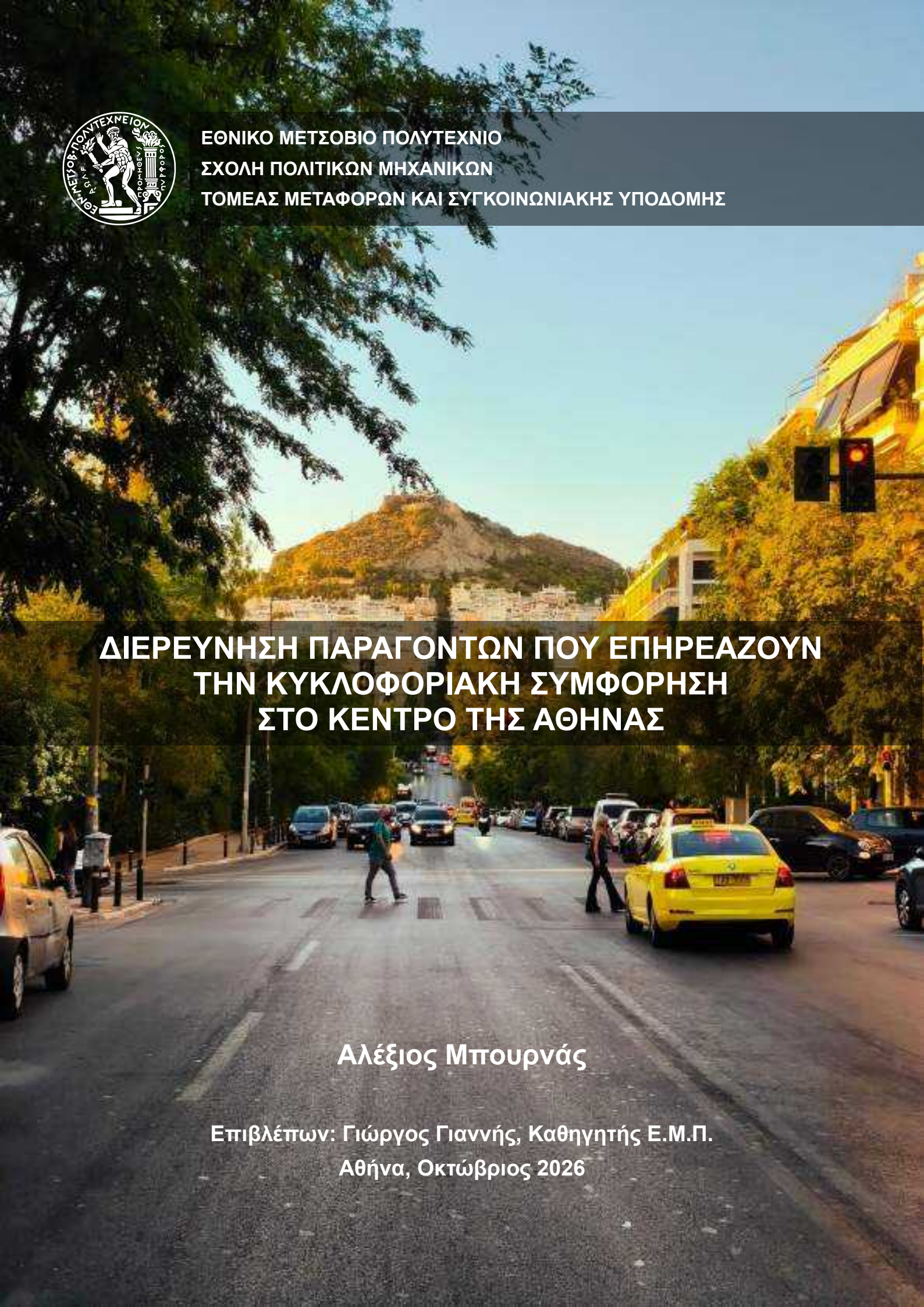


ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

ΔΙΕΡΕΥΝΗΣΗ ΠΑΡΑΓΟΝΤΩΝ ΠΟΥ ΕΠΗΡΕΑΖΟΥΝ ΤΗΝ ΚΥΚΛΟΦΟΡΙΑΚΗ ΣΥΜΦΟΡΗΣΗ ΣΤΟ ΚΕΝΤΡΟ ΤΗΣ ΑΘΗΝΑΣ

Αλέξιος Μπουρνάς

Επιβλέπων: Γιώργος Γιαννής, Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2026

Ευχαριστίες

Με την ολοκλήρωση της Διπλωματικής μου Εργασίας ολοκληρώνεται ο κύκλος των προπτυχιακών σπουδών στη Σχολή Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου.

Θα ήθελα πρωτίστως να ευχαριστήσω θερμά τον κ. Γιώργο Γιαννή για την ανάθεση του συγκεκριμένου θέματος το οποίο αποτελεί και αντικείμενο ιδιαίτερου προσωπικού ενδιαφέροντος που αφορά την αστική οδική κυκλοφορία και τη βιωσιμότητα των μετακινήσεων.

Έπειτα, θα ήθελα να ευχαριστήσω θερμά την ερευνήτρια Έλενα Θεοδωράκη, για την καθοδήγηση που μου παρείχε καθ' όλη τη διάρκεια της συγγραφής, την επίλυση αποριών και την παροχή πολύτιμων συμβουλών και διορθώσεων.

Επίσης, θα ήθελα να ευχαριστήσω όλη την Πολυτεχνειακή κοινότητα, συμφοιτητές και καθηγητές που υπήρξαν μέρος της καθημερινότητας μου τα τελευταία πέντε χρόνια. Μέσα σε αυτό το χρονικό διάστημα, είχα τη δυνατότητα να δικτυωθώ, να συνεργαστώ αλλά πάνω απ' όλα να λάβω πολύτιμες γνώσεις και να αναπτύξω δεξιότητες τις οποίες θα εφαρμόζω στην επαγγελματική μου πορεία.

Τέλος, θέλω να ευχαριστήσω την οικογένεια μου για τη δυνατότητα που μου δώσανε να σπουδάσω σε αυτή τη σχολή και τη στήριξη που μου παρείχαν κατά τη διάρκεια των σπουδών.

Αθήνα, Οκτώβριος 2026
Αλέξιος Μπουρνάς

ΔΙΕΡΕΥΝΗΣΗ ΠΑΡΑΓΟΝΤΩΝ ΠΟΥ ΕΠΗΡΕΑΖΟΥΝ ΤΗΝ ΚΥΚΛΟΦΟΡΙΑΚΗ ΣΥΜΦΟΡΗΣΗ ΣΤΟ ΚΕΝΤΡΟ ΤΗΣ ΑΘΗΝΑΣ

Αλέξιος Μπουρνάς

Επιβλέπων: Γιώργος Γιαννής, Καθηγητής Ε.Μ.Π.

Σύνοψη

Η παρούσα διπλωματική εργασία έχει στόχο τη διερεύνηση της επίδρασης κοινωνικοοικονομικών, κυκλοφοριακών και μετεωρολογικών παραγόντων στη συμφόρηση στο κέντρο της Αθήνας, με εφαρμογή μοντέλων παλινδρόμησης και μηχανικής μάθησης. Τα αποτελέσματα δείχνουν αυξητική τάση της συμφόρησης από το 2022 και εποχικότητα, με το τρίτο τρίμηνο να παρουσιάζει σταθερά τη χαμηλότερη κυκλοφορία. Η σημαντικότερη παράμετρος επιρροής της συμφόρησης αναδείχθηκε ο αριθμός των ΙΧ, ο οποίος αυξήθηκε κατά 396.5 χιλιάδες στην Αττική από το 2021 έως το 2025. Η επίδραση των καιρικών συνθηκών και της σύνθεσης της κυκλοφορίας προέκυψε μικρή. Η έντονη βροχόπτωση μειώνει την ταχύτητα έως και 3.3% της ταχύτητας ελεύθερης ροής (FFS), ενώ η αύξηση του ποσοστού ταξί και λεωφορείων εις βάρος των ΙΧ τη βελτιώνει ελαφρώς (0.18% και 0.44% της FFS αντίστοιχα). Τα μοντέλα μηχανικής μάθησης δείχνουν μείωση της ταχύτητας από 0.06% έως 0.17% της FFS ανά ποσοστιαία μονάδα αύξησης των ΙΧ στο μεγαλύτερο εύρος τιμών, και έως 1.20% σε μία στενή ζώνη γύρω στο 38 έως 42%, η οποία δεν πρέπει να γενικευτεί. Η αντικατάστασή τους από λεωφορεία διατηρεί την ταχύτητα σταθερή ή τη μειώνει ελάχιστα, γεγονός που, σε συνδυασμό με τη μεγαλύτερη μεταφορική ικανότητα των λεωφορείων, υποδηλώνει ότι τα ΜΜΜ θα μπορούσαν να μεταφέρουν περισσότερα άτομα χωρίς επιβάρυνση της ταχύτητας.

Λέξεις Κλειδιά: κυκλοφοριακή συμφόρηση, Αθήνα, αστική οδική κυκλοφορία, μεταβλητότητα/αξιοπιστία χρόνου διαδρομής, ταχύτητα κυκλοφορίας, κοινωνικοοικονομικοί παράγοντες, μετεωρολογικές συνθήκες, σύνθεση της κυκλοφορίας, σημαντικότητα παραγόντων, στατιστική ανάλυση, μοντέλα γραμμικής παλινδρόμησης, μοντέλα μηχανικής μάθησης, μοντέλο ενίσχυσης κλίσης, μοντέλο τυχαίων δασών

INVESTIGATION OF THE FACTORS THAT IMPACT TRAFFIC CONGESTION IN THE CENTER OF ATHENS

Alexios Bournas

Supervisor: George Yannis, Professor NTUA

Abstract

This present research investigates the influence of socioeconomic, traffic and meteorological factors on congestion in central Athens, applying regression and machine learning models. The results show an upward trend in congestion since 2022 and clear seasonality, with the third quarter consistently showing the lowest traffic. The main statistically significant variable influencing congestion was the number of private cars, which increased by 396.5 thousand in Attica from 2021 to 2025. The effects of weather conditions and traffic composition were found to be small. Heavy rainfall reduces speed by up to 3.3% of the free-flow speed (FFS), while increasing the share of taxis and buses at the expense of private cars improves it slightly (by 0.18% and 0.44% of the FFS, respectively). The machine learning models indicate a speed reduction of 0.06% to 0.17% of the FFS per percentage-point increase in private cars over most of the observed range, and up to 1.20% in a narrow zone around 38 to 42%, which should not be generalized. Replacing private cars with buses keeps speed stable or reduces it only minimally, which, combined with the greater passenger capacity of buses, suggests that public transport could carry more people without a loss in speed.

Key Words: traffic congestion, Athens, urban road traffic, travel time variability/reliability, traffic speed, socioeconomic factors, meteorological conditions, traffic composition, feature importance, statistical analysis, linear regression models, machine learning models, gradient boosting model, random forest model

Περίληψη

Στόχος της παρούσας διπλωματικής εργασίας είναι η **διερεύνηση των παραγόντων που επηρεάζουν την κυκλοφοριακή συμφόρηση στο κέντρο της Αθήνας**. Για την προσέγγιση του φαινομένου, συλλέχθηκαν δεδομένα που περιλαμβάνουν κοινωνικοοικονομικούς, κυκλοφοριακούς και μετεωρολογικούς παράγοντες.

Η αρχική προκαταρκτική ανάλυση των δεδομένων ανέδειξε σημαντικά ευρήματα για την κυκλοφοριακή συμφόρηση στην Αθήνα. Από το 2022 έως το 2025, οι συνολικές καθυστερήσεις αυξήθηκαν κατά 72% σε έξι περιφερειακούς οδικούς άξονες και κατά 40% σε 12 κεντρικούς άξονες της Αθήνας το οποίο αναδεικνύει την **αύξηση της έντασης της συμφόρησης**. Εκτός από την ένταση της συμφόρησης, παρατηρήθηκε και διεύρυνση των ωρών αιχμής μέσα στην ημέρα δηλαδή **αύξηση της διάρκειας συμφόρησης**. Από τις εννέα ώρες ανάλυσης ανά εργάσιμη ημέρα (8 – 10:00 & 12 – 19:00), το 2025, το 31% αυτών στις περιφερειακές οδούς παρουσιάζουν ταχύτητες κάτω του 50% της ταχύτητας ελεύθερης ροής και το αντίστοιχο ποσοστό των οδών εντός κέντρου αγγίζει το 52%. Και στις δύο περιπτώσεις υπάρχει άνοδος από το 2022 κατά 89% και 42% αντίστοιχα.

Ταυτόχρονα με την επιδείνωση της συμφόρησης, ο **αριθμός των επιβατικών ΙΧ** οχημάτων που κυκλοφορούνε στην Αττική **έχει αυξηθεί κατά 396.5 χιλιάδες** από το 2021 έως το 2025 (ΕΛΣΤΑΤ) το οποίο έχει ως αποτέλεσμα την επιβάρυνση του οδικού δικτύου. Παράλληλα, μέσω των δεδομένων σύνθεσης της κυκλοφορίας σε δέκα έξι οδούς στο κέντρο της Αθήνας, διαπιστώθηκε η διαχρονική υπεροχή αυτών στην κυκλοφορία με **ποσοστό 50% επί του συνόλου**. Το ποσοστό των λεωφορείων παραμένει στο πολύ χαμηλό επίπεδο του 2 - 5%.

Οι κοινωνικοοικονομικοί παράγοντες **δεν είχαν σαφή επίδραση στη διακύμανση της συμφόρησης** καθώς αυτή, όπως διαπιστώθηκε, καθορίζεται κατά 86.7% από μία σταθερή ανοδική τάση από το έτος 2022 και μία εποχικότητα, με το 3^{ο} τρίμηνο του έτους να εμφανίζει διαχρονικά τις μικρότερες καθυστερήσεις. Ύστερα από αφαίρεση της τάσης και της εποχικότητας των επεξηγηματικών μεταβλητών, η μόνη μεταβλητή με σημαντική συσχέτιση και στους δύο δείκτες Pearson και Spearman είναι ο αριθμός των ΙΧ οχημάτων ενώ οριακή ένδειξη (σημαντική μόνο κατά Spearman) υπάρχει για τους εργαζόμενους και το ποσοστό εργατικού δυναμικού. Τα μοντέλα παλινδρόμησης επιχείρησαν να εξετάσουν την πιθανή βελτίωση του βασικού μοντέλου (τάση + τρίμηνο) με την διαδοχική προσθήκη των σημαντικών μεταβλητών. Από αυτές, μόνο ο συντελεστής του «αριθμού ΙΧ» παρέμεινε θετικός και στατιστικά σημαντικός το οποίο όμως δεν πρέπει να λαμβάνεται ως αιτιότητα αύξησης της συμφόρησης αλλά ως συσχέτιση.

Η επιρροή των μεταβλητών σύνθεσης της κυκλοφορίας **είναι συνολικά μικρή και διαφέρει ανάμεσα στα μοντέλα**. Το μοντέλο γραμμικής παλινδρόμησης ανέδειξε ως σημαντικές μόνο αυτές του ποσοστού ταξί και λεωφορείων στο σύνολο της κυκλοφορίας με αύξηση της μέσης ταχύτητας κατά 0.18% και 0.44%

της FFS αντίστοιχα, για κάθε ποσοστιαία μονάδα αύξησης αυτών εις βάρος των IX. Τα διαγράμματα μερικής εξάρτησης των μοντέλων GBM και RF έδειξαν μείωση της μέσης ταχύτητας, με ρυθμό από 0.06% έως 0.17% της FFS για κάθε ποσοστιαία μονάδα αύξησης των IX στο μεγαλύτερο εύρος τιμών, και έως 1.20% σε μία στενή ζώνη γύρω στο 38 έως 42%, η οποία δεν πρέπει να γενικευτεί. Για τον σωστό υπολογισμό έγινε η υπόθεση αναλογικής μείωσης των υπόλοιπων μεταβλητών σύνθεσης έτσι ώστε το άθροισμα να είναι πάντα 100%.

Στο ερώτημα **ισάριθμης αντικατάστασης IX οχημάτων από λεωφορεία**, τα μοντέλα μηχανικής μάθησης έδειξαν πως η μεταβολή στην ταχύτητα διαφέρει ανά σύνολο ελέγχου των δεδομένων αλλά είναι πρακτικά αμελητέα. Τα τρία μοντέλα κατανέμουν διαφορετικά τη σημαντικότητα στις μεταβλητές. Στην παλινδρόμηση τα ποσοστά οχημάτων έχουν πρακτικά μηδενική σημασία (αθροιστικά περίπου 0.7%), στο GBM συμβάλλουν αθροιστικά περίπου 8%, ενώ στο RF περίπου 20%. Το μοντέλο RF κατανέμει τη σημαντικότητα πιο ισόρροπα μεταξύ των μεταβλητών, ενώ τα GBM και η παλινδρόμηση συγκεντρώνουν σχεδόν όλη τη σημασία στις χωρικές και χρονικές μεταβλητές.

Η επίδραση των καιρικών φαινομένων στην ταχύτητα κυκλοφορίας των οχημάτων στο κέντρο της Αθήνας προέκυψε **μικρή αλλά στατιστικά σημαντική** στο μοντέλο παλινδρόμησης. Οι συντελεστές των μεταβλητών ανέδειξαν **μείωση των ταχυτήτων από 1.3% έως 3.3% της ταχύτητας ελεύθερης ροής** σε συνθήκες μικρής (<2.5 mm/h) και μεγάλης (>7.6 mm/h) έντασης βροχής αντίστοιχα. Η μικρή αυτή μείωση μπορεί να οφείλεται στις ήδη χαμηλές ταχύτητες του αστικού οδικού δικτύου που όπως διαπιστώθηκε στην προκαταρκτική ανάλυση, διατηρούνται σε πολύ χαμηλό επίπεδο. Αντίστοιχα, στο μοντέλο Ενίσχυσης Κλίσης βασισμένο σε Ιστόγραμμα (HGB), η επίδραση των μετεωρολογικών παραγόντων προέκυψε αμελητέα.

Περιεχόμενα

Περίληψη V
Περιεχόμενα VII
Περιεχόμενα Πινάκων X
Περιεχόμενα Διαγραμμάτων XI
Περιεχόμενα Εικόνων XII
1. Εισαγωγή 1
1.1. Γενική ανασκόπηση 1
1.2. Στόχος 2
1.3. Μεθοδολογία 3
1.4. Δομή διπλωματικής εργασίας 3
2. Βιβλιογραφική Ανασκόπηση 5
2.1. Εισαγωγή 5
2.2. Ορισμός και ποσοτικοποίηση της κυκλοφοριακής συμφόρησης 5
2.3. Επίδραση κοινωνικοοικονομικών παραγόντων στην συμφόρηση του αστικού δικτύου 8
2.4. Επίδραση μετεωρολογικών συνθηκών στην κυκλοφορία 9
2.5. Επιλογή μέσου μετακίνησης 10
2.6. Επίδραση των Μέσων Μαζικής Μεταφοράς στη κυκλοφορία 11
2.7. Πρόβλεψη των χρόνων διαδρομής 12
2.7.1. Γενικά 12
2.7.2. Μέθοδοι βασισμένοι σε μοντέλα 13
2.7.3. Μέθοδοι βασισμένοι σε δεδομένα 13
2.7.4. Μεταβλητότητα / Αξιοπιστία χρόνου διαδρομής 13
2.8. Μηχανική Μάθηση στην πρόβλεψη των χρόνων διαδρομής 14
2.9. Σύνοψη 17
3. Θεωρητικό Υπόβαθρο 18
3.1. Εισαγωγή 18
3.2. Βασικές έννοιες στατιστικής 18
3.2.1. Μέγεθος δείγματος (N) 18
3.2.2. Μέση Τιμή (X) 18
3.2.3. Εύρος (R) 18
3.2.4. Διακύμανση Δείγματος (s^2) 18
3.2.5. Τυπική Απόκλιση Δείγματος (s) 19
3.3. Είδη Μεταβλητών 19
3.3.1. Γενικά 19
3.3.2. Ποσοτικές (Συνεχείς) Μεταβλητές 19
3.3.3. Ποιοτικές (Κατηγορικές) Μεταβλητές 19

3.4. Γραμμική Παλινδρόμηση 20
3.4.1. Γενικά 20
3.4.2. Μέθοδος Ελαχίστων Τετραγώνων (OLS) 21
3.4.3. Εκτιμητής διασποράς Newey-West (HAC) 21
3.4.4. Έλεγχος συσχέτισης μεταβλητών 22
3.4.5. Έλεγχος σημαντικότητας μεταβλητών 23
3.5. Αξιολόγηση μοντέλων παλινδρόμησης 23
3.5.1. Ερμηνευτική ισχύς και προσαρμογή 24
3.5.2. Προβλεπτική ακρίβεια 24
3.5.3. Σύγκριση και επιλογή μοντέλων 25
3.5.4. Έλεγχοι καταλοίπων 25
3.5.5. Έλεγχος αυτοσυσχέτισης και πολυσυγγραμικότητας 26
3.6. Μοντέλα Μηχανικής Μάθησης 26
3.6.1. Γενικά 26
3.6.2. Τυχαία Δάση (Random Forests – RF) 26
3.6.3. Αλγόριθμοι Ενίσχυσης Κλίσης (Gradient Boosting Machines) 27
3.7. Αξιολόγηση μοντέλων μηχανικής μάθησης 29
3.8. Σημαντικότητα μεταβλητών 29
4. Συλλογή και Επεξεργασία Στοιχείων 31
4.1. Εισαγωγή 31
4.2. Κοινωνικοοικονομικά δεδομένα 31
4.3. Μετεωρολογικά δεδομένα 33
4.4. Δεδομένα σύνθεσης κυκλοφορίας 33
4.4.1. Καθαρισμός δεδομένων 35
4.4.2. Ανάλυση δεδομένων 35
4.5. Δεδομένα χρόνου διαδρομής και ταχύτητας 41
4.5.1. Μεταβολή έντασης και διάρκειας της συμφόρησης 41
4.5.2. Μεταβλητότητα χρόνων διαδρομής 47
5. Εφαρμογή Μεθοδολογίας και Αποτελέσματα 53
5.1. Εισαγωγή 53
5.2. Συσχέτιση κοινωνικοοικονομικών δεδομένων με τη συμφόρηση 53
5.2.1. Δεδομένα 53
5.2.2. Καθορισμός μεταβλητών 54
5.2.3. Ανάλυση συσχέτισης 54
5.2.4. Μοντέλα παλινδρόμησης 57
5.3. Μοντέλο Γραμμικής Παλινδρόμησης (ταχύτητα - καιρός) 59
5.3.1. Επεξεργασία και συγχώνευση δεδομένων 59
5.3.2. Επεξήγηση και καθορισμός μεταβλητών 60
5.3.3. Έλεγχος συσχέτισης μεταβλητών (πολυσυγγραμικότητα) 61

5.3.4. Ανάλυση και αποτελέσματα μοντέλου 62
5.3.5. Ερμηνεία μοντέλου 65
5.4. Μοντέλο Ενίσχυσης Κλίσης Βασισμένη σε Ιστόγραμμα (ταχύτητα - καιρός) 66
5.4.1. Γενικά 66
5.4.2. Επεξήγηση Μεταβλητών 67
5.4.3. Διαχωρισμός δεδομένων εκπαίδευσης / ελέγχου 67
5.4.4. Επιλογή και ρύθμιση υπερπαραμέτρων 67
5.4.5. Αποτελέσματα και αξιολόγηση μοντέλου 67
5.5. Σύγκριση μοντέλου HGBM με Γραμμική Παλινδρόμηση (ταχύτητα - καιρός) 69
5.6. Μοντέλο Γραμμικής Παλινδρόμησης (ταχύτητα - σύνθεση κυκλοφορίας) . 72
5.6.1. Επεξεργασία και συγχώνευση δεδομένων 72
5.6.2. Επεξήγηση και καθορισμός μεταβλητών 73
5.6.3. Έλεγχος συσχέτισης μεταβλητών (πολυσυγγραμικότητα) 74
5.6.4. Ανάλυση και αποτελέσματα μοντέλου 75
5.6.5. Ερμηνεία μοντέλου 78
5.7. Μοντέλο Ενίσχυσης Κλίσης (GBM) (ταχύτητα - σύνθεση) 79
5.7.1. Γενικά 79
5.7.2. Επεξήγηση μεταβλητών 79
5.7.3. Διαχωρισμός δεδομένων εκπαίδευσης / ελέγχου 79
5.7.4. Επιλογή και ρύθμιση υπερπαραμέτρων 79
5.7.5. Αποτελέσματα και αξιολόγηση μοντέλου 80
5.7.6. Απομόνωση μεταβλητής IX και διαγράμματα μερικής εξάρτησης 81
5.8. Μοντέλο Τυχαίων Δασών (RF) (ταχύτητα - σύνθεση) 84
5.8.1. Γενικά 84
5.8.2. Επεξήγηση μεταβλητών 85
5.8.3. Διαχωρισμός δεδομένων εκπαίδευσης / ελέγχου 85
5.8.4. Επιλογή και ρύθμιση υπερπαραμέτρων 85
5.8.5. Αποτελέσματα και αξιολόγηση μοντέλου 85
5.8.6. Απομόνωση μεταβλητής IX και διαγράμματα μερικής εξάρτησης 87
5.9. Σύγκριση μοντέλων GBM, RF και Γραμμικής Παλινδρόμησης (ταχύτητα - σύνθεση κυκλοφορίας) 89
6. Συμπεράσματα 91
6.1. Σύνοψη αποτελεσμάτων 91
6.2. Περιορισμοί 92
6.3. Προτάσεις 93
7. Βιβλιογραφία 94

Περιεχόμενα Πινάκων

Πίνακας 2.1: Δείκτες ποσοτικοποίησης της συμφόρησης (Aftabuzzaman, 2007) 8
Πίνακας 2.2: Έρευνες για την επίδραση κοινωνικοοικονομικών παραγόντων στην κυκλοφορία 9
Πίνακας 2.3: Έρευνες για την επίδραση των μετεωρολογικών συνθηκών στη ταχύτητα 10
Πίνακας 2.4: Έρευνες για την πρόβλεψη των χρόνων διαδρομής με μεθόδους Μηχανικής Μάθησης 17
Πίνακας 4.1: Κοινωνικοοικονομικά δεδομένα 31
Πίνακας 4.2: Μετεωρολογικά δεδομένα 33
Πίνακας 4.3: Οδικά τμήματα καταγραφής της σύνθεσης κυκλοφορίας 34
Πίνακας 4.4: Καθαρισμός Δεδομένων (Τιμές με Z>= 5) 35
Πίνακας 4.5: Ετήσια μεταβολή ποσοστών οχημάτων στο σύνολο της κυκλοφορίας 2021 - 2025 37
Πίνακας 4.6: Χρονικά διαστήματα καταγραφής δεδομένων από την Google 41
Πίνακας 4.7: Χρονικά διαστήματα καταγραφής δεδομένων από την Tom Tom 41
Πίνακας 4.8: Κεντρικοί οδικοί άξονες ανάλυσης 42
Πίνακας 4.9: Περιφερειακοί οδικοί άξονες ανάλυσης 42
Πίνακας 4.10: Ώρες αιχμής για 12 οδικούς άξονες στο κέντρο της Αθήνας 49
Πίνακας 5.1: Βάση κοινωνικοοικονομικών δεδομένων και συμφόρησης 54
Πίνακας 5.2: Συσχέτιση κοινωνικοοικονομικών μεταβλητών με τάση και τρίμηνο 56
Πίνακας 5.3: Συσχέτιση κοινωνικοοικονομικών μεταβλητών με τη συμφόρηση σε επίπεδα 57
Πίνακας 5.4: Συσχέτιση κοινωνικοοικονομικών μεταβλητών με τη συμφόρηση (αφαίρεση τάσης και εποχικότητας) 57
Πίνακας 5.5: Δείκτες απόδοσης βασικών μοντέλων (μεταβλητές τάσης και τριμήνου) 57
Πίνακας 5.6: Δείκτες απόδοσης μοντέλου M1: «τρίμηνο + IX» 57
Πίνακας 5.7: Συντελεστής, τυπικό σφάλμα και τιμή p της μεταβλητής IX στο M1 μοντέλο 58
Πίνακας 5.8: Δείκτες απόδοσης μοντέλου M4: «τρίμηνο + IX + εργαζόμενοι» 58
Πίνακας 5.9: Συντελεστές, τυπικό σφάλμα και τιμές p των μεταβλητών: IX και εργαζόμενοι στο M4 μοντέλο: «τρίμηνο + IX + εργαζόμενοι» 58
Πίνακας 5.10: Μεταβλητές μοντέλου γραμμικής παλινδρόμησης (ταχύτητα – καιρός) 61
Πίνακας 5.11: Κατηγορίες αναφοράς μοντέλου (ταχύτητα – καιρός) 61
Πίνακας 5.12: Συντελεστές διόγκωσης της διακύμανσης (VIF) στο μοντέλο ταχύτητα – καιρός 61
Πίνακας 5.13: Συντελεστές και τιμές p των μεταβλητών του μοντέλου ταχύτητα - καιρός 63
Πίνακας 5.14: Δείκτες προσαρμογής και έλεγχοι καταλοίπων του μοντέλου ταχύτητα – καιρός 64
Πίνακας 5.15: Έλεγχος συνολικής σημαντικότητας μεταβλητών με Wald Test (ταχύτητα – καιρός) 64
Πίνακας 5.16: Σχετική σημαντικότητα μεταβλητών στο μοντέλο ταχύτητα – καιρός 64
Πίνακας 5.17: Δείκτες αξιολόγησης μοντέλου HGBM (ταχύτητα – καιρός) 68
Πίνακας 5.18: Μεταθετική σημαντικότητα μεταβλητών μοντέλου HGBM (ταχύτητα – καιρός) 68
Πίνακας 5.19: Σχετική σημαντικότητα μεταβλητών μοντέλου HGBM (ταχύτητα – καιρός) 68
Πίνακας 5.20: Αποτελέσματα μοντέλων για διαφορετικά δεδομένα ελέγχου (ταχύτητα – καιρός) 69
Πίνακας 5.21: Σύγκριση σημαντικότητας μεταβλητών μεταξύ του μοντέλου παλινδρόμησης και HGBM (ταχύτητα – καιρός) 70

Πίνακας 5.22: Σύγκριση των δεικτών απόδοσης μεταξύ των μοντέλων παλινδρόμησης και HGBM (ταχύτητα - καιρός)..... 71
Πίνακας 5.23: Αντιστοίχιση οδών ανάμεσα στις δύο βάσεις δεδομένων..... 73
Πίνακας 5.24: Μεταβλητές μοντέλου γραμμικής παλινδρόμησης (ταχύτητα - σύνθεση κυκλοφορίας)..... 74
Πίνακας 5.25: Κατηγορίες αναφοράς μοντέλου γραμμικής παλινδρόμησης (ταχύτητα - σύνθεση κυκλοφορίας)..... 74
Πίνακας 5.26: Συντελεστές Διόγκωσης Διακύμανσης (VIF) στο μοντέλο ταχύτητα - σύνθεση κυκλοφορίας..... 75
Πίνακας 5.27: Συντελεστές και τιμές p των μεταβλητών του μοντέλου ταχύτητα – σύνθεση..... 76
Πίνακας 5.28: Δείκτες προσαρμογής και έλεγχοι καταλοίπων του μοντέλου ταχύτητα – σύνθεση..... 76
Πίνακας 5.29: Έλεγχος συνολικής σημαντικότητας μεταβλητών με Wald Test (ταχύτητα – σύνθεση)..... 77
Πίνακας 5.30: Σχετική σημαντικότητα μεταβλητών στο μοντέλο παλινδρόμησης ταχύτητα – σύνθεση..... 77
Πίνακας 5.31: Αποτελέσματα μοντέλων για διαφορετικά δεδομένα ελέγχου – μοντέλο GBM (ταχύτητα – σύνθεση)..... 81
Πίνακας 5.32: Αποτελέσματα μοντέλων για διαφορετικά δεδομένα ελέγχου – μοντέλο RF (ταχύτητα – σύνθεση)..... 86
Πίνακας 5.33: Σύγκριση δεικτών απόδοσης και σφάλματος στα μοντέλα ταχύτητας - σύνθεσης..... 89
Πίνακας 5.34: Σύγκριση σημαντικότητας μεταβλητών μεταξύ των μοντέλων ταχύτητας – σύνθεσης..... 90

Περιεχόμενα Διαγραμμάτων

Διάγραμμα 4-1: Ετήσια μεταβολή του αριθμού ΙΧ στην Αττική..... 32
Διάγραμμα 4-2: Ετήσια μεταβολή ωριαίου αριθμού οχημάτων στο σύνολο των οδικών τμημάτων το χρονικό διάστημα 8 – 10:00..... 36
Διάγραμμα 4-3: Ποσοστά οχημάτων στο σύνολο της κυκλοφορίας 2021 - 2025..... 37
Διάγραμμα 4-4: Μεταβολή φόρτου επιβατικών ΙΧ οχημάτων 2021 - 2025..... 38
Διάγραμμα 4-5: Μεταβολή φόρτου φορτηγών και λεωφορείων 2021 - 2025..... 38
Διάγραμμα 4-6: Σύνθεση κυκλοφορίας οδών με το μεγαλύτερο ποσοστό επιβατικών το 2025..... 39
Διάγραμμα 4-7: Σύνθεση κυκλοφορίας οδών με το χαμηλότερο ποσοστό ΙΧ το 2025..... 39
Διάγραμμα 4-8: Μεταβολή ροής πεζών 2021 – 2025..... 40
Διάγραμμα 4-9: Κυκλοφορία πεζών ανά οδό το 2025..... 40
Διάγραμμα 4-10: Ετήσια μεταβολή της συνολικής καθυστέρησης σε περιφερειακούς άξονες..... 45
Διάγραμμα 4-11: Ετήσια μεταβολή της συνολικής καθυστέρησης σε κεντρικούς άξονες..... 45
Διάγραμμα 4-12: Μεταβολή ποσοστού συμφορημένων ωρών στους περιφερειακούς άξονες..... 46
Διάγραμμα 4-13: Μεταβολή ποσοστού συμφορημένων ωρών στους οδικούς άξονες στο κέντρο..... 47
Διάγραμμα 4-14: Μεταβλητότητα χρόνου διαδρομής στη Βασιλέως Κωνσταντίνου στις 9:00 το 2025..... 50
Διάγραμμα 4-15: Μεταβλητότητα χρόνου διαδρομής στη Βασιλίσσης Σοφίας στις 14:00 το 2025..... 50
Διάγραμμα 4-16: Μεταβλητότητα χρόνου διαδρομής στη Σταδίου στις 14:00 το 2025..... 51

Διάγραμμα 4-17: Μεταβλητότητα χρόνου διαδρομής στην Αλεξάνδρας στις 15:00 το 2025 51
Διάγραμμα 4-18: Μεταβλητότητα χρόνου διαδρομής στη Βασιλέως Κωνσταντίνου στις 15:00 το 2025 52
Διάγραμμα 5-1: Τυποποιημένες τιμές μέσης τριμηνιαίας καθυστέρησης, εργαζόμενων και ΙΧ (2022 Q1 - 2025 Q4) 55
Διάγραμμα 5-2: Μέση τριμηνιαία καθυστέρηση και αριθμός τουριστών (2022 Q1 - 2025 Q4) 55
Διάγραμμα 5-3: Σχετική σημαντικότητα μεταβλητών στο μοντέλο ταχύτητα – καιρός 65
Διάγραμμα 5-4: Σχετική σημαντικότητα μεταβλητών στο μοντέλο HGBM ταχύτητα – καιρός 69
Διάγραμμα 5-5: Σύγκριση σημαντικότητας μεταβλητών μεταξύ των μοντέλων παλινδρόμησης και HGBM (ταχύτητα – καιρός) 71
Διάγραμμα 5-6: Ροή συγχώνευσης δεδομένων (ταχύτητα – σύνθεση κυκλοφορίας) 72
Διάγραμμα 5-7: Σχετική σημαντικότητα μεταβλητών στο μοντέλο παλινδρόμησης ταχύτητα – σύνθεση 78
Διάγραμμα 5-8: Σχετική σημαντικότητα μεταβλητών στο μοντέλο GBM ταχύτητα – σύνθεση 81
Διάγραμμα 5-9: Επίδραση ποσοστού ΙΧ στην ταχύτητα κυκλοφορίας (μοντέλο GBM) 83
Διάγραμμα 5-10: Επίδραση αντικατάστασης ΙΧ από λεωφορεία στην ταχύτητα κυκλοφορίας (GBM) 84
Διάγραμμα 5-11: Σχετική σημαντικότητα μεταβλητών στο μοντέλο RF ταχύτητα – σύνθεση 87
Διάγραμμα 5-12: Επίδραση ποσοστού ΙΧ στην ταχύτητα κυκλοφορίας (μοντέλο RF) 88
Διάγραμμα 5-13: Επίδραση αντικατάστασης ΙΧ από λεωφορεία στην ταχύτητα κυκλοφορίας (RF) 88
Διάγραμμα 5-14: Σύγκριση σημαντικότητας μεταβλητών μεταξύ των μοντέλων ταχύτητας – σύνθεσης 90

Περιεχόμενα Εικόνων

Εικόνα 1: Σχηματική αναπαράσταση της διαδικασίας bagging 27
Εικόνα 2: Σχηματική αναπαράσταση της διαδοχικής διαδικασίας Gradient Boosting 28
Εικόνα 3: Οδικά τμήματα καταγραφής της σύνθεσης κυκλοφορίας 34
Εικόνα 4: Κεντρικοί οδικοί άξονες ανάλυσης 43
Εικόνα 5: Περιφερειακοί οδικοί άξονες ανάλυσης 44
Εικόνα 6: Ακτίνα επιρροής μετεωρολογικών συνθηκών απο σταθμό ΕΑΑ 60

1. Εισαγωγή

1.1. Γενική ανασκόπηση

Η κυκλοφοριακή συμφόρηση αποτελεί μία από τις σημαντικότερες σύγχρονες προκλήσεις για τη βιωσιμότητα των αστικών κέντρων, με άμεσες και πολυεπίπεδες επιπτώσεις στην οικονομία, το περιβάλλον και την ποιότητα ζωής των πολιτών. Σύμφωνα με στοιχεία της Ευρωπαϊκής Ένωσης (2020), το οικονομικό κόστος από τις καθυστερήσεις στους δρόμους των χωρών της ΕΕ προσεγγίζει ετησίως τα 270 δισεκατομμύρια ευρώ, αποτυπώνοντας τη μεγάλη απώλεια παραγωγικότητας και τη σπατάλη ενεργειακών πόρων. Παράλληλα, η συνεχής λειτουργία των οχημάτων σε χαμηλές ταχύτητες επιτείνει την ατμοσφαιρική ρύπανση, συμβάλλοντας παράλληλα κατά 30% στην ανθρωπογενή αστική θερμότητα (Αυστριακός Οργανισμός Κινητικότητας και Μεταφορών – 2024).

Στο οδικό δίκτυο της Αθήνας, σύμφωνα με στοιχεία της Tom Tom, οι ετήσιες χαμένες ώρες σε συνθήκες αιχμής για έναν μέσο χρήστη ανήλθαν το 2025 στις 143 καταλαμβάνοντας την 17η χειρότερη θέση από τις συνολικά 254 των ευρωπαϊκών πόλεων. Η ανάλυση της παρούσας διπλωματικής ανέδειξε άνοδο των καθυστερήσεων κατά 72% στο σύνολο έξι περιφερειακών οδών και κατά 40% σε δώδεκα άξονες στο κέντρο της Αθήνας κατά την περίοδο 2022 έως 2025. Παράλληλα, διαπιστώθηκε πως από τις εννιά ώρες ανάλυσης (8:00 - 10:00 & 12:00 – 19:00), στις περιφερειακές οδούς, σχεδόν οι τρείς (31%) παρουσιάζουν ταχύτητες κυκλοφορίας κάτω του 50% της ταχύτητας ελεύθερης ροής ενώ στις κεντρικές ξεπερνάνε τις μισές (52%). Το κρίσιμο εύρημα όμως είναι η άνοδος αυτών από το 2022 όπου παρατηρείται αύξηση κατά 89% και 42% αντίστοιχα.

Η κατανόηση της κυκλοφοριακής συμφόρησης προϋποθέτει την ανάλυση των παραγόντων που γεννούν και διαμορφώνουν τη ζήτηση για μετακίνηση. Η ζήτηση αυτή είναι άρρηκτα συνδεδεμένη με τις χρήσεις γης και τον υφιστάμενο πολεοδομικό σχεδιασμό. Ο διαχωρισμός των οικιστικών ζωνών από τους χώρους εργασίας και η επέκταση των δραστηριοτήτων προς τα προάστια, αναγκάζουν τους πολίτες να διανύουν μεγαλύτερες αποστάσεις οι οποίες συνήθως γίνονται πιο εύκολα με τη χρήση των οχημάτων ιδιωτικής χρήσης (ΙΧ). Παράλληλα, η οικονομική ανάπτυξη φέρνει αύξηση του αριθμού των εργαζομένων και των επιχειρήσεων με αποτέλεσμα η ανάγκη για μετακίνηση να γίνεται όλο και πιο έντονη. Το ζήτημα λοιπόν δεν είναι πως θα περιοριστούν οι μετακινήσεις αλλά πώς θα καταστούν αυτές πιο βιώσιμες.

Η βιώσιμη αστική κινητικότητα (sustainable mobility) προκρίνεται διεθνώς ως η μοναδική βιώσιμη διέξοδος για την αποσυμφόρηση των σύγχρονων πόλεων. Η στρατηγική αυτή βασίζεται στην πολυτροπικότητα (multimodality), επιδιώκοντας τη συνέργεια ανάμεσα στα Μέσα Μαζικής Μεταφοράς (ΜΜΜ) και τις αναδυόμενες λύσεις μικροκινητικότητας (micromobility), όπως η χρήση ποδηλάτων, πατινιών αλλά και πεζή μετακίνηση. Χαρακτηριστικά παραδείγματα αποτελούν ευρωπαϊκές πόλεις όπως η Βιέννη, η Ποντεβέδρα και το Παρίσι. Στη Βιέννη, η εφαρμογή μιας ελκυστικής τιμολογιακής πολιτικής στα ΜΜΜ, σε

συνδυασμό με την επέκταση των υποδομών ενεργής μετακίνησης, περιόρισε το μερίδιο των μετακινήσεων των Ι.Χ. από το 40% στο 27% . Αντίστοιχα στην Ποντεβέδρα, μέσω της μαζικής πεζοδρόμησης και της απομάκρυνσης της στάθμευσης από το κέντρο, επιτεύχθηκε μείωση των εσωτερικών οδικών μετακινήσεων κατά 67%, αποδεικνύοντας πως η προσφορά συνάδει με τη ζήτηση. Τέλος, η περίπτωση του Παρισιού αναδεικνύει τη δυναμική της μικροκινητικότητας μέσω της ραγδαίας ανάπτυξης τέτοιων δικτύων τα οποία λειτούργησαν ως καταλύτης για τη μετακίνηση «πρώτου και τελευταίου μιλίου» μειώνοντας δραστικά τη χρήση ΙΧ οχημάτων στο κέντρο της πόλης.

Η μέτρηση της συμφόρησης αποτελεί σημαντικό εργαλείο στη διερεύνηση της βιωσιμότητας των μετακινήσεων στις πόλεις. Διάφορες έρευνες χρησιμοποιούν χωροχρονικά δεδομένα για την πρόβλεψη των χρόνων διαδρομής ή των ταχυτήτων κίνησης χωρίς όμως να εντάσσουν σε αυτές τον κρίσιμο παράγοντα της σύνθεσης της κυκλοφορίας στα αστικά κέντρα. Η παρούσα μελέτη συμπληρώνει αυτό το κενό και εξετάζει τη σημαντικότητα αυτής της παραμέτρου αξιοποιώντας δεδομένα από το 2021 έως το 2025.

1.2. Στόχος

Πρωταρχικός στόχος της παρούσας Διπλωματικής Εργασίας είναι η διερεύνηση των παραγόντων που επηρεάζουν τη συμφόρηση στο κέντρο της Αθήνας. Πιο συγκεκριμένα, διερευνώνται η σύνθεση κυκλοφορίας, οι κοινωνικοοικονομικοί παράγοντες οι οποίοι δημιουργούν ζήτηση για μετακινήσεις αλλά και οι μετεωρολογικοί που μπορεί να επηρεάσουν την συμπεριφορά των μετακινήσεων. Για τις αναλύσεις αυτές εφαρμόστηκαν τόσο οι παραδοσιακές μέθοδοι παλινδρόμησης όσο και προηγμένα μοντέλα μηχανικής μάθησης τα οποία έχουν την δυνατότητα να προβλέψουν μη-γραμμικές σχέσεις και αλληλεπιδράσεις μεταξύ των μεταβλητών. Με τον τρόπο αυτό δίνεται η δυνατότητα σύγκρισης μεταξύ αυτών των μοντέλων και εξέτασης της αποτελεσματικότητας τους.

Για τη διερεύνηση επιρροής των κοινωνικοοικονομικών παραγόντων στη συμφόρηση του οδικού δικτύου της Αθήνας, συλλέχθηκαν δεδομένα από την Ελληνική Στατιστική Αρχή (ΕΛΣΤΑΤ), το Υπουργείο Ανάπτυξης και Επενδύσεων και το Ινστιτούτο του Συνδέσμου Ελληνικών Τουριστικών Επιχειρήσεων – ΙΝΣΕΤΕ για το χρονικό διάστημα 2021 – 2026. Η συμφόρηση μετρήθηκε αξιοποιώντας τη βάση δεδομένων της Google η οποία αποτελείται από μετρήσεις μέσης ταχύτητας σε εξήντα δύο οδικά τμήματα στην Αθήνα τα έτη 2022 έως 2026. Για την ανάλυση επιλέχθηκαν σαράντα δύο οδικοί άξονες εντός και εκτός κέντρου για τα έτη 2022 έως 2025 όπου υπήρχαν πλήρη δεδομένα.

Καθώς τα μετεωρολογικά δεδομένα συλλέχθηκαν μόνο από έναν σταθμό, εκείνον του Εθνικού Αστεροσκοπείου Αθηνών – ΕΑΑ στο Θησείο, η ανάλυση επιρροής των καιρικών συνθηκών στην ταχύτητα κυκλοφορίας, περιορίστηκε σε οδούς εντός του κέντρου της Αθήνας. Για τον σκοπό αυτό επιλέχθηκαν 24 οδικά τμήματα τα οποία βρίσκονται σε ακτίνα τεσσάρων χιλιομέτρων από τον σταθμό καταγραφής. Αν και τα δεδομένα αποτελούνται από μετρήσεις το διάστημα 2021 έως το 2025, τα κυκλοφοριακά δεδομένα της Google περιόρισαν την ανάλυση στο χρονικό διάστημα 2022 – 2025.

Τα δεδομένα σύνθεσης της κυκλοφορίας αποτελούνται από χειροκίνητες καταγραφές σε δέκα έξι οδούς στο κέντρο της Αθήνας από το 2021 έως τον Ιούνιο του 2026. Η συχνότητα αυτών είναι δώδεκα καταγραφές ανά έτος όπου η κάθε μία αποτελείται από δύο χρονικά διαστήματα μέσα σε μία τυπική ημέρα του μήνα: 8:00 – 10:00 ή 8:00 – 11:00 και 11:00 – 15:00. Η ανάλυση περιορίστηκε στο χρονικό διάστημα καταγραφής των κυκλοφοριακών δεδομένων της Google: 2022 – 2025.

Εκτός από τα δεδομένα της Google, συλλέχθηκαν δεδομένα και από την πλατφόρμα Tom Tom τα οποία καλύπτουν μόνο το έτος 2025 και αξιοποιήθηκαν για τη διερεύνηση της μεταβλητότητας των χρόνων διαδρομής στο κέντρο της Αθήνας στο Κεφάλαιο 4.

1.3. Μεθοδολογία

Η μεθοδολογία που εφαρμόστηκε στη διπλωματική εργασία με σκοπό την απάντηση των διερευνητικών ερωτημάτων περιγράφεται παρακάτω.

Αρχικά έγινε οριστικοποίηση του θέματος της μελέτης και ο στόχος αυτής. Στη συνέχεια, μέσω ανασκόπησης της διεθνής βιβλιογραφίας καθορίστηκαν οι μεθοδολογίες που θα εφαρμοστούν και ταυτόχρονα το κενό που καλείται να συμπληρώσει αυτή η έρευνα.

Η συλλογή και επεξεργασία στοιχείων έδωσε μία αρχική εικόνα για τα επίπεδα της συμφόρησης στην Αθήνα αλλά και συμπεράσματα σχετικά με την σύνθεση της κυκλοφορίας στο κέντρο της Αθήνας και τη μεταβλητότητα των χρόνων διαδρομής. Επιπλέον, τα κοινωνικοοικονομικά δεδομένα ανέδειξαν την εξέλιξη της οικονομικής δραστηριότητας της πόλης το οποίο αποτέλεσε κίνητρο για περαιτέρω διερεύνηση.

Η εφαρμογή αναλυτικών μεθοδολογιών συνέβαλε στην βαθύτερη ανάλυση και κατανόηση της επιρροής των παραγόντων στη συμφόρηση. Για την εφαρμογή αυτών, χρησιμοποιήθηκε γλώσσα προγραμματισμού Python όπου αναπτύχθηκαν μοντέλα παλινδρόμησης και μηχανικής μάθησης αξιοποιώντας τη βιβλιοθήκη scikit-learn.

1.4. Δομή διπλωματικής εργασίας

Η δομή της παρούσας εργασίας έχει ως εξής:

Στο Κεφάλαιο 1 γίνεται η εισαγωγή στο γενικότερο πρόβλημα και το πλαίσιο που εντάσσεται αυτή η εργασία. Γίνεται αναφορά στις επιπτώσεις της κυκλοφοριακής συμφόρησης στα αστικά κέντρα, οι παράγοντες που την επιτείνουν αλλά και οι σύγχρονες προσεγγίσεις για την αντιμετώπιση του προβλήματος. Έπειτα, δίνεται μία πρώτη εικόνα για τον στόχο της εργασίας, τα δεδομένα που αξιοποιήθηκαν και οι μεθοδολογίες που εφαρμόστηκαν.

Στο Κεφάλαιο 2 παρουσιάζονται ορισμένες έρευνες σχετικά με το αντικείμενο της συμφόρησης των αστικών κέντρων και των παραγόντων που την επηρεάζουν. Επίσης παρατίθενται και οι μέθοδοι που χρησιμοποιούνται για την

πρόβλεψη των κυκλοφοριακών συνθηκών, απ' όπου προέκυψε και η ανάγκη κάλυψης του βιβλιογραφικού κενού.

Στο Κεφάλαιο 3 γίνεται αναφορά στο θεωρητικό υπόβαθρο των αναλύσεων που πραγματοποιήθηκαν. Εκεί υπάρχουν τα βασικά στατιστικά εργαλεία που χρησιμοποιήθηκαν, οι μέθοδοι παλινδρόμησης και οι προηγμένες μέθοδοι μηχανικής μάθησης που εφαρμόστηκαν.

Στο Κεφάλαιο 4 περιγράφονται τα δεδομένα που συλλέχθηκαν για τις αναλύσεις. Επιπλέον, πραγματοποιείται μια προκαταρκτική στατιστική ανάλυση η οποία δίνει χρήσιμες πληροφορίες και συμπεράσματα για τη μεταβολή των κοινωνικοοικονομικών και κυκλοφοριακών συνθηκών στην Αθήνα λειτουργώντας ώς προθάλαμος για την περαιτέρω ανάλυση μέσω συγκεκριμένων μεθοδολογιών.

Οι μεθοδολογίες αυτές εφαρμόζονται στο Κεφάλαιο 5 της διπλωματικής εργασίας, το οποίο αποτελεί και την κύρια ενότητά της. Αναλύονται διεξοδικά οι μεταβλητές των μοντέλων και ο τρόπος εφαρμογής των μεθοδολογιών. Για τη διερεύνηση επιρροής των κοινωνικοοικονομικών παραγόντων εφαρμόστηκαν αναλύσεις συσχέτισης και μοντέλα παλινδρόμησης. Για τα ερευνητικά ερωτήματα της επιρροής των καιρικών συνθηκών και της σύνθεσης της κυκλοφορίας, αναπτύσσονται δύο προσεγγίσεις: η πρώτη αφορά παραδοσιακά μοντέλα παλινδρόμησης τα οποία είναι απλά στην εφαρμογή και προσφέρουν άμεσα και ερμηνεύσιμα αποτελέσματα και η δεύτερη αφορά πιο σύνθετες μεθόδους μηχανικής μάθησης οι οποίες έχουν τη δυνατότητα να συλλάβουν μη-γραμμικές σχέσεις και αλληλεπιδράσεις μεταξύ των μεταβλητών. Στο τέλος συγκρίνονται μεταξύ τους και καθορίζεται η καταλληλόλητα τους για τον στόχο της διπλωματικής.

Στο Κεφάλαιο 6 παρουσιάζονται συνοπτικά τα αποτελέσματα των αναλύσεων του προηγούμενου κεφαλαίου. Επιπλέον, αναφέρονται οι περιορισμοί της έρευνας και οι προτάσεις για μια πιο ολοκληρωμένη ανάλυση και περεταίρω διερεύνηση.

Το Κεφάλαιο 7 περιλαμβάνει τις βιβλιογραφικές αναφορές που αξιοποιήθηκαν για την εκπόνηση της διπλωματικής εργασίας.

2. Βιβλιογραφική Ανασκόπηση

2.1. Εισαγωγή

Στο κεφάλαιο της βιβλιογραφικής ανασκόπησης παρουσιάζονται συναφείς έρευνες και μεθοδολογίες σχετικά με το αντικείμενο της διπλωματικής εργασίας. Αρχικά δόθηκε ο ορισμός της κυκλοφοριακής συμφόρησης και διερευνήθηκαν οι παράγοντες που προκαλούν αυξημένη κυκλοφορία οχημάτων στα αστικά κέντρα. Στη συνέχεια, εξετάσθηκε ο τρόπος με τον οποίο γίνεται η επιλογή μέσου από τους μετακινούμενους και δόθηκε έμφαση στην αποτελεσματικότητα των Μέσων Μαζικής Μεταφοράς στην αποσυμφόρηση των αστικών οδικών δικτύων. Για την πρόβλεψη των χρόνων διαδρομής, έγινε εμβάθυνση στις μεθόδους μηχανικής μάθησης (machine learning) όντας αξιόπιστες και προσφέροντας μεγάλη ακρίβεια.

Μέσα από τη σύνθεση της βιβλιογραφίας, καθορίστηκε ο στόχος της παρούσας διπλωματικής εργασίας, ο οποίος εστιάζει στην κάλυψη σημαντικών ερευνητικών κενών. Ενώ οι περισσότερες έρευνες στην πρόβλεψη των χρόνων διαδρομής με μεθόδους μηχανικής μάθησης περιορίζονται σε βραχυχρόνιες προβλέψεις σε αυτοκινητόδρομους, χωρίς να λαμβάνουν υπόψη την επιλογή μέσου, η παρούσα εργασία εξετάζει την εφαρμογή τους σε αστικό περιβάλλον. Συγκεκριμένα, διερευνάται ο βαθμός στον οποίο η επιλογή μεταφορικού μέσου επηρεάζει τους χρόνους διαδρομής και τη συμφόρηση στο κέντρο της Αθήνας, ενώ παράλληλα εξετάζεται η επίδραση των καιρικών συνθηκών στην κυκλοφορία. Επιπλέον, η έρευνα επεκτείνεται στη διερεύνηση της συσχέτισης της κυκλοφοριακής συμφόρησης με κοινωνικοοικονομικούς παράγοντες της πόλης. Η διαφοροποίηση της παρούσας εργασίας έγκειται επίσης στο γεγονός ότι η πρόβλεψη μέσω Μηχανικής Μάθησης δεν αφορά βραχυπρόθεσμα διαστήματα, αλλά βασίζεται στην ανάλυση μεγάλων βάσεων δεδομένων που καλύπτουν ιστορικά στοιχεία ετών, προσφέροντας μια μακροπρόθεσμη θεώρηση των αστικών μετακινήσεων.

2.2. Ορισμός και ποσοτικοποίηση της κυκλοφοριακής συμφόρησης

Η κυκλοφοριακή συμφόρηση ορίζεται ως η κατάσταση κατά την οποία η ζήτηση μετακινήσεων υπερβαίνει τη διαθέσιμη χωρητικότητα του οδικού δικτύου, με αποτέλεσμα τη μείωση της ταχύτητας των οχημάτων και την πρόκληση σημαντικών καθυστερήσεων. Η κυκλοφοριακή συμφόρηση σπαταλά χρόνο και ενέργεια, ενώ προκαλεί και ρύπανση. Υπάρχουν γενικά δύο παράγοντες που την επηρεάζουν: (α) παράγοντες μικρο-επιπέδου και (β) παράγοντες μακρο-επιπέδου, οι οποίοι σχετίζονται με τη συνολική ζήτηση για οδική χρήση. Οι παράγοντες μικρο-επιπέδου είναι, για παράδειγμα, η ταυτόχρονη επιθυμία πολλών ανθρώπων να μετακινηθούν, ενώ από την άλλη πλευρά, οι παράγοντες μακρο-επιπέδου είναι κυρίως κοινωνικοοικονομικοί: τα πρότυπα χρήσης γης, οι τάσεις ιδιοκτησίας αυτοκινήτων, η οικονομική δραστηριότητα κ.λπ. Για την επιλογή κατάλληλων μέτρων μείωσης της συμφόρησης, κρίνεται αναγκαίος ο εντοπισμός αυτής μέσω ορισμένων κυκλοφοριακών δεικτών που προτείνονται στη διεθνή βιβλιογραφία.

Η έρευνα των Mohan Rao & Ramachandra Rao, 2012 αναλύει αυτούς τους δείκτες σε τρείς κατηγορίες:

1) Δείκτες ταχύτητας: Μέση Ταχύτητα Μετακίνησης, Δείκτης Μείωσης ταχύτητας

2) Δείκτες Χρόνου Μετακίνησης & Καθυστέρησης: Δείκτης Χρόνου Μετακίνησης, Δείκτης Ρυθμού Μετακίνησης κ.α.

3) Δείκτες Φόρτου & Επιπέδου Εξυπηρέτησης: Επίπεδο Εξυπηρέτησης, Λόγος Φόρτου προς Χωρητικότητα κ.α.

Συνολικά καταλήγει στο συμπέρασμα πως οι δείκτες ταχύτητας είναι πιο κατάλληλοι, διότι συνδέονται άμεσα με τις πραγματικές επιπτώσεις της συμφόρησης (απώλεια χρόνου, αυξημένο λειτουργικό κόστος, κατανάλωση καυσίμου και εκπομπές ρύπων). Ο περιορισμός που επιβάλλουν όμως είναι η ανάγκη προσδιορισμού της ταχύτητας ελεύθερης ροής ως σημείο αναφοράς, κάτι το οποίο διαφέρει για κάθε οδικό τμήμα.

Στην έρευνα του Aftabuzzaman, 2007 , ο συγγραφέας αναλύει επιπλέον δείκτες συμφόρησης οι οποίοι παρουσιάζονται στον Πίνακα 2.1 και τους αξιολογεί με βάση έξι κριτήρια:

1) Απλότητα
2) Δυνατότητα περιγραφή μεγέθους συμφόρησης
3) Δυνατότητα σύγκρισης μεταξύ διαφορετικών πόλεων
4) Διαθέτει συνεχές εύρος τιμών
5) Περιλαμβάνει χρόνο διαδρομής
6) Συσχέτιση με δείκτες αποσυμφόρησης λόγω ΜΜΜ

Κατηγορία	Δείκτης	Περιγραφή / Ορισμός	Πλεονεκτήματα	Περιορισμοί
1. Βασικά Μέτρα (Basic Measures)	Συνολική Καθυστέρηση (Total Delay)	Πρόσθετος χρόνος σε σχέση με την ελεύθερη/αποδεκτή ροή	Κατάλληλος για ανάλυση κόστους-οφέλους και διάρκεια συμφόρησης	Δύσκολες συγκρίσεις μεταξύ περιοχών
	Συμφορημένη Μετακίνηση (Congested Travel)	Φόρτος οχημάτων/επιβατών στα συμφορημένα τμήματα	Δείχνει τη χωρική έκταση	Δεν δείχνει ένταση ούτε χρόνο
	Μήκος Συμφορημένου Δικτύου (Congested Roadway Length)	Συνολικό μήκος δικτύου σε συμφόρηση	Απλός και κατανοητός	Δεν περιλαμβάνει χρόνο
	Ρυθμός Μετακίνησης (Travel Rate)	Χρόνος ανά μονάδα απόστασης (αντίστροφο ταχύτητας)	Βάση για τους δείκτες καθυστέρησης	Από μόνος του δεν μετρά τη συμφόρηση

2. Δείκτες Αναλογίας (Ratio Measures)	Ρυθμός Καθυστέρησης (Delay Rate)	Πραγματικός − αποδεκτός ρυθμός μετακίνησης	Ιεράρχηση βελτιώσεων	Μόνο συγκεκριμένοι τύποι οδών
	Σχετικός Ρυθμός Καθυστέρησης (Relative Delay Rate)	Ρυθμός καθυστέρησης / αποδεκτός ρυθμός	Σύγκριση υποδομών και μέσων	Δύσκολος σε ευρείες περιοχές, δεν συγκρίνει πόλεις
	Λόγος Καθυστέρησης (Delay Ratio)	Ρυθμός καθυστέρησης / πραγματικός ρυθμός	Σύγκριση υποδομών με διαφορετικά χαρακτηριστικά	Δύσκολος σε ευρείες περιοχές, δεν συγκρίνει πόλεις
3. Επιπέδου Εξυπηρέτησης (LOS)	Επίπεδο Εξυπηρέτησης (Level of Service - LOS)	6 κλάσεις (A–F) βάσει ταχύτητας, πυκνότητας ή v/c	Κατανοητό από μη τεχνικό κοινό	Μη συνεχής κλίμακα, παραπλανητικό στα όρια κλάσεων
4. Σύνθετοι Δείκτες (Indices)	Δείκτης Συμφόρησης (Congestion Index)	Καθυστέρηση τμήματος / αποδεκτός χρόνος	Συνεχής κλίμακα, ενσωματώνει χρόνο	Μόνο συγκεκριμένα τμήματα/διαδρομές
	Δείκτης Ρυθμού Μετακίνησης (Travel Rate Index - TRI)	Σύγκριση συνθηκών αιχμής με ελεύθερης ροής	Σύγκριση μητροπολιτικών περιοχών	Κυρίως αυτοκινητόδρομοι και αρτηρίες
	Δείκτης Συμφόρησης Οδοστρώματος (Roadway Congestion Index - RCI)	VMT ανά μίλι λωρίδας	Ευρεία χρήση σε συγκρίσεις πόλεων	Εκφράζει πυκνότητα, όχι πραγματική συμφόρηση
	Δείκτης Κινητικότητας Διαδρόμου (Corridor Mobility Index - CMI)	Ταχύτητα × φόρτος επιβατών σε διάδρομο	Πολυτροπική αξιολόγηση διαδρόμου	Μόνο σε επίπεδο διαδρόμου
	Δείκτης Διάρκειας Λωρίδας-Μιλίου (Lane Mile Duration Index - LMDI)	Συμφορημένα λωρίδα-μίλια × διάρκεια	Αποτυπώνει έκταση και διάρκεια	Δύσκολες συγκρίσεις μεταξύ περιοχών
	Λοιποί Σύνθετοι Δείκτες *(CBI, CSI)*	Δείκτες βάρους και σοβαρότητας συμφόρησης	Συνεχής κλίμακα, σύγκριση πόλεων	Υπολογιστική πολυπλοκότητα

5. Νέος προτεινόμενος Δείκτης	Δείκτης Ελάφρυνσης Συμφόρησης (Congestion Relief Index - CRI)	Πολυτροπικός δείκτης μείωσης συμφόρησης χάρη στα ΜΜΜ	Καλύπτει το κενό των υπαρχόντων δεικτών, ενσωματώνοντας την επίδραση των ΜΜΜ	-

Πίνακας 2.1: Δείκτες ποσοτικοποίησης της συμφόρησης (Aftabuzzaman, 2007)

Η έρευνα καταλήγει στο συμπέρασμα πως κανένας δείκτης μέτρησης της συμφόρησης δεν πληροί και τα έξι παραπάνω κριτήρια. Οι βασικοί δείκτες (1), οι δείκτες αναλογίας (2) και το επίπεδο εξυπηρέτησης (3), είναι απλοί και κατανοητοί αλλά οι περισσότεροι από αυτούς, εκτός από τον δείκτη συνολικής καθυστέρησης, δεν πληρούν περισσότερα από τρία κριτήρια. Οι περισσότεροι από τους σύνθετους δείκτες (4), πληρούν τέσσερα στα έξι κριτήρια αλλά είναι πιο περίπλοκοι από άλλους. Συνολικά κρίνει τον δείκτη συνολικής καθυστέρησης (total delay) τον πιο κατάλληλο για την ποσοτικοποίηση της συμφόρησης όντας απλός και ερμηνεύσιμος. Παράλληλα επισημαίνεται πως κανένας από τους δείκτες δεν περιλαμβάνει την επίδραση των ΜΜΜ και προτείνεται τη χρήση ενός νέου: Δείκτης Ελάφρυνσης Συμφόρησης (Congestion Relief Index - CRI) ο οποίος καλύπτει αυτό το κενό.

2.3. Επίδραση κοινωνικοοικονομικών παραγόντων στην συμφόρηση του αστικού δικτύου

Η πληθυσμιακή πυκνότητα, η έντονη οικονομική δραστηριότητα και η αυξημένη χρήση ιδιωτικών οχημάτων είναι κοινωνικοοικονομικοί παράγοντες μακρο-επιπέδου οι οποίοι δημιουργούν επιπλέον ζήτηση για μετακινήσεις όπως αναλύθηκε στην Ενότητα 2.2. Οι παρακάτω έρευνες μελέτησαν αναλυτικά την επίδραση τέτοιων παραγόντων στην κυκλοφορία των οχημάτων στον αστικό ιστό χρησιμοποιώντας ως εξαρτημένη μεταβλητή δείκτες συμφόρησης. Οι Mokhlesur Rahman et al., n.d. ανέλυσαν 100 αστικές περιοχές των ΗΠΑ χρησιμοποιώντας μοντελοποίηση δομικών εξισώσεων (SEM), εξετάζοντας 25 δομικές και κοινωνικοοικονομικές παραμέτρους και αναλύοντας τη κυκλοφοριακή συμφόρηση σε επίπεδο δικτύου. Στο ίδιο πλαίσιο, η έρευνα των Dadashova et al., 2021 διερεύνησε τα αίτια της συμφόρησης σε 51 αμερικανικές μητροπόλεις μέσω 32 μεταβλητών, που περιλάμβαναν μακροοικονομικά, πληθυσμιακά και μετεωρολογικά δεδομένα. Παράλληλα οι Koźlak & Wach, 2018, κατέδειξαν μέσω αναλύσεων συσχέτισης και παλινδρόμησης τον κρίσιμο ρόλο της οικονομικής ανάπτυξης και της ιδιοκτησίας αυτοκινήτων στις καθυστερήσεις των μετακινήσεων.

Έρευνα	Μέθοδοι Ανάλυσης	Δείκτες συμφόρησης	Παράγοντες αύξησης της κυκλοφορίας
Mokhlesur Rahman et al.	Μοντελοποίηση δομικών εξισώσεων (SEM)	Αναλογία χρόνου ταξιδιού σε ώρες αιχμής έναντι ωρών μη αιχμής (ένταση) Ποσοστό λωρίδων με συμφόρηση (έκταση) Διάρκεια των περιόδων αιχμής σε ώρες (διάρκεια)	Μέγεθος πληθυσμού Εισόδημα & συγκέντρωση απασχόλησης
Dadashova et al., 2021	Πολυμεταβλητή Ανάλυση Χρονοσειρών Διανυσματικά Αυτοπαλίνδρομα Μοντέλα με Εξωγενείς Μεταβλητές (VAR-X):	Δείκτης Χρόνου Ταξιδιού (Travel Time Index - TTI) Δείκτης Χρόνου Προγραμματισμού (Planning Time Index - PTI) Ώρες Συμφόρησης (Congested Hours - CHR)	Αριθμός Απασχολούμενων Αριθμός Οικοδομικών Αδειών Δείκτης Οικονομικών Συνθηκών Ποσοστό Κενών Κατοικιών προς Ενοικίαση Δείκτης Τιμών Καυσίμων
Koźlak & Wach, 2018	Ανάλυση συσχέτισης και πολλαπλή γραμμική παλινδρόμηση	Ετήσια αύξηση του χρόνου μετακίνησης σε σύγκριση με συνθήκες ελεύθερης ροής	Ποσοστό απασχόλησης Μέσος μηνιαίος μισθός Αριθμός επιχειρήσεων Αριθμός επιβατικών οχημάτων Έκταση της πόλης

Πίνακας 2.2: Έρευνες για την επίδραση κοινωνικοοικονομικών παραγόντων στην κυκλοφορία

2.4. Επίδραση μετεωρολογικών συνθηκών στην κυκλοφορία

Οι μετεωρολογικές συνθήκες δύναται να επηρεάσουν την κυκλοφορία με δύο τρόπους. Ο πρώτος αφορά την επίδραση των καιρικών συνθηκών στην ικανότητα ασφαλούς οδήγησης (π.χ. έλλειψη ορατότητας) και ο δεύτερος αφορά την επιλογή μέσου μετακίνησης (Bi et al., 2022). Έρευνες σε αυτοκινητόδρομους του Οντάριο στον Καναδά (And & Hall, 1994) και στο οδικό δίκτυο του Λονδίνου (Tsapakis et al., 2013) διαπίστωσαν πως η ένταση των φαινομένων βροχόπτωσης και χιονόπτωσης προκαλούν σημαντική μείωση στις ταχύτητες ελεύθερης ροής και στη μέγιστη χωρητικότητα των δρόμων. Αντίθετα, τα ήπια φαινόμενα έχουν περιορισμένο αντίκτυπο, ο οποίος συχνά κρίνεται πρακτικά αμελητέος.

Έρευνα	Μέθοδοι Ανάλυσης	Αποτελέσματα
And & Hall, n.d.	Ανάλυση Παλινδρόμησης για να επιλεγούν τα βέλτιστα μοντέλα για τις συνθήκες ελεύθερης ροής Πολλαπλή Παλινδρόμηση με ψευδομεταβλητές για τη σύγκριση των κυκλοφοριακών συνθηκών μεταξύ διαφορετικών καιρικών φαινομένων	Ελαφριά βροχή ή χιόνι: πτώση ταχύτητα ελεύθερης ροής κατά 3 χλμ/ώρα Έντονη βροχή: πτώση ταχύτητας ελεύθερης ροής κατά 5 έως 10 χλμ/ώρα Έντονο χιόνι: σημαντική πτώση των ταχυτήτων ελεύθερης ροής από 35 έως και 50 χλμ/ώρα
Tsapakis et al., 2013	Ποσοστιαία διαφορά του συνολικού χρόνου διαδρομής συγκρίνοντας τις "βασικές συνθήκες" (στεγνό οδόστρωμα, θερμοκρασία >10°C, απουσία βροχής ή χιονιού) με συνθήκες διαφορετικών εντάσεων βροχόπτωσης, χιονόπτωσης και χαμηλών θερμοκρασιών	Ελαφριά βροχή: Αύξηση χρόνου διαδρομής έως και 2.1% Μέτρια βροχή: Αύξηση έως και 3.8% Ισχυρή βροχή: Αύξηση έως και 6% Ελαφρύ χιόνι: Καθυστερήσεις της τάξης του 5.5% έως 7.6% Πυκνό Χιόνι: Καθυστερήσεις από 7.4% έως 11.4% Θερμοκρασία: Αμελητέα επίδραση Μεγαλύτερη επιρροή των καιρικών συνθηκών στις προαστιακές οδούς παρά στις αστικές

Πίνακας 2.3: Έρευνες για την επίδραση των μετεωρολογικών συνθηκών στη ταχύτητα

Η επιλογή μέσου που οφείλεται στις καιρικές συνθήκες διαφοροποιείται από χώρα σε χώρα ανάλογα με το είδος της μετακίνησης και τις προτιμήσεις των μετακινούμενων. Στο Βέλγιο, έρευνα διαπίστωσε πως τα έντονα καιρικά φαινόμενα μειώνουν τον αριθμό οχημάτων στο οδικό δίκτυο καθώς, σύμφωνα με τους συγγραφείς, επηρεάζονται περισσότερο οι διαδρομές αναψυχής παρά οι μετακινήσεις προς την εργασία (Cools et al., 2010). Σε χώρες όπου γίνεται χρήση ποδηλάτου και περπατήματος προς την εργασία, οι θερμές καιρικές συνθήκες ευνοούν αυτού του είδους τις μετακινήσεις (Galich & Nieland, 2023) ενώ οι δυσμενείς καιρικές συνθήκες οδηγούν σε αυξημένη χρήση αυτοκινήτων σύμφωνα με τις έρευνες των Sabir et al., 2007 και Faber et al., 2022 στην Ολλανδία. Αντιθέτως στο Βερολίνο, όταν επικρατεί βροχόπτωση και χαμηλές θερμοκρασίες, οι πολίτες προτιμούν τη χρήση των δημόσιων συγκοινωνιών παρά του αυτοκινήτου σύμφωνα με έρευνα των Nissen et al., 2020.

2.5. Επιλογή μέσου μετακίνησης

Η επιλογή μέσου μετακίνησης όμως δεν εξαρτάται μόνο από τις μετεωρολογικές συνθήκες. Οι κύριοι παράγοντες που την επηρεάζουν είναι: τα χαρακτηριστικά

του μετακινούμενου (διαθεσιμότητα/κατοχή αυτοκινήτου, διαθεσιμότητα χώρου στάθμευσης, δομή νοικοκυριού, εισόδημα), τα χαρακτηριστικά της μετακίνησης (σκοπός της μετακίνησης, ώρα της ημέρας, ημέρα της εβδομάδας) και τα χαρακτηριστικά του μεταφορικού συστήματος τα οποία διακρίνονται σε δύο κατηγορίες. Η πρώτη κατηγορία είναι ποσοτικοί παράγοντες και αφορούν το χρόνο διαδρομής, το κόστος, διαθεσιμότητα θέσεων και κόστος στάθμευσης στον προορισμό και την αξιοπιστία του μέσου. Η δεύτερη αφορά ποιοτικούς παράγοντες όπως η άνεση, η ασφάλεια και οι δυνατότητες που δίνονται στον μετακινούμενο κατά τη χρήση ενός μέσου (π.χ. χρήση κινητού τηλεφώνου) (Ortúzar & Willumsen, 2011).

Ωστόσο, η αξιοπιστία και η άνεση που προσφέρει ένα μέσο μετακίνησης μεταβάλλονται σημαντικά μέσα στην ημέρα επηρεάζοντας τις προτιμήσεις των μετακινούμενων στην επιλογή μέσου. Αυτό αναδεικνύει η μελέτη των Ding & Zhang, 2016 η οποία αναλύει τις δυναμικές μεταβολές στις αποφάσεις επιλογής μεταφορικού μέσου στην πόλη Ναντζίνγκ της Κίνας εστιάζοντας στις διακυμάνσεις που συμβαίνουν εντός της ίδιας ημέρας. Ο συνωστισμός στο οδικό δίκτυο και στα μέσα Μαζικής Μεταφοράς μειώνουν τη μέση ταχύτητα μετακίνησης και την άνεση με αποτέλεσμα να επηρεάζεται άμεσα η χρησιμότητα του κάθε μέσου.

Ένα ακόμα χαρακτηριστικό του μεταφορικού συστήματος που επηρεάζει άμεσα την επιλογή μέσου είναι η απόσταση της μετακίνησης. Η έρευνα των He et al., 2022 στη πόλη Γκουιγιάνγκ της Κίνας δείχνει πως για αποστάσεις άνω των 12.5 χιλιομέτρων, η πιθανότητα επιλογής αυτοκινήτου ξεπερνά το 50% ενώ αντίστοιχο ποσοστό για χρήση δημόσιας συγκοινωνίας παρατηρείται για εύρος 3.5 – 6 χιλιομέτρων. Κοινό εύρημα των παραπάνω ερευνών είναι η επιρροή των χαρακτηριστικών του μετακινούμενου στην επιλογή μέσου. Κοινωνικοοικονομικά χαρακτηριστικά όπως το υψηλό εισόδημα και η ιδιοκτησία αυτοκινήτου, ενισχύουν την προτίμηση προς την οδήγηση ενώ παρατηρείται πως η χρήση δημοσίων συγκοινωνιών προτιμάται από τους ηλικιωμένους και ομάδες χαμηλού και μεσαίου εισοδήματος που δεν έχουν πρόσβαση σε ιδιωτικό όχημα.

2.6. Επίδραση των Μέσων Μαζικής Μεταφοράς στη κυκλοφορία

Αρκετές μελέτες αναδεικνύουν τον καθοριστικό ρόλο των λεωφορείων στην αποσυμφόρηση του αστικού ιστού, επισημαίνοντας ότι η ενίσχυση της χρήσης τους αποτελεί τη βέλτιστη λύση για τη μείωση της χρήσης ιδιωτικών αυτοκινήτων και τη βελτίωση της ταχύτητας του δικτύου. Μέσω προηγμένων μοντέλων προσομοίωσης τεκμηριώνεται ότι τα Μέσα Μαζικής Μεταφοράς προσφέρουν πολυδιάστατα οφέλη - περιβαλλοντικά, οικονομικά και κοινωνικά - παρά τις όποιες τοπικές καθυστερήσεις προκαλούν (Nguyen-Phuoc et al., 2020), (Sultana et al., 2022). Οι ερευνητές συγκλίνουν στην ανάγκη για στρατηγικό σχεδιασμό και επενδύσεις σε υποδομές, όπως οι αποκλειστικές λωρίδες και συστήματα παρακολούθησης σε πραγματικό χρόνο, προκειμένου να διασφαλιστεί η αξιοπιστία και η άνεση των μετακινήσεων, καθιστώντας τις

συγκοινωνίες μια ελκυστική και βιώσιμη εναλλακτική για τους πολίτες (Ery Sugito & Diana Binti Mohamad, 2025).

Ωστόσο, αν και οι αποκλειστικές λωρίδες κυκλοφορίας λεωφορείων μπορεί να μειώνουν τις καθυστερήσεις αυτών, το συνολικό αντίκτυπο στη κυκλοφορία μπορεί να διαφέρει. Οι Guangchuan Yang et al., 2020 μελέτησαν τα αποτελέσματα της εφαρμογής αυτής μέσω ενός μοντέλου μικροπροσομοίωσης συναρτήσει του ποσοστού χρήσης λεωφορείων και της επιβατικής ζήτησης και με σταθερή πληρότητα 60 επιβάτες ανά λεωφορείο και 3 ανά αυτοκίνητο. Το οδικό δίκτυο του μοντέλου αποτελείται από μια αστική οδική αρτηρία τριών χιλιομέτρων, τριών λωρίδων ανά κατεύθυνση, με τρείς ενδιάμεσους σηματοδοτούμενους κόμβους και δύο στάσεις λεωφορείων. Τα αποτελέσματα φανερώνουν την άμεση συσχέτιση της μεταφορικής ζήτησης (επιβάτες/ώρα) και του ποσοστού χρήσης των λεωφορείων με τις καθυστερήσεις των οχημάτων. Αν και τα λεωφορεία κινούνται πάντα χωρίς καθυστερήσεις, απαιτείται ποσοστό χρήσης αυτών άνω του 80% στο εύρος ζήτησης 10 – 14,000 επ./ώρα για να μην επηρεάζονται οι υπόλοιποι χρήστες. Αντίστοιχα, για ζήτηση 6 – 10,000 επ./ώρα απαιτείται ποσοστό πλήρωσης λεωφορείων 60% για την διατήρηση ομαλής ροής στις μειωμένες λωρίδες κίνησης οχημάτων. Η υπεροχή των λεωφορείων ως προς τη μεταφορική ικανότητα της οδού τεκμηριώνεται και σε τεχνικούς οδηγούς σχεδιασμού, σύμφωνα με τους οποίους μία λωρίδα κυκλοφορίας IX μετακινεί περίπου 600 έως 1.600 άτομα ανά ώρα, ενώ μία αποκλειστική λωρίδα λεωφορείου μπορεί να μεταφέρει έως 8.000 άτομα ανά ώρα (NACTO, 2016).

2.7. Πρόβλεψη των χρόνων διαδρομής

2.7.1. Γενικά

Η πρόβλεψη των χρόνων διαδρομής αποτελεί κρίσιμο παράγοντα για την ανάλυση του επιπέδου της κυκλοφοριακής συμφόρησης σε βραχυπρόθεσμο ή μακροπρόθεσμο χρονικό ορίζοντα. Παράλληλα, αποτελεί ένα σύνθετο, μη-γραμμικό χωροχρονικό πρόβλημα το οποίο εξαρτάται από τις εκάστοτε συνθήκες. Για την αντιμετώπιση αυτής της πολυπλοκότητας, απαιτείται η ανάπτυξη εξελιγμένων κυκλοφοριακών ή μη γραμμικών μοντέλων τα οποία βασίζονται σε δεδομένα και έχουν τη δυνατότητα να εκπαιδευτούν από αυτά (Van Lint, 2004).

Οι μέθοδοι που χρησιμοποιούνται για την πρόβλεψη των χρόνων διαδρομής ταξινομούνται ως μέθοδοι βασισμένες σε μοντέλα και μέθοδοι βασισμένες σε δεδομένα και διαφέρουν ως προς το σύνολο δεδομένων που χρησιμοποιούν, το εύρος πρόβλεψης και την ακρίβεια (Bai et al., 2018). Οι μέθοδοι που βασίζονται σε μοντέλα στοχεύουν στη βελτίωση του μοντέλου ή του αλγορίθμου μηχανικής μάθησης χωρίς να προσαρμόζουν τα δεδομένα. Ως αποτέλεσμα, επιτυγχάνεται σταθερή πολυπλοκότητα και αποτελεσματικές προβλέψεις. Οι μέθοδοι που βασίζονται σε δεδομένα, αντίθετα, εκμεταλλεύονται μεγάλο όγκο δεδομένων κάνοντας προ επεξεργασία αυτών για την εξαγωγή των προβλέψεων (Lin et al., 2025).

2.7.2. Μέθοδοι βασισμένοι σε μοντέλα

Οι μέθοδοι βασισμένοι σε μοντέλα, προβλέπουν τους χρόνους διαδρομής χρησιμοποιώντας κυκλοφοριακές παραμέτρους όπως η πυκνότητα, οι φόρτοι και οι ταχύτητες. Δύο κοινές τέτοιες μέθοδοι είναι η θεωρία ουρών (Queuing Theory) και Μοντέλο Μετάδοσης Κυψελών (Cell Transmission Model – CTM) (Qiu & Fan, 2021), (Bai et al., 2018). Το πλεονέκτημα αυτών είναι ότι επιτρέπουν την συμπερίληψη μέτρων διαχείρισης της κυκλοφορίας (π.χ. σηματοδότηση, καταγραφή κυκλοφορίας) και παρέχουν πλήρη πληροφόρηση για τη τοποθεσία και τα αίτια πιθανής συμφόρησης στο οδικό δίκτυο. Επιπλέον, μπορούν να εφαρμοστούν ακόμα και σε διαδρομές όπου δεν υπάρχει εξοπλισμός καταγραφής της κυκλοφορίας. Το κύριο μειονέκτημα των μεθόδων βασισμένων σε μοντέλα είναι η υπολογιστική τους πολυπλοκότητα, ο βαθμός εμπειρογνωμοσύνης που απαιτείται για τον σχεδιασμό, τη συντήρηση τους και το γεγονός ότι απαιτούν προβλέψεις ζήτησης και προσφοράς ως δεδομένα εισόδου στο μοντέλο. Για «εκτός-σύνδεσης» αναλύσεις και εκτίμηση σεναρίων είναι ιδανικά, αλλά σε πραγματικό χρόνο, η ποιότητα αυτών επηρεάζεται σημαντικά από τη ποιότητα των δεδομένων εισόδου και των οριακών συνθηκών (Van Lint, 2004).

2.7.3. Μέθοδοι βασισμένοι σε δεδομένα

Οι μέθοδοι βασισμένοι σε δεδομένα συνδέουν τις παρατηρήσεις των χρόνων διαδρομής ή τις κυκλοφοριακές συνθήκες με τρέχοντα ή προηγούμενα κυκλοφοριακά δεδομένα χωρίς να συνεκτιμάται η επιρροή των συνθηκών όπως γίνεται σε αυτές που βασίζονται σε μοντέλα (Van Lint, 2004). Διακρίνονται σε Παραμετρικές και μη-Παραμετρικές μεθόδους και έχουν στόχο τη δημιουργία συναρτήσεων μεταξύ δεδομένων μεγάλης κλίμακας και της εξαρτημένης μεταβλητής. Οι Παραμετρικές μέθοδοι υποθέτουν πως τα δεδομένα ακολουθούν μια συγκεκριμένη στατιστική κατανομή δομώντας το προγνωστικό μοντέλο εξαρχής και τέτοιες είναι οι εξής: Γραμμική Παλινδρόμηση (Linear Regression), Αυτοπαλινδρομούμενο Ολοκληρωμένο Κινητό Μέσο (Autoregressive Integrated Moving Average – ARIMA), Φιλτράρισμα Kalman (Kalman Filtering). Αντιθέτως, οι μη-Παραμετρικές μέθοδοι χρησιμοποιούν υπάρχοντα δεδομένα με σκοπό να εκπαιδεύσουν το μοντέλο και ύστερα να καθοριστεί ο αριθμός των παραμέτρων του. Ορισμένα τέτοια μοντέλα είναι τα: Παλινδρόμηση Διανυσμάτων Υποστήριξης (Support Vector Regression), Πλησιέστερος Γείτονας (Nearest Neighbor), Συνδυαστική Μάθηση (Ensemble Learning), (Νευρωνικό Δίκτυο) Neural Network και Παλινδρόμηση Μειγμάτων Gauss (Gaussian Mixture Regression) (Lefèvre et al., 2014), (Bai et al., 2018). Το μειονέκτημα όλων των μεθόδων βασισμένων σε δεδομένα, είναι πως το κάθε μοντέλο μπορεί να εφαρμοστεί σε μόνο ένα οδικό τμήμα οπότε η βαθμονόμηση των παραμέτρων γίνεται για όλες τις διαφορετικές τοποθεσίες του δικτύου (Van Lint, 2004). Η μηχανική μάθηση για την οποία γίνεται αναφορά παρακάτω, αποτελεί μέθοδο βασισμένη σε δεδομένα.

2.7.4. Μεταβλητότητα / Αξιοπιστία χρόνου διαδρομής

Η μεταβλητότητα του χρόνου διαδρομής (travel-time variability) μπορεί να επιφέρει απρόοπτες καθυστερήσεις σε ένα οδικό δίκτυο οι οποίες δεν μπορούν

να προβλεφθούν. Αυτή οφείλεται σε μεταβολές των κυκλοφοριακών συνθηκών λόγω καιρικών συνθηκών, βλαβών, ατυχημάτων κ.α., και είναι άμεσα συνδεδεμένες με την αξιοπιστία του χρόνου διαδρομής (travel-time reliability) (Carrion & Levinson, 2012). Συνεπώς αυτή εκφράζει τη σταθερότητα/συνέπεια που έχει ο χρόνος διαδρομής μέσα στις ημέρες ή κατά τη διάρκεια μίας ημέρας και είναι σημαντική για τον προγραμματισμό των δραστηριοτήτων των μετακινούμενων και την αποφυγή καθυστερήσεων. Οι ώρες αιχμής όπου παρατηρείται συμφόρηση δεν μπορούν να θεωρηθούν απρόβλεπτο γεγονός καθώς είναι ένα γνωστό και προβλέψιμο φαινόμενο για τους μετακινούμενους οι οποίοι κινούνται αυτές τις ώρες και προσαρμόζουν την ώρα αναχώρησης τους αναλόγως (Bates et al., 2001).

Οι μέθοδοι αξιολόγησης της αξιοπιστίας που προτείνει το Υπουργείο Μεταφορών των ΗΠΑ είναι: ο 90^{ος} ή 95^{ος} εκατοστιαίος χρόνος διαδρομής (90th or 95th percentile travel times), ο δείκτης buffer (buffer index) και ο δείκτης προγραμματισμένου χρόνου (planning time index). Για την βέλτιστη εφαρμογή αυτών αξιοποιούνται κυκλοφοριακά δεδομένα με διαρκή καταγραφή και συλλογή μέσω Ευφυών Μεταφορικών Συστημάτων (Intelligent Transportations Systems - ITS) ή δεδομένα με άλλες μεθόδους καταγραφής. Κατά συνέπεια, ο υπολογισμός της αξιοπιστίας βασίζεται σε πραγματικές παρατηρήσεις οδικών τμημάτων και εξάγονται συμπεράσματα για τα αντίστοιχα τμήματα.

2.8. Μηχανική Μάθηση στην πρόβλεψη των χρόνων διαδρομής

Η Μηχανική Μάθηση, ορίζεται ως ένα πρόγραμμα υπολογιστή το οποίο μαθαίνει από την εμπειρία πάνω σε ορισμένα προβλήματα με ένα μέτρο απόδοσης, εάν η απόδοσή του στα προβλήματα αυτά, όπως αυτή μετριέται από το μέτρο απόδοσης, βελτιώνεται με την εμπειρία (Mitchell, 1997). Οι πιο διαδεδομένες τεχνικές μηχανικής μάθησης είναι: Τυχαία Δάση (Random Forests – RF), Αλγόριθμοι Ενίσχυσης Κλίσης (Gradient Boosting Machines – GBM), Μηχανή Διανυσμάτων Στήριξης (Support Vector Machines – SVM) και Νευρωνικά Δίκτυα (Neural Networks).

Η μέθοδος των Τυχαίων Δασών (Random Forests) αποτελεί μια προηγμένη τεχνική μηχανικής μάθησης η οποία συνδυάζει πολλαπλά δέντρα αποφάσεων για την επίτευξη υψηλής ακρίβειας σε προβλήματα ταξινόμησης και παλινδρόμησης. Η τυχαία επιλογή των μεταβλητών σε κάθε κόμβο έχει ως αποτέλεσμα τη μείωση της συσχέτισης μεταξύ των δέντρων χωρίς όμως να θυσιάζεται η ισχύς τους. Επιπλέον, τα Τυχαία Δάση παρουσιάζουν μειωμένη τάση υπερπροσαρμογής (overfitting) καθώς αυξάνεται ο αριθμός των δέντρων, ενώ υπερτερούν έναντι άλλων μεθόδων ως προς την ταχύτητα και την ανθεκτικότητα στον θόρυβο των δεδομένων (Breiman, 2001).

Η αποτελεσματικότητα των μοντέλων μηχανικής μάθησης στην βραχυπρόθεσμη πρόβλεψη των χρόνων διαδρομής τεκμηριώνεται από τις μελέτες των Sakib, 2024 και Qiu & Fan, 2021 οι οποίες, παρά τις αποκλίσεις στο εύρος των δεδομένων τους, καταλήγουν στην υπεροχή του αλγορίθμου Τυχαίων Δασών (RF). Στην έρευνα του Sakib, η οποία βασίστηκε σε 3.360 παρατηρήσεις από το Google Maps API σε τμήμα ελεύθερης

λεωφόρου στην πόλη Ντάκα, Μπαγκλαντές, το μοντέλο RF πέτυχε υψηλή ακρίβεια με Mean Absolute Error (MAE) 22.92% και Mean Absolute Percentage Error (MAPE) 12.04%. Αντίστοιχα, οι Qiu & Fan, αναλύοντας δεδομένα ενός πλήρους έτους από τον αυτοκινητόδρομο I-485 στην πόλη Σάρλοτ, Βόρεια Καρολίνα, επιβεβαίωσαν ότι ο RF υπερέχει σε σταθερότητα έναντι των Decision Trees, XGBoost και των νευρωνικών δικτύων Long Short-Term Memory (LSTM).

Μεθοδολογικά, και οι δύο μελέτες διαχώρισαν το 70% του δείγματος για την εκπαίδευση (training) και το 30% για τον έλεγχο (testing) των μοντέλων. Οι μεταβλητές που χρησιμοποιήθηκαν για τα μοντέλα μηχανικής μάθησης είναι χωροχρονικές όπως η ημέρα της εβδομάδας, η ώρα της ημέρας, οι χρόνοι διαδρομής 15, 30, 45 λεπτά πριν και το μήκος του οδικού τμήματος, με την έρευνα των Qiu & Fan να ενσωματώνει μετεωρολογικά δεδομένα (κατηγοριοποιημένα σε κανονικές συνθήκες, βροχή και χιόνι/πάγο). Τα συμπεράσματα των ερευνών αναδεικνύουν τον τρέχοντα χρόνο διαδρομής και τη χρονική στιγμή μέσα στην ημέρα ως τους κρισιμότερους παράγοντες επιρροής, υπογραμμίζοντας ότι η συνδυαστική χρήση ιστορικών και πραγματικών δεδομένων ενισχύει την αξιοπιστία των βραχυπρόθεσμων προβλέψεων. Ωστόσο, οι Qiu & Fan επισημαίνουν ορισμένες ελλείψεις και περιορισμούς στην έρευνα τους. Αρχικά αναφέρουν πως τα μοντέλα πρόβλεψης αναπτύχθηκαν κάτω από κανονικές συνθήκες κυκλοφορίας χωρίς να λαμβάνουν υπόψη απρόσμενα συμβάντα όπως ατυχήματα ή δραστηριότητες σε εργασιακές ζώνες. Επίσης κρίνουν πως πρέπει να διερευνηθεί η δυνατότητα εφαρμογής του μοντέλου σε διαφορετικά τμήματα από αυτό που εκπαιδεύτηκε χρησιμοποιώντας άλλες πηγές δεδομένων και να αξιοποιηθούν δεδομένα σε πραγματικό χρόνο για πιο ακριβείς προβλέψεις.

Την αποτελεσματικότητα των Gradient Boosting Machines (GBM) στην πρόβλεψη των χρόνων διαδρομής αναδεικνύουν οι μελέτες των Zhang & Haghani, 2015 και Cheng et al., 2019 στον αυτοκινητόδρομο I-95 στο Μέριλαντ των ΗΠΑ και στον αυτοκινητόδρομο του αεροδρομίου της πόλης Ναντζίνγκ στην Κίνα αντίστοιχα. Οι έρευνες δείχνουν πως τα GBM μπορούν να επιλέγουν αυτόματα τις σχετικές μεταβλητές, να προσαρμόζουν ακριβή μοντέλα, καθώς και να εντοπίζουν και να μοντελοποιούν τις αλληλεπιδράσεις μεταξύ των παραμέτρων. Επιπλέον, μπορούν να αναγνωρίσουν τις ξαφνικές αλλαγές στις κυκλοφοριακές συνθήκες που οφείλονται σε απρόσμενα συμβάντα ή αλλαγή στις καιρικές συνθήκες. Το μειονέκτημα των μεθόδων αυτών είναι πως επηρεάζονται σημαντικά από τις παραμέτρους τους όπως τον αριθμό των δέντρων, τον ρυθμό μάθησης και την πολυπλοκότητα των δέντρων. Ωστόσο, αποτελούν μία αξιόπιστη και ικανή μέθοδο πρόβλεψης χρόνων διαδρομής η οποία έχει τη δυνατότητα να χειριστεί διαφορετικού τύπου μεταβλητές εισόδου και μεγάλα δεδομένα (big data) στην μοντελοποίηση πολύπλοκων μη-γραμμικών σχέσεων.

Έρευνα	Δεδομένα	Μέθοδοι	Μεταβλητές μοντέλων	Αποτελέσματα
N. Sakib, 2024	Google Maps API κάθε 15 λεπτά για 7 ημέρες (3360 παρατηρήσεις)	• Decision Tree • Random Forest • Gradient Boosting Machine • XGBoost, LightGBM • SHAP για σημαντικότητα μεταβλητών	• Μήκος οδικού τμήματος • Ώρα της ημέρας • Ημέρα εβδομάδας • Προηγούμενοι χρόνοι διαδρομής (15, 30, 45, 60 λεπτά πριν) • Διαφορά χρόνων διαδρομής μεταξύ των παραπάνω	Καλύτερο το RF (MAE = 22.92%, MAPE = 12.04%). Σημαντικότερη μεταβλητή ο χρόνος 15 λεπτά πριν (SHAP)
Qiu B & Fan W, 2021	Δεδομένα χρόνου διαδρομής για ολόκληρο το έτος 2019 (ανά διαστήματα 15 λεπτών) στον αυτοκινητόδρομο I-485 στο Σάρλοτ της Βόρειας Καρολίνας	• Decision Trees • Random Forest • Extreme Gradient Boosting (XGBoost) • Long Short-Term Memory Neural Network (LSTM)	• Ώρα της ημέρας, ημέρα της εβδομάδας, μήνας, καιρικές συνθήκες, μέση ταχύτητα, μήκος οδικού τμήματος • Χρόνος διαδρομής 15, 30 και 45 λεπτά πριν, 1 εβδομάδα πριν	RF: η πιο ακριβής και σταθερή σε όλους τους ορίζοντες. XGBoost: ταχύτερο αλλά πιο δύσκολο στην προσαρμογή. Χρόνος 15 λεπτά πριν: 34.8% συνεισφορά
Zhang & Haghani, 2015	Δεδομένα χρόνου διαδρομής για τμήματα των αυτοκινητοδρόμων I-95 και MD 295 στο Maryland με ομαδοποίηση ανά 5 λεπτά	•Random Forest •Gradient Boosting Machine •ARIMA	• Ώρα της ημέρας, ημέρα της εβδομάδας, ημέρα του μήνα και μήνας • Πρόσφατες παρατηρήσεις χρόνου διαδρομής • Ρυθμούς μεταβολής μεταξύ διαδοχικών χρονικών βημάτων	Στα 5 λεπτά, GBM και RF ξεπερνούν το ARIMA. Στα 15 και 30 λεπτά, το GBM είναι το πιο ανθεκτικό (9 στις 10 περιπτώσεις) και διαχειρίζεται απότομες μεταβολές (ατυχήματα, βροχή)

Cheng et al., 2019	Συλλογή δεδομένων μέσω του λογισμικού προσομοίωσης VISSIM σε τμήμα του αυτοκινητοδρόμου Nanjing Airport 133 προσομοιώσεις (που αντιστοιχούν σε 133 ημέρες κυκλοφορίας μεταξύ 09:00–17:00, με διάστημα δειγματοληψίας 5 λεπτών	• Gradient Boosting Decision Tree (GBDT) • Νευρωνικό δίκτυο Back Propagation (BP) • Support Vector Machines (SVM)	• Τρέχων χρόνος μετακίνησης Ti, προηγούμενοι χρόνοι TTi-1, TT(i-2) και διαφορές χρόνου ΔTi, ΔT(i-1) • Φόρτος κυκλοφορίας, ταχύτητα, πυκνότητα, κατάληψη, αριθμός οχημάτων και παράμετρος κυκλοφοριακής κατάστασης	GBDT: η υψηλότερη ακρίβεια σε όλους τους ορίζοντες και η μικρότερη ευαισθησία στην αύξηση του ορίζοντα. Ο τρέχων χρόνος διαδρομής ήταν η σημαντικότερη μεταβλητή (100%)

Πίνακας 2.4: Έρευνες για την πρόβλεψη των χρόνων διαδρομής με μεθόδους Μηχανικής Μάθησης

2.9. Σύνοψη

Η βιβλιογραφική ανασκόπηση ανέδειξε ότι η κυκλοφοριακή συμφόρηση επηρεάζεται από κοινωνικοοικονομικούς παράγοντες, όπως ο πληθυσμός, το εισόδημα, η απασχόληση και η ιδιοκτησία αυτοκινήτων, αλλά και από τις μετεωρολογικές συνθήκες, με τη βροχή και το χιόνι να μειώνουν την ταχύτητα κίνησης. Οι καιρικές συνθήκες επηρεάζουν παράλληλα την επιλογή μέσου μετακίνησης, η οποία εξαρτάται επιπλέον από τα χαρακτηριστικά του μετακινούμενου, τη φύση της μετακίνησης και τα χαρακτηριστικά του μεταφορικού συστήματος, με την απόσταση να διαδραματίζει καθοριστικό ρόλο. Σε αυτό το πλαίσιο, τα Μέσα Μαζικής Μεταφοράς συμβάλλουν στην αποσυμφόρηση του οδικού δικτύου, υπό την προϋπόθεση επαρκούς ποσοστού χρήσης τους, ιδίως όταν διαθέτουν αποκλειστικές λωρίδες κυκλοφορίας. Όσον αφορά την πρόβλεψη των χρόνων διαδρομής, οι μέθοδοι βασισμένες σε μοντέλα ενδείκνυνται για την ανάλυση σεναρίων αλλά είναι υπολογιστικά απαιτητικές, ενώ οι μέθοδοι βασισμένες σε δεδομένα, στις οποίες ανήκει η μηχανική μάθηση, προσφέρουν υψηλή ακρίβεια και αξιοπιστία. Ειδικότερα, τα Τυχαία Δάση και τα μοντέλα Ενίσχυσης Κλίσης υπερέχουν των παραδοσιακών μεθόδων, καθώς χειρίζονται αποτελεσματικά μη γραμμικές σχέσεις και απότομες μεταβολές. Ωστόσο, οι υπάρχουσες μελέτες περιορίζονται κυρίως σε βραχυπρόθεσμες προβλέψεις σε αυτοκινητόδρομους υπό κανονικές συνθήκες, χωρίς να λαμβάνουν υπόψη την επιλογή μέσου. Η παρούσα διπλωματική εργασία καλύπτει αυτό το κενό, εστιάζοντας στην ανάλυση σε αστικό περιβάλλον, στο κέντρο της Αθήνας, περιλαμβάνοντας δεδομένα πενταετίας, με ενσωμάτωση της σύνθεσης της κυκλοφορίας, των μετεωρολογικών συνθηκών και των κοινωνικοοικονομικών παραγόντων. Οι δείκτες συνολικής καθυστέρησης και ταχύτητας κρίνονται σύμφωνα με τη βιβλιογραφία οι πιο απλοί οπότε χρησιμοποιήθηκαν ως η εξαρτημένη μεταβλητή των μοντέλων. Τέλος, μέσω των αναλύσεων δόθηκε η ευκαιρία σύγκρισης των μεθόδων παλινδρόμησης με αυτές της μηχανικής μάθησης στην πρόβλεψη της συμφόρησης στο αστικό οδικό δίκτυο.

3. Θεωρητικό Υπόβαθρο

3.1. Εισαγωγή

Στο παρών κεφάλαιο παρουσιάζεται το θεωρητικό υπόβαθρο που χρησιμοποιήθηκε για την επεξεργασία και ανάλυση των δεδομένων καθώς και για την ανάπτυξη των στατιστικών και μαθηματικών μοντέλων. Αρχικά αναλύονται οι βασικές έννοιες της περιγραφικής στατιστικής και αναπτύσσεται η λογική των μοντέλων παλινδρόμησης και μηχανικής μάθησης που χρησιμοποιήθηκαν στο Κεφάλαιο 5 της διπλωματικής.

3.2. Βασικές έννοιες στατιστικής

Οι στατιστικοί δείκτες που αναλύονται παρακάτω, εφαρμόστηκαν στην προκαταρκτική στατιστική ανάλυση των δεδομένων με σκοπό την εξαγωγή αρχικών συμπερασμάτων σχετικά με τις κυκλοφοριακές, κοινωνικοοικονομικές και μετεωρολογικές συνθήκες που επικρατούν στην Αττική.

3.2.1. Μέγεθος δείγματος (N)

Το μέγεθος του δείγματος ορίζεται ως ο συνολικός αριθμός των ανεξάρτητων παρατηρήσεων που περιλαμβάνονται στην ανάλυση. Αυτό καθορίζει τη στατιστική ισχύ των ελέγχων υποθέσεων και την ακρίβεια των εκτιμήσεων.

3.2.2. Μέση Τιμή ($\bar{X}$)

Η αριθμητική μέση τιμή αποτελεί έναν δείκτη κεντρικής τάσης και αντιπροσωπεύει το κέντρο βάρους των δεδομένων. Για μία μεταβλητή X με N παρατηρήσεις, ορίζεται ως:

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

Η μέση τιμή παρέχει μια συνολική εκτίμηση για το επίπεδο της εκάστοτε μεταβλητής, ωστόσο είναι ευαίσθητη στην παρουσία ακραίων τιμών οι οποίες δύναται να αλλοιώσουν την αντιπροσωπευτικότητα της.

3.2.3. Εύρος (R)

Το εύρος είναι ένας δείκτης διασποράς και ορίζεται ως η απόλυτη διαφορά ανάμεσα στη μέγιστη και την ελάχιστη παρατηρηθείσα τιμή ενός δείγματος:

$$R = X_{max} - X_{min}$$

Αποτυπώνει τα ακραία όρια στα οποία επεκτείνονται τα δεδομένα, όμως το μειονέκτημα του είναι ότι βασίζεται αποκλειστικά στις δύο ακραίες τιμές, αγνοώντας τη συμπεριφορά των ενδιάμεσων παρατηρήσεων.

3.2.4. Διακύμανση Δείγματος (s^2)

Η διακύμανση (ή διασπορά) είναι ένας θεμελιώδης δείκτης που μετρά τον βαθμό στον οποίο οι τιμές των δεδομένων απλώνονται ή συγκεντρώνονται γύρω από τη μέση τιμή τους και υπολογίζεται ως εξής:

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

Το μειονέκτημα της διακύμανσης στην ερμηνεία είναι ότι εκφράζεται σε τετραγωνισμένες μονάδες μέτρησης της αρχικής μεταβλητής.

3.2.5. Τυπική Απόκλιση Δείγματος (s)

Η τυπική απόκλιση είναι ο πιο ευρέως χρησιμοποιούμενος δείκτης διασποράς και ορίζεται μαθηματικά ως η θετική τετραγωνική ρίζα της διακύμανσης:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

Το πλεονέκτημα της τυπικής απόκλισης έναντι της διακύμανσης είναι ότι εκφράζεται στις ίδιες μονάδες μέτρησης καθιστώντας την πιο ερμηνεύσιμη. Μια μικρή τιμή τυπικής απόκλισης υποδηλώνει ότι οι παρατηρήσεις είναι συγκεντρωμένες πολύ κοντά στη μέση τιμή, ενώ μια μεγάλη τιμή φανερώνει έντονη μεταβλητότητα και ανομοιογένεια στο δείγμα.

3.3. Είδη Μεταβλητών

3.3.1. Γενικά

Στη στατιστική ανάλυση και μοντελοποίηση, οι μεταβλητές ταξινομούνται με βάση τη μαθηματική τους φύση και την κλίμακα μέτρησης τους. Η διάκριση αυτή είναι θεμελιώδης, καθώς καθορίζει τις κατάλληλες μεθόδους στατιστικής ανάλυσης και τον τρόπο εισαγωγής των δεδομένων στα μαθηματικά μοντέλα. Οι μεταβλητές διακρίνονται σε δύο κατηγορίες: τις ποιοτικές (κατηγορικές) και τις ποσοτικές (συνεχείς).

3.3.2. Ποσοτικές (Συνεχείς) Μεταβλητές

Οι συνεχείς μεταβλητές αντιπροσωπεύουν μεγέθη που προκύπτουν από μετρήσεις και μπορούν να λάβουν οποιαδήποτε αριθμητική τιμή εντός ενός δεδομένου διαστήματος, περιλαμβάνοντας και δεκαδικούς αριθμούς. Στις μεταβλητές αυτές έχουν νόημα όλες οι μαθηματικές πράξεις καθώς και οι στατιστικοί δείκτες.

3.3.3. Ποιοτικές (Κατηγορικές) Μεταβλητές

Οι κατηγορικές μεταβλητές αποτελούνται από τιμές που δεν εκφράζουν ποσότητα, αλλά κατατάσσουν τις παρατηρήσεις σε διακριτές ομάδες, κατηγορίες ή ιδιότητες. Διακρίνονται περαιτέρω σε δύο υποκατηγορίες:

- Ονομαστικές Μεταβλητές: Οι κατηγορίες δεν έχουν κάποια φυσική ή μαθηματική ιεράρχηση μεταξύ τους (π.χ. το όνομα ενός οδικού τμήματος)

- Διατακτικές Μεταβλητές: Οι κατηγορίες διαθέτουν μια εγγενή, λογική σειρά ή διαβάθμιση, χωρίς όμως η απόσταση μεταξύ των κατηγοριών να είναι σταθερή ή μετρήσιμη (π.χ. βαθμίδες έντασης φαινομένων)

Μία ειδική περίπτωση κατηγορικών μεταβλητών που διαθέτουν ακριβώς δύο κατηγορίες (π.χ. Ναι/Όχι, Αληθές/Ψευδές), είναι οι δυαδικές μεταβλητές.
Στην παρούσα μελέτη, οι αρχικές συνεχείς μετεωρολογικές μεταβλητές (θερμοκρασία, άνεμος, βροχή) καθώς και η ώρα της ημέρας υπέστησαν μια διαδικασία κατηγοριοποίησης βάση διεθνών προτύπων. Αυτό κρίθηκε αναγκαίο για τους εξής λόγους:

- Αποτύπωση μη γραμμικών σχέσεων: Η επίδραση του καιρού στην οδική συμπεριφορά δεν είναι πάντα γραμμική. Για παράδειγμα, η διαφορά μεταξύ 0 και 2mm βροχής μπορεί να επηρεάσει την ταχύτητα περισσότερο από τη διαφορά 10 και 12mm. Η κατηγοριοποίηση λοιπόν επιτρέπει στο μοντέλο γραμμικής παλινδρόμησης να αποτυπώσει αυτές τις μη γραμμικότητες.
- Ερμηνευσιμότητα: Διευκολύνει την εξαγωγή συμπερασμάτων και την ερμηνεία του μοντέλου.

Στην πολλαπλή γραμμική παλινδρόμηση, οι ανεξάρτητες μεταβλητές πρέπει να είναι ποσοτικές προκειμένου να εκτελεστούν οι μαθηματικές πράξεις του μοντέλου. Για αυτόν τον σκοπό, οι κατηγορικές μεταβλητές μετατράπηκαν σε αριθμητικές μέσω της διαδικασίας κωδικοποίησης ψευδομεταβλητών. Αν μια κατηγορική μεταβλητή διαθέτει k διακριτές κατηγορίες, η μέθοδος δημιουργεί k-1 δυαδικές ψευδομεταβλητές οι οποίες λαμβάνουν την τιμή 1 αν η παρατήρηση ανήκει στη συγκεκριμένη κατηγορία και την τιμή 0 σε κάθε άλλη περίπτωση. Η κατηγορία που παραλείπεται και δεν αντιστοιχεί σε κάποια ψευδομεταβλητή ονομάζεται κατηγορία αναφοράς. Με αυτόν τον τρόπο, οι συντελεστές των υπόλοιπων ψευδομεταβλητών εκφράζουν τη μέση απόκλιση της εκάστοτε κατηγορίας σε σχέση με την κατηγορία αναφοράς, κρατώντας όλους τους άλλους παράγοντες σταθερούς.

3.4. Γραμμική Παλινδρόμηση

3.4.1. Γενικά

Στη στατιστική, η Γραμμική Παλινδρόμηση είναι μια προσέγγιση για τη μοντελοποίηση της σχέσης μεταξύ μιας βαθμωτής εξαρτημένης μεταβλητής Y και μίας ή περισσότερων επεξηγηματικών μεταβλητών (ή ανεξάρτητων μεταβλητών) X. Περίπτωση μιας επεξηγηματικής μεταβλητής ονομάζεται απλή γραμμική παλινδρόμηση ενώ για περισσότερες από μία επεξηγηματικές μεταβλητές, η διαδικασία ονομάζεται πολλαπλή γραμμική παλινδρόμηση και διατυπώνεται από τη σχέση:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Όπου:
- Y: η εξαρτημένη μεταβλητή
- $X_1, X_2, \ldots, X_n$: οι ανεξάρτητες μεταβλητές
- ε: το τυχαίο σφάλμα

- β_0: η σταθερά του μοντέλου
- $\beta_1, \beta_2, \ldots, \beta_n$: οι συντελεστές παλινδρόμησης

3.4.2. Μέθοδος Ελαχίστων Τετραγώνων (OLS)

Η μέθοδος ελαχίστων τετραγώνων (OLS) στην πολλαπλή γραμμική παλινδρόμηση υπολογίζει τις άγνωστες παραμέτρους του μοντέλου (β_i) ελαχιστοποιώντας το άθροισμα των τετραγώνων των διαφορών (καταλοίπων) μεταξύ των πραγματικών και των εκτιμώμενων τιμών. Ο πίνακας των παραμέτρων $\hat{\beta}$ προκύπτει από την εξίσωση:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Όπου: X είναι ο πίνακας των ανεξάρτητων μεταβλητών και y το διάνυσμα της εξαρτημένης μεταβλητής.

Για να είναι έγκυροι οι εκτιμητές της μεθόδου ελαχίστων τετραγώνων, πρέπει να ικανοποιούνται οι παρακάτω βασικές προϋποθέσεις:

1. Γραμμικότητα στις Παραμέτρους:
   Η σχέση μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων μεταβλητών πρέπει να είναι γραμμική ως προς τους συντελεστές.

2. Απουσία Τέλειας Πολυσυγραμμικότητας:
   Οι επεξηγηματικές μεταβλητές δεν πρέπει να έχουν τέλεια γραμμική συσχέτιση μεταξύ τους. Στο μοντέλο που αναπτύχθηκε χρησιμοποιήθηκε ο δείκτης VIF.

3. Μηδενική Μέση Τιμή των Σφαλμάτων:
   Η αναμενόμενη τιμή (μέσος όρος) του όρου σφάλματος ε για οποιονδήποτε συνδυασμό ανεξάρτητων μεταβλητών πρέπει να είναι μηδέν $(E(\varepsilon|X) = 0)$.

4. Ομοσκεδαστικότητα:
   Η διασπορά των καταλοίπων (σφαλμάτων) πρέπει να είναι σταθερή για όλες τις τιμές των ανεξάρτητων μεταβλητών $(\mathrm{Var}(\varepsilon|X) = \sigma^2)$.

5. Απουσία Αυτοσυσχέτισης:
   Τα σφάλματα διαφορετικών παρατηρήσεων δεν πρέπει να συσχετίζονται μεταξύ τους ($\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ για $i \neq j$). Στο μοντέλο που αναπτύχθηκε χρησιμοποιήθηκε ο δείκτης Durbin-Watson (DW).

6. Κανονικότητα των Σφαλμάτων:
   Για μικρά δείγματα, οι όροι σφάλματος πρέπει να ακολουθούν κανονική κατανομή. Στο μοντέλο που αναπτύχθηκε χρησιμοποιήθηκε το στατιστικό τεστ Jarque-Bera (JB).

3.4.3. Εκτιμητής διασποράς Newey-West (HAC)

Όταν τα υπόλοιπα ενός γραμμικού μοντέλου χρονοσειράς παρουσιάζουν αυτοσυσχέτιση ή ετεροσκεδαστικότητα, τα τυπικά σφάλματα του OLS δεν είναι

έγκυρα και τα p-values είναι συνήθως υπερβολικά αισιόδοξα. Για τον λόγο, στο μοντέλο διερεύνησης επιρροής των κοινωνικοοικονομικών δεδομένων στη συμφόρηση, χρησιμοποιείται ο εκτιμητής διασποράς Newey-West (HAC), ο οποίος υπολογίζει τα τυπικά σφάλματα των συντελεστών λαμβάνοντας υπόψη τις αυτοσυνδιακυμάνσεις των υπολοίπων μέχρι μια μέγιστη καθυστέρηση L, με γραμμικά φθίνοντα βάρη Bartlett. Οι εκτιμήσεις των συντελεστών παραμένουν ίδιες με του OLS, ενώ αλλάζουν τα τυπικά σφάλματα, τα p-values και τα διαστήματα εμπιστοσύνης. Στην παρούσα εργασία χρησιμοποιήθηκε L = 2, σύμφωνα με τον κανόνα L = ⌊4(n/100)^(2/9)⌋ για n = 16 παρατηρήσεις (Ενότητα 5.2.3). Επειδή η μέθοδος είναι ασυμπτωτική, με μικρό δείγμα η διόρθωση είναι περιορισμένης αξιοπιστίας και τα αποτελέσματα ερμηνεύονται ως ενδείξεις.

3.4.4. Έλεγχος συσχέτισης μεταβλητών

Η γραμμική σχέση μεταξύ δύο ποσοτικών μεταβλητών, μετρείται με τους συντελεστές συσχέτισης Pearson (r) και Spearman (ρ). Ο συντελεστής Pearson ορίζεται ως:

$$r = \sum (X_i - \bar{X})(y_i - \bar{y}) \Big/ \sqrt{\left[\sum (X_i - \bar{X})^2 \cdot \sum (y_i - \bar{y})^2\right]}$$

δηλαδή ως ο λόγος της συνδιακύμανσης των δύο μεταβλητών προς το γινόμενο των τυπικών τους αποκλίσεων. Λαμβάνει τιμές από −1 (τέλεια αρνητική γραμμική σχέση) έως +1 (τέλεια θετική), ενώ η τιμή 0 δείχνει απουσία γραμμικής σχέσης. Η στατιστική σημαντικότητά του ελέγχεται με τη στατιστική:

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

η οποία ακολουθεί την κατανομή t με n − 2 βαθμούς ελευθερίας, υπό την υπόθεση ότι οι μεταβλητές προέρχονται από κανονική κατανομή. Ο Pearson είναι ευαίσθητος σε ακραίες τιμές και δεν ανιχνεύει μη γραμμικές σχέσεις. Ο συντελεστής του Spearman (ρ) είναι ο συντελεστής του Pearson που εφαρμόζεται στις κατατάξεις των παρατηρήσεων και όχι στις ίδιες τις τιμές, οπότε μετράει τη μονότονη σχέση, δηλαδή το κατά πόσο η μία μεταβλητή αυξάνεται (ή μειώνεται) όταν αυξάνεται η άλλη, χωρίς να απαιτείται η σχέση να είναι γραμμική. Ως μη παραμετρικός δείκτης δεν προϋποθέτει κανονικότητα και είναι λιγότερο ευαίσθητος σε ακραίες τιμές. Η ταυτόχρονη χρήση των δύο δεικτών επιτρέπει να ελεγχθεί αν μια συσχέτιση που προκύπτει με τον Pearson οφείλεται σε λίγες ακραίες παρατηρήσεις ή αν είναι συνεπής και ως προς τη διάταξη των τιμών.

Στη στατιστική μοντελοποίηση, η πολυσυγγραμικότητα εμφανίζεται όταν δύο ή περισσότερες ανεξάρτητες μεταβλητές συνδέονται μεταξύ τους με υψηλό βαθμό γραμμικής συσχέτισης. Το φαινόμενο αυτό παραβιάζει τη βασική προϋπόθεση της ανεξαρτησίας των επεξηγηματικών μεταβλητών. Ο Συντελεστής Διόγκωσης της Διακύμανσης (Variance Inflation Factor – VIF) μετρά την αύξηση της διακύμανσης ενός εκτιμώμενου συντελεστή παλινδρόμησης λόγω της ύπαρξης συσχέτισης μεταξύ των ανεξάρτητων μεταβλητών μέσω της παρακάτω σχέσης:

$$VIF_i = \frac{1}{1 - R_i^2}$$

Όπου:

R_i^2: Ο συντελεστής προσδιορισμού της βοηθητικής γραμμικής παλινδρόμησης, όπου η ανεξάρτητη μεταβλητή X_i παλινδρομεί έναντι όλων των υπόλοιπων ανεξάρτητων μεταβλητών του μοντέλου.

- Για VIF = 1, δεν υπάρχει καμία συσχέτιση.
- Για 1 < VIF < 5, έχουμε χαμηλή έως μέτρια συσχέτιση και είναι αποδεκτό.
- Για VIF > 5, έχουμε υψηλή πολυσυγγραμικότητα και αφαιρείται η αντίστοιχη μεταβλητή από το μοντέλο.

3.4.5. Έλεγχος σημαντικότητας μεταβλητών

Η συνολική σημαντικότητα των επεξηγηματικών μεταβλητών στα μοντέλα γραμμικής παλινδρόμησης: ταχύτητα – καιρός και ταχύτητα – σύνθεση, ελέγχθηκε με τον έλεγχο Wald. Η βασική αρχή του ελέγχου συνίσταται στην αποτίμηση της στατιστικής απόστασης μεταξύ των εκτιμημένων συντελεστών του πλήρους υποδείγματος και των τιμών που επιβάλλει η μηδενική υπόθεση (H_0), λαμβάνοντας υπόψη τα τυπικά σφάλματα των εκτιμητών.

- Μηδενική Υπόθεση (H_0): $\beta_1 = \beta_2 = \ldots = \beta_n = 0$

Ερμηνεία: Το μοντέλο δεν είναι στατιστικά σημαντικό. Καμία από τις επεξηγηματικές μεταβλητές δεν βοηθά στην πρόβλεψη ή την ερμηνεία της εξαρτημένης μεταβλητής. Η ερμηνευτική ισχύς του μοντέλου είναι μηδενική. Αν p > 0.05, η μηδενική υπόθεση δεν απορρίπτεται και η μεταβλητή δεν είναι συνολικά σημαντική.

- Εναλλακτική Υπόθεση (H_1): Τουλάχιστον ένας συντελεστής $\beta_i \neq 0$.

Ερμηνεία: Το μοντέλο είναι συνολικά στατιστικά σημαντικό. Τουλάχιστον μία ανεξάρτητη μεταβλητή (ή ένας συνδυασμός τους) έχει συστηματική και μη τυχαία επίδραση στην εξαρτημένη μεταβλητή. Αν p < 0.05 η μηδενική υπόθεση απορρίπτεται και η μεταβλητή είναι συνολικά σημαντική.

Η εφαρμογή του ελέγχου κρίνεται ιδιαίτερα κρίσιμη στην περίπτωση εισαγωγής κατηγορικών μεταβλητών πολλαπλών επιπέδων καθώς οι μεταβλητές αυτές διασπώνται σε ένα σύνολο επιμέρους μεταβλητών με αποτέλεσμα τα κλασικά κριτήρια t-tests να αποτιμούν τη σημαντικότητα κάθε κατηγορίας μεμονωμένα έναντι της κατηγορίας βάσης.

3.5. Αξιολόγηση μοντέλων παλινδρόμησης

Η αξιολόγηση των μοντέλων παλινδρόμησης έγινε σε πέντε στάδια. Αρχικά ελέγχθηκε η ερμηνευτική ισχύς και η προσαρμογή με τους δείκτες R^2, Adj. R^2, έγινε έλεγχος της προβλεπτικής ακρίβειας με MAE, RMSE, σύγκριση μοντέλων με τους δείκτες: F-test και AIC, και στη συνέχεια πραγματοποιήθηκαν οι διαγνωστικοί έλεγχοι καταλοίπων (Residual Diagnostics) για τον έλεγχο εγκυρότητας. Τέλος πραγματοποιήθηκε έλεγχος αυτοσυσχέτισης και

πολυσυγγραμικότητας με του δείκτες Durbin-Watson (DW) και δείκτη κατάστασης (Condition Number).

3.5.1. Ερμηνευτική ισχύς και προσαρμογή

Ο δείκτης R^2 μετρά το ποσοστό της συνολικής διακύμανσης της εξαρτημένης μεταβλητής που ερμηνεύεται από τις ανεξάρτητες μεταβλητές του μοντέλου και παίρνει τιμές από 0 έως 1. Όσο πιο κοντά είναι στο 1, τόσο καλύτερη είναι η προσαρμογή του μοντέλου στα δεδομένα.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{SSE}{SST}$$

Το R^2 δεν μειώνεται ποτέ με την προσθήκη μεταβλητών, ακόμη και όταν αυτές δεν έχουν πραγματική ερμηνευτική ισχύ, γεγονός που το καθιστά προβληματικό για τη σύγκριση μοντέλων με διαφορετικό αριθμό παραμέτρων, ιδιαίτερα σε μικρά δείγματα. Το προσαρμοσμένο R^2 (Adjusted R^2), αυξάνεται μόνο όταν η νέα μεταβλητή μειώνει το SSE περισσότερο απ' όσο θα αναμενόταν τυχαία, δηλαδή όταν η απόλυτη τιμή της στατιστικής t του συντελεστή της υπερβαίνει τη μονάδα.

$$Adj.R^2 = 1 - \frac{SSE/(n-p)}{SST/(n-1)}$$

όπου n το πλήθος των παρατηρήσεων και p το πλήθος των παραμέτρων του μοντέλου, συμπεριλαμβανομένου του σταθερού όρου.

3.5.2. Προβλεπτική ακρίβεια

Το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error – MAE) υπολογίζει τον μέσο όρο των απόλυτων αποκλίσεων μεταξύ των πραγματικών και των προβλεπόμενων τιμών, προσφέροντας μια άμεσα ερμηνεύσιμη και εύκολα κατανοητή εικόνα του μέσου σφάλματος.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Αντίστοιχα, η Τετραγωνική Ρίζα του Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error – RMSE) μετρά το μέγεθος του σφάλματος δίνοντας μεγαλύτερη βαρύτητα στις μεγαλύτερες αποκλίσεις λόγω του τετραγωνισμού των διαφορών, γεγονός που το καθιστά ιδιαίτερα ευαίσθητο στην παρουσία ακραίων τιμών (outliers) ή ξαφνικών κυκλοφοριακών διακυμάνσεων στο αστικό δίκτυο.

$$RMSE = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right)}$$

Οι παραπάνω δείκτες υπολογίστηκαν τόσο στο σύνολο των δεδομένων, όσο και σε ξεχωριστά σύνολα ελέγχου (test sets) για ορθή σύγκριση με τα μοντέλα Μηχανικής Μάθησης.

3.5.3. Σύγκριση και επιλογή μοντέλων

Ο έλεγχος F (F-test) ελέγχει αν ένα μοντέλο με επιπλέον q ερμηνευτικές μεταβλητές εξηγεί σημαντικά περισσότερο από ένα εμφωλευμένο, απλούστερο μοντέλο. Η μηδενική υπόθεση είναι ότι οι συντελεστές των επιπλέον μεταβλητών είναι ίσοι με μηδέν. Η στατιστική ελέγχου ορίζεται ως:

$$F = [(SSE_1 - SSE_2)/q] \,/\, [SSE_2/(n - p)]$$

όπου SSE_1 και SSE_2 τα αθροίσματα τετραγώνων των υπολοίπων του μικρού και του μεγάλου μοντέλου αντίστοιχα. Υπό την H_0 ακολουθεί την κατανομή F με (q, n − p) βαθμούς ελευθερίας, και μεγάλες τιμές F, δηλαδή μικρά p-values, υποδεικνύουν ότι οι νέες μεταβλητές βελτιώνουν σημαντικά την προσαρμογή.

Το κριτήριο πληροφορίας του Akaike (Akaike Information Criterion – AIC) εκτιμά πόσο καλά θα προέβλεπε ένα μοντέλο νέα δεδομένα, όχι απλώς πόσο καλά προσαρμόζεται στα υπάρχοντα. Ισορροπεί δύο αντίρροπες δυνάμεις: την ποιότητα της προσαρμογής και την πολυπλοκότητα του μοντέλου.

$$AIC = 2k - 2 \cdot ln(L)$$

όπου k το πλήθος των παραμέτρων και L η μέγιστη πιθανοφάνεια του μοντέλου. Όσο μικρότερο είναι το AIC, τόσο καλύτερο το μοντέλο. Κρίσιμη είναι η διαφορά ΔAIC μεταξύ μοντέλων πάνω στα ίδια δεδομένα και εξαρτημένη μεταβλητή. Για ΔAIC < 2 τα μοντέλα είναι ουσιαστικά ισοδύναμα ενώ για ΔAIC > 10, το μοντέλο έχει ισχυρή βελτίωση.

3.5.4. Έλεγχοι καταλοίπων

Οι παρακάτω δείκτες εξετάζουν αν τα σφάλματα (κατάλοιπα) του μοντέλου ακολουθούν την κανονική κατανομή, κάτι που είναι απαραίτητο για την εγκυρότητα των p-values και των διαστημάτων εμπιστοσύνης.

1) Ασυμμετρία / Λοξότητα (Skew):
   Μετρά τη συμμετρία της κατανομής των καταλοίπων γύρω από τον μέσο όρο τους. Για μια τέλεια κανονική κατανομή, η τιμή είναι 0. Θετική τιμή σημαίνει «ουρά» προς τα δεξιά, ενώ αρνητική σημαίνει «ουρά» προς τα αριστερά.
2) Κύρτωση (Kurtosis)
   Μετρά το πόσο «αιχμηρή» ή «πλατιά» είναι η κατανομή. Για την κανονική κατανομή, η τιμή της κύρτωσης είναι 3. Τιμές μεγαλύτερες του 3 (λεπτοκυρτωτική) δείχνουν την ύπαρξη ακραίων τιμών (outliers).
3) Prob (Omnibus)
   Είναι η πιθανότητα (p-value) του Omnibus τεστ, το οποίο συνδυάζει τη λοξότητα και την κύρτωση για να ελέγξει αν τα κατάλοιπα είναι κανονικά κατανεμημένα. Αν p > 0.05 δεχόμαστε την κανονικότητα ενώ αν p < 0.05, η παραδοχή της κανονικότητας παραβιάζεται.

4) Prob (JB) (Jarque-Bera Test)
   Παρόμοιο με το Omnibus, το τεστ Jarque-Bera ελέγχει αν η λοξότητα και η κύρτωση των καταλοίπων ταιριάζουν με αυτές της κανονικής κατανομής.

3.5.5. Έλεγχος αυτοσυσχέτισης και πολυσυγγραμικότητας

1) Durbin – Watson:
   Ελέγχει για την ύπαρξη αυτοσυσχέτισης πρώτης τάξης στα κατάλοιπα. Οι τιμές που λαμβάνει είναι από 0 έως 4, με τιμές κοντά στο 2 να σημαίνουν απουσία αυτοσυσχέτισης και τιμές κοντά στο 0 και 4, θετική και αρνητική αυτοσυσχέτιση αντίστοιχα.
2) Δείκτης Κατάστασης (Cond. No):
   Δείκτης διάγνωσης της πολυσυγγραμικότητας (σχετίζεται στενά με το VIF). Μετρά την ευαισθησία των εκτιμήσεων του μοντέλου σε μικρές αλλαγές στα δεδομένα. Τιμές κάτω από 20 θεωρούνται καλές, μεταξύ 20 και 30 δείχνουν μέτριο πρόβλημα, ενώ τιμές άνω του 30 υποδηλώνουν σοβαρή πολυσυγγραμικότητα.

3.6. Μοντέλα Μηχανικής Μάθησης

3.6.1. Γενικά

Τα μοντέλα μηχανικής μάθησης, ιδίως τα μη-παραμετρικά (όπως τα δέντρα αποφάσεων και τα ensemble μοντέλα), δεν υποθέτουν εκ των προτέρων τη μορφή της σχέσης μεταξύ εξαρτημένης και ανεξάρτητης μεταβλητής όπως κάνουν τα παραδοσιακά μοντέλα παλινδρόμησης. Αντιθέτως, μαθαίνουν τη σχέση απευθείας από τα δεδομένα, οπότε μπορούν να συλλάβουν:

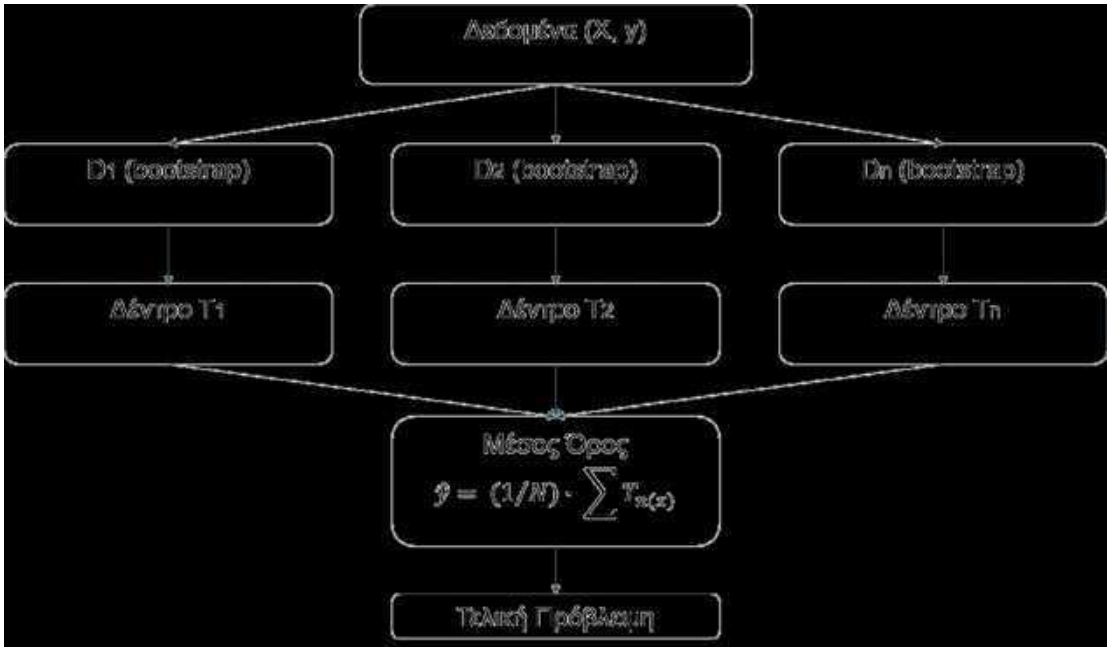
- Μη γραμμικές σχέσεις (π.χ. η επίδραση της βροχής στην ταχύτητα μπορεί να μην είναι ανάλογη της έντασής της).
- Αλληλεπιδράσεις μεταξύ μεταβλητών (π.χ. η βροχή μπορεί να επηρεάζει διαφορετικά την κυκλοφορία σε ώρα αιχμής απ' ό,τι εκτός αιχμής).

Οι τρείς κατηγορίες των μοντέλων μηχανικής μάθησης είναι τα γραμμικά μοντέλα, τα δέντρα αποφάσεων και οι ensemble μέθοδοι. Στην ανάλυση της παρούσας διπλωματικής επιλέχθηκαν οι μέθοδοι: Ενίσχυση Κλίσης (Gradient Boosting – GB) και τα Τυχαία Δάση (Random Forests – RF) καθώς κρίθηκαν οι πιο κατάλληλες με βάση τη βιβλιογραφική ανασκόπηση.

3.6.2. Τυχαία Δάση (Random Forests – RF)

Τα Τυχαία Δάση (Random Forests – RF) είναι ένας ensemble αλγόριθμος μηχανικής μάθησης που ανήκει στην κατηγορία bagging (Bootstrap Aggregating). Η βασική ιδέα είναι η κατασκευή πολλών ανεξάρτητων δέντρων αποφάσεων, το καθένα εκπαιδευμένο σε διαφορετικό δείγμα (bootstrap) των αρχικών δεδομένων (τυχαία επιλογή με επανατοποθέτηση), και η τελική πρόβλεψη να προκύπτει από τον συνδυασμό των επιμέρους δέντρων. Επιπλέον, το RF εισάγει μια δεύτερη πηγή τυχαιότητας πέρα από το bootstrapping: σε κάθε κόμβο διαχωρισμού, εξετάζεται μόνο ένα τυχαίο υποσύνολο μεταβλητών αντί για όλες τις διαθέσιμες, γεγονός που

αποσυσχετίζει τα δέντρα μεταξύ τους και μειώνει τη συνολική διακύμανση του μοντέλου (Breiman, 2001) .



Εικόνα 1: Σχηματική αναπαράσταση της διαδικασίας bagging

3.6.3. Αλγόριθμοι Ενίσχυσης Κλίσης (Gradient Boosting Machines)

Τα μοντέλα Gradient Boosting αποτελούν μια οικογένεια ensemble αλγορίθμων μηχανικής μάθησης που βασίζονται στη διαδοχική συνένωση αδύναμων μαθητών, συνήθως δέντρων αποφάσεων μικρού βάθους. Σε αντίθεση με μεθόδους bagging όπως το Random Forest, όπου τα δέντρα εκπαιδεύονται ανεξάρτητα και ο τελικός στόχος προκύπτει από τον μέσο όρο τους, στο gradient boosting κάθε νέο δέντρο εκπαιδεύεται ώστε να διορθώσει τα σφάλματα των προηγούμενων δέντρων, ελαχιστοποιώντας σταδιακά μια συνάρτηση απώλειας μέσω μιας διαδικασίας ανάλογης της gradient descent. Η τελική πρόβλεψη προκύπτει ως σταθμισμένο άθροισμα των επιμέρους δέντρων, με τον ρυθμό μάθησης να ελέγχει τη συνεισφορά κάθε νέου δέντρου, ώστε να επιτυγχάνεται ισορροπία μεταξύ ακρίβειας και κινδύνου υπερπροσαρμογής. Η προσέγγιση αυτή επιτρέπει στο μοντέλο να συλλαμβάνει σύνθετες, μη γραμμικές σχέσεις και αλληλεπιδράσεις μεταξύ των ανεξάρτητων μεταβλητών, χωρίς να απαιτείται εκ των προτέρων υπόθεση για τη μορφή της συνάρτησης, γεγονός που το καθιστά ιδιαίτερα αποτελεσματικό σε προβλήματα παλινδρόμησης με ετερογενή, κατηγορικά ή αλληλεπιδρώντα δεδομένα, όπως αυτά της κυκλοφοριακής ανάλυσης (Friedman, 2001) . Η διαδικασία μπορεί να αποτυπωθεί από τις παρακάτω εξισώσεις:

1) Προσθετικό Μοντέλο:

$$\hat{y}_i = \sum_{m=1}^{M} \nu \cdot f_m(x_i)$$

Όπου

f_m: το m-οστό δέντρο

M: το συνολικό πλήθος επαναλήψεων

v: ο ρυθμός μάθησης

2) Ελαχιστοποίηση συνάρτησης απώλειας:

$$f_m = arg\min_f \sum_{i=1}^{n} L(y_i,\ F_{m-1}(x_i) + f(x_i))$$

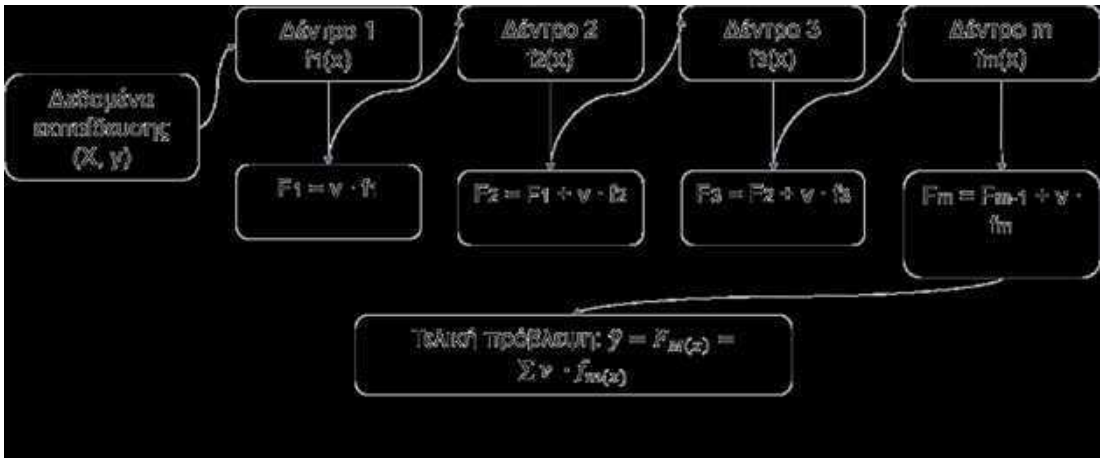
Όπου

F_{m-1}: το μοντέλο μετά από m – 1 βήματα

L: η συνάρτηση απώλειας

3) Ενημέρωση μοντέλου:

$$F_m(x) = F_{m-1}(x) + \nu \cdot f_m(x)$$



Εικόνα 2: Σχηματική αναπαράσταση της διαδοχικής διαδικασίας Gradient Boosting

Μία διαφορετική υλοποίηση που χρησιμοποιείται στην παρούσα εργασία είναι ο HistGradientBoostingRegressor της βιβλιοθήκης scikit-learn, μια παραλλαγή του κλασικού gradient boosting που βασίζεται στην αρχή του histogram-based learning: αντί να εξετάζει όλες τις πιθανές τιμές μιας μεταβλητής για την εύρεση του βέλτιστου σημείου διαχωρισμού σε κάθε κόμβο, ομαδοποιεί τις τιμές σε διακριτά ιστογράμματα (bins) πριν την εκπαίδευση, μειώνοντας δραστικά τον υπολογιστικό φόρτο και επιτρέποντας αποδοτική εκπαίδευση σε μεγάλου όγκου βάσεις δεδομένων, όπως αυτό της παρούσας μελέτης. Πέραν της υπολογιστικής αποδοτικότητας, το HGB προσφέρει εγγενή υποστήριξη

κατηγορικών μεταβλητών, επιτρέποντας τον απευθείας χειρισμό μεταβλητών όπως η ώρα, ο μήνας, η ημέρα της εβδομάδας και οι κατηγορίες καιρικών συνθηκών, χωρίς την ανάγκη κωδικοποίησης τύπου one-hot, η οποία σε δεδομένα με πολλές διακριτές κατηγορίες θα οδηγούσε σε σημαντική αύξηση της διαστασιμότητας και ενδεχόμενη απώλεια πληροφορίας για τη φυσική τους σχέση.

Έτσι, το HGB συνδυάζει τα θεωρητικά πλεονεκτήματα του gradient boosting με πρακτικά χαρακτηριστικά που το καθιστούν κατάλληλο για την ανάλυση κυκλοφοριακών δεδομένων της παρούσας διπλωματικής, όπου οι ανεξάρτητες μεταβλητές ορίστηκαν σε μεγάλο βαθμό κατηγορικές και η κλίμακα των δεδομένων απαιτεί αποδοτικές υπολογιστικά μεθόδους.

3.7. Αξιολόγηση μοντέλων μηχανικής μάθησης

Για την ποσοτική αξιολόγηση της προγνωστικής ικανότητας των μοντέλων Μηχανικής Μάθησης στο σύνολο ελέγχου (test set) χρησιμοποιήθηκαν οι τρεις προαναφερόμενοι δείκτες που χρησιμοποιήθηκαν και στα μοντέλα παλινδρόμησης: ο συντελεστής προσδιορισμού R^2, το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error – MAE) και η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error – RMSE).

3.8. Σημαντικότητα μεταβλητών

Για τη διερεύνηση της σημαντικότητας των επιλεγμένων παραγόντων στην κυκλοφορία και τη συμφόρηση, εφαρμόστηκε η μέθοδος της σημαντικότητας μέσω μεταθέσεων (Permutation Importance). Η συγκεκριμένη τεχνική αποτελεί μια αξιόπιστη προσέγγιση η οποία υπολογίζει τη μείωση της απόδοσης του μοντέλου (εν προκειμένω τη μείωση του R^2) όταν οι τιμές μιας συγκεκριμένης μεταβλητής ανακατεύονται τυχαία (shuffled), καταστρέφοντας τη σχέση της με την εξαρτημένη μεταβλητή. Λόγω της φύσης του μοντέλου HistGradientBoostingRegressor και της χρήσης κατηγορικών μεταβλητών, η μέθοδος αυτή κρίνεται ανώτερη από τις κλασικές προσεγγίσεις: σημαντικότητα Gini (Gini importance) και Μέση Μείωση της Ανομοιογένειας (Mean Decrease in Impurity – MDI), καθώς δεν εισάγει μεροληψία υπέρ μεταβλητών με υψηλό αριθμό ψευδομεταβλητών. Η διαδικασία υπολογισμού της σημαντικότητας περιγράφεται παρακάτω:

Έστω ένα εκπαιδευμένο μοντέλο M, ένα σύνολο δεδομένων ελέγχου D και μια μετρική απόδοσης L (στην περίπτωσή μας ο δείκτης R^2). Η σημαντικότητα I για μια ανεξάρτητη μεταβλητή (feature) j υπολογίζεται ως εξής:

1) Υπολογίζεται η αρχική απόδοση του μοντέλου στα δεδομένα: $L^{orig} = L(M, D)$.

2) Για κάθε επανάληψη k:
   - Δημιουργείται ένα νέο τροποποιημένο σύνολο δεδομένων $D_{j,k}^{perm}$, όπου οι τιμές της μεταβλητής j ανακατεύονται (μετατίθενται) τυχαία, ενώ όλες οι υπόλοιπες μεταβλητές παραμένουν σταθερές.
   - Υπολογίζεται η νέα απόδοση του μοντέλου: $L_{j,k}^{perm} = L(M, D_{j,k}^{perm})$

3) Η σημασία της μεταβλητής j για τη συγκεκριμένη επανάληψη ορίζεται ως η διαφορά:

$$I_{j,k} = L^{orig} - L_{j,k}^{perm}$$

4) Η μέση τιμή σημαντικότητας (Importance Mean) που παρουσιάζει το μοντέλο προκύπτει από τον μαθηματικό μέσο όρο των K επαναλήψεων:

$$\bar{I}_j = \frac{1}{K}\sum_{k=1}^{K} I_{j,k}$$

Για την ευκολότερη ερμηνεία και σύγκριση των αποτελεσμάτων, οι απόλυτες τιμές της μέσης μείωσης της απόδοσης μετατράπηκαν σε σχετική επί τοις εκατό (%) σημαντικότητα. Αρχικά, τυχόν οριακά αρνητικές τιμές σημαντικότητας μηδενίστηκαν, καθώς υποδηλώνουν μεταβλητές χωρίς καμία προγνωστική ισχύ: $I_j^{clipped} = \max(0, \bar{\bar{I}}_j)$

Στη συνέχεια, η σχετική συνεισφορά κάθε μεταβλητής προέκυψε από τον λόγο της ατομικής της τιμής σημαντικότητας προς το άθροισμα των τιμών σημαντικότητας όλων των εξεταζόμενων παραγόντων.

$$Σχετική\ Σημαντικότητα\ (\%) = \left(\frac{I_j^{clipped}}{\sum_{m=1}^{M} I_m^{clipped}}\right) \times 100$$

4. Συλλογή και Επεξεργασία Στοιχείων

4.1. Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζονται τα δεδομένα που αξιοποιήθηκαν για τη συγγραφή της Διπλωματικής Εργασίας. Σκοπός είναι η ανάδειξη των πηγών των δεδομένων και της διαδικασίας συλλογής και ανάλυσης αυτών έτσι ώστε να ελεγχθεί η αξιοπιστία και η ποιότητά τους. Επιπλέον, πραγματοποιείται μία προκαταρκτική στατιστική ανάλυση με σκοπό την καλύτερη κατανόηση των δεδομένων και την εξαγωγή αρχικών συμπερασμάτων.

4.2. Κοινωνικοοικονομικά δεδομένα

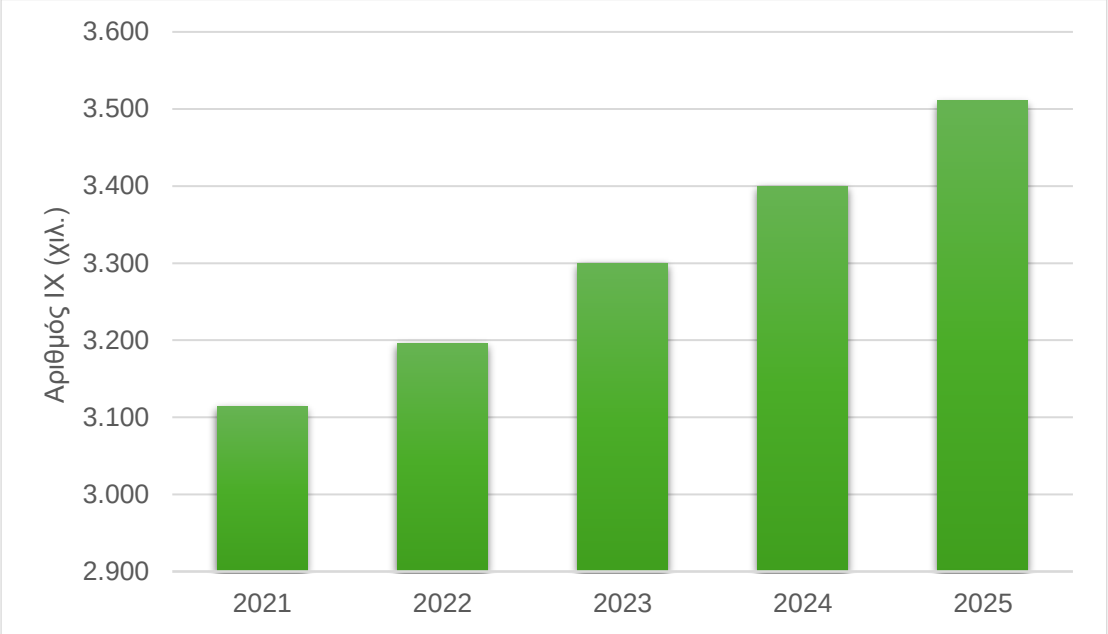
Τα κοινωνικοοικονομικά δεδομένα συλλέχθηκαν από την Ελληνική Στατιστική Αρχή – ΕΛΣΤΑΤ, το Υπουργείο Ανάπτυξης και Επενδύσεων και το Ινστιτούτο του Συνδέσμου Ελληνικών Τουριστικών Επιχειρήσεων – ΙΝΣΕΤΕ για το χρονικό διάστημα 2021 – 2026 με σκοπό τη διερεύνηση επιρροής αυτών στην αστική συμφόρηση (Πίνακας 4.1).

Δεδομένα	Πηγή	Περιγραφή
Εργαζόμενοι	ΕΛΣΤΑΤ	Αριθμός εργαζομένων στην περιφέρεια Αττικής ανά τρίμηνο το διάστημα 01/2021 – 06/2026
Ποσοστό εργατικού δυναμικού	ΕΛΣΤΑΤ	Ποσοστό εργαζομένων στην περιφέρεια Αττικής ανά τρίμηνο το διάστημα 01/2021 – 06/2026
Ποσοστό ανεργίας	ΕΛΣΤΑΤ	Ποσοστό ανεργίας στη περιφέρεια Αττικής ανά τρίμηνο το διάστημα 01/2021 – 06/2026
Νέες οικοδομές	ΕΛΣΤΑΤ	Νέες οικοδομές στην Αττική ανά μήνα
Αριθμός ΙΧ οχημάτων	ΕΛΣΤΑΤ	Αριθμός ΙΧ που βρίσκονται σε κυκλοφορία στην περιφέρεια Αττικής ανά έτος το διάστημα 2021 - 2025
Εκτιμώμενος πληθυσμός	ΕΛΣΤΑΤ	Εκτιμώμενος πληθυσμός της περιφέρειας Αττικής τον Ιανουάριο των ετών: 2021 έως 2025
Τιμή αμόλυβδης 95	Υπουργείο Ανάπτυξης και Επενδύσεων	Μέση μηνιαία τιμή αμόλυβδης 95 οκτανίων (€/lt) στην περιφέρεια Αττικής το διάστημα 01/2021 – 06/2026
Αεροπορικές αφίξεις	ΙΝΣΕΤΕ	Αεροπορικές αφίξεις στο αεροδρόμιο της Αθήνας από το εξωτερικό ανά μήνα το διάστημα 01/2021 – 06/2026
Αριθμός τουριστών	ΙΝΣΕΤΕ	Συνολικός αριθμός τουριστών στην Αττική ανά έτος το διάστημα 2021 - 2025

Πίνακας 4.1: Κοινωνικοοικονομικά δεδομένα

Μέσα από μια προκαταρκτική ανάλυση των δεδομένων, προκύπτουν σημαντικά συμπεράσματα που αναδεικνύουν τη μεταβολή των κοινωνικών και οικονομικών συνθηκών της Αττικής. Παρά τη δημογραφική μείωση του μόνιμου πληθυσμού κατά 73,900 άτομα, η αγορά εργασίας επέδειξε ισχυρή δυναμική, καταγράφοντας αύξηση του εργατικού δυναμικού κατά 201,400 απασχολούμενους την περίοδο Α΄ τρίμηνο 2021 έως Α΄ τρίμηνο 2025. Η αναπτυξιακή τροχιά ενισχύεται από τη δραστηριότητα του κατασκευαστικού κλάδου, καθώς την ίδια περίοδο εκδόθηκαν άδειες για 11,094 νέες οικοδομές. Στον τομέα των μεταφορών, καταγράφηκα αύξηση του στόλου των ιδιωτικής χρήσεως (ΙΧ) οχημάτων κατά 396,509 μονάδες μέσα σε μία πενταετία,

αγγίζοντας το 2025 τα 3.5 εκατομμύρια ΙΧ, μέγεθος που μεταφράζεται στον υψηλό δείκτη πυκνότητας των 0.93 ΙΧ ανά κάτοικο και 2.18 ΙΧ ανά εργαζόμενο.



Διάγραμμα 4-1: Ετήσια μεταβολή του αριθμού ΙΧ στην Αττική

Το πλέον αξιοσημείωτο εύρημα της περιόδου αποτελεί η εκθετική άνοδος του τουριστικού προϊόντος. Οι ετήσιες τουριστικές αφίξεις στην Αττική αυξήθηκαν κατά 7 εκατομμύρια ανθρώπους συγκριτικά με το 2021, με το έτος 2025 να αποτελεί χρονιά αιχμής, καταγράφοντας τον ιστορικό αριθμό των 9.7 εκατομμυρίων συνολικών ετήσιων επισκεπτών. Με την υπόθεση ότι η μηνιαία κατανομή των επισκεπτών ακολουθεί την μηνιαία κατανομή των αεροπορικών αφίξεων, προκύπτει αριθμός τουριστών στην Αθήνα πάνω από 5.6 εκατομμύρια στο χρονικό διάστημα των μηνών: Μάιος – Σεπτέμβριος.

Για την διερεύνηση επιρροής των κοινωνικοοικονομικών συνθηκών στη συμφόρηση του οδικού δικτύου, αξιοποιήθηκαν μόνο οι τιμές των ετών 2022 έως 2025 καθώς τα δεδομένα μέσης ταχύτητας από την Google δεν περιλαμβάνουν το έτος 2021, όπως αναλύεται στην Ενότητα 4.5, και ορισμένα κοινωνικοοικονομικά το έτος 2026.

Οι μηνιαίες τιμές μετατράπηκαν σε τριμηνιαίες είτε ως μέσος όρος των τριών καταγραφών (τιμή καυσίμου) είτε ώς άθροισμα (αριθμός τουριστών). Συγκεκριμένα, ο αριθμός των τουριστών στην Αττική ανά τρίμηνο υπολογίστηκε ως το άθροισμα των μηνιαίων τιμών που υπολογίστηκαν ανάλογα με τις μηνιαίες διεθνείς αεροπορικές αφίξεις και τον αριθμό των συνολικών ετήσιων επισκεπτών. Τέλος, για την αναγωγή των ετήσιων δεδομένων του πληθυσμού και του αριθμού των ΙΧ σε τριμηνιαία, έγινε υπόθεση γραμμικής μεταβολής. Η βάση των κοινωνικοοικονομικών δεδομένων διαμορφώθηκε με τον τρόπο που φαίνεται στον Πίνακα 5.1.

4.3. Μετεωρολογικά δεδομένα

Τα μετεωρολογικά δεδομένα συλλέχθηκαν έπειτα από αίτηση προς το Εθνικό Αστεροσκοπείο Αθηνών (ΕΑΑ). Η καταγραφή αυτών πραγματοποιείται από τον ιστορικό σταθμό του ΕΑΑ στο Θησείο (Λόφος Νυμφών) και οι μετρήσεις που συλλέχθηκαν είναι οι εξής:

Μετεωρολογικά δεδομένα	Συχνότητα καταγραφής
Θερμοκρασία (º C)	Ωριαίες μετρήσεις κάθε ημέρα των ετών 2021 – 2025
Σχετική Υγρασία (%)	Ωριαίες μετρήσεις κάθε ημέρα των ετών 2021 – 2025
Κατακρήμνιση (mm)	Ωριαίες μετρήσεις κάθε ημέρα των ετών 2021 – 2025
Ταχύτητα (m/s) και κατεύθυνση ανέμου	Ωριαίες μετρήσεις κάθε ημέρα των ετών 2021 – 2025
Νέφωση (Octas)	Τις ώρες 8:00, 11:00 και 14:00 κάθε ημέρα εκτός Σαββάτου
Ορατότητα (1-10)	Την ώρα 14:00 κάθε ημέρα εκτός Σαββάτου (2021 – 2025)

Πίνακας 4.2: Μετεωρολογικά δεδομένα

Στη βάση δεδομένων παρατηρήθηκε ένα κενό στις τιμές του ανέμου τις ημερομηνίες: 15/02/2023 στις 08:00 και 09:00, και 19/03/2023 στις 18:00.

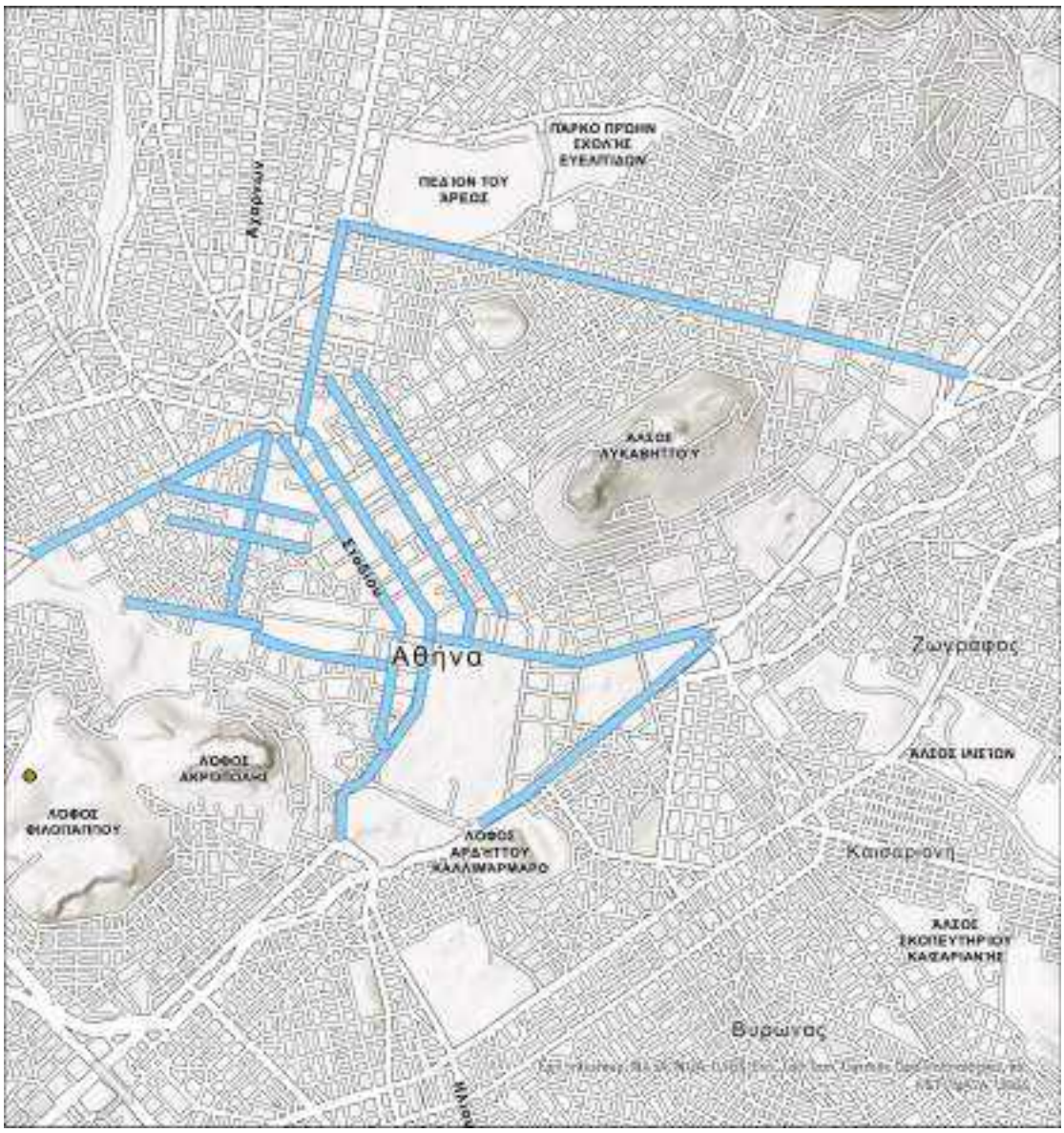
4.4. Δεδομένα σύνθεσης κυκλοφορίας

Η συλλογή δεδομένων σύνθεσης κυκλοφορίας πραγματοποιείται μία φορά τον μήνα σε 16 οδικά τμήματα στο κέντρο της Αθήνας κατά τα χρονικά διαστήματα: 8:00-10:00, 8:00 – 11:00 και 11:00-15:00 για τα έτη 2021 έως 2026. Όπως φαίνεται στον Πίνακα 4.3, εφτά οδικά τμήματα έχουν μετρήσεις προς δύο κατευθύνσεις ενώ τα υπόλοιπα εννιά, μόνο προς μία κατεύθυνση. Συνολικά, στα τέσσερα έτη πραγματοποιούνται σαράντα οχτώ (48) ημερήσιες καταγραφές.

	Οδός	Σημείο Μέτρησης	Κατεύθυνση
1	Αθηνάς	Βορέου Βορέου	Μοναστηράκι Ομόνοια
2	Ακαδημίας	Ασκληπιού	-
3	Αλεξάνδρας	Βουρνάζου Βουρνάζου	Κηφισιάς Πατησίων
4	Βασ. Αμαλίας	Όθωνος	-
5	Βασ. Κων/νου	Ριζάρη Μερκούρη	Συγγρού Χίλτον
6	Βασ. Σοφίας	Ακαδημίας Ακαδημίας	Ευαγγελισμός Σύνταγμα
7	Ερμού	Πλ. Αβησσυνίας Πλ. Αβησσυνίας	Πλατεία Ασωμάτων Πλατεία Μοναστηρακίου

8	Ευριπίδου	Αθηνάς	-
9	Μητροπόλεως	Μητρόπολη απέναντι Μητρόπολη	-
10	Πανεπιστημίου	Τριλογία	-
11	Πατησίων	Μετσόβου Μετσόβου	Λεωφ. Αλεξάνδρας Πλ.Ομονοίας
12	Πειραιώς	Πλ. Κουμουνδούρου Πλ. Κουμουνδούρου	Ιερά Οδός Πλ.Ομονοίας
13	Σοφοκλέους	Αθηνάς	-
14	Σταδίου	Πλ. Κλαυθμώνος	-
15	Σόλωνος	Ασκληπιού	-
16	Φιλελλήνων	Μητροπόλεως	-

Πίνακας 4.3: Οδικά τμήματα καταγραφής της σύνθεσης κυκλοφορίας



Εικόνα 3: Οδικά τμήματα καταγραφής της σύνθεσης κυκλοφορίας

Οι κατηγορίες οχημάτων που καταγράφονται είναι: επιβατικά, ταξί, μοτοσυκλέτες, τρίκυκλα, φορτηγά, βαριά φορτηγά, λεωφορεία, ποδήλατα και πατίνια. Επιπλέον καταγράφεται η κυκλοφορία πεζών, ποδηλάτων, πατινιών και μοτοσυκλετών στο πεζοδρόμιο. Η καταμέτρηση των οχημάτων γίνεται σε χρονικό διάστημα μισής ώρας και των πεζών στο ένα τέταρτο της ώρας, για τα χρονικά διαστήματα που αναφέρθηκαν και ύστερα ανάγονται σε ωριαίες τιμές φόρτου. Καθώς παρατηρήθηκαν μερικές μη-λογικές τιμές, πραγματοποιήθηκε αρχικά καθαρισμός των δεδομένων αυτών.

4.4.1. Καθαρισμός δεδομένων

Ο καθαρισμός δεδομένων (Data cleaning) είναι η διαδικασία εντοπισμού και διόρθωσης σφαλμάτων και ασυνεπειών σε ανεπεξέργαστα δεδομένα με στόχο τη βελτίωση της ποιότητας αυτών. Αρχικά πραγματοποιήθηκε ανίχνευση ακραίων τιμών (outlier detection) με τη μέθοδο «τιμών Z». Αυτή η μέθοδος υπολογίζει την απόσταση μιας τιμής χ από τον μέσο όρο μ σε μονάδες τυπικής απόκλισης σ, δηλαδή:

$$Z = \frac{\chi - \mu}{\sigma}$$

Εντοπίστηκαν 104 ακραίες τιμές στις υπό μελέτη οδούς με δέκα Z ≥ 5, δέκα οχτώ Z με: 4 ≤ Z ≤ 5 και εβδομήντα έξι Z με: 3 ≤ Z ≤ 4. Όλες αυτές οι τιμές θα διορθωθούν σύμφωνα με την εξίσωση:

$$\text{νέα τιμή} = \mu \pm 3 \times \sigma$$

Οδός	Ημερομηνία	Τύπος	Τιμή Ωριαίου Φόρτου	Μ.Ο	Απόκλιση	Z
Μητροπόλεως	16/10/2025	Λεωφορεία	144	9	1567.40 %	10.9
Βασ. Σοφίας	22/7/2021	Φορτηγά	642	58	1012.10 %	9.6
Βασ. Αμαλίας	19/1/2023	Λεωφορεία	780	103	654.10%	9.5
Αθηνάς	18/12/2025	Λεωφορεία	170	15	997.80%	8.8
Φιλελλήνων	17/3/2026	Λεωφορεία	354	99	258.10%	6.7
Σόλωνος	22/11/2022	Επιβατικά	4	415	99.00%	5.6
Βασ. Κων/νου	20/4/2021	Φορτηγά	490	117	317.90%	5.6
Ακαδημίας	28/4/2022	Επιβατικά	54	766	93.00%	5.4
Βασ. Αμαλίας	30/6/2026	Φορτηγά	270	59	360.20%	5
Σόλωνος	22/11/2022	Ταξί	602	223	169.70%	5

Πίνακας 4.4: Καθαρισμός Δεδομένων (Τιμές με Z>= 5)

4.4.2. Ανάλυση δεδομένων

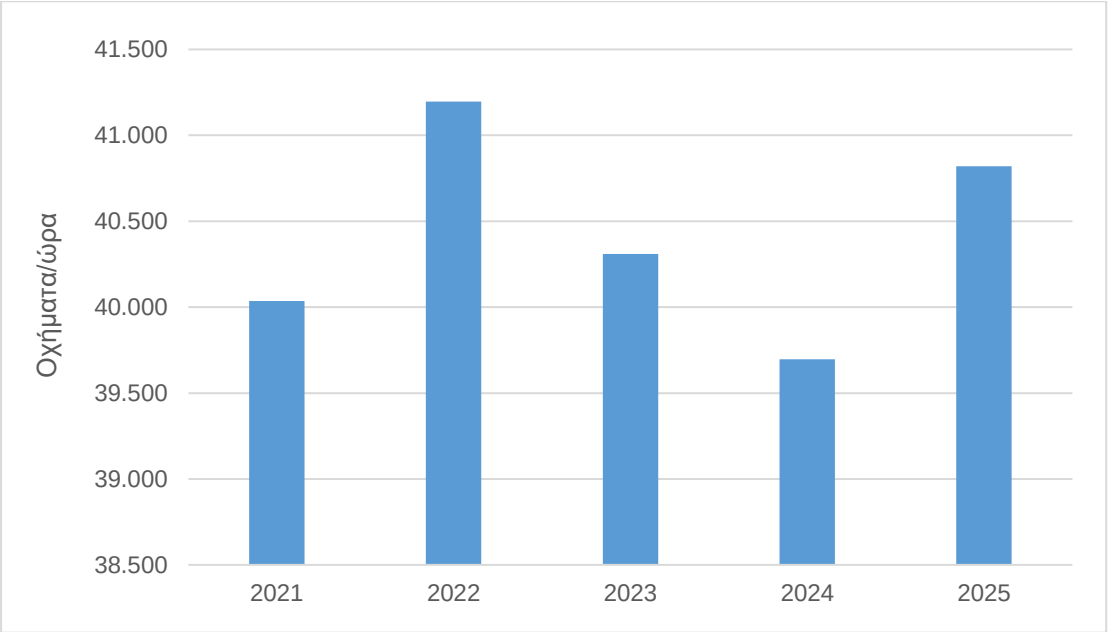
Για την ανάλυση των δεδομένων και την εξαγωγή αρχικών συμπερασμάτων αξιοποιήθηκαν οι καταγραφές των οχημάτων για κάθε έτος (2021 – 2025) το χρονικό διάστημα 8-10:00. Το άθροισμα του Μέσου Ωριαίου Φόρτου οχημάτων στα 23 οδικά τμήματα προκύπτει σύμφωνα με τη σχέση:

$$Μέσος\ Ωριαίος\ Φόρτος = \frac{\sum_{s\in S}\sum_{c\in C}\sum_{d\in D}[F(s,c,d)]}{12}$$

όπου:

- F(s, c, d): ο ωριαίος φόρτος για το οδικό τμήμα s, την κατηγορία οχήματος c, στην ημέρα μέτρησης d, για τη ζώνη ώρας 08:00-10:00
- $S = \{1, 2, \dots, 23\}$: το σύνολο των 23 οδικών τμημάτων
- C ={Επιβατικα, Ταξί, Μοτοσυκλέτες, Φορτηγά, Λεωφορεία, Ποδήλατα, Πατίνια}: οι 7 κατηγορίες οχημάτων.
- D_y = {1, 2, … , 12}: το σύνολο των ημερών μέτρησης του έτους y (12 καταγραφές ανά έτος)

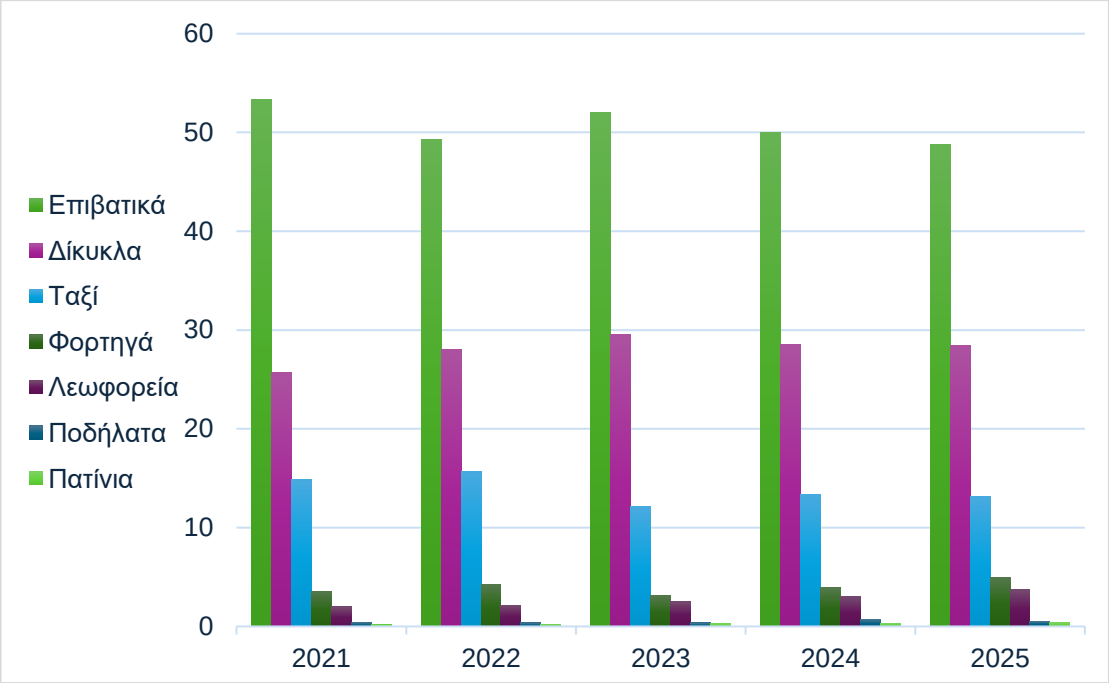
Ουσιαστικά, η τιμή αυτή αντιπροσωπεύει τον ωριαίο αριθμό των οχημάτων που κινούνται συνολικά στα 23 οδικά τμήματα του κέντρου της Αθήνας που αναγράφονται στον Πίνακα 4.3, μία τυπική πρωινή αιχμή του έτους.



Διάγραμμα 4-2: Ετήσια μεταβολή ωριαίου αριθμού οχημάτων στο σύνολο των οδικών τμημάτων το χρονικό διάστημα 8 – 10:00

Σύμφωνα με το Διάγραμμα 4-2, ο μέσος ωριαίος φόρτος οχημάτων κατά την πρωινή αιχμή, δεν έχει σταθερή μεταβολή κατά τα έτη 2021 έως 2025. Το έτος 2022 έχουμε τη μέγιστη τιμή με 41,200 οχήματα/ώρα ενώ το 2024 έχουμε την ελάχιστη με 39,700 οχήματα ανά ώρα, μία μείωση 1,500 οχημάτων (3.6%). Το έτος 2025 έχουμε εκ νέου άνοδο της κυκλοφορίας, στα 40,800 οχήματα/ώρα που αντιστοιχεί σε αύξηση 3% από το 2024. Συνολικά από το έτος 2021, η κυκλοφορία κατά την πρωινή αιχμή (8:00 – 10:00) στο κέντρο της Αθήνας αυξήθηκε κατά 2%.

Διαπιστώθηκε επίσης πως τα επιβατικά οχήματα είναι αυτά που επικρατούν στον αστικό ιστό με ποσοστό περί του 50% του συνόλου της κυκλοφορίας όπως φαίνεται στο Διάγραμμα 4-3.

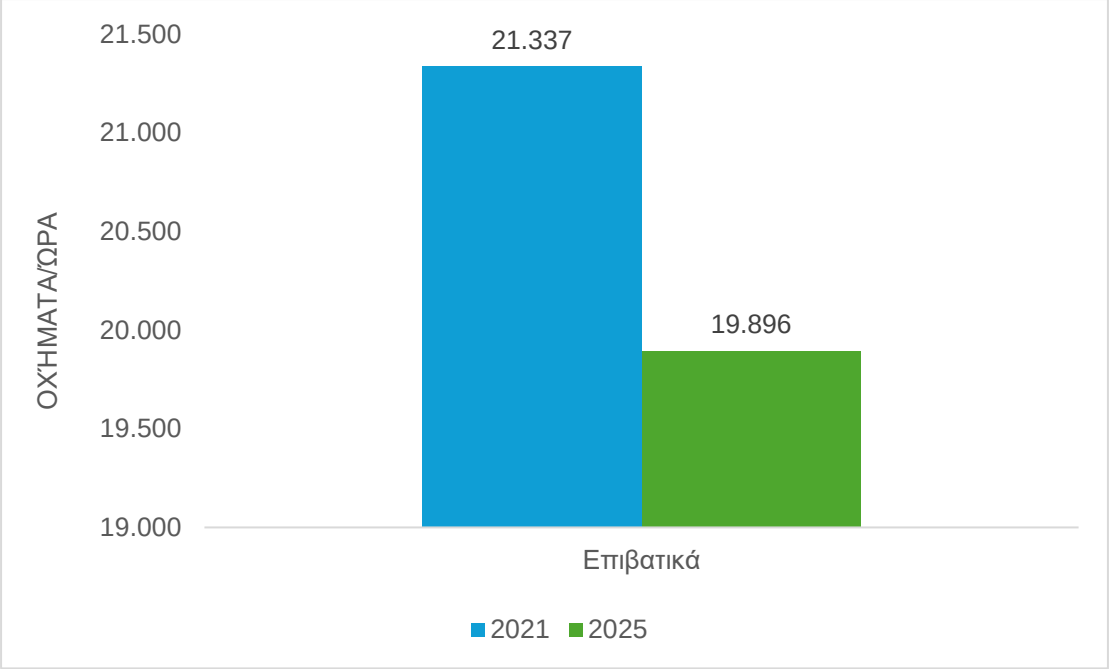


Διάγραμμα 4-3: Ποσοστά οχημάτων στο σύνολο της κυκλοφορίας 2021 - 2025

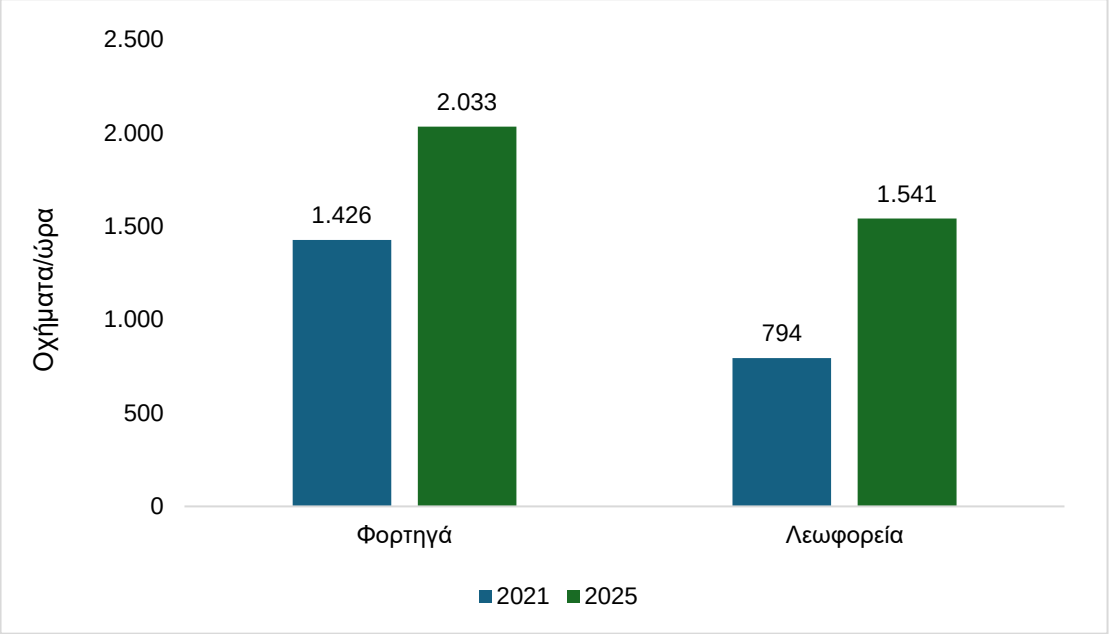
Έτος	Επιβατικά IX	Δίκυκλα	Ταξί	Φορτηγά	Λεωφορεία	Ποδήλατα	Πατίνια
2021	53	26	14.9	3.6	2.0	0.4	0.2
2022	49	28	15.7	4.2	2.2	0.4	0.2
2023	52	30	12.1	3	2.6	0.4	0.3
2024	50	29	13.4	3.9	3.0	0.7	0.3
2025	49	28	13.1	5.0	3.8	0.6	0.4

Πίνακας 4.5: Ετήσια μεταβολή ποσοστών οχημάτων στο σύνολο της κυκλοφορίας 2021 - 2025

Όμως, όπως φαίνεται στα επόμενα διαγράμματα παρατηρείται μία μείωση στον αριθμό των επιβατικών και μία αύξηση στα φορτηγά και τα λεωφορεία στη χρονική περίοδο ανάλυσης (2021 – 2025). Συγκεκριμένα, τα επιβατικά μειώθηκαν κατά 6.8% (21,337 – 19,896 Ι.Χ/ώρα) ενώ τα λεωφορεία αυξήθηκαν κατά 94% (794 – 1,541 λεωφορεία/ώρα) και τα φορτηγά κατά 43% (1,426 – 2,033 φορτηγά/ώρα).

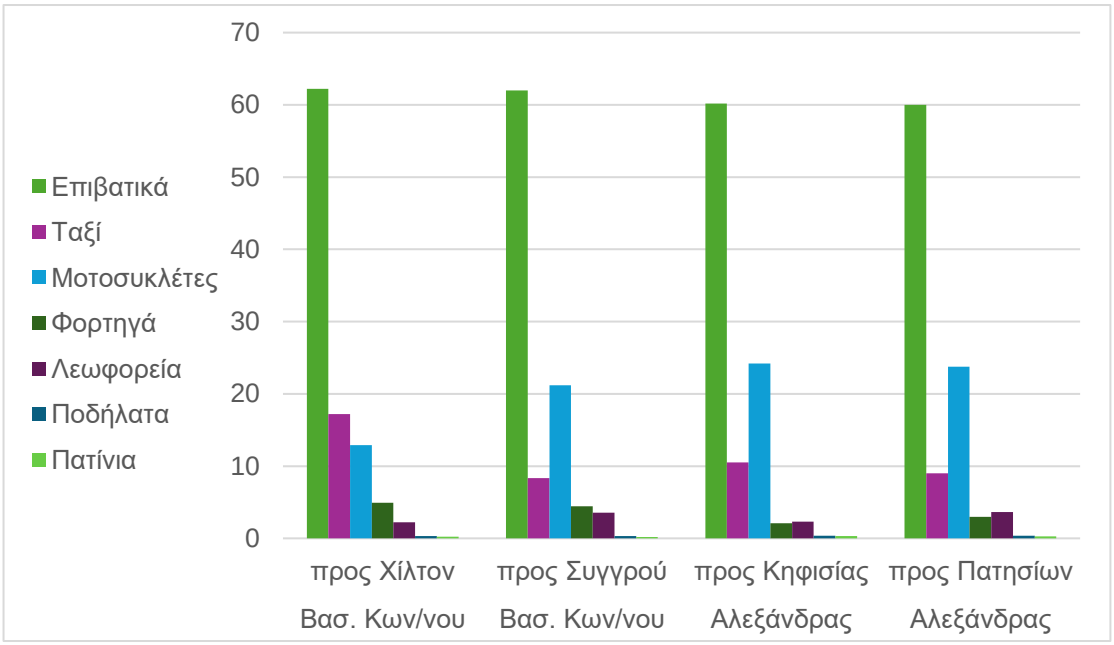


Διάγραμμα 4-4: Μεταβολή φόρτου επιβατικών ΙΧ οχημάτων 2021 - 2025



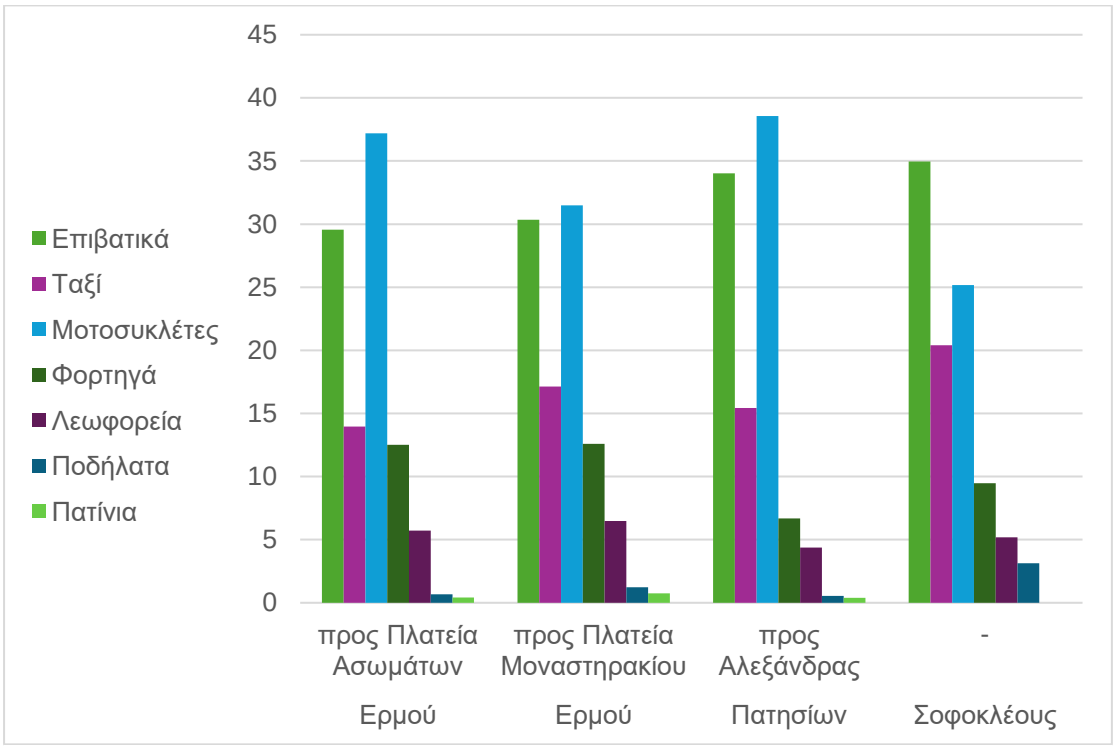
Διάγραμμα 4-5: Μεταβολή φόρτου φορτηγών και λεωφορείων 2021 - 2025

Οι οδοί που εμφανίζουν τη μεγαλύτερη κυκλοφορία επιβατικών ΙΧ οχημάτων είναι οι Βασιλέως Κωνσταντίνου και Αλεξάνδρας με ποσοστά 62 και 60% αντίστοιχα, όπως φαίνεται στο Διάγραμμα 4-6. Η κυκλοφορία λεωφορείων σε αυτά τα τμήματα είναι κάτω από 5%.



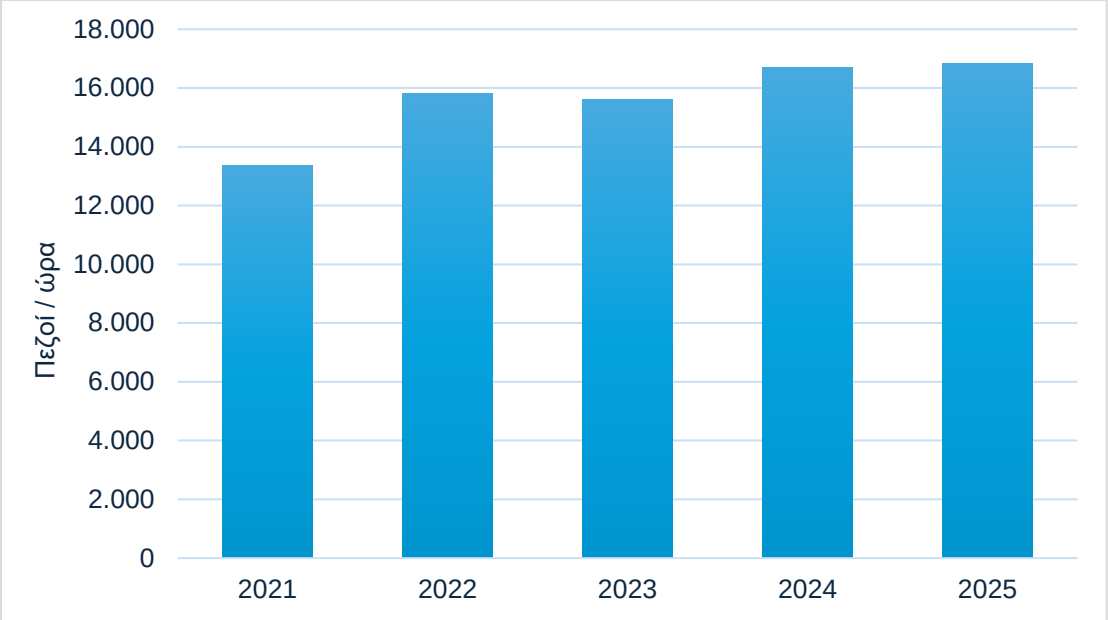
Διάγραμμα 4-6: Σύνθεση κυκλοφορίας οδών με το μεγαλύτερο ποσοστό επιβατικών το 2025

Αντίστοιχα, τα τμήματα με τη μικρότερη κυκλοφορία ΙΧ είναι η Ερμού, η Πατησίων και η Σοφοκλέους με ποσοστά μικρότερα από 35%. Τα οχήματα που αντικαθιστούν τα ΙΧ είναι κυρίως οι μοτοσυκλέτες και τα ταξί.



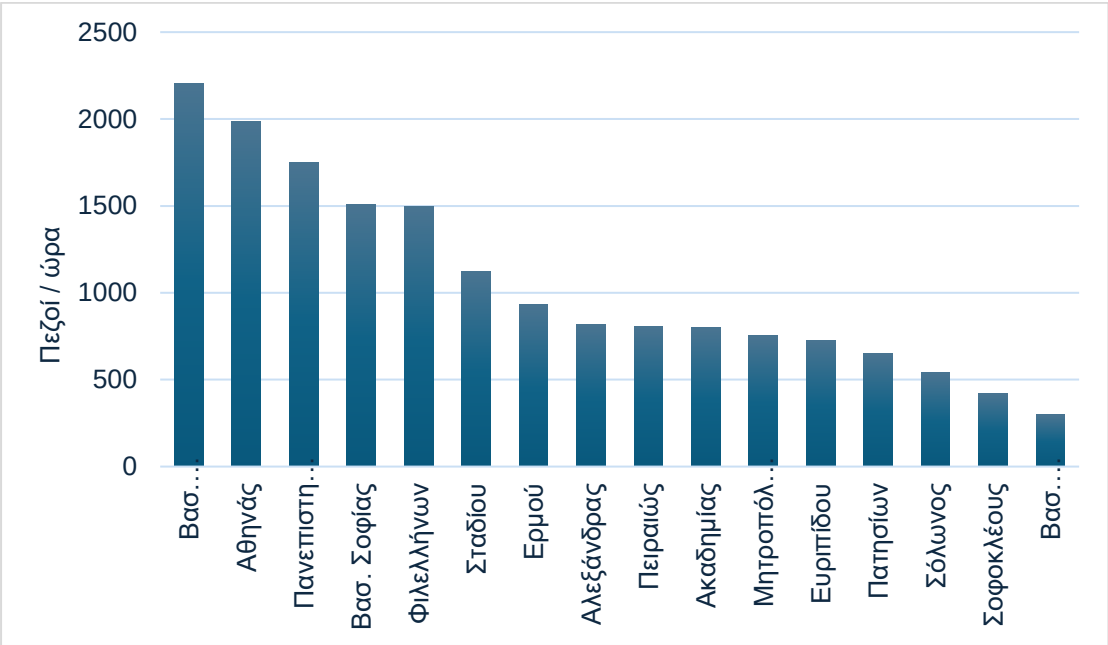
Διάγραμμα 4-7: Σύνθεση κυκλοφορίας οδών με το χαμηλότερο ποσοστό ΙΧ το 2025

Εκτός από την κυκλοφορία των οχημάτων, εξετάσθηκε και η κυκλοφορία των πεζών στα ίδια 16 οδικά τμήματα το χρονικό διάστημα 11:00 – 15:00. Κύριο εύρημα είναι πως οι πεζοί αυξάνονται κατά 3,469 από το έτος 2021 έως το 2025 (αύξηση 26%) με τη τιμή της μέσης ωριαίας κυκλοφορίας στο σύνολο των 16 οδών το 2025 να φτάνει τους 16,821 πεζούς/ώρα.



Διάγραμμα 4-8: Μεταβολή ροής πεζών 2021 – 2025

Οι οδοί με τη μεγαλύτερη κυκλοφορία πεζών είναι οι: Βασ. Αμαλίας, Αθηνάς, Πανεπιστημίου ενώ αυτές με την μικρότερη είναι οι: Σόλωνος, Σοφοκλέους και Βασ. Κωνσταντίνου σύμφωνα με το Διάγραμμα 4-9.



Διάγραμμα 4-9: Κυκλοφορία πεζών ανά οδό το 2025

4.5. Δεδομένα χρόνου διαδρομής και ταχύτητας

Τα δεδομένα χρόνων διαδρομής συλλέχθηκαν από δύο πηγές:

1. από την Google: δεδομένα μέσης ταχύτητας για εξήντα δύο (62) οδικά τμήματα στην Αττική ανά είκοσι λεπτά τα χρονικά διαστήματα που φαίνονται παρακάτω:

Ημερομηνίες	Ημέρες	Ώρες
01/01/2022 – 30/06/2025	Δευτέρα–Παρασκευή	08:00–10:00 & 12:00–19:00
	Σάββατο	12:00–19:00 & 21:00–24:00
	Κυριακή	12:00–19:00
01/07/2025 – 31/05/2026	Δευτέρα–Κυριακή	08:00–10:00 & 12:00–20:00
01/06/2026 – 30/06/2026	Όλες οι ημέρες	Μόνο 00:00
01/07/2026 – 09/07/2026	Όλες οι ημέρες	08:00–10:00 & 12:00–20:00

Πίνακας 4.6: Χρονικά διαστήματα καταγραφής δεδομένων από την Google

2. από την πλατφόρμα Tom Tom: δεδομένα χρόνου διαδρομής, μέσης ταχύτητας, καθυστερήσεων και ουρών για εβδομήντα δύο (72) οδικά τμήματα στην Αττική ανά είκοσι λεπτά τα χρονικά διαστήματα που φαίνονται παρακάτω:

Ημερομηνίες	Ημέρες	Ώρες
21/01/2025 - 31/12/2025	Όλες οι ημέρες	08:00-10:00 & 12:00-20:00

Πίνακας 4.7: Χρονικά διαστήματα καταγραφής δεδομένων από την Tom Tom

4.5.1. Μεταβολή έντασης και διάρκειας της συμφόρησης

Για την διερεύνηση μεταβολής της έντασης και της διάρκειας της συμφόρησης στους οδικούς άξονες της Αττικής, αξιοποιήθηκαν η βάση δεδομένων της Google τα έτη 2022 έως 2025 για τα οποία υπάρχουν πλήρη δεδομένα.

Η τιμής της συνολικής καθυστέρησης είναι ένας δείκτης μέτρησης της έντασης της συμφόρησης ο οποίος είναι απλός στην εφαρμογή, εύκολα ερμηνεύσιμος και ισούται με τη διαφορά του παρατηρούμενου χρόνου διαδρομής από το χρόνο διαδρομής σε συνθήκες ελεύθερης ροής. Ως ταχύτητα ελεύθερης ροής ορίσθηκε το 99^{ο} εκατοστημόριο του συνόλου της βάσης δεδομένων. Η ανάλυση πραγματοποιήθηκε τόσο σε οδικούς άξονες στο κέντρο της Αθήνας (Πίνακας 4.8) όσο και σε περιφερειακούς (Πίνακας 4.9).

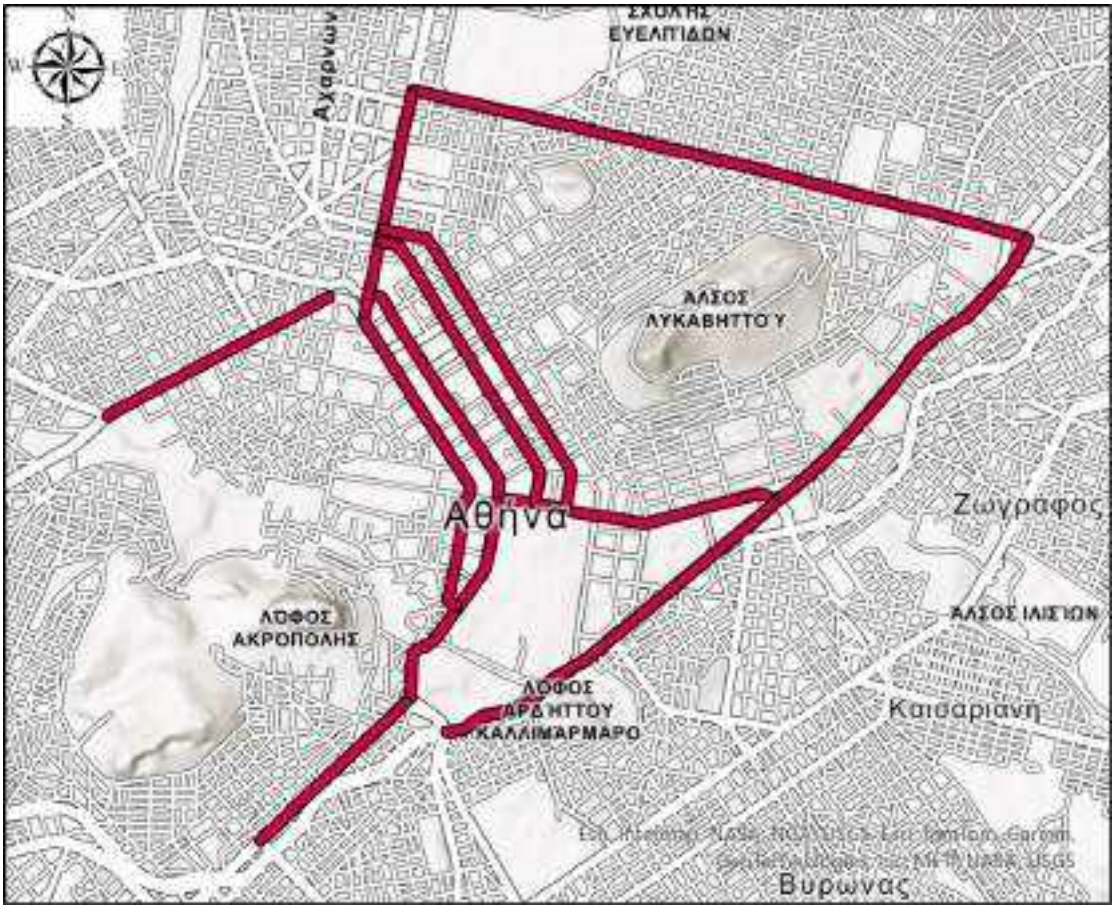
No	Οδικό τμήμα	Μήκος (m)	FFS
1	Βασιλίσσης Σοφίας (Πανεπιστημίου - Βασιλέως Κωνσταντίνου)	1,177	26
	Βασιλίσσης Σοφίας (Βασιλέως Κωνσταντίνου - Πανεπιστημίου)	1,217	26

	Βασιλίσσης Σοφίας (Βασιλέως Κωνσταντίνου - Κηφισίας)	1,808	33
	Βασιλίσσης Σοφίας (Κηφισίας - Βασιλέως Κωνσταντίνου)	1,528	36
2	Λεωφόρος Αλεξάνδρας (Πατησίων - Κηφισίας)	2,669	35
	Λεωφόρος Αλεξάνδρας (Κηφισίας- Πατησίων)	2,673	32
	Λεωφόρος Αλεξάνδρας (Πανόρμου - Χαριλάου Τρικούπη)	1,046	40
	Λεωφόρος Αλεξάνδρας (Χαρ. Τρικούπη- Πανόρμου)	1,024	40
3	Πατησίων (Λεωφόρος Αλεξάνδρας - Σταδίου)	980	30
	Πατησίων (Στούρναρη - Ιουλιανού)	403	26
	Πατησίων (Ιουλιανού - Χαλκοκονδύλη)	585	36
4	Σταδίου (Αιόλου - Βασιλέως Γεωργίου)	893	32
5	Φιλελλήνων (Βασιλέως Γεωργίου - Βασιλίσσης Αμαλίας)	484	27
6	Βασιλίσσης Αμαλίας (Φιλελλήνων - Αθ. Διάκου)	476	27
	Βασιλίσσης Αμαλίας (Αθ. Διάκου - Πανεπιστημίου)	967	35
7	Βασιλέως Κωνσταντίνου (Βασ. Σοφίας - Αρδητού / Αθ. Διάκου)	1,719	33
	Βασιλέως Κωνσταντίνου (Αρδητού / Αθ. Διάκου- Βασ. Σοφίας)	1,734	33
8	Πανεπιστημίου (Βασιλίσσης Σοφίας - Πατησίων)	1,028	35
9	Ακαδημίας (Πατησίων - Βασιλίσσης Σοφίας)	1,345	27
10	Σόλωνος (Βασιλίσσης Σοφίας - Πατησίων)	1,514	24
11	Λεωφ. Συγγρού (Βασιλίσσης Αμαλίας - Φραντζή)	834	30
	Λεωφ. Συγγρού (Φραντζή - Βασιλίσσης Αμαλίας)	817	40
12	Πειραιώς (Κολοκυνθούς - Πλατεία Ομόνοιας)	603	25
	Πειραιώς (Κολοκυνθούς - Ιερό Οδός)	492	37

Πίνακας 4.8: Κεντρικοί οδικοί άξονες ανάλυσης

No	Οδικό τμήμα	Μήκος (m)	FFS
1	Αμφιθέας (Αχιλλέως - Ταξιαρχών)	1,088	48
	Αμφιθέας (Ταξιαρχών - Αχιλλέως)	1,085	43
2	Κηφισός (Ποσειδώνος - Πειραιώς)	961	91
	Κηφισός(Πειραιώς - Ποσειδώνος)	4,997	94
	Κηφισός (Λαχανά - Αττική Οδός)	974	95
	Κηφισός (Αττική Οδός - Λαχανά)	805	97
	Κηφισός (Μούδρου - Αθηνών)	953	93
	Κηφισός (Αθηνών - Μούδρου)	1,010	93
3	Αθηνών (Μαρκόνι - Μητροδώρου)	891	62
	Αθηνών (Μητροδώρου - Μαρκόνι)	900	58
4	Μεσογείων (Κατεχάκη - Κύπρου)	1,003	52
	Μεσογείων (Κύπρου - Κατεχάκη)	963	51
5	Ποσειδώνος (Αλίμου - Αμφιθέας)	1,149	46
	Ποσειδώνος (Αμφιθέας - Αλίμου)	1,200	57
	Ποσειδώνος (Κηφισού - Πλ. Νερού)	1,551	86
	Ποσειδώνος (Νιάρχου - Κηφισού)	3,159	87
6	Βουλιαγμένης (Αγ. Κωνσταντίνου - Πύρρωνος)	900	57
	Βουλιαγμένης (Πύρρωνος - Αγ. Κωνσταντίνου)	904	60

Πίνακας 4.9: Περιφερειακοί οδικοί άξονες ανάλυσης

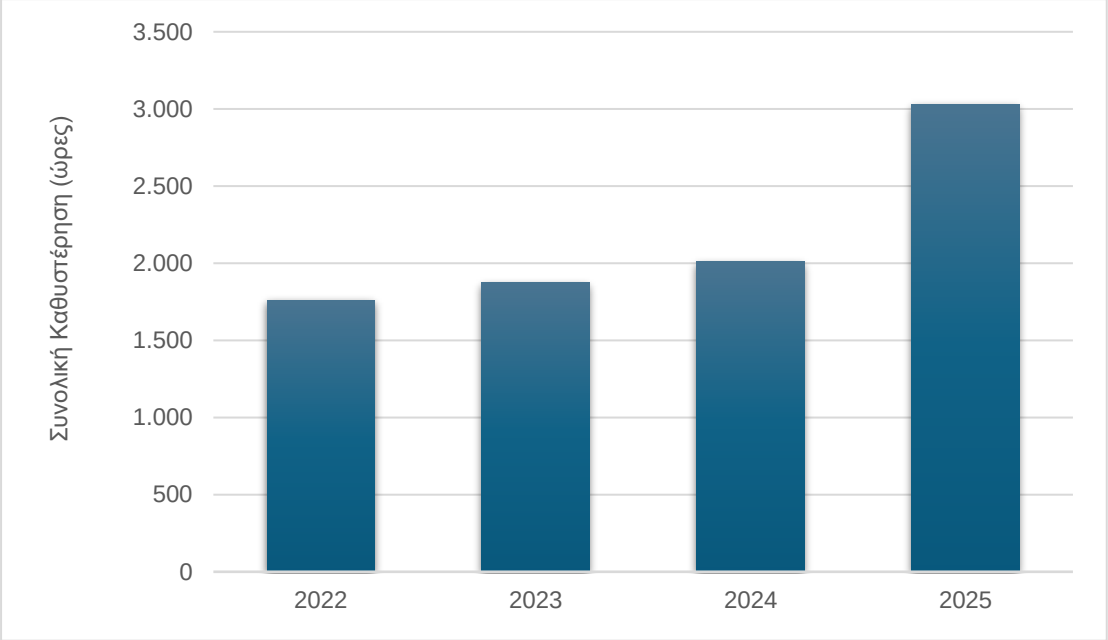


Εικόνα 4: Κεντρικοί οδικοί άξονες ανάλυσης



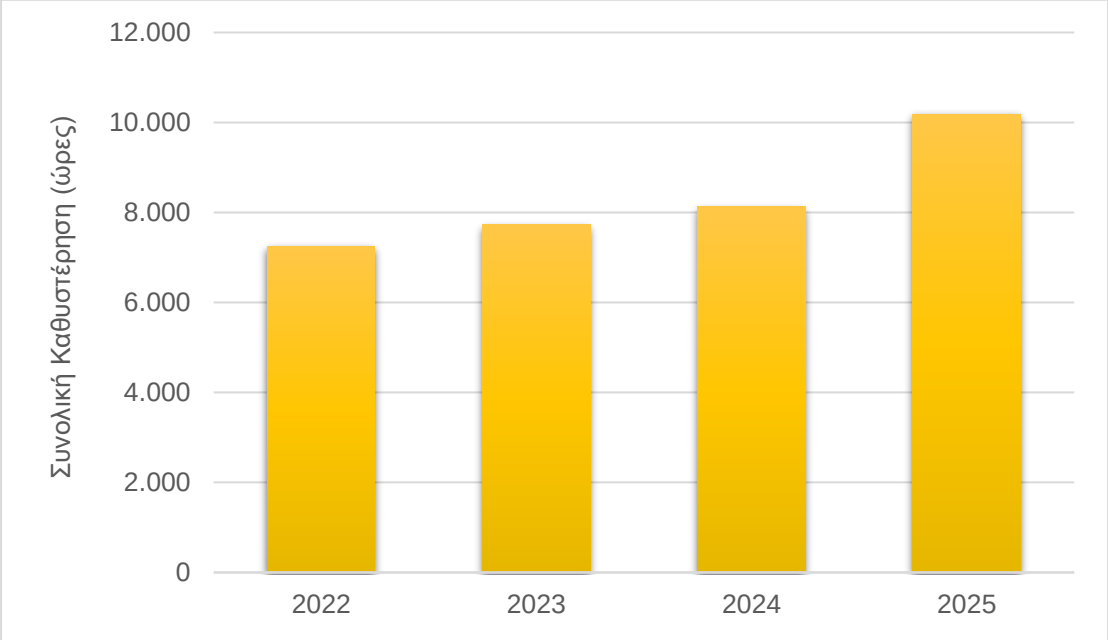
Εικόνα 5: Περιφερειακοί οδικοί άξονες ανάλυσης

Η ανάλυση στους περιφερειακούς άξονες ανέδειξε την επιδείνωση της συμφόρησης από το 2022 έως το 2025 με μεγάλη αύξηση το 2025 (+51% από το 2024). Συνολικά από το 2022 έως το 2025, οι καθυστερήσεις έχουν αυξηθεί κατά 72%. (Διάγραμμα 4-10).



Διάγραμμα 4-10: Ετήσια μεταβολή της συνολικής καθυστέρησης σε περιφερειακούς άξονες

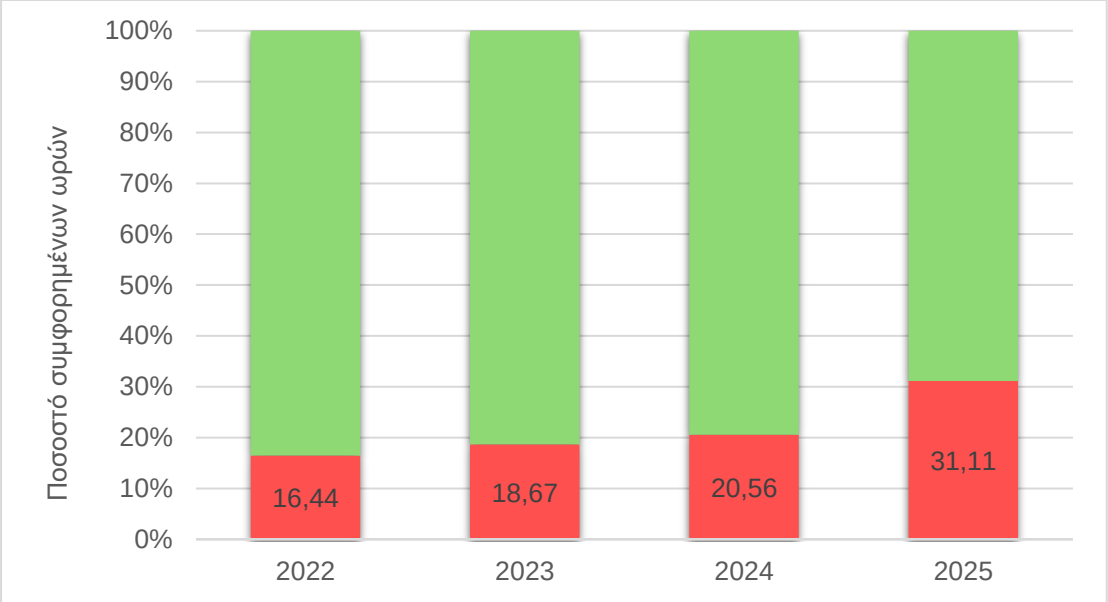
Η ανάλυση στους άξονες του κέντρου της Αθήνας, είχε παρόμοια ευρήματα, με διαρκή αύξηση της συμφόρησης στα έτη 2022 – 2025 και συνολική αύξηση 40% όπως φαίνεται στο Διάγραμμα 4-11.



Διάγραμμα 4-11: Ετήσια μεταβολή της συνολικής καθυστέρησης σε κεντρικούς άξονες

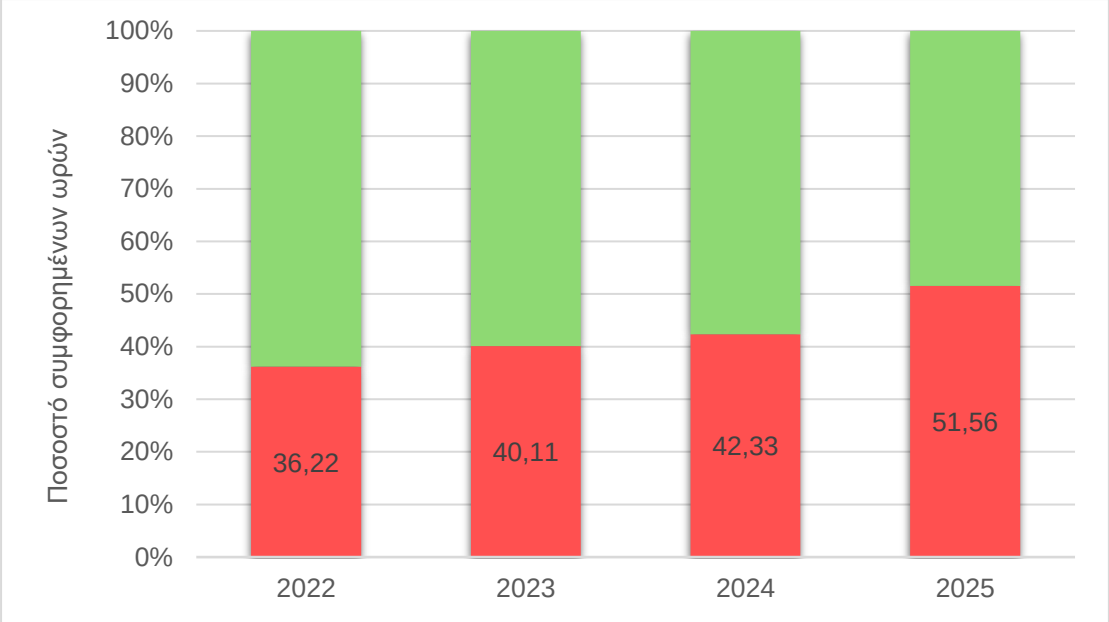
Εκτός από την ένταση της συμφόρησης, ενδιαφέρον έχει και η διάρκεια αυτής η οποία μπορεί να μετρηθεί ως η διάρκεια στην οποία παρουσιάζονται χαμηλές ταχύτητες μέσα στην ημέρα. Θέτοντας ως ταχύτητα συμφόρησης το 50% της FFS και κάτω, προκύπτουν τα αποτελέσματα:

Η διάρκεια των ωρών αιχμής έχει σχεδόν διπλασιαστεί στους περιφερειακούς άξονες στο διάστημα των ετών 2022 – 2025. Συγκεκριμένα το 2025, από τις εννιά ώρες ανάλυσης (8 – 10:00 & 12 – 19:00) ανά εργάσιμη ημέρα, σχεδόν οι τρείς ώρες παρουσιάζουν μέσες ταχύτητες κίνησης <50% της FFS. Η αντίστοιχη τιμή του 2022 είναι 1.5 ώρες (16% της περιόδου ανάλυσης).



Διάγραμμα 4-12: Μεταβολή ποσοστού συμφορημένων ωρών στους περιφερειακούς άξονες

Στους οδικούς άξονες στο κέντρο της Αθήνας, η διάρκεια της συμφόρησης έχει ανοδική τάση με αποκορύφωμα το 2025 όπου 52% της περιόδου ανάλυσης παρουσιάζει πολύ χαμηλές ταχύτητες (4.5 στις 9 ώρες). Από το 2022 υπάρχει αύξηση 42%.



Διάγραμμα 4-13: Μεταβολή ποσοστού συμφορημένων ωρών στους οδικούς άξονες στο κέντρο

4.5.2. Μεταβλητότητα χρόνων διαδρομής

Για τη διερεύνηση μεταβλητότητας των χρόνων διαδρομής στο κέντρο της Αθήνας το 2025, αξιοποιήθηκαν τα δεδομένα από την πλατφόρμα της Tom Tom. Γι' αυτόν τον σκοπό, επιλέχθηκε η μέθοδος μέτρησης της αξιοπιστίας του χρόνου διαδρομής (Travel Time Reliability) μέσω ορισμένων δεικτών που προτείνει και το Υπουργείο Μεταφορών των ΗΠΑ:

1. Χρόνος Διαδρομής του 90ου ή του 95ου Εκατοστημορίου
2. Δείκτης Buffer (Buffer Index)
3. Δείκτης Προγραμματισμένου Χρόνου (Planning Time Index)

Ο Χρόνος Διαδρομής του 90ου ή του 95ου Εκατοστημορίου είναι μία εκτίμηση των καθυστερήσεων που δύναται να υπάρξουν σε περιπτώσεις αυξημένης κυκλοφοριακής συμφόρησης σε συγκεκριμένες διαδρομές. Ο εντοπισμός αυτού του χρόνου γίνεται με την κατάταξη των παρατηρούμενων χρόνων διαδρομής σε αύξουσα σειρά και την εύρεση του 90ου και του 95ου εκατοστημορίου στο σύνολο των τιμών. Για παράδειγμα, αν για μία διαδρομή υπάρχουν 500 μετρήσεις, ο χρόνος διαδρομής του 90ου εκατοστημορίου αντιστοιχεί στην 450η τιμή (0.90×500) και του 95ου στην 475η αντίστοιχα.

Ο Δείκτης Buffer (Buffer Index), χρησιμοποιείται για τον υπολογισμό του επιπλέον χρόνου που πρέπει να προσθέσουν οι μετακινούμενοι κατά τον προγραμματισμό μίας μετακίνησης με σκοπό την έγκαιρη άφιξη τους. Για παράδειγμα, μία τιμή δείκτη Buffer 40% υποδεικνύει πως για μία διαδρομή που έχει μέσο χρόνο διάνυσης είκοσι λεπτά, ο μετακινούμενος πρέπει να προσθέσει επιπλέον οχτώ λεπτά (0.40×20) για να είναι στην ώρα του στο 95% των περιπτώσεων. Ο υπολογισμός αυτού του δείκτη γίνεται με τον παρακάτω τρόπο:

$$Buffer\ Index$$

$$= \frac{Χρόνος\ Διαδρομής\ 95ου\ Εκατοστημορίου\ -\ Μέσος\ Χρόνος\ Διαδρομής}{Μέσος\ Χρόνος\ Διαδρομής}$$

Ο Δείκτης Προγραμματισμένου Χρόνου (Planning Time Index – PTI) υπολογίζει τον συνολικό χρόνο που απαιτείται για να είναι στην ώρα του ένας μετακινούμενος. Για παράδειγμα, για μία διαδρομή που διαρκεί δεκαπέντε λεπτά σε συνθήκες ελεύθερης ροής, ένας δείκτης Προγραμματισμένου Χρόνου ίσος με 1.60 μεταφράζεται σε συνολικό απαιτούμενο χρόνο εικοσιτεσσάρων λεπτών (1.60×15) για την εξασφάλιση έγκαιρης άφιξης στο 95% των περιπτώσεων. Ο υπολογισμός του δείκτη γίνεται σύμφωνα με την εξίσωση:

$$PTI = \frac{Χρόνος\ Διαδρομής\ 95ου\ Εκατοστημορίου}{Χρόνος\ Διαδρομής\ σε\ συνθήκες\ ελεύθερης\ ροής}$$

Ο Δείκτης Χρόνου Διαδρομής (Travel Time Index – TTI) είναι ένας ακόμη δείκτης ο οποίο εκφράζει την αύξηση της κυκλοφορίας κατά τις ώρες αιχμής και υπολογίζεται από τον λόγο:

$$TTI = \frac{Μέσος\ Χρόνος\ Διαδρομής\ στην\ αιχμή}{Χρόνος\ Διαδρομής\ σε\ συνθήκες\ ελύθερης\ ροής}$$

Επιπλέον, υπολογίστηκαν και οι στατιστικοί δείκτες: Αριθμός Μετρήσεων, Μέση Τιμή, Εύρος, Διακύμανση και Τυπική Απόκλιση.

Για την ανάλυση των χρόνων διαδρομής και τον υπολογισμό των παραπάνω δεικτών, εξαιρέθηκαν οι μετρήσεις των Σαββατοκύριακων και επιλέχθηκε ως χρονικό διάστημα η ώρα αιχμής για τις κεντρικές οδούς που αναγράφονται στον Πίνακα 4.8. Αυτή υπολογίστηκε σύμφωνα με τις παρατηρήσεις των μέγιστων καθυστερήσεων (όπως αυτή δίνεται απευθείας από τα δεδομένα) για όλη τη διάρκεια του έτους 2025:

$$h^* = \underset{h \in H}{argmax} \left(\frac{1}{n_h} \sum_{d \in D_h} d \right)$$

$$d^* = \frac{1}{n_{h^*}} \sum_{d \in D_{h^*}} d$$

Όπου:

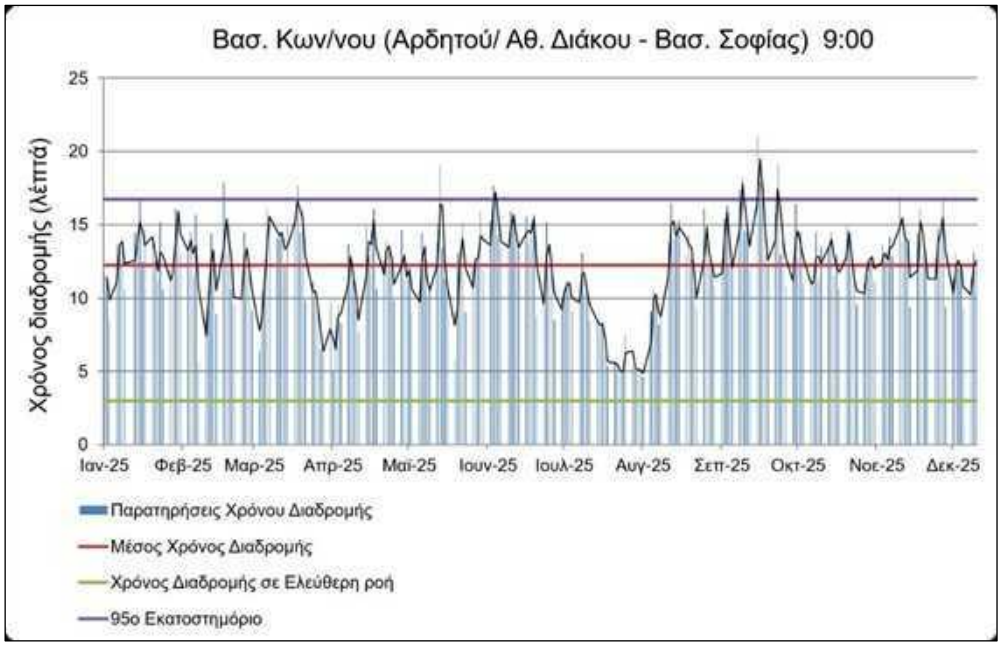
- h^*: η ώρα αιχμής του οδικού τμήματος
- d^*: η αντίστοιχη μέση καθυστέρηση αιχμής
- $D_h = \{d1, d2, \ldots, d_n\}$ το σύνολο των τιμών καθυστέρησης (όπως αναγράφονται στα δεδομένα) που καταγράφηκαν την ώρα $h \in \{0,1,\ldots,23\}$
- H: το σύνολο των ωρών για τις οποίες υπάρχει τουλάχιστον μία μέτρηση ($n_h \geq 1$)
- $n_h = |D_h|$ είναι το πλήθος των μετρήσεων την ώρα h.

Στον παρακάτω πίνακα φαίνονται οι μεγαλύτερες μέσες τιμές καθυστέρησης για τους δώδεκα οδικούς άξονες στο κέντρο της Αθήνας με την αντίστοιχη ώρα καταμέτρησης.

Τμήμα / Κατεύθυνση	Ώρα αιχμής	Μέση Καθυστ.
Βασιλίσσης Σοφίας (Πανεπιστημίου - Βασιλέως Κωνσταντίνου)	8:00	65.6s
Πειραιώς (Κολοκυνθούς - Πλατεία Ομόνοιας)	9:00	152.0s
Βασιλέως Κωνσταντίνου (Αρδητού / Αθ. Διάκου- Βασ. Σοφίας)	9:00	259.2s
Πατησίων (Ιουλιανού - Χαλκοκονδύλη)	9:00	26.8s
Πατησίων (Λεωφόρος Αλεξάνδρας - Σταδίου)	9:00	40.0s
Βασιλίσσης Αμαλίας (Αθ. Διάκου - Πανεπιστημίου)	10:00	101.1s
Λεωφ. Συγγρού (Φραντζή - Βασιλίσσης Αμαλίας)	10:00	183.2s
Σόλωνος (Βασιλίσσης Σοφίας - Πατησίων)	12:00	96.4s
Βασιλίσσης Σοφίας (Βασιλέως Κωνσταντίνου - Κηφισίας)	14:00	154.0s
Βασιλίσσης Σοφίας (Βασιλέως Κωνσταντίνου - Πανεπιστημίου)	14:00	204.0s
Σταδίου (Αιόλου - Βασιλέως Γεωργίου)	14:00	210.8s
Λεωφόρος Αλεξάνδρας (Κηφισίας- Πατησίων)	15:00	269.8s
Βασιλέως Κωνσταντίνου (Βασ. Σοφίας - Αρδητού / Αθ. Διάκου)	15:00	291.3s
Λεωφόρος Αλεξάνδρας (Πανόρμου - Χαριλάου Τρικούπη)	15:00	58.3s
Πανεπιστημίου (Βασιλίσσης Σοφίας - Πατησίων)	16:00	101.5s
Πειραιώς (Κολοκυνθούς - Ιερό Οδός)	16:00	70.6s
Βασιλίσσης Σοφίας (Κηφισίας - Βασιλέως Κωνσταντίνου)	16:00	97.3s
Ακαδημίας (Πατησίων - Βασιλίσσης Σοφίας)	17:00	147.0s
Πατησίων (Στούρναρη - Ιουλιανού)	17:00	5.3s
Λεωφ. Συγγρού (Βασιλίσσης Αμαλίας - Φραντζή)	17:00	8.8s
Λεωφόρος Αλεξάνδρας (Χαρ. Τρικούπη- Πανόρμου)	18:00	105.9s
Βασιλίσσης Αμαλίας (Φιλελλήνων - Αθ. Διάκου)	18:00	14.6s
Φιλελλήνων (Βασιλέως Γεωργίου - Βασιλίσσης Αμαλίας)	18:00	15.7s
Λεωφόρος Αλεξάνδρας (Πατησίων - Κηφισίας)	18:00	160.8s

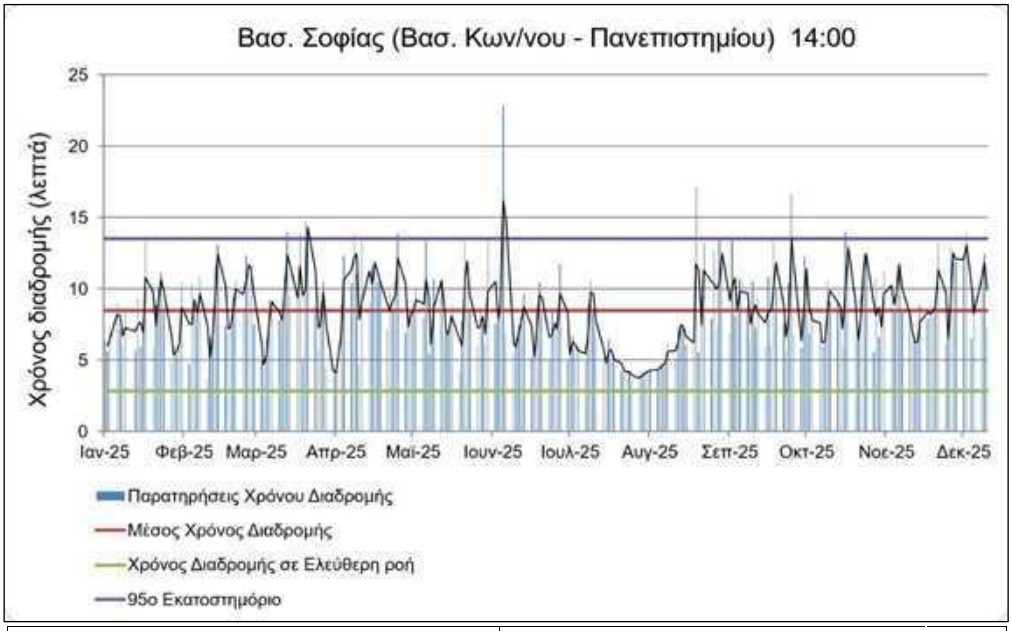
Πίνακας 4.10: Ώρες αιχμής για 12 οδικούς άξονες στο κέντρο της Αθήνας.

Γίνεται αντιληπτό πως οι ώρες αιχμής στο κέντρο της Αθήνας εμφανίζονται καθ' όλη τη διάρκεια της ημέρας, από τις 8:00 έως τις 18:00. Στα παρακάτω διαγράμματα απεικονίζεται η μεταβλητότητα των χρόνων διαδρομής για τα οδικά τμήματα με τις μεγαλύτερες καθυστερήσεις, στις αντίστοιχες ώρες ανάλυσης με τους προαναφερόμενους κυκλοφοριακούς και στατιστικούς δείκτες.



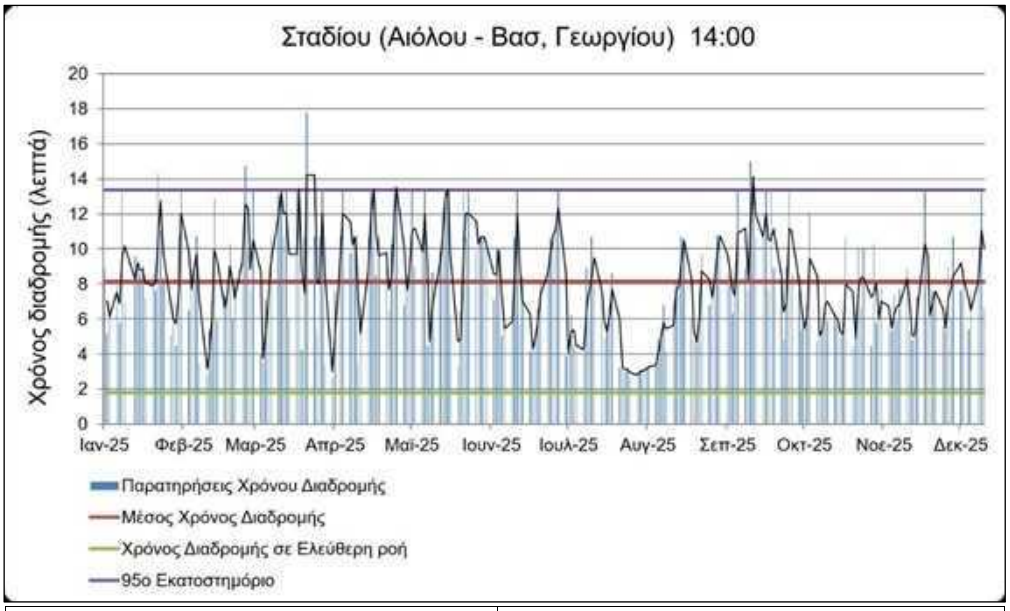
Αριθμός Μετρήσεων	245	95ο Εκατοστημόριο	16.8
Μέση Τιμή	12.2	Δείκτης Buffer	0.37
Εύρος	16.4	Δείκτης Προγραμματισμένου Χρόνου	5.59
Διακύμανση	9.2	Δείκτης Χρόνου Διαδρομής	4.08
Τυπική Απόκλιση	3.0	Μέση Ταχύτητα (χλμ/ώρα)	9

Διάγραμμα 4-14: Μεταβλητότητα χρόνου διαδρομής στη Βασιλέως Κωνσταντίνου στις 9:00 το 2025



Αριθμός Μετρήσεων	243	95ο Εκατοστημόριο	13.5
Μέση Τιμή	8.5	Δείκτης Buffer	0.60
Εύρος	19.2	Δείκτης Προγραμματισμένου Χρόνου	4.80
Διακύμανση	8.8	Δείκτης Χρόνου Διαδρομής	3.01
Τυπική Απόκλιση	3.0	Μέση Ταχύτητα (χλμ/ώρα)	9.8

Διάγραμμα 4-15: Μεταβλητότητα χρόνου διαδρομής στη Βασιλίσσης Σοφίας στις 14:00 το 2025



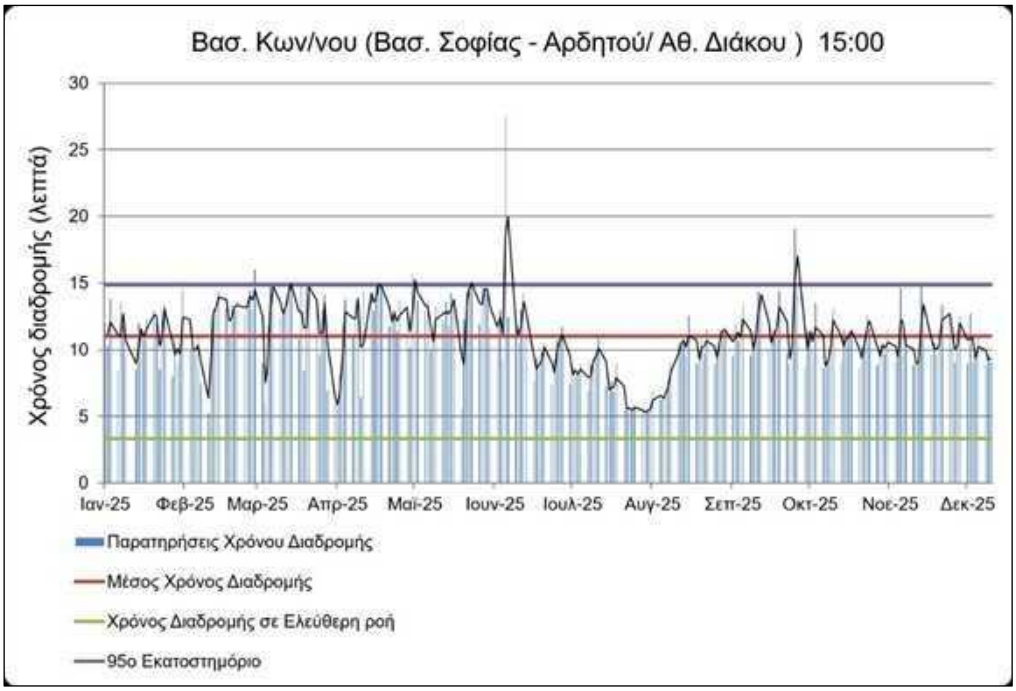
Αριθμός Μετρήσεων	243	95ο Εκατοστημόριο	13.4
Μέση Τιμή	8.2	Δείκτης Buffer	0.64
Εύρος	15.2	Δείκτης Προγραμματισμένου Χρόνου	7.50
Διακύμανση	9.9	Δείκτης Χρόνου Διαδρομής	4.58
Τυπική Απόκλιση	3.1	Μέση Ταχύτητα (χλμ/ώρα)	8

Διάγραμμα 4-16: Μεταβλητότητα χρόνου διαδρομής στη Σταδίου στις 14:00 το 2025



Αριθμός Μετρήσεων	245	95ο Εκατοστημόριο	17.5
Μέση Τιμή	13.6	Δείκτης Buffer	0.29
Εύρος	12.9	Δείκτης Προγραμματισμένου Χρόνου	3.75
Διακύμανση	7.1	Δείκτης Χρόνου Διαδρομής	2.90
Τυπική Απόκλιση	2.7	Μέση Ταχύτητα (χλμ/ώρα)	12.7

Διάγραμμα 4-17: Μεταβλητότητα χρόνου διαδρομής στην Αλεξάνδρας στις 15:00 το 2025



Αριθμός Μετρήσεων	245	95ο Εκατοστημόριο	14.9
Μέση Τιμή	11.0	Δείκτης Buffer	0.35
Εύρος	22.3	Δείκτης Προγραμματισμένου Χρόνου	4.48
Διακύμανση	7.8	Δείκτης Χρόνου Διαδρομής	3.32
Τυπική Απόκλιση	2.8	Μέση Ταχύτητα (χλμ/ώρα)	10

Διάγραμμα 4-18: Μεταβλητότητα χρόνου διαδρομής στη Βασιλέως Κωνσταντίνου στις 15:00 το 2025

Από τα διαγράμματα γίνεται αντιληπτή η σημασία των δεικτών αξιοπιστίας του χρόνου διαδρομής. Στις οδούς Σταδίου και Βασιλίσσης Σοφίας το 5% των παρατηρήσεων των χρόνων διαδρομής στις 14:00 διαφέρει κατά πέντε λεπτά από το μέση τιμή δίνοντας δείκτες Buffer 0.64 και 0.60 αντίστοιχα. Το 5% των παρατηρήσεων αντιστοιχεί σε δώδεκα παρατηρήσεις (0.05*243), δηλαδή 12 εργάσιμες ημέρες στις οποίες οι μετακινούμενοι σε αυτές τις οδούς καθυστέρησαν κατά πέντε λεπτά μόνο σε αυτά τα οδικά τμήματα.

Οι δείκτες Προγραμματισμένου Χρόνου και Χρόνου Διαδρομής συγκρίνουν την ένταση της κυκλοφορίας σε ώρες αιχμής με αυτήν σε συνθήκες ελεύθερης ροής. Η οδός Σταδίου παρουσιάζει τη μεγαλύτερη αύξηση του χρόνου διαδρομής κατά την αιχμή (στις 14:00): κατά μέσο όρο +4.6 φορές τον χρόνο διαδρομής ελεύθερης ροής (δείκτης = 4.6). Για να προγραμματίσει όμως κανείς τη διαδρομή του από την οδό Σταδίου την ίδια ώρα, θα πρέπει να υπολογίζει χρόνο διαδρομής 7.5 φορές αυτόν της ελεύθερης ροής, έτσι ώστε να είναι στην ώρα του κατά 95% των περιπτώσεων (δείκτης = 7.5).

Τέλος, αξίζει να επισημανθεί η μεταβλητότητα των χρόνων διαδρομής των παραπάνω οδών, η οποία συμβάλει στην αναξιοπιστία αυτών και είναι αποτέλεσμα των απρόβλεπτων συνθηκών της κυκλοφορίας στον αστικό ιστό.

5. Εφαρμογή Μεθοδολογίας και Αποτελέσματα

5.1. Εισαγωγή

Μετά την ολοκλήρωση της προκαταρκτικής επεξεργασίας και της περιγραφικής ανάλυσης των δεδομένων, το παρόν κεφάλαιο εστιάζει στην ανάπτυξη και εφαρμογή στατιστικών και μαθηματικών μοντέλων με στόχο την περαιτέρω εμβάθυνση στις αλληλεπιδράσεις των παραγόντων που διαμορφώνουν το επίπεδο της κυκλοφοριακής συμφόρησης στην Αθήνα. Οι μεθοδολογικές προσεγγίσεις που επιλέχθηκαν, βασίστηκαν στα ευρήματα της διεθνούς βιβλιογραφίας και αναλύονται διεξοδικά στο Κεφάλαιο 3 της παρούσας διπλωματικής εργασίας. Εφαρμόστηκαν μοντέλα πολλαπλής γραμμικής παλινδρόμησης με σκοπό τη διερεύνηση της γραμμικής συσχέτισης και της στατιστικής σημαντικότητας των κοινωνικοοικονομικών δεικτών, των μετεωρολογικών φαινομένων, καθώς και της κυκλοφοριακής σύνθεσης. Παράλληλα, έγινε εμβάθυνση μέσω προηγμένων μεθόδων Μηχανικής Μάθησης βασισμένων σε δέντρα αποφάσεων. Συγκεκριμένα, αξιοποιείται ο αλγόριθμος Ενίσχυσης Κλίσης βασισμένος σε Ιστόγραμμα (HistGradientBoostingRegressor – HGB) για τη διερεύνηση επιρροής των μετεωρολογικών συνθηκών στην ταχύτητα κίνησης των οχημάτων και οι αλγόριθμοι Ενίσχυσης Κλίσης (Gradient Boosting) και Τυχαίων Δασών (Random Forest) για την εξέταση της επίδρασης που έχει η σύνθεση της κυκλοφορίας στο κέντρο της Αθήνας.

5.2. Συσχέτιση κοινωνικοοικονομικών δεδομένων με τη συμφόρηση

Στην πρώτη ενότητα διερευνάται η σχέση της συμφόρησης στο οδικό δίκτυο της Αθήνας με κοινωνικοοικονομικούς παράγοντες, με βάση τη διεθνή βιβλιογραφία η οποία συνδέει τη συμφόρηση με παράγοντες ζήτησης, όπως η οικονομική δραστηριότητα, ο πληθυσμός και η ιδιοκτησία οχημάτων.

5.2.1. Δεδομένα

Τα δεδομένα χρόνου διαδρομής τα οποία χρησιμοποιήθηκαν για τον υπολογισμό του δείκτη συμφόρησης συλλέχθηκαν από την βάση δεδομένων της Google για το χρονικό διάστημα Ιανουάριος 2022 – Δεκέμβριος 2025 και τις εξής κοινές ώρες: 8 – 10:00 τις καθημερινές και 12 – 19:00 για όλες τις ημέρες της εβδομάδας. Για την ανάλυση επιλέχθηκαν τα ίδια οδικά τμήματα που αναγράφονται στο Κεφάλαιο 4.5.1 καθώς αποτελούνται από κύριους οδικούς άξονες τόσο στο κέντρο της Αθήνας όσο και γύρο από αυτό. Τα κοινωνικοοικονομικά δεδομένα συλλέχθηκαν σε ετήσια, μηνιαία και τριμηνιαία βάση για την ίδια περίοδο ανάλυσης αλλά για την ανάλυση ανάχθηκαν όλα σε τριμηνιαίες τιμές (Πίνακας 5.1).

5.2.2. Καθορισμός μεταβλητών

Ως εξαρτημένη μεταβλητή του μοντέλου επιλέχθηκε η καθυστέρηση ανά παρατήρηση. Η τιμή του τριμήνου είναι ο μέσος όρος όλων των παρατηρήσεων του τριμήνου, από όλα τα τμήματα μαζί:

$$D_q = \frac{1}{N_q} \sum_{i \in \Omega_q} t_i - t_{FFS}$$

Όπου:
Ω_q είναι το σύνολο των παρατηρήσεων του τριμήνου q από όλα τα οδικά τμήματα (Πίνακες 4.8 & 4.9)
N_q: το πλήθος των Ω_q
t_i: ο χρόνος διαδρομής μίας παρατήρησης i
t_{FFS}: ο χρόνος διαδρομής σε συνθήκες ελεύθερης ροής (ορίσθηκε η μέγιστη παρατηρούμενη ταχύτητα ανά οδικό τμήμα)

Οι ανεξάρτητες μεταβλητές αποτελούνται από τριμηνιαίες τιμές των κοινωνικοοικονομικών δεδομένων που συλλέχθηκαν.

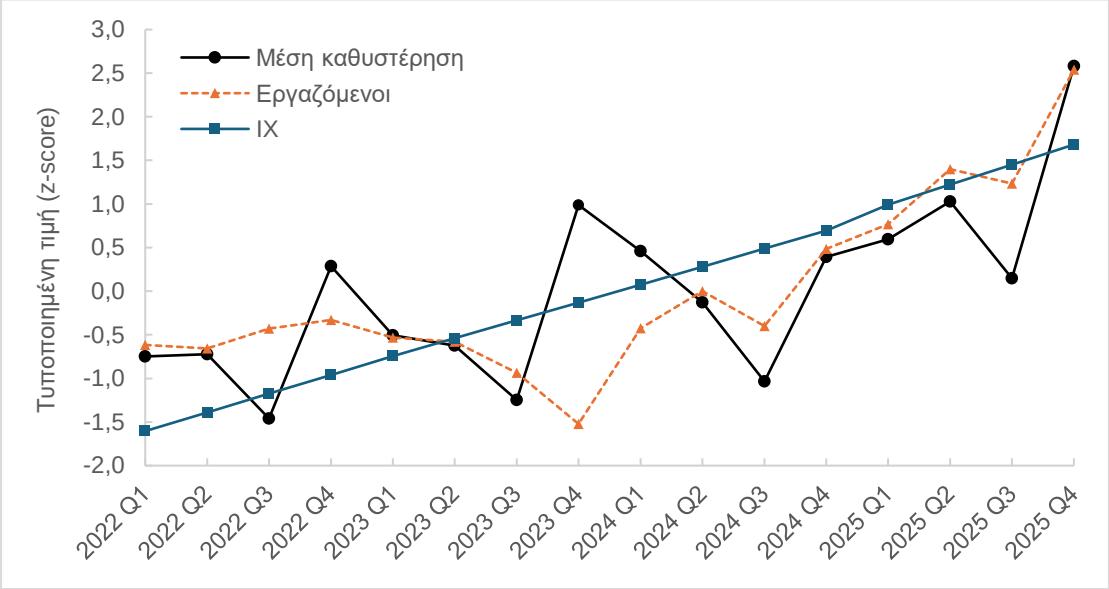
Έτος	Τρίμηνο	Μέση καθυστέρ. (δλ)	Εργαζόμενοι (χιλ.)	Νέες Οικοδομές	Τιμή καυσίμου (€/lt)	Τουρίστες (χιλ.)	ΙΧ (χιλ.)	Πληθυσμός (χιλ.)
2022	1	101.42	1,572.1	451	1.874	504.5	3,196.1	3,807.7
2022	2	101.85	1,570.9	495	2.197	1,589.5	3,222.1	3,803.0
2022	3	89.18	1,577.7	485	2.069	2,289.7	3,248.1	3,798.4
2022	4	119.16	1,580.7	597	1.949	1,240.2	3,274.1	3,793.8
2023	1	105.53	1,574.6	513	1.864	1,079.0	3,300.1	3,789.1
2023	2	103.54	1,573.2	597	1.845	2,336.5	3,325.0	3,785.0
2023	3	92.83	1,562.5	486	1.928	2,835.6	3,349.9	3,780.9
2023	4	131.18	1,544.9	613	1.831	1,637.5	3,374.8	3,776.8
2024	1	122.1	1,577.8	609	1.840	1,291.5	3,399.7	3,772.7
2024	2	112.02	1,590.6	618	1.897	2,577.9	3,424.6	3,770.6
2024	3	96.51	1,578.6	547	1.795	3,165.0	3,449.6	3,768.5
2024	4	120.97	1,605.1	775	1.745	1,750.4	3,474.5	3,766.4
2025	1	124.45	1,613.6	328	1.775	1,372.4	3,510.9	3,764.3
2025	2	131.9	1,632.6	579	1.713	2,883.9	3,538.7	3,762.2
2025	3	116.8	1,627.6	578	1.721	3,481.8	3,566.5	3,760.1
2025	4	158.53	1,666.8	649	1.723	1,967.1	3,594.3	3,758.0

Πίνακας 5.1: Βάση κοινωνικοοικονομικών δεδομένων και συμφόρησης

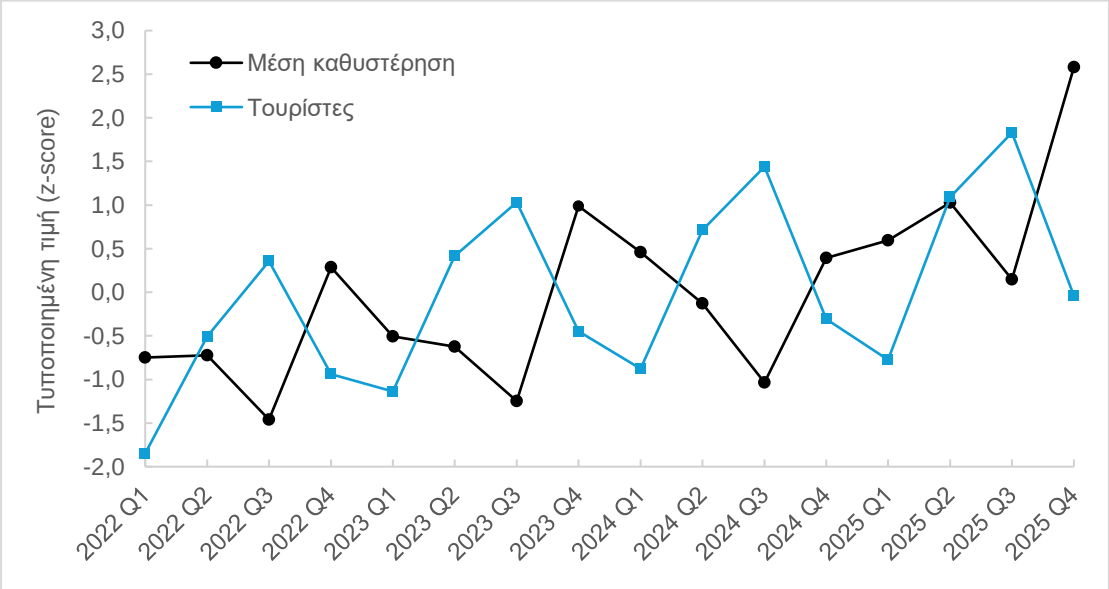
5.2.3. Ανάλυση συσχέτισης

Όπως φαίνεται στα Διαγράμματα 5-1 & 5-2, η μέση καθυστέρηση αυξάνεται ανά έτος αλλά ταυτόχρονα διαθέτει και εποχικότητα: τους καλοκαιρινούς μήνες παρουσιάζει τις ελάχιστες τιμές ενώ τους χειμερινούς, τις μέγιστες. Αντίστοιχα, οι μεταβλητές: αριθμός εργαζομένων και αριθμός ΙΧ αυξάνονται με τα έτη και ο αριθμός τουριστών μεγιστοποιείται κάθε 3ο τρίμηνο του έτους, με αποτέλεσμα

να συσχετίζονται με την μέση καθυστέρηση χωρίς όμως να είναι σαφές αν την επηρεάζουν άμεσα.



Διάγραμμα 5-1: Τυποποιημένες τιμές μέσης τριμηνιαίας καθυστέρησης, εργαζόμενων και IX (2022 Q1 - 2025 Q4)



Διάγραμμα 5-2: Μέση τριμηνιαία καθυστέρηση και αριθμός τουριστών (2022 Q1 - 2025 Q4)

Πριν από τον έλεγχο της σχέσης της συμφόρησης με τις κοινωνικοοικονομικές μεταβλητές, εξετάστηκε πόσο μέρος της διακύμανσης της κάθε σειράς εξηγείται μόνο από τον χρόνο και το τρίμηνο του έτους. Για κάθε σειρά υπολογίστηκε ο συντελεστής συσχέτισης Pearson με τη γραμμική τάση (t = 0, …, 15) και ο συντελεστής προσδιορισμού R^2 (Πίνακας 5.2) της παλινδρόμησης:

$$\textit{σειρά} \sim t + \textit{τρίμηνο}$$

όπου το τρίμηνο εισάγεται ως κατηγορική μεταβλητή (3 dummies, με το Q1 ως κατηγορία αναφοράς).

	r με χρόνο	R^2 (τάση + τρίμηνο)
Συμφόρηση (μέση καθυστέρηση)	0.67	0.867
Εργαζόμενοι	0.77	0.612
Νέες οικοδομές	0.33	0.526
Τιμή καυσίμου	-0.81	0.743
Τουρίστες	0.52	0.974
ΙΧ οχήματα	1.00	0.999
Πληθυσμός	-0.99	0.97
Ποσοστό εργ. δυναμικού	0.60	0.47
Ποσοστό ανεργίας	-0.62	0.636

Πίνακας 5.2: Συσχέτιση κοινωνικοοικονομικών μεταβλητών με τάση και τρίμηνο

Η συμφόρηση έχει μέτρια ανοδική τάση (r = 0.67), αλλά το μοντέλο τάσης και τριμήνου εξηγεί ήδη το 86.7% της διακύμανσής της, άρα μένει μικρό περιθώριο για να προσθέσει κάτι μια κοινωνικοοικονομική μεταβλητή. Οι μεταβλητές: ΙΧ οχήματα (r = 1.00, R^2 = 0.999) και ο Πληθυσμός (r = -0.99, R^2 = 0.97) είναι ουσιαστικά καθαρές γραμμικές τάσεις, η μία ανοδική και η άλλη καθοδική, και δεν περιέχουν σχεδόν καμία ανεξάρτητη πληροφορία. Γι' αυτό οποιαδήποτε συσχέτισή τους με τη συμφόρηση στα επίπεδα δεν διαχωρίζεται από τη γενική τάση του χρόνου. Η χαμηλή συσχέτιση της μεταβλητής: τουρίστες με τον χρόνο (r = 0.52) σε συνδυασμό με R^2 = 0.974 δείχνει ότι η σειρά είναι σχεδόν εξ ολοκλήρου εποχική και όχι τάση. Η τουριστική εποχικότητα όμως μπορεί να εξηγεί και μέρος της εποχικότητας της συμφόρησης, χωρίς αυτό να σημαίνει αιτιότητα. Η τιμή καυσίμου (r = −0.81) και οι εργαζόμενοι (r = 0.77) έχουν επίσης ισχυρή τάση.

Οι υπόλοιπες μεταβλητές έχουν το μεγαλύτερο ποσοστό διακύμανσης που δεν εξηγείται από τάση και εποχικότητα. Είναι οι μεταβλητές για τις οποίες μια συσχέτιση με τα υπόλοιπα της συμφόρησης θα ήταν πιο πληροφοριακή. Με μόλις 16 παρατηρήσεις και 5 παραμέτρους στο μοντέλο, τα R^2 είναι εξαρχής φουσκωμένα και οι τιμές πρέπει να διαβάζονται ως ενδείξεις και όχι ως ακριβείς εκτιμήσεις.

Η συσχέτιση των μεταβλητών με τη συμφόρηση σε επίπεδα (Πίνακας 5.3), δίνει αρκετές σημαντικές μεταβλητές ($p<0.05$) αλλά τα αποτελέσματα είναι παραπλανητικά καθώς υπάρχει τάση αυτών με τον χρόνο.

Μεταβλητή	Pearson r	p	Spearman ρ	p
Εργαζόμενοι	0.675	0.004	0.562	0.024
Νέες οικοδομές	0.401	0.124	0.568	0.022
Τιμή καυσίμου	-0.595	0.015	-0.674	0.004
Τουρίστες	-0.108	0.691	-0.112	0.68
ΙΧ οχήματα	0.676	0.004	0.65	0.006
Πληθυσμός	-0.628	0.009	-0.65	0.006
Ποσοστό εργ. δυναμικού	0.502	0.048	0.522	0.038
Ποσοστό ανεργίας	-0.653	0.006	-0.596	0.015

Πίνακας 5.3: Συσχέτιση κοινωνικοοικονομικών μεταβλητών με τη συμφόρηση σε επίπεδα

Ύστερα από αφαίρεση της τάσης και της εποχικότητας, η μόνη μεταβλητή με σημαντική συσχέτιση και στους δύο δείκτες είναι ο αριθμός των ΙΧ οχημάτων (r = 0.762, ρ = 0.741, p = 0.001). Οριακή ένδειξη (σημαντική μόνο κατά Spearman) υπάρχει για τους εργαζόμενους (ρ = 0.529, p = 0.035) και το ποσοστό εργατικού δυναμικού (ρ = 0.556, p = 0.025) (Πίνακας 5.4).

Μεταβλητή	Pearson r	p	Spearman ρ	p
Εργαζόμενοι	0.433	0.094	0.529	0.035
Νέες οικοδομές	-0.309	0.244	-0.297	0.264
Τιμή καυσίμου	0.242	0.368	0.274	0.305
Τουρίστες	-0.234	0.382	-0.344	0.192
ΙΧ οχήματα	0.762	0.001	0.741	0.001
Πληθυσμός	0.451	0.079	0.435	0.092
Ποσοστό εργ. δυναμικού	0.418	0.107	0.556	0.025
Ποσοστό ανεργίας	-0.112	0.679	-0.224	0.405

Πίνακας 5.4: Συσχέτιση κοινωνικοοικονομικών μεταβλητών με τη συμφόρηση (αφαίρεση τάσης και εποχικότητας)

5.2.4. Μοντέλα παλινδρόμησης

Τα αποτελέσματα των συσχετίσεων καθοδηγούν την ερμηνεία των μοντέλων. Στα επίπεδα, οι μεταβλητές: ΙΧ οχήματα και οι Εργαζόμενοι συσχετίζονται θετικά με τη συμφόρηση (r = 0.676 και 0.675), αλλά αυτό οφείλεται κυρίως στην κοινή ανοδική τάση, αφού οι δύο σειρές έχουν r = 1.00 και 0.77 με τον χρόνο. Γι' αυτό τα μοντέλα με κοινωνικοοικονομικές μεταβλητές αξιολογούνται σε σύγκριση με τα βασικά:

Βασικά μοντέλα	R^2	Adj R^2	AIC
Μόνο τρίμηνο :	0.487	0.358	133.7
Τάση + τρίμηνο :	0.867	0.818	114.1

Πίνακας 5.5: Δείκτες απόδοσης βασικών μοντέλων (μεταβλητές τάσης και τριμήνου)

Το μοντέλο «μόνο τρίμηνο» εξηγεί το 48.7% της διακύμανσης, ενώ η προσθήκη γραμμικής τάσης βελτιώνει δραστικά την προσαρμογή (R^2 = 0.867, προσαρμοσμένο R^2 = 0.818, AIC = 114.1, ΔAIC = 19.6). Το μοντέλο «τάση + τρίμηνο» αποτελεί το πραγματικό σημείο αναφοράς, αφού μια μεταβλητή σαν τα ΙΧ μπορεί να αναπαριστά απλώς τη γενική τάση.

Το μοντέλο M1 (τρίμηνο + ΙΧ οχήματα) δίνει R^2 = 0.878, προσαρμοσμένο R^2 = 0.834 και AIC = 112.7 και έχει τον ίδιο αριθμό παραμέτρων (5) με το βασικό: «τάση + τρίμηνο», οπότε η σύγκριση είναι άμεση. Η βελτίωση είναι μικρή (R^2 από 0.867 σε 0.878, AIC από 114.1 σε 112.7), άρα τα ΙΧ λειτουργούν κυρίως ως υποκατάστατο της γραμμικής τάσης και όχι ως ανεξάρτητος ερμηνευτικός παράγοντας.

Μοντέλο M1 (τρίμηνο + ΙΧ)	R^2	Adj R^2	AIC	Durbin - Watson	F-test
	0.878	0.834	112.7	0.90	p < 0.001

Πίνακας 5.6: Δείκτες απόδοσης μοντέλου M1: «τρίμηνο + ΙΧ»

	coef	std err	p (OLS)	std err HAC	p (HAC)
IX οχήματα	0.0913	0.0153	0.0001	0.0171	0

Πίνακας 5.7: Συντελεστής, τυπικό σφάλμα και τιμή p της μεταβλητής IX στο M1 μοντέλο

Στο M4 (τρίμηνο + IX + εργαζόμενοι) το R^2 αυξάνεται σε 0.897 και το προσαρμοσμένο R^2 σε 0.845, ενώ το AIC μειώνεται οριακά από 112.7 σε 112.1. Η διαφορά των 0.6 μονάδων είναι πολύ μικρή για να θεωρηθεί ουσιαστική βελτίωση, ειδικά με μία επιπλέον παράμετρο και 10 βαθμούς ελευθερίας υπολοίπων. Ο αριθμός IX παραμένει στατιστικά σημαντική μεταβλητή, αλλά ο συντελεστής της μειώνεται κατά περίπου 28% σε σχέση με το M1. Οι εργαζόμενοι δεν είναι σημαντικοί (p (OLS) = 0.212 και p (HAC) = 0.330), σε συμφωνία με τη μόνο οριακή ένδειξη που είχαν στα υπόλοιπα. Το VIF = 2.6 και για τις δύο μεταβλητές δείχνει μέτρια συγγραμμικότητα, που όμως αρκεί για να αυξήσει το τυπικό σφάλμα HAC των IX (από 0.0171 στο M1 σε 0.0245). Ο έλεγχος F ($p < 0.001$) συγκρίνει με το «μόνο τρίμηνο» και όχι με το «τάση + τρίμηνο», οπότε δεν αποδεικνύει ότι οι μεταβλητές προσθέτουν κάτι πέρα από την τάση. Το Durbin-Watson είναι 0.90 στο M1 και 1.12 στο M4, δηλαδή και στα δύο μοντέλα υπάρχει θετική αυτοσυσχέτιση των υπολοίπων. Γι' αυτό τα p του OLS είναι υπερβολικά αισιόδοξα, και με n = 16 ούτε η διόρθωση HAC τα διορθώνει επαρκώς. Συνολικά, η προσθήκη των εργαζομένων στο M4 δεν δικαιολογείται, και τα αποτελέσματα πρέπει να θεωρηθούν ενδείξεις και όχι τεκμήρια αιτιότητας.

Μοντέλο M4 (τρίμηνο + IX + εργαζόμενοι)	R^2	Adj R^2	AIC	Durbin - Watson	F-test
	0.897	0.845	112.1	1.12	$p < 0.001$

Πίνακας 5.8: Δείκτες απόδοσης μοντέλου M4: «τρίμηνο + IX + εργαζόμενοι»

	coef	std err	p (OLS)	std err HAC	p (HAC)	VIF
IX οχήματα	0.0913	0.0153	0.0001	0.0171	0	2.6
Εργαζόμενοι	0.1282	0.0962	0.2123	0.1317	0.3303	2.6

Πίνακας 5.9: Συντελεστές, τυπικό σφάλμα και τιμές p των μεταβλητών: IX και εργαζόμενοι στο M4 μοντέλο: «τρίμηνο + IX + εργαζόμενοι»

Τα αποτελέσματα δείχνουν ότι η μέση καθυστέρηση ανά παρατήρηση αυξάνεται παράλληλα με τον στόλο των IX οχημάτων. Η σχέση αυτή είναι θετική και συνεπής σε όλες τις προσεγγίσεις. Στα επίπεδα, η συσχέτιση είναι r = 0.676, παραμένει ισχυρή στα υπόλοιπα, μετά την αφαίρεση τάσης και εποχικότητας και είναι η μόνη μεταβλητή που τη διατηρεί και στους δύο δείκτες (Pearson & Spearman). Στο M1 ο συντελεστής είναι θετικός και στατιστικά σημαντικός, και παραμένει σημαντικός στο M4 όταν προστεθούν οι εργαζόμενοι, οι οποίοι δεν προσθέτουν ερμηνευτική ισχύ. Η σχέση όμως είναι συνύπαρξη και όχι απόδειξη αιτιότητας. Η προσθήκη των IX στο «τάση + τρίμηνο» βελτιώνει ελάχιστα την προσαρμογή (R^2 από 0.867 σε 0.878, AIC από 114.1 σε 112.7), οπότε τα IX λειτουργούν κυρίως ως υποκατάστατο της γενικής ανοδικής τάσης.

Η ανάλυση έχει σημαντικούς περιορισμούς, που επιβάλλουν να διαβάζονται τα αποτελέσματα ως ενδείξεις. Αρχικά, το δείγμα περιλαμβάνει μόλις 16

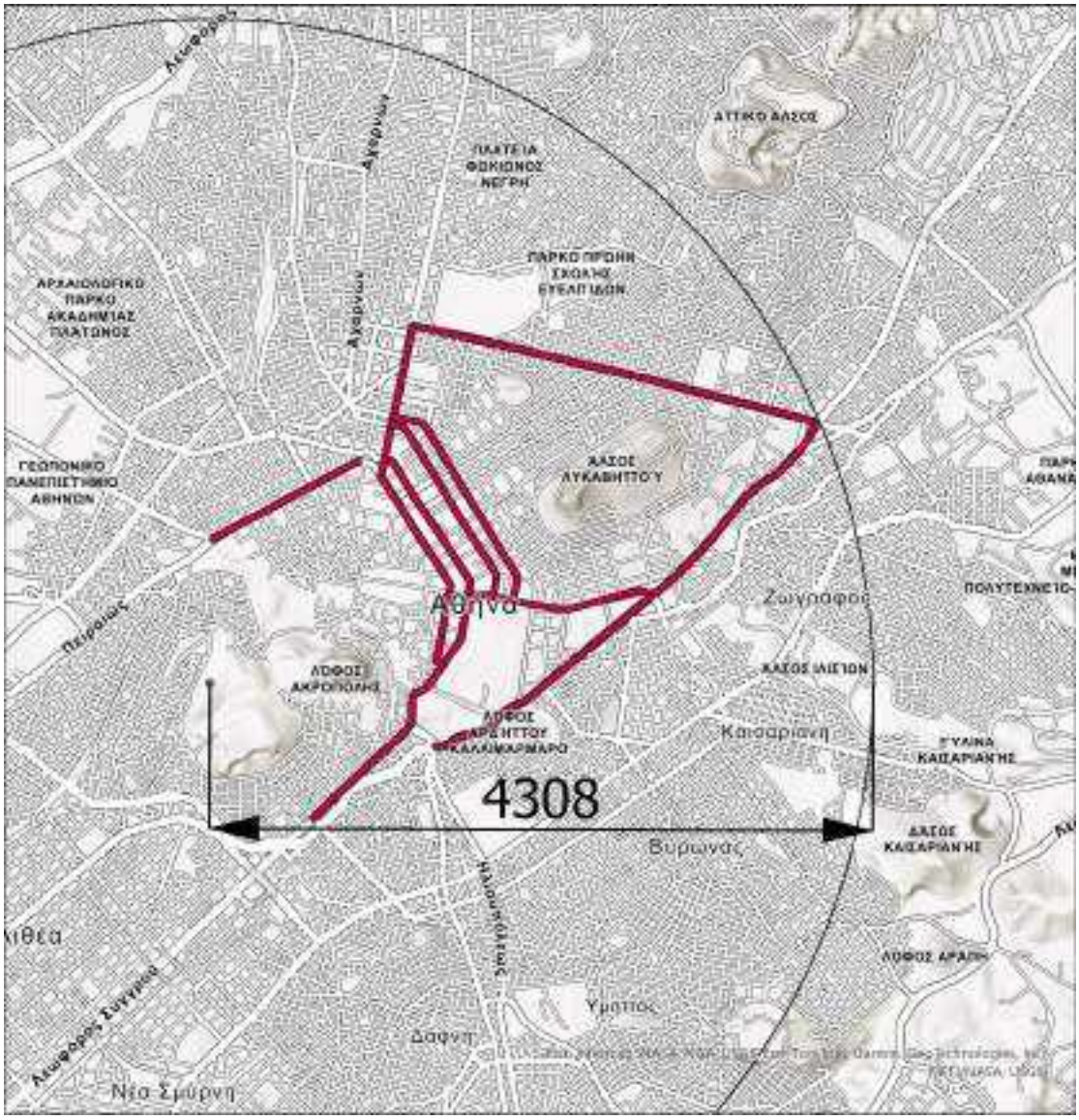
τριμηνιαίες παρατηρήσεις. Μετά την εκτίμηση 5 – 6 παραμέτρων μένουν 10 – 11 βαθμοί ελευθερίας υπολοίπων, οπότε οι εκτιμήσεις είναι αβέβαιες και τα R^2 φουσκωμένα. Επιπλέον, τα ΙΧ οχήματα (όπως και ο πληθυσμός) είναι ετήσιες τιμές με γραμμική παρεμβολή, δηλαδή περιέχουν στην πραγματικότητα μόνο 4 ανεξάρτητες παρατηρήσεις και όχι 16. Η σειρά τους είναι σχεδόν γραμμική τάση με τον χρόνο και δεν διαχωρίζεται από τη γενική ανοδική τάση της συμφόρησης. Οι διαφορές ανάμεσα στα τρίμηνα ενός έτους προέρχονται από την παρεμβολή και όχι από μετρήσεις. Τα υπόλοιπα των μοντέλων παρουσιάζουν θετική αυτοσυσχέτιση, οπότε τα p-values του OLS είναι υπερβολικά αισιόδοξα. Η διόρθωση HAC με n = 16 και με τον συμβατικό αριθμό καθυστερήσεων (L = 2) είναι κι αυτή περιορισμένης αξιοπιστίας. Τέλος, η εποχικότητα, που ελέγχεται με ψευδομεταβλητές τριμήνου, εκτιμάται από μόλις 4 παρατηρήσεις για κάθε τρίμηνο του έτους. Οι τουρίστες, που είναι σχεδόν εξ ολοκλήρου εποχική μεταβλητή, δεν μπορούν να διαχωριστούν αξιόπιστα από τις ψευδομεταβλητές τριμήνου. Για τους λόγους αυτούς, τα αποτελέσματα δεν επιτρέπουν αιτιώδη συμπεράσματα.

5.3. Μοντέλο Γραμμικής Παλινδρόμησης (ταχύτητα - καιρός)

5.3.1. Επεξεργασία και συγχώνευση δεδομένων

Για τη διερεύνηση επιρροής των καιρικών συνθηκών στους χρόνους διαδρομής, επιλέχθηκαν οι δώδεκα οδικοί άξονες στο κέντρο της Αθήνας οι οποίοι χρησιμοποιήθηκαν και στην προκαταρκτική ανάλυση (Κεφάλαιο 4.5.1). Τα δεδομένα ταχύτητας αντλήθηκαν από τη βάση δεδομένων της Google (Πίνακας 4.6) για τα έτη 2022 έως 2025 που αντιστοιχούν σε 954.5 χιλιάδες εγγραφές για τις υπό μελέτη οδούς. Τα μετεωρολογικά δεδομένα αποτελούνται από καθημερινές ωριαίες μετρήσεις θερμοκρασίας, βροχόπτωσης και ταχύτητας ανέμου τα έτη 2021 έως 2025 από τον μετεωρολογικό σταθμό του ΕΑΑ στο Θησείο τα οποία ανέρχονται σε 43,824 παρατηρήσεις. Καθώς οι δύο βάσεις δεδομένων έχουν διαφορετική συχνότητα μετρήσεων, έγινε αντιστοίχιση των μετεωρολογικών δεδομένων σε κάθε καταγραφή της ταχύτητας. Για παράδειγμα, η μέτρηση των 15:20 και των 15:40 πήραν την ίδια τιμή θερμοκρασίας (αυτή στις 15:00) ενώ η μέτρηση στις 16:00 αντιστοιχήθηκε στην μετεωρολογική μέτρηση της ώρας 16:00. Για τα έτη 2021 και 2026 δεν έγινε συγχώνευση καθώς δεν υπάρχουν κοινές μετρήσεις από τις δύο βάσεις.

Για την εφαρμογή της μεθοδολογίας, έγινε η υπόθεση ότι οι μετρημένες καιρικές συνθήκες από τον σταθμό του ΕΑΑ στο Θησείο, επιδρούν ταυτόχρονα στους δώδεκα υπό μελέτη άξονες που βρίσκονται σε ακτίνα 4.3 χιλιομέτρων από τον σταθμό καταγραφής (Εικόνα 6).



Εικόνα 6: Ακτίνα επιρροής μετεωρολογικών συνθηκών απο σταθμό ΕΑΑ

5.3.2. Επεξήγηση και καθορισμός μεταβλητών

Για καλύτερη ερμηνευσιμότητα του μοντέλου ως προς την επίδραση των καιρικών φαινομένων στην κυκλοφορία, η εξαρτημένη μεταβλητή ορίστηκε ως λόγος της ταχύτητας παρατήρησης προς την ταχύτητα ελεύθερης ροής του οδικού τμήματος, ο οποίος αποτελεί και δείκτη συμφόρησης:

$$speed_ratio = \frac{\mu\acute{\epsilon}\sigma\eta\ \tau\alpha\chi\acute{\upsilon}\tau\eta\tau\alpha\ \tau\mu\acute{\eta}\mu\alpha\tau\sigma\varsigma\ \sigma\epsilon\ \mu\acute{\iota}\alpha\ \pi\alpha\rho\alpha\tau\acute{\eta}\rho\eta\sigma\eta}{\tau\alpha\chi\acute{\upsilon}\tau\eta\tau\alpha\ \epsilon\lambda\epsilon\acute{\upsilon}\theta\epsilon\rho\eta\varsigma\ \rho o\acute{\eta}\varsigma\ \tau o\upsilon\ \tau\mu\acute{\eta}\mu\alpha\tau\sigma\varsigma}$$

Η ταχύτητα ελεύθερης ροής του κάθε τμήματος ορίστηκε η τιμή του 99ου εκατοστημορίου των συνολικών παρατηρήσεων για την αποφυγή ακραίων τιμών (Πίνακας 4.8). Ως ανεξάρτητες μεταβλητές, ορίστηκαν οι καιρικές (βροχόπτωση, άνεμος, βροχή, βροχή 1 ώρα πριν και θερμοκρασία), και οι μεταβλητές ελέγχου (όνομα οδικού τμήματος, ώρα της ημέρας, ημέρα της εβδομάδας, μήνας του έτους και έτος) με στόχο την καλύτερη ακρίβεια και αξιοπιστία του μοντέλου (Πίνακας 5.10).

Μεταβλητές	Είδος	Περιγραφή
speed_ratio	συνεχής	ταχύτητα οδικού τμήματος σε μία παρατήρηση/ταχύτητα ελεύθερης ροής του τμήματος.
segment_id	κατηγορική	[1, 2, …, 24] (24 οδικά τμήματα)
Hour	κατηγορική	8, 9, 10, 12, 13, …, 23 (15 ώρες δεδομένων ταχύτητας)
year	κατηγορική	2022, 2023, 2024, 2025
month	κατηγορική	[1, 2, 3, …, 12] Μήνες του έτους (Ιανουάριος – Δεκέμβριος)
weekday_id	κατηγορική	[1, 2, …, 7] Ημέρες εβδομάδας (Δευτέρα – Κυριακή)
wind	κατηγορική	4 κατηγορίες: καθόλου (<2 m/s), μικρής έντασης (2-5 m/s), μεσαίας (5-8 m/s), μεγάλης (>8 m/s)
temperature	κατηγορική	4 κατηγορίες: πολύ κρύο (<10ºC), κρύο (10ºC <...< 20ºC), ζέστη (20ºC <…< 30ºC), πολύ ζέστη (>30ºC)
rain	κατηγορική	4 κατηγορίες: καθόλου βροχή, μικρής έντασης (0-2.5 mm/h), μεσαίας (2.5-7.6 mm/h), μεγάλης (>7.6 mm/h)
rain_lag1	κατηγορική	ένταση βροχής 1 ώρα πριν την παρατήρηση της ταχύτητας (4 κατηγ.)

Πίνακας 5.10: Μεταβλητές μοντέλου γραμμικής παλινδρόμησης (ταχύτητα – καιρός)

Καθώς χρησιμοποιήθηκαν κατηγορικές μεταβλητές, επιλέχθηκαν ως κατηγορίες αναφοράς οι εξής:

Μεταβλητή	Κατηγορία αναφοράς
temperature	πολύ κρύο (<10ºC)
wind	καθόλου (<2 m/s)
rain	καθόλου βροχή
rain_lag1	καθόλου βροχή
hour	8:00
month	1: Ιανουάριος
segment_id	1: Βασ. Σοφίας (Πανεπιστημίου – Βασ. Κωνσταντίνου)
weekday_id	1: Δευτέρα
year	1: 2022

Πίνακας 5.11: Κατηγορίες αναφοράς μοντέλου (ταχύτητα – καιρός)

5.3.3. Έλεγχος συσχέτισης μεταβλητών (πολυσυγγραμικότητα)

Για την ανίχνευση της πολυσυγγραμικότητας στο μοντέλο πολλαπλής γραμμικής παλινδρόμησης, χρησιμοποιήθηκε ο Συντελεστής Διόγκωσης της Διακύμανσης (Variance Inflation Factor – VIF) με τα αποτελέσματα να φαίνονται στον Πίνακα 5.12 (φθίνουσα σειρά):

C(temperature)[T.4_poly_zesti]	9.44
C(temperature)[T.3_zesti]	8.45
C(temperature)[T.2_kruo]	5.5
C(MONTH)[T.7]	4.24
C(MONTH)[T.8]	3.93
…	…

Πίνακας 5.12: Συντελεστές διόγκωσης της διακύμανσης (VIF) στο μοντέλο ταχύτητα – καιρός

Σύμφωνα με τις τιμές του πίνακα, οι μεταβλητές της θερμοκρασίας έχουν υψηλή συσχέτιση (VIF > 5) το οποίο πιθανώς οφείλεται στις μεταβλητές των μηνών και της ώρας μέσα στην ημέρα. Για παράδειγμα, μια μεσημεριανή ώρα του Ιουλίου συνήθως συμβαδίζει με πολύ υψηλές θερμοκρασίες ενώ μία απογευματινή ώρα

του χειμώνα, με χαμηλές θερμοκρασίες. Έτσι λοιπόν, οι μεταβλητές της θερμοκρασίας αφαιρούνται από το μοντέλο.

5.3.4. Ανάλυση και αποτελέσματα μοντέλου

Για την ανάλυση πολλαπλής γραμμικής παλινδρόμησης, χρησιμοποιήθηκε η Μέθοδος Ελαχίστων Τετραγώνων με διόρθωση σφαλμάτων κατά συστάδες (clustered standard errors). Η εξάρτηση των μετρήσεων ταχύτητας από τις προηγούμενες παρατηρήσεις έκρινε αναγκαία αυτή τη ρύθμιση έτσι ώστε οι τιμές p (p-values) να μη βγαίνουν πολύ μικρές και παρερμηνεύουν το μοντέλο. Για παράδειγμα, αν στις 15:00 σε ένα τμήμα υπάρχει συμφόρηση, η μέτρηση στις 15:20 του ίδιου τμήματος δεν είναι ανεξάρτητη πληροφορία διότι πιθανώς η συμφόρηση παραμένει ίδια. Έτσι, ομαδοποιώντας τις παρατηρήσεις ανά ημέρα καταγραφής, επιβλήθηκε στο μοντέλο να μην υποθέτει ανεξαρτησία μέσα στις παρατηρήσεις της ομάδας αυτής, παρά μόνο σε διαφορετικές. Τα αποτελέσματα του μοντέλου παρουσιάζονται σε πίνακα με φθίνουσα σειρά ως προς τις τιμές p:

	coef	std err	z	P>|z|	[0.025	0.975]
C(MONTH)[T.7]	0.001	0.009	0.112	0.911	-0.016	0.018
C(wind)[T.4_entonos]	-0.0011	0.007	-0.148	0.882	-0.015	0.013
C(HOUR)[T.20]	0.0027	0.006	0.477	0.634	-0.008	0.014
C(rain_lag1)[T.2_mikri]	-0.0071	0.007	-0.998	0.318	-0.021	0.007
C(wind)[T.2_ligo]	-0.0034	0.003	-1.297	0.195	-0.008	0.002
C(rain_lag1)[T.3_mesaia]	-0.0163	0.009	-1.821	0.069	-0.034	0.001
C(rain)[T.2_mikri]	-0.013	0.007	-1.945	0.052	-0.026	0
C(rain_lag1)[T.4_megali]	-0.033	0.016	-2.055	0.04	-0.065	-0.002
C(wind)[T.3_mesaios]	-0.0088	0.004	-2.353	0.019	-0.016	-0.001
C(rain)[T.3_mesaia]	-0.0292	0.012	-2.397	0.017	-0.053	-0.005
C(rain)[T.4_megali]	-0.0329	0.013	-2.447	0.014	-0.059	-0.007
C(MONTH)[T.4]	-0.028	0.011	-2.6	0.009	-0.049	-0.007
C(MONTH)[T.6]	-0.0225	0.008	-2.713	0.007	-0.039	-0.006
C(segment_id)[T.23]	-0.0072	0.003	-2.784	0.005	-0.012	-0.002
C(MONTH)[T.3]	-0.028	0.01	-2.892	0.004	-0.047	-0.009
C(MONTH)[T.5]	-0.0324	0.01	-3.159	0.002	-0.052	-0.012
Intercept	0.7768	0.01	76.026	0	0.757	0.797
C(weekday_id)[T.2]	-0.0314	0.007	-4.636	0	-0.045	-0.018
C(weekday_id)[T.3]	-0.0394	0.006	-6.088	0	-0.052	-0.027
C(weekday_id)[T.4]	-0.0474	0.007	-6.906	0	-0.061	-0.034
C(weekday_id)[T.5]	-0.0482	0.007	-6.888	0	-0.062	-0.034
C(weekday_id)[T.6]	0.1181	0.007	17.871	0	0.105	0.131
C(weekday_id)[T.7]	0.2124	0.007	30.807	0	0.199	0.226
C(HOUR)[T.9]	-0.0476	0.001	-49.207	0	-0.05	-0.046
C(HOUR)[T.10]	-0.0269	0.004	-6.829	0	-0.035	-0.019
C(HOUR)[T.12]	-0.0486	0.002	-19.97	0	-0.053	-0.044
C(HOUR)[T.13]	-0.0533	0.003	-19.817	0	-0.059	-0.048
C(HOUR)[T.14]	-0.0719	0.003	-27.171	0	-0.077	-0.067
C(HOUR)[T.15]	-0.0787	0.003	-30.214	0	-0.084	-0.074

C(HOUR)[T.16]	-0.0609	0.003	-23.293	0	-0.066	-0.056
C(HOUR)[T.17]	-0.0656	0.003	-25.761	0	-0.071	-0.061
C(HOUR)[T.18]	-0.0706	0.003	-28.046	0	-0.076	-0.066
C(HOUR)[T.19]	-0.0398	0.005	-8.68	0	-0.049	-0.031
C(HOUR)[T.21]	-0.0338	0.004	-7.667	0	-0.042	-0.025
C(HOUR)[T.22]	0.0353	0.004	7.929	0	0.027	0.044
C(HOUR)[T.23]	0.0695	0.004	15.63	0	0.061	0.078
C(MONTH)[T.2]	-0.0428	0.008	-5.285	0	-0.059	-0.027
C(MONTH)[T.8]	0.1277	0.01	13.225	0	0.109	0.147
C(MONTH)[T.9]	-0.0604	0.008	-7.14	0	-0.077	-0.044
C(MONTH)[T.10]	-0.0618	0.008	-7.27	0	-0.078	-0.045
C(MONTH)[T.11]	-0.0774	0.008	-9.334	0	-0.094	-0.061
C(MONTH)[T.12]	-0.0657	0.011	-6.126	0	-0.087	-0.045
C(YEAR)[T.2023]	-0.0211	0.005	-4.636	0	-0.03	-0.012
C(YEAR)[T.2024]	-0.0334	0.005	-6.59	0	-0.043	-0.023
C(YEAR)[T.2025]	-0.0742	0.005	-16.017	0	-0.083	-0.065
C(segment_id)[T.2]	-0.1011	0.002	-48.478	0	-0.105	-0.097
C(segment_id)[T.3]	-0.0512	0.002	-30.639	0	-0.054	-0.048
C(segment_id)[T.4]	-0.0148	0.002	-7.843	0	-0.019	-0.011
C(segment_id)[T.5]	-0.0929	0.002	-47.234	0	-0.097	-0.089
C(segment_id)[T.6]	-0.1579	0.003	-59.449	0	-0.163	-0.153
C(segment_id)[T.7]	-0.1111	0.002	-50.634	0	-0.115	-0.107
C(segment_id)[T.8]	-0.07	0.002	-32.474	0	-0.074	-0.066
C(segment_id)[T.9]	-0.0814	0.002	-44.852	0	-0.085	-0.078
C(segment_id)[T.10]	-0.1481	0.002	-62.945	0	-0.153	-0.143
C(segment_id)[T.11]	-0.2366	0.003	-81.992	0	-0.242	-0.231
C(segment_id)[T.12]	-0.1161	0.002	-51.92	0	-0.12	-0.112
C(segment_id)[T.13]	-0.1831	0.002	-90.139	0	-0.187	-0.179
C(segment_id)[T.14]	-0.1131	0.002	-61.195	0	-0.117	-0.11
C(segment_id)[T.15]	-0.1325	0.003	-51.231	0	-0.138	-0.127
C(segment_id)[T.16]	-0.1371	0.002	-62.782	0	-0.141	-0.133
C(segment_id)[T.17]	-0.0478	0.002	-25.368	0	-0.051	-0.044
C(segment_id)[T.18]	-0.0794	0.003	-30.705	0	-0.084	-0.074
C(segment_id)[T.19]	0.0844	0.003	26.615	0	0.078	0.091
C(segment_id)[T.20]	0.0706	0.002	38.185	0	0.067	0.074
C(segment_id)[T.21]	-0.2064	0.002	-83.809	0	-0.211	-0.202
C(segment_id)[T.22]	-0.0519	0.003	-19.636	0	-0.057	-0.047
C(segment_id)[T.24]	-0.1119	0.002	-51.422	0	-0.116	-0.108

Πίνακας 5.13: Συντελεστές και τιμές p των μεταβλητών του μοντέλου ταχύτητα - καιρός

R^2	0.476
RMSE	0.142
MAE	0.11
N	954337
Durbin-Watson	0.432
Prob(Omnibus)	0
Prob(JB)	0

Skew	0.051
Kurtosis	3.793
Cond. No.	81.5

Πίνακας 5.14: Δείκτες προσαρμογής και έλεγχοι καταλοίπων του μοντέλου ταχύτητα – καιρός

Ο συντελεστής προσδιορισμού του μοντέλου (R^2 = 0.476) δείχνει μέτρια επεξηγηματική ισχύ και ο δείκτης κατάστασης (Cond. No.) την χαμηλή πολυσυγγραμμικότητα του μοντέλου. Η υψηλή αυτοσυσχέτιση μετρημένη με τον δείκτη Durbin-Watson είναι αναμενόμενη λόγω της δομής χρονοσειράς των δεδομένων, το οποίο διορθώθηκε μέσω ομαδοποίησης (clustering). Ο αριθμός των παρατηρήσεων (N = 954,337) είναι μικρότερος από το σύνολο των δεδομένων της Google (954,553) καθώς το μοντέλο αφαίρεσε 216 κενές τιμές λόγω έλλειψης τριών ωριαίων παρατηρήσεων της μεταβλητής του ανέμου.

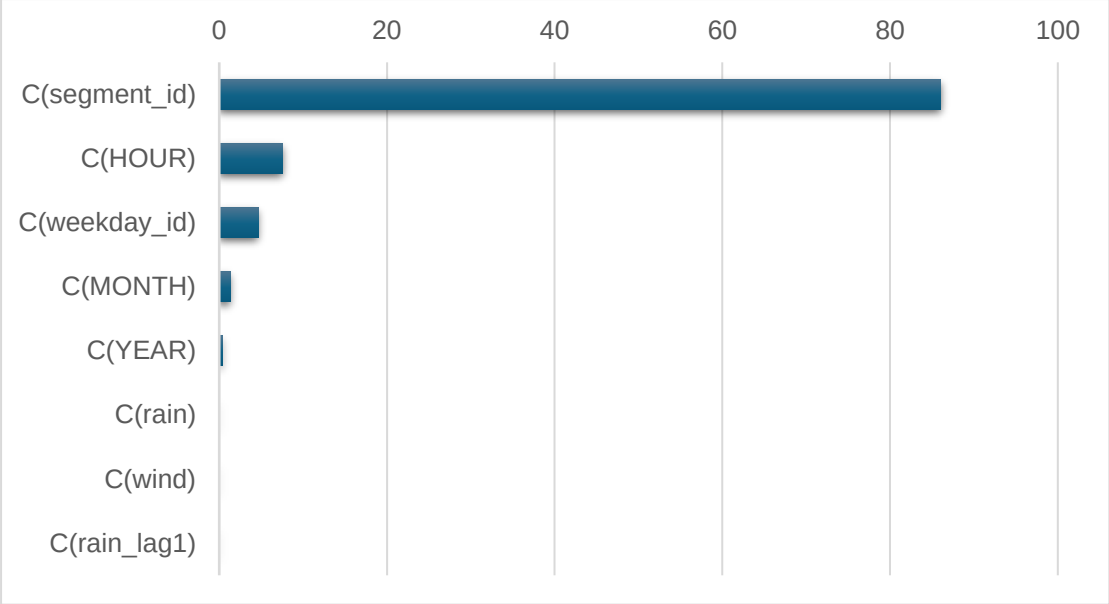
WALD TESTS	statistic	pvalue	df_constraint
C(segment_id)	64319.5	0.00E+00	23
Intercept	5780.0	0.00E+00	1
C(HOUR)	5678.6	0.00E+00	14
C(weekday_id)	3474.0	0.00E+00	6
C(MONTH)	990.5	2.18E-205	11
C(YEAR)	281.8	8.58E-61	3
C(rain)	11.8	7.93E-03	3
C(wind)	7.7	5.18E-02	3
C(rain_lag1)	6.2	1.02E-01	3

Πίνακας 5.15: Έλεγχος συνολικής σημαντικότητας μεταβλητών με Wald Test (ταχύτητα – καιρός)

Ο έλεγχος Wald (Wald Test) αναδεικνύει τη συνολική στατιστική σημαντικότητα των συντελεστών. Στο μοντέλο που αναπτύχθηκε, οι μεταβλητές της βροχής μίας ώρας πριν (rain_lag1) και του ανέμου (wind) αναδείχθηκαν μη σημαντικές ($p > 0.05$). Η σχετική σημαντικότητα (%) υπολογίστηκε διαιρώντας τη στατιστική Wald κάθε μεταβλητής με το άθροισμα όλων των στατιστικών χωρίς τη σταθερά (intercept) (Πίνακας 5.16).

Μεταβλητή	Σημαντικότητα (%)	pvalue
C(segment_id)	86	0.00E+00
C(HOUR)	7.6	0.00E+00
C(weekday_id)	4.6	0.00E+00
C(MONTH)	1.3	2.18E-205
C(YEAR)	0.4	8.58E-61
C(rain)	0.0	7.93E-03
C(wind)	0.0	5.18E-02
C(rain_lag1)	0.0	1.02E-01

Πίνακας 5.16: Σχετική σημαντικότητα μεταβλητών στο μοντέλο ταχύτητα – καιρός



Διάγραμμα 5-3: Σχετική σημαντικότητα μεταβλητών στο μοντέλο ταχύτητα – καιρός

5.3.5. Ερμηνεία μοντέλου

Από τον Πίνακα 5.13 εξάγονται συμπεράσματα για τη στατιστική σημαντικότητα κάθε μεταβλητής αλλά και για την ερμηνεία των συντελεστών.

- Ερμηνεία σταθεράς β_0:
  Η σταθερά του μοντέλου (β_0 = 0.7768) αντιπροσωπεύει την αναμενόμενη τιμή της εξαρτημένης μεταβλητής Y (= παρατηρούμενη ταχύτητα/ταχύτητα ελεύθερης ροής) για τις συνθήκες που περιγράφονται στις κατηγορίες αναφοράς (ημέρα: Δευτέρα, μήνας: Ιανουάριος έτος: 2022, τμήμα: Βασιλίσσης Σοφίας (Πανεπιστημίου - Βασιλέως Κωνσταντίνου), ώρα: 8:00, χωρίς βροχή εκείνη την ώρα, χωρίς βροχή μία ώρα πριν και χωρίς άνεμο). Δηλαδή, για τις παραπάνω συνθήκες, η μέση ταχύτητα των οχημάτων διαμορφώνεται στο 77.7% της ταχύτητας ελεύθερης ροής.

- Ερμηνεία συντελεστών:
  Οι συντελεστές των σημαντικών μεταβλητών ($p<0.05$) δείχνουν τη μεταβολή της μέσης ταχύτητας σε σχέση με την κατάσταση βάσης.

  1. Μετεωρολογικές μεταβλητές:

     - Μικρή ένταση βροχής (<2.5 mm/h): Συντελεστής -0.013. Μειώνει την ταχύτητα κατά 1.3% της ταχύτητας ελεύθερης ροής (Free Flow Speed – FFS).
     - Μεσαία ένταση βροχής (2.5 – 7.6 mm/h): Συντελεστής -0.0292. Μειώνει την ταχύτητα κατά 2.9% της FFS.
     - Μεγάλη ένταση βροχής (>7.6 mm/h): Συντελεστής -0.0329. Προκαλεί τη μέγιστη άμεση επιβράδυνση κατά 3.3% της FFS.
     - Μεγάλη ένταση βροχής 1 ώρα πριν: Συντελεστής -0.033 Μειώνει την ταχύτητα κατά 3.3% της FFS.

2. Μηνιαία περιοδικότητα:

- Αύγουστος: Συντελεστές +0.1277. Καταγράφεται η μέγιστη αύξηση ταχύτητας: +12.7% της FFS.
- Νοέμβριος: Συντελεστής -0.0774. Ο μήνας με τη μεγαλύτερη μείωση στη μέση ταχύτητα: -7.7% της FFS σε σχέση με τον Ιανουάριο.
- Δεκέμβριος: Συντελεστής -0.0657. Μείωση ταχύτητας κατά 6.6% της FFS σε σύγκριση με τον Ιανουάριο.
- Υπόλοιποι μήνες: Αρνητικοί συντελεστές που υποδεικνύουν μείωση της ταχύτητας σε σύγκριση με τον Ιανουάριο

3. Ημερήσια περιοδικότητα:

- Καθημερινές (Τρίτη έως Παρασκευή): Οι συντελεστές κυμαίνονται από -0.314 (Τρίτη) έως -0.0482 (Παρασκευή). Η Παρασκευή παρουσιάζει τη χαμηλότερη μέση ταχύτητα (-4.8% έναντι της Δευτέρας), αποτελώντας την πιο επιβαρυμένη ημέρα των εργάσιμων ημερών.
- Σαββατοκύριακο: Σημαντική βελτίωση των συνθηκών ροής. Το Σάββατο καταγράφει αύξηση ταχύτητας κατά +11.8% (συντελεστής 0.1181) και η Κυριακή κατά +21.2% (συντελεστής 0.2124).

4. Ωριαία Διακύμανση

- Οι συντελεστές των μεταβλητών των ωρών 9:00 – 21:00 είναι αρνητικοί, το οποίο υποδεικνύει επιδείνωση της συμφόρησης σε αυτό το χρονικό διάστημα σε σχέση με τις 8:00.
- Η ώρα 15:00 αποτελεί το σημείο της μέγιστης ημερήσιας συμφόρησης στο κέντρο, μειώνοντας την ταχύτητα κατά 7.87% της FFS.
- Στις 23:00 παρατηρείται αύξηση κατά 6.95%.

5. Οδικά τμήματα

- Το οδικό τμήμα: Σταδίου (Αιόλου - Βασ. Γεωργίου) (C(segment_id)[T.11]) εμφανίζει τον χαμηλότερο συντελεστή (-0.2366), γεγονός που σημαίνει ότι το οχήματα κινούνται σταθερά κατά μέσο όρο 23.6% πιο αργά από τη Βασ. Σοφίας.
- Το οδικό τμήμα: Πατησίων (Στουρνάρη – Ιουλιανού) (C(segment_id)[T.19]) εμφανίζει συντελεστή +0.0844, δηλαδή κίνηση οχημάτων κατά 8.4% ταχύτερα από τον άξονα αναφοράς.

5.4. Μοντέλο Ενίσχυσης Κλίσης Βασισμένη σε Ιστόγραμμα (ταχύτητα - καιρός)

5.4.1. Γενικά

Το μοντέλο υλοποιήθηκε με τον αλγόριθμο Histogram Gradient Boosting Regressor της βιβλιοθήκης scikit-learn, μια υλοποίηση του Gradient Boosting με δέντρα αποφάσεων που ομαδοποιεί τις τιμές των μεταβλητών σε

ιστογράμματα έως 255 κάδους (bins) και υποστηρίζει εγγενώς κατηγορικές μεταβλητές και ελλιπείς τιμές.

5.4.2. Επεξήγηση Μεταβλητών

Οι μεταβλητές που εισήχθησαν στο μοντέλο HGB είναι οι ίδιες με αυτές του μοντέλου Γραμμικής Παλινδρόμησης (Πίνακας 5.10), χωρίς την μεταβλητή year καθώς χρησιμοποιείται μόνο για τον διαχωρισμό εκπαίδευσης και ελέγχου του μοντέλου.

5.4.3. Διαχωρισμός δεδομένων εκπαίδευσης / ελέγχου

Για την αξιολόγηση της προβλεπτικής ικανότητας του μοντέλου εφαρμόστηκε χρονικός διαχωρισμός των δεδομένων. Το σύνολο εκπαίδευσης (training set) περιλαμβάνει τις παρατηρήσεις των ετών 2022 έως 2024 (N_1 = 698,904 εγγραφές), ενώ το σύνολο ελέγχου (test set) περιλαμβάνει τις παρατηρήσεις του έτους 2025 (N_2 = 255,649 εγγραφές). Ο διαχωρισμός αυτός προτιμήθηκε έναντι του τυχαίου διαχωρισμού, διότι οι διαδοχικές μετρήσεις της κυκλοφοριακής κατάστασης είναι χρονικά αυτοσυσχετισμένες. Σε έναν τυχαίο διαχωρισμό, παρατηρήσεις της ίδιας ώρας ή ημέρας θα βρίσκονταν και στα δύο σύνολα, με αποτέλεσμα η επίδοση του μοντέλου να υπερεκτιμάται.

5.4.4. Επιλογή και ρύθμιση υπερπαραμέτρων

Ο ρυθμός μάθησης και ο αριθμός των επαναλήψεων σχετίζονται στενά καθώς ένας μικρότερος ρυθμός μάθησης περιορίζει τη συνεισφορά κάθε δέντρου και απαιτεί περισσότερες επαναλήψεις, αλλά συνήθως βελτιώνει τη γενίκευση. Ο μέγιστος αριθμός δέντρων που ορίσθηκε στο μοντέλο (επαναλήψεις) είναι 300 και ο ρυθμός μάθησης (learning rate) 0.08. Οι υπόλοιπες παράμετροι διατηρήθηκαν στις προεπιλεγμένες τιμές της βιβλιοθήκης, δηλαδή συνάρτηση απώλειας μέσου τετραγωνικού σφάλματος, 31 φύλλα ανά δέντρο, ελάχιστο μέγεθος φύλλου 20 παρατηρήσεις και μηδενική L2 ποινή. Επειδή το σύνολο εκπαίδευσης υπερβαίνει τις 10.000 παρατηρήσεις, ήταν ενεργός ο μηχανισμός πρόωρου τερματισμού (early stopping), ο οποίος διακόπτει την εκπαίδευση όταν η απώλεια σε τυχαίο 10% του συνόλου εκπαίδευσης δεν βελτιώνεται για 10 διαδοχικές επαναλήψεις.

5.4.5. Αποτελέσματα και αξιολόγηση μοντέλου

Στο σύνολο ελέγχου του 2025 το μοντέλο πέτυχε MAE = 0.100, RMSE = 0.132 και R^2 = 0.507. Το μέσο απόλυτο σφάλμα αντιστοιχεί σε περίπου 10% της ταχύτητας ελεύθερης ροής, ενώ το μοντέλο εξηγεί περίπου το 51% της μεταβλητότητας του λόγου ταχύτητας προς ταχύτητα ελεύθερης ροής σε δεδομένα που δεν χρησιμοποιήθηκαν στην εκπαίδευση. Η υψηλότερη τιμή του RMSE σε σχέση με το MAE δείχνει ότι υπάρχουν και μεμονωμένες παρατηρήσεις με μεγαλύτερο σφάλμα, πιθανώς σε περιόδους έντονης συμφόρησης ή έκτακτων συμβάντων.

MAE:	0.100
RMSE:	0.132

R^2:	0.507

Πίνακας 5.17: Δείκτες αξιολόγησης μοντέλου HGBM (ταχύτητα – καιρός)

Σύμφωνα με τη μεταθετική σημαντικότητα (Permutation Importance), ο καθοριστικότερος παράγοντας ήταν η ημέρα της εβδομάδας (weekday_id, 0.60), ακολουθούμενο από το οδικό τμήμα (0.57), τον μήνα (0.21) και την ώρα (0.19). Σε σχετική βάση, το τμήμα και η ημέρα της εβδομάδας εξηγούν μαζί περίπου το 74.5% της συνολικής σημαντικότητας, ο μήνας και η ώρα το 25.5%, και οι καιρικές μεταβλητές πρακτικά μηδέν. Αυτό σημαίνει ότι, στο συγκεκριμένο σύνολο δεδομένων, με την παρούσα κατηγοριοποίηση και καταγραφή, ο καιρός δεν προσφέρει προβλεπτική πληροφορία πέρα από τους χρονικούς παράγοντες.

Μεταβλητή	Σημαντικότητα (Permutation Importance)	Σταθερό Σφάλμα
weekday_id	0.60	0.0039
segment_id	0.57	0.0044
month	0.21	0.0024
hour	0.19	0.0024
rain	0.00	0.0002
wind	0.00	0.0003
rain_lag1	0.00	0.0002
temperature	-0.01	0.0003

Πίνακας 5.18: Μεταθετική σημαντικότητα μεταβλητών μοντέλου HGBM (ταχύτητα – καιρός)

Μεταβλητή	Σχετική Σημαντικότητα (%)
segment_id	38.4
weekday_id	36.1
month	12.0
hour	13.5
wind	0.0
temperature	0.0
rain	0.0
rain_lag1	0.0

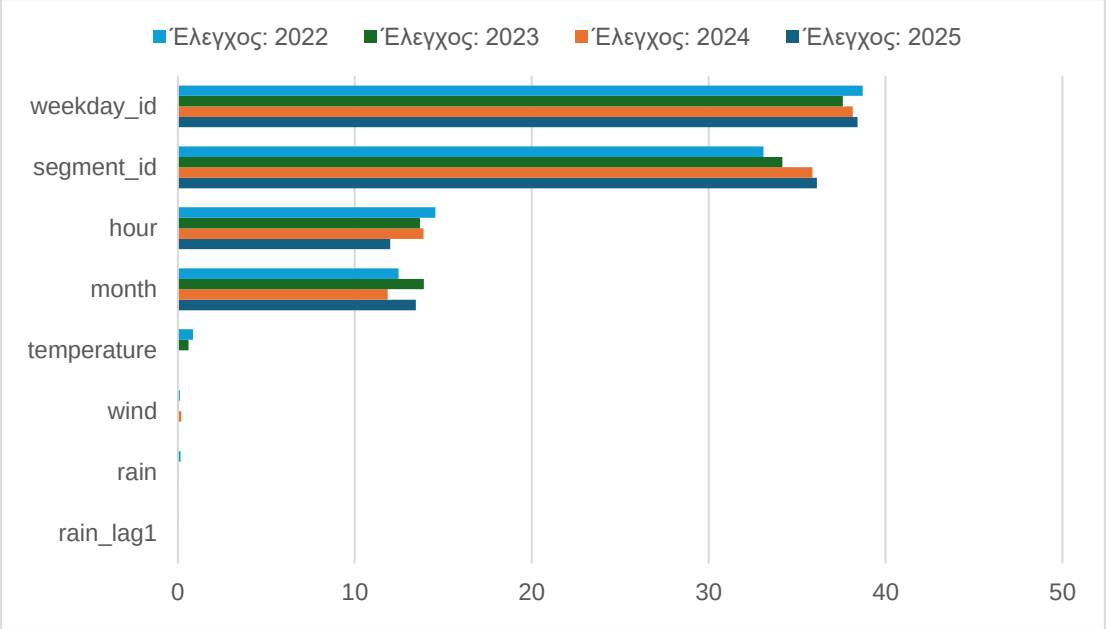
Πίνακας 5.19: Σχετική σημαντικότητα μεταβλητών μοντέλου HGBM (ταχύτητα – καιρός)

Καθώς το κάθε έτος έχει διαφορετικά καιρικά και κυκλοφοριακά χαρακτηριστικά, δοκιμάσθηκε η εισαγωγή διαφορετικού χρονικού διαχωρισμού για τα δεδομένα εκπαίδευσης και ελέγχου.

Έλεγχος: 2025		Έλεγχος: 2024		Έλεγχος: 2023		Έλεγχος: 2022	
Train set	698,904	Train set	721,129	Train set	721,921	Train set	721,705
Test set	255,649	Test set	233,424	Test set	232,632	Test set	232,848
MAE:	0.100	MAE:	0.097	MAE:	0.081	MAE:	0.100
RMSE:	0.132	RMSE:	0.132	RMSE:	0.113	RMSE:	0.136
R^2:	0.507	R^2:	0.543	R^2:	0.660	R^2:	0.545

weekday_id	38	weekday_id	38	weekday_id	38	weekday_id	39
segment_id	36	segment_id	36	segment_id	34	segment_id	33
hour	12	hour	14	hour	14	hour	15
month	13	month	12	month	14	month	12
temperature	0	temperature	0	temperature	1	temperature	1
wind	0	wind	0	wind	0	wind	0
rain	0	rain	0	rain	0	rain	0
rain_lag1	0	rain_lag1	0	rain_lag1	0	rain_lag1	0

Πίνακας 5.20: Αποτελέσματα μοντέλων για διαφορετικά δεδομένα ελέγχου (ταχύτητα – καιρός)



Διάγραμμα 5-4: Σχετική σημαντικότητα μεταβλητών στο μοντέλο HGBM ταχύτητα – καιρός

Ο Πίνακας 5.20, αποδεικνύει πως ανεξάρτητα των δεδομένων εκπαίδευσης και ελέγχου, δεν υπάρχει σαφής επίδραση των καιρικών συνθηκών στη μέση ταχύτητα κίνησης. Αυτό ενδεχομένως οφείλεται στις ήδη χαμηλές ταχύτητες που επικρατούν στο κέντρο της πόλης λόγω αυξημένης συμφόρησης. Εξ άλλου, όπως αναδείχθηκε και στην προκαταρκτική ανάλυση στην ενότητα 4.5.1, οι συμφωνημένες ώρες στο κέντρο προσεγγίζουν το 52% των συνολικών ωρών ανάλυσης (8 – 10:00 & 12 – 19:00). Επιπλέον, ο σταθμός μέτρησης είναι ένας και μακριά από κάποια τμήματα με αποτέλεσμα η μέτρηση είναι χονδρική.

5.5. Σύγκριση μοντέλου HGBM με Γραμμική Παλινδρόμηση (ταχύτητα - καιρός)

Και τα δύο μοντέλα, αναδεικνύουν τα χωρικά χαρακτηριστικά ως τους σημαντικότερους παράγοντες που καθορίζουν την ταχύτητα κίνησης. Η βαρύτητά τους διαφέρει όμως σημαντικά: στη γραμμική παλινδρόμηση, όπου η σημαντικότητα προκύπτει από τους ελέγχους Wald, η σχετική σημαντικότητα της μεταβλητής «segment_id» φτάνει το 86%, ενώ στο HGBM, όπου υπολογίστηκε με μεταθετική σημαντικότητα (permutation importance), περιορίζεται στο 35%. Η διαφορά οφείλεται κυρίως στον τρόπο με τον οποίο

κάθε μοντέλο κατανέμει την επεξηγηματική ισχύ στις χρονικές μεταβλητές. Το HGBM αποδίδει μεγάλη σημασία στην ημέρα της εβδομάδας (weekday_id, 38%), ενώ στη γραμμική παλινδρόμηση η ίδια μεταβλητή συμβάλλει μόλις 5%. Κάτι αντίστοιχο ισχύει για την ώρα: 14% έναντι 8% και για τον μήνα: 13% έναντι 1%.

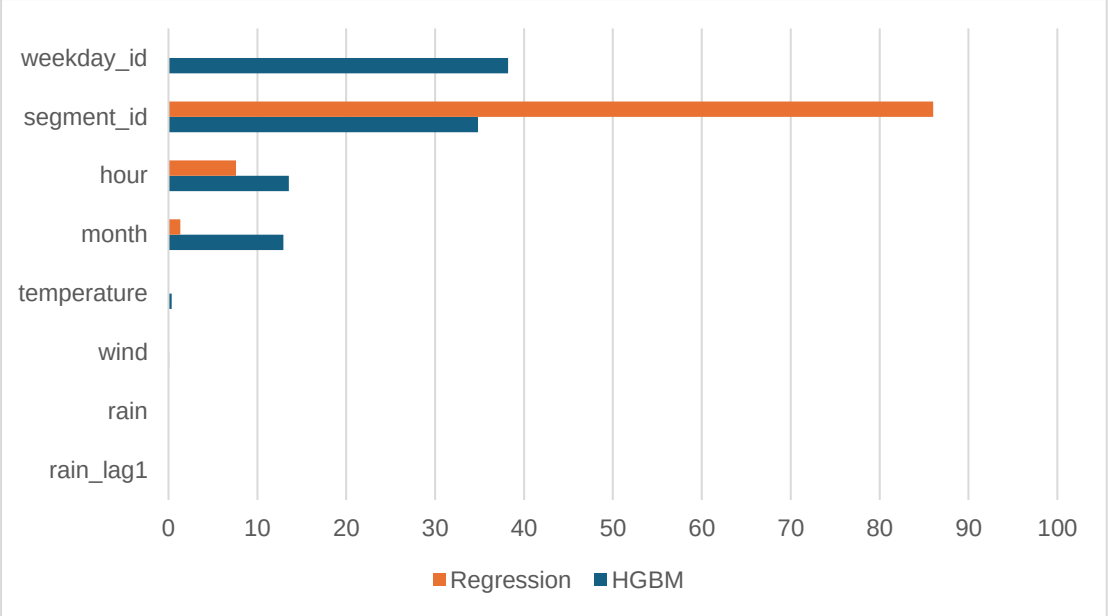
Μεταβλητή	Παλινδρόμηση	HGBM (μέσος όρος ετών ελέγχου)
weekday_id	5	38
segment_id	86	35
month	1	13
hour	8	14
wind	0	0
temperature	0	0
rain	0	0
rain_lag1	0	0

Πίνακας 5.21: Σύγκριση σημαντικότητας μεταβλητών μεταξύ του μοντέλου παλινδρόμησης και HGBM (ταχύτητα – καιρός)

Αυτό μπορεί να σημαίνει ότι το HGBM αποτυπώνει μη γραμμικές σχέσεις και αλληλεπιδράσεις μεταξύ τμήματος και χρονικών μεταβλητών (π.χ. διαφορετικό εβδομαδιαίο ή ωριαίο προφίλ ανά οδό), τις οποίες το γραμμικό μοντέλο με προσθετικούς όρους δεν μπορεί να εκφράσει.

Συνολικά, οι μετεωρολογικές μεταβλητές (wind, temperature, rain, rain_lag1) έχουν πρακτικά μηδενική σημαντικότητα και στα δύο μοντέλα, άρα η επίδραση των καιρικών συνθηκών στον χρόνο διαδρομής είναι αμελητέα σε σχέση με τους χωρικούς και χρονικούς παράγοντες. Το μοντέλο παλινδρόμησης όμως ανέδειξε στατιστικά σημαντική αυτή την επίδραση ($p<0.05$) αν και αρκετά μικρή.

Θα πρέπει τέλος να σημειωθεί ότι τα δύο μέτρα δεν είναι άμεσα ισοδύναμα. Το permutation importance εκφράζει την πτώση της προβλεπτικής ικανότητας του μοντέλου όταν τυχαιοποιείται μία μεταβλητή, ενώ η στατιστική Wald ελέγχει τη στατιστική σημαντικότητα των συντελεστών και εξαρτάται και από τους βαθμούς ελευθερίας κάθε μεταβλητής, με αποτέλεσμα να ευνοεί τις κατηγορικές μεταβλητές με πολλά επίπεδα, όπως το segment_id.



Διάγραμμα 5-5: Σύγκριση σημαντικότητας μεταβλητών μεταξύ των μοντέλων παλινδρόμησης και HGBM (ταχύτητα – καιρός)

Για τη δίκαιη σύγκριση των δύο μοντέλων, οι δείκτες απόδοσης υπολογίστηκαν εκ νέου για το μοντέλο παλινδρόμησης με διαχωρισμό των δεδομένων ελέγχου και εκπαίδευσης ανά έτος.

Έτος ελέγχου	Παλινδρόμηση			HGBM		
	R^2	MAE	RMSE	R^2	MAE	RMSE
2022	0.389	0.124	0.158	0.545	0.100	0.136
2023	0.481	0.108	0.140	0.660	0.081	0.113
2024	0.399	0.117	0.151	0.543	0.097	0.132
2025	0.394	0.113	0.146	0.507	0.100	0.132

Πίνακας 5.22: Σύγκριση των δεικτών απόδοσης μεταξύ των μοντέλων παλινδρόμησης και HGBM (ταχύτητα - καιρός)

Ο Πίνακας 5.22 δείχνει ότι το HGBM υπερτερεί της γραμμικής παλινδρόμησης, με R^2 από 0.507 έως 0.660 στα έτη ελέγχου, έναντι 0.394 έως 0.481. Οι δείκτες MAE και RMSE είναι μικρότεροι στο μοντέλο HGBM για κάθε σύνολο ελέγχου, το οποίο αποδεικνύει την καλύτερη αποτελεσματικότητα στην πρόβλεψη της ταχύτητας κυκλοφορίας. Το αποτέλεσμα συμφωνεί με την ανάλυση σημαντικότητας: το HGBM αποτυπώνει μη γραμμικές σχέσεις και αλληλεπιδράσεις μεταξύ οδού και χρονικών μεταβλητών που το γραμμικό μοντέλο δεν μπορεί να εκφράσει. Η βελτίωση είναι ωστόσο μέτρια. Και τα δύο μοντέλα αφήνουν ανεξήγητο το 34 – 61% της διακύμανσης, κάτι που δείχνει ότι οι κυκλοφοριακές συνθήκες στα αστικά κέντρα ορίζονται σε μεγάλο βαθμό από απρόβλεπτους παράγοντες όπως για παράδειγμα ατυχήματα, έργα, ειδικά γεγονότα κ.α.

5.6. Μοντέλο Γραμμικής Παλινδρόμησης (ταχύτητα - σύνθεση κυκλοφορίας)

5.6.1. Επεξεργασία και συγχώνευση δεδομένων

Τα δεδομένα σύνθεσης της κυκλοφορίας έχουν αρκετά λιγότερες παρατηρήσεις από αυτές του καιρού (2,208 έναντι 35,064) το διάστημα 2022 - 2025, οπότε επιλέχθηκε ξεχωριστό μοντέλο για τη διερεύνηση της επιρροής επιλογής μέσου στην κυκλοφορία. Η συγχώνευση των δεδομένων έγινε ως εξής:

1) Αντιστοίχιση των καιρικών δεδομένων (συχνότητα μέτρησης: κάθε 1 ώρα) σε αυτά της ταχύτητας (συχνότητα μέτρησης: κάθε 20 λεπτά της ώρα) όπως έγινε και στο μοντέλο της Ενότητας 5.3.

2) Καθώς τα δεδομένα σύνθεσης κυκλοφορίας διαχωρίζονται σε τρία χρονικά διαστήματα με μία τιμή μέτρησης ανά διάστημα (8 – 10:00, 8 – 11:00 και 11 – 15:00), τα δεδομένα ταχύτητας (που συνοδεύονται ήδη με τα καιρικά δεδομένα), συμπυκνώνονται σε μία τιμή με βάση τον μέσο όρο αυτών στο αντίστοιχο χρονικό διάστημα.



Διάγραμμα 5-6: Ροή συγχώνευσης δεδομένων (ταχύτητα – σύνθεση κυκλοφορίας)

Οι μετρήσεις σύνθεσης κυκλοφορίας που αξιοποιήθηκαν περιλαμβάνουν τις καταγραφές οχημάτων που κινούνται στις αντίστοιχες οδούς για το χρονικό διάστημα 2021 – 2025 καθώς δεν υπάρχουν δεδομένα μέσων ταχυτήτων και καιρού το 2026. Οι μετρήσεις πεζών και οχημάτων στο πεζοδρόμιο, εξαιρέθηκαν από την ανάλυση.

Η αντιστοίχιση των οδών από τις δύο διαφορετικές βάσεις δεδομένων έγινε σύμφωνα με τον Πίνακα 5.23. Έγινε η υπόθεση πως οι καιρικές συνθήκες που καταγράφονται από τον σταθμό του ΕΑΑ στο Θησείο, επικρατούν ταυτόχρονα

σε όλα τα παρακάτω οδικά τμήματα το οποία βρίσκονται σε ακτίνα 4.3 χιλιόμετρα από αυτόν.

Δεδομένα Σύνθεσης Κυκλοφορίας		Δεδομένα Ταχύτητας από Google
Οδός	Κατεύθυνση	Οδός
Αθηνάς	Ομόνοια	Athinas (Ermou - Stadiou)
	Μοναστηράκι	Athinas (Stadiou - Ermou)
Ακαδημίας	-	Akadimias (Patision - Vas. Sofias)
Αλεξάνδρας	Κηφισίας	Alexandras (Patision - Kifisias)
	Πατησίων	Alexandras (Kifisias - Patision)
Βασ. Αμαλίας	-	Vas. Amalias (Ath. Diakou - Panepistimiou)
Βασ. Κων/νου	Χίλτον	Vas. Konstantinou (Arditou/ Ath. Diakou - Vas. Sofias)
	Συγγρού	Vas. Konstantinou (Vas. Sofias - Arditou/ Ath. Diakou)
Βασ. Σοφίας	Σύνταγμα	Vas. Sofias (Vas. Konstantinou - Panepistimiou)
	Ευαγγελισμός	Vas. Sofias (Panepistimiou - Vas. Konstantinou)
Ερμού	Πλατεία Μοναστηριακίου	Ermou (Ag. Asomaton - Athinas)
	Πλατεία Ασωμάτων	Ermou (Athinas - Ag. Asomaton)
Ευριπίδου	-	Evripidou (Eleftherias Sq. - Athinas)
Μητροπόλεως	Αιόλου	Mitropoleos (Syntagma Sq. - Aiolou)
Πανεπιστημίου	-	Panepistimiou (Vas. Sofias - Patision)
Πατησίων	Πλ. Ομονοίας	Patision (Alexandras - Stadiou)
	Λεωφ. Αλεξάνδρας	Patision (Stournari - Ioulianou)
Πειραιώς	Ιερά οδός	Pireos (Kolokinthous - Iera Odo)
	Πλ. Ομονοίας	Pireos (Kolokinthous - Omonoia Sq.)
Σόλωνος	-	Solonos (Vas. Sofias - Patision)
Σοφοκλέους	-	Sofokleous (Aristidou - P. Tsaldari)
Σταδίου	-	Stadiou (Aiolou - Vas. Georgiou)
Φιλελλήνων	-	Filellinon (Vas. Georgiou - Vas. Amalias)

Πίνακας 5.23: Αντιστοίχιση οδών ανάμεσα στις δύο βάσεις δεδομένων

5.6.2. Επεξήγηση και καθορισμός μεταβλητών

Όπως και στο μοντέλα ταχύτητα καιρού των προηγούμενων ενοτήτων, η εξαρτημένη μεταβλητή ορίσθηκε ως λόγος της ταχύτητας παρατήρησης προς την ταχύτητα ελεύθερης ροής του οδικού τμήματος:

$$speed_ratio = \frac{\mu\acute{\epsilon}\sigma\eta\ \tau\alpha\chi\acute{\upsilon}\tau\eta\tau\alpha\ \tau\mu\acute{\eta}\mu\alpha\tau\sigma\varsigma\ \sigma\epsilon\ \mu\acute{\iota}\alpha\ \pi\alpha\rho\alpha\tau\acute{\eta}\rho\eta\sigma\eta}{\tau\alpha\chi\acute{\upsilon}\tau\eta\tau\alpha\ \epsilon\lambda\epsilon\acute{\upsilon}\theta\epsilon\rho\eta\varsigma\ \rho o\acute{\eta}\varsigma\ \tau o\upsilon\ \tau\mu\acute{\eta}\mu\alpha\tau\sigma\varsigma}$$

Ως ανεξάρτητες μεταβλητές, ορίσθηκαν οι παρακάτω:

Μεταβλητές	Είδος	Περιγραφή
speed_ratio	συνεχής	ταχύτητα οδικού τμήματος σε μία παρατήρηση/ταχύτητα ελεύθερης ροής του τμήματος
segment_id	κατηγορική	[1, 2, …, 23] (23 οδικά τμήματα)
time_slot	κατηγορική	τρία χρονικά διαστήματα: 8 – 10:00, 8 – 11:00 & 11 – 15:00
year	κατηγορική	2022, 2023, 2024, 2025
month	κατηγορική	[1, 2, 3, …, 12] Μήνες του έτους (Ιανουάριος – Δεκέμβριος)
wind	κατηγορική	4 κατηγορίες: καθόλου (<2 m/s), μικρής έντασης (2-5 m/s), μεσαίας (5-8 m/s), μεγάλης (>8 m/s)
rain	κατηγορική	2 κατηγορίες: καθόλου βροχή, βροχή
pct_truck	συνεχής	Ποσοστό φορτηγών επί του συνόλου των οχημάτων
pct_bike	συνεχής	Ποσοστό ποδηλάτων επί του συνόλου των οχημάτων
pct_patini	συνεχής	Ποσοστό πατινιών επί του συνόλου των οχημάτων
pct_twowheel	συνεχής	Ποσοστό μοτοσυκλετών επί του συνόλου των οχημάτων
pct_bus	συνεχής	Ποσοστό λεωφορείων επί του συνόλου των οχημάτων
pct_taxi	συνεχής	Ποσοστό ταξί επί του συνόλου των οχημάτων

Πίνακας 5.24: Μεταβλητές μοντέλου γραμμικής παλινδρόμησης (ταχύτητα - σύνθεση κυκλοφορίας)

Η μεταβλητή του ποσοστού επιβατικών δεν εισήχθη στο μοντέλο καθώς έτσι θα υπήρχε τέλεια πολυσυγγραμικότητα αφού το άθροισμα των ποσοστών σύνθεσης θα ήταν 100%. Με αυτόν τον τρόπο λειτουργεί σαν κατηγορία αναφοράς και οι συντελεστές των υπόλοιπων ερμηνεύονται ως η επίδραση αύξησης/μείωσης 1 ποσοστιαίας μονάδας ενός είδους, στη θέση ενός ισοδύναμου ποσοστού IX.

Οι κατηγορίες αναφοράς των μεταβλητών ορίσθηκαν οι εξής:

Μεταβλητή	Κατηγορία αναφοράς
segment_id	Akadimias (Patision - Vas. Sofias)
time_slot	11:00 – 15:00
year	1: 2022
month	1: Ιανουάριος
wind	1: καθόλου άνεμος
rain	1: χωρίς βροχή

Πίνακας 5.25: Κατηγορίες αναφοράς μοντέλου γραμμικής παλινδρόμησης (ταχύτητα - σύνθεση κυκλοφορίας)

5.6.3. Έλεγχος συσχέτισης μεταβλητών (πολυσυγγραμικότητα)

Ο έλεγχος συσχέτισης μεταβλητών με τη μέθοδο VIF δεν ανέδειξε προβλήματα πολυσυγγραμικότητας μεταξύ των ανεξάρτητων μεταβλητών όπως φαίνεται στον παρακάτω πίνακα:

VIF ανά μεταβλητή (φθίνουσα σειρά)	
C(segment_id)[T.5]	3.17
C(segment_id)[T.7]	2.76
C(segment_id)[T.6]	2.62
C(segment_id)[T.8]	2.59
C(segment_id)[T.20]	2.51
C(segment_id)[T.21]	2.5
C(segment_id)[T.10]	2.47
C(wind)[T.3_mesaios]	2.46
pct_truck	2.45
pct_twowheel	2.43
pct_bus	2.41
pct_patini	2.37
…	…

Πίνακας 5.26: Συντελεστές Διόγκωσης Διακύμανσης (VIF) στο μοντέλο ταχύτητα - σύνθεση κυκλοφορίας

5.6.4. Ανάλυση και αποτελέσματα μοντέλου

Για την ανάλυση πολλαπλής γραμμικής παλινδρόμησης, χρησιμοποιήθηκε η Μέθοδος Ελαχίστων Τετραγώνων με διόρθωση σφαλμάτων κατά συστάδες (clustered standard errors) με ομαδοποίηση ανά ημέρα καταγραφής. Τα αποτελέσματα του μοντέλου φαίνονται στον Πίνακα 5.27.

	coef	std err	z	P>|z|	[0.025	0.975]
C(MONTH)[T.7]	-0.0015	0.025	-0.059	0.953	-0.05	0.047
C(wind)[T.3_mesaios]	0.0019	0.012	0.165	0.869	-0.021	0.025
C(segment_id)[T.11]	-0.0024	0.013	-0.187	0.852	-0.027	0.023
C(MONTH)[T.5]	-0.0055	0.025	-0.218	0.828	-0.055	0.044
C(MONTH)[T.12]	-0.0125	0.037	-0.335	0.738	-0.086	0.061
C(MONTH)[T.6]	-0.0084	0.024	-0.343	0.732	-0.056	0.039
C(segment_id)[T.12]	-0.0046	0.011	-0.41	0.682	-0.026	0.017
C(MONTH)[T.4]	0.013	0.028	0.471	0.638	-0.041	0.067
C(wind)[T.4_entonos]	0.0117	0.016	0.737	0.461	-0.019	0.043
pct_truck	0.0009	0.001	0.835	0.404	-0.001	0.003
C(MONTH)[T.3]	0.0202	0.023	0.882	0.378	-0.025	0.065
pct_bike	0.0019	0.002	0.943	0.346	-0.002	0.006
pct_patini	0.0047	0.005	0.994	0.32	-0.005	0.014
C(MONTH)[T.11]	-0.0266	0.025	-1.064	0.287	-0.076	0.022
C(MONTH)[T.2]	0.0308	0.026	1.206	0.228	-0.019	0.081
C(wind)[T.2_ligo]	-0.0122	0.009	-1.358	0.174	-0.03	0.005
pct_twowheel	-0.0006	0	-1.45	0.147	-0.001	0
C(YEAR)[T.2023]	-0.0145	0.009	-1.611	0.107	-0.032	0.003
C(YEAR)[T.2024]	-0.0282	0.015	-1.89	0.059	-0.057	0.001
C(segment_id)[T.17]	-0.0184	0.01	-1.887	0.059	-0.038	0.001
C(segment_id)[T.14]	0.0228	0.012	1.907	0.057	-0.001	0.046
C(segment_id)[T.20]	-0.0273	0.014	-1.925	0.054	-0.055	0
C(MONTH)[T.9]	-0.0536	0.026	-2.061	0.039	-0.105	-0.003
C(segment_id)[T.2]	-0.0263	0.011	-2.319	0.02	-0.049	-0.004

C(MONTH)[T.10]	-0.0568	0.021	-2.664	0.008	-0.099	-0.015
C(segment_id)[T.21]	-0.0399	0.014	-2.946	0.003	-0.066	-0.013
C(segment_id)[T.8]	0.0554	0.017	3.285	0.001	0.022	0.089
Intercept	0.4871	0.037	13.127	0	0.414	0.56
C(rain)[T.2_vrochi]	-0.0661	0.015	-4.477	0	-0.095	-0.037
C(time_slot)[T.8:00-10:00]	0.093	0.006	15.536	0	0.081	0.105
C(MONTH)[T.8]	0.1298	0.023	5.678	0	0.085	0.175
C(YEAR)[T.2025]	-0.0872	0.011	-8.031	0	-0.109	-0.066
C(segment_id)[T.3]	-0.06	0.009	-6.947	0	-0.077	-0.043
C(segment_id)[T.4]	0.1381	0.014	9.69	0	0.11	0.166
C(segment_id)[T.5]	0.1801	0.015	12.058	0	0.151	0.209
C(segment_id)[T.6]	0.1167	0.017	6.921	0	0.084	0.15
C(segment_id)[T.7]	0.1655	0.02	8.398	0	0.127	0.204
C(segment_id)[T.9]	0.164	0.015	11.171	0	0.135	0.193
C(segment_id)[T.10]	0.1444	0.019	7.712	0	0.108	0.181
C(segment_id)[T.13]	0.2367	0.016	14.457	0	0.205	0.269
C(segment_id)[T.15]	-0.0814	0.011	-7.568	0	-0.103	-0.06
C(segment_id)[T.16]	0.1151	0.013	8.734	0	0.089	0.141
C(segment_id)[T.18]	-0.1131	0.01	-11.53	0	-0.132	-0.094
C(segment_id)[T.19]	-0.0863	0.012	-7.06	0	-0.11	-0.062
C(segment_id)[T.22]	0.0909	0.011	8.503	0	0.07	0.112
C(segment_id)[T.23]	-0.0876	0.009	-9.336	0	-0.106	-0.069
pct_bus	0.0044	0.001	3.618	0	0.002	0.007
pct_taxi	0.0018	0	3.964	0	0.001	0.003

Πίνακας 5.27: Συντελεστές και τιμές p των μεταβλητών του μοντέλου ταχύτητα – σύνθεση

Οι διαγνωστικοί δείκτες δείχνουν ένα μοντέλο με καλή προσαρμογή (R^2 = 0.601), με ελάχιστη αυτοσυσχέτιση (DW = 2.601), και οριακή πολυσυγγραμικότητα (Cond. No. = 924) το οποίο πιθανώς οφείλεται στις σταθερές επιδράσεις (fixed effects) των μεταβλητών: segment_id, month, year.

R^2	0.601
RMSE	0.105
MAE	0.083
N	2208
Durbin-Watson	2.601
Prob(Omnibus)	0.411
Prob(JB)	0.412
Skew	-0.011
Kurtosis	3.137
Cond. No.	924

Πίνακας 5.28: Δείκτες προσαρμογής και έλεγχοι καταλοίπων του μοντέλου ταχύτητα – σύνθεση

Ο έλεγχος Wald (Wald Test) αναδεικνύει τη συνολική στατιστική σημαντικότητα των συντελεστών. Στο μοντέλο που αναπτύχθηκε, οι μεταβλητές των ποσοστών φορτηγών, ποδηλάτων, πατινιών, μοτοσυκλετών και του ανέμου αναδείχθηκαν μη σημαντικές ($p > 0.05$).

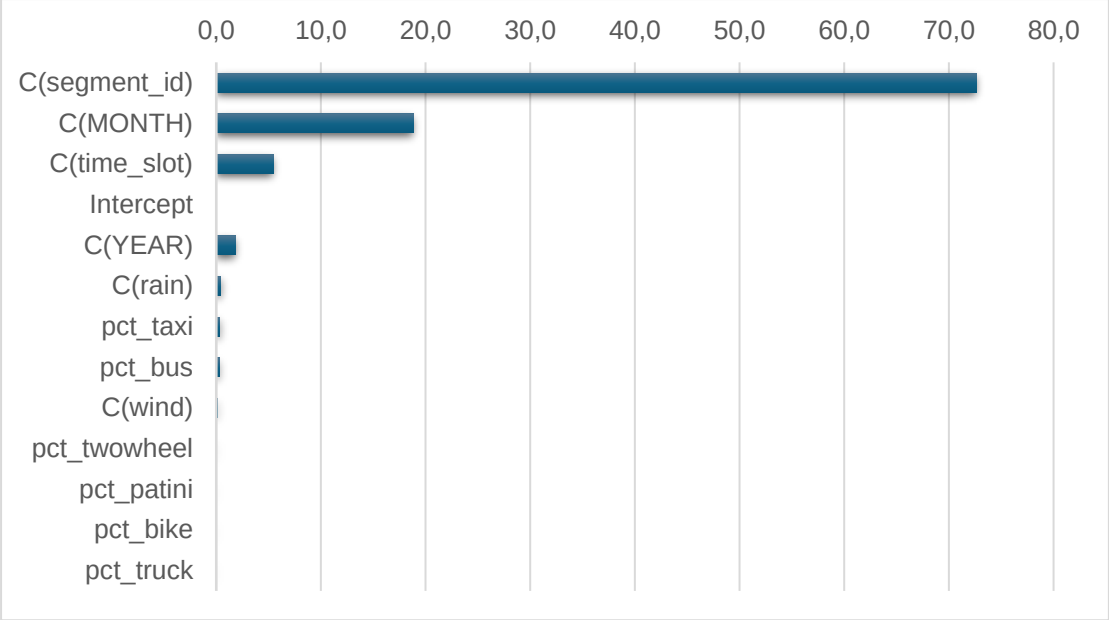
WALD TESTS	statistic	pvalue	df_constraint
C(segment_id)	3233.149	0.00E+00	22
C(MONTH)	837.6257	1.57E-172	11
C(time_slot)	241.3749	1.97E-54	1
Intercept	172.3153	2.31E-39	1
C(YEAR)	79.98294	3.10E-17	3
C(rain)	20.04078	7.58E-06	1
pct_taxi	15.70973	7.38E-05	1
pct_bus	13.09125	2.97E-04	1
C(wind)	5.780246	1.23E-01	3
pct_twowheel	2.102148	1.47E-01	1
pct_patini	0.987127	3.20E-01	1
pct_bike	0.88834	3.46E-01	1
pct_truck	0.697462	4.04E-01	1

Πίνακας 5.29: Έλεγχος συνολικής σημαντικότητας μεταβλητών με Wald Test (ταχύτητα – σύνθεση)

Η σχετική σημαντικότητα (%) υπολογίστηκε διαιρώντας τη στατιστική Wald κάθε μεταβλητής με το άθροισμα όλων των στατιστικών χωρίς τη σταθερά (intercept). Το αποτέλεσμα δείχνει υπεροχή των χωροχρονικών μεταβλητών έναντι των μεταβλητών σύνθεσης κυκλοφορίας με μικρή σημαντικότητα των ποσοστών των λεωφορείων και των ταξί οχημάτων.

Μεταβλητή	Σημαντικότητα (%)	pvalue
C(segment_id)	72.6	0.00E+00
C(MONTH)	18.8	1.57E-172
C(time_slot)	5.4	1.97E-54
Intercept	-	2.31E-39
C(YEAR)	1.8	3.10E-17
C(rain)	0.5	7.58E-06
pct_taxi	0.4	7.38E-05
pct_bus	0.3	2.97E-04
C(wind)	0.1	1.23E-01
pct_twowheel	0.0	1.47E-01
pct_patini	0.0	3.20E-01
pct_bike	0.0	3.46E-01

Πίνακας 5.30: Σχετική σημαντικότητα μεταβλητών στο μοντέλο παλινδρόμησης ταχύτητα – σύνθεση



Διάγραμμα 5-7: Σχετική σημαντικότητα μεταβλητών στο μοντέλο παλινδρόμησης ταχύτητα – σύνθεση

5.6.5. Ερμηνεία μοντέλου

Από τον πίνακα 5.27 εξάγονται συμπεράσματα για τη στατιστική σημαντικότητα κάθε μεταβλητής αλλά και για την ερμηνεία των συντελεστών.

- Ερμηνεία σταθεράς β_0:
  Η σταθερά β_0 = 0.4871 υποδεικνύει πως για τις κατηγορίες αναφοράς, η ταχύτητα διαμορφώνεται στο 48.7% της ταχύτητας ελεύθερης ροής (Free Flow Speed – FFS). Δηλαδή, για το οδικό τμήμα: Ακαδημίας (Πατησίων – Βασ. Σοφίας), το χρονικό διάστημα 11 – 15:00, το έτος 2022, τον μήνα Ιανουάριο, χωρίς άνεμο και βροχή, με 100% επιβατικά οχήματα, η ταχύτητα διαμορφώνεται στο 48.7% της FFS.

- Ερμηνεία συντελεστών β_i:
  Το μοντέλο που αναπτύχθηκε υποδεικνύει μείωση της ταχύτητας σε περίπτωση βροχής κατά 6.6% της FFS (συντελεστής β = -0.0661).

  Η επίδραση της σύνθεσης κυκλοφορίας φαίνεται από τους συντελεστές του μοντέλου. Οι μόνες στατιστικά σημαντικές μεταβλητές αναδείχθηκαν: το ποσοστό των ταξί και των λεωφορείων τα οποία βρέθηκαν να αυξάνουν την μέση ταχύτητα στον αστικό ιστό κατά 0.18% και 0.44% της FFS, για κάθε ποσοστιαία μονάδα αύξησης αυτών εις βάρος των ΙΧ. Για ένα οδικό τμήμα με FFS 40 χλμ/ώρα, αύξηση λεωφορείων κατά 10% έναντι των ΙΧ δύναται να επιφέρει αύξηση της μέσης ταχύτητας κατά 1.8 χλμ/ώρα.

5.7. Μοντέλο Ενίσχυσης Κλίσης (GBM) (ταχύτητα - σύνθεση)

5.7.1. Γενικά

Δεδομένου του σχετικά περιορισμένου μεγέθους δείγματος της σύνθεσης κυκλοφορίας (48 ημέρες καταγραφής), προτιμήθηκε η κλασική υλοποίηση GradientBoostingRegressor, η οποία πραγματοποιεί ακριβείς διαχωρισμούς χωρίς την προηγούμενη ομαδοποίηση των συνεχών μεταβλητών σε ιστογράμματα που εφαρμόζει η ταχύτερη παραλλαγή HistGradientBoosting. Επιπλέον, η μέθοδος αυτή επιτρέπει τον υπολογισμό μη γραμμικών διαγραμμάτων μερικής εξάρτησης (Partial Dependence Plots) για κάθε μεταβλητή σύνθεσης, με σκοπό την οπτικοποίηση της επιρροής αυτών και των πιθανών μη γραμμικών σχέσεων που δεν μπορούν να εντοπιστούν από ένα γραμμικό μοντέλο παλινδρόμησης (OLS).

5.7.2. Επεξήγηση μεταβλητών

Οι μεταβλητές που εισήχθησαν στο μοντέλο GBM είναι οι ίδιες με αυτές του μοντέλου Γραμμικής Παλινδρόμησης (ταχύτητα – σύνθεση) (Πίνακας 5.24), προσθέτοντας τη μεταβλητή της θερμοκρασίας και αφαιρώντας τη μεταβλητή year καθώς χρησιμοποιείται μόνο για τον διαχωρισμό εκπαίδευσης και ελέγχου του μοντέλου.

5.7.3. Διαχωρισμός δεδομένων εκπαίδευσης / ελέγχου

Για τον έλεγχο του μοντέλου, επιλέχθηκε ποσοστό 25% της βάση δεδομένων το οποίο αντιστοιχεί σε 552 παρατηρήσεις δηλαδή 12 ημέρες καταγραφής (ένα ολόκληρο έτος). Οι υπόλοιπες 1,656 παρατηρήσεις (36 ημέρες) χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου. Αντίστοιχα με το μοντέλο HGBM, δοκιμάσθηκαν τέσσερα διαφορετικά σύνολα εκπαίδευσης/ελέγχου (ένα ανά έτος) για να ελεγχθεί η αξιοπιστία των αποτελεσμάτων.

5.7.4. Επιλογή και ρύθμιση υπερπαραμέτρων

Το μοντέλο Gradient Boosting ρυθμίστηκε με τις εξής υπερπαραμέτρους: αριθμός εκτιμητών (δέντρων) = 300, ρυθμός μάθησης = 0.08, μέγιστο βάθος δέντρου = 3 και κατάσταση τυχαιότητας = 42.

Ο αριθμός των εκτιμητών (n_estimators = 30) καθορίζει το πλήθος των διαδοχικών δέντρων αποφάσεων που εκπαιδεύονται, με κάθε νέο δέντρο να προσπαθεί να διορθώσει το υπόλοιπο σφάλμα (residual) των προηγούμενων. Ο αριθμός αυτός επιλέχθηκε ως μια ισορροπημένη τιμή, αρκετά υψηλή ώστε το μοντέλο να μπορεί να συλλάβει σταδιακά και ασθενή σήματα (όπως η επίδραση της σύνθεσης κυκλοφορίας), χωρίς όμως να οδηγεί σε υπερβολικό υπολογιστικό κόστος ή σε κίνδυνο υπερπροσαρμογής.

Ο ρυθμός εκμάθησης (learning_rate=0.08), χαμηλότερος από την προεπιλογή του scikit-learn (0.1), καθορίζει πόσο συμβάλλει η συνεισφορά κάθε δέντρου στην τελική πρόβλεψη· μικρότερη τιμή σημαίνει πιο σταδιακή, συντηρητική

εκμάθηση, η οποία γενικά βελτιώνει τη γενίκευση του μοντέλου, με το κόστος όμως να χρειάζονται περισσότερα δέντρα για να επιτευχθεί η ίδια προσαρμογή στα δεδομένα.

Το μέγιστο βάθος κάθε δέντρου (max_depth = 3) κρατά τα επιμέρους δέντρα σκόπιμα ρηχά («αδύναμους εκτιμητές» - weak learners), ώστε κάθε ένα να συλλαμβάνει μόνο απλές, τοπικές αλληλεπιδράσεις μεταξύ μεταβλητών, αφήνοντας τη συνολική πολυπλοκότητα του μοντέλου να προκύψει αθροιστικά από τη διαδοχική διόρθωση σφαλμάτων — μια πρακτική σύμφωνη με τη θεμελιώδη λογική του boosting και αποτρεπτική υπερπροσαρμογή, ιδιαίτερα σημαντική δεδομένου του σχετικά περιορισμένου μεγέθους του δείγματος (N = 2.208 παρατηρήσεις).

Τέλος, η σταθερή τιμή random_state = 42 διασφαλίζει την αναπαραγωγιμότητα των αποτελεσμάτων, καθιστώντας τη διαδικασία εκπαίδευσης (και τα συνοδά στοχαστικά στοιχεία, όπως η τυχαία υποδειγματοληψία αν ενεργοποιηθεί) πλήρως επαναλήψιμη σε κάθε εκτέλεση.

5.7.5. Αποτελέσματα και αξιολόγηση μοντέλου

Η εκπαίδευση τεσσάρων ξεχωριστών μοντέλων, με κάθε ένα να κρατά διαδοχικά ένα ολόκληρο ημερολογιακό έτος (2022, 2023, 2024, 2025) ως ανεξάρτητο σύνολο ελέγχου, κατέδειξε συνολικά ικανοποιητική προγνωστική ικανότητα με κάποια διακύμανση ανάλογα με το έτος: ο συντελεστής προσδιορισμού κυμάνθηκε από R^2=0.317 (έτος ελέγχου 2025) έως R^2=0.720 (έτος ελέγχου 2023), με το 2025 να παρουσιάζει αισθητά χαμηλότερη επίδοση και στα τρία μέτρα σφάλματος (MAE=0.105, RMSE=0.134) σε σχέση με τα υπόλοιπα έτη.

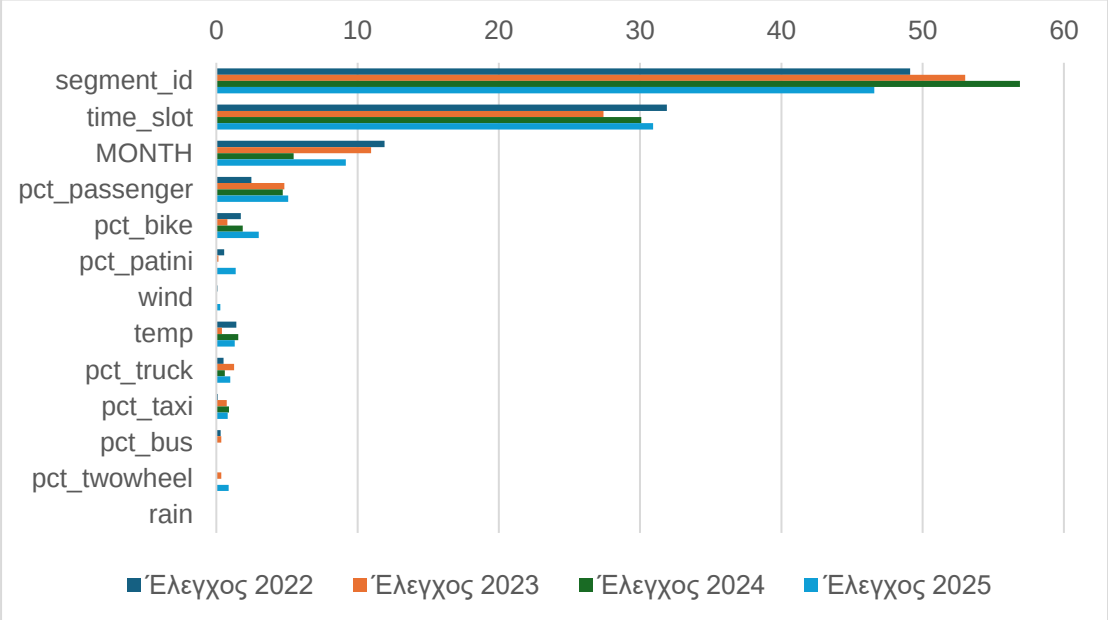
Έλεγχος 2022		Έλεγχος 2023		Έλεγχος 2024		Έλεγχος 2025	
Train Set	1656	Train Set	1656	Train Set	1656	Train Set	1656
Test Set	552	Test Set	552	Test Set	552	Test Set	552
MAE:	0.080	MAE:	0.068	MAE:	0.076	MAE:	0.105
RMSE:	0.102	RMSE:	0.086	RMSE:	0.099	RMSE:	0.134
R^2:	0.599	R^2:	0.720	R^2:	0.645	R^2:	0.317
segment_id	49	segment_id	53	segment_id	57	segment_id	47
time_slot	32	time_slot	27	time_slot	30	time_slot	31
MONTH	12	MONTH	11	MONTH	5	MONTH	9
pct_passenger	2	pct_passenger	5	pct_passenger	5	pct_passenger	5
pct_bike	2	pct_bike	1	pct_bike	2	pct_bike	3
pct_patini	1	pct_patini	0	pct_patini	0	pct_patini	1
wind	0	wind	0	wind	-1	wind	0
temp	1	temp	0	temp	2	temp	1
pct_truck	1	pct_truck	1	pct_truck	1	pct_truck	1
pct_taxi	0	pct_taxi	1	pct_taxi	1	pct_taxi	1
pct_bus	0	pct_bus	0	pct_bus	-1	pct_bus	0
pct_twowheel	0	pct_twowheel	0	pct_twowheel	0	pct_twowheel	1

rain 0	rain 0	rain 0	rain 0

Πίνακας 5.31: Αποτελέσματα μοντέλων για διαφορετικά δεδομένα ελέγχου – μοντέλο GBM (ταχύτητα – σύνθεση)

Παρά τη διακύμανση στην ακρίβεια πρόβλεψης, η ιεράρχηση σημαντικότητας των παραγόντων παραμένει σταθερή σε όλα τα έτη ελέγχου: το οδικό τμήμα (segment_id) κυριαρχεί με συνέπεια (47 - 53%), ακολουθούμενο σταθερά από το time_slot (27 - 32%) και τον μήνα (5 - 12%), ενώ από τις μεταβλητές σύνθεσης κυκλοφορίας το ποσοστό IX (pct_passenger) αναδεικνύεται με συνέπεια ως η σημαντικότερη σε κάθε ένα από τα τέσσερα μοντέλα. Αντίθετα, μεταβλητές όπως το pct_bus, το pct_truck και το rain εμφανίζουν με συνέπεια αμελητέα συνεισφορά σε όλα σχεδόν τα έτη.

Συνολικά, η ανάλυση ανά έτος δείχνει τη στιβαρότητα του βασικού ευρήματος: την κυρίαρχη επίδραση δομικών/χρονικών παραγόντων έναντι της σύνθεσης κυκλοφορίας, με το IX να ξεχωρίζει σταθερά ως η πλέον επιδραστική μεταβλητή σύνθεσης.



Διάγραμμα 5-8: Σχετική σημαντικότητα μεταβλητών στο μοντέλο GBM ταχύτητα – σύνθεση

5.7.6. Απομόνωση μεταβλητής IX και διαγράμματα μερικής εξάρτησης

Για την οπτικοποίηση της επίδρασης της σύνθεσης οχημάτων στη ταχύτητα κυκλοφορίας, χρησιμοποιήθηκαν Γραφήματα Μερικής Εξάρτησης (Partial Dependance Plots) από το μοντέλο για όλα τα έτη ελέγχου. Ο τρόπος εξαγωγής των διαγραμμάτων περιγράφεται ως εξής:

1. Δημιουργία πλέγματος του άξονα X:
   Για την υπό μελέτη μεταβλητή, ο αλγόριθμος παίρνει τις πραγματικές τιμές αυτής της στήλης στο σύνολο εκπαίδευσης και φτιάχνει ένα πλέγμα τιμών – ανάμεσα στο 5ο και 95ο εκατοστημόριο των παρατηρούμενων τιμών, χωρισμένο σε πενήντα (50) ισαπέχοντα σημεία.

2. Για κάθε σημείο του πλέγματος, η μεταβλητή μένει σταθερή σε όλη τη βάση δεδομένων:
   Για κάθε παρατήρηση στο σύνολο εκπαίδευσης, αντικαθιστά τη στήλη της μεταβλητής σε όλες τις γραμμές με την πρώτη τιμή του πλέγματος ακόμη και αν η πραγματική τιμή είναι διαφορετική. Οι υπόλοιπες μεταβλητές σύνθεσης αναπροσαρμόζονται αναλογικά έτσι ώστε να υπάρχει άθροισμα 100%. Τρέχει το μοντέλο με την τροποποιημένη βάση δεδομένων, παίρνοντας μία πρόβλεψη ανά γραμμή και τέλος υπολογίζει τον μέσο όρο όλων αυτών των προβλέψεων.

3. Επανάληψη για κάθε σημείο του πλέγματος
   Για κάθε πενήντα τιμές της μεταβλητής δημιουργεί ολόκληρο το τροποποιημένο dataset (1.656 γραμμές) και ξανατρέχει το μοντέλο. Συνολικά απαιτούνται 50 × 1.656 = 82,800 προβλέψεις για μία καμπύλη.

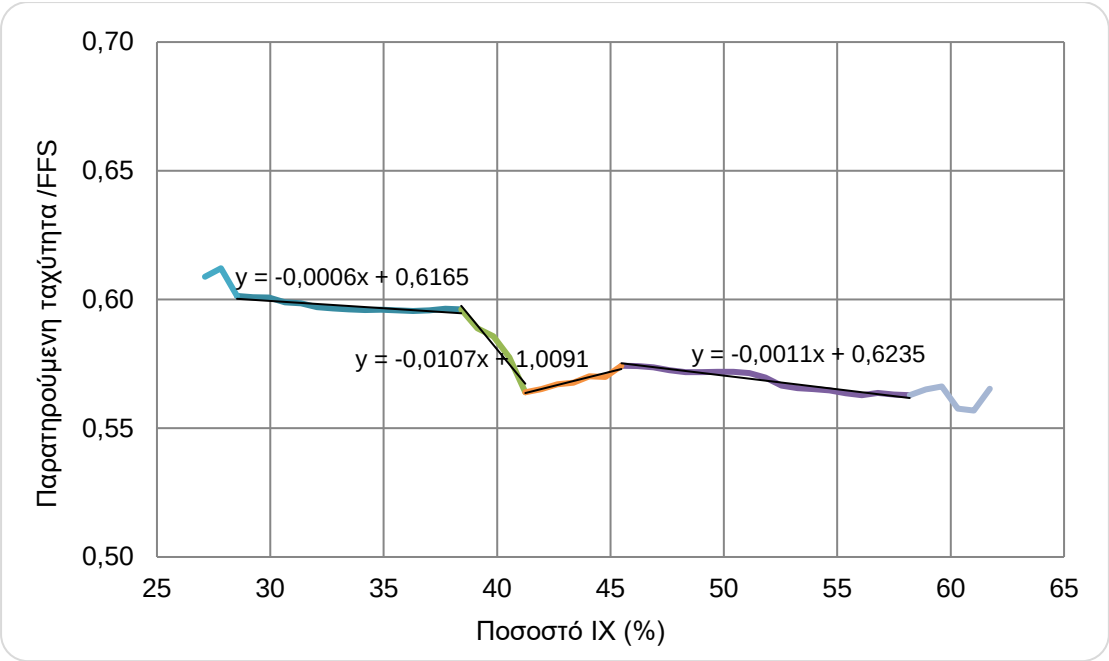
4. Τα 50 σημεία ενώνονται με ευθείες γραμμές.

Για τα τέσσερα σύνολα εκπαίδευσης / ελέγχου, υπολογίζεται η μέση τιμή των σημείων των διαγραμμάτων και σχηματίζεται μία εννιαία καμπύλη.

Για τη μεταβλητή «ποσοστό IX», η οποία αποδείχθηκε η σημαντικότερη εκ των μεταβλητών σύνθεσης, το διάγραμμα ανέδειξε ρυθμό μείωσης της ταχύτητας ίσο με:

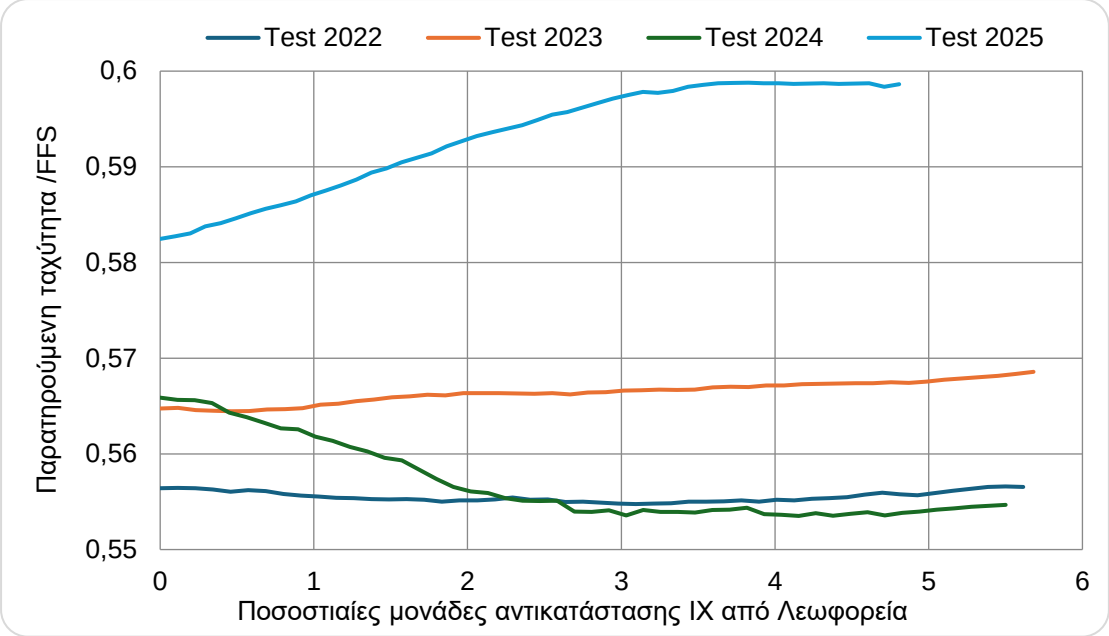
- 0.06% της FFS για 28 – 38%
- 1.07% της FFS για 38 – 41%
- 0.11% της FFS για 45 – 58%

Η μείωση της ταχύτητας είναι ήπια στο μεγαλύτερο μέρος του εύρους του ποσοστού IX και εμφανίζει πιο απότομη πτώση σε στενή ζώνη τιμών (περίπου 38 έως 42%). Δεδομένου ότι η σύνθεση της κυκλοφορίας συσχετίζεται με την οδό και την ώρα καταγραφής και ότι το δείγμα περιορίζεται σε 48 ημέρες, η πτώση αυτή δεν πρέπει να ερμηνευθεί ως όριο ή ως γενικός νόμος.



Διάγραμμα 5-9: Επίδραση ποσοστού IX στην ταχύτητα κυκλοφορίας (μοντέλο GBM)

Για τη διερεύνηση αντικατάστασης IX με λεωφορεία, ο αλγόριθμος που εφαρμόστηκε πραγματοποιεί μια άμεση, ισοδύναμη ανταλλαγή μεταξύ των συγκεκριμένων κατηγοριών οχημάτων: για κάθε τιμή Δ ενός πλέγματος τιμών, αφαιρούνται Δ ποσοστιαίες μονάδες από το ποσοστό IX κάθε παρατήρησης και προστίθενται ακριβώς οι ίδιες μονάδες στο ποσοστό λεωφορείων, ενώ όλες οι υπόλοιπες κατηγορίες (φορτηγά, ταξί, δίκυκλα, ποδήλατα, πατίνια) παραμένουν αμετάβλητες. Με τον τρόπο αυτό το άθροισμα 100% διατηρείται αυτόματα και επακριβώς σε κάθε τροποποιημένη παρατήρηση, χωρίς να απαιτείται πρόσθετη αναλογική αναπροσαρμογή. Το μέγιστο εύρος του πλέγματος τιμών Δ καθορίστηκε με διπλό περιορισμό, ώστε να αποφευχθεί τόσο η εξάντληση του διαθέσιμου ποσοστού IX ανά παρατήρηση όσο και η εξωτερίκευση πέραν του εύρους τιμών που παρατηρήθηκε πραγματικά για το ποσοστό λεωφορείων στο σύνολο εκπαίδευσης, περιορισμός ιδιαίτερα κρίσιμος για μοντέλα βασισμένα σε δέντρα αποφάσεων, τα οποία δεν διαθέτουν ικανότητα αξιόπιστης πρόβλεψης πέραν των ορίων των δεδομένων εκπαίδευσής τους.



Διάγραμμα 5-10: Επίδραση αντικατάστασης IX από λεωφορεία στην ταχύτητα κυκλοφορίας (GBM)

Το Διάγραμμα 5-10, δεν δείχνει σαφή μεταβολή της ταχύτητας κυκλοφορίας όταν αντικαθίσταται ορισμένος αριθμός IX από λεωφορεία. Ανάλογα με το σύνολο δεδομένων ελέγχου του μοντέλου GBM, το διάγραμμα δίνει είτε ελάχιστη μείωση (έτος 2024), είτε σταθερότητα (έτος 2022), είτε μικρή άνοδο της ταχύτητας κυκλοφορίας (έτος 2025 και 2023). Το εύρος τιμών που δύναται να γίνει η αντικατάσταση δεν ξεπέρασε το 6% καθώς στη βάση δεδομένων σύνθεσης κυκλοφορίας το παρατηρούμενο ποσοστό λεωφορείων ανά καταγραφή, δεν ξεπέρασε αυτή την τιμή.

Το εύρημα αυτό δεν δείχνει επιδείνωση της ταχύτητας κυκλοφορίας όταν IX αντικαθίστανται από λεωφορεία. Η μεταφορική ικανότητα των λεωφορείων είναι σημαντικά μεγαλύτερη από των IX: μία λωρίδα με IX μετακινεί περίπου 600 έως 1.600 άτομα ανά ώρα, ενώ μία αποκλειστική λωρίδα λεωφορείου έως 8.000 (NACTO, 2016). Συνεπώς, η αντικατάσταση IX από λεωφορεία μπορεί να εξυπηρετήσει περισσότερα άτομα χωρίς να μειωθεί η ταχύτητα.

5.8. Μοντέλο Τυχαίων Δασών (RF) (ταχύτητα - σύνθεση)

5.8.1. Γενικά

Εκτός από το μοντέλο GBM, χρησιμοποιήθηκε και το Random Forest, ως δεύτερη, ανεξάρτητη οικογένεια μη γραμμικών μοντέλων συνόλου. Ο λόγος ήταν να ελεγχθεί αν τα συμπεράσματα για τη σχετική σημαντικότητα των μεταβλητών εξαρτώνται από τον αλγόριθμο ή αν παραμένουν σταθερά. Τα δύο μοντέλα διαφέρουν στη λογική κατασκευής τους. Το GBM χτίζει τα δέντρα διαδοχικά, με κάθε νέο δέντρο να διορθώνει τα σφάλματα του προηγούμενου, και γι' αυτό είναι πιο ευαίσθητο στον ρυθμό μάθησης και στον αριθμό των επαναλήψεων, με μεγαλύτερο κίνδυνο υπερπροσαρμογής όταν το δείγμα είναι μικρό ή θορυβώδες. Το Random Forest σε αντίθεση, εκπαιδεύει ανεξάρτητα δέντρα σε τυχαία υποσύνολα δεδομένων και μεταβλητών και μέσω του μέσου όρου των προβλέψεων μειώνει τη διακύμανση. Αυτό το κάνει πιο ανθεκτικό και

λιγότερο εξαρτημένο από την επιλογή των υπερπαραμέτρων, ιδιότητα σημαντική στο δείγμα της σύνθεσης κυκλοφορίας, που αποτελείται από 48 ημέρες χειροκίνητων μετρήσεων.

5.8.2. Επεξήγηση μεταβλητών

Για τη σύγκριση του μοντέλου RF με τα υπόλοιπα, οι μεταβλητές που εισήχθησαν είναι οι ίδιες με αυτές του μοντέλου Γραμμικής Παλινδρόμησης και του GBM (ταχύτητα – σύνθεση) (Πίνακας 5.24), προσθέτοντας τη μεταβλητή της θερμοκρασίας και αφαιρώντας τη μεταβλητή year καθώς χρησιμοποιείται μόνο για τον διαχωρισμό εκπαίδευσης και ελέγχου του μοντέλου.

5.8.3. Διαχωρισμός δεδομένων εκπαίδευσης / ελέγχου

Τα δεδομένα διαχωρίστηκαν με τον ίδιο τρόπο όπως στο μοντέλο GBM. Επιλέχθηκε ποσοστό 25% της βάση δεδομένων για τον έλεγχο το οποίο αντιστοιχεί σε 552 παρατηρήσεις δηλαδή 12 ημέρες καταγραφής (ένα ολόκληρο έτος) και οι υπόλοιπες 1,656 παρατηρήσεις (36 ημέρες) χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου. Αντίστοιχα με το μοντέλο GBM της προηγούμενος ενότητας, δοκιμάσθηκαν τέσσερα διαφορετικά σύνολα εκπαίδευσης/ελέγχου (ένα ανά έτος) για να ελεγχθεί η αξιοπιστία των αποτελεσμάτων.

5.8.4. Επιλογή και ρύθμιση υπερπαραμέτρων

Οι υπερπαράμετροι του Random Forest ορίστηκαν εκ των προτέρων με στόχο τον έλεγχο της πολυπλοκότητας των δέντρων, και όχι μέσω εκτεταμένης βελτιστοποίησης. Χρησιμοποιήθηκαν 100 δέντρα για το μεγάλο σύνολο ταχυτήτων και 300 για το μικρότερο σύνολο σύνθεσης κυκλοφορίας, επειδή το Random Forest δεν υπερπροσαρμόζεται με την αύξηση του αριθμού των δέντρων και ο αριθμός αυτός επηρεάζει κυρίως τη σταθερότητα των προβλέψεων. Ο ελάχιστος αριθμός δειγμάτων ανά φύλλο ορίστηκε σε 5 και 3 αντίστοιχα, ώστε τα δέντρα να μην προσαρμόζονται σε μεμονωμένες παρατηρήσεις, και το ποσοστό μεταβλητών που εξετάζεται σε κάθε διαχωρισμό (max_features) σε 0.5, ώστε να μειώνεται η συσχέτιση μεταξύ των δέντρων και να μη κυριαρχεί συστηματικά μία ισχυρή μεταβλητή όπως το οδικό τμήμα. Για την αναπαραγωγιμότητα των αποτελεσμάτων ορίστηκε σταθερή τιμή τυχαιότητας (random_state) ίση με σαράντα δύο (42).

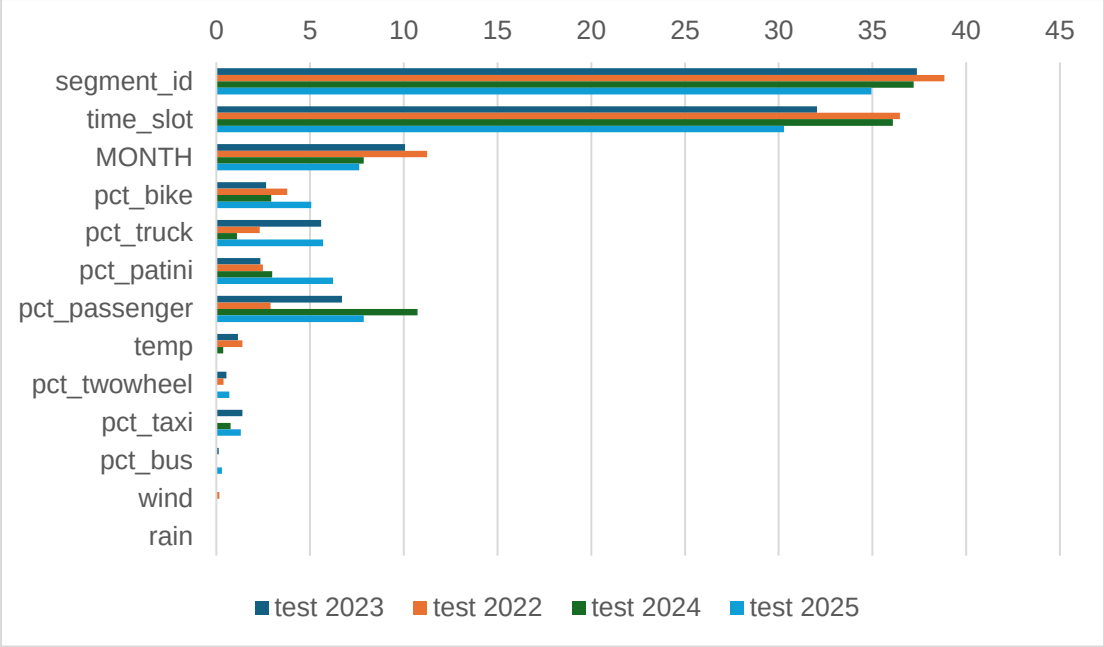
5.8.5. Αποτελέσματα και αξιολόγηση μοντέλου

Το μοντέλο Random Forest που συνδέει τη σύνθεση της κυκλοφορίας με την ταχύτητα επιτυγχάνει μέτρια έως καλή προβλεπτική ικανότητα στα έτη ελέγχου. Η απόδοση εμφανίζει σημαντική διακύμανση ανάμεσα στα έτη: τα καλύτερα αποτελέσματα προκύπτουν για το 2023 και το 2024), ενώ το 2025 είναι το πιο δύσκολο έτος να προβλεφθεί.

Έλεγχος 2022		Έλεγχος 2023		Έλεγχος 2024		Έλεγχος 2025	
Train Set	1656	Train Set	1656	Train Set	1656	Train Set	1656
Test Set	552	Test Set	552	Test Set	552	Test Set	552
MAE:	0.086	MAE:	0.067	MAE:	0.071	MAE:	0.103
RMSE:	0.108	RMSE:	0.086	RMSE:	0.093	RMSE:	0.133
R^2:	0.552	R^2:	0.715	R^2:	0.688	R^2:	0.324
segment_id	39	segment_id	37	segment_id	37	segment_id	35
time_slot	36	time_slot	32	time_slot	36	time_slot	30
MONTH	11	MONTH	10	MONTH	8	MONTH	8
pct_bike	3.8	pct_truck	5.6	pct_passenger	10.7	pct_passenger	7.9
pct_truck	2.3	pct_passenger	6.7	pct_bike	2.9	pct_bike	5.1
pct_patini	2.5	pct_bike	2.7	pct_patini	3.0	pct_patini	6.2
pct_passenger	2.9	pct_patini	2.3	pct_truck	1.1	pct_truck	5.7
temp	1.4	pct_taxi	1.4	pct_taxi	0.8	pct_taxi	1.3
pct_twowheel	0.4	pct_twowheel	0.5	temp	0.4	pct_twowheel	0.7
pct_taxi	0.0	temp	1.2	rain	0.0	pct_bus	0.3
pct_bus	0.0	pct_bus	0.1	pct_twowheel	0.0	temp	0.0
wind	0.2	rain	0.0	pct_bus	0.0	rain	0.0

Πίνακας 5.32: Αποτελέσματα μοντέλων για διαφορετικά δεδομένα ελέγχου – μοντέλο RF (ταχύτητα – σύνθεση)

Σε όλα τα έτη η σχετική σημαντικότητα κυριαρχείται από τρεις μεταβλητές. Το οδικό τμήμα (segment_id) είναι η σημαντικότερη (35 – 39%), ακολουθεί το χρονικό παράθυρο μέτρησης (time_slot, 30 – 36%) και τρίτος έρχεται ο μήνας (8 – 11%). Οι μεταβλητές σύνθεσης της κυκλοφορίας έχουν επιμέρους συμβολή που κυμαίνεται κυρίως από 1% έως 7%, με τα ποδήλατα, τα φορτηγά, τα πατίνια και τα ΙΧ επιβατικά να καταλαμβάνουν συνήθως τις υψηλότερες θέσεις, ενώ τα λεωφορεία, τα δίκυκλα και τα ταξί να συμβάλλουν ελάχιστα. Αθροιστικά, η σύνθεση της κυκλοφορίας εξηγεί περίπου 13 – 19% της σημαντικότητας στα έτη 2022–2024 και αυξάνεται στο 27% το 2025. Οι μετεωρολογικές μεταβλητές (άνεμος, βροχή) έχουν μηδενική συμβολή και η θερμοκρασία αμελητέα. Συνολικά, η ταχύτητα καθορίζεται κυρίως από τα χωρικά και χρονικά χαρακτηριστικά, ενώ η σύνθεση των οχημάτων έχει δευτερεύοντα ρόλο.



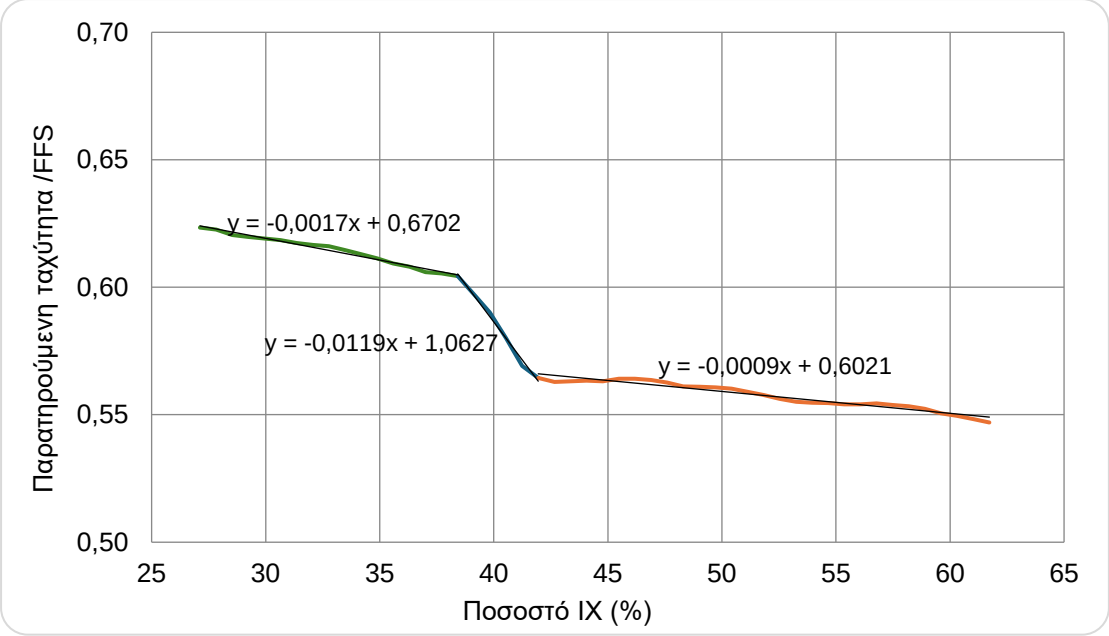
Διάγραμμα 5-11: Σχετική σημαντικότητα μεταβλητών στο μοντέλο RF ταχύτητα – σύνθεση

5.8.6. Απομόνωση μεταβλητής IX και διαγράμματα μερικής εξάρτησης

Για την οπτικοποίηση της επίδρασης των IX στην κυκλοφορία και στο μοντέλο RF, σχεδιάστηκε διάγραμμα μερικής εξάρτησης με τον ίδιο τρόπο που περιγράφεται στην ενότητα **Error! Reference source not found.**. Σε αυτή την περίπτωση, για κάθε ποσοστιαία μονάδα αύξησης του ποσοστού IX ο ρυθμός μείωσης της ταχύτητας προκύπτει:

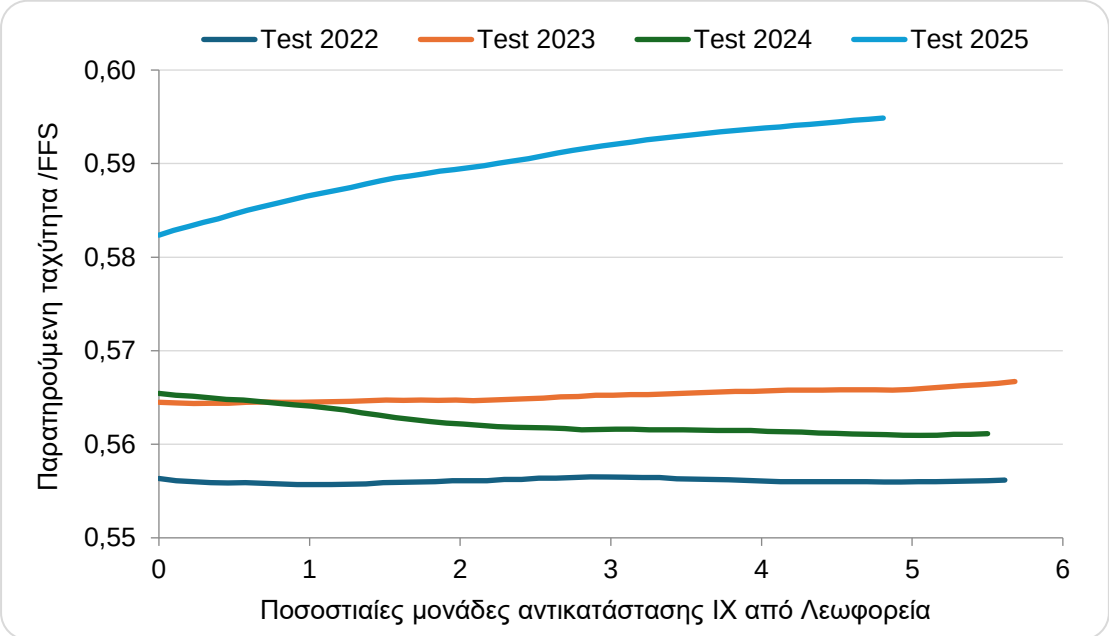
- 0.17% της FFS στις τιμές 27 – 38%
- 1.20% της FFS στις τιμές 38 – 42%
- 0.09% της FFS στις τιμές 42 – 62%

Ισχύει η ίδια επιφύλαξη με το GBM (ενότητα 5.7.6). Η απότομη πτώση στο 38 έως 42% δεν πρέπει να γενικευτεί.



Διάγραμμα 5-12: Επίδραση ποσοστού IX στην ταχύτητα κυκλοφορίας (μοντέλο RF)

Στο ερώτημα αντικατάστασης των IX αποκλειστικά από λεωφορεία, το μοντέλο RF δίνει παρόμοια αποτελέσματα με το GBM όπου δεν υπάρχει σαφής μεταβολή της ταχύτητας κυκλοφορίας.



Διάγραμμα 5-13: Επίδραση αντικατάστασης IX από λεωφορεία στην ταχύτητα κυκλοφορίας (RF)

5.9. Σύγκριση μοντέλων GBM, RF και Γραμμικής Παλινδρόμησης (ταχύτητα - σύνθεση κυκλοφορίας)

Τα μοντέλα μηχανικής μάθησης υπερέχουν της γραμμικής παλινδρόμησης σε όλους τους δείκτες. Τα δύο μοντέλα μηχανικής μάθησης έχουν σχεδόν ισοδύναμη απόδοση, με διαφορά που δεν επιτρέπει ασφαλές συμπέρασμα για την υπεροχή του ενός. Η μικρή βελτίωση σε σχέση με τη γραμμική παλινδρόμηση υποδηλώνει ότι οι μη γραμμικές σχέσεις και οι αλληλεπιδράσεις συμβάλλουν μετρίως στην πρόβλεψη.

Έτος ελέγχου	Παλινδρόμηση			GBM			RF		
	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE
2022	0.410	0.098	0.124	0.599	0.080	0.102	0.552	0.086	0.108
2023	0.528	0.087	0.111	0.720	0.068	0.086	0.715	0.067	0.086
2024	0.469	0.096	0.121	0.645	0.076	0.099	0.688	0.071	0.093
2025	0.353	0.103	0.130	0.317	0.105	0.134	0.324	0.103	0.133

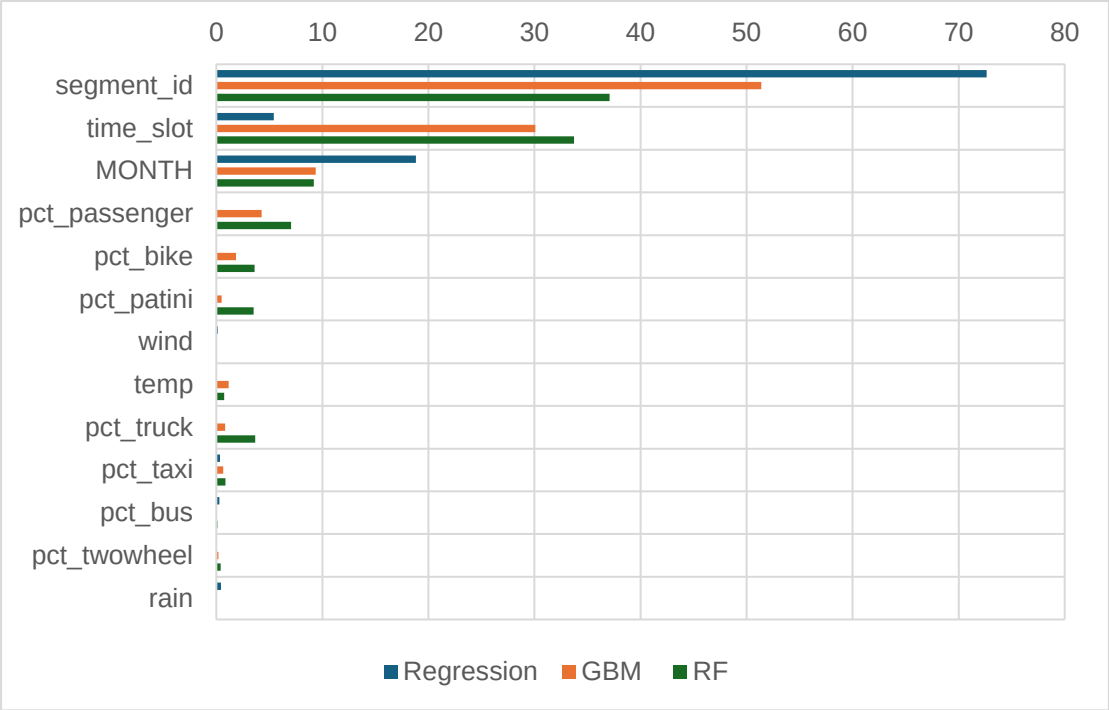
Πίνακας 5.33: Σύγκριση δεικτών απόδοσης και σφάλματος στα μοντέλα ταχύτητας -σύνθεσης

Η σχετική σημαντικότητα των μεταβλητών στα μοντέλα μηχανικής μάθησης υπολογίζεται ώς ο μέσος όρος αυτών και στα τέσσερα σύνολα ελέγχου. Και στα τρία μοντέλα κυρίαρχες μεταβλητές στην προγνωστική ικανότητα είναι το οδικό τμήμα, το χρονικό παράθυρο και ο μήνας. Το οδικό τμήμα είναι η σημαντικότερη μεταβλητή, ακολουθούμενο από τον μήνα και το χρονικό παράθυρο. Οι διαφορές εντοπίζονται στη βαρύτητα των μεταβλητών σύνθεσης οχημάτων. Στην παλινδρόμηση τα ποσοστά οχημάτων έχουν πρακτικά μηδενική σημασία (αθροιστικά περίπου 0.7%), στο GBM συμβάλλουν αθροιστικά περίπου 8%, ενώ στο RF περίπου 20%. Το μοντέλο RF κατανέμει τη σημαντικότητα πιο ισόρροπα μεταξύ των μεταβλητών, ενώ τα GBM και η παλινδρόμηση συγκεντρώνουν σχεδόν όλη τη σημασία στις χωρικές και χρονικές μεταβλητές. Οι καιρικές μεταβλητές έχουν αμελητέα συμβολή και στα τρία μοντέλα το οποίο πιθανώς οφείλεται στις καλές καιρικές συνθήκες τις ημέρες των καταγραφών σύνθεσης οχημάτων (23 από τις 2,028 παρατηρήσεις με βροχή). Συνολικά, το συμπέρασμα ότι η ταχύτητα καθορίζεται κυρίως από τα χωρικά και χρονικά χαρακτηριστικά, με δευτερεύοντα ρόλο της σύνθεσης των οχημάτων, είναι σταθερό ανεξάρτητα από το μοντέλο.

Μεταβλητή	Παλινδρόμηση	GBM	RF
segment_id	73	51	37
time_slot	5.4	30	34
MONTH	19	9.4	9
pct_passenger		4.3	7
pct_bike	0.0	1.8	4
pct_patini	0.0	0.5	4
wind	0.1	-0.2	0
temp		1.2	1
pct_truck	0.0	0.8	4
pct_taxi	0.4	0.6	1
pct_bus	0.3	-0.1	0

pct_twowheel	0.0	0.2	0
rain	0.5	0.0	0
C(YEAR)	1.8	-	-

Πίνακας 5.34: Σύγκριση σημαντικότητας μεταβλητών μεταξύ των μοντέλων ταχύτητας – σύνθεσης



Διάγραμμα 5-14: Σύγκριση σημαντικότητας μεταβλητών μεταξύ των μοντέλων ταχύτητας – σύνθεσης

6. Συμπεράσματα

6.1. Σύνοψη αποτελεσμάτων

Τα αποτελέσματα των αναλύσεων απαντούν στα ερευνητικά ερωτήματα που τέθηκαν στο ξεκίνημα της παρούσας διπλωματικής εργασίας. Όπως αναφέρθηκε και στο Κεφάλαιο 1, στόχος ήταν η **διερεύνηση επιρροής ορισμένων παραγόντων στην συμφόρηση στο κέντρο της Αθήνας** οι οποίοι αποτελούνται από κοινωνικοοικονομικούς, μετεωρολογικούς και κυκλοφοριακούς παράγοντες και συγκεκριμένα την σύνθεση της κυκλοφορίας.

Η αρχική προκαταρκτική ανάλυση των δεδομένων ανέδειξε σημαντικά ευρήματα για την κυκλοφοριακή συμφόρηση στην Αθήνα. Από το 2022 έως το 2025, οι συνολικές καθυστερήσεις αυξήθηκαν κατά 72% σε έξι περιφερειακούς οδικούς άξονες και κατά 40% σε 12 κεντρικούς άξονες της Αθήνας το οποίο αναδεικνύει την **αύξηση της έντασης της συμφόρησης**. Εκτός από την ένταση της συμφόρησης, παρατηρήθηκε και διεύρυνση των ωρών αιχμής μέσα στην ημέρα δηλαδή **αύξηση της διάρκειας συμφόρησης**. Από τις εννέα ώρες ανάλυσης ανά εργάσιμη ημέρα (8 – 10:00 & 12 – 19:00), το 2025, το 31% αυτών στις περιφερειακές οδούς παρουσιάζουν ταχύτητες κάτω του 50% της ταχύτητας ελεύθερης ροής και το αντίστοιχο ποσοστό των οδών εντός κέντρου αγγίζει το 52%. Και στις δύο περιπτώσεις υπάρχει άνοδος από το 2022 κατά 89% και 42% αντίστοιχα (Κεφάλαιο 4.5.1).

Ταυτόχρονα με τη διαχρονική επιδείνωση της συμφόρησης, ο **αριθμός των επιβατικών ΙΧ** οχημάτων που κυκλοφορούνε στην Αττική **έχει αυξηθεί κατά 396.5 χιλιάδες** από το 2021 έως το 2025 (ΕΛΣΤΑΤ) το οποίο έχει ως αποτέλεσμα την επιβάρυνση του οδικού δικτύου. Παράλληλα, μέσω των δεδομένων σύνθεσης της κυκλοφορίας σε δέκα έξι οδούς στο κέντρο της Αθήνας, διαπιστώθηκε η διαχρονική υπεροχή αυτών στην κυκλοφορία με **ποσοστό 50% επί του συνόλου**. Το ποσοστό των λεωφορείων παραμένει στο πολύ χαμηλό επίπεδο του 2 – 5% (Κεφάλαιο 4.4.2).

Οι κοινωνικοοικονομικοί παράγοντες **δεν είχαν σαφή επίδραση στη διακύμανση της συμφόρησης** καθώς αυτή, όπως διαπιστώθηκε, καθορίζεται κατά 86.7% από μία σταθερή ανοδική τάση από το έτος 2022 και μία εποχικότητα, με το 3^{ο} τρίμηνο του έτους να εμφανίζει διαχρονικά τις μικρότερες καθυστερήσεις. Ύστερα από αφαίρεση της τάσης και της εποχικότητας των επεξηγηματικών μεταβλητών, η μόνη μεταβλητή με σημαντική συσχέτιση και στους δύο δείκτες Pearson και Spearman είναι ο αριθμός των ΙΧ οχημάτων ενώ οριακή ένδειξη (σημαντική μόνο κατά Spearman) υπάρχει για τους εργαζόμενους και το ποσοστό εργατικού δυναμικού. Τα μοντέλα παλινδρόμησης επιχείρησαν να εξετάσουν την πιθανή βελτίωση του βασικού μοντέλου (τάση + τρίμηνο) με την διαδοχική προσθήκη των σημαντικών μεταβλητών. Από αυτές, μόνο ο συντελεστής του «αριθμού ΙΧ» παρέμεινε θετικός και στατιστικά σημαντικός το οποίο όμως δεν πρέπει να λαμβάνεται ως αιτιότητα αύξησης της συμφόρησης αλλά ως συσχέτιση (Κεφάλαιο 5.2).

Η επιρροή των μεταβλητών σύνθεσης της κυκλοφορίας **είναι συνολικά μικρή και διαφέρει ανάμεσα στα μοντέλα**. Το μοντέλο γραμμικής παλινδρόμησης

ανέδειξε ως σημαντικές μόνο αυτές του ποσοστού ταξί και λεωφορείων στο σύνολο της κυκλοφορίας με αύξηση της μέσης ταχύτητας κατά 0.18% και 0.44% της FFS αντίστοιχα, για κάθε ποσοστιαία μονάδα αύξησης αυτών εις βάρος των IX (Κεφάλαιο 5.6). Τα διαγράμματα μερικής εξάρτησης των μοντέλων GBM και RF έδειξαν μείωση της μέσης ταχύτητας, με ρυθμό από 0.06% έως 0.17% της FFS για κάθε ποσοστιαία μονάδα αύξησης των IX στο μεγαλύτερο εύρος τιμών, και έως 1.20% σε μία στενή ζώνη γύρω στο 38 έως 42%, η οποία δεν πρέπει να γενικευτεί. Για τον σωστό υπολογισμό έγινε η υπόθεση αναλογικής μείωσης των υπόλοιπων μεταβλητών σύνθεσης έτσι ώστε το άθροισμα να είναι πάντα 100%.

Στο ερώτημα **ισάριθμης αντικατάστασης IX οχημάτων από λεωφορεία**, τα μοντέλα μηχανικής μάθησης έδειξαν πως η μεταβολή στην ταχύτητα διαφέρει ανά σύνολο ελέγχου των δεδομένων αλλά είναι πρακτικά αμελητέα (Κεφάλαια, 5.7 & 5.8). Τα τρία μοντέλα κατανέμουν διαφορετικά τη σημαντικότητα στις μεταβλητές. Στην παλινδρόμηση τα ποσοστά οχημάτων έχουν πρακτικά μηδενική σημασία (αθροιστικά περίπου 0.7%), στο GBM συμβάλλουν αθροιστικά περίπου 8%, ενώ στο RF περίπου 20%. Το μοντέλο RF κατανέμει τη σημαντικότητα πιο ισόρροπα μεταξύ των μεταβλητών, ενώ τα GBM και η παλινδρόμηση συγκεντρώνουν σχεδόν όλη τη σημασία στις χωρικές και χρονικές μεταβλητές.

Η επίδραση των καιρικών φαινομένων στην ταχύτητα κυκλοφορίας των οχημάτων στο κέντρο της Αθήνας προέκυψε **μικρή αλλά στατιστικά σημαντική** στο μοντέλο παλινδρόμησης. Οι συντελεστές των μεταβλητών ανέδειξαν **μείωση των ταχυτήτων από 1.3% έως 3.3% της ταχύτητας ελεύθερης ροής** σε συνθήκες μικρής (<2.5 mm/h) και μεγάλης (>7.6 mm/h) έντασης βροχής αντίστοιχα. Η μικρή αυτή μείωση μπορεί να οφείλεται στις ήδη χαμηλές ταχύτητες του αστικού οδικού δικτύου που όπως διαπιστώθηκε στην προκαταρκτική ανάλυση, διατηρούνται σε πολύ χαμηλό επίπεδο. Αντίστοιχα, στο μοντέλο Ενίσχυσης Κλίσης βασισμένο σε Ιστόγραμμα (HGB), η επίδραση των μετεωρολογικών παραγόντων προέκυψε αμελητέα (Κεφάλαια 5.3 & 5.4).

6.2. Περιορισμοί

Τα ευρήματα της παρούσας έρευνας υπόκεινται σε ορισμένους περιορισμούς που πρέπει να λαμβάνονται υπόψη κατά την ερμηνεία αυτών. Αρχικά, **η ταχύτητα ελεύθερης ροής ορίσθηκε από το ίδιο το σύνολο δεδομένων** το οποίο δεν είναι απόλυτα ακριβές. Ως προς τα κοινωνικοοικονομικά δεδομένα, η ανάλυση βασίστηκε σε μόλις 16 παρατηρήσεις (μία ανά τρίμηνο για τέσσερα έτη), ενώ ορισμένες τριμηνιαίες τιμές προέκυψαν με γραμμική παρεμβολή των ετήσιων τιμών, γεγονός που περιορίζει τη στατιστική ισχύ. Στο μοντέλο ταχύτητας - καιρού, η χρήση ενός μόνο σταθμού καταγραφής για το σύνολο των οδικών τμημάτων δεν επιτρέπει να ληφθούν υπόψη οι τοπικές διαφοροποιήσεις των καιρικών συνθηκών στο εξεταζόμενο δίκτυο.

Τα μοντέλα σύνθεσης κυκλοφορίας βασίστηκαν σε **μικρό δείγμα μόλις 48 ημερών** και σε περιορισμένο αριθμό οδικών τμημάτων, γεγονός που μειώνει τη δυνατότητα γενίκευσης των αποτελεσμάτων. Επιπλέον, η έλλειψη δεδομένων κυκλοφοριακού φόρτου δεν προσφέρει τη δυνατότητα διερεύνησης

επιρροής αυτού σε συνδυασμό με τη σύνθεση. Τα διαγράμματα PDP που παρουσιάστηκαν υπέθεταν άθροισμα σύνθεσης 100%, δηλαδή σταθερό αριθμό οχημάτων με αποτέλεσμα να μην εξετάζεται η πιθανή μείωση του όγκου της κυκλοφορίας αποκλειστικά από συγκεκριμένες κατηγορίες. Συνολικά, μελλοντικές έρευνες με εκτενέστερα και πιο λεπτομερή δεδομένα θα μπορούσαν να επιβεβαιώσουν και να εμπλουτίσουν τα ευρήματα της συγκεκριμένης έρευνας.

6.3. Προτάσεις

Προτείνονται λοιπόν οι ακόλουθες κατευθύνσεις μελλοντικής έρευνας, οι οποίες αφορούν το σύνολο των αναλύσεων.

Για τα κοινωνικοοικονομικά δεδομένα, προτείνεται η συλλογή περισσότερων δεδομένων με **μεγαλύτερη συχνότητα καταγραφής** (π.χ. μηνιαία αντί για τριμηνιαία), καθώς και η επέκταση της χρονικής περιόδου μελέτης. Με τον τρόπο αυτό θα αυξηθεί ο αριθμός των παρατηρήσεων, και θα ενισχυθεί η στατιστική ισχύς και η αξιοπιστία των συσχετίσεων με τη συμφόρηση.

Για την καλύτερη αποτύπωση της επιρροής των καιρικών φαινομένων στη μέση ταχύτητα κίνησης των αστικών οδών, προτείνεται η εγκατάσταση ή αξιοποίηση **επιπλέον σταθμών καταγραφής** πιο κοντά στις οδούς ανάλυσης. Έτσι θα αποτυπώνονται με μεγαλύτερη ακρίβεια οι τοπικές καιρικές συνθήκες που επικρατούν σε κάθε οδικό τμήμα, αντί της χρήσης ενός μόνο σταθμού για ολόκληρο το σύνολο των οδών.

Η αυτόματη καταγραφή της σύνθεσης κυκλοφορίας με **κάμερες τεχνητής νοημοσύνης** (AI), θα βοηθήσει ώστε να συλλεχθεί σημαντικά μεγαλύτερος όγκος δεδομένων, με συνεχή και ομοιόμορφη κάλυψη, και να εξαχθούν πιο αξιόπιστα αποτελέσματα. Η προσέγγιση αυτή θα διευκόλυνε επίσης τη διερεύνηση της σύνθεσης της κυκλοφορίας σε δυσμενείς καιρικές συνθήκες και, κατά συνέπεια, της επιρροής των μετεωρολογικών παραμέτρων σε αυτήν. Επιπλέον, θα ήταν εφικτή η προσθήκη δεδομένων συνολικού φόρτου κυκλοφορίας το οποίο θα επέτρεπε την εξέταση της μείωσης του συνόλου των οχημάτων στην κυκλοφορία ή ορισμένων κατηγοριών.

7. Βιβλιογραφία

Aftabuzzaman, M. (2007). *Measuring Traffic Congestion-A Critical Review*.

And, A. T. I., & Hall, F. L. (1994). *Effect of Adverse Weather Conditions on Speed-Flow-Occupancy Relationships*.

Bai, M., Lin, Y., Ma, M., & Wang, P. (2018). Travel-time prediction methods: A review. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11344 LNCS*, 67–77. https://doi.org/10.1007/978-3-030-05755-8_7

Bates, J., Polak, J., Jones, P., & Cook, A. (2001). *The valuation of reliability for personal travel*. https://www.sciencedirect.com/science/article/abs/pii/S1366554500000119

Bi, H., Ye, Z., & Zhu, H. (2022). Data-driven analysis of weather impacts on urban traffic conditions at the city level. *Urban Climate*, *41*. https://doi.org/10.1016/j.uclim.2021.101065

Breiman, L. (2001). *randomforest2001*.

Carrion, C., & Levinson, D. (2012). Value of travel time reliability: A review of current evidence. *Transportation Research Part A: Policy and Practice*, *46*(4), 720–741. https://doi.org/10.1016/j.tra.2012.01.003

Cheng, J., Li, G., & Chen, X. (2019). Research on travel time prediction model of freeway based on gradient boosting decision tree. *IEEE Access*, *7*, 7466–7480. https://doi.org/10.1109/ACCESS.2018.2886549

Cools, M., Moons, E., & Wets, G. (2010). Assessing the impact of weather on traffic intensity. *Weather, Climate, and Society*, *2*(1), 60–68. https://doi.org/10.1175/2009WCAS1014.1

Dadashova, B., Li, X., Turner, S., & Koeneman, P. (2021). Multivariate time series analysis of traffic congestion measures in urban areas as they relate to socioeconomic indicators. *Socio-Economic Planning Sciences*, *75*. https://doi.org/10.1016/j.seps.2020.100877

Ding, L., & Zhang, N. (2016). Dynamics in Mode Choice Decisions: A Case Study in Nanjing, China. *Procedia Engineering*, *137*, 31–40. https://doi.org/10.1016/j.proeng.2016.01.231

Ery Sugito, & Diana Binti Mohamad. (2025). The Impact of Urban Traffic Density on The Service Quality and Performance of Public Bus Transport. *PROGRESS IN ENERGY AND ENVIRONMENT*, *31*(2), 67–78. https://doi.org/10.37934/progee.31.2.6778

Faber, R. M., Jonkeren, O., de Haas, M. C., Molin, E. J. E., & Kroesen, M. (2022). Inferring modality styles by revealing mode choice heterogeneity in response to weather conditions. *Transportation Research Part A: Policy and Practice*, *162*, 282–295. https://doi.org/10.1016/j.tra.2022.06.003

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. In *Source: The Annals of Statistics* (Vol. 29, Number 5). http://www.jstor.orgURL:http://www.jstor.org/stable/2699986

Galich, A., & Nieland, S. (2023). The Impact of Weather Conditions on Mode Choice in Different Spatial Areas. *Future Transportation*, *3*(3), 1007–1028. https://doi.org/10.3390/futuretransp3030056

Guangchuan Yang, Daobin Wang, & Xuesong Mao. (2020). *Modelling the modal shift effects of converting a general traffic lane into a dedicated bus lane*.

He, M., Li, J., Shi, Z., Liu, Y., Shuai, C., & Liu, J. (2022). Exploring the Nonlinear and Threshold Effects of Travel Distance on the Travel Mode Choice across Different Groups: An Empirical Study of Guiyang, China. *International*

Journal of Environmental Research and Public Health, *19*(23). https://doi.org/10.3390/ijerph192316045
Koźlak, A., & Wach, D. (2018). Causes of traffic congestion in urban areas. Case of Poland. *SHS Web of Conferences*, *57*, 01019. https://doi.org/10.1051/shsconf/20185701019
Lefèvre, S., Sun, C., Bajcsy, R., & Laugier, C. (2014). *Comparison of Parametric and Non-Parametric Approaches for Vehicle Speed Prediction*.
Lin, C. C., Ho, M. C., Hung, C. C., & Hsu, H. H. (2025). A comparative study and simple baseline for travel time prediction. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-02303-5
Mitchell, T. M. . (1997). *Machine Learning*. McGraw-Hill.
Mohan Rao, A., & Ramachandra Rao, K. (2012). MEASURING URBAN TRAFFIC CONGESTION-A REVIEW. *International Journal for Traffic and Transport Engineering*, *2*(4). https://doi.org/10.7708/ijje.2012.2(4).01
Mokhlesur Rahman, M., Najaf, P., Gregory Fields, M., & Thill, J.-C. (2021). *Traffic Congestion and its Urban Scale Factors: Empirical Evidence from American Urban Areas*.
NACTO. (2016). Transit Street Design Guide. National Association of City Transportation Officials*. https://nacto.org/publication/transit-street-design-guide/introduction/why-transit-streets-matter/designing-to-move-people/*
Nguyen-Phuoc, D. Q., Young, W., Currie, G., & De Gruyter, C. (2020). Traffic congestion relief associated with public transport: state-of-the-art. *Public Transport*, *12*(2), 455–481. https://doi.org/10.1007/s12469-020-00231-3
Nissen, K. M., Becker, N., Dahne, O., Rabe, M., Scheffler, J., Solle, M., & Ulbrich, U. (2020). How does weather affect the use of public transport in Berlin? *Environmental Research Letters*, *15*(8). https://doi.org/10.1088/1748-9326/ab8ec3
Ortúzar, J. de D., & Willumsen, L. G. (2011). *Modelling Transport*. John Wiley & Sons.
Qiu, B., & Fan, W. (2021). Machine learning based short-term travel time prediction: Numerical results and comparative analyses. *Sustainability (Switzerland)*, *13*(13). https://doi.org/10.3390/su13137454
Sabir, M., Koetse, M. J., & Rietveld, P. (2007). *The impact of weather conditions on mode choice: Empirical evidence for the Netherlands*.
Sakib, N. (2024). MACHINE LEARNING-BASED TRAVEL TIME PREDICTION MODELS FOR COMPLEX TRAFFIC CONDITIONS. In *CUET*. https://icace2024.cuet.ac.bd
Sultana, T., Sisiopiku, V. P., Khalil, J., & Yan, D. (2022). Potential Benefits of Increased Public Transit Ridership in Medium Sized Cities: A Case Study. *Journal of Transportation Technologies*, *12*(01), 59–79. https://doi.org/10.4236/jtts.2022.121004
Tsapakis, I., Cheng, T., & Bolbol, A. (2013). Impact of weather conditions on macroscopic urban travel times. *Journal of Transport Geography*, *28*, 204–211. https://doi.org/10.1016/j.jtrangeo.2012.11.003
Van Lint, J. W. C. (2004). *Reliable Travel Time Prediction for Freeways*.
Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, *58*, 308–324. https://doi.org/10.1016/j.trc.2015.02.019