



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ «ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

**Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση
ιεραρχικών μεθόδων**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Παναγιώτης Χ. Πετρόχειλος

Επιβλέπων: Ευάγγελος Σπηλιώτης
Επίκουρος καθηγητής Ε.Μ.Π.



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ «ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

**Πρόβλεψη ταξινόμησης αυτοκινήτων στην Ελλάδα με χρήση
ιεραρχικών μεθόδων**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Παναγιώτης Χ. Πετρόχειλος

Επιβλέπων: Ευάγγελος Σπηλιώτης
Επίκουρος καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 2^η Ιουλίου 2025

.....

Σπηλιώτης Ευάγγελος

Επίκουρος καθηγητής,
Ε.Μ.Π.

.....

Μαρινάκης Ευάγγελος

Επίκουρος καθηγητής,
Ε.Μ.Π.

.....

Ασκούνης Δημήτριος

Καθηγητής, Ε.Μ.Π.



.....
Παναγιώτης Πετρόχειλος

Κάτοχος Μεταπτυχιακού Προγράμματος “Τεχνο-οικονομικά Συστήματα” της Σχολής
Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών Ε.Μ.Π.

Copyright © **Παναγιώτης Πετρόχειλος, 2025**

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η παρούσα διπλωματική εργασία ασχολείται με την εξαγωγή, επεξεργασία, ανάλυση και πρόβλεψη δεδομένων ταξινομήσεων αυτοκινήτων στην Ελλάδα για την περίοδο 2000–2024. Στόχος της είναι να αντιμετωπίσει την έλλειψη ανοικτών και δομημένων δεδομένων στον τομέα αυτό, μέσω της ανάπτυξης μιας αυτοματοποιημένης ροής επεξεργασίας αρχείων PDF από τον Σύνδεσμο Εισαγωγέων Αντιπροσώπων Αυτοκινήτων (ΣΕΑΑ). Χρησιμοποιώντας τη γλώσσα προγραμματισμού Python, τα δεδομένα εξάγονται και μετατρέπονται σε επεξεργάσιμες μορφές, ενώ με τη γλώσσα R δημιουργούνται ιεραρχικά μοντέλα πρόβλεψης. Τα δεδομένα οργανώνονται ανά έτος ή μήνα και ταξινομούνται κατά μάρκα, μοντέλο, κατηγορία και γεωγραφική περιοχή. Το προτεινόμενο μοντέλο μπορεί να αξιοποιηθεί για το σχεδιασμό στρατηγικής από εταιρείες του κλάδου και φορείς χάραξης πολιτικής. Η συνδυαστική χρήση των νέων τεχνολογιών αναδεικνύει τη δυνατότητα αξιοποίησης μη δομημένων δεδομένων για την υποστήριξη αξιόπιστων αναλύσεων και προβλέψεων.

Λέξεις-κλειδιά:

Ταξινομήσεις αυτοκινήτων, Πρόβλεψη, Ανάλυση δεδομένων, Μη δομημένα δεδομένα, Ιεραρχικά δεδομένα χρονοσειρών, Ανοιχτά δεδομένα.

Abstract

This thesis explores the extraction, processing, analysis, and forecasting of car registration data in Greece over the period 2000–2024. The work addresses the lack of open and structured datasets by developing an automated pipeline for extracting data from PDF files provided by the Hellenic Association of Motor Vehicle Importers (SEAA). Using Python, data are converted into editable structures, thus making it possible to construct hierarchical time series forecasting models. The data are organized by year or month and classified by country, brand, model, vehicle segment, and geographic region. The proposed methodology provides a valuable tool for industry stakeholders and policymakers in strategic planning and market monitoring. By combining the new technologies, the study highlights how unstructured public data can be transformed into actionable intelligence, demonstrating the potential of open-source tools in economic and industrial research. The results support evidence-based decision-making and offer a replicable framework for similar analyses in other domains lacking structured datasets.

Keywords

Car registration data, Forecasting, Data analysis, Unstructured data, Hierarchical time series, Open data, Open-source technologies.

Περιεχόμενα

Περίληψη	5
Abstract	7
Πρόλογος	12
Κατάλογος σχημάτων	13
Κατάλογος πινάκων	14
1. Εισαγωγή	15
1.1 Αντικείμενο Εργασίας.....	15
1.2 Δομή Εργασίας.....	19
2. Η ιστορία της αυτοκινητοβιομηχανίας στην Ευρώπη και την Ελλάδα (1885–2024).....	21
2.1 Ευρώπη.....	22
2.1.1 Οι απαρχές της αυτοκίνησης (1885–1900).....	22
2.1.2 Η εδραίωση της βιομηχανίας (1900–1945)	23
2.1.3 Η μεταπολεμική άνθηση και η χρυσή εποχή (1945–1975)	23
2.1.4 Περίοδος κρίσης και προσαρμογής (1975–2000)	23
2.1.5 Μετάβαση στον 21ο αιώνα: Παγκοσμιοποίηση και τεχνολογική μετάβαση (2000–2016).....	23
2.1.6 Η πράσινη επανάσταση και η νέα εποχή (2016–2024)	25
2.2 Ελλάδα	28
2.2.1 Τα πρώτα βήματα (1900–1940).....	28
2.2.2 Μεταπολεμική ανάπτυξη και προσπάθεια για ελληνική παραγωγή (1950–1975).....	28
2.2.3 Η έκρηξη της αγοράς και οι εισαγωγές (1980–2009)	30
2.2.4 Η κρίση και η πτώση των πωλήσεων (2010–2015).....	31
2.2.5 Ανάκαμψη και πράσινη μετάβαση (2016–2024).....	32
2.2.6 Πρόσφατα δεδομένα και προκλήσεις (2024–2025)	32
2.3 Σύνοψη	33
Κεφάλαιο 3. Θεωρία προβλέψεων.....	36
3.1 Εισαγωγικές έννοιες.....	36
3.1.1 Προβλέψεις.....	36
3.1.2 Χρονοσειρές	37

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων	
3.1.3 Ορίζοντας πρόβλεψης.....	38
3.2 Κατηγορίες μεθόδων πρόβλεψης	39
3.2.1 Πρόβλεψη στόχου ή προϋπολογισμού	39
3.2.2 Κριτική Πρόβλεψη	39
3.2.3 Στατιστική Πρόβλεψη	39
3.3 Μοντέλα εκθετικής εξομάλυνσης	42
3.3.1 Μοντέλο σταθερού επιπέδου (Simple Exponential Smoothing).....	43
3.3.2 Μοντέλο γραμμικής τάσης (Holt Exponential Smoothing)	45
3.3.3 Μοντέλο μη γραμμικής τάσης (Damped Exponential Smoothing).....	46
3.3.4 Εποχιακή εξομάλυνση	47
3.3.5 Ταξινόμηση μεθόδων εκθετικής εξομάλυνσης - Ο αλγόριθμος ETS.....	49
3.4 Ακρίβεια προβλέψεων.....	52
3.4.1 Στατιστικοί δείκτες	52
3.4.2 Κυλιόμενη πρόβλεψη	54
4. Συλλογή, οργάνωση και επεξεργασία δεδομένων ταξινομήσεων	56
4.1 Εισαγωγή.....	56
4.2 Παραδείγματα αρχείων PDF από ΣΕΑΑ	58
4.3 Περιγραφή μεθοδολογίας.....	60
4.4 Πλεονεκτήματα προτεινόμενης μεθοδολογίας.....	61
4.5 Παραγόμενα σύνολα δεδομένων.....	63
5. Θεωρία ιεραρχικών προβλέψεων.....	66
5.1 Ιεραρχικές προβλέψεις	66
5.1.1 Μέθοδοι ιεραρχικής πρόβλεψης.....	66
5.1.1.1 Μέθοδοι bottom up, top down και middle out.....	66
5.1.1.2 Μέθοδοι βέλτιστης συνάθροισης (Optimal reconciliation)	67
5.1.2 Χρησιμότητα και εφαρμογές.....	68
5.2 Ομαδοποιημένες χρονοσειρές	69
5.2.1 Μέθοδοι συνάθροισης για ομαδοποιημένες χρονοσειρές	69
5.2.2 Πλεονεκτήματα και χρήσεις	70
6. Παρουσίαση προσέγγισης.....	71
6.1 Εισαγωγή.....	71

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων	
6.2 Προετοιμασία δεδομένων.....	71
6.3 Δημιουργία αντικειμένου GTS	71
6.4 Αξιολόγηση μοντέλου	72
6.5 Σύγκριση Μεθόδων Πρόβλεψης	73
6.6 Οπτικοποίηση χρονοσειρών μοντέλου.....	77
7. Παρουσίαση αποτελεσμάτων.....	81
7.1 Επίπεδο συνολικών ταξινομήσεων	81
7.2 Επίπεδο μάρκας.....	82
7.3 Επίπεδο κατηγορίας	83
7.4 Επίπεδο τύπου	84
7.5 Επίπεδο μοντέλου.....	84
7.6 Σύγκριση προβλέψεων για τις τρεις κύριες κατηγορίες.....	85
8. Συμπεράσματα και Προεκτάσεις	88
8.1 Ανασκόπηση.....	88
8.2 Συμπεράσματα	89
8.3 Προεκτάσεις	91
Παράρτημα.....	92
Βιβλιογραφία	97

Πρόλογος

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2024-2025 στα πλαίσια ολοκλήρωσης του Προγράμματος Μεταπτυχιακών Σπουδών «Τεχνοοικονομικά Συστήματα» του τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Θα ήθελα να εκφράσω τις εγκάρδιες ευχαριστίες μου προς τον Επίκουρο Καθηγητή κ. Σπηλιώτη Ευάγγελο για την εμπιστοσύνη που μου έδειξε, αναθέτοντάς μου την παρούσα διπλωματική εργασία, καθώς και για την ευκαιρία που μου προσέφερε να εμβαθύνω στον εξαιρετικά ενδιαφέροντα και απαιτητικό τομέα των προβλέψεων. Η σταθερή καθοδήγηση, η υποστήριξη και το διαρκές επιστημονικό ενδιαφέρον του υπήρξαν καθοριστικής σημασίας σε όλα τα στάδια της εκπόνησης της εργασίας.

Θα ήθελα, επιπλέον, να ευχαριστήσω θερμά τους συναδέλφους μου στον επαγγελματικό χώρο, για την κατανόηση, την ενθάρρυνση και την ηθική στήριξη που μου προσέφεραν κατά τη διάρκεια αυτής της απαιτητικής περιόδου, καθώς επέστρεψα στις σπουδές έπειτα από αρκετά χρόνια επαγγελματικής πορείας.

Τέλος, θα ήθελα να εκφράσω την ειλικρινή μου ευγνωμοσύνη προς την οικογένειά μου, η οποία με στηρίζει αδιάλειπτα σε κάθε μου βήμα, με πίστη, κατανόηση και αμέριστη αγάπη.

Κατάλογος σχημάτων

Σχήμα 1.1 - Το παραγόμενο ιεραρχικό μοντέλο προβλέψεων.....	Σελ. 16
Σχήμα 2.1 Το πρώτο αυτοκινούμενο αυτοκίνητο του Nicolas-Joseph Cugnot.....	Σελ. 18
Σχήμα 2.2 Αυτόνομο ταξί της Waymo One στο San Francisco.....	Σελ. 22
Σχήμα 2.3 Το πρώτο αυτοκίνητο στην Ελλάδα.....	Σελ. 24
Σχήμα 2.4 Το εν μέρει παραγόμενο στην Ελλάδα, Pony της NAMCO.....	Σελ. 26
Σχήμα 2.5 Ευρωπαϊκές και Ελληνικές Ταξινομήσεις.....	Σελ. 30
Σχήμα 3.1: Γραφική αναπαράσταση των μοντέλων εκθετικής εξομάλυνσης.....	Σελ. 39
Σχήμα 3.2 Παράδειγμα rolling origin evaluation.....	Σελ. 51
Σχήμα 4.1 Δείγμα του αρχείου PDF από ΣΕΑΑ ανά μήνα και μάρκα.....	Σελ. 54
Σχήμα 4.2 Δείγμα του αρχείου PDF από ΣΕΑΑ ανά περιοχή και μάρκα.....	Σελ. 55
Σχήμα 4.3 Δείγμα του αρχείου PDF από ΣΕΑΑ ανά μον/λο, μάρκα και κατηγορία.....	Σελ. 56
Σχήμα 6.1 Παραγόμενο αντικείμενο GTS.....	Σελ. 68
Σχήμα 6.2 Αντικείμενα HTS.....	Σελ. 69
Σχήμα 6.3 Συνολικές ταξινομήσεις.....	Σελ. 73
Σχήμα 6.4 Ταξινομήσεις ανά κατηγορία.....	Σελ. 74
Σχήμα 6.5 Ταξινομήσεις ανά τύπο.....	Σελ. 74
Σχήμα 6.6 Ταξινομήσεις ανά μάρκα.....	Σελ. 75
Σχήμα 6.7 Ταξινομήσεις ανά μοντέλο.....	Σελ. 76

Κατάλογος πινάκων

Πίνακας 2.1 Πωλήσεις ηλεκτρονικών αυτοκινήτων στην Ευρωπαϊκή Ένωση.....	Σελ. 23
Πίνακας 3.1 Ταξινόμηση μοντέλων εκθετικής εξομάλυνσης.....	Σελ. 46
Πίνακας 4.1 Δείγμα του παραγόμενου πίνακα ανά μοντέλο, μάρκα, κατηγορία και τύπο....	Σελ. 59
Πίνακας 4.2 Δείγμα του παραγόμενου πίνακα ανά μάρκα και περιοχή.....	Σελ. 60
Πίνακας 4.3 Δείγμα του παραγόμενου πίνακα ανά μάρκα και Μήνα.....	Σελ. 61
Πίνακας 6.1 Εξεταζόμενα μοντέλα και μέθοδοι ιεραρχικών προβλέψεων.....	Σελ. 71
Πίνακας 7.1 Τα πέντε καλύτερα και πέντε χειρότερα αποτελέσματα σε επίπεδο συνολικών ταξινομήσεων.....	Σελ. 77
Πίνακας 7.2 Τα πέντε καλύτερα και πέντε χειρότερα αποτελέσματα σε επίπεδο μάρκας.....	Σελ. 78
Πίνακας 7.3 Τα πέντε καλύτερα και πέντε χειρότερα αποτελέσματα σε επίπεδο κατηγορίας.....	Σελ. 79
Πίνακας 7.4 Τα πέντε καλύτερα και πέντε χειρότερα αποτελέσματα σε επίπεδο τύπου.....	Σελ. 80
Πίνακας 7.5 Τα πέντε καλύτερα και πέντε χειρότερα αποτελέσματα σε επίπεδο μοντέλου...	Σελ. 81
Πίνακας 7.6 Τα αποτελέσματα σύγκρισης σφαλμάτων στο αντικείμενο GTS.....	Σελ. 83
Πίνακας 7.7 Ο αναλυτικός πίνακας του σφάλματος MASE για όλες τις μεθόδους και όλα τα μοντέλα.....	Σελ. 88

1. Εισαγωγή

1.1 Αντικείμενο Εργασίας

Η εργασία έχει ως αντικείμενο τη συστηματική συλλογή, μετατροπή, ανάλυση και πρόβλεψη δεδομένων που αφορούν τις ταξινομήσεις αυτοκινήτων στην Ελλάδα. Η σημασία της εργασίας έγκειται στην ανάγκη παρακολούθησης της οικονομικής δραστηριότητας και της κατανάλωσης μέσα από έναν από τους πλέον ενδεικτικούς δείκτες, αυτόν των νέων ταξινομήσεων. Η παρακολούθηση των νέων ταξινομήσεων είναι πολύ βασική για την χάραξη στρατηγικής των εταιριών του κλάδου της αυτοκινητοβιομηχανίας. Επίσης οι νέες ταξινομήσεις συχνά λειτουργούν ως προπομπός της οικονομικής πορείας, αντανakλώντας την αγοραστική δύναμη των νοικοκυριών, τις επιχειρηματικές προσδοκίες και τη διαθεσιμότητα χρηματοδότησης. Επιπλέον, σχετίζονται με κλάδους όπως η ασφάλιση και η ενέργεια, καθιστώντας την παρακολούθησή τους ουσιώδη για ευρύτερες αναλύσεις.

Ο Σύνδεσμος Εισαγωγέων Αντιπροσώπων Αυτοκινήτων (ΣΕΑΑ) αποτελεί την επίσημη πηγή για τις ταξινομήσεις στην Ελλάδα, ωστόσο τα δεδομένα δημοσιεύονται σε μορφή PDF, χωρίς σταθερή δομή, γεγονός που καθιστά τη συλλογή και την επεξεργασία τους χρονοβόρα και επιρρεπή σε σφάλματα. Τα ευρωπαϊκά δεδομένα από τον αντίστοιχο Ευρωπαϊκό σύνδεσμο (ACEA) είναι πιο εύκολα επεξεργάσιμα καθώς δίνονται κατά κύριο λόγο σε μορφή Excel αλλά δεν παρέχουν ανάλυση σε επίπεδο συνδυασμού μάρκας, κατηγορίας, τύπου μηχανής και μοντέλου αυτοκινήτου.

Για την αντιμετώπιση αυτών των περιορισμών, η εργασία αναπτύσσει μια αυτοματοποιημένη διαδικασία εξαγωγής δεδομένων από αρχεία PDF του ΣΕΑΑ. Η εξαγωγή γίνεται με χρήση της γλώσσας προγραμματισμού Python και των κατάλληλων βιβλιοθηκών, με στόχο τη μετατροπή των δεδομένων σε επεξεργάσιμη μορφή Excel. Στη συνέχεια, τα δεδομένα οργανώνονται σε ιεραρχική δομή και αναλύονται με στόχο την εξαγωγή ουσιαστικών, ποιοτικών και ποσοτικών συμπερασμάτων.

Τα δεδομένα που συλλέγονται, αποθηκεύονται σε διαφορετικά αρχεία Excel, τα οποία περιλαμβάνουν:

- Συνολικές ταξινομήσεις στην Ελλάδα από το 1965 έως το 2024,
- Ταξινομήσεις ανά μοντέλο, μάρκα και κατηγορία από το 2000,
- Ταξινομήσεις ανά γεωγραφική περιοχή από το 2005,
- Αντίστοιχα δεδομένα για τις χώρες της Ευρώπης από το 1990.

Από τα διαθέσιμα αρχεία, εκείνο που περιλαμβάνει τις ταξινομήσεις ανά μοντέλο, μάρκα και κατηγορία (segment) χρησιμοποιείται για την κατασκευή ενός ιεραρχικού μοντέλου χρονοσειρών. Η ανάλυση του βασίζεται σε ετήσια δεδομένα, με στόχο την πρόβλεψη των μελλοντικών ετήσιων ταξινομήσεων σε επίπεδο συνόλου, μάρκας, κατηγορίας, τύπου και μοντέλου.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
Το βασικό αναλυτικό εργαλείο της εργασίας είναι η χρήση ιεραρχικού μοντέλου πρόβλεψης με τη γλώσσα προγραμματισμού R και συγκεκριμένα με τη βιβλιοθήκη *hts*.

Η βιβλιοθήκη *hts* της γλώσσας R χρησιμοποιείται για την ανάλυση και πρόβλεψη ιεραρχικών και ομαδοποιημένων χρονοσειρών. Αναπτύχθηκε από τον Rob J Hyndman και τους συνεργάτες του και αποτελεί ένα ισχυρό εργαλείο για τη διαχείριση δεδομένων που παρουσιάζουν ιεραρχική ή πολυδιάστατη δομή, όπως π.χ. πωλήσεις ανά περιοχή, προϊόν και χρονική περίοδο. Η *hts* επιτρέπει στον χρήστη να ορίσει μια ιεραρχία ή μια ομαδοποίηση των χρονοσειρών και να εφαρμόσει προβλέψεις τόσο στο συνολικό επίπεδο όσο και στα επιμέρους υποεπίπεδα. Βασικό της πλεονέκτημα είναι η δυνατότητα συνδυασμού προβλέψεων (*forecast reconciliation*), έτσι ώστε οι προβλέψεις σε όλα τα επίπεδα της ιεραρχίας να είναι συνεπείς μεταξύ τους [19].

Μέσω της μεθόδου αυτής, επιχειρείται η πρόβλεψη των μελλοντικών ταξινομήσεων σε πολλαπλά επίπεδα ανάλυσης, όπως συνολικές εθνικές ταξινομήσεις, ταξινομήσεις ανά μάρκα, μοντέλο, κατηγορία ή τύπο. Η αξιοπιστία και ακρίβεια των προβλέψεων αξιολογείται στο τελευταίο στάδιο της εργασίας με την χρήση στατιστικών δεικτών σφάλματος.

Η πρόβλεψη σε ιεραρχικές δομές αποτελεί ιδιαίτερη πρόκληση για την επιστήμη των δεδομένων, καθώς απαιτεί τη διατήρηση της συνέπειας μεταξύ επιπέδων. Για παράδειγμα, οι επιμέρους προβλέψεις ανά μοντέλο θα πρέπει να συμφωνούν με τη συνολική πρόβλεψη. Η θεωρητική προσέγγιση των ιεραρχικών χρονοσειρών αναπτύχθηκε ώστε να απαντήσει σε τέτοια προβλήματα, διατηρώντας την ακρίβεια χωρίς να θυσιάζει τη δομική συνοχή των δεδομένων.

Για την αξιολόγηση των μεθόδων πρόβλεψης, επιλέγονται οι τέσσερις (4) τελευταίες ετήσιες παρατηρήσεις ως εκτός δείγματος (*out-of-sample*), ενώ για την εκπαίδευση των μοντέλων χρησιμοποιούνται οι είκοσι (20) προηγούμενες παρατηρήσεις κάθε χρονοσειράς. Οι προβλέψεις παράγονται με την εφαρμογή τεσσάρων διαφορετικών μεθοδολογικών προσεγγίσεων της μεθόδου ιεραρχικών προβλέψεων και των παραλλαγών τους.

Η πρώτη μέθοδος, γνωστή ως *bottom-up*, παράγει προβλέψεις στο χαμηλότερο επίπεδο της ιεραρχίας, δηλαδή για κάθε επιμέρους μοντέλο αυτοκινήτου. Στη συνέχεια, οι προβλέψεις αυτές αθροίζονται προκειμένου να εξαχθούν προβλέψεις για τα ανώτερα επίπεδα: μάρκα, κατηγορία ή τύπο, καθώς και για το συνολικό πλήθος ταξινομήσεων.

Η δεύτερη μέθοδος, η *top-down*, λειτουργεί αντίστροφα. Αρχικά υπολογίζονται προβλέψεις για τις συνολικές ταξινομήσεις, και στη συνέχεια αυτές κατανέμονται στα χαμηλότερα επίπεδα της ιεραρχίας βάσει ποσοστών. Η κατανομή γίνεται είτε με βάση τα ιστορικά ποσοστά είτε ακολουθώντας τη μέθοδο των Athanasopoulos, Ahmed & Hyndman (2009), όπου τα ποσοστά κάθε επιμέρους κατηγορίας καθορίζονται σύμφωνα με το ποσοστό της πρόβλεψής της ως προς την συνολική πρόβλεψη.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
Η τρίτη μέθοδος, γνωστή ως *middle-out*, εφαρμόζει προβλέψεις στο ενδιάμεσο επίπεδο της ιεραρχίας (δηλαδή στο επίπεδο της μάρκας, της κατηγορίας ή του τύπου). Από εκεί, η πληροφορία διαχέεται προς τα πάνω, μέσω συνάθροισης για να προκύψει η συνολική πρόβλεψη και προς τα κάτω, μέσω κατανομής για να προκύψουν προβλέψεις ανά μοντέλο.

Τέλος, η τέταρτη μέθοδος είναι εκείνη της βέλτιστης συνάθροισης (*optimal reconciliation*). Σε αυτή την προσέγγιση, πραγματοποιούνται ανεξάρτητες προβλέψεις σε κάθε επίπεδο της ιεραρχίας, χωρίς να επιβάλλεται εξαρχής κάποια δομική συνέπεια μεταξύ τους. Στη συνέχεια, εφαρμόζεται ένας πίνακας συμφιλίωσης (*reconciliation matrix*), ο οποίος προσαρμόζει τις επιμέρους προβλέψεις με βάση στατιστικά βέλτιστα βάρη. Με αυτόν τον τρόπο επιτυγχάνεται η ελαχιστοποίηση του συνολικού σφάλματος πρόβλεψης και διασφαλίζεται η συνεκτικότητα των προβλέψεων μεταξύ όλων των επιπέδων της ιεραρχίας.

Η εργασία στηρίζεται τεχνικά σε δύο βασικά προγραμματιστικά περιβάλλοντα. Η Python χρησιμοποιείται για την εξαγωγή και μετατροπή των δεδομένων από αρχεία PDF σε δομημένη μορφή, ενώ η R αξιοποιείται για την πρόβλεψη μέσω της μεθόδου των ιεραρχικών χρονοσειρών. Συμπληρωματικά, γίνεται χρήση του Microsoft Excel για τον έλεγχο και τη μορφοποίηση των δεδομένων.

Η σύγκριση των μεθόδων βασίζεται στην ακρίβεια πρόβλεψης που επιτυγχάνεται στις παρατηρήσεις εκτός δείγματος με βάση το μέσο απόλυτο κλιμακωμένο Σφάλμα (MASE) και εξετάζεται τόσο σε επιμέρους επίπεδα (μοντέλο, μάρκα, κατηγορία, τύπο) όσο και στο συνολικό επίπεδο. Μέσω αυτής της προσέγγισης, αξιολογείται η καταλληλότητα κάθε μεθόδου για την ανάλυση και πρόβλεψη πολύπλοκων και πολυεπίδων δεδομένων όπως οι ταξινομήσεις αυτοκινήτων.

Η αξιολόγηση των προβλέψεων πραγματοποιήθηκε σε τέσσερα επίπεδα ανάλυσης: συνολικό (Total), μάρκας (Make), κατηγορίας (Segment) και μοντέλου (Model). Τα αποτελέσματα ανέδειξαν τη μέθοδο Bottom-Up ως τη γενικά πιο αξιόπιστη προσέγγιση, προσφέροντας τις χαμηλότερες τιμές σφάλματος MASE σε όλα τα επίπεδα όπου εφαρμόστηκε. Αντίστοιχα καλές επιδόσεις παρουσίασαν και οι μέθοδοι βέλτιστης συνάθροισης με τα κατάλληλα βάρη.

Αντίθετα, η μέθοδος Top-Down και ορισμένες παραλλαγές της βέλτιστης συνάθροισης με ακατάλληλα βάρη (*ols*, *nseries*) εμφάνισαν σημαντικά υψηλότερα σφάλματα, καταδεικνύοντας την αδυναμία τους να αποδώσουν με ακρίβεια σε πιο λεπτομερή επίπεδα πρόβλεψης.

Συνολικά, οι προβλέψεις είναι πιο ακριβείς όταν βασίζονται σε ανάλυση χαμηλότερου επιπέδου και ανακατασκευή προς τα πάνω (BU), ενώ οι προβλέψεις από ανώτερο επίπεδο προς τα κάτω (TD) αποτυγχάνουν να αποδώσουν με συνέπεια τη δυναμική των επιμέρους σειρών.

1.2 Δομή Εργασίας

Η εργασία είναι δομημένη σε έξι κεφάλαια, τα οποία παρουσιάζουν με συστηματικό τρόπο την πορεία από την εισαγωγή και το θεωρητικό υπόβαθρο έως την ανάλυση των δεδομένων και την πρόβλεψη.

Το πρώτο κεφάλαιο αποτελεί την εισαγωγή στην εργασία και περιλαμβάνει την παρουσίαση του προβλήματος, την τεκμηρίωση της αναγκαιότητας της μελέτης, τους βασικούς στόχους, τα εργαλεία που χρησιμοποιήθηκαν, τις πηγές δεδομένων και τη συνοπτική παρουσίαση της δομής που ακολουθεί.

Το δεύτερο κεφάλαιο παρουσιάζει μια σύντομη ιστορική ανασκόπηση της αυτοκινητοβιομηχανίας στην Ελλάδα και την Ευρώπη. Σκοπός του είναι να προσφέρει το αναγκαίο ιστορικό και οικονομικό πλαίσιο μέσα στο οποίο εντάσσεται η ανάλυση των ταξινομήσεων και να αναδείξει τη σημασία των δεδομένων στον χρόνο.

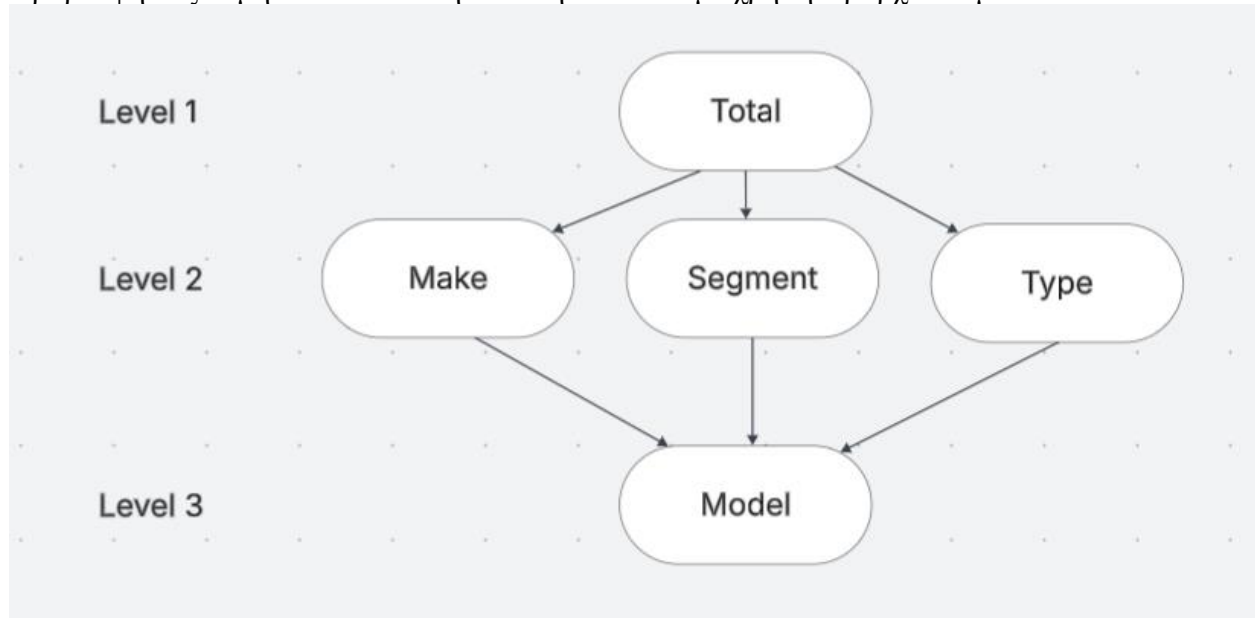
Το τρίτο κεφάλαιο είναι αφιερωμένο στη θεωρητική τεκμηρίωση των μοντέλων πρόβλεψης. Περιγράφονται βασικές έννοιες της θεωρίας των προβλέψεων και της θεωρίας των σφαλμάτων.

Στο τέταρτο κεφάλαιο αναλύεται η συλλογή, οργάνωση και επεξεργασία δεδομένων ταξινομήσεων. Παρουσιάζονται οι τεχνικές επιλογές που έγιναν, οι δυσκολίες που αντιμετωπίστηκαν κατά την επεξεργασία των PDF αρχείων και η τελική μετατροπή των δεδομένων σε αρχεία Excel.

Στο πέμπτο κεφάλαιο παρουσιάζεται αναλυτικότερα η θεωρία των ιεραρχικών προβλέψεων, οι διάφορες μέθοδοι που χρησιμοποιήθηκαν και η γενική χρησιμότητά τους.

Το έκτο κεφάλαιο επικεντρώνεται στη δημιουργία του μοντέλου πρόβλεψης. Γίνεται αναλυτική παρουσίαση του τρόπου με τον οποίο οργανώθηκαν τα δεδομένα, της διαδικασίας εκπαίδευσης του μοντέλου, της εφαρμογής των προβλέψεων και της αξιολόγησής τους.

Το έβδομο και τελευταίο κεφάλαιο περιλαμβάνει τα συνολικά αποτελέσματα της εργασίας. Παρουσιάζονται τα κυριότερα συμπεράσματα που προέκυψαν από την ανάλυση και αξιολογείται η αποτελεσματικότητα της μεθοδολογίας. Επίσης αναφέρονται ενδιαφέρουσες μελλοντικές προεκτάσεις της έρευνας.



Σχήμα 1.1 - Το παραγόμενο ιεραρχικό μοντέλο προβλέψεων

Στο σχήμα, απεικονίζεται η δομή του μοντέλου που δημιουργήθηκε, με πρώτο επίπεδο τις συνολικές ταξινομήσεις, δεύτερο την κατηγοριοποίηση κατά μάρκα, κατηγορία ή τύπο και τρίτο τις ταξινομήσεις σε επίπεδο μοντέλου.

2. Η ιστορία της αυτοκινητοβιομηχανίας στην Ευρώπη και την Ελλάδα (1885–2024)

Η εξέλιξη της αυτοκινητοβιομηχανίας αποτελεί έναν από τους πιο καθοριστικούς παράγοντες της οικονομικής και τεχνολογικής ανάπτυξης των σύγχρονων κοινωνιών. Από την εφεύρεση του πρώτου αυτοκινήτου με κινητήρα εσωτερικής καύσης στα τέλη του 19ου αιώνα, μέχρι την τρέχουσα μετάβαση στην ηλεκτροκίνηση και την αυτόνομη οδήγηση, το αυτοκίνητο δεν επηρέασε μόνο τον τρόπο μετακίνησης, αλλά διαμόρφωσε και τις δομές της βιομηχανίας, της αστικής ανάπτυξης, του εμπορίου και της καθημερινής ζωής.

Η Ευρώπη αποτέλεσε λίκνο της αυτοκινητοβιομηχανίας, με πρωτοπόρες χώρες όπως η Γερμανία, η Γαλλία και η Ιταλία να διαδραματίζουν καίριο ρόλο τόσο στις τεχνολογικές καινοτομίες όσο και στην παγκόσμια εξάπλωση της αυτοκίνησης. Παράλληλα, η ελληνική περίπτωση παρουσιάζει ιδιαίτερο ενδιαφέρον, καθώς η Ελλάδα, χωρίς εγχώρια βαριά αυτοκινητοβιομηχανία, ανέπτυξε κυρίως εμπορικές και εισαγωγικές δραστηριότητες, οι οποίες αντανakλούν τις κοινωνικές, πολιτικές και οικονομικές μεταβολές του 20ού και 21ου αιώνα.

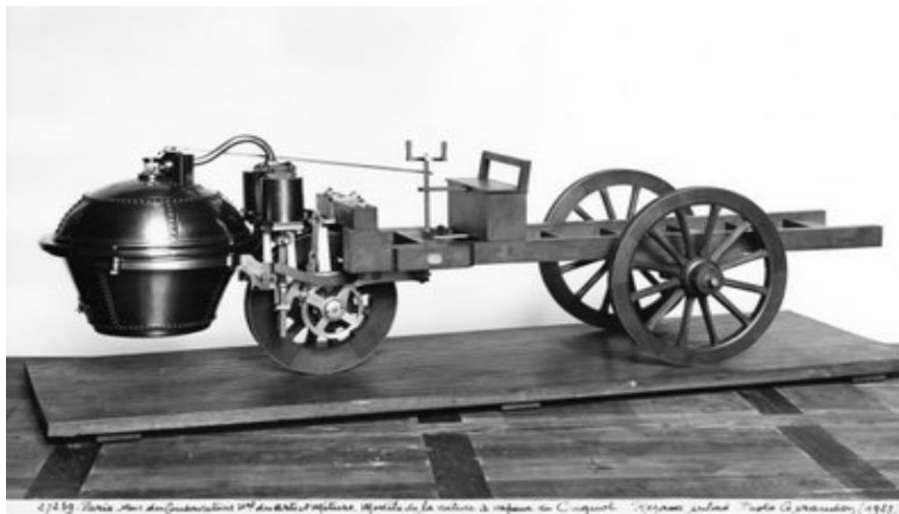
2.1 Ευρώπη

2.1.1 Οι απαρχές της αυτοκίνησης (1885–1900)

Η ιδέα της αυτοκινούμενης άμαξας γεννήθηκε στην Ευρώπη στα τέλη του 18ου και στις αρχές του 19ου αιώνα. Πρωτοπόροι μηχανικοί, όπως ο Nicolas-Joseph Cugnot στη Γαλλία, κατασκεύασαν ατμοκίνητα οχήματα ήδη από το 1769 [[1][1]].

Ωστόσο, η πραγματική απαρχή του αυτοκινήτου εντοπίζεται στα τέλη του 19ου αιώνα, όταν το 1885 ο Karl Benz (Γερμανία) πατεντάρει το πρώτο αμάξι με κινητήρα εσωτερικής καύσης (Patent-Motorwagen). Παράλληλα, ο Gottlieb Daimler και ο Wilhelm Maybach κατασκευάζουν το πρώτο τετράτροχο αυτοκίνητο το 1886 [[1][1]].

Αυτές οι εξελίξεις έθεσαν τη Γερμανία ως το επίκεντρο της πρώιμης αυτοκινητοβιομηχανίας στην Ευρώπη.



Σχήμα 2.1 Το πρώτο αυτοκινούμενο αυτοκίνητο του Nicolas-Joseph Cugnot.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

2.1.2 Η εδραίωση της βιομηχανίας (1900–1945)

Ιδρύθηκαν αρκετές εταιρίες που συνεχίζουν να υπάρχουν μέχρι σήμερα, όπως οι FIAT (1899), Renault (1899), Peugeot, Citroën (1919) [[2]][3]. Η παραγωγή επεκτείνεται και βιομηχανοποιείται: το 1913, η Ford (ΗΠΑ) εισάγει τη γραμμή παραγωγής μαζικής συναρμολόγησης, επηρεάζοντας και τις ευρωπαϊκές βιομηχανίες [3]. Στον Β' Παγκόσμιο Πόλεμο, πολλές ευρωπαϊκές αυτοκινητοβιομηχανίες στράφηκαν στην παραγωγή στρατιωτικών οχημάτων [2].

2.1.3 Η μεταπολεμική άνθηση και η χρυσή εποχή (1945–1975)

Η οικονομική ανάκαμψη, η αστικοποίηση και η αύξηση της μεσαίας τάξης μετά τον Β' Παγκόσμιο Πόλεμο επέφεραν εκρηκτική άνοδο της αυτοκίνησης. Άρχισαν να παράγονται οικονομικά αυτοκίνητα: Renault 4CV, Fiat 500, Citroën 2CV. Ο σκαραβαίος της VW γίνεται το πιο εμπορικό μοντέλο όλων των εποχών [2][4].

2.1.4 Περίοδος κρίσης και προσαρμογής (1975–2000)

Η περίοδος από το 1975 έως το 2000 χαρακτηρίστηκε από σοβαρές προκλήσεις για την παγκόσμια αυτοκινητοβιομηχανία, κυρίως εξαιτίας των πετρελαϊκών κρίσεων της δεκαετίας του 1970, οι οποίες επηρέασαν αρνητικά τόσο την παραγωγή όσο και τη ζήτηση αυτοκινήτων. Παράλληλα, ενισχύθηκε σημαντικά ο ανταγωνισμός από τις ιαπωνικές αυτοκινητοβιομηχανίες, όπως η Toyota, η Honda και η Nissan, οι οποίες προσέφεραν αξιόπιστα και αποδοτικά οχήματα. [[6]].

Οι εξελίξεις αυτές ανάγκασαν τις αυτοκινητοβιομηχανίες να προσαρμοστούν στις νέες συνθήκες της αγοράς.

Ένα από τα βασικά χαρακτηριστικά της περιόδου ήταν η στροφή των καταναλωτών και των εταιρειών σε μικρότερα και οικονομικότερα αυτοκίνητα, λόγω της αύξησης των τιμών των καυσίμων.

Την ίδια εποχή ξεκίνησαν και οι πρώτες σημαντικές συγχωνεύσεις στον ευρωπαϊκό χώρο, όπως η ένωση της Peugeot με τη Citroën για τη δημιουργία του ομίλου PSA Group, καθώς και η ενσωμάτωση της Audi στον όμιλο της Volkswagen. [[3]][[6]]

Τέλος, άρχισαν να εφαρμόζονται πιο αυστηροί περιβαλλοντικοί κανονισμοί, γεγονός που οδήγησε σε αυξημένο ενδιαφέρον για τη μείωση της κατανάλωσης καυσίμου και των εκπομπών ρύπων, προαναγγέλλοντας μια νέα εποχή τεχνολογικών καινοτομιών στον τομέα της αυτοκίνησης [[5]].

2.1.5 Μετάβαση στον 21ο αιώνα: Παγκοσμιοποίηση και τεχνολογική μετάβαση (2000–2016)

Από τις αρχές του 21ου αιώνα, η παγκόσμια αυτοκινητοβιομηχανία εισήλθε σε μια νέα φάση, χαρακτηριζόμενη από τη ραγδαία παγκοσμιοποίηση, τη συγκέντρωση της παραγωγής, τις τεχνολογικές καινοτομίες και την αυξανόμενη ευαισθησία απέναντι στο περιβάλλον [[7]].

Κατά τη διάρκεια αυτής της περιόδου, οι μεγάλες αυτοκινητοβιομηχανίες ενίσχυσαν τις διεθνείς συνεργασίες τους και ενοποίησαν τις γραμμές παραγωγής, με στόχο την εξοικονόμηση κόστους και τη μεγιστοποίηση της αποτελεσματικότητας. Η παγκόσμια αγορά άνοιξε περισσότερο προς την Κίνα, την Ινδία και άλλες αναδυόμενες οικονομίες, γεγονός που μετέβαλε τη γεωγραφία της ζήτησης και της παραγωγής οχημάτων [[6]].

Η τεχνολογική πρόοδος επιταχύνθηκε σημαντικά, με την εισαγωγή προηγμένων ηλεκτρονικών συστημάτων ασφαλείας (όπως ABS, ESP, αισθητήρες παρκαρίσματος και κάμερες), καθώς και με την εξέλιξη της συνδεσιμότητας μέσω GPS, Bluetooth και αργότερα των συστημάτων infotainment. Τα υβριδικά αυτοκίνητα άρχισαν να κερδίζουν έδαφος, με πρωτοπόρους κατασκευαστές όπως η Toyota με το Prius, σηματοδοτώντας την έναρξη της ηλεκτροκίνησης σε ευρύτερη κλίμακα [[7]].

Ταυτόχρονα, ενισχύθηκαν περαιτέρω οι περιβαλλοντικές απαιτήσεις και οι κανονισμοί εκπομπών, κυρίως στην Ευρώπη, ωθώντας τις αυτοκινητοβιομηχανίες σε επενδύσεις για καθαρότερες τεχνολογίες και πιο αποδοτικούς κινητήρες. Η αποδοτικότητα σε καύσιμο, οι μειωμένες εκπομπές CO₂ και η εναρμόνιση με πρότυπα όπως τα Euro 4, Euro 5 και Euro 6 έγιναν κεντρικοί στόχοι του κλάδου [[5]].

Το 2008 ξέσπασε η παγκόσμια χρηματοπιστωτική κρίση, η οποία είχε άμεσες επιπτώσεις στην αυτοκινητοβιομηχανία. Πτώση στις πωλήσεις, οικονομικά προβλήματα και κρατικές παρεμβάσεις (ιδίως στις ΗΠΑ με τη διάσωση της General Motors και της Chrysler) ανέδειξαν τις δομικές αδυναμίες του τομέα. Παράλληλα, η κρίση οδήγησε σε νέες συγχωνεύσεις και στρατηγικές συμμαχίες, όπως η συμμαχία Renault-Nissan με την Mitsubishi [[6]].

Μέχρι το 2016, η αυτοκινητοβιομηχανία είχε αρχίσει να ανακάμπτει δυναμικά, θέτοντας πλέον στο επίκεντρο της στρατηγικής της, την ηλεκτροκίνηση, την αυτόνομη οδήγηση και τις ψηφιακές υπηρεσίες. Η περίοδος αυτή έκλεισε με την αυξανόμενη επιρροή τεχνολογικών εταιρειών (όπως η Google και η Tesla) στον χώρο της κινητικότητας, προαναγγέλλοντας μια νέα εποχή για την παγκόσμια αυτοκίνηση [[8]].

2.1.6 Η πράσινη επανάσταση και η νέα εποχή (2016–2024)

Από το 2016 έως σήμερα, η παγκόσμια αυτοκινητοβιομηχανία βρίσκεται σε μια φάση βαθιάς μεταμόρφωσης, καθώς αντιμετωπίζει νέες τεχνολογικές, περιβαλλοντικές και κοινωνικές προκλήσεις. Στο επίκεντρο της αλλαγής βρίσκονται τέσσερις βασικές τάσεις: η συνδεσιμότητα, η αυτόνομη οδήγηση, η κοινή χρήση των οχημάτων και η ηλεκτροκίνηση.

Η ηλεκτροκίνηση έχει εξελιχθεί στον κυρίαρχο άξονα ανάπτυξης για τις περισσότερες αυτοκινητοβιομηχανίες. Η Tesla έπαιξε καταλυτικό ρόλο [[8]], επιταχύνοντας την παγκόσμια στροφή προς τα ηλεκτρικά οχήματα, ενώ παραδοσιακοί κατασκευαστές όπως η Volkswagen, η Ford και η General Motors επαναπροσδιορίζουν τη στρατηγική τους, ανακοινώνοντας σταδιακή κατάργηση των κινητήρων εσωτερικής καύσης τις επόμενες δεκαετίες. [[9]]. Παράλληλα, νέοι κατασκευαστές από την Κίνα, όπως η NIO και η BYD, αποκτούν ολοένα και μεγαλύτερο μερίδιο αγοράς [0]. Οι κρατικές πολιτικές, τα προγράμματα επιδοτήσεων και οι στόχοι για την κλιματική ουδετερότητα ενισχύουν αυτή τη μετάβαση, ιδίως στην Ευρωπαϊκή Ένωση.

Ταυτόχρονα, η τεχνολογική πρόοδος επέτρεψε την ανάπτυξη της αυτόνομης οδήγησης. Αν και πλήρως αυτόνομα οχήματα δεν έχουν ακόμη τεθεί σε μαζική κυκλοφορία, πολλά μοντέλα διαθέτουν ήδη προηγμένα συστήματα υποβοήθησης οδηγού. Η χρήση αισθητήρων, ραντάρ, καμερών και τεχνητής νοημοσύνης επιτρέπει τη σταδιακή εφαρμογή λύσεων αυτόνομης μετακίνησης. Παράλληλα, η συνεχής σύνδεση των οχημάτων με το διαδίκτυο δίνει τη δυνατότητα για αναβαθμίσεις λογισμικού από απόσταση, έξυπνες πλοηγήσεις, εφαρμογές τηλεματικής και ενσωμάτωση ψηφιακών υπηρεσιών [[8]][[10]].

Η έννοια της ιδιοκτησίας μεταβάλλεται επίσης. Οι νεότερες γενιές δείχνουν μεγαλύτερη προτίμηση σε λύσεις από παρόχους υπηρεσιών κινητικότητας, προσαρμοζόμενες στις νέες ανάγκες της αγοράς [[10]].

Το 2020 η πανδημία COVID-19, η γεωπολιτική κατάσταση και ο αυξανόμενος πληθωρισμός, έφεραν επιπλέον προκλήσεις με σοβαρές επιπτώσεις στην παραγωγή και τη ζήτηση. Η παγκόσμια έλλειψη ημιαγωγών και τα προβλήματα στην εφοδιαστική αλυσίδα προκάλεσαν καθυστερήσεις και αυξήσεις τιμών. Παρ' όλα αυτά παράλληλα, η πανδημία επιτάχυνε και την ψηφιοποίηση των πωλήσεων και την προσαρμογή σε νέες καταναλωτικές συνήθειες [[9]][[11]].

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
Σήμερα, η αυτοκινητοβιομηχανία κινείται σε ένα δυναμικό περιβάλλον, όπου η καινοτομία, η βιωσιμότητα και η τεχνολογία καθορίζουν την πορεία της. Οι εταιρείες καλούνται να διαχειριστούν την ενεργειακή μετάβαση, να αντιμετωπίσουν προκλήσεις που σχετίζονται με την εξόρυξη πρώτων υλών (όπως το λίθιο για τις μπαταρίες) και να προσαρμοστούν στις κλιματικές πολιτικές που θέτουν φιλόδοξους στόχους για το μηδενικό αποτύπωμα άνθρακα. Η επόμενη δεκαετία διαγράφεται κρίσιμη, καθώς το αυτοκίνητο μετατρέπεται σταδιακά από απλό μέσο μεταφοράς σε ένα πολύπλοκο, διασυνδεδεμένο και βιώσιμο προϊόν, που ενσωματώνει τις εξελίξεις στην τεχνητή νοημοσύνη, την ψηφιακή τεχνολογία και την πράσινη ενέργεια [[10]].



Σχήμα 2.2 Αυτόνομο ταξί της Waymo One στο San Francisco.

Η θυγατρική εταιρεία της Google, είναι η πρώτη που έλαβε άδεια να προσφέρει υπηρεσία αυτόνομης οδήγησης το 2019 στην Αμερική.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
Πίνακας 2.1 Πωλήσεις ηλεκτρονικών αυτοκινήτων στην Ευρωπαϊκή Ένωση

Έτος	Πωλήσεις EV στην ΕΕ	Σύνολο Οχημάτων (εκ.)	Ποσοστό EV
2015	~150.000	13,7	1,1%
2020	~740.000	11,6	6,4%
2023	1.630.000	11,2	14,6%

Στον πίνακα φαίνεται καθαρά η τάση αύξησης του ποσοστού των ηλεκτρικών οχημάτων στις συνολικές ταξινομήσεις πανευρωπαϊκά. Το συγκεκριμένο ποσοστό από 1,1 % το 2015 αυξάνεται σε 14,6 % το 2023.

2.2 Ελλάδα

2.2.1 Τα πρώτα βήματα (1900–1940)

Η Ελλάδα εισήλθε στον κόσμο του αυτοκινήτου αρκετά νωρίς, με την πρώτη κυκλοφορία οχήματος να σημειώνεται στις αρχές του 20ού αιώνα [[1]]. Μέχρι το 1920, τα αυτοκίνητα ήταν κυρίως πολυτέλεια των πλουσίων και αριθμούσαν μερικές εκατοντάδες [[13]]. Το 1920 ιδρύθηκε το Υπουργείο Συγκοινωνιών και θεσπίστηκαν οι πρώτοι κανόνες κυκλοφορίας [[13]]. Οι εισαγωγές γινόντουσαν κυρίως από τη Γαλλία, την Αγγλία και τη Γερμανία, με ελάχιστες ποσότητες [[13]].



Σχήμα 2.3 Το πρώτο αυτοκίνητο στην Ελλάδα.

2.2.2 Μεταπολεμική ανάπτυξη και προσπάθεια για ελληνική παραγωγή (1950–1975)

Μετά το τέλος του Β' Παγκοσμίου Πολέμου και τον Εμφύλιο Πόλεμο, η ελληνική οικονομία άρχισε σταδιακά να ανακάμπτει [13]. Η ανάκαμψη αυτή συνοδεύτηκε από αύξηση του βιοτικού επιπέδου και της ζήτησης για μέσα μεταφοράς, ιδιαίτερα από τη δεκαετία του 1960 και έπειτα. Οι εισαγωγές αυτοκινήτων σημείωσαν σημαντική άνοδο εκείνη την περίοδο, αντανακλώντας την αλλαγή των κοινωνικών και οικονομικών συνθηκών [2], [13].

Μέσα σε αυτό το κλίμα, έγιναν αρκετές προσπάθειες για την ανάπτυξη εγχώριας αυτοκινητοβιομηχανίας, με στόχο την κάλυψη μέρους της ζήτησης και την ενίσχυση της

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων βιομηχανικής παραγωγής της χώρας. Η πιο γνωστή τέτοια προσπάθεια ήταν εκείνη της NAMCO (National Motor Company), η οποία παρήγαγε το Pony, ένα μικρό και ιδιαίτερα ανθεκτικό αυτοκίνητο, βασισμένο σε εξαρτήματα της Citroën [[3]], [[14]]. Το Pony κατέκτησε τη φήμη του ως ένα πρακτικό και οικονομικό όχημα, κατάλληλο για τις ανάγκες της ελληνικής υπαίθρου [[14]]. Παράλληλα, η MAVA ανέλαβε τη συναρμολόγηση μοντέλων της Renault, όπως το Renault 4, ενώ η ΕΛΒΟ (Ελληνική Βιομηχανία Οχημάτων) εστίασε κυρίως στην κατασκευή στρατιωτικών οχημάτων και λεωφορείων [[3]], [[14]].

Ωστόσο, παρά τις φιλόδοξες αυτές πρωτοβουλίες, η εγχώρια αυτοκινητοβιομηχανία δεν κατάφερε να αναπτυχθεί σε σημαντικό βαθμό. Η περιορισμένη εσωτερική αγορά, το υψηλό κόστος παραγωγής, η έλλειψη τεχνογνωσίας και η απουσία ουσιαστικής κρατικής στήριξης αποτέλεσαν σοβαρά εμπόδια [[14]], [[15]]. Οι ελληνικές βιομηχανίες δεν μπόρεσαν να ανταγωνιστούν τις μεγάλες διεθνείς αυτοκινητοβιομηχανίες, οι οποίες διέθεταν ισχυρότερη τεχνολογική βάση, μεγαλύτερη παραγωγική δυνατότητα και οικονομίες κλίμακας [[5]], [[6]].

Παρά τις αντιξοότητες, η περίοδος αυτή παραμένει σημαντική για την ιστορία της ελληνικής βιομηχανίας, καθώς ανέδειξε την πρόθεση και την επιθυμία για εγχώρια τεχνολογική ανάπτυξη στον τομέα της αυτοκίνησης.



Σχήμα 2.4 Το εν μέρει παραγόμενο στην Ελλάδα, Pony της NAMCO.

2.2.3 Η έκρηξη της αγοράς και οι εισαγωγές (1980–2009)

Η ένταξη της Ελλάδας στην Ευρωπαϊκή Οικονομική Κοινότητα (ΕΟΚ) το 1981 αποτέλεσε καθοριστικό σημείο καμπής για την ελληνική αγορά αυτοκινήτου [[2]], [[13]]. Η ελευθερία διακίνησης εμπορευμάτων ενίσχυσε ραγδαία τις εισαγωγές οχημάτων, ενώ η εγχώρια βιομηχανική παραγωγή – ήδη σε καθοδική πορεία – άρχισε να υποχωρεί πλήρως [[13]], [[14]]. Οι καταναλωτές απέκτησαν πρόσβαση σε ποικιλία επιλογών από διεθνείς κατασκευαστές, οδηγώντας σε εκρηκτική ανάπτυξη της αγοράς.

Η δεκαετία του 1980 σηματοδότησε την απαρχή μιας εντυπωσιακής ανόδου στις πωλήσεις αυτοκινήτων. Από περίπου 60.000 ταξινομήσεις ετησίως το 1980, η ελληνική αγορά έφτασε τις 250.000 ετήσιες πωλήσεις μέχρι το 2000 [[12]]. Η δεκαετία του 1990 και η αρχή της δεκαετίας του 2000 χαρακτηρίστηκαν από έντονη καταναλωτική ζήτηση, σημαντική αύξηση του στόλου Ι.Χ. στους ελληνικούς δρόμους και διαρκή ενίσχυση των πωλήσεων.

Το 2004, χρονιά των Ολυμπιακών Αγώνων της Αθήνας και κορύφωσης της οικονομικής ευφορίας, σημειώθηκε ρεκόρ πωλήσεων αυτοκινήτων. Η κατανάλωση έφτασε σε ιστορικά υψηλά, ενισχυμένη και από την εύκολη πρόσβαση σε δανεισμό. Ωστόσο, παρά την αυξανόμενη ζήτηση, η Ελλάδα δεν παρήγαγε πλέον οχήματα. Όλες οι ανάγκες της αγοράς καλύπτονταν εξ ολοκλήρου από εισαγωγές, υπογραμμίζοντας την πλήρη εξάρτηση της χώρας από τις διεθνείς αγορές [[15]].

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
Η περίοδος 1980–2009, επομένως, χαρακτηρίστηκε από ραγδαία αύξηση στην κατανάλωση αυτοκινήτων, απόσυρση της εγχώριας παραγωγής και απόλυτη κυριαρχία των εισαγόμενων οχημάτων [[14]].

2.2.4 Η κρίση και η πτώση των πωλήσεων (2010–2015)

Η περίοδος της οικονομικής κρίσης, που ξεκίνησε το 2009 και κορυφώθηκε τα επόμενα χρόνια, είχε καταστροφικές συνέπειες για την ελληνική αγορά αυτοκινήτου [[15]]. Ο κλάδος επλήγη βαθιά, καθώς η αγοραστική δύναμη των καταναλωτών μειώθηκε δραματικά και οι δυνατότητες δανεισμού περιορίστηκαν. Ως αποτέλεσμα, οι πωλήσεις καινούργιων αυτοκινήτων κατέρρευσαν.

Συγκεκριμένα, σε σύγκριση με το 2007, έτος κατά το οποίο οι νέες ταξινομήσεις έφτασαν τις 280.000, το 2012 καταγράφηκαν μόλις 57.000 νέες ταξινομήσεις, σημειώνοντας πτώση της τάξης του 70% [[15]]. Το σοκ στην αγορά ήταν έντονο και οι εταιρείες προσαρμόστηκαν, στρεφόμενες κυρίως στην προσφορά πιο οικονομικών μοντέλων, αλλά και στην αξιοποίηση της αυξανόμενης ζήτησης για μεταχειρισμένα οχήματα.

Κατά την περίοδο αυτή, υπήρξε ραγδαία αύξηση στις εισαγωγές μεταχειρισμένων αυτοκινήτων, κυρίως από χώρες όπως η Γερμανία, η Ολλανδία και το Βέλγιο [[14]]. Οι καταναλωτές αναζητούσαν φθηνότερες λύσεις μετακίνησης και απέφευγαν την αγορά καινούργιου αυτοκινήτου λόγω του υψηλού κόστους απόκτησης και της αβεβαιότητας που επικρατούσε στην οικονομία.

Η εποχή αυτή σηματοδότησε τη μετάβαση από την εποχή της καταναλωτικής ευφορίας σε μια περίοδο έντονης συρρίκνωσης της αγοράς και αλλαγής προτύπων κατανάλωσης στον τομέα του αυτοκινήτου.

2.2.5 Ανάκαμψη και πράσινη μετάβαση (2016–2024)

Μετά την πολυετή κρίση, η ελληνική αγορά αυτοκινήτου άρχισε να παρουσιάζει σημάδια ανάκαμψης, κυρίως μετά το 2016 [[13]]. Παρά τη θετική πορεία, οι πωλήσεις δεν κατάφεραν να επιστρέψουν στα προ κρίσης επίπεδα. Η αγορά παρέμεινε συγκρατημένη, με νέες τάσεις να διαμορφώνουν τον κλάδο.

Μία από τις βασικότερες εξελίξεις ήταν η σημαντική εξάπλωση των προγραμμάτων leasing και η ενίσχυση των εταιρικών πωλήσεων, καθώς μεγάλο μέρος των νέων ταξινομήσεων αφορούσε πλέον εταιρικούς στόλους [[11]]. Παράλληλα, συνεχίστηκε και εντάθηκε η εισαγωγή μεταχειρισμένων οχημάτων από ευρωπαϊκές χώρες [[13]].

Σημαντική εξέλιξη αποτέλεσε η έναρξη της πράσινης μετάβασης. Τα πρώτα βήματα της ηλεκτροκίνησης έγιναν αισθητά, με την εμφάνιση κινήτρων για την αγορά ηλεκτρικών οχημάτων και τη σταδιακή δημιουργία υποδομών φόρτισης, κυρίως στα μεγάλα αστικά κέντρα [[9]]. Παράλληλα, αυξήθηκε κατακόρυφα το ενδιαφέρον των καταναλωτών για SUV και crossover μοντέλα, τα οποία κυριάρχησαν στις νέες ταξινομήσεις της Ελληνικής αγοράς.

Σύμφωνα με στοιχεία του Συνδέσμου Εισαγωγέων Αντιπροσώπων Αυτοκινήτων (ΣΕΑΑ), το 2023 καταγράφηκαν 123.000 νέες ταξινομήσεις Ι.Χ., ενώ οι εισαγωγές μεταχειρισμένων άγγιξαν τις 90.000 [[15]]. Οι αριθμοί αυτοί επιβεβαιώνουν ότι, αν και η αγορά δείχνει δυναμισμό, η εξάρτηση από τις εισαγωγές παραμένει απόλυτη.

2.2.6 Πρόσφατα δεδομένα και προκλήσεις (2024–2025)

Καθώς η ελληνική αγορά εισέρχεται στη νέα εποχή, οι προκλήσεις παραμένουν πολλές και σύνθετες. Η εισαγωγή παραμένει ο μοναδικός τρόπος τροφοδότησης της αγοράς, με τα περισσότερα μεταχειρισμένα οχήματα να προέρχονται από τη Γερμανία, την Ιταλία και την Ολλανδία [0].

Μία από τις βασικότερες προκλήσεις είναι η γήρανση του στόλου. Ο μέσος όρος ηλικίας των οχημάτων που κυκλοφορούν ξεπερνά τα 11 έτη, κάτι που δημιουργεί προβλήματα ασφάλειας, κατανάλωσης καυσίμου και εκπομπών ρύπων [[5]]. Ταυτόχρονα, η μετάβαση στα ηλεκτρικά οχήματα συναντά εμπόδια, όπως η απουσία αξιόπιστης υποδομής φόρτισης στην επαρχία, η φορολογική αστάθεια και οι υψηλοί δασμοί στα εισαγόμενα ηλεκτρικά οχήματα παρά τις προσπάθειες του κράτους για ελάφρυνση τους. [[17]]

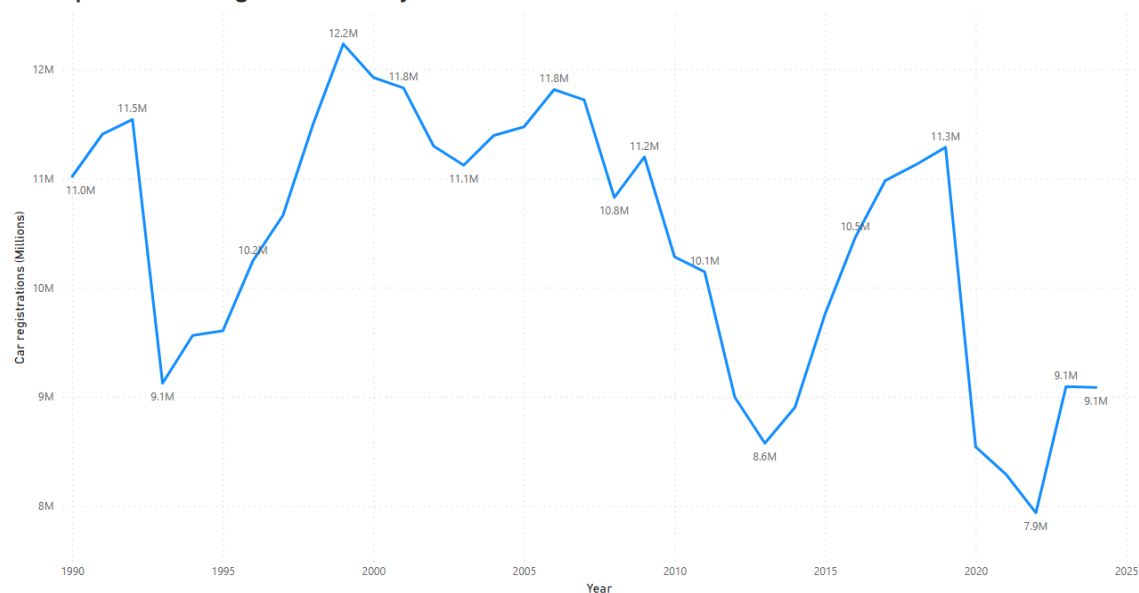
2.3 Σύνοψη

Η ιστορία της αυτοκινητοβιομηχανίας στην Ευρώπη αντικατοπτρίζει την εξέλιξη της ίδιας της ευρωπαϊκής βιομηχανικής κουλτούρας, από την εφεύρεση στη μαζική παραγωγή, την οικονομική ανασυγκρότηση και την πράσινη μετάβαση. Η επόμενη δεκαετία θα κρίνει αν η Ευρώπη θα διατηρήσει την παγκόσμια ηγετική της θέση στη νέα εποχή της ηλεκτροκίνησης και της ψηφιακής μετακίνησης.

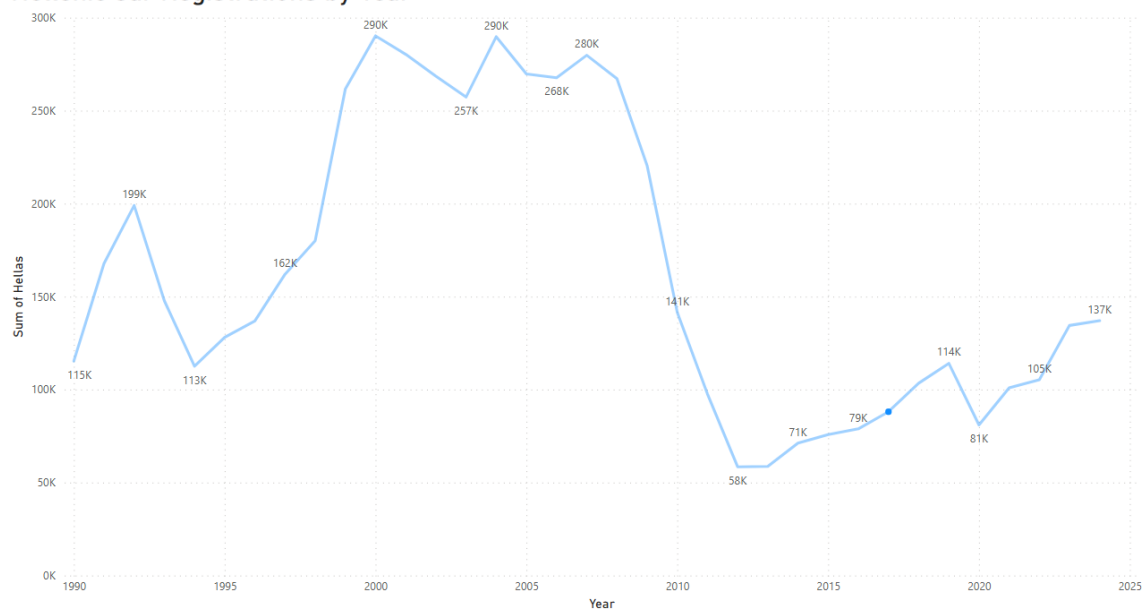
Η Ελλάδα δεν κατάφερε ποτέ να αναπτύξει έναν ισχυρό βιομηχανικό τομέα στον κλάδο του αυτοκινήτου, βασιζόμενη σχεδόν αποκλειστικά στις εισαγωγές. Παρά τις αποσπασματικές προσπάθειες εγχώριας παραγωγής τις προηγούμενες δεκαετίες, η περιορισμένη εσωτερική αγορά, η απουσία τεχνογνωσίας και η κρατική αδράνεια δεν επέτρεψαν την εδραίωση ενός βιώσιμου παραγωγικού μοντέλου.

Οι κρίσεις και η οικονομική αστάθεια επηρέασαν σημαντικά την κατανάλωση και τη διαμόρφωση προτεραιοτήτων στην αγορά. Σήμερα, ο κλάδος του αυτοκινήτου στην Ελλάδα χαρακτηρίζεται από την κυριαρχία των μεταχειρισμένων, την ανάπτυξη του leasing και τα πρώτα, αλλά διστακτικά βήματα προς την ηλεκτροκίνηση.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων European Car Registrations by Year



Hellenic Car Registrations by Year



Σχήμα 2.5 Ευρωπαϊκές και Ελληνικές Ταξινομήσεις

Στα δύο διαγράμματα του σχήματος 2.5 συγκρίνεται χρονολογικά ή πορεία των Ελληνικών ταξινομήσεων σε σχέση με τις Ευρωπαϊκές. Είναι φανερή η έκρηξη των ταξινομήσεων στην Ελλάδα μετά το 1995, η απότομη πτώση τους από το 2011 λόγω της οικονομικής κρίσης και η σταδιακή ανάκαμψη τους από το 2017. Η ανάκαμψη ήρθε αργοπορημένα σε σχέση με την υπόλοιπη Ευρώπη όπου ξεκίνησε να επανέρχεται το 2012. Οι δύο χρονοσειρές βέβαια δεν είναι άμεσα συγκρίσιμες οπτικά καθώς σε επίπεδο Ευρώπης, οι ταξινομήσεις έπεσαν κατά ~ 30% λόγω

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων της κρίσης ενώ στην Ελλάδα κατά ~ 80%. Επίσης μπορούμε να δούμε τόσο σε Ελλάδα όσο και σε Ευρώπη, την επιρροή της παγκόσμιας πανδημίας στις ταξινομήσεις του 2020. Αξίζει να παρατηρήσουμε την αυξητική τάση τα τελευταία τέσσερα χρόνια στην Ελλάδα αντίθετα με την Ευρώπη που δεν έχει επιστρέψει στα επίπεδα προ πανδημίας.

Κεφάλαιο 3. Θεωρία προβλέψεων

3.1 Εισαγωγικές έννοιες

3.1.1 Προβλέψεις

Με τον όρο πρόβλεψη αναφερόμαστε στη διαδικασία συγκέντρωσης διαθέσιμης ιστορικής πληροφορίας και αξιοποίηση αυτής με σκοπό την παραγωγή όσο το δυνατόν ακριβέστερων αποτελεσμάτων για κάποια συγκεκριμένη χρονική στιγμή στο μέλλον, που ακόμα δεν έχουν παρατηρηθεί [[31]]. Λέγοντας μοντέλο πρόβλεψης εννοούμε τη διαδικασία που ακολουθείται για την δημιουργία προβλέψεων. Υπάρχουν πολλά μοντέλα πρόβλεψης, κάθε ένα από τα οποία αντιστοιχεί και σε κάποια τεχνική.

Στον κόσμο των επιχειρήσεων η πρόβλεψη παίζει καθοριστικό ρόλο, καθώς η εκτίμηση που γίνεται σήμερα για τα αναμενόμενα μελλοντικά αποτελέσματα, επηρεάζει τις στρατηγικές αποφάσεις τους. Δεδομένου, ότι το μέλλον χαρακτηρίζεται από έντονη αβεβαιότητα δημιουργείται η ανάγκη στις επιχειρήσεις για παραγωγή καλών εκτιμήσεων ώστε να λαμβάνονται οι ορθές αποφάσεις προς το συμφέρον τους. Για το λόγο αυτό, η προσπάθεια για εύρεση της καταλληλότερης μεθόδου πρόβλεψης πρέπει να είναι μία συνεχής διαδικασία που θα βασίζεται στη σωστή έρευνα και παρακολούθηση.

Η επιλογή της καταλληλότερης μεθόδου πρόβλεψης έχει να κάνει και με τη διαθεσιμότητα των δεδομένων. Πολλές φορές μπορεί να μην έχουμε στη διάθεσή μας επαρκή ιστορικά δεδομένα ή μπορεί τα δεδομένα αυτά να μην ακολουθούν ένα σχετικά σταθερό μοτίβο ώστε να μπορούμε να κάνουμε υποθέσεις για το μέλλον. Στην περίπτωση αυτή, η πρόβλεψη βασίζεται κυρίως σε ποιοτικές μεθόδους. Οι ποιοτικές μέθοδοι δεν αποτελούν απλώς εικασίες. Στηρίζονται στη συσσωρευμένη γνώση, εμπειρία, κριτική ικανότητα αλλά ακόμα και τη διαίσθηση του αναλυτή ενώ πολλές φορές η πρόβλεψη δεν πραγματοποιείται μόνο από ένα άτομο αλλά μπορεί να είναι αποτέλεσμα συνδυασμού πολλών απόψεων.

Όταν υπάρχουν επαρκή δεδομένα μέσα από τα οποία μπορούμε να εξάγουμε χρήσιμη πληροφορία για το πώς θα μπορούσαν αυτά να εξελιχθούν μελλοντικά, τότε μπορούμε να εφαρμόσουμε κάποια ποσοτική μέθοδο πρόβλεψης. Οι ποσοτικές μέθοδοι πρόβλεψης διακρίνονται κυρίως στα αιτιοκρατικά ή επεξηγηματικά μοντέλα και τα μοντέλα χρονοσειρών. Στο αιτιοκρατικό μοντέλο η βασική υπόθεση είναι ότι υπάρχει μία σταθερή σχέση μεταξύ της εξαρτημένης μεταβλητής, δηλαδή του μεγέθους που είναι υπό πρόβλεψη και ορισμένων ανεξάρτητων μεταβλητών που το επηρεάζουν. Στην παρούσα εργασία γίνεται χρήση μεθόδων που αφορούν αποκλειστικά το μοντέλο χρονοσειρών. Για το λόγο αυτό δεν θα αναπτυχθεί περαιτέρω το κομμάτι των αιτιοκρατικών μοντέλων.

Το μοντέλο των χρονοσειρών βασίζεται στην υπόθεση ότι η τιμή του υπό πρόβλεψης μοντέλου ακολουθεί ένα συγκεκριμένο σταθερό πρότυπο συμπεριφοράς το οποίο επαναλαμβάνεται στο χρόνο. Η διαδικασία της πρόβλεψης στηρίζεται στην αναγνώριση αυτού του προτύπου και την

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων προέκτασή του στο μέλλον. Αυτό σημαίνει ότι ένα μοντέλο χρονοσειρών βασίζεται στα ιστορικά δεδομένα και θεωρεί ότι η συμπεριφορά των δεδομένων θα συνεχίσει να είναι ίδια και στο μέλλον. Στις μεθόδους που περιγράφονται με το μοντέλο χρονοσειρών συγκαταλέγονται η αποσύνθεση, η εξομάλυνση και οι αυτοπαλινδρομικές μέθοδοι κινητού μέσου όρου.

3.1.2 Χρονοσειρές

Παραπάνω έγινε αναφορά στις χρονοσειρές. Η χρονοσειρά (ή χρονολογική σειρά) είναι μία ακολουθία διαδοχικών παρατηρήσεων που είναι εξαρτημένες μεταξύ τους. Πρόκειται δηλαδή, για ιστορικά δεδομένα που παρουσιάζονται με σταθερή συχνότητα. Τα δεδομένα μπορεί να εμφανιστούν σε διάφορες μορφές χρονικά. Πιο συγκεκριμένα μπορεί, να έχουμε δεδομένα υψηλής συχνότητας (ωριαία, ημερήσια, εβδομαδιαία) ή χαμηλής συχνότητας (τριμηνιαία, μηνιαία, ετήσια). Συνήθως, αυτό προκύπτει από το σημείο που λαμβάνουμε την απόφαση.

Σε μία χρονοσειρά μπορούμε να αναγνωρίσουμε πολλά χαρακτηριστικά. Τα βασικά ποιοτικά χαρακτηριστικά με τα οποία ασχολούνται οι παραδοσιακές μέθοδοι ανάλυσης είναι: η τάση, η κυκλικότητα, η εποχιακότητα και οι μη κανονικές διακυμάνσεις.

Ως τάση, περιγράφεται η μεταβολή του μέσου επιπέδου των τιμών μιας χρονολογικής σειράς μακροπρόθεσμα. Για να μπορούμε να εντοπίσουμε την ύπαρξη ή μη τάσης στα δεδομένα μιας χρονοσειράς, θα πρέπει να έχουμε στη διάθεσή μας έναν ικανοποιητικό αριθμό ιστορικών παρατηρήσεων. Η τάση δείχνει τη γενικότερη κλίση μιας χρονοσειράς και μπορεί να είναι ανοδική, πτωτική ή σταθερού επιπέδου.

Η εποχιακότητα εκφράζει τις περιοδικές διακυμάνσεις μιας χρονοσειράς που έχουν σταθερό μήκος και διάρκεια μικρότερη του έτους. Ο εντοπισμός της εποχιακότητας, είναι αρκετά εύκολος δεδομένου ότι πρόκειται για μια συμπεριφορά που παρουσιάζεται σε τακτά και αυστηρά συγκεκριμένα διαστήματα. Η εποχιακότητα συνδέεται άμεσα με τη συχνότητα των δεδομένων της χρονοσειράς και βάσει αυτής γίνεται η αναζήτησή της.

Η κυκλικότητα μοιάζει με την εποχιακότητα δεδομένου ότι κι εδώ υπάρχουν διακυμάνσεις που εμφανίζονται κατά περιόδους. Παρόλα αυτά, το μήκος των περιόδων αυτών δεν είναι σταθερό και συνήθως ξεπερνά το ένα έτος.

Οι μη κανονικές διακυμάνσεις μπορεί να αφορούν είτε σε τυχαιότητα είτε σε ασυνέχεια. Ως προς την τυχαιότητα, σε γενικές γραμμές, κάθε χρονοσειρά εμπεριέχει τον τυχαίο παράγοντα. Το ζήτημα είναι σε τι βαθμό εμφανίζεται αυτός. Όσο πιο έντονος είναι ο τυχαίος παράγοντας τόσο πιο δύσκολη γίνεται η διαδικασία της πρόβλεψης. Οι ασυνέχειες εκφράζονται συνήθως είτε με τη μορφή ακραίων μεγάλων ή ακραίων μικρών τιμών που μπορεί να εμφανιστούν μέσα στη χρονοσειρά τα οποία αποκαλούνται ειδικά γεγονότα (outliers ή special events) ή ακόμα και με απότομη μετατόπιση ολόκληρου του επιπέδου των τιμών της χρονοσειράς (level-shifts). Τα ειδικά γεγονότα έχουν παροδικό χαρακτήρα σε αντίθεση με την αλλαγή επιπέδου που έχει πιο μόνιμο. Και στις δύο περιπτώσεις πρόκειται για απότομες αποκλίσεις από τη συνήθη συμπεριφορά της χρονοσειράς που δημιουργούνται λόγω κάποιου εξαιρετικού γεγονότος.

3.1.3 Ορίζοντας πρόβλεψης

Σκοπός της πρόβλεψης όπως προαναφέρθηκε είναι να παραχθούν εκτιμήσεις για μελλοντικά γεγονότα οι οποίες να πλησιάζουν όσο πιο πολύ γίνεται στις πραγματικές τιμές, αξιοποιώντας την υπάρχουσα γνώση που έχουμε για τα ιστορικά δεδομένα. Διαφορετικά, μπορούμε να πούμε ότι προσπαθούμε να προεκτείνουμε το μήκος μιας χρονοσειράς σε κάποιο συγκεκριμένο χρονικό ορίζοντα στο μέλλον. Ανάλογα με τον αριθμό των περιόδων που προσπαθούμε να προβλέψουμε, μία πρόβλεψη μπορεί να χαρακτηριστεί ως βραχυπρόθεσμη (συνήθως 1-3 περίοδοι), μεσοπρόθεσμη (συνήθως 12-15 περίοδοι αν μιλάμε για μηνιαίες παρατηρήσεις) και μακροπρόθεσμη (συνήθως αφορά πρόβλεψη για πάνω από 3 έτη) [[31]]. Η βραχυπρόθεσμη πρόβλεψη είναι απαραίτητη στις επιχειρήσεις για τον προγραμματισμό των αναγκών τους. Η μεσοπρόθεσμη πρόβλεψη σχετίζεται κυρίως με τον προσδιορισμό κάποιων μελλοντικών αναγκών όπως η ενίσχυση του προσωπικού και η αγορά εξοπλισμού ή πρώτων υλών. Τέλος, η μακροπρόθεσμη πρόβλεψη είναι χρήσιμη για τον μακροχρόνιο στρατηγικό σχεδιασμό μιας επιχείρησης [[39]].

Ανάλογα με το ποιος είναι ο ορίζοντας πρόβλεψης που μας ενδιαφέρει, επιλέγουμε και το ποια θα είναι η μορφή των δεδομένων που θα χρησιμοποιήσουμε. Όταν πραγματοποιούνται βραχυπρόθεσμες επιχειρησιακές προβλέψεις, συνήθως, χρησιμοποιούνται μηνιαία, εβδομαδιαία ή ακόμα και ημερήσια δεδομένα. Όταν από την άλλη πλευρά, έχουμε να κάνουμε με μακροπρόθεσμες προβλέψεις προτείνεται η χρήση τριμηνιαίων ή ετήσιων δεδομένων διότι δεν περιέχουν τον ίδιο όγκο λεπτομέρειας, σε σχέση με τους τύπους που χρησιμοποιούνται στη βραχυπρόθεσμη πρόβλεψη, είναι πιο ομαλές και περιγράφουν καλύτερα τη μακροχρόνια συμπεριφορά της χρονοσειράς [[31]].

Επιπλέον, ο ορίζοντας πρόβλεψης επηρεάζει και την επιλογή της μεθόδου πρόβλεψης δεδομένου ότι δεν αποδίδουν όλες το ίδιο για κάθε χρονικό ορίζοντα. Στην επόμενη ενότητα, αναφέρονται κάποια από τα πιο βασικά μοντέλα πρόβλεψης και σε τι χρονικό ορίζοντα δείχνουν να έχουν καλύτερα αποτελέσματα.

3.2 Κατηγορίες μεθόδων πρόβλεψης

Ανάλογα με το σκοπό της πρόβλεψης, τη διαθεσιμότητα και ποιότητα των ιστορικών δεδομένων αλλά και των μέσων μπορούν να υπάρξουν διαφορετικές κατηγορίες πρόβλεψης. Οι τρεις κύριες κατηγορίες είναι η πρόβλεψη στόχου ή προϋπολογισμού, η κριτική πρόβλεψη και η στατιστική πρόβλεψη. Η τελική ή επιχειρησιακή πρόβλεψη είναι συνάρτηση αυτών των τριών κατηγοριών.

3.2.1 Πρόβλεψη στόχου ή προϋπολογισμού

Ουσιαστικά, η μέθοδος αυτή αναφέρεται στην αναπτυξιακή κατάσταση που επιθυμεί να βρεθεί η επιχείρηση μελλοντικά και όχι τόσο στην πορεία που προβλέπεται ότι θα έχει πραγματικά. Με άλλα λόγια, θέτει τους στόχους και την κατεύθυνση της επιχείρησης. Για το λόγο αυτό, οι προβλέψεις στόχου εμπεριέχουν έντονη αισιοδοξία στα αποτελέσματα τους. Συνήθως, η εφαρμογή της γίνεται μετά από εκτίμηση των αποτελεσμάτων της στατιστικής πρόβλεψης.

3.2.2 Κριτική Πρόβλεψη

Βασίζεται στις απόψεις εμπειρογνώμων και ειδικών. Στις κύριες μεθόδους κριτικών προβλέψεων συγκαταλέγονται η απλή κρίση, η μέθοδος Delphi και οι αναλογίες. Είναι ιδιαίτερα χρήσιμες όταν δεν υπάρχει επαρκής αριθμός ιστορικών δεδομένων ενώ μπορούν να λάβουν υπόψιν τους ειδικά γεγονότα. Παρόλα αυτά, έχουν έντονα το στοιχείο της προκατάληψης.

3.2.3 Στατιστική Πρόβλεψη

Οι στατιστικές προβλέψεις αφορούν στην εφαρμογή στατιστικών μοντέλων αιτιοκρατικών ή χρονοσειρών επί μιας σειράς δεδομένων και στοχεύουν στην αυτοματοποιημένη και συστηματική παραγωγή προβλέψεων (Πετρόπουλος Φ., Ασημακόπουλος Β., 2013). Από τα βασικότερα πλεονεκτήματα των στατιστικών μεθόδων πρόβλεψης είναι η ευκολία στη χρήση τους και η ταχύτητα με τη οποία παράγουν προβλέψεις ακόμα και αν πρόκειται για μεγάλο πλήθος χρονοσειρών. Εν τούτοις για να είναι εφικτή η εφαρμογή τους θα πρέπει να υπάρχει ένα ικανοποιητικός αριθμός διαθέσιμων ιστορικών δεδομένων. Επιπλέον, το γεγονός ότι οι προβλέψεις παράγονται με βάση τα ιστορικά δεδομένα σημαίνει ότι οι στατιστικές μέθοδοι θεωρούν ότι τα δεδομένα της κάθε χρονοσειράς θα συνεχίσουν να συμπεριφέρονται με τον ίδιο τρόπο και στο μέλλον, κάτι το οποίο στην πράξη δεν ισχύει πάντα. Επίσης, δεν λαμβάνουν υπόψιν τους την ύπαρξη ειδικών γεγονότων που ενδεχομένως να λάβουν χώρα στο άμεσο μέλλον.

Παρακάτω, γίνεται μία σύντομη αναφορά των κυριότερων στατιστικών μεθόδων πρόβλεψης [[31]]:

- **Naïve (Απλοϊκή μέθοδος):** Πρόκειται για την πιο απλή μέθοδο πρόβλεψης. Η προβλεπόμενη τιμή σε μια συγκεκριμένη χρονική στιγμή ισούται με την πραγματική τιμή της προηγούμενης περιόδου. Η απλότητα της μεθόδου αυτής επιτρέπει τη γρήγορη και εύκολη παραγωγή πρόβλεψης της επόμενης περιόδου. Παρόλα αυτά, ως μέθοδος δεν ενδείκνυται για προβλέψεις μεγάλου χρονικού ορίζοντα δεδομένου ότι δεν λαμβάνει υπόψιν της τα ιστορικά δεδομένα αλλά βασίζεται

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων αποκλειστικά στην τελευταία παρατήρηση. Ωστόσο, χρησιμοποιείται πολύ συχνά ως σημείο αναφοράς προκειμένου να γίνει σύγκριση με άλλες πιο σύνθετες μεθόδους.

- **Κινητοί μέσοι όροι:** Ανήκουν στην οικογένεια των μεθόδων εξομάλυνσης. Λέγεται μέσος όρος γιατί χρησιμοποιείται ο μέσος ενός αριθμού διαδοχικών παρατηρήσεων της χρονοσειράς. Λέγεται κινητός διότι καθώς προστίθεται μία νέα παρατήρηση στη χρονοσειρά, η παλαιότερη παρατήρηση αντικαθίσταται από την καινούρια για τον υπολογισμό του μέσου όρου. Άρα, ο μέσος όρος «κινείται» πάνω στη χρονοσειρά με κάθε νέα πραγματική τιμή που προστίθεται σε αυτήν. Κατά την εφαρμογή αυτής της μεθόδου, έχει μεγάλη σημασία η επιλογή του κατάλληλου αριθμού παρατηρήσεων που θα συμμετέχουν στον υπολογισμό του μέσου όρου. Ο αριθμός αυτός παραμένει σταθερός καθ' όλη τη διάρκεια εφαρμογής της. Οι κινητοί μέσοι όροι απομακρύνουν τα στοιχεία τυχαιότητα και εποχιακότητα από μία χρονοσειρά και γι' αυτό χρησιμοποιούνται στις διαδικασίες αποσύνθεσης όπου αυτή είναι και η κύρια χρησιμότητά τους. Αν και οι κινητοί μέσοι όροι κατατάσσονται στις μεθόδους πρόβλεψης, πρακτικά δεν προτείνονται για αυτή τη χρήση γιατί δεν μπορούν να συμπεριλάβουν την τάση και την εποχιακότητα στα τελικά αποτελέσματα της πρόβλεψης. Στις συγκεκριμένους μεθόδους συγκαταλέγονται ο απλός κινητός μέσος όρος, ο σταθμισμένος κινητός μέσος όρος, ο κεντρικός κινητός μέσος όρος και ο διπλός κινητός μέσος όρος.

- **Μοντέλα εκθετικής εξομάλυνσης:** Οι συγκεκριμένες μέθοδοι έγιναν ιδιαίτερα δημοφιλείς λόγω της ευκολίας, της ελάχιστης απαίτησης σε υπολογιστικό χρόνο και σε ιστορικά δεδομένα προκειμένου να παραχθούν προβλέψεις. Είναι κατάλληλες κυρίως για βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις. Η απόδοσή τους είναι καλύτερη σε δεδομένα που παρουσιάζουν στασιμότητα ή μικρό ρυθμό ανάπτυξης. Από τις πιο γνωστές μεθόδους της εν λόγω κατηγορίας είναι η απλή εκθετική εξομάλυνση σταθερού επιπέδου, η εκθετική εξομάλυνση γραμμικής τάσης και η εκθετική εξομάλυνση μη γραμμικής ή φθίνουσας τάσης. Στην εργασία γίνεται εφαρμογή των μοντέλων εκθετικής εξομάλυνσης και περιγράφονται στη συνέχεια με μεγαλύτερη λεπτομέρεια.

- **Μοντέλα παλινδρόμησης:** Μέσα από τη χρήση τέτοιων μοντέλων, εξετάζεται η σχέση μεταξύ μιας εξαρτημένης μεταβλητής με συγκεκριμένες ανεξάρτητες μεταβλητές. Ανάλογα με το πλήθος των ανεξάρτητων μεταβλητών που συμμετέχουν στη μελέτη αυτή, τα μοντέλα παλινδρόμησης διακρίνονται σε απλά και πολλαπλά.

- **Μέθοδος theta:** Η μέθοδος αυτή βασίζεται στη μεταβολή των τοπικών καμπυλοτήτων μιας χρονοσειράς. Η αρχική χρονοσειρά διαχωρίζεται σε δύο η και περισσότερες συνιστώσες (γραμμές Theta) και πραγματοποιούνται προβλέψεις για την καθεμία ξεχωριστά είτε με την ίδια ή και με διαφορετικές μεθόδους πρόβλεψης. Η τελική πρόβλεψη προκύπτει από το συνδυασμό των παραγόμενων προβλέψεων.

- **Μοντέλα ARIMA:** Τα ολοκληρωμένα αυτοπαλινδρομικά μοντέλα κινητών μέσων όρων αποτελούν στοχαστικά μοντέλα που στόχος τους είναι να περιγράψουν πως εξελίσσεται διαχρονικά κάποιο φυσικό μέγεθος που εξαρτάται από μη ντετερμινιστικούς παράγοντες. Το

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων μοντέλο που προκύπτει βασίζεται στις παρελθοντικές τιμές και αποτελεί ένα γραμμικό συνδυασμό του τυχαίου παράγοντα, των τιμών των μεγεθών και άλλων στοχαστικών παραγόντων.

- **Νευρωνικά δίκτυα:** Ένα νευρωνικό δίκτυο αποτελείται από κόμβους (νευρώνες) εισόδου, εξόδου και (κρυφούς) υπολογιστικούς κόμβους. Η βασική ιδέα είναι ότι το φιλτράρισμα των εισόδων, δηλαδή των ανεξάρτητων μεταβλητών, γίνεται μέσω ενός ή περισσότερων επιπέδων, τα οποία αποτελούνται από κρυφούς κόμβους, πριν παραχθεί η ζητούμενη έξοδος

Τέλος, μία άλλη πρακτική είναι ο συνδυασμός διαφορετικών μεθόδων πρόβλεψης, χρησιμοποιώντας ίσα ή άνισα βάρη για την παραγωγή των τελικών προβλέψεων. Το 1969 οι Bates & Granger έδειξαν ότι ο συνδυασμός διαφορετικών μεθόδων προβλέψεων μπορεί να βελτιώσει την ακρίβεια των προβλέψεων. Ακόμα και αν δύο προβλεπτικές μέθοδοι μεμονωμένα δίνουν μεγάλα προβλεπτικά σφάλματα, ο συνδυασμός τους μπορεί να βελτιώσει το επίπεδο αξιοπιστίας [[32]]. Πολλές φορές οι πιο απλοί συνδυασμοί προβλεπτικών μεθόδων μπορεί να δώσουν καλύτερα αποτελέσματα από τους πιο πολύπλοκους συνδυασμούς. Σε έρευνα τους οι Shen et al (2011) υποστήριξαν ότι οι συνδυασμοί έως και τριών μεμονωμένων μεθόδων είναι αυτοί που μπορούν να δώσουν τα καλύτερα αποτελέσματα [[40]]. Ο Clemen (1989) έγραψε ότι ο συνδυασμός μεθόδων οδηγεί σε μεγαλύτερη ακρίβεια προβλέψεων και θα πρέπει να αποτελεί βασικό κομμάτι των πρακτικών πρόβλεψης [[33]]. Επιπλέον, τα οφέλη του συνδυασμού των προβλέψεων είναι κάτι που επιβεβαιώνεται και από τα αποτελέσματα του διαγωνισμού προβλέψεων M4 όπου οι μέθοδοι με την υψηλότερη απόδοση ήταν από συνδυασμούς κυρίως στατιστικών προσεγγίσεων [[37]].

3.3 Μοντέλα εκθετικής εξομάλυνσης

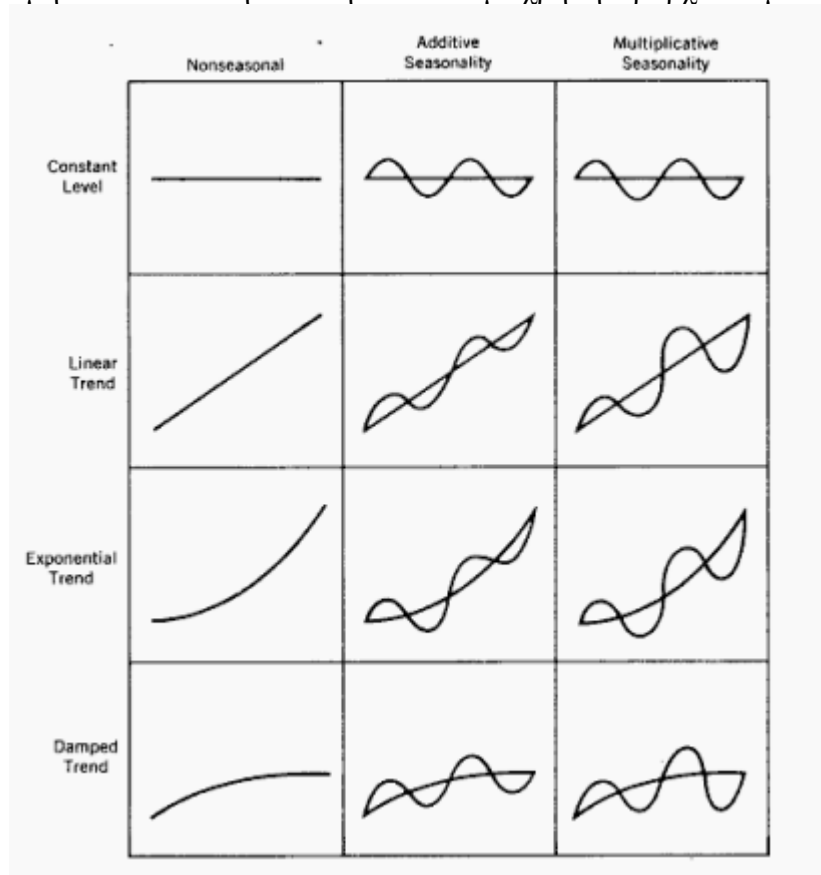
Τα μοντέλα εκθετικής εξομάλυνσης εμφανίσθηκαν για πρώτη φορά κατά τη διάρκεια του 2ου Παγκοσμίου Πολέμου μέσα από την εργασία του Robert G.Brown που εργαζόταν ως αναλυτής στο πολεμικό ναυτικό των Η.Π.Α. [[34]]. Αναπτύχθηκαν κατά τη δεκαετία του 1950 και το 1960 γνώρισαν ευρεία εφαρμογή με την άνθιση της πληροφορικής [[31]].

Ως προβλεπτικά μοντέλα έγιναν ιδιαίτερα δημοφιλείς λόγω της απλότητας των μοντέλων που υιοθετούν, του μειωμένου υπολογιστικού χρόνου και της περιορισμένης απαίτησης σε αριθμό δεδομένων [[31]][[35]]. Για τους λόγους αυτούς, είναι ιδιαίτερα χρήσιμες για την εφαρμογή μαζικών προβλέψεων. Χρησιμοποιούνται, κυρίως, σε περιπτώσεις βραχυπρόθεσμου σχεδιασμού και τα ποσοστά ακρίβειας στα αποτελέσματά τους είναι αρκετά ικανοποιητικά σε σχέση με άλλες πιο πολύπλοκες μεθόδους. Ο λόγος που συμβαίνει αυτό είναι γιατί ως μέθοδοι πρόβλεψης μένουν ανεπηρέαστες από διάφορες ιδιομορφίες της συμπεριφοράς των δεδομένων ή ακραίων τιμών που ενδέχεται να παρουσιαστούν περιστασιακά.

Τα μοντέλα εκθετικής εξομάλυνσης βασίζονται στα ιστορικά δεδομένα και το πρότυπο που αυτά ακολουθούν. Πιο συγκεκριμένα, αυτό που κάνουν είναι να ελαχιστοποιούν την τυχαιότητα και να προεκτείνουν τα στοιχεία του προτύπου που ακολουθούν τα δεδομένα μιας χρονοσειράς, όπως είναι η τάση ή η εποχιακότητα, στο μέλλον. Μέσω της εξομάλυνσης των ιστορικών δεδομένων, απομακρύνουν τις τυχαίες αποκλίσεις από το πραγματικό πρότυπο της χρονοσειράς. Επομένως, η πρόβλεψη πραγματοποιείται πάνω στον εξομαλυμένο τρόπο που συμπεριφέρονται τα δεδομένα.

Η εκθετική εξομάλυνση θεωρείται μια προέκταση των κινητών μέσων όρων όπου βασική αρχή είναι ότι τα πιο πρόσφατα δεδομένα εμπεριέχουν την περισσότερη πληροφορία και γι' αυτό δίνεται μεγαλύτερη βαρύτητα στις πιο πρόσφατες παρατηρήσεις η οποία φθίνει εκθετικά όσο αναφερόμαστε σε παλαιότερες. Αυτός είναι και ο λόγος που ενδείκνυνται για βραχυπρόθεσμες προβλέψεις.

Η βασική κατηγοριοποίηση των μοντέλων εξομάλυνσης προκύπτει από τους συνδυασμούς των τεσσάρων επιπέδων τάσης (constant level, linear level, exponential level, damped level) με τα τρία εποχιακά μοντέλα (no seasonality, additive seasonality, multiple seasonality) [[35]]. Στο παρακάτω σχήμα φαίνεται η γραφική αναπαράσταση των τύπων εκθετικής εξομάλυνσης ανά κατηγορία [[42]]:



Σχήμα 3.1: Γραφική αναπαράσταση των μοντέλων εκθετικής εξομάλυνσης

3.3.1 Μοντέλο σταθερού επιπέδου (Simple Exponential Smoothing)

Το μοντέλο αυτό αναφέρεται και ως απλή εκθετική εξομάλυνση. Η υπόθεση που κάνει το συγκεκριμένο μοντέλο είναι ότι δεν υπάρχει σαφής τάση. Αυτό σημαίνει πως το επίπεδο της χρονοσειράς, δηλαδή, ο μέσος όρος στις τιμές των δεδομένων, είναι σχετικά σταθερό. Επομένως, η πρόβλεψη προκύπτει από την προέκταση μιας οριζόντιας ευθείας γραμμής. Σαν προβλεπτικό μοντέλο θεωρείται κατάλληλο για προβλέψεις ενός βήματος (one-step-ahead-forecasting) ή χρονοσειρές με αυξημένο θόρυβο ή τυχαιότητα.

Η πρόβλεψη γίνεται μέσω του υπολογισμού ενός επιπέδου, οπότε, αυτό που μας ενδιαφέρει είναι ο υπολογισμός του επιπέδου αυτού. Το μοντέλο σταθερού επιπέδου περιγράφεται από τις παρακάτω σχέσεις:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + \alpha \cdot e_t$$

$$F_{t+1} = S_t$$

Όπου,

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
- e :δηλώνει το σφάλμα, δηλαδή, την απόκλιση της πραγματικής τιμής από την πρόβλεψη

- S :δηλώνει το (προβλεπόμενο) επίπεδο της χρονοσειράς

- F :δηλώνει την πρόβλεψη

- t :δηλώνει τη χρονική περίοδο

- α :αποτελεί το συντελεστή εξομάλυνσης της μεθόδου και μπορεί να λάβει τιμές στο διάστημα $[0,1]$

Η πρόβλεψη για την επόμενη χρονική περίοδο ταυτίζεται με το επίπεδο που υπολογίστηκε στην προηγούμενη χρονική περίοδο. Από την άλλη, το επίπεδο μιας περιόδου ισούται με το επίπεδο της προηγούμενης περιόδου προσαυξημένης με το σφάλμα πρόβλεψης. Επομένως, η μέθοδος αυτή, υπολογίζει το σφάλμα σε κάθε χρονική στιγμή προσπαθώντας να φέρει την προβλεπόμενη τιμή κοντά στην πραγματική. Ο συντελεστής α ρυθμίζει το βαθμό (ποσοστό) συμμετοχής του σφάλματος στον υπολογισμό της πρόβλεψης.

Το ζήτημα που προκύπτει ως προς το συγκεκριμένο μοντέλο είναι ποιο θα είναι το πρώτο επίπεδο το οποίο θα μας δώσει και την αρχική πρόβλεψη. Για να μπορέσει να ξεκινήσει ο υπολογισμός του μοντέλου πρόβλεψης θα πρέπει να οριστεί ένα αρχικό επίπεδο (S_0). Συνήθως, ως τιμή αρχικοποίησης μπορεί να χρησιμοποιηθεί:

- Ο μέσος όρος όλων των παρατηρήσεων

- Ο μέσος όρος των n πρώτων παρατηρήσεων

- Η πρώτη παρατήρηση

- Το σταθερό επίπεδο από το μοντέλο της απλής γραμμικής παλινδρόμησης

Όπως εύκολα μπορεί να αντιληφθεί κανείς, έχει μεγάλη σημασία η επιλογή του κατάλληλου αρχικού επιπέδου. Από αυτό επηρεάζεται η ακρίβεια ως προς τις προβλεπόμενες τιμές. Επίσης, το επίπεδο S_0 , επηρεάζει και την επιλογή του συντελεστή εξομάλυνσης.

Η βέλτιστη τιμή του α καθορίζεται από δύο αλληλεξαρτώμενους παράγοντες. Ο πρώτος παράγοντας είναι ο βαθμός θορύβου που εμφανίζει η χρονοσειρά. Όσο μεγαλύτερο είναι το ποσοστό του θορύβου τόσο πιο μικρό πρέπει να είναι το α ώστε το μοντέλο να μην αντιδρά υπερβολικά στο θόρυβο. Ο δεύτερος παράγοντας έχει να κάνει με τη σταθερότητα του μέσου όρου της χρονοσειράς. Αν αυτή είναι μεγάλη τότε και το α θα είναι μικρό. Διαφορετικά, στην περίπτωση που ο μέσος όρος των δεδομένων μεταβάλλεται τότε και το α θα πρέπει να είναι μεγάλο έτσι ώστε το μοντέλο να ακολουθεί τις μεταβολές αυτές.

Ο καθορισμός του βέλτιστου α συνήθως γίνεται αυτόματα με τη χρήση κάποιου υπολογιστικού συστήματος και καθορίζεται από την ελαχιστοποίηση του in-sample σφάλματος. Συνήθως, επιλέγεται η ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος MSE ως κριτήριο έτσι ώστε το μοντέλο να μην δημιουργεί μεγάλα σφάλματα.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
 Όταν το α δεν είναι 0 ή 1 τότε όλες οι παρατηρήσεις συμμετέχουν στην πρόβλεψη. Η βαρύτητα για την κάθε παρατήρηση έχει να κάνει με το μέγεθος του συντελεστή εξομάλυνσης. Όσο πιο μεγάλο είναι το α τόσο μεγαλύτερη έμφαση δίνεται στις τελευταίες παρατηρήσεις ενώ αν πλησιάζει τη μονάδα το σφάλμα της τελευταίας περιόδου είναι και αυτό που επηρεάζει περισσότερο την πρόβλεψη της επόμενης περιόδου. Στην περίπτωση που το α λαμβάνει ακριβώς την τιμή 1 τότε σημαίνει πως λαμβάνεται υπόψιν ολόκληρο το σφάλμα και στην ουσία η SES ταυτίζεται με την απλοϊκή μέθοδο (Naïve). Αντίθετα, αν το α είναι πολύ μικρό η πρόβλεψη δεν εξαρτάται σε μεγάλο βαθμό από το τι συνέβη την προηγούμενη χρονική περίοδο, ενώ, στην ακραία περίπτωση που το α λάβει την τιμή 0, τότε κάθε πρόβλεψη θα ισούται με το αρχικό επίπεδο S_0 .

Στο σημείο αυτό αξίζει να σημειωθεί ότι και το μήκος της χρονοσειράς επηρεάζει το βαθμό επίδρασης της κάθε παρατήρησης στην τελική πρόβλεψη. Αυτό σημαίνει πως όσο μεγαλύτερη είναι η χρονοσειρά τότε για δεδομένο α , οι παρατηρήσεις που χάνονται στο βάθος του χρόνου θα συμμετέχουν λιγότερο στην πρόβλεψη. Στην περίπτωση μιας αρκετά μεγάλης χρονοσειράς, το επίπεδο αρχικοποίησης δεν έχει ιδιαίτερη περίπτωση.

3.3.2 Μοντέλο γραμμικής τάσης (Holt Exponential Smoothing)

Το μοντέλο γραμμικής τάσης προτάθηκε από τον Holt το 1957 και στην ουσία αποτελεί μία επέκταση του μοντέλου σταθερού επιπέδου. Το νέο στοιχείο που λαμβάνει υπόψιν του το συγκεκριμένο μοντέλο είναι η συνιστώσα της τάσης. Επομένως, επιτρέπει την πρόβλεψη σε δεδομένα που εμφανίζουν τάση. Μαθηματικά, περιγράφεται από τις παρακάτω σχέσεις:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + T_{t-1} + \alpha \cdot e_t$$

$$T_t = T_{t-1} + \alpha \cdot \beta \cdot e_t$$

$$F_{t+m} = S_t + m \cdot T_t$$

Όπου,

- e :δηλώνει το σφάλμα, δηλαδή, την απόκλιση της πραγματικής τιμής από την πρόβλεψη
- S :δηλώνει το (προβλεπόμενο) επίπεδο της χρονοσειράς
- F :δηλώνει την πρόβλεψη
- T :δηλώνει την τάση
- t :δηλώνει τη χρονική περίοδο
- m :δηλώνει το χρονικό ορίζοντα πρόβλεψης
- α :αποτελεί το συντελεστή εξομάλυνσης του επιπέδου

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

- β :αποτελεί το συντελεστή εξομάλυνσης της τάσης

Τα α και β λαμβάνουν τιμές στο διάστημα $[0,1]$. Αυτό που είναι επιθυμητό είναι ο βέλτιστος συνδυασμός τους. Η διαδικασία εύρεσης των βέλτιστων παραμέτρων μπορεί να γίνει και πάλι με βάση την ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος στα in-sample δεδομένα.

Από τις παραπάνω σχέσεις γίνεται αντιληπτό ότι, για ορίζοντα πρόβλεψης m , θα έχουμε το επίπεδο της τελευταίας παρατήρησης συν την τάση πολλαπλασιασμένη με τον αριθμό των περιόδων που πραγματοποιείται η πρόβλεψη. Επομένως, για κάθε επιπλέον περίοδο που γίνεται η πρόβλεψη, σε αντίθεση με τη SES, η οποία προβλέπει ένα επίπεδο το οποίο παραμένει σταθερό ανεξάρτητα από τον αριθμό των περιόδων του ορίζοντα πρόβλεψης, εδώ, η πρόβλεψη μεταβάλλεται γραμμικά με το επίπεδο της τάσης.

Ο υπολογισμός της τάσης βασίζεται στην ίδια λογική με την οποία υπολογίζεται το επίπεδο S_0 που περιγράψαμε νωρίτερα και η οποία ισχύει και για το μοντέλο γραμμικής τάσης. Με άλλα λόγια, η τάση μιας περιόδου ισούται με την τάση της προηγούμενης περιόδου προσαυξημένης με ένα ποσοστό σφάλματος. Επομένως, μας ενδιαφέρει τόσο το βέλτιστο επίπεδο πρόβλεψης όσο και το βέλτιστο επίπεδο τάσης.

Επιπλέον, για να μπορέσει να λειτουργεί το μοντέλο πρέπει να καθοριστούν οι αρχικές τιμές του επιπέδου (S_0) και της τάσης (T_0). Για το S_0 ισχύει ό,τι και στο μοντέλο απλής εκθετικής εξομάλυνσης. Για την αρχικοποίηση της τάσης συνήθως χρησιμοποιούνται:

- Η διαφορά μεταξύ των δύο πρώτων παρατηρήσεων ($Y_2 - Y_1$)
- Η διαφορά της n -οστής και πρώτης παρατήρησης διαιρεμένης με $(n-1)$
- Η σταθερά της κλίσης από το μοντέλο της απλής γραμμικής παλινδρόμησης

Για την επιλογή τόσο του αρχικού επιπέδου όσο και της αρχικής τάσης παίζουν ρόλο τα ποιοτικά χαρακτηριστικά της χρονοσειράς. Χρειάζεται ιδιαίτερη προσοχή στην αρχικοποίηση δεδομένου, ότι οι τιμές αυτές επηρεάζουν τις προβλέψεις.

3.3.3 Μοντέλο μη γραμμικής τάσης (Damped Exponential Smoothing)

Εμπειρικά στοιχεία έχουν δείξει ότι το μοντέλο γραμμικής τάσης του Holt τείνει να έχει μία θετική προκατάληψη στις προβλέψεις που παράγει σε σχέση με τις πραγματικές τιμές, ειδικά σε περιπτώσεις μεγάλου χρονικού ορίζοντα πρόβλεψης. Ο λόγος είναι γιατί οι προβλέψεις εμφανίζουν μία σταθερή τάση στο μέλλον επ' αόριστον. Υποκινούμενοι από αυτό, οι Gardner & McKenzie το 1985, εισήγαγαν μία νέα παράμετρο η οποία ελέγχει το ρυθμό αύξησης των τιμών των τάσεων. Με άλλα λόγια, αυξομειώνει τον τρόπο που αλλάζει η τάση στο μέλλον. Η παράμετρος αυτή ονομάζεται παράμετρος διόρθωσης της τάσης (trend – modification parameter) και συμβολίζεται με ϕ . Με τον τρόπο αυτό, το προβλεπτικό μοντέλο μπορεί να προσαρμόζεται και σε μη γραμμικές τάσεις. Πλέον, θεωρούνται τα πιο επιτυχημένα και δημοφιλή μοντέλα σε

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
περιπτώσεις που απαιτείται η πρόβλεψη μεγάλου αριθμού χρονοσειρών. Η μαθηματικές σχέσεις που την περιγράφουν είναι ακόλουθες:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + \varphi \cdot T_{t-1} + \alpha \cdot e_t$$

$$F_{t+m} = s_t + \sum_{i=1}^m \varphi^i \cdot T_t$$

Όπου, το t υποδηλώνει τις τιμές που μπορεί να πάρει ο ορίζοντας πρόβλεψης. Όλοι οι υπόλοιποι συμβολισμοί είναι οι γνωστοί που παρουσιάστηκαν νωρίτερα. Η μορφή του μοντέλου μη γραμμικής εξαρτάται από την τιμή του συντελεστή φ .

- Αν $\varphi = 0$, τότε υπάρχει απουσία της τάσης στον υπολογισμό της πρόβλεψης και επομένως, αναφερόμαστε στο μοντέλο σταθερού επιπέδου
- Αν $0 < \varphi < 1$, τότε προκύπτει το μοντέλο φθίνουσας τάσης (damped exponential smoothing). Εδώ, όπως είναι αντιληπτό από την τελευταία σχέση όσο μεγαλώνει ο ορίζοντας πρόβλεψης (δηλαδή, όσο το i παίρνει μεγαλύτερες τιμές), τόσο πιο πολύ μεγαλώνει το φ και κατ' επέκτασιν μειώνεται η τάση.
- Αν $\varphi=1$, τότε το άθροισμα της τελευταίας σχέσης είναι ίσο με τον αριθμό των περιόδων του χρονικού ορίζοντα πρόβλεψης, m . Άρα, ουσιαστικά, το μοντέλο που προκύπτει είναι το μοντέλο γραμμικής τάσης του Holt.
- Αν $\varphi > 1$, τότε προκύπτει το μοντέλο εκθετικής τάσης, το οποίο λειτουργεί με ανάποδο τρόπο από το μοντέλο φθίνουσας τάσης. Όσο το i αυξάνεται, τόσο πιο πολύ αυξάνει η τάση.

Για τον εντοπισμό της βέλτιστης τιμής του φ προτείνεται, ως κριτήριο επιλογής η ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος, όπως συμβαίνει και με τους συντελεστές εξομάλυνσης α και β . Ως προς τα βέλτιστα S_0 και T_0 , προτείνεται η εφαρμογή της μεθόδου απλής γραμμικής παλινδρόμησης με ανεξάρτητη μεταβλητή το χρόνο, όπου το S_0 ισούται με την σταθερά α και το T_0 ισούται με την κλίση β .

Όπως μπορεί να αντιληφθεί κανείς πολύ εύκολα, το μοντέλο μη γραμμικής τάσης αποτελεί την πιο γενική μορφή του μοντέλου εκθετικής εξομάλυνσης διότι ανάλογα με την τιμή του φ μπορεί να περιγράψει κάθε ένα από τα μοντέλα που προαναφέρθηκαν. Ως προβλεπτικό μοντέλο, είναι ικανό να δώσει αρκετά καλά αποτελέσματα ως προς την ακρίβεια ειδικά όταν αφορά σε μεσοπρόθεσμες και μακροπρόθεσμες προβλέψεις.

3.3.4 Εποχιακή εξομάλυνση

Έως τώρα δεν έχει γίνει αναφορά σε εποχιακή συμπεριφορά των δεδομένων. Τα παραπάνω μοντέλα αφορούν σε χρονοσειρές που εμφανίζουν μόνο κάποιο μοτίβο τάσης. Παρόλα αυτά, η εποχιακότητα είναι ένα χαρακτηριστικό το οποίο μπορεί να εμφανιστεί στα ιστορικά δεδομένα.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
Επομένως, είναι απαραίτητη η εφαρμογή κάποιας μεθόδου που να λαμβάνει υπόψη τις εποχιακές διακυμάνσεις. Αυτό, επιτυγχάνεται εύκολα με την προσθήκη ενός εποχιακού παράγοντα στα μοντέλα εκθετικής εξομάλυνσης.

Ο τρόπος με τον οποίο εισάγεται ο εποχιακός παράγοντας έχει να κάνει με τη φύση των εποχιακών διακυμάνσεων που εμφανίζονται στα ιστορικά δεδομένα, που μπορεί να είναι είτε προσθετικού είτε πολλαπλασιαστικού χαρακτήρα. Στην προσθετικού τύπου εποχιακότητα, ο εποχιακός παράγοντας εισάγεται σε απόλυτους όρους και η αποεποχικοποίηση επιτυγχάνεται με την αφαίρεσή του από τις πραγματικές τιμές. Όταν το μοτίβο είναι πολλαπλασιαστικό, ο εποχιακός παράγοντας περιγράφεται από το λόγο της κάθε τιμής της χρονοσειράς προς το μέσο όρο όλων των τιμών της σε όλο το έτος και η αποεποχικοποίηση γίνεται με διαίρεση της πραγματικής σειράς με τον εποχιακό παράγοντα. Επομένως, η πρόβλεψη εφαρμόζεται στα αποεποχικοποιημένα δεδομένα και έπειτα με την αντίστροφη διαδικασία ανάλογα με τον αν υπάρχει προσθετική ή πολλαπλασιαστική εποχιακότητα, γίνεται η εποχικοποίηση και λαμβάνουμε τις τελικές τιμές των προβλέψεων. Ενδεικτικά, αναφέρονται δύο βασικά μοντέλα με εποχική εξομάλυνση.

Μοντέλο Winters

Ουσιαστικά πρόκειται για το μοντέλο σταθερού επιπέδου με τη διαφορά ότι υπάρχει πολλαπλασιαστική εποχιακότητα.

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + \frac{\alpha \cdot e_t}{I_{t-p}}$$

$$I_t = I_{t-p} + \frac{\gamma \cdot e_t}{S_{t-1}}$$

$$F_{t+m} = S_t \cdot I_{t-p+m}$$

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

Μοντέλο Holt - Winters

Είναι η αντίστοιχη επέκταση του μοντέλου γραμμικής τάσης του Holt με πολλαπλασιαστική εποχιακότητα.

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + T_{t-1} + \frac{\alpha \cdot e_t}{I_{t-p}}$$

$$T_t = T_{t-1} + \frac{\alpha \cdot \beta \cdot e_t}{I_{t-p}}$$

$$I_t = I_{t-p} + \frac{\gamma \cdot e_t}{S_{t-1} + T_{t-1}}$$

$$F_{t+m} = (S_t + m \cdot T_t) \cdot I_{t-p+m}$$

Όπου,

-**I** : είναι ο εξομαλυσμένος εποχιακός παράγοντας

-**γ** : ο συντελεστής εξομάλυνσης των εποχιακών παραγόντων που λαμβάνει τιμές στο διάστημα [0,1]

-**p** : ο αριθμός περιόδων ανά έτος (άρα και ο αριθμός των εποχιακών παραγόντων).

Και στα δύο μοντέλα, η τιμή του επιπέδου (και της τάσης για το μοντέλο Holt - Winters) είναι αποεποχικοποιημένη και μεταβάλλεται κατά ένα ποσοστό του αποεποχικοποιημένου σφάλματος. Ως προς την εξομάλυνση των εποχικών παραγόντων, φαίνεται ότι μεταβάλλονται με ένα ποσοστό του σφάλματος προς το επίπεδο.

Για να λειτουργήσουν τα μοντέλα, θα πρέπει να οριστεί το **p** που απαιτείται για την εκτίμηση των αρχικών εποχιακών παραγόντων. Η αρχικοποιήσεις επιπέδου γίνονται στην περίοδο **p**. Επομένως, η πρόβλεψη ξεκινά την περίοδο **p+1**. Ο εντοπισμός των βέλτιστων τιμών των συντελεστών μπορεί και πάλι να γίνει με βάση την ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος.

Σε γενικές γραμμές, η προσθετική μέθοδος προτιμάται όταν οι εποχιακές διακυμάνσεις παραμένουν σχεδόν σταθερές. Από την άλλη, η πολλαπλασιαστική μέθοδος προτιμάται όταν οι εποχιακές διακυμάνσεις μεταβάλλονται αναλογικά με το επίπεδο της σειράς.

3.3.5 Ταξινόμηση μεθόδων εκθετικής εξομάλυνσης - Ο αλγόριθμος ETS

Δεδομένων όλων όσων έχουν αναφερθεί στην ενότητα αυτή, τα μοντέλα εκθετικής εξομάλυνσης δεν είναι μόνο αυτά τα οποία περιγράψαμε. Σε ένα μοντέλο μπορεί να συναντήσουμε είτε απουσία τάσης, είτε προσθετική ή πολλαπλασιαστική τάση, αυξητική ή φθίνουσα. Επιπλέον, μπορεί στα δεδομένα να μην εμφανίζεται εποχιακότητα ή στις περιπτώσεις που υπάρχει το εποχιακό στοιχείο μπορεί να έχει προσθετικό ή πολλαπλασιαστικό χαρακτήρα. Επομένως, μέσα από όλους αυτούς

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
τους συνδυασμούς μεταξύ τάσης και εποχιακότητας μπορούν να προκύψουν και πολλές
κατηγορίες μοντέλων εκθετικής εξομάλυνσης. Οι κατηγορίες αυτές συνοψίζονται στον παρακάτω
πίνακα.

Πίνακας 3.1 Ταξινόμηση μοντέλων εκθετικής εξομάλυνσης

Trend component	Seasonal component		
	N	A	M
N	(N,N)	(N,A)	(N,M)
A	(A,N)	(A,A)	(A,M)
Ad	(Ad,N)	(Ad,A)	(Ad,M)
M	(M,N)	(M,A)	(M,M)
Md	(Md,N)	(Md,A)	(Md,M)

Όπου, N = None, A = Addictive, Ad = Addictive damped, M = Multiplicative, Md = Multiplicative damped.

Η παραπάνω κατηγοριοποίηση ξεκίνησε από τον Pegel το 1969 όπου εισήγαγε τους τύπους N, A, M για την τάση [17], επεκτάθηκε από τον Gadner (1985) που προσέθεσε τον τύπο Ad [35] και τον Taylor (2003) που εισήγαγε τον τύπο Md [41].

Τα μοντέλα space-state περιγράφονται από κάποιες εξισώσεις οι οποίες εκφράζουν τον τρόπο με τον οποίο τα δεδομένα μιας χρονοσειράς μεταβάλλονται μέσα στο χρόνο. Πιο συγκεκριμένα, μέσα από τις εξισώσεις αυτές φαίνονται και αναλύονται οι μεταβολές των καταστάσεων (επίπεδο, τάση, εποχιακότητα) κάθε χρονική στιγμή. Για κάθε μέθοδο υπάρχουν δύο μοντέλα ανάλογα με το σφάλμα υπολογισμού, το οποίο μπορεί να εμφανίζεται είτε προσθετικά (A) είτε πολλαπλασιαστικά (M). Οι διαφορές ανάμεσα στα προσθετικά και τα πολλαπλασιαστικά μοντέλα είναι ότι λαμβάνουμε τις ίδιες σημειακές προβλέψεις με διαφορετικά διαστήματα πρόβλεψης. Η γενική ονομασία των μεθόδων αυτών είναι ETS και προέρχεται από τα αρχικά των τριών συνιστωσών του Error, Trend και Seasonal. Επομένως, από τους συνδυασμούς μεταξύ των τριών αυτών στοιχείων προκύπτουν συνολικά τριάντα μοντέλα state-space.

Η μέθοδος αυτή είναι πολύ ευέλικτη και αποτελεσματική, καθώς μπορεί να αντιμετωπίσει κάθε διαφορετικό συνδυασμό τάσης και εποχιακότητας και ανάλογα να προσαρμόζεται. Το μεγάλο της πλεονέκτημα είναι ότι μπορεί να αναγνωρίσει τη φύση των δεδομένων μιας χρονοσειράς και να επιλέγει τη μέθοδο πρόβλεψης που εφαρμόζει με βάση τη συμπεριφορά της. Για τις επιχειρήσεις αποτελεί ένα πάρα πολύ χρήσιμο εργαλείο, δεδομένου ότι εμφανίζεται πολλές φορές η ανάγκη για εφαρμογή προβλέψεων που να γίνονται αυτόματα, χωρίς να υπάρχει μεγάλος βαθμός ανθρώπινης παρέμβασης και να λαμβάνονται υπόψιν τα ποιοτικά χαρακτηριστικά των δεδομένων των χρονοσειρών [[39]].

3.4 Ακρίβεια προβλέψεων

3.4.1 Στατιστικοί δείκτες

Η αξιολόγηση και η επιλογή ενός μοντέλου πρόβλεψης βασίζεται στην ακρίβεια των αποτελεσμάτων που δίνει σε σχέση με τις πραγματικές τιμές. Προκειμένου να μετρηθεί η ακρίβεια των προβλέψεων γίνεται χρήση κάποιων στατιστικών δεικτών γύρω από το σφάλμα πρόβλεψης, δηλαδή τη διαφορά μεταξύ πραγματικής και προβλεπόμενης τιμής:

$$e_i = Y_i - F_i, i=1,2,\dots,n$$

Όπου, Y_i είναι η πραγματική τιμή την περίοδο i , F_i είναι η προβλεπόμενη τιμή την περίοδο i . Παρακάτω, αναφέρονται οι βασικότεροι στατιστικοί δείκτες σφάλματος που χρησιμοποιούνται [[31]]:

Μέσο σφάλμα (Mean Error):

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)$$

Το ME αποτελεί ένα μέτρο συστηματικότητας του σφάλματος και μπορεί να δείξει την προκατάληψη του μοντέλου όπου όταν ο δείκτης παίρνει θετικές τιμές η πρόβλεψη θεωρείται απαισιόδοξη ενώ αν παίρνει αρνητικές τιμές δείχνει αισιοδοξία.

Μέσο απόλυτο σφάλμα (Mean Absolute Error):

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - F_i|$$

Το MAE είναι ένα μέσο μέτρο ακρίβειας των προβλέψεων έναντι των πραγματικών τιμών χωρίς να λαμβάνει υπόψιν του την κατεύθυνση της πρόβλεψης.

Μέσο τετραγωνικό σφάλμα (Mean Squared Error):

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

Το MSE είναι και αυτό ένα μέτρο ακρίβειας το οποίο όμως δίνει περισσότερη έμφαση στα μεγάλα σφάλματα και λιγότερη στα μικρά. Όπως αναφέρθηκε σε προηγούμενη ενότητα είναι ιδιαίτερα χρήσιμο μέτρο για τον εντοπισμό των βέλτιστων συντελεστών εξομάλυνσης.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
Ρίζα μέσου τετραγωνικού σφάλματος (Root Mean Squared Error):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2}$$

Διατηρεί τις ίδιες ιδιότητες με το MSE απλά είναι εκφρασμένο στην κλίμακα της χρονοσειράς που μελετάται.

Μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{Y_i} \right| \cdot 100\%$$

Είναι ιδιαίτερα χρήσιμο σε περιπτώσεις που επιθυμούμε προβλέψεις σε χρονοσειρές με διαφορετικό επίπεδο μέσης τιμής. Παρόλα αυτά, δεν μπορεί να εφαρμοστεί σε χρονοσειρές διακοπτόμενης ζήτησης με πολλές μηδενικές τιμές διότι καταλήγει σε απροσδιοριστία.

Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (Symmetric Mean Absolute Percentage Error):

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{\frac{Y_i + F_i}{2}} \right| \cdot 100\% = \frac{1}{n} \sum_{i=1}^n \left| \frac{2(Y_i - F_i)}{Y_i + F_i} \right| \cdot 100\%$$

Είναι μία παραλλαγή του MAPE, το οποίο αν και λέγεται συμμετρικό φαίνεται ότι μεταχειρίζεται διαφορετικά τις προβλέψεις ανάλογα με το αν είναι αισιόδοξες ή απαισιόδοξες.

Μέσο απόλυτο κανονικοποιημένο σφάλμα (Mean Absolute Scaled Error):

$$MASE = \frac{\frac{1}{n} \sum_{i=1}^n |Y_i - F_i|}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|}$$

Το MASE είναι ένα σχετικό σφάλμα που χρησιμοποιεί ως αναφορά την απλοϊκή μέθοδο (naïve). Προκύπτει από το μέσο απόλυτο σφάλμα (MAE) του μοντέλου πρόβλεψης που εφαρμόζεται, κανονικοποιημένο με το MAE που προκύπτει από την εφαρμογή της μεθόδου Naïve στα διαθέσιμα (in-sample) δεδομένα.

Ανεξάρτητα από το ποιος τύπος σφάλματος επιλέγεται για την αξιολόγηση ενός μοντέλου στόχος είναι η ελαχιστοποίηση αυτού.

Στην παρούσα εργασία για τη μέτρηση της ακρίβειας της πρόβλεψης επιλέχθηκε ο στατιστικός δείκτης το MASE. Το 2006 οι Hyndman και Koehler πρότειναν το MASE ως ένα γενικά

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων εφαρμόσιμο εργαλείο μέτρησης της ακρίβειας των προβλέψεων χωρίς τα προβλήματα που παρατηρούνται σε άλλους δείκτες σφαλμάτων. Γενικά, υποστήριξαν τη χρήση των κανονικοποιημένων σφαλμάτων ως μόνιμη προσέγγιση σε περιπτώσεις που θέλουμε να μετρήσουμε την ακρίβεια μεταξύ χρονοσειρών διαφορετικού επιπέδου ενώ ο Hyndman θεωρεί ότι το MASE είναι ο καλύτερος στατιστικός δείκτης μέτρησης ακρίβειας σε περιπτώσεις διακοπτόμενης ζήτησης και όχι μόνο [[38]].

3.4.2 Κυλιόμενη πρόβλεψη

Οι στατιστικοί δείκτες μπορούν να εφαρμοστούν είτε στα διαθέσιμα δεδομένα και τις προβλέψεις που δίνει το μοντέλο για αυτές τις τιμές (in-sample error) είτε στις προβλεπόμενες τιμές εφόσον αυτές γίνουν γνωστές (out-of-sample error). Γενικότερα, για την αξιολόγηση των προβλέψεων προτιμάται από τους αναλυτές η χρήση των out-of-sample τιμών γιατί μας ενδιαφέρει πιο πολύ το πώς εξελίσσονται τα δεδομένα στο μέλλον κάτι το οποίο μπορεί να μην αντικατοπτρίζεται στα ιστορικά δεδομένα [[36]]. Επομένως, μπορεί ένα μοντέλο να υπερπροσαρμόζεται στα ιστορικά δεδομένα δίνοντας μικρότερα σφάλματα αλλά προβλεπτικά να μην λειτουργεί καλά. Επειδή, όμως, τη στιγμή που πραγματοποιείται η πρόβλεψη οι τιμές αυτές δεν είναι ακόμα γνωστές, οι αναλυτές συνηθίζουν να χρησιμοποιούν κάποιες ιστορικές παρατηρήσεις ως out-of-sample. Επομένως, κατά την εφαρμογή ενός μοντέλου προβλέψεων, τα διαθέσιμα δεδομένα χωρίζονται σε δύο μέρη, το σύνολο εκμάθησης (train set) και το σύνολο δοκιμής (test set). Το train set χρησιμοποιείται για την εκτίμηση των παραμέτρων και την επιλογή του μοντέλου ενώ το test set είναι το σύνολο των δεδομένων out-of-sample που θα χρησιμοποιηθούν για την αξιολόγηση της αποτελεσματικότητάς του. Αν η διαδικασία αυτή εφαρμοστεί μία φορά για συγκεκριμένο χρονικό ορίζοντα τότε η μέθοδος αξιολόγησης ονομάζεται fixed origin evaluation. Ωστόσο, αν για τον ίδιο χρονικό ορίζοντα η διαδικασία αυτή επαναληφθεί για περισσότερες από μία φορές, με σταδιακή ολίσθηση του σημείου έναρξης των προβλέψεων, γίνεται λήψη περισσότερων σφαλμάτων και είναι δυνατή η καλύτερη κατανόηση για την απόδοση του μοντέλου. Η μέθοδος αυτή ονομάζεται rolling origin evaluation και είναι ιδιαίτερα χρήσιμη σε περιπτώσεις που οι χρονοσειρές εμφανίζουν κάποια ακραία τιμή ή αλλαγή επιπέδου (outliers ή level shifts) καθώς εξαλείφει την τυχαιότητα.

Στο σχήμα 3.2 φαίνεται ένα παράδειγμα της μεθόδου όπου για την πρόβλεψη της περιόδου 16 χρησιμοποιούνται τα πραγματικά δεδομένα έως και την περίοδο 15, για την πρόβλεψη της περιόδου 17 τα πραγματικά δεδομένα έως και την περίοδο 16 και αντίστοιχα για τις επόμενες περιόδους.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Origin = 15																									
Origin = 16																									
Origin = 17																									
Origin = 18																									
Origin = 19																									
Origin = 20																									
Origin = 21																									
Origin = 22																									
Origin = 23																									
Origin = 24																									

Σχήμα 3.2 Παράδειγμα rolling origin evaluation

4. Συλλογή, οργάνωση και επεξεργασία δεδομένων ταξινομήσεων

4.1 Εισαγωγή

Μία από τις σημαντικότερες προκλήσεις κατά τη συλλογή δεδομένων για ανάλυση είναι η επεξεργασία ημι-δομημένων αρχείων, όπως τα PDF που δημοσιεύει ο Σύνδεσμος Εισαγωγέων Αντιπροσώπων Αυτοκινήτων (ΣΕΑΑ).

Ο Σύνδεσμος Εισαγωγέων Αντιπροσώπων Αυτοκινήτων (ΣΕΑΑ) είναι ο επίσημος φορέας που εκπροσωπεί τις εισαγωγικές εταιρείες αυτοκινήτων στην Ελλάδα. Δημοσιεύει κάθε μήνα και κάθε χρόνο στατιστικά στοιχεία για τις νέες ταξινομήσεις αυτοκινήτων, επαγγελματικών οχημάτων και μοτοσυκλετών. Τα δεδομένα του θεωρούνται αξιόπιστα, γιατί προέρχονται από επίσημες δημόσιες υπηρεσίες και ελέγχονται από τις ίδιες τις εταιρείες του κλάδου.

Τα στοιχεία του ΣΕΑΑ είναι οργανωμένα ιεραρχικά, κάτι που βοηθά στην ανάλυση και την πρόβλεψη των τάσεων της αγοράς. Η ιεραρχία αυτή ξεκινά από το συνολικό αριθμό των ταξινομήσεων και «σπάει» σε επιμέρους επίπεδα, όπως:

- τον τύπο οχήματος (SUV ή όχι, sport, combi),
- την κατηγορία (όπως μικρά, μεσαία, πολυτελή),
- τη μάρκα (π.χ. Toyota, Peugeot),
- το μοντέλο (π.χ. Toyota Yaris, Peugeot 208),
- Την γεωγραφική περιοχή (π.χ. Αττική, Θεσσαλονίκη)

Αυτή η δομή ταιριάζει πολύ καλά με τα ιεραρχικά μοντέλα χρονοσειρών (Hierarchical Time Series – HTS), τα οποία επιτρέπουν την πρόβλεψη σε πολλά επίπεδα ταυτόχρονα. Για παράδειγμα, μπορούμε να προβλέψουμε πρώτα το σύνολο της αγοράς και στη συνέχεια να δούμε πώς μοιράζεται αυτή η πρόβλεψη ανά κατηγορία, μάρκα ή μοντέλο. Ή μπορούμε να ξεκινήσουμε από κάτω προς τα πάνω, προβλέποντας τα μοντέλα και προσθέτοντάς τα για να βγάλουμε τη συνολική εικόνα.

Παρότι τα δεδομένα του ΣΕΑΑ είναι αξιόπιστα, υπάρχει η ανάγκη να μετατραπούν από PDF σε επεξεργάσιμη μορφή, να οργανωθούν σωστά, και να γίνει μια επιπλέον επεξεργασία στην κατηγοριοποίηση (π.χ. να ξεκαθαριστεί αν ένα μοντέλο ανήκει στην κατηγορία SUV ή όχι). Όλα αυτά είναι απαραίτητα βήματα για να αξιοποιηθούν σωστά σε πιο προχωρημένες αναλύσεις και προβλέψεις.

Στο πλαίσιο της παρούσας εργασίας αναπτύχθηκε ένα αυτοματοποιημένο σύστημα στην γλώσσα προγραμματισμού Python, το οποίο:

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

- Κατεβάζει αυτόματα τα απαραίτητα αρχεία PDF από τον ιστότοπο του ΣΕΑΑ.
- Εξάγει τα δεδομένα των πινάκων από τα αρχεία.
- Καθαρίζει και οργανώνει τα δεδομένα.
- Τα μετατρέπει σε ενιαία, επεξεργάσιμη μορφή Excel.

Η λύση αυτή επιτρέπει την ταχεία και αξιόπιστη ανάλυση των δεδομένων για περισσότερα από 20 έτη.

4.2 Παραδείγματα αρχείων PDF από ΣΕΑΑ



PC REGISTRATIONS BY MONTH ΤΑΞΙΝΟΜΗΣΕΙΣ ΕΠΙΒΑΤΙΚΩΝ ΟΧΗΜΑΤΩΝ ΑΝΑ ΜΗΝΑ

DECEMBER '01	YTD																											
Make	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	TOTAL	Make														
OPEL	3154	10.7%	2585	11.3%	3031	12.1%	2448	9.7%	3146	10.9%	2920	10.5%	2971	10.2%	2022	9.3%	1795	9.8%	2202	10.8%	1792	9.5%	958	7.6%	29024	10.4%	OPEL	
HYUNDAI	2740	9.3%	1830	8.0%	1974	7.9%	2375	9.4%	2236	7.8%	2863	10.3%	3249	11.1%	2372	10.9%	1870	10.2%	1693	8.3%	1625	8.6%	1019	8.1%	25846	9.2%	HYUNDAI	
FIAT	2096	7.1%	2031	8.9%	2191	8.8%	2398	9.5%	3260	11.3%	2327	8.4%	2238	7.7%	1684	7.7%	1289	7.0%	1464	7.2%	1039	5.5%	1046	8.3%	23063	8.2%	FIAT	
TOYOTA	3261	11.1%	2031	8.9%	2412	9.6%	1902	7.5%	2165	7.5%	2105	7.6%	2231	7.6%	1818	8.3%	1325	7.2%	1344	6.6%	1232	6.5%	795	6.3%	22621	8.1%	TOYOTA	
VOLKS WAGEN	2634	8.9%	1708	7.5%	1771	7.1%	1725	6.8%	1761	6.1%	1906	6.9%	1841	6.3%	1318	6.0%	1281	7.0%	1427	7.0%	1195	6.3%	977	7.8%	19544	7.0%	VOLKS WAGEN	
CITROEN	1910	6.5%	1738	7.6%	1948	7.8%	1765	7.0%	1874	6.5%	2103	7.6%	2130	6.3%	1303	6.0%	944	5.1%	1527	7.5%	1412	7.5%	353	2.8%	18707	6.7%	CITROEN	
SEAT	1651	5.6%	1815	8.0%	1610	6.4%	1587	6.3%	1800	6.3%	1707	6.2%	2145	7.3%	1549	7.1%	1062	5.8%	1258	6.2%	1195	6.3%	569	4.5%	17948	6.4%	SEAT	
PEUGEOT	1912	6.5%	1200	5.3%	1250	5.0%	1172	4.6%	1273	4.4%	1646	5.9%	1794	6.1%	1444	6.6%	1457	7.9%	1688	8.3%	1798	9.5%	1210	9.6%	17844	6.4%	PEUGEOT	
RENAULT	1025	3.5%	1196	5.3%	1444	5.8%	1460	5.8%	1363	4.7%	1421	5.1%	1457	5.0%	977	4.5%	1028	5.6%	1122	5.5%	1008	5.3%	630	5.0%	14131	5.0%	RENAULT	
NISSAN	1491	5.1%	885	3.9%	935	3.7%	1105	4.4%	1178	4.1%	1111	4.0%	1115	3.8%	931	4.3%	802	4.4%	1077	5.3%	808	4.3%	672	5.4%	12110	4.3%	NISSAN	
FORD	879	3.0%	637	2.8%	824	3.3%	1398	5.5%	1060	3.7%	1297	4.7%	1416	4.8%	1081	4.9%	988	5.4%	752	3.7%	781	4.1%	734	5.8%	11847	4.2%	FORD	
DAEWOO	979	3.3%	669	2.9%	848	3.4%	999	4.0%	1808	6.3%	994	3.6%	1079	3.7%	1087	5.0%	867	4.7%	919	4.5%	628	3.3%	376	3.0%	11253	4.0%	DAEWOO	
SUZUKI	946	3.2%	652	2.9%	768	3.1%	1018	4.0%	1317	4.6%	1096	4.0%	1002	3.4%	808	3.7%	605	3.3%	751	3.7%	750	4.0%	497	4.0%	10210	3.6%	SUZUKI	
SKODA	936	3.2%	507	2.2%	721	2.9%	750	3.0%	781	2.7%	668	2.4%	1044	3.6%	755	3.5%	541	3.0%	390	1.9%	697	3.7%	636	5.1%	8426	3.0%	SKODA	
AUDI	425	1.4%	486	2.1%	500	2.0%	440	1.7%	475	1.7%	565	2.0%	488	1.7%	357	1.6%	346	1.9%	365	1.8%	361	1.9%	278	2.2%	5086	1.8%	AUDI	
ALFA ROMEO	544	1.8%	516	2.3%	513	2.0%	390	1.5%	508	1.8%	420	1.5%	429	1.5%	273	1.2%	245	1.3%	359	1.8%	353	1.9%	185	1.5%	4735	1.7%	ALFA ROMEO	
MERCEDES	393	1.3%	372	1.6%	368	1.5%	449	1.8%	529	1.8%	489	1.8%	479	1.6%	339	1.6%	321	1.8%	357	1.7%	332	1.8%	244	1.9%	4672	1.7%	MERCEDES	
B.M.W.	532	1.8%	333	1.5%	411	1.6%	403	1.6%	376	1.3%	357	1.3%	433	1.5%	343	1.6%	324	1.8%	327	1.6%	334	1.8%	288	2.3%	4461	1.6%	B.M.W.	
HONDA	219	0.7%	264	1.2%	259	1.0%	224	0.9%	281	1.0%	260	0.9%	337	1.2%	177	0.8%	220	1.2%	231	1.1%	250	1.3%	120	1.0%	2842	1.0%	HONDA	
KIA MOTORS	331	1.1%	262	1.2%	97	0.4%	158	0.6%	309	1.1%	217	0.8%	320	1.1%	260	1.2%	179	1.0%	185	0.9%	325	1.7%	146	1.2%	2789	1.0%	KIA MOTORS	
ROVER	319	1.1%	229	1.0%	303	1.2%	244	1.0%	239	0.8%	215	0.8%	267	0.9%	164	0.8%	148	0.8%	139	0.7%	134	0.7%	180	1.4%	2581	0.9%	ROVER	
MITSUBISHI	205	0.7%	247	1.1%	196	0.8%	164	0.6%	205	0.7%	211	0.8%	231	0.8%	201	0.9%	138	0.8%	132	0.6%	169	0.9%	118	0.9%	2217	0.8%	MITSUBISHI	
LANCIA	161	0.5%	121	0.5%	134	0.5%	91	0.4%	105	0.4%	105	0.4%	92	0.3%	88	0.4%	73	0.4%	111	0.5%	72	0.4%	44	0.4%	1197	0.4%	LANCIA	
SMART	125	0.4%	45	0.2%	105	0.4%	144	0.6%	150	0.5%	100	0.4%	105	0.4%	68	0.3%	99	0.5%	109	0.5%	74	0.4%	72	0.6%	1196	0.4%	SMART	
LADA	96	0.3%	31	0.1%	33	0.1%	41	0.2%	95	0.3%	145	0.5%	175	0.6%	92	0.4%	92	0.5%	110	0.5%	89	0.5%	71	0.6%	1070	0.4%	LADA	
DAIHATSU	107	0.4%	81	0.4%	100	0.4%	84	0.3%	107	0.4%	90	0.3%	89	0.3%	51	0.2%	55	0.3%	74	0.4%	74	0.4%	53	0.4%	965	0.3%	DAIHATSU	
MAZDA	84	0.3%	59	0.3%	62	0.2%	75	0.3%	81	0.3%	92	0.3%	89	0.3%	60	0.3%	51	0.3%	74	0.4%	76	0.4%	89	0.7%	892	0.3%	MAZDA	
CHRYSLER	73	0.2%	90	0.4%	56	0.2%	78	0.3%	87	0.3%	67	0.2%	74	0.3%	74	0.3%	43	0.2%	86	0.4%	85	0.5%	56	0.4%	869	0.3%	CHRYSLER	
VOLVO	70	0.2%	60	0.3%	62	0.2%	64	0.3%	94	0.3%	46	0.2%	63	0.2%	34	0.2%	30	0.2%	44	0.2%	47	0.2%	30	0.2%	644	0.2%	VOLVO	
SUBARU	63	0.2%	42	0.2%	39	0.2%	55	0.2%	66	0.2%	64	0.2%	63	0.2%	43	0.2%	48	0.3%	41	0.2%	41	0.2%	55	0.4%	620	0.2%	SUBARU	
SAAB	32	0.1%	30	0.1%	29	0.1%	44	0.2%	19	0.1%	39	0.1%	49	0.2%	23	0.1%	24	0.1%	29	0.1%	30	0.2%	23	0.2%	371	0.1%	SAAB	
JAGUAR	8	0.0%	5	0.0%	5	0.0%	12	0.0%	7	0.0%	17	0.1%	23	0.1%	17	0.1%	18	0.1%	27	0.1%	12	0.1%	10	0.1%	161	0.1%	JAGUAR	
LEXUS	5	0.0%	6	0.0%	8	0.0%	8	0.0%	2	0.0%	15	0.1%	8	0.0%	13	0.1%	1	0.0%	6	0.0%	5	0.0%	8	0.1%	85	0.0%	LEXUS	
PORSCHE	7	0.0%	4	0.0%	9	0.0%	6	0.0%	9	0.0%	7	0.0%	12	0.0%	8	0.0%	3	0.0%	7	0.0%	7	0.0%	2	0.0%	81	0.0%	PORSCHE	
GM	3	0.0%	7	0.0%	7	0.0%	4	0.0%	2	0.0%	2	0.0%	1	0.0%	5	0.0%	5	0.0%	8	0.0%	6	0.0%	6	0.0%	56	0.0%	GM	
MINI																	6	0.0%	10	0.0%	11	0.1%	9	0.1%	36	0.0%	MINI	
ZASTAVA	20	0.1%																							20	0.0%	ZASTAVA	
ISUZU	1	0.0%	1	0.0%	4	0.0%			1	0.0%	1	0.0%	1	0.0%	1	0.0%	2	0.0%					1	0.0%	13	0.0%	ISUZU	
ASIA MOTOR							4	0.0%	6	0.0%	2	0.0%													12	0.0%	ASIA MOTOR	
FERRARI	3	0.0%	1	0.0%							1	0.0%	5	0.0%					2	0.0%					12	0.0%	FERRARI	
MASERATI			1	0.0%	2	0.0%			2	0.0%			2	0.0%					3	0.0%	1	0.0%			11	0.0%	MASERATI	
LOTUS			1	0.0%					2	0.0%			2	0.0%	1	0.0%	1	0.0%	2	0.0%	2	0.0%			11	0.0%	LOTUS	
TATA			1	0.0%	1	0.0%							1	0.0%			3	0.0%					1	0.0%	7	0.0%	TATA	
DACIA			1	0.0%					1	0.0%							2	0.0%	1	0.0%					5	0.0%	DACIA	
PIAGGIO					1	0.0%							1	0.0%											2	0.0%	PIAGGIO	
LAMBORGHINI					1	0.0%																			1	0.0%	LAMBORGHINI	
AIXAM																									1	0.0%	AIXAM	
TOTAL/ ΣΥΝΟΛ	29440		22780		25032		25284		28780		27691		29251		21842		18331		20453		18851		12560		280295			

Σχήμα 4.1 Δείγμα του αρχείου PDF από ΣΕΑΑ ανά μήνα και μάρκα

Το αρχείο PDF δημοσιεύεται μηνιαία μέσω του ιστότοπου του ΣΕΑΑ και αφορά τα επιβατικά αυτοκίνητα. Συνήθως αποτελείται από δύο σελίδες. Στην πρώτη και τελευταία στήλη αναφέρεται η μάρκα. Στις ενδιάμεσες στήλες αναφέρονται οι μηνιαίες ταξινομήσεις και το ποσοστό κάθε μάρκας μέσα στον μήνα. Στην προτελευταία στήλη αναφέρονται οι συνολικές ταξινομήσεις και το ποσοστό μάρκας έως την δεδομένη στιγμή του έτους.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων



PC REGISTRATIONS BY TERRITORY ΜΗΝΙΑΙΕΣ ΤΑΞΙΝΟΜΗΣΕΙΣ ΚΑΤΑ ΠΕΡΙΟΧΗ													
Make	ATHENS	SALONICA	EPIDRUS	THESSALY	THRACE	CRETE	MACEDONIA	AEGEAN ISL	IONIAN ISL	PELOPON.	CENTRAL GR.	TOTAL	Make
TOYOTA	716 11,7%	115 18,9%	17 9,9%	21 10,8%	17 17,7%	66 13,9%	79 22,2%	43 21,6%		41 16,7%	51 30,4%	1166 13,4%	TOYOTA
BMW	608 9,9%	61 10,0%	9 5,3%	26 13,3%		19 4,0%	38 10,7%	1 0,5%	2 2,8%	11 4,5%	1 0,6%	776 8,9%	BMW
SUZUKI	284 4,6%	39 6,4%	18 10,5%	32 16,4%	9 9,4%	35 7,4%	30 8,4%	47 23,6%	20 27,8%	38 15,4%	13 7,7%	565 6,5%	SUZUKI
DACIA	330 5,4%	55 9,0%	11 6,4%	24 12,3%	24 25,0%	24 5,1%	32 9,0%	5 2,5%	5 6,9%	26 10,6%	26 15,5%	562 6,5%	DACIA
RENAULT	445 7,3%	37 6,1%	11 6,4%	9 4,6%	9 9,4%	8 1,7%	20 5,6%			9 3,7%	7 4,2%	555 6,4%	RENAULT
HYUNDAI	242 4,0%	25 4,1%	4 2,3%	10 5,1%	5 5,2%	105 22,1%	24 6,7%	25 12,6%	14 19,4%	14 5,7%	12 7,1%	480 5,5%	HYUNDAI
VOLKSWAGEN	309 5,1%	20 3,3%	12 7,0%	2 1,0%	8 8,3%	19 4,0%	10 2,8%	15 7,5%	1 1,4%	9 3,7%	11 6,5%	416 4,8%	VOLKSWAGEN
VOLVO	313 5,1%	27 4,4%	9 5,3%	3 1,5%		9 1,9%	14 3,9%			4 1,6%		379 4,4%	VOLVO
MINI	297 4,9%	16 2,6%		9 4,6%		9 1,9%	3 0,8%			3 1,2%		337 3,9%	MINI
NISSAN	196 3,2%	9 1,5%	21 12,3%	7 3,6%	5 5,2%	29 6,1%	12 3,4%	11 5,5%	16 22,2%	21 8,5%	4 2,4%	331 3,8%	NISSAN
BYD	216 3,5%	42 6,9%				40 8,4%	5 1,4%		1 1,4%	14 5,7%		318 3,7%	BYD
PEUGEOT	195 3,2%	11 1,8%	4 2,3%	6 3,1%	2 2,1%	8 1,7%	2 0,6%	9 4,5%	3 4,2%	9 3,7%	6 3,6%	255 2,9%	PEUGEOT
AUDI	166 2,7%	16 2,6%	12 7,0%		1 1,0%	22 4,6%	11 3,1%	13 6,5%		5 2,0%	5 3,0%	251 2,9%	AUDI
MERCEDES	214 3,5%	11 1,8%		5 2,6%	4 4,2%	3 0,6%	2 0,6%	2 1,0%		3 1,2%	1 0,6%	245 2,8%	MERCEDES
FORD	197 3,2%	12 2,0%	3 1,8%	1 0,5%		6 1,3%	3 0,8%	7 3,5%		8 3,3%	3 1,8%	240 2,8%	FORD
MG	163 2,7%	5 0,8%	4 2,3%	7 3,6%	2 2,1%	18 3,8%	21 5,9%	4 2,0%	1 1,4%	4 1,6%	3 1,8%	232 2,7%	MG
KIA	129 2,1%	12 2,0%	7 4,1%	4 2,1%	2 2,1%	9 1,9%	15 4,2%	9 4,5%	3 4,2%	5 2,0%	8 4,8%	203 2,3%	KIA
TESLA	148 2,4%	32 5,3%		1 0,5%		1 0,2%			1 1,4%	2 0,8%		185 2,1%	TESLA
JEEP	127 2,1%	6 1,0%	2 1,2%	2 1,0%		12 2,5%	4 1,1%	4 2,0%	1 1,4%	4 1,6%	1 0,6%	163 1,9%	JEEP
OPEL	106 1,7%	13 2,1%	15 8,8%	3 1,5%	3 3,1%	5 1,1%	8 2,2%			4 1,6%	2 1,2%	159 1,8%	OPEL
CITROEN/DS	96 1,6%	7 1,2%	2 1,2%	11 5,6%	2 2,1%	9 1,9%	10 2,8%	2 1,0%		5 2,0%	3 1,8%	147 1,7%	CITROEN/DS
CUPRA	96 1,6%	4 0,7%		3 1,5%		4 0,8%	1 0,3%				5 3,0%	113 1,3%	CUPRA
SKODA	88 1,4%	4 0,7%	3 1,8%	5 2,6%	3 3,1%	2 0,4%				1 0,4%	2 1,2%	108 1,2%	SKODA
FIAT	98 1,6%					2 0,4%	1 0,3%		3 4,2%	3 1,2%		107 1,2%	FIAT
MAZDA	60 1,0%	3 0,5%	1 0,6%			2 0,4%					1 0,6%	67 0,8%	MAZDA
LEXUS	50 0,8%	5 0,8%				1 0,2%			1 1,4%			57 0,7%	LEXUS
HONDA	34 0,6%	6 1,0%		1 0,5%			2 0,6%	2 1,0%				45 0,5%	HONDA
ALFA ROMEO	28 0,5%	2 0,3%	1 0,6%	3 1,5%		2 0,4%	5 1,4%			1 0,4%		42 0,5%	ALFA ROMEO
LAND ROVER	40 0,7%	1 0,2%										41 0,5%	LAND ROVER
LYNK & CO	30 0,5%	1 0,2%	3 1,8%									34 0,4%	LYNK & CO
PORSCHE	27 0,4%	5 0,8%										32 0,4%	PORSCHE
SEAT	18 0,3%	5 0,8%				4 0,8%	2 0,6%				1 0,6%	30 0,3%	SEAT
LEAPMOTOR	14 0,2%	1 0,2%	2 1,2%			1 0,2%	2 0,6%			2 0,8%	2 1,2%	24 0,3%	LEAPMOTOR
SMART	16 0,3%											16 0,2%	SMART
JAGUAR	11 0,2%											11 0,1%	JAGUAR
MITSUBISHI	4 0,1%					1 0,2%						5 0,1%	MITSUBISHI
SUBARU	3 0,0%											3 0,0%	SUBARU
MASERATI	2 0,0%											2 0,0%	MASERATI
LOTUS	1 0,0%											1 0,0%	LOTUS
VOYAH	1 0,0%											1 0,0%	VOYAH
TOTAL	6.118	608	171	195	96	475	356	199	72	246	168	8.704	
												8.240	DECEMBER '23
												5,6%	% D24/23

Σχήμα 4.2 Δείγμα του αρχείου PDF από ΣΕΑΑ ανά περιοχή και μάρκα

Το αρχείο PDF δημοσιεύεται μηνιαία μέσω του ιστότοπου του ΣΕΑΑ και αφορά τα επιβατικά αυτοκίνητα. Συνήθως αποτελείται από δύο σελίδες. Στην πρώτη και τελευταία στήλη αναφέρεται η μάρκα. Στις ενδιάμεσες στήλες αναφέρονται οι συνολικές ταξινομήσεις έως την δεδομένη στιγμή του έτους και το ποσοστό κάθε μάρκας, χωρισμένες ανά γεωγραφική περιοχή. Στην προτελευταία στήλη αναφέρονται οι συνολικές ταξινομήσεις και το ποσοστό κάθε μάρκας έως την δεδομένη στιγμή του Χρόνου.



NEW PC REGISTRATIONS BY SEGMENT ΤΑΞΙΝΟΜΗΣΕΙΣ ΕΠΙΒΑΤΙΚΩΝ ΚΑΤΑ ΚΑΤΗΓΟΡΙΑ

DECEMBER '16

		YTD		DECEMBER '16	
Segment:		TOTAL	Market Share	TOTAL	Market Share
DUAL PURPOSE		8.170	10,4%	627	13,2%
Make:	Range:	Volume:	%	Volume:	%
SUZUKI	VITARA	1.179	14,43%	41	6,5%
BMW	X1-SERIES	1.140	13,95%	172	27,4%
JEEP	RENEGADE	892	10,92%	81	12,9%
AUDI	Q3	580	7,10%	20	3,2%
VOLKSWAGEN	TIGUAN	525	6,43%	61	9,7%
DACIA	DUSTER	508	6,22%	21	3,3%
MINI	COUNTRYMAN	408	4,99%	27	4,3%
FIAT	500X	365	4,47%	7	1,1%
NISSAN	X-TRAIL	272	3,33%	4	0,6%
SUZUKI	SX4	267	3,27%	16	2,6%
VOLVO	XC60	217	2,66%	28	4,5%
VOLVO	XC90	201	2,46%	27	4,3%
HONDA	CR-V	188	2,30%	9	1,4%
BMW	X3-SERIES	181	2,22%	8	1,3%
MERCEDES	GL-SERIES	157	1,92%	17	2,7%
SUZUKI	JIMNY	146	1,79%		
SSANGYONG	TIVOLI	122	1,49%	3	0,5%
FORD	KUGA	105	1,29%	19	3,0%
AUDI	Q5	95	1,16%	3	0,5%
BMW	X4-SERIES	83	1,02%	7	1,1%
RENAULT	KADJAR	81	0,99%	12	1,9%
TOYOTA	RAV4	74	0,91%	5	0,8%
BMW	X5-SERIES	57	0,70%	4	0,6%
AUDI	Q2	35	0,43%	8	1,3%

Σχήμα 4.3 Δείγμα του αρχείου PDF από ΣΕΑΑ ανά μοντέλο, μάρκα και κατηγορία

Το αρχείο PDF δημοσιεύεται μηνιαία μέσω του ιστότοπου του ΣΕΑΑ και αφορά τα επιβατικά αυτοκίνητα. Συνήθως αποτελείται από πολλές σελίδες (~10). Πάνω από τον πίνακα υπάρχει μια καρτέλα μεταδεδομένων που αναφέρει την σχετική κατηγορία, τον συνολικό αριθμό ταξινομήσεων της, το μερίδιο της σε σχέση με τις συνολικές ταξινομήσεις και τον Μήνα στον οποίο βρισκόμαστε. Η καρτέλα επαναλαμβάνεται κάθε φορά που αλλάζουμε κατηγορία. Στην κατηγορία (π.χ. A, B, Sport) αναφέρεται και αν το αυτοκίνητο είναι τύπου SUV ή όχι. Οι στήλες του πίνακα κάτω από αυτήν περιέχουν την μάρκα, το μοντέλο, τις συνολικές ταξινομήσεις του και το ποσοστό του ως προς την κατηγορία, έως την δεδομένη στιγμή του έτους και στις τελευταίες στήλες τα ίδια στοιχεία μηνιαία.

4.3 Περιγραφή μεθοδολογίας

Βήμα 1: Λήψη Αρχείων PDF

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
Η διαδικασία ξεκινά με τη λήψη των αρχείων PDF, τα οποία φιλοξενούνται σε διαφορετικά URLs ανάλογα με το έτος δημοσίευσης. Για παράδειγμα, από το 2022 και μετά, τα αρχεία βρίσκονται σε διαφορετική διαδρομή απ' ό,τι την περίοδο 2004–2019. Ο κώδικας εντοπίζει δυναμικά τη σωστή διεύθυνση και κατεβάζει το αρχείο μέσω της βιβλιοθήκης *requests*.

Βήμα 2: Εξαγωγή Πινάκων

Με χρήση της βιβλιοθήκης *pdfplumber*, ο κώδικας διαβάζει κάθε σελίδα από το εκάστοτε PDF και εξάγει τους πίνακες που περιέχουν τα δεδομένα του, σε επεξεργάσιμη μορφή. Για καλύτερη ακρίβεια, χρησιμοποιούνται ρητά καθορισμένες συντεταγμένες για τις κάθετες γραμμές των πινάκων, ώστε να διασφαλιστεί η σωστή στοίχιση των στηλών.

Βήμα 3: Καθαρισμός Δεδομένων

Αφού εξάγονται οι πίνακες, εφαρμόζεται σειρά καθαρισμών:

- Αφαίρεση γραμμών που περιέχουν τίτλους, σύνολα, ποσοστά ή ελληνικούς χαρακτήρες.
- Μετατροπή των αριθμών σε αριθμητικό τύπο (αφαίρεση κόμματος).
- Προσθήκη στήλης με το έτος η/και τον Μήνα ("YEAR").
- Ενοποίηση διπλών εγγραφών ή διορθώσεις λαθών (π.χ. "VOLKS WAGEN" σε "VOLKSWAGEN").

Βήμα 4: Ενοποίηση και Εξαγωγή

Οι πίνακες από κάθε έτος συγκεντρώνονται σε έναν ενιαίο πίνακα. Τα δεδομένα ταξινομούνται κατά φθίνουσα σειρά έτους και συνολικών ταξινομήσεων, και αποθηκεύονται σε αρχείο Excel για περαιτέρω ανάλυση.

Βήμα 5: Έλεγχος της ποιότητας των δεδομένων

Αφού ολοκληρωθεί η εξαγωγή των δεδομένων, είναι δυνατόν να γίνουν διάφοροι έλεγχοι για την ποιότητα των δεδομένων κι την σωστή αποτύπωση μέσω του EXCEL:

- Έλεγχος αθροίσματος μοντέλων με σύνολο κατηγορίας
- Έλεγχος αθροίσματος Μηνών με Έτος
- Έλεγχος αθροίσματος Περιοχών με σύνολο.
- Έλεγχος ίδιων συνόλων σε διαφορετικά αρχεία.

4.4 Πλεονεκτήματα προτεινόμενης μεθοδολογίας

Η προτεινόμενη προσέγγιση για την αυτοματοποιημένη εξαγωγή δεδομένων από αρχεία PDF και τη μετατροπή τους σε δομημένη μορφή παρουσιάζει σημαντικά πλεονεκτήματα:

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

Αυτοματισμός

Ολόκληρη η διαδικασία έχει αυτοματοποιηθεί, ελαχιστοποιώντας την ανάγκη για ανθρώπινη παρέμβαση. Αυτό εξαλείφει τον κίνδυνο λαθών που ενδέχεται να προκύψουν από χειροκίνητη εισαγωγή και επιταχύνει σημαντικά τη ροή εργασίας.

Επαναχρησιμοποίηση

Ο κώδικας είναι σχεδιασμένος με τρόπο που επιτρέπει την εύκολη προσαρμογή του για τα επόμενα έτη, με ελάχιστες ή και καθόλου αλλαγές. Η προσέγγιση αυτή διασφαλίζει τη βιωσιμότητα της μεθοδολογίας και τη δυνατότητα μακροχρόνιας αξιοποίησής της.

Ακρίβεια

Με τη χρήση κατάλληλων βιβλιοθηκών και τεχνικών εξαγωγής, τα δεδομένα αντλούνται με ακρίβεια, διατηρώντας τη σωστή μορφοποίηση και ελαχιστοποιώντας τα λάθη μεταγραφής ή απώλειας πληροφορίας.

Ταχύτητα

Η αυτοματοποιημένη διαδικασία επιτρέπει τη γρήγορη συγκέντρωση μεγάλου όγκου ιστορικών δεδομένων. Ακόμα και δεδομένα από διαφορετικές χρονιές ή μορφοποιήσεις μπορούν να ενσωματωθούν αποτελεσματικά σε ελάχιστο χρόνο.

Τα παραπάνω χαρακτηριστικά καθιστούν την προσέγγιση ιδιαίτερα χρήσιμη όχι μόνο για ερευνητικούς σκοπούς, αλλά και για οργανισμούς ή επιχειρήσεις που χρειάζονται συνεχή, αξιόπιστη και έγκαιρη πληροφόρηση.

Η ανάπτυξη του συγκεκριμένου αυτοματοποιημένου εργαλείου στην Python συνέβαλε καθοριστικά στη συλλογή και οργάνωση δεδομένων ταξινομήσεων αυτοκινήτων από τον ΣΕΑΑ. Η λύση επέτρεψε τη δημιουργία ενός ενοποιημένου και αναλύσιμου συνόλου δεδομένων για τη χρονική περίοδο 2000–2024, το οποίο μπορεί να αξιοποιηθεί για στατιστική ανάλυση, οπτικοποίηση ή πρόβλεψη μελλοντικών τάσεων στην αγορά αυτοκινήτου.

4.5 Παραγόμενα σύνολα δεδομένων

Πίνακας 4.1 Δείγμα του παραγόμενου πίνακα ανά μοντέλο, μάρκα, κατηγορία και τύπο

Year	Segment	Type	Make	Model	Volume
2024	A	NONSUV	FIAT	PANDA	3579
2024	A	NONSUV	KIA	PICANTO	1888
2024	A	NONSUV	HYUNDAI	i 10	1609
2024	A	NONSUV	FIAT	500	1469
2024	B	NONSUV	MITSUBISHI	SPACE STAR	257
2024	A	NONSUV	VOLKSWAGEN	UP!	55
2024	A	NONSUV	SMART	FORTWO	48
2024	A	NONSUV	LEAPMOTOR	T03	39
2024	A	NONSUV	RENAULT	TWINGO	6
2024	A	NONSUV	TOYOTA	AYGO	5
2024	A	NONSUV	ZHIDOU	D2S	2
2024	A	NONSUV	CITROEN/DS	C1	2
2024	A	NONSUV	PEUGEOT	108	2
2024	A	NONSUV	SUZUKI	WAGON R	1
2024	A	NONSUV	TOYOTA	IQ	1
2024	A	SUV	SUZUKI	IGNIS	2954
2024	A	SUV	TOYOTA	AYGO X	2193
2024	B	SUV	DACIA	SPRING	117
2024	A	SUV	KERABOSS	SUPER K	3

Κατά την μεταποίηση, χρησιμοποιούνται τα διάφορα μεταδομένα του αντίστοιχου PDF τα οποία εισάγονται ως στήλες. Εξάγεται το αρχείο Δεκεμβρίου κάθε έτους και χρησιμοποιείται η στήλη με το σύνολο των ταξινομήσεων έτους έως και αυτόν τον μήνα. Με αυτό τον τρόπο δημιουργείται ένα επεξεργάσιμο Excel με συχνότητα έτους. Οι στήλες του είναι οι εξής: Έτος, Κατηγορία, Τύπος, Μάρκα, Μοντέλο, ετήσιος αριθμός ταξινομήσεων.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
Πίνακας 4.2 Δείγμα του παραγόμενου πίνακα ανά μάρκα και περιοχή

MAKE	AT HE NS	SAL ONI CA	EPI RU S	TH ESS AL Y	TH RA CE	CR ET E	MA CE DO NIA	AE GE AN ISL	ION IAN ISL	PE LO PO N.	CEN TRA L GR.	TOT AL	YEA R	MO NT H
TOYOTA	716	115	17	21	17	66	79	43		41	51	1166	2024	12
BMW	608	61	9	26		19	38	1	2	11	1	776	2024	12
SUZUKI	284	39	18	32	9	35	30	47	20	38	13	565	2024	12
DACIA	330	55	11	24	24	24	32	5	5	26	26	562	2024	12
RENAULT	445	37	11	9	9	8	20			9	7	555	2024	12
HYUNDAI	242	25	4	10	5	105	24	25	14	14	12	480	2024	12
VOLKSWA GEN	309	20	12	2	8	19	10	15	1	9	11	416	2024	12
VOLVO	313	27	9	3		9	14			4		379	2024	12
MINI	297	16		9		9	3			3		337	2024	12
NISSAN	196	9	21	7	5	29	12	11	16	21	4	331	2024	12
BYD	216	42				40	5		1	14		318	2024	12
PEUGEOT	195	11	4	6	2	8	2	9	3	9	6	255	2024	12
AUDI	166	16	12		1	22	11	13		5	5	251	2024	12
MERCEDE S	214	11		5	4	3	2	2		3	1	245	2024	12
FORD	197	12	3	1		6	3	7		8	3	240	2024	12

Το αντίστοιχο PDF έχει μετατραπεί σε Excel σχεδόν αυτούσια αλλά έχουν αφαιρεθεί τα ποσοστά μάρκας ως προς τις συνολικές ταξινομήσεις της περιοχής. Επίσης καθώς το αρχείο εξάγεται στην μηνιαία μορφή του, έχουν προστεθεί δύο στήλες για έτος και μήνα.

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

Πίνακας 4.3 Δείγμα του παραγόμενου πίνακα ανά μάρκα και Μήνα

ΜΑΚ Ε	JAN	FEB	MA R	APR	MA Y	JUN	JUL	AUG	SEP	OCT	NOV	DE C	TOT AL	YEA R
TOY OTA	1948	1790	1478	2532	2228	3083	2309	1882	1681	1765	1656	1166	23518	2024
SUZ UKI	926	763	781	1085	1888	1054	940	660	685	724	565	565	10636	2024
PEU GEO T	1291	992	1034	1255	922	790	1080	598	718	690	756	255	10381	2024
HYU NDAI	805	860	856	998	854	922	899	594	558	540	540	480	8906	2024
CITR OEN/ DS	1376	1027	816	807	769	1080	742	255	320	126	307	147	7772	2024
BMW	529	497	537	372	680	545	580	280	647	679	631	776	6753	2024
OPEL	918	718	775	744	614	570	474	218	441	285	465	159	6381	2024
VOL KSW AGE N	573	384	483	640	562	723	327	535	276	395	791	416	6105	2024
FIAT	387	539	594	516	976	830	411	179	270	166	169	107	5144	2024
KIA	417	391	307	345	578	531	594	330	441	349	285	203	4771	2024
AUDI	430	322	378	343	461	373	449	263	225	413	461	251	4369	2024

Το αντίστοιχο PDF έχει μετατραπεί σε Excel σχεδόν αυτούσια αλλά έχουν αφαιρεθεί τα ποσοστά μάρκας ως προς τις συνολικές ταξινομήσεις του μήνα. Επίσης καθώς εξάγεται το αρχείο Δεκεμβρίου κάθε έτους, έχει προστεθεί μία στήλη που δείχνει το έτος.

5. Θεωρία ιεραρχικών προβλέψεων

Στο παρόν κεφάλαιο παρουσιάζονται κάποια στοιχεία για την θεωρία των ιεραρχικών προβλέψεων.

5.1 Ιεραρχικές προβλέψεις

Οι ιεραρχικές προβλέψεις (hierarchical forecasting) αποτελούν ένα πεδίο της ποσοτικής πρόβλεψης που στοχεύει στην παραγωγή συνεπών και αξιόπιστων προβλέψεων όταν τα δεδομένα έχουν ιεραρχική ή ομαδοποιημένη δομή. Τέτοιες δομές εμφανίζονται συχνά σε πραγματικές εφαρμογές, όπως σε γεωγραφικές κατανομές (π.χ. πωλήσεις ανά χώρα, περιφέρεια, πόλη), σε επιχειρησιακές δομές (π.χ. προϊόν, κατηγορία, συνολικές πωλήσεις), ή σε οικονομικά μεγέθη ανά τομέα της οικονομίας [0].

Η ανάγκη για ιεραρχική πρόβλεψη προκύπτει όταν οι μεμονωμένες προβλέψεις σε χαμηλότερα επίπεδα (π.χ. ανά κατηγορία) πρέπει να είναι συνεπείς με τις προβλέψεις σε υψηλότερα επίπεδα (π.χ. το σύνολο των κατηγοριών). Η συνέπεια αυτή ονομάζεται aggregate consistency, και διασφαλίζει ότι οι προβλέψεις σε όλα τα επίπεδα της ιεραρχίας ευθυγραμμίζονται μεταξύ τους [[20]].

5.1.1 Μέθοδοι ιεραρχικής πρόβλεψης

5.1.1.1 Μέθοδοι bottom up, top down και middle out

Μέθοδος Bottom-up: Σε αυτή τη μέθοδο, δημιουργούνται προβλέψεις για τα χαμηλότερα επίπεδα της ιεραρχίας (π.χ. μεμονωμένα προϊόντα ή περιοχές) και στη συνέχεια προκύπτουν οι συνολικές προβλέψεις με συνάθροιση. Πλεονέκτημά της είναι ότι ενσωματώνει την πληροφόρηση που υπάρχει σε λεπτομερή επίπεδα, ωστόσο μπορεί να είναι ασταθής όταν τα δεδομένα σε χαμηλά επίπεδα είναι θορυβώδη ή σπάνια [0].

Μέθοδος Top-down: Η αντίθετη προσέγγιση, όπου πρώτα γίνονται προβλέψεις για τα συνολικά μεγέθη και κατόπιν διανέμονται σε χαμηλότερα επίπεδα με βάση ιστορικά μερίδια ή άλλες αναλογίες. Παρότι είναι πιο σταθερή, ενδέχεται να αγνοεί σημαντική πληροφορία που υπάρχει στα υποσύνολα των δεδομένων [0].

Υπάρχουν τρεις παραλλαγές της μεθόδου :

Μέσα ιστορικά ποσοστά: Οι συνολικές προβλέψεις κατανέμονται στα επιμέρους επίπεδα βάσει του μέσου ποσοστού συμμετοχής κάθε κατηγορίας στο παρελθόν. Η μέθοδος αυτή είναι απλή και αρκετά σταθερή, αλλά ενδέχεται να μην ανταποκρίνεται σε αλλαγές στη σχετική σημασία των επιμέρους σειρών. Παρακάτω ο σχετικός τύπος [[26]]:

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

$$p_j = \frac{1}{T} \sum_{t=1}^T \frac{y_{j,t}}{y_t}$$

Ποσοστά μέσων τιμών: Μια εναλλακτική είναι η χρήση ποσοστών μέσων τιμών, όπου η κατανομή βασίζεται στους μέσους όρους των ιστορικών τιμών κάθε κατηγορίας σε σχέση με το σύνολο. Η προσέγγιση αυτή προσφέρει μεγαλύτερη σταθερότητα όταν οι επιμέρους σειρές έχουν ήπιες διακυμάνσεις.

$$p_j = \sum_{t=1}^T \frac{y_{j,t} T}{T} \div \sum_{t=1}^T \frac{y_t}{T}$$

Προβλεπόμενα ποσοστά: Οι αναλογίες κατανομής βασίζονται σε προβλέψεις για κάθε επιμέρους κατηγορία, αντί για ιστορικά δεδομένα. Με αυτόν τον τρόπο, λαμβάνονται υπόψη πιθανές μελλοντικές μεταβολές στη δομή των δεδομένων, βελτιώνοντας την ακρίβεια των προβλέψεων στα χαμηλότερα επίπεδα.

Μέθοδος Middle-out: Ενδιάμεση μέθοδος που ξεκινά από κάποιο ενδιάμεσο επίπεδο της ιεραρχίας και κινείται προς τα πάνω και κάτω. Χρησιμοποιείται κυρίως σε περιπτώσεις όπου το μεσαίο επίπεδο παρέχει πιο αξιόπιστες προβλέψεις [[22]].

5.1.1.2 Μέθοδοι βέλτιστης συνάθροισης (Optimal reconciliation)

Οι πλέον σύγχρονες μέθοδοι, όπως αυτές που αναπτύχθηκαν από τον Hyndman και τους συνεργάτες του, βασίζονται σε ένα μαθηματικό πλαίσιο στο οποίο πρώτα παράγονται ανεξάρτητες προβλέψεις για όλα τα επίπεδα και στη συνέχεια αυτές προσαρμόζονται με τη βοήθεια μαθηματικών τεχνικών, ώστε να επιτευχθεί συνέπεια σε όλη την ιεραρχία [[20],[23]].

Η μέθοδοι επιδιώκουν την ελαχιστοποίηση ενός απλού ή ζυγισμένου τετραγωνικού σφάλματος μεταξύ των αρχικών (μη συμφιλιωμένων) και των τελικών προβλέψεων, οι οποίες έχουν εναρμονιστεί έτσι ώστε να τηρούν τη δομή της ιεραρχίας.

Η διαδικασία αυτή διατυπώνεται μαθηματικά ως η ελαχιστοποίηση ενός ζυγισμένου τετραγωνικού σφάλματος, όπου τα βάρη που χρησιμοποιούνται αντανakλούν τη σχετική αξιοπιστία ή σημασία των επιμέρους χρονοσειρών.

Τα βάρη των σφαλμάτων των προβλέψεων καθορίζουν πόση εμπιστοσύνη δείχνουμε στις επιμέρους προβλέψεις όταν προσπαθούμε να τις συναθροίσουμε με την ιεραρχική δομή. Υπάρχουν τέσσερις διαφορετικοί τρόποι για τον υπολογισμό τους:

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

OLS (Ordinary Least Squares): Γίνεται κλασική ελαχιστοποίηση τετραγώνων χωρίς ζύγιση. Επιλέγοντας αυτή την παράμετρο, υποθέτουμε ίση αξιοπιστία για όλες τις προβλέψεις καθώς δεν λαμβάνει υπόψη την αβεβαιότητα ή διακύμανση των σφαλμάτων. Χρησιμοποιείται ως βασική μέθοδος αναφοράς.

WLS (Weighted Least Squares): Τα βάρη είναι αντιστρόφως ανάλογα της διακύμανσης των σφαλμάτων πρόβλεψης. Επιλέγοντας τον συγκεκριμένο υπολογισμό δίνουμε περισσότερο βάρος στις πιο ακριβείς προβλέψεις.

MINT (Minimum Trace): Βασίζεται στον πλήρη πίνακα συνδιακύμανσης των σφαλμάτων. Ελαχιστοποιεί την ίχνη της διακύμανσης των συμφιλιωμένων σφαλμάτων. Είναι η πιο στατιστικά άρτια μέθοδος, αλλά και η πιο υπολογιστικά απαιτητική. Ο πίνακας βαρών (weight matrix) μπορεί να λαμβάνει υπόψη χαρακτηριστικά όπως ο αριθμός των παρατηρήσεων ή η σταθερότητα της κάθε χρονοσειράς. Στην παρούσα εργασία, χρησιμοποιήθηκε η μέθοδος εκτιμητή σμίκρυνσης (shrinkage estimator) καθώς ο αριθμός των χρονοσειρών του τελευταίου επιπέδου είναι πολύ μεγαλύτερος από τα υπόλοιπα [[20]].

nseries: Όταν είναι αδύνατη η εκτίμηση της διασποράς των σφαλμάτων, μπορεί να χρησιμοποιηθεί η προσέγγιση της δομικής κλιμάκωσης για τη συγχώνευση των προβλέψεων με τη μέθοδο βέλτιστης συνάθροισης. Στην περίπτωση αυτή, τα βάρη που χρησιμοποιούνται δεν βασίζονται στα ίδια τα δεδομένα, αλλά αποκλειστικά στη δομή της ιεραρχίας των χρονοσειρών.

Συγκεκριμένα, θεωρείται ότι όλα τα σφάλματα των χρονοσειρών του κατώτερου επίπεδου έχουν την ίδια διασπορά και είναι ασυσχέτιστα μεταξύ τους. Ο πίνακας βαρών προκύπτει από έναν διαγώνιο πίνακα ο οποίος περιλαμβάνει στο διαγώνιό του τον αριθμό των βασικών χρονοσειρών που συμβάλλουν σε κάθε κόμβο της ιεραρχίας. Η προσέγγιση αυτή είναι ιδιαίτερα χρήσιμη όταν οι προβλέψεις έχουν παραχθεί με εμπειρικό τρόπο (π.χ. μέσω εκτίμησης από ειδικούς) και δεν είναι δυνατή η χρήση στατιστικών σφαλμάτων.

5.1.2 Χρησιμότητα και εφαρμογές

Οι ιεραρχικές προβλέψεις είναι ιδιαίτερα χρήσιμες σε τομείς όπως:

- Οικονομικές προβλέψεις: π.χ. εκτιμήσεις ΑΕΠ ανά τομέα ή γεωγραφική περιοχή.
- Εφοδιαστική αλυσίδα και λιανικό εμπόριο: π.χ. πρόβλεψη πωλήσεων ανά κατάσταση ή ανά κατηγορία προϊόντων [[24]].
- Ενέργεια και κοινής ωφέλειας οργανισμοί: π.χ. ζήτηση ηλεκτρικής ενέργειας σε επίπεδο περιφέρειας ή καταναλωτή [[25]].

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
Η χρησιμότητα των ιεραρχικών προβλέψεων δεν έγκειται μόνο στην παραγωγή συνεπών αποτελεσμάτων, αλλά και στη βελτίωση της ακρίβειας μέσω της αξιοποίησης της πληροφορίας που εμπεριέχεται σε όλα τα επίπεδα της ιεραρχίας. Η θεωρία των ιεραρχικών προβλέψεων δείχνει ότι η απλή εφαρμογή bottom-up ή top-down μεθόδων σπάνια οδηγεί στις βέλτιστες προβλέψεις, καθώς αγνοείται η συνολική πληροφορία. Αντίθετα, οι μέθοδοι συνάθροισης επιτρέπουν τη συνδυαστική αξιοποίηση της πληροφορίας, λαμβάνοντας υπόψη τη μεταβλητότητα και τη συνδιακύμανση μεταξύ των επιπέδων [[23]].

Επιπλέον, η δυνατότητα εφαρμογής των μεθόδων σε ομαδοποιημένες προβλέψεις (grouped forecasts), όπου τα δεδομένα δεν έχουν αμιγώς ιεραρχική αλλά πολυδιάστατη ομαδοποίηση (π.χ. γεωγραφία * κατηγορία προϊόντος), καθιστά το πλαίσιο ιδιαίτερα ευέλικτο [[20]].

5.2 Ομαδοποιημένες χρονοσειρές

Οι ομαδοποιημένες χρονοσειρές επεκτείνουν την ιδέα των ιεραρχικών, επιτρέποντας την οργάνωση των δεδομένων όχι μόνο σε ιεραρχική μορφή αλλά και σε πολλαπλές διαστάσεις κατηγοριοποίησης. Σε αυτήν την περίπτωση, οι χρονοσειρές δεν σχηματίζουν δέντρο (tree), όπως συμβαίνει στις αυστηρά ιεραρχικές δομές, αλλά έναν πλέγμα (grid) ταξινομήσεων, με διαφορετικές "όψεις" των δεδομένων [[26]].

Ένα κλασικό παράδειγμα είναι η πρόβλεψη πωλήσεων ενός προϊόντος ταξινομημένου τόσο κατά περιοχή (π.χ. χώρα → περιφέρεια → πόλη), όσο και κατά τύπο προϊόντος (π.χ. κατηγορία → υποκατηγορία → μοντέλο). Οι ίδιες παρατηρήσεις μπορούν να προβληθούν και να συνοψιστούν κατά μία διάσταση (π.χ. πωλήσεις ανά περιοχή) ή κατά την άλλη (π.χ. πωλήσεις ανά τύπο προϊόντος), ή και κατά συνδυασμό αυτών (π.χ. πωλήσεις ανά περιοχή και τύπο προϊόντος) [[27]].

Αυτού του είδους η δομή μπορεί να απεικονιστεί με τη βοήθεια ενός διάγραμμα συναθροίσεων (aggregation structure), το οποίο περιλαμβάνει όλες τις δυνατές ομαδοποιήσεις και τις σχετικές μεταξύ τους σχέσεις. Το ζητούμενο είναι η παραγωγή προβλέψεων που να είναι συνεπείς με όλες αυτές τις ομαδοποιήσεις ταυτόχρονα και αυτό είναι κάτι που αποτελεί σημαντική πρόκληση.

5.2.1 Μέθοδοι συνάθροισης για ομαδοποιημένες χρονοσειρές

Η μεθοδολογία που αναπτύχθηκε για τις ιεραρχικές προβλέψεις επεκτάθηκε ώστε να καλύψει και τις ομαδοποιημένες χρονοσειρές, μέσω γενίκευσης του πλαισίου συνάθροισης. Όσον αφορά τις μεθόδους βέλτιστης συνάθροισης, οι προβλέψεις για κάθε επίπεδο (ή ομαδοποίηση) δημιουργούνται αρχικά ανεξάρτητα, και στη συνέχεια συνδυάζονται με μαθηματικό τρόπο ώστε να προκύψει ένα σύνολο συνεπών προβλέψεων.

Λόγω της φύσεως τους, οι μέθοδοι top down και middle out δεν είναι δυνατόν να χρησιμοποιηθούν στις ομαδοποιημένες χρονοσειρές καθώς δεν υπάρχει αυστηρή ιεραρχία.

5.2.2 Πλεονεκτήματα και χρήσεις

Η χρήση ομαδοποιημένων χρονοσειρών είναι ιδιαίτερα ωφέλιμη όταν:

- Οι χρονοσειρές έχουν πολλαπλές ανεξάρτητες διαστάσεις, και δεν υπάρχει μία "φυσική" ιεραρχία.
- Θέλουμε να εξάγουμε προβλέψεις σε πολλαπλά επίπεδα πληροφορίας ταυτόχρονα.
- Τα δεδομένα έχουν υψηλό επίπεδο αλληλεξάρτησης μεταξύ των διαφορετικών ομαδοποιήσεων [[29]].

Οι χρήσεις τους περιλαμβάνουν εμπορικές προβλέψεις (π.χ. πωλήσεις ανά χώρα και προϊόν), δημογραφικές προβλέψεις (π.χ. πληθυσμός ανά ηλικιακή ομάδα και γεωγραφική περιοχή), και επιχειρησιακή διαχείριση πόρων (π.χ. κατανάλωση ανά κατηγορία πελάτη και περιφέρεια) [[30]].

Οι ομαδοποιημένες και οι ιεραρχικές χρονοσειρές αποτελούν εργαλεία ζωτικής σημασίας για τη δημιουργία προβλέψεων που είναι τόσο ακριβείς όσο και συνεπείς με την πραγματική δομή των δεδομένων. Η σύγχρονη θεωρία με τις συγκεκριμένες μεθόδους, προσφέρει ένα ισχυρό εργαλείο στατιστικής μοντελοποίησης για την αντιμετώπιση αυτών των προβλημάτων.

6. Παρουσίαση προσέγγισης

6.1 Εισαγωγή

Στο παρόν κεφάλαιο αναφέρεται η διαδικασία προετοιμασίας και μετασχηματισμού των δεδομένων ταξινομήσεων αυτοκινήτων, τα οποία εξήχθησαν από τον ΣΕΑΑ, με στόχο τη δημιουργία μοντέλου ομαδοποιημένων χρονοσειρών (grouped time series - GTS). Τέλος, αναφέρεται ή διαδικασία αξιολόγησης των παραγόμενων προβλέψεων από διάφορες μεθόδους του μοντέλου, προκειμένου να εκτιμηθεί η ακρίβεια και η αξιοπιστία του.

6.2 Προετοιμασία δεδομένων

Το αρχείο που χρησιμοποιήθηκε για την δημιουργία του μοντέλου, είναι εκείνο με τις ετήσιες ταξινομήσεις οχημάτων στην Ελλάδα από το 2000 έως και το 2024, ομαδοποιημένες κατά μάρκα (Make), κατηγορία αγοράς (Segment), τύπο οχήματος (Type) και μοντέλο (Model). Η αρχική επεξεργασία περιλαμβάνει:

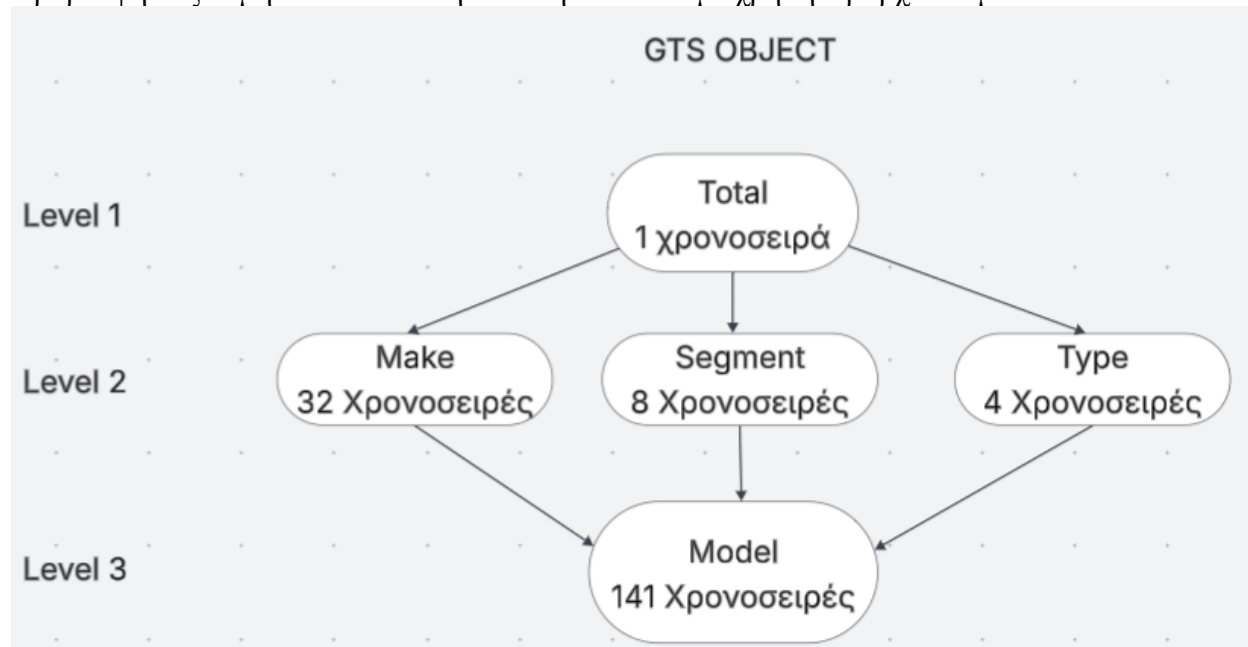
- Φόρτωση του αρχείου Excel.
- Μετατροπή των δεδομένων και δημιουργία ενός πίνακα με τις ταξινομήσεις του κατώτερου επιπέδου, δηλαδή ανά μοντέλο, κρατώντας την κωδικοποίηση ως προς την μάρκα, το κατηγορία και τον τύπο.
- Διαγραφή των μοντέλων που δεν έχουν συνεχόμενες πωλήσεις για τα τελευταία 8 χρόνια για πρακτικότητα και καλύτερη εφαρμογή του μοντέλου.

6.3 Δημιουργία αντικειμένου GTS

Μετά τον καθορισμό της ιεραρχικής δομής, δημιουργείται το αντικείμενο GTS (Grouped time series), το οποίο περιλαμβάνει το σύνολο των χρονοσειρών, καθώς και την ιεραρχία τους. Η συχνότητα είναι ετήσια, ενώ οι χρονοσειρές ξεκινούν από το έτος 2000.

Το αντικείμενο GTS είναι ένα ειδικό αντικείμενο στη γλώσσα R που οργανώνει και διαχειρίζεται χρονοσειρές με ιεραρχική ή ομαδοποιημένη δομή, επιτρέποντας την συστηματική πρόβλεψη και ανάλυση των δεδομένων σε πολλαπλά επίπεδα. Είναι θεμελιώδες για την ιεραρχική πρόβλεψη (hierarchical forecasting) και την ομαδοποιημένη πρόβλεψη (grouped forecasting).

Το αντικείμενο GTS αποτελεί πλέον την οργανωμένη βάση για την εφαρμογή μεθόδων πρόβλεψης χρονοσειρών σε όλα τα επίπεδα της ιεραρχίας, με δυνατότητα συγχρονισμένων προβλέψεων για μοντέλα, τύπους, κατηγορίες και μάρκες.



Σχήμα 6.1 Παραγόμενο αντικείμενο GTS

6.4 Αξιολόγηση μοντέλου

Για την αξιολόγηση της ακρίβειας των προβλέψεων του μοντέλου, οι παρατηρήσεις χωρίζονται σε δεδομένα εκτός δείγματος (out of sample) που περιλαμβάνει τα δεδομένα από το 2021 έως το 2024 και σε δεδομένα εκπαίδευσης (train set) που περιλαμβάνουν τα έτη 2021 έως 2024.

Χρησιμοποιείται η τεχνική Rolling Origin Evaluation κατά την οποία όλα τα πραγματικά διαθέσιμα δεδομένα χρησιμοποιούνται για την πρόβλεψη του επόμενου έτους, όπως αυτή περιγράφηκε στο τρίτο κεφάλαιο.

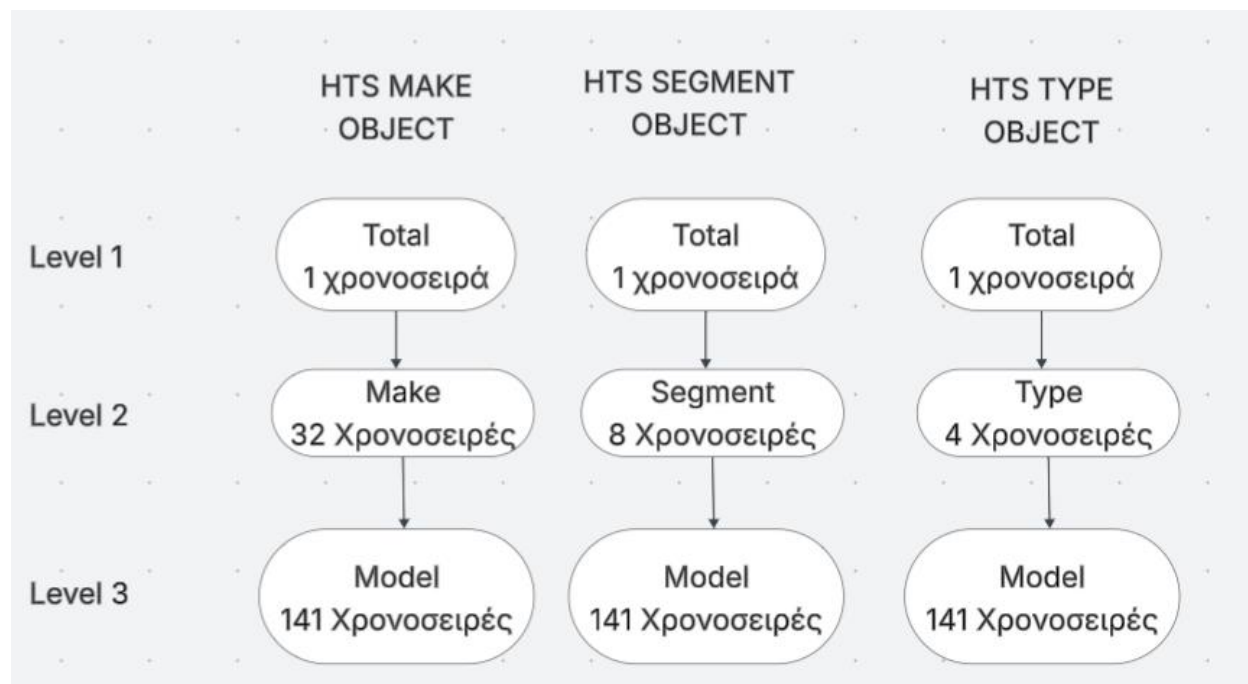
Για την αποτίμηση της απόδοσης, χρησιμοποιείται ο δείκτης MASE (Mean Absolute Scaled Error) όπως αυτός αναφέρεται στο τρίτο κεφάλαιο. Το MASE είναι ένα σχετικό σφάλμα που χρησιμοποιεί ως μέθοδο αναφοράς, την απλοϊκή μέθοδο (naïve) Προκύπτει από το μέσο απόλυτο σφάλμα MAE του μοντέλου πρόβλεψης που εφαρμόζεται, κανονικοποιημένο με το αντίστοιχο MAE που προκύπτει από την εφαρμογή της μεθόδου Naïve στα διαθέσιμα δεδομένα.

Τέλος η μέθοδος πρόβλεψης που χρησιμοποιήθηκε σε όλες τις περιπτώσεις είναι η εκθετική εξομάλυνση (ETS) όπως αυτή αναφέρεται στο τρίτο κεφάλαιο.

6.5 Σύγκριση Μεθόδων Πρόβλεψης

Για την αξιολόγηση της ακρίβειας των προβλέψεων υπολογίστηκε ο δείκτης MASE (Mean Absolute Scaled Error), για κάθε μέθοδο πρόβλεψης σε πέντε επίπεδα της ιεραρχίας: συνολικό επίπεδο (Total), μάρκα (Make), κατηγορία (Segment), τύπος (Type) και μοντέλο (Model).

Για τις μεθόδους Top-Down (TD) και Middle-Out (MO), δημιουργήθηκε η ανάγκη για κατασκευή αυστηρά ιεραρχικών μοντέλων. Κατά συνέπεια δημιουργήθηκε ένα μοντέλο με βάση τη μάρκα (Make), ένα με βάση την κατηγορία (Segment) και ένα με βάση τον τύπο οχήματος (Type).



Σχήμα 6.2 Αντικείμενα HTS

Η ανάλυση των αποτελεσμάτων πραγματοποιήθηκε ανά επίπεδο ιεραρχίας, με στόχο την εξαγωγή γενικών συμπερασμάτων σχετικά με την καταλληλότητα κάθε μεθόδου για προβλέψεις ταξινομήσεων οχημάτων σε διαφορετικά επίπεδα πληροφορίας.

Συνολικά συγκρίθηκαν τριάντα δύο (32) μέθοδοι πρόβλεψης, οι οποίες περιλαμβάνουν τις τέσσερις διαφορετικές μεθόδους ιεραρχικών προβλέψεων και τις διάφορες παραλλαγές τους εφαρμοσμένες, όπου ήταν δυνατόν, για όλα τα μοντέλα που δημιουργήθηκαν. Οι μέθοδοι που αξιολογήθηκαν όπως και οι επεξηγήσεις για τις συντομογραφίες τους φαίνονται παρακάτω:

Μέθοδοι

COMP : Βέλτιστη συνάθροιση,

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

BU: Bottom up,

TD: Top down,

MO: Middle out,

Βάρη top down

tdfp: Προβλεπόμενα ποσοστά,

tdgsa: Μέσα ιστορικά ποσοστά,

tdgsf: Ποσοστά μέσων τιμών,

Βάρη βέλτιστης συνάθροισης

mint: Minimum Trace,

nseries: nseries,

ols: Ordinary Least Squares,

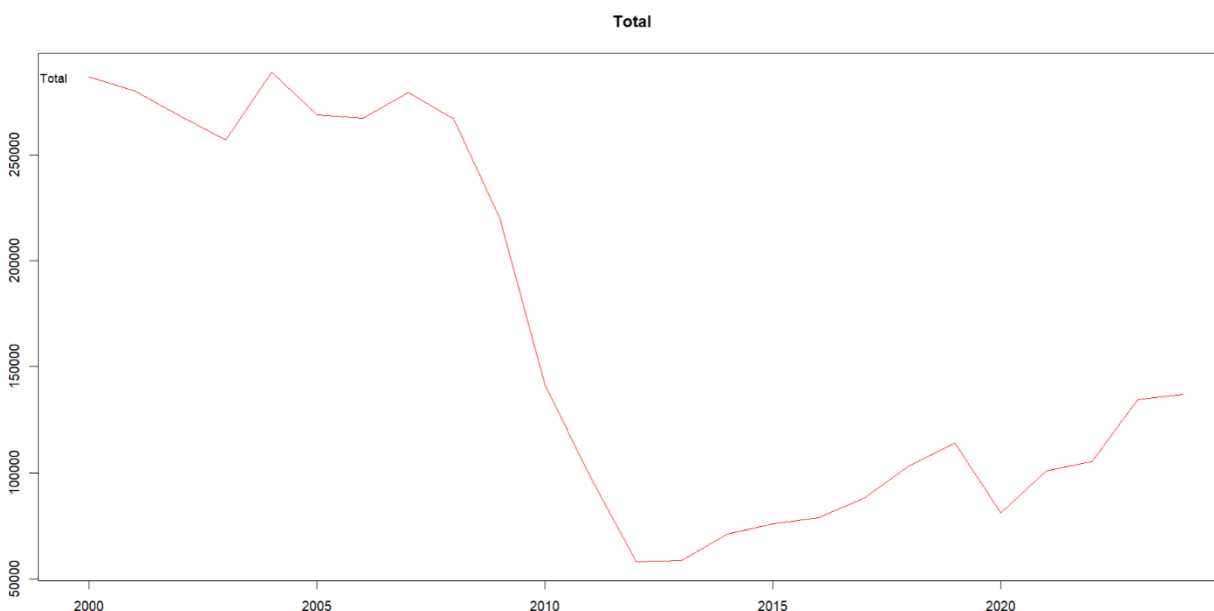
wls: Weighted Least Squares,

Model	Method	Weights
GTS	BU	
GTS	COMP	mint
GTS	COMP	nseries
GTS	COMP	ols
GTS	COMP	wls
HTS MAKE	BU	
HTS MAKE	COMP	mint
HTS MAKE	COMP	nseries
HTS MAKE	COMP	ols
HTS MAKE	COMP	wls
HTS MAKE	MO	
HTS MAKE	TD	tdfp
HTS MAKE	TD	tdgsa
HTS MAKE	TD	tdgsf
HTS SEGMENT	BU	
HTS SEGMENT	COMP	mint
HTS SEGMENT	COMP	nseries
HTS SEGMENT	COMP	ols
HTS SEGMENT	COMP	wls
HTS SEGMENT	MO	
HTS SEGMENT	TD	tdfp
HTS SEGMENT	TD	tdgsa
HTS SEGMENT	TD	tdgsf

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

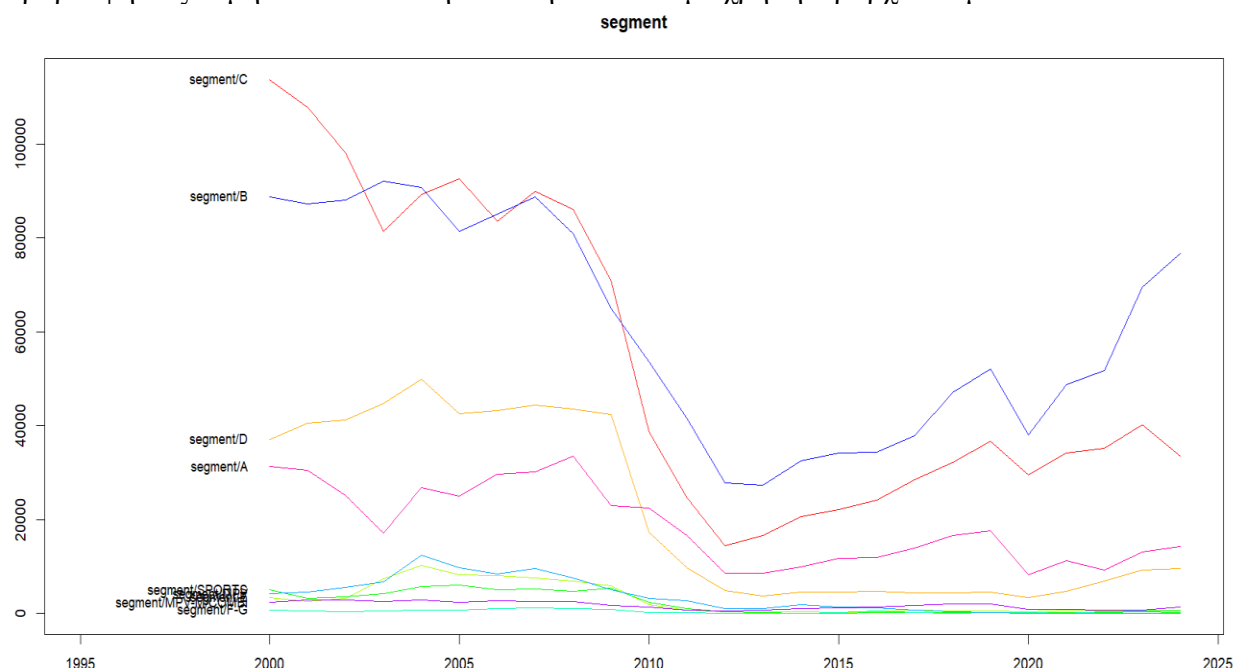
HTS TYPE	BU	
HTS TYPE	COMP	mint
HTS TYPE	COMP	nseries
HTS TYPE	COMP	ols
HTS TYPE	COMP	wls
HTS TYPE	MO	
HTS TYPE	TD	tdfp
HTS TYPE	TD	tdgsa
HTS TYPE	TD	tdgsf

6.6 Οπτικοποίηση χρονοσειρών μοντέλου



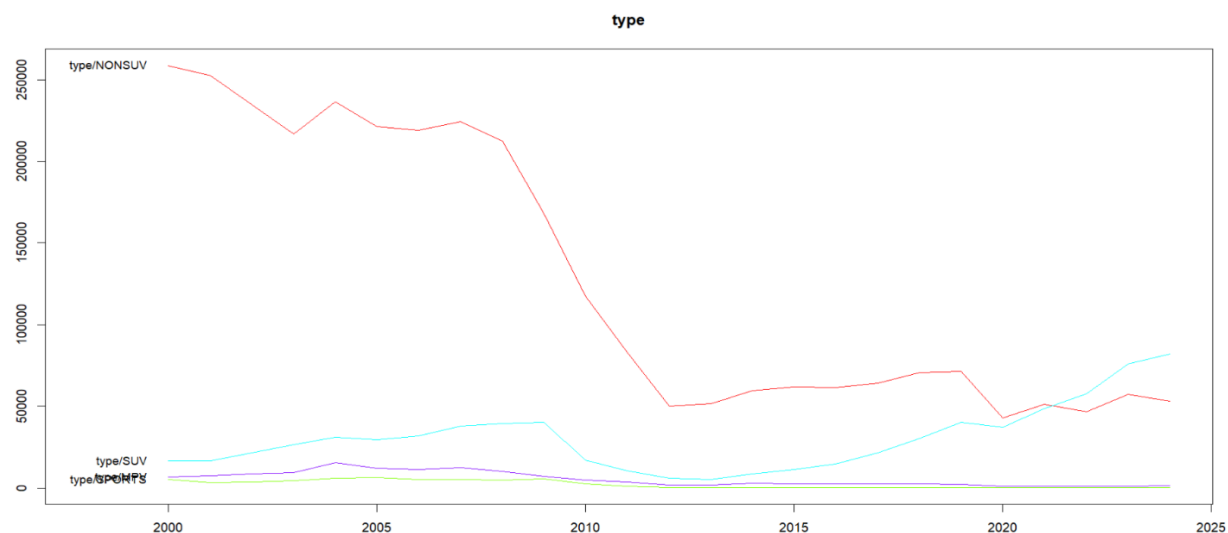
Σχήμα 6.3 Συνολικές ταξινομήσεις

Οι συνολικές ταξινομήσεις όπως οπτικοποιούνται μέσα από το αντικείμενο GTS. Στο σχήμα 6.4 φαίνονται τα ποιοτικά συμπεράσματα που αναφέρθηκαν και στο κεφάλαιο 2. Είναι φανερό το αρχικά πολύ υψηλό νούμερο των ταξινομήσεων στην Ελλάδα, η απότομη πτώση τους από το 2011 λόγω της οικονομικής κρίσης και η σταδιακή ανάκαμψη τους από το 2017. Επίσης φαίνεται η επιρροή της παγκόσμιας κρίσης το 2019 –2020.



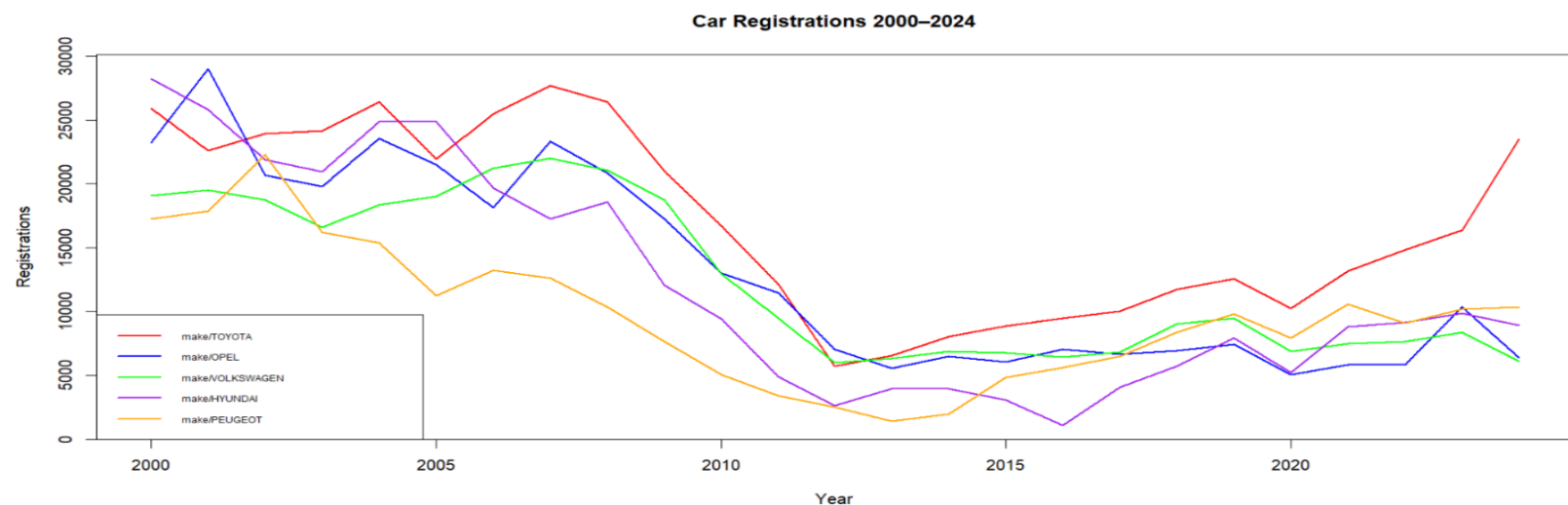
Σχήμα 6.4 Ταξινομήσεις ανά κατηγορία

Οι ταξινομήσεις ανά κατηγορία όπως οπτικοποιούνται μέσα από το αντικείμενο GTS. Οι κατηγορίες είναι: A, B, C, D, F-G, SPORTS, MPV, MPV-M/COMBI. Παρατηρούμε ότι η σταδιακή αύξηση των ταξινομήσεων οδηγείται κυρίως από την κατηγορία B που αντιπροσωπεύει τα μικρομεσαία αυτοκίνητα και λιγότερο αισθητά από την κατηγορία C που αντιπροσωπεύει τα μεσαία αυτοκίνητα.



Σχήμα 6.5 Ταξινομήσεις ανά τύπο

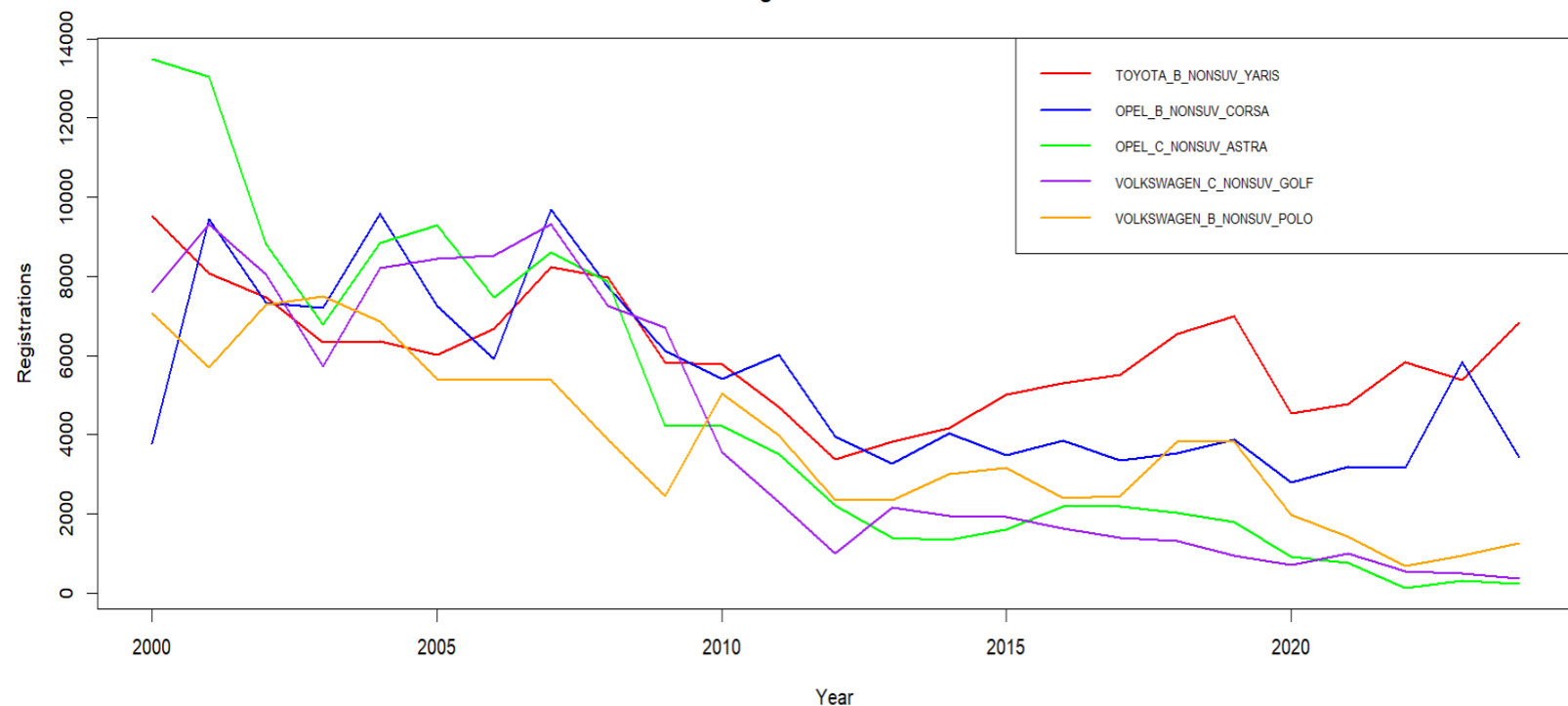
Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων
 Οι ταξινομήσεις ανά τύπο όπως οπτικοποιούνται μέσα από το αντικείμενο GTS. Οι τύποι είναι: SUV, NONSUV, SPORTS, MPV. Παρατηρούμε ότι ενώ πριν το 2010 τα αυτοκίνητα τύπου non-SUV είχαν πολύ μεγαλύτερο μερίδιο ταξινομήσεων σε σχέση με τα SUV, αυτό άλλαξε όταν ξεκίνησε η ανάκαμψη τους μετά το 2027. Σαν συνέπεια από το 2022 τα αυτοκίνητα τύπου SUV έχουν μεγαλύτερο μερίδιο ταξινομήσεων. Αυτό δείχνει την αλλαγή νοοτροπίας των Ελλήνων καταναλωτών και την στροφή τους προς την πρακτικότητα.



Σχήμα 6.6 Ταξινομήσεις ανά μάρκα

Οι ταξινομήσεις ανά μάρκα όπως οπτικοποιούνται μέσα από το αντικείμενο GTS. Λόγω του μεγάλου όγκου των χρονοσειρών της συγκεκριμένης κατηγορίας, έχουν φιλτραριστεί οι πέντε μάρκες με τις περισσότερες ταξινομήσεις : TOYOTA, OPEL, VOLKSWAGEN, HUYNDAL, PEUGEOT. Η υπεροχή των αυτοκινήτων TOYOTA φαίνεται καθαρά τόσο πριν όσο και μετά την οικονομική κρίση. Αυτό οφείλεται στην φήμη που έχει η μάρκα για αξιοπιστία και πρακτικότητα. Αξίζει να σημειωθεί ότι η μόνη εξαίρεση σε αυτή την παρατήρηση είναι το 2012 με την αρχή της οικονομική κρίση όπου βλέπουμε ότι η μάρκα OPEL την ξεπερνάει σε ταξινομήσεις. Αυτό μπορεί να αποδοθεί στην χαμηλότερη τιμή των αυτοκινήτων της συγκεκριμένης μάρκας.

Car Registrations 2000-2024



Σχήμα 6.7 Ταξινομήσεις ανά μοντέλο

Οι ταξινομήσεις ανά μοντέλο όπως οπτικοποιούνται μέσα από το αντικείμενο GTS. Λόγω του μεγάλου όγκου των χρονοσειρών της συγκεκριμένης κατηγορίας, έχουν φιλτραριστεί τα πέντε (5) μοντέλα με τις περισσότερες ταξινομήσεις : TOYOTA YARIS, OPEL CORSA, OPEL ASTRA, VOLKSWAGEN GOLF, VOLKSWAGEN POLO. Και στο σχήμα 5.7 φαίνεται η υπεροχή της TOYOTA με το μοντέλο YARIS αν και είναι ενδιαφέρον ότι το μοντέλο CORSA της OPEL το ξεπέρασε σε ταξινομήσεις το 2023.

7. Παρουσίαση αποτελεσμάτων

Η αξιολόγηση των προβλέψεων πραγματοποιήθηκε σε πέντε επίπεδα ανάλυσης: συνολικό (total), μάρκας (make), κατηγορίας (segment), τύπου (type) και μοντέλου (model), μέσω του δείκτη MASE. Τα μοντέλα που αξιολογήθηκαν είναι το μη ιεραρχικό GTS και τα ιεραρχικά HTS make, segment και type. Σε κάθε μοντέλο, εφαρμόστηκαν όλες οι διαθέσιμες μέθοδοι (Bottom-Up, Optimal Reconciliation, Middle-Out, Top-Down).

7.1 Επίπεδο συνολικών ταξινομήσεων

Στο συνολικό επίπεδο πρόβλεψης ταξινομήσεων, την καλύτερη επίδοση παρουσιάζει η μέθοδος Bottom-Up (BU), καταγράφοντας τη χαμηλότερη τιμή MASE (0,712). Η ίδια τιμή μέσου MASE ισχύει για όλα τα μοντέλα. Ελαφρώς υψηλότερη, αλλά ακόμη πολύ ικανοποιητική, είναι η επίδοση του μοντέλου HTS_Make με τη μέθοδο βέλτιστης συνάθροισης (COMP) και βάρη *mint*, με δείκτη MASE ίσο με 0,729.

Αντίθετα, τις χαμηλότερες επιδόσεις παρουσιάζει η μέθοδος Top-Down (TD), όπου όλες οι παραλλαγές της σε όλα τα μοντέλα καταγράφουν τιμές MASE ίσες με 0,804, ανεξαρτήτως του τρόπου διαίρεσης (*tdfp*, *tdgsa*, *tdgsf*). Η ίδια τιμή του MASE είναι λογική καθώς σε όλες τις περιπτώσεις, ουσιαστικά γίνεται πρόβλεψη από την χρονοσειρά των συνολικών ταξινομήσεων για την ίδια κατηγορία χωρίς περεταίρω υπολογισμούς.

Αυτό που είναι αξιοσημείωτο, είναι ότι είναι καλύτερο να προβλέψουμε τις τάσεις σε επίπεδο μοντέλου και να συναθροίσουμε για να βρούμε την πρόβλεψη για τις συνολικές ταξινομήσεις, παρά να κάνουμε την πρόβλεψη κατευθείαν για αυτές.

Πίνακας 7.1 Τα πέντε καλύτερα και πέντε χειρότερα αποτελέσματα σε επίπεδο συνολικών ταξινομήσεων

Top five					
Ranking	Model	Method	Έτος	Total	Weights
1	GTS_Object	BU	Average	0,712	
2	HTS_Segment	BU	Average	0,712	
3	HTS_Make	BU	Average	0,712	
4	HTS_Type	BU	Average	0,712	
5	HTS_Make	Comp	Average	0,729	mint

Bottom five					
Ranking	Model	Method	Έτος	Total	Weights
1	HTS_Make	TD	Average	0,804	tdfp
2	HTS_Make	TD	Average	0,804	tdgsa
3	HTS_Make	TD	Average	0,804	tdgsf
4	HTS_Segment	TD	Average	0,804	tdfp
5	HTS_Segment	TD	Average	0,804	tdgsa

7.2 Επίπεδο μάρκας

Για το επίπεδο μάρκας (Make), προβλέψεις παρέχονται μόνο από το μοντέλο HTS Make και το μοντέλο GTS.

Την καλύτερη επίδοση παρουσιάζει η μέθοδος Middle-Out (MO) στο μοντέλο HTS_Make, καταγράφοντας τη χαμηλότερη τιμή μέσου MASE ίση με 1,685, δείχνοντας ότι όσον αφορά την μάρκα η απευθείας πρόβλεψή είναι η καλύτερη. Πολύ κοντά βρίσκονται και άλλες μέθοδοι του μοντέλου HTS_Make με την μέθοδο COMP και χρήση βαρών *mint* ή *wls*, που καταγράφουν τιμές MASE 1,690 και 1,694 αντίστοιχα. Η συνέπεια των καλών επιδόσεων στις μεθόδους MO και COMP εντός του μοντέλου HTS_Make καταδεικνύει ότι η κατηγοριοποίηση ανά μάρκα, όταν συνδυάζεται με κατάλληλη μέθοδο πρόβλεψης και στάθμισης, προσφέρει σχετικά πιο ακριβή αποτελέσματα.

Ενδιαφέρον παρουσιάζει επίσης η παρουσία του μοντέλου GTS_Object στη δεύτερη θέση, με τη μέθοδο COMP και χρήση βαρών *mint*, η οποία καταγράφει τιμή MASE 1,687.

Αντίθετα, τις χαμηλότερες επιδόσεις στο επίπεδο μάρκας εμφανίζονται και στα δύο μοντέλα, όταν χρησιμοποιείται είτε η μέθοδος COMP με ακατάλληλα βάρη (*ols*, *nseries*) είτε η μέθοδος TD. Η χειρότερη επίδοση καταγράφεται από τη μέθοδο COMP με βάρη *ols* στο μοντέλο GTS_Object, με δείκτη MASE 4,134. Σημαντικά χαμηλές επιδόσεις εμφανίζουν επίσης οι TD μέθοδοι του HTS_Make, με τιμές MASE άνω του 3,5.

Σε γενικές γραμμές φαίνεται ότι οι ταξινομήσεις σε επίπεδο μάρκας είναι πολύ δύσκολο να προβλεφθούν αφού στην καλύτερη περίπτωση έχουμε πρόβλεψή που είναι κατα 68,7 % χειρότερη από την *naïve*. Αυτό συμβαίνει, όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, λόγω του επηρεασμού τους από την στρατηγική των διάφορων αυτοκινητοβιομηχανιών και γενικώς από τον ανθρώπινο παράγοντα.

Πίνακας 7.2 Τα πέντε καλύτερα και πέντε χειρότερα αποτελέσματα σε επίπεδο μάρκας

Top five						Bottom five					
Ranking	Model	Method	Έτος	Make	Weights	Ranking	Model	Method	Έτος	Make	Weights
1	HTS_Make	MO	Average	1,685		1	GTS_Object	Comp	Average	4,134	ols
2	GTS_Object	Comp	Average	1,688	mint	2	HTS_Make	TD	Average	3,827	tdgsf
3	HTS_Make	Comp	Average	1,690	mint	3	HTS_Make	TD	Average	3,597	tdgsa
4	HTS_Make	Comp	Average	1,694	wls	4	HTS_Make	Comp	Average	2,998	ols
5	HTS_Make	TD	Average	1,694	tdfp	5	GTS_Object	Comp	Average	2,902	nseries

7.3 Επίπεδο κατηγορίας

Για το επίπεδο κατηγορίας (Segment), οι προβλέψεις παρέχονται μόνο από το μοντέλο με την αντίστοιχη κατηγοριοποίηση (HTS Segment) και το μοντέλο GTS.

Την καλύτερη επίδοση παρουσιάζει η μέθοδος BU, με τιμή MASE ίση με 0,609 και στα δύο μοντέλα. Ακολουθούν οι μέθοδοι COMP με πολύ κοντινές επιδόσεις. Ειδικότερα, οι συνδυασμοί με βάρη *mint* και *wls* καταγράφουν δείκτη MASE 0,617 έως 0,622. Οι μέθοδοι COMP με προσεκτική επιλογή βαρών αποδεικνύονται επομένως ιδιαίτερα αποτελεσματικές και ανταγωνιστικές της BU.

Αντιθέτως, τις χαμηλότερες επιδόσεις εμφανίζουν μέθοδοι του μοντέλου HTS_Segment όταν εφαρμόζεται η προσέγγιση TD. Η χειρότερη επίδοση καταγράφεται με τιμή MASE 1,412 στη μέθοδο TD με στάθμιση *tdgsf*, ενώ ελαφρώς καλύτερη είναι η τιμή 1,169 με στάθμιση *tdgsa*. Χαμηλές επιδόσεις καταγράφονται επίσης από μεθόδους COMP με ακατάλληλα βάρη *ols* και *nseries*, τόσο στο μοντέλο HTS_Segment όσο και στο GTS_Object.

Συμπερασματικά, στο επίπεδο κατηγορίας, η μέθοδος BU αναδεικνύεται σαφώς ως η πιο αποτελεσματική στρατηγική, με συνέπεια και ακρίβεια ανεξαρτήτως μοντέλου. Οι μέθοδοι βέλτιστης συνάθροισης με κατάλληλη επιλογή βαρών προσφέρουν επίσης πολύ ικανοποιητικές επιδόσεις.

Γενικά παρατηρούμε ότι, αντίθετα με το επίπεδο μάρκας, λόγω της σταθερότητας του επιπέδου κατηγορίας μπορούμε να δώσουμε μια αρκετά καλή πρόβλεψη. Το επίπεδο κατηγορίας παρουσιάζει σταθερότητα διαχρονικά, αντανakλώντας τις βασικές προτιμήσεις των καταναλωτών. Οι προτιμήσεις αυτές δεν μεταβάλλονται έντονα από εξωτερικούς παράγοντες, όπως το οικονομικό περιβάλλον ή οι βραχυχρόνιες διακυμάνσεις στην αγορά.

Πίνακας 7.3 Τα πέντε καλύτερα και πέντε χειρότερα αποτελέσματα σε επίπεδο κατηγορίας

Top five					
Ranking	Model	Method	Έτος	Segment	Weights
1	GTS_Object	BU	Average	0,609	
2	HTS_Segment	BU	Average	0,609	
3	HTS_Segment	Comp	Average	0,617	mint
4	HTS_Segment	Comp	Average	0,617	wls
5	GTS_Object	Comp	Average	0,618	mint

Bottom five					
Ranking	Model	Method	Έτος	Segment	Weights
1	HTS_Segment	TD	Average	1,412	tdgsf
2	HTS_Segment	TD	Average	1,169	tdgsa
3	GTS_Object	Comp	Average	0,854	ols
4	HTS_Segment	Comp	Average	0,741	ols
5	HTS_Segment	Comp	Average	0,679	nseries

7.4 Επίπεδο τύπου

Για το επίπεδο τύπου (Type), οι προβλέψεις παρέχονται μόνο από το μοντέλο με την αντίστοιχη κατηγοριοποίηση (HTS Type) και το μοντέλο GTS.

Στο επίπεδο πρόβλεψης ταξινομήσεων ανά τύπο, την καλύτερη επίδοση εμφανίζει το μοντέλο HTS_Type με τη μέθοδο MO, επιτυγχάνοντας τιμή MASE ίση με 0,905. Πολύ κοντά ακολουθούν μέθοδοι COMP, τόσο στο ίδιο μοντέλο όσο και στο GTS_Object, με τιμές MASE κυμαινόμενες από 0,923 (HTS_Type με βάρη *wls* και *mint*) έως 0,931 (GTS_Object με βάρη *wls*). Η σταθερή απόδοση αυτών των μεθόδων επιβεβαιώνει τη χρησιμότητα της βέλτιστης συνάθροισης πολλαπλών προβλέψεων όταν γίνεται με κατάλληλη στάθμιση.

Αντίθετα, τις χειρότερες επιδόσεις καταγράφουν οι μέθοδοι TD. Οι παραλλαγές με βάρη *tdgsf* και *tdgsa* παρουσιάζουν πολύ υψηλές τιμές MASE (2,720 και 2,633 αντίστοιχα), καταδεικνύοντας την σαφή αδυναμία της προσέγγισης αυτής να προβλέψει με ακρίβεια στο επίπεδο τύπου. Παράλληλα, και οι μέθοδοι COMP με συντελεστές *ols* ή *nseries*, εμφανίζουν σημαντικά κακές επιδόσεις. Συγκεκριμένα έχουμε μέσο MASE 1,551 για HTS_Type και 1,314 για GTS_Object με βάρη *ols* και στις δύο περιπτώσεις.

Η συνολική εικόνα καταδεικνύει ότι για προβλέψεις σε επίπεδο τύπου, οι βέλτιστες μέθοδοι είναι η middle out και οι μέθοδοι βέλτιστης συνάθροισης, υπό την προϋπόθεση της ορθής επιλογής βαρών. Η μέθοδος Top-Down, φαίνεται όπως και στις προηγούμενες περιπτώσεις να μην μπορεί να αποδώσει με ακρίβεια σε αυτό το επίπεδο ανάλυσης.

Πίνακας 7.4 Τα πέντε καλύτερα και πέντε χειρότερα αποτελέσματα σε επίπεδο τύπου

Top five						Bottom five					
Ranking	Model	Method	Έτος	Type	Weights	Ranking	Model	Method	Έτος	Type	Weights
1	HTS_Type	MO	Average	0,905		1	HTS_Type	TD	Average	2,720	tdgsf
2	HTS_Type	Comp	Average	0,923	wls	2	HTS_Type	TD	Average	2,633	tdgsa
3	HTS_Type	Comp	Average	0,923	mint	3	HTS_Type	Comp	Average	1,551	ols
4	GTS_Object	Comp	Average	0,924	mint	4	GTS_Object	Comp	Average	1,314	ols
5	GTS_Object	Comp	Average	0,931	wls	5	HTS_Type	Comp	Average	1,030	nseries

7.5 Επίπεδο μοντέλου

Στο πιο αναλυτικό επίπεδο πρόβλεψης, δηλαδή σε επίπεδο μοντέλου οχήματος, οι καλύτερες επιδόσεις παρουσιάζονται από τη μέθοδο BU, καθώς και από τη μέθοδο MO καταγράφοντας τιμές

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων MASE ίσες με 1,839. Η ίδια τιμή εμφανίζεται ανεξαρτήτως μοντέλου (HTS_Type, GTS_Object, HTS_Segment, HTS_Make) καθώς στην ουσία αντιπροσωπεύει την απλή πρόβλεψη με μέθοδο εκθετικής εξομάλυνσης από τις επιμέρους χρονοσειρές των μοντέλων για την ίδια κατηγορία. Αυτό υποδηλώνει ότι στο επίπεδο μοντέλου, η ακρίβεια πρόβλεψης εξαρτάται κυρίως από την καλή ανάλυση σε χαμηλό επίπεδο, με τη μέθοδο MO ή την απευθείας πρόβλεψη να οδηγούν σε παρόμοια αποτελέσματα όταν εφαρμόζονται σωστά.

Αντίθετα, τις χειρότερες επιδόσεις καταγράφουν οι προβλέψεις που προέρχονται από τη μέθοδο TD. Όλες οι εκδοχές της μεθόδου TD στο σύνολο των HTS μοντέλων παρουσιάζουν εξαιρετικά υψηλές τιμές MASE, με κορυφαία τη HTS_Type με βάρη *tdgsf*, φτάνοντας τα 4,003, δηλαδή υπερδιπλάσια απόδοση σφάλματος σε σχέση με τις βέλτιστες μεθόδους. Παρόμοια αποτελέσματα παρατηρούνται και στην παραλλαγή της με βάρη *tdgsa*, επιβεβαιώνοντας τη γενικευμένη αδυναμία της TD προσέγγισης να παράξει ακριβείς προβλέψεις στο επίπεδο των μεμονωμένων μοντέλων.

Η συνολική εικόνα στο επίπεδο μοντέλου είναι σαφής: η λεπτομερής πρόβλεψη ανά μοντέλο με BU ή MO προσέγγιση αποτελεί τη βέλτιστη στρατηγική, ενώ η χρήση της TD μεθόδου οδηγεί σε σημαντικά υψηλότερο σφάλμα και θα πρέπει να αποφεύγεται.

Όπως και στην περίπτωση της μάρκας, οι ταξινομήσεις σε επίπεδο μοντέλου είναι πολύ δύσκολο να προβλεφθούν και μάλιστα η καλύτερη επίδοση μεθόδου πρόβλεψής του επιπέδου είναι κατά 84% χειρότερη από την πρόβλεψή με μέθοδο *naive* δηλαδή ακόμα πιο αδύναμη από εκείνη της μάρκας (68% χειρότερη από την *naive*)

Πίνακας 7.5 Τα πέντε καλύτερα και πέντε χειρότερα αποτελέσματα σε επίπεδο μοντέλου

Top five						Bottom five					
Ranking	Model	Method	Έτος	Type	Weights	Ranking	Model	Method	Έτος	Type	Weights
1	HTS_Type	MO	Average	1,839		1	HTS_Type	TD	Average	4,003	<i>tdgsf</i>
2	GTS_Object	BU	Average	1,839		2	HTS_Segment	TD	Average	4,003	<i>tdgsf</i>
3	HTS_Type	BU	Average	1,839		3	HTS_Make	TD	Average	4,003	<i>tdgsf</i>
4	HTS_Segment	BU	Average	1,839		4	HTS_Segment	TD	Average	3,749	<i>tdgsa</i>
5	HTS_Make	BU	Average	1,839		5	HTS_Make	TD	Average	3,749	<i>tdgsa</i>

7.6 Σύγκριση προβλέψεων για τις τρεις κύριες κατηγορίες

Η αξιολόγηση των προβλέψεων με βάση το μέσο MASE για τις κύριες κατηγορίες, δηλαδή σε επίπεδο συνολικών ταξινομήσεων (Total), κατηγορίας (Segment) και τύπου (Type), παρουσιάζει ιδιαιτερότητες ως προς τη δυνατότητα σύγκρισης των διαφορετικών μοντέλων. Ειδικότερα, μόνο το μοντέλο GTS περιλαμβάνει προβλέψεις και για τις τρεις αυτές κατηγορίες, γεγονός που

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων καθιστά τη μεταξύ τους σύγκριση, αξιόπιστη αποκλειστικά εντός του συγκεκριμένου μοντέλου. Ως εκ τούτου, η αξιολόγηση σε αυτό το επίπεδο περιορίζεται στις διαφορετικές παραλλαγές μεθόδων εντός του μοντέλου GTS.

Οι καλύτερες επιδόσεις στο συνδυασμό των κύριων κατηγοριών καταγράφονται από τη μέθοδο BU, με τιμές MASE ίσες με 0,712 για το σύνολο των ταξινομήσεων, 0,609 για την κατηγορία και 0,933 για τον τύπο. Οι τιμές αυτές δείχνουν ξεκάθαρα ότι η απευθείας πρόβλεψη των χρονοσειρών του κατώτατου επιπέδου και η άθροισή τους σε ανώτερο επίπεδο εξακολουθεί να προσφέρει σταθερά και αξιόπιστα αποτελέσματα, ακόμη και σε γενικό επίπεδο.

Ανταγωνιστικές επιδόσεις παρουσιάζουν και οι παραλλαγές της COMP με βάρη *mint* και *wls*, με μέση τιμή MASE για τις τρεις κατηγορίες περίπου 0,76, ελαφρώς υψηλότερη από εκείνη της BU. Η μέθοδος αυτή συνδυάζει πληροφορία από διαφορετικά επίπεδα της ιεραρχίας και φαίνεται να προσφέρει καλή ισορροπία ακρίβειας στις προβλέψεις, ειδικά στα επίπεδα κατηγορίας και τύπου.

Αντίθετα, οι χαμηλότερες επιδόσεις καταγράφονται από τις παραλλαγές βέλτιστης συνάθροισης με βάρη *ols* και *nseries*, όπου οι τιμές MASE αυξάνονται αισθητά. Ειδικά στην περίπτωση των απλών βαρών, η τιμή μέσου MASE φτάνει το 0,798 για τις συνολικές ταξινομήσεις, το 0,854 για την κατηγορία και το 1,314 για τον τύπο. Αυτό είναι λογικό καθώς και οι δύο μέθοδοι δεν λαμβάνουν υπόψη την σταθερότητα και τα δεδομένα στις χρονοσειρές του κατώτατου επιπέδου.

Συμπερασματικά, και στο συνολικό επίπεδο πρόβλεψης, η Bottom-Up μέθοδος παραμένει η πιο ακριβής, παρέχοντας τη χαμηλότερη απόκλιση από τις πραγματικές τιμές. Οι μέθοδοι βέλτιστης συνάθροισης με κατάλληλα βάρη όπως *mint* και *wls* αποδεικνύονται αρκετά αποτελεσματικές, ενώ οι παραλλαγές με βάρη *ols* ή *nseries* θα πρέπει να αποφεύγονται, καθώς οδηγούν σε σημαντικά αυξημένο σφάλμα.

Η γενική εικόνα καταδεικνύει ότι σε συνολικά επίπεδα ανάλυσης (Total, Segment, Type), η πρόβλεψη είναι πιο σταθερή και λιγότερο επιρρεπής σε σφάλμα, με τις κατάλληλες επιλογές μεθόδου να ενισχύουν περαιτέρω την ακρίβεια.

Πίνακας 7.6 Τα αποτελέσματα σύγκρισης σφαλμάτων στο αντικείμενο GTS

Model	Method	Έτος	Total	Make	Segment	Type	Model	AVERAGE	Average WO make and model	Weights
GTS_Object	BU	Average	0,712	1,712	0,609	0,933	1,839	1,161	0,751	
GTS_Object	Comp	Average	0,735	1,688	0,618	0,924	1,842	1,162	0,759	mint
GTS_Object	Comp	Average	0,750	1,694	0,622	0,931	1,843	1,168	0,767	wls
GTS_Object	Comp	Average	0,772	2,902	0,633	0,939	2,356	1,520	0,781	nseries
GTS_Object	Comp	Average	0,798	4,134	0,854	1,314	2,859	1,992	0,989	ols

8. Συμπεράσματα και Προεκτάσεις

8.1 Ανασκόπηση

Η παρούσα εργασία πραγματοποιήθηκε την πρόκληση της συγκέντρωσης, επεξεργασίας, ανάλυσης και πρόβλεψης των δεδομένων ταξινομήσεων αυτοκινήτων στην Ελλάδα και στην Ευρώπη. Στο επίκεντρο της ανάλυσης τέθηκε η αξιοποίηση των δεδομένων του ΣΕΑΑ, τα οποία διατίθενται σε μορφή PDF χωρίς σταθερή δομή, γεγονός που καθιστά τη χρήση τους δύσκολη και χρονοβόρα. Για την αντιμετώπιση αυτών των περιορισμών, αναπτύχθηκε μία αυτοματοποιημένη διαδικασία εξαγωγής και μετατροπής των δεδομένων σε μορφή Excel, με χρήση της γλώσσας Python. Τα δεδομένα οργανώθηκαν σε θεματικά και χρονικά επίπεδα, καλύπτοντας τις ταξινομήσεις στην Ελλάδα από το 1965, τις ανά μοντέλο και κατηγορία ταξινομήσεις από το 2000, καθώς και γεωγραφικές κατανομές από το 2005. Παράλληλα, συγκεντρώθηκαν αντίστοιχα ευρωπαϊκά δεδομένα για σκοπούς συγκριτικής αξιολόγησης.

Με βάση αυτά τα δεδομένα, το κύριο αναλυτικό μέρος της εργασίας επικεντρώθηκε στην πρόβλεψη των μελλοντικών ταξινομήσεων μέσω της μεθοδολογίας των ιεραρχικών χρονοσειρών. Χρησιμοποιώντας τη βιβλιοθήκη `hts` της γλώσσας R, κατασκευάστηκε μία πολυεπίπεδη ιεραρχική δομή, στην οποία οι χρονοσειρές οργανώνονται ανάλογα με τη μάρκα, το μοντέλο, την κατηγορία και τον τύπο του οχήματος. Η πρόβλεψη πραγματοποιήθηκε με τέσσερις διαφορετικές μεθόδους: *bottom-up*, *top-down*, *middle-out* και βέλτιστης συνάθροισης και τις διάφορες παραλλαγές τους. Κάθε προσέγγιση εξετάστηκε με κριτήριο την ακρίβεια των προβλέψεων, η οποία μετρήθηκε με τον δείκτη MASE, τόσο στο συνολικό όσο και στα επιμέρους επίπεδα.

Τα αποτελέσματα της ανάλυσης ανέδειξαν τη μέθοδο *bottom-up* ως την πιο αξιόπιστη και ακριβή, ιδίως όταν η πρόβλεψη εστιάζει σε λεπτομερή επίπεδα ανάλυσης, όπως το μοντέλο ή η κατηγορία. Η *bottom-up* προσέγγιση επέτρεψε τη διατήρηση της συνέπειας της ιεραρχικής δομής, διασφαλίζοντας ταυτόχρονα χαμηλά σφάλματα πρόβλεψης. Παράλληλα, η μέθοδος βέλτιστης συνάθροισης παρουσίασε επίσης ικανοποιητικές επιδόσεις, υπό την προϋπόθεση της χρήσης κατάλληλων στατιστικών βαρών. Αντίθετα, η μέθοδος *top-down*, καθώς και παραλλαγές της βέλτιστης συνάθροισης με ανεπαρκή στάθμιση, αποδείχθηκαν λιγότερο αποδοτικές, κυρίως λόγω της αδυναμίας τους να αποτυπώσουν τη μεταβλητότητα των υποεπιπέδων της ιεραρχίας.

Η συνδυασμένη χρήση των γλωσσών Python και R, επέτρεψε την ανάπτυξη μίας ολοκληρωμένης μεθοδολογικής προσέγγισης. Η εργασία κατέδειξε ότι η κατάλληλη επεξεργασία και μοντελοποίηση σύνθετων ιεραρχικών δεδομένων μπορεί να οδηγήσει σε χρήσιμες προβλέψεις με πρακτική εφαρμογή για τον κλάδο της αυτοκινητοβιομηχανίας, αλλά και για τη γενικότερη κατανόηση της οικονομικής δραστηριότητας. Το έργο αυτό αποτελεί μία προσπάθεια να καλυφθεί ένα σημαντικό αναλυτικό κενό στη συστηματική μελέτη των ταξινομήσεων οχημάτων, προσφέροντας μία τεχνικά τεκμηριωμένη και εμπειρικά αξιολογημένη προσέγγιση πρόβλεψης.

8.2 Συμπεράσματα

Η συνολική αξιολόγηση των αποτελεσμάτων ανέδειξε ουσιαστικές διαφορές μεταξύ των μεθόδων πρόβλεψης, τόσο ως προς την ακρίβεια των προβλέψεων σε διαφορετικά επίπεδα ανάλυσης όσο και ως προς την ικανότητα κάθε προσέγγισης να αξιοποιήσει τη διαθέσιμη πληροφορία.

Πρώτο συμπέρασμα είναι ότι η πρόβλεψη σε συνολικό επίπεδο (Total) αλλά και στα μεσαία επίπεδα ιεραρχίας όπως κατηγορίας και τύπου, είναι σε γενικές γραμμές επιτυχής, με χαμηλές τιμές μέσου MASE (<1) για τις περισσότερες από τις μεθόδους. Ειδικά η μέθοδος Bottom-Up επιτυγχάνει στο επίπεδο συνολικών ταξινομήσεων, μέσο MASE 0,712 ασχέτως μοντέλου δεδομένων και έχει αντίστοιχα πολύ χαμηλές τιμές σε κατηγορία (0,609) και τύπο (0,933). Τα αποτελέσματα αυτά υποδεικνύουν ότι οι προβλέψεις στα επίπεδα που δεν επηρεάζονται από τις ισορροπίες μεταξύ των αυτοκινητοβιομηχανιών και έχουν αρκετή συγκεντρωτική πληροφορία και σταθερότητα στη χρονική τους πορεία, είναι ιδιαίτερα αξιόπιστες.

Αντίθετα, οι προβλέψεις σε επίπεδο μάρκας και μοντέλου παρουσιάζουν σταθερά υψηλότερες τιμές MASE (πάνω από 1,7 για την μάρκα και έως και 4 για το μοντέλο), γεγονός που καταδεικνύει τη δυσκολία πρόβλεψης. Αυτό οφείλεται τόσο στη μεγάλη διασπορά και ετερογένεια αυτών των χρονοσειρών όσο και στη σχετική αστάθεια ή μικρή διάρκεια ζωής κάποιων μοντέλων αυτοκινήτων, η οποία καθιστά δυσκολότερη την εξαγωγή σταθερών προβλεπτικών μοτίβων.

Δεύτερο συμπέρασμα είναι ότι η μέθοδος Bottom-Up υπερτερεί ξεκάθαρα στις περισσότερες περιπτώσεις. Ο λόγος εντοπίζεται στο γεγονός ότι σε αυτή την προσέγγιση αξιοποιούνται πληροφορίες από την πληθώρα χρονοσειρών στο κατώτατο επίπεδο, γεγονός που επιτρέπει την απορρόφηση τοπικών μεταβολών και τη μεταφορά της πληροφορίας προς τα πάνω με πολύ μικρές απώλειες. Η μέθοδος Bottom-Up αναδεικνύεται έτσι ως ιδιαίτερα ισχυρή μέθοδος όταν υπάρχει πλούσια και λεπτομερής ιστορική πληροφορία, όπως συμβαίνει στην περίπτωση των χρονοσειρών επιμέρους μοντέλων αυτοκινήτων.

Αντίστροφα, η μέθοδος Top-Down προσέγγιση αποτυγχάνει να αποδώσει ικανοποιητικά. Οι τιμές MASE της μεθόδου με όλες τις παραλλαγές της είναι υψηλές σε όλα τα επίπεδα και ιδιαίτερα αρνητικές σε κατώτερα επίπεδα όπως η κατηγορία και ο τύπος. Αυτό οφείλεται στο γεγονός ότι η πληροφορία, καθώς αφορά μία μόνο χρονοσειρά των συνολικών ταξινομήσεων, είναι ανεπαρκής για να διαχυθεί με ακρίβεια προς τα κάτω, ιδιαίτερα όταν το πλήθος και η ποικιλία των κατώτερων χρονοσειρών είναι μεγάλο. Η μέθοδος Top-Down δεν μπορεί να αποδώσει τις τοπικές ιδιαιτερότητες των υποκατηγοριών, και επομένως υστερεί εμφανώς.

Τέλος, ξεχωριστή αναφορά αξίζει η μέθοδος βέλτιστης συνάθροισης με βάρη *mint*. Παρουσιάζει σταθερά καλή συνολική απόδοση, με τιμές MASE συγκρίσιμες με τη μέθοδο Bottom-Up. Σημαντικό πλεονέκτημα της συγκεκριμένης μεθόδου είναι ότι, αν εφαρμοστεί στο αντικείμενο GTS, μπορεί να προβλέπει σε όλα τα επίπεδα ιεραρχίας, αξιοποιώντας ομαδοποιήσεις χρονοσειρών χωρίς να απαιτείται αυστηρή ιεραρχική δομή. Αυτή η ευελιξία καθιστά την μέθοδο βέλτιστης συνάθροισης με βάρη *mint* και εφαρμοσμένη στο μοντέλο GTS, μια εξαιρετικά

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων κατάλληλη επιλογή για πολυδιάστατα προβλήματα όπως αυτό της πρόβλεψης ταξινομήσεων αυτοκινήτων.

Συνοψίζοντας, τα αποτελέσματα επιβεβαιώνουν ότι η ακρίβεια των προβλέψεων επηρεάζεται ισχυρά τόσο από τη μεθοδολογία που εφαρμόζεται όσο και από το επίπεδο ανάλυσης. Οι μέθοδοι που αξιοποιούν τη διαθέσιμη πληροφορία σε βάθος, όπως η Bottom-Up ή η βέλτιστη συνάθροιση με βάρη mint, επιτυγχάνουν τις καλύτερες επιδόσεις. Η πολυπλοκότητα και η αστάθεια των χρονοσειρών σε λεπτομερή επίπεδα, όπως η μάρκα και κυρίως το μοντέλο, εξακολουθεί να αποτελεί πρόκληση για την πρόβλεψη, αλλά ο συνολικός αριθμός ταξινομήσεων και τα επίπεδα κατηγορίας και τύπου, εμφανίζονται προβλέψιμα με ικανοποιητική ακρίβεια, προσφέροντας ένα χρήσιμο εργαλείο για μελλοντικό σχεδιασμό και στρατηγικές αποφάσεις στον κλάδο της αυτοκίνησης.

8.3 Προεκτάσεις

Η παρούσα εργασία θέτει τις βάσεις για μία συστηματική προσέγγιση πρόβλεψης ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών χρονοσειρών. Ωστόσο, υπάρχουν σημαντικές προεκτάσεις που μπορούν να εμπλουτίσουν περαιτέρω την ανάλυση και να ενισχύσουν τη χρησιμότητα των αποτελεσμάτων.

Μία κατεύθυνση με ιδιαίτερο ενδιαφέρον είναι η ανάλυση των ταξινομήσεων σε γεωγραφικό επίπεδο. Μέσω της κατηγοριοποίησης των δεδομένων ανά περιφέρεια ή νομό, θα μπορούσε να αναδειχθεί η διαφοροποίηση των καταναλωτικών προτιμήσεων σε τοπικό επίπεδο. Τέτοιου είδους ανάλυση μπορεί να απαντήσει ερωτήματα όπως ποια κατηγορία οχημάτων προτιμάται περισσότερο σε αστικές περιοχές ή πώς διαφέρει η ζήτηση ανάλογα με τον τύπο αυτοκινήτου σε ορεινές περιοχές ή νησιά. Η γεωγραφική κατηγοριοποίηση των προβλέψεων θα μπορούσε να οδηγήσει σε στοχευμένες εμπορικές ή πολιτικές παρεμβάσεις.

Παράλληλα, αξίζει να διερευνηθεί ο ρόλος της εποχιακότητας στην κατανομή των ταξινομήσεων. Οι καταναλωτικές συνήθειες ενδέχεται να επηρεάζονται από παράγοντες όπως η λήξη οικονομικών κινήτρων (π.χ. τέλη κυκλοφορίας), περίοδοι διακοπών ή καιρικές συνθήκες. Η ενσωμάτωση εποχικών δεικτών μπορεί να οδηγήσει σε πιο ρεαλιστικά και ακριβή μοντέλα πρόβλεψης.

Μια ακόμη προέκταση αφορά τη σύγκριση των αποτελεσμάτων της ιεραρχικής πρόβλεψης με άλλες κλασικές και σύγχρονες μεθόδους. Θα ήταν ενδιαφέρον να συγκριθούν οι επιδόσεις της μεθόδου ETS με προσεγγίσεις όπως τα μοντέλα ARIMA, το μοντέλο Theta και οι μέθοδοι μηχανικής μάθησης. Μια τέτοια συγκριτική μελέτη θα επέτρεπε την αξιολόγηση των πλεονεκτημάτων και περιορισμών κάθε μεθόδου και πιθανόν τη δημιουργία υβριδικών μοντέλων πρόβλεψης.

Τέλος, μία πιο προηγμένη και φιλόδοξη επέκταση θα ήταν η δημιουργία μιας δυναμικής ροής ανάλυσης, η οποία να ενσωματώνει εξωτερικά δεδομένα ή ειδικά γεγονότα, όπως αλλαγές στη φορολογία, οικονομικές κρίσεις, νέα κυβερνητικά κίνητρα ή σημαντικές κοινωνικές αλλαγές (π.χ. πανδημίες) ή και την κριτική πρόβλεψη που αποτελεί ένα πολύ σημαντικό κομμάτι στον συγκεκριμένο κλάδο. Η ενσωμάτωση τέτοιων συμβάντων στο μοντέλο, είτε με τη μορφή μεταβλητών είτε με την προσαρμογή της δομής της πρόβλεψης, μπορεί να βελτιώσει σημαντικά την ακρίβεια και τη χρησιμότητα των προβλέψεων σε πραγματικές συνθήκες.

Συνολικά, η εργασία αυτή αποτελεί ένα πρώτο βήμα προς την ορθολογική αξιοποίηση των διαθέσιμων δεδομένων του κλάδου, και προσφέρει πολλαπλά περιθώρια περαιτέρω επέκτασης τόσο ως προς τη μεθοδολογία όσο και ως προς την πρακτική της εφαρμογή.

Παράρτημα

Πίνακας 7.7 Ο αναλυτικός πίνακας του σφάλματος MASE για όλες τις μεθόδους και όλα τα μοντέλα

Model	Method	Έτος	Total	Make	Segment	Type	Car model	Average	Average WO make and model	Weights
GTS_Object	Comp	2021	1,174	1,765	0,618	0,898	2,054	1,302	0,897	mint
GTS_Object	Comp	2022	0,373	2,109	0,386	0,373	1,990	1,046	0,377	mint
GTS_Object	Comp	2023	1,232	1,624	0,875	1,395	1,758	1,377	1,167	mint
GTS_Object	Comp	2024	0,163	1,254	0,595	1,032	1,564	0,922	0,596	mint
GTS_Object	Comp	Average	0,735	1,688	0,618	0,924	1,842	1,162	0,759	mint
HTS_Make	Comp	2021	1,155	1,768	NA	NA	2,058	1,660	1,155	mint
HTS_Make	Comp	2022	0,378	2,113	NA	NA	1,991	1,494	0,378	mint
HTS_Make	Comp	2023	1,235	1,625	NA	NA	1,757	1,539	1,235	mint
HTS_Make	Comp	2024	0,148	1,254	NA	NA	1,561	0,988	0,148	mint
HTS_Make	Comp	Average	0,729	1,690	NA	NA	1,842	1,420	0,729	mint
HTS_Segment	Comp	2021	1,136	NA	0,601	NA	2,056	1,264	0,869	mint
HTS_Segment	Comp	2022	0,391	NA	0,392	NA	1,987	0,923	0,392	mint
HTS_Segment	Comp	2023	1,234	NA	0,871	NA	1,754	1,286	1,052	mint
HTS_Segment	Comp	2024	0,168	NA	0,604	NA	1,563	0,779	0,386	mint
HTS_Segment	Comp	Average	0,732	NA	0,617	NA	1,840	1,063	0,675	mint
HTS_Type	Comp	2021	1,105	NA	NA	0,873	2,052	1,343	0,989	mint
HTS_Type	Comp	2022	0,432	NA	NA	0,398	1,988	0,939	0,415	mint
HTS_Type	Comp	2023	1,214	NA	NA	1,391	1,752	1,453	1,303	mint
HTS_Type	Comp	2024	0,171	NA	NA	1,031	1,566	0,923	0,601	mint
HTS_Type	Comp	Average	0,731	NA	NA	0,923	1,839	1,164	0,827	mint
GTS_Object	Comp	2021	1,212	1,907	0,752	1,119	2,200	1,438	1,027	nseries
GTS_Object	Comp	2022	0,338	3,779	0,403	0,443	2,730	1,538	0,394	nseries
GTS_Object	Comp	2023	1,261	1,777	0,808	1,269	1,933	1,410	1,113	nseries
GTS_Object	Comp	2024	0,277	4,146	0,570	0,926	2,561	1,696	0,591	nseries
GTS_Object	Comp	Average	0,772	2,902	0,633	0,939	2,356	1,520	0,781	nseries
HTS_Make	Comp	2021	1,205	2,205	NA	NA	2,316	1,909	1,205	nseries
HTS_Make	Comp	2022	0,349	2,493	NA	NA	2,155	1,666	0,349	nseries
HTS_Make	Comp	2023	1,271	1,874	NA	NA	1,945	1,697	1,271	nseries
HTS_Make	Comp	2024	0,209	1,873	NA	NA	1,830	1,304	0,209	nseries
HTS_Make	Comp	Average	0,759	2,111	NA	NA	2,062	1,644	0,759	nseries
HTS_Segment	Comp	2021	1,209	NA	0,711	NA	2,353	1,424	0,960	nseries
HTS_Segment	Comp	2022	0,350	NA	0,420	NA	2,113	0,961	0,385	nseries
HTS_Segment	Comp	2023	1,268	NA	0,913	NA	1,855	1,345	1,090	nseries
HTS_Segment	Comp	2024	0,228	NA	0,673	NA	1,731	0,877	0,451	nseries

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

Model	Method	Έτος	Total	Make	Segment	Type	Car model	Average	Average WO make and model	Weights
HTS_Segment	Comp	Average	0,764	NA	0,679	NA	2,013	1,152	0,721	nseries
HTS_Type	Comp	2021	1,168	NA	NA	1,140	2,107	1,472	1,154	nseries
HTS_Type	Comp	2022	0,391	NA	NA	0,442	2,162	0,998	0,416	nseries
HTS_Type	Comp	2023	1,238	NA	NA	1,466	1,873	1,526	1,352	nseries
HTS_Type	Comp	2024	0,280	NA	NA	1,071	2,556	1,302	0,675	nseries
HTS_Type	Comp	Average	0,769	NA	NA	1,030	2,175	1,324	0,899	nseries
GTS_Object	Comp	2021	1,258	2,028	1,051	1,637	2,338	1,662	1,315	ols
GTS_Object	Comp	2022	0,329	5,901	0,628	0,927	3,538	2,265	0,628	ols
GTS_Object	Comp	2023	1,285	1,751	1,255	2,140	2,085	1,703	1,560	ols
GTS_Object	Comp	2024	0,319	6,857	0,482	0,552	3,473	2,337	0,451	ols
GTS_Object	Comp	Average	0,798	4,134	0,854	1,314	2,859	1,992	0,989	ols
HTS_Make	Comp	2021	1,293	2,460	NA	NA	2,444	2,065	1,293	ols
HTS_Make	Comp	2022	0,308	1,990	NA	NA	2,084	1,461	0,308	ols
HTS_Make	Comp	2023	1,313	1,985	NA	NA	2,097	1,798	1,313	ols
HTS_Make	Comp	2024	0,296	5,560	NA	NA	3,004	2,953	0,296	ols
HTS_Make	Comp	Average	0,802	2,998	NA	NA	2,407	2,069	0,802	ols
HTS_Segment	Comp	2021	1,291	NA	0,835	NA	2,526	1,551	1,063	ols
HTS_Segment	Comp	2022	0,307	NA	0,297	NA	2,037	0,880	0,302	ols
HTS_Segment	Comp	2023	1,310	NA	0,997	NA	1,965	1,424	1,154	ols
HTS_Segment	Comp	2024	0,295	NA	0,836	NA	1,912	1,014	0,565	ols
HTS_Segment	Comp	Average	0,801	NA	0,741	NA	2,110	1,217	0,771	ols
HTS_Type	Comp	2021	1,259	NA	NA	1,832	2,212	1,768	1,546	ols
HTS_Type	Comp	2022	0,333	NA	NA	1,212	2,349	1,298	0,773	ols
HTS_Type	Comp	2023	1,286	NA	NA	2,449	1,999	1,911	1,867	ols
HTS_Type	Comp	2024	0,324	NA	NA	0,711	3,218	1,418	0,518	ols
HTS_Type	Comp	Average	0,801	NA	NA	1,551	2,444	1,599	1,176	ols
HTS_Make	TD	2021	1,295	1,820	NA	NA	2,130	1,748	1,295	tdfp
HTS_Make	TD	2022	0,308	2,042	NA	NA	1,954	1,435	0,308	tdfp
HTS_Make	TD	2023	1,314	1,608	NA	NA	1,732	1,551	1,314	tdfp
HTS_Make	TD	2024	0,301	1,307	NA	NA	1,601	1,070	0,301	tdfp
HTS_Make	TD	Average	0,804	1,694	NA	NA	1,854	1,451	0,804	tdfp
HTS_Segment	TD	2021	1,295	NA	0,688	NA	2,157	1,380	0,992	tdfp
HTS_Segment	TD	2022	0,308	NA	0,359	NA	1,996	0,888	0,334	tdfp
HTS_Segment	TD	2023	1,314	NA	0,895	NA	1,757	1,322	1,105	tdfp
HTS_Segment	TD	2024	0,301	NA	0,667	NA	1,589	0,852	0,484	tdfp
HTS_Segment	TD	Average	0,804	NA	0,652	NA	1,875	1,110	0,728	tdfp
HTS_Type	TD	2021	1,295	NA	NA	0,966	2,049	1,437	1,131	tdfp
HTS_Type	TD	2022	0,308	NA	NA	0,294	1,987	0,863	0,301	tdfp
HTS_Type	TD	2023	1,314	NA	NA	1,441	1,757	1,504	1,378	tdfp

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

Model	Method	Έτος	Total	Make	Segment	Type	Car model	Average	Average WO make and model	Weights
HTS_Type	TD	2024	0,301	NA	NA	1,039	1,599	0,980	0,670	tdfp
HTS_Type	TD	Average	0,804	NA	NA	0,935	1,848	1,196	0,870	tdfp
HTS_Make	TD	2021	1,295	3,588	NA	NA	4,318	3,067	1,295	tdgsa
HTS_Make	TD	2022	0,308	4,280	NA	NA	3,857	2,815	0,308	tdgsa
HTS_Make	TD	2023	1,314	3,672	NA	NA	3,773	2,920	1,314	tdgsa
HTS_Make	TD	2024	0,301	2,849	NA	NA	3,050	2,067	0,301	tdgsa
HTS_Make	TD	Average	0,804	3,597	NA	NA	3,749	2,717	0,804	tdgsa
HTS_Segment	TD	2021	1,295	NA	1,086	NA	4,318	2,233	1,191	tdgsa
HTS_Segment	TD	2022	0,308	NA	0,891	NA	3,857	1,685	0,600	tdgsa
HTS_Segment	TD	2023	1,314	NA	1,552	NA	3,773	2,213	1,433	tdgsa
HTS_Segment	TD	2024	0,301	NA	1,146	NA	3,050	1,499	0,724	tdgsa
HTS_Segment	TD	Average	0,804	NA	1,169	NA	3,749	1,907	0,987	tdgsa
HTS_Type	TD	2021	1,295	NA	NA	1,995	4,318	2,536	1,645	tdgsa
HTS_Type	TD	2022	0,308	NA	NA	2,391	3,857	2,185	1,350	tdgsa
HTS_Type	TD	2023	1,314	NA	NA	3,328	3,773	2,805	2,321	tdgsa
HTS_Type	TD	2024	0,301	NA	NA	2,818	3,050	2,056	1,560	tdgsa
HTS_Type	TD	Average	0,804	NA	NA	2,633	3,749	2,395	1,719	tdgsa
HTS_Make	TD	2021	1,295	3,795	NA	NA	4,575	3,222	1,295	tdgsf
HTS_Make	TD	2022	0,308	4,528	NA	NA	4,123	2,986	0,308	tdgsf
HTS_Make	TD	2023	1,314	3,886	NA	NA	4,043	3,081	1,314	tdgsf
HTS_Make	TD	2024	0,301	3,101	NA	NA	3,269	2,224	0,301	tdgsf
HTS_Make	TD	Average	0,804	3,827	NA	NA	4,003	2,878	0,804	tdgsf
HTS_Segment	TD	2021	1,295	NA	1,244	NA	4,575	2,371	1,270	tdgsf
HTS_Segment	TD	2022	0,308	NA	1,116	NA	4,123	1,849	0,712	tdgsf
HTS_Segment	TD	2023	1,314	NA	1,795	NA	4,043	2,384	1,555	tdgsf
HTS_Segment	TD	2024	0,301	NA	1,492	NA	3,269	1,687	0,897	tdgsf
HTS_Segment	TD	Average	0,804	NA	1,412	NA	4,003	2,073	1,108	tdgsf
HTS_Type	TD	2021	1,295	NA	NA	2,012	4,575	2,627	1,654	tdgsf
HTS_Type	TD	2022	0,308	NA	NA	2,467	4,123	2,299	1,388	tdgsf
HTS_Type	TD	2023	1,314	NA	NA	3,421	4,043	2,926	2,368	tdgsf
HTS_Type	TD	2024	0,301	NA	NA	2,982	3,269	2,184	1,642	tdgsf
HTS_Type	TD	Average	0,804	NA	NA	2,720	4,003	2,509	1,762	tdgsf
GTS_Object	Comp	2021	1,183	1,773	0,621	0,908	2,058	1,308	0,904	wls
GTS_Object	Comp	2022	0,362	2,118	0,382	0,365	1,995	1,044	0,370	wls
GTS_Object	Comp	2023	1,259	1,630	0,889	1,411	1,760	1,390	1,186	wls
GTS_Object	Comp	2024	0,197	1,256	0,595	1,039	1,558	0,929	0,610	wls
GTS_Object	Comp	Average	0,750	1,694	0,622	0,931	1,843	1,168	0,767	wls
HTS_Make	Comp	2021	1,171	1,771	NA	NA	2,058	1,667	1,171	wls
HTS_Make	Comp	2022	0,373	2,120	NA	NA	1,995	1,496	0,373	wls

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

Model	Method	Έτος	Total	Make	Segment	Type	Car model	Average	Average WO make and model	Weights
HTS_Make	Comp	2023	1,260	1,629	NA	NA	1,758	1,549	1,260	wls
HTS_Make	Comp	2024	0,165	1,255	NA	NA	1,559	0,993	0,165	wls
HTS_Make	Comp	Average	0,742	1,694	NA	NA	1,842	1,426	0,742	wls
HTS_Segment	Comp	2021	1,139	NA	0,604	NA	2,056	1,266	0,871	wls
HTS_Segment	Comp	2022	0,396	NA	0,390	NA	1,988	0,925	0,393	wls
HTS_Segment	Comp	2023	1,235	NA	0,870	NA	1,755	1,287	1,053	wls
HTS_Segment	Comp	2024	0,174	NA	0,604	NA	1,563	0,781	0,389	wls
HTS_Segment	Comp	Average	0,736	NA	0,617	NA	1,840	1,065	0,677	wls
HTS_Type	Comp	2021	1,103	NA	NA	0,868	2,055	1,342	0,985	wls
HTS_Type	Comp	2022	0,428	NA	NA	0,397	1,988	0,938	0,413	wls
HTS_Type	Comp	2023	1,212	NA	NA	1,391	1,752	1,452	1,302	wls
HTS_Type	Comp	2024	0,187	NA	NA	1,036	1,566	0,929	0,611	wls
HTS_Type	Comp	Average	0,732	NA	NA	0,923	1,840	1,165	0,828	wls
GTS_Object	BU	2021	1,073	1,762	0,568	0,852	2,053	1,262	0,831	
GTS_Object	BU	2022	0,443	2,171	0,408	0,423	1,989	1,087	0,425	
GTS_Object	BU	2023	1,205	1,664	0,859	1,385	1,752	1,373	1,150	
GTS_Object	BU	2024	0,128	1,252	0,601	1,073	1,563	0,923	0,601	
GTS_Object	BU	Average	0,712	1,712	0,609	0,933	1,839	1,161	0,751	
HTS_Make	BU	2021	1,073	1,762	NA	NA	2,053	1,629	1,073	
HTS_Make	BU	2022	0,443	2,171	NA	NA	1,989	1,534	0,443	
HTS_Make	BU	2023	1,205	1,664	NA	NA	1,752	1,540	1,205	
HTS_Make	BU	2024	0,128	1,252	NA	NA	1,563	0,981	0,128	
HTS_Make	BU	Average	0,712	1,712	NA	NA	1,839	1,421	0,712	
HTS_Make	MO	2021	1,261	1,806	NA	NA	2,118	1,728	1,261	
HTS_Make	MO	2022	0,275	2,044	NA	NA	1,955	1,425	0,275	
HTS_Make	MO	2023	1,307	1,607	NA	NA	1,731	1,548	1,307	
HTS_Make	MO	2024	0,195	1,284	NA	NA	1,579	1,019	0,195	
HTS_Make	MO	Average	0,759	1,685	NA	NA	1,846	1,430	0,759	
HTS_Segment	BU	2021	1,073	NA	0,568	NA	2,053	1,231	0,821	
HTS_Segment	BU	2022	0,443	NA	0,408	NA	1,989	0,947	0,426	
HTS_Segment	BU	2023	1,205	NA	0,859	NA	1,752	1,272	1,032	
HTS_Segment	BU	2024	0,128	NA	0,601	NA	1,563	0,764	0,365	
HTS_Segment	BU	Average	0,712	NA	0,609	NA	1,839	1,053	0,661	

Πρόβλεψη ταξινομήσεων αυτοκινήτων στην Ελλάδα με χρήση ιεραρχικών μεθόδων

Model	Method	Έτος	Total	Make	Segment	Type	Car model	Average	Average WO make and model	Weights
HTS_Segment	MO	2021	1,263	NA	0,673	NA	2,146	1,361	0,968	
HTS_Segment	MO	2022	0,290	NA	0,354	NA	1,997	0,880	0,322	
HTS_Segment	MO	2023	1,286	NA	0,883	NA	1,756	1,308	1,085	
HTS_Segment	MO	2024	0,253	NA	0,654	NA	1,581	0,829	0,454	
HTS_Segment	MO	Average	0,773	NA	0,641	NA	1,870	1,095	0,707	
HTS_Type	BU	2021	1,073	NA	NA	0,852	2,053	1,326	0,963	
HTS_Type	BU	2022	0,443	NA	NA	0,423	1,989	0,952	0,433	
HTS_Type	BU	2023	1,205	NA	NA	1,385	1,752	1,447	1,295	
HTS_Type	BU	2024	0,128	NA	NA	1,073	1,563	0,921	0,601	
HTS_Type	BU	Average	0,712	NA	NA	0,933	1,839	1,161	0,823	
HTS_Type	MO	2021	1,151	NA	NA	0,865	1,999	1,338	1,008	
HTS_Type	MO	2022	0,413	NA	NA	0,347	1,986	0,915	0,380	
HTS_Type	MO	2023	1,197	NA	NA	1,374	1,751	1,441	1,286	
HTS_Type	MO	2024	0,410	NA	NA	1,033	1,620	1,021	0,722	
HTS_Type	MO	Average	0,793	NA	NA	0,905	1,839	1,179	0,849	

Βιβλιογραφία

- [1] Wikipedia - History of the automobile
https://en.wikipedia.org/wiki/History_of_the_automobile
- [2] European Automobile Manufacturers Association (ACEA) – Historical Data
<https://www.acea.auto/reliable-data-statistics/>
- [3] FIAT, Renault, Citroën company histories
- [4] Volkswagen Official Site – Beetle History
<https://www.vw.com/en/newsroom/lifestyle-and-heritage/an-ode-to-the-bug.html>
- [5] European Commission – Automotive Industry Environmental Regulations
https://single-market-economy.ec.europa.eu/sectors/automotive-industry/environmental-protection/emissions-automotive-sector_en
- [6] Bloomberg – History of Car Industry Mergers
<https://www.bloomberg.com/>
- [7] Toyota Global – History of the Prius
<https://global.toyota/en/prius20th/evolution/>
- [8] Tesla.com – Company History
<https://www.tesla.com/about>
- [9] European Commission – Green Deal & Mobility
https://transport.ec.europa.eu/transport-themes/mobility-strategy_en
- [10] McKinsey & Company – The Future of Mobility 2030+
<https://www.mckinsey.com/>
- [11] PwC Automotive Industry Outlook 2024
<https://www.pwccn.com/>
- [12] Eurostat – Passenger car registrations in the EU
<https://ec.europa.eu/eurostat/>
- [13] ΕΛΣΤΑΤ – Στατιστικά Οχημάτων & Εισαγωγών
<https://www.statistics.gr/>
- [14] NewsAuto.gr
<https://www.newsauto.gr>
- [15] Autotriti – Αναλύσεις αγοράς:
<https://www.autotriti.gr>
- [16] ACEA (European Automobile Manufacturers Association),
www.acea.auto
- [17] European Commission – “Fit for 55”
<https://www.consilium.europa.eu/en/policies/fit-for-55/>

- [19] Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). *Optimal combination forecasts for hierarchical time series*. Computational Statistics & Data Analysis, 55(9), 2579–2589.
- [20] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.
<https://otexts.com/fpp3/>
- [21] Fliedner, G. (1999). *An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation*. Computers & Operations Research, 26(10-11), 1133–1149.
- [22] Kahn, K. B. (1998). *Revisiting top-down versus bottom-up forecasting*. The Journal of Business Forecasting Methods & Systems, 17(2), 14.
- [23] Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). *Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization*. Journal of the American Statistical Association, 114(526), 804–819.
- [24] Boylan, J. E., & Babai, M. Z. (2016). *Supply chain forecasting: theory, practice, their gap and the future*. European Journal of Operational Research, 252(1), 1–26.
- [25] Taylor, J. W., & McSharry, P. E. (2007). *Short-term load forecasting methods: An evaluation based on European data*. IEEE Transactions on Power Systems, 22(4), 2213–2219.
- [26] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). Chapter 11: Forecasting hierarchical and grouped time series.
<https://otexts.com/fpp3/hierarchical.html>
- [27] Hyndman, R. J., Lee, A., & Wang, E. (2016). *Fast computation of reconciled forecasts for hierarchical and grouped time series*. Working paper. Monash University.
- [28] Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). *Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization*. Journal of the American Statistical Association, 114(526), 804–819.
- [29] Kourentzes, N., Barrow, D. K., & Petropoulos, F. (2019). *Another look at forecast selection and combination: Evidence from forecast pooling*. International Journal of Production Economics, 209, 226–235.
- [30] Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2022). *Forecasting with Temporal Hierarchies*. European Journal of Operational Research, 295(2), 739–755.
- [31] Πετρόπουλος Φ. & Ασημακόπουλος Β. (2013), Επιχειρησιακές Προβλέψεις. Αθήνα, Εκδόσεις: Συμμετρία
- [32] Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. Operational Research Quarterly, 20(4), 451–468

- [33] Clemen, R. (1989). Combining forecasts: A review and annotated bibliography with discussion. *International Journal of Forecasting*, 5, 559–608
- [34] Everette S., Gardner Jr. (2006). Exponential smoothing: The state of the art—Part II, *International Journal of Forecasting* Volume 22, Issue 4, Pages 637– 666
- [35] Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, 4(1), 1–28.
- [36] Leonard J.Tashman (2000), Out-of-sample tests of forecasting accuracy: an analysis and review, Volume 16, Issue 4, Pages 437-450.
- [37] Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54-74
- [38] Rob J. Hyndman (2006), Another look at forecast – Accuracy metrics for intermittent demand, *Foresight*, 4 (4) 43-46
- [39] Rob J Hyndman, Anne B Koehler, Ralph D Snyder, Simone Grose, 2002, A state space framework for automatic forecasting using exponential smoothing methods, Volume 18, Issue 3, Pages 439-454
- [40] S. Shen, G. Li, and H. Song (2011), "Combination forecasts of international tourism demand," *Annals of Tourism Research*, vol. 38 (1), pp. 72-89
- [41] Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19, 715–725.
- [42] Σημειώσεις του μαθήματος «Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές», του Διατμηματικού Προγράμματος Μεταπτυχιακών Σπουδών (ΔΠΜΣ) «Τεχνο-Οικονομικά Συστήματα», του Εθνικού Μετσόβιου Πολυτεχνείου

<http://www.fsu.gr>