Τα γραμμικά δεδομένα θα μπορέσουν πλέον να ενταχθούν σε γραμμικά παλινδρομικά μοντέλα υψηλών αποδόσεων, και τα λογαριθμικά αποτελέσματα θα είναι εύκολα ερμηνεύσιμα με την απλή ανατροπή των λογαρίθμων στην αυθεντική κλίμακα.

Παρομοίως, οι ημερολογιακές μετατροπές μπορούν να φανούν χρήσιμες για την απλοποίηση των δεδομένων και την ανάπτυξη απλούστερων μοντέλων πρόβλεψης με υψηλότερη ακρίβεια. Η μεταμόρφωση αυτού του τύπου αναφέρεται σε χρονολογικές σειρές και φαινόμενα εποχικότητας, ή και μη εποχικά φαινόμενα τα οποία όμως είναι ανιχνεύσιμα. Αν υπάρχει η γνώση σχετικά με την εμφάνιση συγκεκριμένων φαινομένων και μπορούν να αναπαραχθούν χωρίς τη χρήση μοντέλων πρόβλεψης, η αφαίρεση τους μπορεί να

απλοποιήσει τη διαδικασία σε σημαντικό βαθμό. Για παράδειγμα, η εποχικότητα μπορεί να προσδιοριστεί και να αφαιρεθεί από τα δεδομένα και στη συνέχεια να προστεθεί ξανά πάνω στις προβλέψεις με χρήση εποχικών συντελεστών. Επιπλέον, σε κάθε χρονολογική σειρά υπάρχουν τυχαία γεγονότα που δεν αντιπροσωπεύουν τις πραγματικές τιμές που θα μπορούσαν τα δεδομένα να έχουν. Αυτό συμβαίνει διότι για παράδειγμα, προκύπτουν καθυστερήσεις παραλαβής αποθεμάτων από διάφορες αιτίες, υποκαταστήματα κλείνουν, νόμοι καθιερώνονται, άνθρωποι κάνουν λάθη, κλπ. Πολλές φορές, η αφαίρεση των τυχαίων παραγόντων από τα δεδομένα δημιουργεί ένα πιο εύκολα προβλέψιμο δείγμα. Ένας απλός τρόπος να γίνει η μεταμόρφωση μιας χρονολογικής σειράς είναι η αποσύνθεση της και η διεξαγωγή της πρόβλεψης πάνω στην προκειμένη τάση, η οποία δε συμπεριλαμβάνει την εποχικότητα ή το τυχαίο χαρακτηριστικό.

3.4 Δομημένη ανάλυση δεδομένων

Ανάλογα με το μέσο αποθήκευσης τους, τα δεδομένα μπορούν να κατηγοριοποιηθούν σε δομημένα ή μη δομημένα. Τα κύρια χαρακτηριστικά ενός δομημένου συνόλου δεδομένων είναι η εύκολη αναζήτηση, η ακεραιότητα τους και η ευκολία της ερμηνείας τους. Τα δομημένα δεδομένα συνήθως βρίσκονται αποθηκευμένα σε σύνθετες βάσεις δεδομένων ή υπολογιστικά φύλλα με ερμηνεύσιμα αναγνωριστικά, όπως τίτλους στηλών και γραμμών. Αντίθετα, τα μη δομημένα δεδομένα εμφανίζονται σε πιο ελεύθερη μορφή, εκτός πινάκων, διασκορπισμένα και γενικά, είναι δύσκολα ανακτήσιμα.

Οι περισσότεροι τύποι ανάλυσης δεδομένων για διεξαγωγή συμπερασμάτων και προβλέψεων απαιτούν δομημένα δεδομένα. Για να πραγματοποιηθεί ανάλυση πάνω σε μη δομημένα δεδομένα πρέπει πρώτα να γίνουν προσπάθειες δόμησης τους. Συνήθως, σε αυτήν την περίπτωση χρησιμοποιείται εξόρυξη δεδομένων και προσεγγίσεις ανάλυσης κειμένων, ώστε να οργανωθούν έγγραφα, να γίνει σύνδεση των περιεχομένων τους και ανεύρεση μοτίβων, να ομαδοποιηθούν τα δεδομένα και να γίνει χρήση περιληπτικών μεθόδων. (Anasse Bari, 2014) Εφόσον τα δεδομένα μπορούν να κατηγοριοποιηθούν ως δομημένα, μπορεί να γίνει χρήση τους για ανάλυση δεδομένων με μεθόδους όπως αυτούς που παρουσιάζονται στον πίνακα 3.

Πίνακας 3: Τύποι δομημένης ανάλυσης δεδομένων

Μέθοδοι δομημένης ανάλυσης δεδομένων	
Αλγεβρική προσέγγιση	Εδώ υπάγονται οι μέθοδοι όπως η περιγραφική στατιστική, η εξερευνητική ανάλυση και τεχνικές εξόρυξης δεδομένων. Η αλγεβρική προσέγγιση αποτελείται από διαδικασίες επιθεώρησης, καθαρισμού, μετατροπής δεδομένων και ανάπτυξης μοντέλων.
Μπεϋζιανή προσέγγιση	Η προσέγγιση αυτή αφορά στη χρήση γραφημάτων μοντέλων πιθανοτήτων που αντιπροσωπεύουν ένα σύνολο τυχαίων μεταβλητών και τις υπό όρους εξαρτήσεις τους.
Τεχνικές συσταδοποίησης	Οι τεχνικές εξόρυξης δεδομένων και μηχανικής μάθησης που δεν αφορούν εξαρτημένες μεταβλητές είναι τύπου μη καθοδηγούμενης συσταδοποίησης. Ο κύριος στόχος αυτής της προσέγγισης είναι η ανακάλυψη εννοιολογικών μοτίβων και ομάδων εννοιών.
Συνδυαστική προσέγγιση	Αφορά στη συνδυαστική διάταξη αντικειμένων για τα οποία μπορούν να ανακτηθούν σχετικά δεδομένα που να βοηθήσουν στη βέλτιστη αντιπροσώπευση ενός συνόλου δεδομένων.
Τυπική εννοιολογική προσέγγιση	Η τυπική εννοιολογική προσέγγιση (<i>Formal Concept Analysis</i>) είναι μαθηματική θεωρία ανάλυσης δεδομένων με χρήση τυπικών πλαισίων και εννοιολογικών πλεγμάτων.
Λειτουργική ανάλυση	Κλάδος της στατιστικής που κάνει ανάλυση δεδομένων και επιφέρει πληροφορίες σχετικά με την κυρτότητα, επιφάνειες (<i>area charts</i>) ή οτιδήποτε προκαλεί διακυμάνσεις στη συνέχεια των μεταβλητών.
Γεωμετρική ανάλυση	Αναφέρεται στις γεωμετρικές πτυχές της ανάλυσης εικόνων, μοτίβων και σχημάτων.
Παλινδρόμηση	Τεχνική καθοδηγούμενης εκμάθησης. Χρήση γραμμικής εξίσωσης που συσχετίζει μία ή περισσότερες μεταβλητές για τη γενίκευση ενός συνόλου δεδομένων. Η έξοδος της παλινδρόμησης καθορίζεται από το γραμμικό άθροισμα των σταθμισμένων τιμών εισόδου.

Σχηματική προσέγγιση	Η ανάλυση των σχηματικών ιδιοτήτων συγκεκριμένων συνόλων σχημάτων με χρήση στατιστικών μεθόδων.
Διάταξη τοπολογίας	Η προσέγγιση ανάλυσης συνόλων δεδομένων με χρήση τεχνικών τοπολογίας.
Δεντρικές δομές	Δημοφιλής δομή καθοδηγούμενης εκμάθησης. Τα δένδρα αποφάσεων κατασκευάζονται χρησιμοποιώντας μόνο εκείνα τα χαρακτηριστικά τα οποία μπορούν να διαφοροποιήσουν τις έννοιες που απορρέουν από ένα δείγμα δεδομένων. Με τη χρήση αλγορίθμων, παράγονται κανόνες που κατηγοριοποιούν τα δεδομένα και τα εντάσσουν σε δεντρικές δομές.

(Wikipedia: Structured data analysis (statistics), 2017), (Wikipedia: Data analysis, 2017), (Wikipedia: Bayesian network, 2017), (Murphy, 1998), (Richard J. Roiger, 2008), (Wikipedia: Geometric data analysis, 2017), (Priss, 2017), (Wikipedia: Functional data analysis, 2017), (Wikipedia: Statistical shape analysis, 2017), (Wikipedia: Topological data analysis, 2017)

4 Μέθοδοι πρόβλεψης

Η πρόβλεψη παίζει πολύ σημαντικό ρόλο στον επιχειρηματικό κόσμο. Πλέον οι περισσότεροι οργανισμοί του δυτικού κόσμου χρησιμοποιούν την ανάλυση δεδομένων για ανάπτυξη προβλέψεων ώστε να αποκτήσουν ενόραση και καθοδήγηση σχετικά με τις επιχειρηματικές τους αποφάσεις. Γενικά, υπερισχύει η άποψη ότι γεγονότα του παρελθόντος θα επαναληφθούν στο μέλλον, οπότε η χρήση προβλέψεων ενισχύει τον προγραμματισμό των οργανισμών πάνω σε διάφορες επιχειρηματικές διεργασίες ρουτίνας όπως τη διαχείριση προμηθειών, παραγωγής, πωλήσεων, οικονομικών, αλλά και σε καινοτομικές διεργασίες όπως το λανσάρισμα νέων προϊόντων, κλπ. (Douglas A. Lind, 2015)

Υπάρχουν διάφοροι τύποι προβλέψεων και πολλές διαφορετικές μέθοδοι για τη διεξαγωγή τους. Στα προηγούμενα κεφάλαια έγινε η περιγραφή των βασικών αρχών και πλαισίων ανάπτυξης μιας πρόβλεψης. Ένα από τα βασικά βήματα είναι η προετοιμασία των δεδομένων· οι σχετικές διαδικασίες περιγράφηκαν στα παραπάνω κεφάλαια. Εφόσον έχει πραγματοποιηθεί ανάκτηση των δεδομένων, η επεξεργασία, η περίληψη και η κατανόηση τους, θεωρητικά έχουν χτιστεί τα θεμέλια για την ανάπτυξη μιας πρόβλεψης. Σε αυτό το κεφάλαιο παρουσιάζονται τέσσερις στατιστικές μέθοδοι προβλέψεων· οι μέθοδοι των κινούμενων μέσων και της εκθετικής εξομάλυνσης για πρόβλεψη τάσεων, η μέθοδος της γραμμικής παλινδρόμησης και τέλος, η χρήση στοχαστικών μοντέλων ARIMA για ανάλυση χρονολογικών σειρών.

4.1 Κινούμενοι μέσοι

Η πρώτη μέθοδος πρόβλεψης που παρουσιάζεται είναι η μέθοδος των *κινούμενων μέσων*. Η συγκεκριμένη μέθοδος μπορεί να φανεί ιδιαίτερα χρήσιμη διότι είναι εύκολα εφαρμόσιμη. Ο κύριος σκοπός της συγκεκριμένης μεθόδου είναι ο προκαθορισμός της τάσης. Επιπλέον, εκτός από την ανάπτυξη προβλέψεων, οι κινούμενοι μέσοι μπορούν να χρησιμοποιηθούν ως μέσο σύγκρισης ώστε να γίνει μέτρηση των εποχικών αυξομειώσεων. Σε αντίθεση με τη μέθοδο των *ελαχίστων τετραγώνων*, η οποία χρησιμοποιεί τη γραμμική μαθηματική συνάρτηση $\hat{y} = a + \beta x$ για να αναπτύξει την τάση, η μέθοδος των κινούμενων μέσων παρουσιάζει τις υπάρχουσες αυξομειώσεις των δεδομένων με ελαφρώς πιο ομαλή μορφή. Το

παραπάνω προκύπτει καθώς ο αριθμητικός μέσος «κινείται» ανάμεσα τις τιμές μιας χρονολογικής σειράς. Για τη μέθοδο των κινούμενων μέσων ισχύουν τα εξής χαρακτηριστικά: (Douglas A. Lind, 2015)

- i. Για τα βέλτιστα αποτελέσματα της μεθόδου, τα δεδομένα θα πρέπει να διαθέτουν μία κυρίως γραμμική τάση.
- ii. Επίσης, στην κατάλληλη προβολή της τάσης βοηθάει η ύπαρξη συγκεκριμένων μοτίβων που να εμφανίζουν παρόμοιες αυξομειώσεις κατά τη διάρκεια των προκείμενων χρονολογικών κύκλων.
- iii. Ο αριθμός των τιμών των παρατηρήσεων που ενδείκνυται να προστεθούν για τη μέθοδο πρόβλεψης κινούμενων μέσων – ή και των περισσότερων άλλων μεθόδων, εξαρτάται από τη συχνότητα καταγραφής των παρατηρήσεων. Για μηνιαία δεδομένα πολλών χρόνων, τυπικά θα χρησιμοποιηθούν 12 τιμές για κάθε χρόνο, ο οποίος μάλλον παράλληλα αποτελεί έναν χρονολογικό κύκλο.
- iv. Αν η διάρκεια των κύκλων είναι σταθερή με ίσα μήκη, τότε το κυκλικό και απροσδιόριστο χαρακτηριστικό της χρονολογικής σειράς εξαλείφεται εντελώς με τη μέθοδο των κινούμενων μέσων. Σε αυτήν την περίπτωση, το αποτέλεσμα είναι μια ευθεία γραμμή.
- v. Δεδομένα πωλήσεων, παραγωγής και άλλων οικονομικών και επιχειρηματικών σειρών, συνήθως δεν εμφανίζουν αυξομειώσεις που να ανήκουν σε κύκλους με ίσα, σταθερά μήκη. Συνεπώς, η εφαρμογή των κινούμενων μέσων σε αυτές τις χρονολογικές σειρές δε θα αποτελέσει σε μία ευθεία γραμμή, αλλά σε μία τάση με καμπύλες.

Υπάρχει περίπτωση κάποια σύνολα δεδομένων να εμφανίσουν τόσο σποραδικές τιμές, που να μην είναι δυνατή η προβολή μιας τάσης. Σε αυτήν την περίπτωση δεν υπάρχει νόημα να γίνει εφαρμογή της τάσης στα δεδομένα ώστε να προκύψει μια πρόβλεψη. Αντίθετα, είναι δυνατή η χρήση των κινούμενων μέσων ώστε να γίνει παραγωγή μιας τάσης, η οποία θα είναι και η πρόβλεψη. Υπάρχουν δύο βασικοί τύποι κινούμενων μέσων: *ακόλουθοι κινούμενοι μέσοι (trailing moving average – TMA)* και *κεντρικοί κινούμενοι μέσοι (centred moving average – CMA)*. Επιπλέον, από τους παραπάνω μέσους μπορεί να αναπτυχθεί και ο *σταθμισμένος κινούμενος μέσος (weighted moving average – WMA)*, ο οποίος χρησιμοποιείται για να δώσει κάποια βαρύτητα σε συγκεκριμένες τιμές.

4.1.1 Ακόλουθος κινούμενος μέσος

Ο *ακόλουθος κινούμενος μέσος* είναι ίσως ο απλούστερος τύπος κινούμενων μέσων. Ο υπολογισμός του βασίζεται στις περασμένες m παρατηρήσεις. Η εφαρμογή του εξομαλύνει τις περασμένες αυξομειώσεις στη χρονολογική σειρά, καθιστώντας τον πιο εύκολο προσδιορισμό των μοτίβων της τάσης. Η επιλογή των περασμένων περιόδων με τις οποίες γίνεται ο υπολογισμός είναι ανάλογη με την κάθε περίπτωση. Μεγαλύτερα m οδηγούν σε πιο ομαλό κινούμενο μέσο αλλά απαιτούν περισσότερα δεδομένα. Η τιμή \hat{y}_t μπορεί επίσης να χρησιμοποιηθεί ως πρόβλεψη για την $t + 1$. Εκτός από το εύρος των τιμών y_1, y_2, \dots, y_n δεν υπάρχει τρόπος να ενημερωθεί ο κινούμενος μέσος, οπότε θα ήταν κατάλληλο η μέθοδος αυτή να θεωρηθεί ως πρόβλεψη της επόμενης περιόδου. (David P. Doane, 2016)

$$\hat{y}_t = \frac{y_t + y_{t-1} + \dots + y_{t-m+1}}{m}$$

4.1.2 Κεντρικός κινούμενος μέσος

Εναλλακτικός τύπος κινούμενων μέσων είναι ο κεντρικός. Σε αντίθεση με τον ακόλουθο, ο *κεντρικός κινούμενος μέσος* χρησιμοποιεί περασμένες αλλά και μελλοντικές παρατηρήσεις κατά τον υπολογισμό του, το συνολικό πλήθος των οποίων είναι m . Η τιμή \hat{y}_t μπορεί να θεωρηθεί ως ο μέσος όρος της παρούσας παρατήρησης με τις γειτονικές της παρατηρήσεις. Η συγκεκριμένη μέθοδος δεν αποτελεί πρόβλεψη, αλλά εξομάλυνση του δείγματος. Ας σημειωθεί ότι όταν ο m είναι περιττός αριθμός, ο κεντρικός κινούμενος μέσος υπολογίζεται εύκολα. Όταν ο m είναι άρτιος, η συνάρτηση υπολογισμού της μεθόδου γίνεται περισσότερο περίπλοκη, εφόσον ο μέσος όρος αποτυπώνεται ανάμεσα σε δύο παρατηρήσεις και έτσι ο κεντρικός κινούμενος μέσος δε θα ήταν σωστά κεντραρισμένος. Στην περίπτωση που ο m είναι άρτιος, θα χρειαστεί η μέθοδος να εφαρμοστεί δύο φορές. Εναλλακτικά, όπου $m =$ *περιττός*, ισχύει: (David P. Doane, 2016)

$$\hat{y}_t = \frac{y_{t-1} + y_t + y_{t+1}}{m}$$

4.1.3 Σταθμισμένος κινούμενος μέσος

Όλες οι παρατηρήσεις που συμπεριλαμβάνονται στον υπολογισμό ενός κινούμενου μέσου είναι ισόβαρες. Για παράδειγμα, για να υπολογιστεί ο μέσος όρος εξαμήνου από μηνιαία δεδομένα, το άθροισμα των μηνών θα διαιρεθεί προς την τιμή 6 και όλοι οι μήνες θα έχουν ίσο βάρος προς τον μέσο που προκύπτει. Αυτό μπορεί να φαίνεται δίκαιο, αλλά μπορεί κάποιος να θεωρήσει ότι οι πιο παλιές τιμές δεν έχουν την ίδια επιρροή πάνω στην επόμενη τιμή, όσο έχουν οι αμέσως προηγούμενες της. Η επέκταση του παραπάνω προβλήματος είναι η χρήση βαρύτητας στις τιμές που αθροίζονται κατά τον υπολογισμό του μέσου. Έτσι, εφόσον το άθροισμα των βαρών είναι ίσο με τη μονάδα, μπορεί να γίνει χρήση των μελών της μονάδας για να διαμοιραστεί η βαρύτητα των τιμών που αποτελούν τον μέσο. (Douglas A. Lind, 2015) Συνεπώς, αν όπως στο παραπάνω παράδειγμα ο μέσος αποτελείται από έξι μήνες, αρχικά, ο κάθε μήνας έχει ίση βαρύτητα 16,66%. Αν όμως θεωρηθεί ότι οι δύο πιο πρόσφατοι μήνες θα πρέπει να έχουν το μεγαλύτερο μέρος της βαρύτητας, για παράδειγμα μπορεί να τεθεί βαρύτητα 25% ($y_6 \times 0,25$) στον 6^ο μήνα, 20% στον 5^ο μήνα ($y_5 \times 0,2$) και από 13.75% στους πρώτους τέσσερις ($y_{1,2,3,4} \times 0,1375$), ή οποιοσδήποτε άλλος συνδυασμός θεωρηθεί βέλτιστος για τον υπολογισμό του σταθμισμένου μέσου. Τα βάρη μπορούν να ενσωματωθούν στις συναρτήσεις του ακόλουθου και κεντρικού κινούμενου μέσου.

4.2 Εκθετική εξομάλυνση

Η μέθοδος της εκθετικής εξομάλυνσης είναι ένα ιδιαίτερο είδος σταθμισμένων κινούμενων μέσων. Χρησιμοποιείται σε *συνεχής προβλέψεις της επόμενης περιόδου* για δεδομένα που παρουσιάζουν αυξομειώσεις αλλά μη σταθερή τάση. Η συγκεκριμένη μέθοδος είναι ιδιαίτερα κατάλληλη για χρονολογικές σειρές που φέρουν έντονα το τυχαίο χαρακτηριστικό (*random walk*). Για το τυχαίο ισχύει ότι οι τιμές του αποτελούν απροσδιόριστο μέσο ή διακύμανση ενώ παρουσιάζουν ισχυρή χρονική εξάρτηση. Η μέθοδος της εκθετικής εξομάλυνσης δημιουργεί απλά μοντέλα πρόβλεψης που απαιτούν μόνο δύο εισόδους και μία σταθερά. Για τη μέθοδο αυτή ισχύει ο παρακάτω τύπος: (David P. Doane, 2016)

$$F_{t+1} = c \times y_t + (1 - c)F_t$$

Όπου:

- F_{t+1} = η πρόβλεψη για την αμέσως επόμενη παρατήρηση και ο σταθμισμένος μέσος της παρατήρησης y_t και F_t .
- y_t = η πραγματική παρατήρηση για την περίοδο t
- F_t = η προηγούμενη πρόβλεψη για την περίοδο t
- c = η σταθερά εξομάλυνσης. Όσο μεγαλύτερη είναι, τόσο γρηγορότερα οι προβλέψεις προσαρμόζονται στα πιο πρόσφατα δεδομένα. Συγκεκριμένα:
 - Αν $c = 0,05$ τότε $F_{t+1} = 0,05y_t + 0,95F_t$ (έντονη εξομάλυνση, αργή προσαρμογή)
 - Αν $c = 0,2$ τότε $F_{t+1} = 0,2y_t + 0,8F_t$ (μέτρια εξομάλυνση, μέτρια προσαρμογή)
 - Αν $c = 0,5$ τότε $F_{t+1} = 0,5y_t + 0,5F_t$ (ελαφριά εξομάλυνση, γρήγορη προσαρμογή)
 - Αν $c = 1$ τότε $F_{t+1} = y_t$ (καθόλου εξομάλυνση, πρόβλεψη = περασμένη παρατήρηση)

Με τη μέθοδο της εκθετικής εξομάλυνσης ισχύει ότι οι πιο παλιές παρατηρήσεις επηρεάζουν λιγότερο την πρόβλεψη, σε σύγκριση με τις πιο πρόσφατες. Αυτό συμβαίνει γιατί αν αντικατασταθεί το F_t του τύπου με την προηγούμενη πρόβλεψη F_{t-1} , ο τύπος που χρησιμοποιείται από διάφορα προγράμματα λογισμικού για την ανάπτυξη προβλέψεων με εκθετική εξομάλυνση είναι ο εξής:

$$F_{t+1} = c \times y_t + c(1 - c)y_{t-1} + c(1 - c)^2y_{t-2} + c(1 - c)^3y_{t-3} + \dots$$

Ο παραπάνω τύπος δε συνεχίζεται απεριόριστα. Θα τερματίσει όταν φτάσει στην πρώτη πρόβλεψη F_1 . Υπάρχουν διάφοροι τρόποι για τον υπολογισμό της F_1 . οι δύο συχνότεροι είναι η χρήση της πιο πρόσφατης παρατήρησης από το δείγμα, ή ο αριθμητικός μέσος μερικών από τις αρχικές παρατηρήσεις.

4.3 Γραμμική παλινδρόμηση

Μία τεχνική ανάπτυξης πρόβλεψης που επίσης υπάγεται στις τεχνικές εξόρυξης δεδομένων είναι η *στατιστική παλινδρόμηση* (*statistical regression*). Ως τεχνική καθοδηγούμενης

εκμάθησης, η στατιστική παλινδρόμηση πραγματοποιεί προβλέψεις χρησιμοποιώντας μία μαθηματική εξίσωση η οποία συσχετίζει μία ή περισσότερες εισόδους με μία έξοδο. Ένα μοντέλο γραμμικής παλινδρόμησης (*linear regression*) παράγει έξοδο που αποτελείται από το γραμμικό άθροισμα των σταθμισμένων τιμών εισόδου. (Richard J. Roiger, 2008)

Πολλές επιχειρηματικές αποφάσεις βασίζονται στη συσχέτιση δύο ή περισσότερων μεταβλητών. Για παράδειγμα, στον τομέα της διαφήμισης, θεωρείται βέβαιο ότι υπάρχει εξάρτηση μεταξύ των πωλήσεων μιας επιχείρησης και της επένδυσης της σε έξοδα διαφήμισης. Αν αυτό δεν ίσχυε, τότε δεν υπάρχει λόγος να πραγματοποιηθεί διαφήμιση. Πολλές φορές τέτοιου είδους αποφάσεις λαμβάνονται με χρήση ποιοτικών τεχνικών προβλέψεων και βασίζονται στην έμπειρη γνώμη κάποιων ειδικών. Ωστόσο, αν υπάρχουν διαθέσιμα δεδομένα, η γραμμική παλινδρόμηση είναι η καταλληλότερη τεχνική για παραγωγή πρόβλεψης που βασίζεται στη συσχέτιση μεταξύ των μεταβλητών. Η μεταβλητή για την οποία γίνεται η πρόβλεψη ονομάζεται *εξαρτημένη μεταβλητή (Y)*. Η μεταβλητή που χρησιμοποιείται ως προφήτης για την εξαρτημένη μεταβλητή ονομάζεται *ανεξάρτητη (X)*. Για το παραπάνω παράδειγμα, η ανεξάρτητη μεταβλητή θα ήταν τα έξοδα διαφήμισης, ενώ η εξαρτημένη θα ήταν οι πωλήσεις. (David R. Anderson, 2014)

Η γραμμική παλινδρόμηση χωρίζεται στις εξής δύο κατηγορίες, οι οποίες αφορούν τις μεταβλητές εισόδου του μοντέλου: *απλή γραμμική παλινδρόμηση* και *πολλαπλή γραμμική παλινδρόμηση*.

4.3.1 Απλή και πολλαπλή γραμμική παλινδρόμηση

Η παλινδρόμηση μπορεί να χρησιμοποιηθεί για ανάπτυξη πρόβλεψης και ως έρευνα των αιτιών που οδηγούν στα αποτελέσματα της πρόβλεψης. Η εξίσωση της παλινδρόμησης εκφράζει τη γραμμική σχέση μεταξύ των μεταβλητών. Ο σκοπός της είναι η τοποθέτηση μιας γραμμής η οποία να αντιπροσωπεύει κατάλληλα τη σχέση των μεταβλητών, δηλαδή πως η ανεξάρτητη μεταβλητή επηρεάζει την εξαρτημένη. Η απλή γραμμική παλινδρόμηση είναι διαφορετική από την πολλαπλή μονάχα από τον αριθμό των ανεξάρτητων μεταβλητών που περιλαμβάνει. Για την απλή γραμμική παλινδρόμηση υπάρχει μόνο μία ανεξάρτητη μεταβλητή, ενώ για την πολλαπλή υπάρχουν δύο ή περισσότερες ανεξάρτητες μεταβλητές. Και για τις δύο περιπτώσεις η *έξοδος είναι μία* και αφορά στην εξαρτημένη μεταβλητή. Πολλές φορές βέβαια οι δύο παραπάνω τεχνικές δε διαχωρίζονται και θεωρούνται ως μία, η

οποία μπορεί γενικά να ονομασθεί ως γραμμική παλινδρόμηση, ως πολλαπλή γραμμική παλινδρόμηση ή και πιο σύνθετα ως κοινή μέθοδος ελαχίστων τετραγώνων πολλαπλής γραμμικής παλινδρόμησης.

Οι συντελεστές της γραμμικής παλινδρόμησης συνήθως ορίζονται από τη μέθοδο των ελαχίστων τετραγώνων, η οποία είναι ο δημοφιλέστερος τύπος παλινδρόμησης, ωστόσο υπάρχουν και άλλες, πιο περίπλοκες μέθοδοι. Η γραμμική σχέση μεταξύ δύο μεταβλητών περιγράφεται από δύο συντελεστές: την κλίση (*slope*) και την τομή (*intercept*). Οι παραπάνω συντελεστές ορίζονται με σκοπό την ελαχιστοποίηση του αθροίσματος των τετραγωνισμένων σφαλμάτων της πρόβλεψης, δηλαδή το $y - \hat{y}$. Οι συντελεστές παλινδρόμησης αξιολογούνται με χρήση διαστημάτων εμπιστοσύνης και ελέγχους υποθέσεων. Το κριτήριο που χρησιμοποιείται συχνότερα ως μέτρο αξιολόγησης των μεταβλητών και την ικανότητα της ανεξάρτητης μεταβλητής να παράγει πρόβλεψη για την εξαρτημένη, είναι το R^2 , ή αλλιώς ο συντελεστής προσδιορισμού. (Allison, 1999)

Για μοντέλα που περιέχουν μία εξαρτημένη μεταβλητή και μία ανεξάρτητη, ισχύει ο παρακάτω τύπος:

$$\hat{y} = a + \beta x$$

Όπου a ορίζεται ως η τομή και β η κλίση. Με βάση τον παραπάνω τύπο, ισχύει ότι η τιμή της τομής είναι η τιμή που θα έχει το \hat{y} όταν $x = 0$. Επιπλέον, εφόσον η παραπάνω συνάρτηση περιγράφει μια ευθεία γραμμή, ισχύει ότι όταν $\beta = 0$, η γραμμή θα είναι απολύτως οριζόντια. Ουσιαστικά, η κλίση είναι η τιμή που σταθμίζει την επιρροή της ανεξάρτητης μεταβλητής x πάνω στο \hat{y} . Για μοντέλα που περιέχουν n ανεξάρτητες μεταβλητές, η συνάρτηση της παλινδρόμησης είναι ως εξής:

$$\hat{y} = a + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

4.3.2 Στατιστική σημασία και μηδενική υπόθεση

Ο έλεγχος της εξάρτησης των μεταβλητών είναι το πρώτο βήμα που πρέπει να γίνει πριν αναπτυχθεί μια πρόβλεψη παλινδρόμησης. Στο προηγούμενο κεφάλαιο έγινε περιγραφή του συντελεστή συσχέτισης του *Pearson*. Εφόσον τα δεδομένα εμφανίζουν γραμμικότητα, ο

συντελεστής r μπορεί να χρησιμοποιηθεί για τον υπολογισμό της εξάρτησης των μεταβλητών.

Τα αποτελέσματα του r μπορούν να εξεταστούν όσο αναφορά την εγκυρότητα τους και τη στατιστική τους σημασία, με χρήση της μηδενικής υπόθεσης (*null hypothesis*). Η υπόθεση αυτή υποστηρίζει ότι υπάρχει μηδενική εξάρτηση ανάμεσα στις δύο (ή περισσότερες) μεταβλητές που εξετάζονται. Η εξέταση της υπόθεσης έχει σκοπό την απόρριψη της. Αν τελικά αποδειχτεί ότι η μηδενική υπόθεση είναι σωστή, αυτό σημαίνει ότι όντως δεν υπάρχει εξάρτηση ανάμεσα στις μεταβλητές και δε συνίσταται η χρήση της γραμμικής παλινδρόμησης, τουλάχιστον όχι όσο αναφορά τις συγκεκριμένες ανεξάρτητες μεταβλητές.

- $H_0: \rho = 0$ (Δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών)
- $H_1: \rho \neq 0$ (Η συσχέτιση μεταξύ των μεταβλητών δεν είναι μηδενική)

Η εγκυρότητα του συντελεστή συσχέτισης και η προσπάθεια απόρριψης της μηδενικής υπόθεσης μπορεί να ελεγχθεί με μεθόδους όπως το *t-test*. (Douglas A. Lind, 2015) Το *t-test* συμπεριλαμβάνει στιγμές ελευθερίας (*degrees of freedom – df*) που χαρακτηρίζουν τη μεταβλητότητα του δείγματος όταν εξετάζεται από κάποια μέθοδο ελέγχου όπως το *t-test*. Οι στιγμές ελευθερίας ορίζονται ως τη διαφορά του πλήθους των παρατηρήσεων του δείγματος μείον τον αριθμό των απαιτούμενων συσχετίσεων ανάμεσα στις παρατηρήσεις (π.χ. τον αριθμό των εκτιμώμενων παραμέτρων). (Runkel, 2016) Στην περίπτωση της μηδενικής υπόθεσης, οι στιγμές ελευθερίας ορίζονται ως $n - 2$.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Οι τιμές του t θα ανήκουν σε ένα εύρος τιμών, από το οποίο το 5% των ακραίων τιμών θα είναι το υποσύνολο που θα απορρίπτει τη μηδενική υπόθεση. Αν η τιμή του t ανήκει στο 5% του εύρους τότε θεωρείται ότι ανήκει στην περιοχή απόρριψης και ότι ισχύει η H_1 , άρα η μηδενική υπόθεση απορρίπτεται και ο συντελεστής συσχέτισης r έχει στατιστική σημασία. Τα όρια του εύρους των τιμών που μπορεί να πάρει το t υπάρχουν διαθέσιμα στον πίνακα κατανομής του *Student's t*.

4.3.3 Τυπικό σφάλμα εκτίμησης και συντελεστής προσδιορισμού

Εφόσον πραγματοποιηθεί η επιλογή των κατάλληλων μεταβλητών και αναπτυχθεί ένα μοντέλο παλινδρόμησης, το μοντέλο αυτό θα πρέπει να αξιολογηθεί ως προς την ικανότητα του να παράγει έγκυρη πρόβλεψη. Σε προηγούμενο κεφάλαιο παρουσιάστηκαν μέτρα αξιολόγησης ακρίβειας μοντέλων πρόβλεψης, ως προς τα σφάλματα τους. Για μοντέλα παλινδρόμησης υπάρχουν διάφορα μέτρα διαθέσιμα τα οποία ελέγχουν την ικανότητα των *συντελεστών* της παλινδρόμησης να παράγουν κάποια πρόβλεψη με σημαντική ακρίβεια. Τα πιο κοινά μέτρα είναι το *τυπικό σφάλμα εκτίμησης* και ο *συντελεστής προσδιορισμού*.

Το τυπικό σφάλμα εκτίμησης $S_{\hat{y}\cdot x}$ είναι ένα μέτρο προσδιορισμού της διασποράς των παρατηρήσεων γύρω από τη γραμμή της παλινδρόμησης για μία συγκεκριμένη τιμή του X . Περιγράφει πόσο ακριβής είναι η πρόβλεψη του \hat{Y} με βάση το X , ή αντίστροφα, πόσο ανακριβής μπορεί να είναι οι εκτιμήσεις. Παρόμοια με την τυπική απόκλιση, η οποία μετράει την απόκλιση των τιμών από τον αριθμητικό μέσο, το τυπικό σφάλμα εκτίμησης μετράει την απόκλιση των σφαλμάτων από την ευθεία της παλινδρόμησης. Ο υπολογισμός του $S_{\hat{y}\cdot x}$ απαιτεί το άθροισμα των τετραγωνισμένων σφαλμάτων, δηλαδή την τετραγωνισμένη διαφορά της πραγματικής παρατήρησης Y μείον την εκτιμωμένη τιμή της, \hat{Y} . Συγκεκριμένα, το τυπικό σφάλμα εκτίμησης ορίζεται ως εξής: (Douglas A. Lind, 2015)

$$S_{\hat{y}\cdot x} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 2}}$$

Αν η τιμή του τυπικού σφάλματος εκτίμησης είναι μικρή, υποδεικνύεται ότι οι παρατηρήσεις κυμαίνονται σχετικά κοντά στη γραμμή της παλινδρόμησης, άρα οι προκείμενοι συντελεστές μπορούν να χρησιμοποιηθούν ως κατάλληλη πρόβλεψη για την εξαρτημένη μεταβλητή, με χαμηλή απόκλιση. Αντιθέτως, αν η τιμή του τυπικού σφάλματος εκτίμησης είναι μεγάλη, υποστηρίζεται ότι οι παρατηρήσεις είναι πολύ ευρέως διασκορπισμένες και το συγκεκριμένο μοντέλο παλινδρόμησης δε θα αποτελέσει αρκετά ακριβής πρόβλεψη.

Το μειονέκτημα του τυπικού σφάλματος εκτίμησης είναι ότι το $S_{\hat{y}\cdot x}$ μπορεί να πάρει οποιαδήποτε τιμή, οπότε η ερμηνεία του είναι σχετική. Αντίθετα, ο συντελεστής προσδιορισμού R^2 φέρει ποσοστιαίο αποτέλεσμα, άρα η ερμηνεία του είναι ευκολότερα κατανοητή. Ο συντελεστής προσδιορισμού αντιπροσωπεύει το ποσοστό της συνολικής

διακύμανσης των παρατηρήσεων της εξαρτημένης μεταβλητής Y , η οποία υπολογίζεται από τη διακύμανση που εμφανίζεται εντός των παρατηρήσεων της ανεξάρτητης μεταβλητής X . Εκτός από την ευκολία της ερμηνείας του, ο συντελεστής προσδιορισμού προτιμάται ως μέτρο αξιολόγησης ακρίβειας των συντελεστών του μοντέλου πρόβλεψης διότι είναι εύκολα υπολογίσιμος· ο συντελεστής προσδιορισμού ορίζεται ως το τετράγωνο του συντελεστή συσχέτισης. (Douglas A. Lind, 2015)

$$R^2 = r^2$$

Εφόσον ο συντελεστής συσχέτισης από τον ορισμό του δεν μπορεί να πάρει τιμή μεγαλύτερη της μονάδας, ισχύει ότι ο συντελεστής προσδιορισμού θα έχει απαραίτητα τιμή *θετική* και *εντός* του διαστήματος 0 έως 1 . Συνεπώς, το αποτέλεσμα του συντελεστή προσδιορισμού μπορεί να ερμηνευτεί ως «επί τις εκατό». Συγκεκριμένα, η τιμή του R^2 εξηγεί ποιο είναι το ποσοστό των παρατηρήσεων του Y , για το οποίο *ευθύνεται* η μεταβλητή X .

4.3.4 Προϋποθέσεις

Παραπάνω έγινε περιγραφή των ελέγχων της στατιστικής σημασίας των μεταβλητών και της προκειμένης πρόβλεψης με τη μέθοδο της γραμμικής παλινδρόμησης. Για την εγκυρότητα όλων των παραπάνω ελέγχων ισχύουν μερικές προϋποθέσεις. Αν τα παρακάτω χαρακτηριστικά δεν είναι αληθή για τις μεταβλητές και τα σφάλματα της παλινδρόμησης, τα αποτελέσματα μπορεί να είναι μεροληπτικά και αποπλανητικά. (Douglas A. Lind, 2015)

- i. **Γραμμικότητα:** η εξαρτημένη μεταβλητή έχει γραμμική σχέση με την ανεξάρτητη. Για κάθε τιμή του X , προκύπτει μία τιμή Y , η οποία ακολουθεί την κανονική κατανομή.
- ii. **Η διακύμανση των σφαλμάτων είναι η ίδια για μικρές και μεγάλες τιμές του \hat{y} .** Δηλαδή, η τιμή του κάθε $y - \hat{y}$ είναι ανεξάρτητη από πόσο μεγάλο ή μικρό είναι το \hat{y} .
- iii. **Κανονικότητα και ανεξαρτησία των σφαλμάτων.** Οι μέσοι των κανονικών κατανομών που προκύπτουν από κάθε τιμή Y , συμπίπτουν με τη γραμμή παλινδρόμησης. Για τις κατανομές αυτές ισχύει ότι μοιράζονται τις ίδιες τυπικές αποκλίσεις. Παρομοίως, τα

σφάλματα θα πρέπει να ακολουθούν την κανονική κατανομή και ο μέσος τους να είναι ίσος με το μηδέν.

- iv. Μη συσχέτιση των ανεξάρτητων μεταβλητών. Οι τιμές Y είναι στατιστικά ανεξάρτητες. Αυτό σημαίνει ότι κατά την επιλογή ενός δείγματος, μία τιμή X δεν εξαρτάται από κάποια άλλη τιμή X . Αυτή η υπόθεση είναι ιδιαίτερα σημαντική για χρονολογικά δεδομένα. Σε αυτές τις περιπτώσεις, τα σφάλματα μιας συγκεκριμένης χρονολογικής περιόδου συχνά συσχετίζονται με τα σφάλματα άλλων περιόδων.

4.4 Ανάλυση χρονολογικών σειρών

Όλες οι τεχνικές που αφορούν πρόβλεψη χρονολογικών σειρών φέρουν το κοινό χαρακτηριστικό ότι είναι *ενδογενής*. Αυτό σημαίνει ότι για τη διεξαγωγή της πρόβλεψης μελετώνται τα μοτίβα που παρουσιάζονται εντός της ίδιας της χρονολογικής σειράς. Εάν αυτά τα μοτίβα μπορούν να προσδιοριστούν και να αναπαραχθούν στο μέλλον, μπορεί να γίνει η χρήση τους για να αναπτυχθεί μια πρόβλεψη. Τα χαρακτηριστικά – προφήτες για την πρόβλεψη χρονολογικών σειρών δεν απορρέουν από ανεξάρτητες μεταβλητές, όπως στην περίπτωση της γραμμικής παλινδρόμησης, η οποία θεωρείται εξωγενής τεχνική. Αντιθέτως, μελετάται μόνο η εξαρτημένη μεταβλητή στο εσωτερικό της, συγκεκριμένα οι μεταβολές των τιμών της, και πως η ζήτηση εμφανίζεται ανά τις διαχρονικές εκδοχές της μεταβλητής αυτής. Από τις μεταβολές που παρουσιάζονται, οι τεχνικές χρονολογικής ανάλυσης προσδιορίζουν την *τάση* και την *εποχικότητα* και τις διαχωρίζουν από το τυχαίο, απροσδιόριστο χαρακτηριστικό, ή αλλιώς τον *θόρυβο*. Η τάση και η εποχικότητα αντιπροσωπεύει τη ζήτηση και μπορεί να προσδιοριστεί και να αναπαραχθεί *μονάχα* από τη μελέτη της ίδιας της μεταβλητής. Ο θόρυβος θεωρείται απροσδιόριστος και συνήθως επηρεάζεται από εξωγενή χαρακτηριστικά όπως είναι οι αλλαγές τιμών, η διαφήμιση, οι εκπτώσεις, ο ανταγωνισμός, η οικονομική δραστηριότητα κλπ.. Τα εξωγενής χαρακτηριστικά δε μελετώνται κατά την ανάλυση μιας χρονολογικής σειράς, εφόσον για τις συγκεκριμένες τεχνικές σημασία έχει μόνο ο *χρόνος*. (Mentzer, 2005)

Η μέθοδος των κινούμενων μέσων και της εκθετικής εξομάλυνσης για τις οποίες έγινε περιγραφή παραπάνω, μπορούν να θεωρηθούν ως τεχνικές χρονολογικής ανάλυσης. Οι συγκεκριμένες τεχνικές προβλέπουν την τάση μιας μεταβλητής χωρίς να ελέγχουν τους

εξωγενείς παράγοντες. Το μειονέκτημα της χρήσης των κινούμενων μέσων ως μέθοδο πρόβλεψης είναι ότι τα αποτελέσματα τους δεν είναι αντιπροσωπευτικά της χρονολογικής ανάλυσης, αλλά των αριθμητικών μέσων, όπως αυτοί παρουσιάζονται από τις παρατηρήσεις. Με αυτές τις μεθόδους τα εποχικά φαινόμενα εξομαλύνονται, και έτσι τα συστατικά που αποτελούν μια χρονολογική σειρά μπλέκονται και, κατά συνέπεια, οι παραγόμενες προβλέψεις είναι χρήσιμες αθροιστικά και όχι ανά τις ξεχωριστές χρονολογικές εκδοχές τους.

Για την ακριβέστερη προβολή όλων των συστατικών μιας χρονολογικής σειράς υπάρχουν πιο σύνθετες μέθοδοι πρόβλεψης. Η ανάπτυξη *μονομεταβλητών, γραμμικών, στοχαστικών μοντέλων* είναι μία από τις μεθόδους χρονολογικής ανάλυσης για ενδογενείς προβλέψεις. Τα μοντέλα των διαδικασιών της μεθόδου αυτής υποστηρίζονται από το όνομα τους: η ανάλυση προκύπτει από *μία* μεταβλητή, βασίζονται στη γραμμική σχέση της μεταβλητής αυτής με τις χρονολογικές εκδοχές της και φέρουν το στοχαστικό χαρακτηριστικό. *Στοχαστικότητα* είναι η δυνατότητα ανάπτυξης ενός μοντέλου που να περιγράφει κάποιο φαινόμενο με χρήση τυχαίων ενδεχομένων. Με απλά λόγια, τα μοντέλα αυτά αποτελούν προβλέψεις βασισμένες σε στοιχεία σχετικά με την τάση και την εποχικότητα μιας χρονολογικής σειράς, για την οποία επίσης προβλέπεται η παρουσία των τυχαίων ενδεχομένων. Μία ιδιαίτερα δημοφιλής μέθοδος για στοχαστικές διαδικασίες χρονολογικής ανάλυσης είναι αυτή των *αυτοπαλινδρομών ενσωματωμένων κινούμενων μέσων* (*AutoRegressive, Integrated, Moving Averages – ARIMA*), η οποία δημοσιεύτηκε από τους Box και Jenkins, το 1976.

4.4.1 Η μείξη της αυτοπαλινδρόμησης με τους κινούμενους μέσους (ARMA)

Οι μέθοδοι της αυτοπαλινδρόμησης και των κινούμενων μέσων αναπτύχθηκαν κατά τη δεύτερη και τρίτη δεκαετία του 20^{ου} αιώνα αντίστοιχα. Μελετήθηκαν και χρησιμοποιήθηκαν ξεχωριστά για να δώσουν λύση σε παρόμοια προβλήματα. Το 1938, ο *Wold* επιχείρησε την ένωση των δύο αυτών μεθόδων και ανέπτυξε τις θεωρητικές βάσεις για την μέθοδο ARMA. Τα μοντέλα ARMA (*AutoRegressive Moving Averages*) αποτελούνται από τρία συστατικά και ανάλογα με τη φύση της χρονολογικής σειράς, ένα μοντέλο ARMA μπορεί να εφαρμόσει τις εξής μεθόδους, μεμονωμένα ή συνδυαστικά: (Spyros Makridakis S. C., 1983)

- i. Αυτοπαλινδρομες διαδικασίες (AR)
- ii. Διαδικασίες κινούμενων μέσων (MA)
- iii. Μεικτές διαδικασίες (ARMA)

Το έργο του Wold μελετήθηκε από τους Box και Jenkins για χρόνια. 40 χρόνια μετά από τη δημοσίευση του Wold σχετικά με τα μοντέλα ARMA, οι Box και Jenkins δημοσίευσαν το δικό τους έργο προς εξέλιξη των μοντέλων ARMA ώστε να μπορούν περαιτέρω να καλύψουν τους εξής σκοπούς:

- i. Ανάπτυξη διαδικασιών προσδιορισμού και προβλέψεων με περισσότερη ακρίβεια.
- ii. Επέκταση των αποτελεσμάτων αυτών ώστε να περιλαμβάνουν χρονολογικές σειρές που φέρουν το φαινόμενο της εποχικότητας.
- iii. Την απλή αλλά ιδιαίτερα σημαντική επέκταση των αποτελεσμάτων ώστε να περιλαμβάνουν μεταβλητές (*non-stationary*) διαδικασίες, οι οποίες αποτελούν τη μετατροπή του ARMA σε ARIMA. Τα μοντέλα ARMA δεν αποδίδουν αρκετά καλά για μεταβαλλόμενες χρονολογικές σειρές, δηλαδή αυτές που χαρακτηρίζονται από στοιχεία εποχικότητας και δεν αποτελούν απλώς λευκό θόρυβο. Τα μοντέλα ARIMA ενσωματώνουν τα δεδομένα εκμάθησης επιτρέποντας έτσι την παραγωγή έγκυρων προβλέψεων και για μεταβαλλόμενες χρονολογικές σειρές.

Σε ένα μοντέλο πολλαπλής παλινδρόμησης πραγματοποιείται πρόβλεψη της μεταβλητής που μελετάται με τη χρήση ενός γραμμικού συνδυασμού των μεταβλητών – προφήτες. Σε ένα αυτοπαλίνδρομο μοντέλο, η πρόβλεψη πραγματοποιείται με βάση τη γραμμική συσχέτιση των προηγούμενων τιμών της ίδιας μεταβλητής που βρίσκονται στο δείγμα εκμάθησης, αποτελώντας έτσι παλινδρόμηση της μεταβλητής με τον εαυτό της. Γι' αυτό ωστόσο, οι αυτοπαλίνδρομες διαδικασίες, καθώς και οι διαδικασίες ARIMA, ονομάζονται επίσης μονομεταβλητές. (Rob J Hyndman, 2017)

Ένα αυτοπαλίνδρομο μοντέλο της τάξης p ορίζεται ως εξής:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + e_t$$

Όπου η c δηλώνεται ως μια σταθερά και η e_t ως λευκός θόρυβος. Αλλάζοντας τις παραμέτρους $\varphi_1, \dots, \varphi_p$, απορρέουν διαφορετικά χρονολογικά μοτίβα. Οι παράμετροι φ φέρουν κάποιους περιορισμούς. Για μοντέλα πρώτης τάξεως, το φ_1 πρέπει να βρίσκεται εντός του διαστήματος -1 έως 1 . Για μοντέλα δεύτερης τάξεως, το φ_2 πρέπει να βρίσκεται εντός του διαστήματος -1 έως 1 , το άθροισμα των φ_1 και φ_2 να είναι μικρότερο του 1 , καθώς επίσης η διαφορά του φ_1 από το φ_2 να είναι μικρότερο του 1 . Για μοντέλα μεγαλύτερης τάξεως του 2

οι περιορισμοί του φ γίνονται ιδιαίτερα περίπλοκοι, ωστόσο γίνεται ο υπολογισμός αυτών με χρήση κατάλληλου λογισμικού.

Σε στοχαστικά μοντέλα χρονολογικών σειρών, οι διαδικασίες κινούμενων μέσων διαφέρουν από τη μέθοδο εξομάλυνσης – πρόβλεψης των κινούμενων μέσων που αναφέρθηκε στην αρχή του κεφαλαίου. Οι στοχαστικές διαδικασίες περιλαμβάνουν το χαρακτηριστικό της τύχης, και οι κινούμενοι μέσοι που εμφανίζονται στα μοντέλα ARMA και ARIMA εμπεριέχουν τον λευκό θόρυβο e_t , σε πολύ μεγαλύτερη κλίμακα από τα αυτοπαλίνδρομα μοντέλα. Το μοντέλο κινούμενων μέσων της τάξης q ορίζεται ως εξής:

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

Για τον υπολογισμό των παραμέτρων θ των μοντέλων κινούμενων μέσων, ισχύουν περιορισμοί παρόμοιοι με αυτούς των φ των αυτοπαλίνδρομων μοντέλων.

4.4.2 Μεταβαλλόμενες και αμετάβλητες χρονολογικές σειρές

Σημαντική προϋπόθεση πριν την εφαρμογή οποιασδήποτε μεθόδου πρόβλεψης χρονολογικών σειρών είναι η κατανόηση των δεδομένων της. Δηλαδή τι παρουσιάζουν τα δεδομένα, πως σχετίζονται με τον χρόνο και κατά πόσο παίζει ρόλο ο χρόνος στη συμπεριφορά των δεδομένων. Οι χρονολογικές σειρές αποτελούνται από δεδομένα που περιέχουν τη διάσταση του χρόνου. Συνήθως, όταν γίνεται ανάλυση χρονολογικών σειρών, θεωρείται ότι ο χρόνος είναι κρίσιμος και προμηνύει τη συμπεριφορά των δεδομένων. Αυτό όμως δε συμβαίνει πάντα. (Για περισσότερες πληροφορίες σχετικά με τα συστατικά των χρονολογικών σειρών βλ. κεφάλαιο 1.4...)

Κατά κύριο λόγο, η μελέτη χρονολογικών σειρών του πραγματικού κόσμου θα παρουσιάσει κάποιες χαρακτηριστικές ομοιότητες ακόμα και σε διάφορες σειρές από διαφορετικές πηγές. Οι ομοιότητες συνήθως αφορούν τα συστατικά μέρη των χρονολογικών σειρών, και συγκεκριμένα την ικανότητα για τον προσδιορισμό αυτών με μία απλή διαδικασία αποσύνθεσης. Σε μία οποιαδήποτε χρονολογική σειρά, οι πιθανότητες είναι υψηλές ότι θα παρουσιάζεται το φαινόμενο της εποχικότητας διασκορπισμένο κατά τον χρονολογικό κύκλο που τη χαρακτηρίζει. Ακόμα και αν η εποχικότητα δεν μπορεί εύκολα να προσδιοριστεί, μια κοινή χρονολογική σειρά παρουσιάζει τάση: ανοδική ή καθοδική. Το φαινόμενο της εποχικότητας και η παρουσία της τάσης στα δεδομένα, ορίζουν μια χρονολογική σειρά ως

μεταβαλλόμενη (*non-stationary*). Οι μεταβολές που παρουσιάζονται εξαρτώνται από τον χρόνο ο οποίος αντιπροσωπεύει τις παρατηρήσεις. (Rob J Hyndman, 2017)

Στη σπάνια περίπτωση που μία χρονολογική σειρά του πραγματικού κόσμου δεν εμφανίσει στοιχεία τάσης ή εποχικότητας τότε θεωρείται ότι οι μεταβολές των παρατηρήσεων της είναι ανεξάρτητες από τον χρόνο που τις αντιπροσωπεύει. Τέτοιες χρονολογικές σειρές χαρακτηρίζονται κυρίως από λευκό θόρυβο και θεωρούνται αμετάβλητες (*stationary*). Οι αμετάβλητες χρονολογικές σειρές είναι πιο εύκολο να ενταχθούν σε μοντέλα πρόβλεψης με μεγάλη επιτυχία. Είναι πιο εύκολο για μαθηματικά μοντέλα να προβάλουν «τυχαίες» προβλέψεις πάνω σε δεδομένα με τυχαίες μεταβολές, από το να συμπεριλάβουν εποχικές μεταβολές και χρονολογικά εξαρτώμενες τάσεις. Ο επιτυχής προσδιορισμός του χρονολογικού κύκλου μιας χρονολογικής σειράς δε σημαίνει απαραίτητα ότι η χρονολογική σειρά είναι μεταβλητή, αν παράλληλα δεν μπορούν να προσδιοριστούν κάποια μοτίβα στο δείγμα.

Γενικά, μία *αμετάβλητη* χρονολογική σειρά δεν παρουσιάζει μακροπρόθεσμα προβλέψιμα μοτίβα. Η έλλειψη των μοτίβων χαρακτηρίζει μεθόδους πρόβλεψης όπως τους κινούμενους μέσους επαρκείς για την παραγωγή μιας πρόβλεψης με αρκετή ακρίβεια. Παρομοίως, τα μοντέλα ARMA είναι αρκετά καλά για παραγωγή προβλέψεων από αμετάβλητες χρονολογικές σειρές. Ωστόσο, όπως έχει αναφερθεί παραπάνω, οι χρονολογικές σειρές που παρατηρούνται από δεδομένα πραγματικού κόσμου, κατά πάσα πιθανότητα είναι μεταβαλλόμενες. Για τον λόγο αυτό ήταν κρίσιμη η εξέλιξη μεθόδων όπως την ARMA, ώστε να ενσωματώνουν μεταβαλλόμενες χρονολογικές σειρές και να παράγουν αποτελέσματα που φέρουν τάση και εποχικότητα. Η ανάγκη αυτή κατέληξε στην ανάπτυξη των μοντέλων ARIMA.

4.4.3 Ενσωμάτωση μεταβαλλόμενων διαδικασιών

Τυπικά, τα μοντέλα ARIMA των Box και Jenkins διαφοροποιούνται από τα ARMA του Wold εξαιτίας της λέξης «integrated», που σημαίνει «ενσωμάτωση». Μία πιο ελεύθερη μετάφραση περιγράφει τη λέξη αυτή ως «κατάργηση των διακρίσεων». Η παραπάνω απόδοση είναι αρκετά αντιπροσωπευτική της διαδικασίας αυτής, εφόσον τα μοντέλα ARIMA επιδέχονται μεταβαλλόμενες χρονολογικές σειρές και τις ενσωματώνουν στις στοχαστικές διαδικασίες για παραγωγή προβλέψεων.

Η ενσωμάτωση αφορά στη μεταμόρφωση της χρονολογικής σειράς πραγματοποιώντας αφαίρεση των αποκλίσεων μεταξύ των παρατηρήσεων. Συγκεκριμένα, η ενσωμάτωση ενός επιπέδου για μία σειρά περιλαμβάνει τις διαφορές της κάθε παρατήρησης n από την αμέσως προηγούμενη $n - 1$. Εάν τα δεδομένα μετά τη μεταμόρφωση αυτή εξακολουθούν να είναι μεταβαλλόμενα, θα επαναληφθεί η διαδικασία ενσωμάτωσης πάνω στα ήδη διαφοροποιημένα δεδομένα, όσες φορές χρειαστεί μέχρι τα δεδομένα να μην εμφανίζουν πλέον μεταβλητότητα. Ουσιαστικά, με αυτό το τρόπο εξομαλύνονται όλα τα χαρακτηριστικά εποχικότητας. Συνεπώς, οι διαδικασίες *AR* και *MA* θα εφαρμοστούν πάνω στο δείγμα εκμάθησης εφόσον έχει μετατραπεί σε αμετάβλητη διαδικασία.

4.4.4 Η διαδικασία ARIMA των Box και Jenkins

Οι Box και Jenkins κατάφεραν να οργανώσουν και να παρουσιάσουν με κατανοητό τρόπο όλες τις σχετικές πληροφορίες που απαιτούνται ώστε κάποιος να κατανοήσει και να μπορέσει να χρησιμοποιήσει τα μονομεταβλητά μοντέλα χρονολογικών σειρών ARIMA. Η μέθοδος προσέγγισης των Box και Jenkins για τη διαδικασία αυτή χωρίζεται στις εξής φάσεις: (Spyros Makridakis S. C., 1983)

- i. Προσδιορισμός
- ii. Υπολογισμός και έλεγχος
- iii. Εφαρμογή

Κατά την πρώτη φάση, καθορίζεται η γενική κατηγορία της χρονολογικής σειράς: είναι μεταβλητή ή αμετάβλητη; Από αυτό εξαρτάται αν θα πραγματοποιηθούν ενσωματώσεις πάνω στο δείγμα εκμάθησης πριν εφαρμοστούν οι διαδικασίες *AR* και *MA*. Στη συνέχεια θα ερευνηθούν και θα οριστούν τα επίπεδα – ή οι τάξεις, των συστατικών μελών της διαδικασίας. Τα μοντέλα ARIMA απαρτίζονται από μεθόδους αυτοπαλινδρόμησης, ενσωμάτωσης και κινούμενων μέσων. Η δομή της εφαρμογής των συστατικών αυτών μελών σημειώνεται ως $ARIMA(p,d,q)$, όπου το p αντιπροσωπεύει την τάξη του *AR* μοντέλου, το d αντιπροσωπεύει την τάξη της ενσωμάτωσης και το q την τάξη του *MA* μοντέλου, αντίστοιχα. Για παράδειγμα, όπως αναφέρθηκε παραπάνω, αν μια χρονολογική σειρά εξακολουθεί να είναι μεταβαλλόμενη μετά από μία ενσωμάτωση, τότε θα πραγματοποιηθεί έλεγχος της απόδοσης ενσωμάτωσης δευτέρου επιπέδου, δηλαδή μία διαδικασία $ARIMA(p, \underline{2}, q)$.

Ένα απλό μοντέλο αυτοπαλινδρόμησης τάξεως $p = 1$ μπορεί να σημειωθεί ως AR(1) ή ARIMA(1,0,0). Παρομοίως, το κοινό μοντέλο κινούμενων μέσων τάξεως $q = 1$ σημειώνεται ως MA(1) ή ARIMA(0,0,1). Ένα μοντέλο αυτοπαλινδρόμησης $p = 1$ που έχει υποστεί ενσωμάτωση μίας τάξεως $d = 1$ σημειώνεται ως ARIMA(1,1,0). Αν στο προηγούμενο μοντέλο υπάγονται και διπλής εφαρμογής κινούμενοι μέσοι τάξεως $q = 2$, τότε το μοντέλο σημειώνεται ως ARIMA(1,1,2).

Κατά την τρίτη και τελευταία φάση πραγματοποιείται η ανεύρεση των παραγόντων p , d και q που δημιουργούν το κατάλληλο μοντέλο πρόβλεψης. Αυτή είναι και η πιο δύσκολη υπόθεση της όλης διαδικασίας ανεύρεσης και εφαρμογής στοχαστικών μοντέλων πρόβλεψης, όπου χρησιμοποιούνται μαθηματικά μοντέλα ώστε να γίνει ο υπολογισμός των παραμέτρων της διαδικασίας και πραγματοποιούνται διαγνωστικοί έλεγχοι ώστε διερευνηθεί η καταλληλότητα του μοντέλου. Αν το μοντέλο κριθεί ακατάλληλο, τότε η διαδικασία ξεκινάει από την αρχή αλλάζοντας τις τάξεις των συστατικών μελών της διαδικασίας. Εφόσον το μοντέλο κριθεί κατάλληλο, μπορεί να χρησιμοποιηθεί για τη διεξαγωγή μιας πρόβλεψης.

4.4.5 Επιλογή παραμέτρων και αξιολόγηση

Αφού οριστούν οι παράγοντες p , i , q ενός μοντέλου πρέπει να γίνει εκτίμηση των παραμέτρων που θα αποτελέσουν τους συντελεστές του μοντέλου, δηλαδή τα c , ϕ_1, \dots, ϕ_p , $\theta_1, \dots, \theta_q$. Μία τεχνική για τη διεξαγωγή των εκτιμήσεων για τις παραπάνω παραμέτρους είναι η εκτίμηση της μέγιστης πιθανότητας, η οποία βρίσκει τις τιμές των παραμέτρων που μεγιστοποιούν την πιθανότητα να λάβουμε δεδομένα παρόμοια με τις ιστορικές παρατηρήσεις. Η τεχνική αυτή ορίζεται ως την ελαχιστοποίηση του: (Rob J Hyndman, 2017)

$$\sum_{t=1}^T e_t^2$$

Επιπλέον, τα μοντέλα ARIMA μπορούν να αξιολογηθούν με βάση την ικανότητα τους να ελαχιστοποιήσουν τα κριτήρια AIC, AICc και BIC.

$$AIC = -2 \log(L) + 2(p + q + k + 1)$$

$$AIC_c = AIC + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2}$$

$$BIC = AIC + (\log(T) - 2)(p + q + k + 1)$$

Όπου L = η μέγιστη πιθανότητα, $k = 1$ εφόσον η σταθερά $c \neq 0$ και $k = 0$ εφόσον $c = 0$.

5 Μελέτη περίπτωσης – ανάλυση

5.1 Επισκόπηση

Η κατανόηση και η γνώση των επιχειρηματικών διαδικασιών πέρα των αρμοδιοτήτων ενός αναλυτή είναι απαραίτητη ώστε να γίνει αντιληπτή η πληροφορία που πηγάζει από οποιαδήποτε ανάλυση των δεδομένων. Συνεπώς, ως μέρος της ανάλυσης και της διαδικασίας πρόβλεψης, συμπεριλαμβάνεται η εκμάθηση των επιχειρηματικών διεργασιών σε διάφορους τομείς, όσο αναφορά τη διαχείριση και τη λειτουργία της επιχείρησης. Παρακάτω περιγράφεται η δομή και το έργο της επιχείρησης που μελετήθηκε και αναλύονται συγκεκριμένες επιχειρησιακές διαδικασίες που σχετίζονται με τα προβλήματα που απαιτούν επίλυση. Επίσης αναφέρονται και επεξηγούνται τα βασικά βήματα για μία διαδικασία πρόβλεψης, τα οποία χρησιμοποιήθηκαν ως πλάνο για τη δομή της διαδικασίας που ακολουθήθηκε.

5.1.1 Η επιχείρηση

Η *Fairy Glam (FG)* είναι μία μικρή επιχείρηση λιανικής πώλησης τύπου *business to consumer (B2C)* που εδρεύει στο Λίνκολν της Αγγλίας. Η αγορά – στόχος της είναι γονείς με μικρά παιδιά και πουλάει σε περισσότερες από 7 διαφορετικές χώρες. Προσφέρει παιχνίδια για μωρά και πιο μεγάλα παιδιά, και παιδικά διακοσμητικά αντικείμενα τα οποία φέρουν το παραμυθένιο χαρακτηριστικό με νεράιδες, πριγκίπισσες και μαγικά ζωάκια. Τα παιχνίδια έχουν την ιδιότητα της διασκέδασης αλλά κάποια έχουν και εκπαιδευτικά στοιχεία, όπως αριθμούς και σχήματα, ώστε μέσα από το παιχνίδι να βοηθήσουν στη νοητική ανάπτυξη του παιδιού. Επιπλέον η FG πουλάει φορέματα και στολές για κορίτσια και αγόρια και καλύπτουν ηλικίες από νεογνά έως 8+ ετών, που επίσης έχουν την παραμυθένια θεματολογία.

Τα προϊόντα Lucy Locket σχεδιάζονται με στοργή και η ποιότητα είναι τόσο σημαντική για την FG όσο η καλαισθησία. Μετά την παραγωγή τους, τα προϊόντα δοκιμάζονται από τους πιο ειλικρινείς και «αυστηρούς» κριτές, τα μικρά παιδιά της ομάδας της FG! Τα προϊόντα επίσης φέρουν την Ευρωπαϊκή πιστοποίηση ασφάλειας *EN71*.

Για τα Ευρωπαϊκά δεδομένα, η FG κατηγοριοποιείται ως μικρή επιχείρηση. Αποτελείται από συνολικά 6 μόνιμους εργαζόμενους, 5 από τους οποίους είναι οικογενειακά μέλη που εκπροσωπούν 2 γενιές και ρυθμίζουν όλες τις επιχειρησιακές διεργασίες μεταξύ τους, όπως τη σχεδίαση προϊόντων, το μάρκετινγκ, την εξυπηρέτηση πελατών και τις πελατειακές σχέσεις, την τεχνολογία πληροφορικής, τα οικονομικά και την ασφάλεια, ενώ η έκτη εργαζόμενη ασχολείται με τα λογιστικά. Κατά περιόδους εργάζονται επίσης και φοιτητές που κάνουν έρευνα ή πρακτική άσκηση πλήρους απασχόλησης, στους οποίους συνήθως ανατίθενται εργασίες για τις οποίες η ίδια η επιχείρηση δε διαθέτει τη γνώση ή τους πόρους να διεκπεραιώσει. Για το κομμάτι του μάρκετινγκ η FG εμπιστεύεται τους φοιτητές για νέες ιδέες και ενθουσιασμό στα κοινωνικά δίκτυα, ενώ έχει απασχολήσει στο παρελθόν ευρωπαϊούς φοιτητές για να ασχοληθούν με την αγορά της χώρας τους στη μητρική τους γλώσσα, με σκοπό την πιο άμεση στόχευση προς τις συγκεκριμένες αγορές.

5.1.2 Ίδρυση

Η FG ιδρύθηκε το 1994 με την επωνυμία «Lucy Locket». Ξεκίνησε ως μία μικρή οικογενειακή επιχείρηση και έτσι λειτουργεί μέχρι σήμερα. Η οικονομική κρίση που κορυφώθηκε το 2008 επηρέασε άσχημα τις πωλήσεις της επιχείρησης και ως συνέπεια, δύο χρόνια μετά, η Lucy Locket έκλεισε και επανιδρύθηκε το 2010 με την επωνυμία «Fairly Glam». Το όνομα «Lucy Locket» χρησιμοποιείται σήμερα ως το brand name των προϊόντων της FG. Ως δεύτερο brand name ισχύει το «Slimy Toad», το οποίο απεικονίζεται κυρίως σε αγορίστικα παιχνίδια και ενδυμασίες, ενώ το «Lucy Locket» απευθύνεται κυρίως σε κορίτσια. Πλέον η FG αποτελείται από τρεις νόμιμες οντότητες που χρησιμοποιούνται για τις διεθνείς πωλήσεις: μία για τη Γερμανία, μία για την υπόλοιπη Ευρώπη εκτός του Ηνωμένου Βασιλείου και η πρωταρχική FG για το Ηνωμένο Βασίλειο και τις λοιπές διεθνείς πωλήσεις – εκτός Ευρώπης.

5.1.3 Η Αγορά

Η FG απευθύνεται κατά κύριο λόγο κατευθείαν στους πελάτες της μέσω της Amazon.co.uk, Amazon.de, Amazon.fr, Amazon.com, Ebay.co.uk, Ebay.de και Ebay.com.au, καθώς και μέσω της δικής της ιστοσελίδας (lucylocket.com). Μέσω των παραπάνω ηλεκτρονικών καταστημάτων η FG απευθύνεται σε πελάτες παγκοσμίως, ωστόσο κυρίως εστιάζει στην

αγορά του Ηνωμένου Βασιλείου, της Γερμανίας, της Γαλλίας, της Ισπανίας, της Ιταλίας, της Αυστραλίας και, πρόσφατα, των Η.Π.Α. Υπολογίζεται ότι μονάχα το 1% των ετήσιων εσόδων της προέρχεται από πωλήσεις εκτός των προαναφερόμενων χωρών. Λόγο του χαμηλού αυτού ποσοστού, η FG δεν επικεντρώνεται προς το παρόν στις λοιπές διεθνείς αγορές, αλλά σχεδιάζει στο σύντομο μέλλον να εισέλθει άμεσα στην αγορά του Καναδά, της Ελβετίας και της Ινδίας. Όσο αναφορά τις διαδικτυακές πωλήσεις υπολογίζεται ότι το 90% πραγματοποιείται μέσω Amazon. Συνεπώς, η FG εστιάζει και παρακολουθεί κυρίως την Amazon και όχι ιδιαίτερα το τι συμβαίνει στο Ebay.

Στο παρελθόν, η FG συνήθιζε να προμηθεύει αποκλειστικά φυσικά καταστήματα τρίτων με τα προϊόντα της, οπότε διακινούσε εμπόριο B2B. Πλέον η παρουσία της Lucy Locket σε φυσικά καταστήματα έχει ελαττωθεί σημαντικά, αφού τα έσοδα από τις πωλήσεις αυτές αποτελούν λιγότερο από το 5%, ετησίως. Μερικά από τα προϊόντα Lucy Locket διανέμονται στα καταστήματα Lego Land και Hamleys, εντός του Ηνωμένου Βασιλείου.

5.1.4 Προμήθεια και αποστολές

Όλα τα προϊόντα της FG σχεδιάζονται στην Αγγλία και κατασκευάζονται στην Κίνα και στην Ινδονησία. Από εκεί, μέρος του αποθέματος αποστέλλεται κατευθείαν στις αποθήκες της Amazon σε διάφορες χώρες και τα υπόλοιπα κομμάτια αποθηκεύονται στους αποθηκευτικούς χώρους που η επιχείρηση επικοινωνιάζει μέσω μιας ιδιωτικής μεταφορικής εταιρίας. Για περίπου το 90% των συνολικών πωλήσεων της FG, οι αποστολές ολοκληρώνονται απευθείας μέσω της Amazon (Fulfilled by Amazon – FBA), αν και για τα περισσότερα προϊόντα δίνεται η δυνατότητα στον πελάτη να επιλέξει να λάβει το προϊόν από τον παρασκευαστή (Fulfilled by Merchant – FBM), δηλαδή από τα αποθέματα της FG στην Αγγλία. Λόγο του Amazon Prime, οι περισσότεροι πελάτες επιλέγουν FBA, όπου η επιλογή αυτή είναι διαθέσιμη, διότι τα έξοδα αποστολής είναι δωρεάν και η παραλαβή είναι γρήγορη. Γενικά, η FG πληρώνει προμήθεια στην Amazon όχι μόνο για κάθε ξεχωριστή παραγγελία που ολοκληρώνεται επιτυχώς, αλλά και για κάθε προϊόν που αποθηκεύεται και αποστέλλεται στους πελάτες μέσω της Amazon. Βέβαια, παρά το αυξημένο κόστος ολοκλήρωσης των παραγγελιών, η διαδικασία αυτή συμφέρει την επιχείρηση, αφού οι πελάτες αγοράζουν πιο εύκολα μέσω FBA και έτσι και αλλιώς το κόστος αποθήκευσης δεν είναι κάτι που μπορεί να αποφευχθεί.

5.1.5 Εποχικότητα

Λόγο της φύσης των προϊόντων της, η FG παρακολουθεί την απότομη άνοδο των πωλήσεων της για διάφορες περιόδους μέσα στον χρόνο, η οποία άνοδος ωστόσο ακολουθείται από μία εξ' ίσου απότομη κάθοδο, καθώς η ζήτηση επανέρχεται στα φυσιολογικά επίπεδα. Γενικά, εκτός από συγκεκριμένες περιστάσεις, για τα περισσότερα προϊόντα η ζήτηση έχει μία φυσιολογικά ελαφρώς ανοδική συμπεριφορά. Υπάρχουν κάποια προϊόντα που είναι εποχικά από τη φύση τους, όπως π.χ. οι μικρές φουσκωτές πισίνες αγοράζονται περισσότερο κατά τους πιο θερμούς μήνες και κάποια διακοσμητικά είδη γραφείου ή γραφικές ύλες αγοράζονται περισσότερο προς το τέλος του καλοκαιριού, πριν ανοίξουν τα σχολεία. Η εποχικότητα αυτού του είδους είναι αξιοσημείωτη, αλλά σε καμία περίπτωση δε συγκρίνεται με το φαινόμενο της τεράστιας ακμής των πωλήσεων που εμφανίζεται σε συγκεκριμένες χρονικές περιόδους, οι οποίες χωρίζονται στις εξής κατηγορίες:

➤ Σταθερές

- Οι 4-6 εβδομάδες πριν τα *Χριστούγεννα* μπορεί να σημειώσουν άνοδο μέχρι και 400% για κάθε εβδομάδα, για τα περισσότερα προσφερόμενα προϊόντα σε όλες τις χώρες που στοχεύει η FG.
- Οι στολές και τα αξεσουάρ ζητούνται πολύ κατά τις εβδομάδες πριν το *Halloween*.
- Το Ηνωμένο Βασίλειο γιορτάζει κάθε χρόνο την *Παγκόσμια Ημέρα Βιβλίου*, όπου τα παιδιά ενθαρρύνονται να πάνε στο σχολείο με την ενδυμασία του ήρωα του αγαπημένου τους βιβλίου.

➤ Μεταβλητές

- Το *Πάσχα* οι γονείς αγοράζουν δώρα στα παιδιά τους, βέβαια σε μικρότερη ποσότητα από τα Χριστούγεννα. Το φαινόμενο αυτό είναι ιδιαίτερα αισθητό στη Γερμανία.
- Η Γερμανία και η Ισπανία έχουν *καρναβάλι* κάθε χρόνο, οπότε στολές και αξεσουάρ πωλούνται πολύ σε αυτές τις περιόδους.

5.1.6 Διαχείριση αποθεμάτων

Η διαδικασία της παραγγελίας των προϊόντων από τα εργοστάσια των προμηθευτών είναι κάποιες φορές χρονοβόρα και μπορεί να χρειαστούν από 6 έως και 13 εβδομάδες για την παραλαβή των αποθεμάτων. Ένας λόγος για την καθυστέρηση αυτή είναι ότι η αποστολή των εμπορευμάτων από τα εργοστάσια της Κίνας στην Ευρώπη διαρκεί ένα μήνα. Εξ' αιτίας της καθυστέρησης αυτής καθώς και της μεγάλης χρονολογικής απόκλισης που υπάρχει από τις 6 μέχρι τις 13 εβδομάδες που θα μπορούσε να διαρκέσει η παραλαβή των εμπορευμάτων, η FG χρειάζεται να μαντεύει για τη χρονολογική τοποθέτηση και τον όγκο των παραγγελιών της. Πρέπει δηλαδή να φροντίσει να έχει παραγγείλει κατά μέσο όρο 10-11 εβδομάδες πριν τα υπάρχοντα αποθέματα εξαντληθούν.

Στην περίπτωση που η FG αποτύχει στην κατάλληλη οργάνωση των αποθεμάτων της και τα αποθέματα εξαντληθούν από τις αποθήκες της Amazon για κάποιο προϊόν, τότε θα αρχίσει να ολοκληρώνει η ίδια όσες αποστολές προκύψουν από ότι αποθέματα υπάρχουν διαθέσιμα στις εταιρικές αποθήκες. Κάτι τέτοιο θα μειώσει τις πωλήσεις για το προϊόν, εφόσον το FBA δε θα είναι διαθέσιμο και λιγότεροι πελάτες θα προβούν σε αγορά. Η μείωση των πωλήσεων θα ρίξει επίσης την κατάταξη της Lucy Locket, δηλαδή τη θέση στην οποία εμφανίζονται τα προϊόντα της Lucy Locket στα αποτελέσματα μιας αναζήτησης στην Amazon, η οποία θα μηδενιστεί εάν το προϊόν εξαντληθεί πλήρως. Εφόσον η προσφορά για το προϊόν υπάρξει ξανά χρειάζεται χρόνος – περίπου μία εβδομάδα, ώστε η κατάταξη της Lucy Locket να επανέλθει στα δεδομένα που υπήρξε πριν την εξάντληση. Οπότε, στην περίπτωση που υπάρξει εξάντληση των αποθεμάτων, οι πωλήσεις θα ελαττωθούν για τουλάχιστον μία εβδομάδα πριν την εξάντληση και μία εβδομάδα μετά την αποκατάσταση των αποθεμάτων για το προϊόν. Συμπεριλαμβανομένης της περιόδου απουσίας των αποθεμάτων, αν η FG χάσει το «παράθυρο» ευκαιρίας για την έγκαιρη τοποθέτηση της παραγγελίας της ή αν η ζήτηση για κάποιο λόγο υπερβεί το προσδοκώμενο, η επιχείρηση θα χάσει περίπου έναν ολόκληρο μήνα κέρδους από το συγκεκριμένο προϊόν.

Τα περισσότερα προϊόντα ανήκουν σε κατηγορίες προϊόντων που υπάρχουν σε διάφορα σχέδια ή μεγέθη, αν και η αγορά συνήθως δείχνει κάποια προτίμηση σε συγκεκριμένα προϊόντα από την κάθε κατηγορία. Στην περίπτωση εξάντλησης των αποθεμάτων κάποιου προϊόντος, η εμπειρία υποστηρίζει ότι αυτόματα θα αυξηθεί η ζήτηση για τα υπόλοιπα προϊόντα που υπάγονται στην ίδια κατηγορία. Τότε, η υπεύθυνη του μάρκετινγκ θα φροντίσει να «παίξει» με τις τιμές των προϊόντων ώστε είτε να προωθήσει τα προϊόντα που θεωρεί ότι

θα μπορέσουν να καλύψουν τη ζήτηση, είτε να καθυστερήσει την εξάντληση αποθεμάτων συγκεκριμένων προϊόντων, επιτυγχάνοντας παράλληλα τη μεγιστοποίηση των κερδών από το κάθε αυξημένο προϊόν.

Αντιλαμβανόμαστε ότι η διαδικασία διαχείρισης των αποθεμάτων είναι πολύπλοκη και πολυσύνθετη. Η διοίκηση θα πρέπει να γνωρίζει απόλυτα το τι συμβαίνει μέσα στην επιχείρηση και όλες τις επιμέρους διεργασίες. Συνοπτικά, για την έγκαιρη και επιτυχής διαχείριση των αποθεμάτων, πρέπει να ληφθούν υπ' όψη τα εξής:

- Χρόνος παραλαβής εμπορευμάτων από τους προμηθευτές.
- Η ποσότητα που θα παραγγελθεί θα πρέπει να είναι επαρκής για τις επόμενες 10-15 εβδομάδες τουλάχιστον, ώστε το νέο εμπόρευμα να μην εξαντληθεί προτού παραληφθεί το επόμενο νέο εμπόρευμα. Επίσης η ποσότητα πρέπει να περιοριστεί για την ελαχιστοποίηση του κόστους αποθήκευσης των προϊόντων μέχρι την πώληση τους.
- Λόγο της εποχικότητας, θα πρέπει ο όγκος της παραγγελίας να είναι συνετός και εμπρόθεσμα πραγματοποιήσιμος. Επίσης θα πρέπει το κεφάλαιο να είναι διαθέσιμο για την αγορά μεγάλου όγκου εμπορεύματος, αρκετές εβδομάδες πριν τη διαθεσιμότητα των προϊόντων προς πώληση.
- Θα πρέπει να υπολογιστεί η βέλτιστη ισορροπία μεταξύ κόστους αποθήκευσης και διάθεσης κεφαλαίου για αγορά εμπορεύματος, για την άμεση απόδοση των κερδών και την απελευθέρωση των χρηματικών πόρων.
- Για τον κατάλληλο προσδιορισμό της ζήτησης, εκτός από την εποχικότητα πρέπει να συμπεριληφθούν και άλλες παράμετροι όπως:
 - Η παρούσα στρατηγική μαρκετινγκ.
 - Η τιμολόγηση.
 - Η εξάντληση εναλλακτικών αποθεμάτων στη δεδομένη κατηγορία προϊόντων.
 - Η εισαγωγή νέου προϊόντος στην αγορά και οι ενέργειες που θα παρθούν για την προώθηση του.

Ακόμα και ο χρόνος που χρειάζεται μέχρι να καθιερωθεί ένα νέο προϊόν στην αγορά θα επηρεάσει τη ροή της ζήτησης προς τα υπόλοιπα προσφερόμενα προϊόντα. Το μαρκετινγκ είναι ένα κομμάτι που δεν μπορεί να προγραμματιστεί μακροπρόθεσμα, αφού οι τάσεις και οι καιροί αλλάζουν πολύ συχνά. Είτε πρόκειται για την καθιέρωση ενός νέου προϊόντος είτε για τη διαφήμιση των παλιότερων, για το κομμάτι του μάρκετινγκ, που συνήθως στην FG

δουλεύεται από μία υπάλληλο και φοιτητές, είναι πολύ σημαντικό ότι ο προγραμματισμός επανιδρύεται τακτικά και ότι οι τάσεις ακολουθούνται. Για την FG, η επαφή με τους πελάτες είναι πολύ σημαντική και προσπαθούν να λαμβάνουν όσες περισσότερες πληροφορίες μπορούν έτσι ώστε να ρυθμίζουν τη στρατηγική τους ανάλογα με το τι ζητείται.

5.2 Στόχοι και προσδιορισμός του προβλήματος

5.2.1 Η γενική προσέγγιση

Για να μπορέσει κάποιος να κάνει οποιαδήποτε ανάλυση, καταρχάς πρέπει να γνωρίζει τι αναλύει. Πρέπει δηλαδή να γνωρίζει ποιος είναι ο στόχος του έργου που πρέπει να διεκπεραιώσει και να έχει μία γενική ιδέα για το αποτέλεσμα στο οποίο θα πρέπει να καταλήξει. Στα πλαίσια της συγκεκριμένης εργασίας θα καταγραφεί η προσπάθεια επίλυσης του προβλήματος διαχείρισης αποθεμάτων της επιχείρησης. Για να υπολογιστεί το σημείο παραγγελίας, δηλαδή ο αριθμός αποθεμάτων που σηματοδοτεί ότι η επόμενη παραγγελία πρέπει να πραγματοποιηθεί άμεσα, είναι απαραίτητη η γνώση περί των εξής δύο στοιχείων: (Vermorel, 2012)

- *Ζήτηση* – Ο αριθμός προϊόντων που θα καταναλωθούν ή αγοραστούν.
- *Χρόνος παραλαβής παραγγελίας* – Η περίοδος από την ημέρα καταβολής της παραγγελίας μέχρι την ημέρα που τα αποθέματα θα είναι διαθέσιμα προς πώληση.

Ωστόσο, τα δύο παραπάνω στοιχεία φέρουν το χαρακτηριστικό της αβεβαιότητας, αφού η ζήτηση μπορεί να ακμάσει και ο χρόνος παραλαβής παραγγελίας μπορεί να επηρεαστεί από αμέτρητους παράγοντες όπως απεργίες εργαζομένων, καιρικά φαινόμενα, ζημιά σε πρωταρχικό μηχάνημα, κλπ. Για να αντιμετωπιστεί το φαινόμενο της αβεβαιότητας σε αυτή την περίπτωση, οι επιχειρήσεις χρησιμοποιούν το «μαξιλαράκι» αποθεμάτων, δηλαδή το κατώτερο όριο ασφαλείας που θα πρέπει να είναι άμεσα διαθέσιμο για να αποφευχθεί μία εξάντληση αποθεμάτων, κατά την περίοδο αναμονής μέχρι την επόμενη παραλαβή παραγγελίας.

Παραπάνω αναφέρθηκε ότι η FG εμφανίζει μία αδυναμία όσο αναφορά τη διαχείριση των αποθεμάτων της, κυρίως εξαιτίας της τεράστιας εποχικότητας που επηρεάζει τη ροή των πωλήσεων και αναγκάζει τον απότομο και άνισο διαμοιρασμό χρηματικών πόρων προς επένδυση στην αγορά εμπορεύματος. Επειδή κατά τις τελευταίες εβδομάδες του έτους, οι

πωλήσεις ανεβαίνουν κατακόρυφα, υπάρχουν πολύ μικρά περιθώρια για λάθη. Αν μια παραγγελία εμπορεύματος αργήσει να καταφτάσει στις αποθήκες της επιχείρησης κατά δύο εβδομάδες, μπορεί έτσι να σημειωθεί σημαντική μείωση των συνολικών ετήσιων εσόδων, ή να παραμείνει μεγάλος όγκος χρηματικών πόρων σε μορφή εμπορεύματος που δεν κατάφερε να πουληθεί εγκαίρως. Τα προϊόντα της και ο όγκος των δεδομένων της FG δεν είναι τόσο μεγάλος ώστε να είναι απαραίτητο κάτι ιδιαίτερα περίπλοκο όπως ένα σύστημα τεχνητής νοημοσύνης ή ένα έμπειρο σύστημα. Ο όγκος των πληροφοριών είναι αρκετά περιορισμένος ώστε να μπορέσουν τα 2-3 άτομα της διοίκησης να λάβουν όλες τις αποφάσεις μεταξύ τους. Ωστόσο, ένα σύστημα πρόβλεψης της ζήτησης και υποστήριξης αποφάσεων σχετικών με τη διαχείριση των αποθεμάτων της θα διευκολύνει την όλη διαδικασία και θα εκτιμηθεί ιδιαίτερα.

5.2.2 Απαιτήσεις χρηστών και ορισμός προδιαγραφών

Εκτός από την κατανόηση του προβλήματος, ένας αναλυτής για το κάθε έργο του οφείλει να λάβει υπόψη τη διάθεση και της *απαιτήσεις των χρηστών*, δηλαδή των εργαζομένων που τελικά θα χρησιμοποιήσουν το τελικό προϊόν ώστε να αποκομίσουν τις δυνατότητες του. Εκτός από τις δυνατότητες του προϊόντος, θα πρέπει να ληφθούν υπόψη και οι δυνατότητες των ίδιων των χρηστών, σχετικά με τις γνώσεις τους αλλά και τη διάθεση τους περί του έργου και την αξία που η χρήση αυτού θα αποτελέσει για τους ίδιους. Εντός των πλαισίων της μελέτης περίπτωσης, υπήρξε επικοινωνία με όλους τους μελλοντικούς χρήστες του συστήματος. Εξ' αρχής πραγματοποιήθηκαν επανειλημμένες συνεδρίες με τον ιδιοκτήτη της επιχείρησης, ο οποίος εκτός από τις προσωπικές του βλέψεις γνωρίζει επίσης και τις αρμοδιότητες των εργαζομένων του, τις δυσκολίες που αντιμετωπίζουν καθώς και τις λύσεις που αναμένει ότι θα προκύψουν από τη χρήση ενός συστήματος πρόβλεψης της ζήτησης. Σε μία από τις συνεδρίες του τέθηκαν κάποιες ερωτήσεις σχετικά με τις απαιτήσεις του σε μορφή συνέντευξης. Συνοπτικά, οι απαιτήσεις και οι βλέψεις του ιδιοκτήτη της FG διατυπώνονται παρακάτω:

- Να βρεθεί ποιο είναι το κατάλληλο σημείο για το οποίο η παραγγελία νέου εμπορεύματος είναι απαραίτητη, για κάθε προϊόν.
- Να βρεθεί ο αριθμός αποθεμάτων που θεωρούνται «ασφαλές», δηλαδή που δε θα εξαντληθούν μέχρι την επόμενη παραλαβή εμπορεύματος, για κάθε προϊόν.

- Να βρεθεί η συχνότητα που θα πρέπει να αποστέλλονται αποθέματα στις αποθήκες της Amazon ώστε να μην εξαντληθούν και επηρεαστεί η διαδικασία FBA, γεγονός που θα ζημιώνει τις πωλήσεις.
- Να βρεθεί το κατάλληλο εύρος τιμών που θα μεγιστοποιήσει τα κέρδη, ακόμα και αν σημαίνει μείωση πωλήσεων όσο αναφορά τα τεμάχια.
- Να γίνει κατάλληλη διαχείριση των χρηματικών πόρων ώστε να μπορέσει να διατεθεί κεφάλαιο για να παραγγελθεί ο επαρκής αριθμός των προϊόντων που θα ικανοποιήσει τη ζήτηση όπως αυτή παρουσιάζεται για ολόκληρο τον ημερολογιακό χρόνο, αλλά με έμφαση στις περιόδους εποχικότητας.
- Να υπάρξουν πιο ακριβής εκτιμήσεις των ετησίων εσόδων που θα συμβάλλουν στην οικονομική διαχείριση.

Σχετικά με τις τεχνικές προδιαγραφές:

- Δίνεται έμφαση στη σημασία της οπτικής αναπαράστασης της πληροφορίας. Η χρήση γραφημάτων ενθαρρύνεται όπου είναι εφαρμόσιμη.
- Αναφέρθηκε σαν όραμα η εικόνα ενός «ταμπλό» που θα περιέχει πληροφορίες σχετικά με το παρελθόν ενός προϊόντος, καθώς και το μέλλον του. Θα εμφανίζονται γραφήματα που θα αναπαριστούν ιστορικές πωλήσεις για διαφορετικά προϊόντα – ώστε η σύγκριση να είναι εύκολη, και ανά κατηγορία προϊόντων. Όσο αναφορά το μέλλον, η προβλεπόμενη ζήτηση για το κάθε προϊόν θα εμφανίζεται επίσης μέσω γραφημάτων και κατάλληλων πινάκων.
- Τα αποτελέσματα θα πρέπει να είναι δομημένα και κατάλληλα οργανωμένα ώστε να γίνεται εύκολα επισκόπηση και αποκόμιση πληροφοριών με τη χρήση απλών κουμπιών (macros).
- Το ταμπλό θα πρέπει να είναι αρκετά απλό και εύχρηστο, ώστε να μπορέσει να χρησιμοποιηθεί ακόμα και από τους υπαλλήλους της FG που δεν έχουν ευχέρεια στην πληροφορική.

Στο επόμενο κεφάλαιο θα σχεδιαστεί η διαδικασία με την οποία θα προσεγγιστεί η πρόβλεψη της ζήτησης μέσω της ανάλυσης των διαθέσιμων δεδομένων και την ανάπτυξη των κατάλληλων μοντέλων.

6 Σχεδιασμός επίλυσης

Αφού αναφέρθηκαν όσες περισσότερες πληροφορίες υπήρχαν διαθέσιμες σχετικά με την επιχείρηση και τις διεργασίες που φέρουν το πρόβλημα για το οποίο η πρόβλεψη θα μπορούσε να βοηθήσει, το επόμενο βήμα είναι να ξεκινήσει η συλλογή, η ανάλυση των δεδομένων και η ανάπτυξη των μοντέλων. Επειδή όμως η όλη διαδικασία μπορεί να προβεί περίπλοκη, αρχικά θα διαμορφωθεί ένα πλάνο το οποίο θα ακολουθηθεί ώστε η όλη διαδικασία της πρόβλεψης να είναι δομημένη και σωστά διατυπωμένη. Ο σχεδιασμός θα γίνει με βάση τα πέντε βασικά βήματα για την ανάπτυξη μιας πρόβλεψης που αναφέρθηκαν στο δεύτερο κεφάλαιο.

6.1 Κατανόηση του προβλήματος

Στο προηγούμενο κεφάλαιο έγινε περιγραφή των χαρακτηριστικών που είναι σχετικά με τις διαδικασίες που η FG ακολουθεί για μία από τις πιο ουσιώδης λειτουργίες της, δηλαδή τη διαδικασία διαχείρισης αποθεμάτων. Αφού αναλύθηκαν οι διαφορετικές παράμετροι που επηρεάζουν τη διαδικασία αυτή, προσδιορίστηκε επίσης το επιθυμητό αποτέλεσμα, αυτό δηλαδή που θα επιφέρει τη λύση του προβλήματος. Ο σύντομος ορισμός του προβλήματος είναι το πως θα γίνει καλύτερη διαχείριση των αποθεμάτων, όσο αναφορά τον χρόνο τοποθέτησης νέας παραγγελίας για το επόμενο εμπόρευμα, ώστε το παρόν εμπόρευμα να μην προλάβει να εξαντληθεί. Αυτός ο ορισμός όμως στην πραγματικότητα είναι πολυσύνθετος και περιλαμβάνει όλες τις διαδικασίες και τις συνέπειες που απορρέουν από τη διαχείριση των αποθεμάτων της συγκεκριμένης επιχείρησης· σαφώς, η κάθε επιχείρηση λειτουργεί διαφορετικά. Ουσιαστικά, όπως αναφέρθηκε παραπάνω, για την κατάλληλη διαχείριση αποθεμάτων απαιτείται η γνώση της ζήτησης όπως αυτή εμφανίστηκε στο παρελθόν και ακόμα περισσότερο, η δυνατότητα πρόβλεψης αυτής. Οπότε, ο πιο άμεσος ορισμός του προβλήματος, και αυτός που η παρούσα μελέτη θα προσεγγίσει είναι ο *προσδιορισμός της ζήτησης, στο παρελθόν και στο μέλλον*. Ίσως στην πορεία να ανακαλυφθούν και κομμάτια του προβλήματος που δεν ήταν αρχικά αναμενόμενο ότι έπρεπε να διερευνηθούν, ή κομμάτια που για τα οποία η διερεύνηση δεν είναι δυνατή, είτε λόγω αδυναμίας δυνατοτήτων από άποψη λογισμικού, εμπειρίας είτε λόγω απουσίας σχετικών δεδομένων.

6.2 Συλλογή πληροφοριών

Στην περίπτωση της FG, τα καλά νέα είναι ότι η απόκτηση δεδομένων για τα δύο προηγούμενα χρόνια είναι εύκολη υπόθεση, αφού η πλατφόρμα Amazon Seller Central (ASC) διαθέτει δεδομένα για μέχρι δύο χρόνια πίσω, δηλαδή, σύμφωνα με τη χρονολογική τοποθέτηση της συγκεκριμένης μελέτης, από το 2015, μέχρι το 2017. Τα κακά νέα είναι ότι, εφόσον χρειάζονται τουλάχιστον τρία χρόνια ιστορικών δεδομένων, πρέπει να ελεγχθεί η περίπτωση να υπάρχουν αποθηκευμένα δεδομένα για το έτος 2014 ή και πιο παλιά. Η FG πραγματοποιεί αποθήκευση δεδομένων στο εταιρικό της ολοκληρωμένο πληροφοριακό σύστημα, το NetSuite. Εφόσον η επιχειρηματική της δραστηριότητα γίνεται μέσω Amazon, ή τουλάχιστον το κομμάτι της δραστηριότητας που θα αναλυθεί στη συγκεκριμένη μελέτη περίπτωσης, κάποια δεδομένα σχετικά με τις αποδόσεις των προϊόντων της απορρέουν από τη σελίδα αναλυτικών της Amazon, την ASC. Εξαιτίας λοιπόν της χρήσης διαφορετικών συστημάτων για την καταγραφή της εμπορικής δραστηριότητας της επιχείρησης, υπάρχουν πολλαπλά σετ διαθέσιμων δεδομένων πωλήσεων και αποδόσεων για τις ίδιες χρονικές περιόδους και, ανάλογα με την προέλευση τους, φέρουν διαφορετικά χαρακτηριστικά, ή αλλιώς *μεταβλητές*. Η συλλογή και οργάνωση των δεδομένων θα γίνει προσεκτικά και μεθοδικά, ώστε να παραχθεί ένα σύνολο δεδομένων, το οποίο θα χρησιμοποιηθεί για την περαιτέρω ανάλυση από την οποία θα προκύψουν τελικά τα μοντέλα πρόβλεψης.

6.3 Προκαταρκτική – εξερευνητική ανάλυση

Η συλλογή και οργάνωση των δεδομένων είναι το πρώτο κομμάτι της εξονυχιστικής αναζήτησης πληροφοριών από τα διαθέσιμα δεδομένα. Αφού τα δεδομένα οργανώθηκαν και φιλτράρονται με ένα τρόπο ώστε να μπορούν να μελετηθούν καλύτερα, πλέον η εξόρυξη των πληροφοριών είναι ευκολότερη. Σε αυτό το σημείο της ανάλυσης εκτός από την απόκτηση των πληροφοριών από τα δεδομένα είναι επίσης σημαντικό οι συγκεκριμένες πληροφορίες να είναι κατανοητές από τον αναλυτή, γι' αυτό θα χρειαστεί να γίνουν περαιτέρω ενέργειες λήψης πληροφοριών εκτός από την απλή παρατήρηση των δεδομένων. Συγκεκριμένα, και όπου είναι αυτό εφικτό, θα χρειαστεί να γίνει παρατήρηση του περιβάλλοντος, να υπάρξει επικοινωνία με ειδικούς για συγκεκριμένες διαδικασίες ή οτιδήποτε χρειαστεί ώστε να

υπάρχει η δυνατότητα ερμηνείας συγκεκριμένων μοτίβων που μπορεί να παρατηρηθούν, άτοπων τιμών ή απότομων μεταβολών.

Αρχικά, για την ανάλυση των δεδομένων της FG, η οπτική αναπαράσταση των δεδομένων βοηθάει ιδιαίτερα στην αναγνώριση μοτίβων που μπορεί να προκύπτουν από εποχικότητα ή αδυναμία της απαιτούμενης προσφοράς. *Γραφήματα* για διαφορετικά χαρακτηριστικά του συνόλου δεδομένων για το ίδιο προϊόν θα εμφανίσουν συσχετίσεις μεταξύ των δεδομένων. Συγκρίνοντας χρονικές περιόδους μεταξύ τους θα γίνει αντιληπτή η επιρροή συγκεκριμένων ή τυχαίων γεγονότων που συνέβησαν στο παρελθόν και έτσι θα υπάρχει η γνώση της συμπεριφοράς των σχετικών συνεπειών, αν κάτι παρόμοιο προκύψει στο μέλλον.

Παράλληλα, η εφαρμογή απλών περιγραφικών στατιστικών μεθόδων στο οργανωμένο σύνολο θα βοηθήσει στη δόμηση μιας περιληπτικής εικόνας των δεδομένων και σχετικές πληροφορίες που είναι δύσκολα ανιχνεύσιμες με μία πρώτη ματιά θα γίνουν εύκολα αντιληπτές. Για παράδειγμα, ως περιγραφικά στατιστικά θα εφαρμοσθούν το *ελάχιστο* και το *μέγιστο* των τιμών, ο *μέσος*, οι *τυπικές αποκλίσεις*, κλπ.. Περαιτέρω, αν υπάρχουν διαθέσιμα δεδομένα περισσότερων του ενός χρονικού κύκλου, όπως στην περίπτωση των δεδομένων της FG, προτείνεται η χρήση *διαγραμμάτων διασποράς (scatter plots)* και η ανεύρεση των *συντελεστών συσχετίσεων (correlation coefficients)* μεταξύ των διαφορετικών χαρακτηριστικών. Εφαρμόζοντας *μεθόδους αποσύνθεσης (decomposition analysis)* θα προκύψει η βαρύτητα της τάσης των δεδομένων, της εποχικότητας και των χρονικών κύκλων (π.χ. εμφανίζεται ομοιομορφία του μοτίβου κάθε μήνα, τρίμηνο, εξάμηνο; Στην περίπτωση της FG ο χρονικός κύκλος είναι ένα δωδεκάμηνο, και είναι ιδιαίτερα προφανές από τα γραφήματα που θα αναπτυχθούν στο επόμενο κεφάλαιο).

Η προκαταρκτική ανάλυση είναι ουσιώδες και στρατηγικό σημείο μιας διαδικασίας ανάπτυξης ενός μοντέλου πρόβλεψης. Η κατανόηση της συμπεριφοράς των δεδομένων είναι απαραίτητο στοιχείο για την επιλογή του κατάλληλου προγνωστικού μοντέλου, αφού πρέπει να υπάρχουν οι σωστές προσδοκίες ώστε να γίνει μια ποιοτική και αντικειμενική αξιολόγηση των αποτελεσμάτων των μοντέλων, από τα οποία θα γίνει η τελική επιλογή σχετικά με το ποιο μοντέλο θα εφαρμοστεί.

6.4 Επιλογή και προσαρμογή μοντέλων

Ανάλογα με τη διαθεσιμότητα των ιστορικών δεδομένων, υπάρχουν δύο ειδών μοντέλα πρόβλεψης – ποσοτικά και ποιοτικά. Τα ποσοτικά μοντέλα πρόβλεψης χρησιμοποιούν τα διαθέσιμα δεδομένα και εφαρμόζουν μαθηματικές μεθόδους πάνω στις ιστορικές αποδόσεις, ενώ για τα ποιοτικά μοντέλα δεν υπάρχουν αρκετά διαθέσιμα δεδομένα, αλλά υπάρχει αρκετή έμπειρη γνώση των ειδικών και η πρόβλεψη βασίζεται πάνω στην κρίση τους.

Για τα δεδομένα της FG, θα γίνει χρήση ποσοτικών μεθόδων πρόβλεψης. Αρχικά, θα γίνει έλεγχος των δυνατοτήτων πρόβλεψης του Excel. Στη συνέχεια θα εφαρμοστεί απλή γραμμική παλινδρόμηση και πολλαπλή γραμμική παλινδρόμηση πριν και μετά από λογαριθμική μεταμόρφωση του δείγματος. Θα πραγματοποιηθεί ανάπτυξη μοντέλων πρόβλεψης με εκθετική εξομάλυνση, καθώς και με χρήση των κινούμενων μέσων. Τέλος, θα γίνει ανάλυση χρονολογικών σειρών με χρήση στοχαστικών μοντέλων. Για το τελευταίο, θα γίνει η χρήση της γλώσσας προγραμματισμού R, στην οποία θα παραχθούν τα μοντέλα, για τα οποία οι είσοδοι θα εισάγονται από το Excel και τα αποτελέσματα θα επιστρέφονται επίσης στο Excel πάνω σε προκαθορισμένους πίνακες. Για όλα τα προαναφερόμενα μοντέλα η ανάπτυξη τους θα παραμείνει εξολοκλήρου στο Excel.

Αφού οι παραπάνω μέθοδοι τεθούν στην πράξη και απορρεύσουν αποτελέσματα, αυτά θα συγκριθούν ως προς τα πλεονεκτήματα και τα μειονεκτήματα τους.

6.5 Εφαρμογή και αξιολόγηση μοντέλων

Η μέθοδος που τελικά θα επιλεγεί ως η βέλτιστη, θα υποστεί παραμετροποίηση και θα εφαρμοστεί πάνω σε πολλαπλά δείγματα του συνόλου δεδομένων, ώστε να παρουσιαστούν κατάλληλα όλες οι δυνατότητες της συγκεκριμένης μεθόδου. Υπό οποιαδήποτε περίπτωση, η FG αντιμετωπίζει συγκεκριμένα προβλήματα που απαιτούν λύση με χρήση μοντέλων πρόβλεψης. Μετά την εφαρμογή και την αξιολόγηση αυτών, η διοίκηση θα χρησιμοποιήσει τις προβλέψεις ως βοήθημα – και όχι ως κανόνα για τις μελλοντικές της αποφάσεις, ώστε να αποφύγει να επαναλάβει λάθη του παρελθόντος, και να βελτιώσει τις αποδόσεις των προϊόντων της. Ο σκοπός της μελέτης περίπτωσης είναι να ενθαρρυνθεί η κατανόηση των μεθόδων πρόβλεψης που θα εφαρμοσθούν και να εστιαστεί η απόδοση της κάθε μεθόδου πάνω στο συγκεκριμένο πρόβλημα της FG, ενώ παράλληλα θα δοθούν εναλλακτικά

παραδείγματα προβλημάτων για τα οποία οι μέθοδοι που φάνηκαν αδύναμοι να αποτελέσουν λύση για την FG, θα μπορέσουν πιθανόν να αποδώσουν αποτελέσματα με περισσότερη ακρίβεια.

7 Υλοποίηση – προετοιμασία και κατανόηση των δεδομένων

7.1 Συλλογή των δεδομένων

Συγκεκριμένα, η FG ξεκίνησε να αποθηκεύει αυτόματα καθημερινά δεδομένα πωλήσεων που συμπεριλαμβάνουν αριθμό τεμαχίων που πουλήθηκαν και τη χρηματική τους αξία, από τον Ιούνιο του 2012. Νωρίτερα, η αποθήκευση παρόμοιων δεδομένων γινόταν εκούσια και ανά συγκεκριμένες χρονικές περιόδους, που σημαίνει ότι κάποιες είσοδοι μπορεί να αφορούσαν τις προηγούμενες δύο μέρες, την προηγούμενη εβδομάδα, ή ακόμα περισσότερο... Εφόσον δεν υπήρχε σταθερή χρονολογική μορφή των πωλήσεων, τα δεδομένα που θα χρησιμοποιηθούν για την ανάλυση της μελέτης αυτής, θα είναι από τον Ιούνιο του 2012 και μετά. Επιπλέον, πρέπει να ληφθεί υπόψη ότι τα προϊόντα μπορεί να φέρουν μεταβλητές τιμές, αφού η τιμή επηρεάζεται από διάφορους παράγοντες όπως προσφορές και εκπτώσεις, τη διαχείριση αποθεμάτων, κλπ. Τέλος, αν και υπάρχουν διαθέσιμα δεδομένα στη βάση δεδομένων της επιχείρησης σχετικά με πωλήσεις και τεμάχια, δεν υπάρχουν επιπλέον χαρακτηριστικά αποδόσεων της Amazon, όπως επισκέψεις σελίδων, ποσοστά μετατροπών κλπ. Είναι σαφές ότι τα σύνολα δεδομένων τα οποία θα χρησιμοποιηθούν θα πρέπει να είναι συμβατά, οπότε για μοντέλα που εξαρτώνται από συντελεστές πρόβλεψης (predictors), δε θα χρησιμοποιηθούν δεδομένα πωλήσεων πριν από το 2014, για τα οποία δεν υπάρχουν σχετικά χαρακτηριστικά αποδόσεων.

Σαν πρώτο πλάνο, η συλλογή των δεδομένων γίνεται μέσω της ASC. Στο έγγραφο αυτό θα προσομοιωθεί η διαδικασία πρόβλεψης για τη ζήτηση της αγοράς του Ηνωμένου Βασιλείου. Για το πρόβλημα της διαχείρισης αποθεμάτων είναι απαραίτητο να διερευνηθεί η ζήτηση για όλες τις αγορές στις οποίες η FG απευθύνεται. Ωστόσο, για να μη γίνει η διαδικασία της πρόβλεψης ιδιαίτερα περίπλοκη, και αφού η προσέγγιση των λοιπών αγορών θα ήταν η ίδια, οι μέθοδοι πρόβλεψης θα εφαρμοσθούν πάνω στα δεδομένα της amazon.co.uk, τα οποία αφορούν τις πωλήσεις της FG μέσω της πλατφόρμας της amazon.co.uk, μόνο. Όλες οι αναφορές θα πρέπει να καλύπτουν την *ίδια απόσταση ημερών*. Στη συγκεκριμένη περίπτωση η χρονική περίοδος των αναφορών καταχωρήθηκε ως εβδομαδιαία και με έναρξη κάθε Κυριακή και λήξη το επόμενο Σάββατο. Ξεκινώντας λοιπόν από τέλη Φλεβάρη του 2015, αποθηκεύτηκαν εβδομαδιαίες αναφορές σε μορφή αρχείων excel (.csv). Επιπλέον, βρέθηκαν παρόμοιες αναφορές, αποθηκευμένες στο παρελθόν, που ξεκινούσαν από τον Φλεβάρη του

2014, οπότε έτσι εμφανίστηκαν δεδομένα για έναν επιπλέον χρόνο που εναλλακτικά δε θα ήταν προσβάσιμα από την ASC.

Οι αναφορές με τα πηγαία δεδομένα φέρουν ακατέργαστες πληροφορίες. Οι αναφορές αυτές αποθηκεύτηκαν στο Excel ώστε να οργανωθούν και να είναι δυνατή η περαιτέρω ανάλυση των δεδομένων και εξόρυξη των πληροφοριών, αλλά είναι επίσης δυνατόν να γίνει περιήγηση και αναζήτηση συγκεκριμένων προϊόντων ανά οποιαδήποτε χρονική περίοδο κατευθείαν εντός της ASC. Αφού οργανωθούν τα δεδομένα αυτά στο Excel, μπορεί κάποιος να περιηγηθεί μέσα στις καρτέλες και να κάνει εύρεση σε συγκεκριμένα προϊόντα χρησιμοποιώντας χαρακτηριστικά αυτών όπως τον αριθμό τεμαχίου διαχείρισης αποθεμάτων (Stock Keeping Unit – *SKU*), τον αναγνωριστικό αριθμό της Amazon Seller (Amazon Seller Identification Number – *ASIN*), όνομα καταχωρημένου προϊόντος κλπ. Κάθε αναφορά, εκτός από τα αναγνωριστικά χαρακτηριστικά για κάθε προϊόν, φέρει τις εξής πληροφορίες: (Amazon.co.uk - Business Reports Glossary)

- Συνεδρίες (*Sessions*) για το συγκεκριμένο προϊόν. Οι επαναλαμβανόμενες επισκέψεις ενός χρήστη στη σελίδα προϊόντος στο διάστημα 24 ωρών μετράνε ως μία συνεδρία.
- Ποσοστό συνεδριών (*Session %*) που κατέχει το συγκεκριμένο προϊόν σε σύγκριση με τις συνολικές συνεδρίες όλων των προϊόντων
- Επισκέψεις (*Page Views*) για το συγκεκριμένο προϊόν. Δηλαδή πόσες φορές έχει φορτωθεί η σελίδα του προϊόντος, συνολικά για την περίοδο της αναφοράς.
- Ποσοστό επισκέψεων (*Page Views %*) που κατέχει το συγκεκριμένο προϊόν σε σύγκριση με τις συνολικές επισκέψεις σελίδων όλων των προϊόντων.
- Ποσοστό εμφάνισης κουμπιού αγοράς (*Buy Box %*). Υπάρχουν διάφοροι λόγοι για τους οποίους μία σελίδα προϊόντος αν και ενεργή, δε φέρει το κουμπί προσθήκης στο «καλάθι». Ο κύριος λόγος είναι η περίπτωση εξάντλησης αποθεμάτων. Το συγκεκριμένο ποσοστό έχει σημασία εφόσον η περίπτωση απώλειας του κουμπιού αγοράς επηρεάζει άμεσα τις πωλήσεις.
- Τεμάχια που πουλήθηκαν (*Units Ordered*).
- Ποσοστό μετατροπής – κοινώς Conversion Rate (*Unit Session Ratio*) του συγκεκριμένου προϊόντος, δηλαδή το ποσοστό των συνεδριών που καταλήξαν σε αγορά ενός τεμαχίου τουλάχιστον.
- Η συνολική χρηματική αξία των καταγεγραμμένων πωλήσεων για την περίοδο της αναφοράς (*Ordered Product Sales*).

- Συνολικός αριθμός παραγγελιών (*Total Order Items*), όπου εμφανίζεται το συγκεκριμένο προϊόν. Ο συνολικός αριθμός παραγγελιών για το κάθε προϊόν (*Total Order Items*) θα είναι πάντα μικρότερος ή ίσος με τα τεμάχια που πουλήθηκαν (*Units Ordered*).

Title	SKU	Sessions	Session Percentage	Page Views	Page Views Percentage	Buy Box Percentage	Units Ordered	Unit Session Percentage	Ordered Product Sales	Total Order Items
Kids Tea Party Wooden Biscuit Set Counting Game (10 Biscuits) Lucy Locket	BD/2905/04/00/UK	94	0.63%	115	0.58%	100%	10	10.64%	£129.90	10
Kids Boys Yellow Lion Fancy Dress Costume (3-8 Years) Slimy Toad	BY/5214/26/00/UK	70	0.47%	95	0.48%	100%	9	12.86%	£179.91	9

Εικόνα 9: Επιχειρηματικές αναφορές με λεπτομέρειες πωλήσεων και κίνησης με φίλτρο το ASIN, όπως εμφανίζονται στην Amazon Seller Central.

7.2 Οργάνωση των δεδομένων

Τα ξεχωριστά αρχεία .csv επικολλήθηκαν σε ξεχωριστές καρτέλες ενός αρχείου Excel. Για τη μετέπειτα ομαλή λειτουργία των συναρτήσεων του Excel, που θα χρησιμοποιηθούν για την ανάλυση των δεδομένων αυτών, είναι σημαντικό να υπάρχει πλήρης ομοιογένεια των δεδομένων που εμφανίζονται στην κάθε καρτέλα, καθώς και το όνομα των καρτελών να ακολουθεί *συγκεκριμένη δομή* και να έχει *λογική συνέχεια*. Τα ονόματα των καρτελών στη συγκεκριμένη περίπτωση φέρουν την εξής μορφή: “BR – ”, [έτος], HH-MM, “to”, ([έτος]), HH-MM. Στο δεύτερο σκέλος του ονόματος το έτος είναι προαιρετικό και καταχωρείται μόνο όταν η συγκεκριμένη περίοδος αναφοράς βρίσκεται πάνω στην αλλαγή του έτους. Επειδή όλες οι αποθηκευμένες αναφορές αποτελούνται από τα προκαθορισμένα εβδομαδιαία όρια που αναφέρθηκαν παραπάνω, ανεξαρτήτως αν το έτος αλλάζει στη μέση της εβδομαδιαίας αναφοράς, έπρεπε κάπως να καθοριστεί ότι η αλλαγή των ημερομηνιών από τέλος Δεκεμβρίου σε αρχές Ιανουαρίου επισημάνουν και την αλλαγή του χρόνου. Εννοείται ότι αυτό γίνεται για τεχνικούς λόγους ώστε να είναι δυνατή η χρήση συναρτήσεων πάνω στα ονόματα των καρτελών, και όχι για λόγους ανθρώπινης αντίληψης. Εναλλακτικά, το έτος στο δεύτερο σκέλος του ονόματος θα μπορούσε να είναι μονίμως παρόν και όχι προαιρετικό, διότι η παρουσία όπως και η απουσία του δεν πληροί το κριτήριο της ομοιομορφίας δομής, αλλά παράλληλα δεν αποτελεί πρόβλημα που να μην μπορεί να αντιμετωπιστεί με τη χρήση μίας περισσότερο πολύπλοκης συνάρτησης. Επιπλέον, η καταχώρηση των ονομάτων θα πρέπει να γίνει προσεκτικά και χωρίς λάθη, αφού ένα επιπλέον ή λιγότερο κενό θα κάνει

κάποια συνάρτηση να αποτύχει. Ακόμα, υπάρχει διαφορά στη δήλωση της μορφής της ημερομηνίας ως 09-05 για την ενάτη Μαΐου σε σχέση με τη μορφή 5-9, για την ίδια ημερομηνία, οπότε θα πρέπει να δοθεί ιδιαίτερη σημασία στις λεπτομέρειες αυτού του είδους.

BR - 2015 11-01 to 17-01	137	6.08%	189	6.36%	100%	15	10.95%	£134.85
BR - 2015 18-01 to 24-01	59	2.62%	93	3.13%	100%	9	15.25%	£134.91
BR - 2015 25-01 to 31-01	61	2.71%	87	2.93%	100%	9	14.75%	£125.91
BR - 2015 01-02 to 07-02	36	1.60%	48	1.61%	100%	14	38.89%	£125.86
BR - 2015 08-02 to 14-02	121	5.37%	154	5.18%	100%	8	6.61%	£111.92
BR - 2015 15-02 to 21-02	75	3.33%	105	3.53%	99%	7	9.33%	£97.93
BR - 2015 22-02 to 28-02	22	0.98%	26	0.87%	100%	7	31.82%	£34.93
BR - 2015 01-03 to 07-03	90	3.99%	120	4.03%	100%	7	7.78%	£55.93
BR - 2015 08-03 to 14-03	37	1.64%	45	1.51%	100%	7	18.92%	£45.43
BR - 2015 15-03 to 21-03	82	3.64%	107	3.60%	99%	6	7.32%	£89.94
BR - 2015 22-03 to 28-03	47	2.08%	66	2.22%	100%	6	12.77%	£89.94
BR - 2015 29-03 to 04-04	62	2.75%	83	2.79%	100%	6	9.68%	£71.94
BR - 2015 05-04 to 11-04	31	1.37%	39	1.31%	97%	5	16.13%	£49.95
BR - 2015 12-04 to 18-04	22	0.98%	33	1.11%	100%	5	22.73%	£49.95
BR - 2015 19-04 to 25-04	99	4.39%	118	3.97%	100%	5	5.05%	£63.95
BR - 2015 26-04 to 02-05	24	1.06%	33	1.11%	100%	4	16.67%	£19.96
BR - 2015 03-05 to 09-05	18	0.80%	28	0.94%	100%	3	16.67%	£14.97
BR - 2015 10-05 to 16-05	11	0.49%	13	0.44%	100%	3	27.27%	£53.97
BR - 2015 17-05 to 23-05								
BR - 2015 24-05 to 30-05								

Εικόνα 10: Οργάνωση των αποθηκευμένων επιχειρησιακών αναφορών

Παρομοίως, τα κριτήρια της συγκεκριμένης δομής και λογικής συνέχειας θα πρέπει να πληρούνται στο περιεχόμενο της κάθε καρτέλας. Οι στήλες θα πρέπει να ακολουθούν την ίδια σειρά εμφάνισης και να είναι πάντα παρούσες σε όλες τις καρτέλες. Π.χ. τα δεδομένα που αποτελούν τα έσοδα σε χρηματικές μονάδες για το κάθε προϊόν θα πρέπει να βρίσκονται πάντα στη στήλη *M*, χωρίς απολύτως καμία εξαίρεση. Στη συνέχεια θα προστεθούν περισσότερες δυναμικές στήλες σε κάθε καρτέλα για περαιτέρω οργάνωση των πληροφοριών. Οι επιπλέον καρτέλες θα πρέπει ομοίως να επικολληθούν στις κατάλληλες θέσεις, ώστε να υπάρχει ομοιομορφία σε όλες τις καρτέλες.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	(Parent) A (Child) AS Title	SKU	Sessions	Session P	Page View	Page View	Buy Box P	Units Ord	Unit Sessi	Ordered F	Total Order	Items		
2	B00175N8	B00175N8	Beautiful	58-R66F-B	942	6.35%	1,299	6.60%	99%	271	28.77%	£2,436.29	252	
3	B000YWLL	B000YWLL	Pink Fairy	ST/5104/1	1,374	9.27%	1,992	10.12%	100%	248	18.05%	£3,788.02	240	
4	B00FDPC2	B00FDPC2	Pink Toad	ST/2641/1	659	4.44%	879	4.47%	100%	176	26.71%	£1,230.24	155	
5	B00A6RTL	B00A6RTL	Princess T	BD/2578/2	814	5.49%	1,070	5.44%	100%	129	15.85%	£1,821.71	129	

Εικόνα 11: Η αρχική δομή των πηγαίων δεδομένων

Όλες οι καρτέλες θα υποστούν επεξεργασία ώστε να βοηθήσουν στην συμπλήρωση συγκεκριμένων πινάκων περιληπτικών πληροφοριών για τα δεδομένα. Συγκεκριμένα, θα δημιουργηθούν δύο καρτέλες στις οποίες θα αναπτυχθούν δύο πίνακες, ο ένας εκ των οποίων θα εμφανίζει πληροφορίες για ένα αντικείμενο αφού τοποθετηθεί το κλειδί του, και ο άλλος θα εμφανίζει πληροφορίες για ολόκληρη κατηγορία προϊόντων. Και για τις δύο περιπτώσεις θα χρειαστεί να γίνουν κάποιοι υπολογισμοί πάνω στα πηγαία δεδομένα, ώστε οι πίνακες που θα δημιουργηθούν να κάνουν εύρεση και να έχουν ως έξοδο τα σωστά αποτελέσματα.

Συγκεκριμένα, για τον πίνακα προϊόντος, το πρώτο πρόβλημα που θα προκύψει είναι η απόφαση σχετικά με το ποιο κλειδί θα χρησιμοποιηθεί για την εύρεση. Για αυτό υπάρχουν οι παρακάτω επιλογές:

- i. *ASIN* – Μοναδικός κωδικός για το κάθε προϊόν ο οποίος καταχωρείται αυτόματα από την Amazon με την καταχώρηση της σελίδας προϊόντος. Είναι 13ψήφιος κωδικός και αποτελείται από λατινικά κεφαλαία γράμματα και αριθμούς. Χρησιμοποιείται από την Amazon ως κλειδί του καταχωρημένου προϊόντος στη βάση δεδομένων της. Δεν υπάρχει περίπτωση διπλοτυπίας.
- ii. *SKU* – Μοναδικός κωδικός για το κάθε προϊόν ο οποίος καταχωρείται επιλεκτικά από την επιχείρηση. Έχει μορφή *XX/#####/(##/##)*. Τα πρώτα δύο ψηφία αποτελούν γράμματα ως αναγνωριστικό της κατηγορίας στην οποία ανήκει το προϊόν, π.χ. «ST» για «Stationary». Τα επόμενα τέσσερα ψηφία είναι νούμερα μοναδικά για το κάθε προϊόν. Οι τελευταίες δύο δυάδες διψήφιων αριθμών είναι προαιρετικές και συμβολίζουν το χρώμα του προϊόντος, εάν αυτό υπάρχει σε διαφορετικά χρώματα και το μέγεθος του προϊόντος, που εμφανίζεται κατά κύριο λόγο σε ενδύματα και υποδήματα. Προϊόντα που υπάρχουν σε μία μόνο έκδοση φέρουν μονάχα τα πρώτα δύο μέρη του κωδικού, πχ. ST/5104 για τη μουσική κοσμηματοθήκη που είναι νούμερο ένα σε πωλήσεις.
- iii. Όνομα προϊόντος – Το όνομα ή αλλιώς ο «τίτλος» προϊόντος είναι ο ίδιος με τον τίτλο της σελίδας προϊόντος όπως εμφανίζεται στα αποτελέσματα αναζήτησης της Amazon, οπότε στο όνομα χρησιμοποιούνται λέξεις-κλειδιά και κατάλληλη περιγραφή ώστε να υπάρχουν περισσότερες πιθανότητες να εμφανιστεί το προϊόν στον ενδιαφερόμενο χρήστη με όποιον τρόπο είναι πιθανό εκείνος να το αναζητήσει.

Λόγο της μοναδικότητας τους και της προκαθορισμένης δομής τους, πιθανά κλειδιά για το πρόβλημα που θα αντιμετωπιστεί είναι τα ASIN και SKU, και το όνομα προϊόντος θα απορριφθεί. Ύστερα από επισκόπηση παρατηρήθηκε ότι και τα δύο χαρακτηριστικά είναι όντως μοναδικά όσο αναφορά τα διπλότυπα, όπως ήταν αναμενόμενο. Ωστόσο, παρατηρήθηκε επίσης ότι το κάθε ASIN αντιπροσωπεύει ένα μόνο προϊόν, ενώ για κάθε ASIN μπορεί να υπάρχουν ένα ή δύο SKU. Εδώ, ως «προϊόν» εννοείται μια καταχωρημένη σελίδα προϊόντος όπως εμφανίζεται στην Amazon.co.uk, και για την κάθε σελίδα, όπως αναφέρθηκε προηγουμένως, η Amazon καταχωρεί αυτόματα το δικό της αναγνωριστικό κωδικό, δηλαδή το ASIN. Οπότε αφού η Amazon αυτοματοποιεί τη διαδικασία έκδοσης των ASIN και είναι γεγονός ότι για κάθε προϊόν υπάρχει μόνο ένα ASIN, τι συμβαίνει με τα SKU και γιατί εμφανίζονται παραπάνω από ένα για κάθε προϊόν;

Η απάντηση στην παραπάνω ερώτηση έχει να κάνει με τη διαδικασία ολοκλήρωσης της παραγγελίας (*order fulfilment*). Η Amazon χρησιμοποιεί το ASIN ως αναγνωριστικό για το προϊόν, ώστε να ξέρει τι παραγγέλθηκε και πως να διαχειριστεί το οικονομικό κομμάτι της παραγγελίας. Το SKU χρησιμοποιείται από την Amazon ως αναγνωριστικό αναζήτησης του προϊόντος στις αποθήκες της. Αν το προϊόν υπάρχει στις αποθήκες της Amazon και ο πελάτης έχει επιλέξει να το παραλάβει μέσω Amazon, τότε η Amazon θα χρησιμοποιήσει το δικό της καταχωρημένο SKU για τη διαδικασία της ολοκλήρωσης. Αν η ολοκλήρωση της παραγγελίας γίνει από την FG, τότε θα χρησιμοποιηθεί το SKU που καταχωρήθηκε από την FG, η μορφή του οποίου αναλύθηκε παραπάνω. Αυτός είναι ο λόγος που εμφανίζονται μέχρι δύο SKU για το κάθε ASIN, και ως συνέπεια, με την ανάγνωση μιας γραμμής σε μία αποθηκευμένη αναφορά, προκύπτουν πληροφορίες σχετικά με το ποιο προϊόν πουλήθηκε, πως έγινε η ολοκλήρωση της παραγγελίας, τον πλήρες τίτλο προϊόντος και όλα τα υπόλοιπα χαρακτηριστικά που παραχωρεί η Amazon ως στατιστικά δεδομένα απόδοσης του προϊόντος.

Συνεπώς, για τη διαδικασία εύρεσης και οργάνωσης των δεδομένων θα απορριφθεί επίσης το SKU και τελικά, ως κλειδί θα χρησιμοποιηθεί το ASIN. Ωστόσο, ένα ακόμα πρόβλημα που προκύπτει από την ύπαρξη πολλαπλών SKU είναι η λειτουργία της συνάρτησης *VLOOKUP(vlookup_value,table_array,col_index_num,[range_lookup])*. Η συνάρτηση αυτή δέχεται ως είσοδο τον όρο αναζήτησης, δηλαδή το ASIN καθώς επίσης την καρτέλα στην οποία θα γίνει η αναζήτηση και το διάστημα των στηλών, στις οποίες η αναζήτηση θα γίνει κάθετα (*vlookup* = vertical look up, δηλαδή κάθετη αναζήτηση). Τέλος ως είσοδο δέχεται τον αριθμό της στήλης του διαστήματος που καταχωρήθηκε στο προηγούμενο κομμάτι, η

οποία στήλη θα είναι και η έξοδος της συνάρτησης. Κατά τον υπολογισμό της συνάρτησης, η αναζήτηση γίνεται από το πρώτο κελί της πρώτης στήλης του διαστήματος που δηλώθηκε, και αν το περιεχόμενο του κελιού δεν ταιριάζει με τον όρο αναζήτησης, τότε ελέγχεται το επόμενο κελί καθέτως και η διαδικασία συνεχίζεται μέχρι και την τελευταία γραμμή του διαστήματος, εκτός και αν βρεθεί κάποιο ταίρι νωρίτερα. Στην περίπτωση που βρεθεί ταίρι, η αναζήτηση τερματίζει, και ως έξοδος επιστρέφεται η δηλωμένη στήλη της γραμμής στην οποία βρέθηκε το ταίρι. Στην περίπτωση του ASIN, εφόσον μπορεί να υπάρξουν και δύο SKU, σημαίνει ότι το ASIN μπορεί να εμφανιστεί και δεύτερη φορά.

(Child) AS Title	SKU	Sessions	Session P	Page View	Page View	Buy Box P	Units Ord	Unit Sessi	Ordered P	Total Orde
B00175N8 Beautiful	58-R66F-B	489	6.65%	689	6.74%	100%	233	47.65%	£1,861.67	224
B01E70VS1 Enchantec	BD/2901/0	554	7.53%	824	8.06%	100%	178	32.13%	£2,579.22	174
B000YWLL Pink Fairy	ST/5104/1	948	12.89%	1,340	13.11%	100%	162	17.09%	£2,428.38	161
B000YWLL Pink Fairy	3N-ZBEG-I	948	12.89%	1,340	13.11%	100%	18	1.90%	£287.82	16

Εικόνα 12: Παράδειγμα προβλήματος πολλαπλών SKU

Αν τα δεδομένα αφεθούν όπως αποθηκεύτηκαν αρχικά από την ASC, τότε τα αποτελέσματα των VLOOKUP θα αντιπροσωπεύουν μόνο την πρώτη στη σειρά εμφάνιση του ASIN, είτε αυτό αφορούσε ολοκληρώσεις παραγγελιών μέσω της Amazon είτε μέσω της FG. Αν σκοπός της ανάλυσης ήταν η διαφοροποίηση των δύο παραπάνω σεναρίων, τότε θα έπρεπε να ληφθούν τα κατάλληλα μέτρα ώστε ως αποτέλεσμα να επιστρέφεται το κατάλληλο σενάριο. Στη συγκεκριμένη περίπτωση όμως είναι απαραίτητο να ληφθούν υπόψη και τα δύο σενάρια, συγκεκριμένα η πρόσθεση των δύο σεναρίων, που συνολικά αντιπροσωπεύουν τις πωλήσεις του προϊόντος για το χρονικό διάστημα της αποθηκευμένης αναφοράς. Για να προκύψει το άθροισμα των διπλών εγγραφών θα χρειαστεί να γίνει χρήση της συνάρτησης *SUMIF(range,criteria,[sum_range])*, η οποία δέχεται ως είσοδο το διάστημα αναζήτησης, το κριτήριο με βάση το οποίο γίνεται η αναζήτηση και τέλος το διάστημα της στήλης από την οποία θα παραχθούν τα αθροίσματα για όσα ASIN ταιριάζουν με το κριτήριο εντός του διαστήματος που ορίστηκε πρώτο. Ως αποτέλεσμα θα είναι μία νέα στήλη, όπου για τις εγγραφές που ταιριάζουν, η τιμή του αθροίσματος θα είναι η ίδια. Άρα, αν και το VLOOKUP θα γίνει όπως είχε αρχικά προγραμματιστεί, η στήλη της εξόδου του θα είναι η νέα αθροιστική στήλη που φέρει τη συνάρτηση *SUMIF*. Αλλά ποιες στήλες ακριβώς θα χρειαστεί να προστεθούν για το επιθυμητό αποτέλεσμα;

Όπως φαίνεται στην εικόνα 12, από τις εννέα στήλες στα δεξιά που εμφανίζουν τα στατιστικά αποδόσεων, οι πρώτες πέντε στήλες για τις δύο διπλές εγγραφές αποτελούνται από τις ίδιες τιμές. Αυτό συμβαίνει γιατί οι συγκεκριμένες στήλες υπολογίζονται με βάση του ASIN, ενώ οι τελευταίες τέσσερις στήλες που αφορούν καθαρά τις πωλήσεις υπολογίζονται με βάση του SKU, το οποίο ανήκει σε κάποιο ASIN. Συνεπώς, για τις τέσσερις τελευταίες στήλες, θα προστεθούν τέσσερις νέες στήλες που θα περιέχουν τη συνάρτηση *SUMIF*, αλλά η καθεμία με τις δικές τις παραμέτρους. Στην περίπτωση της στήλης «Unit Session %», είναι επίσης εφικτό να παραληφθεί η χρήση της *SUMIF* και να γίνει η απλή διαίρεση της νέας στήλης «SUM Units Ordered» προς τη στήλη «Sessions». Και στις δύο περιπτώσεις το αποτέλεσμα θα είναι το ίδιο. Αφού δημιουργηθούν οι νέες στήλες που θα αντικαταστήσουν τις αντίστοιχες αρχικές στήλες περί πωλήσεων, θα πρέπει να επικολληθούν στην κάθε καρτέλα που περιέχει μία εβδομαδιαία επιχειρηματική αναφορά.

SKU	Sessi	Session	Page	Page		d F Total	SUM Units Orde	SUM Unit	SUM Orde	SUM Total
58-R66F-B	489	6.65%	689	6.74%	=SUMIF(\$B:\$B,\$B2,K:K)	67 224	233.00	0.48	1861.67	224.00
BD/2901/C	554	7.53%	824	8.06%	100%	1/8 32.13%	178.00	0.32	2579.22	174.00
ST/5104/1	948	12.89%	1,340	13.11%	100%	162 17.09%	180.00	0.19	2716.20	177.00
3N-ZBEG-I	948	12.89%	1,340	13.11%	100%	18 1.90%	180.00	0.19	2716.20	177.00

Εικόνα 13: Επίλυση του προβλήματος των πολλαπλών SKU

Έπειτα από τη συγκεκριμένη τροποποίηση των δεδομένων, έχουν ληφθεί όλα τα κατάλληλα μέτρα για τη σύνταξη του περιληπτικού πίνακα προϊόντος. Ωστόσο, απαιτείται περαιτέρω παραμετροποίηση για τη σύνταξη του πίνακα κατηγορίας, αφού μέχρι στιγμής δεν υπάρχει κάποια στήλη στα δεδομένα που να διευκρινίζει την κατηγορία στην οποία ανήκει το κάθε προϊόν. Η ASC δεν αποθηκεύει δεδομένα σχετικά με την κατηγορία του προϊόντος. Σχετικά δεδομένα υπάρχουν διαθέσιμα στο πληροφοριακό σύστημα της επιχείρησης. Λόγο της φύσεως των συγκεκριμένων δεδομένων, η οργάνωσή τους θα πρέπει να προσεγγισθεί ελαφρώς διαφορετικά. Η πρόσβαση στα δεδομένα περί κατηγορίας προϊόντων αποκτάται μέσω μίας αναζήτησης (search) εντός του NetSuite. Θέτοντας τις κατάλληλες παραμέτρους και τα σωστά πεδία, το αποτέλεσμα της αναζήτησης θα είναι αρκετές σελίδες γραμμών που φέρουν τις κατάλληλες πληροφορίες, η κάθε μία από τις γραμμές αποτελεί ένα προϊόν.

NAME A	DESCRIPTION	NAVIGATION OLD ITEM CODE	AZ UK ASIN	FBM SKU UK (TOYS)	FBA SKU UK (TOYS)	PRODUCT MAIN GROUP
ST/5103	Musical Fairy Wardrobe	ST/5103/10/00			U1-Y380-LIHU	Musical boxes
ST/5104	Musical Fairy Jewellery Chest	ST/5104/10/00	B000YWLU05	3N-ZBEG-M0CJ	ST/5104/10/00	Musical boxes
ST/5108	Fairy Skipping Rope	ST/5108/10/00		T5-4P4H-OZH2		Play

Εικόνα 14: Εμφάνιση των αποτελεσμάτων της αναζήτησης του NetSuite περί κατηγορίας προϊόντων

Επόμενο βήμα είναι η αποθήκευση των αποτελεσμάτων σε ένα αρχείο .csv, ώστε να γίνει η συγχώνευση τους με τα υπόλοιπα δεδομένα που έχουν αποθηκευτεί από την ASC. Ο τελικός σκοπός είναι να δημιουργηθεί μια νέα στήλη στις καρτέλες αναφορών η οποία θα αναφέρει σε ποια κατηγορία (product main group) ανήκει το κάθε προϊόν. Αφού το περιεχόμενο του αποθηκευμένου αρχείου αποθηκευτεί σε νέα καρτέλα εντός του αρχείου Excel όπου έχουν οργανωθεί και τα δεδομένα της ASC, αν και δεν είναι απαραίτητο αφού είναι δυνατόν να γίνει διασύνδεση ξεχωριστών αρχείων Excel για τη χρήση των συναρτήσεων *lookup*, είναι σημαντικό να γίνει έλεγχος των δεδομένων από το NetSuite ως προς την ακεραιότητα τους. Τα συγκεκριμένα δεδομένα δεν παράγονται αυτόματα όπως στην περίπτωση των αναφορών της ASC, αλλά κάποιος υπάλληλος της FG έχει καταχωρήσει την κάθε ξεχωριστή εγγραφή. Εφόσον εδώ είναι παρόν ο ανθρώπινος παράγοντας, υπάρχει πολύ περισσότερος χώρος για σφάλματα. Όπως φαίνεται και από την *εικόνα 14*, οι τρεις στήλες που περιέχουν τους κωδικούς ASIN και SKU, έχουν κάποια κενά. Τα ελλιπή δεδομένα θα αποτελέσουν πρόβλημα κατά τα αποτελέσματα της συνάρτησης *VLOOKUP*, διότι αφού τα ASIN και SKU δεν είναι παντού συμπληρωμένα, είναι αδύνατον να επιλεγεί μία συγκεκριμένη στήλη αναζήτησης στην οποία θα βρεθεί ταιρί για το δηλωμένο κριτήριο. Στην περίπτωση που δεν είναι δυνατή η αποκατάσταση της αρτιμέλειας των δεδομένων, θα χρειαστεί να συνταχθεί μια περίπλοκη συνάρτηση που θα περιέχει πολλαπλές *VLOOKUP* και ελέγχους σφαλμάτων ώστε να χρησιμοποιηθούν και τα δύο διαθέσιμα χαρακτηριστικά ταυτοποίησης που υπάρχουν διαθέσιμα για την κάθε γραμμή στις εβδομαδιαίες καρτέλες αναφορών.

B	C	D	E	F	G	H
(Child) ASI	Title	SKU	Category	Session	Session Percentage	Page V
B000YWLU05	Musical Fairy Jewellery Ch	3N-ZBEG-M0CJ	Musical boxes	332	0.08329152	509
B00A6RTLCK	Princess Tin Tea Set	BD/2578/20/00/UK/FBM	H,5,FALSE),TRUE,")	287	0.072002007	402
B003LKUGQ4	Musical Princess Jeweller	ST/5174/20/00/UK	Musical boxes	256	0.064224787	355

Εικόνα 15: Χρήση περίπλοκης αναζήτησης για τη λύση του προβλήματος μη ακεραιότητας των δεδομένων

Αν και πολυσύνθετη, η συνάρτηση που εμφανίζεται στην *εικόνα 15* απαντάει στο πρόβλημα των κενών κελιών των δεδομένων που αποθηκεύτηκαν από το NetSuite, και έτσι η χρήση τους και η ταυτοποίηση του κάθε προϊόντος με την κατηγορία του είναι εφικτή. Η νέα στήλη *E* φέρει την κατηγορία για το κάθε προϊόν, καθιστώντας δυνατή τη χρήση της για να ομαδοποιηθούν οι γραμμές και να χρησιμοποιηθούν για τη σύνταξη νέων στηλών με δεδομένα αθροιστικά ανά κατηγορία προϊόντος. Οι νέες στήλες περί κατηγορίας, θα χρησιμοποιηθούν στον περιληπτικό πίνακα κατηγορίας, άρα θα χρειαστεί να αθροιστούν όλες οι αρχικές στήλες με στατιστικά δεδομένα. Ως αποτέλεσμα, εκτός από τις 13 αρχικές στήλες που αποθηκεύτηκαν κατευθείαν από την ASC και τη στήλη *E* που εμπεριέχει τις κατηγορίες, συντάχθηκαν 13 επιπλέον στήλες με αθροιστικά δεδομένα, τέσσερις από τις οποίες απαντούν στο πρόβλημα των πολλαπλών SKU, ενώ οι υπόλοιπες εννέα φέρουν τα αθροιστικά στατιστικά χαρακτηριστικά αποδόσεων για το κάθε προϊόν, ανά κατηγορία.

O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
SUM Units Ordered	SUM Unit Session Percentage	SUM Ordered Product Sales	SUM Total Order Items	Cat Sessions	Cat Session %	Cat Page Views	Cat Page Views %	Cat Buy Box	Cat Units Ordered	Cat Unit Session %	Cat Ordered Product Sales	Cat Total Order Items
40.00	12%	639.60	40.00	798.00	20%	1173.00	21%	0.96	76.00	10%	1197.24	76.00
33.00	11%	476.07	33.00	545.00	14%	742.00	13%	0.98	56.00	10%	789.84	56.00
24.00	9%	383.76	24.00	798.00	20%	1173.00	21%	0.96	76.00	10%	1197.24	76.00
19.00	9%	322.81	19.00	270.00	7%	354.00	6%	0.99	22.00	8%	374.78	22.00

Εικόνα 16: Συμπληρωματικές στήλες

Οι παραπάνω στήλες, καθώς και η στήλη *E* θα επικολληθούν σε όλες τις καρτέλες εβδομαδιαίων αναφορών για τα 3 έτη για τα οποία υπάρχουν διαθέσιμα δεδομένα και από αυτές τις 27 συνολικά στήλες, θα πηγάζουν τα δεδομένα που θα εμφανίζονται στους δύο περιληπτικούς πίνακες που θα συνταχθούν.

7.3 Φιλτράρισμα των δεδομένων και περιγραφική στατιστική

Αφού τα δεδομένα συλλέχθηκαν, τροποποιήθηκαν και οργανώθηκαν, πλέον μπορούν να χρησιμοποιηθούν για περαιτέρω ανάλυση. Παραπάνω εμφανίζονται οι πηγές από τις οποίες έγινε η συλλογή των δεδομένων, και από αυτά τα δεδομένα κατά κύριο λόγο θα γίνει όλη η ανάλυση που αποτυπώνεται στο παρόν έγγραφο, τουλάχιστον όσο αναφορά την πρόβλεψη της ζήτησης. Η συλλογή των δεδομένων πολλές φορές αποτελεί μεγάλο άθλο, αφού παράλληλα με την κατανόηση του προβλήματος πρέπει να κατανοηθούν και τα διαθέσιμα δεδομένα, τα οποία δεν είναι πάντα εύκολα προσβάσιμα, ή μπορεί να υπάρχουν σε μορφή που απαιτεί ιδιαίτερη επεξεργασία για να μεταμορφωθεί κατάλληλα ώστε να είναι χρήσιμα. Αν και μπορεί να υπάρχει μια γενική ιδέα σχετικά με τη λύση ενός προβλήματος, ακόμα και αν καταναλωθεί σημαντικός χρόνος ώστε να διαμορφωθεί η δομημένη επίλυση, συνήθως στην πορεία εμφανίζονται εναλλακτικές μέθοδοι επίλυσης ή βελτιστοποίησης αυτής. Το ίδιο συμβαίνει με τα δεδομένα και τη μορφή τους· η οργάνωση τους συνήθως αλλάζει κατά τη διάρκεια της επίλυσης του προβλήματος, σε όποιο στάδιο και αν βρίσκεται. Κατά την εξερευνητική ανάλυση, εμφανίστηκαν διάφορα προβλήματα που χρειάστηκε να αντιμετωπιστούν – κάποια από τα οποία αναφέρθηκαν στο προηγούμενο κομμάτι της οργάνωσης δεδομένων. Με την αντιμετώπιση των προβλημάτων αυτών επισημάνθηκε πρόοδος στην εξερευνητική ανάλυση, η οποία μπορεί να αναπτύσσεται απεριόριστα μέχρι που ο όγκος των πληροφοριών στην οθόνη να είναι τόσο μεγάλος που θα καθιστά τον σκοπό της ανούσιο. Γι' αυτό το λόγο, η προκαταρκτική ανάλυση πρέπει να είναι δομημένη και να έχει συγκεκριμένο σκοπό και προδιαγραφές. Καταρχάς, θα αναπτυχθούν οι περιληπτικοί πίνακες που αναφέρθηκαν προηγουμένως. Στη συνέχεια θα δημιουργηθούν γραφήματα που θα αναπαριστούν οπτικά τη συμπεριφορά των προϊόντων σε ιστορική βάση.

7.3.1 Περιληπτικός πίνακας προϊόντος - Σύνταξη

Σκοπός του πίνακα προϊόντος είναι ότι η επισκόπηση του θα δώσει μια γενική πρώτη ιδέα στον παρατηρητή, δηλαδή την πρώτη όψη της συμπεριφοράς του κατά το παρελθόν. Ο τρόπος που παρουσιάζεται το προϊόν χρονολογικά είναι πολύ σημαντικό για την κατανόηση της συμπεριφοράς του. Η κεντρική ιδέα για τον συγκεκριμένο πίνακα είναι να διατηρηθεί η

δομή που έχουν οι επιχειρησιακές αναφορές της ASC αλλά τα πολλαπλά προϊόντα να αντικατασταθούν με τις χρονολογικές εκδοχές του επιλεγμένου προϊόντος.

ASIN (child):	B000YWLU0S	Draw from "NS Items"	1	B000YWLU0S										
Week #	Sheets (Dates)	Title	SKU	Sessions	Session %	Page Views	Page Views %	Buy Box %	Units Ordered	Unit Session %	Ordered Product Sales	Total Order Items		
Week 10-2014	BR - 2014 09-03	Musical Fairy Jewel 3N-ZBEG-		272	9.21%	418	9.76%	96.00%	22	8.09%	£351.78	22		
Week 11-2014	BR - 2014 16-03	Musical Fairy Jewel 3N-ZBEG-		256	8.42%	379	8.76%	96%	17	6.64%	£271.83	17		
Week 12-2014	BR - 2014 23-03	Musical Fairy Jewel 3N-ZBEG-		281	9.44%	404	9.70%	95.00%	22	7.83%	£351.78	22		
Week 13-2014	BR - 2014 30-03	Musical Fairy Jewel 3N-ZBEG-		251	9.52%	368	9.88%	96%	19	7.57%	£303.81	19		

Εικόνα 17: Εμφάνιση ενός προϊόντος ανά τις χρονολογικές εκδοχές του

Αρχικά, θα πρέπει να οριστεί το κελί που θα χρησιμοποιηθεί ως κριτήριο της εύρεσης, δηλαδή το ASIN του προϊόντος το οποίο θα παρουσιάζεται στον πίνακα αυτόν. Για τους σκοπούς της διαδικασίας που αναπτύσσεται, η πιο απλή μορφή του κριτηρίου θα είναι απλώς να οριστεί ένα κελί το οποίο θα αναφέρεται στις παρακάτω φόρμουλες και πάντα το επιθυμητό ASIN να εισάγεται στο προκαθορισμένο κελί. Ωστόσο για να μη χρειαστεί να γίνεται αναζήτηση στα υπόλοιπα συστήματα για να βρεθεί το ASIN για το ενδιαφερόμενο προϊόν, με την παρακάτω διαδικασία θα γίνει επιλογή του προϊόντος από τη λίστα προϊόντων, και το επιλεγμένο προϊόν θα περνάει στον πίνακα προϊόντος με μία απλή *VLOOKUP*. Όπως φαίνεται στην *εικόνα 17*, ως κριτήριο για την αναγνώριση του προϊόντος που επιλέχτηκε χρησιμοποιείται απλώς ο ακέραιος «1». Η λογική πίσω από το κλειδί αυτό είναι ότι η καρτέλα του πίνακα ο οποίος αναπτύσσεται, μπορεί απλώς να αντιγραφεί ώστε να υπάρχουν πολλαπλοί πίνακες προϊόντων. Οι πίνακες θα είναι πανομοιότυποι, με τη μόνη διαφορά το κλειδί που αναφέρεται στην καρτέλα με τη λίστα προϊόντων. Στην πορεία η συγκεκριμένη στρατηγική θα χρησιμοποιηθεί ώστε να αναπτυχθούν γραφήματα που εμφανίζουν ταυτόχρονα διαφορετικά προϊόντα. Όπως φαίνεται στην εικόνα παρακάτω, στον πίνακα που θα ολοκληρωθεί στη συνέχεια θα εμφανιστεί το μουσικό κουτί ST/5104/10/00, επειδή φέρει το κλειδί «1». Στον δεύτερο πίνακα θα εμφανιστεί το μουσικό κουτί ST/5174/20/00, το οποίο έχει τον αριθμό «2», κοκ.. Επίσης με αυτόν τον τρόπο είναι δυνατή η γρήγορη εναλλαγή μεταξύ των προκαθορισμένων προϊόντων εφόσον, αν το κλειδί στον πίνακα οριστεί ως «2» αντί για «1», θα εμφανιστεί το επόμενο προϊόν.

Description	Navision Old	ASIN	FBM SKU	FBA SKU	Product Ma	Individual Item
Fairy Dome Musical Chest	BD/2624/10/00	B00DYU56K1	BD/2624/10/00	BD/2624/10/00	Musical boxes	3
Ballet Musical Chest	BD/2821/10/00	B00S6AEB0Q		BD/2821/10/00	Musical boxes	4
Musical Fairy Jewellery Chest	ST/5104/10/00	B000YWL003	3N-ZBEG-M	ST/5104/10/00	Musical boxes	1
Musical Princess Jewellery Chest	ST/5174/20/00	B003LKUGQ2	ST/5174/20/00	22-RPKC-W9L	Musical boxes	2

Εικόνα 18: Επιλογή προϊόντων για την εμφάνιση τους στον πίνακα προϊόντος

Στη συνέχεια, χρειάζεται να γίνει μια απλή αντιγραφή και επικόλληση της πρώτης γραμμής μιας εβδομαδιαίας αναφοράς, η οποία περιέχει τους τίτλους της κάθε στήλης. Οι πρώτες δύο στήλες σχετικά με το ASIN θα αντικατασταθούν με το «*Week #*» ως αναγνωριστικό για την εβδομάδα την οποία αφορά η κάθε γραμμή και το «*Sheets(Dates)*», όπου θα επικολληθούν οι τίτλοι των καρτελών όπου αποθηκεύτηκαν οι εβδομαδιαίες αναφορές. Η δεύτερη στήλη με τις καρτέλες θα χρησιμοποιηθεί ως μέρος των ορίων για τη *VLOOKUP*, με την οποία θα γίνει η είσοδος των σχετικών δεδομένων του προϊόντος στον πίνακα. Οι τίτλοι των καρτελών θα χρειαστεί να επικολληθούν σε αυτήν τη στήλη ανά σειρά έτσι ώστε να χρησιμοποιηθεί για την εύρεση των δεδομένων. Η φόρμουλα που θα χρησιμοποιηθεί στις στήλες που αναφέρονται στα στοιχεία προϊόντος, αποτελείται από τρεις συναρτήσεις: *IFERROR(value,value_if_error)*, *VLOOKUP* και *INDIRECT(ref_text,[a1])*.

Ενδεικτικά: *IFERROR(VLOOKUP(\$D\$1,INDIRECT("'"&\$C19&"'!B:Q"),D\$2,FALSE),"*

Η *IFERROR*, που εμπεριέχει τις υπόλοιπες συναρτήσεις, χρησιμοποιείται για να αντικαταστήσει το οπτικά δυσάρεστο *#N/A* στην περίπτωση σφάλματος, με απλό κενό. Αυτή η προσέγγιση θα παρουσιαστεί πολλές φορές κατά την όλη διαδικασία. Η *VLOOKUP* χρησιμοποιεί την *INDIRECT* ώστε να κάνει τα όρια εύρεσης της δυναμικά. Με αυτόν τον τρόπο η φόρμουλα σε κάθε γραμμή ψάχνει τα όρια σε διαφορετική καρτέλα. Συγκεκριμένα, με την *INDIRECT*, η *VLOOKUP* για κάθε νέα γραμμή θα ψάξει να βρει την καρτέλα με όνομα όπως αυτό εμφανίζεται στην ίδια γραμμή της φόρμουλας, για τη στήλη «*Sheets(Dates)*». Εάν δε βρεθεί η καρτέλα, η γραμμή θα μείνει κενή. Εναλλακτικά, η φόρμουλα θα ψάξει εντός των προκαθορισμένων ορίων στην εξακριβωμένη καρτέλα και εφόσον βρει ταίρι για το δηλωμένο κριτήριο, θα επιστρέψει τη στήλη που ταιριάζει για το κάθε χαρακτηριστικό του πίνακα. Εδώ χρησιμοποιήθηκαν δείκτες σε κρυφή γραμμή (*δεύτερη γραμμή*) για να πραγματοποιηθεί η δυναμικότητα του *VLOOKUP* όσο αναφορά την επιστρεφόμενη στήλη. Στη φόρμουλα εμφανίζεται το κελί D2, το οποίο εμπεριέχει τον *ακέραιο* αριθμό «2», οπότε

από τα όρια αναζήτησης της *VLOOKUP*, θα επιστραφεί η δεύτερη στη σειρά στήλη, η οποία για το συγκεκριμένο παράδειγμα είναι η στήλη που φέρει τον τίτλο του προϊόντος. Παραλλαγές της συγκεκριμένης φόρμουλας θα χρησιμοποιηθούν για να συμπληρωθούν οι υπόλοιπες στήλες.

Στην περίπτωση που χρειαστεί να γίνει απόκρυψη ορισμένων γραμμών, π.χ. ώστε οι υψηλές τιμές των πωλήσεων πριν τα Χριστούγεννα να αφαιρεθούν ως ακραίες τιμές, θα χρειαστεί να γίνει μια ακόμα τροποποίηση της φόρμουλας. Σαφώς, η απόκρυψη είναι δυνατόν να γίνει χειροκίνητα με χρήση φίλτρων, αλλά σε αυτήν τη περίπτωση οι τιμές απλώς θα κρυφτούν οπτικά και αν ανήκουν στα όρια μιας συνάρτησης όπως τη *MAX(number1,[number2]...)*, θα συνεχίσουν να έχουν υπόσταση. Οπότε, για να υπάρξει η δυνατότητα να γίνει πραγματική αφαίρεση επιλεκτικών εκδοχών, η οποία θα είναι παράλληλα εύκολα ανατρέψιμη, θα χρησιμοποιηθεί μια νέα στήλη με τίτλο «Exclude». Η στήλη αυτή από προεπιλογή θα είναι κενή, αλλά θα φέρει μορφοποίηση με επιλεγμένη γραμματοσειρά τη *Webdings*. Με τη συγκεκριμένη γραμματοσειρά, εφόσον γίνει είσοδος του γράμματος «a» στη στήλη *Exclude*, αυτό θα πάρει τη μορφή ενός ✓. Οι γραμμές για τις οποίες η στήλη *Exclude* περιέχει τον χαρακτήρα ✓, αντί για δεδομένα από τις επιχειρηματικές αναφορές της ASC, θα εμφανίζουν απλώς ένα κενό. Για τη διαδικασία της απόκρυψης, θα γίνει χρήση της παρακάτω συνάρτησης:

IF(logical_test,[value_if_TRUE],[value_if_FALSE])

SKU	Sessions	Session %	Page Views	Page Views %	Buy Box %	Units Ordered	Unit Session %	Ordered Product Sales	Total Order Items	Exclude
3N-ZBEG-	272	9.21%	418	9.76%	96.00%	22	8.09%	€351.78	22	
3N-ZBEG-										✓
3N-ZBEG-										✓
3N-ZBEG-	251	9.52%	368	9.88%	96%	19	7.57%	€303.81	19	
3N-ZBEG-										

Formula: `=IF($O20="a","",IF(ISERROR(VLOOKUP(D1,INDIRECT("'"&$C20&"'!B:ZZ"),N$2,FALSE)),,"",(VLOOKUP(D1,INDIRECT("'"&$C20&"'!B:ZZ"),N$2,FALSE))))`

Εικόνα 19: Απόκρυψη εγγραφών – χρήση της στήλης *Exclude*

7.3.2 Περιληπτικός πίνακας προϊόντος – Στατιστική Σύνοψη

Εφόσον γίνει επαλήθευση ότι τα δεδομένα εμφανίζονται όπως πρέπει – ελέγχοντας ότι οι τιμές που εμφανίζονται στον περιληπτικό πίνακα είναι οι ίδιες με τις τιμές που εμφανίζονται για το προϊόν στην καρτέλα αναφοράς που υπάρχει για την κάθε εβδομάδα, στη συνέχεια θα χρησιμοποιηθούν κάποιες στατιστικές συναρτήσεις για να δομηθεί μία περίληψη της

απόδοσης του προϊόντος. Η περίληψη αυτή θα δομηθεί σε μορφή πίνακα και μπορεί να αποθηκευτεί οπουδήποτε. Εδώ επιλέχτηκε να γίνει η είσοδος της στο πάνω μέρος των δεδομένων του προϊόντος. Η στατιστική περίληψη αποτελείται από τις εξής συναρτήσεις:

- **Ελάχιστο** – $MIN(number1,[number2]...)$
- **Μέγιστο** – $MAX(number1,[number2]...)$
- **Πλήθος** – $COUNT(value1,[value2]...)$
- **Μέσος** – $AVERAGE(number1,[number2]...)$
- **Τυπική απόκλιση** – $STDEV.P(number1,[number2]...)$
- **Τεταρτημόριο** – $QUARTILE(array,quart[1,2 ή 3])$

Εκτός από τις παραπάνω συναρτήσεις, θα γίνει χρήση της $COUNTIF(range,criteria)$ σε φόρμουλες που θα εμφανίζουν το πλήθος των τιμών που εντάσσονται στις εξής κατηγορίες:

- Τιμές που είναι **μεγαλύτερες από δύο** θετικές τυπικές αποκλίσεις.
 $COUNTIF(F17:F500[range], ">=" & F7[mean] + F8[stdev]*2)$
- Τιμές που βρίσκονται **ανάμεσα σε μία με δύο** θετικές τυπικές αποκλίσεις.
 $COUNTIF(F17:F500, ">" & F7 + F8) - COUNTIF(F71:F192, ">=" & F7 + F8*2)$
- Τιμές που βρίσκονται **ανάμεσα σε μία με δύο** αρνητικές τυπικές αποκλίσεις.
 $COUNTIF(F17:F500, "<" & F7 - F8) - COUNTIF(F17:F500, "<=" & F7 - F8*2)$
- Τιμές που είναι **μικρότερες από δύο** αρνητικές τυπικές αποκλίσεις.
 $COUNTIF(F17:F500, "<=" & F7 - F8*2)$

Παρακάτω εμφανίζεται ο πίνακας με τις στατιστικές συναρτήσεις, οι τιμές του οποίου μπορούν να χρησιμοποιηθούν για να βγουν ανάλογα συμπεράσματα από το σύνολο δεδομένων. Το σύνολο δεδομένων πλέον αποτελείται από όλες τις χρονολογικές εκδοχές του προκαθορισμένου προϊόντος, όπως αυτό εμφανίζεται κάθε εβδομάδα. Η απόδοση του συνοψίζεται με τον στατιστικό πίνακα, όπου παρουσιάζονται τιμές που περιγράφουν το προϊόν ανά τα διαφορετικά χαρακτηριστικά του. Αρχικά, γίνεται αντιληπτό ότι το **μέγιστο** μπορεί να υπάρξει ακόμα και έξι φορές μεγαλύτερο από τον **μέσο**. Επίσης το **ελάχιστο** μπορεί να εμφανιστεί με ιδιαίτερα χαμηλές τιμές, αλλά είναι σημαντικό να κατανοηθεί το ότι έχουν υπάρξει περιπτώσεις εξάντλησης αποθεμάτων, όπου η επιχείρηση αδυνατούσε να ανταποκριθεί στη ζήτηση όσο ανέμενε την παραλαβή του νέου εμπορεύματος. Συνεπώς, αν

και οι τιμές που θα εμφανιστούν κοντά στο ελάχιστο είναι πραγματικές, δεν αντιπροσωπεύουν την πραγματική ζήτηση του προϊόντος. Το γεγονός ότι το πλήθος είναι το ίδιο σε όλες τις στήλες προμηνύει ότι το σύνολο δεδομένων μάλλον είναι αρτιμελές.

Τα τεταρτημόρια δίνουν μία πρώτη ματιά στην κατανομή των τιμών· για τα στατιστικά ανιχνευσιμότητας (πέντε πρώτες στήλες), ο μέσος βρίσκεται εντός του διαστήματος του δεύτερου τεταρτημόριου, δηλαδή ανάμεσα στο 50-75% των τιμών. Ωστόσο, συγκρίνοντας τις τιμές, ο μέσος είναι πλησιέστερος στο 75% των τιμών παρά στο 50%, γεγονός που υποδηλώνει ότι στο σύνολο πιθανώς υπάρχουν αρκετά υψηλές ακραίες τιμές που επηρεάζουν τον μέσο. Αυτό επιβεβαιώνεται από τα στατιστικά πωλήσεων (τελευταίες τέσσερις στήλες), όπου είναι προφανές ότι ο μέσος ανήκει στο ανώτερο 75% των τιμών. Εδώ προκύπτει μία ανισότητα μεταξύ των στατιστικών ανιχνευσιμότητας με των στατιστικών πωλήσεων, άρα είναι σαφές ότι για κάποιες εκδοχές του δείγματος, οι πωλήσεις δεν εξαρτώνται τόσο πολύ από την ανιχνευσιμότητα, και για αυτές τις περιπτώσεις, οι συνεδρίες και οι επισκέψεις σελίδας καταλήγουν σε περισσότερες πωλήσεις.

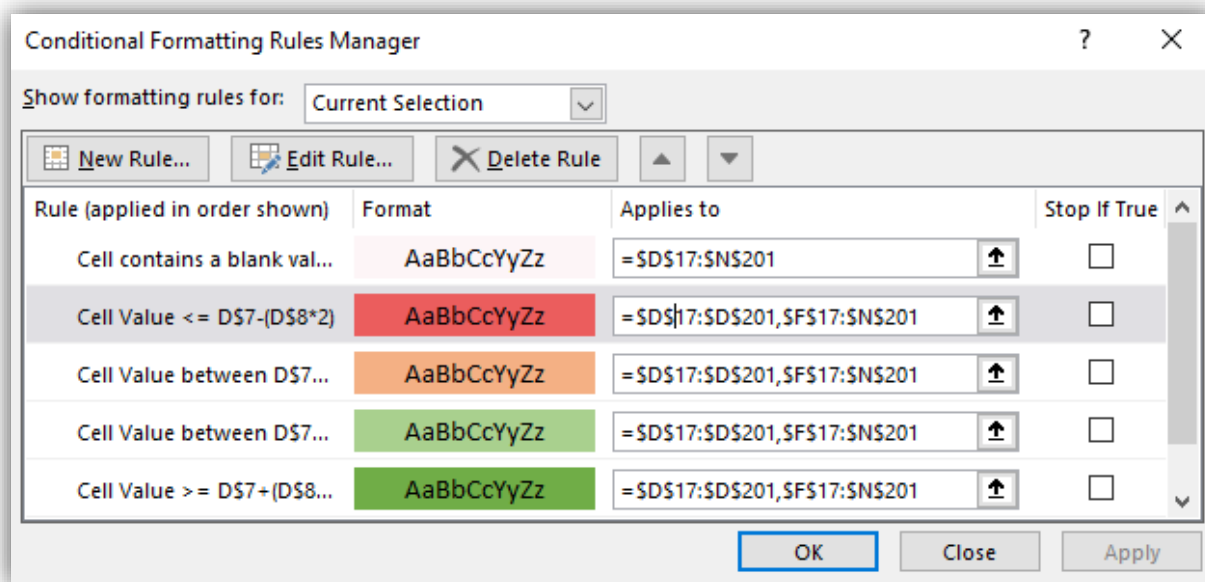
Τέλος, οι τελευταίες τέσσερις γραμμές που περιέχουν τους υπολογισμούς με τις τυπικές αποκλίσεις, δίνουν υπόσταση στις πιθανές ακραίες τιμές, εμφανίζοντας το πλήθος των τιμών που ανήκουν εντός των προκαθορισμένων διαστημάτων τυπικών αποκλίσεων. Συγκρίνοντας τον μέσο με την τυπική απόκλιση, είναι αναμενόμενο ότι δε θα εμφανιστούν τιμές στην τελευταία κατηγορία που αντιπροσωπεύει τιμές που είναι μικρότερες από δύο αρνητικές τυπικές αποκλίσεις, αφού τα στατιστικά δεν μπορούν να πάρουν αρνητικές τιμές. Αυτό δεν ισχύει στην περίπτωση που τα στατιστικά έχουν ποσοστιαία φύση, όπου όντως εμφανίζεται πλήθος, αφού δύο τυπικές αποκλίσεις μείον τον μέσο δε μηδενίζουν τα ποσοστά.

Sheets (Dates)	Title	SKU	Sessions	Session %	Page Views	Page Views %	Buy Box %	Units Ordered	Unit Session %	ered Product	stal Order Ite
MIN	=COUNT(F17:F500)		31.00	0.01	44.00	0.01	0.05	1.00	0.00	13.99	1.00
MAX			2638.00	0.20	3863.00	0.21	1.00	463.00	0.24	6870.45	445.00
COUNT			160.00	160.00	160.00	160.00	160.00	160.00	160.00	160.00	160.00
MEAN	//Excluding Off Stock		566.01	0.11	817.51	0.11	0.96	67.06	0.10	982.39	64.99
STDEV			477.00	0.03	702.08	0.03	0.13	86.36	0.04	1268.57	82.13
# of weeks that	Quartiles	1st: 25%	290.75	0.09	413.00	0.10	0.98	26.00	0.08	390.49	26.00
		2nd: 50%	377.00	0.11	538.00	0.11	0.99	39.00	0.10	565.11	38.00
		3rd: 75%	641.00	0.12	928.75	0.13	1.00	63.25	0.13	946.53	62.25
	Higher than +2 STDEV	11	4	10	5	0	9	7	9	9	
	Between +1 & +2 STDEV	13	19	11	17	0	8	20	7	8	
	Between -1 & -2 STDEV	1	11	1	11	2	0	15	0	0	
Lower than -2 STDEV	0	6	0	6	5	0	4	0	0		

Εικόνα 20: Πίνακας τιμών με στατιστικές συναρτήσεις

7.3.3 Περιληπτικός πίνακας προϊόντος – Μορφοποίηση υπό όρους

Όπως φαίνεται στην εικόνα 20, υπάρχουν πολυπληθές υποσύνολα του δείγματος δεδομένων που ανήκουν εντός των ορίων των υπολογισμών με τις τυπικές αποκλίσεις. Από τον πίνακα στατιστικής σύνοψης απορρέει το πλήθος των τιμών που ανήκουν στο κάθε υποσύνολο. Ποιες είναι όμως οι εβδομάδες που αντιστοιχούν για αυτές τις τιμές; Τι συμπεράσματα μπορούν να διαμορφωθούν από την παρατήρηση των εβδομάδων αυτών; Το Excel διαθέτει ένα χρήσιμο εργαλείο μορφοποίησης που καθιστά δυνατό να γίνει χρήση φόρμουλων για να μορφοποιηθούν κελιά υπό συγκεκριμένους όρους. Το εργαλείο λέγεται «Conditional Formatting», και διαθέτει τρόπους εισαγωγής όρων διαφορετικής φύσεως και πολλαπλές επιλογές μορφοποίησης. Για τον πίνακα προϊόντος θα χρησιμοποιηθεί μορφοποίηση υπό όρους στο σύνολο δεδομένων, για τα στατιστικά απόδοσης του προϊόντος. Οι όροι θα τεθούν με φόρμουλες σχεδόν πανομοιότυπες με τις τελευταίες τέσσερις γραμμές του πίνακα στατιστικής σύνοψης και θα συμπεριφέρονται με τον ίδιο τρόπο. Αντί να προσθέτουν τις τιμές που πληρούν τους όρους και να διαμορφώνουν ένα πλήθος για το κάθε υποσύνολο, θα χρωματίζουν τα κατάλληλα κελιά ώστε αυτά να αντιστοιχούνται στο υποσύνολο για τους κατάλληλους όρους. Με τη χρήση των αντίστοιχων χρωμάτων, θα είναι ξεκάθαρο αν μια εβδομάδα φέρει τιμή που ανήκει σε μία από τις προκαθορισμένες κατηγορίες.



Εικόνα 21: Μορφοποίηση υπό όρους – Εφαρμογή

Παρατηρώντας τα δεδομένα με τη νέα τους μορφή, είναι ενδιαφέρον να σημειωθεί ότι εφόσον ένα κελί φέρει μορφοποίηση που το κατατάσσει στις πράσινες ή κόκκινες κατηγορίες, σχεδόν πάντα θα υπάρχουν και άλλα κελιά στην ίδια γραμμή, ακόμα και κάποια σε γειτονικές γραμμές, που θα ανήκουν στην ίδια κατηγορία. Συσσωρευμένες πράσινες τιμές εμφανίζονται πολύ συχνότερα από ότι οι κόκκινες. Σημειώνοντας τις εβδομάδες που αντιπροσωπεύονται από τις πράσινες τιμές, είναι αισθητό το φαινόμενο της εποχικότητας κατά τις τελευταίες εβδομάδες του κάθε έτους. Αυτό ήταν αναμενόμενο, αφού οι εβδομάδες αυτές συμπίπτουν με τα Χριστούγεννα. Στην περίπτωση εξάντλησης αποθεμάτων, κατά κανόνα εμφανίζονται μία με δύο εβδομάδες πριν την εξάντληση και μετά από την επανακυκλοφόρηση του προϊόντος, για τις οποίες οι τιμές είναι πολύ μειωμένες και ανήκουν στην πορτοκαλί ή κόκκινη κατηγορία. Αν το προϊόν εξαντληθεί εντελώς για μερικές εβδομάδες, δε θα εμφανιστεί η καταχώρηση του στις εβδομαδιαίες πωλήσεις της ASC, οπότε δε θα υπάρχουν διαθέσιμες τιμές για το προϊόν αυτό στον περιληπτικό πίνακα προϊόντος. Ωστόσο, αν υπάρχουν μερικά τελευταία αποθέματα που πωληθούν μέσω FBM, τότε θα υπάρξουν δεδομένα για τις σχετικές εβδομάδες, αλλά κατά πάσα πιθανότητα οι τιμές αυτές θα ανήκουν στην κόκκινη κατηγορία.

Για παράδειγμα, όπως εμφανίζεται στην *εικόνα 22*, κατά τις εβδομάδες 42, 43 και 44 του έτους 2016, τα αποθέματα είχαν σχεδόν εξαντληθεί. Εφόσον ημερολογιακά οι εβδομάδες εξάντλησης έπεσαν πολύ κοντά στην περίοδο των Χριστουγέννων, διαπιστώνεται ότι μάλλον είχε καταχωρηθεί παραγγελία για νέο εμπόρευμα αρκετές εβδομάδες νωρίτερα, και κατά σύμπτωση τα αποθέματα εξαντλήθηκαν λίγες μέρες πριν τη νέα παραλαβή. Είτε η ημερομηνία τελικής παραλαβής ήταν προγραμματισμένη και προκαθορισμένη είτε όχι, είναι σίγουρα σαφές ότι για τις εβδομάδες 42 έως 44 υπήρξε σημαντική μείωση πωλήσεων και εσόδων. Τι συμβαίνει όμως με την εβδομάδα 45; Σίγουρα δεν ανήκει στην κόκκινη ή στην πορτοκαλί κατηγορία, αλλά στην περίπτωση που δεν είχε συμβεί η προηγούμενη εξάντληση αποθεμάτων, θα μπορούσε η εβδομάδα 45 να ανήκει σε μία από τις πράσινες κατηγορίες; Και τι θα γινόταν αν η παραλαβή του νέου εμπορεύματος υπέστη καθυστέρηση για δέκα μέρες;

Week #	Sessions	Session %	Page Views	Page Views %	Buy Box %	Units Ordered	Unit Session %	Ordered Product Sales	Total Order Items
Week 42-2016	274	5.50%	362	5.48%	78%	25	9.12%	£374.75	23
Week 43-2016	150	2.72%	180	2.47%	9%	1	0.67%	£13.99	1
Week 44-2016	130	2.44%	170	2.40%	91%	18	13.85%	£269.82	16
Week 45-2016	728	9.87%	1004	10.15%	98%	62	8.52%	£929.38	60
Week 46-2016	1063	10.83%	1449	11.07%	100%	165	15.52%	£2,368.35	153
Week 47-2016	1485	13.91%	2108	14.60%	100%	276	18.59%	£3,999.24	266
Week 48-2016	1863	16.92%	2662	17.81%	100%	372	19.97%	£5,512.28	352
Week 49-2016	1985	15.18%	2952	16.14%	100%	455	22.92%	£6,870.45	403
Week 50-2016	1848	16.55%	2618	16.90%	100%	440	23.81%	£6,595.60	425
Week 51-2016	948	12.89%	1340	13.11%	100%	162	17.09%	£2,428.38	161

Εικόνα 22: Μορφοποίηση υπό όρους – Αποτελέσματα

Τα παραπάνω ερωτήματα θα μείνουν αναπάντητα, αλλά από τη συγκεκριμένη εικόνα γίνεται αντιληπτό το πόσο σοβαρό και ζημιογόνο μπορεί να αποδειχτεί ένα λάθος του προγραμματισμού παραγγελιών εμπορεύματος, ή μια καθυστέρηση της παραλαβής του, ακόμα και αν οι παράγοντες είναι εξωγενής. Για μία επιχείρηση που φέρει το φαινόμενο της εποχικότητας τόσο έντονα όσο η FG, η απόδοση της κατά την κρίσιμη περίοδο της κορύφωσης της ζήτησης μπορεί να σημάνει την επιβίωση της, καθώς και η αδυναμία της να ανταποκριθεί άμεσα στη ζήτηση αυτή μπορεί να αποτελέσει τη χρεωκοπία της.

7.3.4 Περιληπτικός πίνακας κατηγορίας

Όπως ο πίνακας προϊόντος εξιστορεί κάποια στοιχεία της απόδοσης του προϊόντος που μελετάται, έτσι και ένας περιληπτικός πίνακας κατηγορίας θα εμφανίζει τα ίδια χαρακτηριστικά, αλλά αθροιστικά για όλα τα προϊόντα που υπάγονται σε αυτή. Στο κεφάλαιο [9.2](#), κατά την οργάνωση των δεδομένων, καταγράφηκε η συλλογή δεδομένων περί κατηγορίας προϊόντος και η εισαγωγή κατάλληλων στηλών εντός των εβδομαδιαίων επιχειρηματικών αναφορών που, με χρήση της συνάρτησης *COUNTIF* οργάνωναν τα δεδομένα των μεμονωμένων προϊόντων σε αθροίσματα αντιπροσωπευτικά για την κατηγορία στην οποία τα προϊόντα ανήκαν. Ο περιληπτικός πίνακας προϊόντος έχει ήδη συνταχθεί και μορφοποιηθεί κατάλληλα, και όπως παρουσιάστηκε στο προηγούμενο κεφάλαιο, χρησιμοποιεί τη συνάρτηση *VLOOKUP* για να αναπαράγει τα δεδομένα για την κάθε εβδομάδα. Η επιστρεφόμενη στήλη από τα όρια που δηλώνονται εντός της *VLOOKUP* καθορίζεται από τη δεύτερη γραμμή της καρτέλας, η οποία φέρει ακέραιους αριθμούς που αντιπροσωπεύουν τη θέση της στήλης που πρέπει να επιστραφεί.

Εφόσον οι αθροιστικές στήλες στις αναφορές ακολουθούν τη δομή των αρχικών στηλών, η εναλλαγή μεταξύ της εμφάνισης των δεδομένων μεμονωμένου προϊόντος ή κατηγορίας εξαρτάται απόλυτα από τον ακέραιο αριθμό στη δεύτερη γραμμή. Συνεπώς, θα αντιγραφεί ολόκληρη η καρτέλα που περιέχει τον πίνακα προϊόντος, θα μετονομαστεί ως πίνακας κατηγορίας, θα διαγραφούν οι στήλες αναγνωριστικού προϊόντος, δηλαδή εκείνες που φέρουν τον τίτλο και το SKU, και επιπλέον, θα γίνει η επεξεργασία των ακεραίων στη δεύτερη γραμμή ώστε να εμφανίζουν τις κατάλληλες στήλες. Τέλος, θα τροποποιηθούν ελαφρώς οι *VLOOKUP* που επιστρέφουν τα δεδομένα, όσο αναφορά το εύρος εύρεσης τους, διότι πλέον δεν αναζητούν ASIN αλλά κατηγορίες. Στο κελί όπου στον πίνακα προϊόντος έμπαινε ως είσοδος το ASIN του ενδιαφερόμενου προϊόντος, στον πίνακα κατηγορίας ως κριτήριο θα μπαίνει το όνομα της κατηγορίας.

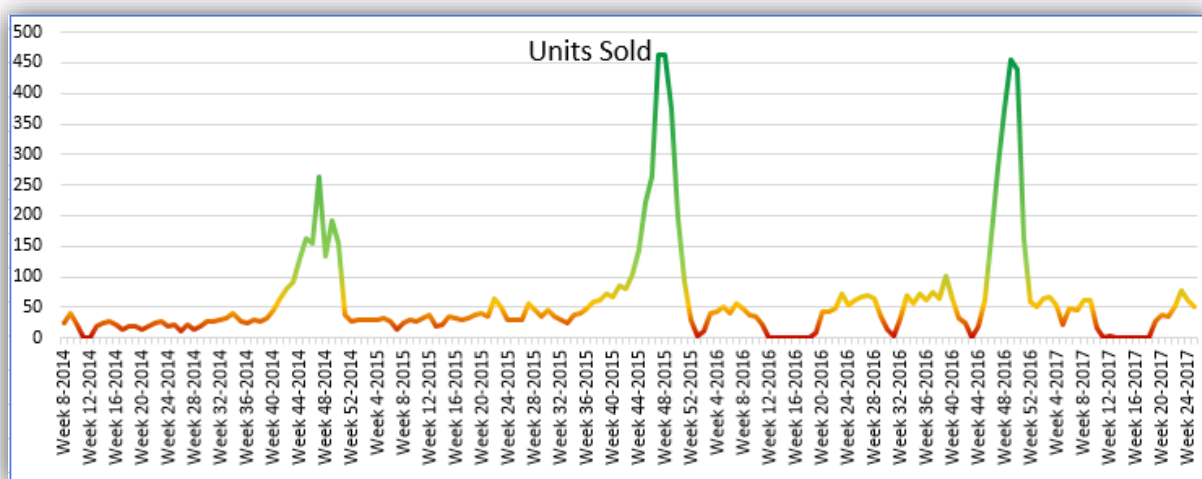
Επειδή προφανώς το πλήθος των διαφορετικών κατηγοριών είναι πολύ μικρότερο από το πλήθος όλων των προϊόντων – ενεργών ή ανενεργών, για να διευκολυνθεί η διαδικασία εναλλαγής μεταξύ των διαθέσιμων κατηγοριών, θα δημιουργηθεί μια λίστα από την οποία η επιλεγόμενη κατηγορία θα χρησιμοποιηθεί ως κριτήριο για τις φόρμουλες του πίνακα. Ο ορισμός της λίστας είναι εύκολη υπόθεση: καταρχάς, πρέπει να οργανωθούν οι τιμές της λίστας, δηλαδή να συμπληρωθούν όλες οι πιθανές κατηγορίες στις οποίες μπορεί να ανήκει ένα προϊόν. Αφού τα ονόματα των κατηγοριών εισαχθούν σε μία στήλη μιας καρτέλας χωρίς κενά (επιθυμητό αλλά όχι απαραίτητο), με τη χρήση του εργαλείου εγκυρότητας δεδομένων (*Data Validation*), θα οριστούν τα όρια στα οποία εμφανίζονται οι τιμές της λίστας και έτσι, το κελί για το οποίο πραγματοποιήθηκε η εγκυρότητα δεδομένων θα έχει τη μορφή μιας λίστας. Πλέον η εναλλαγή μεταξύ των διαθέσιμων κατηγοριών είναι εύκολη και γρήγορη και τα δεδομένα καθώς και η στατιστική σύνοψη αυτών θα εμφανίζονται με τον ίδιο ακριβώς τρόπο όπως στον πίνακα προϊόντος.

Category	Musical boxes					
Column #:	Musical boxes	15	16	17	18	19
Sheets (Dates)	Stationery	Sessions	Session %	Page Views	Page Views %	Buy Box %
MIN	Bedroom	106.00	0.05	145.00	0.05	0.71
MAX	Play	7183.00	0.58	10280.00	0.60	1.00
COUNT	Plush	174.00	174.00	174.00	174.00	174.00
MEAN	Accessories	1402.94	0.27	1981.33	0.28	0.98
	Cosmetics					
	Dresses					
	77 Εκπτώσεις Out Stock					
Week #	Sheets (Dates)	Sessions	Session %	Page Views	Page Views %	Buy Box %
Week 45-2016	06/11/2016 BR - 2016 06-11 to 12-11	2752	37.32%	3777	38.19%	9
Week 46-2016	13/11/2016 BR - 2016 13-11 to 19-11	3910	39.85%	5254	40.15%	9

Εικόνα 23: Εμφάνιση μιας κατηγορίας ανά τις χρονολογικές εκδοχές της

7.4 Απεικόνιση δεδομένων – γραφήματα

Παραπάνω συντάχθηκαν περιληπτικοί πίνακες δεδομένων μεμονωμένου προϊόντος και κατηγορίας. Οι δύο πίνακες αυτοί εμπεριέχουν στατιστικές περιλήψεις που βοηθούν στη διεξαγωγή συμπερασμάτων για τη συμπεριφορά των δεδομένων και την απόδοση του προϊόντος ή της κατηγορίας που παρατηρείται. Για περαιτέρω παρατήρηση και μελέτη συσχέτισης των στατιστικών αποτελεσμάτων με τα πραγματικά δεδομένα, χρησιμοποιήθηκε μορφοποίηση υπό όρους που σκιαγραφεί τα κελιά τα οποία υπάγονται σε προκαθορισμένα υποσύνολα. Από αυτή την πράξη επιβεβαιώθηκε η σημαντική εμφάνιση του φαινομένου της εποχικότητας για το συγκεκριμένο προϊόν. Ένα ακόμα σημαντικό εργαλείο για τη διαδικασία ανάλυσης δεδομένων είναι τα γραφήματα. Τα γραφήματα απεικονίζουν οπτικά τα δεδομένα και αποτελούν τον ευκολότερο τρόπο ανίχνευσης ακραίων τιμών από το σύνολο δεδομένων. Πριν ξεκινήσει μία διαδικασία πρόβλεψης, είναι πολύ σημαντικό να γίνονται ορισμένα μηνύματα αντιληπτά από την αρχή. Με τη μελέτη του πίνακα προϊόντος ή κατηγορίας και την παρατήρηση των γραφημάτων, πρέπει να γίνεται κατανοητή η συμπεριφορά της ζήτησης του προϊόντος. Πρέπει να μπορεί να διαχωρίζεται η πραγματική περίπτωση εμφάνισης μειωμένης ζήτησης από την περίπτωση εξάντλησης αποθεμάτων, κάτι που μπορεί να συμβεί σε απροσδόκητες χρονικές περιόδους και με ακανόνιστη διάρκεια, ή με οποιοδήποτε άλλο σενάριο θα μπορούσε να αντιμετωπίζεται από μία επιχείρηση όπως η FG.



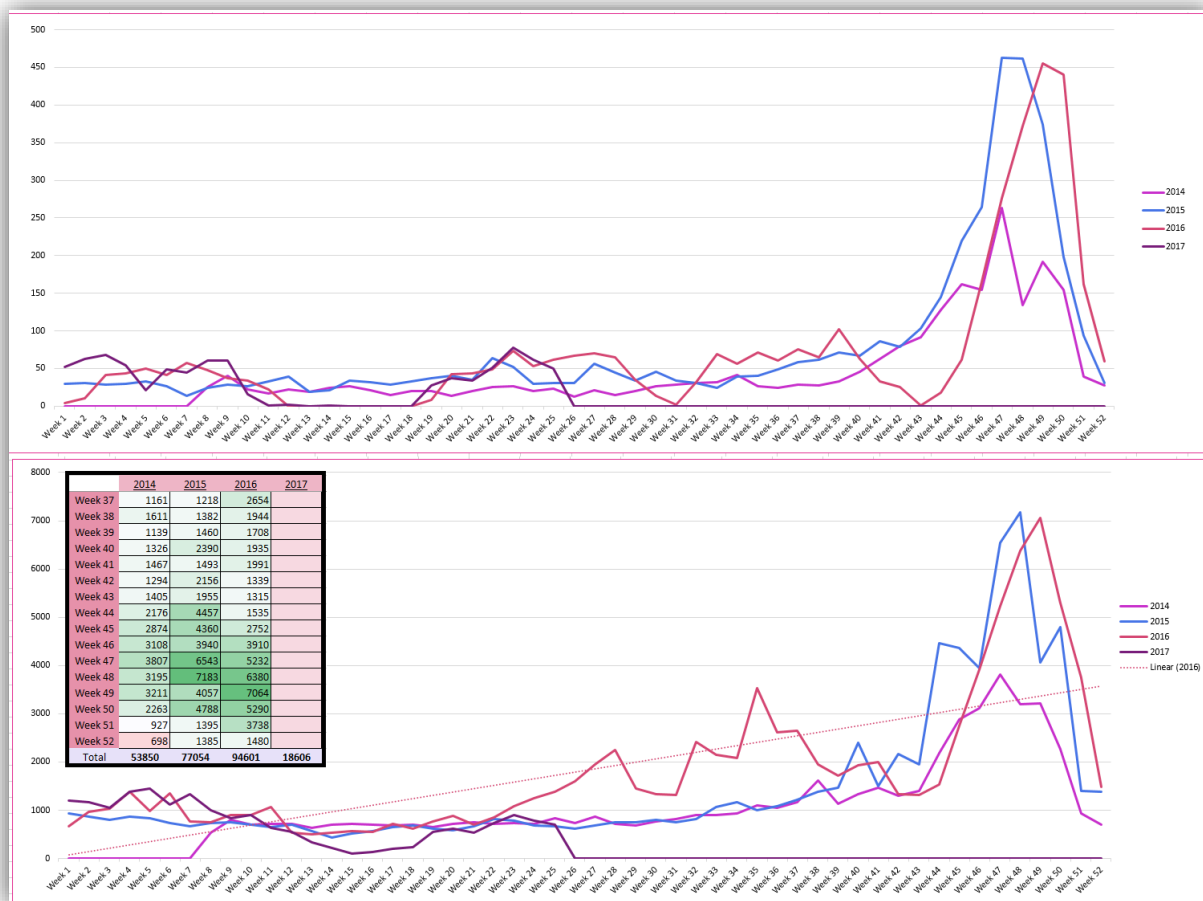
Εικόνα 24: Πωλήσεις τεμαχίων για ένα προϊόν για τα έτη 2014 έως 2017 ανά εβδομάδα

Το παραπάνω γράφημα μπορεί να μοιάζει περισσότερο με ένα ακανόνιστο καρδιογράφημα παρά με τις πωλήσεις τεμαχίων παιχνιδιού, αλλά στην πραγματικότητα αντικατοπτρίζει την παράξενη συμπεριφορά της αγοράς στη βιομηχανία παιχνιδιών, που πηγάζουν από ξέφρενους γονείς που τρέχουν να αγοράσουν Χριστουγεννιάτικα δώρα για τα παιδιά τους και τα παιδιά συγγενών και φίλων. Για πολλά από τα προϊόντα της FG, τα διαγράμματα πωλήσεων έχουν παρόμοιες συμπεριφορές, με παραλλαγές στις ποσότητες που πουλήθηκαν. Σχεδόν όλα τα είδη εμφανίζουν κορύφωση που ξεκινάει από την εβδομάδα 40 και ακμάζει κατά τις δέκα τελευταίες εβδομάδες του έτους. Τα περισσότερα είδη εμφανίζουν νεκρές περιόδους, και για τα περισσότερα από αυτά ευθύνεται κάποια εξάντληση αποθεμάτων. Η ζήτηση κατά την τελευταία εβδομάδα του έτους (μετά τα Χριστούγεννα) πέφτει κατακόρυφα, και συνήθως χρειάζονται τρεις με τέσσερις εβδομάδες ώστε να αρχίσει να επιστρέφει σε κανονικούς ρυθμούς, με την πρώτη εβδομάδα του έτους να σημειώνει συχνά τις χαμηλότερες πωλήσεις ετησίως. Κάποια προϊόντα μπορεί να εμφανίσουν απότομη άνοδο των πωλήσεων σε συγκεκριμένες περιόδους, όπως γραφικά είδη που ακμάζουν σε πωλήσεις πριν την αρχή του σχολικού έτους, στολές και αξεσουάρ όπως φτερά και στέμματα λίγο πριν το Halloween ή τη Βρετανική εθνική ημέρα του βιβλίου.

Για τη διεξαγωγή των παραπάνω συμπερασμάτων πρέπει να γίνει προσεκτική παρατήρηση των γραφημάτων. Το φαινόμενο της εποχικότητας φέρει το χαρακτηριστικό της επανάληψης: η άνοδος που παρατηρείται στις πωλήσεις πριν τα Χριστούγεννα αναμένεται ότι θα επαναληφθεί κάθε χρόνο για όσο το προϊόν συνεχίζει να είναι διαθέσιμο, χωρίς εξαίρεση, ακόμα και αν υπάρχουν σημαντικές ποσοστιαίες διαφορές. Για να γίνει ευκολότερα η σύγκριση των πωλήσεων μιας συγκεκριμένης εβδομάδας με την ίδια εβδομάδα των προηγούμενων ετών, πρέπει να αναπτυχθεί ένα γράφημα με μπάρες που να δέχεται ως είσοδο τα διαφορετικά έτη ως ένα υποσύνολο το καθένα. Για 4 έτη δηλαδή θα εισαχθούν από 52 παρατηρήσεις που θα παρουσιαστούν στο γράφημα, ακόμα και αν δεν υπάρχουν ακριβώς 52 παρατηρήσεις διαθέσιμες για το κάθε έτος. Για να γίνει αυτό εύκολα είναι δυνατόν οι είσοδοι του διαγράμματος να απορρέουν κατευθείαν από τους περιληπτικούς πίνακες. Με τη συγκεκριμένη προσέγγιση όμως την επόμενη εβδομάδα που θα εισαχθούν νέα δεδομένα στους περιληπτικούς πίνακες, αν χρειαστεί εκείνα να συμπεριληφθούν στα διαγράμματα, τότε θα πρέπει να επαναπροσδιοριστούν τα όρια δεδομένων που αποτελούν το γράφημα.

Ένας τρόπος για να αντιμετωπιστεί το πρόβλημα αυτό ώστε να μην απαιτείται χρόνος στη συντήρηση των γραφημάτων, είναι να οργανωθούν τα δεδομένα από τους περιληπτικούς

πίνακες σε ένα νέο πίνακα, όπου περιέχει το στοχευμένο χαρακτηριστικό που θα παρουσιάζεται στο γράφημα π.χ. τεμάχια ή πωλήσεις. Ο πίνακας αυτός θα αποτελείται από 52+1 γραμμές μία για την κάθε εβδομάδα του έτους και η πρώτη για τους τίτλους της κάθε στήλης και 4+1 στήλες, μία για το κάθε έτος και η πρώτη για τον ορισμό της εβδομάδας που αντιστοιχεί σε κάθε γραμμή. Εναλλακτικά, γίνεται ο αριθμός των γραμμών και στηλών να οριστεί ως το αντίστροφο, δημιουργώντας έτσι έναν πλατύ πίνακα αντί για έναν μακρύ, αλλά αυτό εξαρτάται απόλυτα από την προτίμηση του καθενός. Είτε αυτός ο νέος πίνακας εισαχθεί σε νέα καρτέλα είτε στην καρτέλα όπου βρίσκεται ο περιληπτικός πίνακας, θα χρειαστεί να γίνει χρήση της συνάρτησης *CONCAT(text1, ...)* ώστε να διαμορφωθεί το κατάλληλο κριτήριο για τις *VLOOKUP* που θα επιστρέφουν τα δεδομένα. Το κριτήριο θα αποτελείται από μία μίξη των τίτλων των γραμμών και στηλών του νέου πίνακα, δηλαδή τις εβδομάδες και τα έτη, και θα διαμορφώνουν συγκεκριμένα δομημένη φράση, π.χ. «Week 1-2014».

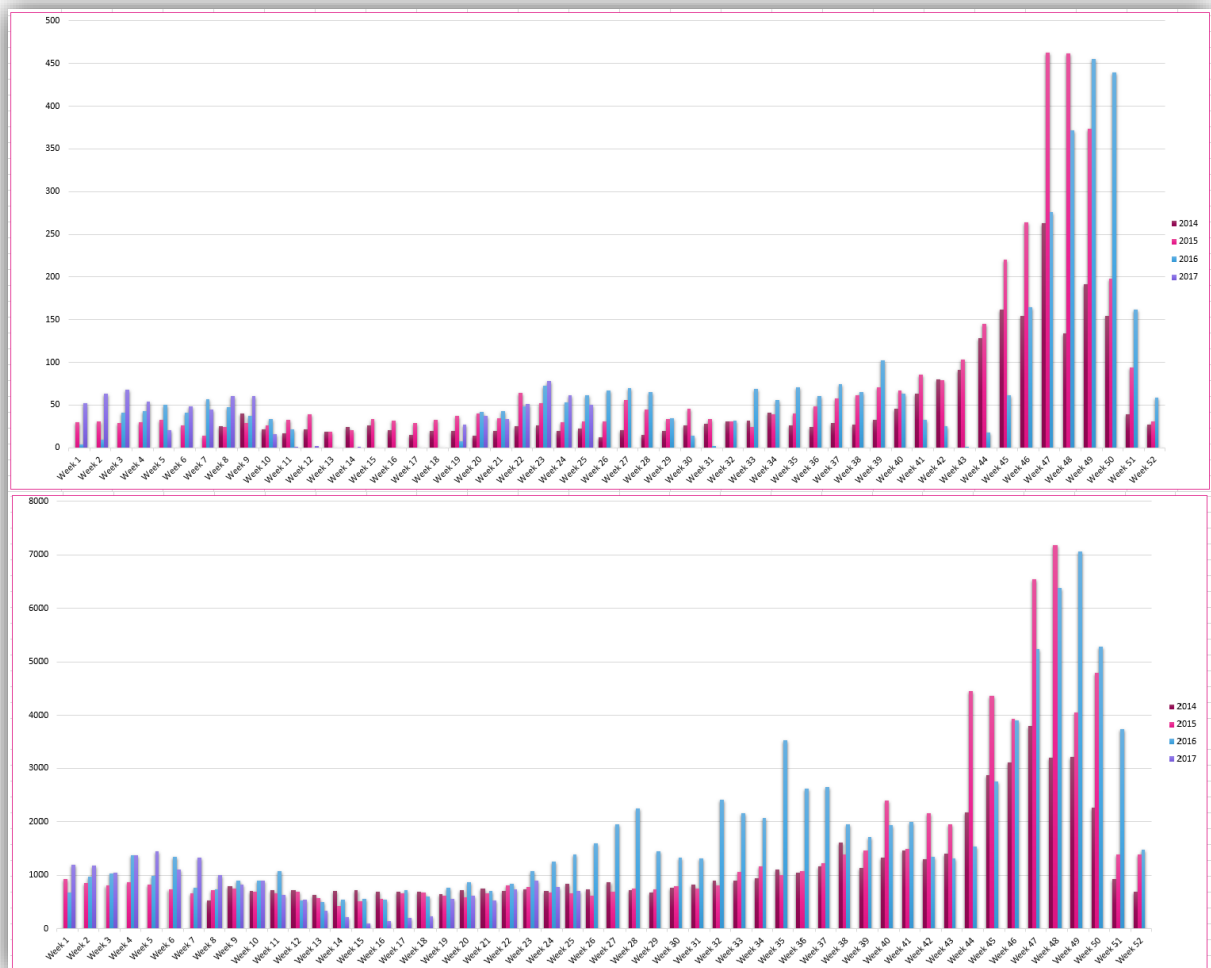


Εικόνα 25: Γραμμικά γραφήματα (*line charts*) προϊόντος (πάνω) και κατηγορίας (κάτω) + πίνακας δεδομένων

Στην *εικόνα 25* παρουσιάζονται δύο γραμμικά γραφήματα, το πρώτο για το νούμερο ένα σε πωλήσεις μουσικό κουτί και το δεύτερο για ολόκληρη την κατηγορία των μουσικών κουτιών. Επειδή οι τιμές της κατηγορίας επηρεάζονται σημαντικά από το πρωτεύον προϊόν της, καθώς και τα υπόλοιπα προϊόντα που υπάγονται σε αυτήν φέρουν παρόμοια μοτίβα όσο αναφορά την εποχικότητα, ή ακόμα και τις περιόδους εξάντλησης αποθεμάτων – όλα τα μουσικά κουτιά κατασκευάζονται στο ίδιο εργοστάσιο, ήταν αναμενόμενο ότι τα δύο αυτά γραφήματα θα μοιάζουν μεταξύ τους. Η σημαντική διαφορά βρίσκεται στον άξονα Υ, που παρουσιάζει τις ποσότητες τεμαχίων που πουλήθηκαν. Σαφώς, οι πωλήσεις του ενός προϊόντος θα είναι κατά κανόνα μικρότερες ή σπανίως ίσες με τις πωλήσεις της κατηγορίας στην οποία ανήκει.

Τα παραπάνω γραμμικά γραφήματα όμως μπορεί να προκαλέσουν μια μικρή σύγχυση, αφού παρουσιάζουν τέσσερις χρονολογικές σειρές οι οποίες συχνά συμπίπτουν και περιπλέκονται. Ακόμα και η ύπαρξη του λογότυπου (*legend*) του γραφήματος που εξηγεί ποια γραμμή αντιπροσωπεύει το κάθε έτος, δε βοηθάει ιδιαίτερα στην εύκολη σύγκριση των χρονολογικών σειρών. Τα γραφήματα με μπάρες θα εμφανίσουν τα έτη με σειρά που μπορεί να καθοριστεί εύκολα, και παρατηρώντας τα μπορεί να γίνει σύγκριση μεταξύ των ετών για κάθε εβδομάδα. Ορίζοντας την εμφάνιση του πιο παλιού έτους να είναι στα αριστερά και του πιο πρόσφατου έτους στα δεξιά, θα δείξει την πρόοδο των πωλήσεων διαχρονικά για κάθε εβδομάδα. Η παρατήρηση μιας αύξουσας κλίσης από τα αριστερά στα δεξιά σημαίνει ότι το προϊόν ή η κατηγορία προϊόντος σημειώνει κάθε χρόνο περισσότερες πωλήσεις από το προηγούμενο. Αν προσδιορισθεί μία χρονική περίοδος όπου όλα τα έτη σημειώνουν μειωμένες πωλήσεις τότε σημαίνει ότι η ζήτηση για αυτή την περίοδο είναι γενικά χαμηλότερη και για αυτό δεν ευθύνεται κάποια εξάντληση αποθεμάτων ή οποιοδήποτε άλλο εμπόδιο. Αυτό βέβαια δεν είναι απόλυτο· αν αντί για χαμηλές τιμές κάποια έτη εμφανίζουν μηδενικές τιμές για τη συγκεκριμένη περίοδο τότε μάλλον τίθεται θέμα μίξης μειωμένης ζήτησης παράλληλα με εξάντληση αποθεμάτων. Αυτό θα παρατηρηθεί κατά τις εβδομάδες 12 με 18, όπου τα αποθέματα εξαντλήθηκαν κατά τα έτη 2016 και 2017, αλλά για τα έτη 2014 και 2015 εμφανίστηκαν πωλήσεις ελαφρώς μειωμένες. Φυσικά, πρέπει επίσης να ληφθεί υπόψη η ανοδική πορεία της ζήτησης κατά τα χρόνια αυτά, οπότε οι πωλήσεις που παρουσιάζονται για τα έτη 2014 και 2015, αν και θεωρούνται χαμηλές για τα δεδομένα του έτους 2017, δε συνεπάγεται ότι θα χαρακτηρίζονταν το ίδιο χαμηλές για τα δεδομένα του 2014.

Επιπλέον, μελετώντας τις τελευταίες μπάρες των γραφημάτων, δηλαδή τις εβδομάδες που εμφανίζουν υψηλή εποχικότητα, γίνεται αντιληπτό ότι η περίοδος των αυξημένων πωλήσεων έχει μεταβληθεί ελαφρώς ανά τα τελευταία χρόνια. Για τα έτη 2014 και 2015, σημειώνεται μία άνοδος των πωλήσεων που ξεκινάει από την εβδομάδα 41, ακμάζει κατά την εβδομάδα 47 και ξεκινάει την πτώση της κατά την εβδομάδα 50. Για το 2017 δεν υπάρχουν διαθέσιμα δεδομένα για τις συγκεκριμένες εβδομάδες, αλλά για το 2016 παρατηρείται άνοδος των πωλήσεων στην εβδομάδα 46, δηλαδή ένα μήνα αργότερα. Σε αυτό σίγουρα συμβάλλει το γεγονός ότι τα αποθέματα εξαντλήθηκαν κατά τη 43^η εβδομάδα, οπότε χρειάστηκαν μία με δύο εβδομάδες ώστε το προϊόν να διεκδικήσει ξανά την ανταγωνιστική του θέση. Ενδιαφέρον αποτελεί το ότι η ακμή των πωλήσεων για αυτό το έτος παρουσιάζεται κατά τη 49^η εβδομάδα, και η πτώση της συμβαίνει κατά την 51^η. Από ότι φαίνεται, η ζήτηση κατά το 2016 παρουσίασε μία καθυστέρηση δύο εβδομάδων σε σύγκριση με τα προηγούμενα χρόνια.

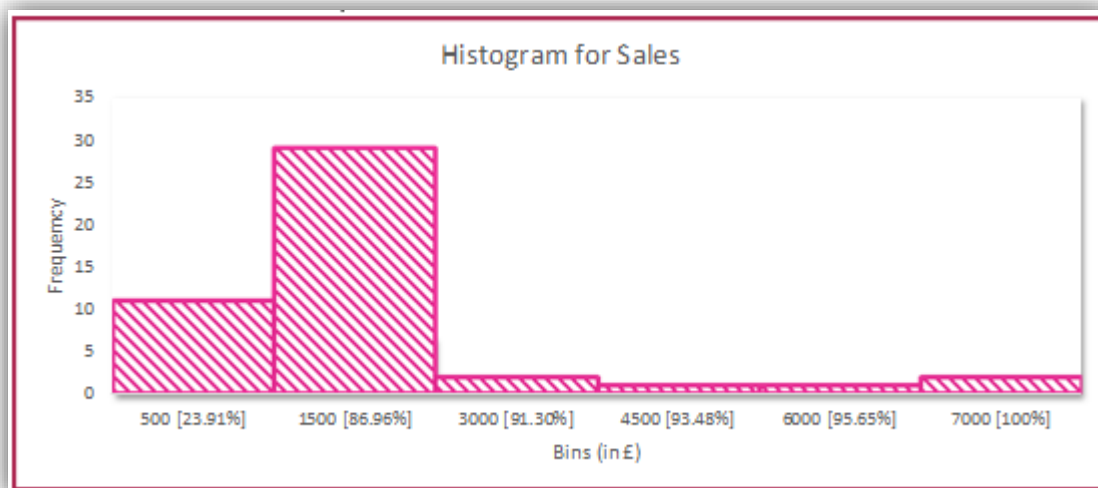


Εικόνα 26: Γραφήματα με μπάρες (*bar charts*) προϊόντος (πάνω) και κατηγορίας (κάτω)

Οι παραπάνω παρατηρήσεις οφείλονται σε διαφορετικούς παράγοντες· η απότομη άνοδος των πωλήσεων κατά το 2016 που παρουσιάζει την περίοδο εποχικότητας ελαφρώς στενότερη από τα προηγούμενα έτη οφείλεται κυρίως σε σύντομη εξάντληση αποθεμάτων. Η μεταβολή δύο εβδομάδων που παρατηρείται κατά για το έτος 2016 μπορεί να πηγάζει από διάφορους εξωγενείς παράγοντες. Ένας από τους παράγοντες θα μπορούσε να είναι η μειωμένη αγοραστική δύναμη ή αποφασιστικότητα εξαιτίας των προβλημάτων που αντιμετωπίζει η οικονομία της Μεγάλης Βρετανίας, και της πτώσης της Αγγλικής Λίρας μετά το δημοψήφισμα του BREXIT. Δεύτερος παράγοντας μάλλον είναι το ότι κατά το 2014 στο Ηνωμένο Βασίλειο και σε άλλες Ευρωπαϊκές χώρες παρουσιάστηκε η Αμερικάνικη μόδα του Black Friday, όπου κατά την τέταρτη Παρασκευή του Νοεμβρίου και μετά την ημέρα των Ευχαριστιών, οι αγορές λανσάρουν τεράστιες εκπτώσεις και προσφορές κάτι που για τις ΗΠΑ σημαίνει την αφετηρία των Χριστουγεννιάτικων αγορών. Η μόδα αυτή φαίνεται να έχει πλέον ξεπεραστεί για το Αγγλικό κοινό, που εκπαιδεύτηκε κατά τα χρόνια να περιμένει υπομονετικά για διάφορες προσφορές και εκπτώσεις κατά όλη τη διάρκεια του έτους, οπότε συμπερασματικά και από την παρατήρηση των πωλήσεων του 2016, η Black Friday δε συγκινεί πλέον την Αγγλική ζήτηση. Τέλος, ακόμα και η ημέρα της εβδομάδας που τυχαίνει να συμπίπτει με την ημέρα των Χριστουγέννων παίζει ρόλο στη συμπεριφορά του καταναλωτή. Πραγματικά, οι λόγοι που μπορεί να επηρεάσουν τη ζήτηση είναι πολλοί και συνήθως εξωγενείς, αφού ακόμα και το αν η ημέρα Χριστουγέννων είναι μεσοβδόμαδα ή πέφτει στο Σαββατοκύριακο, επηρεάζει την κρίση και τη βιασύνη του καταναλωτή για να κάνει τις αγορές του.

Παρόμοια ανάλυση μπορεί να γίνει από την παρατήρηση διαφορετικών προϊόντων της κατηγορίας μουσικών κουτιών, και από τη σύγκριση άλλων προϊόντων μεταξύ τους και με τις αντίστοιχες κατηγορίες τους. Διαφορετικά διαγράμματα διατίθενται ως απάντηση για διαφορετικές σκοπιές μελέτης των δεδομένων. Μία ενδιαφέρουσα οπτική γωνία σχετική με την κατανομή των πωλήσεων είναι προσβάσιμη με χρήση ιστογραμμάτων. Το ιστόγραμμα αντικατοπτρίζει την ομαδοποίηση των πιθανών τιμών σε ίσες κλάσεις και το ύψος των κλάσεων εμφανίζει την ισότητα ή την ανισότητα της κατανομής των τιμών του δείγματος. Για παράδειγμα, η εικόνα 27 εμφανίζει ιστόγραμμα που δημιουργήθηκε με βάση τις πωλήσεις για τις πρόσφατες 52 εβδομάδες. Παρουσιάζεται ότι οι πρώτες δύο κλάσεις οι οποίες περιέχουν 11 και 29 εβδομάδες αντίστοιχα, δηλαδή το 86,96% από τις εβδομάδες που συμπεριλαμβάνονται, σημειώσανε πωλήσεις από £0 to £1500. Οι 6 εβδομάδες που προηγήθηκαν των Χριστουγέννων επιστρέψανε >£1500 την εβδομάδα, αποτελώντας έτσι τον

εντυπωσιακό αριθμό 50,16% των συνολικών πωλήσεων του συγκεκριμένου προϊόντος για τις προηγούμενες 52 εβδομάδες.



Εικόνα 27: Ιστόγραμμα κατανομής πωλήσεων

Η χρήση γραφημάτων είναι πολύτιμο εργαλείο κατά τη διαδικασία ανάλυσης δεδομένων και υπάρχουν απίστευτες δυνατότητες ανάπτυξης γραφημάτων διαθέσιμες από το ίδιο το Excel, από άλλα εμπορικά ή δωρεάν λογισμικά ή ακόμα και από λογισμικά ανοιχτού κώδικα. Από απλά γραφήματα με γραμμές και μπάρες, σε μίξεις γραφημάτων που παρουσιάζουν και τα δύο είδη, σε τρίτους άξονες, pie charts, χαρτογραφήσεις και ιστογράμματα, εφόσον τα δεδομένα οργανωθούν σωστά, οι δυνατότητες απεικόνισης τους είναι πραγματικά τεράστιες.

7.5 Απεικόνιση δεδομένων και αυτοματισμός – Δυναμικό Ταμπλό

Υπάρχουν διάφοροι τρόποι για την οργάνωση και την απεικόνιση ενός συνόλου δεδομένων στο Excel. Στα πλαίσια μιας χρονοβόρας διαδικασίας έρευνας και ανάλυσης είναι σημαντικό να λαμβάνεται πάντα υπόψη ο αυτοματισμός και να εφαρμόζεται όπου είναι δυνατόν, ώστε να μην καταναλώνεται πολύτιμος χρόνος για την περιττή επανάληψη διαδικασιών που θα μπορούσαν να έχουν αυτοματοποιηθεί από την αρχή. Παραπάνω παρουσιάζονται γραφήματα δύο ειδών: γραμμικά γραφήματα και γραφήματα με μπάρες. Ως γραμμικά γραφήματα παρουσιάστηκε αρχικά ένα γράφημα που εμφάνιζε τις πωλήσεις ενός προϊόντος διαχρονικά με όρια τη χρονική απόσταση του συνόλου δεδομένων. Στη συνέχεια εμφανίστηκε ένα

παρόμοιο γράφημα όπου ο άξονας X αποτελούνταν από καθαρά 52 παρατηρήσεις· όσες εβδομάδες υπάρχουν σε ένα έτος. Δόθηκε έμφαση στα γραφήματα με μπάρες και επεξηγήθηκε γιατί στην περίπτωση πολλαπλών χρονολογικών σειρών είναι βέλτιστη η χρήση διαγραμμάτων με μπάρες. Το επόμενο βήμα είναι να παρουσιαστούν διαφορετικές εκδόσεις γραφημάτων για τα διαφορετικά χαρακτηριστικά που περιγράφουν την απόδοση του προϊόντος, ή της κατηγορίας, όπως οι συνεδρίες και οι επισκέψεις σελίδας.

Καταρχάς, όπως και για τα προηγούμενα διαγράμματα που παρουσιάστηκαν, θα χρειαστεί να συνταχθούν νέοι πίνακες ώστε να οργανωθούν τα δεδομένα με τρόπο που θα επιτρέψει την επιθυμητή απεικόνιση τους. Οι πίνακες και τα παραγόμενα διαγράμματα θα αποτελούν κομμάτια ενός δυναμικού ταμπλό, το οποίο ένας παρατηρητής μπορεί να χρησιμοποιήσει ώστε να κάνει σύγκριση χαρακτηριστικών απόδοσης μεταξύ προϊόντων και κατηγοριών. Ενδεικτικά, ως χαρακτηριστικά πωλήσεων θα συμπεριληφθούν οι χρηματικές αποδόσεις και οι πωλήσεις τεμαχίων και ως χαρακτηριστικά ανιχνευσιμότητας, οι συνεδρίες και οι επισκέψεις σελίδων. Συνολικά θα υπάρχουν για τον κάθε πίνακα 4 στήλες συν τη στήλη που θα περιέχει την εβδομάδα στην οποία τα δεδομένα αντιστοιχούν, η οποία θα χρησιμοποιηθεί ως το κριτήριο εύρεσης για τις *VLOOKUP*. Για να περιοριστεί το μέγεθος των πινάκων αλλά και των διαγραμμάτων, από τις 52 εβδομάδες θα εμφανίζονται μόνο οι 15 πιο πρόσφατες. Θα επιτρέπεται στον χρήστη να επιλέξει τον αριθμό των παρατηρήσεων που επιθυμεί από μία λίστα που περιέχει τους ακεραίους αριθμούς από 1 έως 15 και ανάλογα με τον αριθμό που επιλεγθεί στη λίστα, τόσες θα είναι και οι παρατηρήσεις που θα εμφανίζονται εντός των πινάκων, με τις πιο παλιές να αντικαθίστανται με κενές γραμμές.

Τα δεδομένα που εμφανίζονται στους αριστερούς πίνακες στην *εικόνα 28* πηγάζουν από την αρχική καρτέλα που περιέχει τον περιληπτικό πίνακα προϊόντος. Τα δεδομένα είναι ήδη οργανωμένα εκεί, οπότε μια απλή *VLOOKUP* τραβάει τα δεδομένα που αντιστοιχούν στις κατάλληλες εβδομάδες. Οι άλλες δύο σειρές με πίνακες από τα δεξιά της παραπάνω εικόνας τραβάνε τα δεδομένα με τον ίδιο τρόπο από διπλότυπα του αρχικού περιληπτικού πίνακα προϊόντος, για τα οποία έχει οριστεί διαφορετικό κλειδί αναζήτησης ώστε να αντιπροσωπεύουν διαφορετικά προϊόντα. Αυτά περί της οργάνωσης των δεδομένων· το χαρακτηριστικό που θα μετατρέψει τους παραπάνω πίνακες σε αυτόματους θα είναι η δυνατότητα τους να εμφανίζουν πάντα τη 15^η γραμμή για το κάθε πινακάκι ως την περασμένη εβδομάδα.

Product: Pink Fairy & Butterfly Kids Musical Jewellery Box -...					Product: Princess Kids Musical Jewellery Box - Glittery Kids...					Product: Enchanted Fairy Tale Kids Musical Jewellery Box - G...				
Musical boxes					Musical boxes					Musical boxes				
Item Code: ST/5104/10/00					Item Code: ST/5174/20/00					Item Code: BD/2901/00/00				
Last 15 weeks	Sales	Units Sold	Sessions	Page Views	Last 15 weeks	Sales	Units Sold	Sessions	Page Views	Last 15 weeks	Sales	Units Sold	Sessions	Page Views
Week 50-2016	£6,595.60	440	1848	2618	Week 50-2016	£1,916.63	137	830	1187	Week 50-2016	£4,459.36	314	1051	1501
Week 51-2016	£2,428.38	162	948	1340	Week 51-2016	£1,454.96	104	447	636	Week 51-2016	£2,579.22	178	554	824
Week 52-2016	£884.41	59	470	631	Week 52-2016	£237.83	17	169	219	Week 52-2016	£347.76	24	210	282
Week 1-2017	£782.48	52	487	684	Week 1-2017	£181.87	13	154	229	Week 1-2017	£446.19	31	241	352
Week 2-2017	£944.37	63	463	647	Week 2-2017	£475.66	34	251	352	Week 2-2017	£681.03	47	248	353
Week 3-2017	£1,019.32	68	422	579	Week 3-2017	£517.63	37	222	312	Week 3-2017	£623.07	43	199	305
Week 4-2017	£809.46	54	416	611	Week 4-2017	£175.89	11	133	179	Week 4-2017	£521.64	36	205	280
Week 5-2017	£333.79	21	394	541	Week 5-2017	£173.88	12	156	226	Week 5-2017	£677.54	46	228	321
Week 6-2017	£734.51	49	353	497	Week 6-2017	£118.42	8	163	217	Week 6-2017	£344.77	23	172	240
Week 7-2017	£674.55	45	374	546	Week 7-2017	£284.81	19	161	220	Week 7-2017	£389.74	26	174	245
Week 8-2017	£892.89	61	399	595	Week 8-2017	£134.91	9	168	223	Week 8-2017	£494.67	33	179	270
Week 9-2017	£952.99	61	392	558	Week 9-2017	£123.92	8	95	122	Week 9-2017	£584.61	39	217	303
Week 10-2017	£255.84	16	271	384	Week 10-2017					Week 10-2017	£584.61	39	202	298
Week 11-2017	£15.99	1	291	406	Week 11-2017					Week 11-2017	£674.55	45	218	308
Week 12-2017	£31.98	2	260	378	Week 12-2017					Week 12-2017	£595.61	39	197	288
	£10,760.96	714	5940	8397		£3,879.78	272	2119	2935		£9,545.01	649	3244	4669
One year ago...	Sales	Units Sold	Sessions	Page Views	One year ago...	Sales	Units Sold	Sessions	Page Views	One year ago...	Sales	Units Sold	Sessions	Page Views
Week 50-2015	£3,155.52	198	1521	2190	Week 50-2015	£1,135.19	81	546	736	Week 50-2015				
Week 51-2015	£1,426.30	94	639	893	Week 51-2015	£447.68	32	230	307	Week 51-2015				
Week 52-2015	£468.42	31	408	567	Week 52-2015	£195.86	14	157	190	Week 52-2015				
Week 1-2016	£63.96	4	98	132	Week 1-2016	£209.85	15	181	240	Week 1-2016				
Week 2-2016	£150.90	10	223	321	Week 2-2016	£335.76	24	163	233	Week 2-2016				
Week 3-2016	£614.59	41	425	595	Week 3-2016	£141.90	10	174	224	Week 3-2016				
Week 4-2016	£644.57	43	449	631	Week 4-2016	£139.90	10	124	162	Week 4-2016				
Week 5-2016	£750.50	50	447	623	Week 5-2016	£139.90	10	127	177	Week 5-2016				
Week 6-2016	£614.59	41	402	578	Week 6-2016	£69.95	5	95	119	Week 6-2016				
Week 7-2016	£854.43	57	376	535	Week 7-2016	£55.96	4	87	109	Week 7-2016				
Week 8-2016	£754.52	48	358	492	Week 8-2016	£125.91	9	95	122	Week 8-2016				
Week 9-2016	£591.63	37	379	505	Week 9-2016	£153.89	11	150	188	Week 9-2016				
Week 10-2016	£543.66	34	399	570	Week 10-2016	£167.88	12	138	188	Week 10-2016				
Week 11-2016	£351.78	22	268	376	Week 11-2016	£251.82	18	135	183	Week 11-2016				
Week 12-2016					Week 12-2016	£335.76	24	170	250	Week 12-2016				
	£7,829.85	512	4871	6818		£2,772.02	198	2026	2692		£1.00	0	0	0
Two years ago...	Sales	Units Sold	Sessions	Page Views	Two years ago...	Sales	Units Sold	Sessions	Page Views	Two years ago...	Sales	Units Sold	Sessions	Page Views
Week 50-2014					Week 50-2014					Week 50-2014				

Εικόνα 28: Ταμπλό και απεικόνιση δεδομένων – Δυναμικοί πίνακες

Με τη χρήση των συναρτήσεων $WEEKNUM(serial_number,[return_type])$, $YEAR(serial_number)$, $INT(number)$ και $NOW()$, θα αναπτυχθεί μία φόρμουλα που θα χρησιμοποιεί την τωρινή ώρα και ημερομηνία για να βρει την ημερομηνία της πιο πρόσφατης Κυριακής ώστε να διαμορφώσει την περιγραφή της προηγούμενης εβδομάδας σχετικά με τη θέση της στο έτος. Όπως και παραπάνω, θα χρησιμοποιηθεί η εξής δομή: π.χ. «Week 1-2014». Παράλληλα, γίνεται έλεγχος σφαλμάτων ώστε στην περίπτωση που συμπίπτει η αλλαγή του έτους, να ξεκινήσει η αρίθμηση των εβδομάδων από την αρχή, καθώς και να πραγματοποιηθεί αύξηση του έτους κατά μία μονάδα. Επιπλέον χρησιμοποιείται κρυφή στήλη στην οποία ελέγχεται η τιμή της λίστας που επιτρέπει την επιλογή αριθμού προβολής εβδομάδων, ώστε στην περίπτωση που έχουν επιλεγεί λιγότερο από 15 εβδομάδες, να αλλάζει η αρίθμηση της κρυφής στήλης A και ανάλογα με τις τιμές εκείνης, η φόρμουλα της αρίθμησης εβδομάδων να αποκρύπτει τις πρώτες παρατηρήσεις από το σημείο που η κρυφή στήλη μηδενιστεί και πάνω. Με αυτόν τον τρόπο, τα πινακάκια δεδομένων του ταμπλό θα είναι πλέον δυναμικά και δε θα απαιτείται καθόλου χρόνος για την ενημέρωσή τους, αρκεί να γίνεται τακτική ενημέρωση των επιχειρηματικών αναφορών από την ASC.

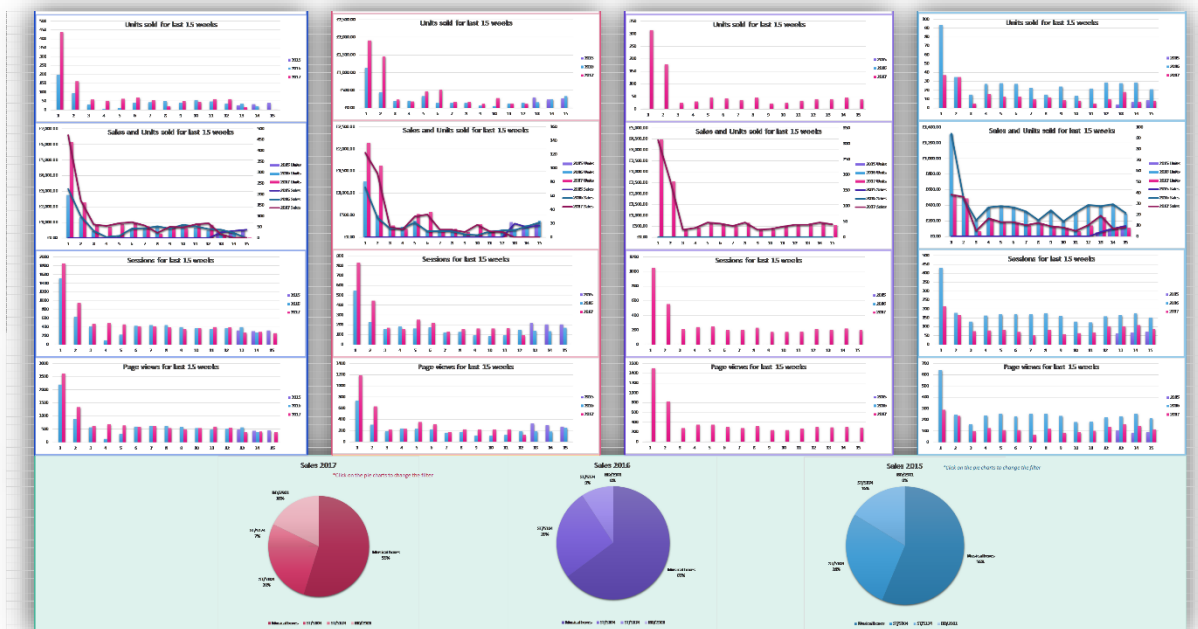
12	Week 14-2016	£636.54	46	536	700
13	Week 15-2016	£779.44	56	563	744
14	Week 16-2016	£866.37	63	550	735
15	Week 17-2016	£1,115.21	79	720	965
	Week 18-2016	£1,215.13	87	608	837
	Week 19-2016	£1,320.04	96	772	1080
	Week 20-2016	£1,491.94	106	874	1217
	Week 21-2016	£1,208.13	87	706	974
	Week 22-2016	£1,251.10	90	846	1129
	Week 23-2016	£1,658.82	118	1078	1497
	Week 24-2016	£1,497.45	105	1251	1657
	Week 25-2016	£1,774.25	125	1389	2051
	Week 26-2016	£1,993.65	135	1600	2228
	Week 27-2016	£2,236.43	157	1946	2750
	Week 28-2016	£1,980.62	138	2240	3046
		£21,025.12	1488	15679	21610

Εικόνα 29: Ταμπλό και απεικόνιση δεδομένων – Δυναμικοί πίνακες – Αυτοματοποίηση

Παραπάνω εμφανίζεται ένα πινακάκι κατηγορίας για τις πιο πρόσφατες εβδομάδες του έτους 2016· σε σύγκριση με την ημερομηνία που τραβήχτηκε το συγκεκριμένο απόκομμα οθόνης. Εδώ τα δεδομένα πηγάζουν από τον περιληπτικό πίνακα κατηγορίας και οι *VLOOKUP* μεταβλήθηκαν ελαφρώς ώστε να επιστρέφουν τις σωστές στήλες. Μόνο μία εκδοχή από τον πίνακα κατηγορίας διατηρήθηκε, εφόσον στο ίδιο ταμπλό εμφανίζονται τέσσερα πινακάκια τα οποία μπορεί να ανήκουν στην ίδια κατηγορία, αλλά μπορεί και όχι. Η ιδέα είναι να επιτραπεί η σύγκριση των προϊόντων που θα έχουν προεπιλεχθεί να προβληθούν στο ταμπλό, και σε διαγράμματα να γίνει όχι μόνο η σύγκριση των τεσσάρων προϊόντων μεταξύ τους αλλά και η σύγκριση με την κατηγορία στην οποία ανήκουν, με την προϋπόθεση ότι όλα ανήκουν στην ίδια.

Αναφορικά με τα διαγράμματα, με τη νέα οργάνωση των δεδομένων στο ταμπλό είναι δυνατόν να αναπτυχθούν διάφορα γραφήματα. Ο σκοπός του ταμπλό είναι να επιτραπεί η σύγκριση διαφορετικών προϊόντων, οπότε θα αναπτυχθούν διπλότυπα για οποιοδήποτε διάγραμμα χρησιμοποιηθεί, ώστε να αντιπροσωπεύονται όλα τα προϊόντα που εμφανίζονται στο ταμπλό. Κατά κύριο λόγο θα χρησιμοποιηθούν γραφήματα με μπάρες, που θα παρουσιάζουν τα δεδομένα για τα διαφορετικά έτη με αύξουσα χρονολογικά σειρά. Υπάρχουν τέσσερα διαθέσιμα χαρακτηριστικά αποδόσεων στο κάθε πινακάκι, οπότε θα

αναπτυχθούν τουλάχιστον από τέσσερα γραφήματα για κάθε προϊόν. Ωστόσο, επειδή οι πωλήσεις σε χρηματικές μονάδες και οι πωλήσεις τεμαχίων θα έχουν σχεδόν πανομοιότυπη απεικόνιση, με κύρια διαφορά τις τιμές του άξονα Υ, θα χρησιμοποιηθεί ένα συνδυαστικό γράφημα με τρίτο άξονα όπου θα αντιπροσωπεύονται οι τιμές και σε χρηματικές μονάδες αλλά και σε τεμάχια. Τέλος, θα αναπτυχθούν γραφήματα πίτας – ένα για το κάθε έτος, όπου θα παρουσιάζονται και τα τέσσερα προϊόντα παράλληλα με την κατηγορία τους. Τα pie charts επίσης εμπεριέχουν δεδομένα για όλα τα χαρακτηριστικά απόδοσης: με την αλλαγή του φίλτρου τους επιτρέπεται η εναλλαγή μεταξύ *Units Sold*, *Sales*, *Sessions* και *Page Views*.



Εικόνα 30: Ταμπλό και απεικόνιση δεδομένων – Γραφήματα

7.6 Πολυμεταβλητή Ανάλυση

Τα δεδομένα που καλείται να αναλύσει ένας ερευνητής προκύπτουν συχνά να είναι πολυμεταβλητής φύσεως, δηλαδή υπάρχει ένα μεγάλο πλήθος από μεταβλητές που περιέχουν πληροφορίες. Στην περίπτωση της FG, η παρουσία των πολλαπλών μεταβλητών είναι και ο λόγος που η ανάλυση γίνεται πάνω σε δεδομένα της ASC και όχι απλώς με τη χρήση των εγγραφών παραγγελιών και πωλήσεων από το NetSuite. Με σκοπό λοιπόν να επωφεληθεί η συγκεκριμένη μελέτη από την πολυμεταβλητότητα των δεδομένων, θα πραγματοποιηθεί πολυμεταβλητή ανάλυση, η οποία αποτελεί βασικό εργαλείο για τη μελέτη δεδομένων αυτού

του είδους καθώς προσφέρει τη δυνατότητα για απόκτηση γνώσης και ερμηνείας περί των σχέσεων μεταξύ των μεταβλητών.

Ένα από αυτά τα εργαλεία της πολυμεταβλητής ανάλυσης είναι ο συντελεστής συσχετίσεων r (Pearson). Ο συντελεστής συσχέτισης μετράει τη δύναμη και την κατεύθυνση της γραμμικής σχέσης μεταξύ δύο μεταβλητών, όπως αυτές εμφανίζονται σε ένα διάγραμμα διασποράς (scatter plot). Ο δειγματικός συντελεστής συσχέτισης αποτελείται από τη διαίρεση της δειγματικής συνδιασποράς ή αλλιώς συνδιακύμανσης (covariance) μεταξύ των δύο μεταβλητών· η μελέτη γίνεται σε ζευγάρια μεταβλητών κάθε φορά, προς το γινόμενο των τετραγωνικών ριζών της δειγματικής διασποράς ή διακύμανσης s της κάθε μεταβλητής. Εφαρμόζοντας τον τύπο σε όλα τα ζευγάρια που μπορούν να προκύψουν από της μεταβλητές, παράγεται ο πίνακας συσχετίσεων.

Οι τιμές του συντελεστή συσχέτισης θα ανήκουν πάντα εντός του διαστήματος -1 με $+1$, και η ερμηνεία τους πηγάζει από τις παρακάτω κατηγορίες:

- Ακριβώς -1 ή $+1$ – Τέλεια αρνητική ή θετική γραμμική συσχέτιση.
- $-0,7$ ή $+0,7$ – Ισχυρή αρνητική ή θετική γραμμική συσχέτιση.
- $-0,5$ ή $+0,5$ – Μέτρια αρνητική ή θετική γραμμική συσχέτιση.
- $-0,3$ ή $+0,3$ – Αδύναμη αρνητική ή θετική γραμμική συσχέτιση.
- 0 – Δεν υπάρχει αρνητική ούτε θετική γραμμική συσχέτιση.

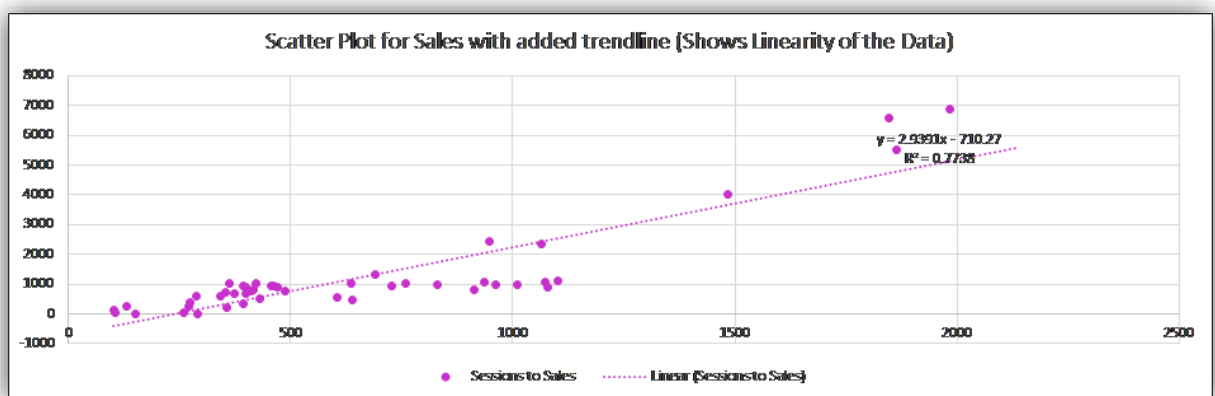
Με τη γνώση των παραπάνω ορίων και με την εφαρμογή τους εντός του εργαλείου μορφοποίησης υπό όρους, παράγεται ο μορφοποιημένος πίνακας συσχετίσεων για όλες τις μεταβλητές που προέρχονται από τις εβδομαδιαίες επιχειρηματικές αναφορές.

Correlation	Sessions	Session %	Page Views	Page Views %	Buy Box %	Units Ordered	Unit Session %	Total Order Item	Sales
Sessions	1.0000								
Session %	0.6744	1.0000							
Page Views	0.9935	0.6771	1.0000						
Page Views %	0.6712	0.9962	0.6773	1.0000					
Buy Box %	0.3103	0.3615	0.3090	0.3675	1.0000				
Units Ordered	0.8802	0.5331	0.8576	0.5290	0.2625	1.0000			
Unit Session %	0.4558	0.4083	0.4324	0.4118	0.5138	0.7150	1.0000		
Total Order Item	0.8816	0.5395	0.8580	0.5345	0.2675	0.9987	0.7168	1.0000	
Sales	0.8797	0.5311	0.8573	0.5265	0.2595	0.9995	0.7098	0.9984	1.0000

Weak Correlation
Moderate Correlation
Strong Correlation
Perfect Correlation

Εικόνα 31: Πίνακας συσχετίσεων

Ο παραπάνω πίνακας συσχετίσεων δημιουργήθηκε με το εργαλείο *Correlation* από τη δωρεάν επέκταση του Excel με όνομα *Data Analysis*. Θέτοντας τις κατάλληλες τιμές στο παράθυρο του εργαλείου, επιστρέφεται ο παραπάνω πίνακας χωρίς την εμφανιζόμενη μορφοποίηση, ο οποίος φέρει απλές τιμές και όχι συναρτήσεις. Αυτό σημαίνει ότι για την ενημέρωση του πίνακα η διαδικασία θα πρέπει να εφαρμοστεί από την αρχή. Εναλλακτικά, η αυτοματοποίηση του πίνακα συσχετίσεων είναι δυνατή μέσω χρήσης της συνάρτησης *CORREL(array1, array2)*. Για το κάθε ζεύγος μεταβλητών γίνεται επίσης να αναπτυχθεί το αντίστοιχο διάγραμμα διασποράς.



Εικόνα 32: Διάγραμμα διασποράς (συνεδρίες και πωλήσεις)

Η χρήση των διαγραμμάτων διασποράς είναι σημαντική για την επαλήθευση του συντελεστή γραμμικής συσχέτισης, εφόσον από το διάγραμμα διασποράς δεν εμφανίζεται καθόλου η εικόνα γραμμικότητας των δεδομένων, τότε ο συντελεστής συσχέτισης δεν έχει στατιστική σημασία. Στην *εικόνα 31* παρατηρούνται οι τιμές των συντελεστών για τα ζεύγη των μεταβλητών. Η συσχέτιση μιας μεταβλητής με τον εαυτό της θα είναι κατά κανόνα τέλεια και θετική. Εφόσον οι μεταβλητές χωρίζονται σε δύο κατηγορίες: αποδόσεων και πωλήσεων, είναι αναμενόμενο ότι θα υπάρχει μεγαλύτερη συσχέτιση μεταξύ των μεταβλητών που υπάγονται στην ίδια κατηγορία. Ωστόσο παρατηρείται ότι σε κάποιες περιπτώσεις, μεταβλητές αντίθετης κατηγορίας μπορεί να υποδεικνύουν ισχυρή συσχέτιση και μεταβλητές της ίδιας κατηγορίας υποδηλώνουν αδύναμη συσχέτιση. Καταρχάς, είναι ενδιαφέρον να σημειωθεί ότι δεν υπάρχει ζεύγος που να υποστηρίζει αρνητική ή ανύπαρκτη συσχέτιση: όλες οι μεταβλητές βοηθούν η μία την απόδοση της άλλης ελαφρώς ή πολύ και οι τιμές ως αποτέλεσμα μετατρέπονται πάντα θετικά.

Συγκεκριμένα παρατηρείται, ότι οι μεταβλητές συνεδριών και επισκέψεων σελίδας φέρουν ισχυρή συσχέτιση μεταξύ τους, σχεδόν τέλεια. Αυτό συμβαίνει διότι μία συνεδρία επιφέρει τουλάχιστον μία επίσκεψη σελίδας, και η επίσκεψη σελίδας εξαρτάται από την ύπαρξη μιας συνεδρίας. Αυτό το φαινόμενο λέγεται συγγραμμικότητα (*collinearity*) όταν αφορά δύο μεταξύ τους μεταβλητές, ή πολλαπλή συγγραμμικότητα (*multicollinearity*), για την περίπτωση που υπάρχουν παραπάνω από δύο μεταβλητές που φέρουν κάποιο βαθμό εξάρτησης μεταξύ τους. Για τα δεδομένα της FG, όπως επίσης και για τα περισσότερα δείγματα δεδομένων πωλήσεων του πραγματικού κόσμου που θα μπορούσαν να μελετηθούν, το φαινόμενο της πολλαπλής συγγραμμικότητας είναι κυρίαρχο. Παράλληλα, αν παρατηρηθεί η συσχέτιση της μεταβλητής συνεδριών με τα ποσοστά της ή των αντίστοιχων επισκέψεων σελίδας με τα δικά της ποσοστά, σημειώνεται χαμηλότερη συσχέτιση, σχεδόν μέτρια. Αυτό συμβαίνει γιατί τα ποσοστά συνεδριών και επισκέψεων σελίδας παράγονται από τη σύγκριση των τιμών αυτών ως προς το άθροισμα των συνολικών συνεδριών ή επισκέψεων σελίδας για όλα τα προϊόντα της επιχείρησης.

Η μεταβλητή *Buy Box %* αναφέρεται στη διαθεσιμότητα του κουμπιού *προσθήκης στο καλάθι* για συνέχεια αγοράς του προβαλλόμενου προϊόντος. Η απουσία του είναι σπάνια και συνήθως συνδυάζεται με περιόδους εξάντλησης αποθεμάτων. Είναι ενδιαφέρον ότι σημειώνει μέτρια ή αδύναμη συσχέτιση με όλες τις υπόλοιπες μεταβλητές, και αυτό συμβαίνει επειδή ο ρόλος του είναι ξεχωριστός από τις υπόλοιπες μεταβλητές και δεν εξαρτάται καθόλου από τη συμπεριφορά του πελάτη και της ζήτησης. Επίσης, οι τιμές του κατά κύριο λόγο παρουσιάζονται να βρίσκονται στο 100% ανεξάρτητα από περιόδους υψηλής εποχικότητας ή χαμηλής ζήτησης, οπότε είναι λογικό ότι η σταθερότητα της συγκεκριμένης μεταβλητής δεν παρουσιάζει ισχυρή συσχέτιση με κάποια άλλη.

Τέλος, για τις μεταβλητές πωλήσεων, παρατηρείται μέτρια ή ισχυρή συσχέτιση με τις μεταβλητές απόδοσης και σχεδόν τέλεια συσχέτιση με τις μεταβλητές της ίδιας κατηγορίας. Η μεταβλητή *Units Ordered* αναφέρεται στα τεμάχια που παραγγέλθηκαν, ενώ η μεταβλητή *Total Order Items* αναφέρεται στον αριθμό των παραγγελιών, που μπορεί να περιέχουν από ένα τεμάχιο τουλάχιστον. Εφόσον οι περισσότερες παραγγελίες αποτελούνται όντως από ένα τεμάχιο, καθώς και επειδή για να παραγγελθεί ένα τεμάχιο πρέπει να υπάρχει αντίστοιχη παραγγελία, οι δύο μεταβλητές μεταξύ τους παρουσιάζουν σχεδόν τέλεια συσχέτιση. Η σχέση αυτών των δύο μεταβλητών με τη μεταβλητή *Sales* εξαρτάται απόλυτα από την τιμή του προϊόντος που μπορεί ελαφρώς να μεταβάλλεται, οπότε πάλι είναι αναμενόμενη η

σχεδόν τέλεια μεταξύ τους συσχέτιση. Η πιο παράξενη μεταβλητή από τον συγκεκριμένο πίνακα ίσως να είναι η *Unit Session %*, που αποτελεί το ποσοστό μετατροπής, δηλαδή τον αριθμό συνεδριών σε σχέση με τα τεμάχια που πουλήθηκαν. Η συγκεκριμένη μεταβλητή εξαρτάται απόλυτα από τις μεταβλητές *τεμάχια* και *συνεδρίες*, αλλά η συσχέτιση της με τα ανάλογα ζεύγη είναι μέτρια προς ισχυρή και μάλιστα πιο χαμηλή από τη συσχέτιση των δύο παραπάνω μεταβλητών μεταξύ τους.

Η κατανόηση των μεταβλητών και των συσχετίσεων τους είναι ιδιαίτερα σημαντική για τη διαδικασία της ανάλυσης δεδομένων και συγκεκριμένα στην περίπτωση της χρήσης των μεταβλητών αυτών για ανάπτυξη μοντέλων πρόβλεψης. Η χρήση των συσχετίσεων θα εμφανιστεί παρακάτω κατά τη διαδικασία ανάπτυξης μοντέλων όπως τη πολλαπλή γραμμική παλινδρόμηση.

8 Υλοποίηση – Ανάπτυξη μοντέλων πρόβλεψης

Σε αυτό το στάδιο θα γίνει χρήση των δεδομένων που προετοιμάστηκαν παραπάνω για την ανάπτυξη μοντέλων από τις διάφορες μεθόδους πρόβλεψης για τις οποίες έγινε περιγραφή στο κεφάλαιο 4. Οι προβλέψεις θα γίνουν για τις εξαρτημένες μεταβλητές *τεμάχια* και *έσοδα*. Ανεξάρτητες μεταβλητές όπως οι *συνεδρίες* και οι *επισκέψεις σελίδας* θα χρησιμοποιηθούν για ανάπτυξη μοντέλων γραμμικής παλινδρόμησης. Οι άλλες τρεις μέθοδοι προβλέψεων – η μέθοδος των κινούμενων μέσων, η εκθετική εξομάλυνση και τα στοχαστικά μοντέλα θα χρησιμοποιηθούν για την ανάπτυξη μονομεταβλητών προβλέψεων, δηλαδή για τα μοντέλα αυτά δε θα χρησιμοποιηθούν ανεξάρτητες μεταβλητές.

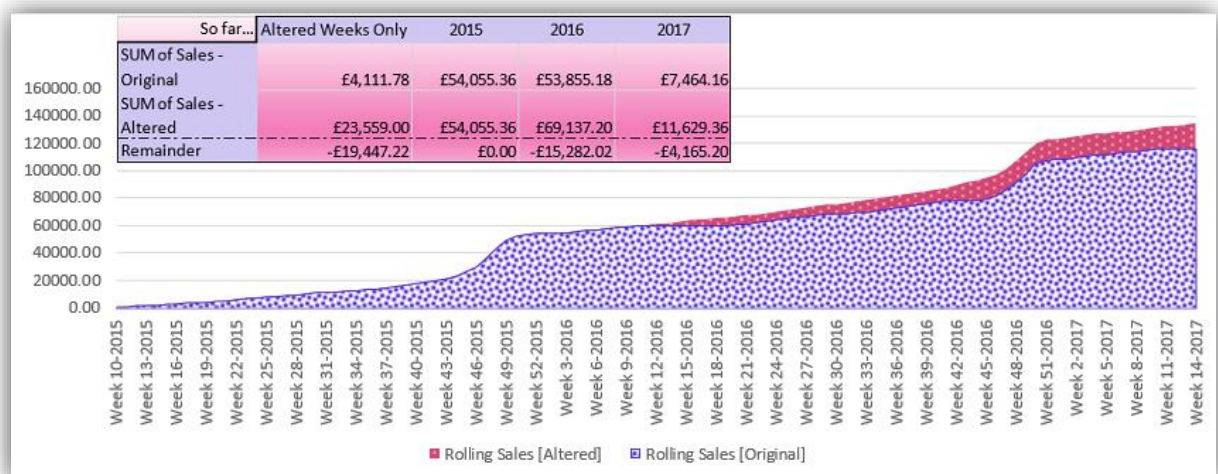
8.1 Ανάκτηση «χαμένων» πωλήσεων

Έχει αναφερθεί ότι ένα από τα βασικά προβλήματα της FG είναι ο προγραμματισμός της διαχείρισης των αποθεμάτων. Τα τεμάχια που παραγγέλλονται από τους κατασκευαστές αργούν να φτάσουν στην κατοχή της FG και να τεθούν διαθέσιμα προς πώληση. Γι' αυτό κυρίως ευθύνεται το γεγονός ότι τα εργοστάσια βρίσκονται στην Ασία και το εμπόρευμα φτάνει στην Αγγλία με καράβι. Πολλές καθυστερήσεις μπορεί να προκύψουν μέχρι να διασφαλισθεί η διαθεσιμότητα του προϊόντος προς πώληση, από τη στιγμή τοποθέτησης μιας παραγγελίας για να ξεκινήσει η παραγωγή του. Για διάφορους ενδογενείς και εξωγενείς παράγοντες, τα δεδομένα για διάφορα προϊόντα εμφανίζουν περιόδους πολύ χαμηλών ή μηδενικών πωλήσεων. Ωστόσο, για το φαινόμενο αυτό δεν ευθύνεται η ζήτηση, αλλά η αδυναμία της προσφοράς να καλύψει τη δυνητική ζήτηση. Το εμπόρευμα αργεί να φτάσει στην Αγγλία και τα υπάρχοντα προϊόντα εξαντλούνται. Έτσι δημιουργούνται περίοδοι εξάντλησης.

Σε αυτό το σημείο τίθενται πολλά ερωτήματα. Πόσο σημαντική είναι η ζημιά από τις εξαντλήσεις στο σύνολο τους ή ξεχωριστά και πως επηρεάζουν την επιχείρηση; Σε τι θα βοηθήσει η γνώση ενός προσεγγιστικού αριθμού των «χαμένων» πωλήσεων; Πως γίνεται να συμπληρωθούν τα κενά στα δεδομένα πωλήσεων για τα οποία ευθύνεται μία εξάντληση αποθεμάτων; Κατά πόσο μια εξάντληση επηρεάζει τη ζήτηση κατά τις επόμενες παρατηρήσεις; Αν δε γίνουν προσπάθειες ανάκτησης των «χαμένων» πωλήσεων, οι

προβλέψεις της ζήτησης θα συμπεριλαμβάνουν μοτίβα εξαντλήσεων όπως αυτά παρουσιάστηκαν στο παρελθόν; Αν στις προβλέψεις συμπεριλαμβάνονται και οι μελλοντικές εξαντλήσεις, οι πιθανές χαμηλότερες προβλέψεις της ζήτησης θα οδηγήσουν σε νέα εξάντληση των αποθεμάτων; Αν οι «χαμένες» πωλήσεις συμπληρωθούν, τι αντίκτυπο θα είχαν στα έσοδα της επιχείρησης;

Οι παραπάνω ερωτήσεις δεν είναι εύκολο να απαντηθούν. Η προσέγγιση των θεμάτων αυτών σε οποιαδήποτε επιχείρηση, εξαρτάται από την άποψη των μετόχων αλλά και του ερευνητή που καλείται να τα αντιμετωπίσει. Για την περίπτωση της FG, αποφασίστηκε ότι η ανάκτηση των συγκεκριμένων τιμών είναι σημαντική. Εφόσον μελετήθηκε η απόκλιση των πραγματικών τιμών από τις δυνητικές πωλήσεις, προτιμήθηκε η εφαρμογή των μοντέλων πρόβλεψης να πραγματοποιηθεί και στα δύο σύνολα δεδομένων· στο *πραγματικό* σύνολο και στο *τροποποιημένο*. Για τη γενική εικόνα δεν έχει σημασία ένας αριθμός που θα μπορούσε να είχε προκύψει, γιατί ανήκει στο παρελθόν, το οποίο φυσικά δεν μπορεί να ανατραπεί. Ωστόσο, εφόσον οι προβλέψεις θα παραχθούν με βάση τα ιστορικά δεδομένα, ίσως να μην είναι καλή πρακτική να γίνει ενσωμάτωση των μοτίβων εξάντλησης σε ήδη περίπλοκους χρονολογικούς κύκλους. Πραγματοποιώντας περαιτέρω διαδικασίες για την ανάκτηση ιστορικών τιμών που δε σημειώθηκαν ποτέ, ίσως οι παράμετροι των μοντέλων που θα αναπτυχθούν για την παραγωγή προβλέψεων να απλοποιηθούν, ώστε αυτές να φέρουν μεγαλύτερη ακρίβεια.



Εικόνα 33: Πίνακας τιμών και γράφημα πρόβλεψης κυλιόμενου αθροίσματος (52 εβδομάδων) πωλήσεων ενός προϊόντος για περιόδους του παρελθόντος όπου η προσφορά αδυνατούσε να καλύψει τη ζήτηση

Όπως φαίνεται στην *εικόνα 33*, το δείγμα δεδομένων που χρησιμοποιήθηκε αφορά στο κυλιόμενο άθροισμα 52 εβδομάδων των πωλήσεων για ένα προϊόν της FG, ανά κάτι παραπάνω από δύο χρόνια, μεταξύ του έτους 2015 και 2017. Παρατηρείται ότι η πρώτη εξάντληση αποθεμάτων στο δείγμα εμφανίζεται περίπου τον Απρίλιο του 2014. Από αυτό το σημείο και μετά, για όσες παρατηρήσεις αντιπροσωπεύουν περίοδο εξάντλησης αποθεμάτων, έχουν ενταχθεί προβλεπόμενες τιμές. Οι τιμές αυτές αθροιστικά συγκρίνονται με το αρχικό δείγμα και, όπως ήταν αναμενόμενο, παρατηρείται μια απόκλιση ανάμεσα στα δείγματα. Συγκεκριμένα, για ολόκληρη τη διάρκεια των παρατηρήσεων, το τροποποιημένο σύνολο παρουσίασε αύξηση των πωλήσεων κατά περίπου 19,500 λίρες. Από το ποσό αυτό, έσοδα αξίας περίπου 15,300 λιρών θα μπορούσαν να είχαν σημειωθεί κατά το έτος 2016, ενώ κατά το πρώτο τρίμηνο του 2017 είχε ήδη χαθεί ένα ποσό περίπου των 4,100 λιρών. Τα δείγματα είχαν μηδενική απόκλιση για τις παρατηρήσεις του έτους 2015, όπου δεν υπήρξε καθόλου εξάντληση αποθεμάτων.

Η συγκεκριμένη πρακτική και η τοποθέτηση της δεν είναι άσχετη. Σίγουρα, είναι πολύ συγκεκριμένη και ιδιαίτερη περίπτωση, αλλά αυτό το μέρος του εγγράφου αφορά μελέτη περίπτωσης. Συνήθως, όταν σκεφτόμαστε για πρόβλεψη κοιτάμε μπροστά, στο μέλλον. Γιατί άλλωστε να γίνει κατανάλωση πόρων για να αναπτυχθεί πρόβλεψη για κάτι που έχει ήδη συμβεί; Όπως αναφέρθηκε προηγουμένως, η χρήση τροποποιημένων δεδομένων μπορεί στην πραγματικότητα να απλοποιήσει τις παραγόμενες προβλέψεις. Το καλό με αυτού του είδους προβλέψεις είναι ότι μπορούν να προσδιοριστούν χρονικά και έχουν όρια αρχής και τέλους. Οι ενδογενείς ή εξωγενείς παράγοντες δεν έχουν ιδιαίτερη σημασία, αφού από μελέτη της χρονολογικής σειράς μπορεί να προσδιοριστεί η τάση και η εποχικότητα για τη συγκεκριμένη περίοδο· οτιδήποτε άλλο είναι απλώς θόρυβος. Η απουσία της περιπλοκότητας για τα συγκεκριμένα δεδομένα, εφόσον τα δείγματα προσδιοριστούν, θέτει τα δεδομένα αυτά κατάλληλα για την ανάπτυξη των ιστορικών προβλέψεων με χρήση της απλούστερης μεθόδου πρόβλεψης: *τους κινούμενους μέσους*.

8.2 Εφαρμογή μεθόδου κινούμενων μέσων

Όπως αναφέρθηκε σε προηγούμενο κεφάλαιο, οι κινούμενοι μέσοι μπορούν να ενταχθούν σε δύο κατηγορίες: *ακόλουθοι κινούμενοι μέσοι* και *κεντρικοί κινούμενοι μέσοι*. Οι παραπάνω δύο κατηγορίες μπορούν επίσης να πάρουν *σταθμισμένη* μορφή. Σε αυτό το σημείο θα

χρησιμοποιηθεί η μέθοδος των κινούμενων μέσων ώστε να πραγματοποιηθεί πρόβλεψη για τις παρατηρήσεις του παρελθόντος που αντιπροσωπεύουν περιόδους εξάντλησης αποθεμάτων. Οι προβλεπόμενες τιμές θα ενταχθούν στα σύνολα δεδομένων και έτσι θα συνταχθούν τα τροποποιημένα σύνολα δεδομένων, τα οποία θα χρησιμοποιηθούν παράλληλα με τα αρχικά για ανάπτυξη μελλοντικών προβλέψεων. Ας σημειωθεί ότι η συγκεκριμένη μέθοδος αφορά την εξομάλυνση των ιστορικών δεδομένων ή αλλιώς πρόβλεψη του ενός βήματος. Η μέθοδος αυτή είναι η πρώτη μέθοδος των κινούμενων μέσων που αναφέρθηκε στο κεφάλαιο 4 και δεν αφορά τα μοντέλα ARIMA(0,0,1) ή MA(1).

Καταρχάς, θα γίνει χρήση του προσθέτου «*Analysis Toolpak*» του Excel, ώστε να παραχθούν οι τιμές των κινούμενων μέσων που απορρέουν από το δείγμα μιας ολόκληρης χρονολογικής σειράς για ένα προϊόν. Έτσι παρουσιάζεται το αποτέλεσμα της εφαρμογής της μεθόδου αυτής. Το πρόσθετο χρησιμοποιεί τη συνάρτηση των κεντρικών κινούμενων μέσων.

Units	151	117	134	90	146	5	2	1	0	0	0	0	17	111	129	144
MA	138	134	114	123	80	51	3	1	0	0	0	6	43	86	128	136

Εικόνα 34: Πίνακας τιμών δείγματος μιας σειράς παρατηρήσεων από πωλήσεις τεμαχίων ενός προϊόντος και η εφαρμογή του εργαλείου κινούμενων μέσων

Όπως φαίνεται στην εικόνα 34, οι τιμές των αποτελεσμάτων κινούμενων μέσων του προσθέτου δεν είναι επαρκής για να αποτελέσουν πρόβλεψη για τις συγκεκριμένες τιμές, εφόσον το μοντέλο λαμβάνει υπόψη τις γειτονικές μηδενικές τιμές. Σίγουρα το φαινόμενο της εξάντλησης αποθεμάτων εξομαλύνεται και φαίνεται να μεταμορφώνεται σε μικρότερη διάσταση, αλλά εξακολουθεί να υποδηλώνει εξάντληση. Για να αντιμετωπιστεί το συγκεκριμένο πρόβλημα, θα χρειαστεί να τροποποιηθεί το μοντέλο κινούμενων μέσων που χρησιμοποιείται, ώστε το μοντέλο να *μη συμπεριλαμβάνει* τις μηδενικές τιμές ή τις χαμηλές γειτονικές τους, που οφείλονται σε εξάντληση αποθεμάτων.

Αρχικά, θα τεθούν τα όρια του υποσυνόλου για το οποίο θα αναπτυχθεί πρόβλεψη. Κατά προτίμηση, εντός των ορίων θα είναι η ακριβώς προηγούμενη τιμή πριν μια μηδενική τιμή, καθώς και η αμέσως επόμενη. Ο λόγος για τη συγκεκριμένη επιλογή είναι ότι η προηγούμενη παρατήρηση γενικά θα είναι χαμηλότερη από ότι ήταν αναμενόμενο, εφόσον η εξάντληση μάλλον συνέβη στη μέση της εβδομαδιαίας περιόδου την οποία η παρατήρηση αυτή αντιπροσωπεύει. Οι πρώτες δύο παρατηρήσεις που εμφανίζουν πωλήσεις μετά από μια

εξάντληση επίσης θα είναι χαμηλές, εφόσον χρειάζονται περίπου δέκα μέρες ώστε οι σελίδες της FG στην Amazon να επανέλθουν στα προηγούμενα επίπεδα εμφάνισης. Αφού τεθούν τα όρια του υποδείγματος για το οποίο θα γίνει πρόβλεψη, θα γίνει χρήση του μέσου των προηγούμενων και επόμενων οκτώ παρατηρήσεων. Υπάρχουν διάφοροι τρόποι υπολογισμού:

- i. Οι τιμές του δείγματος θα είναι το αποτέλεσμα του αριθμητικού μέσου των προηγούμενων 8 και επόμενων 8 παρατηρήσεων, δηλαδή συνολικά 16 τιμών.
- ii. Οι τιμές του δείγματος θα είναι ένας τυχαίος αριθμός ανάμεσα στον αριθμητικό μέσο των προηγούμενων 8 παρατηρήσεων και στον αριθμητικό μέσο των επόμενων 8 παρατηρήσεων.
- iii. Οι τιμές του δείγματος θα είναι το αποτέλεσμα του αριθμητικού μέσου των προηγούμενων 8 παρατηρήσεων πριν από την πρώτη παρατήρηση από το δείγμα τιμών για το οποίο παράγεται η πρόβλεψη.
- iv. Οι τιμές του δείγματος θα είναι ο αριθμητικός μέσος των προηγούμενων 8 παρατηρήσεων από την κάθε παρατήρηση για την οποία παράγεται η πρόβλεψη.

Οι πρώτες δύο περιπτώσεις ίσως να μη θεωρούνται ξεκάθαρα ότι ακολουθούν τη μέθοδο εξομάλυνσης των κινούμενων μέσων, αλλά σίγουρα φέρουν χαρακτηριστικά της. Οι πρώτες δύο περιπτώσεις μοιάζουν με τον *κεντρικό* κινούμενο μέσο, αφού για τον υπολογισμό τους συμπεριλαμβάνονται και τα δύο άκρα. Η τρίτη και τέταρτη περίπτωση ταιριάζουν περισσότερο με τον ακόλουθο κινούμενο μέσο, αφού για τον υπολογισμό τους δε συμπεριλαμβάνονται οι επόμενες παρατηρήσεις. Η δεύτερη περίπτωση είναι όμοια με την πρώτη αλλά συμπεριλαμβάνει το *τυχαίο* χαρακτηριστικό, ο σκοπός του οποίου είναι απλώς να διαφοροποιήσει τις προβλεπόμενες τιμές ώστε να μην επαναλαμβάνονται απαραίτητως.

Η εφαρμογή του πρώτου από τους παραπάνω τρόπους είναι εύκολα υλοποιήσιμη στο Excel με τη χρήση της συνάρτησης *AVERAGE(number1,[number2]...)*. Οι τιμές του δείγματος που αντιπροσωπεύουν εξάντληση αποθεμάτων αντικαθιστούνται από το αποτέλεσμα της συνάρτησης *AVERAGE* που περιλαμβάνει τις προηγούμενες και επόμενες 8 τιμές από το διάστημα τιμών που θα αντικατασταθεί. Με παρόμοιο τρόπο υλοποιείται και ο τρίτος από τους παραπάνω τρόπους υπολογισμού των χαμένων πωλήσεων, απλώς η συνάρτηση περιλαμβάνει μόνο τις προηγούμενες 8 τιμές. Η τέταρτη από τις παραπάνω περιπτώσεις δεν μπορεί να υλοποιηθεί εύκολα μέσω Excel, εκτός και αν ο υπολογισμός για το κάθε κελί πραγματοποιηθεί χειροκίνητα και αφαιρεθεί η συνάρτηση από το προηγούμενο. Αυτό συμβαίνει γιατί ως εμπόδιο για την περίπτωση αυτή εμφανίζεται κάτι που ονομάζεται ως

«κυκλική αναφορά». Μετά από κάθε υπολογισμό, το Excel ελέγχει αν έχουν αλλάξει τα δεδομένα από κελιά που χρησιμοποιούνται σε συναρτήσεις ώστε κατά τον επόμενο αυτόματο ή μη-αυτόματο υπολογισμό να τροποποιήσει τις τιμές που πρέπει να ενημερωθούν. Αν ένα κελί βασίζεται σε υπολογισμό συνάρτησης που περιλαμβάνει κελιά που περιέχουν συναρτήσεις που περιλαμβάνουν τα κελιά της πρώτης συνάρτησης, τότε εμφανίζεται το σφάλμα της κυκλικής αναφοράς.

Ο δεύτερος από τους τρόπους τροποποίησης των δεδομένων που αναφέρονται παραπάνω είναι επίσης εύκολα υλοποιήσιμος από το Excel. Η εισαγωγή του τυχαίου στη διαδικασία αυτή γίνεται με την «παιχνιδιάρικη» συνάρτηση *RANDBETWEEN(bottom; top)*, η οποία εμφανίζει έναν τυχαίο αριθμό εντός των ορίων της. Όπως φαίνεται στην παρακάτω εικόνα, ως όρια της συνάρτησης θέτονται οι αριθμητικοί μέσοι των προηγούμενων και επόμενων 8 παρατηρήσεων. Έτσι οι τροποποιημένες τιμές παραμένουν στα όρια μεταξύ των μέσων των γειτονικών παρατηρήσεων αλλά η ακριβής τιμή τους είναι τυχαία.

Week #	Date	Units O Sea	
Week 2	10/01/2016	10.00	=RANDBETWEEN(AVERAGE(\$DB\$44:\$DB\$51),AVERAGE(\$DB\$61:\$DB\$68))
Week 3	17/01/2016	41.00	0.6510
Week 4	24/01/2016	43.00	0.5797
Week 5	31/01/2016	50.00	0.4270
Week 6	07/02/2016	41.00	0.5375
Week 7	14/02/2016	57.00	0.6092
Week 8	21/02/2016	48.00	0.6510
Week 9	28/02/2016	37.00	0.5853
Week 10	06/03/2016	34.00	0.5614
Week 11	13/03/2016	=RAND	0.6331
Week 12	20/03/2016	57.00	0.6450
Week 13	27/03/2016	54.00	0.4698
Week 14	03/04/2016	56.00	0.4937
Week 15	10/04/2016	57.00	0.5375

Εικόνα 35: Τροποποίηση δεδομένων με χρήση παραλλαγμένης μεθόδου των κεντρικών κινούμενων μέσων και εισαγωγή του τυχαίου

Δυστυχώς η συγκεκριμένη μέθοδος δεν είναι υλοποιήσιμη ευρέως στο δείγμα. Το πρόβλημα είναι ότι για να δουλέψει η συνάρτηση *RANDBETWEEN*, το πρώτο όριο πρέπει να είναι μικρότερο από το δεύτερο. Εφόσον η τάση της χρονολογικής σειράς είναι ανοδική, πολλές παρατηρήσεις πληρούν το παραπάνω κριτήριο. Αν επίσης συμπεριληφθεί στο δεύτερο όριο μία αύξηση 10-20% της τιμής του, τότε αυξάνονται οι πιθανότητες μείωσης της εμφάνισης σφάλματος λόγω λάθος ορισμένων ορίων. Ωστόσο αυτό δε συμβαίνει με βεβαιότητα. Για

παράδειγμα, σίγουρο πρόβλημα θα παρουσιαστεί αν στο πρώτο όριο της συνάρτησης συμπεριλαμβάνονται τιμές από εβδομάδες πριν ή κατά τη διάρκεια των Χριστουγέννων. Ο αριθμητικός μέσος ενός τέτοιου διαστήματος θα είναι πολύ μεγαλύτερος από τον μέσο των εβδομάδων που ακολουθούν την παράκαμψη των πωλήσεων των Χριστουγέννων. Το κάθε πρόβλημα είναι αντιμετωπίσιμο αν μετατραπούν τα όρια της συνάρτησης για κάθε φορά που παρουσιάζεται σχετικό σφάλμα. Ωστόσο, η συγκεκριμένη πρακτική είναι χρονοβόρα και δεν μπορεί να αυτοματοποιηθεί.

Σε επόμενο κεφάλαιο θα γίνει χρήση της *κονσόλας R*, ώστε να αναπτυχθούν και να εφαρμοστούν μοντέλα ARIMA. Ο *κώδικας R* θα βασίζεται πάνω σε καρτέλες αρχείων Excel που θα περιέχουν οργανωμένους πίνακες με δεδομένα από τα οποία οι εισοδοί και έξοδοι του κώδικα θα είναι από προκαθορισμένα κελιά. Η *κονσόλα R* θα συνδέεται με το Excel με κώδικα VBA και με χρήση μιας βιβλιοθήκης R, ο κώδικας θα εξάγει και θα εισάγει δεδομένα στα φύλλα εργασιών, με το πάτημα ενός μακρο-κουμπιού. Θα ήταν ιδανικό αν εκτός από τα αρχικά δεδομένα, στα μοντέλα ARIMA μπορούν να εισαχθούν και τα τροποποιημένα δεδομένα. Ακολουθώντας τη φιλοσοφία της ευκολίας αναπαραγωγής της διαδικασίας, όπως η διεξαγωγή πρόβλεψης θα γίνεται με τη χρήση ενός κουμπιού, με παρόμοιο τρόπο θα παράγονται και τα τροποποιημένα δεδομένα.

Ο *κώδικας R* (βλ. *σχετικό παράρτημα*) θα κάνει χρήση της μεθόδου ακόλουθων κινούμενων μέσων που παράγουν πρόβλεψη ενός βήματος. Δεν τίθεται θέμα κυκλικής αναφοράς με την R, οπότε εφόσον αναπτυχθούν αρκετά περίπλοκοι περιορισμοί και όρια, η υλοποίηση της μεθόδου είναι ιδιαίτερα εύκολη. Καταρχάς, ο κώδικας R θα δέχεται ως είσοδο τη χρονολογική σειρά που φέρει εβδομαδιαία δεδομένα τεμαχίων που πουλήθηκαν ή εσόδων. Τα δεδομένα αυτά εμφανίζονται σε ένα ταμπλό δεδομένων που περιέχουν λίστες με παρατηρήσεις από διαφορετικά προϊόντα, κατηγορίες κλπ.. Επιπλέον εισοδοί θα είναι η εβδομάδα και το έτος της πρώτης παρατήρησης. Στην αρχή του κώδικα επίσης θα δηλωθούν κάποιες μεταβλητές όπως ο *αριθμητικός μέσος ανά στήλη* και η *παρατήρηση λανσαρίσματος προϊόντος*, το περιεχόμενο των οποίων θα υπολογιστεί εντός του ίδιου του κώδικα πριν γίνει χρήση των μεταβλητών αυτών για την τελική τροποποίηση των δεδομένων. Εφόσον οι εισοδοί και οι μεταβλητές είναι έτοιμες, ο κώδικας θα τροποποιήσει όσες τιμές είναι μικρότερες από το 20% του αριθμητικού μέσου της χρονολογικής σειράς. *Οι τροποποιημένες τιμές θα είναι το αποτέλεσμα του αριθμητικού μέσου των προηγούμενων 8 εβδομάδων από τη*

συγκεκριμένη παρατήρηση, συν 10%. Για την τροποποίηση των δεδομένων εξετάζονται οι ακόλουθοι περιορισμοί:

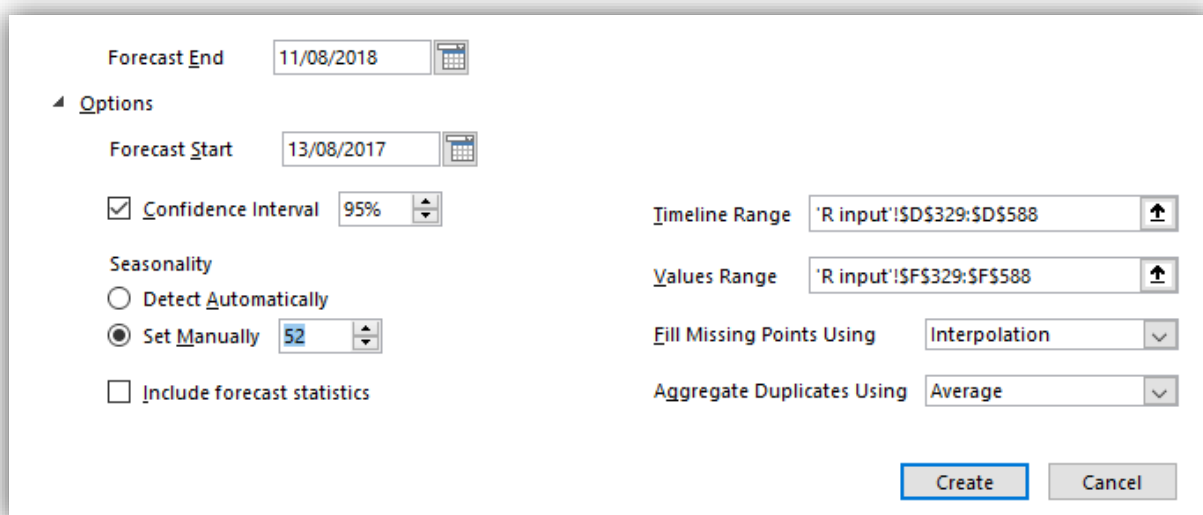
- i. Καμία τροποποίηση δεν μπορεί να συμβεί κατά τις πρώτες 8 παρατηρήσεις, εφόσον πρέπει να υπάρχουν 8 διαθέσιμες τιμές για να υπολογιστεί ο ακόλουθος κινούμενος μέσος.
- ii. Παρομοίως, οι τροποποιήσεις μπορούν να προκύψουν τουλάχιστον 8 παρατηρήσεις μετά την παρατήρηση λανσαρίσματος, πριν από την οποία όλες οι παρατηρήσεις είναι μηδενικές.
- iii. Το 110% του αθροίσματος των προηγούμενων 8 τιμών θα πρέπει να είναι μεγαλύτερο του μηδέν, αν διαιρεθεί με το 8.
- iv. Οι εβδομάδες 1 έως 7 και 52 του έτους δεν μπορούν να τροποποιηθούν με χρήση του ακόλουθου κινούμενου μέσου (λόγο Χριστουγέννων). Εναλλακτικά, αν απαιτούν τροποποίηση, η τιμή που θα πάρουν θα είναι το 30% του αριθμητικού μέσου των παρατηρήσεων της χρονολογικής σειράς.

Date	S all items	All EU platforms					All EU Platforms	Platform specific	AZ DE : DE	AZ UK : UK
		ST/5104	ST/5174	BD/2821	BD/2624	BD/2901	Stationery	AZ DE : DE	Musical boxes	ST/5104
Values shown represent Units Sold for the selected fields										
MIN	14	0	5	0	0	0	25	0	0	0
MEAN	236.62	121.85	64.25	25.33	21.61	138.68	120.16	373.53	118.96	64.08
Outage level:	47	24	13	5	4	28	24	75	24	13
12/02/2017	289	113	35	8	17	116	169	828	105	50
19/02/2017	263	114	29	15	22	93	195	922	102	51
26/02/2017	266	109	20	9	22	106	142	241	78	66
05/03/2017	245	78	56	12	24	129	207	270	75	21
12/03/2017	276	84	54	12	28	155	247	311	107	56
19/03/2017	240	71	49	12	23	139	220	272	80	56
26/03/2017	210	107	25	13	28	151	219	274	71	53
02/04/2017	150	108	15	13	54	78	279	424	30	56
09/04/2017	81	108	39	13	54	133	230	291	29	56
16/04/2017	56	107	39	14	42	135	191	413	79	57
23/04/2017	66	106	41	13	57	141	182	343	75	58
30/04/2017	68	106	44	14	68	146	277	227	75	57

Εικόνα 36: Τροποποιημένες παρατηρήσεις με χρήση κώδικα R

8.3 Εφαρμογή τριπλής εκθετικής εξομάλυνσης

Σε αυτό το σημείο θα γίνει η πρώτη πραγματική μακροπρόθεσμη πρόβλεψη. Σίγουρα, η χρήση κώδικα για την υλοποίηση μεθόδων πρόβλεψης επιτρέπει την παραμετροποίηση της όποιας διαδικασίας, αλλά η κωδικοποίηση απαιτεί πολύ χρόνο και γνώσεις προγραμματισμού. Διάφορα εμπορικά και μη λογισμικά διαθέτουν εργαλεία για διεργασίες που θα ήταν ιδιαίτερα πολύπλοκες και χρονοβόρες για να γραφτούν σε κώδικα. Η απλούστερη μέθοδος μακροχρόνιας πρόβλεψης ίσως είναι η μέθοδος της εκθετικής εξομάλυνσης. Το Excel διαθέτει σχετικές συναρτήσεις όπως τη *FORECAST(x; known_y's; known_x's)* και τη *FORECAST.ETS* που αποτελεί τη βελτίωση της πρώτης στις πιο πρόσφατες εκδόσεις του λογισμικού. Επιπλέον διατίθεται το εργαλείο πρόβλεψης που παρέχει GUI για παραμετροποίηση της πρόβλεψης με σχετικές επιλογές.



The image shows a dialog box for configuring a forecast. It includes fields for 'Forecast End' (11/08/2018) and 'Forecast Start' (13/08/2017). Under the 'Options' section, there are several settings: 'Confidence Interval' is checked and set to 95%; 'Seasonality' is set to 'Set Manually' with a value of 52; 'Timeline Range' is 'R input!\$D\$329:\$D\$588'; 'Values Range' is 'R input!\$F\$329:\$F\$588'; 'Fill Missing Points Using' is set to 'Interpolation'; and 'Aggregate Duplicates Using' is set to 'Average'. At the bottom right, there are 'Create' and 'Cancel' buttons.

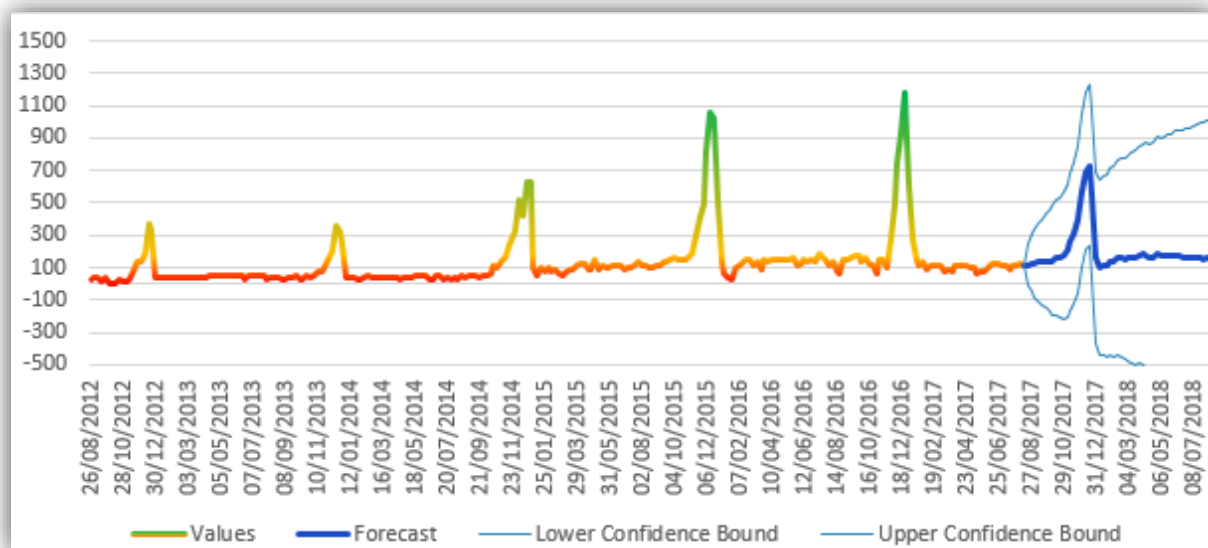
Εικόνα 37: Παραμετροποίηση πρόβλεψης τριπλής εκθετικής εξομάλυνσης – Excel

Το συγκεκριμένο εργαλείο επιστρέφει ως έξοδο ένα πίνακα τιμών που συμπεριλαμβάνει τις προβλεπόμενες τιμές και τα διαστήματα εμπιστοσύνης αυτών, ελέγχους καταλοίπων και το γράφημα της πρόβλεψης. Το διάστημα εμπιστοσύνης ορίζεται από το GUI όπως φαίνεται στην εικόνα 37. Από εκεί επίσης μπορεί να οριστεί ο κύκλος της εποχικότητας αν αυτός είναι γνωστός, ή να αφηθεί αυτόματη. Στην περίπτωση των δεδομένων της FG, είναι γνωστό ότι τα δεδομένα είναι εβδομαδιαία και ότι το σημαντικότερο εποχικό γεγονός συμβαίνει μία φορά τον χρόνο πριν και κατά την περίοδο των Χριστουγέννων. Άρα, ο εποχικός κύκλος των δεδομένων είναι ένα έτος στο οποίο εμφανίζονται 52 παρατηρήσεις: μία για κάθε εβδομάδα.

Timeline	Values	Forecast	Lower Confidence Bound	Upper Confidence Bound	Statistic	Value
23/07/2017	111				Alpha	0.90
30/07/2017	109				Beta	0.00
06/08/2017	125				Gamma	0.10
13/08/2017	113	113	113.00	113.00	MASE	1.66
20/08/2017	118.24975		-10.57	247.07	SMAPE	0.20
27/08/2017	130.13467		-43.26	303.53	MAE	41.93
03/09/2017	124.96277		-83.77	333.69	RMSE	82.05

Εικόνα 38: Πρόβλεψη τριπλής εκθετικής εξομάλυνσης – Excel – πίνακας αποτελεσμάτων

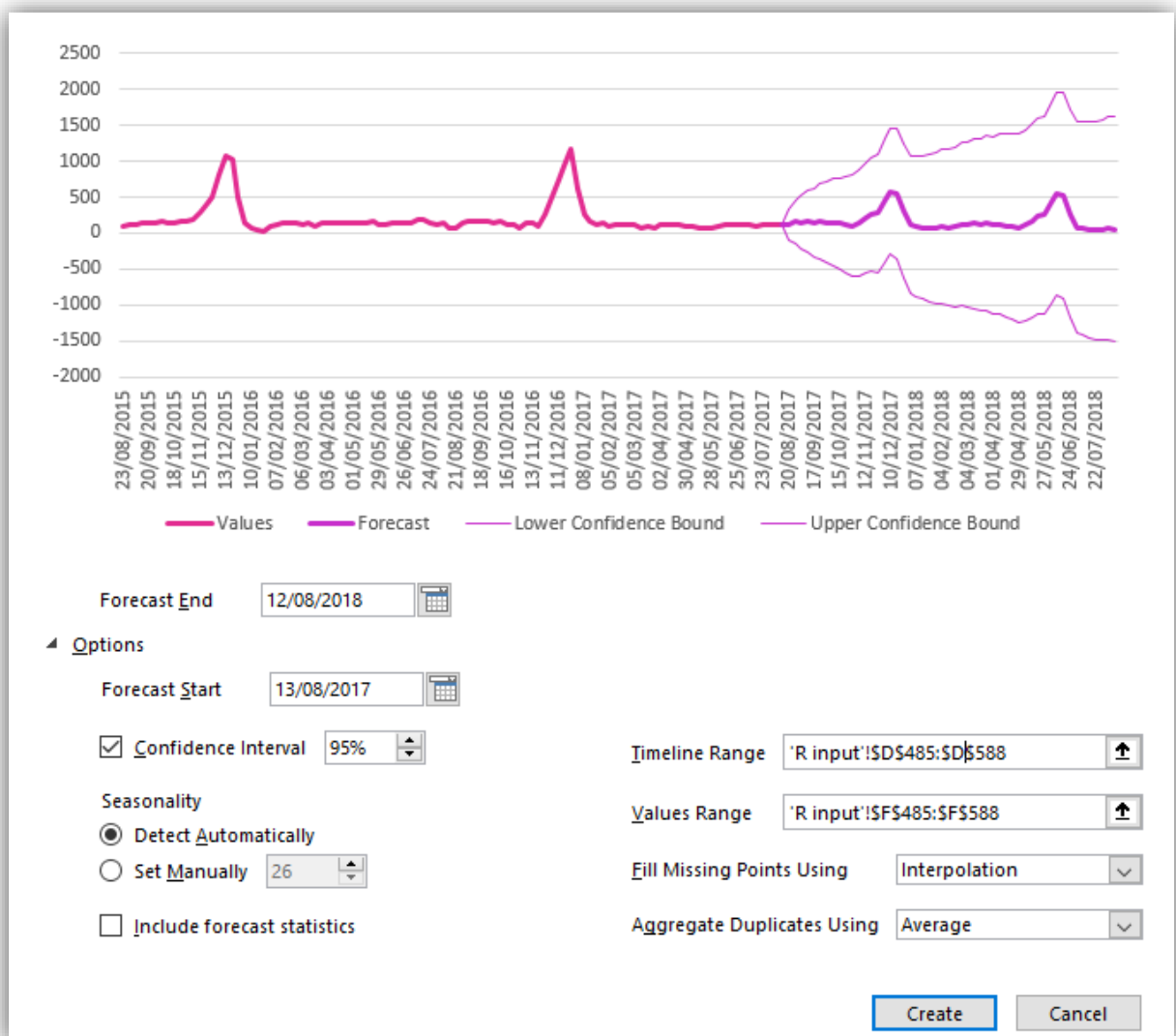
Ο παραπάνω πίνακας καθώς και το παρακάτω γράφημα αποτελούν την έξοδο του εργαλείου πρόβλεψης του Excel. Τα δεδομένα πρόβλεψης ξεκινάνε από την αμέσως επόμενη παρατήρηση ιστορικών δεδομένων. Παράλληλα με τη στήλη της πρόβλεψης, εμφανίζονται επίσης τα διαστήματα εμπιστοσύνης, που στη συγκεκριμένη περίπτωση είχαν οριστεί ως 95%. Αυτό σημαίνει ότι υπάρχει 95% πιθανότητα ότι η πραγματική τιμή για την οποία γίνεται πρόβλεψη θα βρίσκεται εντός των ορίων του διαστήματος εμπιστοσύνης. Η στήλη πρόβλεψης που εμφανίζεται θεωρείται ως η μέση πρόβλεψη.



Εικόνα 39: Πρόβλεψη τριπλής εκθετικής εξομάλυνσης – Excel – γράφημα

Με τον τρόπο που τα αποτελέσματα του πίνακα είναι χωρισμένα σε διαφορετικές στήλες, επιτρέπεται η χρήση πολλαπλών σειρών δεδομένων πάνω στον άξονα X του γραφήματος που εικονίζεται στην εικόνα 39. Έτσι, μορφοποιώντας τις εικονιζόμενες σειρές, είναι ξεκάθαρο

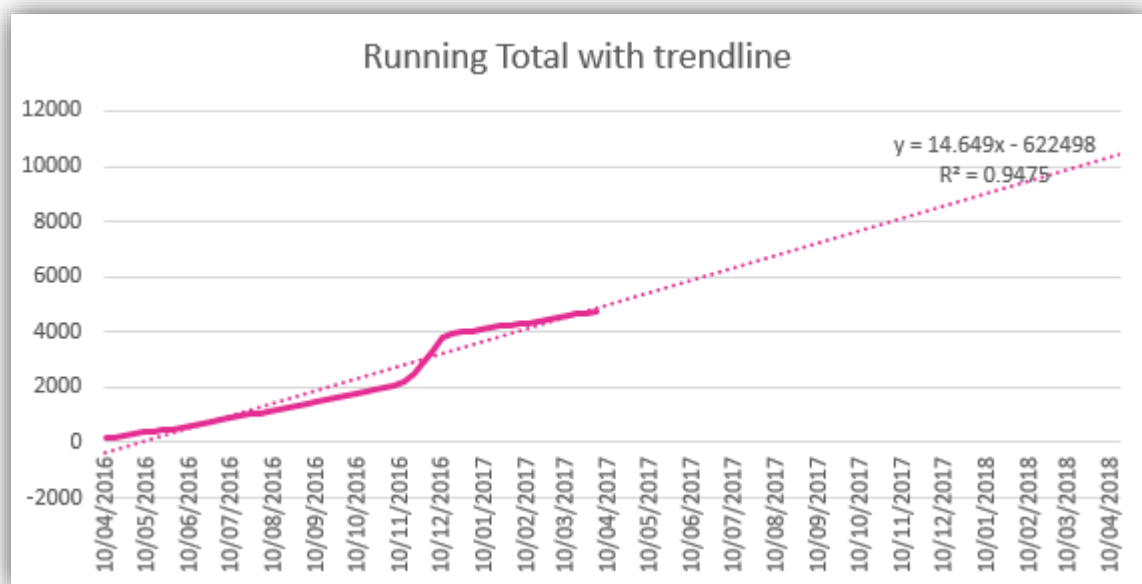
που τελειώνουν τα ιστορικά δεδομένα και που ξεκινούν να σκιαγραφούνται οι τιμές πρόβλεψης μαζί με το αντίστοιχο διάστημα εμπιστοσύνης. Η πρόβλεψη φανερώνει την πιθανή συμπεριφορά της συγκεκριμένης χρονολογικής σειράς για τις επόμενες 52 παρατηρήσεις. Στη συγκεκριμένη περίπτωση έχει οριστεί ο εποχικός κύκλος από πριν. Αν είχε επιλεγθεί αυτόματη αναγνώριση της εποχικότητας για τα συγκεκριμένα δεδομένα, τότε το Excel θα αναγνώριζε επιτυχώς ότι ο κύκλος αποτελείται από 52, ίσως 53 εβδομάδες. Ωστόσο, αν αντί για 5 έτη ιστορικών δεδομένων υπήρχαν διαθέσιμα 2 ή 3, τότε το Excel δε θα είχε την ίδια επιτυχία προσδιορισμού της εποχικότητας.



Εικόνα 40: Πρόβλεψη τριπλής εκθετικής εξομάλυνσης – Excel – αναγνώριση εποχικότητας

Από την εικόνα 39 μπορεί να παρατηρηθεί ότι αν και φανερώνεται αύξηση των πωλήσεων στην κατάλληλη χρονολογική στιγμή, η αύξηση είναι ελαφρώς απογοητευτική. Ο λόγος που οι εποχικές τιμές είναι χαμηλότερες από ότι ήταν αναμενόμενο εξηγείται από το ίδιο το όνομα της μεθόδου πρόβλεψης. Η μέθοδος της εκθετικής εξομάλυνσης κάνει ακριβώς αυτό· εξομαλύνει τα δεδομένα για να παράγει την πρόβλεψη. Η μειωμένες τιμές για τις εποχικές περιόδους δε σημαίνουν απαραίτητα ότι προβλέπεται μείωση των πωλήσεων για τον επόμενο χρόνο, απλώς τα προβλεπόμενα δεδομένα παρουσιάζονται πιο ομαλά.

Ένας τρόπος διαταραχής της ομαλότητας των προβλεπόμενων δεδομένων είναι η χρήση των εποχικών δεικτών. Οι εποχικοί δείκτες χρησιμοποιούνται ώστε να προμηνύουν το βάρος που επιδέχεται η κάθε παρατήρηση, σε σχέση με τις ιστορικές της εμφανίσεις. Ο υπολογισμός των δεικτών γίνεται εύκολα με χρήση πινάκων Pivot. Ας σημειωθεί ότι για να γίνει χρήση των δεικτών πάνω στα δεδομένα πρόβλεψης, τα αποτελέσματα της πρόβλεψης θα πρέπει να έχουν υπολογιστεί από ιστορικά δεδομένα από τα οποία έχει αφαιρεθεί η εποχικότητα. Για να αφαιρεθεί η εποχικότητα από τα ιστορικά δεδομένα, ένας τρόπος που θα χρησιμοποιηθεί αρκετά στο συγκεκριμένο κεφάλαιο είναι το *τρέχουμενο άθροισμα*.

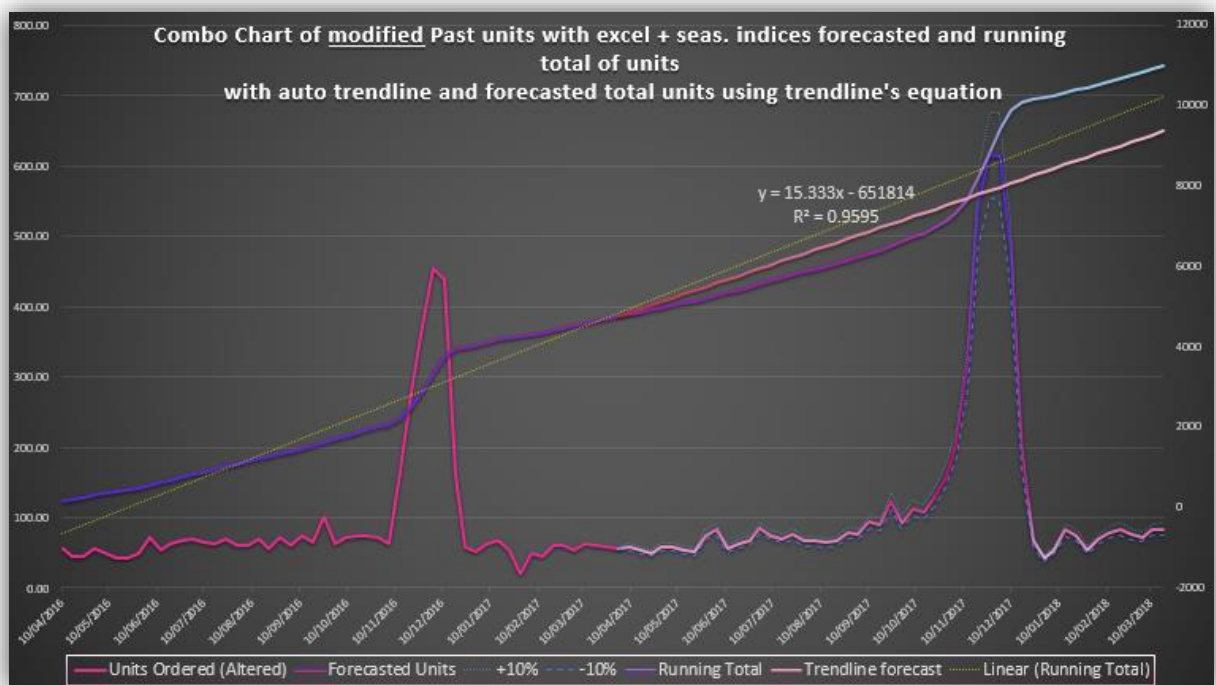


Εικόνα 41: Τρέχουμενο άθροισμα – απεικόνιση τάσης

Το τρέχουμενο άθροισμα για την 5^η εβδομάδα του 2017 συμπεριλαμβάνει τις προηγούμενες 52 παρατηρήσεις αθροιστικά, δηλαδή από την 6^η εβδομάδα του 2016 έως την 5^η εβδομάδα του 2017. Η 6^η εβδομάδα του 2017 προσθέτει την παρατήρηση της τρέχουσας εβδομάδας

και αφαιρεί την παλιότερη παρατήρηση που εμφανίστηκε στο τρεχούμενο άθροισμα της 5^{ης} εβδομάδας. Αν η διαδικασία αυτή εφαρμοστεί σε ολόκληρη τη χρονολογική σειρά, το αποτέλεσμα θα είναι μία νέα χρονολογική σειρά με 52 λιγότερες παρατηρήσεις, που ξεκινάει ένα έτος αργότερα από την πρώτη. Αν όλες οι τιμές της νέας χρονολογικής σειράς διαιρεθούν με το 52, τότε οι τιμές θα είναι κοντά στις αρχικές τους τιμές αλλά θα τους λείπει εντελώς το εποχικό στοιχείο. Η χρονολογική σειρά έχει εξομαλυνθεί εντελώς. Φυσικά, εδώ αναλαμβάνουν οι εποχικοί δείκτες, που αν πολλαπλασιαστούν με τις αντίστοιχες εβδομάδες τους θα αποτελέσουν σε μία χρονολογική σειρά αρκετά παρόμοια με τα αρχικά δεδομένα – χωρίς τις πρώτες 52 παρατηρήσεις.

Η αποσύνθεση της χρονολογικής σειράς σίγουρα απλοποιεί τη διαδικασία και τα αποτελέσματα οποιασδήποτε υλοποιήσιμης μεθόδου πρόβλεψης. Εφόσον η εποχικότητα είναι γνωστή και οι παραγόμενοι δείκτες είναι έμπιστοι και μη-προκατειλημμένοι, η αποσύνθεση της χρονολογικής σειράς πριν την πρόβλεψη είναι μία καλή πρακτική. Όπως φαίνεται στην παρακάτω εικόνα, για την κρίσιμη περίοδο πριν τα Χριστούγεννα προβλέπεται σημαντική αύξηση των πωλήσεων για τη συγκεκριμένη χρονολογική σειρά.



Εικόνα 42: Ανάπτυξη πρόβλεψης εκθετικής εξομαλυνσης πάνω σε δεδομένα τρεχούμενων αθροισμάτων και πολλαπλασιασμός αποτελεσμάτων με εποχικούς δείκτες – γράφημα

8.4 Εφαρμογή πολλαπλής γραμμικής παλινδρόμησης

Η επόμενη μέθοδος πρόβλεψης που θα υλοποιηθεί είναι η αυτή της γραμμικής παλινδρόμησης. Οι προβλέψεις που παράγονται από μοντέλα γραμμικής παλινδρόμησης είναι αποτέλεσμα μιας μαθηματικής εξίσωσης η οποία συσχετίζει μία ή περισσότερες εισόδους, τις σταθμίζει και επιστρέφει έξοδο. Για την ανάπτυξη ενός μοντέλου γραμμικής παλινδρόμησης πρέπει να υπάρχουν διαθέσιμες τουλάχιστον δύο μεταβλητές που να σχετίζονται μεταξύ τους. Η μεταβλητή για την οποία θα γίνει η πρόβλεψη ονομάζεται «εξαρτημένη», ενώ η μεταβλητές που χρησιμοποιούνται στη μαθηματική εξίσωση για την παραγωγή της εξόδου του μοντέλου ονομάζονται «ανεξάρτητες». Αν υπάρχουν παραπάνω από μία ανεξάρτητες μεταβλητές τότε το μοντέλο πρόβλεψης αναπτύσσεται με τη μέθοδο της πολλαπλής γραμμικής παλινδρόμησης. Στην περίπτωση της FG, η μεταβλητή για την οποία θα γίνει πρόβλεψη είναι τα τεμάχια (*Units Sold*), και για τη διεξαγωγή της πρόβλεψης ως ανεξάρτητες μεταβλητές θα χρησιμοποιηθούν οι συνεδρίες (*sessions*), οι επισκέψεις σελίδας (*page views*) και οι μετατροπές (*conversion – unit session %*).

Τα δεδομένα εισόδου της μεθόδου δε χρειάζεται απαραίτητα να αποτελούν χρονολογική σειρά. Αντιθέτως, σε δεδομένα χρονολογιών σειρών μπορεί η διαθεσιμότητα των ανεξάρτητων μεταβλητών να είναι προβληματική. Συγκεκριμένα, οι παραπάνω ανεξάρτητες μεταβλητές από τα δεδομένα της FG δεν είναι προκαθορισμένες, αλλά σημειώνονται παράλληλα με της παρατηρήσεις της εξαρτημένης μεταβλητής. Αυτό σημαίνει ότι και να αναπτυχθεί ένα μοντέλο με τέλεια ακρίβεια, η χρήση αυτού μπορεί να γίνει μόνο θεωρητικά και κατά προσέγγιση. Γνωρίζοντας της επιρροή που θα έχουν οι ανεξάρτητες μεταβλητές συνεδριών, επισκέψεων σελίδας και μετατροπών στα δεδομένα πωληθέντων τεμαχίων, θα προσφέρουν γνώση στην επιχείρηση σχετικά με την ισοστάθμιση των διαφημιστικών τους στόχων, τη διαχείριση των σελίδων των προϊόντων τους, κλπ..

Καταρχάς, προτού ξεκινήσει οποιαδήποτε προσπάθεια ανάπτυξης παλινδρομικών μοντέλων, πρέπει να ελεγχθεί η σχέση των μεταβλητών που υπάρχουν διαθέσιμες. Ας σημειωθεί ότι και για τη μέθοδο της παλινδρόμησης θα επαναληφθεί η πρακτική του τρεχούμενου αθροίσματος που αναφέρθηκε κατά την εφαρμογή της μεθόδου τριπλής εκθετικής εξομάλυνσης. Με αυτό τον τρόπο θα απορροφηθεί το εποχικό στοιχείο από την εξαρτημένη και τις ανεξάρτητες μεταβλητές και η πρόβλεψη θα γίνει πάνω στην τάση των δεδομένων. Για τη μεταβλητή μετατροπών που είναι ποσοστιαίας φύσεως, θα χρησιμοποιηθεί ο τρεχούμενος αριθμητικός μέσος των τιμών της.

Regression Analysis					
Correlation Coefficients (r)					
	Σ of Session	Σ of Page View	Σ of Conversion	Σ of Units	Σ of Sales
Σ of Sessions	1				
Σ of Page View	0.988999	1			
AVG of Conversion	0.687012	0.59871539	1		
Σ of Units	0.940335	0.88578629	0.85703663	1	
Σ of Sales	0.938788	0.886243	0.84605983	0.997966	1

Εικόνα 43: Πίνακας συντελεστών συσχέτισης

Από τον πίνακα συντελεστών συσχέτισης φανερώνεται η ισχυρή σχέση της εξαρτημένης μεταβλητής « Σ of Units» με τις ανεξάρτητες μεταβλητές. Η χρήση των μεταβλητών « Σ of Sessions», « Σ of Page Views» και «AVG of Conversion» για την ανάπτυξη του παλινδρομικού μοντέλου μάλλον θα έχει πολύ καλά αποτελέσματα. Για τη διεξαγωγή της μεθόδου θα χρησιμοποιηθεί ξανά το εργαλείο «Analysis Toolpak» του Excel. Εφόσον γίνει επιλογή των εισόδων του μοντέλου στο εργαλείο, εκείνο θα δώσει ως έξοδο ότι απεικονίζεται στην εικόνα 44, η οποία εμφανίζει τις εξόδους δύο εφαρμογών παλινδρόμησης. Για την πρώτη, ως είσοδο χρησιμοποιήθηκαν τα αρχικά δεδομένα της FG, ενώ για τη δεύτερη χρησιμοποιήθηκαν τα τροποποιημένα δεδομένα που φέρουν μεταμορφώσεις για τις περιόδους εξάντλησης. Στην έξοδο περιγράφονται τα στατιστικά παλινδρόμησης, όπως ο συντελεστής προσδιορισμού R^2 .

Multiple Linear Regression on original Data (on first 79 weeks)							Multiple Linear Regression on modified Data (on first 79 weeks)						
Regression Statistics							Regression Statistics						
Multiple R	0.99538						Multiple R	0.996436					
R Square	0.990782						R Square	0.992884					
Adjusted R Square	0.990413						Adjusted R Square	0.992599					
Standard Error	61.7096						Standard Error	64.27147					
Observations	79						Observations	79					
ANOVA							ANOVA						
	df	SS	MS	F	Significance F		df	SS	MS	F	Significance F		
Regression	3.00	30696931.13	10232310.38	2687.00	0.00	Regression	3.00	43228249.19	14409416.40	3488.27	0.00		
Residual	75.00	285605.61	3808.07			Residual	75.00	309811.65	4130.82				
Total	78.00	30982536.73				Total	78.00	43538060.84					
	Coefficient	Standard Error	t Stat	P-value	Lower 95%	Upper 95%		Coefficient	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-1214.74	365.20	-3.33	0.00	-1942.26	-487.22	Intercept	-1153.17	233.43	-4.94	0.00	-1618.18	-688.15
Σ of Sessions	0.49	0.04	13.87	0.00	0.42	0.57	Σ of Sessions	0.49	0.03	16.32	0.00	0.43	0.55
Σ of Page Views	-0.26	0.03	-8.40	0.00	-0.32	-0.19	Σ of Page Views	-0.25	0.02	-11.05	0.00	-0.30	-0.21
AVG of Conversion	8657.40	2069.32	4.18	0.00	4535.10	12779.71	AVG of Conversion	7829.99	2230.88	3.51	0.00	3385.84	12274.14

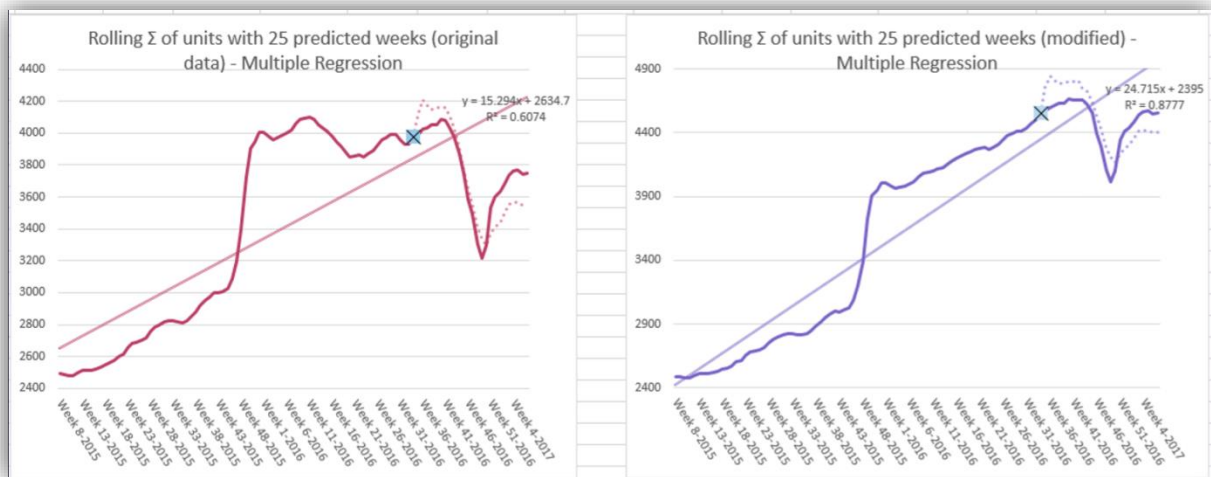
Εικόνα 44: Έξοδος πολλαπλής γραμμικής παλινδρόμησης

Ο συντελεστής προσδιορισμού χρησιμοποιείται συχνά ως μέτρο αξιολόγησης της ικανότητας των ανεξάρτητων μεταβλητών να προβλέψουν την εξαρτημένη. Από την παραπάνω εικόνα είναι ξεκάθαρο ότι οι ανεξάρτητες μεταβλητές μπορούν να προβλέψουν τα πωληθέντα τεμάχια κατά 99,07% για τα μη-τροποποιημένα δεδομένα, και κατά 99,28% για τα τροποποιημένα δεδομένα. Τα ποσοστά αυτά είναι πραγματικά πολύ υψηλά. Οι τιμές των «P-value» και «Significance F» μοιάζουν μηδενικές και υποστηρίζουν τα υψηλά ποσοστά των συντελεστών προσδιορισμού. Το εργαλείο *Analysis Toolpak* του Excel παρέχει πληροφορίες στατιστικής φύσεως σχετικά με την απόδοση του μοντέλου και των καταλοίπων. Το εργαλείο δεν παρέχει τις προβλεπόμενες τιμές, αλλά δίνει τους συντελεστές της παλινδρόμησης «Coefficients». Για να γίνει η εφαρμογή της πολλαπλής παλινδρόμησης πάνω στα δεδομένα, θα πρέπει να αναπτυχθεί συνάρτηση που να πολλαπλασιάζει τις ανεξάρτητες μεταβλητές με τους αντίστοιχους συντελεστές στάθμισης.

Rolling Calculations				
Σ of Sessions	Σ of Page Views	VG of Conversic	Σ of Units	Regr Pred Units
32142	45373	10.47%	3931	
32736	46302	10.48%	3976	3976
33234	47033	10.41%	3993	4126.9
33838	47868	10.37%	4024	4209.0
34437	49136	10.26%	4036	4173.1
34976	50251	10.19%	4053	4148.6

Εικόνα 45: Εφαρμογή συνάρτησης πολλαπλής γραμμικής παλινδρόμησης

Τα δεδομένα της FG αποτελούν χρονολογική σειρά και οι ανεξάρτητες μεταβλητές δεν υπάρχουν διαθέσιμες για μελλοντικές χρονικές στιγμές. Για να γίνει εφαρμογή του μοντέλου παλινδρόμησης πάνω στα δεδομένα θα πρέπει να εισαχθούν τυχαίες τιμές για τις ανεξάρτητες τιμές, αλλιώς να αναπτυχθούν μεμονωμένες προβλέψεις άλλης μεθόδου (π.χ. εκθετικής εξομάλυνσης), ώστε να υπάρχουν διαθέσιμες τιμές των ανεξάρτητων μεταβλητών για να εισαχθούν στο μοντέλο. Για λόγους παρουσίασης, εδώ απλώς χρησιμοποιήθηκαν λιγότερες παρατηρήσεις (οι πρώτες 79) στο δείγμα εκμάθησης, δηλαδή ως είσοδο στο εργαλείο *Analysis Toolpak*. Οι υπόλοιπες παρατηρήσεις θα χρησιμοποιηθούν ως είσοδο στη συνάρτηση της παλινδρόμησης και θα παράγουν τις προβλέψεις. Συγκεκριμένα, οι προβλέψεις θα αντιπροσωπεύουν τις εβδομάδες 35 έως 52 του 2016 και τις 1 έως 7 του 2017.



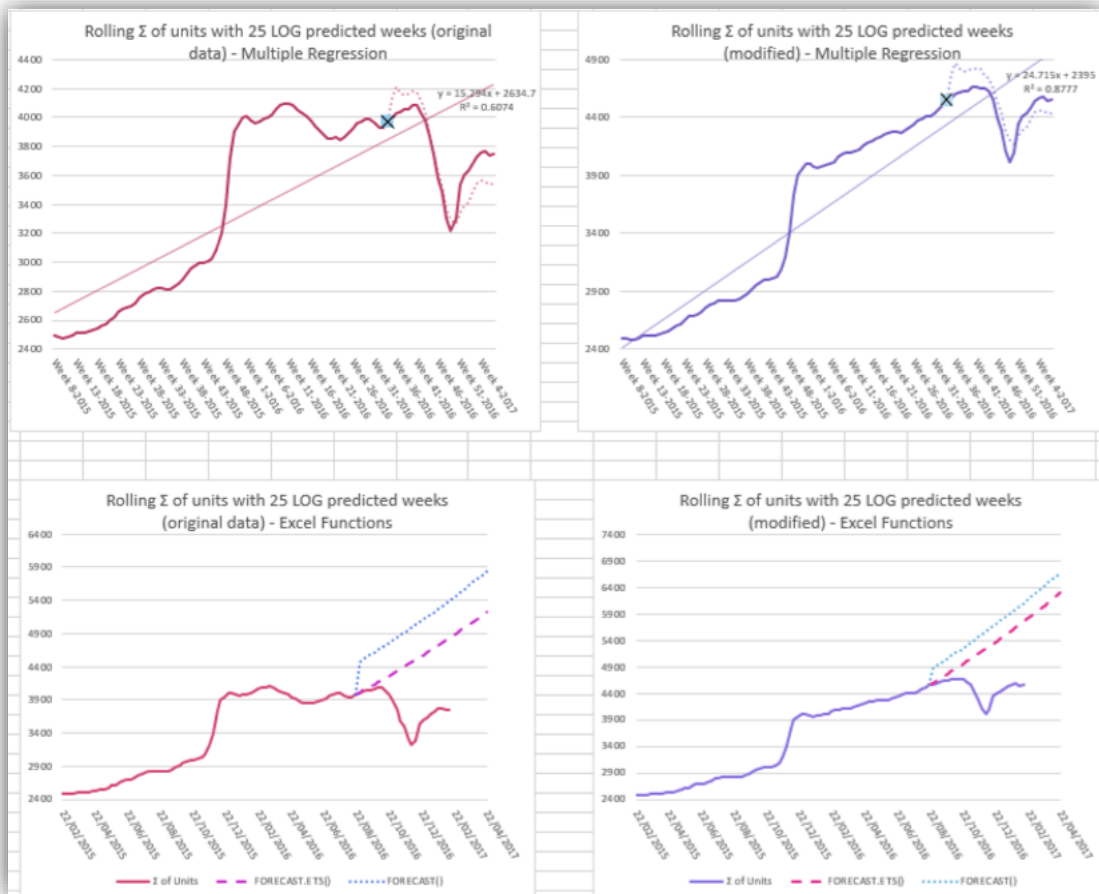
Εικόνα 46: Εφαρμογή συντελεστών παλινδρόμησης σε ιστορικά δεδομένα και σύγκριση των προβλεπόμενων τιμών (διακεκομμένη γραμμή) με τις πραγματικές (+ γραμμική τάση)

Από την παραπάνω εικόνα είναι προφανές ότι η πρόβλεψη δεν είναι τέλεια, αλλά πλησιάζει τις πραγματικές τιμές με πολύ μεγάλη ακρίβεια. Και στις δύο περιπτώσεις παρουσιάζεται απότομη αύξηση των προβλεπόμενων τιμών ακριβώς μετά την έναρξη των προβλέψεων. Οι πραγματικές τιμές σημειώνουν ανοδική τάση, αλλά όχι σε τόσο μεγάλο βαθμό όσο οι προβλεπόμενες. Η παρακμή που σημειώνεται κατά τη μέση της περιόδου πρόβλεψης φαίνεται να προβλέπεται σχεδόν τέλεια. Τέλος, κατά την ανάκαμψη των τιμών μετά τη μεγάλη πτώση, οι προβλέψεις, αν και δείχνουν ανάκαμψη, παρουσιάζουν μικρότερες τιμές από τις πραγματικές. Για τα συγκεκριμένα δεδομένα, η πρόβλεψη αυτή μπορεί να θεωρηθεί ιδιαίτερα ικανοποιητική. Επόμενο βήμα θα ήταν η αντιστροφή του τρεχούμενου αθροίσματος και η επαναφορά της εποχικότητας στα δεδομένα.

Η διαδικασία παλινδρόμησης έχει ολοκληρωθεί. Για να ελεγχτεί το ενδεχόμενο υπολογισμού συντελεστών με μεγαλύτερη ακρίβεια, ένας εύκολος τρόπος είναι η λογαριθμική μεταμόρφωση. Για την υλοποίηση αυτού πρέπει να αναπαραχθούν οι στήλες της εξαρτημένης και των ανεξάρτητων μεταβλητών και να αντικατασταθούν με τη συνάρτηση $LOG(number; [base])$. Οι λογάριθμοι θεωρούνται από τις καλύτερες μαθηματικές μεταμορφώσεις δεδομένων, λόγω της ευκολίας της ερμηνείας τους. Οι μεταβολές των μεταμορφωμένων τιμών είναι σχετικές και ανάλογες με την αρχική κλίμακα. Για μη γραμμικά δεδομένα, οι λογαριθμικές μεταμορφώσεις μπορεί να επιτρέψουν τα δεδομένα αυτά να ενταχθούν σε γραμμικά μοντέλα με επιτυχία. Στο πάνω μέρος της εικόνας 47 παρουσιάζονται τα γραφήματα που αντιπροσωπεύουν προβλέψεις από μοντέλα πολλαπλής

γραμμικής παλινδρόμησης με λογαριθμικά μεταμορφωμένες εισόδους, που στη συνέχεια αναστράφηκαν ($10^{\log(x)}$).

Τα γραφήματα αυτά εμφανίζουν όμοιες προβλέψεις με εκείνες που προήλθαν από συντελεστές παλινδρόμησης που υπολογίστηκαν από μη-λογαριθμικά δεδομένα. Η διαφορά με τα πρώτα είναι τόσο ελάχιστες που μάλλον η λογαριθμική μεταμόρφωση δεν είναι ιδιαίτερα χρήσιμη για τα δεδομένα της FG, τουλάχιστον όχι όσο τα μοντέλα πρόβλεψης εντάσσονται πάνω σε τρεχούμενα αθροίσματα. Στο κάτω μέρος της παρακάτω εικόνας φαίνονται γραφήματα που αντικατοπτρίζουν την πρόβλεψη τριπλής εκθετικής εξομάλυνσης του Excel, πάνω στα ίδια λογαριθμικά μεταμορφωμένα δεδομένα. Είναι προφανές ότι τα αποτελέσματα της μεθόδου σε αυτήν την περίπτωση απέχουν πάρα πολύ από την πραγματικότητα.



Εικόνα 47: Γραφήματα πρόβλεψης πολλαπλής γραμμικής παλινδρόμησης σε λογαριθμικά μεταμορφωμένα δεδομένα (πάνω) και πρόβλεψη τριπλής εκθετικής εξομάλυνσης στα ίδια λογαριθμικά δεδομένα (κάτω)

8.5 Εφαρμογή στοχαστικών μοντέλων ARIMA

Τα μοντέλα ARIMA ίσως αποτελούν την καταλληλότερη μέθοδο πρόβλεψης χρονολογικών σειρών. Οι χρονολογικές σειρές αποτελούνται από δεδομένα που περιέχουν τη διάσταση του χρόνου, ο οποίος θεωρείται κρίσιμος και προμηνύει τη συμπεριφορά των δεδομένων. Τα μοντέλα ARIMA χρησιμοποιούν μαθηματικές εξισώσεις ώστε να προσδιορίσουν και να αναπαράγουν μοτίβα και τάσεις από τις παρατηρήσεις εκμάθησης. Για τη συγκεκριμένη μέθοδο, ως είσοδος απαιτείται μόνο η μεταβλητή για την οποία θα γίνει η πρόβλεψη και το παρελθόν της. Εφόσον για την ανάπτυξη της πρόβλεψης δε συμπεριλαμβάνονται ανεξάρτητες μεταβλητές, εκτός από την *τάση* και την *εποχικότητα* στις χρονολογικές σειρές προσδιορίζεται και το *τυχαίο* χαρακτηριστικό, το οποίο αντικαθιστά τις επιρροές από οποιοδήποτε εξωγενή παράγοντα. Τα μοντέλα ARIMA συνυπολογίζουν το τυχαίο χαρακτηριστικό, γι' αυτό λέγονται και στοχαστικά μοντέλα. Η έλλειψη της αναγκαιότητας τρίτων μεταβλητών είναι σίγουρο πλεονέκτημα της μεθόδου σε σύγκριση με τη μέθοδο της γραμμικής παλινδρόμησης, η οποία απαιτεί την ύπαρξη ανεξάρτητων μεταβλητών. Όπως έχει ήδη αναφερθεί, η μέθοδος της παλινδρόμησης δεν είναι η κατάλληλη για τα δεδομένα της FG, αφού οι ανεξάρτητες μεταβλητές δεν προϋπάρχουν της εξαρτημένης.

Σε αυτό το μέρος του εγγράφου θα γίνει χρήση διάφορων πρακτικών που χρησιμοποιήθηκαν σε άλλα σημεία. Τα μοντέλα πρόβλεψης θα αναπτυχθούν και θα παράγουν έξοδο με κώδικα R. Οι προβλέψεις κάθε φορά θα διεξάγονται από δύο φορές: πάνω στα αρχικά δεδομένα και στα τροποποιημένα δεδομένα. Ο κώδικας θα υπολογίζει το *τρεχούμενο άθροισμα* της μεταβλητής και θα χρησιμοποιεί αυτό για να υπολογίσει και να παράγει την πρόβλεψη, ενώ στη συνέχεια θα αναστρέφει το τρεχούμενο άθροισμα και θα πολλαπλασιάζει τις τιμές με τους *εποχικούς συντελεστές*.

Units <small>Use the arrow to select either Currency or Units</small>										
Seasonal Indices										
Σ of selected items	All platforms					All Platforms	Platform spe	AZ FR : FR	AZ DE : DE	
	ST/5104	ST/5174	BD/2821	BD/2624	BD/2901	Physical box	AZ UK : UK	Physical box	ST/5104	
0.56	0.38	0.69	0.85	1.54	0.41	0.55	0.70	0.49	0.33	
0.62	0.41	0.83	1.02	0.89	0.76	0.62	0.72	0.63	0.41	
0.63	0.56	0.75	0.73	0.55	0.71	0.61	0.67	0.49	0.73	
0.67	0.63	0.57	0.84	0.68	0.95	0.64	0.67	0.59	0.78	

Εικόνα 48: Συντελεστές εποχικότητας

Η υλοποίηση της μεθόδου στοχαστικών μοντέλων ARIMA θα γίνει μέσω της κονσόλας *R*, με χρήση της βιβλιοθήκης «*forecast*» του Rob Hyndman. Οι βιβλιοθήκες περιέχουν κώδικα και εφόσον κληθούν στην κονσόλα υποστηρίζουν τον προγραμματιστή με επιπρόσθετες εντολές. Η *forecast* παρέχει μία εντολή που υπολογίζει αυτόματα το καταλληλότερο μοντέλο ARIMA για τη χρονολογική σειρά. Επιπλέον, απαραίτητες βιβλιοθήκες είναι η «*tseries*» και η «*zoo*». Τέλος, θα χρησιμοποιηθεί η βιβλιοθήκη «*excel.link*», ώστε ο κώδικας που θα γραφτεί να ανταλλάζει εισόδους και εξόδους με το Excel. Με χρήση VBA, ο τελικός κώδικας *R* θα συνδεθεί με μάκρο-κουμπιά εντός των φύλλων εργασιών του Excel και η αναπαραγωγή των προβλέψεων θα γίνεται γρήγορα με το πάτημα μερικών κουμπιών. Συνεπώς, για να μπορεί η διαδικασία να αυτοματοποιηθεί, θα χρειαστεί να υπάρχουν φύλλα εργασίας στο αρχείο με το οποίο θα είναι συνδεδεμένος ο κώδικας *R*, από τα οποία θα υπάρχουν προκαθορισμένες τοποθεσίες με τις οποίες ο κώδικας θα αλληλεπιδρά με το Excel μέσω της κονσόλας *R*.

		Values shown represent Units Sold for the selected fields										Units
Type "s" in the pink boxes on the right to select item to forecast.		Select:	Items					Category	Platforms	Platform/Category	Platform/Item (1)	(Based on)
Week of Year	Date	S all items	ST/5104	ST/5174	BD/2821	BD/2624	BD/2901	Stationery	AZ DE : DE	Musical boxes	ST/5104	I of selected items
Week 34-2012	26/08/2012	46	41	5	0	0	0	28	0	0	41	0.51
Week 35-2012	02/09/2012	43	38	5	0	0	0	38	0	0	38	0.52
Week 36-2012	09/09/2012	14	9	5	0	0	0	30	0	0	9	0.56
Week 37-2012	16/09/2012	48	34	14	0	0	0	61	0	0	34	0.61
Week 38-2012	23/09/2012	15	2	13	0	0	0	43	0	0	2	0.63
Week 39-2012	30/09/2012	21	1	20	0	0	0	46	0	0	1	0.62
Week 40-2012	07/10/2012	22	0	22	0	0	0	65	0	0	0	0.62
Week 41-2012	14/10/2012	28	0	28	0	0	0	63	0	0	0	0.67
Week 42-2012	21/10/2012	29	0	29	0	0	0	71	0	0	0	0.71
Week 43-2012	28/10/2012	46	7	39	0	0	0	31	0	0	7	0.69
Week 44-2012	04/11/2012	86	32	54	0	0	0	60	0	0	32	0.76
Week 45-2012	11/11/2012	133	71	62	0	0	0	74	0	0	71	1.11
Week 46-2012	18/11/2012	198	133	65	0	0	0	108	0	0	133	1.50
Week 47-2012	25/11/2012	214	131	83	0	0	0	165	0	0	131	1.94
Week 48-2012	02/12/2012	346	188	158	0	0	0	302	0	0	188	2.96
Week 49-2012	09/12/2012	650	369	281	0	0	0	448	0	0	369	4.34
Week 50-2012	16/12/2012	629	338	291	0	0	0	472	0	0	338	5.01
Week 51-2012	23/12/2012	96	41	55	0	0	0	246	0	0	41	4.64

Εικόνα 49: Ταμπλό επιλογής δεδομένων για πρόβλεψη – είσοδος *R*

Η εντολή *auto.arima()* στην *R* χρησιμοποιεί μια παραλλαγή του αλγορίθμου των Hyndman και Khandakar, ο οποίος επιστρέφει το μοντέλο ARIMA που επικρατεί έπειτα από μία σειρά ελέγχων και αξιολόγησης παραμέτρων όπως η ελαχιστοποίηση των *AICc*. Οι παράμετροι *p, d, q* του μοντέλου εναλλάσσονται και δοκιμάζονται με διάφορους ελέγχους και επιλέγεται ο συνδυασμός που αποδίδει καλύτερα. (Rob J Hyndman, 2017)

```

R Console (32-bit)
File Edit Misc Packages Windows Help

> cumsum
Time Series:
Start = c(1, 1)
End = c(4, 51)
Frequency = 52
 [1] 2322 2325 2338 2330 2315 2327 2298 2265 2258 2256 2243 2254 2237 2229 2255
[16] 2241 2246 2235 2221 2228 2227 2242 2282 2278 2243 2263 2278 2253 2236 2446
[31] 2477 2472 2479 2491 2508 2516 2549 2571 2581 2600 2599 2600 2609 2611 2620
[46] 2632 2636 2641 2647 2654 2653 2655 2669 2686 2682 2690 2681 2678 2677 2678
[61] 2676 2668 2666 2657 2674 2670 2659 2668 2650 2654 2675 2692 2735 2775 2831
[76] 2926 2982 3094 3025 2873 2706 2494 2486 2507 2518 2520 2521 2533 2510 2490
[91] 2489 2478 2482 2498 2515 2515 2512 2520 2531 2545 2558 2575 2601 2616 2655
[106] 2681 2691 2699 2718 2753 2783 2797 2817 2823 2823 2815 2813 2827 2852 2881
[121] 2916 2954 2975 2998 2997 3009 3026 3084 3194 3394 3722 3904 3948 4003 4007
[136] 3981 3960 3972 3985 4002 4017 4060 4084 4092 4100 4115 4122 4152 4177 4200
[151] 4219 4236 4254 4272 4274 4282 4267 4288 4311 4342 4378 4392 4412 4413 4435
[166] 4471 4507 4552 4569 4600 4612 4629 4632 4663 4659 4655 4654 4625 4552 4394
[181] 4295 4108 4018 4099 4341 4409 4437 4485 4538 4565 4576 4547 4555 4543 4556
[196] 4580 4603 4611 4619 4627 4635 4636 4640 4649 4656 4660 4678

> auto.arima(cumsum)
Series: cumsum
ARIMA(2,1,1) (1,0,0) [52] with drift

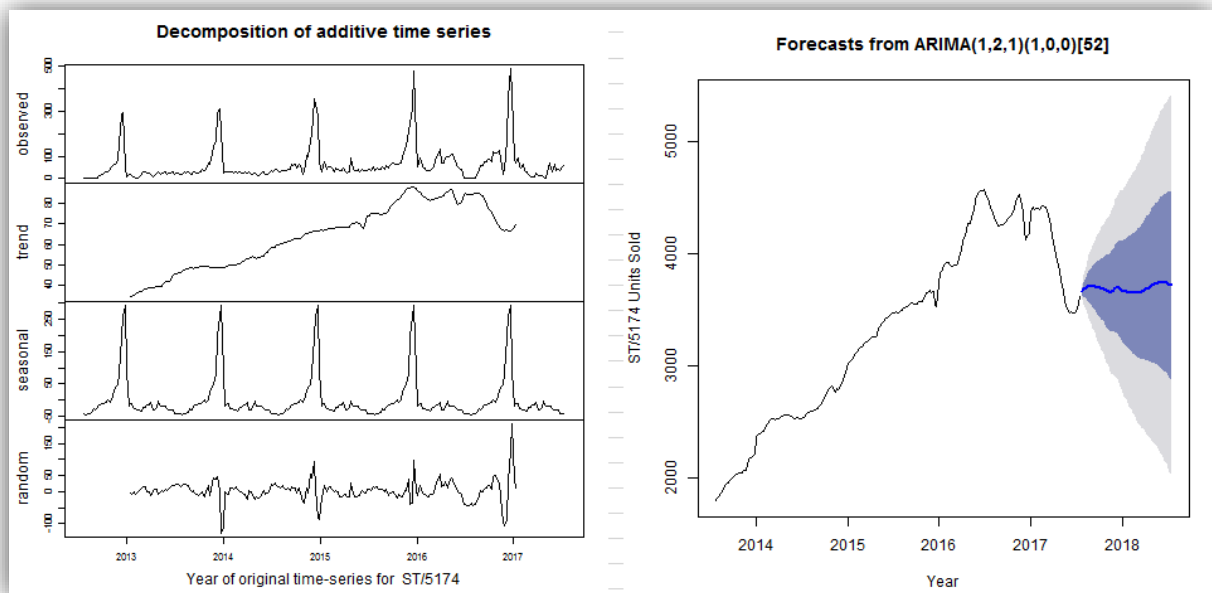
Coefficients:
      ar1      ar2      ma1      sar1      drift
      1.2502 -0.5222 -0.4938 -0.4289 12.2572
s.e.  0.1923  0.1166  0.2175  0.0730  3.5695

sigma^2 estimated as 1340:  log likelihood=-1037.06
AIC=2086.13  AICc=2086.55  BIC=2106.1

```

Εικόνα 50: Κονσόλα R και υπολογισμός `auto.arima()`

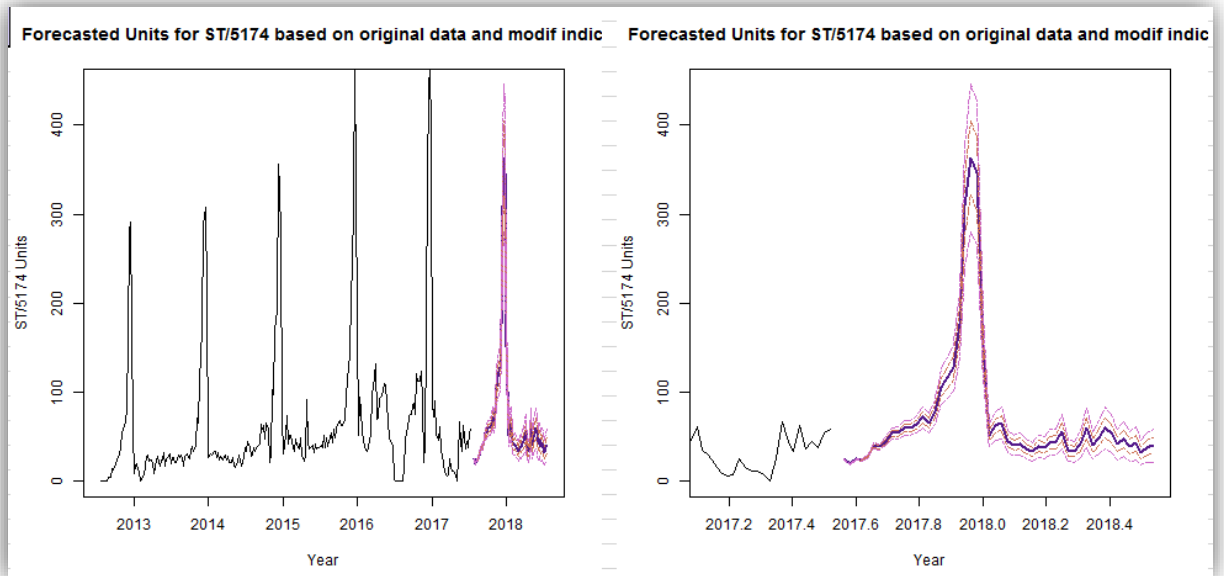
Μετά τον υπολογισμό του βέλτιστου μοντέλου ARIMA, ο κώδικας (βλ. παράρτημα) θα εφαρμόσει το μοντέλο στα δεδομένα και θα παράγει πρόβλεψη για μία προκαθορισμένη χρονική περίοδο. Συνήθως τα стоχαστικά μοντέλα αποδίδουν καλύτερα για βραχυπρόθεσμες προβλέψεις, αλλά μάλλον θα είναι επαρκές να καλύψουν τουλάχιστον έναν εποχικό κύκλο. Στη συνέχεια, ο κώδικας θα αναστρέψει το τρεχούμενο άθροισμα των τιμών και θα επαναφέρει την εποχικότητα στη χρονολογική σειρά. Ο λόγος που προτείνεται η εφαρμογή της αποσύνθεσης της χρονολογικής σειράς και η διεξαγωγή των προβλέψεων πάνω στην τάση της είναι κυρίως για λόγους απλότητας. Όσες λιγότερες παραμέτρους χρειάζεται να συμπεριληφθούν στο μοντέλο, τόσο καλύτερα μάλλον θα αποδώσει. Κατά πάσα πιθανότητα, κανένα μοντέλο δε θα μπορέσει να αναπαράγει τέλεια την απότομη άνοδο των τιμών κατά την περίοδο πριν τα Χριστούγεννα. Η πρακτική αποσύνθεσης των δεδομένων πριν την παραγωγή πρόβλεψης και ο πολλαπλασιασμός των αποτελεσμάτων της πρόβλεψης με εποχικούς συντελεστές, προσφέρει λύση στο παραπάνω πρόβλημα. Με σχετικές εντολές, οι τιμές της πρόβλεψης, ως έξοδος του μοντέλου θα εισαχθούν στο Excel από την κονσόλα R. Για την περαιτέρω κατανόηση του, το άμεσο αποτέλεσμα της πρόβλεψης μπορεί να συνοδευτεί από σχετικά γραφήματα και στατιστικές μετρήσεις.



Εικόνα 51: Αποσύνθεση δεδομένων και πρόβλεψη ARIMA – έξοδος R

R Output													
ARIMA(1,1,0)(1,0,0)[52] with drift ARIMA(1,2,1)													
				Original Data					Modified Data				
Date	Inv QTY	Week number	Forec ast	Lower 67%	Lower 95%	Upper 67%	Upper 95%	Forec ast	Lower 67%	Lower 95%	Upper 67%	Upper 95%	
20/08/2017	426	34-2017	38.18	37.85	37.51	38.52	38.86	40.66	40.33	39.99	41	41.33	
27/08/2017	387.8	35-2017	37.15	36.54	35.93	37.76	38.37	39.37	38.79	38.21	39.94	40.52	
03/09/2017	350.7	36-2017	44.12	43.1	42.06	45.15	46.19	46.47	45.53	44.58	47.41	48.36	
10/09/2017	306.5	37-2017	53.15	51.58	49.99	54.72	56.32	55.67	54.27	52.84	57.08	58.51	
17/09/2017	253.4	38-2017	53.28	51.41	49.51	55.15	57.05	55.48	53.81	52.12	57.15	58.84	
24/09/2017	200.1	39-2017	58.92	56.55	54.15	61.29	63.69	61	58.89	56.76	63.11	65.25	
01/10/2017	141.2	40-2017	59.51	56.84	54.14	62.18	64.88	61.26	58.87	56.45	63.64	66.06	
08/10/2017	81.68	41-2017	62.66	59.59	56.48	65.73	68.84	64.11	61.34	58.53	66.88	69.68	
15/10/2017	19.02	42-2017	71.04	67.28	63.47	74.81	78.62	72.36	68.93	65.45	75.79	79.27	
Outage	22/10/2017	0	43-2017	64.41	60.76	57.07	68.06	71.75	65.29	61.92	58.51	68.66	72.07
Outage	29/10/2017	0	44-2017	77.95	73.26	68.53	82.63	87.37	78.62	74.23	69.79	83	87.44
	05/11/2017	1122	45-2017	104.6	97.97	91.27	111.2	117.9	104.9	98.61	92.25	111.2	117.6
	12/11/2017	1017	46-2017	115.1	107.5	99.75	122.7	130.4	114.5	107.2	99.72	121.8	129.3
	19/11/2017	902.4	47-2017	130.7	121.7	112.6	139.7	148.7	128.7	119.9	111	137.5	146.4
	26/11/2017	771.7	48-2017	182.1	169.2	156.1	195	208	177	164.2	151.2	189.9	202.9
	03/12/2017	589.6	49-2017	318.4	295.1	271.5	341.7	365.3	306.8	283.2	259.3	330.4	354.3
	10/12/2017	271.3	50-2017	381.7	352.8	323.6	410.7	439.9	366.1	336.3	306.2	395.8	426
Outage	17/12/2017	0	51-2017	359	331	302.6	387	415.4	342.1	312.8	283.2	371.5	401.1

Εικόνα 52: Πίνακας αποτελεσμάτων πρόβλεψης με διαστήματα εμπιστοσύνης – έξοδος R



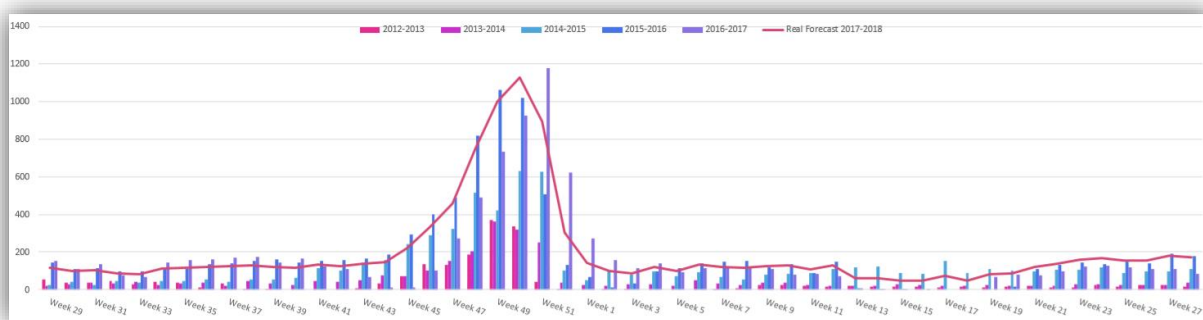
Εικόνα 53: Γραφήματα προβλέψεων ARIMA – έξοδος R

Original Data			Modified Data				
ARIMA(1,1,0)(1,0,0)[52] with drift							
Coefficients		Tests		Coefficients		Tests	
	0.575	Resid mean	0.168		0.407	Resid mean	0.107
	-0.14	Sigma^2	1225		-0.95	Sigma^2	1216
	9.858	log likelihood	-1034		-0.58	log likelihood	-1029
	-0.434	AIC	2076		38.5	AIC	2063
	38.33	AICc	2076			AICc	2064
		BIC	2089			BIC	2073

Εικόνα 54: Συντελεστές συναρτήσεων μοντέλων και κριτήρια αξιολόγησης – έξοδος R

ΕΠΙΛΟΓΟΣ

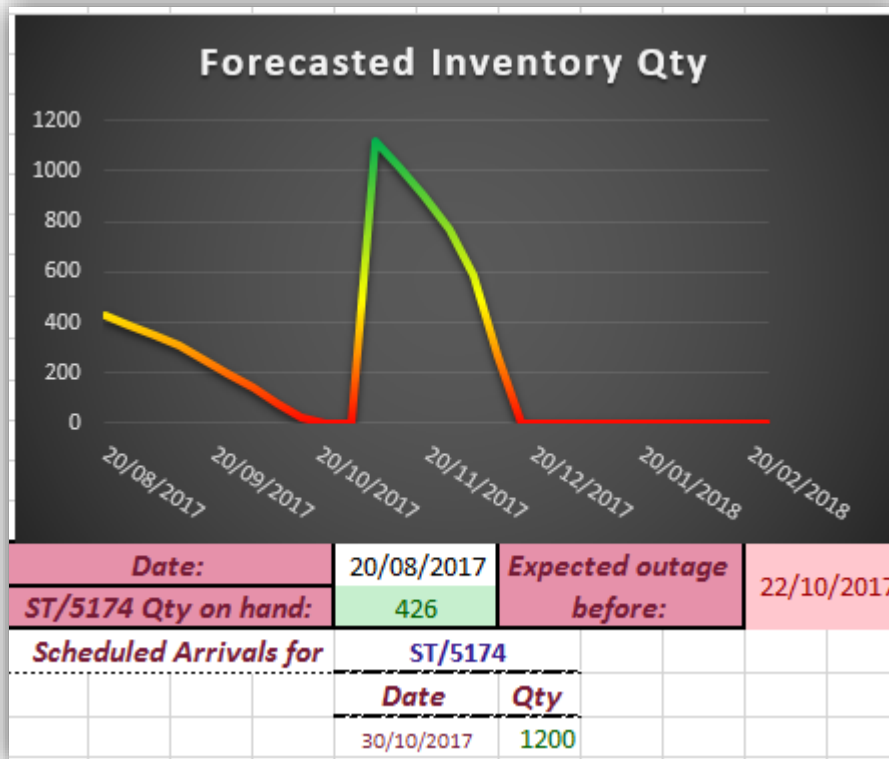
Η FG αντιμετώπιζε ένα πρόβλημα που επιζητούσε την ενόραση που προσφέρει μία πρόβλεψη. Το κύριο θέμα ήταν η διαχείριση των αποθεμάτων, για την οποία κρίθηκε απαραίτητο να προσδιοριστεί η ζήτηση των προϊόντων. Με χρήση του κώδικα *R* που εφαρμόζει τα μοντέλα ARIMA στα δεδομένα, η ανάπτυξη της ζήτησης είναι εύκολη και γρήγορη υπόθεση οπότε, πλέον μπορούν να αναπτυχθούν προβλέψεις για όλα τα προϊόντα που πουλάει η επιχείρηση.



Εικόνα 55: Πρόβλεψη μελλοντικών πωλήσεων ενός προϊόντος

Από την έξοδο του κώδικα σχετικά με τις τιμές πρόβλεψης, η FG γνωρίζει κατά προσέγγιση τη δυνητική ζήτηση για τα προϊόντα της. Χρησιμοποιώντας τη νέα γνώση, κάποιες αποφάσεις της FG θα είναι πλέον καθοδηγούμενες από τα αποτελέσματα των προβλέψεων. Η επιχείρηση μπορεί επίσης να κάνει χρήση των τιμών πρόβλεψης για διάφορους υπολογισμούς. Για παράδειγμα, για να αποφευχθεί μία επόμενη εξάντληση αποθεμάτων, έχοντας γνώση της ενδεχόμενης ζήτησης, γίνεται εύκολα να υπολογιστεί η ημερομηνία εξάντλησης των αποθεμάτων. Έτσι, η FG ίσως να προλάβει να στείλει νέα παραγγελία για εμπόρευμα έγκαιρα, ώστε το εμπόρευμα να αποσταλεί και να συμπληρώσει την προσφορά πριν τα αποθέματα εξαντληθούν. Απαραίτητη προϋπόθεση για τη συγκεκριμένη δραστηριότητα είναι η ύπαρξη της τιμής αποθεμάτων, που να φανερώνει το διαθέσιμο απόθεμα του κάθε προϊόντος για την ημερομηνία της τελευταίας παρατήρησης του δείγματος εκμάθησης. Επιπρόσθετα, τα αποτελέσματα θα επεκτείνονταν από τη γνώση σχετικά με τις μελλοντικές παραλαβές για παραγγελίες που βρίσκονται σε εξέλιξη. Κατά συνέπεια, με τις παραπάνω προϋποθέσεις, θα δημιουργηθεί μια νέα μεταβλητή χρονολογικής σειράς, η οποία για κάθε χρονική περίοδο θα αφαιρεί τα προβλεπόμενα τεμάχια από το απόθεμα της

προηγούμενης παρατήρησης. Η εισαγωγή της μεταβλητής και οι αυξομειώσεις των τιμών της φαίνονται στην *εικόνα 52*. Στην περίπτωση που αναμένεται παραλαβή αποθέματος κατά τη διάρκεια μιας περιόδου, ο αναμενόμενος αριθμός τεμαχίων εμπορεύματος θα προστεθεί στα μελλοντικά αποθέματα, ώστε να συνεχίσουν να αφαιρούνται τα προβλεπόμενα τεμάχια από τα ενημερωμένα αποθέματα, όπως φαίνεται στην *εικόνα 56*.



Εικόνα 56: Πρόβλεψη επόμενης εξάντλησης αποθεμάτων

Από τις τέσσερις μεθόδους πρόβλεψης που εφαρμόστηκαν πάνω στα δεδομένα της FG, θα μπορούσαν να χρησιμοποιηθούν μακροπρόθεσμα οι δύο· η μέθοδος της εκθετικής εξομάλυνσης και τα μοντέλα ARIMA. Και για τις δύο περιπτώσεις προτιμήθηκε η πρακτική της αποσύνθεσης των δεδομένων πριν την είσοδο τους στα μοντέλα, με τη χρήση του τρεχούμενου αθροίσματος. Μετά την παραγωγή των προβλέψεων έγινε επαναφορά των τιμών στα φυσιολογικά επίπεδα και προστέθηκε ξανά η εποχικότητα στα δεδομένα. Ο κύριος λόγος που ακολουθήθηκε η διαδικασία αυτή ήταν για την απλοποίηση των δεδομένων και τη διευκόλυνση των μοντέλων ώστε να παράγουν καλύτερες προβλέψεις. Η εποχικότητα στα δεδομένα της FG είναι σφοδρή και δεν προβλέπεται εύκολα, και μάλλον το ίδιο συμβαίνει σε όλες τις εταιρίες που εμπορεύονται στη βιομηχανία παιχνιδιών.

Η τριπλή εκθετική εξομάλυνση του Excel έκανε καλή δουλειά για τις βραχυπρόθεσμες προβλέψεις, αλλά όσο περνούσε ο χρόνος τόσο πιο ακραίες σημειώνονταν οι προβλεπόμενες τιμές. Αντίθετα, τα μοντέλα ARIMA θα μπορούσαν άνετα να προβλέψουν έναν με δύο χρονολογικούς κύκλους με παρόμοια ακρίβεια σε όλο το εύρος των προβλεπόμενων τιμών της. Η μέθοδος εξομάλυνσης κινούμενων μέσων χρησιμοποιήθηκε με παραλλαγή για να πραγματοποιήσει πρόβλεψη για τιμές κατά τη διάρκεια των ιστορικών δεδομένων. Η μέθοδος προτιμήθηκε για την περίπτωση αυτή ώστε τα αποτελέσματα των προβλέψεων να είναι ιδιαίτερα περιορισμένα: δεν χρειάζεται να συμπεριληφθούν ακραίες τιμές για τη συμπλήρωση των χαμένων πωλήσεων, ή οτιδήποτε τυχαίο. Για το συγκεκριμένο πρόβλημα θα μπορούσε κάλλιστα να γίνει χρήση της εκθετικής εξομάλυνσης ή μοντέλων ARIMA, εφόσον και οι τρεις μέθοδοι αφορούν ανάλυση χρονολογικών σειρών. Τέλος, η πολλαπλή γραμμική παλινδρόμηση θα μπορούσε να θεωρηθεί εντυπωσιακή με τους συντελεστές προσδιορισμού της που υποστηρίζανε 99% ακρίβεια των μοντέλων. Δυστυχώς, τα δεδομένα της FG δεν ήταν κατάλληλα για να επιτραπεί περαιτέρω εξάσκηση με την μέθοδο, εφόσον δεν υπάρχουν ανεξάρτητες μεταβλητές διαθέσιμες που να είναι γνωστές μελλοντικά.

Συμπερασματικά, η επιλογή της μεθόδου πρόβλεψης για οποιοδήποτε πρόβλημα εξαρτάται κυρίως από τη φύση των δεδομένων. Υπάρχει μεγάλο πλήθος διάφορων μεθόδων πρόβλεψης και κατά τη διαδικασία εύρεσης της βέλτιστης μεθόδου, μάλλον θα πρέπει να λαμβάνονται υπόψη τα εξής ερωτήματα: υπάρχουν τα μέσα για να υλοποιηθεί η πρόβλεψη (πχ. λογισμικό, δεδομένα, κλπ.) και θα μπορούσε η φύση των αποτελεσμάτων να αποτελέσει λύση του προβλήματος για το οποίο είναι επιθυμητή η πρόβλεψη; Τα παραπάνω ερωτήματα αφορούν μεθόδους προεκβολής, δηλαδή μεθόδους που βασίζονται σε ιστορικά δεδομένα. Αν το υποσύνολο των προβλέψεων προεκβολής επεκταθεί ώστε να συμπεριλαμβάνει αιτιοκρατικές και ενστικτώδης μεθόδους, τότε οι επιλογές σίγουρα αυξάνονται σημαντικά. Εφόσον βρεθούν μερικές μέθοδοι πρόβλεψης που να ταιριάζουν με το πρόβλημα και τα δεδομένα, η επιλογή του τελικού μοντέλου θα πρέπει να βασίζεται στα αποτελέσματα από σχετικά κριτήρια και ελέγχους αποδόσεων. Ακόμα και μετά την τελική επιλογή μοντέλου και τη διεξαγωγή προβλέψεων, το τέλειο μοντέλο μπορεί να μην είναι το ίδιο τέλειο μετά από κάποιο χρονικό διάστημα. Συνεπώς, αν οι προβλέψεις παράγονται συχνά, καλό θα ήταν να ελέγχεται η απόδοση των μοντέλων και το μοντέλο να συντηρείται με μετατροπές όπου χρειάζεται.

ΠΑΡΑΡΤΗΜΑΤΑ

I. Τεχνολογίες πληροφορικής

Ως το πληροφοριακό της σύστημα, η FG χρησιμοποιεί το NetSuite της Oracle, που είναι εφαρμογή υπολογιστικού νέφους, δηλαδή cloud-based Enterprise Resource Planning (ERP). Εκεί αποθηκεύει τα δεδομένα της και κάνει διαχείριση των διεργασιών της. Πρόσβαση στα δεδομένα από τις πωλήσεις της Amazon αποκτά από την πλατφόρμα αναλυτικών Amazon Seller Central (sellercentral.amazon.co.uk) και σχετικά με τις πωλήσεις στο Ebay από το Seller Hub (ebay.co.uk/sh). Ένας από τους κύριους στόχους και οράματα της FG είναι η αυτοματοποίηση μιας ψηφιακής μορφής του οργανισμού. Ο υπεύθυνος πληροφορικής της επιχείρησης εργάζεται κυρίως στο Excel της Microsoft όπου με κώδικα Visual Basic for Applications (VBA) ρυθμίζει και διαχειρίζεται την αυτοματοποίηση. Χρησιμοποιεί Application Programming Interface (API) για να λάβει όσα περισσότερα δεδομένα μπορεί από τις πλατφόρμες που του το επιτρέπουν. Στο παρελθόν η FG χρησιμοποιούσε το Dynamics NAV ERP της Microsoft, οπότε οι κωδικοί προϊόντων έχουν παραμείνει στη μορφοποίηση που υποστηριζόταν στο παλιό σύστημα.

Στο πλαίσιο της παρούσας εργασίας, η συλλογή των δεδομένων θα γίνει από το ASC και από τη βάση δεδομένων του NetSuite, όπου χρειαστεί. Η οργάνωση και ανάλυση των δεδομένων θα πραγματοποιηθεί στο Microsoft Excel. Για την εφαρμογή κάποιων στατιστικών μοντέλων θα χρησιμοποιηθεί η ενσωματωμένη επέκταση Analysis Toolpak. Για την ανάπτυξη στοχαστικών μοντέλων θα χρησιμοποιηθεί η γλώσσα προγραμματισμού R με τη βασική της κονσόλα που διανέμεται δωρεάν και είναι ανοιχτού κώδικα. Τελικά, ο κώδικας θα ενσωματωθεί στο Excel και θα τρέχει με χρήση μάκρο-εντολών με κώδικα VBA. Εναλλακτικά ήταν δυνατή η χρήση του RStudio που φέρει σημαντική γραφική διεπαφή χρήστη (Graphic User Interface – GUI) και είναι επίσης δωρεάν και ανοιχτού κώδικα.

II. Κώδικας R: Τροποποίηση δεδομένων για ανάκτηση «χαμένων» πωλήσεων

```
library(forecast)
library(tseries)
library(zoo)
library(excel.link)

startyear = xl[b10]
startweek = xl[a10]
data = ts(xlc[E9:N269, xl.sheet.name = "R input"],
frequency=365.25/7, start=startyear+startweek*7/365.25)
count=xl[o588:x588]
mod_data=data
meandata=data[1:ncol(data)]
launch=data[1:ncol(data)]

i=1
while(i<=length(data[1,])){ meandata[i] = c(sum(data[,i])/count[i])
i=i+1 }

limit = round(unlist(meandata)*20/100, 0)

i=1
j=1
pos=1
while(i<=ncol(data)){
  while(data[j,i]==0){ pos=pos+1
  j=j+1 }
  launch[i]=pos
  j=1
  pos=1
  i=i+1 }

i=1
j=1
while(i<=ncol(data)){ ### Loop per column
```

```

while(j<=nrow(data)){ ### Loop per row

### CONDITIONS of modifications in order of appearance:
### 1. No modification can happen within the first 7 rows of a column, these lines are excluded
completely.
### - Modifications will only start occurring 7 rows after the initial launch of an item.
### 2. Weeks 52 and 1:7 of a year cannot be altered based on the average of the Christmas sales,
so they're excluded in the first alteration approach.
### 3. The 110% of the sum of the previous 8 values when divided by 8 must be higher than 0.

if( j>launch[i]+7 &&
(if(j>52-startweek+52*4 && j<52-startweek+52*4+8){
    1>2
} else if(j>52-startweek+52*3 && j<52-startweek+52*3+8){
    1>2
} else if(j>52-startweek+52*2 && j<52-startweek+52*2+8){
    1>2
} else if(j>52-startweek+52 && j<52-startweek+52+8){
    1>2
} else if(j>52-startweek && j<52-startweek+8){
    1>2
} else{
    1<2))
    && sum(mod_data[j-1,i],mod_data[j-2,i],mod_data[j-3,i],mod_data[j-
4,i],mod_data[j-5,i],mod_data[j-6,i],mod_data[j-7,i],mod_data[j-8,i])/8*1.1>0){
        if(data[j,i]<=limit[i]){
### The line above checks if the value needs to be modified based on the limit set above.
            mod_data[j,i]=round(sum(mod_data[j-1,i],mod_data[j-
2,i],mod_data[j-3,i],mod_data[j-4,i],mod_data[j-5,i],mod_data[j-6,i],mod_data[j-7,i],mod_data[j-
8,i])/8*1.1,0) }
            else{ mod_data[j,i]=data[j,i] } }
### If value is modified then it becomes the 110% of the average value of the previous 8 weeks.
### If value is within the weeks after Christmas AND after the initial launch of the item.
### If it needs to be modified, the updated value is 30% of the average value of the column.
        else if(data[j,i]<=limit[i]){
            if((if(j>52-startweek+52*4 && j<52-startweek+52*4+8 && j>launch[i]+7){
                1<2
            } else if(j>52-startweek+52*3 && j<52-startweek+52*3+8 &&
j>launch[i]+7){

```

```

1<2
}else if(j>52-startweek+52*2 && j<52-startweek+52*2+8 &&
j>launch[i]+7){
1<2
}else if(j>52-startweek+52 && j<52-startweek+52+8 && j>launch[i]+7){
1<2
}else if(j>52-startweek && j<52-startweek+8 && j>launch[i]+7){
1<2
}else{
1>2})
{ mod_data[j,i]=round(unlist(meandata[i])*30/100)}
else{
mod_data[j,i]=data[j,i]}
j=j+1}
j=1
i=i+1}

xlc[e328]=mod_data

### Process Complete. ###

```

III. Κώδικας R: ARIMA

```
library(excel.link)
library(forecast)
library(tseries)
library(zoo)

# add data to viriables, data is being pulled from "R input"
xl.sheet.activate("R input") # other commands are .add and .delete
startyear = xl[b10]
startweek = xl[a10]
data = ts(xlc[E9:N269, xl.sheet.name = "R
input"],frequency=365.25/7,start=startyear+startweek*7/365.25)
mod_data = ts(xlc[E328:N588, xl.sheet.name = "R
input"],frequency=365.25/7,start=startyear+startweek*7/365.25)
headers = xl[E8:N8]
sigma_data = ts(rollapply(data,52,sum),frequency=365.25/7,start=startyear+1+startweek*7/365.25)
sigma_mod_data =
ts(rollapply(mod_data,52,sum),frequency=365.25/7,start=startyear+1+startweek*7/365.25)
seasonal_index = xlc[O9:X269,xl.sheet.name = "R input"]
tickbox = xl[E4:N4,xl.sheet.name = "R input"]

# =====
# the code below finds first ticked column for which the forecast with happen
# =====
cell=1
i=1
while(i<=length(tickbox)){ cell[i] = c(tickbox[i])
i=i+1 }

x = 1
i = 1
while(i<=length(cell)){ if(cell[i]=="a"){ x=i
break }else{i = i+1 } }

target=headers[x]

# x stands for the number of column from the musical boxes table. 1 is sum, 2 is Item_1, 3 is Item_2
etc.
```



```

# =====
# forecasting...
# =====
model = auto.arima(sigma_data[,x],approximation=FALSE)
frst = forecast(model,h=52,level=c(67,95)) #### the values within the parenthesis for the "level"
element are the set confidence bounds
xl.sheet.activate("Basic Forecast") # other commands are .add and .delete
xl[e18] = toString(model) # tells which model was used for the forecast above its charts

# fitting the trend into the model and re-attaching seasonality
mean=frst$mean/52*tail(seasonal_index[,x],52)
lower=frst$lower/52*tail(seasonal_index[,x],52)
upper=frst$upper/52*tail(seasonal_index[,x],52)

output=cbind(mean,lower,upper)
xl[e22]=output

# =====
# forecasting MODIFIED...
# =====
mod_model = auto.arima(sigma_mod_data[,x],approximation=FALSE)
mod_frst = forecast(mod_model,h=52,level=c(67,95))
xl.sheet.activate("Basic Forecast") # other commands are .add and .delete
xl[j18] = toString(mod_model) # tells which model was used for the forecast above its charts

# fitting the trend into the model and re-attaching seasonality
mod_mean=mod_frst$mean/52*tail(seasonal_index[,x],52)
mod_lower=mod_frst$lower/52*tail(seasonal_index[,x],52)
mod_upper=mod_frst$upper/52*tail(seasonal_index[,x],52)

mod_output=cbind(mod_mean,mod_lower,mod_upper)
xl[j22]=mod_output

### Process Complete. ###

```

IV. Κώδικας R: Αξιολόγηση μοντέλου και γραφήματα

```
library(excel.link)
library(forecast)
library(tseries)
library(zoo)

# add data to viriables, data is being pulled from "R input"
xl.sheet.activate("R input") # other commands are .add and .delete
startyear = xl[b10]
startweek = xl[a10]
data = ts(xlc[E9:N269, xl.sheet.name = "R
input"],frequency=365.25/7,start=startyear+startweek*7/365.25)
mod_data = ts(xlc[E328:N588, xl.sheet.name = "R
input"],frequency=365.25/7,start=startyear+startweek*7/365.25)
headers = xl[E8:N8]
sigma_data = ts(rollapply(data,52,sum),frequency=365.25/7,start=startyear+1+startweek*7/365.25)
sigma_mod_data =
ts(rollapply(mod_data,52,sum),frequency=365.25/7,start=startyear+1+startweek*7/365.25)
seasonal_index = xlc[O9:X269,xl.sheet.name = "R input"]
checkbox = xl[E4:N4,xl.sheet.name = "R input"]

# =====
# the code below finds first ticked column for which the forecast with happen
# =====
cell=1
i=1
while(i<=length(checkbox)){ cell[i] = c(checkbox[i])
i=i+1 }

x = 1
i = 1
while(i<=length(cell)){ if(cell[i]=="a"){
x=i
break }else{
i = i+1 }}
target=headers[x]
```

```

# =====
# the function below will be used to plot histograms of residuals
# =====
plotForecastErrors <- function(forecasterrors)
{# make a histogram of the forecast errors:
mybinsize <- IQR(forecasterrors)/4
mysd <- sd(forecasterrors)
mymin <- min(forecasterrors) - mysd*5
mymax <- max(forecasterrors) + mysd*3
# generate normally distributed data with mean 0 and standard deviation mysd
mynorm <- rnorm(10000, mean=0, sd=mysd)
mymin2 <- min(mynorm)
mymax2 <- max(mynorm)
if (mymin2 < mymin) { mymin <- mymin2 }
if (mymax2 > mymax) { mymax <- mymax2 }
# make a histogram of the forecast errors, with the normally distributed data overlaid:
mybins <- seq(mymin, mymax, mybinsize)
hist(forecasterrors, col="orchid3", freq=FALSE, breaks=mybins, xlab=paste("Residual Histogram
for",target))
# freq=FALSE ensures the area under the histogram = 1
# generate normally distributed data with mean 0 and standard deviation mysd
myhist <- hist(mynorm, plot=FALSE, breaks=mybins)
# plot the normal curve line on top of the histogram of forecast errors:
points(myhist$mids, myhist$density, type="l", col="purple4", lwd=2)}

# x stands for the number of column from the musical boxes table. 1 is sum, 2 is Item_1, 3 is Item_2
etc.
xl.sheet.activate("Advanced Forecast")
xl[b1] = paste("Data and forecast for",headers[x])
x11()
plot(decompose(data[,x]),xlab=paste("Year of original time-series for ",target))
xl[w40] = current.graphics(type="png",picname="decomp")
plot(decompose(mod_data[,x]),xlab=paste("Year of modified time-series for ",target))
xl[ai65] = current.graphics(type="png",picname="mod_decomp")
dev.off()

# =====
# forecasting....

```

```

# =====
model = auto.arima(sigma_data[,x],approximation=FALSE)
xl[e18] = toString(model) # tells which model was used for the forecast above its charts
frcst = forecast(model,h=52,level=c(67,95))

# fitting the trend into the model and re-attaching seasonality
mean=frcst$mean/52*tail(seasonal_index[,x],52)
lower=frcst$lower/52*tail(seasonal_index[,x],52)
upper=frcst$upper/52*tail(seasonal_index[,x],52)

output=cbind(mean,lower,upper)
xl[e22]=output

# =====
# inserting forecast information and tests output into excel
# =====
# model coefficients
i=1
coef=1
while(i<=length(model$coef)){coef[i] = list(model$coef[i])
i=i+1 }
xl[d77]=coef

# mean value of Residuals fitted to regular values (reversed cumsum) with readjusted seasonality.
mresid=mean(model$residuals/52*tail(as.numeric(seasonal_index[,x]),length(seasonal_index[,x])-
51))
# sigma square, log likelihood, AIC, AICc, BIC, variance coefficients
sigmasq=model$sigma2
llik=model$loglik
aicresid=model$aic
aiccresid=model$aicc
bicresid=model$bic
xl[h77]= rbind(mresid,sigmasq,llik,aicresid,aiccresid,bicresid)

x11()
plotForecastErrors(frcst$residuals)
xl[a85]=current.graphics(type="png",picname="resid_hist")

```

```

# plotting the forecast
plot(frcst, xlab="Year", ylab= paste(target,"Units Sold"))
xl[w65] = current.graphics(type="png", picname = "sigma_forecast")

ts.plot(data[,x], ylab= paste(target,"Units"), xlab="Year",
ylim=c(0,max(frcst$upper[,2]/52*tail(seasonal_index[,x],52))),
xlim=c(startyear+startweek*7/365.25,startyear+6+(startweek-1)*7/365.25))
lines(frcst$mean/52*tail(seasonal_index[,x],52),col="purple4",lwd = 2)
lines(frcst$upper[,1]/52*tail(seasonal_index[,x],52),col="salmon3",lwd = 1,lty=6)
lines(frcst$lower[,1]/52*tail(seasonal_index[,x],52),col="salmon3",lwd = 1,lty=6)
lines(frcst$upper[,2]/52*tail(seasonal_index[,x],52),col="orchid3",lwd = 1,lty=6)
lines(frcst$lower[,2]/52*tail(seasonal_index[,x],52),col="orchid3",lwd = 1,lty=6)
title(main=paste("Forecasted Units for",target,"based on original data and MOD indices"))
xl[w90] = current.graphics(type="png",picname="forecast")

# same plot as above but only shows last observations and predicted values
ts.plot(data[,x], ylab= paste(target,"Units"), xlab="Year",
ylim=c(0,max(frcst$upper[,2]/52*tail(seasonal_index[,x],52))),
xlim=c(startyear+4+(startweek+30)*7/365.25,startyear+6+(startweek-1)*7/365.25))
lines(frcst$mean/52*tail(seasonal_index[,x],52),col="purple4",lwd = 2)
lines(frcst$upper[,1]/52*tail(seasonal_index[,x],52),col="salmon3",lwd = 1,lty=6)
lines(frcst$lower[,1]/52*tail(seasonal_index[,x],52),col="salmon3",lwd = 1,lty=6)
lines(frcst$upper[,2]/52*tail(seasonal_index[,x],52),col="orchid3",lwd = 1,lty=6)
lines(frcst$lower[,2]/52*tail(seasonal_index[,x],52),col="orchid3",lwd = 1,lty=6)
title(main= paste("Forecasted Units for",target,"based on original data and MOD indices"))
xl[ai90] = current.graphics(type="png",picname="forecast_zoom")

# =====
# forecasting MODIFIED...
# =====
mod_model = auto.arima(sigma_mod_data[,x],approximation=FALSE)
mod_frcst = forecast(mod_model,h=52,level=c(67,95))
xl.sheet.activate("Basic Forecast") # other commands are .add and .delete
xl[j18] = toString(mod_model) # tells which model was used for the forecast above its charts

# fitting the trend into the model and re-attaching seasonality
mod_mean=mod_frcst$mean/52*tail(seasonal_index[,x],52)
mod_lower=mod_frcst$lower/52*tail(seasonal_index[,x],52)
mod_upper=mod_frcst$upper/52*tail(seasonal_index[,x],52)

```

```

mod_output=cbind(mod_mean,mod_lower,mod_upper)
xl[j22]=mod_output

# =====
# inserting forecast information and tests output into excel
# =====

# model coefficients
i=1
mod_coef=1
while(i<=length(mod_model$coef)){mod_coef[i] = list(mod_model$coef[i])
i=i+1 }
xl[l77]=mod_coef

# mean value of Residuals fitted to regular values (reversed cumsum) with readjusted seasonality.
mod_mresid=mean(mod_model$residuals/52*tail(as.numeric(seasonal_index[,x]),length(seasonal_in
dex[,x])-51))

# sigma square, log likelihood, AIC, AICc, BIC, variance coefficients
mod_sigmasq=mod_model$sigma2
mod_llik=mod_model$loglik
mod_aicresid=mod_model$aic
mod_aicresid=mod_model$aicc
mod_bicresid=mod_model$bic
xl[o77]= rbind(mod_mresid,mod_sigmasq,mod_llik,mod_aicresid,mod_aicresid,mod_bicresid)

x11()
plotForecastErrors(mod_frcst$residuals)
xl[m85]=current.graphics(type="png",picname="mod_resid_hist")

# plotting the forecast
plot(frcst, xlab="Year", ylab= paste(target,"MOD Units Sold"))
xl[ai65] = current.graphics(type="png", picname = "mod_sigma_forecast")

ts.plot(mod_data[,x], ylab= paste(target,"MOD Units"), xlab="Year",
ylim=c(0,max(mod_frcst$upper[,2]/52*tail(seasonal_index[,x],52))),
xlim=c(startyear+startweek*7/365.25,startyear+6+(startweek-1)*7/365.25))
lines(mod_frcst$mean/52*tail(seasonal_index[,x],52),col="purple4",lwd = 2)
lines(mod_frcst$upper[,1]/52*tail(seasonal_index[,x],52),col="salmon3",lwd = 1,lty=6)

```

```

lines(mod_frcst$lower[,1]/52*tail(seasonal_index[,x],52),col="salmon3",lwd = 1,lty=6)
lines(mod_frcst$upper[,2]/52*tail(seasonal_index[,x],52),col="orchid3",lwd = 1,lty=6)
lines(mod_frcst$lower[,2]/52*tail(seasonal_index[,x],52),col="orchid3",lwd = 1,lty=6)
title(main=paste("Forecasted Units for",target,"based on MOD data and MOD indices"))
xl[w115] = current.graphics(type="png",picname="mod_forecast")

# same plot as above but only shows last observations and predicted values
ts.plot(mod_data[,x], ylab= paste(target,"MOD Units"), xlab="Year",
ylim=c(0,max(mod_frcst$upper[,2]/52*tail(seasonal_index[,x],52))),
xlim=c(startyear+4+(startweek+30)*7/365.25,startyear+6+(startweek-1)*7/365.25))
lines(mod_frcst$mean/52*tail(seasonal_index[,x],52),col="purple4",lwd = 2)
lines(mod_frcst$upper[,1]/52*tail(seasonal_index[,x],52),col="salmon3",lwd = 1,lty=6)
lines(mod_frcst$lower[,1]/52*tail(seasonal_index[,x],52),col="salmon3",lwd = 1,lty=6)
lines(mod_frcst$upper[,2]/52*tail(seasonal_index[,x],52),col="orchid3",lwd = 1,lty=6)
lines(mod_frcst$lower[,2]/52*tail(seasonal_index[,x],52),col="orchid3",lwd = 1,lty=6)
title(main= paste("Forecasted Units for",target,"based on MOD data and MOD indices"))
xl[ai115] = current.graphics(type="png",picname="mod_forecast_zoom")

### Process Complete. ###

```

ΒΙΒΛΙΟΓΡΑΦΙΑ

- 1) *Wikimedia File:Correlation examples.png*. (2010, 02 02). Ανάκτηση από Wikimedia Commons, the free media repository:
https://commons.wikimedia.org/wiki/File:Correlation_examples.png
- 2) *Wikimedia File:Standard deviation diagram.svg*. (2015, 01 15). Ανάκτηση από Wikimedia Commons, the free media repository:
https://commons.wikimedia.org/wiki/File:Standard_deviation_diagram.svg
- 3) *Wikipedia: Bayesian network*. (2017, 07 17). Ανάκτηση από Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Bayesian_network
- 4) *Wikipedia: Data analysis*. (2017, 08 04). Ανάκτηση από Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Data_analysis
- 5) *Wikipedia: Descriptive Statistics*. (2017, 08 06). Ανάκτηση 08 03, 2017, από Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Descriptive_statistics
- 6) *Wikipedia: Functional data analysis*. (2017, 06 11). Ανάκτηση από Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Functional_data_analysis
- 7) *Wikipedia: Geometric data analysis*. (2017, 06 10). Ανάκτηση από Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Geometric_data_analysis
- 8) *Wikipedia: Statistical shape analysis*. (2017, 05 30). Ανάκτηση από Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Statistical_shape_analysis
- 9) *Wikipedia: Structured data analysis (statistics)*. (2017, 04 04). Ανάκτηση 04 15, 2017, από wikipedia.org: [en.wikipedia.org/wiki/Structured_data_analysis_\(statistics\)](https://en.wikipedia.org/wiki/Structured_data_analysis_(statistics))
- 10) *Wikipedia: Topological data analysis*. (2017, 08 10). Ανάκτηση από Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Topological_data_analysis
- 11) Allison, P. D. (1999). *Multiple Regression - a primer*. Pine Forge Press, Inc.
- 12) *Amazon.co.uk - Business Reports Glossary*. (n.d.). Ανάκτηση από Amazon:
https://www.amazon.co.uk/gp/help/customer/display.html/ref=help_search_1-1?ie=UTF8&nodeId=201369630

- 13) Anasse Bari, M. C. (2014). *Predictive Analytics For Dummies*.
- 14) David P. Doane, L. E. (2016). Time-Series Analysis - Exponential Smoothing. Στο L. E. David P. Doane, *Applied Statistics in Business and Economics Fifth Edition*. McGraw Hill Education.
- 15) David P. Doane, L. E. (2016). Time-Series Analysis - Moving Averages. Στο L. E. David P. Doane, *Applied Statistics in Business and Economics Fifth Edition*. McGraw Hill Education.
- 16) David R. Anderson, D. J. (2014). *Statistics for Business and Economics, Third Edition*. Cengage Learning EMEA.
- 17) Douglas A. Lind, W. G. (2015). Correlation and linear regression. Στο W. G. Douglas A. Lind, *Statistical Techniques in Business & Economics*. Mc Graw Hill Education.
- 18) Douglas A. Lind, W. G. (2015). Time series and forecasting. Στο W. G. Douglas A. Lind, *Statistical Techniques in Business and Economics*. McGraw Hill Education.
- 19) Εμίρης, Δ. (2012). Προβλέψεις. Πειραιάς.
- 20) Gilchrist, W. (1976). An introduction to forecasting. Στο W. Gilchrist, *Statistical Forecasting*.
- 21) Hyndman, R. J. (2011). *Cyclic and seasonal time series*. Ανάκτηση από Hyndsight: <https://robjhyndman.com/hyndsight/cyclicts/>
- 22) Keith Ord, R. F. (2013). How to measure forecasting accuracy. Στο R. F. Keith Ord, *Principles of business forecasting* (σσ. 41-48).
- 23) Keith Ord, R. F. (2013). Types of Data. Στο R. F. Keith Ord, *Principles of Business Forecasting*.
- 24) Matteson, D. S. (2017). *Introduction to Time Series Analysis*. Ανάκτηση από DataCamp: <https://www.datacamp.com/courses/introduction-to-time-series-analysis>
- 25) Mentzer, J. T. (2005). Time series forecasting techniques. Στο J. T. Mentzer, *Sales forecasting management - a demand management approach*. SAGE Publications Inc.
- 26) Murphy, K. (1998). *A Brief Introduction to Graphical Models and Bayesian Networks*. Ανάκτηση από <https://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html>
- 27) Murray R. Spiegel, L. J. (2014). *Statistics - Fifth Edition*. McGraw-Hill Education.

- 28) Priss, U. (2017). *Formal Concept Analysis*. Ανάκτηση από Uta Priss:
<http://www.upriss.org.uk/fca/fca.html>
- 29) Richard J. Roiger, M. W. (2008). Εξειδικευμένες τεχνικές - Ανάλυση χρονοσειρών.
Στο M. W. Richard J. Roiger, *Εξόρυξη πληροφορίας - ένας εισαγωγικός οδηγός με παραδείγματα*. Εκδόσεις Κλειδάριθμος.
- 30) Richard J. Roiger, M. W. (2008). *Εξόρυξη πληροφορίας - ένας εισαγωγικός οδηγός με παραδείγματα*. Εκδόσεις Κλειδάριθμος.
- 31) Rob J Hyndman, G. A.s.l. (2017). *Forecasting: Principles and Practice - ARIMA models - ARIMA modelling in R*. Ανάκτηση από OTexts - Online open-access textbooks: <https://www.otexts.org/fpp/8/7>
- 32) Rob J Hyndman, G. A.s.l. (2017). *Forecasting: Principles and Practice - ARIMA models - Autoregressive models*. Ανάκτηση από OTexts - Online open-access textbooks: <https://www.otexts.org/fpp/8/3>
- 33) Rob J Hyndman, G. A.s.l. (2017). *Forecasting: Principles and Practice - ARIMA models - Estimation and order selection*. Ανάκτηση από OTexts - Online open-access textbooks: <https://www.otexts.org/fpp/8/6>
- 34) Rob J Hyndman, G. A.s.l. (2017). *Forecasting: Principles and Practice - ARIMA models - Stationarity and differencing*. Ανάκτηση από OTexts - Online open-access textbooks: <https://www.otexts.org/fpp/8/1>
- 35) Rob J Hyndman, G. A.s.l. (2017). *Forecasting: Principles and Practice - The forecaster's toolbox - Evaluating forecast accuracy*. Ανάκτηση από OTexts - Online open-access textbooks: <https://www.otexts.org/fpp/2/5>
- 36) Rob J Hyndman, G. A.s.l. (2017). *Forecasting: Principles and Practice - The forecaster's toolbox - Numerical data summaries*. Ανάκτηση από OTexts - Online open-access textbooks: <https://www.otexts.org/fpp/2/2>
- 37) Rob J. Hyndman, G. A. (2017). *Forecasting: Principles and Practice - The forecaster's toolbox - Transformations and adjustments*. Ανάκτηση από OTexts - Online open-access textbooks: <https://www.otexts.org/fpp/2/4>

- 38) Runkel, P. (2016, 04 08). *What Are Degrees of Freedom in Statistics?* Ανάκτηση από The Minitab Blog: <http://blog.minitab.com/blog/statistics-and-quality-data-analysis/what-are-degrees-of-freedom-in-statistics>
- 39) Spyros Makridakis, S. C. (1983). *Forecasting: Methods and Applications - 2nd Ed.* John Wiley & Sons Inc.
- 40) Spyros Makridakis, S. C. (1998). *Forecasting: Methods and Applications - 3rd Ed.*
- 41) Vermorel, J. (2012). *Calculate safety stocks with sales forecasting.* Ανάκτηση από lokad.com: <https://www.lokad.com/calculate-safety-stocks-with-sales-forecasting>
- 42) Walkenbach, J. (2007). Conveying a message with a chart. Στο J. Walkenbach, *Microsoft Office Excel 2007 Charts.* Wiley Publishing, Inc.
- 43) Walkenbach, J. (2007). Introducing Excel Charts. Στο J. Walkenbach, *Microsoft Office Excel 2007 Charts.* Wiley Publishing, Inc.
- 44) Παππής, Κ. (2006). Προβλέψεις. Στο Κ. Παππής, *Προγραμματισμός παραγωγής.* Εκδόσεις Σταμούλη Α.Ε.

Πνευματικά δικαιώματα

Copyright © ΤΕΙ Δυτικής Ελλάδας. Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν. 1599/1988 και τα άρθρα 2,4,6 παρ. 3 του Ν. 1256/1982, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής εργασίας και δεν προσβάλλει κάθε μορφής πνευματικά δικαιώματα τρίτων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον.

Ελένη Μαρία Λεπεσιώτη, 2017