



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών
«Κυβερνοασφάλεια και Επιστήμη Δεδομένων»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	Τμηματοποίηση πελατών και πρόβλεψη πωλήσεων για μια ηλεκτρονική πλατφόρμα πωλήσεων Customer Segmentation and Sales Forecast for E-Shop
Όνοματεπώνυμο Φοιτητή	Αλεξάνδρα Παπανίκου
Πατρώνυμο	Χαρίσιος
Αριθμός Μητρώου	ΜΠΚΕΔ21041
Επιβλέπων	Αποστόλου Δημήτριος, Καθηγητής

Ημερομηνία Παράδοσης **Ιούλιος 2024**

Στην ολοκλήρωση της παρούσας μεταπτυχιακής διατριβής, ιδιαίτερα σημαντική ήταν η συμβολή του Διδάσκοντα του ΠΜΣ κ. Ανδρέα Ζάρα, που προσέφερε επιστημονική και συμβουλευτική καθοδήγηση σε όλα τα στάδια εκπόνησής της.

Τριμελής Εξεταστική Επιτροπή

Δημήτριος Αποστόλου
Καθηγητής

Άγγελος Πικράκης
Επίκουρος Καθηγητής

Διονύσιος
Σωτηρόπουλος
Επίκουρος Καθηγητής

Η παρούσα μεταπτυχιακή διατριβή είναι αφιερωμένη στην αδερφή μου Κατερίνα και στη μνήμη του παππού μου Ευάγγελου.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω την αμέριστη εκτίμησή μου κ. Ανδρέα Ζάρα, Διδάσκοντα στο Π.Μ.Σ. για την υποστήριξή του, την καθοδήγησή του και τις πολύτιμες συμβουλές του καθ' όλη τη διάρκεια αυτής της διατριβής. Η γνώση και η εμπειρία του αποτέλεσαν πηγή έμπνευσης και καθοδήγησης για μένα.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για την αδιάκοπη υποστήριξη, την κατανόηση και την ενθάρρυνσή τους καθ' όλη τη διάρκεια των σπουδών μου. Η αγάπη και η υποστήριξή τους ήταν και θα παραμείνουν για πάντα πηγή δύναμης και έμπνευσης για μένα.

Πίνακας Περιεχομένων

Κατάλογος Εικόνων	7
Κατάλογος Πινάκων.....	9
Κατάλογος Συντομογραφιών.....	10
Περίληψη	11
Abstract.....	11
1. Εισαγωγή.....	12
1.1 Γενικά	12
1.2 Σκοπός Διπλωματικής	13
2 Big Data.....	14
2.1 Χαρακτηριστικά Big Data.....	14
2.2 Τομείς Παραγωγής Δεδομένων.....	14
2.3 Πηγές Παραγωγής Δεδομένων.....	15
2.4 Τεχνολογίες Αποθήκευσης Big Data.....	15
3. Big Data & Ανάλυση	16
3.1 Big Data σε Επιχειρήσεις και Οργανισμούς	16
3.2 Επιχειρηματική Αναλυτική (Business Analytics)	16
3.2.1 Είδη Επιχειρηματικής Αναλυτικής	17
3.3 Επιχειρηματική Ευφυΐα (Business Intelligence)	18
4. Customer Analytics	19
4.1 Ηλεκτρονικό Εμπόριο.....	19
4.2 Πλεονεκτήματα και Τεχνικές Customer Analytics	20
4.3 Τμηματοποίηση Πελατών (Customer Segmentation).....	20
5. Πρόβλεψη Πωλήσεων και Χρονοσειρές.....	22
5.1 Πρόβλεψη Πωλήσεων	22
5.2 Χρονοσειρές.....	23
5.2.1 Ανάλυση Χρονοσειρών	23
5.3 Μέθοδοι Πρόβλεψης.....	24
5.3.1 Ποιοτικές Μέθοδοι Πρόβλεψης.....	24
5.3.2 Ποσοτικές Μέθοδοι Πρόβλεψης.....	26
6. Διευρενητική Ανάλυση Δεδομένων (EDA).....	32
7. Τμηματοποίηση Πελατών.....	40
7.1 Διάσταση Recency	40
7.2 Διάσταση Frequency	41
7.3 Διάσταση Monetary	42
7.4 Δημιουργία Τεταρτημορίων.....	42
7.5 Αποτελέσματα Ανάλυσης Τμηματοποίησης Πελατών.....	43
7.6 Προτάσεις Στρατηγικών και Καμπανιών Μάρκετινγκ	44
8. Πρόβλεψη Πωλήσεων - Αποτελέσματα Δεδομένων Εκπαίδευσης και Ελέγχου	46
8.1 1 ^ο Επίπεδο Ιεραρχίας Γεωγραφική Περιοχή – Προβλεπόμενες Παραγγελίες	1
8.2 1 ^ο Επίπεδο Ιεραρχίας Γεωγραφική Περιοχή - Σφάλματα.....	7
8.4 2 ^ο Επίπεδο Ιεραρχίας Πόλη Καταστήματος – Προβλεπόμενες Παραγγελίες	8
8.2 2 ^ο Επίπεδο Ιεραρχίας Πόλη Καταστήματος - Σφάλματα.....	10

8.5 3 ^ο Επίπεδο Ιεραρχίας – Προβλεπόμενες Παραγγελίες.....	1
8.6 3 ^ο Επίπεδο Ιεραρχίας Κατάστημα - Σφάλματα	7
8.7 Αποτελέσματα	8
9. Πρόβλεψη Πωλήσεων - Αποτελέσματα Πρόβλεψης Πωλήσεων	9
9.1 Πρόβλεψη Πωλήσεων 1 ^ο Επίπεδο Ιεραρχίας.....	9
9.2 Πρόβλεψη Πωλήσεων 2 ^ο Επίπεδο Ιεραρχίας.....	16
9.3 Πρόβλεψη Πωλήσεων 3 ^ο Επίπεδο Ιεραρχίας.....	18
9.4 Αποτελέσματα	25
10. Συμπεράσματα.....	27
Βιβλιογραφία	28

Κατάλογος Εικόνων

Εικόνα 1 Όγκος των δεδομένων που δημιουργήθηκαν, καταγράφηκαν, αντιγράφηκαν και καταναλώθηκαν 2010-2020 , με προβλέψεις για το 2025 (Statista, 2023).....	12
Εικόνα 2 Σύνολο γραμμών και στηλών.....	32
Εικόνα 3 Πρώτες γραμμές συνόλου δεδομένων.....	32
Εικόνα 4 Τελευταίες γραμμές συνόλου δεδομένων.....	33
Εικόνα 5 Τύποι συνόλου δεδομένων.....	33
Εικόνα 6 Ελλιπείς τιμές.....	34
Εικόνα 7 Κατανομή order_value και food_score.....	35
Εικόνα 8 Κατανομή delivery_score και service_score.....	35
Εικόνα 9 Κατανομή αξίας παραγγελίας.....	36
Εικόνα 10 Συχνότητα μεταβλητών Boolean.....	37
Εικόνα 11 Συσχέτιση μεταβλητών.....	38
Εικόνα 12 Αριθμός Πελατών / Κατάστημα και Αριθμός Παραγγελιών / Κατάστημα.....	39
Εικόνα 13 Αριθμός Παραγγελιών / Είδος Προϊόντος.....	39
Εικόνα 14 Recency σημείο αναφοράς τελευταίας ημερομηνίας.....	41
Εικόνα 15 Δημιουργία στήλης καταγραφής ημερομηνίας παραγγελίας.....	41
Εικόνα 16 Υπολογισμός διάστασης Recency.....	41
Εικόνα 17 Υπολογισμός διάστασης Frequency.....	42
Εικόνα 18 Υπολογισμός διάστασης Monetary.....	42
Εικόνα 19 Διαστάσεις RFM.....	42
Εικόνα 20 Quantiles.....	42
Εικόνα 21 Quartiles Διαστάσεων.....	43
Εικόνα 22 RFM Βαθμολογία.....	43
Εικόνα 23 Αποτελέσματα Ανάλυσης Τμηματοποίησης Πελατών.....	44
Εικόνα 24 Γράφημα Αποτελεσμάτων Ανάλυσης Τμηματοποίησης Πελατών.....	44
Εικόνα 25 Δημιουργία στήλης κατηγοριοποίησης καταστημάτων.....	46
Εικόνα 26 Ιεραρχική Δομή.....	47
Εικόνα 27 Απεικόνιση Ιεραρχικής Δομής.....	48
Εικόνα 28 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (1).....	1
Εικόνα 29 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (2).....	1
Εικόνα 30 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (3).....	2
Εικόνα 31 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (4).....	2
Εικόνα 32 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (5).....	3
Εικόνα 33 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (6).....	3
Εικόνα 34 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (7).....	4
Εικόνα 35 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (8).....	4
Εικόνα 36 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (9).....	5
Εικόνα 37 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (10).....	5
Εικόνα 38 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (11).....	6
Εικόνα 39 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (12).....	6
Εικόνα 40 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (13).....	7
Εικόνα 41 2ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (1).....	9
Εικόνα 42 2ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (2).....	9

Εικόνα 43 2ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (3)	10
Εικόνα 44 2ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (4)	10
Εικόνα 45 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (1)	1
Εικόνα 46 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (2)	1
Εικόνα 47 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (3)	2
Εικόνα 48 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (4)	2
Εικόνα 49 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (5)	3
Εικόνα 50 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (6)	3
Εικόνα 51 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (7)	4
Εικόνα 52 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (8)	4
Εικόνα 53 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (9)	5
Εικόνα 54 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (10)	5
Εικόνα 55 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (11)	6
Εικόνα 56 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (12)	6
Εικόνα 57 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (13)	7
Εικόνα 58 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (1)	10
Εικόνα 59 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (2)	10
Εικόνα 60 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (3)	11
Εικόνα 61 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (4)	11
Εικόνα 62 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (5)	12
Εικόνα 63 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (6)	12
Εικόνα 64 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (7)	13
Εικόνα 65 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (8)	13
Εικόνα 66 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (9)	14
Εικόνα 67 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (10)	14
Εικόνα 68 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (11)	15
Εικόνα 69 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (12)	15
Εικόνα 70 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (13)	16
Εικόνα 71 2ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (1)	17
Εικόνα 72 2ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (2)	17
Εικόνα 73 2ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (3)	18
Εικόνα 74 2ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (4)	18
Εικόνα 75 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (1)	19
Εικόνα 76 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (2)	20
Εικόνα 77 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (3)	20
Εικόνα 78 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (4)	21
Εικόνα 79 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (5)	21
Εικόνα 80 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (6)	22
Εικόνα 81 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (7)	22
Εικόνα 82 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (8)	23
Εικόνα 83 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (9)	23
Εικόνα 84 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (10)	24
Εικόνα 85 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (11)	24
Εικόνα 86 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (12)	25

Εικόνα 87 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (13).....	25
--	----

Κατάλογος Πινάκων

Πίνακας 1 Αριθμός καταστημάτων σε κάθε κατηγορία	47
Πίνακας 2 Κατηγοριοποίηση Καταστημάτων	47
Πίνακας 3 Προβλεπόμενες Παραγγελίες Δεδομένων Εκπαίδευσης 1 ^ο Επίπεδο Ιεραρχίας	3
Πίνακας 4 Προβλεπόμενες Παραγγελίες Δεδομένων Ελέγχου 1 ^ο Επίπεδο Ιεραρχίας	4
Πίνακας 5 1 ^ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Εκπαίδευσης.....	7
Πίνακας 6 1ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Ελέγχου	8
Πίνακας 7 Προβλεπόμενες Παραγγελίες Δεδομένων Εκπαίδευσης 2 ^ο Επίπεδο Ιεραρχίας	8
Πίνακας 8 Προβλεπόμενες Παραγγελίες Δεδομένων Ελέγχου 2 ^ο Επίπεδο Ιεραρχίας	8
Πίνακας 9 2 ^ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Εκπαίδευσης.....	11
Πίνακας 10 2 ^ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Ελέγχου	11
Πίνακας 11 Προβλεπόμενες Παραγγελίες Δεδομένων Εκπαίδευσης 3 ^ο Επίπεδο Ιεραρχίας	5
Πίνακας 12 Προβλεπόμενες Παραγγελίες Δεδομένων Ελέγχου 3 ^ο Επίπεδο Ιεραρχίας	5
Πίνακας 13 3 ^ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Εκπαίδευσης.....	7
Πίνακας 14 3 ^ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Ελέγχου	8
Πίνακας 15 Προβλεπόμενες Πωλήσεις 1 ^ο Επίπεδο Ιεραρχίας (01/06/2021 έως 07/06/2021)	9
Πίνακας 16 Προβλεπόμενες Πωλήσεις 2 ^ο Επίπεδο Ιεραρχίας (01/06/2021 έως 07/06/2021)	16
Πίνακας 17 Προβλεπόμενες Πωλήσεις 3 ^ο Επίπεδο Ιεραρχίας (01/06/2021 έως 07/06/2021)	19

Κατάλογος Συντομογραφιών

Συντομογραφία	Περιγραφή
API	Application Programming Interface
ARIMA	Autoregressive Integrated Moving Average
BI	Business Intelligence
CSAT	Customer Satisfaction
EDA	Exploratory Data Analysis
EDI	Electronic Data Interchange
EMA	Exponential Moving Average
GPS	Global Positioning System
GFS	Google File System
HDFS	Hadoop Distributed File System
IoT	Internet of Things
MA	Moving Average
NLP	Natural Language Processing
RFM	Recency, Frequency, Monetary
SARIMA	Seasonal Autoregressive Integrated Moving Average
SMA	Simple Moving Average
VoC	Voice of the Customer

Περίληψη

Η παρούσα διπλωματική εργασία έχει ως στόχο της αξιοποίηση τεχνολογιών μεγάλων δεδομένων (Big Data), έτσι ώστε να επιτευχθεί η ανάλυση δεδομένων, η τμηματοποίηση πελατών και η πρόβλεψη πωλήσεων μιας ηλεκτρονικής πλατφόρμας πωλήσεων. Η τμηματοποίηση των πελατών διενεργείται με τη χρήση της μεθόδου RFM (Recency, Frequency, Monetary), όπου οι πελάτες ομαδοποιούνται ανάλογα με την πρόσφατη δραστηριότητά τους στην πλατφόρμα, τη συχνότητα των αγορών τους και την οικονομική τους αξία, το οποίο δίνει σημαντικές πληροφορίες τη συμπεριφορά τους. Για την πρόβλεψη των πωλήσεων γίνεται χρήση του μοντέλου Holt-Winters στο οποίο λαμβάνονται υπόψιν οι εποχικές τάσεις και οι διακυμάνσεις των δεδομένων. Το συγκεκριμένο μοντέλο μέσω της εκθετικής εξομάλυνσης, προσφέρει αξιόπιστες προβλέψεις που βοηθούν τις επιχειρήσεις να δημιουργήσουν πιο αποδοτικές στρατηγικές και να διαχειριστούν τα αποθέματα και τους πόρους τους πιο αποτελεσματικά. Τα αποτελέσματα της πρόβλεψης αξιολογούνται και συγκρίνονται με τις πραγματικές πωλήσεις για την επαλήθευση της ακρίβειας του μοντέλου. Τα συνολικά αποτελέσματα της ανάλυσης δείχνουν ότι η χρήση της ανάλυσης RFM και του μοντέλου πρόβλεψης Holt-Winters είναι εφικτό να βελτιώσει σημαντικά τη λήψη αποφάσεων στις επιχειρήσεις, καθώς παρέχεται μια πιο ολοκληρωμένη εικόνα για τη συμπεριφορά των πελατών και τις μελλοντικές τάσεις των πωλήσεων.

Abstract

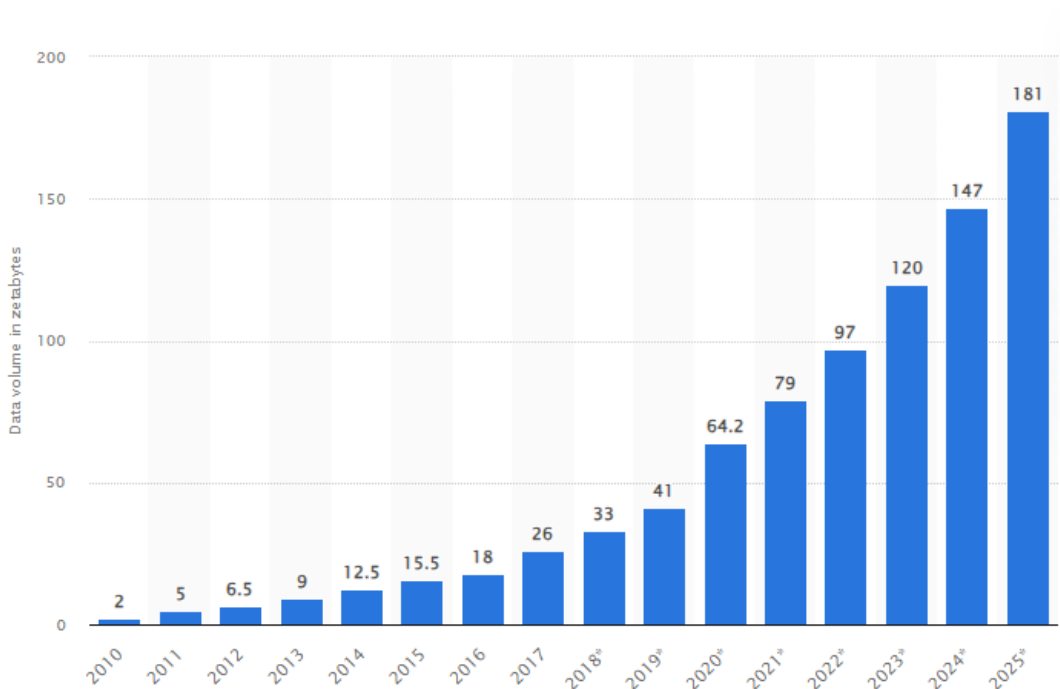
This thesis aims to leverage Big Data technologies to achieve data analysis, customer segmentation, and sales forecasting for an e-commerce sales platform. Customer segmentation is performed using the RFM (Recency, Frequency, Monetary) method, where customers are grouped based on their recent activity on the platform, the frequency of their purchases, and their monetary value, providing significant insights into their behavior. For sales forecasting, the Holt-Winters model is used, which considers seasonal trends and data fluctuations. This model, through exponential smoothing, offers reliable forecasts that help businesses create more efficient strategies and manage their inventory and resources more effectively. The forecasting results are evaluated and compared with actual sales to verify the model's accuracy. The overall analysis results show that the use of RFM analysis and the Holt-Winters forecasting model can significantly improve decision-making in businesses by providing a more comprehensive view of customer behavior and future sales trends.

1. Εισαγωγή

1.1 Γενικά

Στη σύγχρονη εποχή είναι βέβαιο ότι ζούμε σε έναν κόσμο πλημμυρισμένο από δεδομένα, καθώς καθημερινά παράγονται, συλλέγονται και διαδίδονται με εκθετικούς ρυθμούς μεγάλες ποσότητες πληροφοριών. Η διαρκής παραγωγή δεδομένων οφείλεται στην ανάπτυξη της τεχνολογίας, την αυξημένη υπολογιστική ισχύ, τη σημαντική πρόοδο στους τομείς της επικοινωνίας και της συνδεσιμότητας, την επένδυση στην έρευνα και ανάπτυξη, καθώς και την παροχή λογισμικού ανοικτού κώδικα. Το 2023 είχε υπολογιστεί ότι ο όγκος των δεδομένων που θα παραγόταν κάθε ημέρα ήταν 328,77 εκατομμύρια terabytes (Business2Community. (2023) How Much Data is Created Every Day).

Το 2023 η Statista δημοσίευσε έρευνα που αφορούσε στον όγκο των δεδομένων που δημιουργήθηκαν, καταγράφηκαν, αντιγράφηκαν και καταναλώθηκαν από το 2010 έως το 2020 σε όλο τον κόσμο. Επίσης, στη συγκεκριμένη έρευνα συμπεριλήφθηκαν προβλέψεις για τα έτη από το 2021 έως το 2025. Στη μελέτη είχε εκτιμηθεί ότι ο συνολικός όγκος των δεδομένων που δημιουργείται μέσα σε ένα έτος αυξάνεται κατά 8.950% (Statista, 2023, Amount of data created, consumed, and stored worldwide from 2010 to 2025).



Εικόνα 1 Όγκος των δεδομένων που δημιουργήθηκαν, καταγράφηκαν, αντιγράφηκαν και καταναλώθηκαν 2010-2020 , με προβλέψεις για το 2025 (Statista, 2023)

Με την αύξηση της παραγωγής δεδομένων εμφανίζονται διαρκείς εξελίξεις και αλλαγές σχεδόν σε κάθε πτυχή της προσωπικής και επαγγελματικής ζωής. Αυτό έχει ως αποτέλεσμα τη μεταβολή του τρόπου ζωής και λήψης αποφάσεων, της λειτουργίας των επιχειρήσεων, της εργασίας και της αλληλεπίδρασης του ανθρώπου με το περιβάλλον και τους συνανθρώπους του.

Ο όγκος των διαθέσιμων δεδομένων παρέχει μεγάλη αξία, καθώς δημιουργεί ευκαιρίες και προκλήσεις για τις οντότητες που έχουν τη δυνατότητα να τον διαχειριστούν και να προβούν στις κατάλληλες αναλύσεις. Επιπλέον, για τους οργανισμούς που χρησιμοποιούν προηγμένα εργαλεία και τεχνικές ανάλυσης αποτελεί σημαντικό παράγοντα για την επίτευξη ανταγωνιστικού πλεονεκτήματος, τη λήψη αποφάσεων αλλά και στρατηγικών σχεδίων. Τα δεδομένα δεν αποτελούν απλά έναν πόρο, αλλά μια κινητήρια δύναμη η οποία διαμορφώνει σε μεγάλο βαθμό το παρόν και το μέλλον.

Η ραγδαία αύξηση των δεδομένων έχει οδηγήσει στην εμφάνιση του όρου των Big Data. Τα Big Data είναι δεδομένα που χαρακτηρίζονται από τον τεράστιο όγκο τους, την πολύ γρήγορη παραγωγή τους και την πολυπλοκότητά τους. Η εμφάνιση του όρου αυτού έχει δημιουργήσει με τη σειρά του την ανάγκη για ανάπτυξη ισχυρών τεχνολογιών αποθήκευσης και διαχείρισης του μεγάλου όγκου δεδομένων. Η έννοια των Μεγάλων Δεδομένων εμφανίστηκε στις αρχές της δεκαετίας του 2000, όταν ο τότε αναλυτής του κλάδου Doug Laney κατέγραψε τον ορισμό των Big Data με βάση τα 3V's.

1.2 Σκοπός Διπλωματικής

Στη σημερινή εποχή η τεχνολογία εξελίσσεται με ραγδαίους ρυθμούς και το γεγονός αυτό οδηγεί τις επιχειρήσεις και τους οργανισμούς στην εύρεση νέων τεχνικών και μεθοδολογιών για την απόκτηση ανταγωνιστικού πλεονεκτήματος, τη διατήρηση πελατών, τη δημιουργία καινοτόμων προϊόντων και υπηρεσιών και την αύξηση των κερδών τους. Τα τελευταία χρόνια οι υποδομές του ηλεκτρονικού εμπορίου έχουν εξελιχθεί σημαντικά με αποτέλεσμα οι επιχειρήσεις να δίνουν μεγάλη βαρύτητα στο κομμάτι αυτό.

Σκοπός της παρούσας διπλωματικής είναι η λήψη πληροφοριών με την αξιοποίηση τεχνικών και μεθοδολογιών της επιχειρηματικής αναλυτικής και πρόβλεψης πωλήσεων κάνοντας χρήση δεδομένων από ηλεκτρονική πλατφόρμα πωλήσεων. Πιο συγκεκριμένα, θα χρησιμοποιηθεί μέθοδος τμηματοποίησης πελατών με βάση τη συμπεριφορά τους, καθώς και πρόβλεψη πωλήσεων παρέχοντας έτσι σημαντικές πληροφορίες για τη διατήρηση και την ενίσχυση της συνεργασίας με συγκεκριμένα καταστήματα.

2 Big Data

2.1 Χαρακτηριστικά Big Data

Στη σημερινά εποχή τα Big Data κατέχουν έναν από τους κυρίαρχους ρόλους στον τεχνολογικό και επιχειρηματικό κόσμο. Στην αρχή τα Big Data είχαν οριστεί από τρία χαρακτηριστικά “3V’s”, τα οποία ήταν ο όγκος (volume), η ταχύτητα (velocity) και η ποικιλία (variety). Με στόχο να γίνει καλύτερη αποτύπωση της πολυπλοκότητας και των προκλήσεων των Big Data έχουν προστεθεί τέσσερις επιπλέον διαστάσεις η ακρίβεια (veracity), η αξία (value), η μεταβλητότητα (variability) και η οπτικοποίηση (visualization).

Volume (Όγκος): Η διάσταση του όγκου αφορά στον όγκο των δεδομένων που παράγεται ανά δευτερόλεπτο. Αυτό, όπως έχει ήδη αναφερθεί, οφείλεται στο γεγονός του πολλαπλασιασμού των ψηφιακών συσκευών, των μέσων κοινωνικής δικτύωσης, του διαδικτύου των πραγμάτων (IoT) και της ραγδαίας ανάπτυξης της τεχνολογίας.

Velocity (Ταχύτητα): Το χαρακτηριστικό της ταχύτητας έχει σχέση με τον ρυθμό με τον οποία παράγονται, επεξεργάζονται και αναλύονται τα δεδομένα. Για τους περισσότερους τομείς η επεξεργασία και η ανάλυση δεδομένων σε πραγματικό χρόνο είναι πολύ σημαντική για τη λήψη αποφάσεων και τη δημιουργία στρατηγικών. Η πρόκληση για τη διάσταση της ταχύτητας είναι η άμεση διαχείριση των μεγάλων και ταχύτατων ροών δεδομένων.

Variety (Ποικιλία): Η διάσταση της ποικιλίας χρησιμοποιείται για τους διαφορετικούς τύπους δεδομένων που παράγονται ταυτόχρονα από ποικίλες πηγές. Οι τύποι δεδομένων που παράγονται είναι οι ακόλουθοι:

Veracity (Ακρίβεια): Το χαρακτηριστικό της ακρίβειας αναφέρεται στην αξιοπιστία των δεδομένων. Η ακρίβεια των δεδομένων αποτελεί ένα από τα πιο σημαντικά ζητήματα, καθώς τα μη αξιόπιστα δεδομένα οδηγούν στην εξαγωγή λανθασμένων πληροφοριών αλλά και τη μη ορθή λήψη αποφάσεων. Η αξιοπιστία των δεδομένων διαμορφώνεται από την ποιότητα, την πηγή και τη διακυβέρνησή τους.

Value (Αξία): Η διάσταση της αξίας έχει σχέση με τα διάφορα οφέλη και τις γνώσεις που παράγονται μέσα από την ανάλυση δεδομένων. Ο κύριος στόχος της ανάλυσης δεδομένων είναι η ανάκτηση σημαντικών πληροφοριών οι οποίες με τη σειρά τους οδηγούν στη λήψη τεκμηριωμένων και ορθών αποφάσεων και ταυτόχρονα συμβάλλουν στην απόκτηση ανταγωνιστικού πλεονεκτήματος.

Variability (Μεταβλητότητα): Το χαρακτηριστικό της μεταβλητότητας αναφέρεται στις ασυνέπειες και τις διακυμάνσεις που παρουσιάζουν τα δεδομένα. Η μεταβλητότητα στα δεδομένα είναι πιθανό να οφείλεται στους διάφορους τύπους των δεδομένων, που έχουν ήδη αναφερθεί, στην ποιότητά τους και σε ποικίλες αλλαγές που προκύπτουν στις πηγές από τις οποίες προέρχονται. Το γεγονός αυτό δημιουργεί πολυπλοκότητα στην ανάλυσή τους.

Visualization (Οπτικοποίηση): Η διάσταση της οπτικοποίησης αναφέρεται στην απεικόνιση των δεδομένων σε οπτική μορφή. Η οπτικοποίηση έχει ως σκοπό να συμβάλλει στην κατανόηση και την αποκωδικοποίηση μεγάλου και σύνθετου όγκου δεδομένων. Μέσω της οπτικοποίησης είναι εφικτό να επιτευχθεί με ταχύτατους ρυθμούς ανάλυση δεδομένων και λήψη τεκμηριωμένων και βέλτιστων αποφάσεων. Επιπρόσθετα, με τη χρήση της οπτικοποίησης είναι δυνατός ο εντοπισμός της μη ακρίβειας των δεδομένων και τυχόν σφαλμάτων.

2.2 Τομείς Παραγωγής Δεδομένων

Στον σύγχρονο κόσμο, τα δεδομένα παράγονται από ποικίλους τομείς και διαδραματίζουν σημαντικό ρόλο στην ψηφιακή εποχή. Ο κάθε τομέας είναι σίγουρο πως συνεισφέρει με τον δικό του τρόπο στην παραγωγή δεδομένων. Όταν τα δεδομένα διαχειρίζονται και αναλύονται σωστά, οδηγούν σε καινοτομίες, βέλτιστη λήψη αποφάσεων, βελτιώσεις της καθημερινότητας και δημιουργία τεκμηριωμένων στρατηγικών. Μερικοί από τους τομείς που παράγονται δεδομένα είναι οι ακόλουθοι:

1. Τομέας Υγείας
2. Χρηματοοικονομικός & Τραπεζικός Τομέας

3. Τομέας Τηλεπικοινωνιών
4. Δημόσιος Τομέας
5. Τομέας Ηλεκτρονικού Εμπορίου
6. Τομέας Μεταφορών
7. Τομέας Εκπαίδευσης

2.3 Πηγές Παραγωγής Δεδομένων

Στη σημερινή εποχή, όπως έχει ήδη αναφερθεί, τα δεδομένα αποτελούν σπουδαίο κομμάτι για τις επιχειρήσεις, καθώς τους παρέχουν σημαντική γνώση για την καλύτερη λήψη αποφάσεων, την υιοθέτηση στρατηγικών και τη λειτουργία τους. Τα δεδομένα παράγονται από διαφορετικές πηγές και κάθε μία από αυτές προσφέρει εξίσου σημαντικές πληροφορίες. Μερικές από τις πηγές που παράγονται δεδομένα είναι οι ακόλουθες:

1. Μέσα Κοινωνικής Δικτύωσης
2. Συστήματα POS
3. Συστήματα Συναλλαγών
4. Κεντρικά Τραπεζικά Συστήματα
5. Συστήματα Επιχειρησιακού Σχεδιασμού (ERP)

2.4 Τεχνολογίες Αποθήκευσης Big Data

Η εμφάνιση των Big Data έχει δημιουργήσει την ανάγκη για ανάπτυξη ισχυρών τεχνολογιών αποθήκευσης και διαχείρισης μεγάλου όγκου δεδομένων.

Με την πλήρη κατανόηση και χρήση τέτοιου είδους τεχνολογιών οι επιχειρήσεις έχουν τη δυνατότητα να αποθηκεύουν, να διαχειρίζονται και να αναλύουν τα δεδομένα τους με ακόμη πιο βέλτιστο και αποδοτικό τρόπο. Μερικές από τις τεχνολογίες αποθήκευσης δεδομένων είναι οι ακόλουθες:

1. Σχεσιακές Βάσεις Δεδομένων (SQL Databases)
2. NoSQL Βάσεις Δεδομένων
3. Κατανεμημένα Συστήματα Αρχείων (Distributed File Systems)
4. Αποθήκες Δεδομένων (Data Warehouse Solutions)
5. Cloud-Based Λύσεις

3. Big Data & Ανάλυση

3.1 Big Data σε Επιχειρήσεις και Οργανισμούς

Στη σημερινή εποχή τα Big Data αποτελούν σημαντικό πλεονέκτημα για τις επιχειρήσεις και τους οργανισμούς. Όπως έχει ήδη αναφερθεί τα Big Data είναι τεράστιοι όγκοι πληροφοριών που συλλέγονται από ποικίλες πηγές όπως κοινωνικά δίκτυα, διενέργεια συναλλαγών, χρήση έξυπνων συσκευών και αισθητήρων και άλλα.

Μέσω της επεξεργασίας και της ανάλυσης των δεδομένων οι επιχειρήσεις καταφέρνουν να συλλέγουν πολύτιμη πληροφορία και γνώση για τις διαδικασίες, τον τρόπο λειτουργίας, τους πελάτες τους και τις τάσεις της αγοράς. Επίσης, έχουν τη δυνατότητα να εντοπίζουν νέες ευκαιρίες, να ελαχιστοποιούν τους κινδύνους, να λαμβάνουν καλύτερες αποφάσεις, να μειώνουν τα κόστη τους και να βελτιώνουν την ποιότητα των προϊόντων και υπηρεσιών τους.

Πιο συγκεκριμένα, οι επιχειρήσεις και οι οργανισμοί αξιοποιώντας το πλεονέκτημα της ανάλυσης των Big Data μπορούν να επιτύχουν τα ακόλουθα:

1. Την ενίσχυση των διαδικασιών λήψης αποφάσεων που αποτελεί σημαντικό συστατικό στοιχείο για τις επιχειρήσεις του σύγχρονου ανταγωνιστικού κόσμου.
2. Τη βελτίωση της εξατομικευμένης εμπειρίας πελατών, γεγονός το οποίο παρέχει ανταγωνιστικό πλεονέκτημα στις επιχειρήσεις, καθώς με την ενίσχυση της ικανοποίησής τους αυξάνεται η πιθανότητα διατήρησής τους.
3. Την αποτελεσματική λειτουργία των συσκευών, καθώς οι επιχειρήσεις συλλέγουν δεδομένα από αισθητήρες που έχουν τοποθετηθεί στις συσκευές και μέσω της ανάλυσής τους μπορούν να εντοπίσουν πιθανά προβλήματα και να μειώσουν τον κίνδυνο εμφάνισης βλαβών ή ακόμη και διακοπής της λειτουργίας τους (Manyika et al., 2011). Επιπλέον, μέσω της ανάλυσης αυτών των δεδομένων παράγονται πληροφορίες που αφορούν στην έγκαιρη προειδοποίηση για συντήρηση των συσκευών, γεγονός το οποίο αυξάνει τη διάρκεια ζωής τους και κατά συνέπεια μειώνει και το κόστος για τις επιχειρήσεις.
4. Τη διενέργεια της προγνωστικής ανάλυσης που παρέχει στις επιχειρήσεις τη δυνατότητα να κάνουν πρόβλεψη μελλοντικών τάσεων και συμπεριφορών.
5. Τη βελτίωση των στρατηγικών μάρκετινγκ, καθώς με την ανάλυση μεγάλου όγκου δεδομένων έχει αλλάξει ο τρόπος με τον οποίο οι επιχειρήσεις προσεγγίζουν τους καταναλωτές.
6. Την ανάπτυξη της καινοτομίας που αποτελεί σημαντικό συστατικό στοιχείο για τις επιχειρήσεις. Η ανάλυση μεγάλου όγκου δεδομένων έχει κυρίαρχο ρόλο σε αυτή τη διαδικασία. Μέσω της ανάλυσης αυτών των δεδομένων, οι επιχειρήσεις μπορούν να αποκτήσουν βαθιά κατανόηση των αναγκών των πελατών και των τάσεων της αγοράς (Waller & Fawcett, 2013). Τέτοιου είδους γνώση είναι πολύτιμη για την εύρεση των κενών που υπάρχουν στα υπάρχοντα προϊόντα και υπηρεσίες, αλλά και για τον εντοπισμό των επερχόμενων τάσεων που μπορούν να επηρεάσουν τις προτιμήσεις των καταναλωτών.
7. Τη διαχείριση κινδύνων και την ανίχνευση απάτης, καθώς η ανάλυση μεγάλων δεδομένων κατέχει ουσιαστικό ρόλο, αφού συμβάλλει σε μεγάλο βαθμό στην προστασία των χρηματοπιστωτικών ιδρυμάτων και των πελατών τους.

3.2 Επιχειρηματική Αναλυτική (Business Analytics)

Η επιχειρηματική αναλυτική αποτελεί μία τεχνική ανάλυσης δεδομένων κατά τη διάρκεια της οποίας διενεργείται εξόρυξη δεδομένων, στατιστική ανάλυση αλλά και προγνωστική μοντελοποίηση, για να αναλυθούν και να αποσαφηνιστούν επιχειρηματικά δεδομένα με απώτερο στόχο τη λήψη τεκμηριωμένων αποφάσεων.

Οι επιχειρήσεις και οι οργανισμοί κάνοντας χρήση της επιχειρηματικής αναλυτικής έχουν τη δυνατότητα να μετατρέπουν τα δεδομένα από διάφορες πηγές σε χρήσιμες και επικυρωμένες πληροφορίες,

οι οποίες θα τους οδηγήσουν στον ανασχεδιασμό των διαδικασιών τους, τη βελτίωση της λήψης αποφάσεων, την απόκτηση ανταγωνιστικού πλεονεκτήματος και την αύξηση της παραγωγικότητας και των εσόδων (Evans, 2017).

3.2.1 Είδη Επιχειρηματικής Αναλυτικής

Στην επιχειρηματική αναλυτική περιλαμβάνονται διάφορα είδη ανάλυσης τα οποία χρησιμοποιούνται από τις επιχειρήσεις για την ανάλυση δεδομένων και κατ' επέκταση τη λήψη αποφάσεων και τη βελτίωση του στρατηγικού τους σχεδιασμού. Τα είδη αυτά αποτελούν σημαντικό συστατικό στοιχείο για τη μετατροπή των δεδομένων σε χρήσιμες πληροφορίες. Τα κυριότερα είδη της επιχειρηματικής αναλυτικής είναι το descriptive analytics, το diagnostic analytics, το predictive analytics και το prescriptive analytics.

Descriptive Analytics (Περιγραφική Αναλυτική)

Το descriptive analytics (περιγραφική αναλυτική) αποτελεί ένα από τα πιο βασικά είδη της επιχειρηματικής αναλυτικής. Κάνοντας χρήση της περιγραφικής αναλυτικής είναι εφικτό να δοθεί απάντηση στο ερώτημα «Τι συμβαίνει;». Επιλέγονται διάφορες τεχνικές για τη συλλογή και την ανάλυση δεδομένων μέσω των οποίων δίνεται η δυνατότητα στις επιχειρήσεις να κατανοήσουν μία κατάσταση συγκεκριμένης χρονικής περιόδου, όπως για παράδειγμα την εποχιακή ζήτηση προϊόντων.

Diagnostic Analytics (Διαγνωστική Αναλυτική)

Το diagnostic analytics (διαγνωστική αναλυτική) χρησιμοποιείται για την ανάλυση δεδομένων με στόχο την αντίληψη παρελθοντικών αποτελεσμάτων και γεγονότων. Κάνοντας χρήση της διαγνωστικής αναλυτικής είναι εφικτό να δοθεί απάντηση στο ερώτημα «Γιατί συμβαίνει αυτό;». Η διαγνωστική αναλυτική προχωράει ένα βήμα παραπάνω, δηλαδή δεν στοχεύει στο τι έχει συμβεί, όπως στην περιγραφική αναλυτική, αλλά στοχεύει στη διερεύνηση του λόγου για τον οποίο έχει συμβεί μια κατάσταση. Σ' αυτού του είδους την ανάλυση περιλαμβάνεται η ανίχνευση των βασικών αιτιών των τάσεων, των ανωμαλιών και των μοτίβων σε μεγάλους όγκους δεδομένων.

Predictive Analytics (Προγνωστική Αναλυτική)

Το predictive analytics (προγνωστική αναλυτική) αποτελεί σημαντικό συστατικό στοιχείο της επιχειρηματικής αναλυτικής με τη χρήση του οποίου αξιοποιούνται ιστορικά δεδομένα, αλγόριθμοι στατιστικής αλλά και τεχνικές μηχανικής μάθησης με σκοπό τη μελλοντική πρόβλεψη. Κάνοντας χρήση της προγνωστικής αναλυτικής είναι εφικτό να δοθεί απάντηση στο ερώτημα «Τι θα συμβεί στο μέλλον;». Η προγνωστική αναλυτική βοηθά τις επιχειρήσεις και τους οργανισμούς να επιτύχουν στην πρόβλεψη των τάσεων, την κατανόηση ενδεχόμενων κινδύνων και τον εντοπισμό σημαντικών ευκαιριών.

Prescriptive Analytics (Καθοδηγητική Αναλυτική)

Το prescriptive analytics (καθοδηγητική αναλυτική) ξεπερνάει τα συμπεράσματα που αφορούν στην πρόβλεψη και μέσω αυτού παρέχονται προτάσεις για τις βέλτιστες επόμενες ενέργειες των επιχειρήσεων με στόχο την επίτευξη του επιθυμητού αποτελέσματος. Κάνοντας χρήση της καθοδηγητικής αναλυτικής είναι εφικτό να δοθεί απάντηση στο ερώτημα «Ποια είναι η επόμενη κίνηση;». Με έναν συνδυασμό μαθηματικών μοντέλων και ποικίλων αλγορίθμων, η καθοδηγητική αναλυτική συμβάλει στη βελτιστοποίηση της λήψης αποφάσεων και του στρατηγικού σχεδιασμού.

Οι πιο βασικοί τομείς που υιοθετείται η επιχειρηματική αναλυτική είναι οι ακόλουθοι:

1. Οικονομικός και τραπεζικός
2. Υγεία
3. Διανικό εμπόριο
4. Παραγωγή
5. Τηλεπικοινωνίες
6. Μεταφορές
7. Μάρκετινγκ
8. Δημόσιος Τομέας

3.3 Επιχειρηματική Ευφυΐα (Business Intelligence)

Η επιχειρηματική ευφυΐα (BI) αποτελεί μία διαδικασία ανάλυσης δεδομένων και παροχής χρήσιμων πληροφοριών στις επιχειρήσεις και τους οργανισμούς με στόχο τη λήψη βέλτιστων επιχειρηματικών αποφάσεων. Το BI αφορά στη χρήση τεχνολογιών, στρατηγικών και πρακτικών για την επίτευξη της συλλογής, της ενσωμάτωσης, της ανάλυσης και της παρουσίασης δεδομένων. Η επιχειρηματική ευφυΐα διαθέτει εργαλεία, μεθοδολογίες και εφαρμογές που δίνουν τη δυνατότητα στις επιχειρήσεις να συλλέγουν δεδομένα από διάφορες πηγές είτε εσωτερικές είτε εξωτερικές, να κάνουν την απαραίτητη προετοιμασία για την ανάλυσή τους, να διενεργούν τα κατάλληλα ερωτήματα και να δημιουργούν αναφορές.

Στην επιχειρηματική ευφυΐα ακολουθούνται συγκεκριμένα βήματα τα οποία είναι ζωτικής σημασίας για τη μετατροπή των δεδομένων σε πληροφορίες που μπορούν να εκμεταλλευτούν από τις επιχειρήσεις και τους οργανισμούς. Κατά τη διαδικασία της ανάλυσης δεδομένων με τη χρήση της επιχειρηματικής ευφυΐας διασφαλίζεται ότι τα δεδομένα συλλέγονται με ακρίβεια, αποθηκεύονται ορθά και αναλύονται με αποτελεσματικό τρόπο. Τα κυριότερα βήματα της επιχειρηματικής ευφυΐας είναι τα ακόλουθα:

1. Η συλλογή δεδομένων αποτελεί το βασικότερο βήμα της επιχειρηματικής ευφυΐας. Κατά τη διάρκεια της συγκεκριμένης φάσης συλλέγονται δεδομένα από διάφορες πηγές.
2. Ο καθαρισμός δεδομένων αποτελεί επίσης ένα σημαντικό βήμα της διαδικασίας που ακολουθείται στην επιχειρηματική ευφυΐα. Στο παρόν βήμα από το σύνολο δεδομένων που έχει συλλεχθεί ανιχνεύονται και διαγράφονται μη ακριβείς ή ελλιπείς εγγραφές δεδομένων.
3. Η αποθήκευση δεδομένων αποτελεί το τρίτο βήμα της διαδικασίας και περιλαμβάνει διάφορες τεχνολογίες και μεθόδους τόσο για την αποθήκευση όσο και για τη διαχείριση μεγάλου όγκου δεδομένων που συλλέγεται από ποικίλες πηγές
4. Η ανάλυση δεδομένων αποτελεί το βήμα στο οποίο γίνεται αυτοματοποιημένα η μετατροπή ακατέργαστων δεδομένων σε σημαντικές και αξιόπιστες πληροφορίες με τη χρήση ποσοτικών και ποιοτικών τεχνικών ανάλυσης.
5. Η δημιουργία αναφορών είναι το τελευταίο βήμα της διαδικασίας της επιχειρηματικής ευφυΐας όπου με την ολοκλήρωση όλων των προηγούμενων βημάτων επιτυγχάνεται η δημιουργία αναφορών που περιέχουν γραφήματα και άλλες μορφές οπτικοποίησης των δεδομένων.

4. Customer Analytics

4.1 Ηλεκτρονικό Εμπόριο

Το ηλεκτρονικό εμπόριο (e-commerce) αποτελεί τη διαδικασία κατά την οποία αγοράζονται και πωλούνται προϊόντα και υπηρεσίες μέσω του διαδικτύου. Με τον συγκεκριμένο τύπο εμπορίου τις τελευταίες δεκαετίες έχουν επέλθει αξιοσημείωτες αλλαγές στη βιομηχανία λιανικής και χονδρικής πώλησης, τη συμπεριφορά και τις ανάγκες των καταναλωτών.

Οι διαρκείς τεχνολογικές εξελίξεις διαδραματίζουν σημαντικό ρόλο στην ανάπτυξη του ηλεκτρονικού εμπορίου. Πλέον με τη χρήση υπολογιστών και κινητών συσκευών οι καταναλωτές μπορούν να πραγματοποιήσουν τις αγορές τους οποιαδήποτε στιγμή και σε όποιο μέρος και αν βρίσκονται. Το ηλεκτρονικό εμπόριο είναι αξιόπιστο και αποτελεσματικό, καθώς οι συναλλαγές πραγματοποιούνται με ασφάλεια μέσω συστημάτων πληρωμών, υπάρχουν διάφοροι τρόποι παραλαβής των παραγγελιών και η διαδικασία αποστολής έχει αυτοματοποιηθεί με αποτέλεσμα να έχουν μειωθεί σε μεγάλο βαθμό τα σφάλματα.

Με τις τεχνολογικές εξελίξεις που δημιουργούν διαφορετικές συνθήκες έχουν αλλάξει οι προτιμήσεις και οι ανάγκες των καταναλωτών. Οι καταναλωτές αναζητούν ευκολία, ταχύτητα και αυξημένη ποικιλία προϊόντων και υπηρεσιών. Μέσω του ηλεκτρονικού εμπορίου είναι εφικτή η ικανοποίηση αυτών των απαιτήσεων, καθώς μπορούν να συγκριθούν τιμές και χαρακτηριστικά προϊόντων, να ληφθούν πληροφορίες για τα προϊόντα μέσω των κριτικών άλλων ατόμων και να ολοκληρωθούν οι αγορές πολύ γρήγορα και εύκολα.

Οι αρχές του ηλεκτρονικού εμπορίου εντοπίζονται περίπου στη δεκαετία του 1960 όπου με την ανάπτυξη της ηλεκτρονικής ανταλλαγής δεδομένων (EDI) οι επιχειρήσεις είχαν τη δυνατότητα να μεταφέρουν έγγραφα με ηλεκτρονικό τρόπο, γεγονός που καθιστούσε την ανταλλαγή των πληροφοριών γρήγορη και αποτελεσματική. Με τη χρήση της τεχνολογίας αυτής βελτιώθηκε αρκετά η διαχείριση της εφοδιαστικής αλυσίδας και η επικοινωνία μεταξύ των επιχειρήσεων (Laudon & Traver, 2019).

Στην πραγματικότητα όμως το ηλεκτρονικό εμπόριο ξεκίνησε να διαμορφώνεται προς τα τέλη της δεκαετίας του 1990, με την άφιξη του παγκόσμιου ιστού και την αύξηση των υπολογιστών. Η κατασκευή του παγκόσμιου ιστού και η ευρεία διάδοση και χρήση των υπολογιστών είχαν ως αποτέλεσμα τη δημιουργία ενός φιλικού προς τον χρήστη περιβάλλοντος, στο οποίο πραγματοποιούνταν αγορές και πωλήσεις μέσω του διαδικτύου. Οι εταιρίες Amazon και eBay οι οποίες δημιουργήθηκαν το 1994 και το 1995 αντίστοιχα, αποτέλεσαν τους πρωτοπόρους στον χώρο του ηλεκτρονικού εμπορίου.

Στις αρχές της δεκαετίας του 2000 το ηλεκτρονικό εμπόριο σημείωσε σημαντική ανάπτυξη. Στο γεγονός αυτό συντέλεσαν αρκετοί παράγοντες οι οποίοι οδήγησαν στην υιοθέτησή του από τις επιχειρήσεις και τους καταναλωτές. Εκείνη την περίοδο οι άνθρωποι απέκτησαν μεγαλύτερη και πιο εύκολη πρόσβαση στο διαδίκτυο και ταυτόχρονα βελτιώθηκε η διαδικτυακή ασφάλεια που εξασφάλισε την προστασία των συναλλαγών και την αύξηση της εμπιστοσύνης των καταναλωτών στον κόσμο του ηλεκτρονικού εμπορίου. Η αύξηση της ασφάλειας ώθησε όλο και περισσότερους καταναλωτές στο να πραγματοποιούν τις αγορές τους ηλεκτρονικά (Chaffey, 2015). Οι επιχειρήσεις παρατηρώντας την αύξηση των ηλεκτρονικών αγορών εκμεταλλεύτηκαν τις δυνατότητες που παρείχε το ηλεκτρονικό εμπόριο και δημιούργησαν πλατφόρμες διεθνούς ηλεκτρονικού εμπορίου μέσω των οποίων μπορούσαν να κάνουν πωλήσεις των προϊόντων και υπηρεσιών τους σε όλο τον κόσμο. Συνεπώς, δημιουργήθηκαν νέες ευκαιρίες ανάπτυξης και επέκτασης για αυτές.

Επιπλέον, με την ανάπτυξη των smartphone το ηλεκτρονικό εμπόριο έγινε ακόμη πιο προσβάσιμο και ευέλικτο για τους καταναλωτές. Η δημιουργία εφαρμογών για κινητά συνέβαλαν στην εύκολη αναζήτηση και αγορά προϊόντων και υπηρεσιών. Σημαντική βελτίωση της εμπειρίας του χρήστη ήταν και η δυνατότητα παρακολούθησης παραγγελίας σε πραγματικό χρόνο (Turban et al., 2018).

Τα μέσα κοινωνικής δικτύωσης έχουν παίξει καθοριστικό ρόλο στην εξέλιξη του ηλεκτρονικού εμπορίου, καθώς αυξάνουν την αλληλεπίδραση μεταξύ επιχειρήσεων και καταναλωτών και δημιουργούν καινούργιες ευκαιρίες για την εμπορική δραστηριότητα. Πλατφόρμες όπως το Facebook και το Instagram αποτελούν πλέον σημαντικά κανάλια μάρκετινγκ και πωλήσεων. Τέτοιου είδους πλατφόρμες κάνουν χρήση του μεγάλου όγκου δεδομένων των χρηστών που κατέχουν λογαριασμούς σε αυτές, με στόχο να προσεγγίσουν πιθανούς πελάτες μέσω στοχευμένων διαφημίσεων και προωθητικών ενεργειών. Οι Τμηματοποίηση πελατών και πρόβλεψη πωλήσεων
για μια ηλεκτρονική πλατφόρμα πωλήσεων

διαφημίσεις είναι εφικτό να προσαρμοστούν στις προτιμήσεις και τα ενδιαφέροντα των χρηστών, με αποτέλεσμα να αυξάνεται η πιθανότητα αλληλεπίδρασης και αγορών (Kim & Peterson, 2017).

Η πανδημία COVID-19 αύξησε ραγδαία τη χρήση και υιοθέτηση του ηλεκτρονικού εμπορίου, καθώς οι καταναλωτές λόγω των περιοριστικών μέτρων και της ανάγκης για προστασία και ασφάλεια επέλεξαν τις ηλεκτρονικές αγορές. Οι επιχειρήσεις με τη σειρά τους επένδυσαν πάρα πολύ στον ψηφιακό μετασχηματισμό και την αξιοποίηση νέων τεχνολογιών, όπως η τεχνητή νοημοσύνη και η μηχανική μάθηση για να βελτιώσουν την εμπειρία των καταναλωτών και την απόδοσή τους. Το ηλεκτρονικό εμπόριο πλέον αποτελεί αναπόσπαστο κομμάτι της καθημερινότητας και των επιχειρηματικών δραστηριοτήτων (Deloitte, 2020).

Το ηλεκτρονικό εμπόριο δημιουργεί μεγάλο όγκο δεδομένων ο οποίος μπορεί να αξιοποιηθεί από τις επιχειρήσεις για τη βελτίωση της στρατηγικής και την ενίσχυση της εμπειρίας του καταναλωτή. Η ανάλυση αυτών των δεδομένων είναι σημαντική για την κατανόηση των καταναλωτικών συμπεριφορών, την αύξηση της απόδοσης των επιχειρήσεων, της δημιουργίας καινοτόμων προϊόντων και υπηρεσιών και την πρόβλεψη των τάσεων της αγοράς.

4.2 Πλεονεκτήματα και Τεχνικές Customer Analytics

Η ανάλυση πελατών παρέχει στις επιχειρήσεις και τους οργανισμούς πολλά πλεονεκτήματα μέσω των οποίων είναι εφικτό να βελτιωθεί σε μεγάλο βαθμό η απόδοσή τους και η ικανοποίηση των πελατών. Τα πιο βασικά πλεονεκτήματα της ανάλυσης πελατών είναι τα ακόλουθα:

1. Η κατανόηση των πελατών και η τμηματοποίησή τους.
2. Η βελτίωση της εξατομικευμένης εμπειρίας του πελάτη.
3. Η επίτευξη αυξημένης διατήρησης πελατών.
4. Η λήψη αποδοτικών επιχειρηματικών αποφάσεων αποτελεί στοιχείο ζωτικής σημασίας για τις επιχειρήσεις.
5. Η βελτιστοποίηση των καμπανιών διαφήμισης.
6. Η απόκτηση ανταγωνιστικού πλεονεκτήματος.

Στην ανάλυση πελατών χρησιμοποιούνται ποικίλες τεχνικές ανάλυσης δεδομένων των πελατών. Οι πιο βασικές τεχνικές της ανάλυσης πελατών είναι οι ακόλουθες:

1. Τμηματοποίηση Πελατών (Customer Segmentation)
2. Πρόβλεψη Διαφυγόντων Πελατών (Churn Prediction)
3. Ανάλυση RFM
4. Ανάλυση Καλαθιού Αγοράς (Market Basket Analysis)
5. A/B Testing
6. Κοινωνική Ακρόαση (Social Listening)
7. Engagement Metrics
8. Χαρτογράφηση Πελατειακής Διαδρομής (Customer Journey Mapping)
9. Voice Of The Customer (VoC)
10. Ικανοποίηση Πελατών (Customer Satisfaction (CSAT) Scores)

4.3 Τμηματοποίηση Πελατών (Customer Segmentation)

Η τμηματοποίηση πελατών είναι μία τεχνική ανάλυσης δεδομένων κατά την οποία οι πελάτες των επιχειρήσεων και οργανισμών διαχωρίζονται σε ομάδες που έχουν κοινά χαρακτηριστικά. Ο πιο βασικός στόχος της τμηματοποίησης πελατών είναι να βοηθήσει τις επιχειρήσεις να δημιουργήσουν πιο αποτελεσματικές καμπάνιες μάρκετινγκ οι οποίες θα εξατομικεύονται για κάθε ομάδα, έτσι ώστε να έχουν τη δυνατότητα να καλύψουν τις ανάγκες και τις προτιμήσεις των πελατών τους. Αυτό θα έχει ως αποτέλεσμα την αύξηση της ικανοποίησης των πελατών, της διατήρησής τους και τη βελτίωση της επιχειρηματικής απόδοσης. Η τμηματοποίηση πελατών αποτελείται από τους ακόλουθους τύπους:

Δημογραφική τμηματοποίηση πελατών στην οποία οι πελάτες διαχωρίζονται με βάση διάφορα χαρακτηριστικά της ζωής τους, όπως για παράδειγμα το φύλο, την ηλικία, το εισόδημα, την οικογενειακή Τμηματοποίηση πελατών και πρόβλεψη πωλήσεων
για μια ηλεκτρονική πλατφόρμα πωλήσεων

κατάσταση, την εκπαίδευση και άλλα. Αποτελεί μία από τις πιο απλές μεθόδους τμηματοποίησης πελατών, καθώς τα δημογραφικά δεδομένα είναι εφικτό να συλλεχθούν πιο εύκολα από δεδομένα άλλων τύπων (Wedel & Kamakura, 2000).

Γεωγραφική τμηματοποίηση πελατών στην οποία οι πελάτες ομαδοποιούνται με βάση τη γεωγραφική τους θέση, δηλαδή την περιοχή διαμονής τους, την τοποθεσία τους, αλλά και τους πολιτιστικούς παράγοντες και τη γλώσσα που χρησιμοποιούν (Kotler & Keller, 2016).

Ψυχογραφική τμηματοποίηση πελατών στην οποία οι πελάτες διαιρούνται με βάση τις αξίες, τον τρόπο ζωής τους, τα ενδιαφέροντά τους και τα χαρακτηριστικά της προσωπικότητάς τους. Οι επιχειρήσεις αναλύουν δεδομένα από παραγγελίες των πελατών τους και απαντήσεις σε διάφορες έρευνες που έχουν διενεργήσει, έτσι ώστε να κατανοήσουν τον τρόπο ζωής τους και τις ανάγκες τους. Αυτό έχει ως αποτέλεσμα την πιο βαθιά αντίληψη των κινήτρων των πελατών τους και κατ'επέκταση τη δημιουργία πιο στοχευμένων στρατηγικών μάρκετινγκ (Yankelovich & Meer, 2006).

Τμηματοποίηση πελατών βάσει συμπεριφοράς στην οποία οι πελάτες τμηματοποιούνται ανάλογα με τη συμπεριφορά που εμφανίζουν στο διαδίκτυο. Οι επιχειρήσεις αναλύουν δεδομένα από παρελθοντικές αγορές των πελατών τους, τη συχνότητα αγοράς προϊόντων και υπηρεσιών του ίδιου πελάτη και το ενδιαφέρον που εμφανίζει για το κάθε προϊόν.

Τεχνολογική τμηματοποίηση πελατών στην οποία οι πελάτες διαχωρίζονται βάσει της τεχνολογίας και των εφαρμογών που αξιοποιούν, τον τρόπο με τον οποίο εντοπίζουν την επιχείρηση στο διαδίκτυο και τα κανάλια αλληλεπίδρασης με την επιχείρηση.

Τα οφέλη που παρέχει η τμηματοποίηση των πελατών στις επιχειρήσεις είναι αρκετά. Τα πιο βασικά από αυτά είναι τα ακόλουθα:

Η δημιουργία εξατομικευμένων στρατηγικών μάρκετινγκ και η κατάλληλη αλληλεπίδραση μέσω διάφορων μηνυμάτων με τις ομάδες πελατών που έχουν εντοπίσει.

Η επιλογή των καταλληλότερων καναλιών επικοινωνίας για την επίτευξη βελτιωμένης αλληλεπίδρασης.

Ο εντοπισμός αδυναμιών στα προϊόντα και τις υπηρεσίες και η δημιουργία νέων που θα καλύπτουν τις ανάγκες και τις απαιτήσεις των πελατών.

Η βελτίωση της εξυπηρέτησης των πελατών και η δημιουργία μακροχρόνιων πελατειακών σχέσεων.

5. Πρόβλεψη Πωλήσεων και Χρονοσειρές

5.1 Πρόβλεψη Πωλήσεων

Η πρόβλεψη πωλήσεων είναι ένα από τα βασικά εργαλεία που αξιοποιούν οι επιχειρήσεις και οι οργανισμοί όλων των κλάδων και μεγεθών. Μέσω της πρόβλεψης πωλήσεων παρέχεται η κατάλληλη πληροφορία για τον σχεδιασμό στρατηγικών, την αποδοτική διαχείριση των πόρων και τη λήψη αποφάσεων μέσω των οποίων επιτυγχάνεται τελικά η αύξηση της απόδοσης των επιχειρήσεων αλλά και η κερδοφορία τους. Οι επιχειρήσεις έχουν τη δυνατότητα να προβλέψουν μελλοντικές πωλήσεις προϊόντων και υπηρεσιών κάνοντας χρήση ιστορικών δεδομένων, ανάλυσης των τάσεων της αγοράς και άλλων παραγόντων και πληροφοριών.

Τα οφέλη που δημιουργούνται για τις επιχειρήσεις από την πρόβλεψη πωλήσεων είναι πολλά. Αρχικά, μέσω της πρόβλεψης πωλήσεων επιτυγχάνεται η αποδοτική διαχείριση των αποθεμάτων, καθώς όταν οι επιχειρήσεις αποκοτούν πληροφορίες για το τι πρόκειται να συμβεί μπορούν να προετοιμαστούν κατάλληλα. Πιο συγκεκριμένα, βάσει της πληροφορίας που λαμβάνουν προγραμματίζουν ανάλογα τον αριθμό διαθέσιμων αποθεμάτων, έτσι ώστε να μην αντιμετωπίσουν ελλείψεις ή να μην βρεθούν αντίστοιχα με πολύ μεγαλύτερο αριθμό αποθεμάτων από αυτόν που στην πραγματικότητα χρειάζονται. Συνεπώς, είναι εφικτό να αποφευχθεί η αντιμετώπιση καταστάσεων όπως μείωση πωλήσεων λόγω χαμηλού αποθέματος ή περιττά έξοδα αποθήκευσης λόγω υψηλού αποθέματος.

Επιπλέον, η πρόβλεψη πωλήσεων συμβάλλει στην καλύτερη οργάνωση της διαδικασίας παραγωγής και της αγοράς προμηθειών. Οι επιχειρήσεις έχοντας γνώση για τη μελλοντική ζήτηση προγραμματίζουν με τον πιο βέλτιστο τρόπο τις δραστηριότητες παραγωγής των προϊόντων τους και μεριμνούν να είναι εξοπλισμένες με όλα τα απαραίτητα υλικά για την παραγωγή τους. Με τον τρόπο αυτό μπορούν να ανταποκρίνονται άμεσα στις απαιτήσεις της αγοράς, γεγονός που αυξάνει την ικανοποίηση των πελατών τους και τη μακροχρόνια διατήρησή τους αλλά και την απόκτηση ανταγωνιστικού πλεονεκτήματος.

Επιπρόσθετα, όπως έχει ήδη αναφερθεί η πρόβλεψη πωλήσεων υποστηρίζει τις επιχειρήσεις κατά τη διαδικασία του στρατηγικού σχεδιασμού αλλά και της λήψης αποφάσεων. Με την πρόβλεψη πωλήσεων και την ανάλυση δεδομένων οι επιχειρήσεις έχουν τη δυνατότητα να κατανοήσουν τις ανάγκες των πελατών τους, καθώς και το διάστημα που θα χρειαστεί να ικανοποιηθούν οι εκάστοτε ανάγκες, να εντοπίσουν τις τάσεις της αγοράς αλλά και νέες ευκαιρίες ανάπτυξης. Συνεπώς, οι επιχειρήσεις βάσει των πληροφοριών που λαμβάνουν προχωρούν στη δημιουργία νέων προϊόντων και υπηρεσιών και ταυτόχρονα είναι κατάλληλα προετοιμασμένες για να καλύψουν ανά πάσα στιγμή τις ανάγκες των πελατών τους και να ανταποκριθούν αποδοτικά και αποτελεσματικά στις ανάγκες της αγοράς.

Ένα ακόμη από τα οφέλη της πρόβλεψης πωλήσεων είναι η βελτίωση της διαχείρισης των ταμειακών ροών, που αποτελεί στοιχείο ζωτικής σημασίας για τις καθημερινές λειτουργίες των επιχειρήσεων και τους μακροπρόθεσμους στρατηγικούς στόχους τους. Με την αξιοποίηση της πρόβλεψης πωλήσεων οι επιχειρήσεις έχουν τη δυνατότητα να εκτιμήσουν με ακρίβεια τα μελλοντικά τους έσοδα. Έχοντας τη συγκεκριμένη γνώση μπορούν να διαχειριστούν με βέλτιστο τρόπο τους οικονομικούς τους πόρους και να αποφύγουν τυχόν οικονομικά θέματα που μπορεί να δημιουργήσουν προβλήματα στη λειτουργία και τις διαδικασίες τους. Ταυτόχρονα, γνωρίζοντας τα μελλοντικά τους έσοδα μπορούν να επενδύσουν σε νέα προϊόντα, υπηρεσίες και τεχνολογίες γεγονός που τις καθιστά να είναι πιο ευέλικτες και να προσαρμόζονται άμεσα στις ανάγκες της αγοράς και των καταναλωτών και κατά συνέπεια τους οδηγεί στην αύξηση της κερδοφορίας τους και την ανάπτυξή τους.

Τέλος, η πρόβλεψη πωλήσεων αποτελεί βασική παράμετρο στην αξιολόγηση των καμπανιών μάρκετινγκ. Οι επιχειρήσεις συγκρίνουν τον αριθμό των πραγματικών πωλήσεων με τον αριθμό των πωλήσεων που είχαν προβλεφθεί και εκτιμούν την ανάλυση αποκλίσεων. Μέσω της συγκεκριμένης σύγκρισης έχουν τη δυνατότητα να εντοπίσουν τις πιο αποδοτικές στρατηγικές καμπανιών μάρκετινγκ, αλλά και των καμπανιών που χρήζουν βελτίωσης. Πιο συγκεκριμένα, όταν μία καμπάνια μάρκετινγκ έχει αποδώσει καλύτερα από τις προβλεπόμενες πωλήσεις, τότε οι επιχειρήσεις έχοντας ως γνώμονα αυτή την πληροφορία μπορούν να επενδύσουν στη συγκεκριμένη καμπάνια και να χρησιμοποιούν παρόμοιες στρατηγικές στόχευσης στο μέλλον, γεγονός που θα αυξήσει την κερδοφορία τους και το ανταγωνιστικό τους πλεονέκτημα.

Για την πρόβλεψη πωλήσεων υπάρχουν διαθέσιμες ποικίλες μεθοδολογίες. Οι επιχειρήσεις επιλέγουν τη μεθοδολογία που τους εξυπηρετεί περισσότερο λαμβάνοντας υπόψιν τις ανάγκες τους την εκάστοτε στιγμή που απαιτείται να διενεργηθεί η πρόβλεψη. Οι μεθοδολογίες πρόβλεψης πωλήσεων χωρίζονται σε δύο κατηγορίες τις ποσοτικές και τις ποιοτικές, οι οποίες θα αναλυθούν εκτενώς στις επόμενες παραγράφους.

5.2 Χρονοσειρές

5.2.1 Ανάλυση Χρονοσειρών

Οι χρονοσειρές είναι ακολουθίες από δεδομένα τα οποία συλλέγονται σε συγκεκριμένα χρονικά διαστήματα. Πιο συγκεκριμένα, είναι ένα σύνολο τυχαίων μεταβλητών x_t , με $t \in T$, όπου T είναι μια χρονική περίοδος ή ένα υποσύνολο του χρόνου. Αν το T είναι συνεχές ή διακριτό, τότε λαμβάνει αντίστοιχο χαρακτηρισμό και η χρονοσειρά.

Οι χρονοσειρές αποτελούν μία από τις βασικές μεθοδολογίες που χρησιμοποιούνται στην ανάλυση δεδομένων και αξιοποιούνται κυρίως για τον έλεγχο των διάφορων μεταβολών σε αυτά. Οι μεταβολές στα δεδομένα προκύπτουν λόγω της χρονικής παρέλευσης. Στην ανάλυση χρονοσειρών εξετάζονται μοτίβα και τάσεις τα οποία μπορεί να εμφανίζονται την εκάστοτε χρονική περίοδο που αφορά στο εξεταζόμενο σύνολο δεδομένων (Box, Jenkins, & Reinsel, 2015).

Οι τιμές που λαμβάνουν οι χρονοσειρές προκύπτουν από τον συνδυασμό ποικίλων παραγόντων, που είναι γνωστές ως συνιστώσες και οι οποίες ακολουθούν συστηματική ή μη συστηματική συμπεριφορά. Πιο συγκεκριμένα, οι συνιστώσες που χαρακτηρίζονται από συστηματική συμπεριφορά ακολουθούν κάποιο πρότυπο, ενώ αντίστοιχα όσες χαρακτηρίζονται από μη συστηματική συμπεριφορά είναι τυχαίες. Οι συνιστώσες που επηρεάζουν τις χρονοσειρές είναι οι ακόλουθες:

1. Η τάση (trend) αποτελεί τη συνιστώσα μέσω της οποίας λαμβάνονται πληροφορίες σχετικά με την κατεύθυνση των δεδομένων σε μακροχρόνια βάση, δηλαδή το αν εμφανίζουν ανοδική, καθοδική ή σταθερή τάση (Chatfield, 2004). Η ανοδική τάση αφορά στη συνεχή αύξηση των τιμών των δεδομένων με την πάροδο του χρόνου. Αντίστοιχα, η καθοδική τάση παρουσιάζει συνεχή μείωση των τιμών με τον χρόνο. Στην περίπτωση της σταθερής τάσης οι τιμές των δεδομένων παραμένουν στα ίδια επίπεδα με μικρές αλλαγές και χωρίς να εμφανίζουν ανοδική ή καθοδική πορεία.
2. Η εποχικότητα (seasonality) είναι η συνιστώσα που αφορά σε περιοδικές μεταβολές οι οποίες επαναλαμβάνονται σε συγκεκριμένα χρονικά διαστήματα (Kendall & Ord, 1990). Τα βασικά χαρακτηριστικά της είναι η περιοδικότητα, από τη στιγμή που οι διακυμάνσεις στα δεδομένα εμφανίζονται ανά τακτά χρονικά διαστήματα όπως εβδομαδιαία, μηνιαία ή ετήσια και η προβλεψιμότητα, καθώς οι διακυμάνσεις ακολουθούν ένα προκαθορισμένο μοτίβο το οποίο τις καθιστά προβλέψιμες.
3. Η κυκλική συνιστώσα (cyclic component) αποτελεί τη συνιστώσα στην οποία περιλαμβάνονται διακυμάνσεις που αφορούν σε μεγαλύτερα χρονικά διαστήματα από αυτά της εποχικότητας. Οι διακυμάνσεις στην κυκλική συνιστώσα δεν χαρακτηρίζονται από σταθερό και προδιαγεγραμμένο μοτίβο.
4. Η τυχαία συνιστώσα (random component) αναφέρεται σε διακυμάνσεις δεδομένων οι οποίες δεν είναι προβλέψιμες και είναι απόρροια τυχαίων παραγόντων και γεγονότων. Χαρακτηριστικό παράδειγμα αποτελούν τα σφάλματα κατά τη διαδικασία συλλογής δεδομένων τα οποία δημιουργούν τυχαίες διακυμάνσεις.
5. Τα σφάλματα (errors) είναι οι αποκλίσεις που εμφανίζονται ανάμεσα στις προβλεπόμενες τιμές και τις πραγματικές τιμές της χρονοσειράς. Τέτοιου είδους αποκλίσεις οφείλονται σε ποικίλους λόγους και παράγοντες (Hyndman & Athanasopoulos, 2018). Είναι σύνηθες τα σφάλματα να συμβολίζονται με e_t όπου t είναι η χρονική στιγμή. Είναι σημαντικό να αναφερθεί ότι τα σφάλματα δεν είναι προβλέψιμα.

5.3 Μέθοδοι Πρόβλεψης

Για τη διεξαγωγή προβλέψεων με τη χρήση χρονοσειρών ακολουθούνται συγκεκριμένα βήματα, τα οποία συμβάλλουν στην επίτευξη ακριβών και αξιόπιστων προβλέψεων. Τα πιο βασικά βήματα για τη διενέργεια προβλέψεων είναι τα ακόλουθα:

1. Ο καθορισμός των κύριων ερωτημάτων και του σκοπού διεξαγωγής της πρόβλεψης. Κατά τη διαδικασία πρόβλεψης είναι πολύ σημαντικό να προκαθοριστούν οι βασικές μεταβλητές του προβλήματος που πρόκειται να αναλυθεί αλλά και ο τρόπος με τον οποίο οι συγκεκριμένες μεταβλητές θα επηρεάσουν το πρόβλημα.
2. Η επιλογή της χρονικής περιόδου πρόβλεψης, για να διασφαλιστεί πως υπάρχουν διαθέσιμα δεδομένα για την επίτευξη της διεξαγωγής της.
3. Η επιλογή του ενδεδειγμένου μοντέλου πρόβλεψης, βήμα που αποτελεί σημείο ζωτικής σημασίας στην ανάλυση χρονοσειρών. Το ορθό μοντέλο πρόβλεψης έχει τη δυνατότητα να ενισχύσει σε μεγάλο βαθμό την ακρίβεια της πρόβλεψης και κατ' επέκταση να οδηγήσει στη λήψη βέλτιστων αποφάσεων.
4. Η επίτευξη συμβατότητας μεταξύ δεδομένων και επιλεγμένου μοντέλου πρόβλεψης. Η συλλογή και η προετοιμασία των δεδομένων είναι εξίσου ένα σημαντικό βήμα για την επιτυχή ανάλυση χρονοσειρών.
5. Η εφαρμογή του μοντέλου δεδομένων κατά τη διάρκεια της οποίας το σύνολο των δεδομένων πρέπει να διαχωριστεί σε ένα σύνολο εκπαίδευσης (training) κι ένα σύνολο δοκιμής (testing).
6. Εκτίμηση της πρόβλεψης που αποτελεί το τελευταίο βήμα της διαδικασίας και το οποίο είναι το ίδιο σημαντικό με τα προηγούμενα, έτσι ώστε να επιβεβαιωθεί ότι οι προβλέψεις που παράχθηκαν από το μοντέλο είναι ακριβείς και αξιόπιστες.

Τα μοντέλα πρόβλεψης που έχει τη δυνατότητα να επιλέξει ο αναλυτής για να διενεργήσει προβλέψεις είναι αρκετά. Οι κύριες μέθοδοι για τη διεξαγωγή προβλέψεων είναι δύο η ποιοτική και η ποσοτική.

5.3.1 Ποιοτικές Μέθοδοι Πρόβλεψης

Όταν τα ποσοτικά δεδομένα είναι ελλιπή ή δεν υπάρχουν καθόλου τότε εφαρμόζονται οι ποιοτικές μέθοδοι πρόβλεψης οι οποίες αξιοποιούν την ανθρώπινη κρίση και εμπειρία με σκοπό να προβλέψουν μελλοντικά γεγονότα. Οι συγκεκριμένες μέθοδοι πρόβλεψης χρησιμοποιούνται όταν το είδος των προβλέψεων δεν είναι εφικτό να γίνει με τη χρήση ιστορικού συνόλου δεδομένων ή όταν κυκλοφορούν νέα προϊόντα και υπηρεσίες για τα οποία δεν υφίσταται κάποια ιστορική πληροφορία. Οι ειδικοί στον τομέα αυτό εκμεταλλευόμενοι την εμπειρία τους μπορούν να διεξάγουν προβλέψεις που δεν είναι εύκολο να ολοκληρωθούν με αριθμητικά μοντέλα. Σε περιπτώσεις αβεβαιότητας και διακύμανσης οι ποιοτικές μέθοδοι είναι πολύ σημαντικές για τη λήψη πληροφοριών και συμπερασμάτων (Makridakis, 1998).

Ένα από τα χαρακτηριστικά των ποιοτικών μεθόδων πρόβλεψης είναι η δυνατότητα που παρέχουν να αξιοποιούνται στα μοντέλα τους μη μετρήσιμες παράμετροι, όπως είναι οι τάσεις, οι αλλαγές στις προτιμήσεις των καταναλωτών και άλλα. Στις συγκεκριμένες μεθόδους λαμβάνεται υπόψη πλήθος απόψεων μέσω των οποίων αναλύονται σύνθετα θέματα που είναι πιθανό να επηρεάσουν το μέλλον. Οι ποιοτικές μέθοδοι χαρακτηρίζονται από ευελιξία και προσαρμοστικότητα, καθώς είναι δυνατό να χρησιμοποιηθούν σε διαφορετικά σενάρια και να παρέχουν πολυδιάστατη προσέγγιση στις προβλέψεις (Gilliland, 2010). Η ευελιξία και η προσαρμοστικότητα είναι στοιχεία ζωτικής σημασίας όταν οι προβλέψεις πρέπει να γίνουν για προϊόντα ή υπηρεσίες για τα οποία δεν υπάρχουν διαθέσιμα ιστορικά δεδομένα.

Οι ποιοτικές μέθοδοι πρόβλεψης επιτυγχάνονται με την αξιοποίηση ποικίλων τεχνικών συλλογής και ανάλυσης δεδομένων, όπως για παράδειγμα συνεντεύξεις από ειδικούς και ομάδες εστίασης. Οι συγκεκριμένες τεχνικές παρέχουν τη δυνατότητα συλλογής δεδομένων και ανίχνευσης βασικών μοτίβων που είναι πιθανό να επηρεάσουν τις προβλέψεις. Για να ολοκληρωθούν με επιτυχία οι ποιοτικές μέθοδοι πρόβλεψης είναι πολύ σημαντική η ορθή επιλογή και εκπαίδευση των συμμετεχόντων στη διαδικασία των συνεντεύξεων, αλλά και η δομημένη διαδικασία συλλογής και ανάλυσης δεδομένων (J. Scott Armstrong, 2001).

Στις ποιοτικές μεθόδους πρόβλεψης ανάλογα με τα ερωτήματα για τα οποία πρέπει να δοθούν απαντήσεις αξιοποιούνται διάφορες τεχνικές. Οι πιο γνωστές είναι οι ακόλουθες:

Η τεχνική Delphi χρησιμοποιείται για τη συλλογή δεδομένων από μια ομάδα ειδικών. Στην τεχνική αυτή η ανωνυμία των ατόμων που συμμετέχουν στη διαδικασία ορίζεται ως απαραίτητη προϋπόθεση. Στη συγκεκριμένη τεχνική καθορίζεται μια σειρά από ερωτήσεις και γύρους, όπου οι ερωτήσεις του κάθε γύρου βασίζονται στις απαντήσεις του προηγούμενου. Ουσιαστικά πρόκειται για μια επαναληπτική διαδικασία μέσω της οποίας υπάρχει η δυνατότητα αναθεώρησης των απαντήσεων που έχουν δοθεί, γεγονός που οδηγεί σε πιο αξιόπιστες και ακριβείς προβλέψεις. Η τεχνική Delphi έχει αρκετά πλεονεκτήματα. Ένα από αυτά είναι πως λόγω της ανωνυμίας που υπάρχει ενδυναμώνεται η ελεύθερη έκφραση και μειώνεται η επιρροή ισχυρών προσωπικοτήτων που είναι πολύ πιθανό να επηρεάσουν τα αποτελέσματα των προβλέψεων και των συζητήσεων (Rowe & Wright, 1999). Το πόσο αποτελεσματική και αξιόπιστη θα είναι η τεχνική αυτή, οφείλεται αρκετά στην επιλογή των ατόμων και τη δομή των ερωτήσεων. Τα άτομα που θα συμμετέχουν είναι απαραίτητο να είναι ειδικοί στον τομέα που πρόκειται να εξεταστεί και αντίστοιχα οι ερωτήσεις πρέπει να είναι ξεκάθαρες και να κατευθύνουν στη λήψη πληροφοριών για συγκεκριμένους σκοπούς (Hsu & Sandford, 2007).

Η τεχνική Grass Roots χρησιμοποιείται για τη συλλογή δεδομένων από άτομα με μη διοικητικές θέσεις σε μια επιχείρηση, δηλαδή τους εργαζομένους που είναι πιο κοντά στους καταναλωτές. Με τη χρήση της τεχνικής αυτής αξιοποιούνται η εμπειρία και οι γνώσεις των εργαζομένων που έχουν μεγαλύτερη επαφή με τις τάσεις και τα δεδομένα των αγορών. Η διαδικασία που ακολουθείται είναι η συλλογή δεδομένων από τους πωλητές, του υπαλλήλους εξυπηρέτησης πελατών και όσους εργαζόμενους έχουν σημαντική επαφή με την καθημερινή λειτουργία της επιχείρησης. Όλες οι πληροφορίες που συλλέγονται επικοινωνούνται στα ανώτερα στελέχη της διοίκησης, όπου στη συνέχεια αναλύονται και χρησιμοποιούνται για τη διεξαγωγή προβλέψεων. Μέσω της τεχνικής αυτής εξασφαλίζεται ότι για τις προβλέψεις γίνεται χρήση ρεαλιστικών δεδομένων και λαμβάνεται υπόψη η πραγματική κατάσταση και οι τάσεις της αγοράς (Makridakis., 1998). Ένα από τα βασικά πλεονεκτήματα της τεχνικής Grass Roots είναι ο άμεσος εντοπισμός των αλλαγών και των τάσεων των αγορών από τη στιγμή που τα άτομα που συμμετέχουν στη διαδικασία έχουν αρκετή γνώση της κατάστασης και αντιλαμβάνονται με ταχύτατους ρυθμούς τις εξελίξεις. Για να ολοκληρωθούν με επιτυχία οι προβλέψεις με τη χρήση της συγκεκριμένης τεχνικής είναι απαραίτητο να υπάρχει ανοιχτή επικοινωνία, εμπιστοσύνη αλλά και προθυμία των εργαζομένων να επικοινωνήσουν τις πληροφορίες που έχουν στη διάθεσή τους (J. Scott Armstrong, 2001).

Με την τεχνική της ιστορικής αναλογίας συγκρίνονται γεγονότα ή καταστάσεις του παρόντος με αντίστοιχα ή παρόμοια γεγονότα και καταστάσεις του παρελθόντος. Στη συγκεκριμένη τεχνική η βασική υπόθεση είναι ότι οι τάσεις αλλά και τα μοτίβα του παρελθόντος θα επανεμφανιστούν και στο μέλλον. Τα βήματα της τεχνικής αυτή είναι η αναζήτηση, η ταυτοποίηση και η ανάλυση καταστάσεων του παρελθόντος που εμφανίζουν κοινά χαρακτηριστικά με καταστάσεις του παρόντος, καθώς και η μελέτη και ανάλυση των παραγόντων ή συνθηκών που είχαν επίδραση σ αυτές. Για να χαρακτηριστεί επιτυχής η πρόβλεψη με τη χρήση της εν λόγω τεχνικής είναι πολύ σημαντική η επιλογή και η κατανόηση των παραγόντων που επηρέασαν τις καταστάσεις του παρελθόντος (J. Scott Armstrong, 2001). Αν και η τεχνική αυτή έχει αρκετά πλεονεκτήματα ταυτόχρονα έχει και κάποιους περιορισμούς. Ένα από τους κυριότερους είναι η δυσκολία που υπάρχει στο να ανιχνευτούν παρόμοιες ιστορικές καταστάσεις, καθώς με το πέρασμα του χρόνου οι συνθήκες μεταβάλλονται σε μεγάλο βαθμό.

Η έρευνα αγοράς αξιοποιείται για να ανιχνευτούν και να κατανοηθούν οι τάσεις, οι προτιμήσεις και οι ανάγκες των καταναλωτών. Συλλέγονται δεδομένα από ποικίλες πηγές όπως ερωτηματολόγια, συνεντεύξεις και άλλα, με στόχο την εύρεση της υπάρχουσας κατάστασης της αγοράς. Η τεχνική έρευνας αγοράς αποτελεί βασικό συστατικό στοιχείο για τον καθορισμό στρατηγικών, την ανάπτυξη καινοτόμων προϊόντων και υπηρεσιών, τη βελτίωση των υπαρχόντων και την απόκτηση ανταγωνιστικού πλεονεκτήματος, καθώς οι επιχειρήσεις μέσω αυτής λαμβάνουν σημαντικές πληροφορίες για την επίτευξη όσων αναφέρθηκαν (Kotler & Keller, 2016). Τα βήματα που ακολουθούνται κατά τη διάρκεια της έρευνας αγοράς είναι ο ορισμός των στόχων και η δημιουργία του προτύπου έρευνας. Στη συνέχεια, γίνεται συλλογή δεδομένων μέσω ερευνών και ερωτηματολογίων και τέλος ανάλυση των δεδομένων που παράχθηκαν από τις έρευνες αυτές με σκοπό να διεξαχθούν οι προβλέψεις μελλοντικές συμπεριφορών. Το κύριο πλεονέκτημα της έρευνας αγοράς είναι ότι λαμβάνονται άμεσα πληροφορίες από την ομάδα των καταναλωτών και λόγω αυτού οι προβλέψεις είναι πιο αξιόπιστες και ακριβείς.

5.3.2 Ποσοτικές Μέθοδοι Πρόβλεψης

Οι ποσοτικές μέθοδοι πρόβλεψης είναι ένα από τα βασικά εργαλεία που χρησιμοποιούνται για την ανάλυση και πρόβλεψη μελλοντικών τάσεων και μοτίβων. Οι συγκεκριμένες μέθοδοι στηρίζονται στη συλλογή και επεξεργασία ιστορικών συνόλων δεδομένων και την αξιοποίηση μαθηματικών και στατιστικών αλγορίθμων με απώτερο σκοπό τη λήψη χρήσιμων πληροφοριών αναφορικά με μελλοντικές προβλέψεις.

Οι ποσοτικές μέθοδοι πρόβλεψης λόγω του γεγονότος ότι οι προβλέψεις που διεξάγονται βασίζονται σε αριθμητικά δεδομένα και μαθηματικές και στατιστικές αναλύσεις είναι πιο ακριβείς και αξιόπιστες σε σχέση με τις ποιοτικές μεθόδους. Οι επιχειρήσεις χρησιμοποιούν ποσοτικές μεθόδους για να διενεργήσουν προβλέψεις που αφορούν στη ζήτηση των προϊόντων και υπηρεσιών τους, να βελτιστοποιήσουν τη διαδικασία διαχείρισης των αποθεμάτων τους και να δημιουργήσουν αποτελεσματικότερες στρατηγικές (Hyndman & Athanasopoulos, 2018). Είναι σημαντικό να αναφερθεί ότι οι συγκεκριμένες μέθοδοι πρόβλεψης κατέχουν κύριο ρόλο στις διαδικασίες διαχείρισης κινδύνου και λήψης στρατηγικών αποφάσεων.

Στις ποσοτικές μεθόδους πρόβλεψης αξιοποιούνται διάφορα μοντέλα και μέθοδοι πρόβλεψης, όπως η μέθοδος του μέσου όρου (Averaging Forecasting Method), το μοντέλο Autoregressive Integrated Moving Average (ARIMA), το μοντέλο Seasonal Autoregressive Integrated Moving Average (SARIMA), το μοντέλο Holt, το μοντέλο Holt Winters και άλλα.

Μέθοδος του Μέσου Όρου (Averaging Forecasting Method)

Η μέθοδος του μέσου όρου (Averaging Forecasting Method) αποτελεί μία στατιστική τεχνική η οποία αξιοποιείται για να προβλεφθούν μελλοντικές τιμές σε μια συγκεκριμένη χρονική σειρά συνόλου δεδομένων. Η μέθοδος αυτή βασίζεται στο γεγονός ότι οι μελλοντικές τιμές είναι εφικτό να προβλεφθούν με περισσότερη ακρίβεια όταν υπολογίζεται ο μέσος όρος των αντίστοιχων ιστορικών τιμών. Η μέθοδος του μέσου όρου χρησιμοποιείται σε περιπτώσεις όπου τα σύνολα δεδομένων δεν εμφανίζουν μεγάλες τάσεις ή εποχικές διακυμάνσεις. Για τον υπολογισμό του μέσου όρου γίνεται χρήση διαφορετικών τύπων, όπως ο απλός κινητός μέσος όρος (Simple Moving Average – SMA), ο εκθετικός κινητός μέσος όρος (Exponential Moving Average – EMA) και ο σταθμισμένος κινητός μέσος (Weighted Moving Average – WMA).

Απλός Κινητός Μέσος Όρος (Simple Moving Average – SMA)

Ο απλός κινητός μέσος όρος (Simple Moving Average – SMA) αξιοποιώντας δεδομένα συγκεκριμένης χρονικής περιόδου υπολογίζει την τιμή του μέσου όρου για το εκάστοτε σύνολο δεδομένων. Η διαφορά με τον παραδοσιακό υπολογισμό του μέσου όρου είναι ότι με το που ληφθεί ένα νέο σύνολο δεδομένων, το παλαιότερο σύνολο δεδομένων δε συνεχίζει να χρησιμοποιείται. Συνεπώς, οι υπολογισμοί του μέσου όρου βασίζονται στα πιο πρόσφατα δεδομένα. Ο τύπος που χρησιμοποιείται για τον υπολογισμό του απλού κινητού μέσου όρου είναι ο ακόλουθος:

$$SMA = \frac{1}{N} \sum_{t=1}^N X_t$$

όπου:

N είναι ο αριθμός των περιόδων και

X_t είναι η τιμή τη χρονική στιγμή t (Hyndman & Athanasopoulos, 2018).

Εκθετικός Κινητός Μέσος Όρος (Exponential Moving Average - EMA)

Ο εκθετικός κινητός μέσος όρος (Exponential Moving Average - EMA) λειτουργεί με παρόμοιο τρόπο με τον απλό κινητό μέσο όρο. Η διαφορά τους είναι ότι η εξίσωση του εκθετικού κινητού μέσου όρου είναι πιο περίπλοκη από αυτή του απλού, καθώς υπολογίζονται οι πιο πρόσφατες τιμές με εκθετικά μειούμενα βάρη. Ο τύπος που χρησιμοποιείται για τον υπολογισμό του εκθετικού κινητού μέσου όρου είναι ο ακόλουθος:

$$EMA_t = \alpha X_t + (1 - \alpha) EMA_{t-1}$$

όπου:

α είναι ο παράγοντας εξομάλυνσης ($0 < \alpha \leq 1$) (Makridakis, Wheelwright & Hyndman, 1998).

Σταθμισμένος Κινητός Μέσος Όρος (Weighted Moving Average - WMA)

Ο σταθμισμένος κινητός μέσος όρος (Weighted Moving Average - WMA) χρησιμοποιείται για την εύρεση των τάσεων και την πρόβλεψη πιθανών διακυμάνσεων των τιμών. Στον συγκεκριμένο τύπο

υπολογισμού μέσου όρου η μεγαλύτερη βαρύτητα κατανέμεται στις πιο πρόσφατες παρατηρήσεις. Ο τύπος που χρησιμοποιείται για τον υπολογισμό του σταθισμένου κινητού μέσου όρου είναι ο ακόλουθος:

$$WMA = \frac{\sum_{t=1}^N W_t X_t}{\sum_{t=1}^N W_t}$$

όπου:

W_t είναι τα βάρη που αποδίδονται σε κάθε τιμή X_t (Chatfield, 2000).

Μοντέλο ARIMA (AutoRegressive Integrated Moving Average)

Το μοντέλο ARIMA είναι ένα αυτοπαλινδρομικό μοντέλο κινητών μέσων όρων και αποτελεί ένα δυνατό εργαλείο που χρησιμοποιείται στην ανάλυση και πρόβλεψη χρονοσειρών. Στο συγκεκριμένο μοντέλο περιέχονται τριών ειδών συνιστώσες, η αυτοπαλινδρόμηση (AR), η ολοκλήρωση (I) και ο κινητός μέσος όρος (MA). Με την αυτοπαλινδρόμηση ελέγχεται η σχέση ανάμεσα στις τρέχουσες τιμές και τις ιστορικές τιμές μιας χρονοσειράς, με τον κινητό μέσο όρο υπολογίζεται η σχέση των τρεχουσών τιμών με προηγούμενα σφάλματα και με την ολοκλήρωση επιτυγχάνεται η στασιμότητα της χρονοσειράς, καθώς αφαιρούνται οι τάσεις μέσω της διαφοροποίησης των δεδομένων (Box, Jenkins, & Reinsel, 2015). Για να εφαρμοστεί το μοντέλο ARIMA η διαδικασία έχει ως αφετηρία την ταυτοποίηση των προηγούμενων τιμών και σφαλμάτων (p και q) και τον αριθμό των διαφορών που είναι απαραίτητος για να επιτευχθεί η στασιμότητα (d). Οι συγκεκριμένες παράμετροι είναι εφικτό να καθοριστούν μέσω διαγραμμάτων αυτοσυσχέτισης και μερικής αυτοσυσχέτισης. Έπειτα, η εκτίμηση των παραμέτρων γίνεται με τη χρήση μεθόδων όπως η μέγιστη πιθανότητα και τέλος η καταλληλότητα το μοντέλου εξετάζεται με την αξιοποίηση διαγνωστικών ελέγχων (Hyndman & Athanasopoulos, 2018). Αν και το συγκεκριμένο μοντέλο έχει αρκετά πλεονεκτήματα, παρουσιάζει μερικούς περιορισμούς. Πιο συγκεκριμένα, για να επιτευχθεί η ακριβής ταυτοποίηση και εκτίμηση των παραμέτρων του χρειάζεται μεγάλος όγκος δεδομένων. Επίσης, είναι μη ενδεδειγμένο για ανάλυση χρονοσειρών που εμφανίζουν υψηλές διακυμάνσεις και η επιλογή του βέλτιστου μοντέλου ARIMA είναι πιθανό να αποτελέσει μια περίπλοκη διαδικασία (Hamilton, 1994). Ακολουθεί η ανάλυση της κάθε συνιστώσας του μοντέλου ARIMA και ο τρόπος με τον οποίο συνδυάζονται.

Η συνιστώσα της αυτοπαλινδρόμησης (AutoRegressive - AR) υπολογίζει το πως η υφιστάμενη τιμή της χρονοσειράς εξαρτάται από προηγούμενες τιμές.

$$Y_t = c + \sum_{i=1}^p \varphi_i Y_{t-i} + \varepsilon_t$$

όπου:

Y_t είναι η τιμή της χρονικής σειράς στη χρονική στιγμή t ,

c είναι μια σταθερά,

φ_i είναι οι παράμετροι του μοντέλου AR που πρέπει να εκτιμηθούν και

ε_t είναι το τυχαίο σφάλμα ή θόρυβος.

Η συνιστώσα της ολοκλήρωσης (Integrated - I) είναι η διαφορά στις τιμές της χρονοσειράς για την επίτευξη στασιμότητας. Το I(d) μοντέλο σημαίνει ότι η χρονοσειρά έχει διαφοροποιηθεί d φορές. Για παράδειγμα, αν $d=1$ τότε η διαφοροποιημένη σειρά είναι

$d=1$, τότε η διαφοροποιημένη σειρά είναι:

$$Y'_t = Y_t - Y_{t-1}$$

Η συνιστώσα του κινητού μέσου όρου (Moving Average - MA) αναλύει τη σχέση που υπάρχει ανάμεσα στην παρούσα τιμή της χρονοσειράς και των προηγούμενων σφαλμάτων (τυχαίων σφαλμάτων). Το MA(q) μοντέλο εκφράζεται ως:

$$Y_t = c + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

όπου:

θ_i είναι οι παράμετροι του μοντέλου Moving Average που πρέπει να εκτιμηθούν και

ε_t είναι το τυχαίο σφάλμα ή θόρυβος.

Ο συνδυασμένος τύπος ARIMA(p,d,q) ενσωματώνει τις συνιστώσες AR, I και MA. Το πλήρες μοντέλο γράφεται ως εξής:

$$Y_t = c + \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

όπου:

p είναι η τάξη της συνιστώσας AR,

d είναι ο αριθμός των διαφορών για την επίτευξη στασιμότητας και

q είναι η τάξη της συνιστώσας MA.

Μοντέλο SARIMA (Seasonal AutoRegressive Integrated Moving Average)

Το μοντέλο SARIMA αποτελεί την επέκταση του μοντέλου ARIMA με τη διαφορά ότι εμπεριέχει εποχικές συνιστώσες, γεγονός που το καθιστά ιδανικό για τη μοντελοποίηση χρονοσειρών που εμφανίζουν επαναλαμβανόμενα μοτίβα σε συγκεκριμένες χρονικές περιόδους. Το μοντέλο SARIMA συμπεριλαμβάνει όλες τις συνιστώσες του ARIMA και προσθέτει εποχικές συνιστώσες, έτσι ώστε να λαμβάνονται υπόψη τις εποχικές διακυμάνσεις των δεδομένων. Η διαδικασία της εφαρμογής του συγκεκριμένου μοντέλου αρχίζει με την ανάλυση της εποχικότητας του συνόλου των δεδομένων, όπου με τον τρόπο αυτό προσδιορίζονται η περίοδος και τα επαναλαμβανόμενα μοτίβα. Στη συνέχεια, διενεργείται η ταυτοποίηση των παραμέτρων για τις εποχικές αλλά και τις μη εποχικές συνιστώσες του μοντέλου. Οι συγκεκριμένες παράμετροι είναι εφικτό να καθοριστούν με τη χρήση διαγραμμάτων αυτοσυσχέτισης και μερικής αυτοσυσχέτισης, αλλά και μέσω δοκιμών στασιμότητας και εποχιακής διαφοροποίησης. Ένα από τα βασικά βήματα της παρούσας διαδικασίας είναι η εκτίμηση των παραμέτρων του μοντέλου, το οποίο επιτυγχάνεται κυρίως μέσω μεθόδων μέγιστης πιθανότητας ή ελάχιστων τετραγώνων. Το μοντέλο SARIMA έχει αρκετά πλεονεκτήματα όμως παρουσιάζει και κάποιους περιορισμούς. Με την προσθήκη των εποχικών παραμέτρων αυξάνεται η πολυπλοκότητα και η εκτίμηση των παραμέτρων γίνεται πιο απαιτητική. Τέλος, για την ακριβή ανάλυση και πρόβλεψη απαιτείται μεγάλος όγκος δεδομένων, ειδικά σε περιπτώσεις όπου οι εποχικές συνιστώσες είναι περίπλοκες. Ακολουθεί η ανάλυση της κάθε συνιστώσας του μοντέλου SARIMA και ο τρόπος με τον οποίο συνδυάζονται.

Η αυτοπαλινδρόμηση (AutoRegressive - AR):

$$AR(p): Y_t = c + \sum_{i=1}^p \varphi_i Y_{t-i} + \varepsilon_t$$

όπου:

φ_i είναι οι παράμετροι αυτοπαλινδρόμησης και

ε_t είναι το τυχαίο σφάλμα ή θόρυβος.

Η ολοκλήρωση (Integrated - I):

$$I(d): (1 - B)^d Y_t$$

όπου:

B είναι ο τελεστής καθυστέρησης και

d ο βαθμός διαφοροποίησης.

Ο κινητός μέσος όρος (Moving Average - MA):

$$MA(q): Y_t = c + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

όπου:

θ_i είναι οι παράμετροι του κινητού μέσου όρου.

Στο μοντέλο SARIMA οι εποχικές συνιστώσες προσθέτουν εποχική αυτοπαλινδρόμηση (SAR), εποχική ολοκλήρωση (SI) και εποχικό κινητό μέσο όρο (SMA).

Η εποχική αυτοπαλινδρόμηση (SAR):

$$SAR(P): \Phi_P(B^S) Y_t = \Phi_P Y_{t-s}$$

όπου:

s είναι η εποχική περίοδος και

Φ_P οι εποχικές παράμετροι αυτοπαλινδρόμησης.

Η εποχική ολοκλήρωση (SI):

$$SI(D): (1 - B^s)^D Y_t$$

όπου:

D είναι ο βαθμός εποχικής διαφοροποίησης.

Ο εποχικός κινητός μέσος όρος (SMA):

Τμηματοποίηση πελατών και πρόβλεψη πωλήσεων
για μια ηλεκτρονική πλατφόρμα πωλήσεων

$$\text{SMA}(Q): Y_t = \varepsilon_t + \sum_{i=1}^Q \theta_i \varepsilon_{t-is}$$

όπου:

θ_i είναι οι εποχικές παράμετροι του κινητού μέσου όρου.

Το ολοκληρωμένο μοντέλο SARIMA(p, d, q)(P, D, Q)[s] συνδυάζει τις παραπάνω συνιστώσες και μπορεί να γραφεί ως εξής:

$$(1 - \sum_{i=1}^p \phi_i B^i)(1 - B)^d(1 - \sum_{i=1}^p \phi_i B^{si}) (1 - B^s)^D Y_t = (1 + \sum_{i=1}^q \theta_i B^i) (1 + \sum_{i=1}^Q \theta_i B^{si}) \varepsilon_t$$

Μοντέλο Holt

Το μοντέλο Holt ή αλλιώς μοντέλο εκθετικής εξομάλυνσης με προσαρμογή στην τάση αποτελεί μία προέκταση της απλής εκθετικής εξομάλυνσης και εφαρμόζεται σε περιπτώσεις που οι παρατηρήσεις της χρονοσειράς δείχνουν την ύπαρξη γραμμικής τάσης. Το συγκεκριμένο μοντέλο αναπτύχθηκε από τον Charles Holt το 1957 και αξιοποιείται για την πρόβλεψη χρονοσειρών οι οποίες δεν χαρακτηρίζονται από σταθερότητα, αλλά εμφανίζουν διαρκή αύξηση ή μείωση. Η κύρια ιδέα του μοντέλου αυτού είναι ότι γίνεται χρήση δύο εξομαλυντικών παραμέτρων, μία που αφορά στην εξομάλυνση των τιμών της χρονοσειράς και μία στην εξομάλυνση της τάσης, γεγονός που βοηθά στην προσαρμογή του μοντέλου στις κυμαινόμενες συνθήκες που παρατηρούνται στη χρονοσειρά που αναλύεται (Holt, 2004). Η εξίσωση της εξομάλυνσης για το επίπεδο εξετάζει την τιμή της υφιστάμενης παρατήρησης και την προβλεπόμενη τιμή της προηγούμενης περιόδου ρυθμισμένη για την τάση. Η εξίσωση που αφορά στην τάση ενημερώνει την εκτίμηση της τάσης λαμβάνοντας υπόψη τη διαφορά ανάμεσα στα υφιστάμενα και τα προηγούμενα επίπεδα. Ο υπολογισμός αυτός δίνει τη δυνατότητα στο μοντέλο Holt να ανταποκρίνεται στις βραχυπρόθεσμες μεταβολές και τις μακροπρόθεσμες τάσεις, γεγονός που προσφέρει περισσότερη ακρίβεια και ευελιξία από το μοντέλο της απλής εκθετικής εξομάλυνσης (Hyndman & Athanasopoulos, 2018). Με τη χρήση του συγκεκριμένου μοντέλου οι επιχειρήσεις έχουν τη δυνατότητα να διενεργήσουν προβλέψεις για τη ζήτηση των προϊόντων τους και με τις πληροφορίες που θα λάβουν να ρυθμίσουν κατάλληλα τα αποθέματά τους και να κάνουν πιο ακριβή προγραμματισμό της παραγωγής τους (Gardner, 1985). Το μοντέλο Holt παρά τα πλεονεκτήματά του παρουσιάζει συγκεκριμένους περιορισμούς. Δεν είναι κατάλληλο για δεδομένα που εμφανίζουν μη γραμμικές τάσεις ή έντονες εποχικές διακυμάνσεις. Σε τέτοιου είδους περιπτώσεις, προτείνεται να χρησιμοποιηθεί το μοντέλο της εκθετικής εξομάλυνσης με προσαρμογή στην τάση και την εποχικότητα, γνωστό ως μοντέλο Holt-Winters, το οποίο συμπεριλαμβάνει και τις εποχικές συνιστώσες. Οι βασικοί τύποι που χρησιμοποιούνται στη Μέθοδο Holt είναι οι ακόλουθοι:

Η εξίσωση εξομάλυνσης επιπέδου:

$$l_t = \alpha Y_t + (1 - \alpha)(l_{t-1} + b_{t-1})$$

όπου:

l_t είναι η εκτίμηση του επιπέδου στη χρονική στιγμή t ,

α είναι παράμετρος εξομάλυνσης για το επίπεδο ($0 < \alpha < 1$),

Y_t είναι η πραγματική τιμή στη χρονική στιγμή t ,

l_{t-1} είναι η εκτίμηση του επιπέδου την προηγούμενη χρονική στιγμή $t-1$ και

b_{t-1} είναι η εκτίμηση της τάσης την προηγούμενη χρονική στιγμή $t-1$.

Η εξίσωση εξομάλυνσης τάσης:

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$$

όπου:

b_t είναι η εκτίμηση της τάσης τη χρονική στιγμή t ,

β είναι η παράμετρος εξομάλυνσης για την τάση ($0 < \beta < 1$),

l_t είναι η εκτίμηση του επιπέδου τη χρονική στιγμή t και

l_{t-1} είναι η εκτίμηση του επιπέδου την προηγούμενη χρονική στιγμή $t-1$.

Η εξίσωση πρόβλεψης:

$$\hat{Y}_{t+h} = l_t + hb_t$$

όπου:

\hat{Y}_{t+h} είναι η προβλεπόμενη τιμή για h χρονικές στιγμές στο μέλλον,

l_t είναι η εκτίμηση του επιπέδου τη χρονική στιγμή t ,
 h είναι ο αριθμός των περιόδων στο μέλλον για την πρόβλεψη και
 b_t είναι η εκτίμηση της τάσης τη χρονική στιγμή t .

Μοντέλο Holt-Winters

Το μοντέλο Holt-Winters η αλλιώς μοντέλο εκθετικής εξομάλυνσης με προσαρμογή στην τάση και την εποχικότητα αποτελεί την προέκταση του μοντέλου Holt, με τη διαφορά ότι συμπεριλαμβάνει στην πρόβλεψη και τη συνιστώσα της εποχικότητας. Το μοντέλο αυτό δημιουργήθηκε από τον Charles Holt και τον Peter Winters τη δεκαετία του 1960 και είναι κατάλληλο για την πρόβλεψη χρονοσειρών που εμφανίζουν τάσεις και εποχικές διακυμάνσεις. Η συνιστώσα της εποχικότητας υπάρχει περίπτωση να είναι προσθετική ή πολλαπλασιαστική, δηλαδή οι εποχικές διακυμάνσεις προστίθενται στην κύρια τάση ή οι εποχικές διακυμάνσεις να πολλαπλασιάζονται με την κύρια τάση αντίστοιχα (Hyndman & Athanadoroulos, 2018). Στο μοντέλο Holt-Winters η εξομάλυνση του επιπέδου, της τάσης και της εποχικότητας επιτυγχάνεται με το συνδυασμό τριών εξισώσεων. Η εξίσωση επιπέδου λαμβάνει υπόψη την υφιστάμενη τιμή, την προηγούμενη τάση και την εποχική συνιστώσα. Η εξίσωση τάσης διαμορφώνεται ανάλογα με την αλλαγή στο επίπεδο και η εξίσωση εποχικότητας ενημερώνεται βάσει της υπάρχουσας τιμή και του επικαιροποιημένου επιπέδου. Ο συνδυασμός των τριών εξισώσεων συμβάλλει στη δυναμική ενάρμοση του μοντέλου με τις αλλαγές του επιπέδου, της τάσης και της εποχικότητας (Gardner, 1985). Το συγκεκριμένο μοντέλο αξιοποιείται κυρίως για την ανάλυση χρονοσειρών που παρουσιάζουν περιοδικές διακυμάνσεις όπως για παράδειγμα οι μηνιαίες πωλήσεις. Το μοντέλο Holt-Winters παρά τα πλεονεκτήματά του παρουσιάζει συγκεκριμένους περιορισμούς. Για τη χρήση του είναι απαραίτητη μεγάλο όγκος δεδομένων, έτσι ώστε να είναι εφικτό να υπολογιστούν με ακρίβεια οι εποχικές συνιστώσες. Επιπρόσθετα, το μοντέλο μπορεί να παρουσιάσει ευαισθησία σε αλλαγές των εποχικών τάσεων και για τον λόγο αυτό να επηρεαστεί η ακρίβεια των προβλέψεων (Hyndman & Athanadoroulos, 2018; Gardner, 1985). Οι βασικοί τύποι που χρησιμοποιούνται στη Μέθοδο Holt-Winters είναι οι ακόλουθοι:

Η εξίσωση εξομάλυνσης επιπέδου:

$$l_t = \alpha \frac{Y_t}{S_{t-L}} + (1 - \alpha)(l_{t-1} + b_{t-1})$$

όπου:

l_t είναι η εκτίμηση του επιπέδου τη χρονική στιγμή t ,
 α είναι η εξομαλυντική παράμετρος για το επίπεδο ($0 < \alpha < 1$),
 Y_t είναι η πραγματική τιμή στη χρονική στιγμή t ,
 S_{t-L} είναι η εποχική συνιστώσα της περιόδου πριν από L χρονικές στιγμές,
 l_{t-1} είναι η εκτίμηση του επιπέδου την προηγούμενη χρονική στιγμή $t-1$ και
 b_{t-1} είναι η εκτίμηση της τάσης την προηγούμενη χρονική στιγμή $t-1$.

Η εξίσωση εξομάλυνσης τάσης:

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$$

όπου:

b_t είναι η εκτίμηση της τάσης τη χρονική στιγμή t ,
 β είναι η παράμετρος εξομάλυνσης για την τάση ($0 < \beta < 1$),
 l_t είναι η εκτίμηση του επιπέδου τη χρονική στιγμή t και
 l_{t-1} είναι η εκτίμηση του επιπέδου την προηγούμενη χρονική στιγμή $t-1$.

Η εξίσωση εξομάλυνσης εποχικότητας:

$$S_t = \gamma \frac{Y_t}{l_t} + (1 - \gamma) S_{t-L}$$

όπου,

S_t είναι η εποχική συνιστώσα τη χρονική στιγμή t ,
 γ είναι η εξομαλυντική παράμετρος για την εποχικότητα ($0 < \gamma < 1$),
 Y_t είναι η πραγματική τιμή στη χρονική στιγμή t ,
 l_t είναι η εκτίμηση του επιπέδου τη χρονική στιγμή t και

S_{t-L} είναι η εποχική συνιστώσα της περιόδου πριν από L χρονικές στιγμές.

Η εξίσωση πρόβλεψης:

$$\hat{Y}_{t+h} = (l_t + hb_t) S_{t+h-L(k+1)}$$

όπου,

\hat{Y}_{t+h} είναι η προβλεπόμενη τιμή για h χρονικές στιγμές στο μέλλον,

l_t είναι η εκτίμηση του επιπέδου τη χρονική στιγμή t ,

b_t είναι η εκτίμηση της τάσης τη χρονική στιγμή t και

$S_{t+h-L(k+1)}$ είναι η εποχική συνιστώσα για την χρονική στιγμή $t+h$.

6. Διερευνητική Ανάλυση Δεδομένων (EDA)

Στόχος της παρούσας διπλωματικής είναι η τμηματοποίηση πελατών με τη χρήση της RFM ανάλυσης και η πρόβλεψη πωλήσεων με την αξιοποίηση του μοντέλου Holt-Winters.

Για να επιτευχθεί η εύρεση των διαφορετικών ομάδων πελατών και να διενεργηθεί πρόβλεψη πωλήσεων για συγκεκριμένα καταστήματα θα χρειαστεί να προετοιμαστεί κατάλληλα το σύνολο δεδομένων. Το συγκεκριμένο θα επιτευχθεί με την εφαρμογή της μεθόδου διερευνητικής ανάλυσης δεδομένων (EDA - Exploratory Data Analysis), με στόχο την μετατροπή και εξερεύνηση των δεδομένων κάνοντας χρήση τεχνικών οπτικοποίησης και στατιστικής. Με την ολοκλήρωση της διερευνητικής ανάλυσης δεδομένων και λαμβάνοντας υπόψιν τις πληροφορίες που παράχθηκαν μέσω αυτής θα γίνει η τμηματοποίηση πελατών με την αξιοποίηση της RFM ανάλυσης και τέλος θα διενεργηθεί η πρόβλεψη πωλήσεων με τη χρήση του αλγορίθμου Holt-Winters.

Τα δεδομένα που χρησιμοποιήθηκαν για το πρακτικό μέρος αφορούν την περίοδο δύο μηνών από 01/04/2021 έως 31/05/2021.

Η διερευνητική ανάλυση δεδομένων αποτελεί μία μέθοδο που χρησιμοποιείται στην ανάλυση δεδομένων και την οπτικοποίησή τους, με στόχο την εύρεση των βασικών χαρακτηριστικών τους. Με τη χρήση της διερευνητικής ανάλυσης δεδομένων είναι εφικτό να κατανοηθούν μοτίβα και σχέσεις μεταξύ των δεδομένων και να ανιχνευτούν ενδεχόμενες τάσεις και ακραίες τιμές. Στη διαδικασία της ανάλυσης δεδομένων η διερευνητική ανάλυση είναι ένα από τα πρώτα βήματα που διενεργούνται και κατέχει κυρίαρχο ρόλο στην διατύπωση υποθέσεων και στον τρόπο μοντελοποίησης των δεδομένων και κατ'επέκταση της ανάλυσης. Η διερευνητική ανάλυση δεδομένων αποτελείται από πέντε βασικούς τύπους:

1. Μονομεταβλητή ανάλυση (Univariate analysis), η οποία αποτελεί την πιο απλή μορφή της διερευνητικής ανάλυσης δεδομένων, καθώς τα δεδομένα στα οποία γίνεται ανάλυση αποτελούνται μόνο από μία μεταβλητή. Στόχος είναι η περιγραφή των δεδομένων και η ανίχνευση μοτίβων σε αυτά.
2. Διμεταβλητή ανάλυση (Bivariate analysis), η οποία χρησιμοποιείται για την εύρεση των σχέσεων μεταξύ των μεταβλητών.
3. Πολυμεταβλητή ανάλυση (Multivariate analysis), η οποία χρησιμοποιείται για την κατανόηση των σχέσεων ανάμεσα σε δύο ή περισσότερες μεταβλητές του συνόλου των δεδομένων.
4. Ανάλυση χρονοσειρών (Time series analysis), κατά την οποία αναλύεται μια ακολουθία σημειών που έχει συλλεχθεί σε ένα συγκεκριμένο χρονικό διάστημα.
5. Ανάλυση χαμένων τιμών (Missing values analysis), η οποία χρησιμοποιείται για τον εντοπισμό των τιμών που λείπουν από το σύνολο δεδομένων.

Στη συνέχεια, παρουσιάζονται αναλυτικά τα βήματα της διερευνητικής ανάλυσης και τα αποτελέσματα αυτής. Το πρώτο βήμα της εξερεύνησης των δεδομένων είναι η εύρεση του αριθμού των γραμμών και των στηλών του συνόλου των δεδομένων. Στο σύνολο δεδομένων της ηλεκτρονικής πλατφόρμας πωλήσεων το σύνολο των γραμμών είναι 581.946 και το σύνολο των στηλών 20.

```
In [3]: data.shape
```

```
Out[3]: (581946, 20)
```

Εικόνα 2 Σύνολο γραμμών και στηλών

```
In [4]: data.head()
```

```
Out[4]:
```

	customer_id	order_id	order_timestamp	shop_id	shop_is_chain	order_value	
0	572abeaa18b4cbf193fb0c375e78bc0d	ee71b15d83b2e6033ccb14f053fb0cde	2021-04-24 19:11:25 UTC	bbca124ee3ee991bfb36277f2910a66a	True	7.575	Ri
1	6cf012d298f991da5469e0e9611fe2c1	a2df8551c40495007e1cde42b9d82bf	2021-04-24 11:59:52 UTC	ec2f46d836b5a5f14f44193186970b63	False	7.600	Ri
2	2a710488e6269883737121fc7dc3a61b	a99034c8ae90e05aaace516f3d8d04bb3	2021-04-18 09:49:36 UTC	3b79459351228b9d823b0ec9e40cf952	True	6.200	Ri
3	c2bc3b580812615d8dc6934b7aa68997	6e1678c4b70441b610a41ecd27b8724f	2021-04-19 09:25:07 UTC	e5978fcb9838d7a9701c309d6f6ea7a7	False	2.700	Ri
4	87ff27d1b4a2726b82e917795603d349	1a64ee45bb91d3f26fd2203c25d33d60	2021-04-19 10:31:36 UTC	2ea3746d0213b19b382219a682dc76a0	False	2.400	Ri

Εικόνα 3 Πρώτες γραμμές συνόλου δεδομένων

Τμηματοποίηση πελατών και πρόβλεψη πωλήσεων για μια ηλεκτρονική πλατφόρμα πωλήσεων


```
In [5]: data.tail()
```

```
Out[5]:
```

	customer_id	order_id	order_timestamp	shop_id	shop_is_chain	order_value
581941	9b23abe4b25bf326774ac506045fd414	e7c3e5defe0acc8b23689316acd028de	2021-05-20 18:57:52 UTC	d85bc768df554582e5e570d734d79a30	False	20.4
581942	6e5e306728a8c720ce2d8ad344f80b02	d46b125dc9076e85cf48e7096bfe15a8	2021-05-12 14:23:49 UTC	e24ae1f35e757ae5451d45d8359664a4	False	5.5
581943	f205f88f10357a4a38286d72c3d5df39	21946e53aa094fc59bee88ae52133adc	2021-05-02 17:05:04 UTC	edbef787da72ea5b6de32daaadaa4803	True	16.6
581944	305919674912655716908fa5c7c1340c	41b8816f18e2c1d4a6dc888a744359d8	2021-05-08 12:52:08 UTC	0e9d43c00859e5edb6f46ef302716fa1	False	4.7
581945	e9631e08ac24cc298408c55d8a8d2b21	dd92f3910c328b1ab6768ff442f4990a	2021-05-26 19:49:12 UTC	535b0ef9af2b49f0a6ce6dddeb5a2892	False	20.0

Εικόνα 4 Τελευταίες γραμμές συνόλου δεδομένων

Στη διερευνητική ανάλυση δεδομένων σημαντικό βήμα είναι και η κατανόηση του τύπου των δεδομένων και διάφορων πληροφοριών σχετικά με αυτά, όπως ο αριθμός των εγγραφών σε κάθε στήλη, τα μηδενικά και μη μηδενικά δεδομένα και άλλα. Από την ακόλουθη εικόνα λαμβάνεται η πληροφορία ότι οι μεταβλητές `geographical_region_of_user`, `food_score`, `delivery_score` και `service_score` έχουν ελλιπείς τιμές. Επίσης, οι μεταβλητές `customer_id`, `order_id`, `order_timestamp`, `shop_id`, `shop_cuisine`, `vertical`, `geographical_region_of_shop`, `shop_city`, `geographical_region_of_shop`, `last_order_at` και `first_order_at` είναι κατηγορικές. Οι μεταβλητές `order_value`, `food_score`, `delivery_score` και `service_score` είναι αριθμητικές και οι `shop_is_chain`, `delivered_by_efood`, `is_online_payment`, `order_has_discount_coupon` και `order_from_pinata` είναι Boolean.

```
In [6]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 581946 entries, 0 to 581945
Data columns (total 20 columns):
#   Column                                     Non-Null Count  Dtype
---  ---
0   customer_id                               581946 non-null object
1   order_id                                  581946 non-null object
2   order_timestamp                           581946 non-null object
3   shop_id                                    581946 non-null object
4   shop_is_chain                             581946 non-null bool
5   order_value                               581946 non-null float64
6   vertical                                   581946 non-null object
7   delivered_by_efood                        581946 non-null bool
8   shop_cuisine                              581946 non-null object
9   is_online_payment                         581946 non-null bool
10  order_has_discount_coupon                 581946 non-null bool
11  order_from_pinata                        581946 non-null bool
12  geographical_region_of_user               580759 non-null object
13  shop_city                                 581946 non-null object
14  geographical_region_of_shop               581946 non-null object
15  food_score                               8669 non-null  float64
16  delivery_score                            8790 non-null  float64
17  service_score                             8789 non-null  float64
18  last_order_at                            581946 non-null object
19  first_order_at                           581946 non-null object
dtypes: bool(5), float64(4), object(11)
memory usage: 69.4+ MB
```

Εικόνα 5 Τύποι συνόλου δεδομένων

Στην ακόλουθη εικόνα παρουσιάζεται για κάθε μεταβλητή το άθροισμα των ελλιπών τιμών. Αν χρειαστεί για τα επόμενα βήματα οι συγκεκριμένες μεταβλητές θα χειριστούν με κατάλληλο τρόπο.

```
In [7]: data.isnull().sum()
Out[7]: customer_id      0
order_id      0
order_timestamp  0
shop_id      0
shop_is_chain  0
order_value   0
vertical     0
delivered_by_elfood  0
shop_cuisine  0
is_online_payment  0
order_has_discount_coupon  0
order_from_pinata  0
geographical_region_of_user  1187
shop_city     0
geographical_region_of_shop  450
food_score    573277
delivery_score 573156
service_score 573157
last_order_at  0
first_order_at 0
dtype: int64
```

Εικόνα 6 Ελλιπείς τιμές

Ένα ακόμη βήμα είναι αποτελεί η στατιστική περίληψη των δεδομένων τύπου int, float, καθώς μπορούν να ληφθούν διάφορες πληροφορίες σχετικά με το σύνολο των δεδομένων.

```
In [9]: data.describe().T
Out[9]:
```

	count	mean	std	min	25%	50%	75%	max
order_value	581946.0	8.962665	7.455989	0.0	4.4	7.0	11.1	719.2
food_score	8669.0	4.526243	1.074910	1.0	5.0	5.0	5.0	5.0
delivery_score	8790.0	4.597270	1.028457	1.0	5.0	5.0	5.0	5.0
service_score	8789.0	4.583457	1.062886	1.0	5.0	5.0	5.0	5.0

Ο μέσος όρος του food_score είναι περίπου 4.5. Στη συγκεκριμένη μεταβλητή παρατηρείται μεγάλη διαφορά μεταξύ μέγιστης και ελάχιστης τιμής, καθώς οι ελάχιστες τιμές δίνουν το συμπέρασμα ότι το score 1.0 αποτελεί outlier.

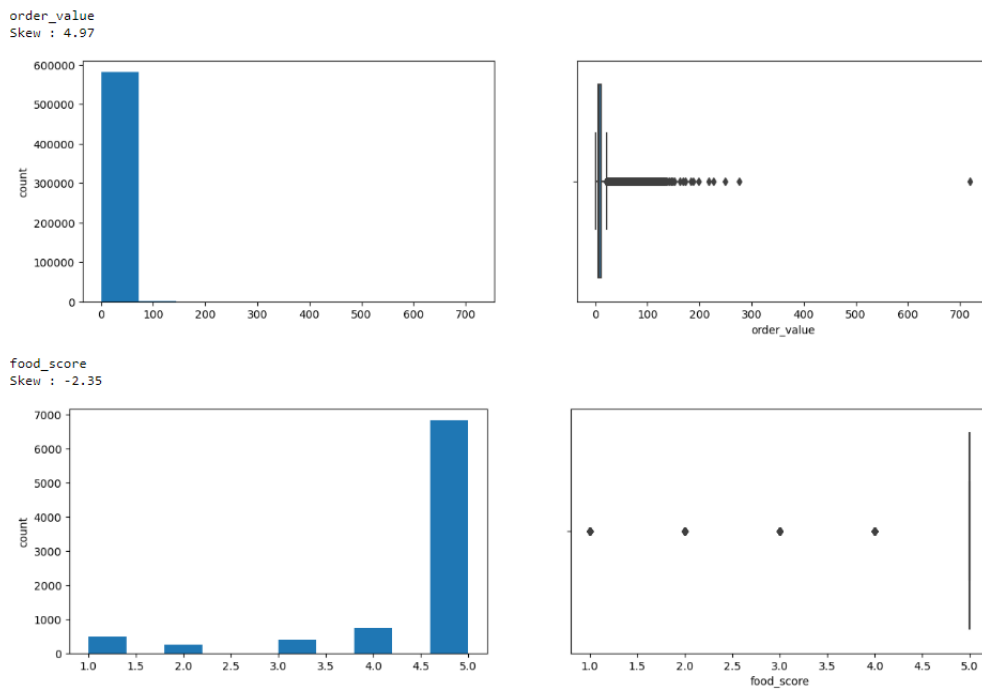
Ο μέσος όρος του delivery_score είναι περίπου 4.59. Στη συγκεκριμένη μεταβλητή παρατηρείται μεγάλη διαφορά μεταξύ μέγιστης και ελάχιστης τιμής, καθώς οι ελάχιστες τιμές δίνουν το συμπέρασμα ότι το score 1.0 αποτελεί outlier.

Ο μέσος όρος του service_score είναι περίπου 4.58. Στη συγκεκριμένη μεταβλητή παρατηρείται μεγάλη διαφορά μεταξύ μέγιστης και ελάχιστης τιμής, καθώς οι ελάχιστες τιμές δίνουν το συμπέρασμα ότι το score 1.0 αποτελεί outlier.

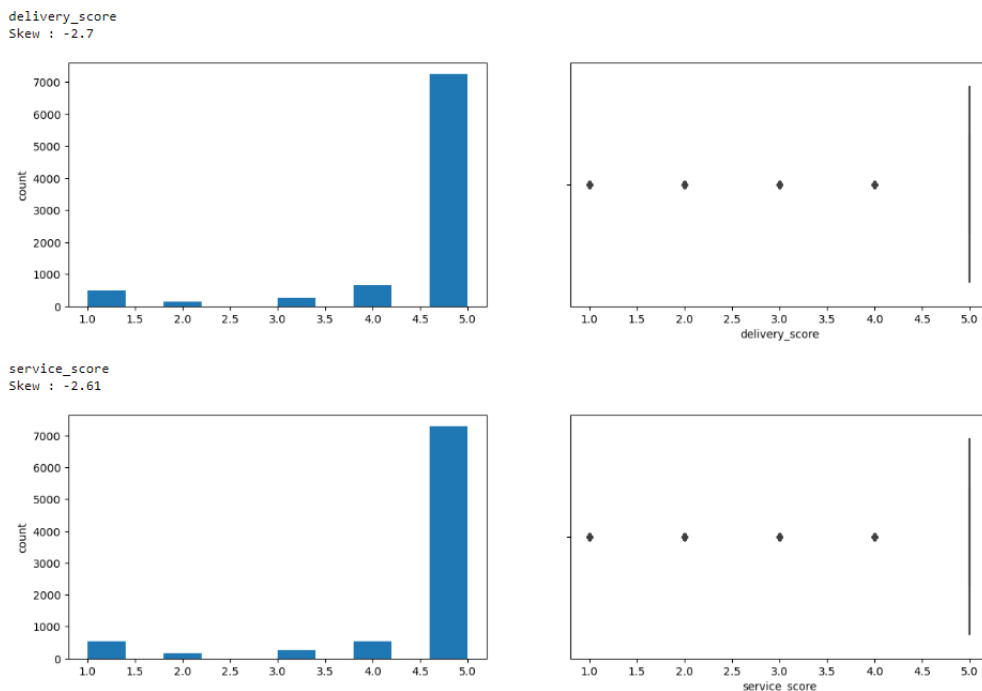
Στις εικόνες 14 και 15 παρουσιάζεται η κατανομή των αριθμητικών μεταβλητών και συγκεκριμένα των order_value, food_score, delivery_score και service_score.

Για τη μεταβλητή order_value παρατηρείται ότι η πλειονότητα της αξίας των παραγγελιών βρίσκεται μεταξύ 0 και 100. Υπάρχουν πολύ λίγες τιμές παραγγελιών πάνω από 100. Λαμβάνοντας υπόψιν τα δύο γραφήματα της μεταβλητής order_value είναι σαφές ότι η κατανομή της αξίας των παραγγελιών δεν είναι κανονική, αλλά έχει ισχυρή θετική κλίση, καθώς οι περισσότερες παραγγελίες έχουν σχετικά χαμηλή αξία και ένας μικρός αριθμός παραγγελιών έχει υψηλή αξία.

Για τη μεταβλητή food_score παρατηρείται ότι η πλειονότητα της βαθμολογίας φαγητού έχει συγκεντρωθεί στο ανώτερο άκρο της κλίμακας και ειδικότερα στη βαθμολογία 5.0. Υπάρχουν πολύ λίγες βαθμολογίες κάτω από 5.0. Λαμβάνοντας υπόψιν τα δύο γραφήματα της μεταβλητής food_score είναι σαφές ότι η κατανομή της αξίας των παραγγελιών δεν είναι κανονική, αλλά έχει ισχυρή αρνητική κλίση, καθώς οι περισσότερες βαθμολογίες έχουν σχετικά υψηλή αξία και ένας μικρός αριθμός βαθμολογιών έχει χαμηλή αξία. Το ίδιο ισχύει και για τις μεταβλητές delivery_score και service_score.

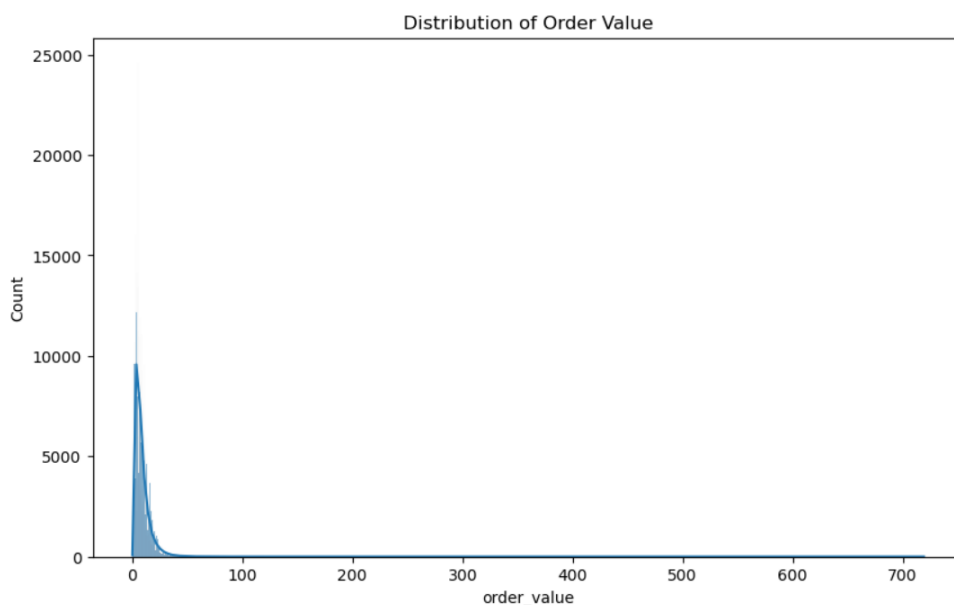


Εικόνα 7 Κατανομή order_value και food_score



Εικόνα 8 Κατανομή delivery_score και service_score

Στην Εικόνα 17 παρουσιάζεται η κατανομή της αξία των παραγγελιών που όπως και στην Εικόνα 14 παρατηρείται ότι η πλειονότητα της αξίας των παραγγελιών βρίσκεται μεταξύ 0 και 100. Υπάρχουν πολύ λίγες τιμές παραγγελιών πάνω από 100.



Εικόνα 9 Κατανομή αξίας παραγγελίας

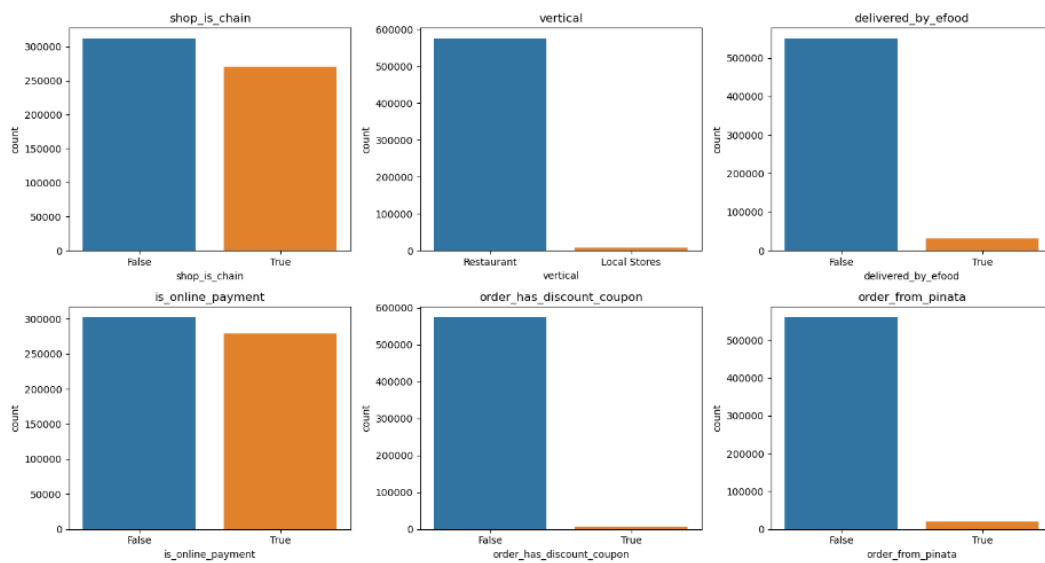
Στην Εικόνα 17 έχουν συγκεντρωθεί έξι γραφήματα στα οποία απεικονίζεται η συχνότητα για τις μεταβλητές τύπου Boolean (`shop_is_chain`, `delivered_by_efood`, `is_online_payment`, `order_has_discount_coupon` και `order_from_pinata`). Η πληροφορία που λαμβάνεται για τις μεταβλητές είναι η ακόλουθη:

1. Για τη μεταβλητή `shop_is_chain` είναι ότι ο αριθμός των καταστημάτων που δεν είναι αλυσίδες είναι ελαφρώς μεγαλύτερος από τον αριθμό των καταστημάτων που είναι αλυσίδες.
2. Για τη μεταβλητή `delivered_by_efood` η πλειονότητα των παραγγελιών δεν παραδόθηκε από την ηλεκτρονική πλατφόρμα πωλήσεων.
3. Για τη μεταβλητή `is_online_payment` ο αριθμός των παραγγελιών που πληρώθηκαν online είναι μικρότερος από τον αριθμό των παραγγελιών που δεν πληρώθηκαν online.
4. Για τη μεταβλητή `order_has_discount_coupon` η πλειονότητα των παραγγελιών δεν είχε εκπτωτικό κουπόνι.
5. Για τη μεταβλητή `order_from_pinata` η πλειονότητα των παραγγελιών δεν έγινε με τη χρήση της Pinata.

Περίληψη

Τα βασικά συμπεράσματα είναι τα ακόλουθα:

1. Οι περισσότερες παραγγελίες έχουν γίνει σε καταστήματα που δεν είναι αλυσίδες.
2. Πολύ μικρός αριθμός παραγγελιών παραδόθηκε από την ηλεκτρονική πλατφόρμα πωλήσεων.
3. Οι περισσότερες παραγγελίες εξοφλήθηκαν με αντί για online πληρωμή.
4. Σε ελάχιστες παραγγελίες χρησιμοποιήθηκαν εκπτωτικά κουπόνια.
5. Σε ελάχιστες παραγγελίες έγινε χρήση του Pinata.



Εικόνα 10 Συχνότητα μεταβλητών Boolean

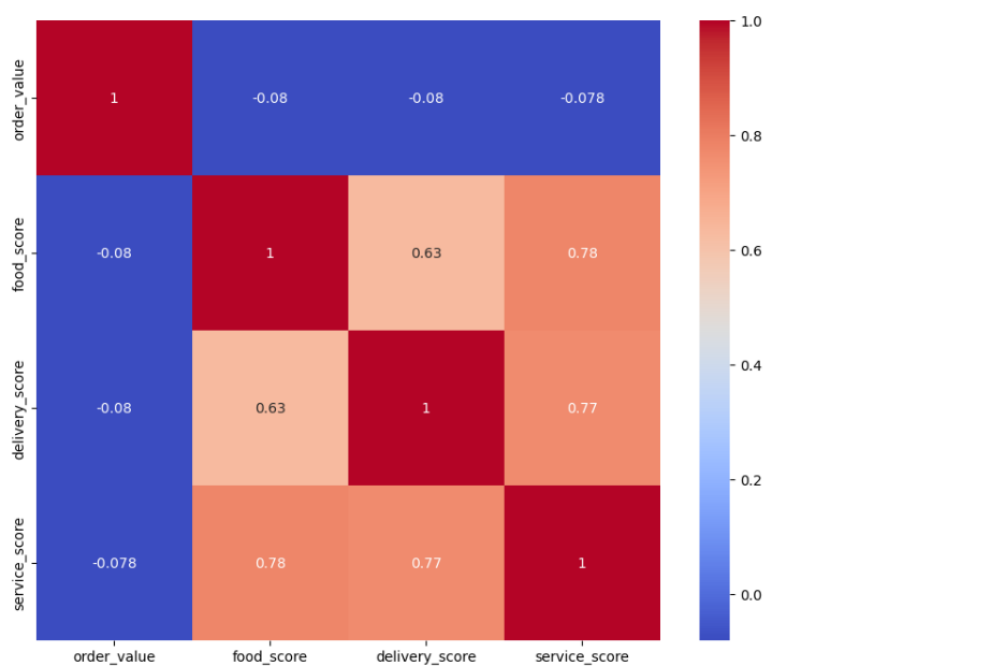
Ο χάρτης θερμότητας (heat map) με τη χρήση χρωμάτων απεικονίζει τον βαθμό συσχέτισης μεταξύ των μεταβλητών. Η κλίμακα στα δεξιά δείχνει το επίπεδο συσχέτισης, όπου το κόκκινο χρώμα αντιπροσωπεύει υψηλή θετική συσχέτιση (κοντά στο 1) και το μπλε υψηλή αρνητική συσχέτιση (κοντά στο -1).

Η μεταβλητή `order_value` δεν παρουσιάζει ισχυρή συσχέτιση με καμία από τις υπόλοιπες μεταβλητές. Συνεπώς, οι μεταβλητές `food_score`, `delivery_score`, και `service_score` δεν σχετίζονται ιδιαίτερα με την αξία της παραγγελίας.

Η μεταβλητή `food_score` έχει ισχυρή θετική συσχέτιση με τη μεταβλητή `service_score` (0.78) και τη μεταβλητή `delivery_score` (0.63). Συνεπώς, οι πιο υψηλές βαθμολογίες για το φαγητό είναι πιθανό να συνοδεύονται και από υψηλές βαθμολογίες που αφορούν στην εξυπηρέτηση και την παράδοση.

Η μεταβλητή `delivery_score` παρουσιάζει ισχυρή θετική συσχέτιση με τη μεταβλητή `service_score` (0.77) και με τη μεταβλητή `food_score` (0.63). Το γεγονός αυτό δείχνει ότι οι πελάτες που δίνουν υψηλές βαθμολογίες για την παράδοση τείνουν να δίνουν υψηλές βαθμολογίες και για την εξυπηρέτηση και το φαγητό.

Η μεταβλητή `service_score` εμφανίζει ισχυρή θετική συσχέτιση με τη μεταβλητή `food_score` (0.78) και τη μεταβλητή `delivery_score` (0.77). Άρα οι πελάτες που είναι ευχαριστημένοι με την εξυπηρέτηση του καταστήματος είναι επίσης ευχαριστημένοι με το φαγητό και την παράδοση.

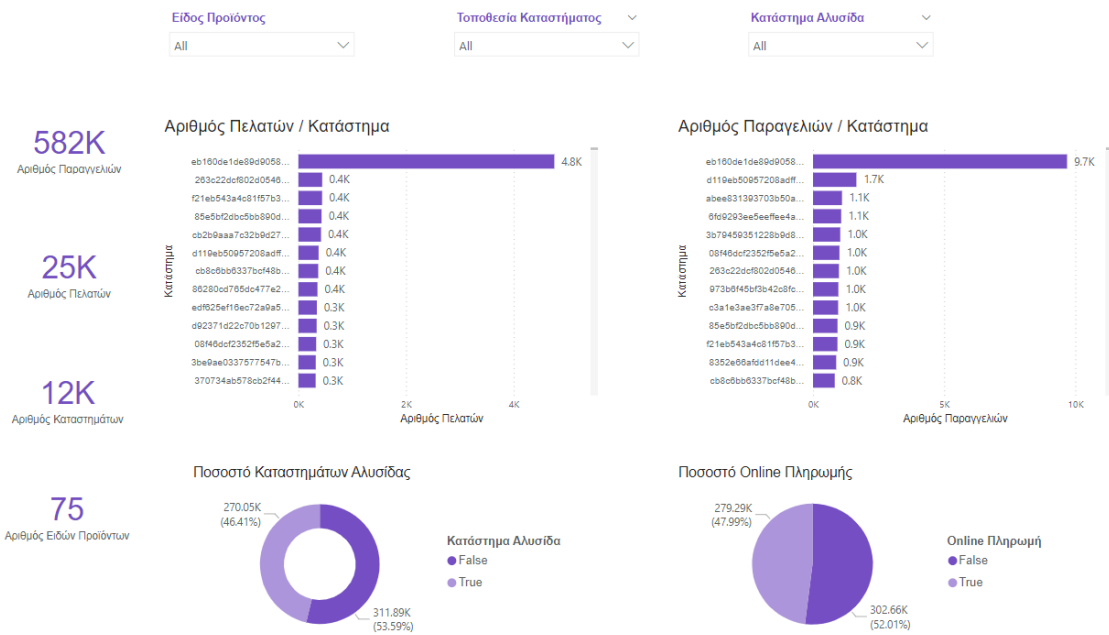


Εικόνα 11 Συσχέτιση μεταβλητών

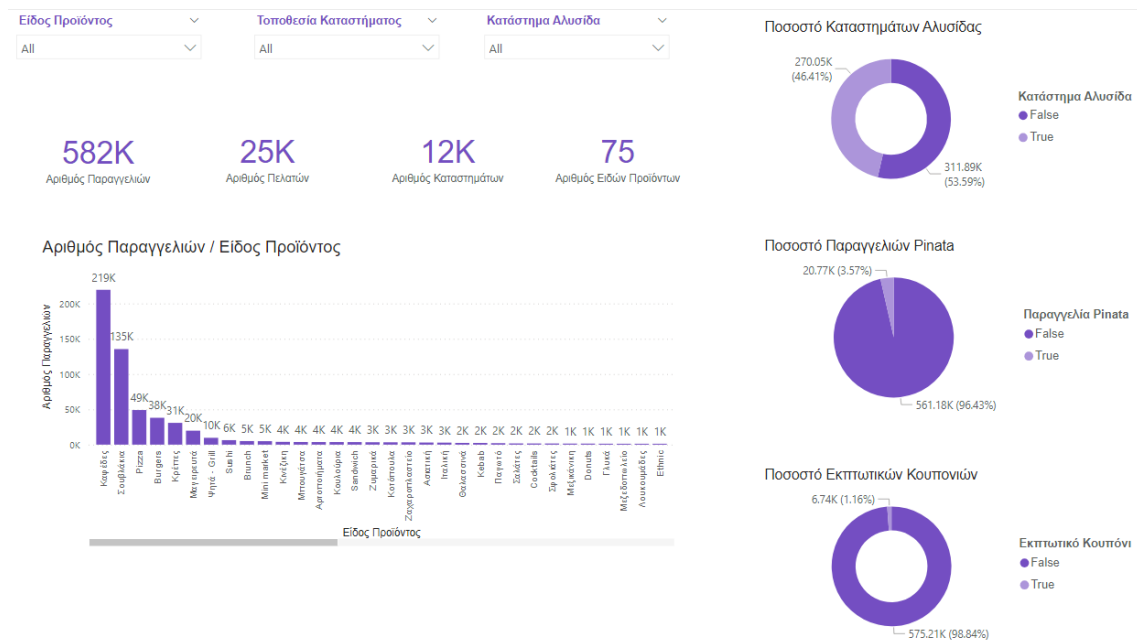
Στην Εικόνα 19 παρουσιάζεται αναφορά η οποία έχει γίνει με τη χρήση του λογισμικού Microsoft Power BI. Παρουσιάζεται ο συνολικός αριθμός παραγγελιών ο οποίος είναι περίπου 582.000, ο αριθμός πελατών που είναι περίπου 25.000, ο αριθμός καταστημάτων που είναι περίπου 12.000 και ο αριθμός ειδών προϊόντων που είναι 75. Επίσης, στο γράφημα Αριθμός Πελατών / Κατάστημα απεικονίζονται τα πρώτα 13 καταστήματα που έχουν τους περισσότερους πελάτες, όπου το πρώτο κατάστημα αριθμεί συνολικά στους 4.755 πελάτες. Στο γράφημα Αριθμός Παραγγελιών / Κατάστημα απεικονίζονται τα πρώτα 13 καταστήματα που έχουν τις περισσότερες παραγγελίες. Στο πρώτο κατάστημα έχουν γίνει συνολικά 9.684 παραγγελίες.

Στην Εικόνα 20 παρουσιάζεται αναφορά η οποία έχει γίνει επίσης με τη χρήση του λογισμικού Microsoft Power BI. Στο γράφημα Αριθμός Παραγγελιών / Είδος Προϊόντος απεικονίζονται τα προϊόντα που έχουν τη μεγαλύτερη ζήτηση. Τις πρώτες θέσεις κατέχουν οι καφέδες, τα σουβλάκια, οι πίτσες και τα burgers.

Επιπλέον, υπάρχουν γραφήματα τα οποία παρουσιάζουν το ποσοστό των παραγγελιών που πληρώθηκαν online ή με μετρητά, το ποσοστό των καταστημάτων που ανήκουν σε αλυσίδα, το ποσοστό των παραγγελιών που είχαν κάποιο κουπόνι και το ποσοστό των παραγγελιών που έγιναν με τη χρήση Pinata.



Εικόνα 12 Αριθμός Πελατών / Κατάστημα και Αριθμός Παραγγελιών / Κατάστημα



Εικόνα 13 Αριθμός Παραγγελιών / Είδος Προϊόντος

7. Τμηματοποίηση Πελατών

Για την τμηματοποίηση πελατών της ηλεκτρονικής πλατφόρμας πωλήσεων χρησιμοποιήθηκε η μεθοδολογία ανάλυσης RFM (Recency, Frequency, Monetary), μέσω της οποίας οι πελάτες διαχωρίζονται σε ομάδες ανάλογα με τη συμπεριφορά που έχουν δείξει στο παρελθόν. Η ανάλυση RFM καθορίζεται από τρεις διαστάσεις. Η πρώτη διάσταση είναι το Recency (R) που υπολογίζει την πιο πρόσφατη αγορά του πελάτη, η δεύτερη είναι το Frequency (F) στην οποία γίνεται μέτρηση της συχνότητας των αγορών του πελάτη και η τρίτη είναι το Monetary (M) στην οποία υπολογίζεται η χρηματική αξία των αγορών. Με τη χρήση της RFM ανάλυσης εντοπίζονται οι πελάτες που παρουσιάζουν αυξημένη πιθανότητα να κάνουν κι άλλες αγορές, καθώς έχουν προμηθευτεί πρόσφατα προϊόντα ή υπηρεσίες από μία επιχείρηση, κάνουν συχνές αγορές ή έχουν ξοδέψει συγκεκριμένη αξία χρήματων για αγορές (Lamb et al. 2011).

Η αξία που παρέχει η ανάλυση RFM στις επιχειρήσεις είναι ευρέως διαδεδομένη, καθώς με τη χρήση της είναι εφικτό να ανιχνευτούν πελάτες που θα ανταποκριθούν καλύτερα σε καμπάνιες μάρκετινγκ, γεγονός το οποίο θα αυξήσει το συνολικό ποσοστό ανταπόκρισης. Οι πελάτες οι οποίοι έχουν ολοκληρώσει μια αγορά από μια επιχείρηση σχετικά πρόσφατα εμφανίζουν μεγαλύτερη πιθανότητα ανταπόκρισης σε επερχόμενη καμπάνια μάρκετινγκ, σε σχέση με πελάτες που έχουν κάνει κάποια αγορά στην ίδια επιχείρηση πριν από μεγάλο χρονικό διάστημα. Επιπρόσθετα, οι πελάτες που πραγματοποιούν συχνές αγορές είναι πιο σίγουρο ότι θα ανταποκριθούν σε οποιαδήποτε προωθητική ενέργεια σε αντίθεση με τους μη συχνούς αγοραστές και το ίδιο ισχύει και για τους πελάτες που δαπανούν μεγάλα ποσά στις αγορές τους.

Είναι βέβαιο ότι οι επιχειρήσεις πρέπει να παρακολουθούν τη συχνότητα, την ένταση και την επανάληψη των αγορών, γιατί όπως έχει ήδη αναφερθεί είναι από τις κύριες διαστάσεις που μπορούν να προσδιορίσουν τη σχέση του πελάτη με μια επιχείρηση. Μέσω της RFM ανάλυσης μπορούν να εντοπιστούν αλλαγές στις καταναλωτικές συνήθειες των πελατών μιας επιχείρησης, γεγονός το οποίο αν γίνει αντιληπτό έγκαιρα είναι εύκολο να δημιουργηθούν στρατηγικές μάρκετινγκ που θα συμβάλλουν στη διατήρηση της σχέσης πελάτη και επιχείρησης (Tsiptsis & Chorianopoulos, 2009).

Το μοντέλο RFM ανάλυσης που έχει προταθεί από τους Tsiptsis και Chorianopoulos (2009) είναι ότι γίνεται μετατροπή των μεταβλητών Recency, Frequency και Monetary σε συγκεκριμένες τιμές 1,2,3,4, 5. Συνεπώς, οι πελάτες λαμβάνουν βαθμολογία από τις τιμές που αναφέρθηκαν για την κάθε μία από τις τρεις διαστάσεις. Για την τμηματοποίηση των πελατών βάσει της RFM ανάλυσης ακολουθούνται συγκεκριμένα βήματα. Αρχικά, συλλέγονται οι αγορές του κάθε πελάτη και στη συνέχεια γίνεται ταξινόμηση αυτών ανάλογα με το πόσο πρόσφατη ήταν η τελευταία αγορά που πραγματοποίησαν. Οι πελάτες που έκαναν την πιο πρόσφατη αγορά βαθμολογούνται με τη μέγιστη τιμή που είναι το 5. Έπειτα, ταξινομούνται οι πελάτες που πραγματοποιούν πιο συχνά αγορές και βαθμολογούνται αντίστοιχα και τέλος γίνεται ταξινόμηση και βαθμολογία των πελατών λαμβάνοντας υπόψιν την αξία των αγορών που πραγματοποίησαν. Στο τέλος της συγκεκριμένης διαδικασίας ο κάθε πελάτης έχει λάβει για την κάθε διάσταση από μία βαθμολογία και ανάλογα με αυτή διαχωρίζεται στην αντίστοιχη ομάδα. Με τον τρόπο αυτό οι επιχειρήσεις διαμορφώνουν κατάλληλες και εξατομικευμένες στρατηγικές και καμπάνιες μάρκετινγκ για την κάθε ομάδα.

Για την τμηματοποίηση πελατών της ηλεκτρονικής πλατφόρμας πωλήσεων χρησιμοποιήθηκαν δεδομένα που αφορούν στην περίοδο δύο μηνών από 01/04/2021 έως 31/05/2021. Για τη δημιουργία των διάφορων ομάδων πελατών χρησιμοποιήθηκαν τεταρτημόρια, δηλαδή για την κάθε διάσταση εκχώρηση βαθμολογίας από 1 έως 4, όπου το 4 είναι η καλύτερη τιμή και αντίστοιχα το 1 είναι η χειρότερη. Η τελική βαθμολογία για τον διαχωρισμό των πελατών σε ομάδες θα προκύψει από τον συνδυασμό της βαθμολογίας των τριών διαστάσεων. Αξίζει να αναφερθεί ότι τα πεντημόρια, δηλαδή βαθμολογία από 1 έως 5 παρέχουν καλύτερο βαθμό ανάλυσης, όμως λόγω της μικρής έκτασης των δεδομένων είναι πιο δύσκολο να δημιουργηθούν ομάδες βάσει αυτής της βαθμολογίας και για τον λόγο αυτό έγινε χρήση τεταρτημορίων.

7.1 Διάσταση Recency

Για τον υπολογισμό της διάστασης Recency επιλέχθηκε ως σημείο αναφοράς η ημερομηνία 31/05/2021, έτσι ώστε να εντοπιστεί πόσες ημέρες πριν ήταν η τελευταία αγορά που έκανε ο πελάτης.


```
In [10]: #last date available in the dataset
data['order_timestamp'].max()

Out[10]: Timestamp('2021-05-31 23:59:51+0000', tz='UTC')
```

```
In [11]: now = dt.date(2021,5,31)
print(now)

2021-05-31
```

Εικόνα 14 Recency σημείο αναφοράς τελευταίας ημερομηνίας

Επιπλέον, στο σύνολο των δεδομένων δημιουργήθηκε μία επιπλέον στήλη στην οποία καταγράφεται μόνο η ημερομηνία της παραγγελίας.

```
In [13]: #create a new column called date which contains the date of order timestamp only
data['date'] = pd.DatetimeIndex(data['order_timestamp']).date

In [14]: data.head()

Out[14]:
```

from_pinata	geographical_region_of_user	shop_city	geographical_region_of_shop	food_score	delivery_score	service_score	last_order_at	first_order_at	date
True	Αθήνα	Αθήνα	Χαλάνδρι-Πολύδροσο	NaN	NaN	NaN	2021-05-29	2017-06-13	2021-04-24
False	Αθήνα	Αθήνα	Αθήνα-Κολωνός	NaN	NaN	NaN	2021-05-23	2021-04-02	2021-04-24
False	Αθήνα	Αθήνα	Ελληνικό	NaN	NaN	NaN	2021-05-29	2018-09-19	2021-04-18
False	Αθήνα	Αθήνα	Χαιδάρι-Σκαρामαγκάς	NaN	NaN	NaN	2021-05-28	2018-10-22	2021-04-19
False	Βόλος	Βόλος	Βόλος	NaN	NaN	NaN	2021-05-27	2017-04-24	2021-04-19

Εικόνα 15 Δημιουργία στήλης καταγραφής ημερομηνίας παραγγελίας

Το επόμενο βήμα είναι η δημιουργία ενός πίνακα στον οποίο για κάθε πελάτη θα υπάρχει η τελευταία ημερομηνία που πραγματοποίησε μια αγορά και θα υπολογίζεται η διάσταση Recency βάσει των ημερών τελευταίας παραγγελίας από το σημείο αναφοράς 31/05/2021.

```
In [18]: #calculate recency
recency_data['Recency'] = recency_data['order_timestamp'].apply(lambda x: (now - x).days)

In [20]: recency_data.head()

Out[20]:
```

	customer_id	order_timestamp	Recency
0	0001496e540173a40362759ed10b3ce2	2021-05-31	0
1	0003b7de618c1acfc67eac91b853b287	2021-05-28	3
2	00045e9d1c96d7e2353f886704e51c88	2021-05-19	12
3	000e223791e890e85938d71bee32cf5a	2021-05-31	0
4	0016652ea3c3a1176dd2385e89de9ff6	2021-05-30	1

Εικόνα 16 Υπολογισμός διάστασης Recency

7.2 Διάσταση Frequency

Για τη διάσταση Frequency δημιουργήθηκε μία νέα μεταβλητή στην οποία υπολογίστηκε ο αριθμός των παραγγελιών που πραγματοποίησε ο πελάτης στο διάστημα από 01/04/2021 έως 31/05/2021.

```
In [22]: #calculate frequency of purchases
frequency_data = data.groupby(by=['customer_id'], as_index=False)['order_id'].count()
frequency_data.columns = ['customer_id', 'Frequency']
frequency_data.head()

Out[22]:
```

	customer_id	Frequency
0	0001496e540173a40362759ed10b3ce2	34
1	0003b7de618c1acfc67eac91b853b287	40
2	00045e9d1c96d7e2353f886704e51c88	13
3	000e223791e890e85938d71bee32cf5a	49
4	0016652ea3c3a1176dd2385e89de9ff6	51

Εικόνα 17 Υπολογισμός διάστασης Frequency

7.3 Διάσταση Monetary

Για τη διάσταση Monetary δημιουργήθηκε μια νέα μεταβλητή στην οποία υπολογίστηκε η συνολική αξία των παραγγελιών του κάθε πελάτη.

```
In [30]: monetary_data = data.groupby(by='customer_id', as_index=False).agg({'order_value_x': 'sum'})
monetary_data.columns = ['customer_id', 'Monetary']
monetary_data.head()
```

```
Out[30]:
```

	customer_id	Monetary
0	0001496e540173a40362759ed10b3ce2	318.20
1	0003b7de618c1acfc67eac91b853b287	341.75
2	00045e9d1c96d7e2353f886704e51c88	43.80
3	000e223791e890e85938d71bee32cf5a	122.50
4	0016652ea3c3a1176dd2385e89de9ff6	494.96

Εικόνα 18 Υπολογισμός διάστασης Monetary

7.4 Δημιουργία Τεταρτημορίων

Τα τελικά αποτελέσματα του υπολογισμού των τριών διαστάσεων για τον κάθε πελάτη παρουσιάζονται στον ακόλουθο πίνακα. Αφορούν στα 5 πρώτα id πελατών.

```
In [36]: rfm_data.tail()
```

```
Out[36]:
```

	customer_id	Recency	Frequency	Monetary
	ffe41c4dcc53a7ffd316f22ddb97be7	2	12	163.35
	fff262fbc647360700cee69566e26989	2	16	162.95
	fff356794e97eaa07f63f4fbb4d0cd2	2	15	248.20
	fffac61b056a50e5cf7173b94c939df	1	20	122.20
	ffffac8ce4cdfb1e5bc28beb7d46f7d5	5	18	128.35

Εικόνα 19 Διαστάσεις RFM

Όπως έχει αναφερθεί για τη δημιουργία των διάφορων ομάδων πελατών χρησιμοποιήθηκαν τα τεταρτημόρια.

```
In [40]: quantiles = rfm_data.quantile(q=[0.25,0.5,0.75])
quantiles
```

```
Out[40]:
```

	Recency	Frequency	Monetary
0.25	0.0	13.0	115.45
0.50	2.0	19.0	168.70
0.75	4.0	28.0	254.90

Εικόνα 20 Quantiles

Στη στήλη Recency απεικονίζεται το πόσο πρόσφατα έχει πραγματοποιήσει ο πελάτης την τελευταία του αγορά από το σημείο αναφοράς. Η τιμή 0 βρίσκεται στο σημείο 0.25 το οποίο σημαίνει ότι το 25,00% των πελατών έχει πραγματοποιήσει κάποια αγορά πολύ πρόσφατα, το 50,00% λιγότερο πρόσφατα και το 75,00% έχει αρκετό καιρό να πραγματοποιήσει κάποια αγορά.

Στη στήλη Frequency απεικονίζεται το πόσο συχνά ο πελάτης κάνει αγορές. Στο σημείο 0.25 βρίσκεται η τιμή 13, το οποίο σημαίνει ότι το 25,00% των πελατών έχει πραγματοποιήσει τουλάχιστον 13 αγορές. Αντίστοιχα, το 50,00% των πελατών έχει κάνει τουλάχιστον 19 αγορές και το 75,00% των πελατών τουλάχιστον 28 αγορές.

Στη στήλη Monetary απεικονίζεται το συνολικό χρηματικό ποσό που έχει δαπανήσει σε αγορές στην ηλεκτρονική πλατφόρμα πωλήσεων ο πελάτης. Στο σημείο 0.25 βρίσκεται η τιμή 115.45, το οποίο σημαίνει

ότι το 25,00% των πελατών έχει δαπανήσει τουλάχιστον 115.45 €. Αντίστοιχα, το 50,00% των πελατών έχει δαπανήσει τουλάχιστον 168.70 € και το 75,00% των πελατών τουλάχιστον 254.90 €.

Στη συνέχεια, θα δημιουργηθούν δύο κλάσεις ομάδων βάσει των διαστάσεων Frequency και Monetary, καθώς η διάσταση Recency δεν έχει τόσο καλή απόδοση αφού το 75,00% των πελατών έχει καιρό να πραγματοποιήσει κάποια αγορά στην ηλεκτρονική πλατφόρμα πωλήσεων.

```
In [45]: rfm_segmentation.head()
```

```
Out[45]:
```

customer_id	Recency	Frequency	Monetary	R_Quartile	F_Quartile	M_Quartile
0001496e540173a40362759ed10b3ce2	0	34	318.20	4	4	4
0003b7de618c1acfc67eac91b853b287	3	40	341.75	2	4	4
00045e9d1c96d7e2353f886704e51c88	12	13	43.80	1	1	1
000e223791e890e85938d71bee32cf5a	0	49	122.50	4	4	2
0016652ea3c3a1176dd2385e89de9ff6	1	51	494.96	3	4	4

Εικόνα 21 Quartiles Διαστάσεων

Τέλος, για τη δημιουργία των ομάδων από τη στιγμή που έχει υπολογιστεί η βαθμολογία για τον κάθε πελάτη, συνδυάζονται οι βαθμολογίες των τριών διαστάσεων για να υπολογιστεί η τελική RFM βαθμολογία του κάθε πελάτη αντίστοιχα. Η καλύτερη βαθμολογία για τη διάσταση Recency είναι 4 και αφορά στους πελάτες που έχουν κάνει πιο πρόσφατα παραγγελία από την ημερομηνία στόχο που έχει οριστεί. Η καλύτερη βαθμολογία για τη διάσταση Frequency είναι 4 και αφορά στους πελάτες που κάνουν πιο συχνά παραγγελίες και η καλύτερη βαθμολογία για τη διάσταση Monetary είναι επίσης 4 και αφορά στους πελάτες που δαπανούν συνολικά αρκετά χρήματα στις παραγγελίες.

```
In [46]: rfm_segmentation['RFMScore'] = rfm_segmentation.R_Quartile.map(str) \
+ rfm_segmentation.F_Quartile.map(str) \
+ rfm_segmentation.M_Quartile.map(str)
```

```
rfm_segmentation.head()
```

```
Out[46]:
```

customer_id	Recency	Frequency	Monetary	R_Quartile	F_Quartile	M_Quartile	RFMScore
0001496e540173a40362759ed10b3ce2	0	34	318.20	4	4	4	444
0003b7de618c1acfc67eac91b853b287	3	40	341.75	2	4	4	244
00045e9d1c96d7e2353f886704e51c88	12	13	43.80	1	1	1	111
000e223791e890e85938d71bee32cf5a	0	49	122.50	4	4	2	442
0016652ea3c3a1176dd2385e89de9ff6	1	51	494.96	3	4	4	344

Εικόνα 22 RFM Βαθμολογία

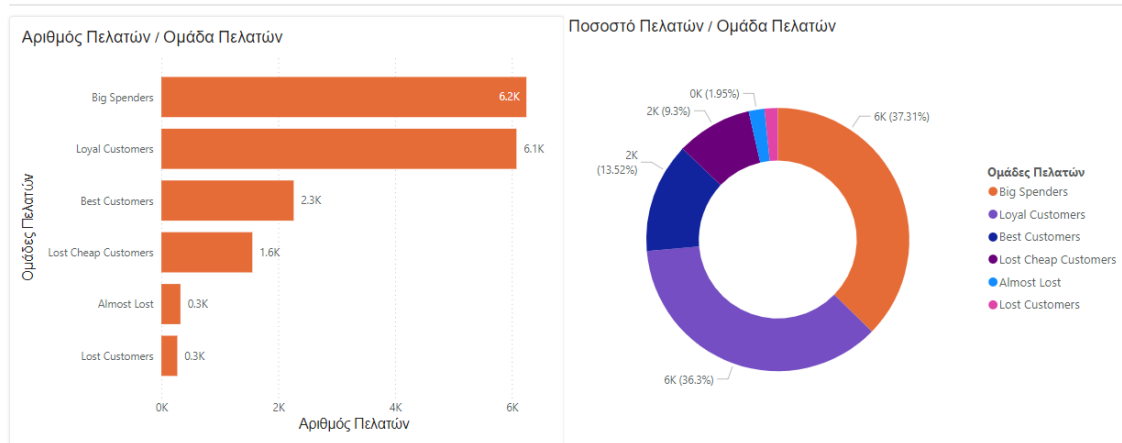
7.5 Αποτελέσματα Ανάλυσης Τμηματοποίησης Πελατών

Τα αποτελέσματα της ανάλυσης τμηματοποίησης πελατών παρουσιάζονται στην Εικόνα 30. Πιο συγκεκριμένα, στην ηλεκτρονική πλατφόρμα πωλήσεων οι πελάτες που έχουν συγκεντρώσει και στις τρεις διαστάσεις βαθμολογία 4 είναι 2.265 και αποτελούν του καλύτερους πελάτες. Οι πιο πιστοί πελάτες, δηλαδή αυτοί που κάνουν συχνά αγορές είναι 6.079 και αυτοί που έχουν ξοδέψει συνολικά τα περισσότερα χρήματα είναι 6.249. Οι πελάτες που έχουν κάνει συχνά παραγγελίες και έχουν ξοδέψει αρκετά χρήματα, όμως έχουν να κάνουν αρκετές ημέρες παραγγελία από την ημερομηνία στόχο κατατάσσονται σε αυτούς που σχεδόν έχουν εγκαταλείψει την ηλεκτρονική πλατφόρμα πωλήσεων. Τέλος, οι πελάτες που έχουν κάνει συχνά παραγγελίες και έχουν ξοδέψει αρκετά χρήματα, όμως έχουν να κάνουν ακόμη περισσότερες ημέρες παραγγελία από την ημερομηνία στόχο κατατάσσονται σε αυτούς που έχουν εγκαταλείψει την ηλεκτρονική πλατφόρμα πωλήσεων και αυτοί που έχουν συγκεντρώσει βαθμολογία 1 και στις τρεις διαστάσεις είναι αυτοί οι πελάτες που έχουν εγκαταλείψει την πλατφόρμα, δεν έκανα συχνά παραγγελίες και δεν έχουν ξοδέψει αρκετά χρήματα σε αυτές που είχαν κάνει.

```
In [48]: print("Best Customers: ", len(rfm_segmentation[rfm_segmentation['RFMScore']=='444']))
print("Loyal Customers: ", len(rfm_segmentation[rfm_segmentation['F_Quartile']==4]))
print("Big Spenders: ", len(rfm_segmentation[rfm_segmentation['M_Quartile']==4]))
print("Almost Lost: ", len(rfm_segmentation[rfm_segmentation['RFMScore']=='244']))
print("Lost Customers: ", len(rfm_segmentation[rfm_segmentation['RFMScore']=='144']))
print("Lost Cheap Customers: ", len(rfm_segmentation[rfm_segmentation['RFMScore']=='111']))

Best Customers: 2265
Loyal Customers: 6079
Big Spenders: 6249
Almost Lost: 326
Lost Customers: 272
Lost Cheap Customers: 1557
```

Εικόνα 23 Αποτελέσματα Ανάλυσης Τμηματοποίησης Πελατών



Εικόνα 24 Γράφημα Αποτελεσμάτων Ανάλυσης Τμηματοποίησης Πελατών

7.6 Προτάσεις Στρατηγικών και Καμπανιών Μάρκετινγκ

Με τα αποτελέσματα της συγκεκριμένης ανάλυσης και την ομαδοποίηση των πελατών στις διάφορες ομάδες (Best Customers, Loyal Customers, Big Spenders, Almost Lost, ;Lost Customers και Lost Cheap Customers) ακολουθούν συγκεκριμένες προτάσεις στρατηγικών και καμπανιών μάρκετινγκ για τη διατήρηση των υπάρχοντων πελάτων, την αύξηση της ικανοποίησής τους, αλλά και την επαναφορά των χαμένων πελατών.

Οι προτάσεις στρατηγικών για τη διατήρηση και την αύξηση της ικανοποίησης της ομάδας πελατών «Best Customers» είναι οι ακόλουθες:

1. Δημιουργία προγραμμάτων επιβράβευσης με αποκλειστικά οφέλη, όπως δωρεάν υπηρεσίες παράδοσης για τα καταστήματα που έχουν χρέωση της υπηρεσίας. Επιπρόσθετα, παροχή πόντων επιβράβευσης για κάθε παραγγελία. Οι πόντοι επιβράβευσης θα μπορούν να εξαργυρωθούν σε επόμενες παραγγελίες.
2. Παροχή εξατομικευμένων προσφορών για τα αγαπημένα καταστήματα του κάθε πελάτη, λαμβάνοντας υπόψιν τις προηγούμενες παραγγελίες που έχει κάνει.
3. Προσφορές που αφορούν σε νέα συνεργαζόμενα καταστήματα και νέες γεύσεις για την προσέλκυση τους να δοκιμάσουν κάτι καινούργιο.

Οι προτάσεις στρατηγικών για τη διατήρηση και την αύξηση της ικανοποίησης της ομάδας πελατών «Loyal Customers» είναι οι ακόλουθες:

1. Προγραμματισμός συχνής επικοινωνίας μέσω αποστολής ενημερώσεων για νέες προσφορές, συνεργασίες με καινούργια καταστήματα και εκπτώσεις καθώς και δημιουργία ενημερωτικών δελτίων στα οποία θα περιλαμβάνονται διάφορες προτάσεις.

2. Δημιουργία προγραμμάτων επιβράβευσης μέσω των οποίων θα παρέχονται εξατομικευμένες εκπτώσεις σε πελάτες που παραπέμπουν στην ηλεκτρονική πλατφόρμα πωλήσεων νέους πελάτες.
3. Διοργάνωση διαγωνισμών που θα λαμβάνουν βραβεία αξίας οι πελάτες με τις περισσότερες παραγγελίες.
4. Παροχή επιπρόσθετων πόντων για παραγγελίες που θα γίνονται κατά τη διάρκεια ειδικών ημερών.

Οι προτάσεις στρατηγικών για τη διατήρηση και την αύξηση της ικανοποίησης της ομάδας πελατών «Big Spenders» είναι οι ακόλουθες:

1. Παροχή εξατομικευμένων προσφορών για premium καταστήματα αλλά και προσωποποιημένων προτάσεων βάσει των προτιμήσεων των πελατών.
2. Προώθηση επιπλέον συμπληρωματικών προϊόντων που ταιριάζουν σε κάθε παραγγελία.
3. Προσφορά δωρεάν προϊόντων για μεγάλες παραγγελίες.

Οι προτάσεις στρατηγικών για την επαναφορά της ομάδας πελατών «Almost Lost» είναι οι ακόλουθες:

1. Αποστολή εκπτωτικών κουπονιών με μεγάλα ποσοστά έκπτωσης, έτσι ώστε να τους δοθεί κίνητρο να επιστρέψουν.
2. Προσφορά δωρεάν υπηρεσίας παράδοσης ή ενός προϊόντος για την επόμενη παραγγελία.
3. Αποστολή προσωποποιημένων μηνυμάτων για να τους υπενθυμίζονται τα οφέλη που παρέχει η πλατφόρμα αλλά και προσφορές για τα αγαπημένα τους καταστήματα.
4. Αποστολή ερευνών που θα έχουν ως στόχο την κατανόηση των λόγων απομάκρυνσης και τη συλλογή προτάσεων για βελτίωση των προϊόντων και υπηρεσιών.

Οι προτάσεις στρατηγικών για την επαναφορά της ομάδας πελατών «Lost Customers» είναι οι ακόλουθες:

1. Παροχή σημαντικών εκπτώσεων και ειδικών κουπονιών για να επιτευχθεί πιο εύκολα η προσέλκυση των πελατών που έχουν χαθεί.
2. Δημιουργία χρονικά περιορισμένων προσφορών για να έχει άμεσα αποτελέσματα η συγκεκριμένη ενέργεια.
3. Αποστολή ερευνών που θα έχουν ως στόχο την κατανόηση των λόγων απομάκρυνσης και τη συλλογή προτάσεων για βελτίωση των προϊόντων και υπηρεσιών, καθώς και επιβράβευση για τη συμμετοχή τους στις έρευνες με εκπτώσεις ή δωρεάν προϊόντα.
4. Αποστολή ενημερώσεων για τις νέες αφίσες καταστημάτων, καθώς και τις βελτιώσεις στις υπηρεσίες και τα προϊόντα.

Οι προτάσεις στρατηγικών για την επαναφορά της ομάδας πελατών «Lost Cheap Customers» είναι οι ακόλουθες:

1. Αποστολή προσφορών και κουπονιών με χαμηλές τιμές για να προσελκύσουν την επιστροφή τους.
2. Προώθηση οικονομικών προσφορών μέσω email ή κοινωνικών δικτύων.
3. Αποστολή ερευνών που θα έχουν ως στόχο την κατανόηση των λόγων απομάκρυνσης και τη συλλογή προτάσεων για βελτίωση των προϊόντων και υπηρεσιών καθώς και επιβράβευση για τη συμμετοχή τους στις έρευνες με εκπτώσεις ή πόντους.

8. Πρόβλεψη Πωλήσεων - Αποτελέσματα Δεδομένων Εκπαίδευσης και Ελέγχου

Ο στόχος της παρούσας ανάλυσης είναι η πρόβλεψη πωλήσεων μιας ηλεκτρονικής πλατφόρμας πωλήσεων, όπου για τη διεξαγωγή της πρόβλεψης χρησιμοποιήθηκαν δεδομένα δύο μηνών. Για την πρόβλεψη πωλήσεων αξιοποιήθηκε η μέθοδος Holt-Winters που όπως έχει ήδη αναφερθεί αποτελεί μια μέθοδο πρόβλεψης για δεδομένα που εμφανίζουν εποχικότητα και βασίζεται στις τάσεις των δεδομένων και την εποχική τους συμπεριφορά. Τα βήματα που ακολουθήθηκαν για τη συγκεκριμένη διαδικασία είναι τα ακόλουθα:

Η συλλογή δεδομένων όπου χρησιμοποιήθηκαν δεδομένα πωλήσεων για την περίοδο από 1 Απριλίου 2021 έως 31 Μαΐου 2021. Αυτά τα δεδομένα περιλαμβάνουν τις παραγγελίες για κάθε μέρα αυτής της περιόδου.

Η κατηγοριοποίηση των καταστημάτων, όπου ανάλογα με τον συνολικό αριθμό των παραγγελιών του κάθε καταστήματος, ομαδοποιήθηκαν σε τρεις κατηγορίες:

1. Big (Μεγάλα): Καταστήματα από 800 παραγγελίες και άνω στο διάστημα των δύο μηνών.
2. Medium (Μεσαία): Καταστήματα από 300 μέχρι 799 παραγγελίες στο διάστημα των δύο μηνών.
3. Small (Μικρά): Καταστήματα μέχρι 299 παραγγελίες στο διάστημα των δύο μηνών

```
In [4]: # Calculate the count of each shop_id
category_counts = data['shop_id'].value_counts()

# Define a function to determine the size based on the count
def determine_size(shop_id):
    count = category_counts[shop_id]
    if count > 800:
        return 'Big'
    elif count <= 299:
        return 'Small'
    else:
        return 'Medium'

# Apply the function to create a new 'size' column
data['size'] = data['shop_id'].apply(determine_size)
```

```
In [5]: data.head()
```

```
Out[5]:
```

nt	...	geographical_region_of_user	shop_city	geographical_region_of_shop	food_score	delivery_score	service_score	last_order_at	first_order_at	date	size	
se	...		Αθήνα	Αθήνα	Χαλάνδρι-Πολύδροσο	NaN	NaN	NaN	2021-05-29	2017-06-13	2021-04-24	Small
se	...		Αθήνα	Αθήνα	Αθήνα-Κολωνός	NaN	NaN	NaN	2021-05-23	2021-04-02	2021-04-24	Small
se	...		Αθήνα	Αθήνα	Ελληνικό	NaN	NaN	NaN	2021-05-29	2018-09-19	2021-04-18	Big
je	...		Αθήνα	Αθήνα	Χαϊδάρι-Σκαρामαγκάς	NaN	NaN	NaN	2021-05-28	2018-10-22	2021-04-19	Small
se	...		Βόλος	Βόλος	Βόλος	NaN	NaN	NaN	2021-05-27	2017-04-24	2021-04-19	Small

Εικόνα 25 Δημιουργία στήλης κατηγοριοποίησης καταστημάτων

Η συγκεκριμένη κατηγοριοποίηση συμβάλλει στην εστίαση της πρόβλεψης στα καταστήματα με τον μεγαλύτερο όγκο πωλήσεων, καθώς όσο περισσότερα δεδομένα είναι διαθέσιμα τόσο πιο αξιόπιστη θα είναι η πρόβλεψη.

Η εκπαίδευση του μοντέλου, όπου έγινε διαχωρισμός των δεδομένων σε training και test δεδομένα. Με τα training δεδομένα επιτυγχάνεται η εκπαίδευση του μοντέλου πρόβλεψης. Το μοντέλο από τα training δεδομένα αποκτά γνώση για τις σχέσεις μεταξύ των μεταβλητών (π.χ. ιστορικό πωλήσεων, εποχιακές μεταβολές) και τη μεταβλητή στόχου (πωλήσεις). Τα test δεδομένα χρησιμοποιούνται για την αξιολόγηση της απόδοσης του μοντέλου και δεν αξιοποιούνται στη διαδικασία της εκπαίδευσης. Τα training δεδομένα είναι μέχρι τις 18/05/2021 και τα test δεδομένα είναι από 19/05/2021 μέχρι και 31/05/2021.

Η πρόβλεψη πωλήσεων, όπου μετά το πέρας της εκπαίδευσης του μοντέλου, προβλέφθηκαν οι πωλήσεις για διάστημα επτά ημερών και συγκεκριμένα από 01 Ιουνίου μέχρι και 07 Ιουνίου 2021. Η πρόβλεψη έγινε για τα 13 καταστήματα τα οποία ανήκουν στην κατηγορία Big, καθώς αυτά παρουσιάζουν τον μεγαλύτερο όγκο πωλήσεων.

Στον ακόλουθο πίνακα παρουσιάζεται βάσει της κατηγοριοποίησης ο αριθμός των καταστημάτων στην κάθε κατηγορία.

Κατηγοριοποίηση Καταστήματος	Αριθμός Καταστημάτων
Big	13
Medium	182
Small	12112

Πίνακας 1 Αριθμός καταστημάτων σε κάθε κατηγορία

Στον ακόλουθο πίνακα παρουσιάζονται κατηγοριοποιημένα τα πρώτα 15 καταστήματα με τον αριθμό παραγγελιών τους.

Id Καταστήματος	Αριθμός Παραγγελιών	Μέγεθος Καταστήματος
eb160de1de89d9058fcb0b968dbbbd68	9684	Big
d119eb50957208adffa28742636b3600	1652	Big
abee831393703b50a6699790083a577f	1103	Big
6fd9293ee5effee4ab1fd29713b84a4	1072	Big
3b79459351228b9d823b0ec9e40cf952	1036	Big
08f46dcf2352f5e5a272957ecb971b7e	1006	Big
263c22dcf802d054642fa5aa32da24e0	986	Big
973b6f45bf3b42c8fc7cf58040109868	964	Big
c3a1e3ae3f7a8e705c86d8e4ad337b39	956	Big
85e5bf2dbc5bb890d23f2ab86c92648a	932	Big
f21eb543a4c81f57b31316fd2a3b89cd	930	Big
8352e66afdd11dee472475518ecafaa1	882	Big
cb8c6bb6337bcf48b6e64f9903b4aae2	826	Big
eda2d9d70af0335292faf1dbfd55afcf	788	Medium
b07658682176644cf165373c6785eb60	771	Medium

Πίνακας 2 Κατηγοριοποίηση Καταστημάτων

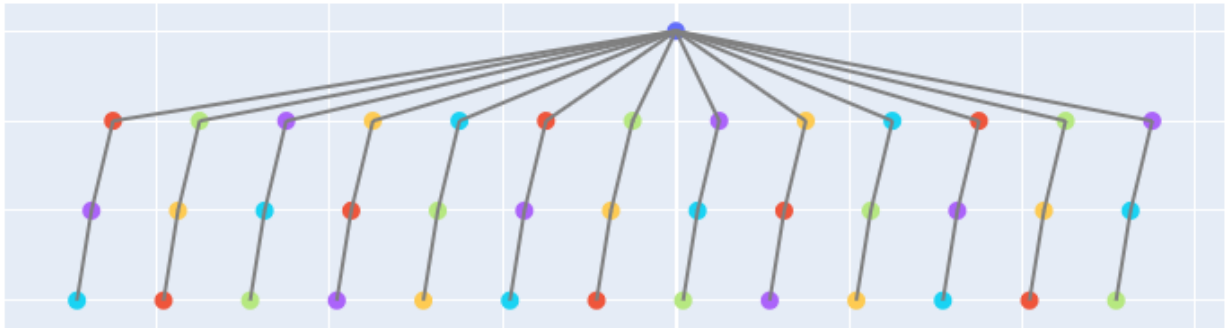
Για την πρόβλεψη πωλήσεων έγινε χρήση της ιεραρχικής μεθόδου μέσω της οποίας επιτυγχάνεται η ανάλυση και η πρόβλεψη δεδομένων σε πολλαπλά επίπεδα ανάλυσης, όπως γεωγραφικά επίπεδα, επίπεδα προϊόντων ή τμήματα αγοράς. Με τη χρήση της ιεραρχικής μεθόδου διευκολύνεται η διάσπαση του όγκου των δεδομένων σε μικρότερα υποσύνολα, στα οποία είναι εφικτό να γίνει ξεχωριστή ανάλυση και πρόβλεψη πριν διενεργηθεί η συνολική πρόβλεψη. Η συγκεκριμένη μέθοδος συμβάλει στην κατανόηση της συμπεριφοράς των δεδομένων στα διάφορα επίπεδα και διασφαλίζει ότι οι προβλέψεις είναι ακριβείς και συνεπείς.

Η ιεραρχική δομή που ακολουθήθηκε είναι:

1. πρώτο επίπεδο η γεωγραφική περιοχή των καταστημάτων,
2. δεύτερο επίπεδο η πόλη των καταστημάτων και
3. τρίτο επίπεδο τα καταστήματα.

```
[11]: spec = [['geographical_region_of_shop'],
           ['geographical_region_of_shop', 'shop_city'],
           ['geographical_region_of_shop', 'shop_city', 'shop_id']]
```

Εικόνα 26 Ιεραρχική Δομή



Εικόνα 27 Απεικόνιση Ιεραρχικής Δομής

Τα αποτελέσματα των δεδομένων εκπαίδευσης και ελέγχου με τη χρήση του μοντέλου Holt Winters παρουσιάζονται και για τις τρεις ιεραρχίες στις εικόνες που ακολουθούν. Το πρώτο επίπεδο της ιεραρχίας είναι η γεωγραφική περιοχή των καταστημάτων της ηλεκτρονικής πλατφόρμας πωλήσεων, το δεύτερο επίπεδο είναι η πόλη των καταστημάτων και το τρίτο επίπεδο είναι τα καταστήματα.

Στα δεδομένα εκπαίδευσης η μπλε γραμμή απεικονίζει τον πραγματικό αριθμό παραγγελιών στο διάστημα από 1 Απριλίου 2021 έως 18 Μαΐου 2021. Σε αυτά τα δεδομένα εμφανίζονται αρκετές διακυμάνσεις στον αριθμό παραγγελιών. Το γεγονός αυτό μπορεί να οφείλεται σε διάφορους παράγοντες, όπως η εποχικότητα ή άλλα γεγονότα. Η πορτοκαλί γραμμή απεικονίζει τις προβλέψεις για τα δεδομένα εκπαίδευσης. Είναι φανερό ότι οι προβλέψεις έχουν ακολουθήσει σε ικανοποιητικό βαθμό τις τάσεις των πραγματικών δεδομένων. Συνεπώς, το μοντέλο, έχει τη δυνατότητα να αντιληφθεί τις εποχικές και τυχαίες συνιστώσες της ζήτησης σε μεγάλο βαθμό.

Στα δεδομένα ελέγχου η κόκκινη γραμμή απεικονίζει τον πραγματικό αριθμό παραγγελιών στο διάστημα από 19 Μαΐου 2021 έως 31 Μαΐου 2021. Τη συγκεκριμένη περίοδο παρατηρείται πιο μεγάλη μεταβλητότητα σε σχέση με τις προβλέψεις, γεγονός που είναι πιθανό να συμβαίνει καθώς μπορεί να υπάρχουν εξωτερικοί παράγοντες που δεν έχουν ληφθεί υπόψιν. Η πράσινη γραμμή απεικονίζει τις προβλέψεις για τα δεδομένα εκπαίδευσης. Εμφανίζονται μικρότερες διακυμάνσεις στα δεδομένα και ο αριθμός των παραγγελιών είναι μικρότερος από αυτόν των πραγματικών δεδομένων.

Τα γενικά συμπεράσματα είναι ότι το μοντέλο Holt Winters έχει συλλάβει σε ικανοποιητικό βαθμό τις τάσεις και τις εποχικές συνιστώσες κατά την περίοδο εκπαίδευσης. Κατά την περίοδο ελέγχου οι προβλέψεις του αριθμού παραγγελιών εμφανίζουν κάποιες αποκλίσεις από τον πραγματικό αριθμό παραγγελιών. Οι αποκλίσεις αυτές είναι πιθανό να οφείλονται στην ανάγκη για ενσωμάτωση μεγαλύτερου όγκου δεδομένων στο υπάρχον σύνολο δεδομένων που εξετάζεται.

8.1 1^ο Επίπεδο Ιεραρχίας Γεωγραφική Περιοχή – Προβλεπόμενες Παραγγελίες

Στον ακόλουθο πίνακα παρουσιάζεται ο αριθμός των προβλεπόμενων παραγγελιών για την 1^ο επίπεδο ιεραρχίας (γεωγραφική περιοχή), που αφορά στα δεδομένα εκπαίδευσης.

	Θεσσαλονίκη-Χαριλάου-Μαρτίου-Βουλγάρη	Νέα Σμύρνη	Άνο Γλυφάδα	Αθήνα-Κάτω Πατήσια	Πάτρα 2	Αθήνα-Γίσιμα-Μαβίλη	Ζωγράφου	Θεσσαλονίκη-Κέντρο Αυτικά-Λικαστήρια-Παλιός Σταθμός-Αμπελόκηποι-Ξηροκρήνη-Μενεμένη	Αθήνα-Θησείο-Ψυρρή-Μεταξουργείο	Καλλιθέα-Ανατολικά	Ηράκλειο Κρήτης 2	Ηράκλειο Κρήτης 2	Αθήνα-Νεάπολη-Λυκαβηττός
2021-04-01	22	14	15	17	16	18	17	19	20	13	36	169	12
2021-04-02	19	16	20	20	16	18	20	16	18	15	32	180	20
2021-04-03	20	13	16	19	13	17	15	18	16	12	29	170	14
2021-04-04	22	15	16	19	15	16	16	21	16	15	35	189	14
2021-04-05	21	13	19	21	15	19	16	20	20	13	24	195	14
2021-04-06	22	17	21	22	15	17	20	19	20	13	32	204	17
2021-04-07	18	18	19	19	17	13	15	19	22	13	26	141	12
2021-04-08	20	17	18	21	17	16	16	20	20	12	33	173	13
2021-04-09	23	17	18	20	16	16	17	21	17	15	33	178	17
2021-04-10	24	17	14	18	20	20	16	23	16	10	33	153	18
2021-04-11	21	18	19	20	22	15	19	19	18	12	32	173	14
2021-04-12	21	14	13	19	19	12	14	18	17	14	24	155	19
2021-04-13	19	15	15	16	18	17	17	18	19	13	33	161	12
2021-04-14	16	17	20	19	18	18	20	15	17	15	30	171	20
2021-04-15	17	13	16	18	16	17	14	17	15	13	26	162	14

Μεταπτυχιακή Διατριβή

Αλεξάνδρα Παπανίκου

	Θεσσαλονίκη- Χαριλάου-Μαρτίου- Βουλγάρη	Νέα Σμύρνη	Άνο Γλυφά δα	Αθήνα- Κάτο Πατήσια	Πάτ ρα 2	Αθήνα- Γίσιμα- Μαβίλη	Ζωγράφου	Θεσσαλονίκη-Κέντρο Σταθμός-Αμπελόκηποι-Ξηροκρήνη-Μενεμένη	Αθήνα-Θησείο- Ψυρρή- Μεταξουργείο	Καλλιθέα- Ανατολικά	Ηράκλειο Κρήτης 2	Ηράκλειο Κρήτης 2	Αθήνα- Νεάπολη- Λυκαβηττός
2021-04-16	20	15	16	18	19	16	15	20	15	16	33	180	14
2021-04-17	18	13	19	20	19	18	15	19	19	14	22	185	14
2021-04-18	19	17	21	21	18	16	19	18	19	13	30	196	17
2021-04-19	15	19	19	18	18	12	15	18	21	13	24	135	12
2021-04-20	17	18	18	20	15	15	16	19	19	12	31	166	13
2021-04-21	20	17	18	19	13	16	16	20	16	15	30	172	17
2021-04-22	21	17	14	17	15	20	15	22	15	10	30	147	18
2021-04-23	18	19	19	19	12	14	19	18	17	12	29	166	14
2021-04-24	18	15	13	18	16	11	14	17	16	15	21	147	19
2021-04-25	16	15	14	16	17	17	16	17	19	14	31	154	13
2021-04-26	13	17	20	18	14	17	19	14	17	16	28	165	20
2021-04-27	15	13	15	18	10	16	13	16	14	14	24	155	14
2021-04-28	17	16	16	17	12	15	15	19	15	16	31	175	15
2021-04-29	15	13	19	19	11	18	14	18	18	14	19	180	14
2021-04-30	16	17	21	21	10	15	18	17	18	14	28	191	17
2021-05-01	13	19	19	17	9	11	14	17	20	14	22	133	12
2021-05-03	14	18	18	19	6	15	15	18	18	13	28	164	13

	Θεσσαλονίκη- Χαριλάου-Μαρτίου- Βουλγάρη	Νέα Σμύρνη	Άνο Γλυφά δα	Αθήνα- Κάτο Πατήσια	Πάτ ρα 2	Αθήνα- Γίλσια- Μαβύλη	Ζωγρ άφου	Θεσσαλονίκη-Κέντρο Σταθμός-Αμπελόκηποι-Ξηροκρήνη-Μενεμένη	Αθήνα-Θησείο- Ψυρρή- Μεταξουργείο	Καλλιθέα- Ανατολικά	Ηράκλειο Κρήτης 2	Ηράκλειο Κρήτης 2	Αθήνα- Νεάπολη- Λυκαβηττός
2021- 05-04	18	17	17	18	7	15	16	18	15	16	28	170	17
2021- 05-05	18	17	14	16	10	19	15	21	15	11	28	145	18
2021- 05-06	15	19	19	18	8	13	18	17	16	13	27	165	14
2021- 05-07	15	15	13	17	8	11	13	16	15	16	19	145	19
2021- 05-08	14	16	14	15	8	16	15	16	18	15	28	150	13
2021- 05-09	11	18	20	17	8	16	18	13	16	16	25	161	20
2021- 05-10	12	14	15	17	8	15	13	15	13	14	21	152	14
2021- 05-11	14	16	16	16	12	14	14	18	14	17	28	171	15
2021- 05-12	13	14	19	18	13	17	14	17	17	15	17	176	14
2021- 05-13	13	18	21	20	13	15	18	16	17	15	25	187	17
2021- 05-14	10	19	19	16	14	11	13	16	19	14	19	123	12
2021- 05-15	11	19	18	18	9	14	15	17	17	14	26	155	13
2021- 05-16	15	18	17	17	9	14	15	17	14	16	25	160	17
2021- 05-17	16	18	14	15	14	18	14	20	14	12	26	136	18
2021- 05-18	13	19	19	NaN	12	13	17	16	15	14	25	155	14
2021- 05-19	12	16	13	NaN	15	10	13	15	NaN	16	17	135	19

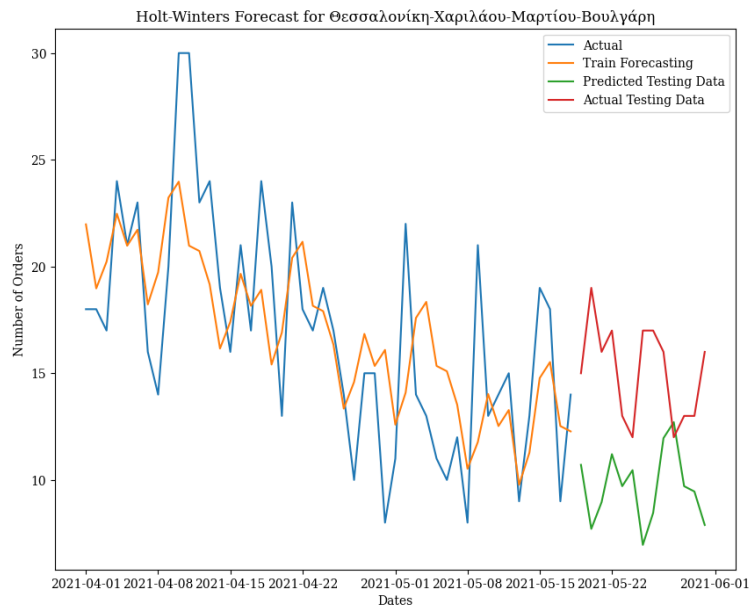
Πίνακας 3 Προβλεπόμενες Παραγγελίες Δεδομένων Εκπαίδευσης 1^ο Επίπεδο Ιεραρχίας

Στον ακόλουθο πίνακα παρουσιάζεται ο αριθμός των προβλεπόμενων παραγγελιών για το 1^ο επίπεδο ιεραρχίας (γεωγραφική περιοχή), που αφορά στα δεδομένα ελέγχου.

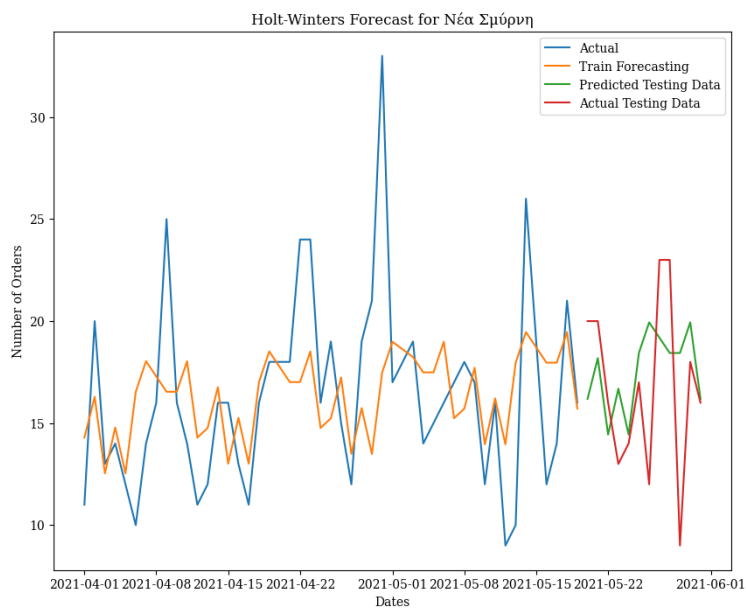
Τμηματοποίηση πελατών και πρόβλεψη πωλήσεων
για μια ηλεκτρονική πλατφόρμα πωλήσεων

	Θεσσαλονίκη- Χαριλάου-Μαρτίου- Βουλγάρη	Νέα Σμύρνη	Άνω Γλυφά δα	Αθήνα- Κάτω Πατήσια	Πάτ- ρα 2	Αθήνα- Γίσιμα- Μαβίλη	Ζωγράφου	Θεσσαλονίκη-Κέντρο Αυτικά-Λικαστήρια-Παλιός Σταθμός-Αμπελόκηποι-Ξηροκρήνη-Μενεμένη	Αθήνα-Θησείο- Ψυρρή- Μεταξουργείο	Καλλιθέα- Ανατολικά	Ηράκλειο Κρήτης 2	Μελίσιμα -Πεντέλη	Αθήνα- Νεάπολη- Λυκαβηττός
2021- 05-20	11	16	14	17	16	15	15	15	15	15	26	141	13
2021- 05-21	8	18	20	16	15	15	18	12	17	17	23	152	20
2021- 05-22	9	14	15	14	13	14	12	14	15	15	19	142	14
2021- 05-23	11	17	16	16	15	13	13	17	12	18	26	161	15
2021- 05-24	10	14	19	16	15	16	13	16	13	16	15	167	14
2021- 05-25	10	18	21	15	15	14	17	15	16	15	23	177	17
2021- 05-26	7	20	19	17	15	10	13	15	16	15	17	116	12
2021- 05-27	8	19	18	19	13	13	14	16	19	14	24	147	13
2021- 05-28	12	18	17	15	12	13	14	16	17	17	23	152	17
2021- 05-29	13	18	14	17	16	17	13	19	14	12	23	128	18
2021- 05-30	10	20	19	16	14	12	17	15	13	14	22	147	14
2021- 05-31	9	16	13	14	16	9	12	14	15	17	14	128	19

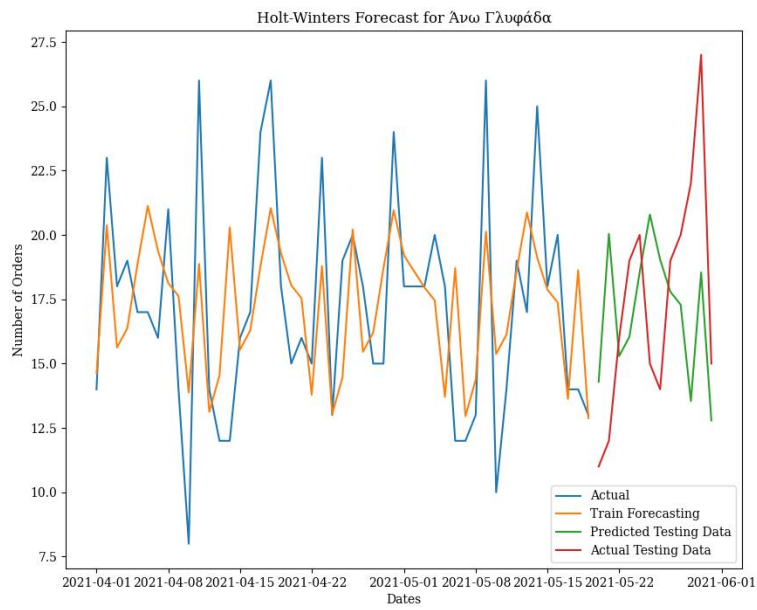
Πίνακας 4 Προβλεπόμενες Παραγγελίες Δεδομένων Ελέγχου 1^ο Επίπεδο Ιεραρχίας



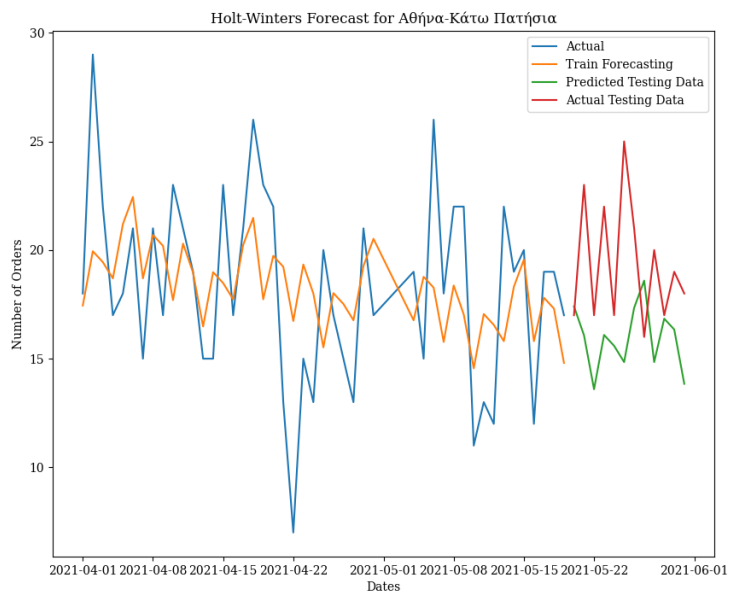
Εικόνα 28 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (1)



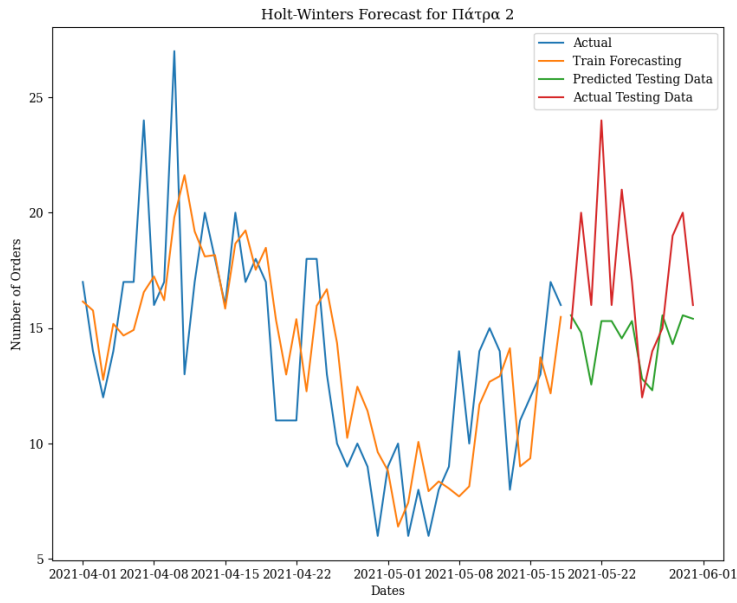
Εικόνα 29 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (2)



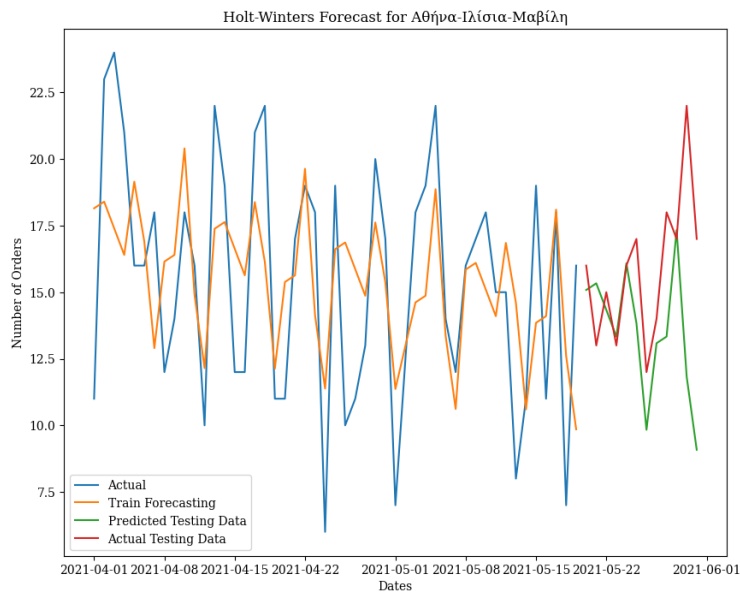
Εικόνα 30 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (3)



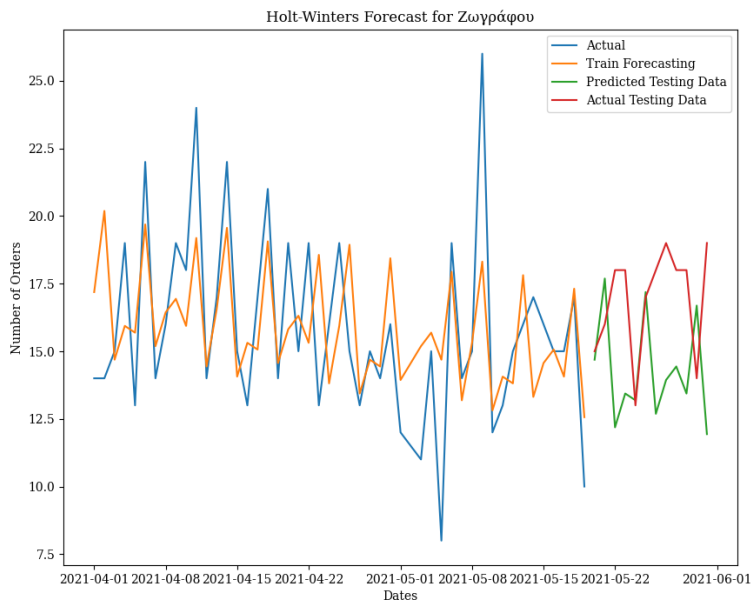
Εικόνα 31 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (4)



Εικόνα 32 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (5)

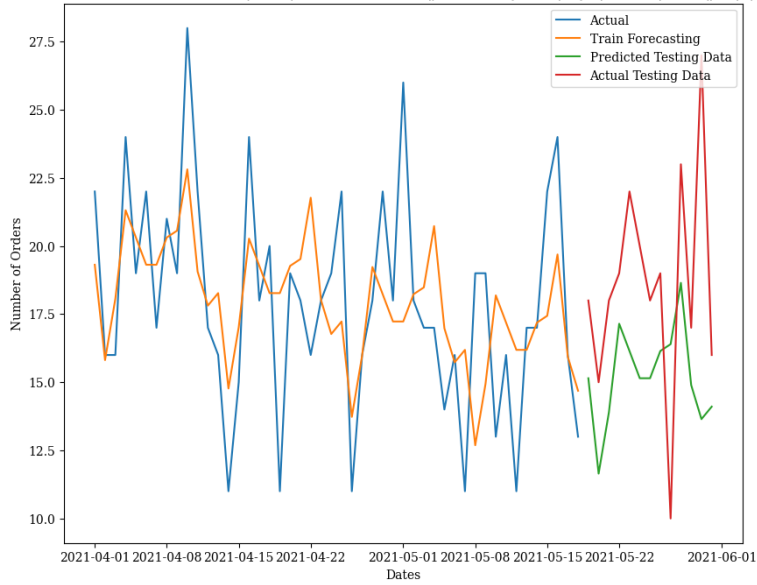


Εικόνα 33 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (6)

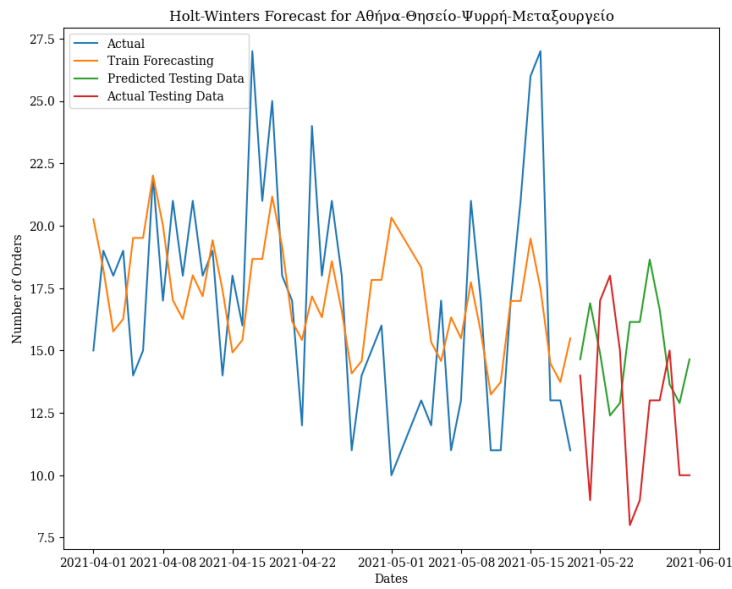


Εικόνα 34 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (7)

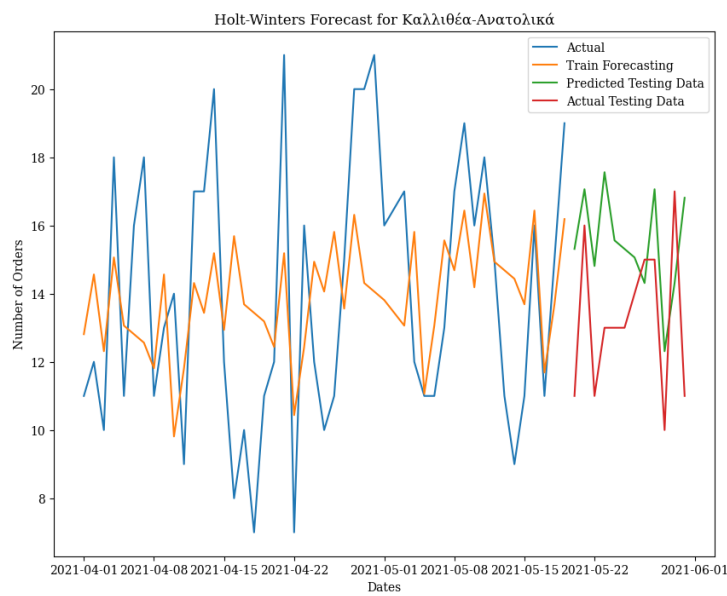
Holt-Winters Forecast for Θεσσαλονίκη-Κέντρο Δυτικά-Δικαστήρια-Παλιός Σταθμός-Αμπελόκηποι-Ξηροκρήνη-Μενεμένη



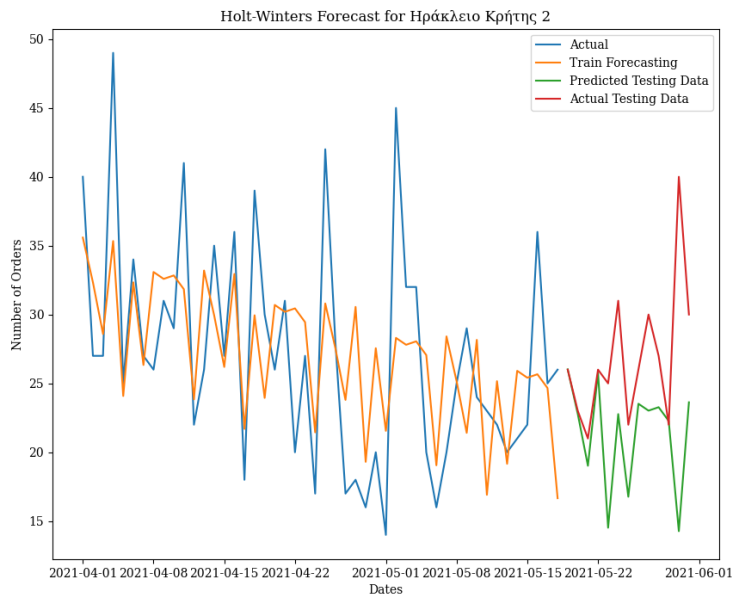
Εικόνα 35 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (8)



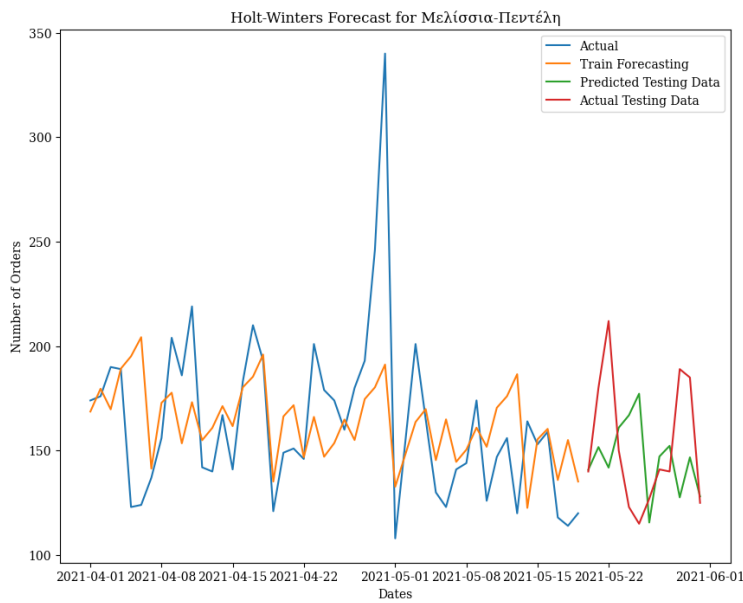
Εικόνα 36 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (9)



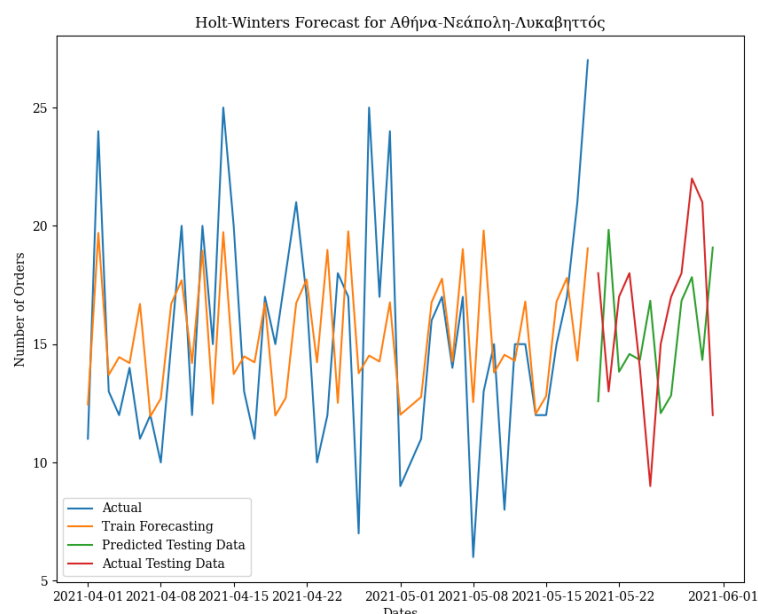
Εικόνα 37 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (10)



Εικόνα 38 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (11)



Εικόνα 39 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (12)



Εικόνα 40 1ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (13)

8.2 1^ο Επίπεδο Ιεραρχίας Γεωγραφική Περιοχή - Σφάλματα

Τα δεδομένα που παρουσιάζονται στους πίνακες αφορούν στο μέσο απόλυτο σφάλμα (MAE) και το μέσο τετραγωνικό σφάλμα (MSE) που χρησιμοποιείται για την αξιολόγηση της ακρίβειας των προβλέψεων ενός μοντέλου, τόσο για τα δεδομένα εκπαίδευσης όσο και για τα δεδομένα ελέγχου.

	MAE	MSE
Θεσσαλονίκη-Χαριλάου-Μαρτίου-Βουλγάρη	4.619260	31.633813
Νέα Σμύρνη	4.143411	26.012802
Άνω Γλυφάδα	4.880419	30.697867
Αθήνα-Κάτω Πατήσια	4.808373	32.303155
Πάτρα 2	4.489707	30.331525
Αθήνα-Ιλίσια-Μαβίλη	3.627175	21.144629
Ζωγράφου	3.853261	20.807575
Θεσσαλονίκη-Κέντρο Δυτικά-Δικαστήρια-Παλιός Σταθμός - Αμπελόκηποι-Ξηροκρήνη-Μενεμένη	4.977427	34.278366
Αθήνα-Θησείο-Ψυρρή-Μεταξουργείο	4.751059	32.524227
Καλλιθέα-Ανατολικά	3.709238	23.082691
Ηράκλειο Κρήτης 2	12.471030	191.73894
Μελίσσια-Πεντέλη	149.838074	22764.203
Αθήνα-Νεάπολη-Λυκαβηττός	3.031474	702
		15.257473

Πίνακας 5 1^ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Εκπαίδευσης

Οι τοποθεσίες με υψηλά σφάλματα είναι τα Μελίσσια – Πεντέλη και το Ηράκλειο Κρήτης 2. Το γεγονός αυτό είναι πιθανό να οφείλεται στο ότι το μοντέλο δε διαθέτει επαρκή δεδομένα για τις συγκεκριμένες περιοχές ή να υφίσταται κάποια διακύμανση στις τοπικές συνθήκες.

	MAE	MSE
Θεσσαλονίκη-Χαριλάου-Μαρτίου-Βουλγάρη	6.576583	50.581499
Νέα Σμύρνη	3.093648	15.704803

	MAE	MSE
Άνω Γλυφάδα	3.999934	27.119497
Αθήνα-Κάτω Πατήσια	2.930557	10.950034
Πάτρα 2	3.761482	18.157967
Αθήνα-Ιλίσια-Μαβίλη	4.166667	20.722180
Ζωγράφου	4.020834	21.996097
Θεσσαλονίκη-Κέντρο Δυτικά-Δικαστήρια-Παλιός Στ...	2.513641	9.770943
Αθήνα-Θησείο-Ψυρρή-Μεταξουργείο	3.871366	21.906061
Καλλιθέα-Ανατολικά	3.510416	20.069055
Ηράκλειο Κρήτης 2	4.999979	41.116644
Μελίσσια-Πεντέλη	130.247242	17297.462429
Αθήνα-Νεάπολη-Λυκαβηττός	4.430763	24.869516

Πίνακας 6 1ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Ελέγχου

Οι τοποθεσίες με υψηλά σφάλματα είναι τα Μελίσσια – Πεντέλη, το Ηράκλειο Κρήτης 2 και η Θεσσαλονίκη-Χαριλάου-Μαρτίου-Βουλγάρη. Το γεγονός αυτό είναι πιθανό να οφείλεται στο ότι το μοντέλο δε διαθέτει επαρκή δεδομένα για τις συγκεκριμένες περιοχές ή να υφίσταται κάποια διακύμανση στις τοπικές συνθήκες.

8.4 2^ο Επίπεδο Ιεραρχίας Πόλη Καταστήματος – Προβλεπόμενες Παραγγελίες

Στον ακόλουθο πίνακα παρουσιάζεται ενδεικτικά ο αριθμός των προβλεπόμενων παραγγελιών για τις πέντε πρώτες ημέρες του συνόλου δεδομένων εκπαίδευσης, που αφορά στο 2^ο επίπεδο ιεραρχίας (πόλη καταστήματος).

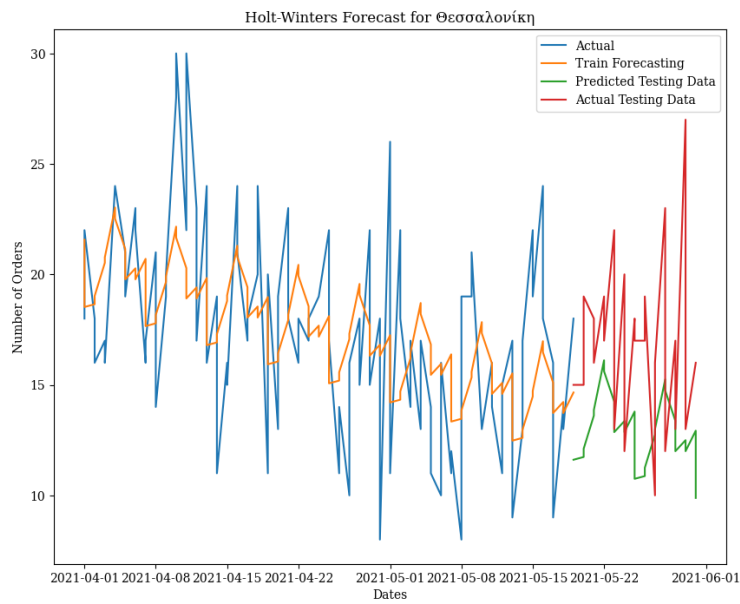
	Θεσσαλονίκη	Αθήνα	Πάτρα	Ηράκλειο Κρήτη
2021-04-01	22	33	16	36
2021-04-02	19	30	16	32
2021-04-03α	19	60	13	29
2021-04-04	19	29	15	35
2021-04-05	21	28	15	24

Πίνακας 7 Προβλεπόμενες Παραγγελίες Δεδομένων Εκπαίδευσης 2^ο Επίπεδο Ιεραρχίας

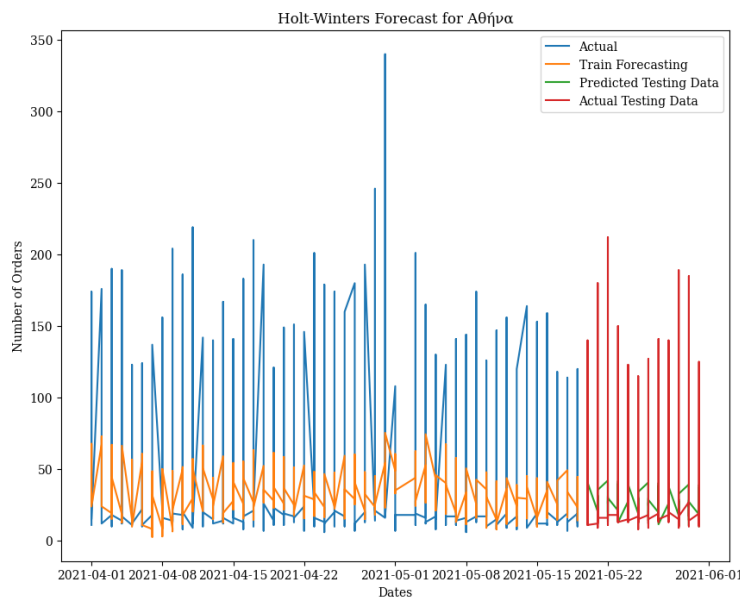
Στον ακόλουθο πίνακα παρουσιάζεται ενδεικτικά ο αριθμός των προβλεπόμενων παραγγελιών για τις πέντε πρώτες ημέρες του συνόλου δεδομένων ελέγχου, που αφορά στο 2^ο επίπεδο ιεραρχίας (πόλη καταστήματος).

	Θεσσαλονίκη	Αθήνα	Πάτρα	Ηράκλειο Κρήτη
2021-05-19	12	29	16	26
2021-05-20	12	18	15	23
2021-05-21	12	31	13	19
2021-05-22	14	22	15	26
2021-05-23	14	25	15	15

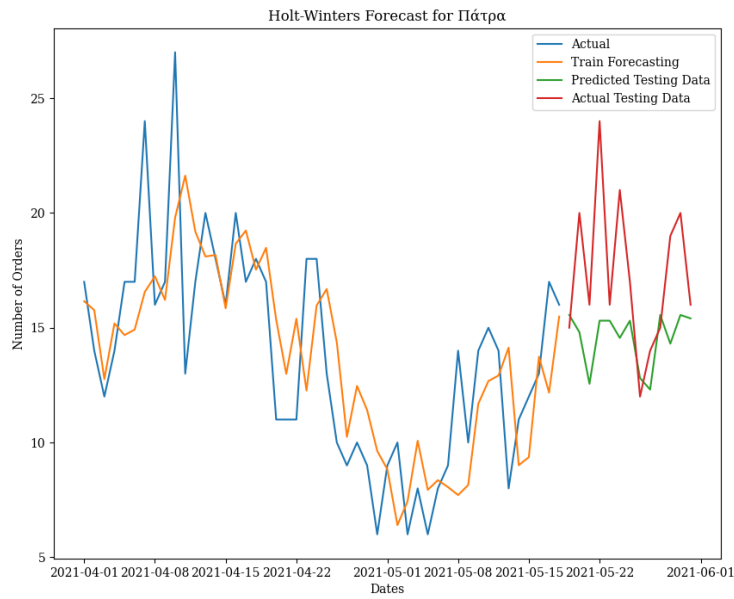
Πίνακας 8 Προβλεπόμενες Παραγγελίες Δεδομένων Ελέγχου 2^ο Επίπεδο Ιεραρχίας



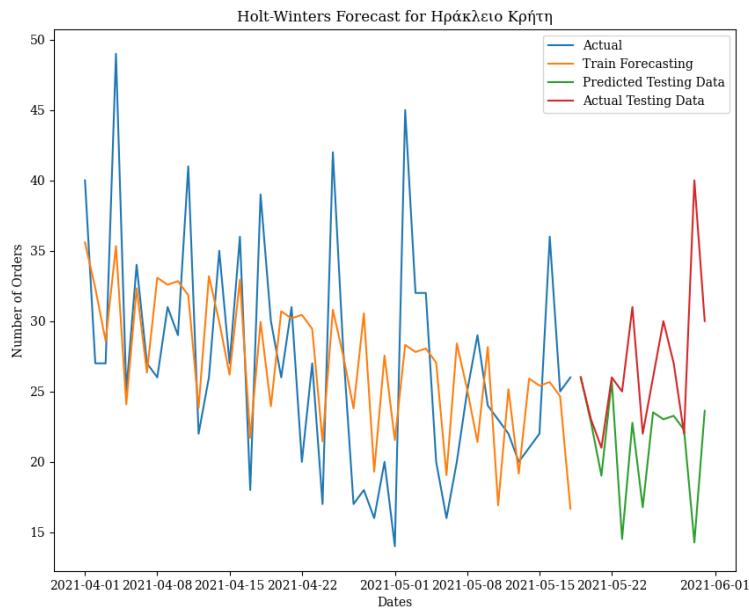
Εικόνα 41 2ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (1)



Εικόνα 42 2ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (2)



Εικόνα 43 2ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (3)



Εικόνα 44 2ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (4)

8.2 2^ο Επίπεδο Ιεραρχίας Πόλη Καταστήματος - Σφάλματα

Τα δεδομένα που παρουσιάζονται στους πίνακες αφορούν στο μέσο απόλυτο σφάλμα (MAE) και το μέσο τετραγωνικό σφάλμα (MSE) που χρησιμοποιείται για την αξιολόγηση της ακρίβειας των προβλέψεων ενός μοντέλου, τόσο για τα δεδομένα εκπαίδευσης όσο και για τα δεδομένα ελέγχου.

	MAE	MSE
Θεσσαλονίκη	8.649346	125.986103
Αθήνα	17.179755	496.094128
Πάτρα	13.509769	241.440271

	MAE	MSE
Ηράκλειο Κρήτη	5.262723	42.125956

Πίνακας 9 2^ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Εκπαίδευσης

Οι πόλεις με υψηλά σφάλματα είναι η Θεσσαλονίκη, η Αθήνα και η Πάτρα. Το γεγονός αυτό είναι πιθανό να οφείλεται στο ότι το μοντέλο δε διαθέτει επαρκή δεδομένα, στη διακύμανση των δεδομένων ή σε διαφορετικούς εξωτερικούς παράγοντες.

	MAE	MSE
Θεσσαλονίκη	13.575148	208.608926
Αθήνα	7.873526	117.976490
Πάτρα	12.282215	173.754728
Ηράκλειο Κρήτη	5.535000	75.405319

Πίνακας 10 2^ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Ελέγχου

Οι πόλεις με υψηλά σφάλματα είναι η Θεσσαλονίκη, η Αθήνα και η Πάτρα. Το γεγονός αυτό είναι πιθανό να οφείλεται στο ότι το μοντέλο δε διαθέτει επαρκή δεδομένα, στη διακύμανση των δεδομένων ή σε διαφορετικούς εξωτερικούς παράγοντες.

8.5 3^ο Επίπεδο Ιεραρχίας – Προβλεπόμενες Παραγγελίες

Στον ακόλουθο πίνακα παρουσιάζεται ο αριθμός των προβλεπόμενων παραγγελιών για το 3^ο επίπεδο ιεραρχίας (κατάστημα), που αφορά στα δεδομένα εκπαίδευσης.

	08f46dcf2352f5e5a272957ceb971b7e	263c22def802d054642fa5aa32da24e0	3b79459351228b9d823b0ec9e40cf952	6fd9293ee5eefee4ab1fd29713b84a4	8352e66afdd11dee472475518ecafaa1	85e5bf2dbc5bb890d23f2ab86c92648a	973b6f45bf3b42c8fc7cf58040109868	abec811393703b50a6699790083a577f	e3a1e3ae3f7a8e705c86d8e4ad337b39	eb8c6bb6337bcf48b6e64f9903b4aae2	d119eb50957208adffa28742636b3600	eb160de1de89d9058fcb0b968dbbbd68	f21eb543a4c81f57b31316fd2a3b89cd
2021-04-01	22	14	15	17	16	18	17	19	20	13	36	169	12
2021-04-02	19	16	20	20	16	18	20	16	18	15	32	180	20
2021-04-03	20	13	16	19	13	17	15	18	16	12	29	170	14
2021-04-04	22	15	16	19	15	16	16	21	16	15	35	189	14
2021-04-05	21	13	19	21	15	19	16	20	20	13	24	195	14
2021-04-06	22	17	21	22	15	17	20	19	20	13	32	204	17
2021-04-07	18	18	19	19	17	13	15	19	22	13	26	141	12
2021-04-08	20	17	18	21	17	16	16	20	20	12	33	173	13
2021-04-09	23	17	18	20	16	16	17	21	17	15	33	178	17
2021-04-10	24	17	14	18	20	20	16	23	16	10	33	153	18
2021-04-11	21	18	19	20	22	15	19	19	18	12	32	173	14

	08f46dcf2352f5e5a272957ceb971b7e	263c22dcf802d054642fa5aa32da24e0	3b79459351228b9d823b0ec9e40cf952	6fd9293ee5eefee4ab1fd29713b84a4	8352e66afdd11dec472475518ecafaa1	85e5bf2dbc5bb890d23f2ab86c92648a	973b6f45bf3b42c8fc7cf58040109868	abee831393703b50a6699790083a577f	c3a1e3ae3f7a8e705c86d8e4ad337b39	eb8c6bb6337b4f48b6e64f9903b4aac2	d119eb50957208adffa28742636b3600	eb160de1de89d9058fcb0b968dbbbd68	f21eb543a4c81f57b31316fd2a3b89cd
2021-04-12	21	14	13	19	19	12	14	18	17	14	24	155	19
2021-04-13	19	15	15	16	18	17	17	18	19	13	33	161	12
2021-04-14	16	17	20	19	18	18	20	15	17	15	30	171	20
2021-04-15	17	13	16	18	16	17	14	17	15	13	26	162	14
2021-04-16	20	15	16	18	19	16	15	20	15	16	33	180	14
2021-04-17	18	13	19	20	19	18	15	19	19	14	22	185	14
2021-04-18	19	17	21	21	18	16	19	18	19	13	30	196	17
2021-04-19	15	19	19	18	18	12	15	18	21	13	24	135	12
2021-04-20	17	18	18	20	15	15	16	19	19	12	31	166	13
2021-04-21	20	17	18	19	13	16	16	20	16	15	30	172	17
2021-04-22	21	17	14	17	15	20	15	22	15	10	30	147	18
2021-04-23	18	19	19	19	12	14	19	18	17	12	29	166	14

	08f46dcf2352f5e5a272957ceb971b7e	263c22dcf802d054642fa5aa32da24e0	3b79459351228b9d823b0ec9e40cf952	6fd9293ee5eefee4ab1fd29713b84a4	8352e66afdd11dec472475518ecafaa1	85e5bf2dbc5bb890d23f2ab86e92648a	973b6f45bf3b42c8fc7cf58040109868	abee831393703b50a6699790083a577f	c3a1e3ae3f7a8e705c86d8e4ad337b39	eb8c6bb6337b4f48b6e64f9903b4aac2	d119eb50957208adffa28742636b3600	eb160de1de89d9058fcb0b968dbbbd68	f21eb543a4c81f57b31316fd2a3b89cd
2021-04-24	18	15	13	18	16	11	14	17	16	15	21	147	19
2021-04-25	16	15	14	16	17	17	16	17	19	14	31	154	13
2021-04-26	13	17	20	18	14	17	19	14	17	16	28	165	20
2021-04-27	15	13	15	18	10	16	13	16	14	14	24	155	14
2021-04-28	17	16	16	17	12	15	15	19	15	16	31	175	15
2021-04-29	15	13	19	19	11	18	14	18	18	14	19	180	14
2021-04-30	16	17	21	21	10	15	18	17	18	14	28	191	17
2021-05-01	13	19	19	17	9	11	14	17	20	14	22	133	12
2021-05-03	14	18	18	19	6	15	15	18	18	13	28	164	13
2021-05-04	18	17	17	18	7	15	16	18	15	16	28	170	17
2021-05-05	18	17	14	16	10	19	15	21	15	11	28	145	18
2021-05-06	15	19	19	18	8	13	18	17	16	13	27	165	14

	08f46dcf2352f5e5a272957ceb971b7e	263c22dcf802d054642fa5aa32da24e0	3b79459351228b9d823b0ec9e40cf952	6fd9293ee5eefee4ab1fd29713b84a4	8352e66afdd11dec472475518ecafaa1	85e5bf2dbc5bb890d23f2ab86c92648a	973b6f45bf3b42c8fc7cf58040109868	abee831393703b50a6699790083a577f	c3a1e3ae3f7a8e705c86d8e4ad337b39	eb8c6bb6337b4f48b6e64f9903b4aac2	d119eb50957208adffa28742636b3600	eb160de1de89d9058fcb0b968dbbbd68	f21eb543a4c81f57b31316fd2a3b89cd
2021-05-07	15	15	13	17	8	11	13	16	15	16	19	145	19
2021-05-08	14	16	14	15	8	16	15	16	18	15	28	150	13
2021-05-09	11	18	20	17	8	16	18	13	16	16	25	161	20
2021-05-10	12	14	15	17	8	15	13	15	13	14	21	152	14
2021-05-11	14	16	16	16	12	14	14	18	14	17	28	171	15
2021-05-12	13	14	19	18	13	17	14	17	17	15	17	176	14
2021-05-13	13	18	21	20	13	15	18	16	17	15	25	187	17
2021-05-14	10	19	19	16	14	11	13	16	19	14	19	123	12
2021-05-15	11	19	18	18	9	14	15	17	17	14	26	155	13
2021-05-16	15	18	17	17	9	14	15	17	14	16	25	160	17
2021-05-17	16	18	14	15	14	18	14	20	14	12	26	136	18
2021-05-18	13	19	19	NaN	12	13	17	16	15	14	25	155	14

08f46dcf2352f5e5a272957ecb971b7e	263c22dcf802d054642fa5aa32da24e0	3b79459351228b9d823b0ec9e40cf952	6fd9293ee5eefee4ab1fd29713b84a4	8352e66afdd11dee472475518ecafaa1	85e5bf2dbc5bb890d23f2ab892648a	973b6f45bf3b42e8fc7cf58040109868	abee831393703b50a6699790083a577f	c3a1e3ae3f7a8e705c86d8e4ad337b39	cb8c6bb6337bcf48b6e64f9903b4aac2	d119eb50957208adffa28742636b3600	eb160de1de89d9058fcb0b968dbbbd68	f21eb543a4c81f57b31316fd2a3b89cd	
2021-05-19	12	16	13	NaN	15	10	13	15	NaN	16	17	135	19

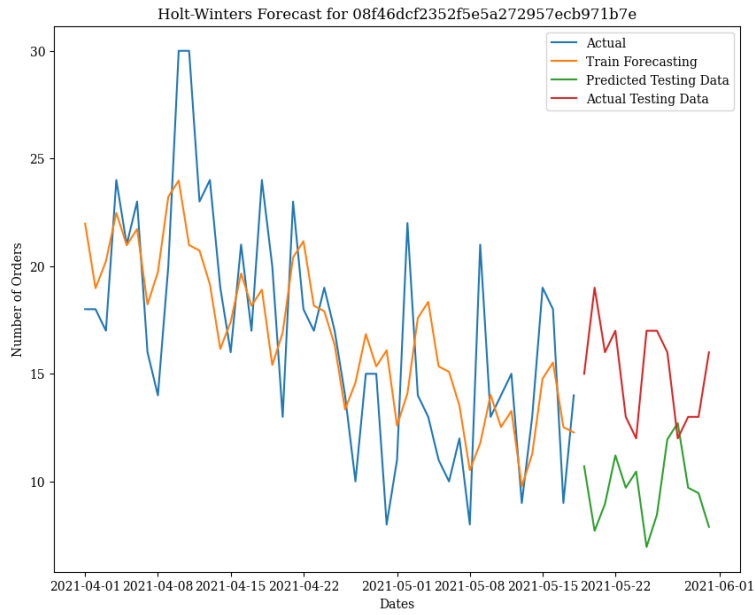
Πίνακας 11 Προβλεπόμενες Παραγγελίες Δεδομένων Εκπαίδευσης 3^ο Επίπεδο Ιεραρχίας

Στον ακόλουθο πίνακα παρουσιάζεται ο αριθμός των προβλεπόμενων παραγγελιών για το 3^ο επίπεδο ιεραρχίας (κατάστημα), που αφορά στα δεδομένα ελέγχου.

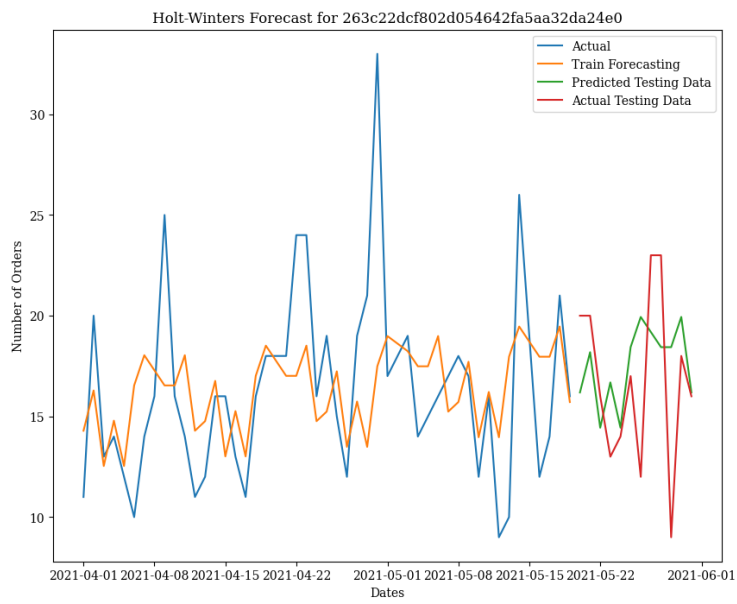
08f46dcf2352f5e5a272957ecb971b7e	263c22dcf802d054642fa5aa32da24e0	3b79459351228b9d823b0ec9e40cf952	6fd9293ee5eefee4ab1fd29713b84a4	8352e66afdd11dee472475518ecafaa1	85e5bf2dbc5bb890d23f2ab892648a	973b6f45bf3b42e8fc7cf58040109868	abee831393703b50a6699790083a577f	c3a1e3ae3f7a8e705c86d8e4ad337b39	cb8c6bb6337bcf48b6e64f9903b4aac2	d119eb50957208adffa28742636b3600	eb160de1de89d9058fcb0b968dbbbd68	f21eb543a4c81f57b31316fd2a3b89cd	
2021-05-20	11	16	14	17	16	15	15	15	15	26	141	13	
2021-05-21	8	18	20	16	15	15	18	12	17	17	23	152	20
2021-05-22	9	14	15	14	13	14	12	14	15	19	142	14	
2021-05-23	11	17	16	16	15	13	13	17	12	18	26	161	15
2021-05-24	10	14	19	16	15	16	13	16	13	16	15	167	14
2021-05-25	10	18	21	15	15	14	17	15	16	15	23	177	17
2021-05-26	7	20	19	17	15	10	13	15	16	15	17	116	12
2021-05-27	8	19	18	19	13	13	14	16	19	14	24	147	13
2021-05-28	12	18	17	15	12	13	14	16	17	17	23	152	17
2021-05-29	13	18	14	17	16	17	13	19	14	12	23	128	18
2021-05-30	10	20	19	16	14	12	17	15	13	14	22	147	14
2021-05-31	9	16	13	14	16	9	12	14	15	17	14	128	19

Πίνακας 12 Προβλεπόμενες Παραγγελίες Δεδομένων Ελέγχου 3^ο Επίπεδο Ιεραρχίας

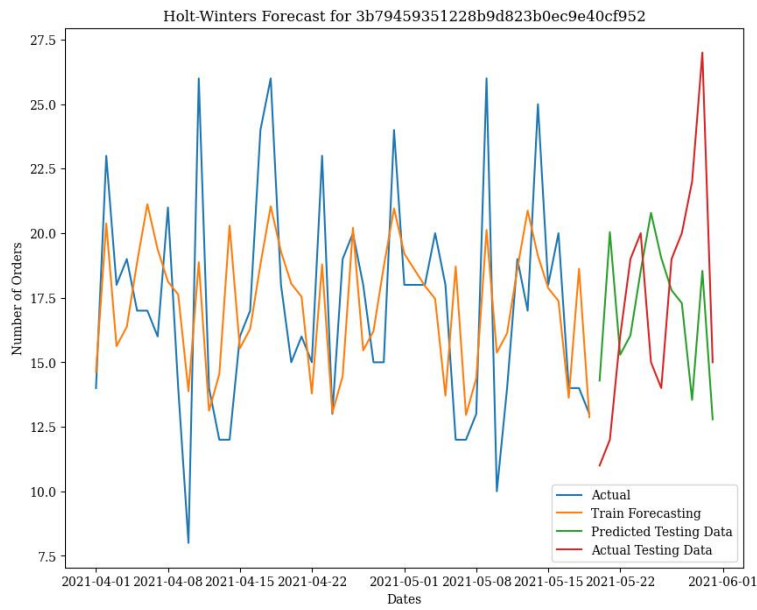
Τμηματοποίηση πελατών και πρόβλεψη πωλήσεων για μια ηλεκτρονική πλατφόρμα πωλήσεων



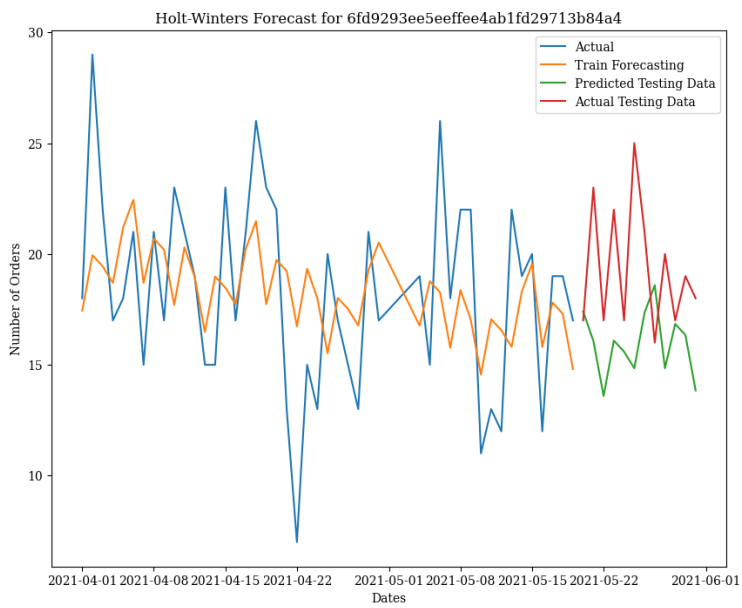
Εικόνα 45 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (1)



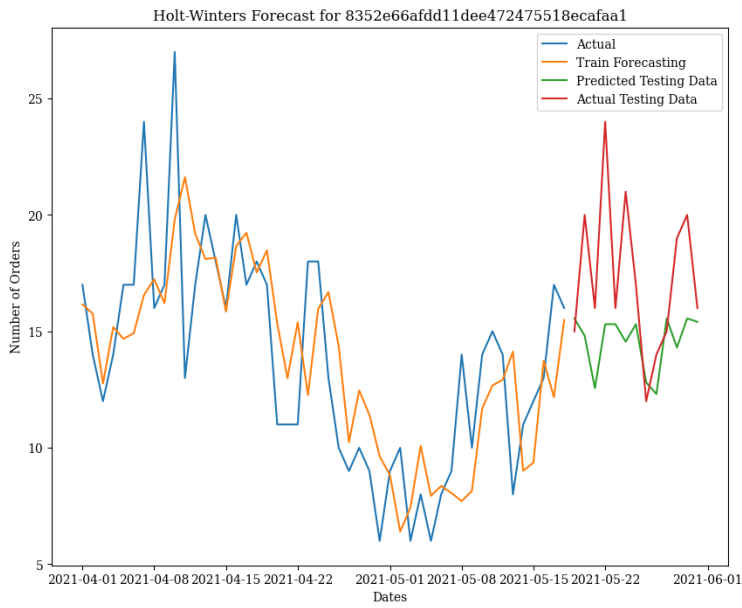
Εικόνα 46 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (2)



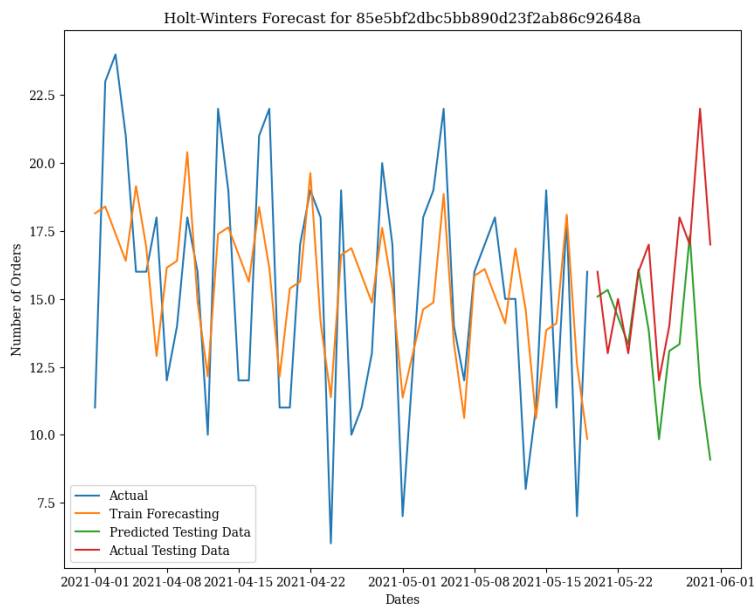
Εικόνα 47 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (3)



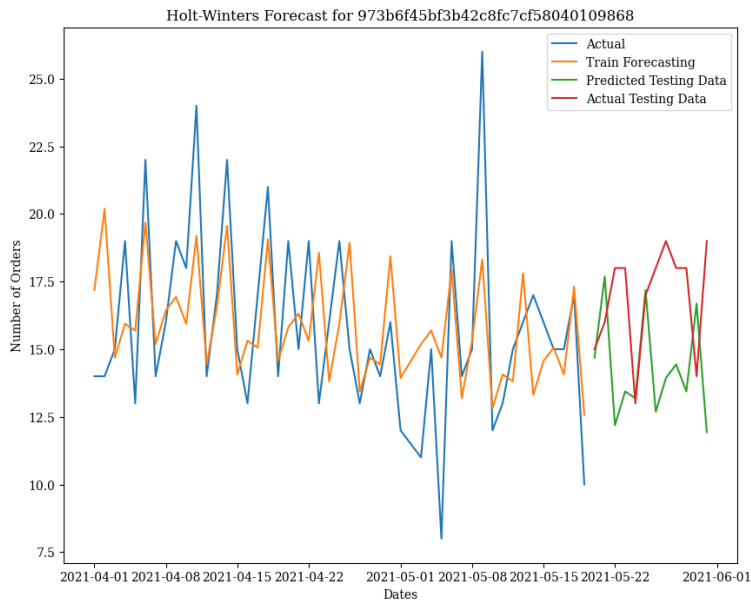
Εικόνα 48 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (4)



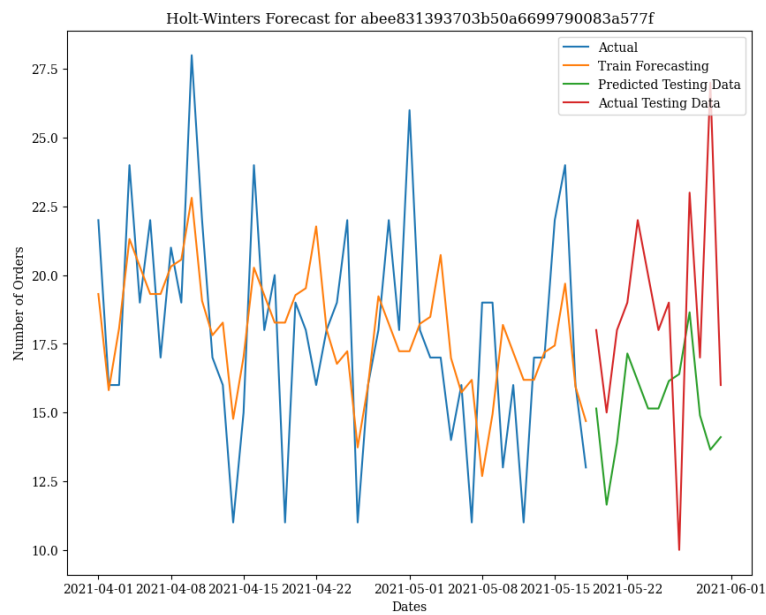
Εικόνα 49 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (5)



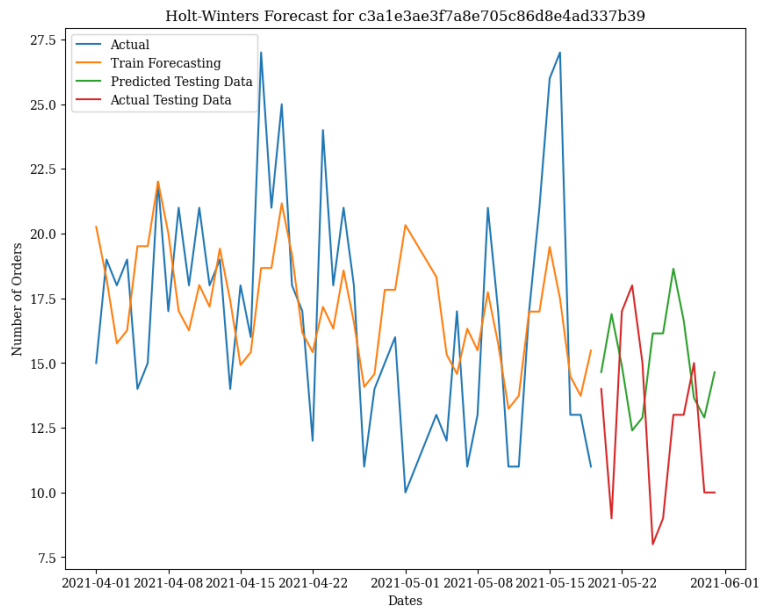
Εικόνα 50 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (6)



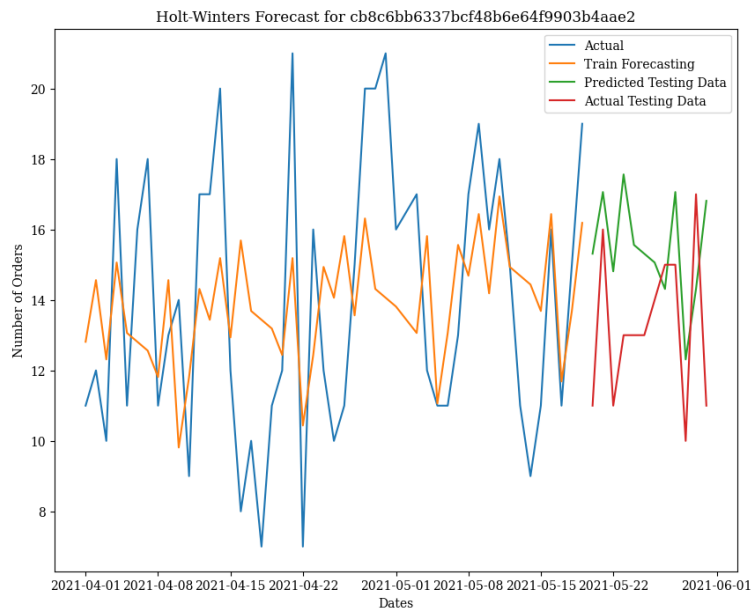
Εικόνα 51 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (7)



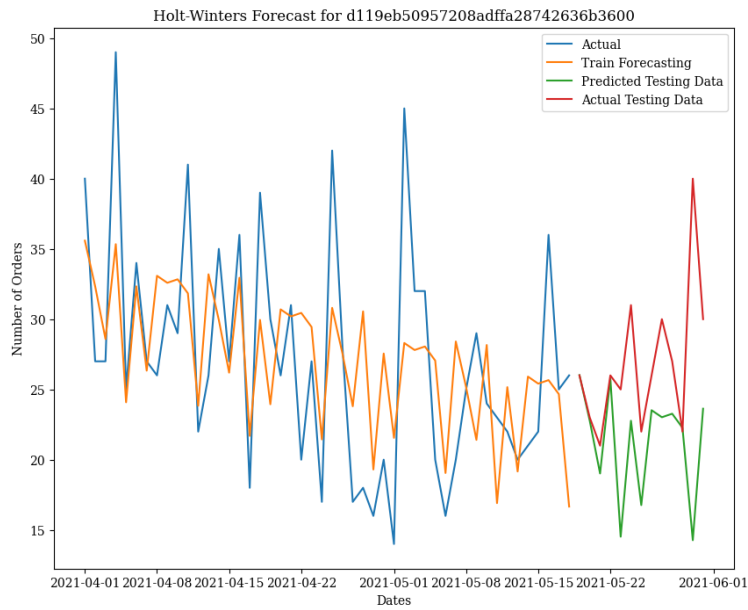
Εικόνα 52 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (8)



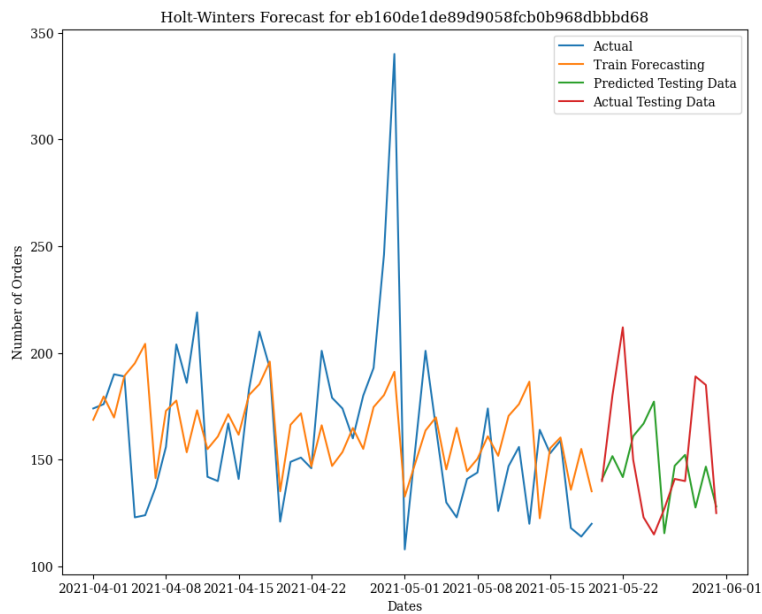
Εικόνα 53 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (9)



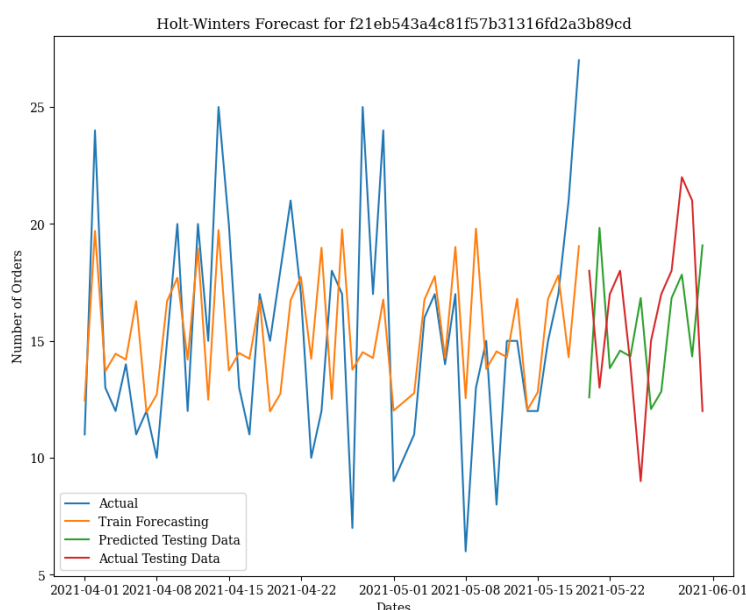
Εικόνα 54 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (10)



Εικόνα 55 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (11)



Εικόνα 56 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (12)



Εικόνα 57 3ο Επίπεδο Ιεραρχίας Δεδομένα Εκπαίδευσης και Ελέγχου (13)

8.6 3^ο Επίπεδο Ιεραρχίας Κατάστημα - Σφάλματα

Τα δεδομένα που παρουσιάζονται στους πίνακες αφορούν στο μέσο απόλυτο σφάλμα (MAE) και το μέσο τετραγωνικό σφάλμα (MSE) που χρησιμοποιείται για την αξιολόγηση της ακρίβειας των προβλέψεων ενός μοντέλου, τόσο για τα δεδομένα εκπαίδευσης όσο και για τα δεδομένα ελέγχου.

Column2	MAE	MSE
08f46dcf2352f5e5a272957ecb971b7e	4.619260	31.633813
263c22dcf802d054642fa5aa32da24e0	4.143411	26.012802
3b79459351228b9d823b0ec9e40cf952	4.880419	30.697867
6fd9293ee5effee4ab1fd29713b84a4	4.808373	32.303155
8352e66afdd11dee472475518ecafaa1	4.489707	30.331525
85e5bf2dbc5bb890d23f2ab86c92648a	3.627175	21.144629
973b6f45bf3b42c8fc7cf58040109868	3.853261	20.807575
abee831393703b50a6699790083a577f	4.977427	34.278366
c3a1e3ae3f7a8e705c86d8e4ad337b39	4.751059	32.524227
cb8c6bb6337bcf48b6e64f9903b4aae2	3.709238	23.082691
d119eb50957208adffa28742636b3600	12.471030	191.738940
eb160de1de89d9058fcb0b968dbbbd68	149.838074	22764.203702
f21eb543a4c81f57b31316fd2a3b89cd	3.031474	15.257473

Πίνακας 13 3^ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Εκπαίδευσης

Τα καταστήματα με υψηλά σφάλματα είναι το d119eb50957208adffa28742636b3600 και το eb160de1de89d9058fcb0b968dbbbd68. Το γεγονός αυτό είναι πιθανό να οφείλεται στο ότι το μοντέλο δε διαθέτει επαρκή δεδομένα, στη διακύμανση των δεδομένων ή σε διαφορετικούς εξωτερικούς παράγοντες.

	MAE	MSE
08f46dcf2352f5e5a272957ecb971b7e	6.576583	50.581499
263c22dcf802d054642fa5aa32da24e0	3.093648	15.704803

3b79459351228b9d823b0ec9e40cf952	3.999934	27.119497
6fd9293ee5eeffe4ab1fd29713b84a4	2.930557	10.950034
8352e66afdd11dee472475518ecafaa1	3.761482	18.157967
85e5bf2dbc5bb890d23f2ab86c92648a	4.166667	20.722180
973b6f45bf3b42c8fc7cf58040109868	4.020834	21.996097
abee831393703b50a6699790083a577f	2.513641	9.770943
c3a1e3ae3f7a8e705c86d8e4ad337b39	3.871366	21.906061
cb8c6bb6337bcf48b6e64f9903b4aae2	3.510416	20.069055
d119eb50957208adffa28742636b3600	4.999979	41.116644
eb160de1de89d9058fcb0b968dbbbd68	130.247242	17297.462429
f21eb543a4c81f57b31316fd2a3b89cd	4.430763	24.869516

Πίνακας 14 3^ο Επίπεδο Ιεραρχίας Σφάλματα – Δεδομένα Ελέγχου

Τα καταστήματα με υψηλά σφάλματα είναι το 08f46dcf2352f5e5a272957ecb971b7e, το d119eb50957208adffa28742636b3600 και το eb160de1de89d9058fcb0b968dbbbd68. Το γεγονός αυτό είναι πιθανό να οφείλεται στο ότι το μοντέλο δε διαθέτει επαρκή δεδομένα, στη διακύμανση των δεδομένων ή σε διαφορετικούς εξωτερικούς παράγοντες.

8.7 Αποτελέσματα

Τα αποτελέσματα των δεδομένων εκπαίδευσης και ελέγχου με τη χρήση του μοντέλου Holt Winters παρουσιάστηκαν και για τις τρεις ιεραρχίες. Το πρώτο επίπεδο της ιεραρχίας είναι η γεωγραφική περιοχή των καταστημάτων της ηλεκτρονικής πλατφόρμας πωλήσεων, το δεύτερο επίπεδο είναι η πόλη των καταστημάτων και το τρίτο επίπεδο είναι τα καταστήματα.

Στα δεδομένα εκπαίδευσης η μπλε γραμμή απεικονίζει τον πραγματικό αριθμό παραγγελιών στο διάστημα από 1 Απριλίου 2021 έως 18 Μαΐου 2021. Σε αυτά τα δεδομένα εμφανίζονται αρκετές διακυμάνσεις στον αριθμό παραγγελιών. Το γεγονός αυτό μπορεί να οφείλεται σε διάφορους παράγοντες, όπως η εποχικότητα ή άλλα γεγονότα. Η πορτοκαλί γραμμή απεικονίζει τις προβλέψεις για τα δεδομένα εκπαίδευσης. Είναι φανερό ότι οι προβλέψεις έχουν ακολουθήσει σε ικανοποιητικό βαθμό τις τάσεις των πραγματικών δεδομένων. Συνεπώς, το μοντέλο, έχει τη δυνατότητα να αντιληφθεί τις εποχικές και τυχαίες συνιστώσες της ζήτησης σε μεγάλο βαθμό.

Στα δεδομένα ελέγχου η κόκκινη γραμμή απεικονίζει τον πραγματικό αριθμό παραγγελιών στο διάστημα από 19 Μαΐου 2021 έως 31 Μαΐου 2021. Τη συγκεκριμένη περίοδο παρατηρείται πιο μεγάλη μεταβλητότητα σε σχέση με τις προβλέψεις, γεγονός που είναι πιθανό να συμβαίνει καθώς μπορεί να υπάρχουν εξωτερικοί παράγοντες που δεν έχουν ληφθεί υπόψιν. Η πράσινη γραμμή απεικονίζει τις προβλέψεις για τα δεδομένα εκπαίδευσης. Εμφανίζονται μικρότερες διακυμάνσεις στα δεδομένα και ο αριθμός των παραγγελιών είναι μικρότερος από αυτόν των πραγματικών δεδομένων.

Τα γενικά συμπεράσματα είναι ότι το μοντέλο Holt Winters έχει συλλάβει σε ικανοποιητικό βαθμό τις τάσεις και τις εποχικές συνιστώσες κατά την περίοδο εκπαίδευσης. Κατά την περίοδο ελέγχου οι προβλέψεις του αριθμού παραγγελιών εμφανίζουν κάποιες αποκλίσεις από τον πραγματικό αριθμό παραγγελιών. Οι αποκλίσεις αυτές είναι πιθανό να οφείλονται στην ανάγκη για ενσωμάτωση μεγαλύτερου όγκου δεδομένων στο υπάρχον σύνολο δεδομένων που εξετάζεται.

9. Πρόβλεψη Πωλήσεων - Αποτελέσματα Πρόβλεψης Πωλήσεων

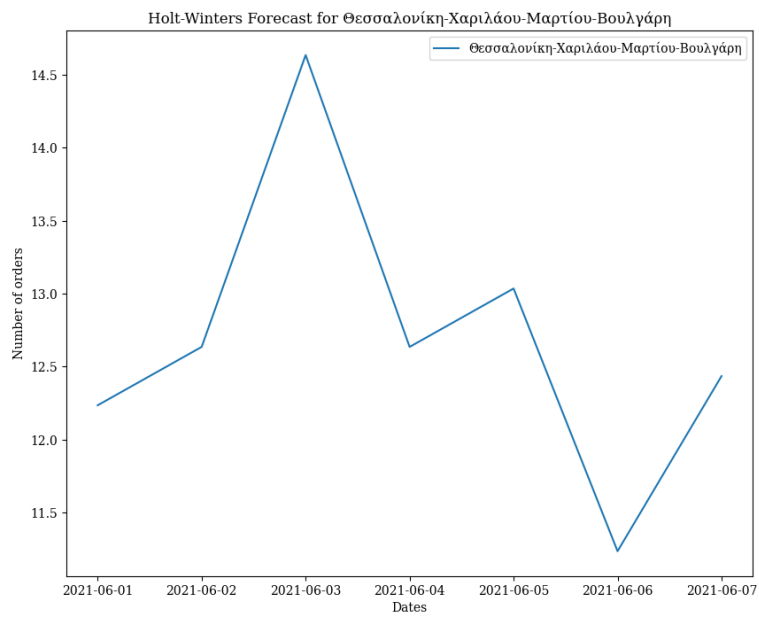
Η πρόβλεψη πωλήσεων έχει διενεργηθεί για το χρονικό διάστημα εφτά ημερών από 01/06/2021 έως 07/06/2021 με τη χρήση του μοντέλου Holt-Winters. Τα αποτελέσματα της πρόβλεψης πωλήσεων παρουσιάζονται και για τις τρεις ιεραρχίες στις εικόνες που ακολουθούν. Το πρώτο επίπεδο της ιεραρχίας είναι η γεωγραφική περιοχή των καταστημάτων της ηλεκτρονικής πλατφόρμας πωλήσεων, το δεύτερο επίπεδο είναι η πόλη των καταστημάτων και το τρίτο επίπεδο είναι τα καταστήματα.

9.1 Πρόβλεψη Πωλήσεων 1^ο Επίπεδο Ιεραρχίας

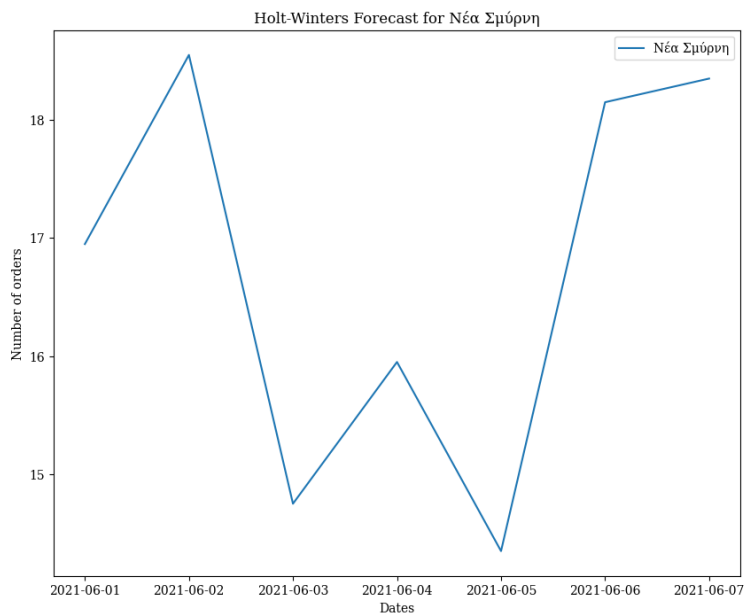
Στον ακόλουθο πίνακα παρουσιάζεται ο αριθμός των προβλεπόμενων παραγγελιών για το 1^ο επίπεδο ιεραρχίας, που αφορά στο διάστημα από 01/06/2021 έως 07/06/2021.

	Θεσσαλονίκη- Χαριλάου- Μαρτίου- Βουλγάρη	Νέα Σμύρνη	Α ν ω Γ λ υ φ α	Αθήνα- Κάτω Πατήσι α	Π ά τ ρα 2	Αθήνα- Λίσα- Μαβί η	Ζω γράφ ο υ	Θεσσαλονίκη-Κέντρο Λυκαστήρια-Παλιάς Αμπελόκηποι-Ξηροκρήνη- Μενεμένη	Αθήνα- Θησείο- Ψυρρή- Μεταξουργείο	Καλλι θέα- Ανατο λικά	Ηράκ λειο Κρήτ ης 2	Μελί σια- Πεντέ λη	Αθήνα- Νεάπολη- Λυκαβητ τός
20 21- 06- 01	12	17	4	18	1	16	1	13	12	14	24	132	14
20 21- 06- 02	13	19	9	19	1	15	1	16	13	16	20	150	19
20 21- 06- 03	15	15	6	15	1	15	1	19	13	13	27	148	15
20 21- 06- 04	13	16	7	18	1	14	1	18	11	16	18	151	16
20 21- 06- 05	13	14	9	17	1	17	1	17	11	14	25	153	15
20 21- 06- 06	11	18	0	18	1	15	1	17	12	14	19	159	16
20 21- 06- 07	12	18	8	19	1	11	1	18	12	14	25	112	13

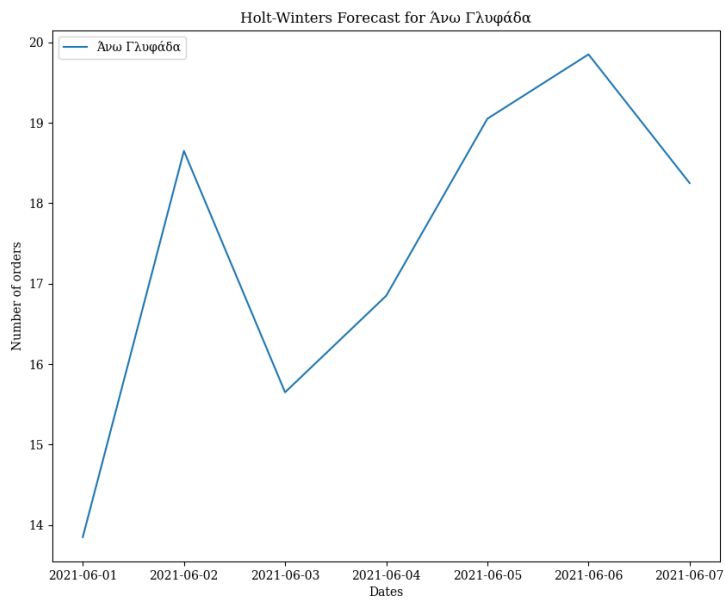
Πίνακας 15 Προβλεπόμενες Πωλήσεις 1^ο Επίπεδο Ιεραρχίας (01/06/2021 έως 07/06/2021)



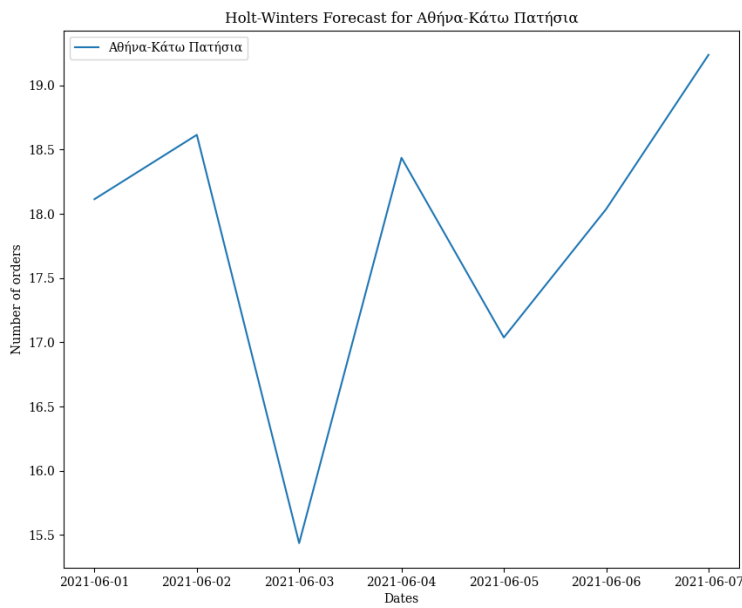
Εικόνα 58 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (1)



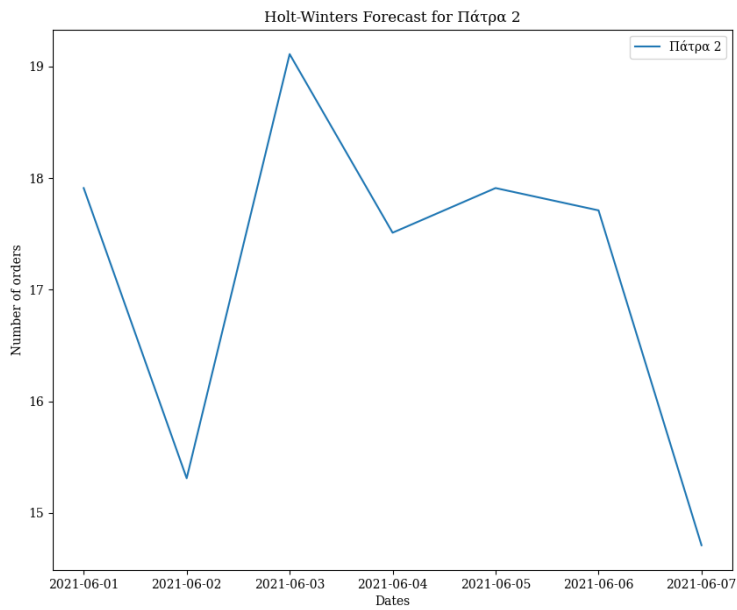
Εικόνα 59 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (2)



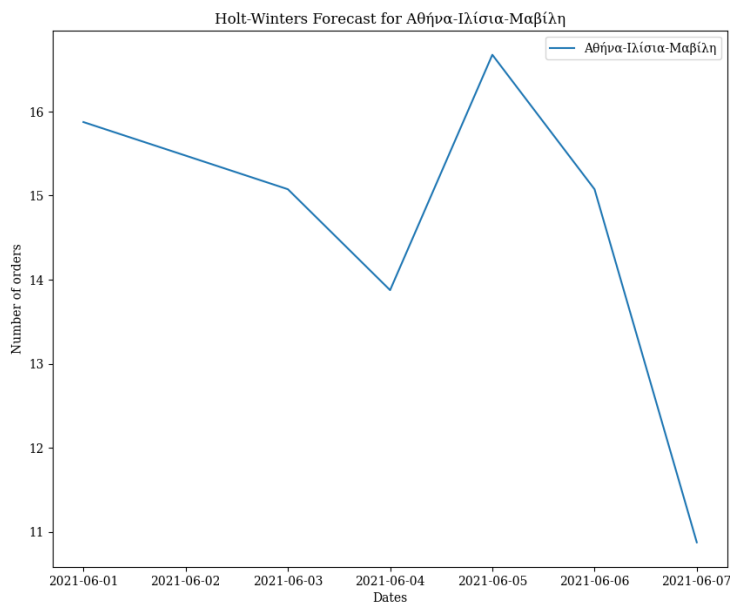
Εικόνα 60 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (3)



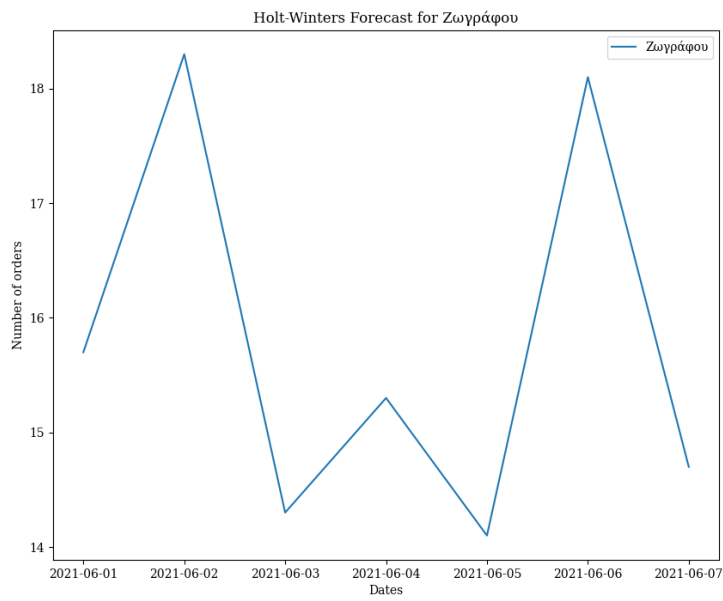
Εικόνα 61 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (4)



Εικόνα 62 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (5)

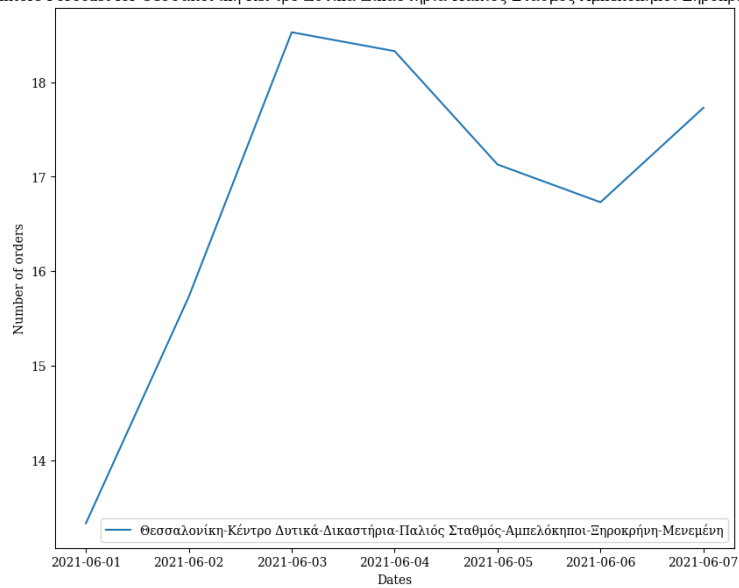


Εικόνα 63 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (6)

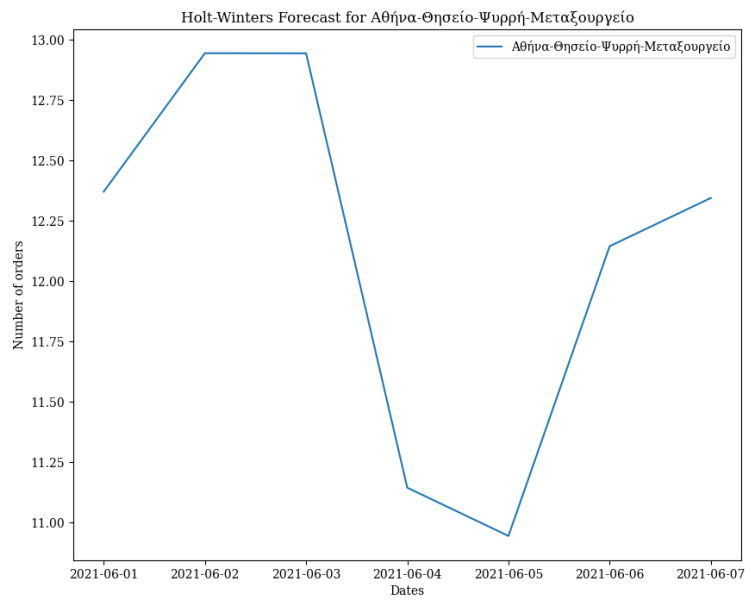


Εικόνα 64 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (7)

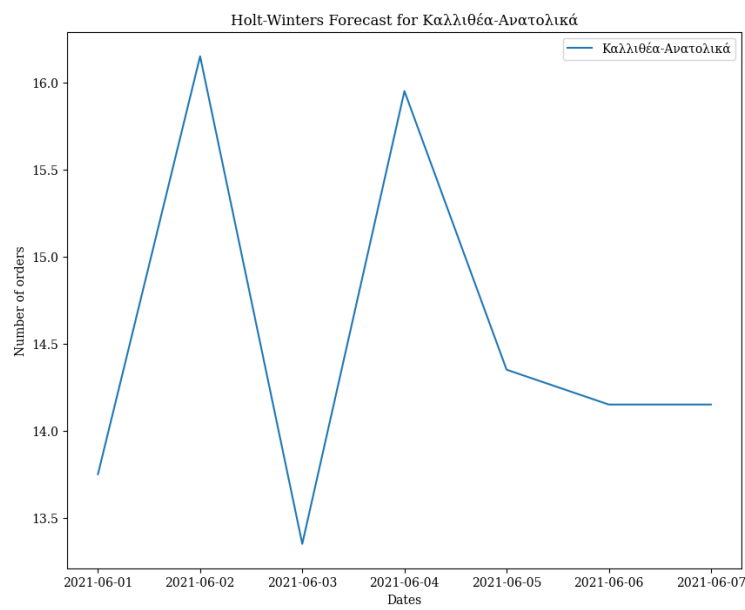
Holt-Winters Forecast for Θεσσαλονίκη-Κέντρο Δυτικά-Δικαστήρια-Παλιός Σταθμός-Αμπελόκηποι-Ξηροκρήνη-Μευεμένη



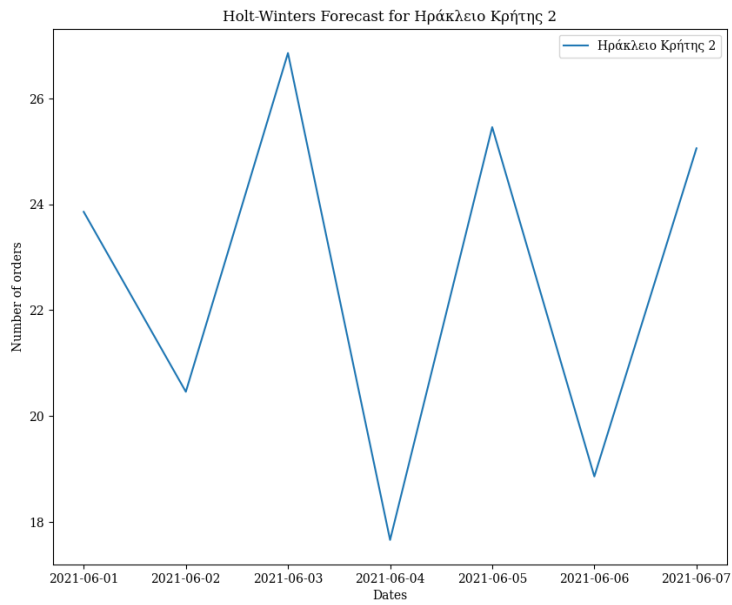
Εικόνα 65 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (8)



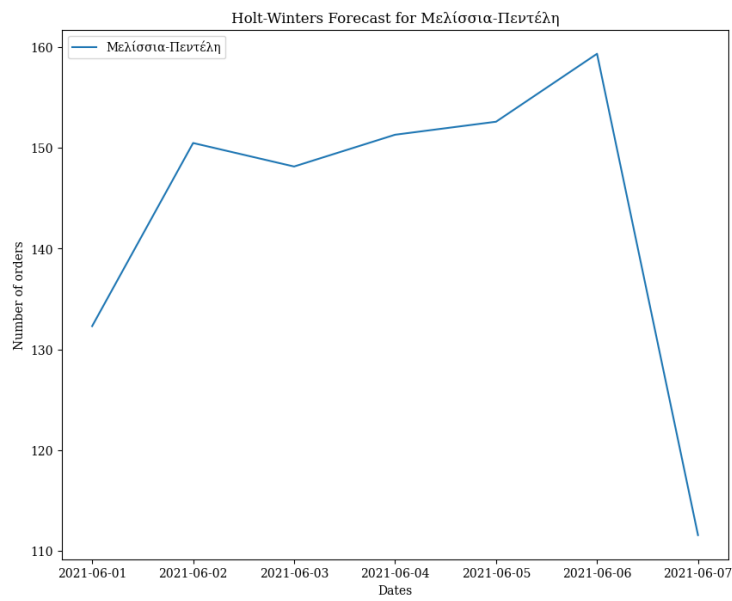
Εικόνα 66 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (9)



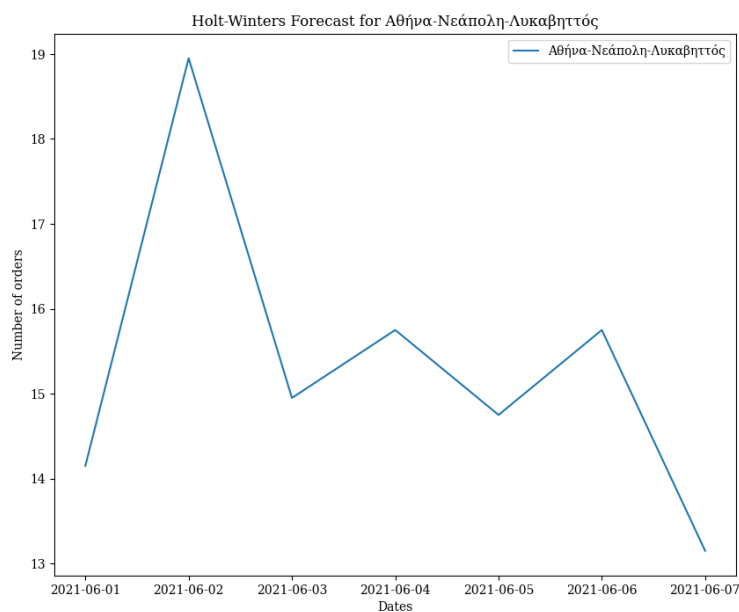
Εικόνα 67 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (10)



Εικόνα 68 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (11)



Εικόνα 69 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (12)



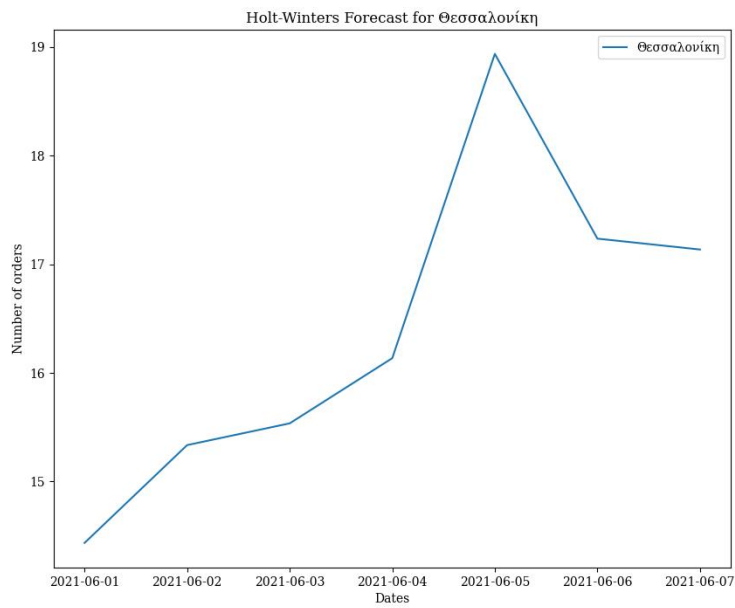
Εικόνα 70 1ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (13)

9.2 Πρόβλεψη Πωλήσεων 2ο Επίπεδο Ιεραρχίας

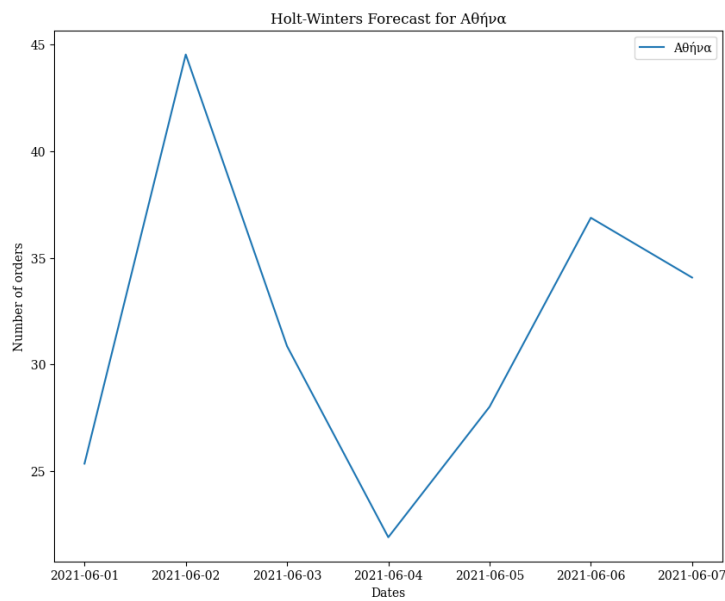
Στον ακόλουθο πίνακα παρουσιάζεται ο αριθμός των προβλεπόμενων παραγγελιών για το 2ο επίπεδο ιεραρχίας, που αφορά στο διάστημα από 01/06/2021 έως 07/06/2021.

	Θεσσαλονίκη	Αθήνα	Πάτρα	Ηράκλειο Κρήτη
2021-06-01	14	25	18	24
2021-06-02	15	45	15	20
2021-06-03	16	31	19	27
2021-06-04	16	22	18	18
2021-06-05	19	28	18	25
2021-06-06	17	37	18	19
2021-06-07	17	34	15	25

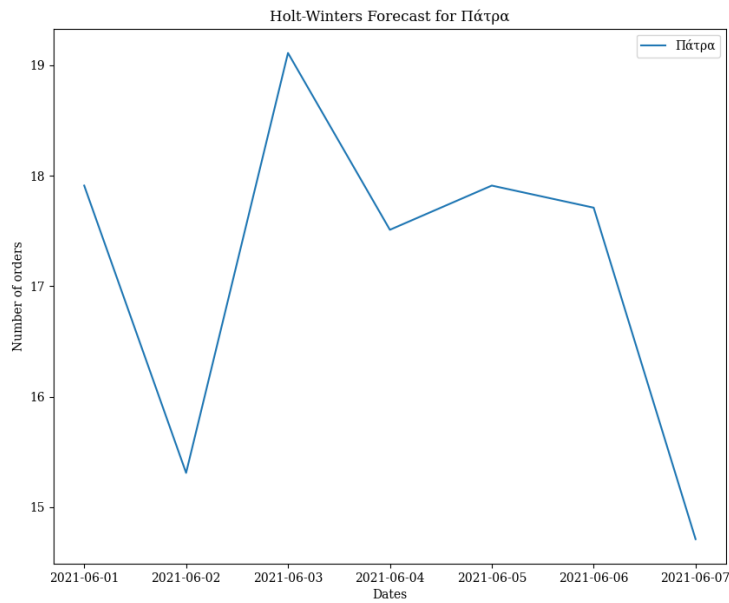
Πίνακας 16 Προβλεπόμενες Πωλήσεις 2ο Επίπεδο Ιεραρχίας (01/06/2021 έως 07/06/2021)



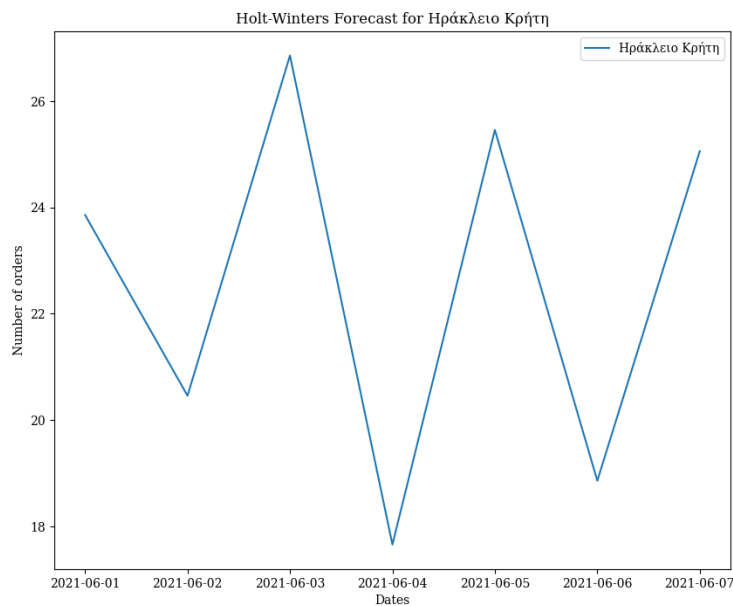
Εικόνα 71 2ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (1)



Εικόνα 72 2ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (2)



Εικόνα 73 2ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (3)



Εικόνα 74 2ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (4)

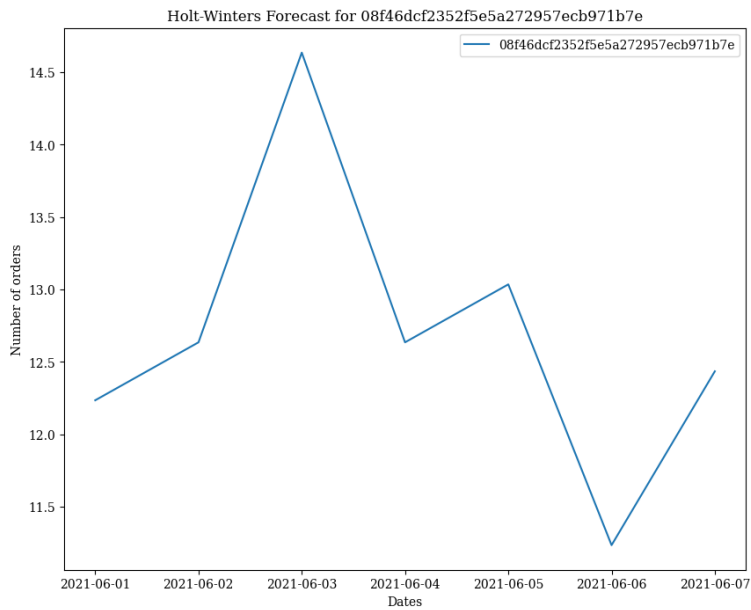
9.3 Πρόβλεψη Πωλήσεων 3ο Επίπεδο Ιεραρχίας

Στον ακόλουθο πίνακα παρουσιάζεται ο αριθμός των προβλεπόμενων παραγγελιών για το 3ο επίπεδο ιεραρχίας, που αφορά στο διάστημα από 01/06/2021 έως 07/06/2021.

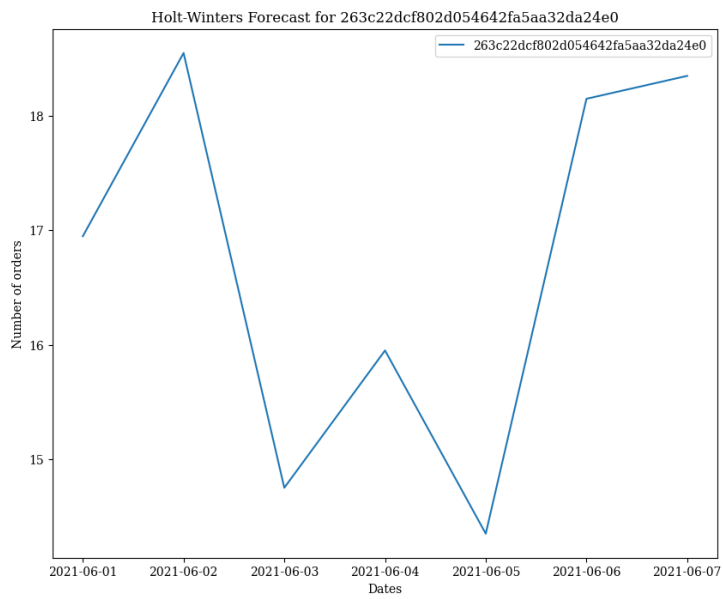
08f46dc	263c22	3b7945	6fd929	8352e6	6afd11	85e5bf2	973b6f	abee831	c3a1e3a	cb8c6b	d119eb	eb160de	f21eb54
f2352f5	dcf802d	935122	3ee5eef	1dee47	dbc5bb	45bf3b	393703	e3f7a8e	b6337b	08adffa	1de89d	3a4e81f	
e5a272	054642f	8b9d82	fee4ab1	247551	890d23f	42c8fc7	b50a66	705c86	cf48b6e	287426	9058fcb	57b313	
957ecb	a5aa32	3b0ee9e	fd2971	8ecafaa	2ab86c	cf58040	997900	d8e4ad	64f9903	36b360	0b968d	16fd2a3	
971b7e	da24e0	40cf952	3b84a4	1	92648a	109868	83a577f	337b39	b4aac2	0	bbbd68	b89cd	
20													
21	12	17	14	18	18	16	16	13	12	14	24	132	14

20	21	-	06	-	01	-	06	-	01	-	06	-	01			
20	21	-	02	13	19	19	19	15	15	18	16	13	16	20	150	19
20	21	-	03	15	15	16	15	19	15	14	19	13	13	27	148	15
20	21	-	04	13	16	17	18	18	14	15	18	11	16	18	151	16
20	21	-	05	13	14	19	17	18	17	14	17	11	14	25	153	15
20	21	-	06	11	18	20	18	18	15	18	17	12	14	19	159	16
20	21	-	07	12	18	18	19	15	11	15	18	12	14	25	112	13

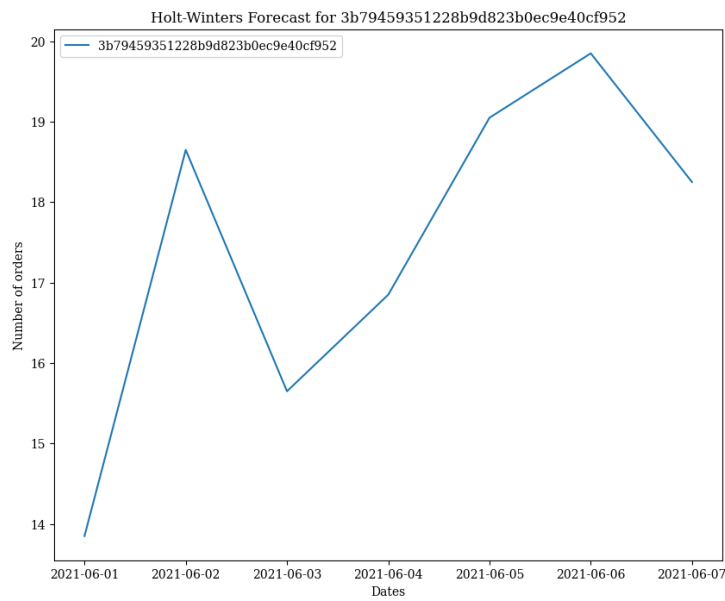
Πίνακας 17 Προβλεπόμενες Πωλήσεις 3^ο Επίπεδο Ιεραρχίας (01/06/2021 έως 07/06/2021)



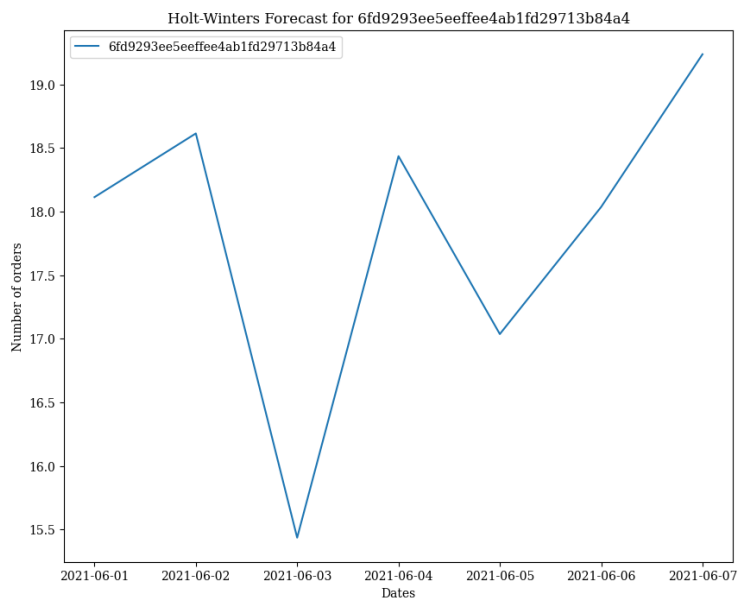
Εικόνα 75 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (1)



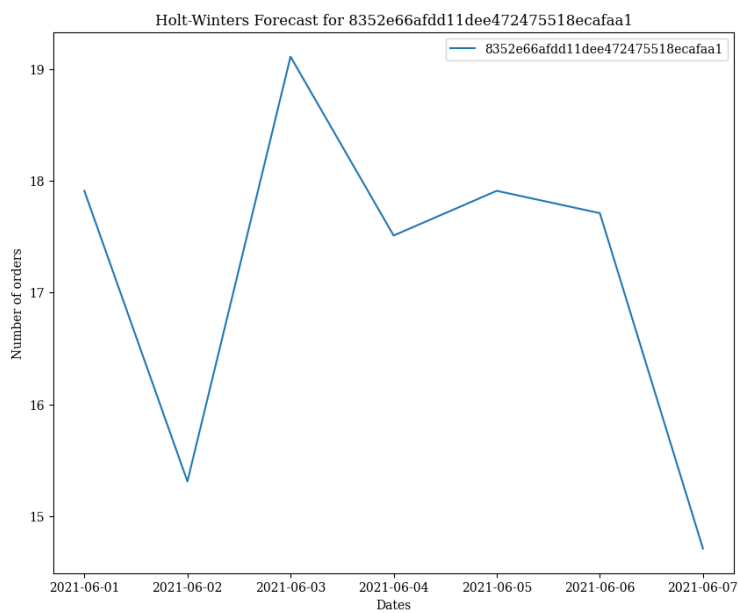
Εικόνα 76 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (2)



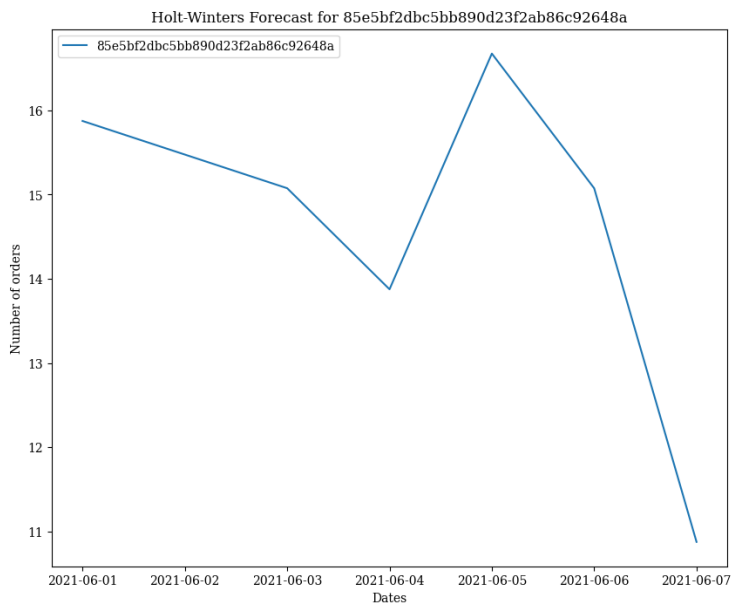
Εικόνα 77 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (3)



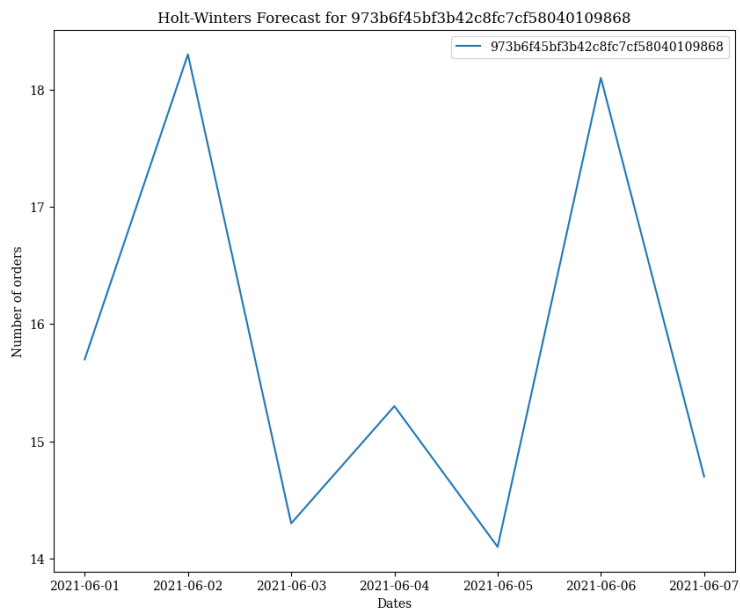
Εικόνα 78 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (4)



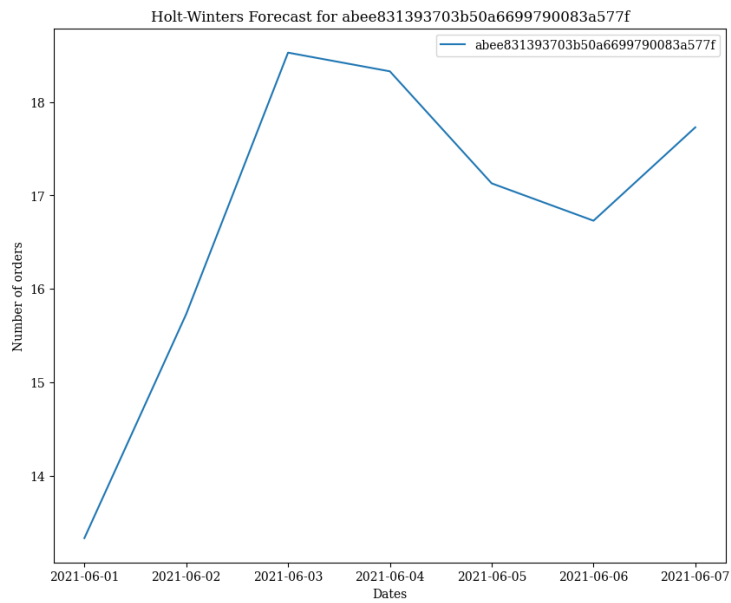
Εικόνα 79 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (5)



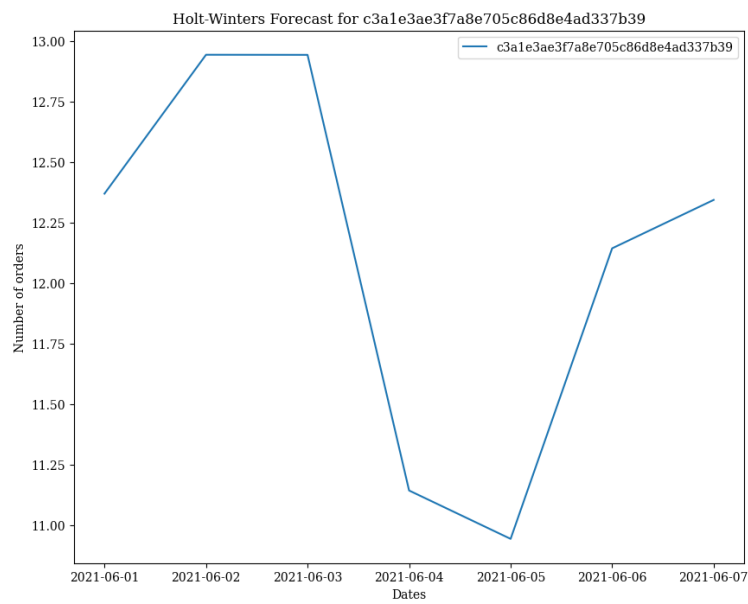
Εικόνα 80 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (6)



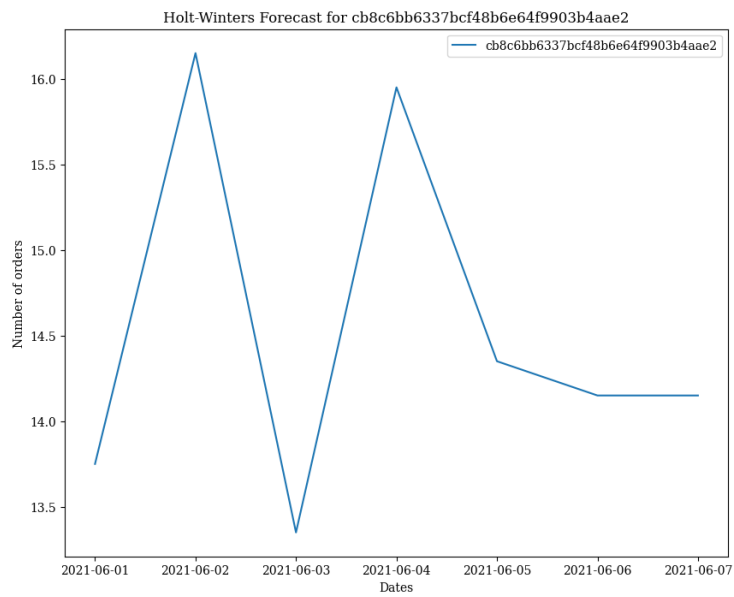
Εικόνα 81 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (7)



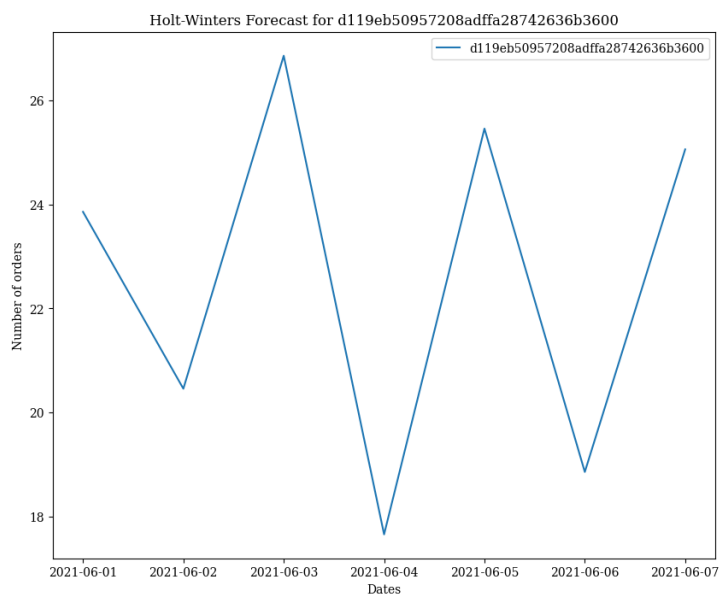
Εικόνα 82 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (8)



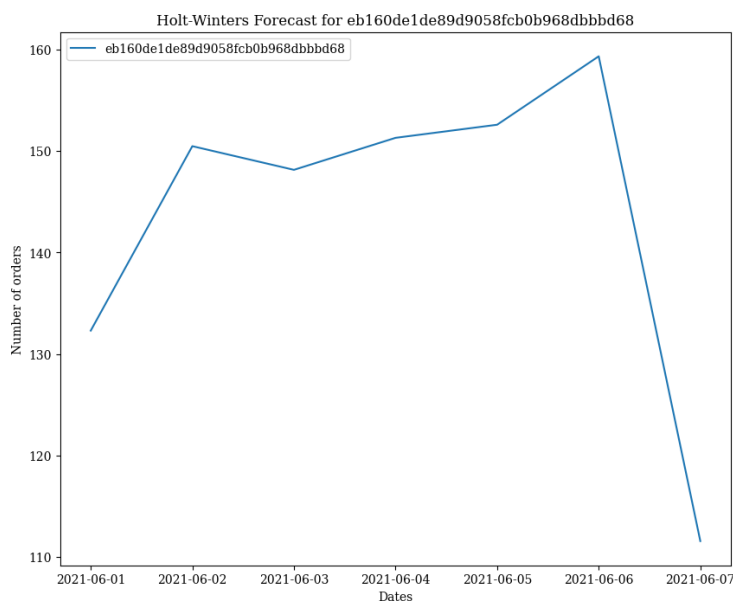
Εικόνα 83 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (9)



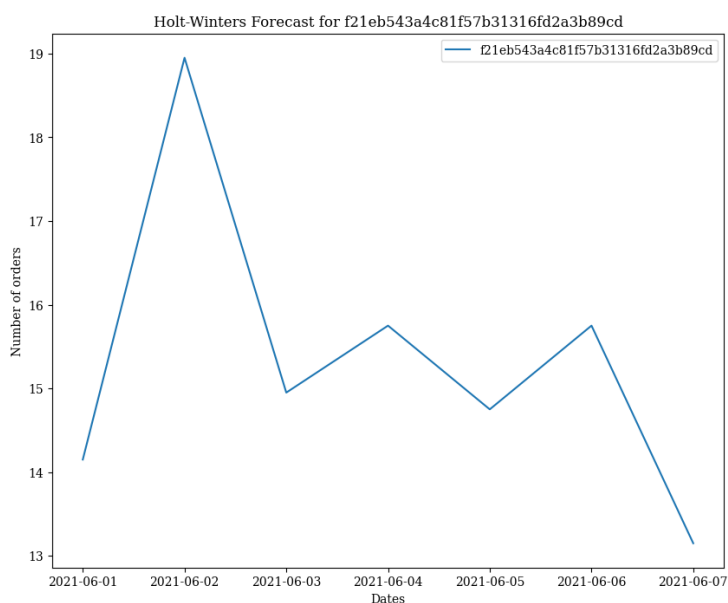
Εικόνα 84 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (10)



Εικόνα 85 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (11)



Εικόνα 86 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (12)



Εικόνα 87 3ο Επίπεδο Ιεραρχίας Πρόβλεψη Πωλήσεων (13)

9.4 Αποτελέσματα

Τα αποτελέσματα της πρόβλεψης πωλήσεων μπορούν να αξιοποιηθούν για την παροχή πληροφοριών σχετικά με τις τάσεις της αγοράς, την αποδοτική διαχείριση των αποθεμάτων και την καλύτερη προετοιμασία για την κάλυψη των αναγκών των πελατών. Σε περιπτώσεις όπου στις προβλέψεις πωλήσεων εμφανίζονται:

1. Υψηλή ζήτηση τα καταστήματα πρέπει να φροντίζουν να έχουν στη διάθεσή τους επαρκή αριθμό αποθεμάτων, έτσι ώστε να είναι σε θέση να καλύψουν τη ζήτηση. Επιπλέον, όταν για την αύξηση των αποθεμάτων απαιτείται και η ενίσχυση της παραγωγής, τότε η πρόσληψη προσωρινού προσωπικού αποτελεί ένα αρκετά βοηθητικό μέτρο. Ακόμη, είναι απαραίτητη και η

βελτιστοποίηση της λειτουργίας και των διαδικασιών των καταστημάτων, για να έχουν τη δυνατότητα να ανταποκριθούν έγκαιρα και αποτελεσματικά στις απαιτήσεις.

2. Χαμηλή ζήτηση τα καταστήματα χρειάζεται να προβούν στη δημιουργία προωθητικών ενεργειών, ή εκπτώσεων, για να αυξήσουν τη ζήτηση. Επιπρόσθετα, είναι απαραίτητη και η διαχείριση των αποθεμάτων, έτσι ώστε να μην έχουν στη διάθεσή τους τεράστιο αριθμό αποθεμάτων, ο οποίος θα επιφέρει έξοδα αποθήκευσης αλλά και κίνδυνο αλλοίωσης τους. Μία ακόμη σημαντική ενέργεια για την αύξηση της ζήτησης προϊόντων είναι η δημιουργία νέων προϊόντων για να προσελκυστούν νέοι πελάτες και να διατηρηθεί το ενδιαφέρον των υφιστάμενων.

10. Συμπεράσματα

Λαμβάνοντας υπόψιν όσα έχουν αναφερθεί στα προηγούμενα κεφάλαια, καθώς και της ανάλυσης που έχει διενεργηθεί στο σύνολο δεδομένων, ένα κοινό συμπέρασμα είναι ότι η ανάλυση δεδομένων και η χρήση μεθόδων και τεχνολογιών που βασίζονται σε αυτή, παρέχει πλήθος πολύτιμων πληροφοριών στις επιχειρήσεις και τους οργανισμούς. Τη γνώση αυτή μπορούν να εκμεταλλευτούν για να βελτιώσουν τις στρατηγικές μάρκετινγκ, την κατανόηση των αναγκών των πελατών και γενικά των καταναλωτών, τη λήψη αποφάσεων, τη βελτίωση των διαδικασιών, την πρόβλεψη της ζήτησης και την αποτελεσματική διαχείριση των πόρων και αποθεμάτων.

Η μέθοδος RFM (Recency, Frequency, Monetary) για την τμηματοποίηση των πελατών της ηλεκτρονικής πλατφόρμας πωλήσεων αποτελεί βασικό συστατικό στοιχείο για την ανάλυση των συνηθειών των καταναλωτών. Η λήψη πληροφοριών για την πιο πρόσφατα ολοκληρωμένη αγορά, τη συχνότητα των αγορών και τη χρηματική αξία των αγορών βοήθησε στον διαχωρισμό των πελατών σε διάφορες ομάδες. Με τον τρόπο αυτό οι επιχειρήσεις λαμβάνουν σημαντική πληροφορία την οποία μπορούν να αξιοποιήσουν για τη δημιουργία στοχευμένων καμπανιών μάρκετινγκ, τη βελτιστοποίηση των στρατηγικών τους, την εύρεση μεθόδων για τη διατήρηση των υφιστάμενων και την επαναφορά των χαμένων πελατών τους, καθώς και την αύξηση της ικανοποίησης των αναγκών των καταναλωτών τους.

Για την πρόβλεψη πωλήσεων που διενεργήθηκε για τα μεγάλα καταστήματα, δηλαδή τα καταστήματα με τον μεγαλύτερο όγκο πωλήσεων, καθώς όσο περισσότερα δεδομένα είναι διαθέσιμα τόσο πιο αξιόπιστη είναι και η πρόβλεψη, αξιοποιήθηκε το μοντέλο Holt-Winters. Στο μοντέλο η ανάλυση βασίζεται στην εποχικότητα και τις τάσεις που εμφανίζονται στο σύνολο δεδομένων που εξετάζεται. Σε αρκετά από τα καταστήματα που εξετάστηκαν παρατηρήθηκε αυξομείωση των παραγγελιών, γεγονός που υποδηλώνει την ύπαρξη προωθητικών ενεργειών, ειδικών προσφορών ή άλλων παραγόντων που επηρεάζουν τις συνήθειες των πελατών. Τα δεδομένα αυτά είναι πολύ χρήσιμα για τις επιχειρήσεις, καθώς έχουν τη δυνατότητα να προσαρμόσουν την παραγωγή και τα αποθέματά τους, για να καλύψουν τις επερχόμενες ανάγκες των πελατών τους. Επίσης, λαμβάνουν πληροφορίες για τις αποτυχημένες ή επιτυχημένες προωθητικές ενέργειες και τις προσφορές. Συνεπώς, μπορούν να δημιουργήσουν πιο αποτελεσματικές στρατηγικές στόχευσης των πελατών τους.

Η αξιοποίηση των Big Data, της τμηματοποίησης RFM και της πρόβλεψης πωλήσεων με το μοντέλο Holt-Winters είναι βέβαιο ότι παρέχει στις επιχειρήσεις ένα σημαντικό εργαλείο για τη λήψη ενημερωμένων και βέλτιστων αποφάσεων. Η ανάλυση μεγάλων όγκων δεδομένων συμβάλλει στην εξαγωγή χρήσιμων συμπερασμάτων τα οποία δεν θα ήταν εμφανή με τις παραδοσιακές μεθόδους ανάλυσης. Επιπρόσθετα, η ευελιξία και η ακρίβεια που προσφέρουν οι συγκεκριμένες τεχνικές και μέθοδοι βελτιώνουν σε μεγάλο βαθμό την ανταγωνιστικότητα των επιχειρήσεων.

Τα αποτελέσματα της ανάλυσης οδηγούν στο συμπέρασμα ότι η χρήση τέτοιου είδους μεθόδων και τεχνολογιών από τις επιχειρήσεις είναι εφικτό να οδηγήσει σε σημαντική αύξηση των πωλήσεων και της κερδοφορίας. Η δυνατότητα πρόβλεψης μελλοντικών τάσεων και η κατανόηση των αναγκών των πελατών βοηθούν στην ανάπτυξη πιο αποδοτικών στρατηγικών μάρκετινγκ και διαχείρισης αποθεμάτων. Όλα αυτά έχουν ως αποτέλεσμα, οι επιχειρήσεις που αξιοποιούν τις συγκεκριμένες τεχνολογίες να μπορούν να εκμεταλλευτούν έγκαιρα και αποτελεσματικά τις ευκαιρίες της αγοράς και να αντιμετωπίσουν τις προκλήσεις.

Τέλος, βάσει όσων αναφέρθηκαν οι τεχνικές και μεθοδολογίες των Big Data είναι απαραίτητα εργαλεία για τις επιχειρήσεις και την αποδοτική λειτουργία τους. Η ορθή χρήση τέτοιου είδους μεθόδων μπορεί να οδηγήσει σε βελτιωμένες επιχειρηματικές διαδικασίες, αυξημένη ικανοποίηση των πελατών και απόκτηση ανταγωνιστικού πλεονεκτήματος. Συνεπώς, προτείνεται στις επιχειρήσεις να επενδύουν σε τεχνολογίες και μεθόδους που διαχειρίζονται και αναλύουν μεγάλους όγκους δεδομένων και να υιοθετούν στις στρατηγικές τους τις διαθέσιμες πρακτικές, έτσι ώστε να επιτυγχάνουν μακροπρόθεσμη ανάπτυξη και επιτυχία.

Βιβλιογραφία

1. Statista, 2023, Amount of data created, consumed, and stored worldwide from 2010 to 2025, <https://www.statista.com/statistics/871513/worldwide-data-created/>
2. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
3. Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77-84.
4. Evans, J. R. (2017). *Business analytics: Methods, models, and decisions*. Pearson.
5. Chaffey, D. (2015). *Digital Business and E-Commerce Management*. Pearson.
6. Deloitte. (2020). *The Future of Retail: From Doing Digital to Being Digital*. Retrieved from Deloitte.
7. Kim, A. J., & Peterson, R. A. (2017). "A Meta-analysis of Online Trust Relationships in E-commerce." *Journal of Interactive Marketing*, 38, 44-54.
8. Laudon, K. C., & Traver, C. G. (2019). *E-Commerce 2019: Business, Technology, and Society*. Pearson.
9. Turban, E., Outland, J., King, D., Lee, J. K., Liang, T. P., & Turban, D. C. (2018). *Electronic Commerce 2018: A Managerial and Social Networks Perspective*. Springer.
10. Kotler, P., & Keller, K. L. (2016). *Marketing Management*. Pearson.
11. Wedel, M., & Kamakura, W. A. (2000). *Market Segmentation: Conceptual and Methodological Foundations*. Springer Science & Business Media.
12. Yankelovich, D., & Meer, D. (2006). "Rediscovering Market Segmentation." *Harvard Business Review*, 84(2), 122-131.
13. Lamb, C. W., Hair, J. F., & McDaniel, C. (2011). *Marketing*, 11th edition. Mason: SouthWestern Cengage Learning.
14. Tsipstsis, K., & Chorianopoulos, A. (2009). *Data Mining Techniques in CRM: Inside Customer Segmentation*. West Sussex: Wiley.
15. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control*. Wiley.
16. Chatfield, C. (2004). *The Analysis of Time Series: An Introduction*. CRC Press
17. Kendall, M. G., & Ord, J. K. (1990). *Time Series*. Edward Arnold.
18. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
19. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and Applications*. John Wiley & Sons.
20. Gilliland, M. (2010). *The Business Forecasting Deal: Exposing Myths, Eliminating Bad Practices, Providing Practical Solutions*. John Wiley & Sons.
21. Armstrong, J. S. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Springer.
22. Hsu, C. C., & Sandford, B. A. (2007). The Delphi Technique: Making Sense Of Consensus. *Practical Assessment, Research, and Evaluation*, 12(1), 10.
23. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and Applications*. John Wiley & Sons.
24. Chatfield, C. (2000). *Time-Series Forecasting*. CRC Press.
25. Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting*, 15(4), 353-375.
26. Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
27. Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1), 5-10.
28. Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, 4(1), 1-28.