

ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΤΜΗΜΑΤΟΣ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

**ΑΝΑΠΤΥΞΗ ΜΟΝΤΕΛΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ ΠΡΟΒΛΕΨΗ
ΠΩΛΗΣΕΩΝ**

Διπλωματική Εργασία

του

Κωνσταντίνου Γρηγοριάδη

Θεσσαλονίκη, Ιούνιος 2023

**ΑΝΑΠΤΥΞΗ ΜΟΝΤΕΛΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ ΠΡΟΒΛΕΨΗ
ΠΩΛΗΣΕΩΝ**

Κωνσταντίνος Γρηγοριάδης

Πτυχίο Μηχανικών Πληροφορικής Τεχνολογικής Εκπαίδευσης Πανεπιστημίου Δυτικής
Μακεδονίας

Διπλωματική Εργασία

υποβαλλόμενη για τη μερική εκπλήρωση των απαιτήσεων του

**ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΤΙΤΛΟΥ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ
ΠΛΗΡΟΦΟΡΙΚΗ**

Επιβλέπων Καθηγητής
Ιωάννης Ρεφανίδης

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 20/06/23

Ιωάννης Ρεφανίδης
Καθηγητής

Ευτύχης Πρωτοπαπαδάκης
Επίκουρος Καθηγητής

Ηλίας Σακελλαρίου
Επίκουρος Καθηγητής

.....

.....

.....

Κωνσταντίνος Γρηγοριάδης

.....

Περίληψη

Από την αρχαιότητα το να γνωρίζει κανείς τα «μελλούμενα» ήταν ένα μεγάλο πλεονέκτημα. Με την ανάπτυξη των μαθηματικών και κυρίως των υπολογιστών, σήμερα υπάρχει η δυνατότητα να «προβλέψει» κανείς, με επιστημονικό τρόπο, το πως θα εξελιχθούν κάποια πράγματα. Σε μια επιχείρηση βέβαια εκείνο που ενδιαφέρει περισσότερο είναι οι πωλήσεις που μπορεί να πετύχει.

Στην παρούσα εργασία γράφτηκαν κώδικες (σε γλώσσα Python) για την πρόβλεψη των πωλήσεων της εταιρείας IC:Company. Τα στοιχεία που είναι διαθέσιμα από την εταιρεία περιλαμβάνουν πωλήσεις για 34 μήνες (από τον Ιανουάριο του 2013 μέχρι και τον Οκτώβριο του 2015) 22.170 διαφορετικών προϊόντων χωρισμένων σε 84 κατηγορίες που πωλούνται σε 60 διαφορετικά σημεία.

Τα στοιχεία αυτά ελέγχθηκαν ως προς την συνέπεια τους.

Οι μέθοδοι πρόβλεψης που χρησιμοποιήθηκαν ήταν η Μέθοδος πρόβλεψης τελευταίας τιμής (σαν σύγκριση), η Απλή γραμμική παλινδρόμηση (με χρήση 2 έως 4 προηγούμενων μηνών) και η Τετραγωνική παλινδρόμηση (με χρήση 4 έως 8 μηνών).

Τα δεδομένα εξετάστηκαν και για την εμφάνιση εποχικότητας (αυτοσυσχέτιση) έτσι οι προβλέψεις που έγιναν συγκρίθηκαν με και χωρίς την χρήση εποχικότητας πριν εξαχθούν τα τελικά συμπεράσματα.

Επίσης δημιουργήθηκαν κώδικες που αξιολογούν αυτές τις μεθόδους κάνοντας συγκρίσεις μεταξύ τους. Η σύγκριση γίνεται με την χρήση του Μέσου Απόλυτου Ποσοστιαίου σφάλματος (MAPE). Η αξιολόγηση βέβαια περιλαμβάνει συντελεστές βαρύτητας που λαμβάνουν υπόψη και τις εισπράξεις.

Όλοι οι κώδικες προσαρμόζονται αυτόματα αν η εταιρεία προσθέσει και άλλα στοιχεία στα δεδομένα της. Οι προβλέψεις μπορεί να αφορούν στον επόμενο μήνα αλλά και σε περισσότερους μήνες με επιλογή του χρήστη.

Κατά την διάρκεια της εργασίας έγιναν δοκιμές ταξινομώντας τα δεδομένα με διαφορετικό τρόπο όπως για παράδειγμα να χρησιμοποιούνται τα στοιχεία των 24 μηνών για την πρόβλεψη των υπολοίπων, καθώς και κάποιες άλλες επιπλέον δυνατότητες.

Τέλος από την ανάλυση των προϊόντων καθώς και από τις δοκιμές που έγιναν προέκυψαν κάποια συμπεράσματα που μπορούν να βοηθήσουν ένα μελλοντικό ερευνητή να οδηγηθεί σε ακόμα καλύτερες προβλέψεις.

Λέξεις Κλειδιά: Πρόβλεψη, παλινδρόμηση, αυτοσυσχέτιση, εποχικότητα, Μέσο απόλυτο ποσοστιαίο σφάλμα, συντελεστές βαρύτητας.

Abstract

Since ancient times, knowing the "future" has been a great advantage. With the development of mathematics and especially computers, it is now possible to "predict", in a scientific way, how things will turn out. In a business, of course, what matters most is the sales it can achieve.

In this paper codes (in Python language) were written to forecast the sales of 1C:Company. The data available from the company include sales for 34 months (from January 2013 to October 2015) of 22,170 different products divided into 84 categories sold in 60 different locations.

These data were checked for consistency.

The forecasting methods used were the Last Value Forecast Method (as a comparison), Simple Linear Regression (using 2 to 4 previous months) and Quadratic Regression (using 4 to 8 months).

The data were also examined for the occurrence of seasonality (autocorrelation) so the predictions made were compared with and without the use of seasonality before final conclusions were drawn.

Codes were also created to evaluate these methods by making comparisons between them. The comparison is made using the Mean Absolute Percentage Error (MAPE). The assessment of course includes weighting factors that also take into account receipts.

All codes are automatically adjusted if the company adds other elements to its data. The predictions can be for the next month or for several months at the user's option.

During the work, tests were made by sorting the data in different ways, such as using the 24 months of data to predict the rest, and some other additional possibilities.

Finally, the analysis of the products and the tests that were carried out led to some conclusions that can help a future researcher to make even better predictions.

Keywords: Prediction, regression, autocorrelation, seasonality, mean absolute percentage error, weighting coefficients.

Πρόλογος – Ευχαριστίες

Ευχαριστώ τον Καθηγητή κ. Ιωάννη Ρεφανίδη για τις πολύτιμες συμβουλές του κατά την εκπόνηση αυτής της Διπλωματικής Εργασίας. Καθώς επίσης ευχαριστώ και τους γονείς μου Ιωάννη Γρηγοριάδη και Ζόμπολα Μαριάννα για την ηθική τους στήριξη και κατανόηση.

Περιεχόμενα

Περιεχόμενα	vii
Κατάλογος Εικόνων	xii
Κατάλογος Πινάκων	xiv
Συντομογραφίες	xv
1 Εισαγωγή	1
1.1 Προβλέψεις	1
1.2 Προβλέψεις πωλήσεων	1
1.3 Χαρακτηριστικά των προβλέψεων	1
1.4 Μέθοδοι πρόβλεψης	2
1.5 Η γλώσσα Python	3
1.6 Διάρθρωση της μελέτης	4
2 Μέθοδοι παλινδρόμησης - σφάλματα	6
2.1 Παλινδρόμηση	6
2.1.1 Γενικά	6
2.1.2 Απλή Γραμμική παλινδρόμηση	6
2.1.3 Τετραγωνική παλινδρόμηση	8
2.1.4 Άλλες μορφές παλινδρόμησης	9
2.2 Τάση και Εποχικότητα (Trend, seasonality)	10
2.3 Αυτοσυσχέτιση (autocorrelation)	11
2.3.1 Γενικά	11
2.3.2 Τάση και εποχικότητα σε διαγράμματα ACF	12
2.4 Σφάλματα κατά την Πρόβλεψη	13
3 Περιγραφή δεδομένων, στατιστικά	14
3.1 Η Εταιρεία	14
3.2 Τα πρωτογενή αρχεία	14
4 Μεθοδολογία, διαδικασία εξαγωγής αποτελεσμάτων	22
4.1 Γενικά	22
4.2 Έλεγχοι που έγιναν στα δεδομένα	22
4.2.1 Γενικά	22
4.2.2 Έλεγχος ύπαρξης κενών στο αρχείο των προϊόντων	22
4.2.3 Έλεγχος ύπαρξης κενών στο αρχείο των κατηγοριών	22
4.3 Επεξεργασία πρωτογενών αρχείων.	22

4.3.1	Γενικά	22
4.3.2	Εξαγωγή πωλήσεων ανά προϊόν και μήνα	23
4.3.3	Εξαγωγή πωλήσεων ανά προϊόν για τον τελευταίο μήνα	24
4.3.4	Εξαγωγή πωλήσεων ανά κατηγορία και μήνα	24
4.3.5	Διάρκεια ζωής των προϊόντων	24
4.3.6	Διάρκεια ζωής των κατηγοριών	25
4.4	Προβλέψεις πωλήσεων για τις κατηγορίες	26
4.4.1	Γενικά	26
4.4.2	Μέθοδος πρόβλεψης τελευταίας αξίας (LVFM)	26
4.4.3	Απλή Γραμμική Παλινδρόμηση (SLR)	27
4.4.4	Τετραγωνική παλινδρόμηση (QR)	28
4.5	Εποχικότητα	29
4.5.1	Γενικά	29
4.5.2	Αυτοσυσχέτιση	30
4.5.3	Εφαρμογή εποχικότητας	32
4.5.3.1	Γενικά	32
4.5.3.2	Γραμμική παλινδρόμηση (SLR) με εποχικότητα	33
4.5.3.3	Τετραγωνική παλινδρόμηση (QR) με εποχικότητα	33
4.6	Μέσο ποσοστιαίο σφάλμα (MAPE)	34
4.6.1	Γενικά	34
4.6.2	Συντελεστές Βαρύτητας	35
4.6.3	Υπολογισμός MAPE για τις προβλέψεις στις κατηγορίες	37
4.6.3.1	Μέσες τιμές MAPE για προβλέψεις χωρίς εποχικότητα	37
4.6.3.2	Μέσες τιμές MAPE για προβλέψεις με εποχικότητα	37
4.6.3.3	Μέσες τιμές MAPE με βαρύτητα 1 χωρίς εποχικότητα	38
4.6.3.4	Μέσες τιμές MAPE με βαρύτητα 1 και εποχικότητα	38
4.6.3.5	Μέσες τιμές MAPE με βαρύτητα 2 χωρίς εποχικότητα	39
4.6.3.6	Μέσες τιμές MAPE με βαρύτητα 2 και εποχικότητα	39
4.6.3.7	Απλός μέσος όρος για όλους τους μήνες	40
4.6.3.8	Σταθμισμένος μέσος όρος για όλους τους μήνες	41
4.7	Προβλέψεις πωλήσεων για τα προϊόντα	42
4.7.1	Γενικά	42
4.7.2	Εφαρμογή της LVFM για τα προϊόντα	42

4.7.3 Εφαρμογή της SLR4 για τα προϊόντα	43
4.7.4 MAPE για LVFM και SLR4 χωρίς βαρύτητα	43
4.7.5 MAPE για LVFM και SLR4 με βαρύτητα 1	43
4.7.6 MAPE για LVFM και SLR4 με βαρύτητα 2	44
4.7.7 Απλό μέσο MAPE LVFM και SLR4 χωρίς βαρύτητα	44
4.7.8 Απλό μέσο MAPE LVFM και SLR4 με βαρύτητα 1	44
4.7.9 Απλό μέσο MAPE LVFM και SLR4 με βαρύτητα 2	44
4.7.10 Σταθμισμένο MAPE LVFM και SRL4 χωρίς βαρύτητα	45
4.7.11 Σταθμισμένο MAPE LVFM και SLR4 με βαρύτητα 1	45
4.7.12 Σταθμισμένο MAPE LVFM και FLR4 με βαρύτητα 2	45
4.8 Επιπλέον δυνατότητες	45
4.8.1 Γενικά	45
4.8.2 Πλήθος προϊόντων ανά κατηγορία	46
4.8.3 Δημιουργία αρχείων επικεφαλίδων	46
4.8.4 Πρόβλεψη με καθυστέρηση	46
4.8.4.1 Γενικά	46
4.8.4.2 SLR με καθυστέρηση (delay) για κάθε προϊόν	46
4.8.4.3 SLR με καθυστέρηση (delay) για κάθε κατηγορία	47
4.8.4.4 SLR με καθυστέρηση και εποχικότητα για κάθε προϊόν	47
4.8.4.5 SLR με καθυστέρηση και εποχικότητα για κατηγορίες	48
4.8.4.6 QR με καθυστέρηση για κάθε προϊόν	48
4.8.4.7 QR με καθυστέρηση για τις κατηγορίες	49
4.8.4.8 QR με καθυστέρηση και εποχικότητα για κάθε προϊόν	49
4.8.4.9 QR με καθυστέρηση και εποχικότητα για κατηγορίες	50
4.8.5 MAPE για προβλέψεις με καθυστέρηση	51
4.8.5.1 Γενικά	51
4.8.5.2 MAPE για προβλέψεις προϊόντων με καθυστέρηση	51
4.8.5.3 MAPE για προϊόντα με καθυστέρηση και εποχικότητα	51
4.8.5.4 MAPE για κατηγορίες με καθυστέρηση	52
4.8.5.5 MAPE κατηγορίες με καθυστέρηση και εποχικότητα	52
4.8.6 Συνάρτηση αυτοσυσχέτισης για επιλεγμένη κατηγορία	53
4.8.7 Συνάρτηση αυτοσυσχέτισης για επιλεγμένο προϊόν	53
4.8.8 Προβλέψεις με δεδομένα 24 μηνών για προϊόντα	54

Ανάπτυξη μοντέλων μηχανικής μάθησης για πρόβλεψη πωλήσεων	
4.8.8.1 Γενικά	54
4.8.8.2 SLR2 – SLR12 με χρήση 24 μηνών χωρίς εποχικότητα	55
4.8.8.3 SLR2 – SLR12 με χρήση 24 μηνών και εποχικότητα	56
4.8.8.4 QR3 – QR12 με χρήση 24 μηνών χωρίς εποχικότητα	57
4.8.8.5 QR3 – QR12 με χρήση 24 μηνών και εποχικότητα	57
4.8.9 MAPE για προβλέψεις με χρήση 24 μηνών	58
4.8.9.1 MAPE με χρήση 24 μηνών χωρίς εποχικότητα	58
4.8.9.2 MAPE με χρήση 24 μηνών και εποχικότητα	59
4.8.10 Προβλέψεις με χρήση 24 μηνών για κατηγορίες	60
4.8.10.1 Γενικά	60
4.8.10.2 SLR2 – SLR12 με χρήση 24 μηνών χωρίς εποχικότητα	61
4.8.10.3 SLR2 – SLR12 με χρήση 24 μηνών και εποχικότητα	61
4.8.10.4 QR3 – QR12 με χρήση 24 μηνών χωρίς εποχικότητα	62
4.8.10.5 QR3 – QR12 χρήση 24 μηνών και εποχικότητα	62
4.8.10.6 MAPE με χρήση 24 μηνών χωρίς εποχικότητα	63
4.8.10.7 MAPE με χρήση 24 μηνών και εποχικότητα	64
4.9 Εισπράξεις	65
4.9.1 Συνολικές εισπράξεις ανά μήνα	65
4.9.2 Συνολικές εισπράξεις ανά κατηγορία και μήνα	66
4.9.3 Μέση τιμή εισπραξης ανά προϊόν και μήνα	66
5 Επίλογος	68
5.1 Σύνοψη και συμπεράσματα	68
5.1.1 Διάρκεια ζωής προϊόντων και κατηγοριών	68
5.1.2 Αυτοσυσχέτιση και εποχικότητα	69
5.1.3 Συντελεστές βαρύτητας στον υπολογισμό του MAPE	70
5.1.4 Κυκλικότητα	71
5.1.5 Περιορισμοί στις προβλέψεις	72
5.1.6 MAPE για τις κατηγορίες	73
5.1.6.1 Αναλυτικά MAPE για κάθε μήνα και μέθοδο	73
5.1.6.2 Απλοί μέσοι όροι MAPE για κάθε μέθοδο	75
5.1.6.3 Σταθμισμένο μέσο MAPE για κάθε μέθοδο	76
5.1.6.4 Προβλέψεις για τις κατηγορίες με LVFM και SLR4	77
5.1.7 MAPE για τα προϊόντα	78

5.1.7.1 Αναλυτικά MAPE για κάθε μήνα και μέθοδο	78
5.1.7.2 Προβλέψεις για τα προϊόντα με SLR4	81
5.1.8 Τελικά συμπεράσματα	83
5.2 Όρια και περιορισμοί της έρευνας	83
5.3 Μελλοντικές Επεκτάσεις	84
6 Βιβλιογραφία	87

Κατάλογος Εικόνων

Εικόνα 1.5-1 Εγκατάσταση της Python	4
Εικόνα 1.5-2 Ολοκλήρωση της διαδικασίας.....	4
Εικόνα 2.1-1 Ευθεία ελαχίστων τετραγώνων	7
Εικόνα 2.2-1 Πωλήσεις για την κατηγορία 12 (κονσόλες PS4).....	10
Εικόνα 2.2-2 Πωλήσεις για κατηγορία 3 (αξεσουάρ για κονσόλες PS4)	11
Εικόνα 2.3-1 Συνάρτηση αυτοσυσχέτισης (ACF) για τις συνολικές πωλήσεις.....	12
Εικόνα 3.2-1 Σύγκριση κατηγοριών 8 και 80	20
Εικόνα 4.3-1 Πωλήσεις για την κατηγορία 39 (περιορισμένη διάρκεια ζωής)	26
Εικόνα 4.4-1 Προβλέψεις πωλήσεων για την κατηγορία 41 με LVFM.....	27
Εικόνα 4.4-2 Προβλέψεις πωλήσεων για την κατηγορία 6 με SLR 2,3,4	28
Εικόνα 4.4-3 Προβλέψεις πωλήσεων για την κατηγορία 2 με QR 4,5,6,7,8.....	29
Εικόνα 4.5-1 Συνολικές πωλήσεις ανά μήνα	30
Εικόνα 4.5-2 Πωλήσεις για την κατηγορία 25.....	31
Εικόνα 4.5-3 Συνάρτηση αυτοσυσχέτισης για την κατηγορία 25	31
Εικόνα 4.5-4 Πωλήσεις για την κατηγορία 21.....	31
Εικόνα 4.5-5 Συνάρτηση αυτοσυσχέτισης για την κατηγορία 21	32
Εικόνα 4.5-6 Προβλέψεις πωλήσεων για την κατηγορία 6 με SLR 2,3,4	33
Εικόνα 4.5-7 Προβλέψεις πωλήσεων για την κατηγορία 2 με QR 4,5,6,7,8.....	34
Εικόνα 4.8-1 Πρόβλεψη πωλήσεων για την κατηγορία 73 με SLR4 και delay=3	48
Εικόνα 4.8-2 Πρόβλεψη πωλήσεων για την κατηγορία 40 με QR8 και delay=3.....	50
Εικόνα 4.8-3 Συνάρτηση ACF για την κατηγορία 63 (Δώρα – Μαζεμένα παιχνίδια).....	53
Εικόνα 4.8-4 Συνάρτηση ACF για το προϊόν 5872 (μπάλα γυμναστικής)	54
Εικόνα 4.8-5 Πρόβλεψη για το προϊόν 4181 της κατηγορίας 75 (προγράμματα)	57
Εικόνα 4.8-6 Πρόβλεψη για το προϊόν 4391 της κατηγορίας 30 (Παιχνίδια PC)	58
Εικόνα 4.8-7 Πρόβλεψη για κατηγορία 5 με χρήση 24 μηνών και SLR12.....	62
Εικόνα 4.8-8 Πρόβλεψη για κατηγορία 4 με χρήση 24 μηνών και QR12.....	63
Εικόνα 4.9-1 Εισπράξεις ανά μήνα	65
Εικόνα 5.1-1 Στοιχεία για την διάρκεια ζωής των προϊόντων	68
Εικόνα 5.1-2 Στοιχεία για την διάρκεια ζωής των κατηγοριών.....	68
Εικόνα 5.1-3 Προβλέψεις (SLR4) με και χωρίς εποχικότητα για την κατηγορία 49.....	70
Εικόνα 5.1-4 Σύγκριση πωλήσεων κατηγοριών 12 και 58	70

<i>Εικόνα 5.1-5 Σύγκριση εισπράξεων κατηγοριών 12 και 58</i>	71
<i>Εικόνα 5.1-6 Μέσο MAPE με εποχικότητα και βαρύτητα 2 (LVFM και SLR 2,3,4)</i>	74
<i>Εικόνα 5.1-7 Μέσο MAPE με εποχικότητα και βαρύτητα 2 (QR 4,5,6,7,8)</i>	74
<i>Εικόνα 5.1-8 Προβλέψεις με LVFM και SLR4 για κατηγορία 43</i>	77
<i>Εικόνα 5.1-9 Προβλέψεις με LVFM και SLR4 για κατηγορία 55</i>	77
<i>Εικόνα 5.1-10 Προβλέψεις με LVFM και SLR4 για κατηγορία 32</i>	78
<i>Εικόνα 5.1-11 Προβλέψεις με LVFM και SLR4 για κατηγορία 37</i>	78
<i>Εικόνα 5.1-12 Σύγκριση MAPE για LVFM με διαφορετικούς συντελεστές βαρύτητας</i>	80
<i>Εικόνα 5.1-13 Πωλήσεις προϊόντων 10207, 10209,10210, 10212</i>	81
<i>Εικόνα 5.1-14 Προβλέψεις για το προϊόν 10210 με LVFM και SLR2,3,4</i>	81
<i>Εικόνα 5.1-15 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 12677</i>	81
<i>Εικόνα 5.1-16 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 5380</i>	82
<i>Εικόνα 5.1-17 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 17717</i>	82
<i>Εικόνα 5.1-18 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 491</i>	82
<i>Εικόνα 5.1-19 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 3280</i>	82
<i>Εικόνα 5.1-20 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 1905</i>	83
<i>Εικόνα 5.3-1 Μοντελοποίηση με QR12 στην κατηγορία 62</i>	84
<i>Εικόνα 5.3-2 Μοντελοποίηση με QR12 της κατηγορίας 19</i>	85
<i>Εικόνα 5.3-3 Σύγκριση κατηγοριών 64, 65, 69 και 70 (δώρα – παιχνίδια)</i>	85

Κατάλογος Πινάκων

<i>Πίνακας 2.1-1 Άλλες μορφές παλινδρόμησης</i>	9
<i>Πίνακας 3.2-1 Απόσπασμα πρωτογενούς αρχείου πωλήσεων ανά ημέρα</i>	15
<i>Πίνακας 3.2-2 Εμφάνιση πρωτογενούς αρχείου πωλήσεων ανά ημέρα στο Excel</i>	16
<i>Πίνακας 3.2-3 Απόσπασμα ονομαστικού καταλόγου προϊόντων</i>	17
<i>Πίνακας 3.2-4 Κατάλογος κατηγοριών προϊόντων</i>	18
<i>Πίνακας 3.2-5 Κατάλογος σημείων πώλησης</i>	19
<i>Πίνακας 3.2-6 Σύγκριση κατηγοριών προϊόντων 8 και 80</i>	20
<i>Πίνακας 4.3-1 Απόσπασμα με διάρκειες ζωής προϊόντων</i>	25
<i>Πίνακας 4.6-1 Μηδενικές πωλήσεις, μη μηδενικές προβλέψεις</i> Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.	
<i>Πίνακας 4.9-1 Συνολικές εισπράξεις ανά κατηγορία</i>	66
<i>Πίνακας 4.9-2 Μέσες τιμές είσπραξης ανά προϊόν και μήνα</i>	66
<i>Πίνακας 5.1-1 Λευκός θόρυβος (True)</i>	69
<i>Πίνακας 5.1-2 Κατηγορίες χωρίς εποχικότητα με βάση την αυτοσυσχέτιση</i>	69
<i>Πίνακας 5.1-3 Διάρκεια ζωής προϊόντων</i>	71
<i>Πίνακας 5.1-4 Για πόσες κατηγορίες δεν μπορεί να γίνει πρόβλεψη</i>	72
<i>Πίνακας 5.1-5 MAPE % για τις κατηγορίες με βαρύτητα 2 και εποχικότητα</i>	73
<i>Πίνακας 5.1-6 Απλοί μέσοι όροι MAPE % για τις κατηγορίες</i>	75
<i>Πίνακας 5.1-7 Σταθμισμένοι μέσοι όροι MAPE % για τις κατηγορίες</i>	76
<i>Πίνακας 5.1-8 MAPE για τα προϊόντα</i>	79

Συντομογραφίες

SLR: Simple Linear Regression (Απλή γραμμική παλινδρόμηση)

SLR_n: όπου n ακέραιος αριθμός. Απλή γραμμική παλινδρόμηση με χρήση δεδομένων από n προηγούμενες περιόδους.

QR: Quadratic Regression (Τετραγωνική παλινδρόμηση)

QR_n: όπου n ακέραιος αριθμός. Τετραγωνική παλινδρόμηση με χρήση n δεδομένων από n προηγούμενες περιόδους.

ACF: Auto Correlation Function – συνάρτηση αυτοσυσχέτισης

LVFM: Last Value Forecasting Method (Μέθοδος Πρόβλεψης Τελευταίας Αξίας)

MAD: Mean Absolute Deviation (μέση απόλυτη απόκλιση)

MSE: Mean Square Error (Το μέσο τετραγωνικό σφάλμα)

RMSE: Root Mean Square Error Ρίζα του μέσου τετραγωνικού σφάλματος)

MAPE: Mean Average Percentage Error ή MAPE (Μέσο απόλυτο ποσοστιαίο σφάλμα)

1 Εισαγωγή

1.1 Προβλέψεις

Από την αρχαιότητα το να γνωρίζει κανείς τα «μελλούμενα» ήταν ένα πολύ μεγάλο πλεονέκτημα. Από τους πρωτόγονους «μάντιες» και τα αρχαία «μαντεία», όπως αυτό των Δελφών στην αρχαία Ελλάδα, μέχρι και τους σημερινούς μάντιες και αστρολόγους, οι άνθρωποι πάντα ήθελαν και συνεχίζουν να θέλουν να γνωρίζουν το τι πρόκειται να συμβεί. Από την πρόβλεψη του «καιρού» και τα προεκλογικά «γκάλοπ» μέχρι τα ωράρια των μέσων μαζικής μεταφοράς, ειδικά όσον αφορά στις ώρες άφιξης στον προορισμό τους, οι προβλέψεις είναι μέρος της καθημερινότητας όλων των ανθρώπων.

Όσον αφορά τις επιχειρήσεις, είτε αυτές παράγουν προϊόντα, είτε εμπορεύονται, είτε ακόμα παρέχουν υπηρεσίες η «πρόβλεψη» ειδικά των μελλοντικών πωλήσεων προσφέρει πολλά οικονομικά οφέλη. Από την καλύτερη διάρθρωση της παραγωγής και τον οικονομικότερο έλεγχο των αποθεμάτων τους μέχρι και την καλύτερη αξιοποίηση του προσωπικού τους.

Μια πρόβλεψη σήμερα δεν είναι «μαντεία» με την μεταφυσική έννοια του όρου. Χρησιμοποιεί κάποιες μεθόδους. Κάποιες από αυτές είναι υποκειμενικές και στηρίζονται σε εκτιμήσεις προσώπων που γνωρίζουν, ενώ κάποιες άλλες αντικειμενικές και βασίζονται στην χρήση μαθηματικών και στατιστικής.

Οι τελευταίες αυτές μέθοδοι απαιτούν την χρήση ηλεκτρονικών υπολογιστών που να μπορούν να επεξεργαστούν μεγάλους όγκους δεδομένων. Στην παρούσα εργασία επιλέχτηκαν κάποιες μέθοδοι πρόβλεψης και για την εφαρμογή τους αναπτύχθηκε λογισμικό στην γλώσσα Python.

1.2 Προβλέψεις πωλήσεων

Σε μια επιχείρηση υπάρχουν πολλά πράγματα που χρειάζονται πρόβλεψη. Το πόσα και ποια βέβαια εξαρτάται από το είδος της επιχείρησης. Από τις συναλλαγματικές ισοτιμίες σε μια εμπορική εισαγωγική ή εξαγωγική επιχείρηση μέχρι τις μετεωρολογικές προβλέψεις σε επιχείρηση παραγωγής αγροτικών προϊόντων.

Μια από τις σημαντικότερες ίσως προβλέψεις είναι αυτή που αναφέρεται στις πωλήσεις μιας επιχείρησης.

Η πρόβλεψη πωλήσεων είναι η διαδικασία εκτίμησης των μελλοντικών εσόδων με την πρόβλεψη της ποσότητας των προϊόντων ή υπηρεσιών που θα πουλήσει μια μονάδα πωλήσεων (η οποία μπορεί να είναι ένας μεμονωμένος πωλητής, μια ομάδα πωλήσεων ή μια εταιρεία) την επόμενη εβδομάδα, μήνα, τρίμηνο ή έτος. (Dealhup, 2022)

1.3 Χαρακτηριστικά των προβλέψεων

Οι προβλέψεις γενικά και επομένως και οι προβλέψεις πωλήσεων ειδικότερα έχουν κάποια γενικά χαρακτηριστικά. Οι Chopra και Meindl (Chopra&Meindl, 2004), παραθέτουν αυτά τα γενικά χαρακτηριστικά των προβλέψεων ως εξής:

1. Οι προβλέψεις είναι πάντοτε ανακριβείς και, ως εκ τούτου, θα πρέπει να περιλαμβάνουν τόσο την αναμενόμενη αξία της πρόβλεψης όσο και ένα μέτρο του σφάλματος πρόβλεψης.

2. Οι μακροπρόθεσμες προβλέψεις είναι συνήθως λιγότερο ακριβείς από τις βραχυπρόθεσμες προβλέψεις, δηλαδή οι μακροπρόθεσμες προβλέψεις έχουν μεγαλύτερη τυπική απόκλιση σφάλματος σε σχέση με τον μέσο όρο από ό,τι οι βραχυπρόθεσμες προβλέψεις. Η πρόβλεψη της ζήτησης για ένα μήνα στο μέλλον είναι δυσκολότερη από

την πρόβλεψη της ζήτησης για μια ημέρα στο μέλλον.

3. Οι συγκεντρωτικές προβλέψεις είναι συνήθως πιο ακριβείς από τις αναλυτικές προβλέψεις, καθώς τείνουν να έχουν μικρότερη τυπική απόκλιση σφάλματος σε σχέση με τον μέσο όρο.

4. Γενικά, όσο πιο ψηλά στην αλυσίδα εφοδιασμού βρίσκεται μια εταιρεία (ή όσο πιο μακριά από τον καταναλωτή), τόσο μεγαλύτερη είναι η στρέβλωση των πληροφοριών που λαμβάνει.

1.4 Μέθοδοι πρόβλεψης

Οι μέθοδοι πρόβλεψης διακρίνονται σε δύο βασικές κατηγορίες:

- (i) ποιοτικές και
- (ii) ποσοτικές.

Στις **ποιοτικές μεθόδους** πρόβλεψης περιλαμβάνονται μέθοδοι στις οποίες η πρόβλεψη γίνεται από έναν ή περισσότερους ειδικούς με βάση τη γνώση, την εμπειρία και το ένστικτο τους. Τέτοιες μέθοδοι είναι:

(α) Έρευνα αγοράς (Market Research): Η έρευνα αγοράς αναφέρεται στη συλλογή δεδομένων και πληροφοριών για τους καταναλωτές, όπως οι προτιμήσεις ή οι ανάγκες τους. Τα δεδομένα αυτά συλλέγονται με συνεντεύξεις ή με κατάλληλα ερωτηματολόγια. Η μέθοδος μπορεί να αξιοποιηθεί για νέα κυρίως προϊόντα. (AMA, 2023)

(β) Συμβούλιο στελεχών (Expert opinion): Οι προβλέψεις προκύπτουν από ανταλλαγή απόψεων σε συναντήσεις στελεχών, πωλητών, πελατών ή άλλων ειδικών. (Armstrong, 2008)

(γ) Μέθοδος των Δελφών (Delphi Method): Χρησιμοποιεί ερωτηματολόγια που απαντώνται από ειδικούς. Εφαρμόζεται για την πρόβλεψη μελλοντικών αναγκών, εξελίξεων και συνεπειών, σε πολλούς τομείς και στο χώρο της εκπαίδευσης. (Adriano Bernardo Renzia, 2015)

(δ) Μέθοδος σεναρίων (Scenario forecasting): Οι προβλέψεις με βάση τα σενάρια επιτρέπει τη δημιουργία ενός ευρέος φάσματος πιθανών προβλέψεων και την αναγνώριση ορισμένων ακραίων περιπτώσεων. Είναι ουσιαστικά μια εφεδρική μέθοδος πρόβλεψης πωλήσεων. Η σύνταξη σεναρίων χρησιμοποιείται όταν το μακροπρόθεσμο σχέδιο πωλήσεων μιας εταιρείας είναι δύσκολο να προβλεφθεί λόγω οικονομικών ή περιβαλλοντικών παραγόντων. (Hyndman & Athanopoulos, 2021)

(ε) Μέθοδος Grass Roots: Η πρόβλεψη βασίζεται σε δεδομένα που προέρχονται και πάλι από άτομα που έχουν άμεση σχέση με το αντικείμενο όπως είναι οι πωλητές. (Akkar, 2016)

(στ) Μέθοδος ιστορικής αναλογίας (Historical Analogy): Χρησιμοποιείται στον σχεδιασμό νέων προϊόντων. Τα δεδομένα προέρχονται από τις παλαιότερες πωλήσεις παρόμοιων. (Κουτσοθανάση, 2015)

Οι **ποσοτικές μέθοδοι** διακρίνονται:

(α) Μέθοδοι προεκβολής (ανάλυση χρονοσειρών): Βασίζονται πάνω στην ιδέα ότι οι ιστορικές παρατηρήσεις μπορούν να χρησιμοποιηθούν για την πρόβλεψη μελλοντικών παρατηρήσεων.

(β) Προσομοίωση: Αυτή παρέχει την δυνατότητα εισαγωγής ορισμένων υποθέσεων σχετικά με την πρόβλεψη

(γ) Αιτιακές μέθοδοι: Σε αυτές γίνεται η παραδοχή ότι η ζήτηση εξαρτάται από ορισμένους εξωτερικούς (μακροοικονομικούς) παράγοντες, και χρησιμοποιούν ανάλυση παλινδρόμησης.

Οι **μέθοδοι προεκβολής** (ανάλυση χρονοσειρών) μπορούν να ταξινομηθούν και ανάλογα με το μοντέλο που χρησιμοποιείται.

(1) Μέθοδος πρόβλεψης τελευταίας τιμής (LVFM, Last Value forecasting method ή Naïve

method): Είναι περισσότερο μέθοδος εκτίμησης κατά την οποία τα πραγματικά στοιχεία της προηγούμενης περιόδου χρησιμοποιούνται ως πρόβλεψη της τρέχουσας περιόδου, χωρίς να προσαρμίζονται ή να επιχειρείται ο προσδιορισμός των αιτίων. Χρησιμοποιείται συνήθως για σύγκριση με τις προβλέψεις που παράγονται από πιο εξελιγμένες μεθόδους. (Dhakal, 2017)

(2) *Αποδόμηση χρονοσειρών (Deconstruction of time series)*: Είναι μια διαδικασία αποσύνθεσης μιας χρονοσειράς στα παρακάτω στοιχεία: Τάση - γενική κλίση με την πάροδο του χρόνου. Εποχικές - συμπεριφορές που καταγράφονται σε μεμονωμένες εποχικές περιόδους. Υπόλοιπο - όλα όσα δεν καταγράφονται από την τάση και τις εποχικές συνιστώσες. (Dagum, 2013)

(3) *Ομαλή αφαίρεση ανωμαλιών (εξομάλυνση) για πιο σαφή μοτίβα (Removal of anomalies for clear patterns)*: Η εξομάλυνση δεδομένων είναι μια στατιστική τεχνική που περιλαμβάνει την αφαίρεση ακραίων τιμών από ένα σύνολο δεδομένων χρονοσειρών για να γίνει ένα μοτίβο πιο ορατό. (Roberto Corizzo, 2019)

(4) *Μέθοδος του κινούμενου μέσου όρου (Averaging Forecasting Methods)*: Το κύριο χαρακτηριστικό της μεθόδου των κινητών μέσων όρων είναι ότι παράγει μια πρόβλεψη για μια συγκεκριμένη χρονική περίοδο με τη μέση τιμή των παρατηρούμενων τιμών των δεδομένων (δηλαδή των πραγματικών τιμών της εξαρτημένης μεταβλητής) για τις πιο πρόσφατες η χρονικές περιόδους. (Winkler, 1983)

(5) *Σταθμισμένος Κινούμενος Μέσος Όρος (Weighted Moving Average Method)*: Ένας τέτοιος μέσος όρος δίνει μεγαλύτερη βαρύτητα στα πρόσφατα δεδομένα και μικρότερη στα δεδομένα του παρελθόντος. Αυτό γίνεται με τον πολλαπλασιασμό της τιμής κάθε ράβδου με έναν συντελεστή στάθμισης. (Erycha Puspitasari, 2023)

(6) *Εκθετική εξομάλυνση (Exponential Smoothing Method)*: Είναι ουσιαστικά μια οικογένεια μοντέλων πρόβλεψης. Όλες χρησιμοποιούν σταθμισμένους μέσους όρους προηγούμενων παρατηρήσεων για την πρόβλεψη νέων τιμών. Η ιδέα είναι να δοθεί μεγαλύτερη σημασία στις πρόσφατες τιμές της σειράς. Έτσι, καθώς οι παρατηρήσεις γίνονται παλαιότερες στο χρόνο, η σημασία αυτών των τιμών μειώνεται εκθετικά. Σε αυτήν την κατηγορία υπάγονται και οι μέθοδοι Holt (Εκθετική Εξομάλυνση με Προσαρμογή στην Τάση) και Holt-Winters (Εκθετική Εξομάλυνση με προσαρμογή στην τάση και στην εποχικότητα). (Παρασκευάς, 2017) (Ψωϊνός, 1983)

1.5 Η γλώσσα Python

Για το λογισμικό που συντάχτηκε στην παρούσα εργασία χρησιμοποιήθηκε η αντικειμενοστραφής (object oriented) γλώσσα προγραμματισμού Python. Συγκεκριμένα χρησιμοποιήθηκε η έκδοση 3.11.0 για windows.

Επιπλέον χρησιμοποιήθηκαν οι εξής βιβλιοθήκες:

(i) **Pandas** που χρειάζεται για την διαχείριση database πινάκων, διάβασμα από .csv αρχεία ή φόρτωση database πινάκων σε .csv.

(ii) **Datetime** που χρειάζεται για να αναγνωρίζει ο κώδικας ημερομηνίες.

(iii) **Math** που χρησιμοποιείται για τον υπολογισμό τον θορύβου στην Auto-correlation.

(iv) **statsmodels** που χρειάζεται για τον υπολογισμό του Auto-correlation.

(v) **matplotlib** που χρειάζεται για να εμφανίζονται στην οθόνη τα γραφήματα Auto-correlation.

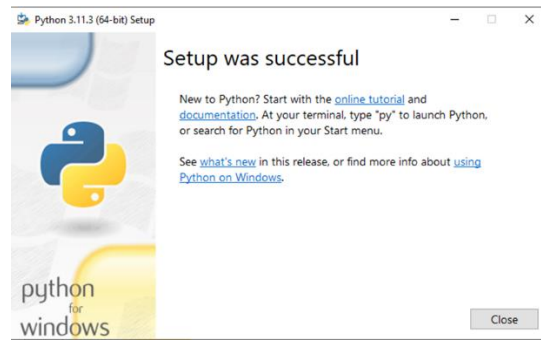
(vi) **numpy** που χρησιμοποιείται στον υπολογισμό της Quadratic Regression.

Για να «φορτωθεί» η Python σε κάποιο υπολογιστή πρώτα πρέπει να συνδεθεί κανείς στην ιστοσελίδα <https://www.python.org/downloads/> και να κατεβάσει την έκδοση που επιθυμεί (προς το παρόν η πιο σύγχρονη έκδοση είναι 3.11.3. Αμέσως μετά να εκτε-

λέσει το .exe που έχει κατεβάσει.



Εικόνα 1.5-1 Εγκατάσταση της Python



Εικόνα 1.5-2 Ολοκλήρωση της διαδικασίας

Επιλέξτε το «Use admin privileges when installing py.exe» και έπειτα επιλέξετε το «Install Now». Περιμένετε να ολοκληρωθεί η διαδικασία και μετά κάντε click στο «close». Η python έχει φορτωθεί στον υπολογιστή σας..

Εάν δεν είναι φορτωμένο το **pip**, από την python που χρησιμοποιείτε, φορτώνεται ως εξής:

Ανοίγει κανείς το cmd των windows και φορτώνει το pip με την εντολή:

```
C:> py get-pip.py
```

Η με την εντολή:

```
python -m pip install -U pip
```

Το **Pandas** φορτώνεται ευκολά μέσω cmd ενεργοποιώντας το pip από PyPI με την εξής εντολή:

```
python -m pip install pandas
```

Η **Datetime** είναι ήδη ενσωματωμένη με την python οπότε δεν χρειάζεται κάποιο είδος φορτώσεις.

Η **Math** είναι ήδη ενσωματωμένη με την python οπότε δεν χρειάζεται κάποιο είδος φορτώσεις.

Το **statsmodels** φορτώνεται ευκολά μέσω cmd με την εξής εντολή:

```
python -m pip install statsmodels
```

Το **matplotlib** φορτώνεται ευκολά μέσω cmd με την εξής εντολή:

```
python -m pip install -U matplotlib
```

Το **numpy** φορτώνεται ευκολά μέσω cmd με την εξής εντολή:

```
python -m pip install numpy
```

1.6 Διάρθρωση της μελέτης

Αντικείμενο της παρούσας εργασίας είναι η σύνταξη λογισμικού για την πρόβλεψη των πωλήσεων της εταιρείας «IC» που παράγει και εμπορεύεται ένα μεγάλο πλήθος διαφορετικών προϊόντων. Από λογισμικό για PC και μουσικά CD, ταινίες DVD μέχρι και είδη δώρων ακόμη και παροχή υπηρεσιών μεταφοράς ή και εισιτήρια για ειδικά event.

Η εργασία χωρίζεται σε κεφάλαια ως εξής:

Στο 1^ο κεφάλαιο δίνεται μια γενική και θεωρητική προσέγγιση των μεθόδων πρόβλεψης, καθώς και κάποια στοιχεία και οδηγίες για την εγκατάσταση της γλώσσας προγραμματισμού Python.

Στο 2^ο κεφάλαιο γίνεται μια παρουσίαση των μεθόδων πρόβλεψης που θα χρησιμοποιηθούν στην εργασία αυτή, αλλά και των εργαλείων με τα οποία θα γίνει ο έλεγχος των προβλέψεων αυτών.

Στο 3^ο κεφάλαιο παρουσιάζονται τα στοιχεία που ευγενικά παραχωρήθηκαν από

την εταιρεία «IC:Companу», καθώς και το λογισμικό των επεξεργασιών που έπρεπε να γίνουν ώστε αυτά να μπορούν να χρησιμοποιηθούν.

Το 4^ο κεφάλαιο αναφέρεται στο λογισμικό που συντάχθηκε για να εφαρμοστούν κάποιες συγκεκριμένες διαδικασίες πρόβλεψης καθώς και λογισμικό για τον έλεγχο και την αξιολόγηση των προβλέψεων. Εδώ επίσης παρατίθενται και λογισμικά για κάποιες επιπλέον δυνατότητες.

Τέλος 5^ο κεφάλαιο παρουσιάζονται τα αποτελέσματα και τα συμπεράσματα καθώς και κάποιες προτάσεις για βελτιώσεις τόσο των μεθόδων πρόβλεψης όσο και του ίδιου του λογισμικού.

2 Μέθοδοι παλινδρόμησης - σφάλματα

2.1 Παλινδρόμηση

2.1.1 Γενικά

Με την ανάλυση παλινδρόμησης (regression analysis) εξετάζουμε τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών με σκοπό την πρόβλεψη των τιμών της μιας, μέσω των τιμών της άλλης (ή των άλλων). Σε κάθε πρόβλημα παλινδρόμησης διακρίνουμε δύο είδη μεταβλητών: τις ανεξάρτητες ή ελεγχόμενες και τις εξαρτημένες ή απόκρισης. Βέβαια κανείς δεν αποκλείει την περίπτωση η ανεξάρτητη μεταβλητή να είναι ο χρόνος.

Ας θεωρήσουμε δύο μεταβλητές X , Y . Αν οι μεταβλητές αυτές συνδέονται με μια σχέση της μορφής $Y=F(X)$ μέσω της οποίας για κάθε τιμή της X μπορούμε να προβλέψουμε ακριβώς την τιμή της Y , δηλαδή, αν οι τιμές της Y δεν υπόκεινται σε σφάλματα, τότε λέμε ότι οι δύο μεταβλητές συνδέονται με μια συναρτησιακή-προσδιοριστική (deterministic) σχέση. Ανάλογα με το ποια είναι η συνάρτηση $F(x)$ διακρίνουμε διάφορες μορφές παλινδρόμησης π.χ. γραμμική όπου η $F(X)$ ευθεία γραμμή, ή τετραγωνική (παραβολή), ή κυβική (τρίτου βαθμού) που αναλύονται στις παρακάτω ενότητες. (Στούμπου, 2021)

Ένας βασικός συντελεστής που κρίνει την αξία του μοντέλου $Y=F(X)$, που προσαρμόζουμε στα δεδομένα μας λέγεται **συντελεστής συσχέτισης r** . Εάν το r είναι κοντά στην μονάδα, τότε πρόκειται για καλή προσαρμογή. Ο συντελεστής αυτός δίνεται γενικά από την εξίσωση:

$$r = \sqrt{1 - \frac{\sum_{i=1}^n (Y_i - F(X_i))^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}}$$

Όπου:

n : το πλήθος των δεδομένων που χρησιμοποιούνται για την πρόβλεψη.

r : ο συντελεστής συσχέτισης και εκφράζει το κλάσμα (ή το ποσοστό) της μεταβλητότητας της τυχαίας μεταβλητής Y που εξηγείται από την X . Ο δείκτης μπορεί να πάρει τιμές ανάμεσα σε 0 και 1 (ή 0 έως 100, όταν εκφράζεται ως ποσοστό επί τοις 100). Όσο πιο κοντά στη μονάδα είναι ο συντελεστής τόσο ισχυρότερη είναι η συναρτησιακή σχέση $Y = F(x)$ εξάρτησης των μεταβλητών X και Y , υπό την προϋπόθεση βέβαια ότι η σχέση αυτή (μοντέλο) είναι το κατάλληλο.

$\hat{Y}_i = F(X_i)$: οι προβλεπόμενες τιμές της μεταβλητής Y .

Y_i : οι πραγματικές τιμές της μεταβλητής Y .

\bar{Y}_i : η μέση τιμή των πραγματικών τιμών της μεταβλητής Y_i .

2.1.2 Απλή Γραμμική παλινδρόμηση

Σκοπός της μεθόδου της Απλής Γραμμικής Παλινδρόμησης (Simple Linear Regression, **SLR**) είναι η προσαρμογή μιας ευθείας γραμμής, η οποία προσαρμόζεται όσο το δυνατόν καλύτερα στη συμπεριφορά των δεδομένων μας. Αν ορίσουμε ως X την ανεξάρτητη μεταβλητή (χρόνος) και Y την εξαρτημένη μεταβλητή (πωλήσεις), τότε η ευθεία θα έχει τη μορφή:

$$\hat{Y}_i = \alpha \cdot X_i + \beta + \varepsilon_i$$

Όπου: \hat{Y}_i : η εξαρτημένη μεταβλητή εδώ οι πωλήσεις

X_i : η ανεξάρτητη μεταβλητή, εδώ ο χρόνος (μήνες)

a : η κλίση της ευθείας και

β : η αρχική τιμή της εξαρτημένης μεταβλητής για $X_i = 0$.

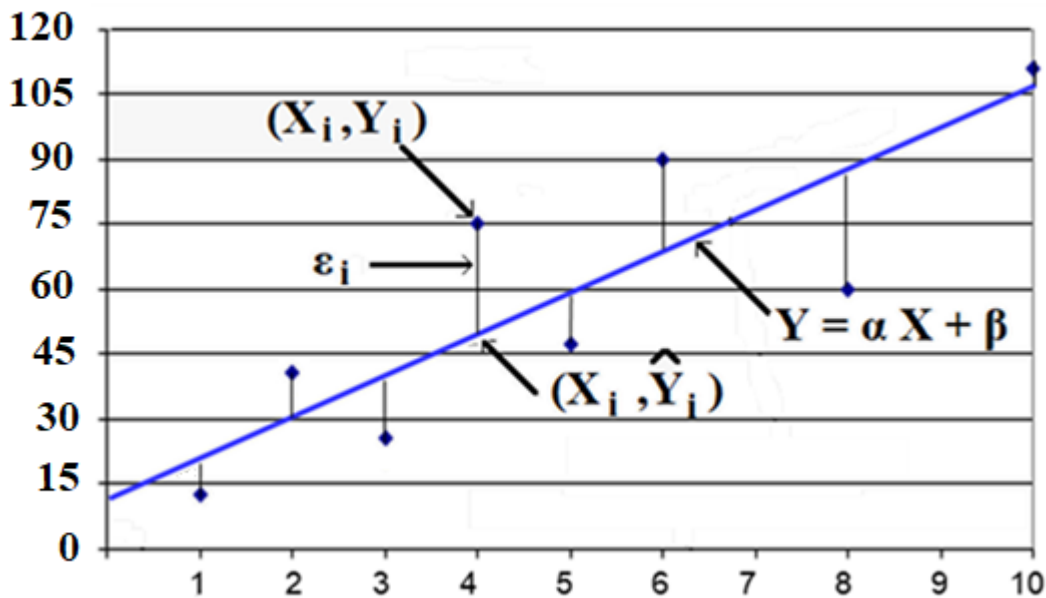
ε_i : είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν κανονική κατανομή $N(0, \sigma^2)$ (σ^2 άγνωστη διασπορά) και συνήθως καλούνται «σφάλματα» των μετρήσεων. Μπορεί να θεωρηθεί ότι τα σφάλματα $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ εμπεριέχουν όλους τους άλλους παράγοντες (εκτός του X) που επηρεάζουν την τιμή της μεταβλητής Y .

Τα σφάλματα προκύπτουν από την σχέση:

$$\varepsilon_i = Y_i - \hat{Y}_i$$

Είναι δηλαδή η διαφορά της πραγματικής από την προβλεπόμενη τιμή της εξαρτημένης μεταβλητής. Με άλλα λόγια είναι:

$$\varepsilon_i = Y_i - a \cdot X_i - \beta$$



Εικόνα 2.1-1 Ευθεία ελαχίστων τετραγώνων

Για τον προσδιορισμό των συντελεστών a και β χρησιμοποιείται η μέθοδος των ελαχίστων τετραγώνων. Ο στόχος είναι να βρεθούν οι εκτιμώμενες τιμές των a και β οι οποίες θα παρέχουν την «καλύτερη» προσαρμογή για τα δεδομένα. Σαν «καλύτερη» προσαρμογή νοείται η γραμμή που ελαχιστοποιεί το άθροισμα των τετραγώνων των σφαλμάτων.

$$F(X_i, Y_i) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - a \cdot X_i - \beta)^2$$

(Βλέπε εικόνα 2.1-1).

Για να πάρει η σύνθετη αυτή συνάρτηση την ελάχιστη της τιμή, σύμφωνα με τα μαθηματικά θα πρέπει να μηδενίζονται οι μερικές παράγωγοι ως προς τους συντελεστές a και β , δηλαδή:

$$\frac{\partial F(X_i, Y_i)}{\partial a} = \frac{\partial F(X_i, Y_i)}{\partial \beta} = 0$$

Έτσι οι τιμές των a και β που ικανοποιούν την παραπάνω συνθήκη δίνονται από τις παρακάτω σχέσεις:

$$a = \frac{\sum_{i=1}^n (X_i \cdot Y_i) - n \cdot \bar{X} \cdot \bar{Y}}{\sum_{i=1}^n X_i^2 - n \cdot \bar{X}^2} = \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

και

$$\beta = \bar{Y} - a\bar{X}$$

Όπου:

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

και

$$\bar{Y} = \frac{1}{n} \cdot \sum_{i=1}^n Y_i$$

οι μέσες τιμές της εξαρτημένης (Y) και της ανεξάρτητης (X) μεταβλητής.

Από το πλήθος n των περιόδων που θα χρησιμοποιηθούν δεδομένα θα έχουμε αντίστοιχα SLR3, SLR4 κ.ο.κ.

Ο συντελεστής συσχέτισης στην γραμμική παλινδρόμηση μπορεί να βρεθεί και από τις σχέσεις:

$$r = \frac{\sqrt{\sum_{i=1}^n (\hat{Y}_i - \bar{Y}_i)^2}}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}} = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{(n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2)(n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2)}}$$

Όπου:

\hat{Y}_i : οι προβλεπόμενες τιμές της μεταβλητής Y .

Y_i : οι πραγματικές τιμές της μεταβλητής Y .

\bar{Y}_i : η μέση τιμή των πραγματικών τιμών της μεταβλητής Y .

Μια άλλη παράμετρος είναι ο αριθμός των δεδομένων (n) που θα χρησιμοποιηθούν για την πρόβλεψη. (Boutsikas, 2004),

2.1.3 Τετραγωνική παλινδρόμηση

Υπάρχουν περιπτώσεις όπου τα δεδομένα δεν ακολουθούν ακριβώς μια ευθεία γραμμή, αλλά μια παραβολή, δηλαδή ένα πολυώνυμο δευτέρου βαθμού (Quadratic Regression, QR).

Γενικά η παραβολή είναι της μορφής:

$$\hat{Y}_i = a \cdot X_i^2 + \beta \cdot X_i + c + \varepsilon_i$$

Όπου \hat{Y}_i : η εξαρτημένη μεταβλητή εδώ οι πωλήσεις

X_i : η ανεξάρτητη μεταβλητή, εδώ ο χρόνος (μήνες)

a, β, c : συντελεστές της παραβολής.

ε_i : τα «σφάλματα» των μετρήσεων, δηλαδή η διαφορά της πραγματικής από την προβλεπόμενη τιμή της εξαρτημένης μεταβλητής:

$$\varepsilon_i = Y_i - \hat{Y}_i$$

Με άλλα λόγια είναι:

$$\varepsilon_i = Y_i - a \cdot X_i^2 - \beta \cdot X_i - c$$

Για τον προσδιορισμό των συντελεστών a , b και c χρησιμοποιείται και πάλι η μέθοδος των ελαχίστων τετραγώνων. Ο στόχος και πάλι είναι να βρεθούν οι εκτιμώμενες τιμές των a , b και c οι οποίες ελαχιστοποιούν το άθροισμα των τετραγώνων των σφαλμάτων, δηλαδή της συνάρτησης:

$$(\varepsilon_i^2) = (Y_i - a \cdot X_i^2 - \beta \cdot X_i - c)^2$$

Για να προσδιοριστεί η εξίσωση της παραβολής μιας τετραγωνικής παλινδρόμησης μπορεί να χρησιμοποιηθεί και πάλι η μέθοδος των ελαχίστων τετραγώνων. Βασιζόμενοι σε ένα σύνολο δεδομένων, πρέπει να προσδιοριστούν οι τιμές των a , b και c έτσι ώστε το άθροισμα των τετραγώνων των κάθετων αποστάσεων μεταξύ κάθε δεδομένου σημείου (X, Y) και της εξίσωσης της παραβολής, να είναι ελάχιστο.

Οι τιμές των a , b και c που ικανοποιούν την παραπάνω συνθήκη δίνονται από την λύση του παρακάτω συστήματος τριών εξισώσεων με τρεις αγνώστους:

$$\begin{aligned} \alpha \cdot \sum_{i=1}^n X_i^4 + b \cdot \sum_{i=1}^n X_i^3 + c \cdot \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n X_i^2 \cdot Y_i \\ \alpha \cdot \sum_{i=1}^n X_i^3 + b \cdot \sum_{i=1}^n X_i^2 + c \cdot \sum_{i=1}^n X_i &= \sum_{i=1}^n X_i \cdot Y_i \\ \alpha \cdot \sum_{i=1}^n X_i^2 + b \cdot \sum_{i=1}^n X_i + c \cdot n_i &= \sum_{i=1}^n Y_i \end{aligned}$$

Η λύση του συστήματος στην παρούσα εργασία γίνεται με την χρήση της εντολής $x = \text{np.linalg.solve}(A, b)$ της `python`. ([Github, 2021](#))

Ο συντελεστής, r που καθορίζει πόσο καλά μια δευτεροβάθμια εξίσωση μπορεί να ταιριάζει στα δεδομένα λέγεται συντελεστής συσχέτισης. Εάν το r είναι κοντά στο 1, τότε πρόκειται για καλή προσαρμογή. Ο συντελεστής αυτός δίνεται από την εξίσωση:

$$r = \sqrt{1 - \frac{\sum_{i=1}^n (Y_i - a \cdot X_i^2 - \beta \cdot X_i - c)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

([Hotmath, 2020](#)).

2.1.4 Άλλες μορφές παλινδρόμησης

Ανάλογα με το ποια συνάρτηση εκτιμά κανείς ότι ακολουθεί η χρονοσειρά μπορούν να χρησιμοποιηθούν διάφορες συναρτήσεις όπως φαίνονται στον παρακάτω πίνακα 2.1-1. Φυσικά αυτές δεν είναι και οι μόνες που μπορούν να χρησιμοποιηθούν. Μπορεί κανείς να «δοκιμάσει» οποιαδήποτε συνάρτηση αρκεί μαθηματικά να εκτιμήσει τις απαραίτητες παραμέτρους της με γνώμονα τον μηδενισμό των πρώτων παραγώγων στην εφαρμογή της μεθόδου των ελαχίστων τετραγώνων.

Πίνακας 2.1-1 Άλλες μορφές παλινδρόμησης

Είδος	Εξίσωση	Υπολογισμός συντελεστών
Κυβική	$\hat{Y}_i = a \cdot X_i^3 + b \cdot X_i^2 + c \cdot X_i + d + \varepsilon_i$	$a \cdot \sum_{i=1}^n X_i^3 + b \cdot \sum_{i=1}^n X_i^2 + c \cdot \sum_{i=1}^n X_i + n \cdot d = \sum_{i=1}^n Y_i$ $a \cdot \sum_{i=1}^n X_i^4 + b \cdot \sum_{i=1}^n X_i^3 + c \cdot \sum_{i=1}^n X_i^2 + d \cdot \sum_{i=1}^n X_i = \sum_{i=1}^n X_i \cdot Y_i$ $a \cdot \sum_{i=1}^n X_i^5 + b \cdot \sum_{i=1}^n X_i^4 + c \cdot \sum_{i=1}^n X_i^3 + d \cdot \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i^2 \cdot Y_i$ $a \cdot \sum_{i=1}^n X_i^6 + b \cdot \sum_{i=1}^n X_i^5 + c \cdot \sum_{i=1}^n X_i^4 + d \cdot \sum_{i=1}^n X_i^3 = \sum_{i=1}^n X_i^3 \cdot Y_i$

Δύναμης	$\hat{Y}_i = a \cdot \hat{X}_i^b + \varepsilon_i$	$a = \exp\left(\frac{1}{n} \sum_{i=1}^n \ln Y_i - \frac{\ln b}{n} \sum_{i=1}^n X_i\right)$ $b = \frac{n \sum_{i=1}^n (\ln X_i \cdot \ln Y_i) - \sum_{i=1}^n \ln X_i \cdot \sum_{i=1}^n \ln Y_i}{n \sum_{i=1}^n \ln^2 X_i - (\sum_{i=1}^n \ln X_i)^2}$
Εκθετική ab	$\hat{Y}_i = a \cdot b^{X_i} + \varepsilon_i$	$a = \exp\left(\frac{1}{n} \sum_{i=1}^n \ln Y_i - \frac{\ln b}{n} \sum_{i=1}^n X_i\right)$ $b = \exp\left\{\frac{n \sum_{i=1}^n X_i \ln Y_i - \sum_{i=1}^n X_i \cdot \sum_{i=1}^n \ln Y_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}\right\}$
Υπερβολική	$\hat{Y}_i = a + \frac{b}{X_i} + \varepsilon_i$	$a = \frac{1}{n} \sum_{i=1}^n Y_i - \frac{b}{n} \sum_{i=1}^n \frac{1}{X_i}$ $b = \frac{n \sum_{i=1}^n \frac{Y_i}{X_i} - \sum_{i=1}^n \frac{1}{X_i} \sum_{i=1}^n Y_i}{n \sum_{i=1}^n \frac{1}{X_i^2} - (\sum_{i=1}^n \frac{1}{X_i})^2}$
Εκθετική	$\hat{Y}_i = e^{a+bX_i} + \varepsilon_i$	$a = \frac{1}{n} \sum_{i=1}^n \ln Y_i - \frac{b}{n} \sum_{i=1}^n X_i$ $b = \frac{n \sum_{i=1}^n X_i \ln Y_i - \sum_{i=1}^n X_i \cdot \sum_{i=1}^n \ln Y_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$

(Planetcalc, n.d.)

2.2 Τάση και Εποχικότητα (Trend, seasonality)

Σε κάθε δεδομένη χρονική στιγμή, μια χρονοσειρά γενικά μπορεί να αναλυθεί σε τέσσερα μέρη: τάση, εποχικότητα κυκλικότητα και τα υπόλοιπα. Δηλαδή:

$$\text{Χρονοσειρά} = \text{Τάση} + \text{Εποχικότητα} + \text{Κυκλικότητα} + \text{Υπόλοιπα}$$

Η **τάση (trend)** σε μια χρονοσειρά αντιπροσωπεύει τη μακροπρόθεσμη μεταβολή της στάθμης της. Η αλλαγή αυτή μπορεί να είναι είτε ανοδική (αύξηση της στάθμης) είτε καθοδική (μείωση της στάθμης). (Chatfield, 2000)

Η **εποχικότητα (seasonality)** είναι ο βαθμός στον οποίο μια χρονοσειρά μεταβάλλεται με συνέπεια μέσα στην χρονική περίοδο ενός έτους. (Cerqueira, 2023)



Εικόνα 2.2-1 Πωλήσεις για την κατηγορία 12 (κονσόλες PS4).

Ένας **κύκλος (κυκλικότητα, circularity)** εμφανίζεται όταν τα δεδομένα παρουσιάζουν αυξήσεις και μειώσεις που δεν έχουν σταθερή συχνότητα. Οι διακυμάνσεις αυτές οφείλονται συνήθως στις οικονομικές συνθήκες και συχνά συνδέονται με τον «επιχειρηματικό κύκλο». Η διάρκεια αυτών των διακυμάνσεων είναι συνήθως ίση ή μεγαλύτερη από 2 χρόνια. Επομένως για να ελέγξει κανείς αν μπορεί να ερευνηθεί και να χρησιμοποιηθεί θα πρέπει να διαθέτει στοιχεία για περισσότερα από δύο χρόνια.

Τα **υπόλοιπα (remainder)** είναι ό,τι έχει απομείνει από τα δεδομένα μιας χρονοσειράς μετά την αφαίρεση των συνιστωσών της τάσης, του κύκλου και της εποχικότητας. Πρόκειται για την τυχαία διακύμανση των δεδομένων της χρονοσειράς, την οποία δεν μπορούν να εξηγήσουν οι παραπάνω συνιστώσες, συνήθως αποκαλείται «λευκός θόρυβος» και μάλλον μόνο στατιστικός έλεγχος μπορεί να γίνει για αυτόν.

(Hyndman & Athanasopoulos, 2.8 Autocorrelation, 2018)



Εικόνα 2.2-2 Πωλήσεις για κατηγορία 3 (αξεσουάρ για κονσόλες PS4)

Ένας τρόπος μαθηματικού ελέγχου της ύπαρξης τάσης ή και εποχικότητας είναι η αυτοσυσχέτιση που αναλύεται στην επόμενη ενότητα.

2.3 Αυτοσυσχέτιση (autocorrelation)

2.3.1 Γενικά

Όπως η συσχέτιση μετρά την έκταση μιας γραμμικής σχέσης μεταξύ δύο μεταβλητών, η αυτοσυσχέτιση μετρά τη γραμμική σχέση μεταξύ υστερημένων (καθυστερημένων) τιμών μιας χρονοσειράς.

Υπάρχουν αρκετοί συντελεστές αυτοσυσχέτισης οι οποίοι αντιστοιχούν σε κάθε ένα τμήμα του διαγράμματος υστέρησης. Για παράδειγμα, ο συντελεστής r_1 μετράει την συσχέτιση μεταξύ Y_t και Y_{t-1} , ο r_2 μεταξύ Y_t και Y_{t-2} , και ούτω καθεξής.

$$r_k = \frac{\sum_{t=k+1}^T (Y_t - \bar{Y}) \cdot (Y_{t-k} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

Όπου:

r_k : είναι ο συντελεστής συσχέτισης δύο στοιχείων της χρονοσειράς που απέχουν χρονικά k περιόδους (μήνες)

n : το πλήθος των τιμών της χρονοσειράς

T : είναι το μήκος της χρονοσειράς

Y_t : οι τρέχουσες πωλήσεις

Y_{t-k} : οι υστερημένες πωλήσεις

\bar{Y} : η μέση τιμή των πωλήσεων

Οι συντελεστές αυτοσυσχέτισης αποτελούν τη συνάρτηση αυτοσυσχέτισης ή ACF (Auto-correlation Function). (Tinungki, 2019)

2.3.2 Τάση και εποχικότητα σε διαγράμματα ACF

Όταν τα δεδομένα παρουσιάζουν τάση, οι συντελεστές αυτοσυσχέτισης για μικρές υστερήσεις τείνουν να είναι μεγάλες και θετικές. Αυτό οφείλεται στο ότι οι παρατηρήσεις σε κοντινές χρονικές στιγμές είναι επίσης ανάλογου μεγέθους. Έτσι, η ACF των χρονοσειρών αυτών τείνει να έχει θετικές τιμές που μειώνονται αργά καθώς αυξάνονται οι υστερήσεις.

Όταν τα δεδομένα είναι εποχιακά, οι αυτοσυσχετίσεις θα είναι μεγαλύτερες για τις εποχιακές υστερήσεις (σε πολλαπλάσια της εποχιακής συχνότητας) από ό,τι για τις άλλες υστερήσεις.

Όταν τα δεδομένα παρουσιάζουν και τάση και εποχικότητα, τότε μπορούμε να παρατηρήσουμε έναν συνδυασμό των παραπάνω.

Στην εικόνα 2.3-1 παρουσιάζεται η συνάρτηση αυτοσυσχέτισης (ACF) των συνολικών πωλήσεων της εταιρείας. Οι δυο πορτοκαλί γραμμές αντιστοιχούν στα όρια του λευκού θορύβου. Ενώ το μήκος της χρονοσειράς είναι $T=34$, οι συντελεστές αυτοσυσχέτισης είναι 33 καθώς δεν συμπεριλαμβάνετε ο συντελεστής r_0 καθώς αυτός είναι πάντα μονάδα.



Εικόνα 2.3-1 Συνάρτηση αυτοσυσχέτισης (ACF) για τις συνολικές πωλήσεις

Για σειρές λευκού θορύβου, αναμένουμε κάθε αυτοσυσχέτιση να είναι κοντά στο μηδέν. Φυσικά, δεν θα είναι ακριβώς ίσες με το μηδέν, καθώς υπάρχει κάποια τυχαία διακύμανση. Για μια σειρά λευκού θορύβου, αναμένουμε το 95% των αιχμών στο ACF να βρίσκονται εντός $\pm 2/\sqrt{T}$ όπου T είναι το μήκος της χρονοσειράς. Είναι σύνηθες να απεικονίζονται αυτά τα όρια σε ένα γράφημα του ACF (οι πορτοκαλί διακεκομμένες γραμμές παραπάνω). Εάν μία ή περισσότερες μεγάλες αιχμές βρίσκονται εκτός αυτών των ορίων ή εάν σημαντικά περισσότερες από το 5% των αιχμών βρίσκονται εκτός αυτών των

ορίων, τότε η σειρά πιθανόν να μην είναι λευκός θόρυβος. Στις συνολικές πωλήσεις, όπου $T=34$ τα όρια είναι $\pm \frac{2}{\sqrt{34}} = \pm 0,343$. Όπως φαίνεται στο σχήμα 2.3-1 τρεις από τους 33 συντελεστές αυτοσυσχέτισης βρίσκονται εκτός αυτών των ορίων. Αυτό σημαίνει ένα ποσοστό της τάξης $\frac{3}{33} \times 100\% = 9,1\% > 5\%$ είναι εκτός των ορίων. Άρα δεν πρόκειται για «λευκό θόρυβο». (Hyndman & Athanasopoulos, 2.9 White noise, 2018)

2.4 Σφάλματα κατά την Πρόβλεψη

Οι συγκρίσεις που αναφέρονται παραπάνω πραγματοποιούνται με τη βοήθεια των μέτρων που θα ορίσουμε σε αυτό το κεφάλαιο. Τα πιο συνηθισμένα μέτρα αποτελεσματικότητας μίας πρόβλεψης είναι:

- (1) Η μέση απόλυτη απόκλιση η οποία συμβολίζεται ως **MAD** (Mean Absolute Deviation),
- (2) Το μέσο τετραγωνικό σφάλμα που συμβολίζεται ως **MSE** (Mean Square Error)
- (3) Η ρίζα του μέσου τετραγωνικού σφάλματος η οποία συμβολίζεται με **RMSE** (Root Mean Square Error) και
- (4) Το μέσο ποσοστιαίο σφάλμα, γνωστό ως **MAPE** (Mean Average Percentage Error).

Έστω ότι διαθέτουμε δείγμα με τις πραγματικές τιμές μίας μεταβλητής για n περιόδους. Τότε, οι τιμές των μέτρων υπολογίζονται από τις ακόλουθες σχέσεις:

$$MAD = \frac{1}{n} \sum_{i=1}^n |A_t - F_t|$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (A_t - F_t)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_t - F_t)^2}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

Όπου:

n : το πλήθος των περιόδων και επομένως των ζευγαριών πραγματικής και προβλεπόμενη τιμής

A_t : η πραγματική τιμή

F_t : η προβλεπόμενη τιμή

Στον υπολογισμό του MAPE υπάρχει πρόβλημα όταν συμβαίνει οι πραγματικές πωλήσεις σε μια περίοδο είναι μηδενικές. Επειδή στην περίπτωση αυτή η διαίρεση είναι αδύνατη μπορεί εναλλακτικά να χρησιμοποιηθεί η σχέση:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{A_t - F_t}{A_t + 1} \right|$$

(Ma. Del Rocio Castillo E, 2020) (Jedox, 2022) .

3 Περιγραφή δεδομένων, στατιστικά

3.1 Η Εταιρεία

Τα αρχεία περιλαμβάνουν 22170 προϊόντα ταξινομημένα σε 84 κατηγορίες. Τα προϊόντα αυτά πωλούνται σε 60 καταστήματα (σημεία πώλησης). Σύμφωνα με την ίδια:

«Η 1C Company ειδικεύεται στην ανάπτυξη, διανομή, έκδοση και υποστήριξη λογισμικού. Εξυπηρετεί τους πελάτες της μέσω ενός εκτεταμένου δικτύου συνεργατών που εκτείνεται σε 25 χώρες, συμπεριλαμβανομένων περισσότερων από 7.000 συνεργατών 1C:Franchisee. Η 1C πιστοποιεί αυτούς τους συνεργάτες για την παροχή υπηρεσιών ολοκλήρωσης για την αυτοματοποίηση επιχειρήσεων. Μεταξύ των πιο γνωστών λύσεων λογισμικού που έχει αναπτύξει η 1C Company είναι το 1C:Enterprise, ένα σύστημα προϊόντων αυτοματοποίησης επιχειρήσεων. Περιλαμβάνει επιχειρηματικές λύσεις κατηγορίας ERP που βελτιώνουν την αποδοτικότητα της επιχείρησης και του οργανισμού και βασίζονται σε μια καινοτόμο τεχνολογική πλατφόρμα. Γνωστή για την εξαιρετική της ευελιξία που απορρέει από τα περιβάλλοντα on-premise, cloud και mobile, η 1C:Enterprise αυτοματοποιεί με επιτυχία τον επιχειρησιακό προγραμματισμό πόρων, τις λειτουργίες, τη λογιστική, τα οικονομικά, το HR και τις δραστηριότητες διαχείρισης σε περισσότερες από 1.500.000 εταιρείες, συμπεριλαμβανομένων μεγάλων εταιρειών και κυβερνητικών ιδρυμάτων. Το σύστημα κατέχει ηγετική θέση στη ρωσική αγορά ERP όσον αφορά τον αριθμό των αυτοματοποιημένων θέσεων εργασίας και έχει υψηλό δυναμικό εξαγωγών». (1C:Company, 2019)

3.2 Τα πρωτογενή αρχεία

Τα πρωτογενή αρχεία που ευγενικά παραχωρήθηκαν από την εταιρεία 1C βρίσκονται στην διεύθυνση: (1C:Company, Kaggle, 2018) και περιλαμβάνουν 22170 προϊόντα ταξινομημένα σε 84 κατηγορίες.

Τα προϊόντα αυτά πωλούνται σε 60 καταστήματα (σημεία πώλησης). Τα επιμέρους αρχεία που παραχωρήθηκαν είναι:

- 1. sales_train.csv:** Περιλαμβάνει στοιχεία για τις πωλήσεις των προϊόντων
- 2. items.csv:** συμπληρωματικές πληροφορίες σχετικά με τα προϊόντα.
- 3. item_categories.csv:** συμπληρωματικές πληροφορίες σχετικά με τις κατηγορίες στις οποίες χωρίζονται τα προϊόντα.
- 4. shops.csv:** συμπληρωματικές πληροφορίες σχετικά με τα σημεία πώλησης (καταστήματα) όπου διατίθενται τα προϊόντα της εταιρείας.

Αναλυτικότερα τα πρωτογενή αρχεία που διαθέτει η εταιρεία περιλαμβάνουν:

1. sales_train.csv: Περιέχει ημερήσια ιστορικά δεδομένα πωλήσεων από τον Ιανουάριο του 2013 έως τον Οκτώβριο του 2015. Στο αρχείο αυτό οι πωλήσεις εμφανίζονται ανά ημέρα και ανά σημείο πώλησης και συμπεριλαμβάνει 2.935.849 γραμμές δεδομένων. Τα δεδομένα διαχωρίζονται με «κόμμα» καθώς το αρχείο είναι της μορφής .csv (**Comma Separated Values**) αρχείο. Απόσπασμα αυτού του αρχείου παρουσιάζεται στην πρωτογενή του μορφή στον πίνακα 3.2-1. Οι ημερομηνίες έχουν την εμφάνιση dd.mm.yy.

Τα αρχεία csv μπορούν να «ανοιχτούν» και με την βοήθεια του excel πατώντας
Δεδομένα → Από κείμενο/CSV → Επιλογή του αρχείου CSV → Μετασχηματισμός δεδομένων → Χρήση της πρώτης γραμμής ως κεφαλίδων → Επιλογή πρώτης στήλης → Ημερομηνία

Το αρχείο έχει την μορφή που φαίνεται στον πίνακα 3.2-1.

Στις «κεφαλίδες» αυτού του αρχείου εμφανίζονται:

date είναι η ημερομηνία Ημέρα/μήνας/Έτος

Πίνακας 3.2-1 Απόσπασμα πρωτογενούς αρχείου πωλήσεων ανά ημέρα

sales_train.csv (94.6 MB)

Detail Compact Column

date	# date_block_num	shop_id	item_id	# item_price	# item_cnt_day
1034 unique values					
02.01.2013	0	59	22154	999.0	1.0
03.01.2013	0	25	2552	899.0	1.0
05.01.2013	0	25	2552	899.0	-1.0
06.01.2013	0	25	2554	1709.05	1.0
15.01.2013	0	25	2555	1099.0	1.0
10.01.2013	0	25	2564	349.0	1.0
02.01.2013	0	25	2565	549.0	1.0
04.01.2013	0	25	2572	239.0	1.0
11.01.2013	0	25	2572	299.0	1.0
03.01.2013	0	25	2573	299.0	3.0
03.01.2013	0	25	2574	399.0	2.0
05.01.2013	0	25	2574	399.0	1.0
07.01.2013	0	25	2574	399.0	1.0
08.01.2013	0	25	2574	399.0	2.0
10.01.2013	0	25	2574	399.0	1.0
11.01.2013	0	25	2574	399.0	2.0
13.01.2013	0	25	2574	399.0	1.0
16.01.2013	0	25	2574	399.0	1.0
26.01.2013	0	25	2574	399.0	1.0
27.01.2013	0	25	2574	399.0	1.0
09.01.2013	0	25	2593	279.0	1.0
16.01.2013	0	25	2604	299.0	1.0
27.01.2013	0	25	2604	299.0	1.0
27.01.2013	0	25	2607	279.0	1.0
29.01.2013	0	25	2607	279.0	1.0
27.01.2013	0	25	2609	1699.0	1.0
06.01.2013	0	25	2548	1708.95	1.0
26.01.2013	0	25	2611	299.0	1.0
02.01.2013	0	25	2546	299.0	1.0
06.01.2013	0	25	2515	1649.0	1.0
07.01.2013	0	25	2705	1799.0	1.0
11.01.2013	0	25	2705	1799.0	1.0
26.01.2013	0	25	2705	1799.0	1.0

date_block_num είναι κωδικοποιημένοι με αρίθμηση από την εταιρεία οι μήνες π.χ. 0 είναι ο Ιανουάριος του 2013, 1 ο Φεβρουάριος του 2013 έως και 33 που αντιστοιχεί στον τελευταίο μήνα για τον οποίο υπάρχουν δεδομένα και είναι ο Οκτώβριος του 2015

shop_id κωδικός του σημείου πώλησης

item_id κωδικός προϊόντος από 0 έως 22169 (22170 προϊόντα).

item_price τιμή προϊόντος ανά τεμάχιο (σε Ρούβλια)

item_cnt_day ποσότητα του συγκεκριμένου προϊόντος που πουλήθηκε την συγκεκριμένη ημέρα (το «-», π.χ. «-1», μπροστά από μια ποσότητα σημαίνει επιστροφή προϊόντος). Η εμφάνιση του παραπάνω αρχείου στο Excel είναι όπως στον πίνακα 3.2-2.

Πίνακας 3.2-2 Εμφάνιση πρωτογενούς αρχείου πωλήσεων ανά ημέρα στο Excel

	A	B	C	D	E	F	G
1	date	date_block_num	shop_id	item_id	item_price	item_cnt_day	
2	2/1/2013	0	59	22154	999	1	
3	3/1/2013	0	25	2552	899	1	
4	5/1/2013	0	25	2552	899	-1	
5	6/1/2013	0	25	2554	1709,05	1	
6	15/1/2013	0	25	2555	1099	1	
7	10/1/2013	0	25	2564	349	1	
8	2/1/2013	0	25	2565	549	1	
9	4/1/2013	0	25	2572	239	1	
10	11/1/2013	0	25	2572	299	1	
11	3/1/2013	0	25	2573	299	3	
12	3/1/2013	0	25	2574	399	2	
13	5/1/2013	0	25	2574	399	1	
14	7/1/2013	0	25	2574	399	1	
15	8/1/2013	0	25	2574	399	2	
16	10/1/2013	0	25	2574	399	1	
17	11/1/2013	0	25	2574	399	2	
18	13/1/2013	0	25	2574	399	1	
19	16/1/2013	0	25	2574	399	1	
20	26/1/2013	0	25	2574	399	1	
21	27/1/2013	0	25	2574	399	1	
22	9/1/2013	0	25	2593	279	1	
23	16/1/2013	0	25	2604	299	1	
24	27/1/2013	0	25	2604	299	1	
25	27/1/2013	0	25	2607	279	1	
26	29/1/2013	0	25	2607	279	1	
27	27/1/2013	0	25	2609	1699	1	
28	6/1/2013	0	25	2548	1708,95	1	
29	26/1/2013	0	25	2611	299	1	
30	2/1/2013	0	25	2546	299	1	
31	6/1/2013	0	25	2515	1649	1	
32	7/1/2013	0	25	2705	1799	1	
33	11/1/2013	0	25	2705	1799	1	
34	26/1/2013	0	25	2705	1799	1	
35	2/1/2013	0	25	2715	899	1	
36	12/1/2013	0	25	2715	898,5	1	
37	2/1/2013	0	25	2716	149	1	
38	23/1/2013	0	25	2716	149	1	
39	2/1/2013	0	25	2719	2699	1	



2. items.csv:

Το πρωτογενές αρχείο που περιέχει τον ονομαστικό κατάλογο με τα προϊόντα (items.csv) έχει την εμφάνιση που φαίνεται στον Πίνακα 3.2-3 όπου δίνεται ενδεικτικά μια σελίδα του αρχείου αυτού.

Πίνακας 3.2-3 Απόσπασμα ονομαστικού καταλόγου προϊόντων

items.csv (1.57 MB)

Detail Compact Column

▲ item_name	☰ item_id	☰ item_category_id
22170 unique values	 0 22.2k	 0 83
! ВО ВЛАСТИ НАВАЖДЕНИЯ (ПЛАСТ.) D	0	40
!ABBY FineReader 12 Professional Edition Full [PC, Цифровая версия]	1	76
***В ЛУЧАХ СЛАВЫ (UNV) D	2	40
***ГОЛУБАЯ ВОЛНА (Univ) D	3	40
***КОРОБКА (СТЕКЛО) D	4	40
***НОВЫЕ АМЕРИКАНСКИЕ ГРАФФИТИ (UNI) D	5	40
***УДАР ПО ВОРОТАМ (UNI) D	6	40
***УДАР ПО ВОРОТАМ-2 (UNI) D	7	40
***ЧДЙ С МУССОЛИНИ D	8	40
***ШУГАРЛЭНДСКИЙ ЭКСПРЕСС (UNI) D	9	40
*ЗА ГРАНЬЮ СМЕРТИ D	10	40

Για κάθε item (προϊόν) εμφανίζεται στο παραπάνω αρχείο η ονομασία του (item_name) στα Ρωσικά, ο κωδικός του προϊόντος (item_id) που ξεκινά από το 0 και φτάνει μέχρι το 22169 καθώς και ο κωδικός της κατηγορίας στην οποία υπάγεται (item_category_id).

3. item_categories.csv: Στο αρχείο αυτό δίνονται οι 84 κατηγορίες στις οποίες η εταιρεία έχει κατατάξει τα προϊόντα της (με κωδικούς item_category_id από 0 έως 83). Στον πίνακα 3.2-4 οι τίτλοι (είδος κατηγορίας) έχουν μεταφραστεί από τα Ρωσικά.

Πίνακας 3.2-4 Κατάλογος κατηγοριών προϊόντων

Είδος κατηγορίας	Κωδικός Item_category_id	Αριθμός προϊόντων	Είδος κατηγορίας	Κωδικός Item_category_id	Αριθμός προϊόντων
Ακουστικά/ηχεία	0	4	Βιβλία – Βιβλία τέχνης, εγκυκλοπαίδειες	42	12
Αξεσουάρ - PS2	1	2	Βιβλία- Ακουστικά Βιβλία	43	471
Αξεσουάρ - PS3	2	75	Βιβλία - Ακουστικά βιβλία (ψηφιακά)	44	161
Αξεσουάρ - PS4	3	34	Βιβλία - Ακουστικά βιβλία 1C	45	195
Αξεσουάρ- PSP	4	15	Βιβλία - Επιχειρηματική λογοτεχνία	46	12
Αξεσουάρ - PSVita	5	29	Βιβλία - Κόμικς, Μάνγκα	47	151
Αξεσουάρ - XBOX 360	6	40	Βιβλία - Βιβλιογραφία υπολογιστών	48	5
Αξεσουάρ - XBOX ONE	7	24	Βιβλία - 1Γ Διδακτικό υλικό	49	167
Εισιτήρια (Εικ.)	8	9	Βιβλία - Cart postal	50	2
Παράδοση αγαθών	9	12	Βιβλία – Ενημερωτική λογοτεχνία	51	1
Κονσόλες παιχνιδιών - PS2	10	1	Βιβλία - Οδηγοί	52	3
Κονσόλες παιχνιδιών - PS3	11	41	Βιβλία - Μυθολογία	53	3
Κονσόλες παιχνιδιών - PS4	12	20	Βιβλία - The figure	54	90
Κονσόλες παιχνιδιών - PSP	13	6	Μουσική - CD τοπικής παραγωγής	55	2365
Κονσόλες παιχνιδιών - PSVita	14	18	Μουσική – παραγωγή επώνυμου CD	56	397
Κονσόλες παιχνιδιών - XBOX 360	15	37	Μουσική - MP3	57	427
Κονσόλες παιχνιδιών - XBOX ONE	16	20	Μουσική - βινύλιο	58	790
Κονσόλες παιχνιδιών - Άλλα	17	7	Μουσική – Μουσικό video	59	294
Παιχνίδια - PS2	18	6	Μουσική – Εκδόσεις δώρων	60	52
Παιχνίδια - PS3	19	628	Δώρα- Χαρακτηριστικά	61	598
Παιχνίδια - PS4	20	175	Δώρα - Gadgets, ρομπότ, σπορ	62	295
Παιχνίδια- PSP	21	125	Δώρα - Μαζεμένα παιχνίδια	63	366
Παιχνίδια - PSVita	22	79	Δώρα - Επιτραπέζια παιχνίδια	64	290
Παιχνίδια - XBOX 360	23	501	Δώρα - Επιτραπέζια παιχνίδια (compact)	65	163
Παιχνίδια - XBOX ONE	24	123	Δώρα- Cart postal, αυτοκόλλητα	66	178
Παιχνίδια - Αξεσουάρ για παιχνίδια	25	140	Δώρα - Ανάπτυξη	67	394
Παιχνίδια Android - Αριθμητικά	26	3	Δώρα - Πιστοποιητικά, υπηρεσίες	68	27
MAC Games - Η φιγούρα	27	8	Δώρα - Αναμνηστικά	69	317
Παιχνίδια PC – Έκτακτες εκδόσεις	28	240	Δώρα - Αναμνηστικά (σε μεντεσέ)	70	228
Παιχνίδια PC – Συλλεκτικές εκδόσεις	29	46	Δώρα - Τσάντες, άλμπουμ, mouse pads	71	6
Παιχνίδια PC – Τυπικές εκδόσεις	30	756	Δώρα - Ειδώλια	72	666
Παιχνίδια PC - Ψηφιακά	31	1125	Λογισμικό - 1C: Enterprise 8	73	36
Κάρτες πληρωμών (κινηματογράφος, μουσική, παιχνίδια)	32	6	Προγράμματα - MAC (ψηφιακά)	74	8
Κάρτες πληρωμών - Live!	33	18	Προγράμματα - Για το σπίτι και το γραφείο	75	277
Κάρτες πληρωμών - Live! (Εικόνα)	34	8	Προγράμματα - Για το σπίτι και το γραφείο (ψηφιακά)	76	333
Κάρτες πληρωμών - PSN	35	4	Προγράμματα - Εκπαιδευτικά	77	220
Κάρτες - Windows (Digital)	36	6	Προγράμματα- Εκπαιδευτικά (ψηφιακά)	78	278
Κινηματογράφος - Blu-ray	37	1780	Υπηρεσίες	79	1
Κινηματογράφος- Blu-ray 3D	38	310	Υπηρεσία - Εισιτήρια	80	6
Κινηματογράφος - Blu-ray 4K	39	7	Καθαροί μεταφορείς (ακίδα)	81	7
Κινηματογράφος – DVD	40	5035	Καθαρά μέσα ενημέρωσης (αποσπασματικά)	82	8
Κινηματογράφος – Συλλεκτική έκδοση	41	332	Στοιχεία Ισχύος	83	15

4. shops.csv : Στο αρχείο αυτό εμφανίζονται τα 60 σημεία πώλησης (κωδικός shop_id). Στον πίνακα 3.2-5 εμφανίζεται μετάφραση του πίνακα αυτού από τα Ρωσικά στα ελληνικά.

Πίνακας 3.2-5 Κατάλογος σημείων πώλησης

Όνομα καταστήματος	shop_id	Όνομα καταστήματος	shop_id
yakutsk Ordzhonikidze, 56 Γαλλία	0	Εμπορικό κέντρο Perlovsky της Μόσχας	30
franchise του κεντρικού εμπορικού κέντρου του Γιακούτσκ	1	Μόσχα Εμπορικό κέντρο Semenovsky	31
Εμπορικό κέντρο Adygeya Mega	2	Μόσχα Εμπορικό κέντρο Serebryany Dom	32
Εμπορικό και ψυχαγωγικό κέντρο Balashikha Oktyabr-Kinomir	3	Εμπορικό κέντρο Mytishchi XL-3	33
Εμπορικό κέντρο Volga	4	Εμπορικό κέντρο RIO, Νίζνι Νόβγκοροντ	34
Εμπορικό κέντρο μαρμελάδας Vologda	5	Εμπορικό κέντρο N. Novgorod Fantastica	35
Voronezh (Plekhanovskaya 13)	6	Εμπορικό κέντρο Novosibirsk Galereya	36
Εμπορικό κέντρο Voronezh Maximir	7	Εμπορικό κέντρο του Novosibirsk Mega	37
Εμπορικό και ψυχαγωγικό κέντρο Voronezh City-Park Grad	8	Εμπορικό κέντρο Omsk Mega	38
Εξωτερικό εμπόριο	9	Εμπορικό κέντρο RostovNaDon MegaCentre Horizon Shopping Mall	39
Zhukovsky st. Chkalova 39m?	10	RostovNaDon MegaCentre Horizon Εμπορικό κέντρο Ostroynoy	40
Zhukovsky St. Chkalova 39m ²	11	Εμπορικό κέντρο RostovNaDon Mega	41
Ηλεκτρονικό κατάστημα CS	12	SPb TC Nevsky Centre	42
Εμπορικό κέντρο Kazan Behetle	13	SPb TC Sennaya	43
Εμπορικό κέντρο Kazan Park House II	14	Εμπορικό κέντρο Samara Melody	44
Εμπορικό και ψυχαγωγικό κέντρο Kaluga XXI Century	15	Εμπορικό κέντρο Park House στη Σαμάρα	45
Εμπορικό κέντρο Kolomna Rio	16	Sergiev Posad TC "7Ya"	46
Εμπορικό κέντρο Krasnoyarsk Vzletka Plaza	17	Εμπορικό κέντρο Surgut City	47
Εμπορικό κέντρο Krasnoyarsk June	18	Εμπορικό κέντρο Tomsk Emerald City	48
Εμπορικό κέντρο Kursk Pushkinskiy	19	Εμπορικό κέντρο Tyumen Crystal	49
Μόσχα "Πώληση"	20	Εμπορικό κέντρο Tyumen Goodwin	50
Μόσχα MTRC Afi Mall	21	Tyumen TC «Πράσινη παραλία»	51
Κατάστημα Μόσχας C21	22	Ufa TC Central	52
Εμπορικό κέντρο Budyonovsky της Μόσχας (περίπτερο A2)	23	Ufa TC «Οικογένεια» 2	53
Εμπορικό κέντρο Budyonovsky της Μόσχας (περίπτερο K7)	24	Εμπορικό κέντρο Khimki Mega	54
Εμπορικό κέντρο Atrium της Μόσχας	25	1C-Online Ψηφιακή Αποθήκη	55
Εμπορικό κέντρο Moscow Areal (Belyaevno)	26	Εμπορικό και ψυχαγωγικό κέντρο Chekhov Karnaval	56
Εμπορικό κέντρο Μόσχας MEGA Belaya Dacha II	27	Yakutsk Ordzhonikidze, 56	57
Εμπορικό κέντρο MEGA Teply Stan Μόσχα II	28	Κεντρικό εμπορικό κέντρο του Γιακούτσκ	58
Εμπορικό κέντρο New Century της Μόσχας (Novokosino)	29	Εμπορικό κέντρο Yaroslavl Altair	59

Από τον έλεγχο των κατηγοριών προέκυψε ένα σφάλμα στις καταχωρήσεις της εταιρείας. Συγκεκριμένα : Κατά την ανάλυση των στοιχείων βρέθηκε ότι οι κατηγορίες:

(α) 8 (εισιτήρια)

(β) 80 (Υπηρεσία – εισιτήρια)

Αυτές παρουσίαζαν παρόμοια συμπεριφορά στον υπολογισμό του σφάλματος στις προβλέψεις ειδικά των τελευταίων μηνών. Μετά από έλεγχο και απομονώνοντας τις δυο αυτές κατηγορίες ανακαλύφθηκε ότι εμπεριείχαν τα ίδια προϊόντα (Ίδια ακριβώς ονομασία προϊόντων αλλά διαφορετικά item_id και ελάχιστα διαφορετικό όγκο πωλήσεων). Στον παρακάτω πίνακα 3.2-6 φαίνονται οι πωλήσεις ανά μήνα για τις δυο αυτές κατηγορίες.

Πίνακας 3.2-6 Σύγκριση κατηγοριών προϊόντων 8 και 80

Μήνας	Κατηγορία 8	Κατηγορία 80	Μήνας	Κατηγορία 8	2.Κατηγορία 80
Ιαν. 2013	0	0	Ιούν. 2014	0	0
Φεβ. 2013	0	0	Ιουλ. 2014	116	0
Μαρ. 2013	0	0	Αύγ. 2014	335	0
Απρ. 2013	0	0	Σεπ. 2014	5263	0
Μάιος 2013	0	0	Οκτ. 2014	1873	0
Ιούν. 2013	0	0	Νοέμ. 2014	0	0
Ιουλ. 2013	0	0	Δεκ. 2014	0	0
Αύγ. 2013	0	0	Ιαν. 2015	0	0
Σεπ. 2013	0	0	Φεβ. 2015	0	0
Οκτ. 2013	0	0	Μαρ. 2015	0	0
Νοέμ. 2013	0	0	Απρ. 2015	45	45
Δεκ. 2013	0	0	Μάιος 2015	147	147
Ιαν. 2014	0	0	Ιούν. 2015	117	117
Φεβ. 2014	0	0	Ιουλ. 2015	193	195
Μαρ. 2014	0	0	Αύγ. 2015	483	483
Απρ. 2014	0	0	Σεπ. 2015	5128	5149
Μάιος 2014	0	0	Οκτ. 2015	-1	0



Εικόνα 3.2-1 Σύγκριση κατηγοριών 8 και 80

Στην εικόνα 3.2-1 εμφανίζονται οι πωλήσεις ανά μήνα για τις δυο αυτές κατηγορίες όπως δίνονται από στοιχεία της εταιρείας.

Είναι φανερό ότι οι πωλήσεις από τον Απρίλιο του 2015 μέχρι τον Οκτώβριο του 2015 είναι πανομοιότυπες και είναι αυτές που αναφέρονται σε items με την ίδια ονομασία. Προφανώς έχουν καταχωρηθεί λάθος από την ίδια την εταιρεία. Και οι δυο κατηγορίες αναφέρονται σε εισιτήρια για το ίδιο event («Game World» που γίνεται κάθε χρόνο στην Ρωσία και διαρκεί 4 ημέρες όπου η τέταρτη μέρα είναι πάντα η πρώτη Κυριακή του Οκτωβρίου). Η κατηγορία 8 εμπεριέχει τις πωλήσεις εισιτηρίων για το ίδιο event του 2014 καθώς επίσης και του 2015 ενώ η 80 εμπεριέχει τα εισιτήρια του ίδιου event αλλά μόνο για το 2015, με μικρές αποκλίσεις εξαιτίας των καταχωρήσεων από διαφορετικά σημεία πώλησης.

Τα παραπάνω οδήγησαν στο να μην ληφθούν υπόψη οι πωλήσεις που αναφέρονται στην κατηγορία 80.

4 Μεθοδολογία, διαδικασία εξαγωγής αποτελεσμάτων

4.1 Γενικά

Εξαιτίας της μορφής που έχουν τα πρωτογενή αρχεία, αυτά έπρεπε να υποστούν μια επεξεργασία έτσι ώστε να μπορούν να χρησιμοποιηθούν. Η επεξεργασία αυτή περιλαμβάνει κάποια στάδια που περιγράφονται αναλυτικά στις επόμενες ενότητες.

4.2 Έλεγχοι που έγιναν στα δεδομένα

4.2.1 Γενικά

Πριν ξεκινήσει η διαδικασία των προβλέψεων έπρεπε να γίνουν κάποιοι έλεγχοι στα δεδομένα. Οι έλεγχοι αυτοί περιγράφονται στις παρακάτω ενότητες.

4.2.2 Έλεγχος ύπαρξης κενών στο αρχείο των προϊόντων

Πρώτα – πρώτα τα δεδομένα έπρεπε να εξεταστούν ως προς την ύπαρξη κενών, τόσο στα επιμέρους προϊόντα όσο και στις κατηγορίες. Εξαιτίας του τεράστιου αριθμού των καταγραφών στο ιστορικό αρχείο της εταιρείας ο έλεγχος δεν μπορούσε να γίνει οπτικά. Για τον σκοπό αυτό συντάχθηκε ο κώδικας: **CheckForItemsIDInCSV.py**.

Στον κώδικα αυτό ο έλεγχος γίνεται στο πρωτογενές αρχείο **items.csv** για το αν τυχόν δεν έχει καταχωρηθεί από την εταιρεία, σε αυτό κάποιο προϊόν. Ο έλεγχος γίνεται συγκρίνοντας τον αύξοντα αριθμό της γραμμής (από 0 έως 22169) με τον κωδικό **item_id** του πρωτογενούς αρχείου.

Ο κώδικας αυτός δημιουργεί το αρχείο:

(α) AreIDCorrect.csv. Αν υπάρχει κάποια παράλειψη και οι δυο αριθμοί δεν συμφωνούν, τότε στο παραπάνω csv αρχείο θα εμφανιστεί η τιμή False. Αν συμφωνούν εμφανίζεται η τιμή True. Στο αρχείο αυτό δεν εμφανίστηκε τιμή false όποτε είναι καταχωρημένα σωστά όλα τα προϊόντα με το κατάλληλο ID.

4.2.3 Έλεγχος ύπαρξης κενών στο αρχείο των κατηγοριών

Στην συνέχεια συντάχθηκε ο κώδικας **CheckForItemsCategoryIDInCSV.py**. Με αυτόν γίνεται ο ίδιος έλεγχος στο πρωτογενές αρχείο **item_categories.csv** που περιέχει τις κατηγορίες των προϊόντων. Ο έλεγχος αφορά στο αν τυχόν δεν έχει καταχωρηθεί σε αυτό κάποια κατηγορία προϊόντων. Ο έλεγχος γίνεται συγκρίνοντας και πάλι, τον αύξοντα αριθμό της γραμμής (από 0 έως 83) με τον κωδικό **item_category_id** του πρωτογενούς αρχείου.

Ο κώδικας αυτός δημιουργεί το αρχείο:

(α) AreCategoryIDCorrect.csv. Αν υπάρχει κάποια παράλειψη και οι δυο αριθμοί δεν συμφωνούν, τότε στο παραπάνω csv αρχείο θα εμφανιστεί η τιμή False. Αν συμφωνούν εμφανίζεται η τιμή True. Στο αρχείο αυτό δεν εμφανίστηκε τιμή false όποτε είναι καταχωρημένα σωστά όλες οι κατηγορίες προϊόντων με το κατάλληλο ID.

4.3 Επεξεργασία πρωτογενών αρχείων.

4.3.1 Γενικά

Στα πρωτογενή αρχεία οι πωλήσεις φαίνονται με την μορφή τεμαχίων ανά ημέρα που πουλήθηκαν σε κάθε ένα σημείο πώλησης ξεχωριστά. Από το αρχείο αυτό πρέπει να

εξαχθούν οι συνολικές πωλήσεις για κάθε προϊόν από όλα τα σημεία πώλησης και ανά μήνα. Το ίδιο πρέπει να γίνει και με τις κατηγορίες.

Μια άλλη επεξεργασία αφορά στην εξαγωγή των εισπράξεων από τις συγκεκριμένες πωλήσεις, καθώς και μια μέση τιμή πώλησης ανά προϊόν και μήνα.

Εξαιτίας της φύσης των προϊόντων που διαθέτει η εταιρεία, με λίγες εξαιρέσεις έχουν «ημερομηνία απελευθέρωσης» στην αγορά (release day) και «ημερομηνία λήξης». Μερικοί από τους λόγους που συμβαίνει αυτό είναι:

(α) Οι κονσόλες παιχνιδιών, τα αξεσουάρ τους και τα παιχνίδια ανανεώνονται ανά τακτά χρονικά διαστήματα.

(β) Τα μουσικά CD που παύουν να είναι δημοφιλή μετά από κάποιο χρονικό διάστημα.

(γ) Στις καινούργιες ταινίες DVD που συνηθώς πωλούν μόνο ένα χρονικό διάστημα μετά την κυκλοφορία τους.

Πρέπει λοιπόν, από τα πρωτογενή αρχεία της εταιρείας να εξαχθούν οι παραπάνω ημερομηνίες και να βρεθεί η «διάρκεια ζωής» του κάθε προϊόντος.

Οι επεξεργασίες αυτές περιγράφονται αναλυτικά στις παρακάτω ενότητες.

4.3.2 Εξαγωγή πωλήσεων ανά προϊόν και μήνα

Με τον κώδικα `All_in_Both_Income_Sales.py` γίνεται επεξεργασία των στοιχείων του πρωτογενούς αρχείου `sales_train.csv`. Δέχεται σαν είσοδο τα εξής αρχεία:

(i) Το πρωτογενές αρχείο `items.csv`.

(ii) Το πρωτογενές αρχείο `sales_train.csv`.

Συγκεκριμένα διαβάζοντας μια-μια τις γραμμές του αρχείου, αθροίζονται οι πωλήσεις για κάθε μήνα και για κάθε προϊόν. Επίσης υπολογίζεται και ο συνολικός αριθμός πωλήσεων για όλους τους μήνες που αναγράφονται στο `sales_train.csv` εκτός από τον τελευταίο. Στην περίπτωση μας εμφανίζονται συνολικά οι πρώτοι 33 από τους 34 μήνες που περιέχονται στο πρωτογενές αρχείο, δηλαδή από τον Ιανουάριο του 2013 έως και τον προτελευταίο μηνά που συμφώνα με το πρωτογενές διαθέσιμο αρχείο είναι ο Σεπτέμβριος του 2015. Ο τελευταίος μήνας που είναι εδώ, ο 34^{ος} (Οκτώβριος 2015, κρατιέται ξεχωριστά σαν μήνας ελέγχου των προβλέψεων).

Αν στο αρχείο συμπληρωθούν από την εταιρεία, περισσότεροι μήνες τότε ο κώδικας αυτόματα θα το εντοπίσει και θα διαμορφώσει καταλληλά τα csv αρχεία εξόδου. Επιπλέον υπολογίζονται αντίστοιχα οι εισπράξεις για κάθε μήνα και κάθε προϊόν. Συγκεκριμένα αυτές υπολογίζονται αθροίζοντας τα γινόμενα πωλήσεων των προϊόντων επί την εισπραξη ανά πώληση που αναγράφεται στα πρωτογενές αρχείο. Αυτό γίνεται επειδή σε κάθε προϊόν μπορεί να αλλάξει η τιμή και μέσα στον ίδιο τον μήνα λόγω πιθανών εκπτώσεων, προσφορών ή και πτώσης της τιμής λόγω νέων κυκλοφοριών ομοειδών προϊόντων, από κατάσταση σε κατάσταση. Επίσης υπολογίζεται η μέση τιμή πώλησης ενός προϊόντος σε κάθε μήνα. Ο υπολογισμός αυτός γίνεται διαιρώντας τις συνολικές εισπράξεις για το προϊόν και τον μήνα με τον αριθμό των πωλήσεων.

Ο κώδικας αυτός δημιουργεί τα αρχεία:

(α) **ForAllItemsFINAL.csv**. Περιέχει τις ανά μήνα πωλήσεις για όλους τους μήνες, εκτός από τον τελευταίο και για κάθε προϊόν ξεχωριστά. Καθώς και τις συνολικές πωλήσεις ανά προϊόν και μήνα.

(β) **ForAllItemsFINALincome.csv**. Περιέχει τις ανά μήνα εισπράξεις για όλους τους μήνες, εκτός από τον τελευταίο και για κάθε προϊόν ξεχωριστά. Καθώς και τις συνολικές εισπράξεις από τις πωλήσεις ανά προϊόν.

(γ) **ForAllItemsFINALincome_items.csv**. Αυτό το αρχείο περιέχει τις μέσες τιμές των προϊόντων για κάθε μήνα εκτός από τον τελευταίο. Εάν ένα προϊόν δεν «πούλησε» σε κάποιον μήνα τότε στο αρχείο κρατιέται η μέση τιμή του προηγούμενου μήνα.

4.3.3 Εξαγωγή πωλήσεων ανά προϊόν για τον τελευταίο μήνα

Ο κώδικας **Last_month.py** δέχεται σαν είσοδο τα εξής αρχεία:

- (i) Το πρωτογενές αρχείο **items.csv**.
- (ii) Το πρωτογενές αρχείο **sales_train.csv**.
- (iii) **ForAllItemsFINALincome_items.csv**.

Αμέσως μετά εντοπίζει τις πωλήσεις όλων των προϊόντων μόνο για τον τελευταίο μήνα από το πρωτογενές αρχείο **sales_train.csv** (στην περίπτωση μας τον Οκτώβριο του 2015). Καθώς και τις εισπράξεις που προέρχονται από τις πωλήσεις αυτές. Επίσης εμπεριέχετε και η μέση τιμή πώλησης του κάθε προϊόντος κατά τον τελευταίο μήνα.

Ο κώδικας αυτός δημιουργεί το αρχείο:

ForAllItemsLastMonth.csv. Περιέχει τις πωλήσεις όλων των προϊόντων μόνο για τον τελευταίο μήνα, τις εισπράξεις από τις πωλήσεις αυτές καθώς και τις μέσες τιμές πωλήσεως του κάθε προϊόντος στο μήνα αυτόν.

4.3.4 Εξαγωγή πωλήσεων ανά κατηγορία και μήνα

Ο κώδικας **AllinCategoriesBothIncomeAndSales.py** δέχεται σαν είσοδο τα εξής αρχεία:

- (i) Το πρωτογενές αρχείο **item_categories.csv**.
- (ii) **ForAllItemsFINALincome.csv**
- (iii) **ForAllItemsFINAL.csv**
- (iv) **ForAllItemsLastMonth.csv**

Αυτός ο κώδικας κάνει το ίδιο ακριβώς ότι και ο **All_in_Both_Income_Sales.py**, αλλά με τις 84 κατηγορίες των προϊόντων και επιπλέον εμπεριέχει και τον τελευταίο μήνα (στην περίπτωση μας τον Οκτώβριο του 2015). Επίσης δεν υπολογίζεται η μέση τιμή πώλησης, για την κατηγορία, καθώς σε κάθε κατηγορία υπάγονται πολλαπλά διαφορετικά προϊόντα και το κάθε ένα έχει φυσικά, διαφορετική τιμή. Οπότε δεν έχει νόημα να βρεθεί μέση τιμή.

Ο κώδικας αυτός δημιουργεί τα αρχεία:

- (α) **ForAllCategoriesFINAL.csv**. Περιέχει τις ανά μήνα πωλήσεις για όλους τους μήνες για κάθε κατηγορία ξεχωριστά.
- (β) **ForAllCategoriesFINALincome.csv**. Περιέχει τις ανά μήνα εισπράξεις για όλους τους μήνες για κάθε κατηγορία ξεχωριστά.

4.3.5 Διάρκεια ζωής των προϊόντων

Ο κώδικας **ProductLife.py** υπολογίζει την προσεγγιστική διάρκεια ζωής κάθε προϊόντος. Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) **ForAllItemsFINAL.csv**.
- (ii) **ForAllItemsLastMonth.csv**

Στα πρωτογενή αρχεία της εταιρείας υπάρχει ο κωδικός **date_block_num** που είναι ουσιαστικά ο αύξοντα αριθμός για τους μήνες που περιέχονται στα αρχεία αυτά. Ο αριθμός αυτός έχει μεταφερθεί στο αρχείο που χρησιμοποιείται σαν είσοδος. Η διάρκεια ζωής προσδιορίζεται από τον μήνα όπου εμφανίζονται για πρώτη φορά πωλήσεις του συγκεκριμένου προϊόντος, όπως αυτές φαίνονται στο αρχείο εισόδου, καθώς και από τον τελευταίο μήνα όπου εμφανίζονται πωλήσεις. Η διάρκεια ζωής προκύπτει από την αφαίρεση των **date_block_num** των δυο αυτών μηνών. Εδώ περιλαμβάνεται και ο μήνας ελέγχου (στην περίπτωση μας ο Οκτώβριος του 2015).

Ο κώδικας αυτός δημιουργεί το αρχείο: **ProductLifes.csv**.

Στο αρχείο αυτό κρατούνται τρεις τιμές με το `date_block_num`, αυτές είναι:

- (1) ο μήνας όπου εμφανίζονται για πρώτη φορά πωλήσεις,
- (2) ο μήνας όπου εμφανίζονται για τελευταία φορά πωλήσεις και
- (3) η διάρκεια ζωής του προϊόντος σαν η διαφορά των δύο.

Επιπλέον εκτυπώνονται στην οθόνη μερικά χρήσιμα για τα προϊόντα στοιχεία που συνδέονται με την διάρκεια ζωής τους. Τα στοιχεία αυτά θα αναφερθούν στα αποτελέσματα.

Ένα απόσπασμα του αρχείου αυτού φαίνεται στον πίνακα 4.3-1. Στην πρώτη στήλη (`Item_id`) φαίνεται ο κωδικός του προϊόντος, στην δεύτερη (`Item_category_id`) ο κωδικός της κατηγορίας στην οποία ανήκει το προϊόν. Οι επόμενες δυο στήλες (`First Sale Date` και `Last Sale Date`) είναι οι μήνες όπου πρωτοεμφανίζονται οι πωλήσεις και που σταματούν. Στις επόμενες δυο στήλες (`First Sale Num` και `Last Sale Num`) είναι οι κωδικοί των αντίστοιχων μηνών. Οι διαφορά των δυο αυτών στηλών δίνει την τελευταία `Number of Months`) που είναι η διάρκεια ζωής του συγκεκριμένου προϊόντος.

Πίνακας 4.3-1 Απόσπασμα με διάρκειες ζωής προϊόντων

<code>item_id</code>	<code>item_category_id</code>	<code>First Sale Date</code>	<code>Last Sale Date</code>	<code>First Sale Number</code>	<code>Last Sale Number</code>	<code>Number of months</code>
0		40 Sep 2014	Sep 2014	20	21	1
1		76 Apr 2014	Oct 2014	15	22	7
2		40 Aug 2014	Nov 2014	19	23	4
3		40 Jul 2014	Aug 2014	18	20	2
4		40 Sep 2014	Sep 2014	20	21	1
5		40 Dec 2014	Dec 2014	23	24	1
6		40 Jul 2014	Jul 2014	18	19	1
7		40 Dec 2014	Dec 2014	23	24	1
8		40 Aug 2014	Sep 2014	19	21	2
9		40 Aug 2014	Aug 2014	19	20	1
10		40 Jan 2015	Jan 2015	24	25	1
11		40 Nov 2014	Nov 2014	22	23	1
12		55 Feb 2013	Feb 2013	1	2	1
13		40 Sep 2014	Sep 2014	20	21	1
14		40 Dec 2014	Dec 2014	23	24	1
15		40 Nov 2014	Nov 2014	22	23	1
16		40 Sep 2014	Sep 2014	20	21	1
17		40 Sep 2014	Sep 2014	20	21	1
18		40 Aug 2014	Aug 2014	19	20	1

4.3.6 Διάρκεια ζωής των κατηγοριών

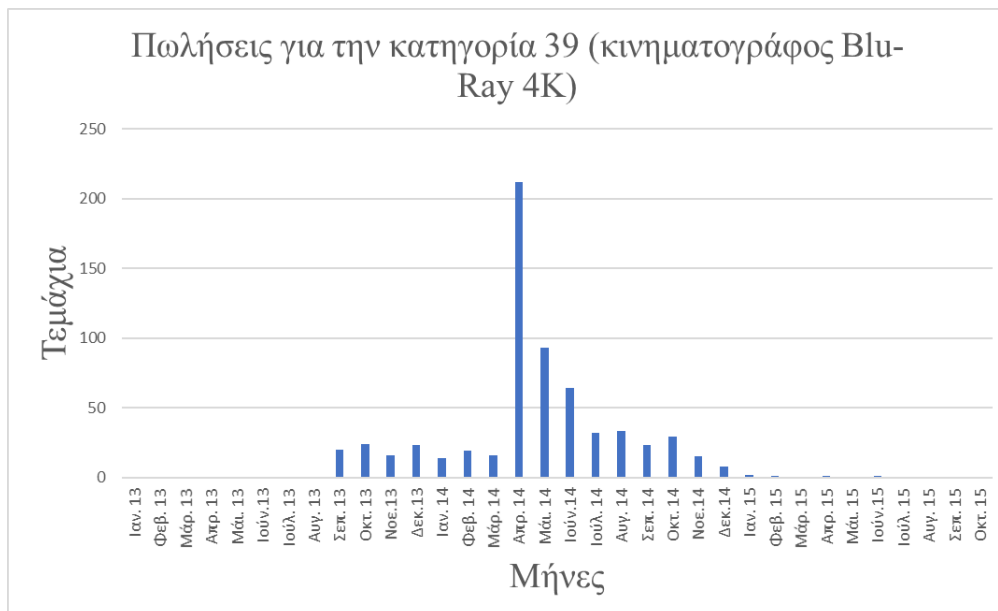
Ο κώδικας `ProductLifeCategories.py` υπολογίζει την προσεγγιστική διάρκεια ζωής κάθε κατηγορίας προϊόντων. Δέχεται σαν είσοδο το εξής αρχείο:

ForAllCategoriesFINAL.csv.

Η διάρκεια ζωής προσδιορίζεται από τον μήνα όπου εμφανίζονται για πρώτη φορά πωλήσεις σε προϊόντα της συγκεκριμένης κατηγορίας και του τελευταίου μήνα όπου εμφανίζονται πωλήσεις. Εδώ περιλαμβάνεται και ο μήνας ελέγχου (στην περίπτωση μας ο Οκτώβριος του 2015).

Ο κώδικας αυτός δημιουργεί το αρχείο:

ProductLifesCategories.csv. Στο αρχείο αυτό κρατιούνται τρεις τιμές εκμεταλλευόμενοι τον `date_block_num` που είναι ο κωδικός αριθμός που αποδίδει σε κάθε μήνα η εταιρεία. Συγκεκριμένα ο μήνας όπου εμφανίζονται για πρώτη φορά πωλήσεις προϊόντων στην συγκεκριμένη κατηγορία, ο μήνας όπου εμφανίζονται για τελευταία φορά πωλήσεις και η διάρκεια ζωής της κατηγορίας σαν η διαφορά των δύο προηγούμενων. Για παράδειγμα η κατηγορία 39 (κινηματογράφος Blu-Ray 4K) πρόκειται για 7 ταινίες Blu-Ray παρουσίασαν πωλήσεις από τον Σεπτέμβριο του 13 (8^{ος}) μέχρι τον Ιούνιο του 15 (30^{ος}), άρα η «διάρκεια ζωής» είναι 22 μήνες.



Εικόνα 4.3-1 Πωλήσεις για την κατηγορία 39 (περιορισμένη διάρκεια ζωής)

4.4 Προβλέψεις πωλήσεων για τις κατηγορίες

4.4.1 Γενικά

Από τις μεθόδους πρόβλεψης επιλέχθηκαν να δοκιμαστούν:

- (α) Μέθοδος Πρόβλεψης Τελευταίας Αξίας (Last Value Forecasting Method, LVFM).
- (β) Απλή Γραμμική Παλινδρόμηση (Simple Linear Regression, SLR).
- (γ) Τετραγωνική παλινδρόμηση (Quadratic Regression, QR).

Ενώ ανάλογα με το είδος των δεδομένων που θα χρησιμοποιηθούν έχουμε δυο κατηγορίες προβλέψεων:

- (i) Προβλέψεις που γίνονται με τα δεδομένα για κάθε κατηγορίας προϊόντων ξεχωριστά.
- (ii) Προβλέψεις που γίνονται με τα δεδομένα για κάθε προϊόν ξεχωριστά.

Οι μέθοδοι αυτές αρχικά θα δοκιμαστούν στις κατηγορίες. Από τις δοκιμές αυτές θα επιλεγεί τελικά η μέθοδος που θα εφαρμοστεί και για το κάθε προϊόν ξεχωριστά. Ας μην ξεχνάμε ότι υπάρχουν 22170 διαφορετικά προϊόντα και μόνο 84 κατηγορίες. Έτσι οι κώδικες που αφορούν στις κατηγορίες εκτελούνται πολύ πιο γρήγορα από τους αντίστοιχους κώδικες για τα προϊόντα.

Η δοκιμή των μεθόδων αυτών παρουσιάζεται στις επόμενες ενότητες.

4.4.2 Μέθοδος πρόβλεψης τελευταίας αξίας (LVFM)

Η μέθοδος LVFM είναι γνωστή και σαν **Naive** (Αφελής) μέθοδος. Σε αυτήν σαν πρόβλεψη πωλήσεων για την επόμενη περίοδο λαμβάνεται η παρούσα τιμή. Φυσικά ο προγραμματισμός αυτής της μεθόδου είναι πολύ απλός.

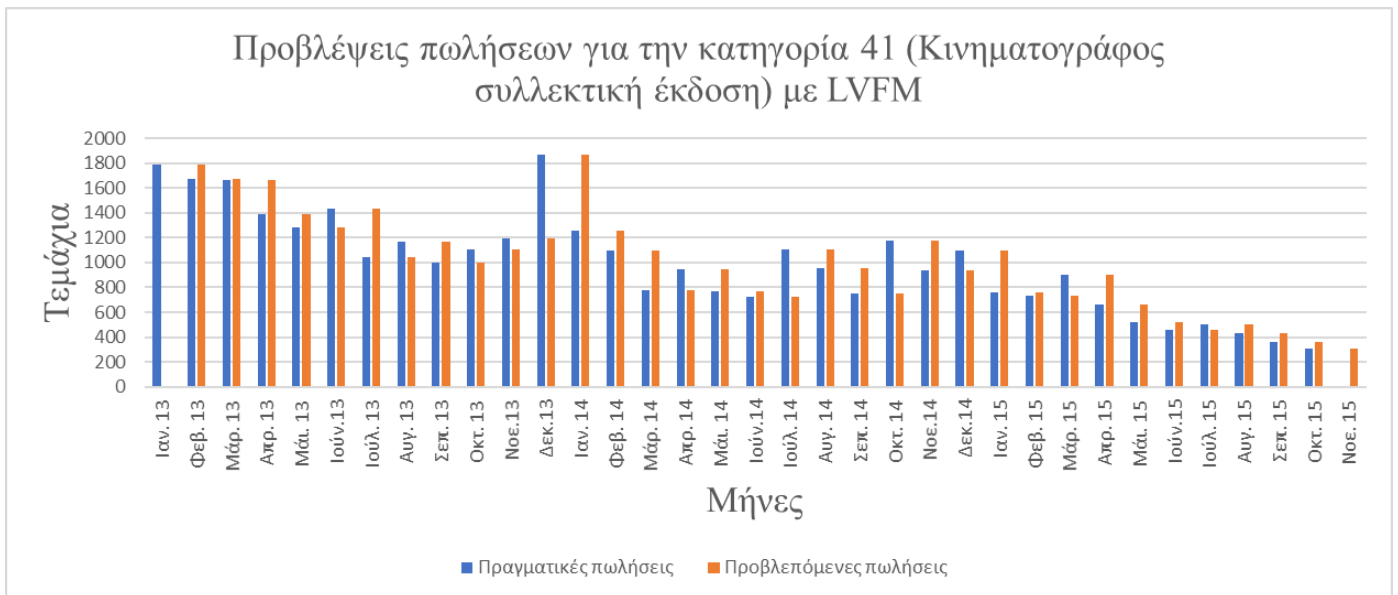
Ο κώδικας `LastValueForecastingMethodForAllMonthscat.py` εφαρμόζει την Μέθοδο Πρόβλεψης Τελευταίας Αξίας (Last Value Forecasting Method). Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) `ForAllCategoriesFINAL.csv`.
- (ii) `ProductLifesCategories.csv`.

Χρησιμοποιεί δηλαδή την τελευταία παρατήρηση στη χρονοσειρά ως πρόβλεψη για την επόμενη περίοδο (μήνα) για κάθε κατηγορία ξεχωριστά.

Ο κώδικας αυτός δημιουργεί το αρχείο:

LastValueForecastingMethodForAllMonthscat.csv. Σε αυτό το αρχείο στην ουσία έχει γίνει μεταφορά των πωλήσεων για κάθε κατηγορία σε κάθε μήνα, κατά ένα μήνα. Π.χ. Οι πραγματικές πωλήσεις του Σεπτεμβρίου 2015 είναι η πρόβλεψη για τον Οκτώβριο του 2015. Στο παρακάτω εικόνα 4.4-1 φαίνονται οι προβλεπόμενες πωλήσεις με αυτή την μέθοδο για την κατηγορία 41.



Εικόνα 4.4-1 Προβλέψεις πωλήσεων για την κατηγορία 41 με LVFM

4.4.3 Απλή Γραμμική Παλινδρόμηση (SLR)

Ο κώδικας **SimpleLinearRegressionFinalcatRepeat_0.py** εφαρμόζει την μέθοδο της Απλής Γραμμικής Παλινδρόμησης (Simple Linear Regression). Χρησιμοποιεί την Μέθοδο Ελαχίστων Τετραγώνων για να βρει την ευθεία η οποία προσαρμόζεται όσο το δυνατόν καλύτερα στην συμπεριφορά των δεδομένων μας. Δέχεται σαν είσοδο τα εξής αρχεία:

(i) ProductLifesCategories.csv.

(ii) ForAllCategoriesFINAL.csv.

Στον κώδικα αυτό εφαρμόζετε για κάθε κατηγορία προϊόντος ξεχωριστά για τους 33 πρώτους μήνες και προβλέπει μέχρι και τον 34^ο (Οκτώβριο του 2015) σταδιακά. Η ευθεία μπορεί να χρησιμοποιεί από 2 έως 4 προηγούμενους μήνες για να κάνει πρόβλεψη για τον επόμενο. Δηλαδή με την χρήση 2 μηνών (SLR2) χρησιμοποιεί για παράδειγμα τα δεδομένα του Ιανουαρίου και του Φεβρουαρίου 2013 για να κάνει πρόβλεψη για τον Μάρτιο του 2013. Μετά με τα στοιχεία του Φεβρουαρίου και του Μαρτίου προβλέπει για τον Απρίλιο κ.ο.κ. Ενώ με την χρήση 3 προηγούμενων περιόδων (μηνών, SLR3), χρησιμοποιεί τα δεδομένα του Ιανουαρίου, Φεβρουαρίου και Μαρτίου 2013 για να κάνει πρόβλεψη για τον Απρίλιο του 2013 κ.ο.κ.

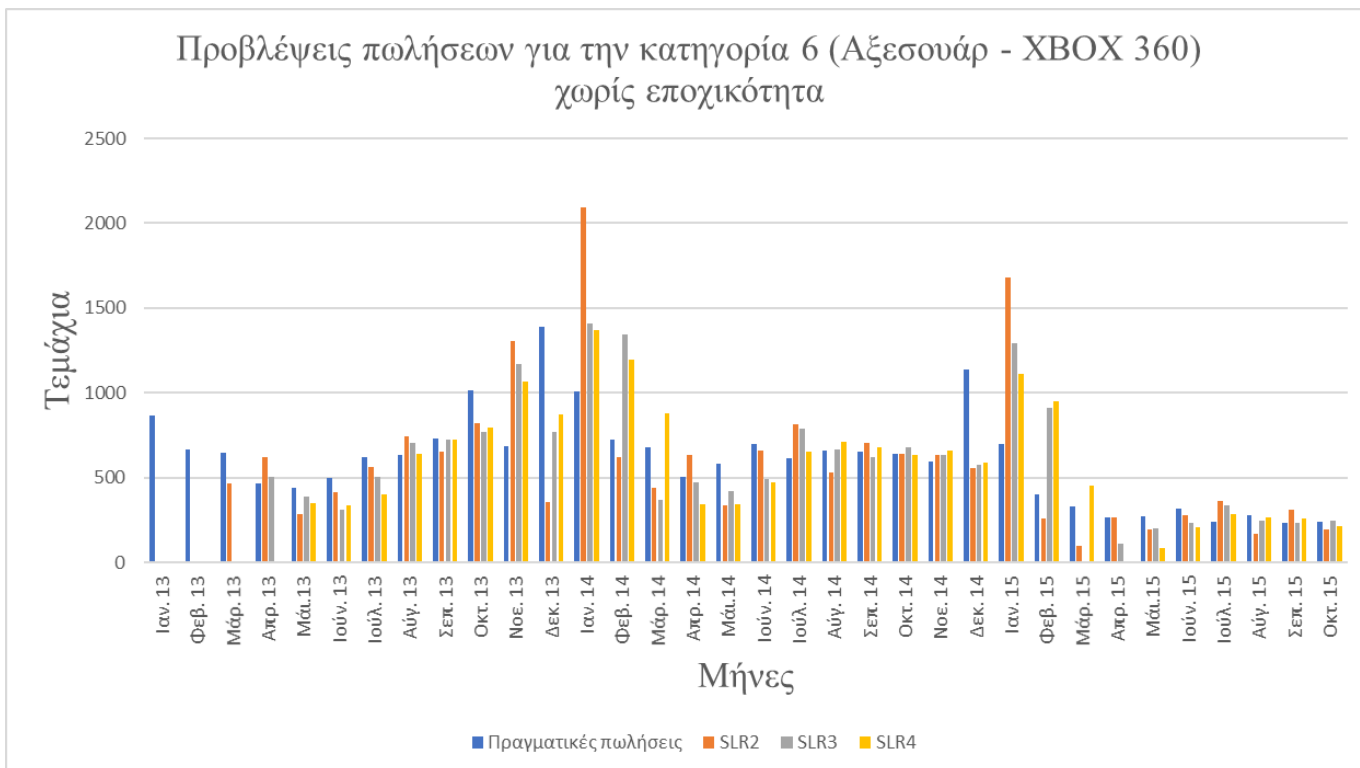
Ο κώδικας αυτός δημιουργεί τα αρχεία:

(α) **SimpleLinearRegression2cat_0.csv**. Σε αυτό το αρχείο γίνονται προβλέψεις από τον 3^ο μήνα έως τον τελευταίο μήνα που είναι ο 34^{ος} που υπάρχει στα πρωτογενή αρχεία (αν στο πρωτογενές αρχείο προσθέσει η εταιρεία και άλλα στοιχεία ο κώδικας εντοπίζει κάθε φορά τον τελευταίο μηνά αυτόματα), χρησιμοποιώντας κάθε φορά τους δυο προηγούμενους μήνες, για κάθε κατηγορία ξεχωριστά. Π.χ. για την πρόβλεψη του Μαΐου 2014 χρησιμοποιούνται ο Μάρτιος και Απρίλιος του 2014.

(β) SimpleLinearRegression3cat_0.csv. Σε αυτό το αρχείο γίνονται προβλέψεις από τον 4^ο μήνα έως τον 34^ο που είναι και ο τελευταίος μήνας που υπάρχει στα πρωτογενή αρχεία, χρησιμοποιώντας κάθε φορά τους τρεις προηγούμενους μήνες, για κάθε κατηγορία ξεχωριστά.

(γ) SimpleLinearRegression4cat_0.csv. Εδώ χρησιμοποιώντας αντίστοιχα τους 4 προηγούμενους μήνες.

Για την **εποχικότητα** που αναφέρετε στον τίτλο των εικόνων 4.4-2 και 4.4-3 θα αναφερθούμε στην επόμενη ενότητα.



Εικόνα 4.4-2 Προβλέψεις πωλήσεων για την κατηγορία 6 με SLR 2,3,4

4.4.4 Τετραγωνική παλινδρόμηση (QR)

Ο κώδικας **QuadraticRegressionAllCategoriesFinalRepeat_0.py** εφαρμόζει το τετραγωνικό (παραβολικό) μοντέλο (Quadratic Regression). Χρησιμοποιεί την Μέθοδο Ελαχίστων Τετραγώνων για να βρει την παραβολή δευτέρου βαθμού η οποία προσαρμόζεται όσο το δυνατόν καλύτερα στην συμπεριφορά των δεδομένων μας. Δέχεται σαν είσοδο τα εξής αρχεία:

(i) ProductLifesCategories.csv.

(ii) ForAllCategoriesFINAL.csv.

Στον κώδικα αυτό εφαρμόζετε για κάθε κατηγορία προϊόντος ξεχωριστά για τους πρώτους 33 μήνες και προβλέπει μέχρι και τον 34^ο (Οκτώβριο του 2015) σταδιακά. Η παραβολή μπορεί να χρησιμοποιεί 4 έως 8 προηγούμενους μήνες για να κάνει πρόβλεψη για τους επόμενους. Μπορεί να επεκταθεί ώστε να χρησιμοποιεί περισσότερους μήνες.

Ο κώδικας αυτός δημιουργεί τα αρχεία

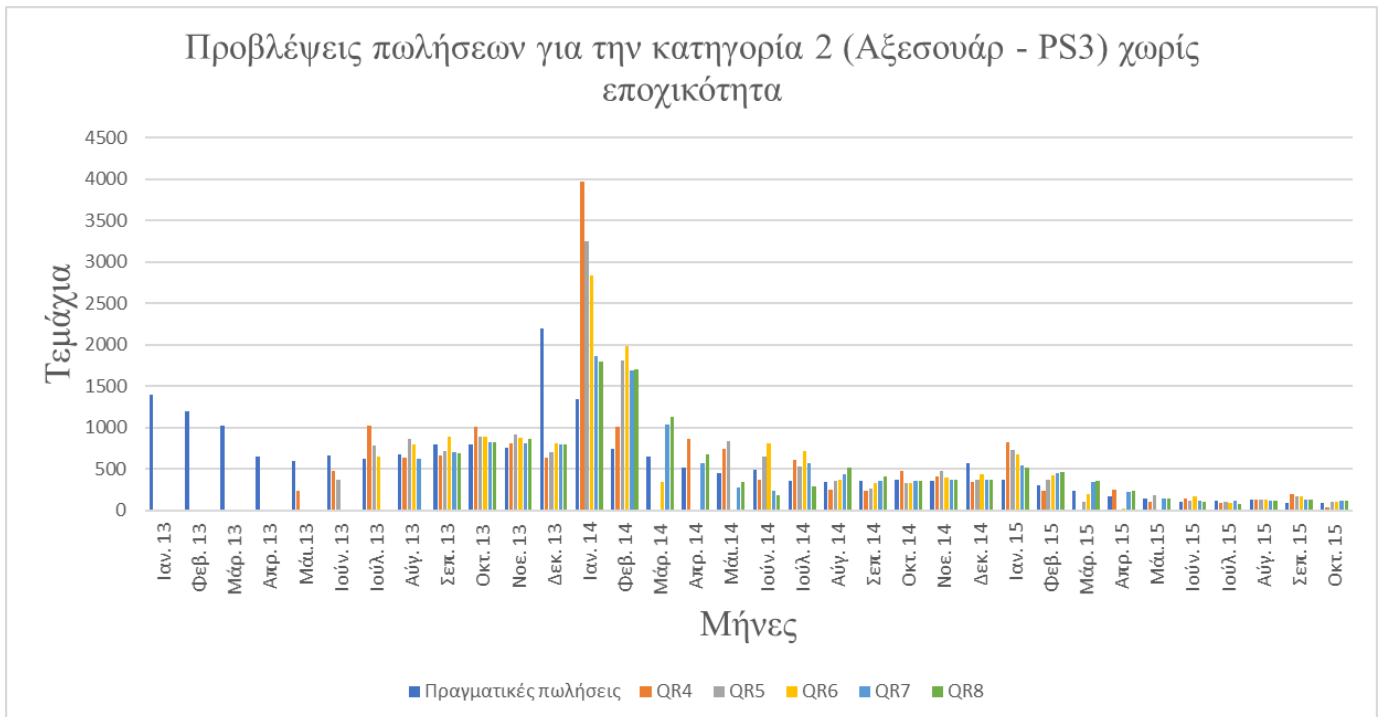
(α) QuadraticRegressionWith4MonthsCAT_0.csv. Σε αυτό το αρχείο γίνονται προβλέψεις από τον 5^ο μήνα έως τον 34^ο που είναι και ο τελευταίος μήνας που υπάρχει στα πρωτογενή αρχεία, χρησιμοποιώντας κάθε φορά τους τέσσερις προηγούμενους μήνες, για κάθε κατηγορία προϊόντων ξεχωριστά. Π.χ. για την πρόβλεψη του Μαΐου 2014 χρησιμοποιούνται ο Φεβρουάριος, ο Μάρτιος και Απρίλιος του 2014.

(β) QuadraticRegressionWith5MonthsCAT_0.csv. Σε αυτό το αρχείο γίνονται προβλέψεις από τον 6^ο μήνα έως και τον 34^ο που είναι και ο τελευταίος μήνας που υπάρχει στα πρωτογενή αρχεία, χρησιμοποιώντας κάθε φορά τους πέντε προηγούμενους μήνες, για κάθε προϊόν ξεχωριστά.

(γ) QuadraticRegressionWith6MonthsCAT_0.csv. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 6 προηγούμενους μήνες.

(δ) QuadraticRegressionWith7MonthsCAT_0.csv. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 7 προηγούμενους μήνες.

(ε) QuadraticRegressionWith8MonthsCAT_0.csv. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 8 προηγούμενους μήνες.



Εικόνα 4.4-3 Προβλέψεις πωλήσεων για την κατηγορία 2 με QR 4,5,6,7,8

4.5 Εποχικότητα

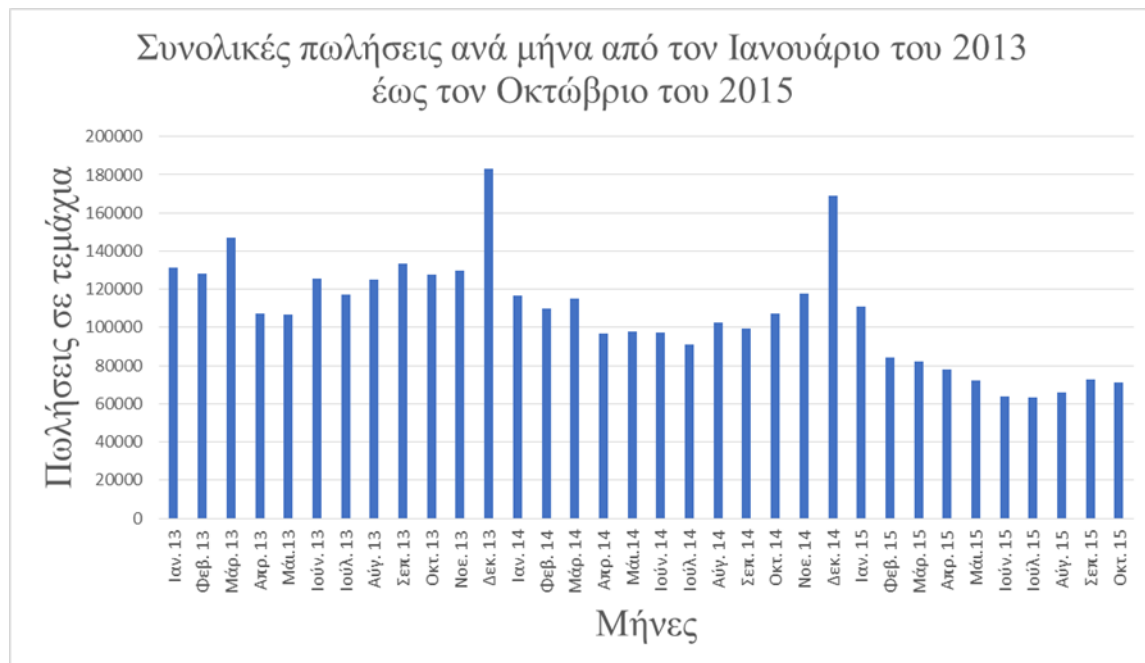
4.5.1 Γενικά

Τα προϊόντα που πουλά η παραπάνω επιχείρηση περιλαμβάνουν μεγάλο αριθμό από είδη δώρων και αναμνηστικών, μουσική σε CD, MP3 (ακόμα και βινύλιο για τους λάτρεις της παράδοσης), κινηματογραφικές ταινίες DVD, Blue-ray, βιβλία ψηφιακά ακουστικά και κυρίως παιχνιδομηχανές (κονσόλες παιχνιδιών) καθώς και αξεσουάρ για αυτές, επίσης παιχνιδιών και αξεσουάρ για PC. Τέλος περιλαμβάνουν λογισμικά είτε εκπαιδευτικά είτε για το γραφείο, το σπίτι κ.λπ.

Είναι λογικό ότι οι πωλήσεις των προϊόντων θα αυξάνονται τους μήνες Νοέμβριο Δεκέμβριο και Ιανουάριο (εορτές Χριστουγέννων). Αυτό προδιαθέτει για την ύπαρξη εποχικότητας. Ταυτόχρονα όμως τα προϊόντα αυτά «έχουν περιορισμένη διάρκεια ζωής» καθώς εμφανίζονται συνέχεια καινούργια προϊόντα στην αγορά.

Αυτή η συνεχής ανανέωση των προϊόντων θα μπορούσε βέβαια να χαρακτηριστεί από «κυκλικότητα». Το πρόβλημα είναι ότι για τον προσδιορισμό της θα απαιτούνταν στοιχεία για αρκετά περισσότερα από δυο χρόνια. Έτσι δεν μπορεί προς το παρόν να εκτιμηθεί.

Οι συνολικές πωλήσεις ανά μήνα φαίνονται στην εικόνα 4.5 - 1. Στην εικόνα αυτή φαίνεται να υπάρχει εποχικότητα με έξαρση των πωλήσεων ειδικά κάθε Δεκέμβριο. Ταυτόχρονα φαίνεται και μια πτωτική τάση.



Εικόνα 4.5-1 Συνολικές πωλήσεις ανά μήνα

Ο έλεγχος για την ύπαρξη τάσης και εποχικότητας όμως γίνεται με την χρήση της αυτοσυσχέτισης που παρουσιάζεται στην παρακάτω ενότητα.

4.5.2 Αυτοσυσχέτιση

Για την αυτοσυσχέτιση χρησιμοποιήθηκε η έτοιμη συνάρτηση: `statsmodels.api.tsa.acf(table, nlags = numofmonths)` της `python` από την βιβλιοθήκη της `statsmodels.api` όπου το `table` = ο πίνακας δεδομένων και `numofmonths` ο αριθμός των περιόδων (Απόσταση)

Για την αυτοσυσχέτιση χρησιμοποιείται ο κώδικας:

AutoCorrelationFinalCategories_0.py.

Αυτός δέχεται σαν είσοδο δεδομένων το αρχείο: **ForAllCategoriesFINAL.csv.** και βρίσκει την Αυτοσυσχέτιση (Auto Correlation) σε όλες τις κατηγορίες των προϊόντων. Με απλά λόγια υπολογίζει τους συντελεστές αυτοσυσχέτισης, δηλαδή την συνάρτηση αυτοσυσχέτισης (ACF) για όλους τους μήνες όπου υπάρχουν δεδομένα.

Ο κώδικας αυτός δημιουργεί τα εξής αρχεία:

(α) **AutoCorrelationCategories.csv.** Στο οποίο αναγράφονται οι συντελεστές αυτοσυσχέτισης των σημείων για κάθε μια κατηγορία.

(β) **AutoCorrelationCategoriesPassNoiseLimit.csv.** Στο οποίο αναγράφονται μόνο οι συντελεστές συσχέτισης για κάθε μια κατηγορία, που ξεπερνούν τα όρια λευκού θορύβου.

(γ) **AutoCorrelationCategoriesIsItNoice.csv.** Στο οποίο αναγράφεται σε ποιες κατηγορίες είναι λευκός θόρυβος και ποιες δεν είναι λευκός θόρυβος με `true/false` αντίστοιχα. Αρχεία που θα χρησιμοποιηθούν αργότερα.

Στις παρακάτω εικόνες 4.5-2 και 4.5-3 δίνονται οι γραφικές παραστάσεις των συναρτήσεων αυτοσυσχέτισης δυο κατηγοριών της 25 (Παιχνίδια - Αξεσουάρ για παιχνίδια και της 21 (Παιχνίδια – PSP) καθώς επίσης και οι πωλήσεις τους.



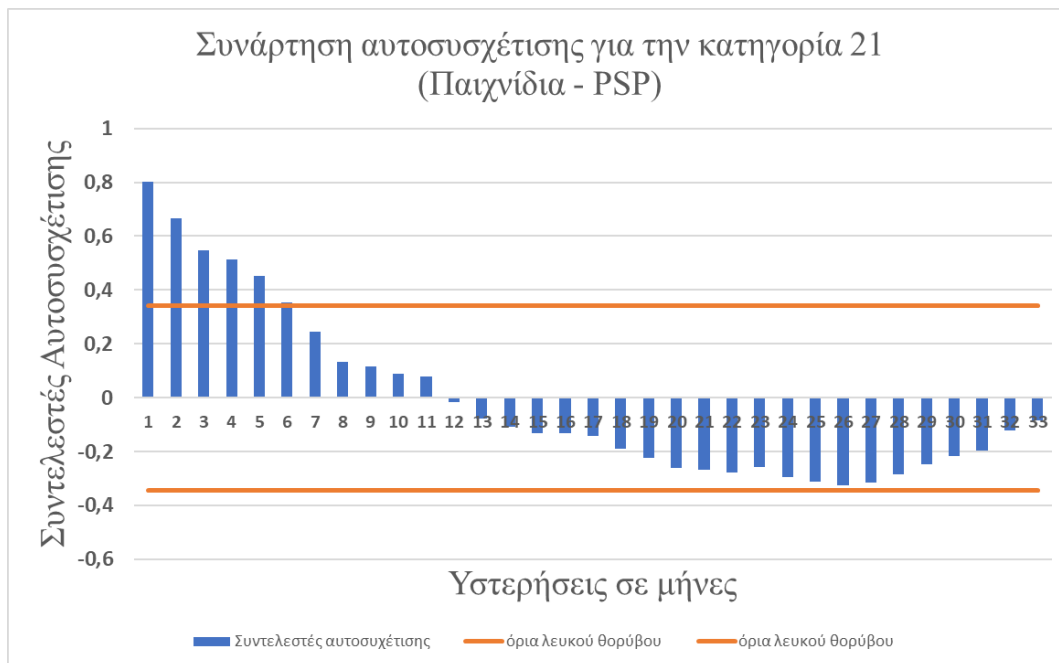
Εικόνα 4.5-2 Πωλήσεις για την κατηγορία 25



Εικόνα 4.5-3 Συνάρτηση αυτοσυσχέτισης για την κατηγορία 25



Εικόνα 4.5-4 Πωλήσεις για την κατηγορία 21



Εικόνα 4.5-5 Συναρτήση αυτοσυσχέτισης για την κατηγορία 21

4.5.3 Εφαρμογή εποχικότητας

4.5.3.1 Γενικά

Για την εποχικότητα η διόρθωση των προβλέψεων για κάθε μήνα γίνεται λαμβάνοντας υπόψη το σχετικό σφάλμα από τον αντίστοιχο μήνα του προηγούμενου έτους.

Το σφάλμα αυτό δίνεται από την σχέση (βλέπε ενότητα 2.4):

$$\frac{R_t - F_t}{R_t + 1}$$

Οπού:

R_t οι πραγματικές πωλήσεις για τον μήνα t του προηγούμενου έτους

και F_t οι προβλεπόμενες πωλήσεις για τον μήνα t του προηγούμενου έτους.

Το +1 στον παρονομαστή είναι για την περίπτωση μηδενικών πραγματικών πωλήσεων. Αν δεν γινόταν δεν θα μπορούσε να γίνει η διαίρεση (μηδενικός παρονομαστής).

Έτσι οι προβλεπόμενες πωλήσεις για τον ίδιο μήνα του επόμενου έτους θα είναι:

$$F_{(t+T)Final} = \left(1 + a \cdot \frac{R_t - F_t}{R_t + 1} \right) \cdot F_{t+T}$$

Όπου F_{t+T} είναι οι προβλεπόμενες πωλήσεις για τον συγκεκριμένο μήνα του επομένου έτους, χωρίς την εποχικότητα

$F_{(t+T)Final}$ η τελική πρόβλεψη και

a : ένας επιπλέον συντελεστής διόρθωσης που μετά από δοκιμές επιλέχτηκε να είναι:

$$a = \frac{1}{3} = 0,333$$

Η επιλογή αυτή έγινε με κριτήριο την ελαχιστοποίηση του μέσου ποσοστιαίου σφάλματος MAPE που αναλύεται στις παρακάτω ενότητες.

4.5.3.2 Γραμμική παλινδρόμηση (SLR) με εποχικότητα

Ο κώδικας `SimpleLinearRegressionFinalcatSessionRepeat_0.py` χρησιμοποιεί την Μέθοδο Ελαχίστων Τετραγώνων για να βρει την ευθεία η οποία προσαρμόζεται όσο το δυνατόν καλύτερα στην συμπεριφορά των δεδομένων μας. Δέχεται σαν είσοδο τα εξής αρχεία:

(i) `ProductLifesCategories.csv`.

(ii) `ForAllCategoriesFINAL.csv`.

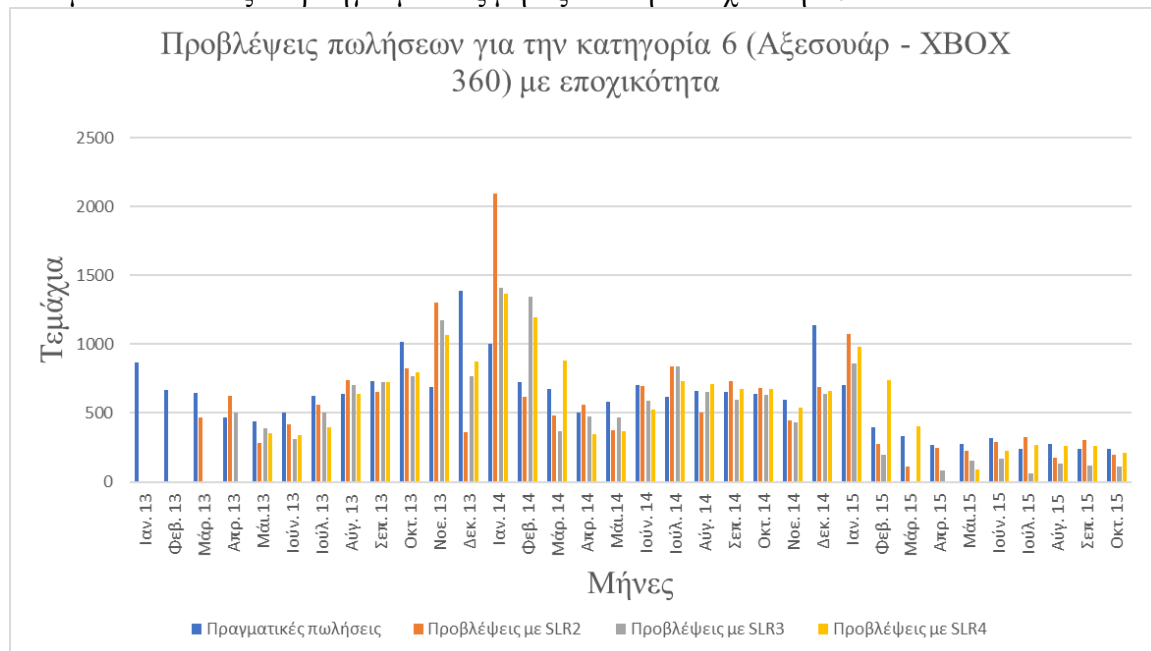
Στον κώδικα αυτό εφαρμόζετε για κάθε κατηγορία προϊόντος ξεχωριστά για τους 33 πρώτους μήνες και προβλέπει μέχρι και τον 34^ο (Οκτώβριο του 2015) σταδιακά, αλλά ταυτόχρονα χρησιμοποιεί και την **εποχικότητα** για κάθε κατηγορία προϊόντων ξεχωριστά.

Ο κώδικας αυτός δημιουργεί τα αρχεία:

(α) `SimpleLinearRegression2catSession_0.csv.csv`. Σε αυτό το αρχείο γίνονται προβλέψεις από τον 3^ο μήνα έως τον 34^ο που είναι και ο τελευταίος μήνας που υπάρχει στα πρωτογενή αρχεία, χρησιμοποιώντας κάθε φορά τους δυο προηγούμενους μήνες, για κάθε κατηγορία ξεχωριστά και επιπλέον χρησιμοποιείτε και η εποχικότητα.

(β) `SimpleLinearRegression3catSession_0.csv`. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 3 προηγούμενους μήνες και την εποχικότητα.

(γ) `SimpleLinearRegression4catSession_0.csv`. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 4 προηγούμενους μήνες και την εποχικότητα.



Εικόνα 4.5-6 Προβλέψεις πωλήσεων για την κατηγορία 6 με SLR 2,3,4

4.5.3.3 Τετραγωνική παλινδρόμηση (QR) με εποχικότητα

Ο κώδικας `QuadraticRegressionAllCategoriesFinalSessionsRepeat_0.py` κάνει το ίδιο με τον προηγούμενο αλλά χρησιμοποιώντας την τετραγωνική παλινδρόμηση και εδώ χρησιμοποιείται και η **εποχικότητα** για κάθε κατηγορία προϊόντων ξεχωριστά. Δέχεται σαν είσοδο τα εξής αρχεία:

(i) `ProductLifesCategories.csv`.

(ii) `ForAllCategoriesFINAL.csv`.

Ο κώδικας αυτός δημιουργεί τα αρχεία

(α) QuadraticRegressionWith4MonthsCATSesson_0.csv. Σε αυτό το αρχείο γίνονται προβλέψεις από τον 5^ο μήνα έως τον 34^ο που είναι και ο τελευταίος μήνας που υπάρχει στα πρωτογενή αρχεία, χρησιμοποιώντας κάθε φορά τους τέσσερις προηγούμενους μήνες, για κάθε κατηγορία ξεχωριστά.

(β) QuadraticRegressionWith5MonthsCATSesson_0.csv.

Εδώ χρησιμοποιώντας αντίστοιχα τους 5 προηγούμενους μήνες.

(γ) QuadraticRegressionWith6MonthsCATSesson_0.csv.

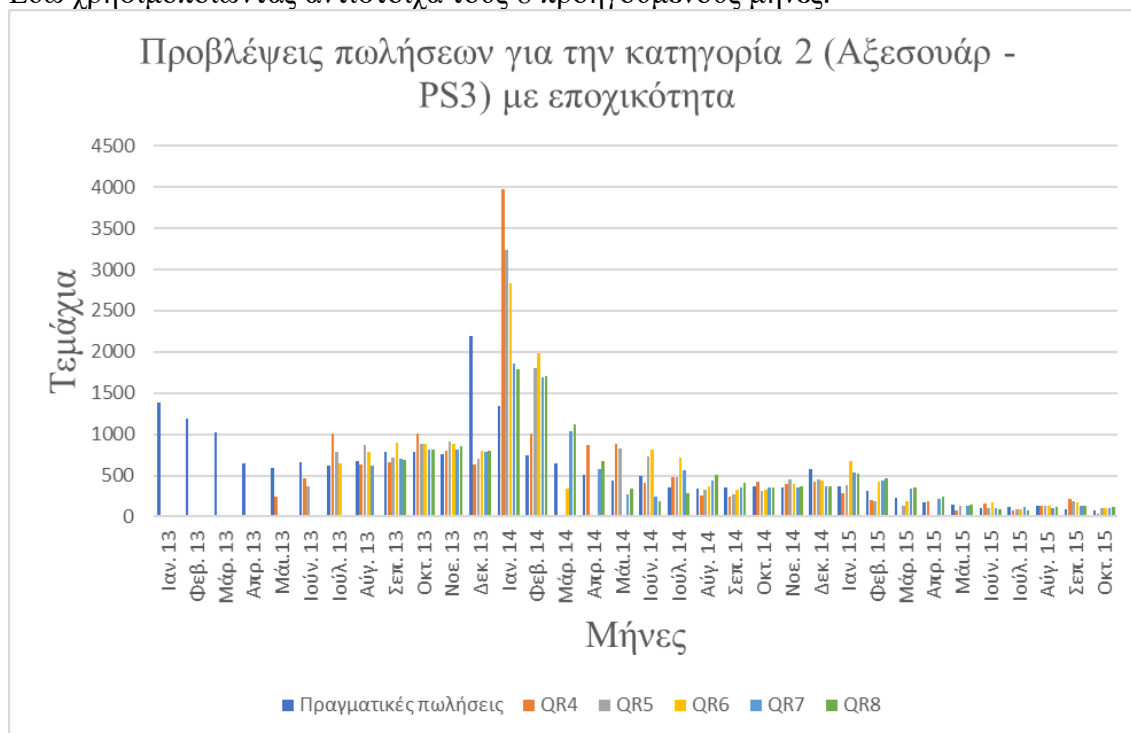
Εδώ χρησιμοποιώντας αντίστοιχα τους 6 προηγούμενους μήνες.

(δ) QuadraticRegressionWith7MonthsCATSesson_0.csv.

Εδώ χρησιμοποιώντας αντίστοιχα τους 7 προηγούμενους μήνες.

(ε) QuadraticRegressionWith8MonthsCATSesson_0.csv.

Εδώ χρησιμοποιώντας αντίστοιχα τους 8 προηγούμενους μήνες.



Εικόνα 4.5-7 Προβλέψεις πωλήσεων για την κατηγορία 2 με QR 4,5,6,7,8

4.6 Μέσο ποσοστιαίο σφάλμα (MAPE)

4.6.1 Γενικά

Το μέσο ποσοστιαίο σφάλμα (**Mean Average Percentage Error** ή **MAPE**) υπολογίζεται από την σχέση που είδαμε στην παράγραφο 2.4.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t + 1} \right|$$

Όπου:

n : το πλήθος των δειγμάτων

A_t : η πραγματική τιμή

F_t : η προβλεπόμενη τιμή

Στην παραπάνω σχέση το τμήμα $\left| \frac{A_t - F_t}{A_t + 1} \right|$ είναι το σχετικό σφάλμα για κάθε πρόβλεψη, ο πολλαπλασιασμός με το 100% το μετατρέπει σε «επί τοις εκατό» ποσοστό, ενώ η διαίρεση με το πλήθος n το μετατρέπει σε μέση τιμή.

4.6.2 Συντελεστές Βαρύτητας

Γενικά στον υπολογισμό των σφαλμάτων υπάρχει ένα στατιστικό πρόβλημα. Τα προϊόντα της εταιρείας φυσικά δεν πωλούνται όλα στην ίδια τιμή. Είναι αυτονόητο ότι οι προβλέψεις πωλήσεων σίγουρα συνδέονται και με τις αναμενόμενες εισπράξεις από την πώληση των προϊόντων. Επομένως και το μέσο ποσοστιαίο σφάλμα θα πρέπει με κάποιο τρόπο να διαφοροποιείται ανάλογα με τις εισπράξεις από τις πωλήσεις της κάθε κατηγορίας ή προϊόντος. Με απλά λόγια είναι διαφορετικό να κάνει κανείς μεγάλο σφάλμα στην πρόβλεψη πώλησης αντικειμένων χαμηλής αξίας από ότι να κάνει το ίδιο σφάλμα στην πρόβλεψη πώλησης αντικειμένων μεγάλης αξίας. Αντί λοιπόν για τον υπολογισμό της απλής μέσης τιμής του MAPE μπορεί να χρησιμοποιηθεί ένας μέσος όρος με **συντελεστές βαρύτητας**.

Σαν **συντελεστές βαρύτητας** για τον υπολογισμό του MAPE για τα προϊόντα μπορεί να χρησιμοποιηθεί η σχέση:

$$w_i = \frac{\text{Εισπράξεις ανά προϊόν και μήνα}}{\text{Συνολικές εισπράξεις του μήνα από όλα τα προϊόντα}}$$

Ενώ για τις κατηγορίες:

$$w_i = \frac{\text{Εισπράξεις ανά κατηγορία και μήνα}}{\text{Συνολικές εισπράξεις του μήνα από όλες τις κατηγορίες}}$$

Για το συνολικό MAPE, για κάθε μήνα και μέθοδο πρόβλεψης, η σχέση:

$$MAPE_j = \sum_{i=m}^k \frac{|x_i - y_i|}{x_i + 1} \cdot 100 \cdot w_i$$

Οπού j οι μήνες

i τα προϊόντα από το πρώτο που έχει πωληθεί (m) μέχρι το τελευταίο (k)

x_i οι πραγματικές πωλήσεις του προϊόντος i τον συγκεκριμένο μήνα j

y_i οι προβλεπόμενες πωλήσεις του προϊόντος i τον συγκεκριμένο μήνα j

w_i ο συντελεστής βαρύτητας του προϊόντος i τον συγκεκριμένο μήνα j

Ενώ για τον υπολογισμό του συνολικού MAPE κάθε μεθόδου παίρνοντας υπόψιν όλους τους μήνες, να χρησιμοποιηθεί ο **σταθμισμένος μέσος** όρος με συντελεστές:

$$w_j = \frac{\text{Συνολικές εισπράξεις όλου του μήνα j}}{\text{Συνολικές εισπράξεις για όλους τους μήνες και για όλα τα προϊόντα}}$$

Φυσικά το άθροισμα όλων των συντελεστών βαρύτητας πρέπει να είναι μονάδα.

Σε κάποιους μήνες είναι δυνατόν, ενώ δεν υπάρχουν πωλήσεις να προβλέπονται από την αντίστοιχη μέθοδο πρόβλεψης. Οπότε το σφάλμα MAPE χωρίς την βαρύτητα είναι μη μηδενικό, αλλά αφού δεν υπάρχουν εισπράξεις στον αντίστοιχο μήνα ο συντελεστής βαρύτητας θα είναι $w_i = 0$, έτσι μηδενίζεται και το σφάλμα MAPE, όποτε ενδεχομένως να μην προκύπτουν απόλυτα ειλικρινά αποτελέσματα.

Αυτό στις κατηγορίες σπανία συμβαίνει, αλλά συμβαίνει πιο συχνά στα προϊόντα. Συγκεκριμένα στις προβλέψεις ανά κατηγορία και μήνα οι συχνότητες εμφάνισης των περιπτώσεων όπου έχουμε μηδενικές πωλήσεις και πρόβλεψη διαφορετική από το μηδέν φαίνονται στον παρακάτω πίνακα 4.6-1. Στο ίδιο πίνακα στο τέλος του φαίνονται τα αντίστοιχα για τα προϊόντα. Είναι φανερό ότι στα προϊόντα συμβαίνει πολύ συχνότερα. Αυτό έχει σαν αποτέλεσμα ο υπολογισμός των MAPE με τον συγκεκριμένο συντελεστή να είναι λιγότερο ειλικρινής.

Πίνακας 4.6-1 Μηδενικές πωλήσεις, μη μηδενικές προβλέψεις

ΚΑΤΗΓΟΡΙΕΣ					
Μέθοδος	Πλήθος κατηγοριών ή προϊόντων που έγινε πρόβλεψη	Πλήθος μηνών που έγινε πρόβλεψη	Πλήθος προβλέψεων	Περιπτώσεις όπου πωλήσεις = 0 και υπάρχει πρόβλεψη	Ποσοστό %
LVFM	83	33	2739	0	0
SLR2	83	32	2656	0	0
SLR3	83	31	2573	1	0,03886514
SLR4	83	30	2490	0	0
QR4	83	30	2490	4	0,16064257
QR5	83	29	2407	5	0,20772746
QR6	83	28	2324	5	0,2151463
QR7	83	27	2241	4	0,17849174
QR8	83	26	2158	2	0,09267841
ΠΡΟΪΟΝΤΑ					
LVFM	22164	33	731412	45483	6,21851979
SLR4	22164	30	664920	46266	6,9581303

Έτσι στα προϊόντα μπορεί να χρησιμοποιηθεί ένας λίγο διαφορετικός συντελεστής βαρύτητας (**Συντελεστής βαρύτητας 1**), ο οποίος είναι:

$$w_i = \frac{\text{Μέση τιμή είσπραξης του προϊόντος στον αντίστοιχο μήνα}}{\text{Άθροισμα των μέσων τιμών είσπραξης όλων των προϊόντων στον αντίστοιχο μήνα}}$$

Στον αριθμητή χρησιμοποιείται η μέση τιμή του προϊόντος γιατί στον ίδιο μήνα, μερικές φορές, διαφορετικά καταστήματα προσφέρουν τα ίδια προϊόντα σε διαφορετικές τιμές.

Ο υπολογισμός γίνεται με τον κώδικα **ForAllItemsFINALincome_items.csv** που περιγράφεται παραπάνω.

Αν σε κάποιο μήνα δεν υπάρχει τιμή τότε αντικαθιστάτε με την μέση τιμή που είχε στον προηγούμενο μήνα. Αυτή παραμένει μέχρι να εμφανιστεί σε κάποιο επόμενο μηνά καινούργια μέση τιμή.

Αντίστοιχα μπορεί να εφαρμοστεί και στις κατηγορίες, αν και εκεί θα υπάρχει το πρόβλημα των πολλών και πολύ διαφορετικών τιμών προϊόντων μέσα στην ίδια κατηγορία, οπότε η μέση τιμή δεν θα είναι αντιπροσωπευτική.

Μια άλλη προσέγγιση για την άρση του προβλήματος των μηδενικών πραγματικών πωλήσεων ενώ υπάρχουν προβλεπόμενες, είναι να υπολογιστούν λίγο διαφορετικά οι συντελεστές βαρύτητας (**Συντελεστής βαρύτητας 2**) ειδικά για τις κατηγορίες.

$$w_i = \frac{\max(\hat{Y}_i, Y_i)}{\sum_{i=1}^n \max(\hat{Y}_i, Y_i)}$$

Όπου: w_i : ο συντελεστής βαρύτητας της i κατηγορίας στον συγκεκριμένο μήνα.

\hat{Y}_i : οι προβλεπόμενες πωλήσεις της i κατηγορίας στον συγκεκριμένο μήνα.

Y_i : οι πραγματικές πωλήσεις της i κατηγορίας στον συγκεκριμένο μήνα.

n : το πλήθος των κατηγοριών.

Στα προϊόντα μπορεί να επανέλθει η μέση τιμή είσπραξης ανά προϊόν και ο αντίστοιχος συντελεστής βαρύτητας να περιέχει και την είσπραξη από το αντίστοιχο προϊόν σε κάθε μήνα. Συγκεκριμένα για τα προϊόντα ο συντελεστής βαρύτητας 2 μπορεί να γίνει:

$$w_i = \frac{\max(\hat{Y}_i, Y_i) \cdot C_{ave,i}}{\sum_{i=1}^n (\max(\hat{Y}_i, Y_i) \cdot C_{ave,i})}$$

Όπου: w_i : ο συντελεστής βαρύτητας του i προϊόντος στον συγκεκριμένο μήνα.

\hat{Y}_i : οι προβλεπόμενες πωλήσεις του i προϊόντος στον μήνα.

Y_i : οι πραγματικές πωλήσεις του i προϊόντος στον μήνα.

n : πλήθος των προϊόντων στον συγκεκριμένο μήνα.

$C_{ave,i}$: η μέση τιμή είσπραξης ανά προϊόν και ανά μήνα

4.6.3 Υπολογισμός MAPE για τις προβλέψεις στις κατηγορίες

4.6.3.1 Μέσες τιμές MAPE για προβλέψεις χωρίς εποχικότητα

Ο υπολογισμός γίνεται με τον κώδικα **Mapecat_0.py**. Αυτός ο κώδικας παίρνει σαν είσοδο τα εξής αρχεία:

- (i) ProductLifesCategories.csv
- (ii) ForAllCategoriesFINAL.csv
- (iii) LastValueForecastinMethodForAllMonthscat.csv
- (iv) SimpleLinearRegression2cat_0.csv
- (v) SimpleLinearRegression3cat_0.csv
- (vi) SimpleLinearRegression4cat_0.csv
- (vii) QuadraticRegressionWith4MonthsCAT_0.csv
- (viii) QuadraticRegressionWith5MonthsCAT_0.csv
- (ix) QuadraticRegressionWith6MonthsCAT_0.csv
- (x) QuadraticRegressionWith7MonthsCAT_0.csv
- (xi) QuadraticRegressionWith8MonthsCAT_0.csv

Από αυτά τα αρχεία ο κώδικας θα εκτελέσει τον υπολογισμό του **MAPE** για κάθε κατηγορία προϊόντος σε κάθε μήνα και σε κάθε μια από τις εξής μεθόδους:

- (i) Last Value Forecasting Method.
- (ii) Simple Linear Regression με χρήση από 2 έως 4 μήνες δεδομένων.
- (iii) Quadratic Regression με χρήση από 4 έως 8 μήνες δεδομένων.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) MapeAllcat_0.csv. Στο οποίο καταγράφονται οι μέσοι όροι για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCounterscat_0.csv. Στο οποίο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.6.3.2 Μέσες τιμές MAPE για προβλέψεις με εποχικότητα

Ο υπολογισμός γίνεται με τον κώδικα **MapecatSessions_0.py**. Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifesCategories.csv
- (ii) ForAllCategoriesFINAL.csv
- (iii) LastValueForecastinMethodForAllMonthscat.csv

- (iv) SimpleLinearRegression2catSesson_0.csv
- (v) SimpleLinearRegression3catSesson_0.csv
- (vi) SimpleLinearRegression4catSesson_0.csv
- (vii) QuadraticRegressionWith4MonthsCATSesson_0.csv
- (viii) QuadraticRegressionWith5MonthsCATSesson_0.csv
- (ix) QuadraticRegressionWith6MonthsCATSesson_0.csv
- (x) QuadraticRegressionWith7MonthsCATSesson_0.csv
- (xi) QuadraticRegressionWith8MonthsCATSesson_0.csv

Με αυτά τα αρχεία θα εκτελέσει όπως και ο προηγούμενος κώδικας το MAPE αλλά αυτή την φορά θα εμπεριέχεται και η εποχικότητα μέσα στα αρχεία που χρησιμοποιούνται.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) MapeAllcatSessons_0.csv. Στο οποίο καταγράφονται οι απλοί μέσοι όροι για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCounterscatSessons_0.csv. Στο οποίο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.6.3.3 Μέσες τιμές MAPE με βαρύτητα 1 χωρίς εποχικότητα

Εδώ ο υπολογισμός γίνεται με τους συντελεστές βαρύτητας 1. Ο υπολογισμός γίνεται με τον κώδικα **Mapecat_0W.py**. Αυτός ο κώδικας δέχεται σαν εισόδους τα εξής αρχεία:

- (i) ProductLifesCategories.csv
- (ii) ForAllCategoriesFINAL.csv
- (iii) LastValueForecastinMethodForAllMonthscat.csv
- (iv) SimpleLinearRegression2cat_0.csv
- (v) SimpleLinearRegression3cat_0.csv
- (vi) SimpleLinearRegression4cat_0.csv
- (vii) QuadraticRegressionWith4MonthsCAT_0.csv
- (viii) QuadraticRegressionWith5MonthsCAT_0.csv
- (ix) QuadraticRegressionWith6MonthsCAT_0.csv
- (x) QuadraticRegressionWith7MonthsCAT_0.csv
- (xi) QuadraticRegressionWith8MonthsCAT_0.csv
- (xii) ForAllCategoriesFINALincome.csv

Αυτός θα εκτελέσει τον υπολογισμό όχι ενός απλού μέσου όρου όλων των κατηγοριών, αλλά του μέσου MAPE με συντελεστές βαρύτητας 1 για κάθε κατηγορία, δεν λαμβάνει υπόψιν την εποχικότητα.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) MapeAllcat_0W.csv. Στο οποίο καταγράφονται τα αποτελέσματα για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCounterscat_0W.csv. Στο οποίο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.6.3.4 Μέσες τιμές MAPE με βαρύτητα 1 και εποχικότητα

Εδώ ο υπολογισμός γίνεται πάλι με τους συντελεστές βαρύτητας 1. Ο υπολογισμός γίνεται με τον κώδικα **MapecatSessons_0W.py**. Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifesCategories.csv

- (ii) ForAllCategoriesFINAL.csv
- (iii) LastValueForecastinMethodForAllMonthscat.csv
- (iv) SimpleLinearRegression2cat_0.csv
- (v) SimpleLinearRegression3cat_0.csv
- (vi) SimpleLinearRegression4cat_0.csv
- (vii) QuadraticRegressionWith4MonthsCAT_0.csv
- (viii) QuadraticRegressionWith5MonthsCAT_0.csv
- (ix) QuadraticRegressionWith6MonthsCAT_0.csv
- (x) QuadraticRegressionWith7MonthsCAT_0.csv
- (xi) QuadraticRegressionWith8MonthsCAT_0.csv
- (xii) ForAllCategoriesFINALincome.csv

Αυτός ο κώδικας θα κάνει τον υπολογισμό του MAPE αλλά αυτήν την φορά θα συνυπολογιστεί επιπλέον η μέση τιμή για τις κατηγορίες με την βαρύτητα 2, ενώ λαμβάνει υπόψιν και την εποχικότητα.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) MapeAllcatSessons_0W.csv. Στο όποιο καταγράφονται τα αποτελέσματα για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCounterscatSessons_0W.csv Στο όποιο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.6.3.5 Μέσες τιμές MAPE με βαρύτητα 2 χωρίς εποχικότητα

Εδώ ο υπολογισμός γίνεται με τους συντελεστές βαρύτητας 2.

Ο υπολογισμός γίνεται με τον κώδικα **Mapecat_1W.py** Αυτός ο κώδικας δέχεται σαν εισόδους τα ίδια αρχεία με το προηγούμενο κώδικα (**Mapecat_0W.py**). Αυτό θα εκτελέσει τον υπολογισμό του μέσου MAPE αλλά αυτήν την φορά θα συνυπολογιστεί επιπλέον με την βαρύτητα 2 που επιλέχτηκε για τις κατηγορίες, δεν λαμβάνει υπόψιν την εποχικότητα.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) MapeAllcat_1W.csv. Στο όποιο καταγράφονται τα αποτελέσματα για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCounterscat_1W.csv. Στο όποιο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.6.3.6 Μέσες τιμές MAPE με βαρύτητα 2 και εποχικότητα

Εδώ ο υπολογισμός γίνεται με τους συντελεστές βαρύτητας 2. Ο υπολογισμός γίνεται με τον κώδικα **MapecatSessons_1W.py**. Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία τα ίδια αρχεία με τον προηγούμενο κώδικα (**MapecatSessons_0W.py**).

Αυτός θα εκτελέσει τον υπολογισμό του MAPE αλλά αυτήν την φορά θα συνυπολογιστεί επιπλέον με την βαρύτητα 2, για τις κατηγορίες, επιπλέον λαμβάνει υπόψιν την εποχικότητα.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) MapeAllcatSessons_1W.csv. Στο όποιο καταγράφονται τα αποτελέσματα για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCounterscatSessons_1W.csv. Στο όποιο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.6.3.7 Απλός μέσος όρος για όλους τους μήνες

(α) Χωρίς βαρύτητα χωρίς εποχικότητα

Ο κώδικας **Mape_Average_Cat_Without_De.py** υπολογίζει τον απλό μέσο όρο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, χωρίς εποχικότητα. Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) MapeAllcat_0.csv
- (ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο: **MethodsCatAverageW_de.csv**. Το αρχείο αυτό θα εμφανίσει τον απλό μέσο όρο MAPE των μηνών για τις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

(β) Χωρίς βαρύτητα με εποχικότητα

Ο κώδικας **Mape_Average_Cat_Sessons_W_De.py** υπολογίζει τον απλό μέσο όρο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, με εποχικότητα. Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) MapeAllcatSessons_0.csv
- (ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο: **MethodsCatAverageSessonsWDe_0.csv**. Το αρχείο αυτό θα εμφανίσει τον απλό μέσο όρο MAPE για τις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

(γ) Με βαρύτητα 1 χωρίς εποχικότητα

Ο κώδικας **Mape_Average_Cat_Without_De_W_0.py** υπολογίζει τον απλό μέσο όρο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, χωρίς εποχικότητα. Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) MapeAllcat_0W.csv
- (ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο: **MethodsCatAverageWDe_W0.csv**. Το αρχείο αυτό θα εμφανίσει τον απλό μέσο όρο MAPE για τις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

(δ) Με βαρύτητα 1 με εποχικότητα

Ο κώδικας **Mape_Average_Cat_Sessons_W_De_W_0.py** υπολογίζει τον απλό μέσο όρο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, χωρίς εποχικότητα. Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) MapeAllcatSessons_0W.csv
- (ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο: **MethodsCatAverageSessonsWDeW_0.csv**. Το αρχείο αυτό θα εμφανίσει τον απλό μέσο όρο MAPE για τις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

(ε) Με βαρύτητα 2 χωρίς εποχικότητα

Ο κώδικας **Mape_Average_Cat_Without_De_W_1.py** υπολογίζει τον απλό μέσο όρο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, με εποχικότητα. Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) MapeAllcat_1W.csv
- (ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο: **MethodsCatAverageWDe_W1.csv**. Το αρχείο αυτό θα εμφανίσει τον απλό μέσο όρο MAPE για τις εξής μεθόδους: LVFM, SLR2

έως SLR4 και QR4 έως QR8.

(στ) Με βαρύτητα 2 με εποχικότητα

Ο κώδικας **Mape_Average_Cat_Sessons_W_De_W_1.py**, υπολογίζει τον απλό μέσο όρο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, με εποχικότητα. Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

(i) MapeAllcatSessons_1W.csv

(ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο: **MethodsCatAverageSessonsWDeW_1.csv**. Το αρχείο αυτό θα εμφανίσει τον απλό μέσο όρο MAPE για τις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

4.6.3.8 Σταθμισμένος μέσος όρος για όλους τους μήνες

(α) Χωρίς βαρύτητα χωρίς εποχικότητα

Ο κώδικας **Mape_Average_Cat_De_0.py** υπολογίζει το μέσο σταθμισμένο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, χωρίς την εποχικότητα.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

(i) MapeAllcat_0.csv.

(ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο: **MethodsAverageCat_De_0.csv**. Το αρχείο αυτό θα εμφανίσει το μέσο σταθμισμένο MAPE για τις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

(β) Χωρίς βαρύτητα με εποχικότητα

Ο κώδικας **Mape_Average_Cat_Sessons_De_0.py** υπολογίζει το μέσο σταθμισμένο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, με την εποχικότητα.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

(i) MapeAllcatSessons_0.csv.

(ii) ForAllCategoriesFINALincome.csv.

Ο κώδικας αυτός δημιουργεί το εξής αρχείο:

MethodsAverageCatSessons_De_0.csv.

Το αρχείο αυτό θα εμφανίσει το μέσο σταθμισμένο MAPE στις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

(γ) Με βαρύτητα 1 χωρίς εποχικότητα

Ο κώδικας **Mape_Average_Cat_De_W_0.py** υπολογίζει το μέσο σταθμισμένο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, χωρίς την εποχικότητα.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

(i) MapeAllcat_0W.csv

(ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο: **MethodsAverageCat_De_W_0.csv**.

Το αρχείο αυτό θα εμφανίσει το μέσο σταθμισμένο MAPE στις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

(δ) Με βαρύτητα 1 με εποχικότητα

Ο κώδικας **Mape_Average_Cat_Sessons_De_W_0.py** υπολογίζει το μέσο σταθμισμένο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, χωρίς την εποχικότητα.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) MapeAllcatSessions_0W.csv
- (ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο:

MethodsAverageCatSessions_De_W_0.csv. Το αρχείο αυτό θα εμφανίσει το μέσο σταθμισμένο MAPE στις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

(ε) Με βαρύτητα 2 χωρίς εποχικότητα

Ο κώδικας **Mape_Average_Cat_De_W_1.py** υπολογίζει το μέσο σταθμισμένο MAPE για τις μεθόδους που χρησιμοποιούνται στις κατηγορίες, χωρίς την εποχικότητα.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) MapeAllcat_1W.csv
- (ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο: **MethodsAverageCat_De_W_1.csv.** Το αρχείο αυτό θα εμφανίσει το μέσο σταθμισμένο MAPE στις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

(στ) Με βαρύτητα 2 με εποχικότητα

Ο κώδικας **Mape_Average_Cat_Sessions_De_W_1.py.** Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) MapeAllcatSessions_1W.csv
- (ii) ForAllCategoriesFINALincome.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο: **MethodsAverageCatSessions_De_W_1.csv.** Το αρχείο αυτό θα εμφανίσει το μέσο σταθμισμέ-νο MAPE στις εξής μεθόδους: LVFM, SLR2 έως SLR4 και QR4 έως QR8.

4.7 Προβλέψεις πωλήσεων για τα προϊόντα

4.7.1 Γενικά

Όπως θα παρατηρήσουμε από τα αποτελέσματα οι δυο πιο αποτελεσματικές μέθοδοι είναι η Last Value Forecasting Method και η SLR 4 με χρήση εποχικότητας, όποτε χρησιμοποιήθηκαν μόνο αυτές οι δυο για την πρόβλεψη πάνω στα προϊόντα. Παρόλα αυτά οι κώδικες που δίνονται στην παρακάτω ενότητα μπορούν εύκολα να επεκταθούν ώστε να αυξηθεί ο αριθμός των μηνών που χρησιμοποιούνται για πρόβλεψη (SLR5, 6 κ.λπ) ή γενικότερα να αλλαχθούν (π.χ. από SLR4 σε SLR3).

4.7.2 Εφαρμογή της LVFM για τα προϊόντα

Ο κώδικας **LastValueForecastingMethodForAllMonths.py** εφαρμόζει την Μέθοδο Πρόβλεψης Τελευταίας Αξίας (Last Value Forecasting Method). Χρησιμοποιεί δηλαδή την τελευταία παρατήρηση στη χρονοσειρά ως πρόβλεψη για την επόμενη περίοδο (μήνα) για κάθε προϊόν ξεχωριστά. Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ForAllItemsFINAL.csv.
- (ii) ProductLifes.csv.

Ο κώδικας αυτός δημιουργεί το αρχείο:

LastValueForecastingMethodForAllMonths.csv.

Σε αυτό το αρχείο στην ουσία έχει γίνει μεταφορά των πωλήσεων για κάθε προϊόν σε κάθε μήνα, κατά ένα μήνα. Π.χ. Οι πραγματικές πωλήσεις του Σεπτεμβρίου 2015 είναι η πρόβλεψη για τον Οκτώβριο του 2015 κ.ο.κ.

4.7.3 Εφαρμογή της SLR4 για τα προϊόντα

Ο κώδικας **SimpleLinearRegressionFinalsessionSLR4_0.py** εφαρμόζει την μέθοδο SLR4 για τα προϊόντα δηλαδή θα χρησιμοποιήσει μόνο τους 4 προηγούμενους μήνες σαν δεδομένα, για να κάνει προβλέψεις. Δέχεται σαν είσοδο τα εξής αρχεία:

(i) ForAllItemsFINAL.csv.

(ii) ProductLifes.csv.

Ο κώδικας αυτός δημιουργεί μόνο ένα αρχείο που είναι το:

SimpleLinearRegression4session_33%.csv.

Σε αυτό το αρχείο γίνονται προβλέψεις από τον 5^ο μήνα έως τον 34^ο που είναι και ο τελευταίος μήνας που υπάρχει στα πρωτογενή αρχεία, χρησιμοποιώντας κάθε φορά τους δυο προηγούμενους μήνες, για κάθε κατηγορία ξεχωριστά και επιπλέον χρησιμοποιείτε και η εποχικότητα.

4.7.4 MAPE για LVFM και SLR4 χωρίς βαρύτητα

Ο κώδικας **MapeSLR4Items.py** υπολογίζει τα MAPE για κάθε προϊόν και μήνα για τις δύο μεθόδους, που όπως θα δούμε τελικά θα επιλεγθούν.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

(i) ProductLifes.csv

(ii) ForAllItemsFINAL.csv

(iii) ForAllItemsLastMonth.csv

(iv) LastValueForecastingMethodForAllMonths.csv

(v) SimpleLinearRegression4session_33%.csv

Ο κώδικας αυτός εκτελεί τον υπολογισμό του MAPE, αλλά είναι μόνο για την Last Value Forecasting Method, καθώς και την SLR4, με εποχικότητα και χωρίς συντελεστές βαρύτητας αντίστοιχα.

Ο κώδικας αυτός δημιουργεί τα εξής αρχεία:

(α) **MapeSLR4Items.csv.** Στο οποίο καταγράφονται τα αποτελέσματα για κάθε ένα μήνα για κάθε μέθοδο (LVFM και SLR4), ξεχωριστά.

(β) **MapeSLR4ItemsCounters.csv.** Στο οποίο καταγράφονται πόσα από τα προϊόντα αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.7.5 MAPE για LVFM και SLR4 με βαρύτητα 1

Ο κώδικας **MapeSLR4ItemsW_0.py** υπολογίζει το MAPE και πάλι μόνο για την Last Value Forecasting Method και την SLR4 με εποχικότητα αλλά αυτή την φορά με συντελεστές βαρύτητας 1.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

(i) ProductLifes.csv

(ii) ForAllItemsFINAL.csv

(iii) ForAllItemsLastMonth.csv

(iv) LastValueForecastingMethodForAllMonths.csv

(v) SimpleLinearRegression4session_33%.csv

(vi) ForAllItemsFINALincome_items.csv

Ο κώδικας αυτός δημιουργεί τα εξής αρχεία:

(α) **MapeSLR4ItemsW.csv.** Στο οποίο καταγράφονται τα αποτελέσματα για κάθε ένα μήνα ανά κάθε μέθοδο, ξεχωριστά.

(β) **MapeSLR4ItemsCountersW.csv.** Στο οποίο καταγράφονται πόσα από τα προϊόντα αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.7.6 MAPE για LVFM και SLR4 με βαρύτητα 2

Ο κώδικας **MapeSLR4ItemsW_2.py** υπολογίζει το MAPE και πάλι μόνο για την Last Value Forecasting Method και την SLR4 με εποχικότητα αλλά αυτή την φορά με συντελεστές βαρύτητας 2.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes.csv
- (ii) ForAllItemsFINAL.csv
- (iii) ForAllItemsLastMonth.csv
- (iv) LastValueForecastingMethodForAllMonths.csv
- (v) SimpleLinearRegression4season_33%.csv
- (vi) ForAllItemsFINALincome_items.csv

Ο κώδικας αυτός δημιουργεί τα εξής αρχεία:

- (α) **MapeSLR4ItemsW_2.csv**. Στο οποίο καταγράφονται τα αποτελέσματα για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.
- (β) **MapeSLR4ItemsCountersW_2.csv**. Στο οποίο καταγράφονται πόσα από τα προϊόντα αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.7.7 Απλό μέσο MAPE LVFM και SLR4 χωρίς βαρύτητα

Ο κώδικας **Mape_Average_Items_Sessons_0.py** υπολογίζει το μέσο όρο MAPE για τις δύο μεθόδους που θα επιλεγθούν για να χρησιμοποιηθούν τελικά, με εποχικότητα αλλά χωρίς βαρύτητα.

Αυτός ο κώδικας δέχεται σαν είσοδο το εξής αρχείο: MapeSLR4Items.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο:

MethodsItemsAverageSessons.csv.

Το οποίο θα εμφανίσει τον απλό μέσο όρο MAPE στις δυο μεθόδους (LVFM και SLR4).

4.7.8 Απλό μέσο MAPE LVFM και SLR4 με βαρύτητα 1

Ο κώδικας **Mape_Average_Items_SessonsW_1.py** υπολογίζει το μέσο όρο MAPE για τις δύο μεθόδους που θα επιλεγθούν για να χρησιμοποιηθούν τελικά, με εποχικότητα και με βαρύτητα 1.

Αυτός ο κώδικας δέχεται σαν είσοδο το εξής αρχείο:

MapeSLR4ItemsW.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο:

MethodsItemsAverageSessonsW_1.csv.

Το οποίο θα εμφανίσει τον απλό μέσο όρος MAPE στις δυο μεθόδους (LVF και SLR4).

4.7.9 Απλό μέσο MAPE LVFM και SLR4 με βαρύτητα 2

Ο κώδικας **Mape_Average_Items_SessonsW_2.py** υπολογίζει το μέσο όρο MAPE για τις δύο μεθόδους που θα επιλεγθούν για να χρησιμοποιηθούν τελικά, με εποχικότητα, και με βαρύτητα 2.

Αυτός ο κώδικας δέχεται σαν είσοδο το εξής αρχείο:

MapeSLR4ItemsW_2.csv

Ο κώδικας αυτός δημιουργεί το εξής αρχείο:

MethodsItemsAverageSessonsW_2.csv.

Το όποιο θα εμφανίσει τον απλό μέσο όρο MAPE στις δυο μεθόδους (LVFM και SLR4).

4.7.10 Σταθμισμένο MAPE LVFM και SRL4 χωρίς βαρύτητα

Ο κώδικας `Mape_Average_Items_Sessons_De_0.py` υπολογίζει το μέσο σταθμισμένο MAPE για τις δύο μεθόδους που θα επιλεγθούν για να χρησιμοποιηθούν τελικά, με εποχικότητα, χωρίς βαρύτητα.

Αυτός ο κώδικας δέχεται σαν είσοδο το εξής αρχείο:

`MapeSLR4Items.csv`

Ο κώδικας αυτός δημιουργεί το εξής αρχείο:

`MethodsItemsAverageSessons_De_0.csv`.

Το όποιο θα εμφανίσει το μέσο σταθμισμένο MAPE στις δυο μεθόδους (LVFM και SLR4).

4.7.11 Σταθμισμένο MAPE LVFM και SLR4 με βαρύτητα 1

Ο κώδικας `Mape_Average_Items_Sessons_W_DE_1.py` υπολογίζει το μέσο σταθμισμένο MAPE για τις δύο μεθόδους που θα επιλεγθούν για να χρησιμοποιηθούν τελικά, με εποχικότητα και με βαρύτητα 1.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

(i) `MapeSLR4ItemsW.csv`

(ii) `ForAllItemsFINALincome.csv`

(iii) `ForAllItemsLastMonth.csv`

Ο κώδικας αυτός δημιουργεί το εξής αρχείο:

`MethodsAverageSessonsW_De_1.csv`.

Το όποιο θα εμφάνιση το μέσο σταθμισμένο MAPE στις δυο μεθόδους (LVFM και SLR4).

4.7.12 Σταθμισμένο MAPE LVFM και FLR4 με βαρύτητα 2

Ο κώδικας `Mape_Average_Items_Sessons_W_DE_2.py` υπολογίζει το μέσο σταθμισμένο MAPE για τις δύο μεθόδους που θα επιλεγθούν για να χρησιμοποιηθούν τελικά, με εποχικότητα και με βαρύτητα 2.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

(i) `MapeSLR4ItemsW.csv`

(ii) `ForAllItemsFINALincome.csv`

(iii) `ForAllItemsLastMonth.csv`

Ο κώδικας αυτός δημιουργεί το εξής αρχείο:

`MethodsAverageSessonsW_De_2.csv`.

Το όποιο θα εμφανίσει το μέσο σταθμισμένο MAPE στις δυο μεθόδους (LVFM και SLR4).

4.8 Επιπλέον δυνατότητες

4.8.1 Γενικά

Εκτός από τις προβλέψεις που γίνονται κάθε φορά για την επόμενη άγνωστη περίοδο, υπάρχουν και οι εξής δυνατότητες.

(1) Μερικές γενικές εφαρμογές όπως η εύρεση του πλήθους των προϊόντων που περιέχει

η κάθε κατηγορία.

(2) Δημιουργία επικεφαλίδων διαφόρων πινάκων που εμφανίζονται σε επιμέρους αρχεία.

(3) Προβλέψεις με καθυστέρηση.

(4) Υπολογισμός του MAPE για τις προβλέψεις με καθυστέρηση.

(5) Προβλέψεις με χρήση δεδομένων 24 μηνών τόσο για τις κατηγορίες όσο και για τα προϊόντα.

(6) Υπολογισμός του MAPE για τις προβλέψεις με την χρήση 24 μηνών.

(7) Συναρτήσεις αυτοσυσχέτισης (ACF) για επιλεγμένη από τον χρήστη κατηγορία ή προϊόν.

4.8.2 Πλήθος προϊόντων ανά κατηγορία

Ο κώδικας **ItemsperCategories.py** βρίσκει το πλήθος των προϊόντων που περιέχει η κάθε κατηγορία. Τα πλήθη αυτά φαίνεται στον πίνακα 3.2-4

Ο κώδικας δέχεται σαν είσοδο δεδομένων τα αρχεία:

(i) Το πρωτογενές αρχείο: items.csv

(ii) ForAllItemsFINAL.csv

(iii) ForAllItemsLastMonth.csv

Αυτός ο κώδικας υπολογίζει πόσα προϊόντα βρίσκονται σε κάθε κατηγορία, καθώς δημιουργεί ένα πίνακα ο οποίος είναι ταξινομημένος με προτεραιότητα ανά κατηγορία και μετά ανά προϊόν.

Αυτός ο κώδικας θα δημιουργήσει τα εξής αρχεία.

(α) **NumItemsInCategories.csv**. Όπου εμφανίζονται πόσα προϊόντα βρίσκονται σε κάθε κατηγορία.

(β) **ItemsperCategories.csv**. Ο ταξινομημένος πίνακας στον οποίο αναφερθήκαμε προηγούμενος. Τα αποτελέσματα παρουσιάζονται στον πίνακα 3.2-4.

4.8.3 Δημιουργία αρχείων επικεφαλίδων

Ο κώδικας **SetDatesIncolumns.py** είναι ένας κώδικας με πολλές παρόμοιες μεθόδους που έχουν σαν σκοπό να περάσουν τους τίτλους του κάθε μήνα στα αρχεία εξόδων csv στους κώδικες που είναι απαραίτητο να γίνουν. Όλοι οι κώδικες που χρειάζονται τους προαναφερόμενους τίτλους καλούν αυτόν των κώδικα και την ονομασία της μεθόδου για την κάθε περίπτωση αντίστοιχα.

4.8.4 Πρόβλεψη με καθυστέρηση

4.8.4.1 Γενικά

Σε ορισμένες περιπτώσεις μια εταιρεία μπορεί να χρειάζεται να έχει στην διάθεση τις προβλέψεις για κάποιες μελλοντικές περιόδους. Για παράδειγμα όταν πλησιάζουν οι εορτές των Χριστουγέννων, όπου οι πωλήσεις παρουσιάζουν αύξηση, θα ήταν καλό να προβλέπει τις πωλήσεις της εορταστικής περιόδου ώστε να κάνει εγκαίρως τις παραγγελίες της σε πρώτες ύλες. Οι κώδικες που αναλύονται στις παρακάτω ενότητες κάνουν ακριβώς αυτό. Να μπορεί κανείς να προβλέπει με τα δεδομένα που θα έχει μέχρι εκείνη την στιγμή τι πωλήσεις θα έχει μετά από κάποιες περιόδους.

4.8.4.2 SLR με καθυστέρηση (delay) για κάθε προϊόν

Ο κώδικας **SimpleLinearRegressionFinalWithDelay_0.py** είναι παρόμοιος με τον **SimpleLinearRegressionFinalRepeat_0.py**, μόνο που εκτελεί την πρόβλεψη για κάθε προϊόν ξεχωριστά με επιλογή **g** μηνών σαν δεδομένων και καθυστέρηση **delay**

μηνών, όπου το **g** επιλέγεται από τον χρήστη, από 2 έως 4, και είναι για το πόσοι μήνες θα χρησιμοποιηθούν σαν δεδομένοι. Ενώ το **delay** θα επιλέγεται και αυτό από τον χρήστη και αντιπροσωπεύει το πόσους μήνες μετά θα ξεκινήσει να προβλέπει. Π.χ. Με την **Simple Linear Regression** με χρήση **g=2** και **delay=3**, χρησιμοποιώντας τα δεδομένα των δυο πρώτων μηνών, αγνοεί τους τρεις επόμενους και κάνει πρόβλεψη κατευθείαν για τον 6^ο μήνα. Αντίστοιχα για τον επόμενο μήνα χρησιμοποιείται ο δεύτερος και τρίτος μήνας των δεδομένων αγνοώντας και πάλι τους τρεις επομένους μήνες, γίνεται η πρόβλεψη κατευθείαν για τον 7^ο μήνα. Ομοίως και για τους υπόλοιπους μήνες. Ο κώδικας μπορεί να προσαρμοστεί τόσο για μεγαλύτερα **g** όσο και μεγαλύτερα **delay**. Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes.csv.
- (ii) ForAllItemsFINAL.csv.

Ο κώδικας αυτός μπορεί να δημιουργήσει τα εξής αρχεία:

(α) SimpleLinearRegression2WithDelay.csv. Σε αυτό το αρχείο γίνονται προβλέψεις από τον 6^ο μήνα έως τον 34^ο που είναι και ο τελευταίος μήνας που υπάρχει στα πρωτογενή αρχεία, αγνοώντας κάθε φορά τους τρεις προηγούμενους μήνες και μετά από αυτούς χρησιμοποιώντας τους δυο προηγούμενους, για κάθε προϊόν ξεχωριστά.

(β) SimpleLinearRegression3WithDelay.csv. Εδώ χρησιμοποιώντας αντίστοιχα τους 3 μήνες σαν δεδομένα.

(γ) SimpleLinearRegression4WithDelay.csv. Εδώ χρησιμοποιώντας αντίστοιχα τους 4 μήνες σαν δεδομένα.

4.8.4.3 SLR με καθυστέρηση (delay) για κάθε κατηγορία

Ο κώδικας **SimpleLinearRegressionFinalcatWithDelay_0.py** κάνει το ίδιο με το προηγούμενο, αλλά για κάθε κατηγορία αντί για κάθε προϊόν ξεχωριστά. Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifesCategories.csv.
- (ii) ForAllCategoriesFINAL.csv.

Ο κώδικας αυτός μπορεί να δημιουργήσει τα εξής αρχεία:

(α) SimpleLinearRegression2catWithDelay.csv Σε αυτό το αρχείο γίνονται προβλέψεις από τον 6^ο μήνα έως τον 34^ο που είναι και ο τελευταίος μήνας που υπάρχει στα πρωτογενή αρχεία, αγνοώντας κάθε φορά τους τρεις προηγούμενους μήνες και μετά από αυτούς χρησιμοποιώντας τους δυο προηγούμενους, για κάθε κατηγορία ξεχωριστά.

(β) SimpleLinearRegression3catWithDelay.csv. Εδώ χρησιμοποιώντας αντίστοιχα τους 3 μήνες σαν δεδομένα.

(γ) SimpleLinearRegression4catWithDelay.csv. Εδώ χρησιμοποιώντας αντίστοιχα τους 4 μήνες σαν δεδομένα.

4.8.4.4 SLR με καθυστέρηση και εποχικότητα για κάθε προϊόν

Ο κώδικας **SimpleLinearRegressionFinalsessionWithDelay_0.py** κάνει τα ίδια με τον **SimpleLinearRegressionFinalWithDelay_0.py**, αλλά με την επιπλέον εφαρμογή την εποχικότητας. Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes.csv.
- (ii) ForAllItemsFINAL.csv.

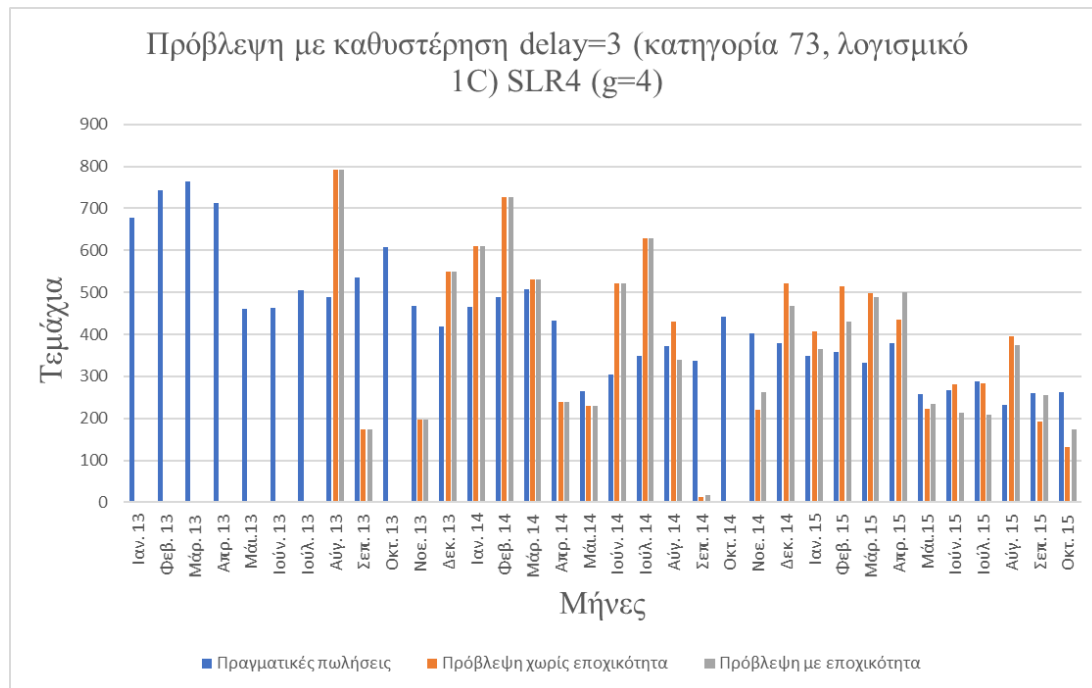
Ο κώδικας αυτός μπορεί να δημιουργήσει τα εξής αρχεία:

(α) SimpleLinearRegression2sessionWithDelay.csv. Σε αυτό το αρχείο γίνονται προβλέψεις από τον 6^ο μήνα έως τον 34^ο που είναι και ο τελευταίος μήνας που υπάρχει στα πρωτογενή αρχεία, αγνοώντας κάθε φορά τους τρεις προηγούμενους μήνες και μετά

από αυτούς χρησιμοποιώντας τους δυο προηγούμενους, για κάθε προϊόν ξεχωριστά. Μόνο που σε αυτό το αρχείο έχει προστεθεί η εποχικότητα.

(β) SimpleLinearRegression3sessionWithDelay.csv. Εδώ χρησιμοποιώντας αντίστοιχα τους 3 μήνες σαν δεδομένα.

(γ) SimpleLinearRegression4sessionWithDelay.csv. Εδώ χρησιμοποιώντας αντίστοιχα τους 4 μήνες σαν δεδομένα.



Εικόνα 4.8-1 Πρόβλεψη πωλήσεων για την κατηγορία 73 με SLR4 και delay=3

4.8.4.5 SLR με καθυστέρηση και εποχικότητα για κατηγορίες

Ο κώδικας **SimpleLinearRegressionFinalcatSessionWithDelay.py** κάνει τα ίδια με τον **SimpleLinearRegressionFinalcatWithDelay_0.py**, αλλά με την επιπλέον εφαρμογή την εποχικότητας. Δέχεται σαν είσοδο τα εξής αρχεία:

(i) ProductLifesCategories.csv.

(ii) ForAllCategoriesFINAL.csv.

Ο κώδικας αυτός μπορεί να δημιουργήσει τα εξής αρχεία:

(α) SimpleLinearRegression2catSessionWithDelay.csv.

Σε αυτό το αρχείο γίνονται προβλέψεις από τον 6^ο μήνα έως τον 34^ο που είναι και ο τελευταίος μήνας που υπάρχει στα πρωτογενή αρχεία, αγνοώντας κάθε φορά τους τρεις προηγούμενους μήνες και μετά από αυτούς χρησιμοποιώντας τους δυο προηγούμενους, για κάθε κατηγορία ξεχωριστά. Μόνο που σε αυτό το αρχείο έχει προστεθεί η εποχικότητα.

(β) SimpleLinearRegression3catSessionWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα τους 3 μήνες σαν δεδομένα.

(γ) SimpleLinearRegression4catSessionWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα τους 4 μήνες σαν δεδομένα.

4.8.4.6 QR με καθυστέρηση για κάθε προϊόν

Ο κώδικας **QuadraticRegressionAllItemsFinalWithDelay_0.py** χρησιμοποιεί την μέθοδο **QR** όπως και ο κώδικας:

QuadraticRegressionAllCategoriesFinalRepeat_0.py,

όμως έχει διάφορες και αυτές είναι οι εξής:

α) Δεν εκτελείτε για τις κατηγορίες προϊόντων αλλά για κάθε προϊόν ξεχωριστά.

β) Το **g** που επιλέγετε από τον χρήστη (από 4 έως 8) και είναι για το πόσοι μήνες θα χρησιμοποιηθούν σαν δεδομένα για την πρόβλεψη ή αλλιώς ποια ακριβώς μέθοδος πρόβλεψης θα χρησιμοποιηθεί (QR4, QR5 κ.λπ.).

γ) Εμπεριέχει την καθυστέρηση **delay**, που μπορεί να δοθεί από τον χρήστη όπως και στον κώδικα **SimpleLinearRegressionFinalWithDelay_0.py**. Αυτή καθορίζει το για μετά από πόσους μήνες θα αρχίσει να γίνεται η πρόβλεψη.

Π.χ. Με την Quadratic Regression με χρήση 4 μηνών, χρησιμοποιώντας τα δεδομένα των τεσσάρων πρώτων μηνών, αγνοούμε τους τρεις επόμενους και γίνεται πρόβλεψη κατευθείαν για τον 8^ο μήνα. Αντίστοιχα για τον επόμενο μήνα χρησιμοποιούμε τον 2^ο, 3^ο, 4^ο και 5^ο μήνα των δεδομένων αγνοώντας και πάλι τους τρεις επομένους μήνες, γίνεται η πρόβλεψη κατευθείαν για τον 9^ο μήνα. Ομοίως και για τους υπόλοιπους μήνες. Δέχεται σαν είσοδο τα εξής αρχεία:

(i) ProductLifes.csv.

(ii) ForAllItemsFINAL.csv.

Ο κώδικας αυτός δημιουργεί τα εξής αρχεία:

(α) QuadraticRegressionWith4MonthsWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 4 μήνες πριν από την καθυστέρηση.

(β) QuadraticRegressionWith5MonthsWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 5 μήνες πριν από την καθυστέρηση.

(γ) QuadraticRegressionWith6MonthsWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 6 μήνες πριν από την καθυστέρηση.

(δ) QuadraticRegressionWith7MonthsWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 7 μήνες πριν από την καθυστέρηση.

(ε) QuadraticRegressionWith8MonthsWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 8 μήνες πριν από την καθυστέρηση.

4.8.4.7 QR με καθυστέρηση για τις κατηγορίες

Ο κώδικας **QuadraticRegressionAllCategoriesFinalWithDelay_0.py** κάνει τα ίδια ακριβώς που έκανε και ο προηγούμενος, όχι για τα επιμέρους προϊόντα αλλά για τις κατηγορίες τους. Δέχεται σαν είσοδο τα εξής αρχεία:

(i) ProductLifesCategories.csv.

(ii) ForAllCategoriesFINAL.csv.

Ο κώδικας αυτός μπορεί να δημιουργήσει τα εξής αρχεία:

(α) QuadraticRegressionWith4MonthsCATWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 4 μήνες πριν από την καθυστέρηση.

(β) QuadraticRegressionWith5MonthsCATWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 5 μήνες πριν από την καθυστέρηση.

(γ) QuadraticRegressionWith6MonthsCATWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 6 μήνες πριν από την καθυστέρηση.

(δ) QuadraticRegressionWith7MonthsCATWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 7 μήνες πριν από την καθυστέρηση.

(ε) QuadraticRegressionWith8MonthsCATWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 8 μήνες πριν από την καθυστέρηση.

4.8.4.8 QR με καθυστέρηση και εποχικότητα για κάθε προϊόν

Ο κώδικας **QuadraticRegressionAllItemsFinalSessionsAndDelay_0.py** κάνει ότι και ο **QuadraticRegressionAllItemsFinalWithDelay_0.py**, αλλά λαμβάνεται υπόψη η

εποχικότητα. Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes.csv.
- (ii) ForAllItemsFINAL.csv.

Ο κώδικας αυτός μπορεί να δημιουργήσει τα εξής αρχεία:

(α) QuadraticRegressionWith4MonthsSessionsAndDelay.csv.

Εδώ χρησιμοποιώντας δεδομένα από αντίστοιχα τους 4 μήνες πριν από την καθυστέρηση.

(β) QuadraticRegressionWith5MonthsSessionsAndDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 5 μήνες πριν από την καθυστέρηση.

(γ) QuadraticRegressionWith6MonthsSessionsAndDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 6 μήνες πριν από την καθυστέρηση.

(δ) QuadraticRegressionWith7MonthsSessionsAndDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 7 μήνες πριν από την καθυστέρηση.

(ε) QuadraticRegressionWith8MonthsSessionsAndDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 8 μήνες πριν από την καθυστέρηση.

4.8.4.9 QR με καθυστέρηση και εποχικότητα για κατηγορίες

Ο κώδικας **QuadraticRegressionAllCategoriesFinalSessionsWithDelay.py** κάνει το ίδιο με τον **QuadraticRegressionAllCategoriesFinalWithDelay_0.py** αλλά με την διαφορά, ότι λαμβάνει υπόψιν την εποχικότητα. Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifesCategories.csv.
- (ii) ForAllCategoriesFINAL.csv.

Ο κώδικας αυτός μπορεί να δημιουργήσει τα εξής αρχεία:

(α) QuadraticRegressionWith4MonthsCATSessionWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 4 μήνες πριν από την καθυστέρηση.

(β) QuadraticRegressionWith5MonthsCATSessionWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 5 μήνες πριν από την καθυστέρηση.

(γ) QuadraticRegressionWith6MonthsCATSessionWithDelay.csv.

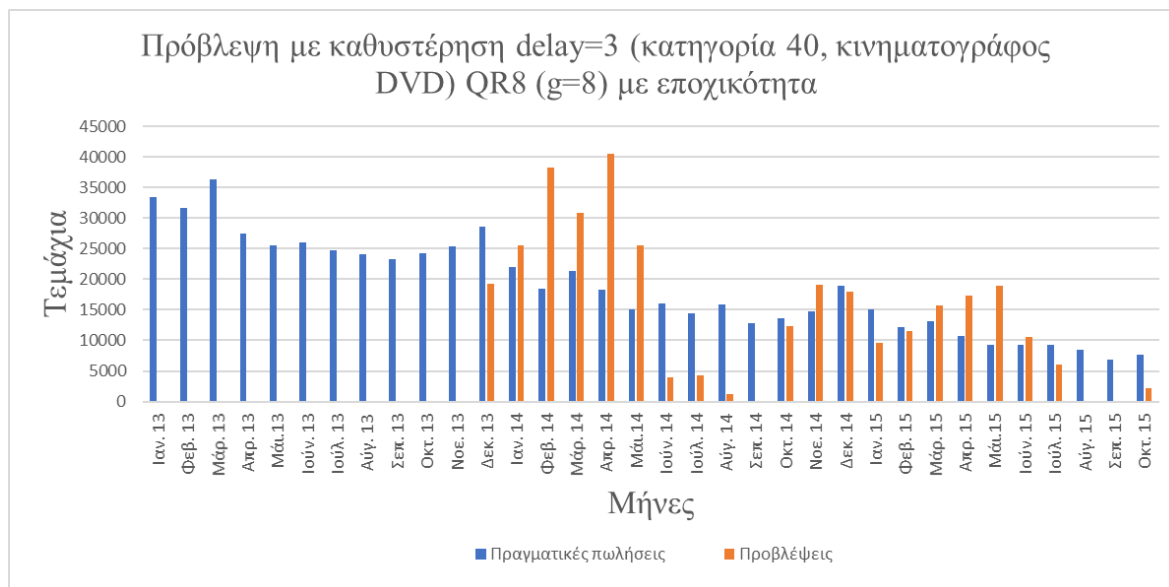
Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 6 μήνες πριν από την καθυστέρηση.

(δ) QuadraticRegressionWith7MonthsCATSessionWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 7 μήνες πριν από την καθυστέρηση.

(ε) QuadraticRegressionWith8MonthsCATSessionWithDelay.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 8 μήνες πριν από την καθυστέρηση.



Εικόνα 4.8-2 Πρόβλεψη πωλήσεων για την κατηγορία 40 με QR8 και delay=3

4.8.5 MAPE για προβλέψεις με καθυστέρηση

4.8.5.1 Γενικά

Ο υπολογισμός του MAPE εδώ γίνεται όπως και για τις υπόλοιπες προβλέψεις. Απλά αλλάζουν τα αρχεία εισόδου σε κάθε αντίστοιχο κώδικα.

4.8.5.2 MAPE για προβλέψεις προϊόντων με καθυστέρηση

Χρησιμοποιείται ο κώδικας **MapeWithDelay_0.py**. Αυτός παίρνει σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes.csv
- (ii) ForAllItemsFINAL.csv
- (iii) ForAllItemsLastMonth.csv
- (iv) LastValueForecastinMethodForAllMonths.csv
- (v) SimpleLinearRegression2WithDelay.csv
- (vi) SimpleLinearRegression3WithDelay.csv
- (vii) SimpleLinearRegression4WithDelay.csv
- (viii) QuadraticRegressionWith4MonthsWithDelay.csv
- (ix) QuadraticRegressionWith5MonthsWithDelay.csv
- (x) QuadraticRegressionWith6MonthsWithDelay.csv
- (xi) QuadraticRegressionWith7MonthsWithDelay.csv
- (xii) QuadraticRegressionWith8MonthsWithDelay.csv

Από αυτά τα αρχεία ο κώδικας **MapeWithDelay_0.py** θα κάνει τον υπολογισμό του **MAPE** για κάθε προϊόν σε κάθε μήνα, με την καθυστέρηση που έδωσε ο χρήστης στα προαναφερόμενα αρχεία, σε κάθε μια από τις εξής μεθόδους:

- (i) Last Value Forecasting Method
- (ii) Simple Linear Regression από 2 μήνες έως 4 μήνες δεδομένων.
- (iii) Quadratic Regression από 4 μήνες έως 8 μήνες δεδομένων.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

- (α) MapeAllWithDelay_0. Στο οποίο καταγράφονται τα αποτελέσματα για κάθε ένα μήνα ανά κάθε μέθοδο, ξεχωριστά.
- (β) MapeAllMonthCountersWithDelay_0.csv. Στο οποίο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.8.5.3 MAPE για προϊόντα με καθυστέρηση και εποχικότητα

Χρησιμοποιείται ο κώδικας **MapeSessionsAndWithDelay_0.py**. Αυτός παίρνει σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes.csv
- (ii) ForAllItemsFINAL.csv
- (iii) ForAllItemsLastMonth.csv
- (iv) LastValueForecastinMethodForAllMonths.csv
- (v) SimpleLinearRegression2sessionWithDelay.csv
- (vi) SimpleLinearRegression3sessionWithDelay.csv
- (vii) SimpleLinearRegression4sessionWithDelay.csv
- (viii) QuadraticRegressionWith4MonthsSessionsAndDelay.csv
- (ix) QuadraticRegressionWith5MonthsSessionsAndDelay.csv
- (x) QuadraticRegressionWith6MonthsSessionsAndDelay.csv
- (xi) QuadraticRegressionWith7MonthsSessionsAndDelay.csv

(xii) QuadraticRegressionWith8MonthsSessionsAndDelay.csv

Με αυτά τα αρχεία θα εκτελέσει όπως και ο προηγούμενος κώδικας το MAPE αλλά αυτή την φορά θα εμπεριέχεται και η εποχικότητα μέσα στα δεδομένα που χρησιμοποιούνται.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) MapeAllSessionsWithDelayFrom_0.csv. Στο οποίο καταγράφονται τα αποτελέσματα για κάθε ένα μήνα ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCountersSessionsWithDelay_0.csv. Στο οποίο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.8.5.4 MAPE για κατηγορίες με καθυστέρηση

Χρησιμοποιείται ο κώδικας **MapeCatWithDelay_0.py**. Αυτός παίρνει σαν είσοδο τα εξής αρχεία:

(i) ProductLifesCategories.csv

(ii) ForAllCategoriesFINAL.csv

(iii) LastValueForecastinMethodForAllMonthscat.csv

(iv) SimpleLinearRegression2catWithDelay.csv

(v) SimpleLinearRegression3catWithDelay.csv

(vi) SimpleLinearRegression4catWithDelay.csv

(vii) QuadraticRegressionWith4MonthsCATWithDelay.csv

(viii) QuadraticRegressionWith5MonthsCATWithDelay.csv

(ix) QuadraticRegressionWith6MonthsCATWithDelay.csv

(x) QuadraticRegressionWith7MonthsCATWithDelay.csv

(xi) QuadraticRegressionWith8MonthsCATWithDelay.csv

Αυτό θα εκτελέσει τον υπολογισμό του MAPE αλλά αυτήν την φορά θα εκτελεστεί για τις κατηγορίες, με την καθυστέρηση που έδωσε ο χρήστης στα προαναφερόμενα αρχεία, χωρίς εποχικότητα.

(α) MapeAllCatWithDelay_0.csv. Στο οποίο καταγράφονται τα αποτελέσματα για κάθε ένα μήνα ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCountersCatWithDelay_0.csv. Στο οποίο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.8.5.5 MAPE κατηγορίες με καθυστέρηση και εποχικότητα

Χρησιμοποιείται ο κώδικας **MapeCatSessionsAndWithDelay_0.py**.

Αυτός δέχεται σαν εισόδους τα εξής αρχεία:

(i) ProductLifesCategories.csv

(ii) ForAllCategoriesFINAL.csv

(iii) LastValueForecastinMethodForAllMonthscat.csv

(iv) SimpleLinearRegression2catsessionWithDelay.csv

(v) SimpleLinearRegression3catsessionWithDelay.csv

(vi) SimpleLinearRegression4catsessionWithDelay.csv

(vii) QuadraticRegressionWith4MonthsCATSessionWithDelay.csv

(viii) QuadraticRegressionWith5MonthsCATSessionWithDelay.csv

(ix) QuadraticRegressionWith6MonthsCATSessionWithDelay.csv

(x) QuadraticRegressionWith7MonthsCATSessionWithDelay.csv

(xi) QuadraticRegressionWith8MonthsCATSessionWithDelay.csv

Ο κώδικας αυτός θα εκτελέσει τον υπολογισμό του MAPE αλλά αυτήν την φορά

θα εκτελεστεί για τις κατηγορίες, με την καθυστέρηση που έδωσε ο χρήστης στα προαναφερόμενα αρχεία, με εποχικότητα.

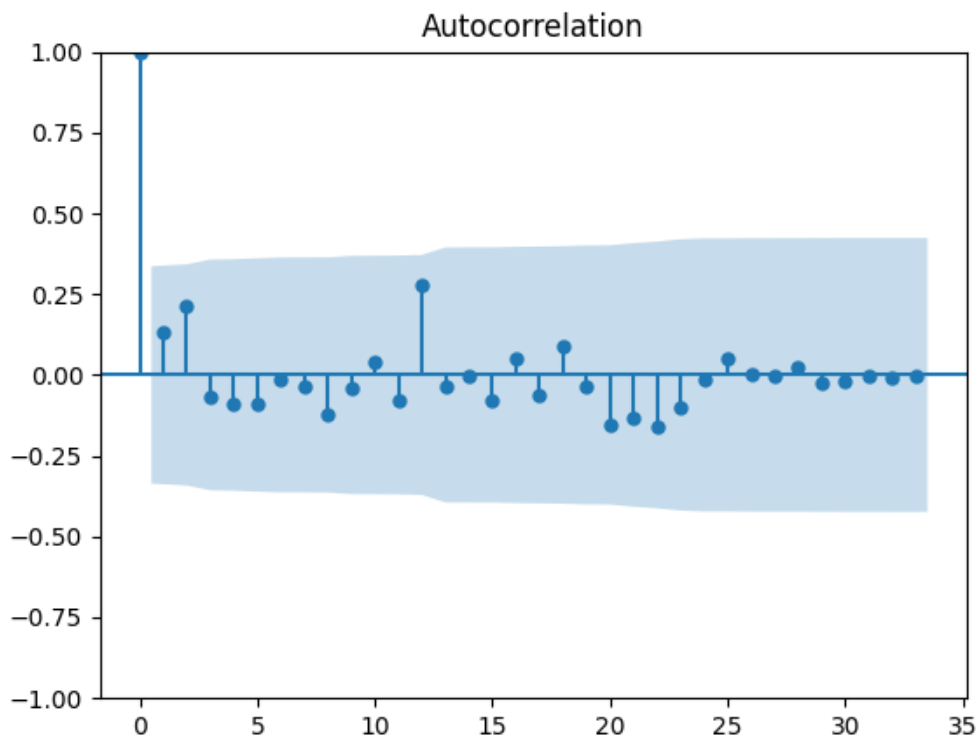
(α) `MapeAllSessionsWithDelayFrom_0.csv`. Στο οποίο καταγράφονται τα αποτελέσματα για κάθε ένα μήνα ανά κάθε μέθοδο, ξεχωριστά.

(β) `MapeAllMonthCountersSessionsWithDelay_0.csv`. Στο οποίο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.8.6 Συνάρτηση αυτοσυσχέτισης για επιλεγμένη κατηγορία

Ο κώδικας `AutoCorreletionStatsforAnyCat_0.py` δέχεται σαν είσοδο δεδομένων το αρχείο: `ForAllCategoriesFINAL.csv`

Αυτός ο κώδικας βρίσκει τους συντελεστές Αυτοσυσχέτισης (Auto Correlation) για μία κατηγορία προϊόντων. Η επιλογή γίνεται από τον χρήστη με τον αντίστοιχο κωδικό ID της κατηγορίας. Το πρόγραμμα εμφανίζει στην οθόνη του υπολογιστή το διάγραμμα της συνάρτησης αυτοσυσχέτισης (ACF) μιάς επιλεγμένης κατηγορίας. 63 (Δώρα – Μαζεμένα παιχνίδια).



Εικόνα 4.8-3 Συνάρτηση ACF για την κατηγορία 63 (Δώρα – Μαζεμένα παιχνίδια)

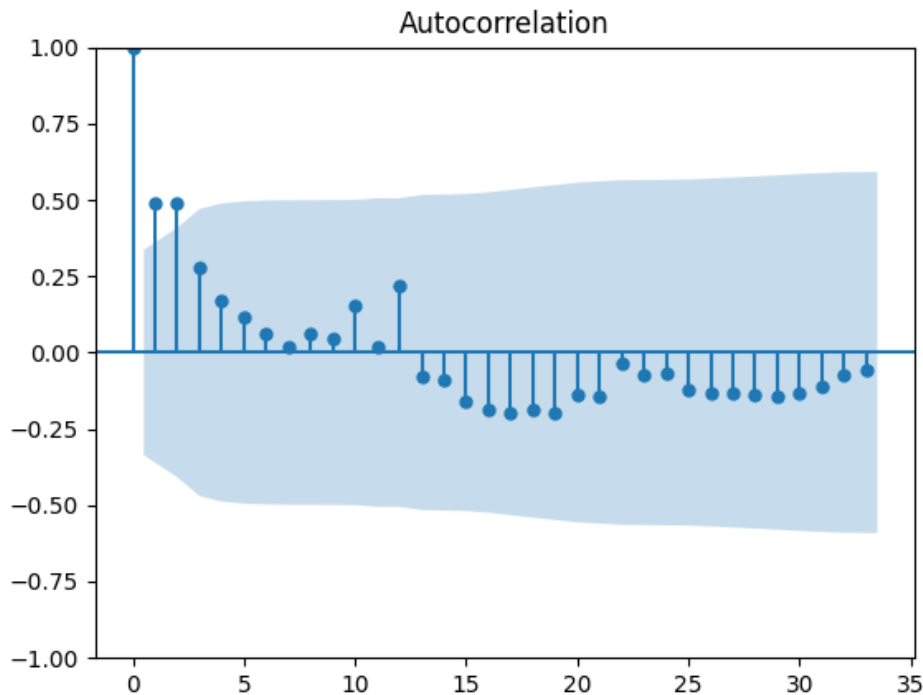
4.8.7 Συνάρτηση αυτοσυσχέτισης για επιλεγμένο προϊόν

Ο κώδικας `AutoCorreletionStatsforAnyItem_0.py` δέχεται σαν είσοδο δεδομένων τα αρχεία:

(i) `ForAllItemsFINAL.csv`

(ii) `ForAllItemsLastMonth.csv`

Αυτός ο κώδικας βρίσκει τους συντελεστές Αυτοσυσχέτισης (Auto Correlation) για ένα συγκεκριμένο προϊόν. Η επιλογή γίνεται από τον χρήστη με τον αντίστοιχο κωδικό ID του προϊόντος. Το πρόγραμμα εμφανίζει στην οθόνη του υπολογιστή το διάγραμμα της συνάρτησης αυτοσυσχέτισης (ACF) ενός επιλεγμένου προϊόντος.



Εικόνα 4.8-4 Συνάρτηση ACF για το προϊόν 5872 (μπάλα γυμναστικής)

4.8.8 Προβλέψεις με δεδομένα 24 μηνών για προϊόντα

4.8.8.1 Γενικά

Για την εφαρμογή αυτή θα πρέπει πρώτα να εκτελεστούν οι παρακάτω κώδικες ώστε να διαμορφωθούν τα κατάλληλα αρχεία.

(α) Κώδικας AllinBothIncomeAndSales24.py

Ο κώδικας αυτός είναι παρόμοιος με το κώδικα `All_in_Both_Income_Sales.py`, ωστόσο έχει τις εξής διαφορές:

- (α) Εκτελείτε όλους τους μήνες εκτός από αυτών που βρίσκονται στην τελευταία χρόνια, στην περίπτωση μας μόνο για δυο πρώτα χρόνια (2013 και 2014).
- (β) Δεν εντοπίζει τον μέσο όρο κάθε μήνα, καθώς και τον συνολικό μέσο όρο.

Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) Το πρωτογενές `items.csv`.
- (ii) Το πρωτογενές `sales_train.csv`.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) ForAllItemsFINAL24.csv. Περιέχει τις ανά μήνα πωλήσεις για όλους τους μήνες, εκτός του τελευταίου χρόνου (στην περίπτωση μας για 2013-2014 και το 2015 είναι εκτός), για κάθε προϊόν ξεχωριστά. Καθώς και τις συνολικές πωλήσεις ανά προϊόν και μήνα.

(β) ForAllItemsFINAL24income.csv. Περιέχει τις ανά μήνα εισπράξεις για όλους τους μήνες εκτός του τελευταίου χρόνου, για κάθε προϊόν ξεχωριστά. Καθώς και τις συνολικές εισπράξεις από τις πωλήσεις ανά προϊόν μέσα στο χρονικό διάστημα που προαναφέρθηκε.

(β) Κώδικας Last Year Real Sales.py

Ο κώδικας αυτός αντιστοιχεί στον προηγούμενο κώδικα, μόνο που είναι για τον τους μήνες του τελευταίου χρόνου. Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) Το πρωτογενές `items.csv`.
- (ii) Το πρωτογενές `sales_train.csv`.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) ForAllItemsLastYear.csv. Περιέχει τις ανά μήνα πωλήσεις για όλους τους μήνες του τελευταίου χρόνου (στην περίπτωση μας για 2015), για κάθε προϊόν ξεχωριστά. Καθώς και τις συνολικές πωλήσεις ανά προϊόν και μήνα.

(β) ForAllItemsLastYearincome.csv. Περιέχει τις ανά μήνα εισπράξεις για όλους τους μήνες του τελευταίου χρόνου, για κάθε προϊόν ξεχωριστά. Καθώς και τις συνολικές εισπράξεις από τις πωλήσεις ανά προϊόν μέσα στο χρονικό διάστημα που προαναφέρθηκε.

(γ) Κώδικας ProductLife24.py

Ο κώδικας αυτός είναι αντίστοιχός με τον κώδικα **ProductLife.py**. Συγκεκριμένα υπολογίζει την προσεγγιστική διάρκεια ζωής κάθε προϊόντος για όλους τους μήνες, εκτός αυτών που βρίσκονται στο τελευταίο χρόνο. Δεν περιλαμβάνεται ο χρόνος ελέγχου (στην περίπτωση μας, το 2015).

Δέχεται σαν είσοδο το εξής αρχείο: **ForAllItemsFINAL24.csv**.

Ο κώδικας αυτός δημιουργεί το αρχείο:

(α) ProductLifes24.csv. Όπως και στο **ProductLifes.csv**, σε αυτό το αρχείο αυτό κρατούνται τρεις τιμές με το date_block_num, αυτές είναι:

- (1) ο μήνας όπου εμφανίζονται για πρώτη φορά πωλήσεις,
- (2) ο μήνας όπου εμφανίζονται για τελευταία φορά πωλήσεις και
- (3) η διάρκεια ζωής του προϊόντος σαν η διαφορά των δύο.

Αμέσως μετά μπορούν να εκτελεστούν οι κώδικες που εφαρμόζουν τις παρακάτω μεθόδους πρόβλεψης.

4.8.8.2 SLR2 – SLR12 με χρήση 24 μηνών χωρίς εποχικότητα

Ο κώδικας **SimpleLinearRegressionFinal24Repeat.py** εκτελεί την μέθοδο SLR όπως ο κώδικας **SimpleLinearRegressionFinalcatSessionRepeat_0.py**, αλλά με τις παρακάτω διαφορές:

- (i) Εκτελούνται για τα προϊόντα και όχι για τις κατηγορίες των προϊόντων.
- (ii) Εκτελεί τις SLR2 – SLR12 και όχι μέχρι την SLR4.
- (iii) Οι προβλέψεις γίνεται με τον ίδιο τρόπο για όλα τα χρόνια εκτός του τελευταίου (στην περίπτωση μας για το 2015). Οπού στον τελευταίο χρόνο οι προβλέψεις γίνονται με βάση την τελευταία εξίσωση της μεθόδου που εφαρμόστηκε πριν αλλάξει η χρονιά. Π.χ. στην SLR 3 χρησιμοποιούνται οι πρώτοι 3 μήνες του 2013 για να γίνει η πρόβλεψη για τον 4^ο μήνα. Για τον 5^ο μήνα χρησιμοποιούνται οι 3 προηγούμενοι του μήνες (από τον 2^ο έως τον 4^ο). Για τον Ιανουάριο του 2015 και μετά στο ίδιο παράδειγμα χρησιμοποιούμε τα στοιχεία του Οκτώβριου, Νοέμβριου και Δεκέμβριου του 2014 για πρόβλεψη για όλους τους μήνες του 2015 με την ίδια εξίσωση. Το ίδιο γίνεται για όλες τις μεθόδους, απλά αλλάζει ο αριθμός των μηνών του 2014 που χρησιμοποιούνται. Π.χ. για τις **SLR 4** χρησιμοποιούνται τα στοιχεία του Σεπτεμβρίου, Οκτώβριου, Νοέμβριου και Δεκέμβριου **2014**, για να προβλεφθούν οι πωλήσεις για όλους τους μήνες του **2015**.

Ο Κώδικας αυτός δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes24.csv.
- (ii) ForAllItemsLastYear.csv.
- (iii) ForAllItemsFINAL24.csv.

Ο κώδικας αυτός δημιουργεί τα αρχεία:

(α) SimpleLinearRegression2_24.csv. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 2 μήνες για την πρόβλεψη.

(β) SimpleLinearRegression3_24.csv. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 3 μήνες για την πρόβλεψη.

- (γ) **SimpleLinearRegression4_24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 4 μήνες για την πρόβλεψη.
- (δ) **SimpleLinearRegression5_24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 5 μήνες για την πρόβλεψη.
- (ε) **SimpleLinearRegression6_24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 6 μήνες για την προβλέψει.
- (στ) **SimpleLinearRegression7_24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 7 μήνες για την πρόβλεψη.
- (ζ) **SimpleLinearRegression8_24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 8 μήνες για την πρόβλεψη.
- (η) **SimpleLinearRegression9_24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 9 μήνες για την πρόβλεψη.
- (θ) **SimpleLinearRegression10_24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 10 μήνες για την πρόβλεψη.
- (ι) **SimpleLinearRegression11_24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 11 μήνες για την πρόβλεψη.
- (ια) **SimpleLinearRegression12_24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 12 μήνες για την πρόβλεψη.

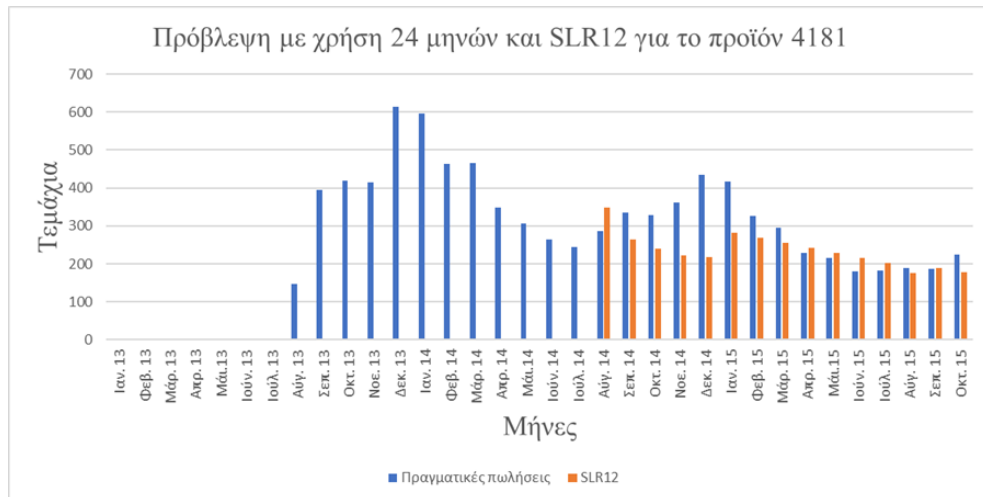
4.8.8.3 SLR2 – SLR12 με χρήση 24 μηνών και εποχικότητα

Ο κώδικας **SimpleLinearRegressionFinal24sessionRepeat_0.py** εκτελεί ακριβώς ότι και ο προηγούμενος ωστόσο εμπλέκει και την εποχικότητα. Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes24.csv.
- (ii) ForAllItemsLastYear.csv.
- (iii) ForAllItemsFINAL24.csv.

Ο κώδικας αυτός δημιουργεί τα αρχεία:

- (α) **SimpleLinearRegression2session24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 2 μήνες για την πρόβλεψη.
- (β) **SimpleLinearRegression3session24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 3 μήνες για την πρόβλεψη.
- (γ) **SimpleLinearRegression4session24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 4 μήνες για την πρόβλεψη.
- (δ) **SimpleLinearRegression5session24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 5 μήνες για την πρόβλεψη.
- (ε) **SimpleLinearRegression6session24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 6 μήνες για την πρόβλεψη.
- (στ) **SimpleLinearRegression7session24.csv**.
Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 7 μήνες για την πρόβλεψη.
- (ζ) **SimpleLinearRegression8session24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 8 μήνες για την πρόβλεψη.
- (η) **SimpleLinearRegression9session24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 9 μήνες για την πρόβλεψη.
- (θ) **SimpleLinearRegression10session24.csv**. Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 10 μήνες για την πρόβλεψη.
- (ι) **SimpleLinearRegression11session24.csv**.
Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 11 μήνες για την πρόβλεψη.
- (ια) **SimpleLinearRegression12session24.csv**.
Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 12 μήνες για την πρόβλεψη.



Εικόνα 4.8-5 Πρόβλεψη για το προϊόν 4181 της κατηγορίας 75 (προγράμματα)

4.8.8.4 QR3 – QR12 με χρήση 24 μηνών χωρίς εποχικότητα

Ο κώδικας **QuadraticRegressionAllItemsFinal24Repeat.py** είναι αντίστοιχος με τον κώδικα **SimpleLinearRegressionFinal24Repeat.py**. Ωστόσο είναι για την μέθοδο QR αντί της μεθόδου SLR.

Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes24.csv.
- (ii) ForAllItemsLastYear.csv.
- (iii) ForAllItemsFINAL24.csv.

Ο κώδικας αυτός δημιουργεί τα αρχεία:

(α) QuadraticRegressionWith3Months24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 3 μήνες για την πρόβλεψη.

(β) QuadraticRegressionWith4Months24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 4 μήνες για την πρόβλεψη.

(γ) QuadraticRegressionWith5Months24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 5 μήνες για την πρόβλεψη.

(δ) QuadraticRegressionWith6Months24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 6 μήνες για την πρόβλεψη.

(ε) QuadraticRegressionWith7Months24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 7 μήνες για την πρόβλεψη.

(στ) QuadraticRegressionWith8Months24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 8 μήνες για την πρόβλεψη.

(ζ) QuadraticRegressionWith9Months24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 9 μήνες για την πρόβλεψη.

(η) QuadraticRegressionWith10Months24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 10 μήνες για την πρόβλεψη.

(θ) QuadraticRegressionWith11Months24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 11 μήνες για την πρόβλεψη.

(ι) QuadraticRegressionWith12Months24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 12 μήνες για την πρόβλεψη.

4.8.8.5 QR3 – QR12 με χρήση 24 μηνών και εποχικότητα

Ο κώδικας **QuadraticRegressionAllItemsFinal24SessionsRepeat.py** είναι αντί-

στοιχος με τον προηγούμενο μόνο που εμπλέκεται σ' αυτόν και η εποχικότητα.

Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes24.csv.
- (ii) ForAllItemsLastYear.csv.
- (iii) ForAllItemsFINAL24.csv.

Ο κώδικας αυτός δημιουργεί τα αρχεία:

(α) QuadraticRegressionWith3MonthsSessions24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 3 μήνες για την πρόβλεψη.

(β) QuadraticRegressionWith4MonthsSessions24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 4 μήνες για την πρόβλεψη.

(γ) QuadraticRegressionWith5MonthsSessions24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 5 μήνες για την πρόβλεψη.

(δ) QuadraticRegressionWith6MonthsSessions24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 6 μήνες για την πρόβλεψη.

(ε) QuadraticRegressionWith7MonthsSessions24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 7 μήνες για την πρόβλεψη.

(στ) QuadraticRegressionWith8MonthsSessions24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 8 μήνες για την πρόβλεψη.

(ζ) QuadraticRegressionWith9MonthsSessions24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 9 μήνες για την πρόβλεψη.

(η) QuadraticRegressionWith10MonthsSessions24.csv.

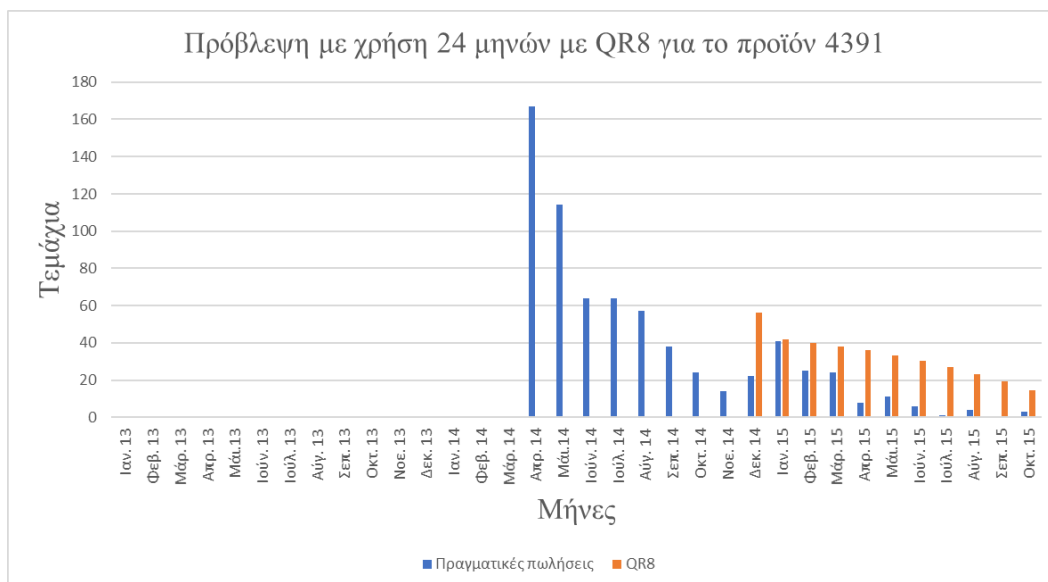
Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 10 μήνες για την πρόβλεψη.

(θ) QuadraticRegressionWith11MonthsSessions24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 11 μήνες για την πρόβλεψη.

(ι) QuadraticRegressionWith12MonthsSessions24.csv.

Εδώ χρησιμοποιώντας αντίστοιχα δεδομένα από τους 12 μήνες για την πρόβλεψη.



Εικόνα 4.8-6 Πρόβλεψη για το προϊόν 4391 της κατηγορίας 30 (Παιχνίδια PC)

4.8.9 MAPE για προβλέψεις με χρήση 24 μηνών

4.8.9.1 MAPE με χρήση 24 μηνών χωρίς εποχικότητα

Ο κώδικας `Mape_24.py` είναι αντίστοιχος με τον κώδικα `Mapecat_0.py`, με την διαφορά ότι η τελευταία χρόνια δεν χρησιμοποιούνται σαν δεδομένα εισόδου (εξαιρείται

η Last Value forecasting Method) και είναι για τα προϊόντα και όχι για τις κατηγορίες προϊόντων.

Αυτός ο κώδικας δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes24.csv
- (ii) ForAllItemsFINAL24.csv
- (iii) ForAllItemsFINALLastYear.csv
- (iv) LastValueForecastingMethodForAllMonths.csv
- (v) SimpleLinearRegression2_24.csv
- (vi) SimpleLinearRegression3_24.csv
- (vii) QuadraticRegressionWith3Months24.csv
- (viii) SimpleLinearRegression4_24.csv
- (ix) QuadraticRegressionWith4Months24.csv
- (x) SimpleLinearRegression5_24.csv
- (xi) QuadraticRegressionWith5Months24.csv
- (xii) SimpleLinearRegression6_24.csv
- (xiii) QuadraticRegressionWith6Months24.csv
- (xiv) SimpleLinearRegression7_24.csv
- (xv) QuadraticRegressionWith7Months24.csv
- (xvi) SimpleLinearRegression8_24.csv
- (xvii) QuadraticRegressionWith8Months24.csv
- (xviii) SimpleLinearRegression9_24.csv
- (xix) QuadraticRegressionWith9Months24.csv
- (xx) SimpleLinearRegression10_24.csv
- (xxi) QuadraticRegressionWith10Months24.csv
- (xxii) SimpleLinearRegression11_24.csv
- (xxiii) QuadraticRegressionWith11Months24.csv
- (xxiv) SimpleLinearRegression12_24
- (xxv) QuadraticRegressionWith12Months24.csv

Από αυτά τα αρχεία ο κώδικας θα εκτελέσει τον υπολογισμό του MAPE για κάθε κατηγορία προϊόντος σε κάθε μήνα και σε κάθε μια από τις εξής μεθόδους:

- (α) Last Value Forecasting Method
- (β) Simple Linear Regression από 2 μήνες έως 12 μήνες δεδομένων.
- (γ) Quadratic Regression από 3 μήνες έως 12 μήνες δεδομένων.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) MapeAll24.csv. Στο οποίο καταγράφονται τα αποτελέσματα για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCounters24.csv. Στο οποίο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.8.9.2 MAPE με χρήση 24 μηνών και εποχικότητα

Ο κώδικας **MapeSessions24.py** είναι αντίστοιχός με τον προηγούμενο κώδικα **Mape_24.py**, μόνο που εμπεριέχεται και η εποχικότητα στον υπολογισμό.

Αυτός ο κώδικας παίρνει σαν είσοδο τα εξής αρχεία:

- (i) ProductLifes24.csv
- (ii) ForAllItemsFINAL24.csv
- (iii) ForAllItemsFINALLastYear.csv
- (iv) LastValueForecastingMethodForAllMonths.csv
- (v) SimpleLinearRegression2session24.csv.csv
- (vi) SimpleLinearRegression3session24.csv.csv

- (vii) QuadraticRegressionWith3MonthsSessions24.csv
- (viii) SimpleLinearRegression4session24.csv
- (ix) QuadraticRegressionWith4MonthsSessions24.csv
- (x) SimpleLinearRegression5session24.csv
- (xi) QuadraticRegressionWith5MonthsSessions24.csv
- (xii) SimpleLinearRegression6session24.csv
- (xiii) QuadraticRegressionWith6MonthsSessions24.csv
- (xiv) SimpleLinearRegression7session24.csv
- (xv) QuadraticRegressionWith7MonthsSessions24.csv
- (xvi) SimpleLinearRegression8session24.csv
- (xvii) QuadraticRegressionWith8MonthsSessions24.csv
- (xviii) SimpleLinearRegression9session24.csv
- (xix) QuadraticRegressionWith9MonthsSessions24.csv
- (xx) SimpleLinearRegression10session24.csv
- (xxi) QuadraticRegressionWith10MonthsSessions24.csv
- (xxii) SimpleLinearRegression11session24.csv
- (xxiii) QuadraticRegressionWith11MonthsSessions24.csv
- (xxiv) SimpleLinearRegression12session24.csv
- (xxv) QuadraticRegressionWith12MonthsSessions24.csv

Από αυτά τα αρχεία ο κώδικας θα εκτελέσει τον υπολογισμό του MAPE για κάθε προϊόν σε κάθε μήνα και σε κάθε μια από τις εξής μεθόδους:

- (α) Last Value Forecasting Method
- (β) Simple Linear Regression από 2 μήνες έως 12 μήνες δεδομένων.
- (γ) Quadratic Regression από 3 μήνες έως 12 μήνες δεδομένων.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

- (α) **MapeAllSessions24**. Στο όποιο καταγράφονται τα αποτελέσματα για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.
- (β) **MapeAllMonthCountersSessions24.csv**. Στο όποιο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.8.10 Προβλέψεις με χρήση 24 μηνών για κατηγορίες

4.8.10.1 Γενικά

Για την εφαρμογή αυτή θα πρέπει πρώτα να εκτελεστούν οι παρακάτω κώδικες ώστε να διαμορφωθούν τα κατάλληλα αρχεία.

(1) AllinBothCategoriesIncomeAndSales24.py

Ο κώδικας **AllinBothCategoriesIncomeAndSales24.py** είναι αντίστοιχος με τους κώδικες **AllinBothIncomeAndSales24.py** και **Last_Year_Real_Sales.py** με διαφορά ότι στις κατηγορίες προϊόντων και όχι για τα προϊόντα.

Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) Το πρωτογενές item_categories.csv.
- (ii) ForAllItemsFINAL24income.csv.
- (iii) ForAllItemsFINAL24.csv
- (iv) ForAllItemsLastYearincome.csv
- (v) ForAllItemsLastYear.csv

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

- (α) ForAllCategoriesFINAL24.csv.
- (β) ForAllCategoriesFINAL24income.csv

- (γ) ForAllCategoriesFINALLastYear.csv
- (δ) ForAllCategoriesFINALLastYearincome.csv

(2) ProductLifeCategories24.py

Ο κώδικας αυτός είναι αντίστοιχός με τον κώδικα **ProductLife24.py** με διαφορά ότι είναι για τις κατηγορίες προϊόντων και όχι για τα προϊόντα.

Δέχεται σαν είσοδο το εξής αρχείο: ForAllCategoriesFINAL24.csv

Αμέσως μετά θα δημιουργήσει το εξής αρχείο: ProductLifesCategories24.csv.

4.8.10.2 SLR2 – SLR12 με χρήση 24 μηνών χωρίς εποχικότητα

Ο κώδικας **SimpleLinearRegressionFinal24catRepeat_0.py** είναι αντίστοιχος με τον κώδικα: **SimpleLinearRegressionFinal24Repeat.py**, με διαφορά ότι είναι για τις κατηγορίες προϊόντων και όχι για τα προϊόντα.

Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifesCategories24.
- (ii) ForAllCategoriesFINAL24.csv.
- (iii) ForAllCategoriesFINALLastYear.csv

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

- (α) SimpleLinearRegression2cat24_0.csv
- (β) SimpleLinearRegression3cat24_0.csv
- (γ) SimpleLinearRegression4cat24_0.csv
- (δ) SimpleLinearRegression5cat24_0.csv
- (ε) SimpleLinearRegression6cat24_0.csv
- (στ) SimpleLinearRegression7cat24_0.csv
- (ζ) SimpleLinearRegression8cat24_0.csv
- (η) SimpleLinearRegression9cat24_0.csv
- (θ) SimpleLinearRegression10cat24_0.csv
- (ι) SimpleLinearRegression11cat24_0.csv
- (ια) SimpleLinearRegression12cat24_0.csv

4.8.10.3 SLR2 – SLR12 με χρήση 24 μηνών και εποχικότητα

Ο κώδικας **SimpleLinearRegressionFinal24catSessonRepeat_0.py** είναι αντίστοιχος με τον προηγούμενο κώδικα: **SimpleLinearRegressionFinal24catRepeat_0.py**, με διαφορά ότι περιέχει και η εποχικότητα.

Δέχεται σαν είσοδο τα εξής αρχεία:

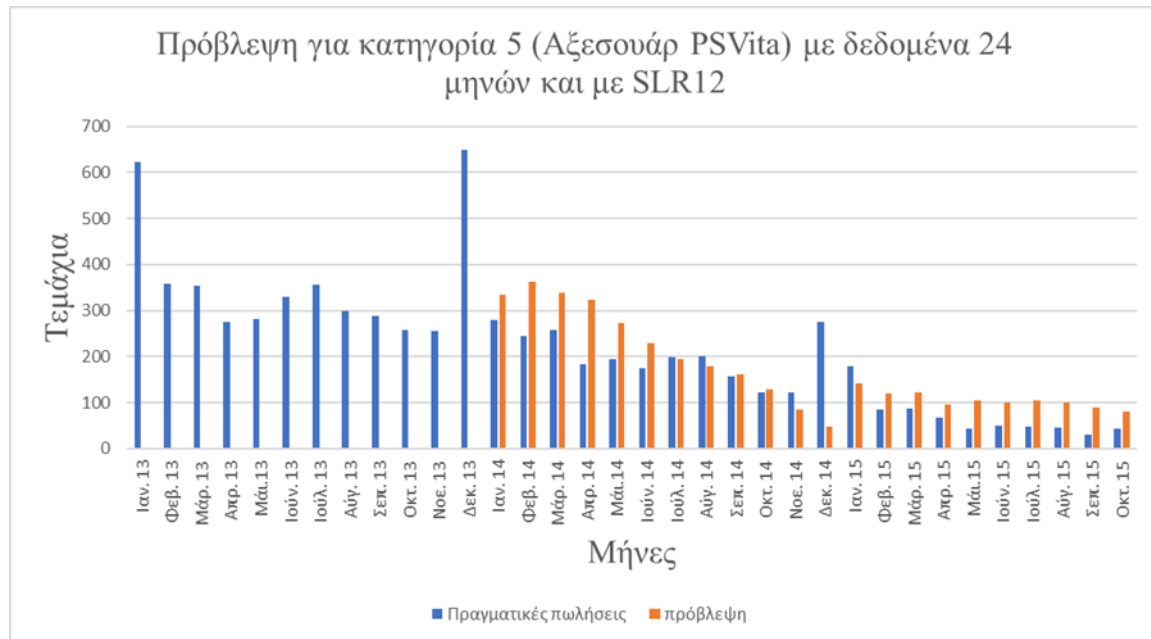
- (i) ProductLifesCategories24.
- (ii) ForAllCategoriesFINAL24.csv.
- (iii) ForAllCategoriesFINALLastYear.csv

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

- (α) SimpleLinearRegression2catSesson24_0.csv
- (β) SimpleLinearRegression3catSesson24_0.csv
- (γ) SimpleLinearRegression4catSesson24_0.csv
- (δ) SimpleLinearRegression5catSesson24_0.csv
- (ε) SimpleLinearRegression6catSesson24_0.csv
- (στ) SimpleLinearRegression7catSesson24_0.csv
- (ζ) SimpleLinearRegression8catSesson24_0.csv
- (η) SimpleLinearRegression9catSesson24_0.csv
- (θ) SimpleLinearRegression10catSesson24_0.csv

(ι) SimpleLinearRegression11catSesson24_0.csv

(ια) SimpleLinearRegression12catSesson24_0.csv



Εικόνα 4.8-7 Πρόβλεψη για κατηγορία 5 με χρήση 24 μηνών και SLR12

4.8.10.4 QR3 – QR12 με χρήση 24 μηνών χωρίς εποχικότητα

Ο κώδικας **QuadraticRegressionAllItemsFinal24catRepeat_0.py** είναι αντίστοιχος με τον κώδικα: **QuadraticRegressionAllItemsFinal24Repeat**, είναι για τις κατηγορίες προϊόντων και όχι για τα προϊόντα.

Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifesCategories24.
- (ii) ForAllCategoriesFINAL24.csv.
- (iii) ForAllCategoriesFINALLastYear.csv

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

- (α) QuadraticRegressionWith3MonthsCAT24_0.csv
- (β) QuadraticRegressionWith4MonthsCAT24_0.csv
- (γ) QuadraticRegressionWith5MonthsCAT24_0.csv
- (δ) QuadraticRegressionWith6MonthsCAT24_0.csv
- (ε) QuadraticRegressionWith7MonthsCAT24_0.csv
- (στ) QuadraticRegressionWith8MonthsCAT24_0.csv
- (ζ) QuadraticRegressionWith9MonthsCAT24_0.csv
- (η) QuadraticRegressionWith10MonthsCAT24_0.csv
- (θ) QuadraticRegressionWith11MonthsCAT24_0.csv
- (ι) QuadraticRegressionWith12MonthsCAT24_0.csv

4.8.10.5 QR3 – QR12 χρήση 24 μηνών και εποχικότητα

Ο κώδικας **QuadraticRegressionAllItemsFinal24catSessionsRepeat_0.py** είναι αντίστοιχος με τον προηγούμενο κώδικα:

QuadraticRegressionAllItemsFinal24catRepeat_0.py, με διαφορά ότι εμπεριέχεται και η εποχικότητα.

Δέχεται σαν είσοδο τα εξής αρχεία:

- (i) ProductLifesCategories24.
- (ii) ForAllCategoriesFINAL24.csv.
- (iii) ForAllCategoriesFINALLastYear.csv

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

- (α) QuadraticRegressionWith3MonthsCATSesson24_0.csv
- (β) QuadraticRegressionWith4MonthsCATSesson24_0.csv
- (γ) QuadraticRegressionWith5MonthsCATSesson24_0.csv
- (δ) QuadraticRegressionWith6MonthsCATSesson24_0.csv
- (ε) QuadraticRegressionWith7MonthsCATSesson24_0.csv
- (στ) QuadraticRegressionWith8MonthsCATSesson24_0.csv
- (ζ) QuadraticRegressionWith9MonthsCATSesson24_0.csv
- (η) QuadraticRegressionWith10MonthsCATSesson24_0.csv
- (θ) QuadraticRegressionWith11MonthsCATSesson24_0.csv
- (ι) QuadraticRegressionWith12MonthsCATSesson24_0.csv



Εικόνα 4.8-8 Πρόβλεψη για κατηγορία 4 με χρήση 24 μηνών και QR12

4.8.10.6 MAPE με χρήση 24 μηνών χωρίς εποχικότητα

Ο κώδικας **MapeCat24_0.py** είναι αντίστοιχος με τον κώδικα **Mapecat_0.py**, με την διαφορά ότι η τελευταία χρόνια δεν χρησιμοποιούνται σαν δεδομένα είσοδού (εξαιρείται η Last Value forecasting Method)

Αυτός ο κώδικας παίρνει σαν είσοδο τα εξής αρχεία:

- (i) ProductLifesCategories24.csv
- (ii) ForAllCategoriesFINAL24.csv
- (iii) ForAllCategoriesFINALLastYear.csv
- (iv) LastValueForecastingMethodForAllMonthscat.csv
- (v) SimpleLinearRegression2cat24_0.csv
- (vi) SimpleLinearRegression3cat24_0.csv
- (vii) QuadraticRegressionWith3MonthsCAT24_0.csv"
- (viii) SimpleLinearRegression4cat24_0.csv

- (ix) QuadraticRegressionWith4MonthsCAT24_0.csv
- (x) SimpleLinearRegression5cat24_0.csv
- (xi) QuadraticRegressionWith5MonthsCAT24_0.csv
- (xii) SimpleLinearRegression6cat24_0.csv
- (xiii) QuadraticRegressionWith6MonthsCAT24_0.csv
- (xiv) SimpleLinearRegression7cat24_0.csv
- (xv) QuadraticRegressionWith7MonthsCAT24_0.csv
- (xvi) SimpleLinearRegression8cat24_0.csv
- (xvii) QuadraticRegressionWith8MonthsCAT24_0.csv
- (xviii) SimpleLinearRegression9cat24_0.csv
- (xix) QuadraticRegressionWith9MonthsCAT24_0.csv
- (xx) SimpleLinearRegression10cat24_0.csv
- (xxi) QuadraticRegressionWith10MonthsCAT24_0.csv
- (xxii) SimpleLinearRegression11cat24_0.csv
- (xxiii) QuadraticRegressionWith11MonthsCAT24_0.csv
- (xxiv) SimpleLinearRegression12cat24_0.csv
- (xxv) QuadraticRegressionWith12MonthsCAT24_0.csv

Από αυτά τα αρχεία ο κώδικας θα εκτελέσει τον υπολογισμό του MAPE για κάθε κατηγορία προϊόντος σε κάθε μήνα και σε κάθε μια από τις εξής μεθόδους:

- (α) Last Value Forecasting Method
- (β) Simple Linear Regression από 2 μήνες έως 12 μήνες δεδομένων.
- (γ) Quadratic Regression από 3 μήνες έως 12 μήνες δεδομένων.

Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

(α) MapeAllcat24_0.csv. Στο οποίο καταγράφονται τα αποτελέσματα για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCounterscat24_0.csv. Στο οποίο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.8.10.7 MAPE με χρήση 24 μηνών και εποχικότητα

Ο κώδικας **MapeCatSessions24_0.py** είναι αντίστοιχος με τον προηγούμενο κώδικα **Mapecat24_0.py**, με την διαφορά ότι εμπεριέχεται και η εποχικότητα στους υπολογισμούς.

Αυτός ο κώδικας παίρνει σαν είσοδο τα εξής αρχεία:

- (i) ProductLifesCategories24.csv
- (ii) ForAllCategoriesFINAL24.csv
- (iii) ForAllCategoriesFINALLastYear.csv
- (iv) LastValueForecastingMethodForAllMonthscat.csv
- (v) SimpleLinearRegression2catSesson24_0.csv
- (vi) SimpleLinearRegression3catSesson24_0.csv
- (vii) QuadraticRegressionWith3MonthsCATSesson24_0.csv"
- (viii) SimpleLinearRegression4catSesson24_0.csv
- (ix) QuadraticRegressionWith4MonthsCATSesson24_0.csv
- (x) SimpleLinearRegression5catSesson24_0.csv
- (xi) QuadraticRegressionWith5MonthsCATSesson24_0.csv
- (xii) SimpleLinearRegression6catSesson24_0.csv
- (xiii) QuadraticRegressionWith6MonthsCATSesson24_0.csv
- (xiv) SimpleLinearRegression7catSesson24_0.csv
- (xv) QuadraticRegressionWith7MonthsCATSesson24_0.csv

- (xvi) SimpleLinearRegression8catSesson24_0.csv
- (xvii) QuadraticRegressionWith8MonthsCATSesson24_0.csv
- (xviii) SimpleLinearRegression9catSesson24_0.csv
- (xix) QuadraticRegressionWith9MonthsCATSesson24_0.csv
- (xx) SimpleLinearRegression10catSesson24_0.csv
- (xxi) QuadraticRegressionWith10MonthsCATSesson24_0.csv
- (xxii) SimpleLinearRegression11catSesson24_0.csv
- (xxiii) QuadraticRegressionWith11MonthsCATSesson24_0.csv
- (xxiv) SimpleLinearRegression12catSesson24_0.csv
- (xxv) QuadraticRegressionWith12MonthsCATSesson24_0.csv

Από αυτά τα αρχεία ο κώδικας θα εκτελέσει τον υπολογισμό του MAPE για κάθε κατηγορία προϊόντος σε κάθε μήνα και σε κάθε μια από τις εξής μεθόδους:

- (α) Last Value Forecasting Method
 - (β) Simple Linear Regression από 2 μήνες έως 12 μήνες δεδομένων.
 - (γ) Quadratic Regression από 3 μήνες έως 12 μήνες δεδομένων.
- Αμέσως μετά θα δημιουργήσει τα εξής αρχεία:

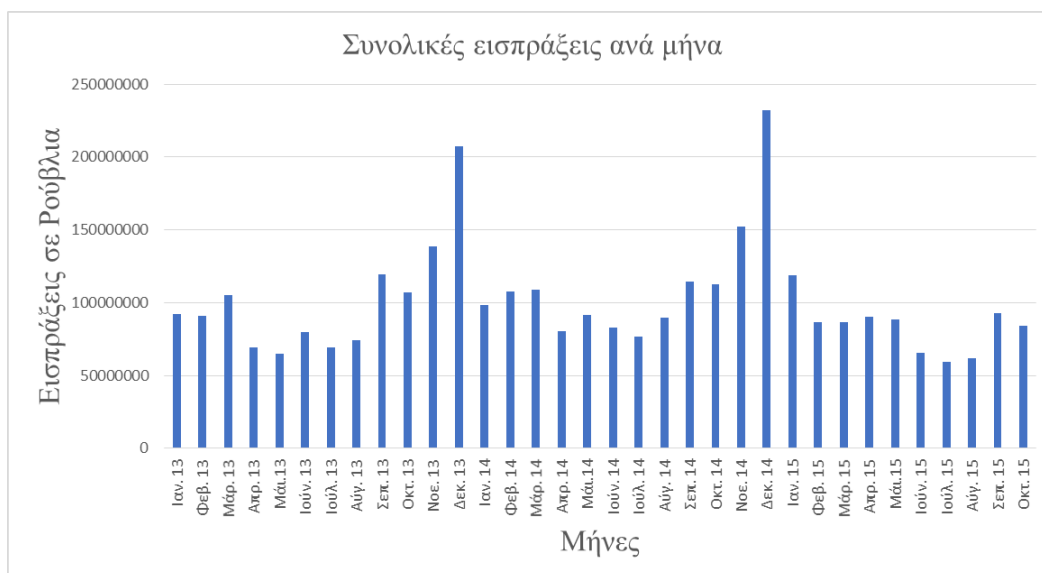
(α) MapeAllcatSesson24_0.csv. Στο όποιο καταγράφονται τα αποτελέσματα για κάθε ένα μηνά ανά κάθε μέθοδο, ξεχωριστά.

(β) MapeAllMonthCounterscatSesson24_0.csv. Στο όποιο καταγράφονται πόσες από τις κατηγορίες αφαιρούνται από το σύνολο, σε κάθε έναν από τους μήνες λόγω μηδενικών πωλήσεων.

4.9 Εισπράξεις

4.9.1 Συνολικές εισπράξεις ανά μήνα

Οι τιμές πώλησης ανά «τεμάχιο» για κάθε προϊόν εμφανίζονται στο πρωτογενές αρχείο `sales_train.csv`. Ο κώδικας `All_in_Both_Income_Sales.py` βρίσκει τις μέσες τιμές πώλησης ανά προϊόν και μήνα. Η μέση τιμή είναι απαραίτητο να βρεθεί καθώς μέσα στον ίδιο μήνα κάποιο προϊόν μπορεί να πωλείται σε διαφορετική τιμή (προσφορά) από τα διάφορα σημεία πώλησης ή ακόμα και για κάποιες μέρες μέσα στον ίδιο μήνα (εκπτώσεις). Επίσης στο ίδιο αρχείο φαίνονται και οι συνολικές εισπράξεις για το συγκεκριμένο προϊόν.



Εικόνα 4.9-1 Εισπράξεις ανά μήνα

Από τα παραπάνω αρχεία προκύπτει ότι οι συνολικές εισπράξεις είναι 3.022.163.678 ρούβλια (36.265.964 €, ισοτιμία 1 P = 0,012 €). Ενώ στον παρακάτω πίνακα 4.9-1 φαίνονται οι εισπράξεις ανά κατηγορία πάντα σε ρούβλια.

4.9.2 Συνολικές εισπράξεις ανά κατηγορία και μήνα

Ο κώδικας `AllinCategoriesBothIncomeAndSales.py` υπολογίζει τις συνολικές εισπράξεις για κάθε κατηγορία και μήνα. Τα αποτελέσματα εμφανίζονται στο αρχείο `ForAllCategoriesFINALIncome.csv`. Στην τελευταία στήλη (SUM) αυτού του αρχείου εμφανίζονται οι συνολικές εισπράξεις ανά κατηγορία που παρουσιάζονται στον πίνακα 4.9-1.

Πίνακας 4.9-1 Συνολικές εισπράξεις ανά κατηγορία

Κατηγορία	Εισπράξεις	Κατηγορία	Εισπράξεις	Κατηγορία	Εισπράξεις	Κατηγορία	Εισπράξεις
0	263	21	18.906.847	42	4.619.705	63	45.453.093
1	431	22	32.940.026	43	8.217.934	64	40.106.424
2	43.564.198	23	259.599.263	44	37.619	65	24.727.569
3	62.553.410	24	40.808.014	45	1.348.056	66	442.343
4	1.332.134	25	16.501.961	46	4.502	67	30.374.888
5	6.384.797	26	18.318	47	3.414.709	68	332.243
6	38.220.394	27	5.422	48	1.654	69	30.209.074
7	14.330.276	28	136.400.205	49	11.503.386	70	16.561.651
8	19.308.359	29	60.218.311	50	58	71	931.521
9	6.530.909	30	257.489.627	51	129	72	68.951.505
10	1.998	31	23.823.282	52	300	73	26.236.663
11	81.117.219	32	6.706.100	53	847	74	199.808
12	292.238.749	33	10.794.260	54	1.199.628	75	87.085.530
13	1.625.566	34	8.068.419	55	100.021.854	76	7.392.451
14	28.235.448	35	47.367.947	56	9.499.615	77	792.493
15	71.880.694	36	16.000	57	9.198.937	78	683.492
16	52.226.664	37	90.468.309	58	23.216.716	79	17.198.168
17	699.792	38	37.558.272	59	2.825.823	80	8.360.326
18	3.828	39	281.485	60	4.956.141	81	175.741
19	412.027.464	40	169.937.941	61	22.316.108	82	336.264
20	373.000.257	41	36.722.958	62	28.733.034	83	1.189.199

Πίνακας 4.9-2 Μέσες τιμές εισπράξης ανά προϊόν και μήνα

Item_id	989	990	991	992	993	994	995	996	997	998	999	1000	1001
Item_category_id	67	67	67	67	67	67	67	67	67	67	67	67	67
Jan 2013	68,27273	82,53571	91,29032	86,9	91,88235	99	79,2	65,73333	88,83333	87,26829	85,65625	86,44118	90,8
Feb 2013	71,6875	82,19231	75,68293	67,73684	71,30556	88,75	99	77,27273	90,8	66,60465	68,95556	60,1087	70,36111
Mar 2013	94,93333	95,27273	88,65517	91,25	97,04762	99	99	93,875	99	91,48889	97,13636	92,33333	99
Apr 2013	94,9	92,4	87,57692	99	83,63636	99	99	88,75	99	91,76	89,34483	90,42857	81,5
May 2013	86,625	85	99	74,25	99	66	66	49,5	99	99	89,04167	86,93333	90
Jun 2013	99	89,53846	96,4375	99	96,07143	66	99	99	99	92,63636	97,36	94,44444	93,14286
Jul 2013	99	99	99	99	86,625	99	99	99	99	99	92,63636	99	91,38462
Aug 2013	88	99	93,53333	99	99	99	99	99	99	91,6	89,96774	96,84211	96,07143
Sep 2013	99	71,92857	102,8667	99	99	99	99	99	99	99	95,84615	84,09091	94,9
Oct 2013	99	87,28571	99	99	93,875	99	99	99	99	96,07143	87,31579	96,26667	87,28571
Nov 2013	99	83,625	91,3125	99	99	99	99	99	78,5	94,9	96,58824	83,91667	99
Dec 2013	68,55	72,375	67,35484	63,91304	73,70588	69	72,75	70,16667	61,66667	67,73913	63,70588	75,05556	71,61538
Jan 2014	69	58,88889	67,42857	63,28571	57,5	69	69	69	69	63,5	62	65,33333	69
Feb 2014	69	67,77778	69	51,75	69	39,42857	69	63,25	69	69	53,42857	69	61
Mar 2014	69	56	69	69	69	69	69	61,33333	69	69	69	69	69
Apr 2014	69	69	69	69	69	69	69	69	69	69	69	69	69
May 2014	69	66,25	69	69	65,25	69	69	55,2	58	69	64,33333	69	61,5
Jun 2014	69	63,5	69	63,5	69	69	69	69	58	69	69	69	69
Jul 2014	69	69	51,75	69	69	69	69	69	58	69	69	86,25	69
Aug 2014	52	61,66667	69	58	69	60,75	58	60,75	28	69	65,33333	61,66667	69
Sep 2014	58	58	69	58	69	58	58	58	28	69	65,33333	61,66667	69
Oct 2014	58	58	58	58	58	58	58	49,71429	28	69	58	58	58
Nov 2014	58	58	58	58	58	58	58	58	28	69	58	58	58
Dec 2014	34,8	58	58	58	58	58	58	43,5	28	69	58	58	58
Jan 2015	34,8	58	58	58	58	58	58	58	28	69	58	58	58
Feb 2015	34,8	58	58	58	58	28	28	28	28	69	58	28	58
Mar 2015	34,8	58	58	58	58	28	28	28	28	69	58	28	58
Apr 2015	34,8	58	58	58	58	28	28	28	28	69	58	28	58
May 2015	34,8	58	58	58	58	14	28	28	28	69	58	28	58
Jun 2015	34,8	58	58	58	58	14	28	28	28	69	58	28	58
Jul 2015	34,8	58	58	58	58	14	28	28	28	69	58	28	58
Aug 2015	34,8	58	58	58	58	14	28	28	28	69	58	28	58
Sep 2015	34,8	58	58	58	58	14	28	28	28	69	58	28	58
SUM	13822	17537	25302	11961	21699	5128	6569	13045	4158	28287	31834	22476	18850

4.9.3 Μέση τιμή εισπράξης ανά προϊόν και μήνα

Στο παρακάτω πίνακα δίνεται ένα μικρό απόσπασμα από το αρχείο:

ForAllItemsFINALincome_items.csv.

Αυτό το αρχείο περιέχει, όπως έχει αναφερθεί στην ενότητα 4.3.2 τις μέσες τιμές των προϊόντων για κάθε μήνα εκτός από τον τελευταίο. Εάν ένα προϊόν δεν «πούλησε» σε κάποιον μήνα τότε στο αρχείο κρατιέται η μέση τιμή του προηγούμενου μήνα. Για τον τελευταίο μήνα το αντίστοιχο αρχείο είναι το **ForAllItemsLastMonth.csv**.

Το αρχείο αυτό δεν μπορεί εδώ να παρουσιαστεί ολόκληρο καθώς περιέχει 22170 στήλες.

5 Επίλογος

5.1 Σύνοψη και συμπεράσματα

5.1.1 Διάρκεια ζωής προϊόντων και κατηγοριών

Πριν αναφερθούμε στα συμπεράσματα θα πρέπει να σταθούμε για λίγο σε μερικά ποιοτικά στατιστικά στοιχεία, που προέκυψαν από την ανάλυση των στοιχείων:

(1) Από τον κώδικά **ProductLife.py** προκύπτουν τα εξής στοιχεία:

Από τα 22170 προϊόντα μόνο τα 1014 παρουσιάζουν πωλήσεις σε όλους τους μήνες όπου υπάρχουν δεδομένα (ποσοστό 4,57%).

841 δεν παρουσίασαν πωλήσεις (ποσοστό 3,79%)

2467 παρουσίασαν πωλήσεις μόνο για έναν μήνα (ποσοστό 11,13%)

12592 σταμάτησαν να πωλούν τουλάχιστον 5 μήνες πριν τον τελευταίο μήνα των δεδομένων (Οκτώβριος 2015) (ποσοστό 56,78%)

302 προϊόντα παρουσίασαν πωλήσεις μόνο τον τελευταίο μήνα των δεδομένων (Οκτώβριος 2015). Πρόκειται για καινούργια προϊόντα (ποσοστό 1,36%)

17293 προϊόντα έχουν διάρκεια ζωής μεγαλύτερη από 3 μήνες (ποσοστό 78 %)

Τέλος η μέση διάρκεια ζωή ενός προϊόντος είναι **16,85 μήνες**.

```
===== RESTART: C:\Python\Python311\ProductLife.py =====
the csv file ProductLifes have been created
The number of items that have been sold from the beginning to the end is : 8110
That did not sell is:  841 3.7934145241317094 %
That only sell for 1 month is:  0 0.0 %
Items that have been dead for the past 5 months 0 0.0 %
Items that have been sold for the last 2 months 584 2.634190347316193 %
Items that have been sold for the last month 302 1.3622011727559766 %
The mean of life Duration is:  26.482670523017596
And the number of items that have more that 3 months life is:  20745 93.5723951285521 %
```

Εικόνα 5.1-1 Στοιχεία για την διάρκεια ζωής των προϊόντων

(2) Από τον κώδικα **product LifeCategories.py** προκύπτουν τα εξής στοιχεία:

Από τις 84 κατηγορίες οι 45 παρουσιάζουν πωλήσεις σε όλους τους μήνες όπου υπάρχουν δεδομένα (ποσοστό 53,57%).

0 δεν παρουσίασαν πωλήσεις (ποσοστό 0,00%)

4 παρουσίασαν πωλήσεις μόνο για έναν μήνα (ποσοστό 4,76%)

15 σταμάτησαν να πωλούν τουλάχιστον 5 μήνες πριν τον τελευταίο μήνα των δεδομένων (Οκτώβριος 2015) (ποσοστό 17,86%)

1 κατηγορία παρουσίασε πωλήσεις μόνο τον τελευταίο μήνα των δεδομένων (Οκτώβριος 2015). Πρόκειται για καινούργια κατηγορία (ποσοστό 1,19%)

78 κατηγορίες έχουν διάρκεια ζωής μεγαλύτερη από 3 μήνες (ποσοστό 92,86 %)

Τέλος η μέση διάρκεια ζωή ενός προϊόντος είναι 27,53 μήνες.

```
===== RESTART: C:\Python\Python311\ProductLifeCategories.py =====
the csv file ProductLifesCategories have been created
The number of categories that have been sold from the beginning to the end is : 45
That did not sell is:  0 0.0 %
That only sell for 1 month is:  4 4.761904761904762 %
Categories that have been dead for the past 5 months 15 17.857142857142858 %
Categories that have been sold for the last 2 months 1 1.1904761904761905 %
Categories that have been sold for the last month 1 1.1904761904761905 %
The mean of life Duration is:  27.525641025641026
And the number of Categories that have more that 3 months life is:  78 92.85714285714286 %
```

Εικόνα 5.1-2 Στοιχεία για την διάρκεια ζωής των κατηγοριών

5.1.2 Αυτοσυσχέτιση και εποχικότητα

Όπως αναφέρθηκε στο κεφάλαιο 4 από τον κώδικα:

AutoCorrelationFinalCategories_0.py προκύπτει το αρχείο:

AutoCorrelationCategoriesIsItNoise.csv

Στο αρχείο αυτό οι κατηγορίες που παρουσιάζουν περισσότερους από το 5% των συντελεστών αυτοσυσχέτισης να ξεπερνούν τα όρια του λευκού θορύβου μαρκάρονται σαν «FALSE». Δηλαδή δεν είναι λευκός θόρυβος και άρα παρουσιάζουν εποχικότητα. Το «TRUE» απαντά στην ερώτηση «είναι λευκός θόρυβος;»

Το αρχείο αυτό έχει την μορφή που φαίνεται στην παρακάτω εικόνα.

Πίνακας 5.1-1 Λευκός θόρυβος (True)

item_category_id	Noise	item_category_id	Noise	item_category_id	Noise	item_category_id	Noise
0	TRUE	21	FALSE	42	FALSE	63	TRUE
1	TRUE	22	FALSE	43	FALSE	64	FALSE
2	FALSE	23	TRUE	44	TRUE	65	TRUE
3	FALSE	24	FALSE	45	FALSE	66	FALSE
4	FALSE	25	FALSE	46	TRUE	67	FALSE
5	FALSE	26	FALSE	47	TRUE	68	FALSE
6	FALSE	27	TRUE	48	TRUE	69	TRUE
7	TRUE	28	TRUE	49	TRUE	70	FALSE
8	TRUE	29	TRUE	50	TRUE	71	FALSE
9	TRUE	30	FALSE	51	TRUE	72	FALSE
10	TRUE	31	FALSE	52	TRUE	73	FALSE
11	FALSE	32	FALSE	53	TRUE	74	FALSE
12	TRUE	33	FALSE	54	FALSE	75	FALSE
13	FALSE	34	FALSE	55	FALSE	76	FALSE
14	TRUE	35	FALSE	56	TRUE	77	FALSE
15	TRUE	36	TRUE	57	FALSE	78	FALSE
16	FALSE	37	FALSE	58	FALSE	79	FALSE
17	FALSE	38	FALSE	59	FALSE	80	TRUE
18	TRUE	39	TRUE	60	FALSE	81	FALSE
19	FALSE	40	FALSE	61	FALSE	82	FALSE
20	FALSE	41	FALSE	62	FALSE	83	FALSE

Από τον παραπάνω πίνακα φαίνεται ότι από τις 84 κατηγορίες προϊόντων οι 54 παρουσιάζουν εποχικότητα.

Από τις υπόλοιπες 30 κατηγορίες, οι 12 παρουσιάζουν πολύ μικρές συνολικές πωλήσεις (λιγότερες από 22 τεμάχια στους 34 μήνες).

Άλλες 6 κατηγορίες παρουσιάζουν πωλήσεις μόνο τους τελευταίους μήνες, ενώ η κατηγορία 80 θα πρέπει εξαιρεθεί καθώς αποτελεί λανθασμένη καταχώρηση από την εταιρεία, όπως έχει ήδη αναφερθεί.

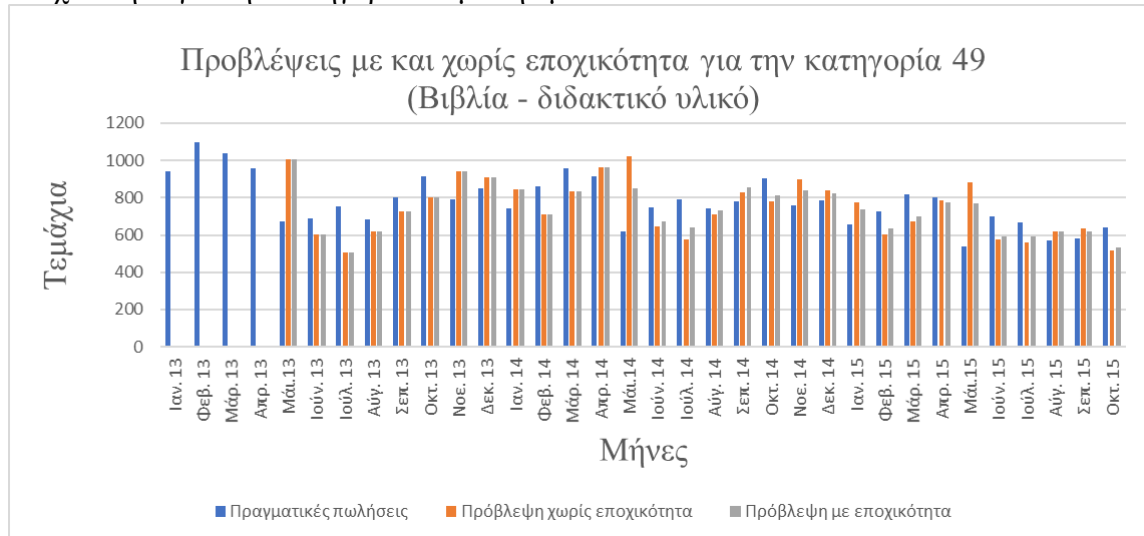
Έτσι απομένουν 11 κατηγορίες που φαίνεται να μην χαρακτηρίζονται από εποχικότητα. Οι κατηγορίες αυτές αναφέρονται στον παρακάτω πίνακα.

Πίνακας 5.1-2 Κατηγορίες χωρίς εποχικότητα με βάση την αυτοσυσχέτιση

a/a	Κωδικός	Είδος	MAPE χωρίς εποχικότητα	MAPE με εποχικότητα
1	9	Παράδοση αγαθών	95,17101172	84,36911809
2	14	Κονσόλες – PSVita	69,77658919	51,77472293
3	15	Κονσόλες XBOX 360	64,40623037	52,17503011
4	23	Παιχνίδια XBOX 360	33,94648681	32,65964329
5	28	Παιχνίδια PC έκτακτες εκδόσεις	53,9550586	52,59226576
6	29	Παιχνίδια PC συλλεκτικές εκδόσεις	165,8614277	131,073234
7	49	Διδακτικό υλικό	18,24322903	14,9552115
8	56	Μουσική επάνυμα CD	52,49910272	46,81689545
9	63	Δώρα παιχνίδια	55,52429503	52,21614145
10	65	Επιτραπέζια παιχνίδια	23,82796562	21,02488624
11	69	Δώρα αναμνηστικά	52,15481387	44,109189

Στον πίνακα 5.1-2 φαίνονται τα MAPE για τις προβλέψεις με και χωρίς εποχικότητα. Είναι φανερό ότι εφαρμόζοντας την εποχικότητα και σε αυτές της κατηγορίες το σφάλμα γίνεται μικρότερο.

Στην παρακάτω εικόνα 5.1-3 φαίνονται ενδεικτικά οι προβλέψεις με και χωρίς εποχικότητα για την κατηγορία 49 με την μέθοδο SLR4.

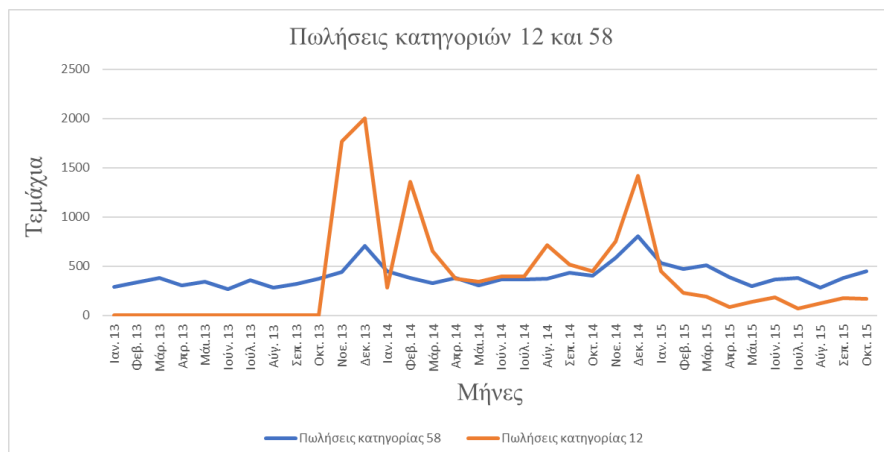


Εικόνα 5.1-3 Προβλέψεις (SLR4) με και χωρίς εποχικότητα για την κατηγορία 49

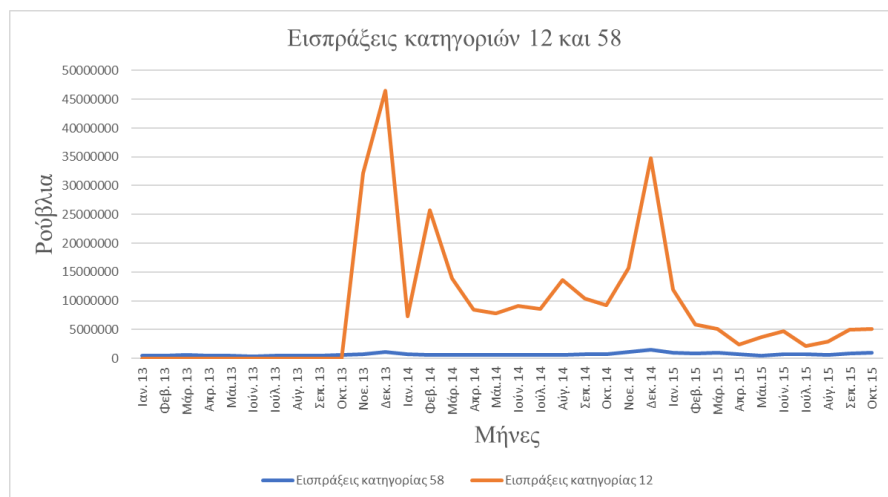
Το «κακό» βέβαια με την εποχικότητα είναι ότι βελτιώνει τις προβλέψεις μετά την πάροδο τουλάχιστον ενός έτους έτσι όσο περισσότερα δεδομένα προστίθενται τόσο βελτιώνονται οι προβλέψεις εξαιτίας της διόρθωσης από αυτήν. Μια άλλη πτυχή του θέματος «εποχικότητα» συνδέεται με την χρήση του συντελεστή διόρθωσης α . Δοκιμές που έγιναν με διάφορες τιμές του α έδειξαν ότι το μικρότερο MAPE εμφανίζεται όταν $\alpha = 0,333$ ή $\alpha = \frac{1}{3}$.

5.1.3 Συντελεστές βαρύτητας στον υπολογισμό του MAPE

Όσο για τους συντελεστές βαρύτητας ας συγκρίνουμε δυο κατηγορίες που έχουν περίπου τις ίδιες συνολικές πωλήσεις. Την κατηγορία 12 (Κονσόλες παιχνιδιών – PS4) με συνολικές πωλήσεις 13230 τεμάχια και την κατηγορία 58 (Μουσική - βινύλιο) με συνολικές πωλήσεις 13631 τεμάχια. Στις παρακάτω εικόνες φαίνονται οι πωλήσεις ανά μήνα για τις δυο αυτές κατηγορίες στην 5.1-4 και στην επόμενη 5.1-5 οι εισπράξεις ανά μήνα από τις δυο αυτές κατηγορίες.



Εικόνα 5.1-4 Σύγκριση πωλήσεων κατηγοριών 12 και 58



Εικόνα 5.1-5 Σύγκριση εισπράξεων κατηγοριών 12 και 58

Οι συνολικές εισπράξεις για την κατηγορία 12 είναι 292.238.749 ρούβλια (3.506.864 €) ενώ για την κατηγορία 58 είναι 23.216.716 ρούβλια (278.600 €). Με άλλα λόγια με τις ίδιες περίπου πωλήσεις οι εισπράξεις για την κατηγορία 12 είναι 12,6 φορές μεγαλύτερες. Μια λάθος πρόβλεψη, λοιπόν έστω και κατά ένα «τεμάχιο» στην κατηγορία 12 θα «στοίχιζε» 12,6 φορές περισσότερο από το ίδιο λάθος στην κατηγορία 58. Για τον λόγο αυτό είναι απαραίτητο να χρησιμοποιηθούν οι συντελεστές βαρύτητας στον υπολογισμό του MAPE και όχι μια απλή μέση τιμή του MAPE.

5.1.4 Κυκλικότητα

Για την κυκλικότητα θα μπορούσε ίσως να χρησιμοποιηθεί εξαιτίας του ότι τα προϊόντα από την στιγμή που πρωτοεμφανίζονται έχουν περιορισμένη διάρκεια ζωής. Η μέση διάρκεια ζωής είναι όπως είδαμε 16,85 μήνες. Όμως συνεχώς εμφανίζονται και νέα προϊόντα για τα οποία συνήθως δεν υπάρχουν επαρκή στοιχεία. Επομένως χρειάζονται πολύ περισσότερα δεδομένα για να ελέγξει κανείς την ύπαρξη ή όχι κυκλικότητας.

Πίνακας 5.1-3 Διάρκεια ζωής προϊόντων

item_id	item_category_id	First Sale Date	Last Sale Date	First Sale Number	Last Sale Number	Number of months
0		40 Sep 2014	Sep 2014	20	21	1
1		76 Apr 2014	Oct 2014	15	22	7
2		40 Aug 2014	Nov 2014	19	23	4
3		40 Jul 2014	Aug 2014	18	20	2
4		40 Sep 2014	Sep 2014	20	21	1
5		40 Dec 2014	Dec 2014	23	24	1
6		40 Jul 2014	Jul 2014	18	19	1
7		40 Dec 2014	Dec 2014	23	24	1
8		40 Aug 2014	Sep 2014	19	21	2
9		40 Aug 2014	Aug 2014	19	20	1
10		40 Jan 2015	Jan 2015	24	25	1
11		40 Nov 2014	Nov 2014	22	23	1
12		55 Feb 2013	Fed 2013	1	2	1
13		40 Sep 2014	Sep 2014	20	21	1
14		40 Dec 2014	Dec 2014	23	24	1
15		40 Nov 2014	Nov 2014	22	23	1
16		40 Sep 2014	Sep 2014	20	21	1
17		40 Sep 2014	Sep 2014	20	21	1
18		40 Aug 2014	Aug 2014	19	20	1
19		40 Jan 2013	Jan 2013	0	1	1
20		40 Aug 2014	Aug 2014	19	20	1
21		40 Sep 2014	Sep 2014	20	21	1
22		40 Apr 2015	Apr 2015	27	28	1
23		40 Fed 2015	Fed 2015	25	26	1
24		40 Apr 2015	Apr 2015	27	28	1
25		40 Aug 2014	Aug 2014	19	20	1

Στον πίνακα αυτό στην πρώτη στήλη Item_id είναι ο κωδικός του προϊόντος, στην 2^η Item_category_id είναι ο κωδικός της κατηγορίας στην οποία ανήκει το προϊόν. Στις επόμενες δυο στήλες είναι οι μήνες όπου εμφανίστηκαν οι πρώτες και οι τελευταίες πωλήσεις του Προϊόντος. Στις επόμενες δυο οι κωδικοί των αντίστοιχων μηνών (αύξοντας αριθμός μήνα) και στην τελευταία στήλη η διαφορά τους που είναι σε μήνες η διάρκεια ζωής του προϊόντος.

5.1.5 Περιορισμοί στις προβλέψεις

Στον παρακάτω πίνακα 5.1-3 φαίνεται το πλήθος των κατηγοριών για τις οποίες ΔΕΝ μπορούσαν να γίνουν προβλέψεις σε κάποιους μήνες.

Πίνακας 5.1-4 Για πόσες κατηγορίες δεν μπορεί να γίνει πρόβλεψη

Methods	LVFM	SLR2	SLR3	SLR4	QR4	QR5	QR6	QR7	QR8
Jan 2013	84	84	84	84	84	84	84	84	84
Feb 2013	28	84	84	84	84	84	84	84	84
Mar 2013	27	28	84	84	84	84	84	84	84
Apr 2013	27	27	28	84	84	84	84	84	84
May 2013	26	27	27	28	28	84	84	84	84
Jun 2013	20	26	27	27	27	28	84	84	84
Jul 2013	20	20	26	27	27	27	28	84	84
Aug 2013	20	20	20	26	26	27	27	28	84
Sep 2013	19	20	20	20	20	26	27	27	28
Oct 2013	18	19	20	20	20	20	26	27	27
Nov 2013	17	18	19	20	20	20	20	26	27
Dec 2013	15	17	18	19	19	20	20	20	26
Jan 2014	14	15	17	18	18	19	20	20	20
Feb 2014	14	14	15	17	17	18	19	20	20
Mar 2014	14	14	14	15	15	17	18	19	20
Apr 2014	14	14	14	14	14	15	17	18	19
May 2014	14	14	14	14	14	14	15	17	18
Jun 2014	14	14	14	14	14	14	14	15	17
Jul 2014	14	14	14	14	14	14	14	14	15
Aug 2014	10	14	14	14	14	14	14	14	14
Sep 2014	10	10	14	14	14	14	14	14	14
Oct 2014	7	10	10	14	14	14	14	14	14
Nov 2014	7	7	10	10	10	14	14	14	14
Dec 2014	7	7	7	10	10	10	14	14	14
Jan 2015	5	7	7	7	7	10	10	14	14
Feb 2015	5	5	7	7	7	7	10	10	14
Mar 2015	5	5	5	7	7	7	7	10	10
Apr 2015	3	5	5	5	5	7	7	7	10
May 2015	2	3	5	5	5	5	7	7	7
Jun 2015	2	2	3	5	5	5	5	7	7
Jul 2015	2	2	2	3	3	5	5	5	7
Aug 2015	2	2	2	2	2	3	5	5	5
Sep 2015	1	2	2	2	2	2	3	5	5
Oct 2015	1	1	2	2	2	2	2	3	5

Οι λόγοι που γίνεται αυτό είναι:

- (1) Οι διάφορες μέθοδοι ανάλογα με το πλήθος των δεδομένων που χρησιμοποιούν ξεκινούν τις προβλέψεις σε διαφορετικούς μήνες. Για παράδειγμα η SLR4 ξεκινά 4 μήνες μετά από την εμφάνιση των πρώτων δεδομένων. Δηλαδή αν οι πρώτες πωλήσεις ενός προϊόντος ή μιας κατηγορίας εμφανίζονται τον 6^ο μήνα (Ιούνιος 2013), η SLR4 θα αρχίσει να προβλέπει 4 μήνες μετά, δηλαδή τον 10^ο μήνα (Οκτώβριος 2013). Στους 9 πρώτους μήνες (από τον 1^ο μέχρι τον 9^ο (0 έως 8)) θα εμφανίζονται «-» στα αρχεία της αντίστοιχης μεθόδου στην αντίστοιχη κατηγορία.
- (2) Η LVFM είναι η μόνη που ξεκινά ένα μήνα μετά την εμφάνιση των πρώτων δεδομένων, άρα περιλαμβάνει τις περισσότερες προβλέψεις.
- (3) Για την κατηγορία 80 δεν γίνονται προβλέψεις για τους λόγους που προαναφέρθηκαν.

(4) Αν ένα προϊόν ή μια κατηγορία ξεκινά τις πωλήσεις του π.χ. τρεις μήνες πριν τον Οκτώβριο του 2015 όπως συμβαίνει με την κατηγορία 36 η LVFM θα δώσει προβλέψεις για τον Σεπτέμβριο και τον Οκτώβριο του 2015, έτσι στον πίνακα φαίνεται η LVFM να μην κάνει προβλέψεις μόνο για μία κατηγορία (πρόκειται για την 80). Το ίδιο και η SLR2 που κάνει πρόβλεψη για την κατηγορία 36 μόνο για τον Οκτώβριο του 2015 (δεν κάνει φυσικά για την 80). Η SLR4 δεν κάνει πρόβλεψη για την κατηγορία 36 και φυσικά την 80.

5.1.6 MAPE για τις κατηγορίες

5.1.6.1 Αναλυτικά MAPE για κάθε μήνα και μέθοδο

Το σημαντικότερο κριτήριο για να συγκρίνει κανείς τις διάφορες μεθόδους πρόβλεψης μεταξύ τους είναι το MAPE. Εδώ θα πρέπει να πούμε και πάλι ότι αναλυτικά MAPE για τις κατηγορίες, για κάθε μήνα και κάθε μέθοδο πρόβλεψης υπολογίζουν 6 κώδικες που δημιουργούν τα αρχεία:

(1) Μέσο MAPE χωρίς βαρύτητα χωρίς εποχικότητα (MapeAllcat_0.csv).

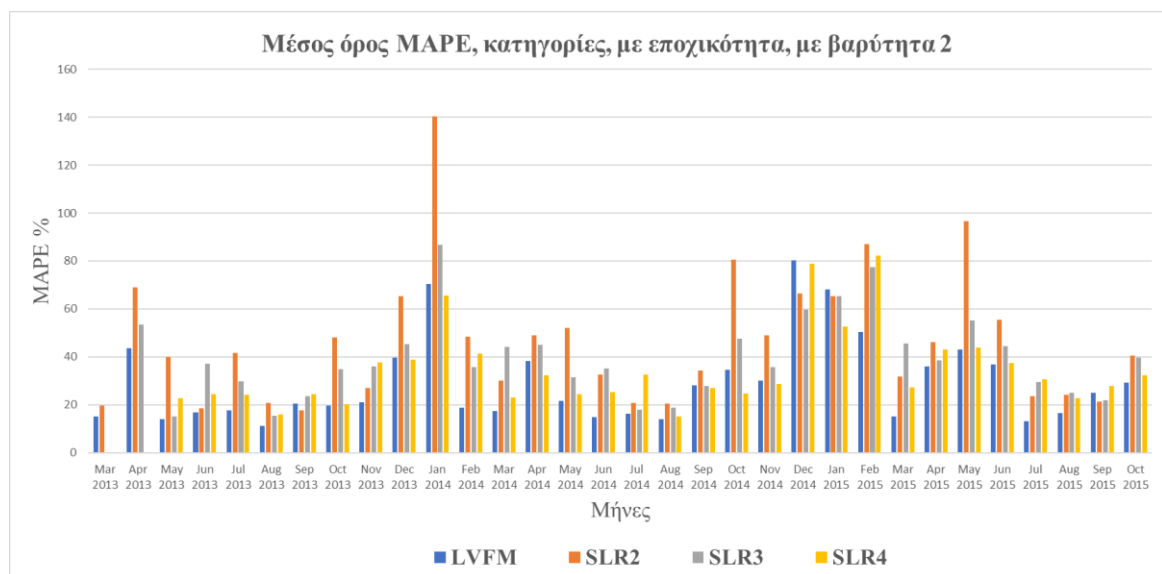
Πίνακας 5.1-5 MAPE % για τις κατηγορίες με βαρύτητα 2 και εποχικότητα

Με βαρύτητα 2 - με εποχικότητα									
Μήνες	LVFM	SLR2	SLR3	SLR4	QR4	QR5	QR6	QR7	QR8
Jan 2013	-	-	-	-	-	-	-	-	-
Feb 2013	12,61496	-	-	-	-	-	-	-	-
Mar 2013	15,11544	19,62058	-	-	-	-	-	-	-
Apr 2013	43,69532	69,06725	53,48441	-	-	-	-	-	-
May 2013	14,11561	39,96311	15,02451	22,66644	38,76635	-	-	-	-
Jun 2013	16,74495	18,58396	37,26434	24,5027	41,89571	39,62481	-	-	-
Jul 2013	17,72212	41,7604	29,86756	24,18259	76,88186	40,71563	34,38074	-	-
Aug 2013	11,19339	20,67	15,40887	15,92087	22,76286	23,85656	21,19927	27,25519	-
Sep 2013	20,49337	17,60954	23,56369	24,39968	26,19054	27,79164	32,13141	24,39882	29,34422
Oct 2013	19,56104	48,22483	34,74879	20,09144	74,4984	47,84461	35,2087	39,26108	28,17319
Nov 2013	20,98347	26,89529	35,90242	37,82199	33,88803	45,07647	42,86834	40,80841	45,17722
Dec 2013	39,69863	65,26775	45,2479	38,89518	65,10508	54,93061	47,21747	47,46283	44,17689
Jan 2014	70,44309	140,3897	86,82739	65,57505	148,2076	111,9506	91,95105	84,56024	70,58115
Feb 2014	18,79434	48,47571	35,74277	41,22933	39,67629	26,26007	36,76263	43,22009	42,35924
Mar 2014	17,4045	30,1008	44,27202	23,02186	73,01501	52,99756	39,10026	26,6288	26,19873
Apr 2014	38,22175	48,99745	45,16144	32,34482	95,3531	47,52021	31,74034	33,82544	21,98994
May 2014	21,6979	52,07541	31,51957	24,46004	42,92131	43,2775	39,45581	37,216	32,88158
Jun 2014	14,95882	32,71331	35,30021	25,41226	51,14203	45,92386	41,85837	57,63761	40,43428
Jul 2014	16,2312	20,84303	18,09491	32,75704	22,87517	31,23914	33,58097	47,48678	58,44095
Aug 2014	14,08641	20,41503	18,77159	15,26944	25,89216	20,04411	20,39356	19,42389	26,8638
Sep 2014	28,22179	34,35144	27,73097	27,06642	40,02916	32,60142	37,36317	35,25215	35,0989
Oct 2014	34,56741	80,42752	47,53762	24,74776	58,72393	42,71026	30,05012	21,92559	25,51662
Nov 2014	29,94743	48,83439	35,78535	28,5782	53,68901	39,56674	41,45905	35,87549	33,52577
Dec 2014	80,26051	66,38281	59,78287	78,87022	27,88896	38,26321	31,98426	32,65351	33,15389
Jan 2015	68,09544	65,3511	65,25062	52,54864	1682,204	1605,561	1634,456	47,99903	41,62925
Feb 2015	50,32342	87,15228	77,43062	82,17519	1264,049	1218,051	1134,254	1149,637	76,32999
Mar 2015	15,00351	31,80009	45,60489	27,40527	663,269	708,2166	777,7218	811,5985	818,6334
Apr 2015	35,8932	46,2038	38,68292	43,10313	70,55751	85,9174	71,17312	72,44454	50,64166
May 2015	42,92443	96,62941	55,04353	43,9664	111,3629	75,35814	75,46353	82,22406	70,44643
Jun 2015	36,72689	55,47162	44,56968	37,31127	67,01673	66,8283	58,67669	61,27198	60,31065
Jul 2015	13,08971	23,64471	29,56443	30,66051	46,71806	39,0405	43,3517	41,05148	44,50524
Aug 2015	16,49106	24,04238	25,00047	22,61698	67,17369	53,53181	35,74164	31,30127	42,04147
Sep 2015	25,01369	21,29972	21,97731	27,79809	64,37124	67,07828	53,77269	40,02877	40,4002
Oct 2015	29,19985	40,40789	39,66629	32,39853	66,7338	76,87339	78,64726	57,01259	45,87376
Average	28,77378	46,36476	39,34935	34,25991	172,0953	165,8156	166,1416	112,943	72,48955

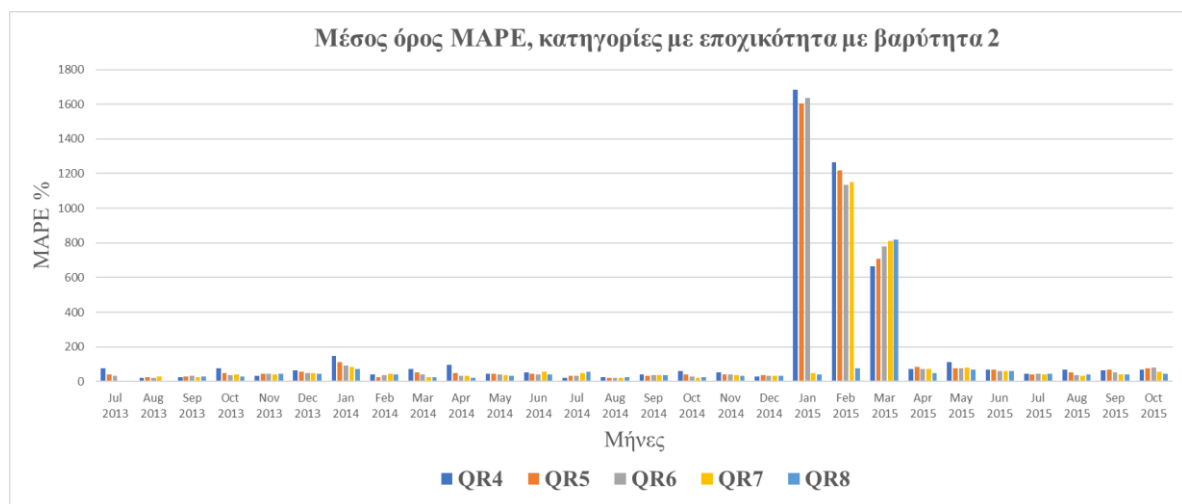
- (2) Μέσο MAPE χωρίς βαρύτητα με εποχικότητα (**MapeAllcatSessions_0.csv**)
- (3) Μέσο MAPE με βαρύτητα 1 χωρίς εποχικότητα (**MapeAllcat_0W.csv**)
- (4) Μέσο MAPE με βαρύτητα 1 και εποχικότητα (**MapeAllcatSessions_0W.csv**)
- (5) Μέσο MAPE με βαρύτητα 2 χωρίς εποχικότητα (**MapeAllcat_1W.csv**)
- (6) Μέσο MAPE με βαρύτητα 2 και εποχικότητα (**MapeAllcatSessions_1W.csv**)

Στον παρακάτω πίνακα 5.1-4 δίνεται ενδεικτικά μια από τις 6 περιπτώσεις (Με βαρύτητα 2 και εποχικότητα). Στην τελευταία γραμμή εμφανίζονται οι απλοί (μη σταθμισμένοι) μέσοι όροι όλων των μηνών ανά μέθοδο.

Στις παρακάτω εικόνες (5.1-6 και 5.1-7) εμφανίζονται τα MAPE του παραπάνω πίνακα (5.1-4) για την LVFM μαζί με τις SLR (5.1-6) και ξεχωριστά οι QR (5.1-7).



Εικόνα 5.1-6 Μέσο MAPE με εποχικότητα και βαρύτητα 2 (LVFM και SLR 2,3,4)



Εικόνα 5.1-7 Μέσο MAPE με εποχικότητα και βαρύτητα 2 (QR 4,5,6,7,8)

Είχαμε αναφερθεί στο πρόβλημα που υπάρχει όταν σε μια πρόβλεψη η πραγματική πώληση είναι μηδενική ενώ γίνεται πρόβλεψη για αυτήν. Έτσι αντί να παίρνει κανείς μέσο

όρο για όλα τα προϊόντα ή όλες τις κατηγορίες να χρησιμοποιηθούν συντελεστές βαρύτητας.

Δυο τέτοιοι μπορούν να χρησιμοποιηθούν. Ο ένας χρησιμοποιεί τις μέσες τιμές πώλησης (Βαρύτητα 1) ενός προϊόντος και ο άλλος τον μέγιστο ανάμεσα στην πρόβλεψη και την πραγματική πώληση (Βαρύτητα 2) (Βλέπε ενότητα 4.6.2 και για τις δυο βαρύτητες). Στις κατηγορίες δεν υπάρχει πρόβλημα καθώς οι δυο βαρύτητες δίνουν παραπλήσιες τιμές, όπως φαίνεται και στον πίνακα 5.1-4. Και οι δύο είναι περισσότερο «δίκαιοι» από τον απλό μέσο όρο (βλέπε ενότητα 5.1.3).

Από τις εικόνες 5.1-6 και 5.1-7 είναι φανερό ότι τα μεγαλύτερα σφάλματα ανεξάρτητα από την χρησιμοποιούμενη μέθοδο εμφανίζονται τον Ιανουάριο όπου σε όλες τις μεθόδους η πρόβλεψη για τον Ιανουάριο επηρεάζεται πιο πολύ από τις αυξημένες πωλήσεις του Δεκεμβρίου. Με την χρήση όμως της βαρύτητας 2, εμφανίζεται και ένα ακόμα πρόβλημα ειδικά στις QR.

Επειδή σ' αυτές και κατά τους μήνες Ιανουάριο Φεβρουάριο και Μάρτιο ειδικά του τελευταίου έτους (εδώ 2015) εμφανίζονται τα μεγαλύτερα σφάλματα στην πρόβλεψη των πωλήσεων, το σφάλμα αυτό μεγεθύνεται με την βαρύτητα 2, καθώς ο συντελεστής στον αριθμητή περιέχει το μέγιστο ανάμεσα στις προβλεπόμενες και τις πραγματικές πωλήσεις. Όμως και αυτό έχει την αξία του καθώς μπορεί κανείς να εστιάσει πιο εύκολα στους προβληματικούς μήνες.

5.1.6.2 Απλοί μέσοι όροι MAPE για κάθε μέθοδο

Οι απλοί μέσοι όροι MAPE όλων των μηνών, για κάθε μέθοδο ξεχωριστά βρίσκονται φυσικά σε 6 αρχεία. Αυτά είναι τα εξής:

- (1) Απλοί μέσοι όροι των μηνών, για κάθε μέθοδο ξεχωριστά, χωρίς βαρύτητα χωρίς εποχικότητα. (**MethodsCatAverageW_de.csv**).
- (2) Απλοί μέσοι όροι των μηνών, για κάθε μέθοδο ξεχωριστά, χωρίς βαρύτητα με εποχικότητα. (**MethodsCatAverageSessonsWDe_0.csv**).
- (3) Απλοί μέσοι όροι των μηνών, για κάθε μέθοδο ξεχωριστά, με βαρύτητα 1 χωρίς εποχικότητα. (**MethodsCatAverageWDe_W0.csv**).
- (4) Απλοί μέσοι όροι των μηνών, για κάθε μέθοδο ξεχωριστά, με βαρύτητα 1 με εποχικότητα. (**MethodsCatAverageWDe_W0.csv**).
- (5) Απλοί μέσοι όροι των μηνών, για κάθε μέθοδο ξεχωριστά, με βαρύτητα 2 χωρίς εποχικότητα. (**MethodsCatAverageWDe_W1.csv**).
- (6) Απλοί μέσοι όροι των μηνών, για κάθε μέθοδο ξεχωριστά, με βαρύτητα 2 με εποχικότητα. **MethodsCatAverageSessonsWDeW_1.csv**

Στον πίνακα 5.1-5 εμφανίζονται οι απλοί μέσοι όροι για όλους τους μήνες σε κάθε μέθοδο και των 6 περιπτώσεων.

Πίνακας 5.1-6 Απλοί μέσοι όροι MAPE % για τις κατηγορίες

Μέθοδοι	Χωρίς βαρύτητα		Με βαρύτητα 1		Με βαρύτητα 2	
	Χωρίς εποχικότητα	Με εποχικότητα	Χωρίς εποχικότητα	Με εποχικότητα	Χωρίς εποχικότητα	Με εποχικότητα
LVFM	47,18	47,18	32,55	32,55	28,77	28,77
SLR2	60,62	60,08	52,00	48,36	50,51	46,36
SLR3	55,33	55,03	46,49	44,28	42,01	39,35
SLR4	54,80	54,54	39,43	37,71	36,26	34,26
QR4	242,6	241,7	63,72	60,34	159,2	172,9

QR5	250,1	249,7	55,73	53,03	152,9	165,8
QR6	254,6	248,3	51,90	49,64	153,9	166,14
QR7	187,4	179,9	52,89	49,58	115,2	112,94
QR8	137,5	129,5	49,72	46,56	78,22	72,49

Από τον πίνακα αυτό γίνεται φανερό ότι τα μικρότερα MAPE % και σε όλους τους τρόπους υπολογισμού το παρουσιάζουν η LVFM και η SLR4.

5.1.6.3 Σταθμισμένο μέσο MAPE για κάθε μέθοδο

Οι σταθμισμένοι μέσοι όροι MAPE όλων των μηνών, για κάθε μέθοδο ξεχωριστά βρίσκονται φυσικά σε άλλα 6 αρχεία. Αυτά είναι τα εξής:

- (1) Μέσο MAPE χωρίς βαρύτητα χωρίς εποχικότητα
(MethodsAverageCat_De_0.csv.csv).
- (2) Μέσο MAPE χωρίς βαρύτητα με εποχικότητα
(MethodsAverageCatSessions_De_0.csv).
- (3) Μέσο MAPE με βαρύτητα 1 χωρίς εποχικότητα
(MethodsAverageCat_De_W_0.csv).
- (4) Μέσο MAPE με βαρύτητα 1 με εποχικότητα
(MethodsAverageCatSessions_De_W_0.csv).
- (5) Μέσο MAPE με βαρύτητα 2 χωρίς εποχικότητα
(MethodsAverageCat_De_W_1.csv).
- (6) Μέσο MAPE με βαρύτητα 2 με εποχικότητα
(MethodsAverageCatSessions_De_W_1.csv).

Στον πίνακα 5.1-6 παρουσιάζονται οι σταθμισμένοι μέσοι όροι για τις 6 προηγούμενες περιπτώσεις. Και από αυτόν τον πίνακα γίνεται και πάλι φανερό ότι τα μικρότερα MAPE % από όλους τους τρόπους υπολογισμού το παρουσιάζουν η LVFM και η SLR4.

Πίνακας 5.1-7 Σταθμισμένοι μέσοι όροι MAPE % για τις κατηγορίες

Μέθοδοι	Χωρίς βαρύτητα		Με βαρύτητα 1		Με βαρύτητα 2	
	Χωρίς εποχικότητα	Με εποχικότητα	Χωρίς εποχικότητα	Με εποχικότητα	Χωρίς εποχικότητα	Με εποχικότητα
LVFM	60,52	60,52	33,72	33,72	31,03	31,03
SLR2	66,91	67,52	51,05	47,12	49,90	45,76
SLR3	61,03	62,19	44,06	41,92	39,83	37,57
SLR4	62,14	63,28	36,62	35,01	34,45	32,92
QR4	216,7	215,7	56,71	53,26	139,7	153,2
QR5	220,9	220,1	49,27	46,58	130,7	143,4
QR6	215,3	210,6	44,24	42,25	127,5	139,2
QR7	141,8	137,2	43,33	40,73	84,99	83,11
QR8	104,24	99,67	39,64	37,23	57,34	53,24

Σε αυτόν τον πίνακα 5.1-8 προκύπτει ένα λίγο διαφορετικό αποτέλεσμα. Συγκεκριμένα στο υπολογισμό του σταθμισμένου μέσου όρου των μηνών και στην περίπτωση που δεν χρησιμοποιείται συντελεστής βαρύτητας, δηλαδή βγαίνει ένας απλός μέσος όρος για τους μήνες, φαίνεται η SLR3 να έχει μικρότερο μέσο MAPE από την SLR4. Όταν όμως «μπαίνουν στο παιχνίδι» οι συντελεστές βαρύτητας και πάλι η SLR4 έχει μικρότερο MAPE από την SLR3. Όταν χρησιμοποιούνται λοιπόν συντελεστές βαρύ-

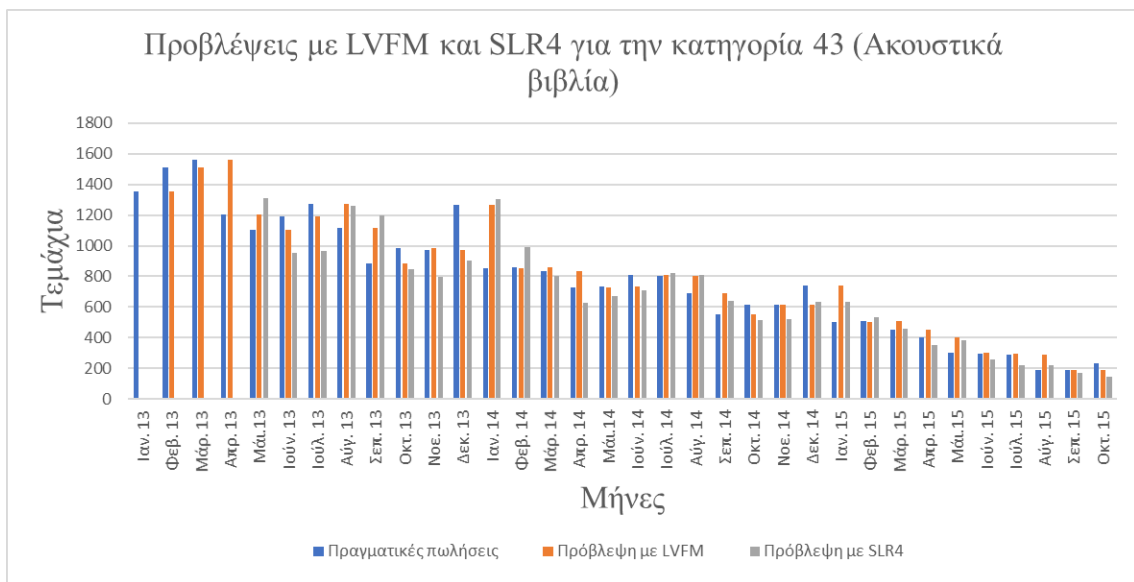
τητας μεταξύ των κατηγοριών καταλήγουμε και πάλι στο ίδιο συμπέρασμα, δηλαδή ότι τα μικρότερα MAPE % από όλους τους τρόπους υπολογισμού τα παρουσιάζουν η LVFM και η SLR4.

Όπως αναλύθηκε όμως στην ενότητα 5.1.3 είναι απαραίτητο να χρησιμοποιηθούν οι συντελεστές βαρύτητας στον υπολογισμό του MAPE και όχι μια απλή μέση τιμή του.

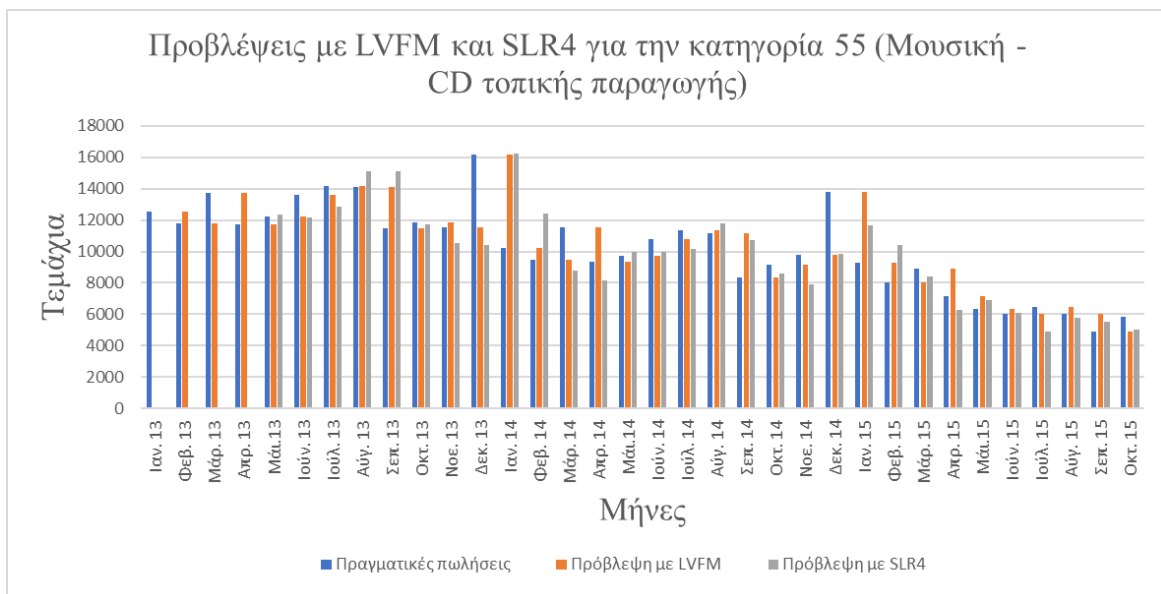
5.1.6.4 Προβλέψεις για τις κατηγορίες με LVFM και SLR4

Τελικά οι μέθοδοι που παρουσιάζουν τα μικρότερα συνολικά MAPE στις κατηγορίες είναι η LVFM και η SLR4 με εποχικότητα με συντελεστή $\alpha = 0,333$. Στις παρακάτω εικόνες παρουσιάζονται ενδεικτικά μερικά αποτελέσματα, με αυτές τις μεθόδους, για διάφορες κατηγορίες.

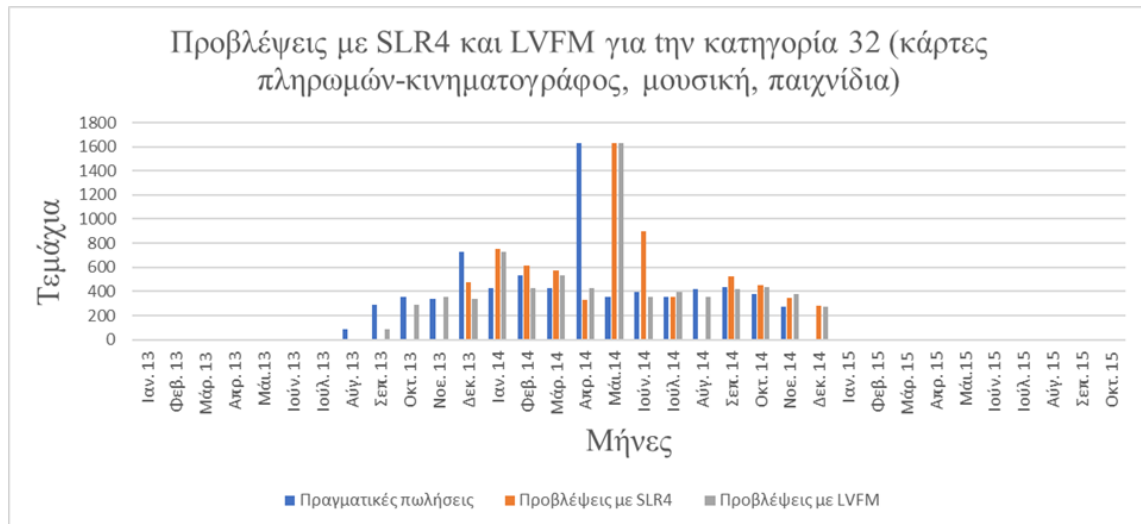
Αυτές οι δυο μέθοδοι θα χρησιμοποιηθούν και για τα προϊόντα για να επιλεγεί τελικά η μέθοδος που μπορεί να εφαρμοστεί.



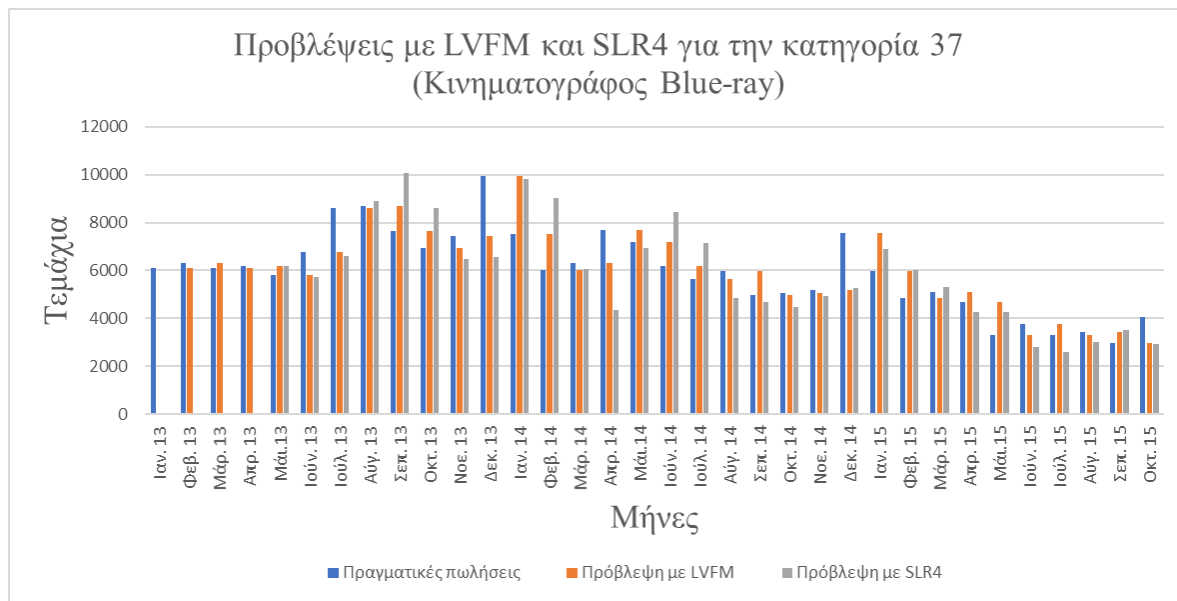
Εικόνα 5.1-8 Προβλέψεις με LVFM και SLR4 για κατηγορία 43



Εικόνα 5.1-9 Προβλέψεις με LVFM και SLR4 για κατηγορία 55



Εικόνα 5.1-10 Προβλέψεις με LVFM και SLR4 για κατηγορία 32



Εικόνα 5.1-11 Προβλέψεις με LVFM και SLR4 για κατηγορία 37

5.1.7 MAPE για τα προϊόντα

5.1.7.1 Αναλυτικά MAPE για κάθε μήνα και μέθοδο

Για τα προϊόντα και για κάθε μήνα και τις δυο μεθόδους πρόβλεψης, συγκεκριμένα για SLR4 και LVFM, το MAPE υπολογίζουν άλλοι 3 κώδικες που δημιουργούν τα αρχεία:

- (1) Μέσο MAPE χωρίς βαρύτητα με εποχικότητα (**MapeSLR4Items.csv**)
- (2) Μέσο MAPE με βαρύτητα 1 με εποχικότητα (**MapeSLR4ItemsW.csv**)
- (3) Μέσο MAPE με βαρύτητα 2 με εποχικότητα (**MapeSLR4ItemsW_2.csv**).

Στον παρακάτω πίνακα 5.1-9 δίνονται και οι τρεις περιπτώσεις για τις δύο μεθόδους. Στην προτελευταία γραμμή εμφανίζονται οι απλοί (μη σταθμισμένοι) μέσοι όροι όλων των μηνών ανά μέθοδο. Τέλος στην τελευταία γραμμή εμφανίζονται οι σταθμισμένοι μέσοι όροι.

Πίνακας 5.1-8 MAPE για τα προϊόντα

	Χωρίς βαρύτητα		Με βαρύτητα 1		Με βαρύτητα 2	
	LVFM	SLR4	LVFM	SLR4	LVFM	SLR4
Jan 2013	-	-	-	-	-	-
Feb 2013	78,42344	-	86,02407	-	81,97789	-
Mar 2013	59,01665	-	64,28703	-	132,4381	-
Apr 2013	79,61527	-	92,75421	-	233,5843	-
May 2013	57,90321	64,64081	58,10422	53,86929	93,68874	70,48231
Jun 2013	54,2992	59,16716	55,03611	52,14372	85,46834	62,30728
Jul 2013	54,07862	58,52725	58,04002	55,4294	152,1524	94,43668
Aug 2013	55,1271	61,81129	53,57243	54,94293	79,33047	60,62757
Sep 2013	58,89601	69,14072	54,44936	53,03587	89,15471	76,81592
Oct 2013	50,79969	56,82851	59,44004	53,76585	206,2503	130,61
Nov 2013	52,03912	51,09052	66,08179	52,92756	204,949	121,422
Dec 2013	45,53376	48,85462	48,93719	41,69726	144,8702	66,45507
Jan 2014	62,52806	62,48134	83,29606	65,74334	277,242	135,1935
Feb 2014	47,46866	59,91769	50,7221	61,95507	115,5134	198,01
Mar 2014	37,9264	45,62559	39,81863	43,71656	99,77008	106,1123
Apr 2014	46,92381	42,6233	54,64141	36,92222	245,3596	92,73671
May 2014	36,14327	33,08558	36,78361	32,55564	204,3637	171,0637
Jun 2014	34,69779	34,11913	32,82656	32,86317	89,39827	112,7915
Jul 2014	35,78444	35,84379	34,77552	32,17609	93,8078	77,65031
Aug 2014	32,18764	32,26839	29,95507	29,88954	58,22812	59,29215
Sep 2014	35,9034	31,33505	35,84942	31,39484	106,2938	82,12561
Oct 2014	29,09731	27,51825	34,32918	28,94972	215,1928	77,55256
Nov 2014	49,65225	26,47224	47,71557	30,95947	508,0684	87,30911
Dec 2014	30,54055	26,76592	34,50386	28,66028	116,9135	70,31761
Jan 2015	39,12697	30,8707	75,45486	48,08968	282,3035	179,7794
Feb 2015	36,95312	32,40217	56,63783	64,85201	436,0556	619,8925
Mar 2015	27,3525	28,05349	32,22601	33,21349	95,39852	146,3105
Apr 2015	35,34947	30,0199	40,3185	30,92383	162,7457	128,0494
May 2015	28,16262	25,08979	37,37916	27,61241	1613,863	97,75961
Jun 2015	23,38836	21,63415	28,30234	24,02859	294,4345	91,45411
Jul 2015	22,99887	19,53967	31,03373	24,77192	240,3501	161,5879
Aug 2015	22,88252	21,27899	27,63078	26,57902	111,4432	97,82477
Sep 2015	31,53001	30,92643	38,23499	33,75111	203,107	176,5497
Oct 2015	45,40393	21,13748	30,99507	28,24617	430,0471	122,6934
Απλός μέσος όρος	43,5677	39,63566	48,79263	40,5222	227,3868	125,8404
Σταθμισμένος μέσος όρος	44,18347	36,85027	52,21861	38,6326	231,4156	120,5732

Οι απλοί μέσοι όροι MAPE βρίσκονται στα αρχεία:

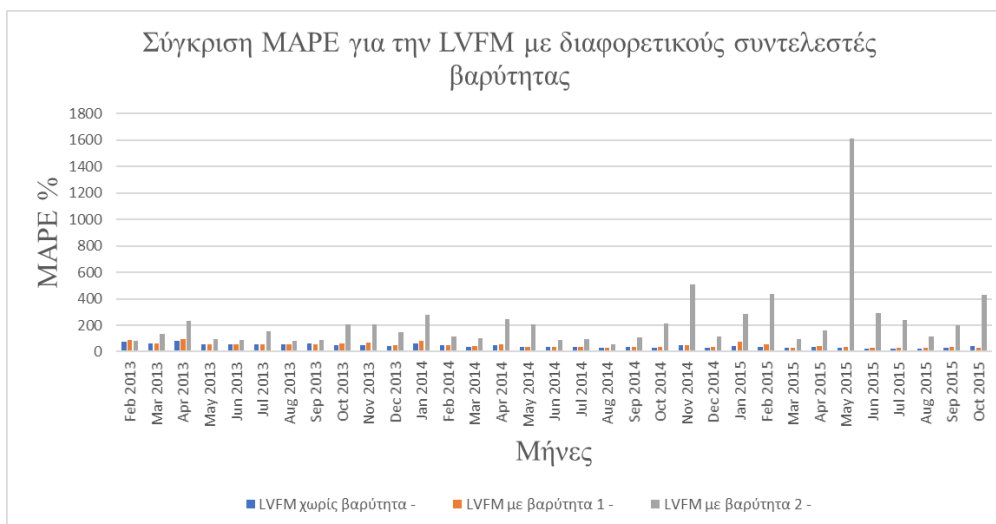
- (1) Μέσο MAPE χωρίς βαρύτητα και με εποχικότητα
(MethodsItemsAverageSessons_0.csv)
- (2) Μέσο MAPE με βαρύτητα 1 και με εποχικότητα
(MethodsItemsAverageSessonsW_1.csv)
- (3) Μέσο MAPE με βαρύτητα 2 και με εποχικότητα
(MethodsItemsAverageSessonsW_2.csv)

Οι σταθμισμένοι μέσοι όροι MAPE βρίσκονται στα αρχεία:

- (1) Σταθμισμένο μέσο MAPE χωρίς βαρύτητα για τα προϊόντα και με εποχικότητα
(MethodsItemsAverageSessons_De_0.csv).
- (2) Σταθμισμένο μέσο MAPE με βαρύτητα 1 για τα προϊόντα και με εποχικότητα
(MethodsAverageSessonsW_De_1.csv)
- (3) Σταθμισμένο μέσο MAPE με βαρύτητα 2 για τα προϊόντα με εποχικότητα
(MethodsAverageSessonsW_De_2.csv)

Εδώ οι σταθμισμένοι μέσοι όροι ανέδειξαν κάποιο άλλο πρόβλημα. Συγκεκριμένα οι σταθμισμένοι μέσοι όροι με βαρύτητα 2 εμφανίζονται και οι δυο αυξημένοι. Περισσότερο βέβαια για την LVFM. Αυτό βέβαια χρειάστηκε να ερευνηθεί.

Στην εικόνα 5.1-11 έχουμε την γραφική παράσταση ενός μέρους του πίνακα 5.1-9. Συγκεκριμένα ο μέσος όρος με βαρύτητα 2 εμφανίζεται ξαφνικά τεράστιος (1613,853%). Η τεράστια αυτή αύξηση στο MAPE, δεν εμφανίζεται αλλού.

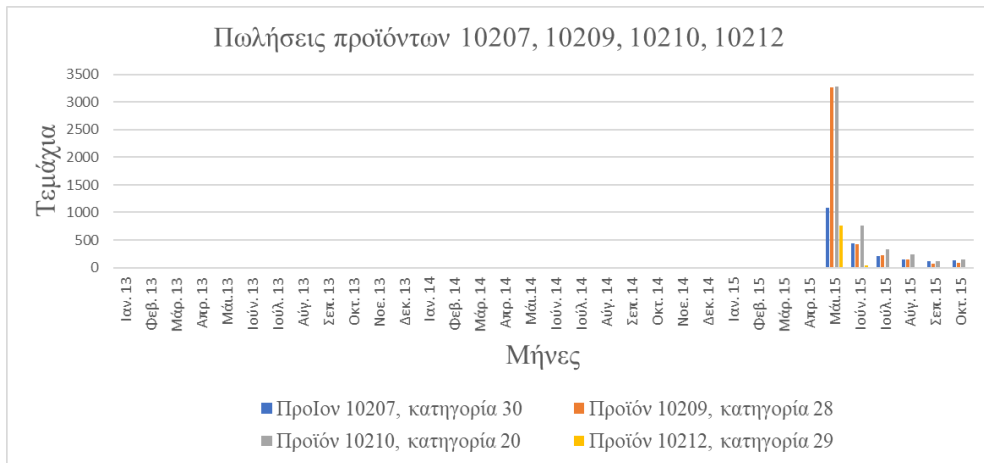


Εικόνα 5.1-12 Σύγκριση MAPE για LVFM με διαφορετικούς συντελεστές βαρύτητας

Η έρευνα «έδειξε» αρχικά το προϊόν με κωδικό 10210 (πρόκειται για τη Ρωσική έκδοση ενός δημοφιλούς παιχνιδιού για το PS4 «The Witcher 3: wild hunt»). Πιο συγκεκριμένα τον μήνα Μάιο του 2015 (από μηδενικές πωλήσεις εκτινάχθηκε στις 3275) κυκλοφόρησε για πρώτη φορά σε εκδόσεις για πολλές διαφορετικές κονσόλες όπου τελικά εμφανίζεται σαν διαφορετικά προϊόντα (από κωδικό 10205 έως 10215 που ανήκουν σε διαφορετικές κατηγορίες αλλά όλα παρουσίασαν αυτήν την συμπεριφορά (απότομη αύξηση στις πωλήσεις). Φυσικά αυτό παρασέρνει και αρκετά άλλα προϊόντα που συνδέονται με τις κονσόλες και τα αξεσουάρ τους (όλα σε διαφορετικές κατηγορίες). Έτσι η αύξηση στο MAPE δεν μπορούσε να εμφανιστεί στις κατηγορίες. Η SLR4 άρχισε να προβλέπει με 4 μήνες καθυστέρηση και όχι άμεσα όπως η LVFM. Όλα αυτά βέβαια σε συνδυασμό με το γεγονός ότι η βαρύτητα 2 μεγεθύνει τα σφάλματα που αφορούν λάθος πρόβλεψη σε «τεμάχια» οδήγησαν στο πρόβλημα (ίσως πάλι στον εντοπισμό του προβλήματος).

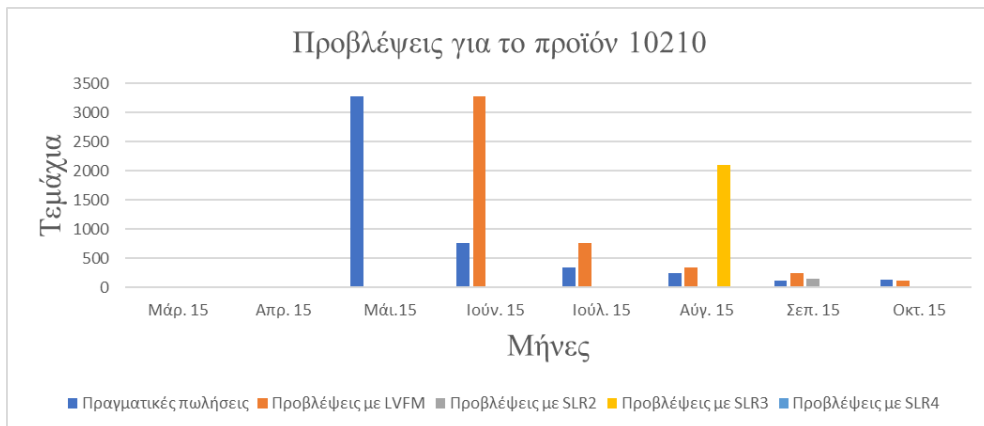
Στην εικόνα 5.1-12 δίνονται ενδεικτικά οι πωλήσεις 4 προϊόντων που αναφέρο-

ρονται στο ίδιο παιχνίδι αλλά για διαφορετικές κονσόλες.



Εικόνα 5.1-13 Πωλήσεις προϊόντων 10207, 10209, 10210, 10212

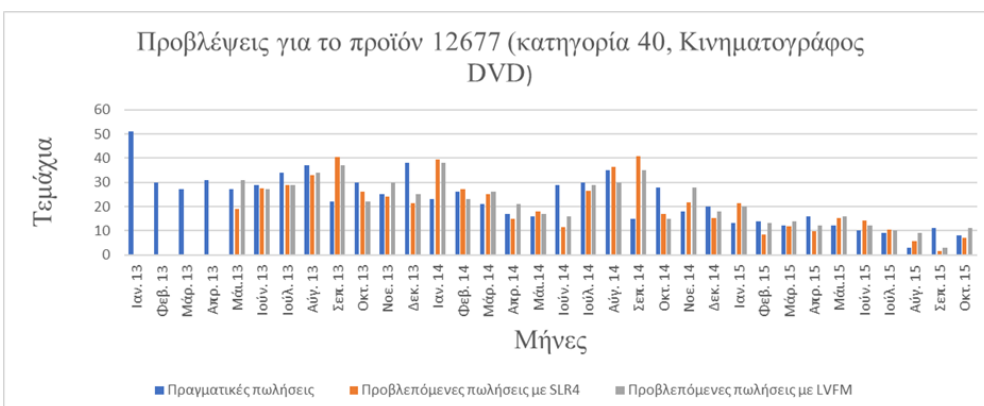
Στην εικόνα 5.1-13 δίνονται οι προβλέψεις για το προϊόν 10210 με τις LVFM, SLR2, SLR3 και SLR4 (η τελευταία δίνει μηδενικές προβλέψεις και αυτές μόνο για τους δυο τελευταίους μήνες).



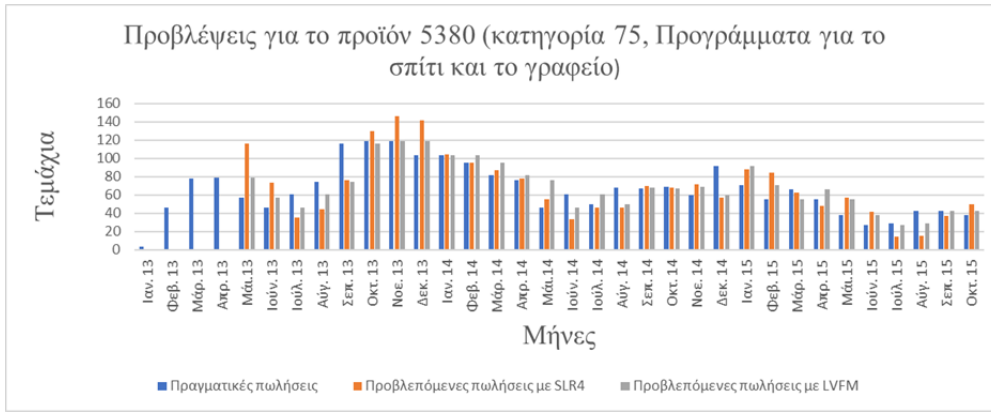
Εικόνα 5.1-14 Προβλέψεις για το προϊόν 10210 με LVFM και SLR2,3,4

5.1.7.2 Προβλέψεις για τα προϊόντα με SLR4

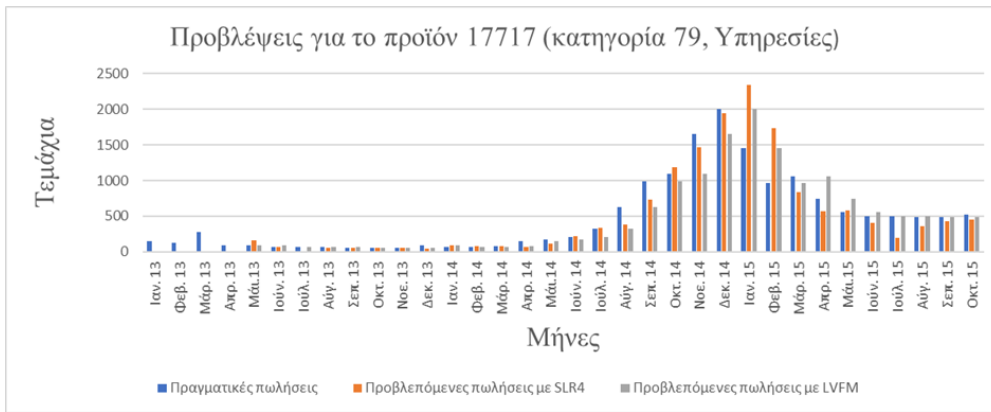
Τελικά η καλύτερη μέθοδος που μπορεί να χρησιμοποιηθεί για τις προβλέψεις είναι η SLR4. Στις παρακάτω εικόνες και για λόγους σύγκρισης φαίνονται οι προβλέψεις για κάποια προϊόντα τόσο με την SLR4 όσο και με την LVFM.



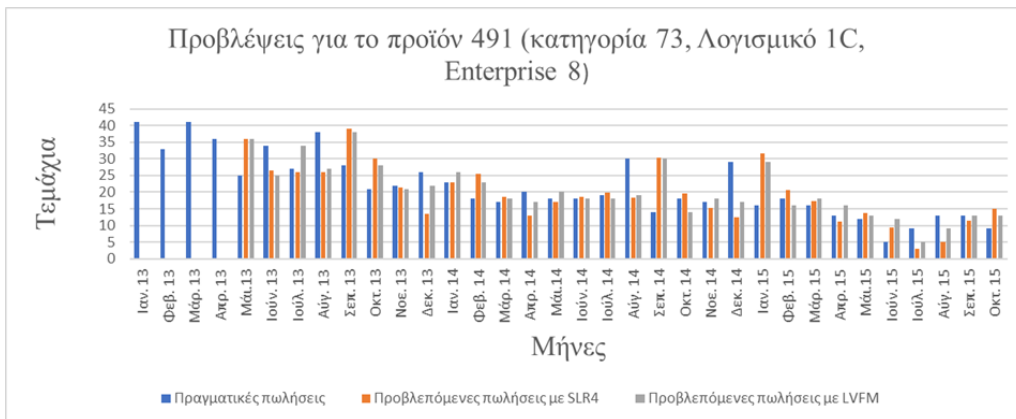
Εικόνα 5.1-15 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 12677



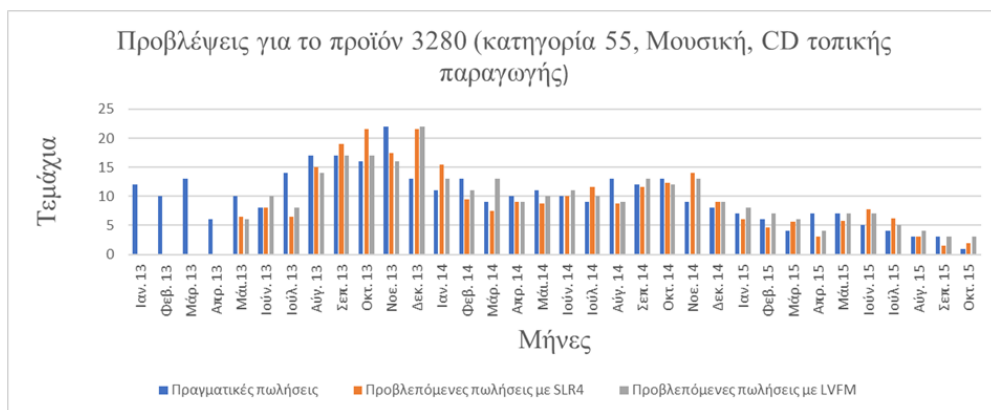
Εικόνα 5.1-16 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 5380



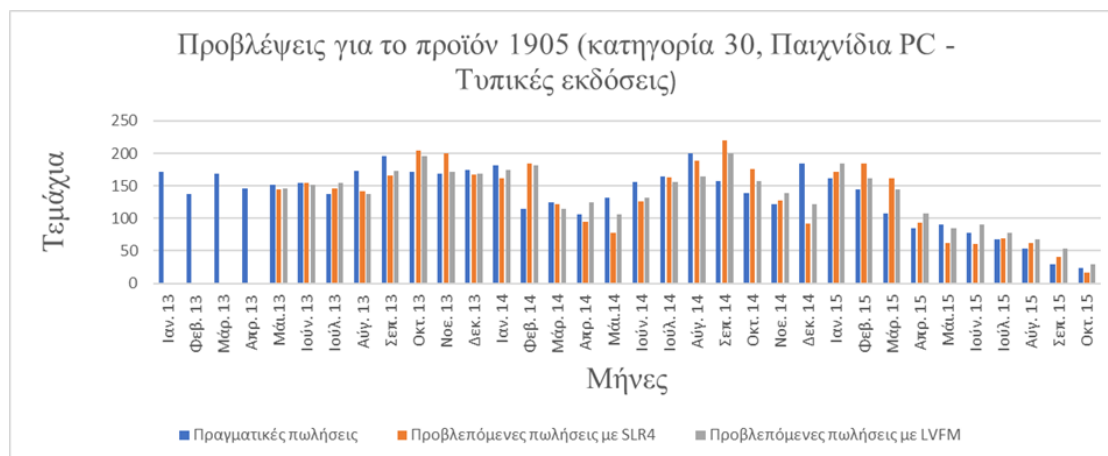
Εικόνα 5.1-17 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 17717



Εικόνα 5.1-18 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 491



Εικόνα 5.1-19 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 3280



Εικόνα 5.1-20 Προβλέψεις με SLR4 και LVFM για το προϊόν με κωδικό 1905

5.1.8 Τελικά συμπεράσματα

(1) Τα αποτελέσματα είναι φανερό ότι δεν είναι τα καλύτερα δυνατά. Θα πρέπει να δοκιμαστούν στην πράξη και άλλες μέθοδοι. Μερικές από αυτές παρουσιάζονται στην παρακάτω ενότητα.

(2) Στις κατηγορίες καλύτερες προβλέψεις φαίνεται να δίνει η LVFM και αμέσως μετά η SLR4 με εποχικότητα. Στα προϊόντα όμως τις καλύτερες προβλέψεις τις δίνει η SLR4 με εποχικότητα.

Έτσι η τελική επιλογή για τις προβλέψεις στα προϊόντα είναι τελικά η μέθοδος SLR4 με εποχικότητα.

(3) Για τον υπολογισμό των MAPE προτείνεται η βαρύτητα 1 για τον υπολογισμό του μέσου MAPE είτε των κατηγοριών, είτε των προϊόντων και με σταθμισμένο μέσο όρο για όλους τους μήνες.

5.2 Όρια και περιορισμοί της έρευνας

Ένας σημαντικός περιορισμός ήταν η έλλειψη δεδομένων για περισσότερους από τους 34 μήνες. Όπως είδαμε η μέση διάρκεια ζωής των προϊόντων ήταν οι 17 περίπου μήνες (16,85). Με δεδομένα 34 μηνών είναι λογικό να μην μπορούν να προκύψουν επαρκή στοιχεία τόσο για την **εποχικότητα** όσο και για μια ενδεχόμενη ύπαρξη **κυκλικότητας**.

Ένα άλλο πρόβλημα ήταν ότι στα δεδομένα της εταιρείας δεν υπήρχαν καθόλου στοιχεία για την εμφάνιση στην αγορά ενός καινούργιου προϊόντος από την εταιρεία ή για το αν και πότε αποσύρεται από την αγορά.

Τα στοιχεία αυτά έπρεπε να βρεθούν από τις πωλήσεις (πότε εμφανίζουν τις πρώτες πωλήσεις) και πότε σταματούν να πωλούν. Έτσι όταν εμφανίζεται ένα καινούργιο προϊόν είναι αδύνατον να κάνει κανείς πρόβλεψη με τα ανύπαρκτα προηγούμενα στοιχεία. Γενικά για τα πρωτοεμφανιζόμενα προϊόντα θα πρέπει να ακολουθήσει κανείς τελείως διαφορετικές τακτικές πρόβλεψης. (van Steenbergen & Mes, 2020).

Στο προηγούμενο πρόβλημα θα πρέπει να προστεθεί και το ενδεχόμενο του «preorder». Μόλις μια εταιρεία κάνει γνωστό ότι πρόκειται σε λίγους μήνες να κυκλοφορήσει μια καινούργια κονσόλα ή ενός «δημοφιλούς» παιχνιδιού, πολλοί πελάτες προπληρώνουν για να την αποκτήσουν.

Τέλος εδώ αξίζει να αναφερθεί και πάλι ότι, από τα στοιχεία της εταιρείας, μόνο τον τελευταίο μήνα (Οκτώβριος 2015) εμφανίστηκαν 302 νέα προϊόντα. Αν η ημερομηνία εμφάνισης τους ήταν γνωστή εκ των προτέρων θα μπορούσε να γίνει πρόβλεψη έστω και με τις υπάρχουσες μεθόδους.

5.3 Μελλοντικές Επεκτάσεις

Εδώ αξίζει να σημειώσει κανείς ότι έγιναν και κάποιες άλλες δοκιμές στην προσπάθεια να γίνουν πιο σωστές προβλέψεις. Κάποιες από αυτές τις δοκιμές ίσως αν ερευνηθούν περισσότερο να δώσουν πιο ακριβή αποτελέσματα.

(1) Σε δόκιμη που έγινε με χρήση της **cubic regression** δεν έδωσε καθόλου ικανοποιητικά αποτελέσματα. Αλλά ενδεχόμενος μια ανώτερου βαθμού πολυωνυμική παλινδρόμηση να αποδώσει καλύτερα αποτελέσματα.

(2) Έγιναν δόκιμες με κώδικα ο οποίος εφαρμόζει την μέθοδο πρόβλεψης **Holt - Winters (Εκθετική Εξομάλυνση με προσαρμογή στην τάση και στην εποχικότητα)**. Η μέθοδος χρησιμοποιεί τρεις παραμέτρους εξομάλυνσης, τις α, β και γ , οι οποίες βοηθούν στην εξομάλυνση των τιμών της χρονοσειράς, της τάσης και της εποχικότητας αντίστοιχα.

α = εξομάλυνση των τιμών της χρονοσειράς

β = εξομάλυνση των τιμών της τάσης

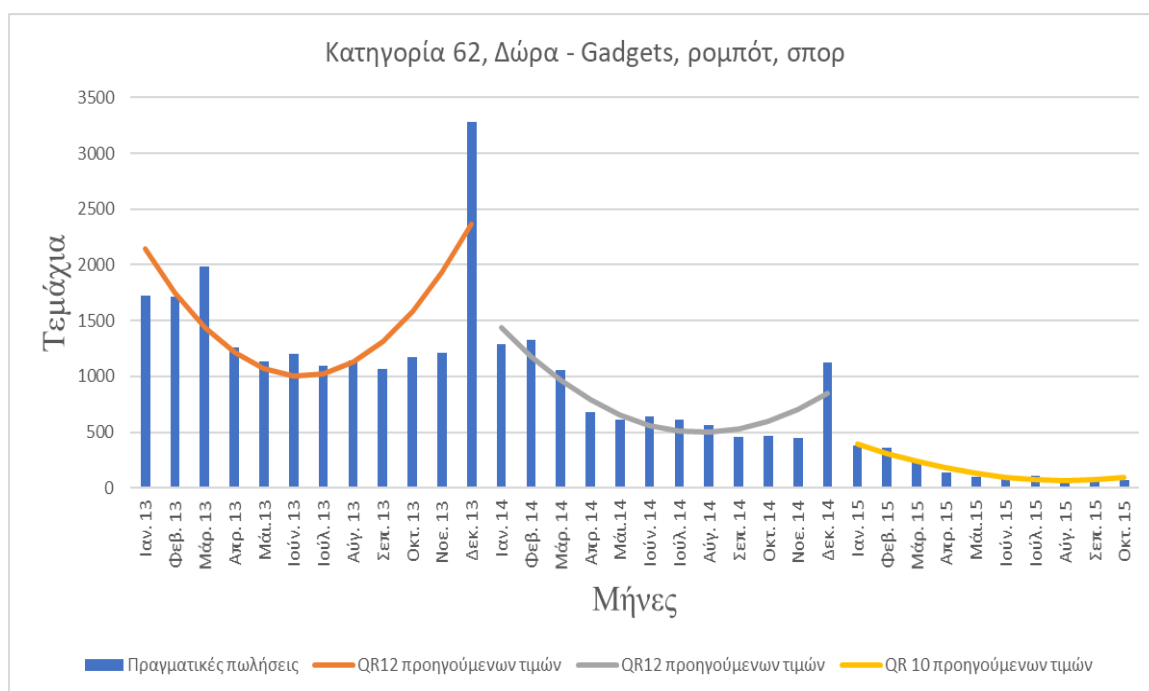
γ = εξομάλυνση των τιμών της εποχικότητας

(Kotsialos, Papageorgiou, & Poulimenos, 2005)

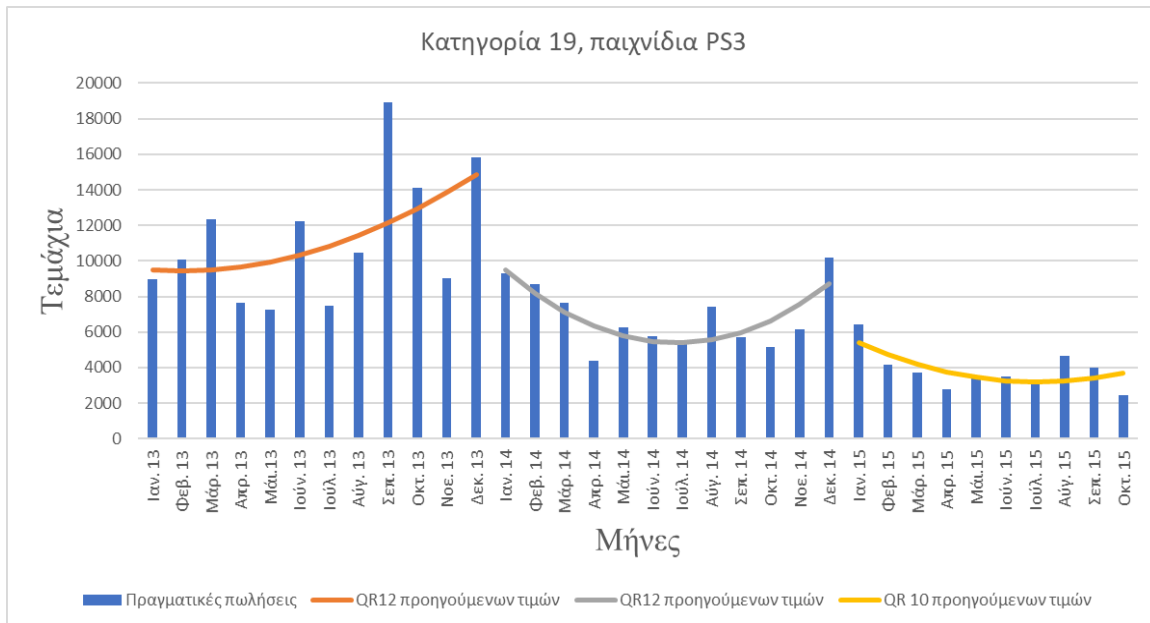
Αν και οι δοκιμές έδειξαν σχετικά καλή προσαρμογή στα δεδομένα, υπάρχει πρόβλημα με την επιλογή των τριών παραμέτρων εξομάλυνσης. Έγινε μια δοκιμή αυτοί να προσδιορίζονται με κριτήριο την ελαχιστοποίηση των MAPE, από τους προηγούμενους μήνες, αλλά υπήρχε το πρόβλημα ότι οι συντελεστές άλλαζαν καθώς άλλαζε ο αριθμός των δεδομένων, το προϊόν ή η κατηγορία.

(3) Εξαιτίας της εποχικότητας μια **απλή παρατήρηση** που ίσως βοηθήσει σε μια μελλοντική έρευνα είναι η εξής: Αν ομαδοποιήσει κανείς τα δεδομένα των πρώτων 24 μηνών σε 2 δωδεκάμηνα και εφαρμοστεί η QR12 για αυτά τα δύο δωδεκάμηνα, παίρνουμε τις καμπύλες της εικόνας 5.3-1 και 5.3-2. Η τρίτη καμπύλη αφορά την εφαρμογή της QR10 για τους τελευταίους μήνες. Φαίνεται η QR12 να αποδίδει με κάποια προσέγγιση την εποχικότητα.

Αυτός εξάλλου ήταν και ένας από τους λόγους που δοκιμάστηκε η μέθοδος QR, αν και τα αποτελέσματα δεν ήταν καθόλου ικανοποιητικά σε αρκετές από τις κατηγορίες, με εξαίρεση την QR8 με βαρύτητα 1, με και χωρίς εποχικότητα (Πίνακας 5.1.7).



Εικόνα 5.3-1 Μοντελοποίηση με QR12 στην κατηγορία 62

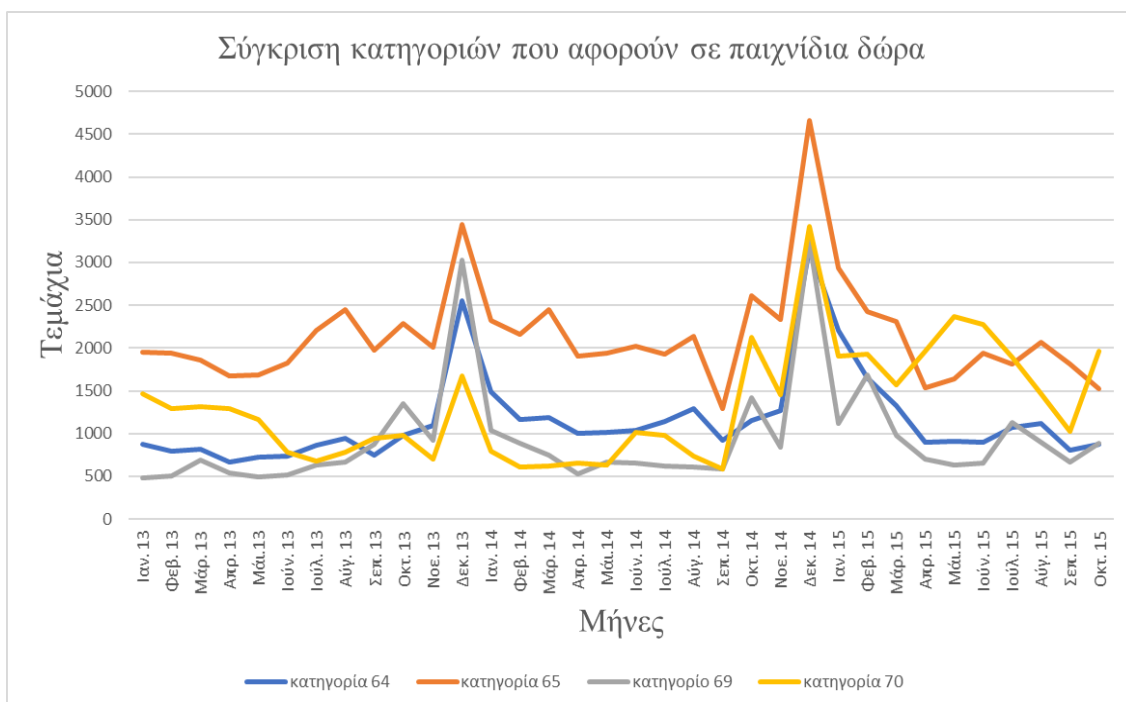


Εικόνα 5.3-2 Μοντελοποίηση με QR12 της κατηγορίας 19

(4) Μια άλλη πιθανή λύση που θα μπορούσε να ερευνηθεί δυστυχώς απαιτεί μια άλλη μεγαλύτερη ίσως ομαδοποίηση των κατηγοριών και ενδεχομένως και των προϊόντων που θα βασίζεται όμως σε κάποια ποιοτικά κριτήρια.

Ένα γενικό συμπέρασμα που ίσως θα μπορούσε να βοηθήσει ένα μελλοντικό ερευνητή, είναι ότι εξαιτίας της φύσης των προϊόντων υπάρχουν κάποιες ομοιότητες στις πωλήσεις ανά μήνα τόσο των κατηγοριών όσο και των προϊόντων.

Σαν παράδειγμα στο παρακάτω σχήμα φαίνονται οι πωλήσεις τεσσάρων (4) διαφορετικών κατηγοριών. Βέβαια όλες τους αναφέρονται σε δώρα - παιχνίδια. Τουλάχιστον οπτικά οι κατηγορίες αυτές φαίνονται να «συγχρονίζονται». Αυτό είναι όμως κάτι που μένει να μελετηθεί στο μέλλον. Απαιτεί βέβαια μια εκτεταμένη σύγκριση των πωλήσεων μεταξύ των διαφόρων κατηγοριών και ενδεχομένως των προϊόντων.



Εικόνα 5.3-3 Σύγκριση κατηγοριών 64, 65, 69 και 70 (δώρα – παιχνίδια)

(5) Μια άλλη προσέγγιση που ίσως θα μπορούσε να δώσει πολύ καλύτερα αποτελέσματα θα απαιτούσε και πάλι ένα διαφορετικό τρόπο ομαδοποίησης και επιπλέον χαρακτηρισμού των προϊόντων. Από το πρόβλημα που αναφέρθηκε στην ενότητα 5.1.7.1 φαίνεται να υπάρχει μια πιο σύνθετη αλληλεξάρτηση μεταξύ προϊόντων και κατηγοριών.

Για παράδειγμα σε κάθε καινούργιο προϊόν δεν αρκεί να αποδίδεται ένας κωδικός προϊόντος και ένας κωδικός κατηγορίας στην οποία ανήκει. Θα πρέπει να έχει μια ημερομηνία πρώτης κυκλοφορίας, έστω και κωδικοποιημένη. Προφανώς η εταιρεία γνωρίζει «εκ των προτέρων» αν στους επόμενους μήνες θα κυκλοφορήσει κάποια συγκεκριμένα προϊόντα.

Το επόμενο στοιχείο που σίγουρα γνωρίζει η εταιρεία είναι αν το προϊόν που πρόκειται να διαθέσει στην αγορά, μπορεί να συνδεθεί με κάποια άλλα προϊόντα. Για παράδειγμα αν είναι παιχνίδι για κάποια κονσόλα παιχνιδιών, με ποια κονσόλα συνδέεται (δηλαδή με ποια άλλη κατηγορία ή ποιο άλλο προϊόν) ή ακόμα με ποια συγκεκριμένα αξεσουάρ συνδέεται. Ας μην ξεχνάμε ότι η αγορά του παιχνιδιού μπορεί να συμπαρασείρει και την απαραίτητη αγορά κάποιων αξεσουάρ που συνδέονται με αυτό (π.χ. ειδικά χειριστήρια, ειδικά γυαλιά κ.λπ).

Αν είναι κάποιο δώρο αναμνηστικό με ποια γιορτή συνδέεται (ημερολογιακά) (Χριστούγεννα, Αγίου Βαλεντίνου, Αποκριά). Αν είναι κάποιο μουσικό CD σε τι είδους μουσική αναφέρεται;

Θα πρέπει να κατηγοριοποιηθούν τα διάφορα είδη μουσικής π.χ. τοπική μουσική, ξένη μουσική, κλασική κ.λπ. Προφανώς οι πωλήσεις επηρεάζονται από το είδος της μουσικής.

Οι κινηματογραφικές ταινίες μπορούν να κατηγοριοποιηθούν ανάλογα ως προς την μορφή τους (DVD, blue-ray, blue-ray - 3d, blue-ray - 4k) ως προς το είδος τους (Action, animation... κ.λπ), ως προς την εποχή κυκλοφορίας τους, εάν είναι block buster, εάν είναι συλλεκτικές κ.α.)

Βέβαια θα πρέπει να φιλτράρονται τα προϊόντα που δεν πωλούνται πια καθώς και αυτά που δεν πουλήσαν καθόλου (μηδενικές πωλήσεις) για κάποιο συγκεκριμένο χρονικό διάστημα.

Αυτό βέβαια συνεπάγεται ένα σύνθετο αρχείο δεδομένων και κατάλληλο προγραμματισμό για την διαχείριση του. Όλα τα παραπάνω μπορούν να επεξεργαστούν και να υλοποιηθούν με την βοήθεια των **νευρωνικών δικτύων** (Neural Networks) μηχανικής μάθησης (machine learning)

6 Βιβλιογραφία

- 1C:Company. (2018, February). *Kaggle*. Ανάκτηση από https://www.kaggle.com/competitions/competitive-data-science-predict-future-sales/data?select=sales_train.csv
- 1C:Company. (2019). Ανάκτηση από <https://1c.com/>
- Adriano Bernardo Renzia, S. F. (2015). *science direct*. Ανάκτηση από <https://www.sciencedirect.com>:
<https://www.sciencedirect.com/science/article/pii/S2351978915008276>
- Akkar, P. D. (2016, Δεκέμβριος). *Grass Root Algorithm Optimize Neural Networks for Classification Problem*. Ανάκτηση από asrjetsjournal.org:
<https://core.ac.uk/download/pdf/235050107.pdf>
- AMA, A. (2023). *ama.org*. Ανάκτηση από <https://www.ama.org/topics/market-research/>
- Armstrong, J. S. (2008, july 20). *papers.ssrn.com*. Ανάκτηση από https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1164602
- Boutsikas, M. (2004). *unnipi.gr*. Ανάκτηση από Τμήμα Στατιστικής & ασφάλειας επιστήμης Πανεπιστήγμιο Πειραιά:
http://www.unipi.gr/faculty/mbouts/statprog/SPSS_lesson9-10.pdf
- Cerqueira, V. (2023, March 14). *towardsdatascienc*. Ανάκτηση από <https://towardsdatascience.com/understanding-time-series-trend-addfd9d7764e>
- Chatfield, C. (2000). *books.google.gr*. Ανάκτηση από https://books.google.gr/books?hl=en&lr=&id=PFHMBQAAQBAJ&oi=fnd&pg=PR7&dq=Times+series+forecasting+methods+papers&ots=fZcpiYeDvj&sig=MDq88mAJIpEDliOosJPvwDvRFqs&redir_esc=y#v=onepage&q=Times%20series%20forecasting%20methods%20papers&f=false
- Chopra&Meindl. (2004). *Studocu*. Ανάκτηση από <https://www.studocu.com/en-us/document/university-of-maryland-eastern-shore/supply-chain-management/chapter-7-text/40812999>
- Dagum, E. B. (2013, May). *Time Series Modelling and decomposition*. Ανάκτηση από https://www.researchgate.net/publication/307663962_Time_Series_Modelling_and_Decomposition
- Dealhub, E. (2022, July 31). *dealhub.io*. Ανάκτηση από <https://dealhub.io/glossary/sales-forecasting/>
- Dhokal, C. P. (2017, January - March). *A Naïve Approach for Comparing a Forecast*. Ανάκτηση από https://www.researchgate.net/publication/326972994_A_Naive_Approach_for_Comparing_a_Forecast_Model

- Erycha Puspitasari, N. E. (2023, February). *Inventory Forecasting Analysis using The Weighted Moving Average Method in Go Public Trading Companies*. Ανάκτηση από <https://equatorsience.com/index.php/jabter/article/view/160/124>
- Github. (2021, Απρίλιος 29). *problemsolvingwithpython.com*. Ανάκτηση από <https://problemsolvingwithpython.com/05-NumPy-and-Arrays/05.08-Systems-of-Linear-Equations/>
- Hotmath. (2020, Δεκέμβριος). *varsitytutors.com*. Ανάκτηση από https://www.varsitytutors.com/hotmath/hotmath_help/topics/quadratic-regression
- Hyndman, R. J., & Athanasopoulos, G. (2018, April). 2.9 White noise. Στο R. J. Hyndman, & G. Athanasopoulos, *Forecasting: Principles and Practice (2nd ed)*. Ανάκτηση από <https://otexts.com/fpp2/wn.html>
- Hyndman, R. J., & Athanasopoulos, G. (2018, April). 2.8 Autocorrelation. Στο R. J. Athanasopoulos, *Forecasting: Principles and Practice (2nd ed)*. Ανάκτηση από <https://otexts.com/fpp2/autocorrelation.html>
- Hyndman, R. J., & Athanasopoulos, G. (χ.χ.). *Otexts*. Ανάκτηση από <https://otexts.com/fpp2/wn.html>
- Hyndman, R., & Athanasopoulos, G. (2021, Φεβρουάριος). 6.5 Πρόβλεψη σεναρίου. Στο R. Hyndman, & G. Athanasopoulos, *Πρόβλεψη: Αρχές και Πρακτική (3rd ed)*. Ανάκτηση από Monash University, Australia: <https://otexts.com/fppgr/scenarios.html>
- Jedox. (2022). Ανάκτηση από <https://www.jedox.com/en/blog/error-metrics-how-to-evaluate-forecasts/>
- Kotsialos, A., Papageorgiou, M., & Poulimenos, A. (2005, August). Long-term sales forecasting using Holt–Winters and neural network methods. *Journal of forecasting*. Ανάκτηση από https://www.researchgate.net/publication/5141722_Long-term_sales_forecasting_using_Holt-Winters_and_neural_network_methods
- Ma. Del Rocio Castillo E, M. E. (2020, June). *Evaluation of Several Error Measures Applied to the Sales Forecast System of Chemicals Supply Enterprises*. Ανάκτηση από https://www.researchgate.net/publication/342592090_Evaluation_of_Several_Error_Measures_Applied_to_the_Sales_Forecast_System_of_Chemicals_Supply_Enterprises
- Planetcalc. (χ.χ.). *planetcalc.com*. Ανάκτηση από <https://planetcalc.com/5992/>
- Roberto Corizzo, M. C. (2019, July). Anomaly Detection and Repair for Accurate Predictions in Geo-distributed Big Data. *Big Data Research*, σσ. Volume 1,, Pages 18-35. Ανάκτηση από www.sciencedirect.com/science/article/abs/pii/S2214579618302119

- Tinungki, G. M. (2019). *IOP Conference Series: Earth and Environmental Science*.
Ανάκτηση από <https://iopscience.iop.org/article/10.1088/1755-1315/235/1/012097/pdf>
- van Steenberg, R., & Mes, M. K. (2020, December). Forecasting demand profiles of new products. *Decision Support Systems*, σσ. Volume 139, December 2020, 113401.
- Winkler, S. M. (1983, September). *AVERAGES OF FORECASTS: SOME EMPIRICAL RESULTS*.
Ανάκτηση από https://www.researchgate.net/publication/227445401_Averages_of_Forecasts_Some_Empirical_Results
- Κουτσοθανάση, Β. (2015). dione.lib.unipi.g. Στο Β. Κουτσοθανάση, *ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ, ΑΝΑΛΥΣΗ* (σ. 13). Πειραιάς. Ανάκτηση από https://dione.lib.unipi.gr/xmlui/bitstream/handle/unipi/8834/Koustoathanasi_Vasiliki.pdf?sequence=1&isAllowed=y
- Παρασκευάς, Δ. (2017, Φεβρουάριος). *ikee.lib.auth.gr*. Ανάκτηση από <https://ikee.lib.auth.gr/record/288156/files/GRI-2017-18804.pdf>
- Στούμπου, Χ. (2021, Οκτώβριος). *Μέθοδοι πρόβλεψης και εφαρμογές*. Ανάκτηση από <https://dspace.lib.ntua.gr/xmlui/bitstream/handle/123456789/54451/HARA%20TOUMPOU%20THESIS.pdf?sequence=1>
- Ψωϊνός, Δ. Π. (1983). *Οργάνωση και διοίκηση εργοστασίων, Εισαγωγή στις βασικές μεθόδους*. Θεσσαλονίκη: ΖΗΤΗ.