



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ & ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ**

ΠΑΡΑΓΟΝΤΕΣ ΕΠΙΡΡΟΗΣ ΤΗΣ ΥΠΕΡΒΑΣΗΣ ΤΩΝ ΟΡΙΩΝ ΤΑΧΥΤΗΤΑΣ ΣΤΟ ΟΔΙΚΟ ΔΙΚΤΥΟ ΤΗΣ ΑΘΗΝΑΣ

Διπλωματική Εργασία



ΔΑΦΝΗ ΚΥΠΡΟΥΛΗ

Επιβλέπων: Γιώργος Γιαννής, Καθηγητής ΕΜΠ

Αθήνα, Ιούλιος 2025

Ευχαριστίες

Καταρχήν, θα ήθελα να ευχαριστήσω θερμά τον κ. Γιώργο Γιαννή, Καθηγητή του τμήματος Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου, ο οποίος ανέλαβε την ανάθεση και επίβλεψη της παρούσας Διπλωματικής Εργασίας. Θα ήθελα να τον ευχαριστήσω ακόμη, για την καθοδήγησή του καθ' όλη την διάρκεια εκπόνησης της Εργασίας αυτής.

Επίσης, θα ήθελα να ευχαριστήσω εξίσου θερμά τον Δημήτρη Νικολάου, Διδάκτορα Πολιτικό Μηχανικό και Ερευνητή του Εθνικού Μετσόβιου Πολυτεχνείου, για την συνεχή και σημαντικότερη συνεισφορά του σε όλο το διάστημα της έρευνας και της συγγραφής, καθώς επίσης και για τα εποικοδομητικά σχόλια τα οποία οδήγησαν στην βελτιστοποίηση του τελικού αποτελέσματος της Εργασίας.

Ιδιαίτερες ευχαριστίες απευθύνω στην εταιρία OSeven Telematics, για τα δεδομένα που μοιράστηκε κατά το αρχικό στάδιο της Εργασίας και αποτέλεσαν πυρήνα της παρούσας έρευνας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου για τη συνεχή στήριξη που μου παρείχαν με τον δικό τους μοναδικό τρόπο σε όλη την διάρκεια αυτής της διαδρομής. Η υποστήριξή τους υπήρξε καθοριστική για να φτάσω στο τέλος ενός μακροχρόνιου αλλά γεμάτου εμπειρίες και γνώση ταξιδιού, το οποίο ξεκίνησε με την εισαγωγή μου στη Σχολή Πολιτικών Μηχανικών του Ε.Μ.Π. και ολοκληρώνεται με την παρούσα Διπλωματική Εργασία.

Αθήνα, Ιούλιος 2025

Δάφνη Κυπρούλη

Παράγοντες Επιρροής της Υπέρβασης των Ορίων Ταχύτητας στο Οδικό Δίκτυο της Αθήνας

Δάφνη Κυπρούλη

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

Σύνοψη

Στόχο της παρούσας Διπλωματικής Εργασίας αποτελεί η εξέταση των παραγόντων επιρροής της υπέρβασης των ορίων ταχύτητας στο οδικό δίκτυο της Αθήνας. Η έρευνα αξιοποιεί δεδομένα που συλλέχθηκαν από την OSeven Telematics και το OpenStreetMaps με πληροφορίες όπως, δείκτες μορφολογικών χαρακτηριστικών οδών, δείκτες μέτρων ασφάλειας αλλά και συμπεριφοράς οδηγών.

Αναπτύχθηκαν στατιστικά μοντέλα και αλγόριθμοι μηχανικής μάθησης για δύο σενάρια, με στόχο την πρόβλεψη της υπέρβασης των ορίων ταχύτητας και την κατανόηση των παραγόντων που την επηρεάζουν. Το πρώτο επικεντρώθηκε στον έλεγχο παρουσίας ή απουσίας της υπέρβασης των ορίων ταχύτητας, ενώ το δεύτερο αφορούσε στην ανάλυση της υπέρβασης με ένα συγκεκριμένο όριο. Συνολικά τα μοντέλα που δημιουργήθηκαν ήταν δέκα. Μέσω των διαφορετικών αυτών μοντέλων, έγινε δυνατή η ανάδειξη της σημασίας συγκεκριμένων παραγόντων και η δυνατότητα τους να προβλέπουν την πιθανότητα υπέρβασης των ορίων ταχύτητας.

Τα αποτελέσματα έδειξαν ότι η υπέρβαση των ορίων ταχύτητας έχει στατιστικά σημαντική συσχέτιση με διάφορες μεταβλητές και η βελτίωση της συμπεριφοράς του οδηγού θα οδηγήσει κατ' επέκταση στη μείωση των ατυχημάτων. Οι κύριοι παράγοντες που επηρεάζουν την υπέρβαση των ορίων ταχύτητας στα εξεταζόμενα οδικά τμήματα, βρέθηκε να είναι ο αριθμός των διαδρομών, το μήκος του οδικού τμήματος και το ποσοστό χρήσης κινητού τηλεφώνου, ενώ την μικρότερη επιρροή παρουσιάζει η μεταβλητή της κλίσης των οδών.

Λέξεις κλειδιά: οδική ασφάλεια, υπέρβαση ορίου ταχύτητας, μοντέλα πρόβλεψης, χρήση κινητού τηλεφώνου, τύπος οδού, συμπεριφορά οδηγού.

Factors Influencing Speed Limit Violations on Athens Road Network

Daphne Kyprouli

Supervisor: George Yannis, Professor NTUA

Abstract

The objective of this Diploma Thesis is to examine the factors influencing speed limit violations across the entire road network of Athens. The study utilizes data collected by OSeven Telematics and OpenStreetMaps, which includes information such as road geometry indicators, safety measurements and driving behavior metrics.

Statistical models and machine learning algorithms were developed for two scenarios aiming to predict speeding violations and understand the factors influencing them. The first scenario focused on determining the presence or absence of speeding violations, while the second analyzed speeding relative to a specific threshold. Overall, ten models were created. Through these different models, it was possible to highlight the importance of specific factors and their ability to predict the probability of speed limit violations.

The results demonstrated that speeding has a statistically significant correlation with various variables and improving driving behavior will consequently lead to a reduction of road crashes. The main factors affecting the likelihood of speeding in the examined road sections are the number of trips, the road section length and the percentage of mobile phone use, while slopes presented the least impact.

Keywords: road safety, speed limit violations, prediction models, mobile phone usage, road type, driver behavior.

Περίληψη

Η **υπέρβαση του ορίου ταχύτητας** αποτελεί έναν από τους σημαντικότερους παράγοντες που επηρεάζουν την πιθανότητα ατυχημάτων, τόσο σε συχνότητα όσο και σε σοβαρότητα. Η ανάλυση της συγκεκριμένης παραμέτρου κρίνεται απαραίτητη για την κατανόηση και αντιμετώπιση των επικίνδυνων οδηγικών συμπεριφορών με στόχο την βελτίωση της συνολικής οδηγικής ασφάλειας.

Η παρούσα Διπλωματική Εργασία επικεντρώνεται στην **εξέταση των παραγόντων επιρροής της υπέρβασης των ορίων ταχύτητας στο οδικό δίκτυο της Αθήνας**. Πιο συγκεκριμένα, επιδιώκεται συνεισφορά στη βελτίωση της οδικής ασφάλειας, μέσα από τη δημιουργία ενός εργαλείου που θα αναγνωρίζει τους βασικούς παράγοντες που επηρεάζουν την υπέρβαση των ορίων ταχύτητας και κατ' επέκταση την πρόκληση οδικών ατυχημάτων και θα επιτρέπει την λήψη στοχευμένων μέτρων.

Αρχικά, πραγματοποιήθηκε η **βιβλιογραφική ανασκόπηση** (Κεφάλαιο 2), η οποία εστίασε στους παράγοντες που επηρεάζουν την υπέρβαση των ορίων ταχύτητας και την πιθανότητα οδικού ατυχήματος με βάση σύγχρονες μεθοδολογίες.

Στη συνέχεια, περιεγράφηκαν τα **στατιστικά εργαλεία και οι μέθοδοι ανάλυσης** (Κεφάλαιο 3) που χρησιμοποιήθηκαν, καθώς και τα κριτήρια αξιολόγησης των μοντέλων. Ειδικότερα, παρουσιάστηκαν οι βασικές αρχές της στατιστικής, όπως η ανάλυση της παλινδρόμησης, οι οποίες αξιοποιήθηκαν για την μοντελοποίηση των δεδομένων, αλλά και οι διαδικασίες αξιολόγησης των μεταβλητών για τον έλεγχο της αξιοπιστίας και της εγκυρότητας των αποτελεσμάτων. Επιπλέον, παρουσιάστηκε και το θεωρητικό υπόβαθρο των αλγορίθμων μηχανικής μάθησης που εφαρμόστηκαν.

Έπειτα, ακολούθησε η **συλλογή και επεξεργασία των δεδομένων** (Κεφάλαιο 4), όπου αξιοποιήθηκαν δεδομένα οδηγικής συμπεριφοράς υπό πραγματικές συνθήκες, τα οποία συλλέχθηκαν μέσω της εταιρίας OSeven Telematics. Τα δεδομένα αυτά, αντλήθηκαν από αισθητήρες κινητών τηλεφώνων και περιλάμβαναν κρίσιμες μεταβλητές, οι οποίες στη συνέχεια φιλτραρίστηκαν και υποβλήθηκαν σε επεξεργασία για την εξαγωγή των κρίσιμων δεικτών οδικής ασφάλειας. Επιπλέον, συλλέχθηκαν και επεξεργάστηκαν δεδομένα σχετικά με τα γεωμετρικά χαρακτηριστικά των οδών του εξεταζόμενου οδικού δικτύου, αντλώντας πληροφορίες μέσω του OpenStreetMaps.

Τέλος, στο κεφάλαιο της ανάλυσης των δεδομένων (Κεφάλαιο 5), αναπτύχθηκαν **στατιστικά μοντέλα για τη διερεύνηση των σχέσεων** μεταξύ των παραμέτρων. Πρώτα, έγινε η επιλογή των βασικών μεταβλητών μέσα από τον έλεγχο συσχέτισης και στη συνέχεια πραγματοποιήθηκε περιγραφική στατιστική ανάλυση των βασικών δεικτών. Μετά από κατάλληλη επεξεργασία, εφαρμόστηκαν οι επιλεγμένες στατιστικές μέθοδοι παλινδρόμησης και μηχανικής μάθησης για να διερευνηθεί η επίδραση κάθε παραμέτρου στην επιρροή της υπέρβασης των ορίων ταχύτητας από τους οδηγούς, ενώ παράλληλα δοκιμάστηκαν διαφορετικοί συνδυασμοί μεταβλητών για την ανάπτυξη των μοντέλων. **Τα μοντέλα που επιλέχθηκαν ήταν πέντε και χωρίστηκαν σε δύο σενάρια ελέγχου.**

Το πρώτο μοντέλο επικεντρώθηκε στον έλεγχο παρουσίας ή απουσίας της υπέρβασης των ορίων ταχύτητας, ενώ το δεύτερο αφορούσε στην ανάλυση της υπέρβασης με ένα συγκεκριμένο όριο. Η εφαρμογή των μοντέλων σε κάθε σενάριο επέτρεψε την διερεύνηση της επίδρασης διαφορετικών τύπων συμπεριφοράς

ταχύτητας στα αποτελέσματα της ανάλυσης και βελτίωση της ακρίβειας πρόβλεψης των μοντέλων.

Παρακάτω, δίνεται ο **συγκεντρωτικός πίνακας των αποτελεσμάτων** της ανάλυσης για τις μεθόδους που χρησιμοποιήθηκαν και για τα δύο σενάρια:

Σενάριο 1					
(0 = δεν καταγράφηκε υπέρβαση των ορίων ταχύτητας, 1 = καταγράφηκε υπέρβαση των ορίων ταχύτητας τουλάχιστον ενός δευτερολέπτου)					
	Binary Logistic	Random Forest	Decision Tree	SVM	Naive Bayes
Ορθότητα	0.778	0.784	0.771	0.781	0.767
Ευαισθησία	0.355	0.409	0.330	0.319	0.323
Ειδικότητα	0.942	0.928	0.941	0.958	0.938

Πίνακας: Συγκεντρωτικός πίνακας αποτελεσμάτων Σεναρίου 1

Σενάριο 2					
(0 = καταγράφηκε υπέρβαση των ορίων ταχύτητας μικρότερη των 15.55 δευτερολέπτων, 1 = καταγράφηκε υπέρβαση των ορίων ταχύτητας τουλάχιστον 15.55 δευτερολέπτων)					
	Binary Logistic	Random Forest	Decision Tree	SVM	Naive Bayes
Ορθότητα	0.941	0.945	0.938	0.939	0.916
Ευαισθησία	0.355	0.392	0.265	0.275	0.486
Ειδικότητα	0.989	0.989	0.993	0.994	0.951

Πίνακας: Συγκεντρωτικός πίνακας αποτελεσμάτων Σεναρίου 2

Κατά τα διάφορα στάδια εκπόνησης της παρούσας Διπλωματικής Εργασίας προέκυψε μία **σειρά συμπερασμάτων** που συνδέεται άμεσα με τους αρχικούς στόχους και τα ερωτήματα της, τα οποία συνοψίζονται ως εξής:

- Η διεθνής βιβλιογραφία ανέδειξε ότι, η **υπέρβαση των ορίων ταχύτητας** αποτελεί αδιαμφισβήτητα **κρίσιμο παράγοντα που συνδέεται άμεσα με την πιθανότητα ατυχήματος και τη σοβαρότητα των τραυματισμών**. Βρέθηκε ακόμη, ότι η αύξηση της ταχύτητας μειώνει τον χρόνο αντίδρασης και αυξάνει τη σφοδρότητα της σύγκρουσης.
- Η υπέρβαση των ορίων ταχύτητας **συνδέεται θετικά με μεταβλητές** όπως, ο **αριθμός διαδρομών ανά εξεταζόμενο οδικό τμήμα** (trip_count) και το **μήκος οδικού τμήματος** (length). Αυτό εξηγείται από το γεγονός ότι τόσο ο αριθμός διαδρομών ανά οδικό τμήμα, όσο και το μήκος του τμήματος μπορούν να θεωρηθούν δείκτες έκθεσης στον κίνδυνο.
- Παρατηρείται **αρνητική συσχέτιση της χρήσης κινητού τηλεφώνου** κατά την οδήγηση **με την παράμετρο της υπέρβασης των ορίων ταχύτητας**. Επομένως, η χρήση κινητού μπορεί να σχετίζεται με πιο προσεκτική οδήγηση, λόγω της απαίτησης ταυτόχρονων πολλαπλών κινήσεων και της αντιστάθμισης του κινδύνου. Ωστόσο, η απόσπαση της προσοχής ενδέχεται να προκαλεί άλλες επικίνδυνες συμπεριφορές.
- Η **υπέρβαση των ορίων ταχύτητας σχετίζεται θετικά με τις απότομες επιβραδύνσεις και τις απότομες επιταχύνσεις**. Η θετική συσχέτιση μεταξύ τους, οφείλεται στις αυξημένες απαιτήσεις αντίδρασης που προκύπτουν σε υψηλές ταχύτητες. Αναλυτικά, όταν το όχημα κινείται γρηγορότερα από το επιτρεπτό όριο, ο οδηγός έχει λιγότερο χρόνο να προσαρμοστεί στις συνθήκες

της οδού, οδηγώντας συχνά σε απότομες επιβραδύνσεις για την αποφυγή κινδύνων. Παράλληλα, η υπερβολική ταχύτητα συχνά σχετίζεται με επιθετική οδήγηση, που περιλαμβάνει ξαφνικές επιταχύνσεις για τη διατήρηση υψηλών ταχυτήτων ή την προσπέραση.

- Οι απότομες επιταχύνσεις και επιβραδύνσεις συνδέονται με αυξημένο κίνδυνο υπέρβασης ταχύτητας στο πρώτο σενάριο, ενώ στο δεύτερο οι επιταχύνσεις δεν παραμένουν εξίσου σημαντικές. Αυτό φανερώνει ότι **η επιθετική οδήγηση είναι πιο καθοριστική για μικρές υπερβάσεις**. Συγκεκριμένα, η διαφορά στη συσχέτιση μεταξύ των δύο σεναρίων δείχνει ότι οι συμπεριφορές αυτές, συνδέονται κυρίως με σύντομες υπερβάσεις ταχύτητας, καθώς η επιθετική οδήγηση οδηγεί σε στιγμιαίες αλλαγές ταχύτητας. Αντίθετα, στις παρατεταμένες υπερβάσεις, οι επιταχύνσεις χάνουν τη σημασία τους, υποδεικνύοντας ότι αυτές σχετίζονται περισσότερο με σταθερές συμπεριφορές, όπως η συστηματική αδιαφορία για τα όρια ταχύτητας ή η επίδραση εξωτερικών παραγόντων, όπως τα χαρακτηριστικά της οδού.
- Ο ταξινομητής **Random Forest** σύμφωνα με τον δείκτη MeanDecreaseGini, έδειξε ότι και στα δύο σενάρια την **πιο κρίσιμη μεταβλητή αποτελεί ο αριθμός διαδρομών ανά εξεταζόμενο οδικό τμήμα** και σταθερά σημαντικές εξίσου παραμένουν το μήκος οδικού τμήματος αλλά και η χρήση κινητού τηλεφώνου κατά την οδήγηση. Τον χαμηλότερο ρόλο διατηρεί η μεταβλητή της κατηγορίας κλίσης της οδού.
- Ο ταξινομητής **Random Forest** παρουσίασε τις **καλύτερες επιδόσεις** ως προς την πρόβλεψη της υπέρβασης των ορίων ταχύτητας και στα δύο σενάρια που εξετάστηκαν.
- **Παρατηρείται σταθερά χαμηλή ευαισθησία** σε όλες τις μεθόδους και τα σενάρια. Γεγονός το οποίο, **υποδηλώνει δυσκολία στον εντοπισμό των θετικών περιπτώσεων**, πιθανώς λόγω ανισορροπίας στα δεδομένα, ή περιορισμένων χαρακτηριστικών που διαφοροποιούν το δείγμα.
- Συγκριτικά, **στο Σενάριο 2 όλες οι μέθοδοι έχουν υψηλότερη ακρίβεια σε σχέση με το Σενάριο 1**. Επομένως, το σενάριο αυτό είναι πιο αποτελεσματικό, αφού οδηγεί σε καλύτερη συνολική απόδοση για την εκτίμηση του αρχικού προβλήματος.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ	1
1.1 Γενική Ανασκόπηση	1
1.2 Στόχος.....	2
1.3 Μεθοδολογία.....	3
1.4 Δομή.....	4
2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ	6
2.1 Εισαγωγή	6
2.2 Όριο ταχύτητας.....	6
2.3 Λόγοι υπέρβασης των ορίων ταχύτητας.....	7
2.4 Παράγοντες επιρροής της οδικής ασφάλειας.....	8
2.5 Σύνοψη	11
3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....	13
3.1 Εισαγωγή	13
3.2 Βασικές έννοιες στατιστικής.....	13
3.3 Βασικές Μέθοδοι Στατιστικής Ανάλυσης.....	14
3.4 Αλγόριθμοι Μηχανικής Μάθησης	15
3.5 Στατιστικοί έλεγχοι και κριτήρια αποδοχής μοντέλου	16
4. ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ.....	20
4.1 Εισαγωγή	20
4.2 Συλλογή δεδομένων.....	20
4.2.1 Δεδομένα συμπεριφοράς οδηγού υπό πραγματικές συνθήκες	20
4.2.2 Διαδικασία συλλογής στοιχείων	22
4.3 Επεξεργασία δεδομένων	23
4.3.1 Διαμόρφωση δεδομένων.....	23
4.3.2 Σύνθεση αλγορίθμου	32
4.4 Σύνοψη - Παρατηρήσεις.....	33
5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ - ΑΠΟΤΕΛΕΣΜΑΤΑ	34
5.1 Εισαγωγή	34
5.2 Έλεγχος συσχέτισης μεταβλητών.....	34
5.3 Ανάπτυξη Μοντέλων Παλινδρόμησης στο Λογισμικό της R.....	36
5.3.1 Μοντέλο Δυναμικής Λογιστικής Παλινδρόμησης.....	37
5.3.2 Δέντρο Απόφασης	42
5.3.3 Μοντέλο Τυχαίων Δασμών.....	43

5.3.4 Μοντέλο Παλινδρόμησης Μηχανή Διανυσμάτων Υποστήριξης (SVM)	47
5.3.5 Μοντέλο Παλινδρόμησης Naive Bayes	49
5.4 Απεικόνιση Χάρτη	51
6. ΣΥΜΠΕΡΑΣΜΑΤΑ	53
6.1 Σύνοψη Αποτελεσμάτων	53
6.2 Συνολικά Συμπεράσματα	54
6.3 Προτάσεις Βελτίωσης	55
6.4 Προτάσεις για Περαιτέρω Έρευνα	56
7. ΒΙΒΛΙΟΓΡΑΦΙΑ	58

1. ΕΙΣΑΓΩΓΗ

1.1 Γενική Ανασκόπηση

Στη σύγχρονη κοινωνία, η οδική μετακίνηση αποτελεί αναπόσπαστο κομμάτι της καθημερινότητας των ανθρώπων και ενισχύει την οικονομική και κοινωνική τους δραστηριότητα. Ωστόσο, παρά τη συνεισφορά της, οι οδικές μεταφορές είναι άρρηκτα συνδεδεμένες με σημαντικούς κινδύνους, καθώς τα οδικά ατυχήματα παραμένουν μια από τις κύριες αιτίες θανάτου και τραυματισμών σε παγκόσμιο επίπεδο. **Η υπέρβαση του ορίου ταχύτητας (speeding) αναγνωρίζεται ως ένας από τους σημαντικότερους παράγοντες**, που συμβάλλουν στην πρόκληση αυτών των ατυχημάτων και αυξάνουν την σοβαρότητα των επιπτώσεων.

Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας (World Health Organization, 2023) **περίπου 1,19 εκατομμύρια άνθρωποι ετησίως χάνουν την ζωή τους σε οδικά ατυχήματα**, ενώ περισσότεροι από 3.200 πεθαίνουν καθημερινά. Τα θύματα των ατυχημάτων αυτών κυρίως αφορούν **σε ηλικίες 5 έως 29 ετών** και κοστίζουν ένα σημαντικό ποσοστό (περίπου 3%) του ΑΕΠ στις περισσότερες χώρες. Η υπέρβαση του ορίου ταχύτητας αποτελεί βασική αιτία των ατυχημάτων, καθώς μειώνει τον χρόνο αντίδρασης, αυξάνει την απόσταση ακινητοποίησης και οδηγεί σε συγκρούσεις με μεγάλη ανταλλαγή ενέργειας και ορμή, προκαλώντας σοβαρότερους τραυματισμούς ή και θανάτους.

Στην Ελλάδα, παρά τη μείωση των οδικών ατυχημάτων τις τελευταίες δεκαετίες, συγκεκριμένα κατά 54% από το 2010 έως το 2020, το πρόβλημα παραμένει έντονο. Πιο αναλυτικά, το 2021 καταγράφηκαν 57 θάνατοι ανά εκατομμύριο πληθυσμού, συγκριτικά με τον ευρωπαϊκό μέσο όρο των 45 (European Commission 2022) και έτσι **η χώρα εξακολουθεί να καταγράφει υψηλότερους δείκτες σε σύγκριση με τις υπόλοιπες της Ευρωπαϊκής Ένωσης** (European Transport Safety Council, 2021). Ιστορικά η χώρα παρουσιάζει υψηλή συχνότητα παραβίασης των ορίων ταχύτητας και άλλων επικίνδυνων οδηγικών συμπεριφορών.

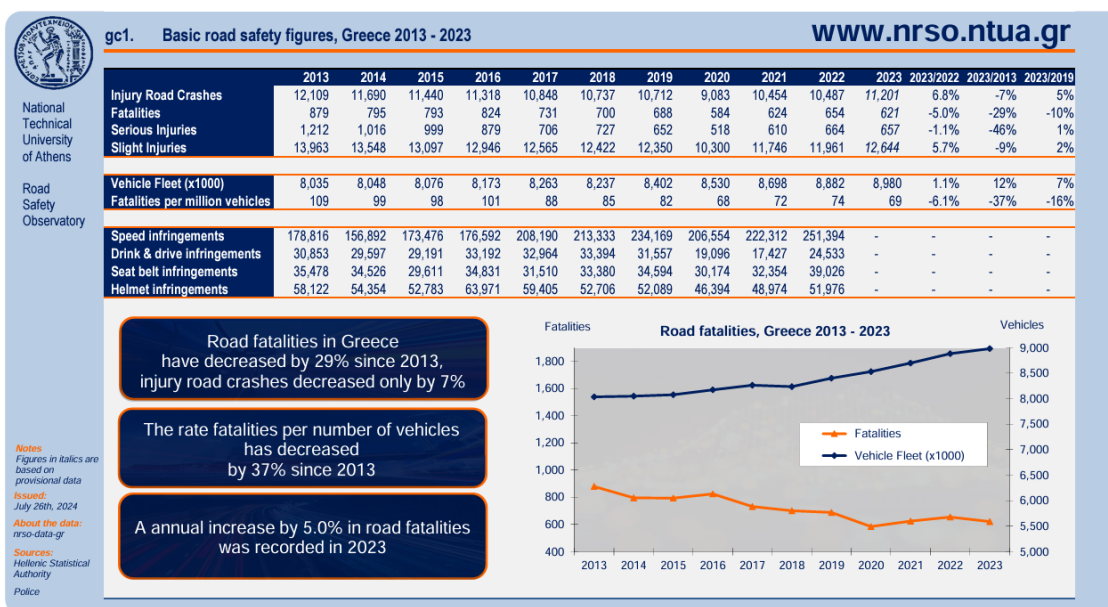
Αξίζει να σημειωθεί ότι, ως **οδικό ατύχημα** ορίζεται κάθε συμβάν το οποίο συντελείται σε οδούς δημόσιας χρήσης και οφείλεται σε ένα ή περισσότερα οχήματα, εκ των οποίων τουλάχιστον ένα βρισκόταν σε κίνηση, και προκαλεί τραυματισμό ή θάνατο ενός ή παραπάνω ατόμων. **Θανατηφόρο οδικό ατύχημα** χαρακτηρίζεται εκείνο στο οποίο ένας ή περισσότεροι συμμετέχοντες αποβιώνουν εξαιτίας αυτού, σε διάστημα τριάντα ημερών από το συγκεκριμένο συμβάν. Οι **τραυματίες**, δηλαδή τα άτομα που είτε υπέστησαν σωματική κάκωση ή γενικά βλάβη της υγείας τους, διαχωρίζονται σε δύο κατηγορίες, τους ελαφρά και τους βαριά τραυματίες. Ως **ελαφρά**, θεωρούνται όσοι αποκόμισαν απλή σωματική κάκωση χωρίς να τεθεί σε κίνδυνο η ζωής τους. Αντίθετα, **βαριά τραυματίες** θεωρούνται όσοι δέχτηκαν σοβαρή βλάβη στην υγεία τους και κινδυνεύουν να χάσουν την ζωή τους ή την διανοητική επαφή με το περιβάλλον τους.

Ακόμη, σύμφωνα με τη Στατική Επετηρίδα της Ελληνικής Αστυνομίας για το έτος 2023 στο σύνολο της χώρας, δίνεται ότι σημειώθηκαν 263.942 παραβάσεις για υπερβολική ταχύτητα, 44.383 για οδήγηση χωρίς χρήση προστατευτικών μέσων και 2.833 για χρήση κινητού τηλεφώνου κατά την οδήγηση. Από τα παραπάνω συμβάντα συγκεκριμένα στην περιοχή της Αττικής καταγράφηκαν 20.107 παραβάσεις για υπερβολική ταχύτητα, 10.406 για μη χρήση ζώνης ασφαλείας, 13.723 για μη χρήση κράνους πρόσκρουσης στα δίκυκλα και 5.466 για απόσπαση προσοχής λόγω χρήσης κινητού τηλεφώνου. Από τα παραπάνω φαίνεται ότι, **οι παραβάσεις των ορίων**

ταχύτητας αποτελούν μία από τις συχνότερες και πιο επικίνδυνες αιτίες οδικών ατυχημάτων στην Ελλάδα, ενώ η υπερβολική ταχύτητα επισημαίνεται ως κύριος παράγοντας που επιτείνει τις συνέπειές τους.

Επιπλέον, το Ινστιτούτο Οδικής Ασφάλειας «Πάνος Μυλωνάς», αναφέρει ότι περίπου το 40 - 50% των οδηγών παραδέχονται πως έχουν παραβιάσει τα όρια ταχύτητας, ενώ η υπερβολική ταχύτητα ευθύνεται για το 12% του συνόλου των οδικών συμβάντων στη χώρα. Επιπρόσθετα, η υπερβολική ταχύτητα συμβάλλει σημαντικά σε συγκρούσεις όπου εμπλέκεται μόνο ένα όχημα, κατηγορία τροχαίων που στην Ελλάδα συνδέεται με τον μεγαλύτερο αριθμό θανάτων πανευρωπαϊκά (ETSC, 2017).

Στην παρακάτω εικόνα παρουσιάζονται δεδομένα για τα οδικά ατυχήματα, τις παραβάσεις και τα θύματα στην Ελλάδα την περίοδο 2013 - 2023, σύμφωνα με στοιχεία που συγκεντρώθηκαν από το Εθνικό Μετσόβιο Πολυτεχνείο (ΕΜΠ). Τα στοιχεία αυτά επιβεβαιώνουν τη συνεχή παρουσία σοβαρών οδικών παραβάσεων, όπως η υπερβολική ταχύτητα, και υπογραμμίζουν την ανάγκη για συστηματική βελτίωση της οδικής ασφάλειας.



Εικόνα 1.1: Στατιστικά Οδικής Ασφάλειας στην Ελλάδα (2013–2023)

Με βάση τα παραπάνω, εντοπίζεται η επιτακτική ανάγκη μείωσης των ατυχημάτων αυτών, μέσω της ανάπτυξης του πεδίου της οδικής ασφάλειας. Αυτό μπορεί να εστιάσει στην κατανόηση αρχικά των αιτιών και στην συνέχεια στην εφαρμογή αποτελεσματικών μέτρων, με ιδιαίτερη προσοχή σε μέτρα καταπολέμησης της υπέρβασης των ορίων ταχύτητας. Η υπέρβαση ταχύτητας, ως ένας από τους κύριους παράγοντες πρόκλησης ατυχημάτων, καθιστά απαραίτητη τη συστηματική διερεύνηση και ανάλυσή της, με στόχο τη βελτίωση των υποδομών, τη θέσπιση αυστηρότερων μέτρων και την ενίσχυση της οδηγικής εκπαίδευσης.

1.2 Στόχος

Ο κύριος στόχος της παρούσας Διπλωματικής Εργασίας είναι η εξέταση των παραγόντων που επηρεάζουν την υπέρβαση των ορίων ταχύτητας, στο οδικό δίκτυο της Αθήνας.

Πιο συγκεκριμένα, **επιδιώκεται η συνεισφορά στη βελτίωση της οδικής ασφάλειας**, μέσα από τη δημιουργία ενός εργαλείου που θα αναγνωρίζει τους βασικούς παράγοντες που επηρεάζουν την υπέρβαση των ορίων ταχύτητας και κατ' επέκταση την πρόκληση οδικών ατυχημάτων και θα επιτρέπει την λήψη στοχευμένων μέτρων. Προκειμένου να επιτευχθεί ο στόχος αυτός, χρησιμοποιήθηκαν δεδομένα που συλλέχθηκαν από την OSeven Telematics, τα οποία περιλάμβαναν πληροφορίες για μορφολογικά χαρακτηριστικά οδών, δείκτες μέτρων ασφαλείας και οδηγικής συμπεριφοράς. Έπειτα, **πραγματοποιήθηκε ανάλυση των δεδομένων αυτών, με την εφαρμογή μαθηματικών στατιστικών μοντέλων, για την διερεύνηση της επίδρασης κάθε παραμέτρου στον κίνδυνο ατυχήματος**. Χρησιμοποιήθηκαν πέντε διαφορετικά μοντέλα, τα οποία χωρίστηκαν σε δύο σενάρια ελέγχου. Το πρώτο σενάριο επικεντρώθηκε στην παρουσία ή απουσία της υπέρβασης των ορίων ταχύτητας, ενώ το δεύτερο αφορούσε στην ανάλυση της υπέρβασης με ένα συγκεκριμένο όριο.

Τα **συμπεράσματα της εργασίας** αναμένεται να αποτελέσουν χρήσιμο εργαλείο για την κατανόηση των παραγόντων που επηρεάζουν την υπέρβαση των ορίων ταχύτητας. Μέσω των αποτελεσμάτων που θα προκύψουν από την ανάλυση, θα είναι δυνατή η ανάπτυξη στοχευμένων μέτρων για τη βελτίωση της οδικής ασφάλειας και η υιοθέτηση πολιτικών που θα ενισχύουν την συμμόρφωση με τα όρια ταχύτητας. Μακροπρόθεσμα, η εργασία αυτή μπορεί να παρέχει θετική επίδραση στη μείωση των οδικών ατυχημάτων, προστατεύοντάς ανθρώπινες ζωές.

1.3 Μεθοδολογία

Για την εκπόνηση της παρούσας Εργασίας και την επίτευξη του τελικού της στόχου ακολουθήθηκε η συγκεκριμένη μεθοδολογία, τα στάδια της οποίας παρουσιάζονται στην συνέχεια.

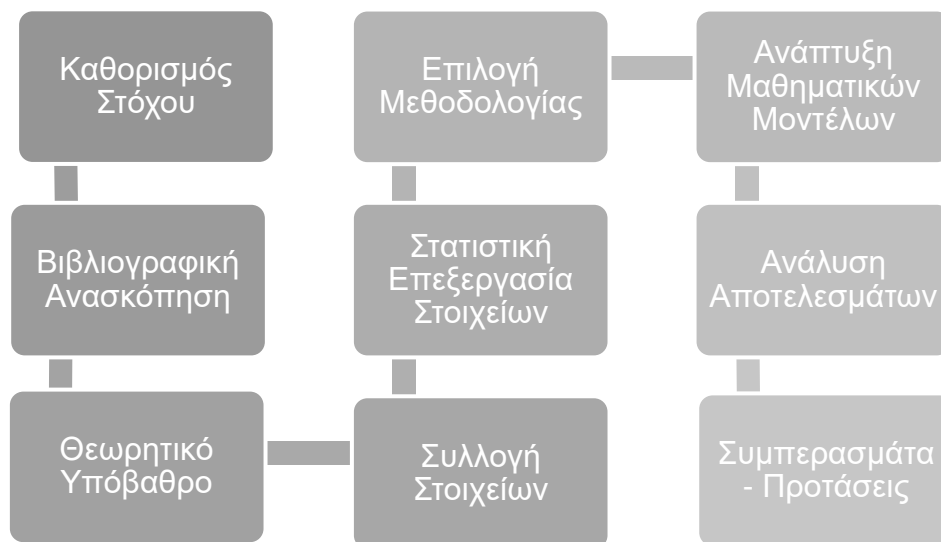
Αρχικά, οριστικοποιήθηκε το **θέμα της εργασίας** καθώς και ο επιδιωκόμενος στόχος της. Έπειτα, πραγματοποιήθηκε εκτενής **βιβλιογραφική ανασκόπηση**, τόσο σε ελληνική όσο και σε διεθνή βιβλιογραφία, μέσα από μελέτη αντίστοιχων ερευνών και επιστημονικών κειμένων, που θα μπορούσαν να φανούν χρήσιμες για την συγκεκριμένη έρευνα. Από την διαδικασία αυτή, καταβλήθηκε προσπάθεια να αποκτηθεί μια σχετική εμπειρία στην επεξεργασία τέτοιων θεμάτων και να αποφασιστεί η μέθοδος βάση στην οποία θα πραγματοποιηθεί στη συνέχεια η επεξεργασία των στοιχείων.

Επόμενο βήμα αποτέλεσε η **συλλογή των στοιχείων**, τα οποία θα επιτρέπουν την συσχέτιση της υπέρβασης των ορίων ταχύτητας με την ύπαρξη οδικών ατυχημάτων. Τα δεδομένα οδηγικής συμπεριφοράς συλλέχθηκαν μέσω της OSeven Telematics, από αισθητήρες κινητών τηλεφώνων σε πραγματικό χρόνο (Smartphones) και υποβλήθηκαν σε επεξεργασία για την εξαγωγή κρίσιμων δεικτών οδικής ασφάλειας.

Ύστερα, επιλέχθηκαν οι **μέθοδοι στατιστικής επεξεργασίας** των στοιχείων στο ειδικό λογισμικό στατιστικής ανάλυσης της γλώσσας προγραμματισμού (R), για να διερευνηθεί η επίδραση κάθε παραμέτρου στον κίνδυνο ατυχήματος. Τα μαθηματικά μοντέλα που αναπτύχθηκαν ήταν πέντε και χωρίστηκαν σε δύο σενάρια ελέγχου, με εξαρτημένη μεταβλητή τις περιπτώσεις υπέρβασης των ορίων ταχύτητας. Ανεξάρτητες μεταβλητές των μοντέλων αυτών αποτέλεσαν οι άλλες μεταβλητές που επιλέχθηκαν νωρίτερα. Κατόπιν, ακολούθησε η εφαρμογή των μοντέλων αυτών και η αξιολόγηση της ακρίβειας και της προβλεπτικής του ικανότητας μέσω κατάλληλων μετρικών.

Τέλος, έγινε η **αξιολόγηση και η ερμηνεία των αποτελεσμάτων** και εξήχθησαν τα αντίστοιχα συμπεράσματα για την κατανόηση των παραγόντων που επηρεάζουν την υπέρβαση των ορίων ταχύτητας. Μέσα από την ανάλυση αυτή, προέκυψαν πληροφορίες για το υπό εξέταση ζήτημα, αλλά και διατυπώθηκαν αξιολογες προτάσεις για περαιτέρω έρευνα.

Στην συνέχεια παρουσιάζονται σχηματικά τα στάδια της μεθοδολογίας που ακολουθήθηκαν για τις ανάγκες της παρούσας Διπλωματικής Εργασίας:



Εικόνα 1.2: Διάγραμμα ροής σταδίων εκπόνησης Διπλωματικής Εργασίας

1.4 Δομή

Στη συγκεκριμένη υποενότητα παρουσιάζεται η δομή της Διπλωματικής Εργασίας μέσω της συνοπτικής αναφοράς του περιεχομένου του εκάστοτε κεφαλαίου της.

Το **πρώτο κεφάλαιο αποτελεί μια εισαγωγή**, καθώς είναι και η βάση να γίνει αντιληπτό και κατανοητό το αντικείμενο της Διπλωματικής Εργασίας. Επίσης, γίνεται σαφής ο στόχος της εργασίας και επισημαίνεται η μεθοδολογία που θα ακολουθηθεί μέσα από ένα διάγραμμα ροής, καθώς και η δομή της.

Στο **δεύτερο κεφάλαιο παρουσιάζονται τα αποτελέσματα που προέκυψαν από εκτενή βιβλιογραφική ανασκόπηση** ερευνών με αντικείμενο τους λόγους υπέρβασης των ορίων ταχύτητας καθώς και άλλες μεταβλητές που σχετίζονται με την ύπαρξη οδικών ατυχημάτων.

Το **τρίτο κεφάλαιο αποτελεί το θεωρητικό υπόβαθρο**, στο οποίο βασίζεται η στατιστική ανάλυση των δεδομένων. Πιο συγκεκριμένα, περιγράφονται όλες οι μαθηματικές και στατιστικές έννοιες, οι οποίες βοηθούν στην καλύτερη κατανόηση και επεξήγηση των μοντέλων που χρησιμοποιήθηκαν.

Στο **τέταρτο κεφάλαιο αναφέρεται η διαδικασία συλλογής και επεξεργασίας των δεδομένων** που χρησιμοποιήθηκαν για την επίτευξη των στόχων της παρούσας εργασίας. Αρχικά, περιγράφεται η πηγή της συλλογής των δεδομένων και οι μεταβλητές που επεξεργάστηκαν για να χρησιμοποιηθούν στην ανάλυση.

Στο **πέμπτο κεφάλαιο εφαρμόστηκαν οι επιλεγμένες στατιστικές μέθοδοι** για την κατασκευή των τελικών μοντέλων, αφού πρώτα έγινε η επιλογή των βασικών μεταβλητών. Συγκεκριμένα τα μοντέλα ήταν πέντε και τα σενάρια ελέγχου δύο.

Παρουσιάζονται επίσης οι στατιστικοί έλεγχοι αξιοπιστίας των αποτελεσμάτων για την αποδοχή των μοντέλων και την αξιοπιστία των αποτελεσμάτων.

Στο **έκτο κεφάλαιο αναλύονται τα κυριότερα συμπεράσματα**, τα οποία προέκυψαν από τα προηγούμενα κεφάλαια και πιο συγκεκριμένα από την περιγραφική ανάλυση των στοιχείων και την εξαγωγή των τελικών μοντέλων. Ακόμη, στο ίδιο κεφάλαιο επισημαίνονται προτάσεις σχετικά με την αξιοποίηση των αποτελεσμάτων της Διπλωματικής Εργασίας, με στόχο την βελτίωση της οδικής ασφάλειας. Δίνονται επίσης και προτάσεις για περαιτέρω έρευνες που σχετίζονται με το συγκεκριμένο θέμα.

Τέλος, στο **έβδομο κεφάλαιο παρατίθενται οι βιβλιογραφικές αναφορές** σε αλφαβητική σειρά, οι οποίες χρησιμοποιήθηκαν για την εκπόνηση της Διπλωματικής Εργασίας. Ειδικότερα, πρόκειται για έναν κατάλογο, ο οποίος περιλαμβάνει έρευνες σχετικές με το θέμα της παρούσας εργασίας, καθώς και οποιεσδήποτε άλλες πηγές λήφθηκαν υπόψιν κατά την διαδικασία σύνταξης.

2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1 Εισαγωγή

Το παρόν κεφάλαιο αποτελεί τη βιβλιογραφική ανασκόπηση, η οποία πραγματοποιήθηκε στο πλαίσιο της εκπόνησης της παρούσας Διπλωματικής Εργασίας. Περιλαμβάνονται **δημοσιευμένες έρευνες** της ελληνικής και διεθνούς βιβλιογραφίας, με ανάλυση στη μεθοδολογία και το αντικείμενο τους. Συγκεκριμένα, αναζητήθηκαν έρευνες που επικεντρώνονται στην **ανάλυση και συσχέτιση της υπέρβασης των ορίων ταχύτητας με την πιθανότητα οδικού ατυχήματος** αλλά και στους παράγοντες επιρροής της υπέρβασης των ορίων ταχύτητας. Μέσω της παρουσίασης των ερευνών θα προκύψει ο στόχος της παρούσας μελέτης, η κατάλληλη μεθοδολογία για την επίτευξή του και η εύρεση παραμέτρων που δεν έχουν μελετηθεί επαρκώς ώστε να αποτελέσουν αντικείμενο έρευνας.

2.2 Όριο ταχύτητας

Ως υπερβολική ταχύτητα ορίζεται η οδήγηση σε συνθήκες, μεγαλύτερες από το δηλωμένο όριο ταχύτητας (Liu and Chen, 2009). Πολυάριθμες μελέτες έχουν δείξει ότι **η υπερβολική ταχύτητα θεωρείται ιδιαίτερα βασικός παράγοντας επιρροής**, τόσο της συχνότητας όσο και της σοβαρότητας των οδικών ατυχημάτων σε όλο τον κόσμο. Επίσης, έχει χαρακτηριστεί ως μία από τις μοιραίες πέντε συνθήκες, οι οποίες προκαλούν τα ατυχήματα οδικής κυκλοφορίας. Οι συνθήκες αυτές είναι, η ταχύτητα, η οδήγηση υπό την επήρεια αλκοόλ, η κόπωση, η απόσπαση της προσοχής και η μη χρήση των μέτρων ασφαλείας, όπως η χρήση ζώνης ή κράνους.

Αρχικά, ο Solomon (1964) ανέπτυξε μια σχέση της μορφής U, στην οποία απέδειξε ότι **τα οχήματα που ταξίδευαν με περίπου 10km/h πάνω από την μέση ταχύτητα της εκάστοτε οδού, είχαν το χαμηλότερο ποσοστό ατυχημάτων**. Ενώ τα οχήματα που ταξιδεύουν πολύ πιο γρήγορα ή πιο αργά από το όριο ταχύτητας, είχαν σημαντικά υψηλότερα ποσοστά ατυχημάτων. Έτσι, ήταν από τους πρώτους που ανέπτυξαν μια σχέση μεταξύ της υπερβολικής ταχύτητας και των ποσοστών των ατυχημάτων και οδηγήθηκε στο συμπέρασμα ότι τα οχήματα με ταχύτητα λίγο πάνω από την μέση, είναι τα ασφαλέστερα. Στην συνέχεια, και με βάση τις παραπάνω έρευνες, οι Nilsson (2004) και Elvik (2013) έδειξαν ότι οι χαμηλότερες μέσες ταχύτητες κυκλοφορίας έχουν ως αποτέλεσμα τη μειωμένη πιθανότητα ατυχημάτων, ενισχύοντας την αρχική θεωρία και δημιουργώντας μια γραμμική σχέση ταχύτητας και πιθανότητας ατυχημάτων.

Σε μια πιο εκτεταμένη έρευνα, η οποία έλαβε χώρα συγκεκριμένα για την περιοχή των Ηνωμένων Πολιτειών, δείχνει ότι η **υπερβολική ταχύτητα παίζει σημαντικό ρόλο στα θανατηφόρα ατυχήματα, εμπλεκόμενη σε περίπου 60% αυτών** των περιστατικών. Ένας λόγος είναι ότι οι υψηλές ταχύτητες οδηγούν σε μεγαλύτερες ανταλλαγές ενέργειας, εάν συμβούν συγκρούσεις, γεγονός που αυξάνει την πιθανότητα τραυματισμού (Munden, 1967). Ακόμη, είναι εύκολο να γίνει κατανοητό ότι στις ταχύτητες αυτές, ο χρόνος αντίδρασης στις αλλαγές του περιβάλλοντα χώρου είναι μικρότερος, η απόσταση ακινητοποίησης είναι μεγαλύτερη και η ικανότητα ελιγμών μειώνεται, επομένως η πιθανότητα σύγκρουσης αυξάνεται σημαντικά.

Πιο αναλυτικά αξίζει να σημειωθεί ότι στη συγκεκριμένη μελέτη, αποφάνθηκε ότι το 63% των συμμετεχόντων είχαν οδικά ατυχήματα, όπου το 67% αυτών των ατυχημάτων προκαλούνται από ανθρώπινα λάθη, το 29% από τις συνθήκες της οδού και το 4% από

ελαττώματα οχημάτων. Αυτή η παρατήρηση δίνει ένα επιπλέον στοιχείο για το βάρος της υπαιτιότητας του οδηγού και των υπολοίπων χαρακτηριστικών που συμβάλλουν.

Επίσης, διαπιστώθηκε ότι, **όπου τα μέγιστα όρια είναι στα 110km/h συνέβησαν περίπου 30% περισσότεροι θάνατοι από όταν ήταν στα 90-100km/h**. Ωστόσο, οι αυξήσεις στις ταχύτητες ταξιδιού έτειναν να είναι λιγότερο έντονες στα υψηλότερα αναρτημένα όρια (Davis et al., in press). Οι συγκεκριμένες καταγραφές δημιουργούν το ερώτημα για τον προσδιορισμό του κατάλληλου ορίου ταχύτητας, έτσι ώστε να επιτευχθεί ασφαλέστερη οδήγηση με λιγότερα ατυχήματα και παράλληλα να αποφευχθούν οι συχνές παραβάσεις του συγκεκριμένου ορίου.

Σε μια άλλη έρευνα του Nilsson (1982) και στην συνέχεια του Taylor (2002), φάνηκε ότι μια **αλλαγή στη μέση ταχύτητα της τάξης του 1km/h, θα έχει ως αποτέλεσμα μια αλλαγή στους αριθμούς των ατυχημάτων** που κυμαίνονται μεταξύ 2% για οδούς με όριο ταχύτητας 120km/h, 4% για οδούς με όριο 50km/h, 2,5% και 5,5% για αστικές και επαρχιακές οδούς αντίστοιχα.

2.3 Λόγοι υπέρβασης των ορίων ταχύτητας

Οι λόγοι υπέρβασης των ορίων ταχύτητας μπορούν να αποδοθούν σε διάφορους παράγοντες που σχετίζονται με τον οδηγό, το όχημα και τις συνθήκες της οδού. Συγκεκριμένα ως προς τα χαρακτηριστικά του οδηγού, έχει παρατηρηθεί ότι **η ηλικία, το φύλο και το εισόδημα** επηρεάζουν σημαντικά. Όσον αφορά την ηλικία, διαπιστώθηκε ότι οι νεότεροι οδηγοί, ιδιαίτερα στις ηλικίες 25-34, είναι πιθανό να υπερβούν τα όρια ταχύτητας 1,5 φορά περισσότερο σε σχέση με τους μεγαλύτερους οδηγούς. Το ποσοστό αυτό μειώνεται με την αύξηση της ηλικίας, με τους άνω των 60 να είναι οι πιο προσεκτικοί οδηγοί στα όρια ταχύτητας. Επίσης, έχει καταγραφεί ότι οι άντρες οδηγοί έχουν ελαφρώς μεγαλύτερες πιθανότητες να υπερβούν τα όρια ταχύτητας σε σχέση με τις γυναίκες, συγκεκριμένα κατά 1,1 φορά περισσότερο. Το εισόδημα σχετίζεται επίσης με την ταχύτητα, καθώς τα άτομα με υψηλότερα ετήσια εισοδήματα τείνουν να παραβιάζουν τα επιτρεπόμενα όρια πιο συχνά. Μια παρατήρηση που αξίζει να σημειωθεί, είναι αναφορικά με τους βαθμούς ποινής που προσδίδονται σε περιπτώσεις παράβασης των κανόνων οδικής κυκλοφορίας. Έχει διαπιστωθεί ότι οι βαθμοί ποινής έχουν άμεση συσχέτιση με την οδηγική συμπεριφορά, καθώς όσο περισσότεροι οι βαθμοί, τόσο μεγαλύτερη η πιθανότητα υπέρβασης της ταχύτητας μελλοντικά (Perez et al., 2021).

Στην συνέχεια ως προς την επιρροή του οχήματος στην παράμετρο της υπέρβασης των ορίων ταχύτητας, αναφέρονται διάφοροι λόγοι που σχετίζονται κυρίως με τα χαρακτηριστικά και τις δυνατότητες του οχήματος (Zijun Liang, Yun Xiao, 2020). Πιο αναλυτικά, **ο τύπος του οχήματος** είναι καθοριστικός, αφού προφανώς τα σπορ αυτοκίνητα ή τα αυτοκίνητα αγωνιστικών επιδόσεων συνδέονται συχνότερα με υπερβάσεις ταχύτητας, καθώς μπορούν να φτάσουν σε υψηλότερες ταχύτητες και με μεγάλες επιταχύνσεις, σε αντίθεση με οχήματα όπως φορτηγά ή βαρέα τα οποία είναι λιγότερο ευέλικτα. Αντίστοιχα, τα οχήματα με αυτόματο κιβώτιο ταχυτήτων υπερβαίνουν ευκολότερα τα όρια ταχύτητας, διότι δεν έχουν την αίσθηση της αύξησης της ταχύτητας που προσφέρει η αλλαγή ταχυτήτων στα χειροκίνητα οχήματα και ο έλεγχος μπορεί να χαθεί ευκολότερα. Με την εξέλιξη των συστημάτων υποστήριξης οδήγησης, δηλαδή με τις προειδοποιήσεις για υπέρβαση της ταχύτητας, την αυτόματη ρύθμιση σταθερής ταχύτητας κ.ά., πλέον αποτρέπεται η αλόγιστη επιτάχυνση σε αντίθεση με τα παλαιότερα οχήματα τα οποία δεν έχουν αντίστοιχες τεχνολογίες.

Ιδιαίτερα σημαντικός παράγοντας για την εμφάνιση ατυχημάτων που σχετίζονται με την ταχύτητα, αποτελούν **οι συνθήκες της οδού**, αυτό περιλαμβάνει τις καιρικές συνθήκες, τον τύπο οδού, την κατάσταση της οδού αλλά και τα όρια ταχύτητας. Γενικά έχει παρατηρηθεί ότι οι καλές οδικές συνθήκες συχνά οδηγούν σε χρήση υπερβολικής ταχύτητας σε οδούς μεγάλων ταχυτήτων όπως αυτοκινητοδρόμους, καθώς οι οδηγοί αισθάνονται μεγαλύτερη ασφάλεια και άνεση να κινούνται σε αυτούς. Σε καλά συντηρημένες και ευθύγραμμες οδούς, με επαρκή ορατότητα, οι οδηγοί τείνουν να υπερεκτιμούν τις ικανότητες τους και τις δυνατότητες του οχήματος. Από την άλλη, σε οδούς με πολλαπλές στροφές και απότομες κλίσεις που απαιτούν πιο προσεκτική οδήγηση, συχνά οι οδηγοί αμφισβητούν τα υπάρχοντα όρια ταχύτητας θεωρώντας τα συντηρητικά και τα υπερβαίνουν, αυξάνοντας την πιθανότητα ατυχημάτων. Σύμφωνα με κάποιες έρευνες που έχουν διεξαχθεί, εντοπίστηκε ότι πολλοί οδηγοί υπερβαίνουν το όριο ταχύτητας κατά ένα μικρό ποσοστό, περίπου σε 10km/h, κυρίως όταν οι κατηγορίες των ορίων ταχύτητας είναι πιο αυστηρές (Zijun Liang, Yun Xiao, 2020). Επίσης, τα ευρήματα της έρευνας έδειξαν ότι οι χαμηλές τιμές των ανακοινωθέντων ορίων ταχύτητας, η καθυστέρηση λόγω κυκλοφοριακής συμφόρησης και η έλλειψη επαρκούς επιβολής τους, ήταν οι τρεις κύριοι λόγοι που κάνουν τους οδηγούς να υπερβαίνουν τα όρια ταχύτητας. Τέλος επισημαίνεται ότι, οι συνεπιβάτες στο όχημα μπορούν να επηρεάσουν τον οδηγό να μειώσει ταχύτητα και να οδηγήσει πιο προσεκτικά, με γνώμονα την ασφάλεια τους.

Η υπερβολική ταχύτητα σε αρκετές μελέτες έχει συνδεθεί και με **ψυχολογικούς παράγοντες**, όπως η ικανοποίηση του εγωισμού, κατά την οποία οι οδηγοί αναζητούν ευχαρίστηση μέσω της ταχύτητας, η ανάληψη κινδύνων, οι χρονικές πιέσεις, η περιφρόνηση της οδήγησης, αφού θεωρείται χαμένος χρόνος, αλλά και η απροσεξία, όπου τα όρια ταχύτητας παραλείπονται λόγω αμέλειας (Gabany et al., 1998).

2.4 Παράγοντες επιρροής της οδικής ασφάλειας

Κάποιες από τις βασικότερες αιτίες οδικών ατυχημάτων είναι οι εξής:

1. Υπερβολική ταχύτητα
2. Απόσπαση προσοχής, κυρίως λόγω χρήσης κινητού
3. Μη χρήση ζώνης ασφαλείας ή κράνους
4. Επικίνδυνη και επιθετική οδήγηση, με ξαφνικές επιβραδύνσεις και επιταχύνσεις
5. Οδήγηση υπό την επήρεια αλκοόλ ή ναρκωτικών ουσιών
6. Κακές καιρικές συνθήκες

Η επιρροή της **υπερβολικής ταχύτητας (1)** στην πιθανότητα οδικού ατυχήματος έχει αναλυθεί εκτενώς παραπάνω.

Γενικά, οι ανθρώπινοι παράγοντες αποτελούν την βασική αιτία των οδικών ατυχημάτων συνολικά στο 65-95% (Sabey-Taylor 1980, Salmon et al. 2011). Μεταξύ αυτών των παραγόντων, η **απροσεξία και η απόσπαση της προσοχής (2)** συμβάλλει σε ποσοστό 15 έως 30% (MacEnvoy et al., 2007, Wang et al 1996). Η απόσπαση της προσοχής του οδηγού μπορεί να περιλαμβάνει τέσσερα διακριτά στοιχεία, την οπτική (π.χ. διαφημιστικές πινακίδες), την ακουστική (π.χ. ραδιόφωνο), την κινητήρια (π.χ. χρήση κινητού τηλεφώνου) και τη γνωστική (π.χ. συνομιλία με επιβάτη). Ακόμη, μπορούν να διαχωριστούν σε καταστάσεις που συμβαίνουν εξωτερικά ή εσωτερικά του οχήματος (Yannis et al., 2010). Όλα τα παραπάνω, θα έχουν αντίκτυπο στην προσοχή του οδηγού, δηλαδή στην απομάκρυνση του από το τιμόνι και από την οπτική επαφή με την οδό, στη συμπεριφορά του, δηλαδή στην ταχύτητα και στην θέση του οχήματος αλλά και στο χρόνο αντίδρασης σε πιθανότητα

ατυχήματος. Σε έρευνα που πραγματοποίησε ο MacEnoy (et al, 2005), τα αποτελέσματα υποδεικνύουν ότι η χρήση του κινητού τηλεφώνου μπορεί να είναι η σημαντικότερη πηγή απόσπασης της προσοχής στο όχημα για τους οδηγούς. Αν και οι οδηγοί τείνουν να μειώνουν την ταχύτητα τους κατά την διάρκεια της χρήσης του κινητού τηλεφώνου και η μειωμένη ταχύτητα συνδέεται γενικά με χαμηλότερο κίνδυνο ατυχήματος, εκείνοι που χρησιμοποιούν το κινητό τους τηλέφωνο κατά την διάρκεια που οδηγούν παρουσιάζουν έως και 4 φορές μεγαλύτερο κίνδυνο ατυχήματος, κυρίως λόγω του αυξημένου φόρτου εργασίας και της καθυστερημένης αντίδρασης.

Επιπλέον, πολλές μελέτες έχουν βρει ότι η συνομιλία σε ένα τηλέφωνο hands-free κατά την οδήγηση δεν είναι ασφαλέστερη από την χρήση ενός τηλεφώνου χειρός (Matthews et al. 2003). Γενικά, περίπου το ένα τρίτο των μετακινήσεων περιλαμβάνει κάποιο είδος αλληλεπίδρασης του οδηγού με κινητό τηλέφωνο, με συνηθέστερη την συνομιλία ή ακρόαση του τηλεφώνου. Πιο συγκεκριμένα με βάση μια έρευνα που διεξήγαγαν οι Farmer (et. al., 2015), εξετάστηκε η σχέση μεταξύ του κινδύνου ατυχήματος και του χρόνου χρήσης του κινητού τηλεφώνου κατά την οδήγηση και βρέθηκε ότι το 12% του χρόνου οδήγησης αφορά σε χρήση του κινητού με ποσοστά 6,5% για ομιλία/ακρόαση, 0,5% για εντοπισμό/απάντηση, 0,8% για κλήση, 0,3% για προβολή και 3,8% πληκτρολόγηση/συμμετοχή. Όταν το όχημα ακινητοποιείται, τα ποσοστά αυτά αυξάνονταν και οι οδηγοί απαντούσαν ή καλούσαν πιο συχνά σε σύγκριση με τους οδηγούς στα κινούμενα αυτοκίνητα. Όμως ακόμα και σε αυτές τις περιπτώσεις που τα οχήματα σταματούν την κυκλοφορία ή βρίσκονται σταθμευμένα στην άκρη του δρόμου, ο κίνδυνος σύγκρουσης εξακολουθεί να υφίσταται. Στην ίδια έρευνα, παρατηρήθηκε επίσης ότι τα ποσοστά συμβάντων συντριβής και χρήσης κινητών τηλεφώνων ήταν υψηλότερα για γυναίκες οδηγούς και για οδηγούς νεαρών ηλικιών 16-24 ετών εντός κατοικημένων περιοχών.

Με βάση προηγούμενες μελέτες, τα τρία πέμπτα των τροχαίων ατυχημάτων συμβαίνουν ως αποτέλεσμα παραγόντων συμπεριφοράς. Η χρήση προστατευτικού εξοπλισμού, όπως οι **ζώνες ασφαλείας και τα κράνη (3)** είναι η πιο σημαντική προστατευτική συμπεριφορά κατά την οδήγηση. Έχει παρατηρηθεί μια εξαιρετικά σημαντική μείωση 31% των πιθανοτήτων θανάτου μεταξύ των χρηστών προστατευτικού εξοπλισμού. Οι πιθανότητες θανάτου είναι υψηλότερες κατά περίπου 50% μεταξύ των μη χρηστών της ζώνης ασφαλείας από ό,τι μεταξύ των χρηστών και υψηλότερη μεταξύ των οδηγών παρά μεταξύ των επιβατών αυτοκινήτων (Febres et. al., 2020). Η συγκεκριμένη έρευνα έδειξε ότι η υψηλή πιθανότητα εκδήλωσης σοβαρού και θανατηφόρου ατυχήματος με ποσοστό περίπου 13%, συμβαίνει όταν ο οδηγός δεν χρησιμοποιεί ζώνη ασφαλείας. Εκτός από αυτό, άλλες μεταβλητές που συνδέονται άμεσα με την σοβαρότητα τραυματισμού είναι η απόσπαση προσοχής και ο τύπος δρόμου, με πιθανότητα κινδύνου 7,04% και 5,53% αντίστοιχα.

Επίσης, όσον αφορά τον δημογραφικό παράγοντα, οι άνδρες οδηγοί και οι οδηγοί κάτω των 25 και άνω των 60 ετών είναι πιο πιθανό να υποστούν σοβαρό ή/και θανατηφόρο τραυματισμό σε οδικό ατύχημα. Συμπερασματικά, υπολογίζεται ότι μια αύξηση της χρήσης προστατευτικού εξοπλισμού κατά 1 ποσοστιαία μονάδα, σώζει 136 ζωές ετησίως και ότι μια αύξηση 1% στη χρήση μειώνει κατά περίπου 0,13% τους ετήσιους θανάτους.

Σε μια μελέτη των Manouchehrifar (et al., 2018), διαπιστώθηκε ότι σε όλες τις τοποθεσίες το υψηλότερο ποσοστό ατυχημάτων συσχετίζεται με μοτοσικλέτες. Οι πιο συχνά τραυματισμένες περιοχές ήταν το κεφάλι και τα άνω άκρα με την πλειονότητα των θυμάτων να είναι εκείνοι που δεν είχαν δέσει τις ζώνες ασφαλείας και δεν είχαν

χρησιμοποιήσει κράνος πρόσκρουσης. Επιπρόσθετα αξίζει να σημειωθεί ότι, η απουσία αερόσακου είχε ως αποτέλεσμα ακόμη σοβαρότερους τραυματισμούς. Έτσι, μπορεί να γίνει κατανοητό ότι η μη χρήση προστατευτικού εξοπλισμού θα αύξανε τον κίνδυνο τραυματισμού περισσότερο από 5 φορές σε σύγκριση με την χρήση ζώνης και κράνους. Το τεστ 'Fisher's Exact', το οποίο χρησιμοποιείται για να εξετάσει την συσχέτιση μεταξύ δύο μεταβλητών, αποκάλυψε ότι η τιμή πιθανότητας είναι 94% ($P < 0,05$), που σημαίνει ότι υπάρχει ισχυρή στατιστική συσχέτιση μεταξύ της χρήσης εξοπλισμού ασφαλείας και της θέσης του τραύματος.

Επίσης, βρέθηκε μια σύνδεση στην μεταβλητή **του τύπου της οδού** και τον κίνδυνο τραυματισμού και θανάτου των οδηγών που δεν χρησιμοποιούν ζώνες ασφαλείας, κυρίως για τους οδούς υψηλής και μέσης ταχύτητας. Σε άρθρο των Sun (et. al., 2021), που διερευνούν την συσχέτιση αυτών των δύο μεταβλητών αναφέρεται ότι το ποσοστό θανάτων από οδικά ατυχήματα σε διοικητικές οδούς είναι το μεγαλύτερο, ακολουθούμενο από αυτό σε λειτουργικές. Μεταξύ του διοικητικού βαθμού τα περισσότερα ατυχήματα συμβαίνουν σε επαρχιακές οδούς, διότι η ποιότητα και το περιβάλλον τους είναι χαμηλό και υπάρχουν πολλές διασταυρώσεις, ενώ οι λιγότεροι θάνατοι παρατηρούνται σε δημοτικές οδούς. Από την άλλη, μεταξύ του λειτουργικού βαθμού τα περισσότερα οδικά ατυχήματα και θάνατοι σημειώνονται σε οδούς τέταρτης κατηγορίας, δηλαδή οδούς χαμηλότερης κυκλοφοριακής σημασίας, συνήθως τοπικές ή αγροτικές. Όσο υψηλότερος είναι ο βαθμός της λειτουργικής οδού τόσο λιγότερα ατυχήματα και θάνατοι πραγματοποιούνται, καθώς παρόλο που ο όγκος της κυκλοφορίας των αυτοκινητοδρόμων και των οδών πρώτης κατηγορίας είναι πολύ μεγαλύτερος, διαθέτουν πληρέστερα μέτρα ασφάλειας και σύστημα παρακολούθησης με περιορισμένη ταχύτητα και ραντάρ. Ένας ακόμη λόγος που συμβαίνει αυτό είναι διότι σε οδούς κάτω της πρώτης κατηγορίας επιτρέπεται η κίνηση όλων των ειδών των οχημάτων. Τελικά, μεταξύ όλων των οδών ο χαμηλότερος αριθμός RTA (τροχαίων οδικών ατυχημάτων) είναι στις αστικές οδούς ταχείας κυκλοφορίας.

Όσον αφορά την παράμετρο της **επικίνδυνης και επιθετικής οδήγησης (4)** σε μια μελέτη του Osafune et. al., 2016, αναλύθηκαν τα τέσσερα είδη τυπικής επικίνδυνης συμπεριφοράς τα οποία ορίστηκαν ως εξής:

- Ασυνήθιστη αύξηση ή μείωση της ταχύτητας: *απότομη αυξομείωση της ταχύτητας σε σύντομο χρονικό διάστημα, διακύμανση μεγαλύτερη από 30km/h σε 10 δευτερόλεπτα*
- Έντονες και απότομες στροφές του τιμονιού: *περισσότερο από 0,7 rad σε ταχύτητες μεγαλύτερες των 30km/h*
- Κίνηση σε ύφανση: *οδήγηση σε σχήμα S, εναλλαγές λωρίδων*
- Λειτουργία του κινητού τηλεφώνου κατά την οδήγηση

Γενικά, η περίπτωση που η ποσότητα διαμήκους επιτάχυνσης ή επιβράδυνσης υπερβαίνει το 1,0 m/s² θα θεωρείται ξαφνική. Οι τραυματισμοί από επιβράδυνση συμβαίνουν όταν ένα όχημα σταματά απότομα ή επιβραδύνει απότομα, με αποτέλεσμα οι επιβάτες να βιώσουν δυνάμεις που μπορεί να οδηγήσουν σε σοβαρούς τραυματισμούς οδηγώντας συχνά σε τραυματικές εγκεφαλικές βλάβες (Encyclopaedia Britannica). Αυτοί οι τραυματισμοί είναι πιο πιθανό να συμβούν όταν υπάρχουν γρήγορες αλλαγές στην ταχύτητα, δίνοντας έμφαση στους κινδύνους που συνδέονται με τα επιθετικά πρότυπα οδήγησης.

Η **οδήγηση υπό την επήρεια αλκοόλ ή ναρκωτικών (5)** συσχετίζεται σημαντικά με αυξημένο κίνδυνο οδικών ατυχημάτων, καθώς οι ουσίες αυτές βλάπτουν τις γνωστικές λειτουργίες, τους χρόνους αντίδρασης και τις σωματικές ικανότητες που είναι

απαραίτητες για την ασφαλή οδήγηση (Khatir et. al. 2022). Πάνω από το 55% των ατόμων που εμπλέκονται σε σοβαρά ή θανατηφόρα ατυχήματα βρέθηκαν θετικοί σε αλκοόλ ή ναρκωτικά στο σώμα τους. Ορισμένες ουσίες όπως η μεθαδόνη και η μορφίνη συνδέονται ιδιαίτερα με αυξημένο κίνδυνο ατυχήματος και ο συνδυασμός των ουσιών με αλκοόλ ή άλλα ναρκωτικά επιδεινώνει σημαντικά τον κίνδυνο ατυχήματος. Εντοπίστηκαν επιπλέον δημογραφικές τάσεις που δείχνουν ότι πιο επιρρεπείς σε αυτές τις καταστάσεις είναι οι νεότερες ηλικίες. Στοιχεία από τη Μεγάλη Βρετανία αποκάλυψαν μια αξιοσημείωτη αύξηση των θανάτων που αφορούσαν οδηγούς με προβλήματα στα ναρκωτικά, ξεπερνώντας αυτούς που αφορούσαν το αλκοόλ. Από το 2014 έως το 2022, ο αριθμός των νεκρών οδηγών από χρήση ναρκωτικών ουσιών αυξήθηκε πάνω από 70%, υποδηλώνοντας μια αυξανόμενη ανησυχία σχετικά με περιστατικά οδήγησης που σχετίζονται με ναρκωτικά (Department for Transport, 2023).

Τέλος, έχει παρατηρηθεί σημαντική συσχέτιση και με τις **καιρικές συνθήκες (6)** και τα οδικά ατυχήματα λόγω των επιπτώσεων τους στην απόδοση του οχήματος και την συμπεριφορά του οδηγού. Η κακοκαιρία αυξάνει σημαντικά την πιθανότητα των ατυχημάτων, δηλαδή περίπου το 21% του συνόλου των ατυχημάτων σχετίζονται με τις καιρικές συνθήκες με περίπου 1,2 εκατομμύρια ατυχήματα κάθε χρόνο λόγω συνθηκών όπως βροχή, χιόνι, πάγο και ομίχλη. Πιο συγκεκριμένα, το 46% περίπου των ατυχημάτων συμβαίνουν σε συνθήκες βροχής, λόγω του υγρού οδοστρώματος που προκαλεί υδρολίσθηση των τροχών, μειωμένη ορατότητα αλλά και πρόσφυση των ελαστικών. Σε συνθήκες χιονιού και πάγου σημειώνεται το ποσοστό του 18%, καθώς οι χιονισμένοι οδοί επηρεάζουν τον έλεγχο των οχημάτων και τις αποστάσεις ακινητοποίησης, ενώ ο παγετός προκαλεί ολίσθηση και απώλεια ελέγχου πέδησης. Η ομίχλη μειώνει δραματικά την ορατότητα και προκαλεί δυσκολίες στην ανάγνωση πινακίδων και τον εντοπισμό οχημάτων και εμποδίων. Συνεπώς, οι δυσμενείς καιρικές συνθήκες επηρεάζουν την συμπεριφορά του οδηγού ακόμα και στην αντίθετη περίπτωση, όπου γίνονται υπερβολικά προσεκτικοί και οδηγούνται σε μικρότερες ταχύτητες και αυξημένη κυκλοφοριακή συμφόρηση (Kuchcik et. al., 2023).

2.5 Σύνοψη

Στο παρόν κεφάλαιο εξετάστηκαν διεξοδικά παράγοντες που σχετίζονται με την οδική ασφάλεια και την υπέρβαση των ορίων ταχύτητας, με βάση τη βιβλιογραφία και τα αποτελέσματα από πρόσφατες σχετικές μελέτες. Οι παράγοντες αυτοί αναλύονται τόσο σε επίπεδο οδηγικής συμπεριφοράς, όσο και σε επίπεδο επιρροής τους στην σοβαρότητα των ατυχημάτων αυτών. Τα σημαντικότερα **συμπεράσματα** που προκύπτουν συνοψίζονται παρακάτω.

Σύμφωνα με την βιβλιογραφική ανασκόπηση προέκυψε ότι έχουν πραγματοποιηθεί αρκετές έρευνες για τη διερεύνηση της επίδρασης πολλαπλών παραγόντων στην πιθανότητα ατυχημάτων, με ιδιαίτερη έμφαση στην υπερβολική ταχύτητα. Ωστόσο, συγκεκριμένα στην Ελλάδα έχουν πραγματοποιηθεί ελάχιστες παρόμοιες έρευνες και κυρίως έχουν επικεντρωθεί συγκεκριμένα στην παράμετρο του κινητού τηλεφώνου. Επίσης, **διαπιστώνεται ένα κενό στην έρευνα όσον αφορά την μελέτη της ταυτόχρονης συσχέτισης πολλαπλών μεταβλητών**, αντί της αποσπασματικής ανάλυσης της κάθε μεταβλητής ξεχωριστά. Αυτό το κενό στην έρευνα πιθανότατα να οφείλεται στην πολυπλοκότητα της ταυτόχρονης ανάλυσης πολλαπλών παραγόντων και στον υψηλό βαθμό δυσκολίας στην εξαγωγή συμπερασμάτων. Η μελέτη αυτή απαιτεί προηγμένες στατιστικές τεχνικές και λογισμικά. Μπορεί όμως να προκύψουν

εσφαλμένα ή μεροληπτικά συμπεράσματα λόγω της έλλειψης ταυτόχρονης ανάλυσης των μεταβλητών, καθώς η απομονωμένη εξέταση κάθε παράγοντα δεν λαμβάνει υπόψη τις αλληλεπιδράσεις που μπορεί να επηρεάζουν την πραγματική έκταση και ένταση του κινδύνου.

Πιο αναλυτικά, διαπιστώθηκε αρχικά ότι η **ταχύτητα αποτελεί καθοριστικό παράγοντα** για τη πιθανότητα πρόκλησης οδικών ατυχημάτων, **καθώς και μία από τις πέντε κύριες αιτίες οδικών ατυχημάτων**, μαζί με την οδήγηση υπό την επήρεια αλκοόλ, την κόπωση, την απόσπαση προσοχής και τη μη χρήση ζώνης ή κράνους. Το επίπεδο της ταχύτητας μπορεί να επηρεάσει διαφορετικά την πιθανότητα ατυχήματος, ανάλογα με το εκάστοτε όριο ταχύτητας σε κάθε οδό. Μελέτες, όπως του Solomon, έδειξαν ότι τα οχήματα με ταχύτητα ελαφρώς πάνω από τη μέση είναι πιο ασφαλή, ενώ τα πολύ υψηλά ή χαμηλά όρια ταχύτητας συνδέονται με αυξημένα ποσοστά ατυχημάτων. Επιπλέον, έρευνες των Nilsson και Elvik επιβεβαίωσαν ότι η **μείωση της μέσης ταχύτητας μειώνει τον κίνδυνο ατυχημάτων και τη σοβαρότητα των τραυματισμών**. Το γεγονός αυτό αναδεικνύει τη σημασία του ορθού καθορισμού των ορίων ταχύτητας, έτσι ώστε να μειωθεί ο κίνδυνος ατυχημάτων και να κατοχυρωθεί η ασφάλεια των οδηγών και των επιβατών.

Τα ανθρώπινα λάθη βρέθηκαν να είναι ο κρισιμότερος παράγοντας στην εμφάνιση των ατυχημάτων, ενώ μικρότερο ποσοστό αποδίδεται στις συνθήκες της οδού και τα ελαττώματα των οχημάτων. Νεότεροι και άνδρες οδηγοί, καθώς και όσοι έχουν υψηλότερο εισόδημα, παρουσιάζουν μεγαλύτερη πιθανότητα παραβίασης των ορίων. Τα σπορ οχήματα και τα αυτόματα κιβώτια διευκολύνουν την υπέρβαση της ταχύτητας. Παράλληλα, οι καλές οδικές συνθήκες ενισχύουν την αίσθηση ασφάλειας, ενθαρρύνοντας την υπέρβαση των ορίων, ενώ η κυκλοφοριακή συμφόρηση και η ανεπαρκής επιβολή των κανόνων αποτελούν πρόσθετους λόγους παραβίασης.

Παράλληλα όπως ήταν αναμενόμενο, βρέθηκε μία **έντονη συσχέτιση ανάμεσα στην χρήση ζώνης ασφαλείας και κράνους και στην σοβαρότητα των τραυματισμών** κατά τα ατυχήματα, αλλά και στη χρήση κινητού τηλεφώνου κατά την οδήγηση και την πιθανότητα εμπλοκής σε ατυχήματα. Τα παραπάνω συμπεράσματα καταδεικνύουν την επιτακτική ανάγκη επιβολής αυστηρότερων κανονισμών και ποινών, με στόχο τη διαμόρφωση ενός ασφαλέστερου περιβάλλοντος οδήγησης και την καθιέρωση των βέλτιστων συμπεριφορών οδήγησης από τους χρήστες της οδού.

Επιπλέον, η **χρήση κινητού τηλεφώνου κατά την οδήγηση αποδείχθηκε ένας από τους πιο επικίνδυνους παράγοντες**, καθώς μειώνει σημαντικά την προσοχή του οδηγού και αυξάνει τους χρόνους αντίδρασης του. Το γεγονός αυτό οδηγεί σε αυξημένη πιθανότητα εμπλοκής σε ατυχήματα. Οι παραπάνω έρευνες δείχνουν ότι η οδήγηση υπό την επίδραση περισπασμών, όπως είναι το κινητό τηλέφωνο, απαιτούν ιδιαίτερη προσοχή και επιβολή επιτακτικών μέτρων με σκοπό την μείωση της συχνότητας εμφάνισης τους.

Οι συγκεκριμένες μεταβλητές μαζί με άλλες συναφείς παραμέτρους, στην συνέχεια της διπλωματικής μου εργασίας θα αποτελέσουν κεντρικό σημείο ανάλυσης. Θα διερευνηθεί η συσχέτιση τους και ο βαθμός επίδρασης τους στην συχνότητα των ατυχημάτων που προκαλούνται από την υπέρβαση του ορίου ταχύτητας. Επομένως, η βιβλιογραφική ανασκόπηση παρέχει τις βάσεις για τον καθορισμό των βασικών παραμέτρων του μοντέλου πρόβλεψης που πρόκειται να αναπτυχθεί στην συνέχεια.

3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

3.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται το θεωρητικό υπόβαθρο που αξιοποιήθηκε για την ανάλυση του αντικειμένου της παρούσας Διπλωματικής Εργασίας. Αρχικά, γίνεται μια σύντομη περιγραφή για τις βασικές έννοιες της στατιστικής και μοντέλων που χρησιμοποιήθηκαν για την ανάλυση των δεδομένων. Επίσης, επεξηγείται η διαδικασία ομαδοποίησης των δεδομένων που συμπεριλήφθηκαν στην ανάλυση, τα κριτήρια αποδοχής ή απόρριψης των παραπάνω μοντέλων και ορισμένες εντολές που χρησιμοποιήθηκαν στον αλγόριθμο.

3.2 Βασικές έννοιες στατιστικής

Ο όρος **πληθυσμός** (population) αναφέρεται στο σύνολο των παρατηρήσεων ενός χαρακτηριστικού που εξετάζεται στην στατιστική έρευνα. Ο πληθυσμός μπορεί να είναι πραγματικός ή θεωρητικός.

Ο όρος **δείγμα** (sample) αναφέρεται σε ένα υποσύνολο του πληθυσμού, λόγω της αδυναμίας εξέτασης του συνόλου του πληθυσμού. Όλα τα στοιχεία ενός δείγματος ανήκουν και στον πληθυσμό, χωρίς όμως να ισχύει αναγκαστικά και το αντίστροφο. Τα συμπεράσματα που θα προκύψουν από την μελέτη του δείγματος θα ισχύουν με ικανοποιητική ακρίβεια για το σύνολο του πληθυσμού, εφόσον το δείγμα είναι αντιπροσωπευτικό αυτού.

Ως **μεταβλητές** (variables) ορίζονται τα χαρακτηριστικά που ενδιαφέρουν να μετρηθούν και να καταγραφούν σε ένα δείγμα και διακρίνονται στις παρακάτω κατηγορίες:

- **Ποιοτικές μεταβλητές** (qualitative variables): Είναι οι μεταβλητές, όπου οι δυνατές τιμές τους είναι κατηγορίες διαφορετικές μεταξύ τους. Η χρήση των αριθμών για την αναπαράσταση των τιμών μιας τέτοιας μεταβλητής είναι αποκλειστικά συμβολική και δεν έχει την έννοια της μέτρησης.
- **Ποσοτικές μεταβλητές** (quantitative variables): Είναι οι μεταβλητές με τιμές αριθμούς, οι οποίες έχουν την έννοια της μέτρησης και διακρίνονται σε διακριτές και συνεχείς. Διακριτή θεωρείται μια μεταβλητή, όταν η μικρότερη μη μηδενική διαφορά που μπορεί να έχουν δύο τιμές της είναι σταθερή ποσότητα. Συνεχής θεωρείται μια μεταβλητή όταν δυο τιμές της μπορούν να διαφέρουν κατά οποιαδήποτε μικρή ποσότητα, δηλαδή να πάρει οποιαδήποτε τιμή σε ένα διάστημα τιμών.

Εξαρτημένη μεταβλητή (dependent variable) είναι εκείνη η οποία μετράμε ή παρατηρούμε σε μία δοκιμή και τροποποιείται ως αποτέλεσμα του χειρισμού της ανεξάρτητης.

Ανεξάρτητη μεταβλητή (independent variable) είναι εκείνη που ελέγχεται ή τροποποιείται σε μία μελέτη με σκοπό την εξέταση της επίδρασης της εξαρτημένης και αποτελεί την αιτία.

Μέση τιμή (mean), ορίζεται το άθροισμα του συνόλου των τιμών, ως προς το πλήθος αυτών, με βάση την παρακάτω σχέση:

$$\bar{x} = \frac{(x_1 + x_2 + \dots + x_v)}{v} = \frac{1}{v} \sum_{i=1}^v x_i$$

Διάμεσος (median), είναι η τιμή που χωρίζει ένα σύνολο δεδομένων σε δύο ίσα μέρη, έτσι ώστε το 50% των παρατηρήσεων να βρίσκεται κάτω από αυτήν και το υπόλοιπο 50% πάνω, όταν τα δεδομένα είναι ταξινομημένα σε αύξουσα ή φθίνουσα σειρά.

Διακύμανση (variance), ορίζεται ο μέσος όρος των τετραγωνικών διαφορών από το μέσο όρο. Αν η διακύμανση είναι ίση με 0 τότε όλες οι τιμές της μεταβλητής είναι ίσες, σε κάθε άλλη περίπτωση, όσο μεγαλύτερη είναι η διαφορά μεταξύ των τιμών τόσο μεγαλύτερη θα είναι και η διακύμανση. Όταν λάβει τιμή μεγαλύτερη από τη μέση τιμή ονομάζεται υπερδιασπορά και είναι μεγαλύτερη από 1, ενώ αν είναι μικρότερη από 1 είναι γνωστή ως υποδιαστολή.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Ο **συντελεστής συσχέτισης** (correlation coefficient) εκφράζει το βαθμό και τον τρόπο που δυο τυχαίες μεταβλητές συσχετίζονται. Επιπλέον, λαμβάνει τιμές μεταξύ του -1 και του 1 και είναι ανεξάρτητος από την μονάδα μέτρησης των μεταβλητών. Τιμές κοντά στο 1 δηλώνουν ισχυρή συσχέτιση, τιμές κοντά στο -1 δηλώνουν ισχυρή αρνητική συσχέτιση, ενώ τιμές κοντά στο 0 δηλώνουν γραμμική ανεξαρτησία.

3.3 Βασικές Μέθοδοι Στατιστικής Ανάλυσης

Για να εξεταστεί η σχέση μεταξύ των μεταβλητών έτσι ώστε να είναι δυνατή η πρόβλεψη μιας από τις υπόλοιπες μεταβλητές, εφαρμόζεται η στατιστική ανάλυση παλινδρόμησης (regression analysis). Ως εξαρτημένη μεταβλητή ορίζεται εκείνη της οποίας η τιμή πρόκειται να προβλεφθεί, ενώ ως ανεξάρτητη εκείνη η οποία χρησιμοποιείται για την πρόβλεψη της εξαρτημένης. Οι ανεξάρτητες μεταβλητές παίρνουν καθορισμένες τιμές, ενώ η εξαρτημένη τυχαίες.

Γραμμική Παλινδρόμηση (linear regression)

Όταν η εξαρτημένη μεταβλητή είναι συνεχής και ακολουθεί κανονική κατανομή, γίνεται χρήση της μεθόδου της γραμμικής παλινδρόμησης. Αρχικά, γίνεται ο υπολογισμός της συνάρτησης χρησιμότητας κάποιου ορισμένου γεγονότος, σε σχέση με τους παράγοντες που το επηρεάζουν καταλήγοντας σε ένα γραμμικό μαθηματικό πρότυπο. Στη συνέχεια, υπολογίζεται η πιθανότητα πραγματοποίησης του γεγονότος, η οποία αποτελεί το πρότυπο πρόβλεψης της πιθανότητας. Στην περίπτωση που η εξαρτημένη μεταβλητή εξαρτάται μόνο γραμμικά από μια ανεξάρτητη μεταβλητή, η παλινδρόμηση είναι απλή, ενώ στην περίπτωση που εξαρτάται από περισσότερες η παλινδρόμηση είναι πολλαπλή.

Παρακάτω δίνονται οι τύποι της απλής και πολλαπλής παλινδρόμησης αντίστοιχα:

$$y_i = \alpha + \beta * x_i + \epsilon_i$$

$$y_i = \beta_0 + \beta_1 * x_{1i} + \beta_2 * x_{2i} + \dots + \beta_n * x_{ni} + \epsilon_i$$

Διαδική Λογιστική Παλινδρόμηση (Binary Logistic Regression)

Η συγκεκριμένη παλινδρόμηση αποτελεί μια στατιστική τεχνική, η οποία χρησιμοποιείται για την ανάλυση και μοντελοποίηση της σχέσης μεταξύ μιας εξαρτημένης δυαδικής μεταβλητής και μίας ή και περισσότερων ανεξαρτήτων μεταβλητών. Η εξαρτημένη μεταβλητή παίρνει δύο πιθανές τιμές, συνήθως 0 και 1, οι

οποίες αντιπροσωπεύουν δύο διαφορετικές καταστάσεις (πχ. επιτυχία, αποτυχία ή ναι, όχι κ.ά.).

Σε αντίθεση με την γραμμική παλινδρόμηση, όπου η εξαρτημένη μεταβλητή είναι συνεχής, στη λογιστική παλινδρόμηση η πρόβλεψη δεν είναι η άμεση τιμή της εξαρτημένης, αλλά η πιθανότητα εμφάνισης της. Ο στόχος επομένως είναι, να προβλεφθεί η πιθανότητα το Y να πάρει την τιμή 1, με βάση τις τιμές των ανεξαρτήτων μεταβλητών X .

Παρακάτω δίνονται οι εξισώσεις της συνάρτησης χρησιμότητας (U) πιθανότητας (P):

$$U_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e_i$$

β_0 : Σταθερός όρος, ο οποίος αντιπροσωπεύει την βασική χρησιμότητα.

$\beta_1, \beta_2, \dots, \beta_k$: Συντελεστές που εκφράζουν την βαρύτητα των ανεξάρτητων μεταβλητών στην χρησιμότητα.

$X_1, X_2, \dots, X_k$: Ανεξάρτητες μεταβλητές.

e_i : Στοχαστικός όρος που περιλαμβάνει μη παρατηρήσιμους παράγοντες.

$$P_i = \frac{e^{U_i}}{1 + e^{U_i}}$$

P_i : Η πιθανότητα i να επιλεγεί συμβεί μια συγκεκριμένη η υπόθεση.

U_i : Η χρησιμότητα που αντιστοιχεί στην επιλογή της υπόθεσης.

Naive Bayes Regression

Η Naive Bayes είναι μια μέθοδος ταξινόμησης που βασίζεται στο θεώρημα του Bayes και θεωρεί ότι όλα τα χαρακτηριστικά είναι ανεξάρτητα μεταξύ τους, κάτι που στην πράξη σπανίως ισχύει, αλλά αποδεικνύεται εξαιρετικά αποτελεσματικό σε διάφορες περιπτώσεις. Το συγκεκριμένο μοντέλο είναι γρήγορο και κατάλληλο για μεγάλα σύνολα δεδομένων.

Επιπλέον, το θεώρημα Bayes συνδυάζει την πιθανότητα για κάθε κατηγορία από το σύνολο ενός συμβάντος, με την πιθανότητα να παρατηρηθούν συγκεκριμένα χαρακτηριστικά ή αποδείξεις και δίνει τελικά την πιθανότητα για την κατηγορία του συμβάντος. Παρακάτω εκφράζεται το θεώρημα του Bayes:

$$P(C|X) = \frac{P(X|C) * P(C)}{P(X)}$$

$P(C|X)$: Είναι η πιθανότητα της κατηγορίας δεδομένων C των χαρακτηριστικών X .

$P(X|C)$: Είναι η πιθανότητα να εμφανιστούν τα χαρακτηριστικά X δεδομένης κατηγορίας C .

$P(C)$: Είναι η πιθανότητα κατηγορίας C .

$P(X)$: Είναι η συνολική πιθανότητα να εμφανιστούν τα χαρακτηριστικά C .

3.4 Αλγόριθμοι Μηχανικής Μάθησης

Τυχαία Δάση (Random Forest Regression)

Η παλινδρόμηση αυτή είναι ένας αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται τόσο για προβλήματα παλινδρόμησης, όσο και για προβλήματα ταξινόμησης. Η μέθοδος αυτή δημιουργεί πολλά δέντρα απόφασης από τυχαία υποσύνολα των

δεδομένων και των χαρακτηριστικών και συνδυάζει τις προβλέψεις τους για να βελτιώσει την ακρίβεια.

Στην περίπτωση της ταξινόμησης, η τελική πρόβλεψη βασίζεται στην πλειοψηφική ψήφο των δέντρων, ενώ στην παλινδρόμηση προκύπτει από το μέσο όρο των προβλέψεών τους. Αξίζει να σημειωθεί ότι η μέθοδος είναι αποτελεσματική για μεγάλα και σύνθετα δεδομένα.

Δέντρο Απόφασης (Decision Tree Regression)

Η μέθοδος αυτή χρησιμοποιείται όπως και η προηγούμενη για προβλήματα παλινδρόμησης και ταξινόμησης. Η βασική ιδέα πίσω από αυτήν είναι η διάσπαση των δεδομένων σε υποσύνολα με βάση τις τιμές των ανεξαρτήτων μεταβλητών, δημιουργώντας ένα δενδρικό διάγραμμα όπου κάθε κόμβος αντιπροσωπεύει μια απόφαση.

Οι αποφάσεις βασίζονται σε κανόνες που μεγιστοποιούν την διαφορά μεταξύ των τιμών της εξαρτημένης μεταβλητής. Μεταφορικά, κάθε κόμβος του δέντρου αντιστοιχεί σε μια ερώτηση που αφορά κάποιο χαρακτηριστικό των δεδομένων και κάθε κλαδί σε μία απάντηση που οδηγεί σε περαιτέρω διαίρεση ή σε κάποιο τελικό αποτέλεσμα. Στην ταξινόμηση, το δέντρο προσπαθεί να ομαδοποιήσει τα δεδομένα σε κατηγορίες, ενώ στην παλινδρόμηση, το δέντρο προσπαθεί να βρει συνεχείς τιμές. Τα δέντρα απόφασης έχουν το πλεονέκτημα της απλότητας και της ερμηνευσιμότητας.

Support Vector Machine Παλινδρόμηση (SVM Regression)

Η συγκεκριμένη μέθοδος μηχανικής μάθησης που χρησιμοποιείται για προβλήματα παλινδρόμησης, βασίζεται στην εύρεση μιας γραμμής ή επιφάνειας που να προβλέπει τις τιμές της εξαρτημένης μεταβλητής με ακρίβεια. Στόχος είναι να βρεθεί μια υπερεπίπεδη επιφάνεια που να χωρίζει τα δεδομένα με τέτοιο τρόπο ώστε οι περισσότερες προβλέψεις να βρίσκονται εντός ενός προκαθορισμένου περιθωρίου σφάλματος, γνωστό ως ε-περιθώριο (e-tube).

Επομένως, ο σκοπός δεν είναι απλώς να ελαχιστοποιηθεί το συνολικό σφάλμα, αλλά και να εξασφαλιστεί ότι οι προβλέψεις βρίσκονται όσο το δυνατόν πιο κοντά στη γραμμή που καθορίζει την σχέση των χαρακτηριστικών, με τις τιμές που αποκλίνουν εκτός του ε-περιθωρίου να επιβαρύνονται με την χρήση ενός ορίου απώλειας. Ακόμη, η SVM είναι αποτελεσματική σε πολύπλοκα δεδομένα, αφού μπορεί να χρησιμοποιήσει πυρήνες για να μοντελοποιήσει μη γραμμικές σχέσεις.

$$f(x) = w^T * x + b$$

3.5 Στατιστικοί έλεγχοι και κριτήρια αποδοχής μοντέλου

Μετά τη διαμόρφωση των μοντέλων απαιτείται μία διαδικασία αξιολόγησης προκειμένου να καθοριστεί ο βαθμός στον οποίο εξυπηρετεί την ανάλυση. Χρειάζεται συνεπώς να εκτιμηθεί η επιρροή των ανεξάρτητων μεταβλητών στην εξαρτημένη μεταξύ τους συσχέτιση, ώστε να μην χρησιμοποιηθούν μεταβλητές με μεγάλη συσχέτιση και επηρεαστεί το αποτέλεσμα του μοντέλου. Άρα οι ανεξάρτητες μεταβλητές θα πρέπει να είναι γραμμικώς ανεξάρτητες μεταξύ τους.

Παρακάτω παρουσιάζονται οι συνηθέστεροι έλεγχοι στο πλαίσιο ενός μοντέλου παλινδρόμησης για την αποδοχή των μοντέλων:

Συντελεστές παλινδρόμησης

Σχετικά με τους συντελεστές της εξίσωσης της εκάστοτε παλινδρόμησης, ως κριτήριο αποδοχής του μοντέλου απαιτείται η λογική ερμηνεία των πρόσημων τους, αν αυτή δεν υπάρχει τότε το μοντέλο πρέπει να απορριφθεί.

Συγκεκριμένα, γίνεται έλεγχος σε όλες τις ανεξάρτητες μεταβλητές. Θετικό πρόσημο του συντελεστή β συνεπάγεται με αύξηση της εξαρτημένης μεταβλητής σε κάθε αύξηση της ανεξάρτητης. Στην περίπτωση που το πρόσημο είναι αρνητικό συμβαίνει το αντίθετο, δηλαδή μείωση της εξαρτημένης μεταβλητής με αύξηση της ανεξάρτητης και άρα υπάρχει αρνητική συσχέτιση μεταξύ των μεταβλητών.

Συσχέτιση Παραμέτρων

Όταν ο συντελεστής συσχέτισης έχει τιμές κοντά στο 1, υπάρχει ισχυρή θετική συσχέτιση, ενώ όταν έχει τιμές κοντά στο -1, τότε υπάρχει ισχυρή αρνητική συσχέτιση. Στην περίπτωση που παίρνει τιμές κοντά στο 0, τότε υπάρχει γραμμική ανεξαρτησία μεταξύ των μεταβλητών.

Ελαστικότητα

Ως ελαστικότητα εκφράζεται ο δείκτης, οποίος υποδεικνύει την ευαισθησία της εξαρτημένης μεταβλητής στην μεταβολή μίας ή και περισσότερων ανεξαρτήτων μεταβλητών. Αποτελεί ένα αδιάστατο μέγεθος και δεν εξαρτάται από τις μονάδες μέτρησης των μεταβλητών. Επιπλέον, εκφράζεται ως ποσοστιαία μεταβολή της εξαρτημένης μεταβλητής που προκαλεί 1% μεταβολή της ανεξάρτητης.

Για γραμμικά μοντέλα και συνεχείς μεταβλητές, η ελαστικότητα προκύπτει από την παρακάτω σχέση:

$$e_i = \frac{\Delta Y_i}{\Delta X_i} * \frac{X_i}{Y_i} = \beta_i \frac{X_i}{Y_i}$$

Για γραμμικά μοντέλα και διακριτές μεταβλητές χρησιμοποιείται η έννοια της ψευδοελαστικότητας, η οποία περιγράφει την μεταβολή στην τιμή της πιθανότητας επιλογής κατά την μετάβαση από την μία τιμή της διακριτής μεταβλητής στην άλλη.

Στατιστική σημαντικότητα

Ένας ακόμη έλεγχος για την αξιολόγηση του προτύπου είναι ο έλεγχος t-test, με τον συντελεστή t, ο οποίος προσδιορίζει την στατιστική σημαντικότητα των ανεξάρτητων μεταβλητών για την επιλογή των κατάλληλων μεταβλητών που θα συμπεριληφθούν στο τελικό πρότυπο. Προϋπόθεση για την αποδοχή του μοντέλου είναι το επίπεδο εμπιστοσύνης να έχει υψηλή τιμή.

Ο συντελεστή t-ratio αναφέρεται σε καθεμιά από τις μεταβλητές ξεχωριστά και αποτελεί το αποτέλεσμα της διαίρεσης της εκτιμώμενης για τον συντελεστή τιμής προς την τυπική απόκλιση της.

Ο τύπος του t-test είναι ο εξής:

$$t_{stat} = \frac{\beta_i}{s.e}$$

,όπου s.e το τυπικό σφάλμα (standard error).

Από την παραπάνω σχέση προκύπτει ότι η μείωση στο τυπικό σφάλμα επιφέρει αύξηση του συντελεστή t και έτσι ευνοείται η επάρκεια (efficiency). Ακόμη, όσο μεγαλύτερη είναι η τιμή κατά απόλυτη τιμή, τόσο μεγαλύτερη θα είναι και η επιρροή της συγκεκριμένης μεταβλητής στο τελικό αποτέλεσμα. Αν η επιρροή αυτή είναι σημαντική, τότε η συγκεκριμένη μεταβλητή πρέπει να συμπεριληφθεί στην στατιστική ανάλυση, αν όχι, τότε θα πρέπει να αποκλειστεί.

Το εύρος τιμών του συντελεστή κυμαίνεται από $-\infty$ έως $+\infty$ και πιο συγκεκριμένα, παρακάτω παρατίθενται οι κρίσιμες τιμές του συντελεστή για ορισμένα επίπεδα εμπιστοσύνης, με συνηθέστερο εκείνο του 95% και τιμή t ίση με περίπου 1,7.

B.E	Επίπεδο Εμπιστοσύνης				
	0,900	0,950	0,975	0,990	0,995
80	1,2960	1,9710	2,0000	2,3900	2,6600
120	1,2890	1,6580	1,9800	2,3580	2,6170
∞	1,2820	1,6450	1,9600	2,3260	2,5760

Πίνακας 3.1: Κρίσιμες τιμές συντελεστών t_{stat}

Συντελεστής Προσαρμογής R^2

Ο δημοφιλέστερος στατιστικός δείκτης, μέσω του οποίου ελέγχεται η συνολική ποιότητα του μοντέλου, είναι ο συντελεστής προσαρμογής ή αλλιώς το κριτήριο καλής προσαρμογής R^2 .

Ο συντελεστής εκφράζει το ποσοστό της μεταβλητότητας μιας μεταβλητής από μία άλλη και λαμβάνει τιμές μεταξύ 0 και 1, με τιμές κοντά στο 0 να αντιστοιχούν σε ανεπαρκές μοντέλο, ενώ κοντά στο 1 να δείχνουν τέλεια προσαρμογή στα δεδομένα και ισχυρή σχέση μεταξύ των δύο μεταβλητών.

Συνήθως, η τιμή του δεν ξεπερνάει το 0,45 και επομένως θεωρείται πλήρως αποδεκτό καθώς στα πραγματικά προβλήματα τα δεδομένα έχουν σημαντικές διακυμάνσεις.

Ο συντελεστής ορίζεται από την σχέση:

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{\bar{Y}})^2}{\sum_{i=1}^n (Y_i - \bar{\bar{Y}})^2} = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{\bar{Y}})^2}$$

Τιμή p-value

Ο δείκτης αυτός αποτελεί μια στατιστική τιμή, η οποία προκύπτει στα αποτελέσματα κάθε μοντέλου για κάθε ανεξάρτητη μεταβλητή και δείχνει την πιθανότητα να προκύψει ένα αποτέλεσμα ίσο ή μεγαλύτερο από το παρατηρούμενο, αν η μηδενική υπόθεση είναι αληθινή.

Επιθυμητό είναι η τιμή να είναι μικρότερη από 0.05, καθώς μικρότερες τιμές υποδηλώνουν ισχυρότερη απόδειξη υπέρ της εναλλακτικής υπόθεσης. Επίσης, αν η τιμή του είναι μεγαλύτερη από 0.10 τότε το αποτέλεσμα πιθανόν δεν είναι στατιστικά διαφορετικό από το μηδέν. Ακόμη, μια μεγάλη διαφορά μεταξύ των δύο τιμών των μεταβλητών, αντιστοιχεί σε χαμηλότερη τιμή p .

Στατιστικό σφάλμα

Το σφάλμα περιγράφει τη διαφορά μεταξύ μια τιμής που λαμβάνεται από μια διαδικασία συλλογής δεδομένων και της πραγματικής τιμής για τον πληθυσμό και χωρίζονται σε σφάλματα δειγματοληψίας και μη δειγματοληπτικά σφάλματα.

Όσο μεγαλύτερο είναι το σφάλμα, τόσο λιγότερο αντιπροσωπευτικά είναι τα δεδομένα του πληθυσμού. Το πλήθος των στατιστικών σφαλμάτων εξαρτάται από το μέγεθος του δείγματος που επιλέχθηκε για την μελέτη.

Ορθότητα (Accuracy)

Ορθότητα είναι η απόσταση μίας μέτρησης από την πραγματικής της τιμή, δηλαδή αναφέρεται στην εγγύτητα των υπολογισμών ή των εκτιμήσεων στις πραγματικές τιμές που προορίζονταν να μετρήσουν τα στατιστικά στοιχεία.

Αποτελεί σημαντική παράμετρο για τη στατιστική ανάλυση, καθώς βοηθά στην λήψη συμπερασμάτων από τα δεδομένα. Επομένως, απαιτούνται υψηλά επίπεδα της συγκεκριμένης παραμέτρου για τον ορθό προσδιορισμό της αξίας ενός πληθυσμού.

Υπάρχουν αρκετοί τρόποι μέτρησης της, ένας από τους συνηθέστερους είναι ο υπολογισμός ενός ποσοστού σφάλματος. Υψηλό ποσοστό υποδεικνύει ότι η τιμή απέχει πολύ από την πραγματική και επομένως είναι ανακριβής.

Ευαισθησία (Sensitivity)

Ως ευαισθησία ορίζεται το ποσοστό των αληθώς θετικών αποτελεσμάτων (true positives), δηλαδή η ικανότητα ενός μοντέλου να ανιχνεύει σωστά τα θετικά δείγματα. Για παράδειγμα, ένα μοντέλο με υψηλή ευαισθησία ανιχνεύει όλα τα θετικά δείγματα, ενώ ένα με χαμηλή ευαισθησία ανιχνεύει μόνο μέρος αυτών, αυξάνοντας τα ψευδώς αρνητικά αποτελέσματα (False Negatives), τα οποία συμβαίνουν όταν αποτυγχάνουμε να απορρίψουμε μία ψευδή αρνητική υπόθεση (σφάλματα τύπου II).

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

Ειδικότητα (Specificity)

Ως ειδικότητα ορίζεται ως η πιθανότητα αρνητικού αποτελέσματος με την προϋπόθεση ότι το δείγμα είναι πράγματι αρνητικό (True Negatives), δηλαδή η ικανότητα ενός μοντέλου να απορρίπτει σωστά τα ψευδώς θετικά αποτελέσματα. Τα ψευδώς θετικά αποτελέσματα (False Positives) συμβαίνουν, όταν απορρίπτεται μία αληθής μηδενική υπόθεση, δηλαδή όταν θεωρούμε ότι η υπόθεση ισχύει ενώ δεν ισχύει (σφάλματα τύπου I).

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives} + \text{False positives}}$$

4. ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ

4.1 Εισαγωγή

Στο συγκεκριμένο κεφάλαιο, περιγράφεται η **διαδικασία συλλογής και επεξεργασίας των απαραίτητων στοιχείων**, με σκοπό την τελική παρουσίαση μιας επαρκούς και ολοκληρωμένης εικόνας για την ποιότητα και την αξιοπιστία των δεδομένων που χρησιμοποιήθηκαν. Μετά την ολοκλήρωση της βιβλιογραφικής ανασκόπησης αλλά και του θεωρητικού υπόβαθρου που οδήγησε στην επιλογή των κατάλληλων μεθόδων στατιστικής ανάλυσης και μηχανικής μάθησης, ορίστηκαν οι διάφορες τεχνικές μοντελοποίησης που θα χρησιμοποιηθούν θα συγκριθούν ώστε να επιλεγεί τελικά η καλύτερη δυνατή.

Ειδικότερα, το παρόν κεφάλαιο χωρίζεται σε δύο υποενότητες. Στην πρώτη, η οποία αφορά τη συλλογή των στοιχείων, αναλύεται το πείραμα που πραγματοποιήθηκε σε πραγματικές συνθήκες οδήγησης με πηγή δεδομένων σύγχρονες κινητές συσκευές. Στη δεύτερη, επεξεργάζονται τα στοιχεία του πειράματος και παρουσιάζονται σε συγκεντρωτικούς πίνακες και σε γραφήματα. Τα αποτελέσματα της επεξεργασίας προκύπτουν με την βοήθεια του λογισμικού στατιστικής ανάλυσης R, ακολουθούμενα με τον απαραίτητο σχολιασμό. Τέλος, μέσω της διαδικασίας που θα αναλυθεί παρακάτω γίνεται επίτευξη του στόχου της Διπλωματικής Εργασίας.

4.2 Συλλογή δεδομένων

Η καταγραφή και ανάλυση της οδηγικής συμπεριφοράς υπό πραγματικές συνθήκες αποτελεί κρίσιμο βήμα για την κατανόηση και βελτίωση της οδικής ασφάλειας. Τα δεδομένα που συλλέγονται μέσω προηγμένων τεχνολογιών, όπως οι εφαρμογές της OSeven Telematics, προσφέρουν πολύτιμες πληροφορίες για την αξιολόγηση της συμπεριφοράς των οδηγών, την ανάπτυξη προγνωστικών μοντέλων και την αναγνώριση επικίνδυνων περιοχών. Η αξιοποίηση αυτών των δεδομένων συμβάλλει στη μείωση των οδικών ατυχημάτων και στη διαμόρφωση πιο ασφαλών και βιώσιμων οδικών συστημάτων.

4.2.1 Δεδομένα συμπεριφοράς οδηγού υπό πραγματικές συνθήκες

Τα δεδομένα συμπεριφοράς οδηγού υπό πραγματικές συνθήκες βασίστηκαν **στην εφαρμογή της εταιρείας OSeven Telematics**. Η OSeven Telematics είναι μια εταιρεία τηλεματικής, η οποία ανέπτυξε μια εφαρμογή συμβατή με διάφορες κατηγορίες έξυπνων κινητών, με σκοπό να συλλέγει δεδομένα που αφορούν τον τρόπο οδήγησης και την συμπεριφορά των οδηγών. Με τα δεδομένα αυτά δημιουργεί εξατομικευμένα προφίλ συμπεριφορών των οδηγών μέσω ειδικών αλγορίθμων. Στόχος της είναι να ενημερώνει τους οδηγούς για τα σημεία όπου μπορούν να βελτιώσουν την οδήγηση τους, προάγοντας την βιώσιμη οδήγηση και μειώνοντας τον αριθμό των ατυχημάτων.

Τα δεδομένα **συλλέγονται μέσω εφαρμογών σε έξυπνες συσκευές**. Η καταγραφή τους ξεκινάει αυτόματα όταν αναγνωρίζεται ένας από τους αισθητήρες των κινητών τηλεφώνων σε μια κατάσταση οδήγησης και σταματάει όταν αναγνωρίζεται μια κατάσταση μη οδήγησης. Η τελευταία συνθήκη αναγνωρίζεται από διάστημα αδράνειας περισσότερων από πέντε λεπτών, καθώς θεωρείται ότι ο οδηγός έκανε μια ολιγόλεπτη στάση και μετά συνέχισε την διαδρομή του. Στην περίπτωση όπου η παραπάνω θεώρηση δεν ισχύει απορρίπτεται από τις μεθόδους τεχνικής μάθησης.

Τα καταγεγραμμένα δεδομένα προέρχονται από τους διάφορους αισθητήρες των έξυπνων κινητών τηλεφώνων (smartphone) και τους αλγορίθμους συγχώνευσης (fusion algorithm), οι οποίοι παρέχονται από το Android (Google) και το iOS (Apple). Η εφαρμογή που έχει αναπτυχθεί, **καταγράφει τη συμπεριφορά του χρήστη** χρησιμοποιώντας τους αισθητήρες επιταχυνσιόμετρου (accelerometer), γυροσκόπιο (gyroscope), μαγνητόμετρο (magnetometer), και GPS (ταχύτητα, πορεία, γεωγραφικό μήκος και πλάτος). Επιπλέον, από το iOS και το Android παρέχεται ο ρυθμός εκτροπής, ο βαθμός έντασης και κύλισης, η γραμμική επιτάχυνση και η επιτάχυνση της βαρύτητας. Για να διαβαστούν τα δεδομένα που κατέγραψαν οι αισθητήρες και να αποθηκευτούν προσωρινά στη συσκευή, χρησιμοποιείται μια ποικιλία APIs (Application Programming Interface). Στη συνέχεια τα στοιχεία αυτά στέλνονται μέσω σύνδεσης ίντερνετ στην κεντρική βάση δεδομένων της OSeven και ταυτόχρονα διαγράφονται από την μνήμη του κινητού τηλεφώνου, ώστε να είναι έτοιμο να καταγράψει στην επόμενη διαδρομή. Στην παρακάτω εικόνα παρουσιάζεται το διάγραμμα ροής των δεδομένων:



Εικόνα 4.1: Διάγραμμα Ροής Δεδομένων OSeven Telematics

Όλα τα δεδομένα αποθηκεύονται χρησιμοποιώντας προηγμένες τεχνικές κρυπτογράφησης και προστασίας δεδομένων, σύμφωνα με τους εθνικούς νόμους και τις οδηγίες της Ευρωπαϊκής Ένωσης (ΕΕ), για την διαφύλαξη των προσωπικών δεδομένων (GDPR) και τη διασφάλιση της ανωνυμίας. Οι εφαρμογές που χρησιμοποιούνται υποστηρίζουν τον έλεγχο της ταυτότητας του χρήστη και την κρυπτογράφηση, για να αποτρέπεται η πρόσβαση σε μη εξουσιοδοτημένα δεδομένα από τρίτους.

Μετά την αποθήκευση των στοιχείων, γίνεται η κεντρική επεξεργασία και η διαλογή των δεδομένων, ώστε να παραμείνουν μόνο εκείνα που αποτελούν σημαντικές πληροφορίες για την οδική ασφάλεια και την οικολογική οδήγηση. Αυτό υλοποιείται με τη χρήση αλγορίθμων μηχανικής μάθησης (Machine Learning, ML). Η επεξεργασία των δεδομένων περνά από διάφορα στάδια. Αρχικά, γίνεται το φιλτράρισμα των δεδομένων και η εξομάλυνση τους, στη συνέχεια εντοπίζονται τα συμβάντα έντονων συμπεριφορών, όπως οι απότομες επιταχύνσεις, τα φρεναρίσματα, οι στροφές, οι περιοχές υπερβολικής ταχύτητας αλλά και οι επικίνδυνες ενέργειες κατά την οδήγηση, όπως η χρήση κινητού τηλεφώνου. Τέλος, σημειώνεται ο τρόπος μετακίνησης και ο καθορισμός των επιβατών του κάθε οχήματος.

Μετά την πολυεπίπεδη διαδικασία ανάλυσης των παραπάνω δεδομένων στον αλγόριθμο μηχανικής μάθησης, τα ακατέργαστα δεδομένα μετατρέπονται σε πληροφορίες υψηλής ακρίβειας για την οδηγική συμπεριφορά. **Παράγονται δείκτες έκθεσης κινδύνου**, όπως το ποσοστό οδήγησης σε ώρες αιχμής, η συνολική διάρκεια οδήγησης, οι καιρικές συνθήκες, η κατανάλωση καυσίμου, **αλλά και δείκτες οδηγική συμπεριφοράς**, όπως ο αριθμός των σοβαρών συμβάντων, τα ποσοστά των απότομων φρεναρισμάτων, επιταχύνσεων, στροφών και γενικότερα η επιθετική οδηγική συμπεριφορά κ.ά. Επίσης, τα δεδομένα μέσω ενός συστήματος, βαθμονομούν τον κάθε χρήστη είτε για κάθε του ταξίδι, είτε για το σύνολο αυτών και αξιολογείται η συμπεριφορά του με βάση τους παραπάνω και άλλους δείκτες.

Παράλληλα, ο χρήστης με το τέλος του ταξιδιού του μπορεί να δει τη διαδρομή που πραγματοποίησε στο χάρτη, αλλά και σε ποια οδικά τμήματα εμφάνισε επιθετική συμπεριφορά ή έκανε σφάλματα, ώστε να αποφευχθούν παρόμοιες ενέργειες στο μέλλον.

4.2.2 Διαδικασία συλλογής στοιχείων

Στο πλαίσιο της Διπλωματικής αυτής Εργασίας, τα δεδομένα συμπεριφοράς οδηγού συλλέχθηκαν με την υποστήριξη της εταιρίας OSeven Telematics, η οποία μου παρείχε στοιχεία που καλύπτουν την περιοχή της Αθήνας.

Για το εξεταζόμενο οδικό δίκτυο **πραγματοποιήθηκε μια αρχική ανάλυση, όλων των τμημάτων που προήλθαν από το OpenStreetMap (OSM)**, προκειμένου να εξαχθούν τα γεωμετρικά χαρακτηριστικά. Το OSM είναι μια ελεύθερη και ανοιχτή πλατφόρμα χαρτών, η οποία επιτρέπει στους ίδιους τους χρήστες να δημιουργούν και να επεξεργάζονται γεωγραφικά και χωρικά δεδομένα. Οι χρήστες του OSM μπορούν να προσθέτουν πληροφορίες όπως δρόμους, κτίρια, γεωγραφικά χαρακτηριστικά και άλλα, οι οποίες είναι διαθέσιμες για δημόσια χρήση, δημιουργώντας έτσι έναν ελεύθερο και ακριβή χάρτη του κόσμου. Ακόμα, μέσω της εφαρμογής για έξυπνα κινητά τηλέφωνα, συλλέχθηκαν δεδομένα συμπεριφοράς των οδηγών και συσχετίστηκαν με τα αντίστοιχα οδικά τμήματα του OSM.

Τα δεδομένα αυτά περιλάμβαναν **35.282 οδικά τμήματα** και περιείχαν σημαντικές μεταβλητές, καθώς και δείκτες οδικής ασφάλειας και συμπεριφοράς. Παρακάτω θα αναλυθούν ανά κατηγορίες κάποια από τα στοιχεία που μου δόθηκαν.

Στις **βασικές μεταβλητές** συγκαταλέγονται οι:

- αναγνωριστικό (osm_id)
- ονοματολογία (name)
- γεωγραφικό κέντρο (centroidlon, centroidlat)
- μήκος οδικού τμήματος (length)
- τύπος οδού (highway)
- αριθμός λωρίδων (lanes)
- επιφάνεια (surface)
- όριο ταχύτητας (maxspeed)
- ευθυγραμμία (efficiency)

Δίνονται επίσης, διάφοροι **δείκτες μορφολογικών χαρακτηριστικών** της διαδρομής όπως:

- απότομη αλλαγή κατεύθυνσης (tot_angle)
- ρυθμός γωνίας (angle_rate)
- κλίση οδού (slope)
- κλάση κλίσης (slope_class)
- αριθμός διαδρομών (trip_count)

Δείκτες ασφάλειας και χρήσης προστατευτικών μέτρων:

- ποσοστό χρήσης ζώνης ασφαλείας (PC_D_Seatbelt_Yes%)
- ποσοστό χρήσης κράνους πρόσκρουσης (PTW_D_Helmet_Yes%)

Δείκτες οδηγικής συμπεριφοράς:

- ποσοστό οδηγών που δεν χρησιμοποιούν το κινητό τηλέφωνο κατά την οδήγηση (PC_D_Phone_No%)
- ποσοστό οδηγών που δεν υπερβαίνουν το όριο ταχύτητας (PC_Speeding_No%)
- συνολικός αριθμός περιστατικών υπερβάσεων ταχύτητας (speeding_count)
- συνολικός αριθμός περιστατικών χρήσης κινητού (mobile_usage_count)
- συνολικός αριθμός απότομων επιταχύνσεων (harsh_acc_count)
- συνολικός αριθμός απότομων επιβραδύνσεων (harsh_braking_count)

Δείκτες ατυχημάτων και σοβαρότητας για τα έτη 2016-2020:

- συνολικός αριθμός ατυχημάτων στον Δήμο που ανήκει η οδός (Crashes2016-2020)
- συνολικός αριθμός νεκρών σε οδικά ατυχήματα στον Δήμο που ανήκει η οδός (Fat2016-2020)
- συνολικός αριθμός σοβαρών τραυματισμών ή θανάτων στον Δήμο που ανήκει η οδός (KSI2016-2020)

Τέλος, δίνονται ορισμένοι δείκτες περιβαλλοντικού αποτυπώματος:

- συνολική εκπομπή διοξειδίου του άνθρακα (CO2_count)
- συνολική εκπομπή μονοξειδίου του άνθρακα (CO_count)
- συνολική εκπομπή υδρογονανθράκων (HC_count)
- συνολική εκπομπή οξειδίων αζώτου (NOx_count)

Ο στόχος της συλλογής των δεδομένων είναι η ανάπτυξη ενός μοντέλου πρόβλεψης της υπέρβασης των ορίων ταχύτητας οδικό δίκτυο της Αθήνας, το οποίο θα μπορεί να εφαρμοστεί αντίστοιχα σε οποιαδήποτε άλλη περιοχή, έτσι ώστε να κατανοηθούν και αξιολογηθούν οι παράγοντες την επηρεάζουν και κατ' επέκταση μπορούν να συμβάλλουν στα ατυχήματα. **Τα δεδομένα αυτά θα αξιοποιηθούν με σκοπό την δημιουργία χαρτών** που αναδεικνύουν τις περιοχές υψηλού κινδύνου. Ένα παράδειγμα εφαρμογής της προσέγγισης αυτής είναι στην ειδική ιστοσελίδα **Safe Roads Maps** (SmartMaps), το οποίο προσφέρει χάρτες επικινδυνότητας για διαφορετικές περιοχές, με στόχο την βελτίωση της οδικής ασφάλειας.

4.3 Επεξεργασία δεδομένων

Στη συγκεκριμένη ενότητα περιγράφεται η διαδικασία επεξεργασίας των στοιχείων που συλλέχθηκαν νωρίτερα και αναλύεται η προκαταρκτική ανάλυση που έγινε σε περιβάλλον Excel. Έπειτα, παρουσιάζονται ορισμένα διαγράμματα που δημιουργήθηκαν με σκοπό την καλύτερη δυνατή κατανόηση και οπτικοποίηση των δεδομένων.

4.3.1 Διαμόρφωση δεδομένων

Στα δεδομένα που συλλέχθηκαν και αναφέρθηκαν αναλυτικά στην προηγούμενη ενότητα, **έγινε αρχικά η επιλογή των δεδομένων που ήταν κρίσιμα** για την ανάλυση και την διαμόρφωση του τελικού συνόλου. Είναι σημαντικό να σημειωθεί ότι πέρα από τις μεταβλητές που παρουσιάστηκαν νωρίτερα, υπήρχε ένα ευρύτερο σύνολο μεταβλητών που περιλάμβανε επιπλέον στοιχεία. Ωστόσο, ορισμένες από τις μεταβλητές αυτές, δεν διέθεταν επαρκή δεδομένα για το σύνολο των οδών, είτε δεν ήταν σχετικές με τους στόχους της ανάλυσης, επομένως και απορρίφθηκαν.

Αυτή η διαδικασία διαμόρφωσης εξασφαλίζει ότι το τελικό σύνολο δεδομένων είναι περιεκτικό και αξιόπιστο για την επερχόμενη στατιστική ανάλυση. Συγκεκριμένα οι μεταβλητές που επιλέχθηκαν τελικά ήταν οι εξής:

- **osm_id, highway, length** και **centroidlon, centroidlat** για τις γεωγραφικές πληροφορίες των οδών,
- **tot_angle, efficiency, angle_rate, slope** και **slope_class** για τα χαρακτηριστικά των οδών,
- **PC_D_Seatbelt_Yes%, PTW_D_Helmet_Yes%, PC_D_Phone_No%, PC_Speeding_No%** για την συμπεριφορά των οδηγών μέσω παρατηρήσεων πεδίου,
- **Crashes2016-2020, Fat2016-2020, KSI2016-2020** για τα στατιστικά των οδικών ατυχημάτων,
- **Speeding_count, speeding_di, mobile_usage_count, harsh_acc_count** και **harsh_braking_count** για πληροφορίες επικίνδυνης συμπεριφοράς οδηγού υπό πραγματικές συνθήκες,
- **CO2_count, CO_count, HC_count, NOx_count** για περιβαλλοντικούς δείκτες,
- **trip_count** για τον αριθμό των διαδρομών στα εξεταζόμενα οδικά τμήματα.

Σε ένα φύλλο Excel δημιουργήθηκε ένας **ενοποιημένος πίνακας** που περιλάμβανε όλες τις παραπάνω μεταβλητές στον οριζόντιο άξονα και τις 35.282 εγγραφές στον κατακόρυφο άξονα. Ενδεικτικό τμήμα του ενοποιημένου πίνακα φαίνεται στην παρακάτω εικόνα:

osm_id	highway	length	Centroidlon	Centroidlat	tot_angle	efficiency	angle_rate	slope	slope_class	PC_D_Seatbelt_Yes%
1001813325	residential	59,5	23,72779635	38,0415106	0,0	1,00	0,0	0,0	0-3: flat	0,9
100183614	residential	332,5	23,67893245	37,99947103	176,2	0,96	0,6	1,7	0-3: flat	0,9
1002987583	residential	48,5	23,7420903	38,00663970	0,0	1,00	0,0	4,3	3-5: mild	0,8
1003002285	residential	82,7	23,74764336	38,00699376	19,7	0,99	0,2	4,2	3-5: mild	0,8
1003003221	residential	45,1	23,7468365	38,00783565	0,0	1,00	0,0	8,2	8-10: hard	0,8
100301476	residential	94,4	23,67993505	37,99329615	0,0	1,00	0,0	1,3	0-3: flat	0,9
100301477	residential	105,1	23,68783822	37,99218388	28,6	0,99	0,3	1,0	0-3: flat	0,9
100301478	residential	10,3	23,6818543	37,99697445	0,0	1,00	0,0	1,6	0-3: flat	0,9
1003015926	residential	19,5	23,7307152	38,03923735	0,0	1,00	0,0	1,4	0-3: flat	0,9
1003026378	residential	18,0	23,73166935	38,04278190	0,0	1,00	0,0	1,1	0-3: flat	0,9
1003035597	residential	91,0	23,73161209	38,04471775	3,4	1,00	0,0	0,3	0-3: flat	0,9
1003035598	residential	48,3	23,73092085	38,04504005	0,0	1,00	0,0	0,4	0-3: flat	0,9
1003043415	living_street	54,0	23,72959782	38,04882566	0,1	1,00	0,0	0,6	0-3: flat	0,9
1003079295	residential	92,6	23,72739818	38,06108244	130,4	0,90	1,6	1,6	0-3: flat	0,9
100322953	residential	190,8	23,67953759	37,98306438	129,2	0,97	0,7	0,2	0-3: flat	0,9
100327809	tertiary	96,7	23,6824339	37,99181800	0,0	1,00	0,0	1,1	0-3: flat	0,9
100327813	residential	201,2	23,68387609	37,9899694	2,6	1,00	0,0	1,1	0-3: flat	0,9

Πίνακας 4.1: Ενοποιημένος πίνακας τελικών μεταβλητών

Έπειτα, για να πραγματοποιηθεί μια **αρχική ανάλυση** των δεδομένων, εφαρμόστηκαν διάφορες ενέργειες στις μεταβλητές **με κριτήριο την κατηγορία των οδών (highway) και την κατηγορία κλίσης κάθε οδού (slope_class)**. Πιο αναλυτικά, οι τύποι οδών και οι κατηγορίες κλίσης κατανεμήθηκαν σε διακριτές ομάδες, ώστε να αναλυθεί η κατανομή και η συχνότητα κάθε τύπου. Με αυτόν τον τρόπο, γίνεται εφικτός ο εντοπισμός των κυρίαρχων τύπων οδών και των επικρατέστερων κατηγοριών κλίσης στο σύνολο των δεδομένων.

Για κάθε τύπο οδού, όπως **living_street** (οδός ήπιας κυκλοφορίας), **motorway** (αυτοκινητόδρομος), **primary** (κύρια οδός), **residential** (οδός κατοικιών), **secondary** (δευτερεύουσα οδός), **tertiary** (τριτεύουσα οδός), **trunk** (κεντρική αρτηρία) και **link**

(σύνδεσμος οδού), καταγράφηκε τόσο ο αριθμός των εγγραφών όσο και το ποσοστό συμμετοχής επί του συνόλου των δεδομένων. Αυτή η ανάλυση έδειξε ότι **οι οδοί κατοικιών (residential) αποτελούν την κυρίαρχη κατηγορία, καλύπτοντας το 66,4% των δεδομένων**, ενώ οι υπόλοιποι τύποι έχουν εμφανώς μικρότερα ποσοστά.

Row Labels	Count of highway	Row Labels	Count of highway
living_street	776	living_street	2,2%
motorway	524	motorway	1,5%
motorway_link	547	motorway_link	1,6%
primary	1625	primary	4,6%
primary_link	229	primary_link	0,6%
residential	23409	residential	66,4%
secondary	3081	secondary	8,7%
secondary_link	287	secondary_link	0,8%
tertiary	4615	tertiary	13,1%
tertiary_link	108	tertiary_link	0,3%
trunk	66	trunk	0,2%
trunk_link	14	trunk_link	0,0%
Grand Total	35281	Grand Total	100,0%

Πίνακας 4.2: Πίνακας κατανομής τύπου οδού

Αντίστοιχα, για τις παρακάτω κατηγορίες κλίσης ακολουθήθηκε η ίδια διαδικασία.

- 0-3: επίπεδη (**flat**) κλίση από 0 έως 3%, σχεδόν επίπεδη με εύκολη κίνηση,
- 3-5: ήπια (**mild**) κλίση από 3 έως 5%, ελαφρώς αισθητή αλλά όχι δύσκολη για την πλειονότητα των οχημάτων,
- 5-8: μέτρια (**medium**) κλίση από 5 έως 8%, αισθητή και με απαίτηση προσπάθειας για συγκεκριμένα οχήματα και ποδήλατα,
- 8-10: δύσκολη (**hard**) κλίση από 8 έως 10%, αρκετά απότομη και με δυσκολίες κίνησης σε συνθήκες κακής πρόσφυσης ή και μεγάλων οχημάτων,
- >10: ακραία (**extreme**) κλίση μεγαλύτερη από 10%, πολύ απότομη για όλες τις κατηγορίες.

Row Labels	Count of slope_class	Row Labels	Count of slope_class
>10: extreme	386	>10: extreme	1,1%
0-3: flat	26946	0-3: flat	76,4%
3-5: mild	5408	3-5: mild	15,3%
5-8: medium	2112	5-8: medium	6,0%
8-10: hard	429	8-10: hard	1,2%
Grand Total	35281	Grand Total	100,0%

Πίνακας 4.3: Πίνακας κατανομής κλίσης οδού

Δηλαδή, αναλύθηκε η συχνότητα και το ποσοστό τους στο σύνολο των δεδομένων και βρέθηκε ότι **οι περισσότερες οδοί έχουν ελάχιστη κλίση (flat) με ποσοστό 76,4%**, ενώ οι υπόλοιπες κατηγορίες έχουν σημαντικά μικρότερες συμμετοχές.

Με αυτές τις αρχικές ενέργειες, επιτεύχθηκε μια σαφής κατανόηση της κατανομής διάφορων κατηγοριών και θα μελετηθούν στη συνέχεια πιθανές συσχετίσεις μεταξύ των χαρακτηριστικών αυτών και της επικινδυνότητας του οδικού δικτύου ή/και της συμπεριφοράς των οδηγών.

Στη συνέχεια, πραγματοποιήθηκε μια **περιγραφική στατιστική ανάλυση των βασικών μεταβλητών** και υπολογίστηκαν οι τέσσερις βασικοί στατιστικοί δείκτες, ελάχιστη τιμή (**min**), μέση τιμή (**mean**), διάμεση τιμή (**median**) και μέγιστη τιμή (**max**), ώστε να αποκτηθεί μια σαφής εικόνα για τα χαρακτηριστικά κάθε μεταβλητής και για το εύρος της. Η ελάχιστη τιμή αντιπροσωπεύει την κατώτατη τιμή που παρατηρήθηκε και η μέγιστη το ανώτατο όριο των τιμών κάθε μεταβλητής, επισημαίνοντας έτσι τα ακραία στοιχεία που απαιτούν διερεύνηση. Η μέση τιμή αποτελεί τον μέσο όρο και προσδιορίζει το γενικό μέσο επίπεδο κάθε μεταβλητής, ενώ η διάμεση τιμή είναι η μεσαία τιμή, όταν οι τιμές ταξινομηθούν από την μικρότερη στην μεγαλύτερη και αποτελεί το κεντρικό σημείο των τιμών.

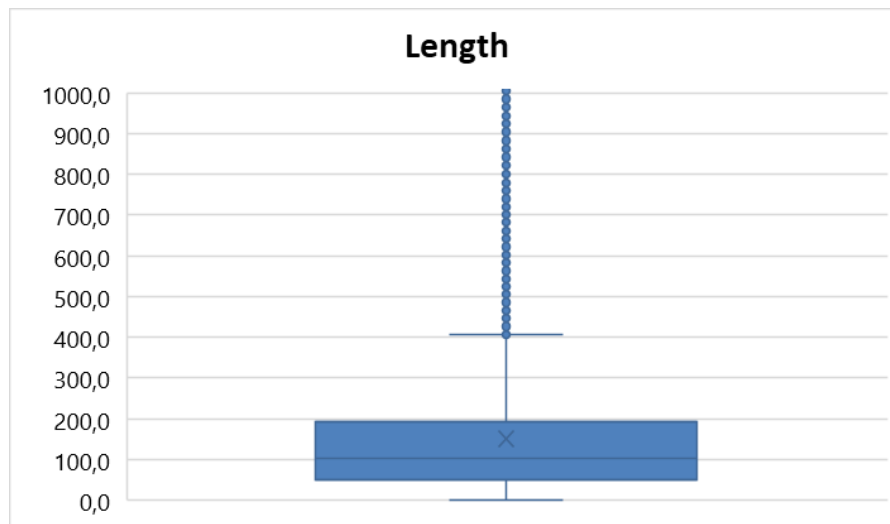
Variable	Min	Mean	Median	Max
length	0,1	151,1	102,3	8468,6
efficiency	0,1	1,0	1,0	1,0
PC_D_Seatbelt_Yes%	0,7	0,8	0,9	1,0
PTW_D_Helmet_Yes%	0,7	0,8	0,8	1,0
PC_D_Phone_No%	0,7	0,9	0,9	1,0
PC_Speeding_No%	0,0	0,3	0,3	0,8
Crashes2016-2020	3,3	629,7	463,5	6317,7
Fat2016-2020	0,5	12,9	11,7	69,8
speeding_count	0,0	15,6	0,0	6143,0
mobile_usage_count	0,0	18,9	2,0	1481,0
harsh_acc_count	0,0	0,5	0,0	59,0
harsh_braking_count	0,0	0,7	0,0	59,0
trip_count	0,0	126,1	27,0	3292,0

Πίνακας 4.4: Περιγραφικά Στατιστικά των Μεταβλητών

Η ανάλυση αυτή εφαρμόστηκε στις μεταβλητές μήκος οδικού τμήματος (**length**), ευθυγράμμια (**efficiency**), ποσοστό χρήσης ζώνης ασφαλείας (**PC_D_Seatbelt_Yes%**), ποσοστό χρήσης κράνους πρόσκρουσης (**PTW_D_Helmet_Yes%**), ποσοστό που δεν χρησιμοποιούσε κινητό τηλέφωνο (**PC_D_Phone_No%**), ποσοστό οδηγών που δεν υπερβαίνουν τα όρια ταχύτητας (**PC_Speeding_No%**), συνολικός αριθμός ατυχημάτων στον Δήμο που ανήκει η οδός (**Crashes2016-2020**), συνολικός αριθμός νεκρών σε οδικά ατυχήματα στον Δήμο που ανήκει η οδός (**Fat2016-2020**), συνολικός αριθμός περιστατικών υπερβάσεων ταχύτητας (**speeding_count**), συνολικός αριθμός περιστατικών χρήσης κινητού τηλεφώνου (**mobile_usage_count**), συνολικός αριθμός απότομων επιταχύνσεων (**harsh_acc_count**), συνολικός αριθμός περιστατικών επιβράδυνσης (**harsh_braking_count**) και αριθμός διαδρομών (**trip_count**) και φαίνεται στον πίνακα παρακάτω:

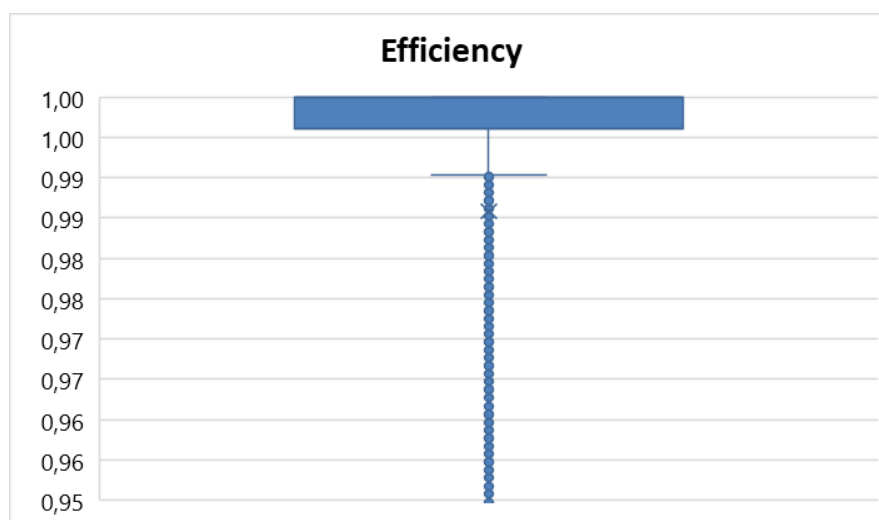
Το εύρος από την ελάχιστη έως την μέγιστη τιμή δείχνει την διακύμανση των τιμών στα δεδομένα και η απόκλιση της διάμεσης από την μέση τιμή υποδεικνύει την ύπαρξη ακραίων τιμών που επηρεάζουν τον μέσο όρο.

Στο πλαίσιο της ανάλυσης, μετά τον υπολογισμό των βασικών στατιστικών δεικτών, **έγινε η ανάλυση σε boxplots διαγράμματα** για τις προαναφερόμενες μεταβλητές. Τα boxplots βοηθούν στην οπτική αναπαράσταση της κατανομής κάθε μεταβλητής και παρέχουν πληροφορίες σχετικά με τα κεντρικά σημεία, την διασπορά και τα ακραία δεδομένα (outliers) που ενδεχομένως να απαιτούν περεταίρω διερεύνηση. Τα αποτελέσματα αυτά θα συμπίπτουν με εκείνα των βασικών δεικτών που εντοπίστηκαν νωρίτερα.



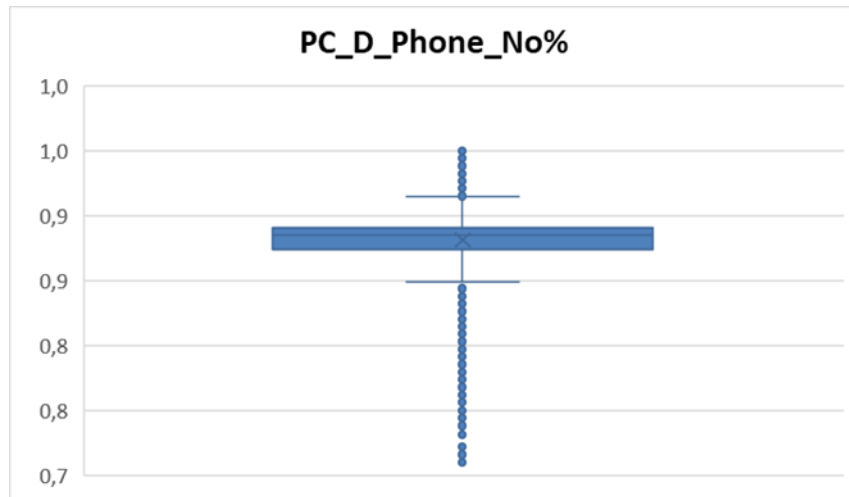
Διάγραμμα 4.1: Κατανομή Μήκους Οδικού Τμήματος

Το διάγραμμα δείχνει την κατανομή των τιμών του μήκους των εξεταζόμενων οδικών τμημάτων. Η μέση τιμή είναι συγκεντρωμένη γύρω από τα 150 μ., ενώ παρατηρούνται ακραίες τιμές που ξεπερνούν τα 900 μ.

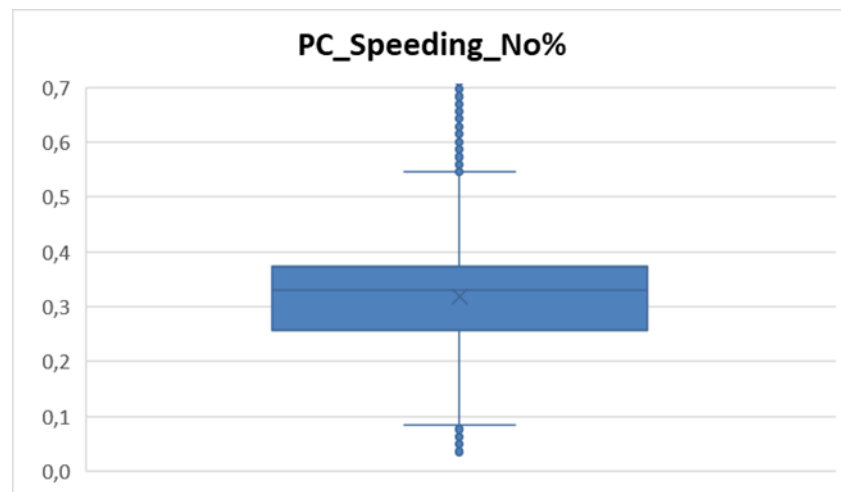


Διάγραμμα 4.2: Κατανομή Δείκτη Ευθυγραμμίας Οδικού Τμήματος

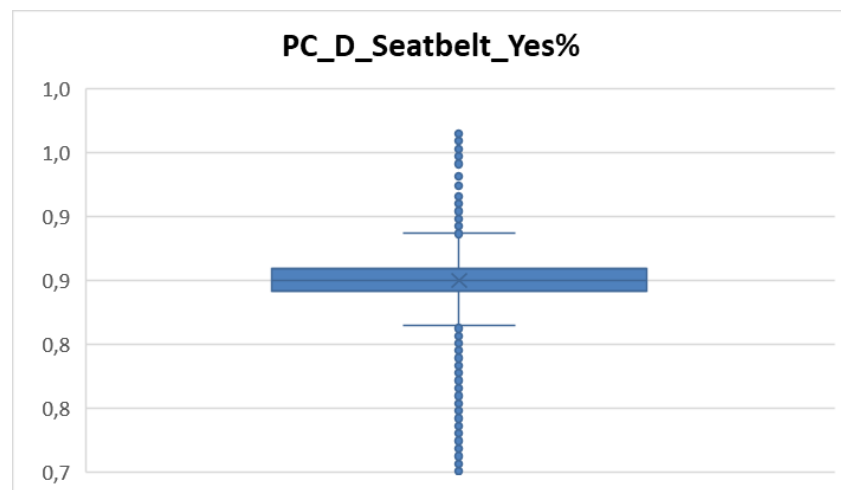
Η πλειονότητα των τιμών είναι κοντά στη μονάδα, γεγονός που υποδηλώνει ότι η **πλειοψηφία των εξεταζόμενων οδικών τμημάτων είναι ευθυγραμμίες.**



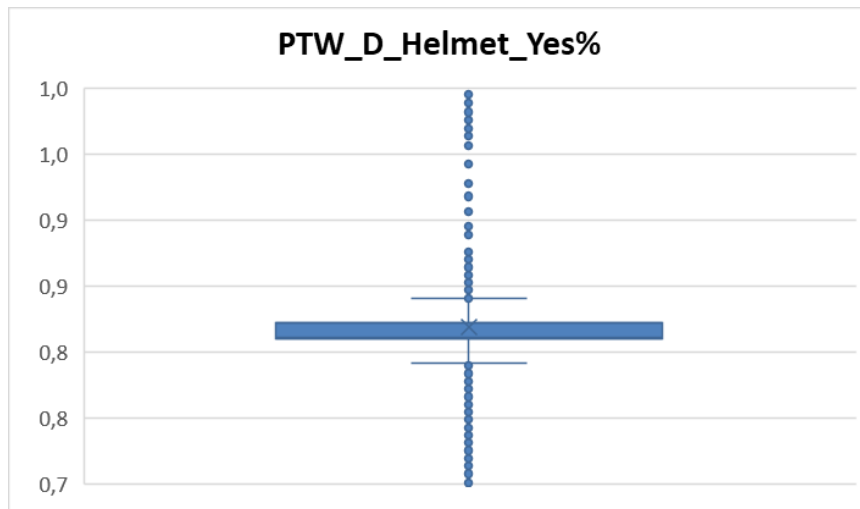
Διάγραμμα 4.3: Κατανομή μη χρήσης κινητού



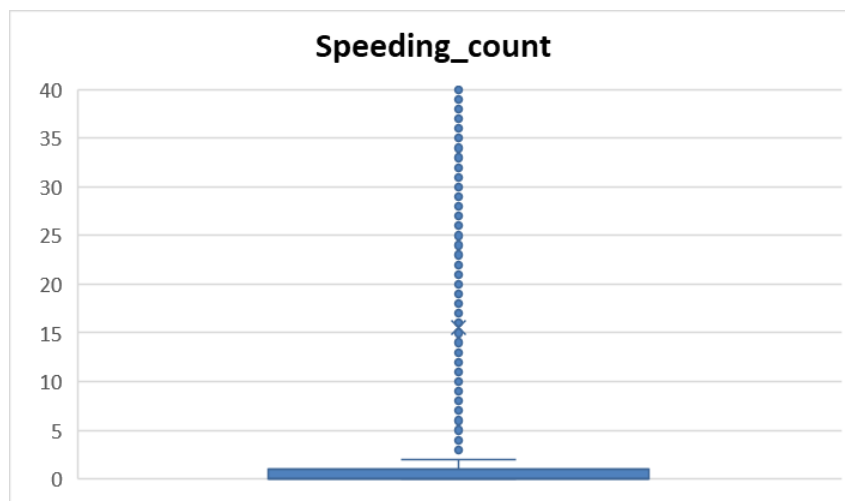
Διάγραμμα 4.4: Κατανομή μη υπέρβασης του ορίου ταχύτητας



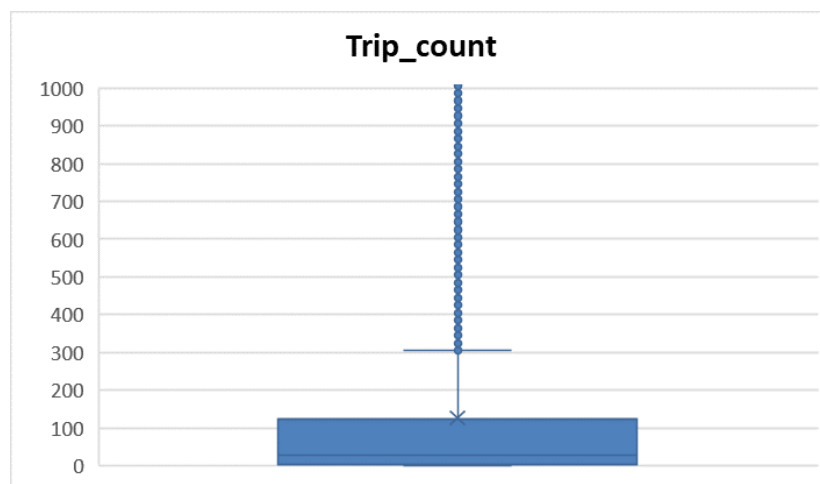
Διάγραμμα 4.5: Κατανομή χρήσης ζώνης ασφαλείας



Διάγραμμα 4.6: Κατανομή χρήσης κράνους πρόσκρουσης

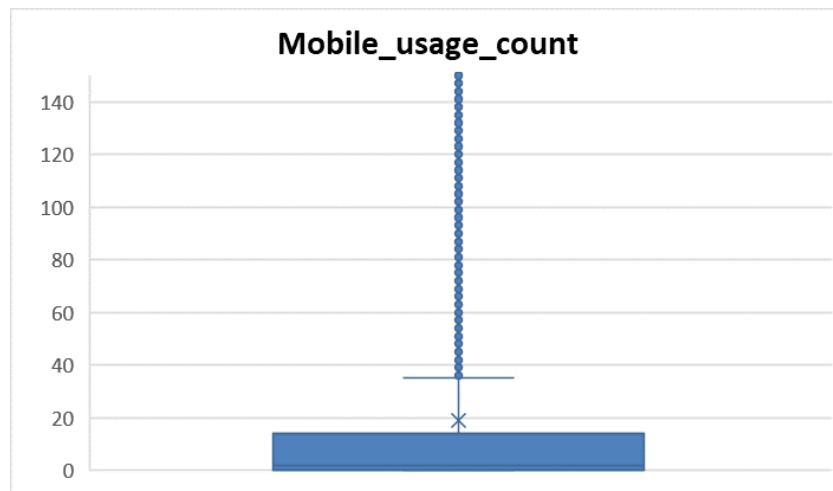


Διάγραμμα 4.7: Κατανομή περιστατικών υπερβολικής ταχύτητας

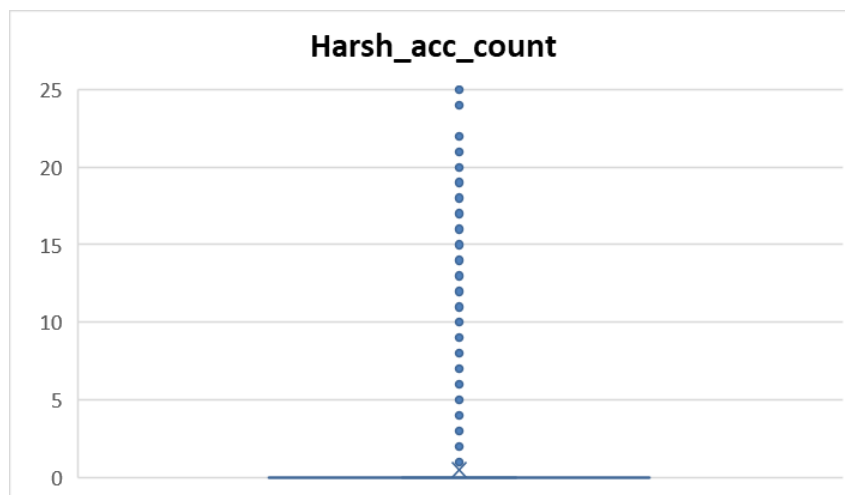


Διάγραμμα 4.8: Κατανομή συνολικού αριθμού διαδρομών

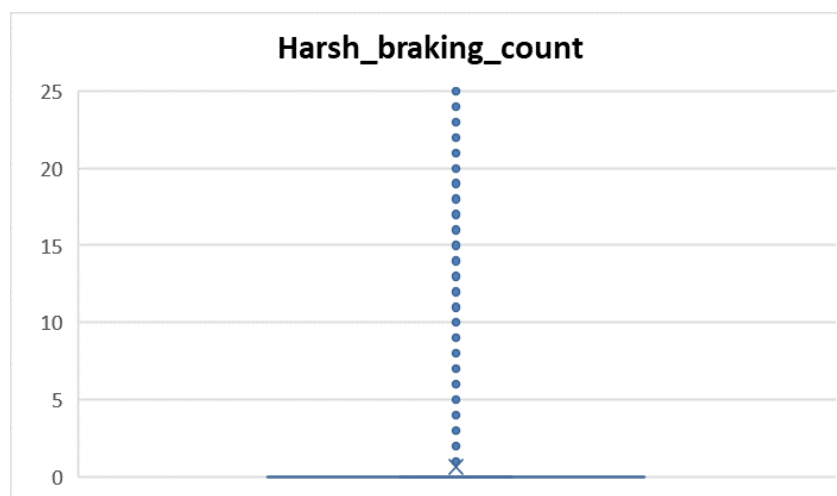
Η κατανομή είναι **έντονα ασύμμετρη**, με τις περισσότερες τιμές να βρίσκονται κάτω από το 100. Ωστόσο, οι ακραίες τιμές εμφανίζουν εξαιρετικά υψηλές τιμές, οι οποίες αυξάνουν σημαντικά το μέγιστο, υποδεικνύοντας ότι ορισμένες διαδρομές παρουσιάζουν πολύ μεγαλύτερη συχνότητα.



Διάγραμμα 4.9: Κατανομή συχνότητας χρήσης κινητού τηλεφώνου



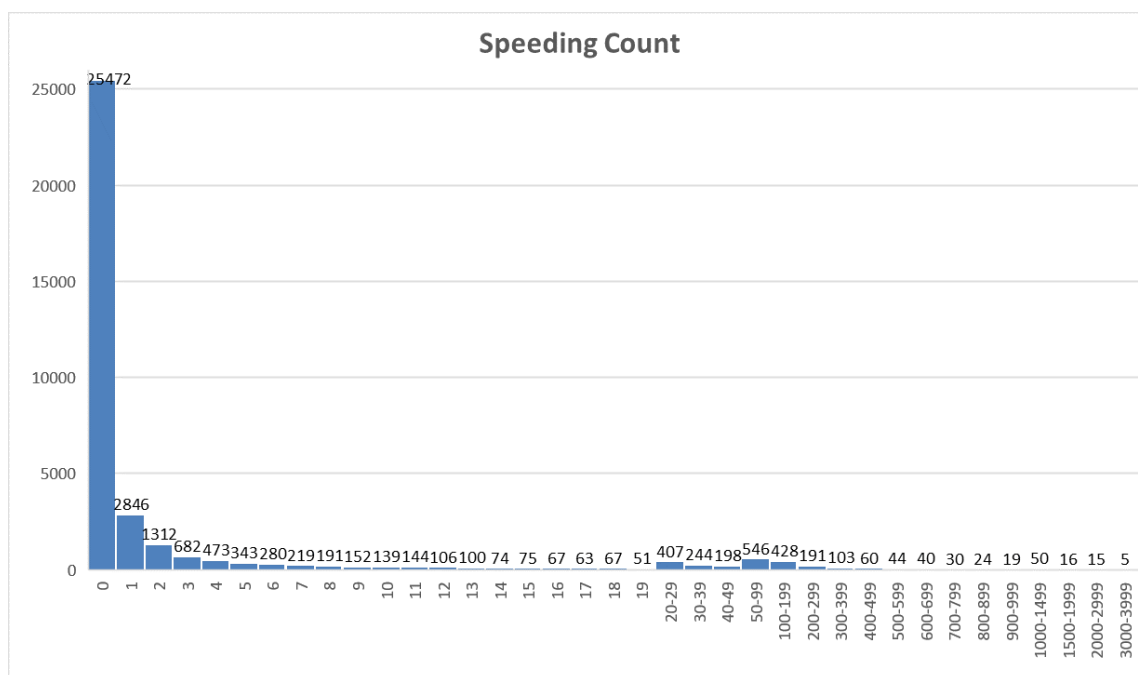
Διάγραμμα 4.10: Κατανομή συχνότητας απότομων επιταχύνσεων



4.11: Κατανομή συχνότητας απότομων επιβραδύνσεων

Στο διάγραμμα για τις επιταχύνσεις και για τις επιβραδύνσεις παρατηρείται ότι, η πλειοψηφία των περιπτώσεων βρίσκεται σε χαμηλά επίπεδα, ενώ οι υψηλές τιμές αναφέρονται σε περιστατικά επιθετικών συμβάντων, που ενδέχεται να συνδέονται με αυξημένο κίνδυνο ατυχήματος.

Ακολούθως, γίνεται ένας διαχωρισμός των περιπτώσεων υπέρβασης του ορίου ταχύτητας (*speeding_count*), ο οποίος αποσκοπεί στην ανάλυση της συχνότητας εμφάνισης σε διάφορα εύρη τιμών, ώστε να βρεθούν οι τιμές που παρουσιάζουν μεγαλύτερη συχνότητα. Ακόμη, με τη δημιουργία ενός ιστογράμματος θα παρατηρηθεί η κατανομή που ακολουθείται. Η ταξινόμηση των τιμών, θα βοηθήσει στην στατιστική ανάλυση που θα επιλεγεί στην συνέχεια, επιτρέποντας τον σωστό καθορισμό των κατηγοριών της μεταβλητής '*speeding_count*' για πιο ακριβή αποτελέσματα και συμπεράσματα σχετικά με τις υπερβάσεις ταχύτητας.



Διάγραμμα 4.12: Ιστόγραμμα παραβάσεων του ορίου ταχύτητας στα εξεταζόμενα οδικά τμήματα

Υπενθυμίζεται ότι η μεταβλητή '*speeding_count*', αναφέρεται στον αριθμό των περιπτώσεων υπέρβασης του ορίου ταχύτητας και κάθε τιμή αντιπροσωπεύει το πλήθος των παραβάσεων που καταγράφηκαν σε συγκεκριμένο εύρος ταχύτητας πάνω από το όριο.

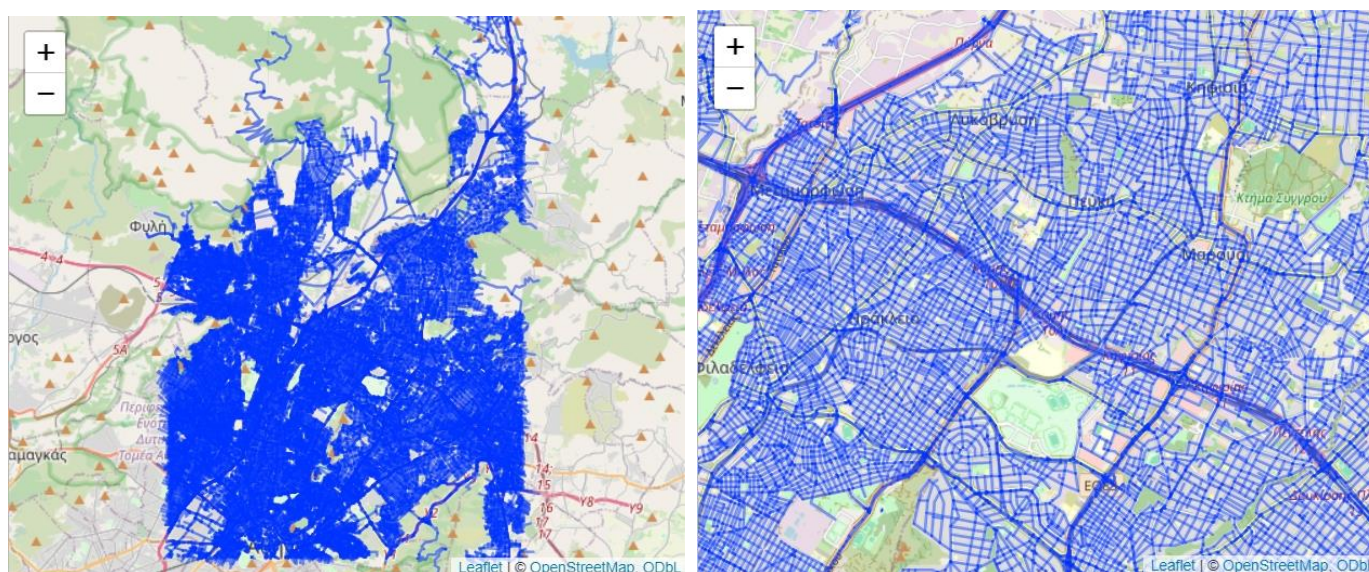
Στο παραπάνω διάγραμμα απεικονίζεται ο αριθμός των παραβάσεων του ορίου ταχύτητας, ομαδοποιημένος σε διάφορες κατηγορίες. Κάθε κατηγορία αντιπροσωπεύει μια συγκεκριμένη συχνότητα παραβιάσεων. **Παρατηρείται ότι 25.742 περιπτώσεις δεν παρουσίασαν καμία υπέρβαση του ορίου ταχύτητας**, ενώ 2.846 είχαν μια υπέρβαση των ορίων ταχύτητας διάρκειας ενός δευτερολέπτου. Οι υπόλοιπες κατηγορίες που αντιπροσωπεύουν με τον ίδιο τρόπο μεγαλύτερης διάρκειας υπερβάσεις, εμφανίζουν σημαντικά μικρότερα ποσοστά. Οι μετρήσεις αυτές, υποδηλώνουν ότι **η πλειοψηφία των οδηγών είτε τηρεί τα όρια ταχύτητας είτε τα υπερβαίνει για μικρές χρονικές περιόδους**, με σχετικά λίγους να υπερβαίνουν σημαντικά τα όρια.

4.3.2 Σύνθεση αλγορίθμου

Στο περιβάλλον της R, φορτώθηκε ένα χωρικό σύνολο δεδομένων σε μορφή *GeoJSON*, το οποίο περιλάμβανε τα δεδομένα του οδικού δικτύου της Αθήνας που συλλέχθηκαν όπως αναλύθηκε νωρίτερα. Αρχικά, έγινε η εισαγωγή των δεδομένων με την βιβλιοθήκη *jsonlite*, έτσι ώστε να μπορέσει να διαβαστεί το αρχείο και να μετατραπεί σε μορφή κατάλληλη για επεξεργασία.

Στην συνέχεια, οι μεταβλητές των συντεταγμένων των δρόμων μετατράπηκαν σε μορφή *WKT (Well- Known Text)*, η οποία είναι ιδανική για χωρική ανάλυση. Επίσης, το σύνολο των δεδομένων μετατράπηκε σε αντικείμενο τύπου *SF (Simple Features)*, για τον ίδιο λόγο.

Με την βιβλιοθήκη *leaflet*, **πραγματοποιήθηκε η οπτικοποίηση του οδικού δικτύου**, με τα δεδομένα να εμφανίζονται ως γραμμές σε έναν διαδραστικό χάρτη.



Εικόνα 4.2: Χάρτες οδών Αττικής

Έπειτα, **εφαρμόστηκαν τεχνικές επεξεργασίας**, όπως η επιλογή συγκεκριμένων μεταβλητών και η μετονομασία ορισμένων για να είναι πιο κατανοητές. Οι μεταβλητές *'highway'* (τύπος οδού) και *'slope_class'* (κατηγορία κλίσης οδού), μετατράπηκαν σε **κατηγορικές μεταβλητές** (factors), δηλαδή σε ένα τύπο μεταβλητής που χρησιμοποιείται για να αναπαραστήσει δεδομένα τα οποία χωρίζονται σε διακριτές κατηγορίες.

Αμέσως, **δημιουργήθηκαν δύο νέες δυαδικές μεταβλητές** για τον τύπο του δρόμου, όπου η μια καθορίζει αν είναι αυτοκινητόδρομος (motorway) και η άλλη αν είναι συνδετικός αυτοκινητόδρομος (motorway_link). Αυτές οι μεταβλητές θα βοηθήσουν στην ανάλυση για να ελεγχθεί πώς ο τύπος του δρόμου επηρεάζει την ασφάλεια και τις παραβάσεις ταχύτητας. Με τον ίδιο τρόπο δημιουργήθηκε μια νέα δυαδική μεταβλητή με την ονομασία ***speeding_di***, η οποία έχει ως σκοπό να **κατηγοριοποιήσει τις οδούς ανάλογα με την ύπαρξη ή μη περιστασιακών υπέρβασης ταχύτητας**. Αναλυτικότερα, η μεταβλητή παίρνει την τιμή 0, αν δεν καταγράφηκαν καθόλου περιστατικά και την τιμή 1 στην αντίθετη περίπτωση (που καταγράφηκαν από ένα και πάνω περιστατικά). Αυτή η διαδικασία βοηθά στη μετατροπή της συνεχούς μεταβλητής *'speeding_count'* σε δυαδική μορφή, διευκολύνοντας έτσι την εφαρμογή μεθόδων ταξινόμησης.

Αυτό το στάδιο της επεξεργασίας δεδομένων αποτελεί θεμέλιο για τη στατιστική ανάλυση που θα ακολουθήσει, διασφαλίζοντας ότι τα δεδομένα είναι οργανωμένα και σωστά για να χρησιμοποιηθούν σε μοντέλα ταξινόμησης και αναλύσεις.

4.4 Σύνοψη - Παρατηρήσεις

Η επεξεργασία δεδομένων αναδεικνύει σημαντικές αλληλεπιδράσεις μεταξύ μεταβλητών και τα γραφήματα προσφέρουν μια ολοκληρωμένη εικόνα της οδικής συμπεριφοράς στην περιοχή μελέτης. Συγκεκριμένα παρατηρήθηκε ότι:

- **Οι οδοί με κατηγορία 'residential' (οδοί κατοικιών), αντιπροσωπεύουν το 66,4% του συνόλου** αποτελώντας την κυρίαρχη κατηγορία.
- **Οι οδοί 'motorways' (αυτοκινητόδρομος) και 'primary' (κύρια οδός)** καλύπτουν χαμηλότερα ποσοστά, αλλά **έχουν κρίσιμη σημασία για την ανάλυση** λόγω της υψηλής ταχύτητας που παρατηρείται συνήθως σε αυτούς τους τύπους δρόμων.
- Στην κατηγορία κλίσης 'επίπεδη - flat (0-3%)' ανήκουν το 76,4% των οδών και μόλις το 2,1% στην κατηγορία 'ακραία - extreme (>10%)'. Αυτή η ανισομερής κατανομή υποδεικνύει **ότι η πλειονότητα των οδών δεν έχει σημαντικές κλίσεις**.
- Στο **72% των εξεταζόμενων οδικών τμημάτων**, δεν παρουσιάστηκαν περιστατικά υπέρβασης ταχύτητας (**speeding_count = 0**).
- Οι δείκτες του **μήκους οδού 'length', ευθυγραμμίας 'efficiency' και ποσοστού χρήσης της ζώνης ασφαλείας 'PC_D_Seatbelt_Yes%** παρουσίασαν **έντονη ασυμμετρία (skewness)** στα περιγραφικά στατιστικά που μελετήθηκαν, υποδεικνύοντας ότι μικρός αριθμός παρατηρήσεων επηρεάζει τις μέσες τιμές.
- Στα διαγράμματα boxplots **παρατηρήθηκαν ακραίες τιμές (outliers)** στις μεταβλητές των **περιστατικών υπέρβασης των ορίων ταχύτητας (speeding_count), απότομων επιταχύνσεων (harsh_acc_count), και απότομων επιβραδύνσεων (harsh_braking_count)**, όπου τελικά **συνδέονται με περιστατικά επιθετικής οδήγησης**.
- **Η πλειοψηφία των οδηγών τηρεί τα όρια ταχύτητας** ή τα υπερβαίνει για σύντομες περιόδους, ενώ οι σημαντικές παραβιάσεις είναι περιορισμένες αλλά κρίσιμες για την οδική ασφάλεια.

5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ - ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Εισαγωγή

Στο συγκεκριμένο κεφάλαιο περιλαμβάνονται τα βήματα που ακολουθήθηκαν κατά την εφαρμογή της μεθοδολογίας, η διαδικασία ανάπτυξης των κατάλληλων μαθηματικών μοντέλων και των αλγορίθμων μηχανικής μάθησης, καθώς και η παρουσίαση των αποτελεσμάτων της ανάλυσης της Διπλωματικής Εργασίας. Ύστερα από τη βιβλιογραφική ανασκόπηση συναφών ερευνών, την παρουσίαση του θεωρητικού υποβάθρου που χρησιμοποιήθηκε για την ανάλυση των στοιχείων και της επεξεργασίας αυτών, επιλέχθηκε η κατάλληλη μεθοδολογία. **Οι μεθοδολογίες που επιλέχθηκαν** για την ανάλυση των στατιστικών στοιχείων, όπως παρουσιάστηκαν εκτενώς στο κεφάλαιο 3, είναι οι εξής: **Διαδική Λογιστική Παλινδρόμηση** (Binary Logistic Regression), **Δέντρο Απόφασης** (Decision Tree), **Τυχαία Δάση** (Random Forest), **Μηχανή Διανυσμάτων Υποστήριξης** (SVM), **Ναϊβ Μπέιζ** (Naive Bayes). Αναπόσπαστο μέρος των αποτελεσμάτων αποτελούν και οι στατιστικοί έλεγχοι, που απαιτούνται για την αποδοχή ή απόρριψη των μοντέλων πρόβλεψης.

Οι παραπάνω παλινδρομήσεις εφαρμόστηκαν σε **δύο διαφορετικά σενάρια**, προκειμένου να αξιολογηθεί η αποτελεσματικότητα των μοντέλων υπό διαφορετικές συνθήκες υπέρβασης των ορίων ταχύτητας. Τα σενάρια αυτά είναι τα εξής:

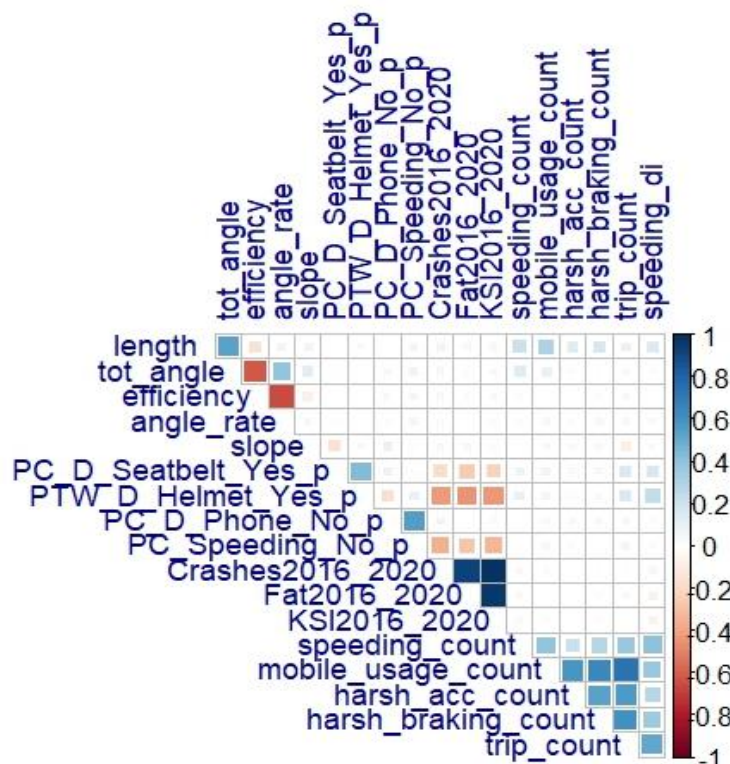
- **Σενάριο 1:** Δείκτης υπέρβασης ταχύτητας για, 0 = δεν καταγράφηκε υπέρβαση των ορίων ταχύτητας στο εξεταζόμενο οδικό τμήμα, 1 = καταγράφηκε υπέρβαση των ορίων ταχύτητας στο εξεταζόμενο οδικό τμήμα τουλάχιστον ενός δευτερολέπτου.
- **Σενάριο 2:** Δείκτης υπέρβασης ταχύτητας για, 0 = καταγράφηκε υπέρβαση των ορίων ταχύτητας στο εξεταζόμενο οδικό τμήμα μικρότερη των 15.55 δευτερολέπτων, 1 = καταγράφηκε υπέρβαση των ορίων ταχύτητας στο εξεταζόμενο οδικό τμήμα τουλάχιστον 15.55 δευτερολέπτων. Επισημαίνεται ότι η τιμή των 15.55 δευτερολέπτων, αποτελεί τη μέση τιμή της διάρκειας της υπέρβασης των ορίων ταχύτητας στα εξεταζόμενα οδικά τμήματα. Με το σενάριο αυτό, δίνεται έμφαση σε πιο σοβαρές περιπτώσεις, που θεωρούνται πιθανότερο να συνδέονται με αυξημένο κίνδυνο.

Η εφαρμογή των μοντέλων σε κάθε σενάριο επιτρέπει τη διερεύνηση της επίδρασης διαφορετικών τύπων συμπεριφοράς ταχύτητας στα αποτελέσματα της ανάλυσης και βελτίωση της ακρίβειας πρόβλεψης των μοντέλων.

5.2 Έλεγχος συσχέτισης μεταβλητών

Αρχικά, πριν ξεκινήσει η δημιουργία των μοντέλων για τον προσδιορισμό της επικινδυνότητας μέσω της πρόβλεψης της υπέρβασης των ορίων ταχύτητας, πρέπει να εξεταστεί η **συσχέτιση των μεταβλητών και η αλληλεπίδραση μεταξύ των παραμέτρων που θα περιλαμβάνονται στο μοντέλο ανάλυσης**. Επιδιώκεται η μέγιστη δυνατή συσχέτιση μεταξύ εξαρτημένων και ανεξάρτητων μεταβλητών. Αντίθετα, για τις ανεξάρτητες μεταβλητές η υψηλή συσχέτιση δεν είναι επιθυμητή. Όταν κατά απόλυτη τιμή η γραμμική σχέση των δύο μεταβλητών είναι κοντά στην μονάδα, η συσχέτιση είναι ισχυρή, ενώ όταν είναι κοντά στο μηδέν δεν είναι στατιστικά σημαντική.

Στην πράξη, όταν η απόλυτη τιμή του δείκτη συσχέτισης κατά Pearson r είναι μικρότερη ή ίση του 0.5, τότε οι μεταβλητές θεωρείται ότι έχουν μικρή συσχέτιση.



Εικόνα 5.1: Διάγραμμα Συσχέτισης Μεταβλητών

Στο παραπάνω διάγραμμα συσχέτισης, αποτυπώνονται με **μπλέ** χρώμα **οι θετικές συσχετίσεις** και με **κόκκινο** **οι αρνητικές**. Με την ένταση του χρώματος υποδυκνύεται το μέγεθος της συσχέτισης.

Υπάρχει **ισχυρή θετική συσχέτιση** μεταξύ του συνολικού αριθμού των ατυχημάτων (**Crashes2016_2020**) και των νεκρών (**Fat2016_2020**), όπως φαίνεται με την έντονη μπλέ απόχρωση. Αυτό είναι και λογικό, καθώς μεγαλύτερος αριθμός συνολικών ατυχημάτων συνδέεται άμεσα και με μεγαλύτερο αριθμό θανάτων.

Επίσης, **θετική συσχέτιση** έχουν και οι μεταβλητές του ποσοστού χρήσης ζώνης ασφαλείας και κράνους πρόσκρουσης, **PC_D_Seatbelt_Yes_p** και **PTW_D_Helmet_Yes_p**, κάτι το οποίο φανερώνει ότι οι περιοχές ή οι πληθυσμοί που τηρούν την χρήση ζώνης, παρουσιάζουν και υψηλά ποσοστά χρήσης κράνους. Αυτή η σύνδεση αντικατοπτρίζει μια γενικότερη **τάση συμμόρφωσης σε μέτρα ασφαλείας**, η οποία μας δίνει ένα παραπάνω δεδομένο για τη συμπεριφορά των χρηστών της οδού στο οδικό δίκτυο που θα εξετάσουμε.

Από την άλλη, εντοπίζεται **αρνητική συσχέτιση** μεταξύ του συνολικού αριθμού σοβαρών τραυματισμών ή θανάτων (**KSI2016_2020**) και του ποσοστού των οδηγών που δεν υπερβαίνουν τα όρια ταχύτητας (**PC_Speeding_No_p**), αυτό μας κάνει να καταλαβαίνουμε, ότι οι περιοχές με υψηλό ποσοστό οδηγών που δεν ξεπερνούν τα όρια ταχύτητας, έχουν και λιγότερους σοβαρούς τραυματισμούς και θανάτους. Η συσχέτιση αυτή **τονίζει την σημασία της τήρησης των ορίων ταχύτητας για την βελτίωση της οδικής ασφάλειας**.

Παρατηρούνται **ισχυρές θετικές** μεταξύ των μεταβλητών της υπέρβασης ή μη των ορίων ταχύτητας (**speeding_di**) και του αριθμού των διαδρομών (**trip_count**), το οποίο υποδηλώνει ότι οι περισσότερες διαδρομές σχετίζονται με μεγαλύτερο αριθμό υπερβάσεων ταχύτητας, γεγονός που είναι αναμενόμενο, όπως και με το δείκτη του μήκους της οδού (**length**), παρατηρείται θετική συσχέτιση που φανερώνει ότι οδοί με μεγαλύτερο μήκος παρουσιάζουν συχνότερες υπερβάσεις των ορίων ταχύτητας.

Ακόμη, **οι μεταβλητές των απότομων συμπεριφορών οδήγησης** (**harsh_acc_count** και **harsh_braking_count**), **εμφανίζουν θετικές συσχετίσεις** που δείχνουν ότι οι επιθετικές συμπεριφορές συνοδεύονται από παραβάσεις ταχύτητας. Αυτό αντικατοπτρίζει ένα πιο επικίνδυνο προφίλ οδηγικής συμπεριφοράς. Αυτό επιβεβαιώνεται και με τη θετική συσχέτιση των των περιστατικών χρήσης κινητού τηλεφώνου κατά την οδήγηση (**Mobile_usage_count**) και των περιστατικών απότομης επιτάχυνσης (**Harsh_acc_count**), κάτι που ενδεχομένως δείχνει ότι η χρήση κινητού τηλεφώνου κατά την οδήγηση συνδέεται με αυξημένες επιθετικές επιταχύνσεις, πιθανώς λόγω της διάσπασης της προσοχής των οδηγών.

Τέλος, παρατηρείται και μια **συσχέτιση** με τη μεταβλητή των περιστατικών χρήσης κινητού τηλεφώνου κατά την οδήγηση (**mobile_usage_count**), η οποία πρέπει να διερευνηθεί για να εντοπιστεί η ένταση της, καθώς συσχετίζεται **ασθενώς με τις υπόλοιπες μεταβλητές**. Αυτό μπορεί να σημαίνει είτε ότι δεν επηρεάζει άμεσα άλλες παραμέτρους της συγκεκριμένης ανάλυσης, είτε ότι λειτουργεί ανεξάρτητα, είτε ότι έχει συμπληρωθεί μικρή καταγραφή σχετικών συμβάντων.

Αυτή η ανάλυση αποτελεί σημαντικό βήμα για την κατανόηση των σχέσεων μεταξύ μεταβλητών και συμβάλλει στην επιλογή εκείνων που θα συμπεριληφθούν στα τελικά μοντέλα πρόβλεψης.

5.3 Ανάπτυξη Μοντέλων Παλινδρόμησης στο Λογισμικό της R

Στην παρούσα ενότητα περιγράφεται η **διαδικασία ανάλυσης** για την ανάπτυξη των στατιστικών μοντέλων και των μοντέλων μηχανικής μάθησης. Σκοπός της ανάλυσης με τις μεθόδους παλινδρόμησης που έχουν επιλεγεί είναι να υπολογιστούν στατιστικά σημαντικά μοντέλα και να εντοπιστούν παράγοντες που επηρεάζουν περισσότερο την εξαρτημένη μεταβλητή.

Αρχικά, **τα δεδομένα χωρίστηκαν σε ένα σύνολο εκπαίδευσης (80%) και σύνολο δοκιμής (20%), μέσω μιας τυχαίας κατανομής**. Η εξαρτημένη μεταβλητή είναι η **speeding_di** και μετατρέπεται σε κατηγορική (factor), καθώς πρόκειται για δυαδική μεταβλητή που δηλώνει ύπαρξη ή μη υπέρβασης ταχύτητας, ανάλογα με το όριο που έχει οριστεί σε κάθε σενάριο.

Οι ανεξάρτητες μεταβλητές που επιλέχθηκαν, μετά από τον έλεγχο στατιστικής σημαντικότητας και πολλαπλές δοκιμές ήταν οι:

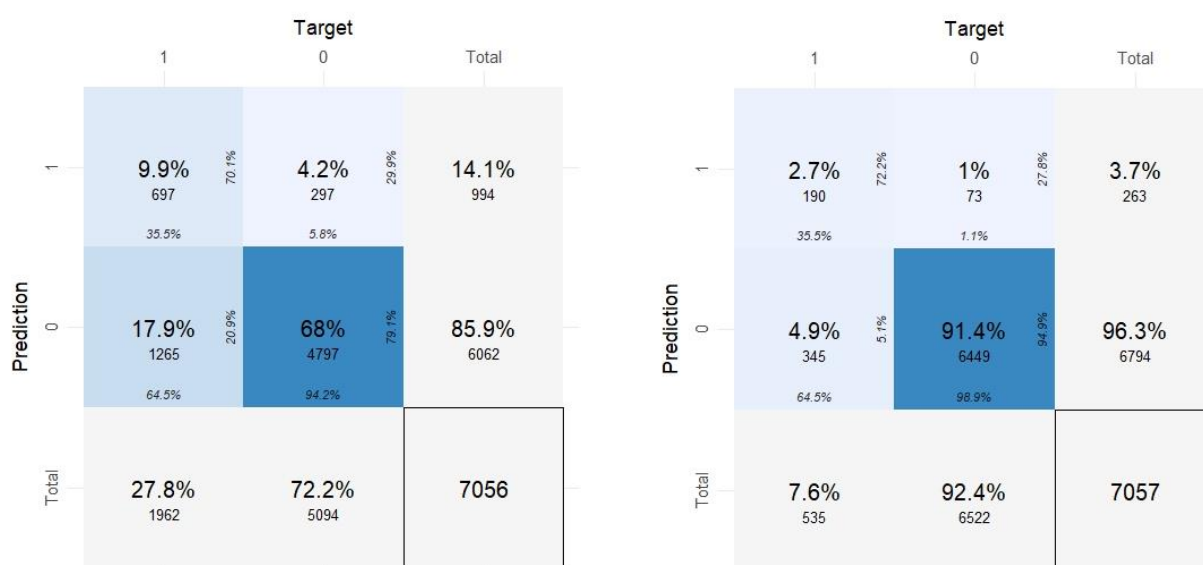
- length (μήκος οδού),
- efficiency (δείκτης ευθυγραμμίας οδού),
- slope_class (κατηγορία κλίσης οδού),
- highway (τύπος οδού),
- mobile_usage_count (χρήση κινητού (sec)),
- harsh_acc_count (αριθμός απότομων επιταχύνσεων),
- harsh_braking_count (αριθμός απότομων επιβραδύνσεων),
- trip_count (αριθμός διαδρομών).

5.3.1 Μοντέλο Δυαδικής Λογιστικής Παλινδρόμησης

Στο συγκεκριμένο κεφάλαιο, εφαρμόστηκε δυαδική λογιστική παλινδρόμηση (Binary Logistic Regression) για την ανάλυση της πιθανότητας ύπαρξης υπέρβασης του ορίου ταχύτητας (speeding) σε σχέση με διάφορους παράγοντες. Σκοπός της ανάλυσης είναι η **διερεύνηση των σχέσεων μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών**, καθώς και η **αξιολόγηση της ακρίβειας του μοντέλου πρόβλεψης**. Η συγκεκριμένη επιλέχθηκε επειδή είναι κατάλληλη για την πρόβλεψη δυαδικών εξαρτημένων μεταβλητών και επιτρέπει την εκτίμηση των πιθανοτήτων εμφάνισης ενός γεγονότος.

Εκτιμήθηκαν οι συντελεστές του μοντέλου, μέσω της συνάρτησης *glm* στο R-studio. Τα αποτελέσματα της εκτίμησης παρουσιάζονται μέσα από την περίληψη του μοντέλου με την εντολή *summary(bl_model)* και αναγράφονται τα σημαντικά στατιστικά για να υποδείξουν την συνεισφορά της κάθε ανεξάρτητης μεταβλητής στην πιθανότητα υπέρβασης της ταχύτητας.

Για την περαιτέρω αξιολόγησή του μοντέλου, δημιουργήθηκε ένας πίνακας σύγχυσης (*confusion_matrix*), ο οποίος εμφανίζει και συγκρίνει τις πραγματικές τιμές με τις προβλεπόμενες τιμές του μοντέλου και δίνεται παρακάτω:



Πίνακας 5.2-3: Πίνακες αξιολόγησης απόδοσης Σεναρίου 1 και 2 (Confusion Matrix)

	Πραγματικά Θετικά (Positive)	Πραγματικά Αρνητικά (Negative)
Προβλέφθηκε Θετικό	True Positive (TP)	False Positive (FP)
Προβλέφθηκε Αρνητικό	False Negative (FN)	True Negative (TN)

Πίνακας 5.4: Επεξήγηση τιμών Πίνακα Σύγχυσης

Μέσω εκείνου αναδεικνύεται η ακρίβεια και η δυνατότητα πρόβλεψης τόσο των θετικών όσο και των αρνητικών τιμών, όπου υπολογίζονται μέσω των δεικτών της ορθότητας (Accuracy), της ευαισθησίας (Sensitivity) και της ειδικότητας (Specificity).

Τα αποτελέσματα είναι τα εξής:

Binary Logistic	Σενάριο 1	Σενάριο 2
Ορθότητα	0.778	0.941
Ευαισθησία	0.355	0.355
Ειδικότητα	0.942	0.989

Πίνακας 5.5: Αποτελέσματα βασικών στατιστικών δεικτών

Πιο αναλυτικά, η **ορθότητα** (Accuracy) στο πρώτο σενάριο είναι ικανοποιητική αλλά όχι πολύ υψηλή, ενώ στο δεύτερο είναι εξαιρετική, κάτι που υποδηλώνει ότι το μοντέλο αποδίδει πολύ καλά, λόγω της επιτυχίας της επεξεργασίας των στοιχείων ή λόγω ισχυρών μεταβλητών.

Η **ευαισθησία** (Sensitivity) και στα δύο σενάρια παραμένει στα ίδια επίπεδα, περίπου στο 35%. Αυτό δείχνει ότι τα μοντέλα δεν είναι όσο αποτελεσματικά επιθυμούμε στον εντοπισμό των πραγματικών υπερβάσεων του ορίου ταχύτητας. Ενδεχομένως, αυτό να οφείλεται στην ανισορροπία των δεδομένων, είτε σε ανεπαρκή χαρακτηριστικά για την κατηγορία ελέγχου.

Η **ειδικότητα** (Specificity) και στα δύο σενάρια εμφανίζει πολύ καλή τιμή, που σημαίνει ότι το μοντέλο είναι πολύ αποτελεσματικό στην πρόβλεψη περιπτώσεων 'οδοί χωρίς υπέρβαση των ορίων ταχύτητας', αφού τις εντοπίζει σχεδόν πάντα.

Άλλα αποτελέσματα που λήφθηκαν είναι σχετικά με τις εκτιμήσεις των παραμέτρων (estimates). **Στις μεταβλητές που περιλαμβάνουν υποκατηγορίες θα γίνει σύγκριση με τις βασικές κατηγορίες αναφοράς** που επιλέγονται σε κάθε περίπτωση. Συγκεκριμένα, οι μεταβλητή της κατηγορίας κλίσης οδού (slope_class) θα έχει ως μεταβλητή αναφοράς το 'slope_class: extreme >10' και οι μεταβλητή του τύπου της οδού (highway) θα λαμβάνει μεταβλητή αναφοράς το 'highway type: living_street'. Αυτή η προσέγγιση μας επιτρέπει να τις αξιολογήσουμε με την σχετική επίδραση που θα έχει κάθε κατηγορία σε σύγκριση με τις κατηγορίες αναφοράς, λαμβάνοντας αποτελέσματα για να ερμηνεύσουμε την σημασία τους στο πλαίσιο της οδικής ασφάλειας.

Σενάριο 1

Πιο αναλυτικά, η σταθερά (Intercept) δείχνει το σημείο εκκίνησης της πιθανότητας, τη στιγμή που όλες οι ανεξάρτητες μεταβλητές είναι μηδέν. Η τιμή είναι με αρνητική εκτίμηση και αναδεικνύεται ως εξαιρετικά σημαντική.

Το μήκος του οδικού τμήματος (length) έχει θετική εκτίμηση, που δείχνει ότι η αύξηση του μήκους **συνδέεται με την πιθανότητα υπέρβασης της ταχύτητας** και είναι επίσης εξαιρετικά σημαντικό. **Αντίστοιχα και η ευθυγραμμία** (efficiency) έχει θετική εκτίμηση και είναι στατιστικά σημαντική, απλά σε μικρότερο βαθμό.

Όσον αφορά την μεταβλητή της **κατηγορίας κλίσης της οδού** (slope_class), όλες οι κατηγορίες έχουν θετικές εκτιμήσεις και είναι στατιστικά σημαντικές με $Pr < 0.05$ ή και μικρότερο. Στην πρώτη κατηγορία '0-3%' η εκτίμηση υποδεικνύει σημαντικά υψηλή πιθανότητα υπέρβασης ταχύτητας σε επίπεδες περιοχές σε σύγκριση με την κατηγορία

>10%. Αυτό ενδέχεται να οφείλεται στη μικρότερη αίσθηση κινδύνου ή στην ευκολότερη ανάπτυξη ταχύτητας σε μικρές κλίσεις. Στις επόμενες κατηγορίες η πιθανότητα μειώνεται σημαντικά, συγκριτικά με τις επίπεδες περιοχές, αλλά παραμένει αυξημένη σε σχέση με την κατηγορία αναφοράς. Σε κλίσεις μεγαλύτερες του 10%, η αίσθηση κινδύνου είναι εμφανής και η τεχνική δυσκολία του οδικού περιβάλλοντος οδηγεί σε μικρή πιθανότητα υπέρβασης.

Για τον τύπο οδού (highway), όλες οι κατηγορίες έχουν θετικές εκτιμήσεις, που δείχνουν ότι οι τύποι αυτοί σχετίζονται με αυξημένη πιθανότητα υπέρβασης της ταχύτητας.

Στις κατηγορίες αυτοκινητόδρομος (motorway), σύνδεσμος αυτοκινητόδρομου (motorway_link), κύρια οδός (primary), σύνδεσμος κύριας οδού (primary_link), δευτερεύουσα οδός (secondary), σύνδεσμος δευτερεύουσας οδού (secondary_link), τριτεύουσα οδός (tertiary), σύνδεσμος τριτεύουσας οδού (tertiary_link), κεντρική αρτηρία (trunk), και σύνδεσμος κεντρικής αρτηρίας (trunk_link) εμφανίζεται μεγάλη στατιστική σημαντικότητα, με εξαίρεση να αποτελεί η κατηγορία της οδού κατοικιών (residential), η οποία δεν είναι σημαντική, για $Pr = 0.55$ και φανερώνει ότι η πιθανότητα υπέρβασης της ταχύτητας είναι συγκρίσιμη με εκείνη της κατηγορίας οδού ήπιας κυκλοφορίας (living_street).

Πιο αναλυτικά, στις κατηγορίες αυτοκινητόδρομος και σύνδεσμος αυτοκινητόδρομου οι εκτιμήσεις είναι μεγάλες, όπως είναι αναμενόμενο καθώς οι οδοί είναι σχεδιασμένοι για υψηλές ταχύτητες. Ακόμη, στις κατηγορίες δευτερεύουσας οδού και συνδέσμου αυτής, οι εκτιμήσεις δείχνουν ότι η πιθανότητα υπέρβασης είναι υψηλότερη στις συνδετήριες οδούς από ότι στους δευτερεύοντες άξονες. Τέλος, γίνεται αντιληπτό, ότι οι υψηλές εκτιμήσεις για τις κατηγορίες αυτοκινητόδρομος και κύρια οδός, δημιουργούν την ανάγκη ελέγχου ταχύτητας σε αυτά τα περιβάλλοντα, καθώς οι οδηγοί τείνουν να αναπτύσσουν ταχύτητες σημαντικά υψηλότερες από τις οδούς χαμηλότερης κυκλοφορίας.

Στην μεταβλητή περιπτώσεων χρήσης κινητού τηλεφώνου, η αρνητική εκτίμηση δείχνει ότι η αύξηση στην χρήση κινητού, οδηγεί στην μείωση της πιθανότητας υπέρβασης της ταχύτητας και είναι στατιστικά σημαντική.

Επιπλέον, στις **απότομες συμπεριφορές οδήγησης**, δηλαδή στις απότομες επιταχύνσεις και επιβραδύνσεις υπολογίζονται θετικές συσχετίσεις και πολύ σημαντικές στατιστικά.

Τέλος, αντίστοιχα και στην **μεταβλητή του αριθμού των διαδρομών, υπάρχει θετική εκτίμηση**, η οποία υποδηλώνει ότι η αύξηση του αριθμού των διαδρομών σχετίζεται με μεγαλύτερη πιθανότητα υπέρβασης του ορίου ταχύτητας, όπως και θα περιμέναμε.

Παρακάτω, δίνονται αναλυτικά τα στοιχεία που λάβαμε από το μοντέλο της δυαδικής λογιστικής παλινδρόμησης, όπου στην πρώτη στήλη εμφανίζεται το όνομα της κάθε μεταβλητής. Η δεύτερη περιλαμβάνει τις εκτιμήσεις των συντελεστών, που προσδιορίζουν την επίδραση στην εξαρτημένη μεταβλητή και η τρίτη στήλη δείχνει την τιμή p . Η έλλειψη αστερίσκων ($p > 0.05$), σημαίνει ότι η μεταβλητή δεν θεωρείται στατιστικά σημαντική σε επίπεδο εμπιστοσύνης 95%, ενώ ο αριθμός των αστερίσκων υποδηλώνει την ένταση της σημαντικότητας.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Name	Estimate	Pr(> z)
(Intercept)	-37.359.162	< 2e-16 ***
length	0.0023779	< 2e-16 ***
efficiency	0.6267127	0.039356 *
slope_class0-3	10.487.670	1.96e-07 ***
slope_class3-5	0.9506093	3.30e-06 ***
slope_class5-8	0.7791038	0.000223 ***
slope_class8-10	0.7992437	0.001297 **
highwaymotorway	30.019.741	< 2e-16 ***
highwaymotorway_link	20.316.959	< 2e-16 ***
highwayprimary	10.810.562	1.15e-14 ***
highwayprimary_link	18.766.039	< 2e-16 ***
highwayresidential	0.0699595	0.554660
highwaysecondary	0.4190342	0.001174 **
highwaysecondary_link	15.609.873	< 2e-16 ***
highwaytertiary	0.7164621	6.10e-09 ***
highwaytertiary_link	0.7323863	0.007577 **
highwaytrunk	16.009.920	1.72e-05 ***
highwaytrunk_link	14.009.141	0.020153 *
mobile_usage_count	-0.0039140	4.19e-12 ***
harsh_acc_count	0.0283877	0.036268 *
harsh_braking_count	0.0781338	2.24e-12 ***
trip_count	0.0031773	< 2e-16 ***

Πίνακας 5.6: Εκτιμήσεις και Στατιστική Σημαντικότητα Σεναρίου 1

Σενάριο 2

Στην σταθερά και στην μεταβλητή **του μήκους της οδού**, τα αποτελέσματα είναι παρόμοια με το προηγούμενο σενάριο. Αντίθετα, η μεταβλητή **του δείκτη της ευθυγραμμίας** παρουσιάζει αρκετά μεγάλη θετική συσχέτιση και δεν είναι στατιστικά σημαντική αφού έχει $Pr = 0.105$.

Οι κατηγορίες της **μεταβλητής της κλίσης της οδού** έχουν όλες εκτιμήσεις πολύ κοντά στο μηδέν και είναι στατιστικά μη σημαντικές με $p > 0.05$. Αυτό δείχνει ότι η μεταβλητή της κλίσης, ενδεχομένως να μην έχει ισχυρή επίδραση στο μοντέλο, όταν συγκρίνεται με την κατηγορία '>10'. Οι τιμές πολύ κοντά στο μηδέν, δείχνουν μικρή ή αμελητέα συσχέτιση σε σχέση με την μεταβλητή αναφοράς. Η κατηγορία '5-8', έχει αρνητικό πρόσημο που δείχνει μικρή αρνητική συσχέτιση με το αποτέλεσμα. Οι μικρές διαφορές στις εκτιμήσεις μπορεί να οφείλονται σε τυχαίες διακυμάνσεις ή σε χαρακτηριστικά του δείγματος.

Για την **κατηγορία της οδού**, όλες οι κατηγορίες με εξαίρεση τις οδός κατοικιών (residential) και σύνδεσμος κεντρικής αρτηρίας (trunk_link), έχουν θετικές και αυξημένες συσχετίσεις και είναι εξαιρετικά σημαντικές στατιστικά, όπως και προηγούμενως. Οι δύο τελευταίες κατηγορίες παρουσιάζουν αρνητική συσχέτιση και

αρκετά μεγάλη τιμή $p > 0.05$, η οποία τις καθιστά στατιστικά μη σημαντικές σε σύγκριση με την κατηγορία αναφοράς. Συμπεραίνουμε, ότι **οι δρόμοι με υψηλή σημασία, όπως οι αυτοκινητόδρομοι και οι κύριοι δρόμοι, επηρεάζουν έντονα τα αποτελέσματα**. Αυτό μπορεί να οφείλεται στην αυξημένη κίνηση, ταχύτητα ή και την φύση της οδηγική συμπεριφοράς σε δρόμους ταχείας κυκλοφορίας.

Η εκτίμηση για την **μεταβλητή των περιστατικών χρήσης κινητού τηλεφώνου** (mobile_usage_count) **παραμένει αρνητική**, πιθανώς λόγω πιο προσεκτικής οδήγησης σε αυτές τις περιπτώσεις και στατιστικά σημαντική. Επίσης, **ο αριθμός των διαδρομών** (trip_count) **διατηρεί τη θετική** και στατιστικά σημαντική του επίδραση, ενισχύοντας την προηγούμενη ανάλυση.

Η **μεταβλητή της απότομης επιτάχυνσης** (harsh_acc_count), έχει **αρνητική εκτίμηση**, δεν είναι στατιστικά σημαντική σε αυτό το σενάριο, και δεν υπάρχει ισχυρή ένδειξη ότι η αύξηση στις απότομες επιταχύνσεις σχετίζεται με την εξαρτημένη μεταβλητή. Αντίθετα, όσον αφορά τις **απότομες επιβραδύνσεις** (harsh_braking_count), η εκτίμηση είναι θετική και στατιστικά σημαντική. Το μέγεθος της επίδρασης, όμως, παρόλο που είναι θετικό, παραμένει σχετικά μικρό.

Ο αντίστοιχος πίνακας με την σύνοψη των αποτελεσμάτων δίνεται παρακάτω:

Name	Estimate	Pr(> z)
(Intercept)	-62.446.160	2.05e-13 ***
length	0.0028611	< 2e-16 ***
efficiency	10.822.690	0.10508
slope_class0-3	0.1233906	0.75311
slope_class3-5	0.3118467	0.43216
slope_class5-8	-0.0704221	0.86465
slope_class8-10	0.2702110	0.58874
highwaymotorway	41.106.006	< 2e-16 ***
highwaymotorway_link	31.497.470	1.73e-13 ***
highwayprimary	20.847.150	8.02e-07 ***
highwayprimary_link	22.768.847	7.75e-07 ***
highwayresidential	-0.1735704	0.67717
highwaysecondary	20.414.313	9.63e-07 ***
highwaysecondary_link	25.713.696	1.91e-08 ***
highwaytertiary	18.642.344	7.03e-06 ***
highwaytertiary_link	21.618.771	0.00027 ***
highwaytrunk	38.601.927	4.75e-14 ***
highwaytrunk_link	-95.131.345	0.95428
mobile_usage_count	-0.0037988	2.81e-12 ***
harsh_acc_count	-0.0168832	0.20498
harsh_braking_count	0.0610185	7.96e-10 ***
trip_count	0.0031843	< 2e-16 ***

Πίνακας 5.7: Εκτιμήσεις και Στατιστική Σημαντικότητα Σεναρίου 2

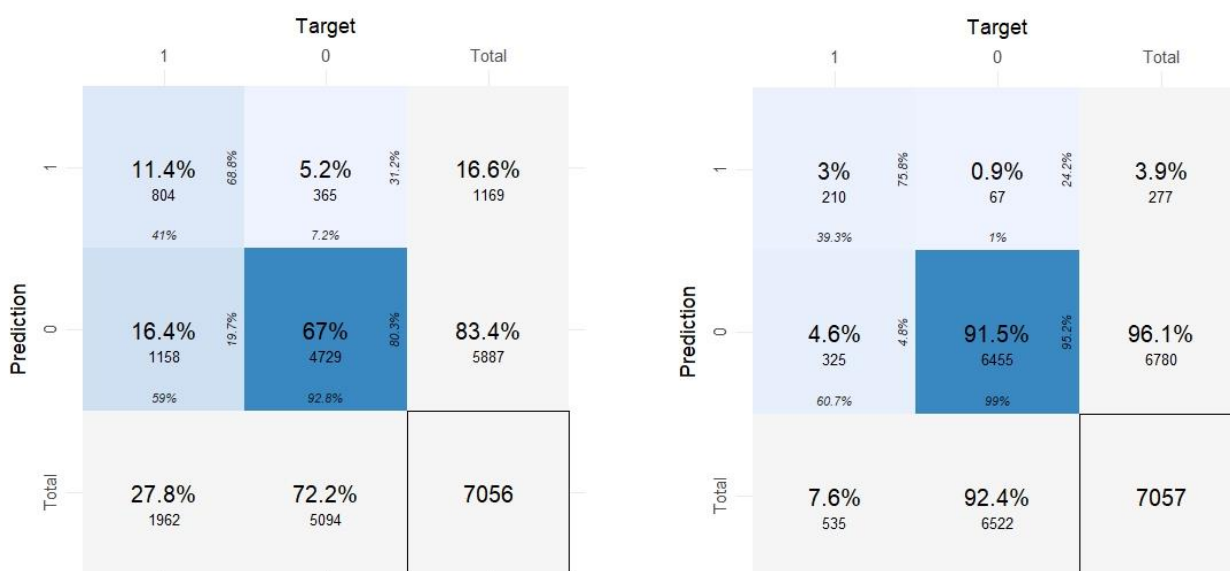
5.3.2 Δέντρο Απόφασης

Στη συνέχεια εφαρμόζεται η Decision Tree Classification και αναλύεται παρακάτω η διαδικασία. Η συγκεκριμένη ανάλυση επικεντρώνεται στην πρόβλεψη της μεταβλητής στόχου και εστιάζει στην ανάπτυξη μοντέλου που βελτιστοποιεί τις υπέρ-παραμέτρους.

Αρχικά, για την εκπαίδευση και αξιολόγηση του μοντέλου, **εφαρμόστηκε η μέθοδος πενταπλής επικύρωσης Cross-Validation (CV)**. Η συγκεκριμένη μέθοδος, διαιρεί το σύνολο δεδομένων σε πέντε ισομεγέθη υποσύνολα. Σε κάθε βήμα, τέσσερα από αυτά χρησιμοποιούνται για την εκπαίδευση και το ένα που περισσεύει για την αξιολόγηση. Με τον τρόπο αυτό, εξασφαλίζεται η αξιοπιστία του μοντέλου, καθώς μειώνεται η πιθανότητα υπερεκπαίδευσης (overfitting) και αυξάνεται η γενικευσιμότητα των αποτελεσμάτων στα δεδομένα.

Για να βρεθούν οι κατάλληλες τιμές που θα χρησιμοποιηθούν στις υπέρ-παραμέτρους των τιμών βάθους `maxdepth`, επιλέχθηκε η τεχνική `Grid Search`, στην οποία το βάθος εξετάστηκε για τις τιμές 1,3,5,7 και 9. Οι τιμές αυτές επηρεάζουν άμεσα την απόδοση του μοντέλου. Συγκεκριμένα, μικρές τιμές βάθους (π.χ. `maxdepth = 1`), οδηγούν σε λιγότερο σύνθετα μοντέλα, τα οποία όμως ενδέχεται να μην αποτυπώνουν σωστά την σχέση μεταξύ των ανεξάρτητων μεταβλητών (underfitting). Ενώ, μεγαλύτερες τιμές βάθους (π.χ. `maxdepth = 9`), δημιουργούν πιο σύνθετα μοντέλα που μπορούν να προσαρμοστούν καλύτερα στα δεδομένα, αλλά ενδέχεται να υπάρχει αυξημένος κίνδυνος υπερεκπαίδευσης (overfitting).

Στην συνέχεια, το μοντέλο ταξινομήθηκε με την μέθοδο `rpart` (Recursive Partitioning and Regression Trees) και χρησιμοποιήθηκε για την πρόβλεψη της εξαρτημένης μεταβλητής `speeding_di` στο `test set`. Όπως και στις προηγούμενες παλινδρομήσεις, τα αποτελέσματα των προβλέψεων αξιολογήθηκαν μέσω του πίνακα σύγχυσης (Confusion Matrix) και απεικονίστηκαν γραφικά:



Πίνακας 5.8-9: Πίνακες αξιολόγησης απόδοσης Σεναρίου 1 και 2 (Confusion Matrix)

Decision Tree	Σενάριο 1	Σενάριο 2
Ορθότητα	0.771	0.938
Ευαισθησία	0.330	0.265
Ειδικότητα	0.941	0.993

Πίνακας 5.10: Αποτελέσματα βασικών στατιστικών δεικτών

Η **ορθότητα** (Accuracy) του αλγορίθμου στο πρώτο σενάριο είναι 77,11%, που υποδεικνύει τον αριθμό των προβλέψεων που είναι σωστές. Πρόκειται για μια ικανοποιητική τιμή, αλλά όχι αρκετή συγκριτικά με το δεύτερο σενάριο, στο οποίο αυξάνεται σημαντικά σε 93,77%. Άρα στη δεύτερη περίπτωση καταλαβαίνουμε, ότι το μοντέλο λειτουργεί πολύ πιο αποτελεσματικά σε γενικό επίπεδο.

Η **ευαισθησία** (Sensitivity) είναι στο πρώτο σενάριο στο 33,03%, κάτι που τονίζει ότι το μοντέλο αδυνατεί και πάλι να εντοπίσει αρκετές από τις περιπτώσεις που ανήκουν στην θετική κλάση, δηλαδή στην ταχύτητα εκτός ορίου. Στο δεύτερο σενάριο μειώνεται ακόμη περισσότερο σε 26,54%. Οι τιμές αυτές φανερώνουν πρόβλημα στην ισορροπία των δεδομένων ή στην επιλογή παραμέτρων και πρέπει να διερευνηθεί.

Τέλος, η **ειδικότητα** (Specificity), κυμαίνεται σε μεγάλα ποσοστά και στα δύο σενάρια, ξεκινάει από 94,09% στο πρώτο και αυξάνεται σε 99,28% στο δεύτερο. Και τα δύο είναι ποσοστά εξαιρετικά, τα οποία καταφέρνουν σχεδόν τέλεια να διακρίνουν τις αρνητικές περιπτώσεις και αυτό αποτελεί σημαντικό πλεονέκτημα.

Συμπερασματικά, **το σενάριο 2 υπερτερεί σημαντικά**, κυρίως ως προς την ακρίβεια, καθώς αυξάνεται περίπου 15%, έναντι του σεναρίου 1. Επίσης, παρουσιάζει υψηλότερο ποσοστό στην ειδικότητα σε ποσοστό σχεδόν 100%, γεγονός που το καθιστά ιδανικό σε περιπτώσεις όπου είναι σημαντική η ανίχνευση αρνητικών περιστατικών, ενώ αντίθετα το σενάριο 1 υπερτερεί ελαφρώς στην ευαισθησία, κάτι που είναι χρήσιμο σε περιπτώσεις ανίχνευσης κρίσιμων θετικών περιστατικών (π.χ. επικίνδυνη ταχύτητα). Λόγω των χαμηλών ποσοστών στην ευαισθησία και στα δύο σενάρια, **κρίνεται απαραίτητη η βελτιστοποίηση του αλγορίθμου**, είτε μέσω επανένταξης των χαρακτηριστικών (features), είτε μέσω χρήσης τεχνικών εξισορρόπησης δεδομένων, όπως το oversampling/undersampling. Επομένως, συνολικά το δεύτερο μοντέλο είναι γενικά πιο αποτελεσματικό στις προβλέψεις του.

Άλλα αποτελέσματα που λήφθηκαν από το συγκεκριμένο μοντέλου, αφορούν την **τιμή του Kappa**, η οποία υποδηλώνει την συμφωνία μεταξύ των πραγματικών και των προβλεπόμενων τιμών πέρα από την συμφωνία που θα μπορούσε να συμβεί τυχαία. Και στα δύο σενάρια η τιμή είναι γύρω στο 0,3 και **δείχνει μια μέτρια συμφωνία**, κάτι που είναι αναμενόμενο για δεδομένα με χαμηλή ευαισθησία.

Επίσης, έγινε ένα **McNemar's Test P-Value** το οποίο συγκρίνει την συμμετρία μεταξύ των λανθασμένων θετικών και αρνητικών προβλέψεων (false positives και false negatives). Οι τιμές που υπολογίστηκαν ήταν πολύ χαμηλές ($<2.2e-16$) και αυτό μαρτυρά **στατιστικά σημαντική ασυμμετρία**, δηλαδή ότι το μοντέλο έχει μεγαλύτερη πιθανότητα να κάνει το ένα είδος λάθους σε σχέση με το άλλο (false negatives ή false positives).

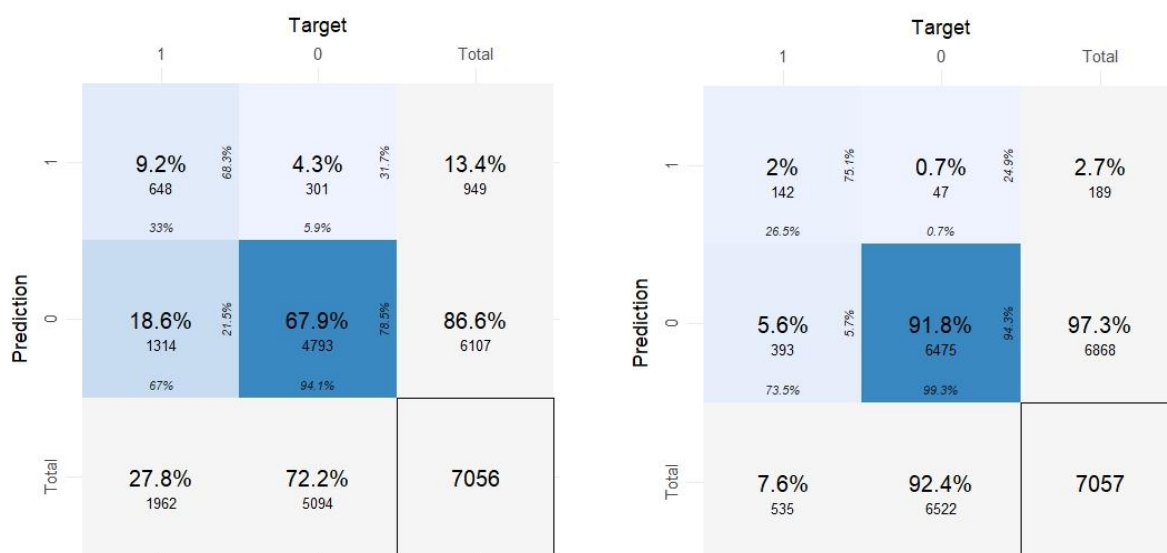
5.3.3 Μοντέλο Τυχαίων Δασών

Στη συγκεκριμένη ανάλυση εφαρμόστηκε η ταξινόμηση Random Forest, για την πρόβλεψη της εξαρτημένης μεταβλητής speeding_di, χρησιμοποιώντας διάφορες ανεξάρτητες μεταβλητές από το σύνολο δεδομένων.

Παρακάτω εξηγείται η διαδικασία αναλυτικά. Η συγκεκριμένη μέθοδος χρησιμοποιήθηκε λόγω της ικανότητας της να χειρίζεται **σύνθετες σχέσεις μεταξύ μεταβλητών**, να παρέχει αξιόπιστες προβλέψεις και να προσφέρει εργαλεία για την αξιολόγηση της σημασίας των χαρακτηριστικών.

Αρχικά, χρησιμοποιήθηκε η συνάρτηση `randromForest()`, για να δημιουργηθεί το μοντέλο, με εξαρτημένη και ανεξάρτητες μεταβλητές όπως έχουν ήδη οριστεί νωρίτερα. Ο αριθμός των δέντρων (`ntrree`), τέθηκε σε 500, προκειμένου να διασφαλιστεί ότι το μοντέλο θα έχει επαρκή αριθμό επαναλήψεων για να παράγει σταθερά και αξιόπιστα αποτελέσματα.

Στην συνέχεια με την συνάρτηση `print(rf_model)`, παρέχονται οι βασικές πληροφορίες για το μοντέλο και με την `plot(rf_model)` παρουσιάζεται το σφάλμα του, το οποίο βοηθά στην εκτίμηση της βελτίωσης της απόδοσης με την αύξηση των δέντρων. Επίσης, χρησιμοποιήθηκε η συνάρτηση `predict()`, για να γίνουν οι προβλέψεις στο σύνολο δεδομένων και ο πίνακας σύγχυσης (`confusion Matrix`), για την αξιολόγηση της απόδοσης του μοντέλου. Η οπτικοποίηση του πίνακα έγινε με τη χρήση της συνάρτησης `plot_confusion_matrix()`.



Πίνακας 5.11-12: Πίνακες αξιολόγησης απόδοσης Σεναρίου 1 και 2 (Confusion Matrix)

Τέλος, η συνάρτηση `varImpPlot()`, παρουσιάζει γραφικά τη σχετική σημασία κάθε μεταβλητής στο μοντέλο, υποδεικνύοντας ποιες είχαν την μεγαλύτερη συμβολή στην πρόβλεψη της `speeding_di`.

Τα αποτελέσματα είναι τα εξής:

Random Forest	Σενάριο 1	Σενάριο 2
Ορθότητα	0.784	0.945
Ευαισθησία	0.409	0.392
Ειδικότητα	0.928	0.989

Πίνακας 5.13: Αποτελέσματα βασικών στατιστικών δεικτών

Η **ορθότητα** (Accuracy) και στα δυο σενάρια βρίσκεται σε ικανοποιητικά επίπεδα με σημαντική διαφορά μεταξύ τους. Στο δεύτερο συγκεκριμένα κυμαίνεται σε πολύ υψηλό επίπεδο της τάξης του 95% περίπου και φανερώνει ότι το μοντέλο είναι πολύ πιο

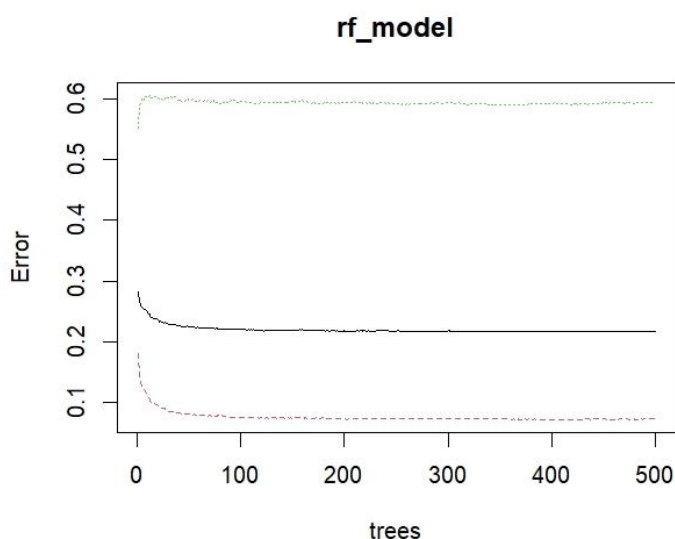
αποτελεσματικό. Αυτό μπορεί να οφείλεται σε βελτιώσεις στο σύνολο δεδομένων, στην επιλογή χαρακτηριστικών ή στις παραμέτρους του μοντέλου.

Η **ευαισθησία** (Sensitivity) και στα δύο σενάρια παραμένει στα ίδια επίπεδα, περίπου στο 40%. Το χαμηλό αυτό ποσοστό υποδεικνύει ότι το μοντέλο δυσκολεύεται να εντοπίσει σωστά τις περιπτώσεις υπέρβασης της ταχύτητας (false negatives). Αυτό μπορεί να σημαίνει ότι οι θετικές περιπτώσεις δεν αντιπροσωπεύονται επαρκώς ή ότι το μοντέλο προτιμά να αποφεύγει τους ψευδούς θετικούς.

Η **ειδικότητα** (Specificity) είναι υψηλή και στα δύο σενάρια, κάτι που φανερώνει ότι το μοντέλο αποδίδει πολύ καλά στην πρόβλεψη των αρνητικών περιπτώσεων. Η βελτίωση στο Σενάριο 2 είναι σημαντική και υποδηλώνει καλύτερη ικανότητα του μοντέλου να αποφεύγει θετικές προβλέψεις.

Συνολικά, το **δεύτερο σενάριο δείχνει πιο ισχυρό** από άποψη συνολικής ακρίβειας και ειδικότητας, **όμως η χαμηλή ευαισθησία δείχνει ότι απαιτείται περαιτέρω βελτίωση** για ισορροπημένη απόδοση.

Στη συνέχεια, παρακάτω **δίνεται η γραφική παράσταση** που δείχνει την μεταβολή **του σφάλματος του μοντέλου**, καθώς αυξάνεται ο αριθμός των δέντρων (trees) που δημιουργούνται.

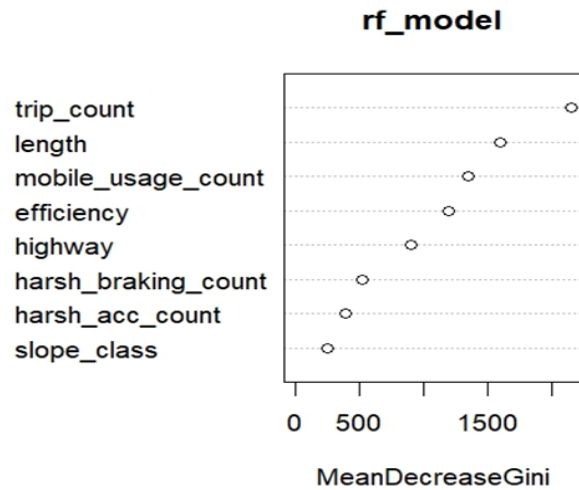


Διάγραμμα 5.1: Διάγραμμα Συσχέτισης Μεταβλητών

Στον οριζόντιο άξονα φαίνονται οι **αριθμοί των δέντρων** που δημιουργήθηκαν στο μοντέλο και στον κατακόρυφο το **σφάλμα ταξινόμησης**, δηλαδή το ποσοστό των περιπτώσεων που το μοντέλο δεν προέβλεψε σωστά. Παρατηρείται ότι **το συνολικό σφάλμα μειώνεται γρήγορα**, καθώς ο αριθμός των δέντρων αυξάνεται και σταθεροποιείται περίπου μετά τα 150 δέντρα. Αυτό δείχνει ότι η αύξηση των δέντρων πέρα από αυτό το σημείο δεν προσφέρει σημαντική βελτίωση στο μοντέλο. Επομένως, **η επιλογή των 500 δέντρων είναι ασφαλής**, αλλά η σταθεροποίηση του σφάλματος μετά τα 150 δέντρα δείχνει ότι και λιγότερα δέντρα θα ήταν αποδοτικά.

Στην συνέχεια, δίνονται τα **διαγράμματα σημαντικότητας των μεταβλητών** της ταξινόμησης Random Forest, που χρησιμοποιούνται για να αξιολογηθεί η συνεισφορά κάθε μεταβλητής στην ακρίβεια του μοντέλου κάθε σεναρίου. Η σημαντικότητα μετράται συνήθως με βάση την μείωση της αβεβαιότητας (MeanDecreaseGini), ή της απόκλισης.

Γενικά, οι υψηλές τιμές υποδεικνύουν ότι η μεταβλητή έχει μεγαλύτερη επιρροή στην πρόβλεψη του εξαρτημένου μεταβλητού παράγοντα, δηλαδή της υπέρβασης της ταχύτητας, ενώ αντίστοιχα οι χαμηλές τιμές το αντίστροφο. Η ανάλυση αυτή παρέχει μια βαθύτερη κατανόηση του ρόλου κάθε παράγοντα και βοηθάει στον εντοπισμό των κρίσιμων μεταβλητών.

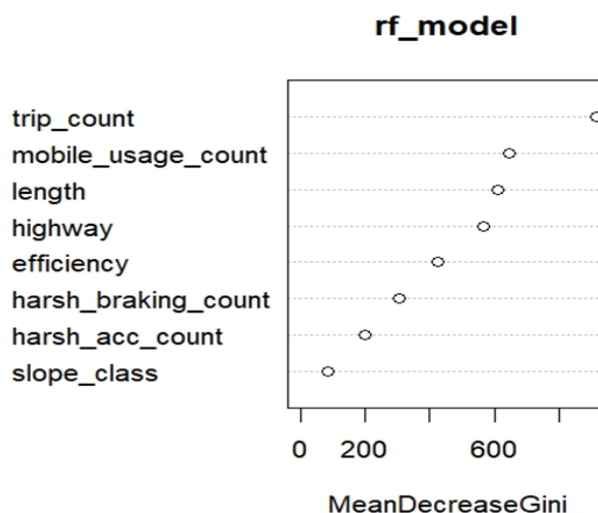


Διάγραμμα 5.2: Σημαντικότητα Μεταβλητών Σεναρίου 1

Σενάριο 1

Στο παραπάνω διάγραμμα εμφανίζονται με σειρά σημαντικότητας οι μεταβλητές που επηρεάζουν το μοντέλο. Αρχικά, **η μεταβλητή του αριθμού των διαδρομών (trip_count), καταγράφει την υψηλότερη τιμή με διαφορά**, φανερώνοντας ότι ο αριθμός των ταξιδιών αποτελεί τον πιο κρίσιμο παράγοντα.

Ακόμη, οι μεταβλητές: μήκος οδού (length), περιπτώσεις χρήσης κινητού τηλεφώνου (mobile_usage_count), ευθυγραμμία οδού (efficiency) και τύπος οδού (highway), εμφανίζουν επίσης υψηλή σημαντικότητα και συνεισφορά στο μοντέλο. Ενώ, οι μεταβλητές των περιπτώσεων απότομης επιβράδυνσης (harsh_braking_count), περιπτώσεων απότομης επιτάχυνσης (harsh_acc_count) και κατηγορίας κλίσης οδού (slope_class), έχουν χαμηλότερη σημαντικότητα συγκριτικά με τις υπόλοιπες μεταβλητές και δεν είναι τόσο απαραίτητες στην πρόβλεψη.



Διάγραμμα 5.3: Σημαντικότητα Μεταβλητών Σεναρίου 2

Σενάριο 2

Αντίστοιχα, στο παραπάνω διάγραμμα εμφανίζονται με σειρά σημαντικότητας οι μεταβλητές που επηρεάζουν το μοντέλο για το δεύτερο σενάριο. **Η μεταβλητή του αριθμού των διαδρομών (trip_count) παραμένει η πιο σημαντική**, υπογραμμίζοντας την σταθερή σημασία του αριθμού ταξιδιών στην πρόβλεψη.

Στη συνέχεια, οι μεταβλητές των περιπτώσεων χρήσης κινητού τηλεφώνου (mobile_usage_count), μήκους διαδρομής (length) και τύπου οδού (highway) εμφανίζουν αυξημένη σημαντικότητα σε παρόμοια επίπεδα με το προηγούμενο σενάριο και έτσι επιβεβαιώνεται η σχέση τους με την υπέρβαση της ταχύτητας. Οι μεταβλητές της ευθυγραμμίας (efficiency) και των περιπτώσεων απότομης επιβράδυνσης (harsh_braking_count), παρουσιάζουν μέτρια συνεισφορά, ενώ οι μεταβλητές των περιπτώσεων απότομης επιτάχυνσης (harsh_acc_count) και κατηγορίας κλίσης οδού (slope_class) καταγράφουν την χαμηλότερη σημαντικότητα και σε αυτό το σενάριο.

Συνολικά, παρατηρείται ότι η κατάταξη των μεταβλητών είναι **παρόμοια μεταξύ των δύο σεναρίων, με ορισμένες διαφορές** στη σημαντικότητα συγκεκριμένων παραμέτρων. Συγκεκριμένα, στο σενάριο 2 παρατηρείται αυξημένη σημαντικότητα της μεταβλητής της χρήσης κινητού τηλεφώνου, ενώ εκείνες του μήκους οδού και της αποδοτικότητας έχουν ελαφρώς μικρότερη συνεισφορά συγκριτικά με το σενάριο 1. Αριθμητικά κυμαίνονται στα ίδια επίπεδα, επομένως **οι διαφορές δεν είναι σημαντικές**. Αξιοσημείωτο είναι επίσης ότι και στα δύο σενάρια **οι παράγοντες που σχετίζονται με τις απότομες επιταχύνσεις και την κλίση οδού, έχουν ελάχιστη επιρροή**, γεγονός που δείχνει μικρή συσχέτιση με την υπέρβαση της ταχύτητας, συγκριτικά με τις υπόλοιπες.

Συνεπώς, η ανάλυση δείχνει ότι **οι παράμετροι του αριθμού των διαδρομών (trip_count) και των περιπτώσεων χρήσης κινητού τηλεφώνου κατά την οδήγηση (mobile_usage_count) παίζουν τον πιο κρίσιμο ρόλο στην πρόβλεψη της υπέρβασης ταχύτητας**, ανεξαρτήτως σεναρίου. Ωστόσο, η διαφοροποίηση στη σημαντικότητα των υπόλοιπων παραμέτρων υποδηλώνει ότι διαφορετικές συνθήκες ή δείγματα δεδομένων μπορεί να επηρεάζουν τη συνεισφορά κάθε παράγοντα.

5.3.4 Μοντέλο Παλινδρόμησης Μηχανή Διανυσμάτων Υποστήριξης (SVM)

Στο κεφάλαιο αυτό, εφαρμόστηκε η μέθοδος Support Vector Machine (SVM), η οποία είναι ικανή να ταξινομεί δεδομένα σε προβλήματα δυαδικών πολύ-ταξινόμησης και να προσφέρει ευελιξία στην διαχείριση μη γραμμικών σχέσεων μέσω πυρήνων (kernels). Το συγκεκριμένο μοντέλο εκπαιδεύτηκε με την χρήση του radial kernel, και στην συνέχεια υπολογίστηκαν τα αποτελέσματα μέσω της συνάρτησης `summary(svm_model)`, που παρουσίασε πληροφορίες, όπως ο αριθμός των υποστηρικτικών διανυσμάτων (support vectors) αλλά και οι παράμετροι εκπαίδευσης.

Τα αποτελέσματα αυτά δίνονται στον πίνακα παρακάτω:

Παράμετροι	SVM
SVM-Type	C-classification
SVM-Kernel	Radial
Cost	1
Number of SV	13286
Number of Classes	2

Πίνακας 5.14: Αποτελέσματα μοντέλου SMV

Ο **τύπος του μοντέλου** (SVM-Type) υποδεικνύει την κατηγορία που ανήκει, συγκεκριμένα εδώ την 'C', η οποία αντιπροσωπεύει την παράμετρο κόστους (Cost) και ρυθμίζει τον βαθμό στον οποίο το μοντέλο επιτρέπει αποκλίσεις κατά την ταξινόμηση (errors). Ένα μικρότερο C μπορεί να οδηγήσει σε απλό μοντέλο (υπερβολική γενίκευση), ενώ ένα μεγαλύτερο C μπορεί να μειώσει τα σφάλματα αλλά να οδηγήσει σε υπερ-προσαρμογή (overfitting).

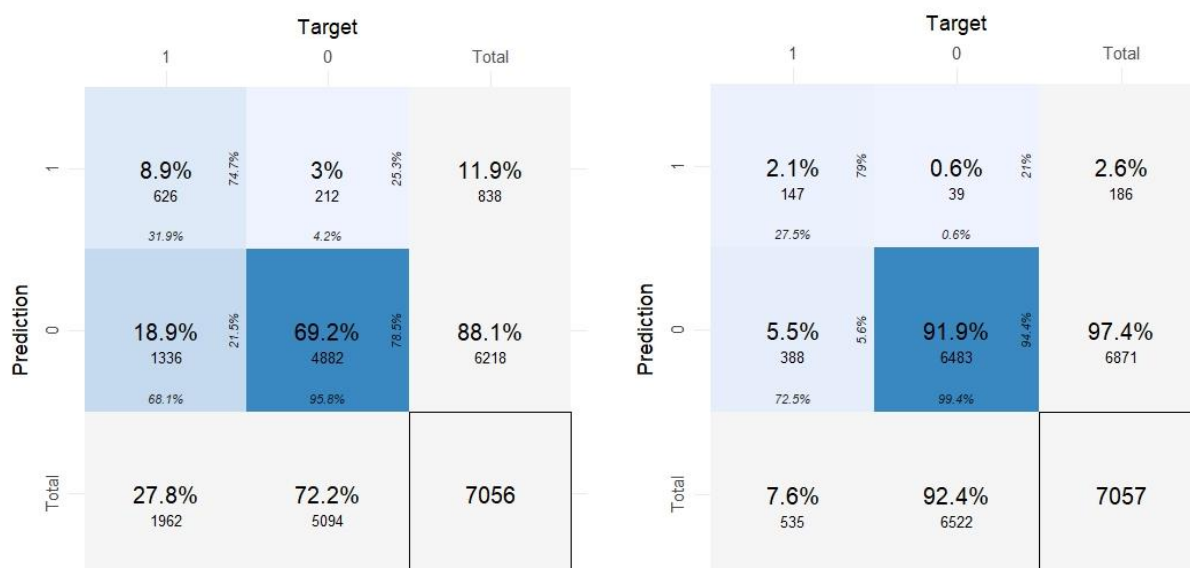
Ο **τύπος του πυρήνα** (SVM – Kernel) 'radial kernel', είναι ένας μη γραμμικός πυρήνας που επιτρέπει τη μοντελοποίηση σύνθετων σχέσεων και για δεδομένα τα οποία όπως προαναφέρθηκε δεν μπορούν να διαχωριστούν γραμμικά.

Η **παράμετρος του κόστους** (Cost), ρυθμίζει την ισορροπία μεταξύ της μεγιστοποίησης του περιθωρίου (margin) και της ελαχιστοποίησης των σφαλμάτων ταξινόμησης. Η τιμή 1 είναι μια μέτρια τιμή, η οποία φανερώνει ότι το μοντέλο προσπαθεί να επιτύχει καλή ισορροπία, χωρίς να δίνει υπερβολική έμφαση στα δεδομένα εκπαίδευσης ή στην γενίκευση.

Ο **αριθμός των διανυσμάτων στήριξης** (Number of SV) που χρησιμοποίησε το μοντέλο ήταν 13.286. Το δεδομένο αυτό καθορίζει τα όρια της απόφασης (decision boundary). Συγκεκριμένα είναι 6.634 για την κλάση 0 (μη υπέρβαση) και 6.652 για την κλάση 1 (υπέρβαση καθορισμένου ορίου), δηλαδή περίπου ίσος αριθμός και για τις δύο κατηγορίες. Ο αριθμός είναι σχετικά υψηλός, κάτι που υποδηλώνει ότι τα δεδομένα είναι σύνθετα ή ότι υπάρχει ανάγκη για περαιτέρω προ-επεξεργασία.

Ο **αριθμός των κλάσεων** (Number of Classes) , όπως έχουν ήδη αναφερθεί είναι 2 για την υπέρβαση και μη υπέρβαση των ορίων ταχύτητας. Από την αξιολόγηση φάνηκε ότι οι δύο κλάσεις είναι σαφώς ορισμένες και το μοντέλο τις διαχειρίζεται εξίσου καλά, δίνοντας δίκαιο και ισορροπημένο αποτέλεσμα.

Στην συνέχεια, χρησιμοποιήθηκαν οι προβλέψεις του μοντέλου για το σύνολο δοκιμών (test set) και δημιουργήθηκε ο πίνακας σύγχυσης για το κάθε σενάριο. Τα ευρήματα από την τελική ανάλυση αναδεικνύουν την αποτελεσματικότητα του μοντέλου και δίνονται παρακάτω.



Πίνακες 5.15-16: Πίνακες αξιολόγησης απόδοσης Σεναρίου 1 και 2 (Confusion Matrix)

SVM	Σενάριο 1	Σενάριο 2
Ορθότητα	0.781	0.939
Ευαισθησία	0.319	0.275
Ειδικότητα	0.958	0.994

Πίνακας 5.17: Αποτελέσματα βασικών στατιστικών δεικτών

Όμοια με προηγουμένως, η **ορθότητα** (Accuracy) στο πρώτο σενάριο έχει μια ικανοποιητική τιμή περίπου στο 78%, χωρίς όμως να εξασφαλίζει απόλυτα τη σωστή απόδοση. Στο δεύτερο σενάριο αυξάνεται σημαντικά στο 94% περίπου, βελτιώνοντας την συνολική του απόδοση.

Η **ευαισθησία** (Sensitivity) εξακολουθεί να κινείται σε χαμηλά επίπεδα. Στο πρώτο σενάριο το μοντέλο ανίχνευσε το 31,91% των θετικών περιπτώσεων και στο δεύτερο μειώθηκε ακόμη περισσότερο στο 27,48%.

Η **ειδικότητα** (Specificity) ανιχνεύει αρχικά το 95,83% των αρνητικών περιπτώσεων και στη συνέχεια το 99,40%, που σημαίνει ότι σχεδόν όλες οι περιπτώσεις ανιχνεύονται σωστά. Το μοντέλο φαίνεται να εστιάζει υπερβολικά στις αρνητικές τιμές, κάτι που μπορεί να αποτελέσει πρόβλημα αν οι θετικές περιπτώσεις είναι πιο σημαντικές.

Συνολικά, για άλλη μια φορά **και στα δύο σενάρια η ευαισθησία είναι χαμηλή**, ενώ η ειδικότητα είναι ικανοποιητική και ταξινομεί σωστά τις αρνητικές τιμές. Το σενάριο 1 παρουσιάζει μια πιο ισορροπημένη συμπεριφορά μεταξύ ευαισθησίας και ειδικότητας, αν και η συνολική του ακρίβεια είναι χαμηλότερη. Είναι πιο κατάλληλο στην περίπτωση που δίνεται μεγαλύτερη έμφαση στην ανίχνευση θετικών περιπτώσεων. Από την άλλη **το δεύτερο σενάριο υπερέχει σημαντικά σε ορθότητα, αλλά σε βάρος της ευαισθησίας** και επομένως είναι καταλληλότερο αν οι αρνητικές περιπτώσεις έχουν μεγαλύτερη σημασία. Επομένως, η επιλογή του κατάλληλου σεναρίου εξαρτάται από το πρόβλημα που εξετάζεται και από το αν προτεραιότητα είναι η ευαισθησία ή η ορθότητα.

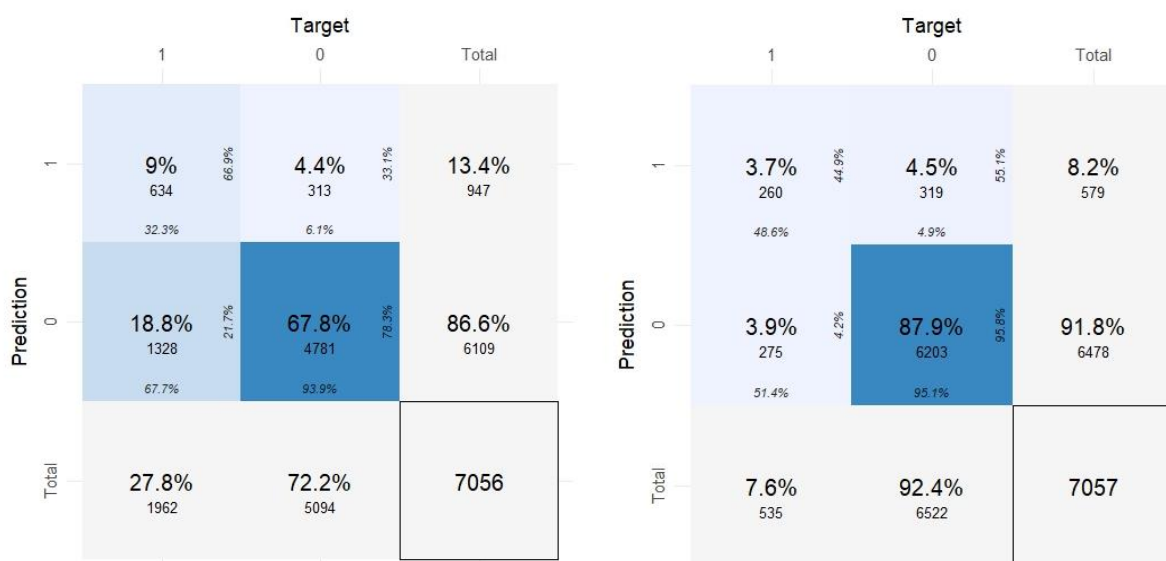
5.3.5 Μοντέλο Παλινδρόμησης Naive Bayes

Η τελευταία μέθοδος που εφαρμόστηκε ήταν η Naive Bayes, που είναι ιδιαίτερα χρήσιμη για τις περιπτώσεις όπου οι ανεξάρτητες μεταβλητές έχουν κατηγορηματική μορφή ή οι παραδοχές της ανεξαρτησίας είναι αποδεκτές. Στη συγκεκριμένη ανάλυση, η μέθοδος επιλέχθηκε για την πρόβλεψη της πιθανότητας υπέρβασης των ορίων ταχύτητας, καθώς παρέχει ταχύτητα υπολογισμού και δυνατότητα χρήσης σε δεδομένα με πολλές κατηγορίες.

Για την εφαρμογή της μεθόδου, δημιουργήθηκε αρχικά το μοντέλο (nb_model) μέσω της συνάρτησης *naiveBayes* από την βιβλιοθήκη *e1071*. Αμέσως, παρουσιάστηκε η περίληψη του, με τη χρήση της εντολής *summary(nb_model)*, στην οποία εμφανίζονται τα στατιστικά στοιχεία που βοηθούν στη βαθύτερη κατανόηση του. Στη συνέχεια, το μοντέλο αυτό χρησιμοποιήθηκε για την πρόβλεψη στο σύνολο δοκιμών (test), με την βοήθεια της συνάρτησης *predict*.

Για την αξιολόγηση της απόδοσης του μοντέλου, δημιουργήθηκε ο πίνακας σύγχυσης (Confusion Matrix), μέσω της εντολής *confusionMatrix(pred, test\$speeding_di, positive='1')* και οπτικοποιήθηκε με την *plot_confusion_matrix*, η οποία παρέχει μια γραφική απεικόνιση που αναδεικνύει τις σωστές προβλέψεις στη διαγώνιο και τα σφάλματα στους υπόλοιπους πίνακες.

Τα αποτελέσματα αυτά δίνονται στον παρακάτω πίνακα:



Πίνακας 5.18-19: Πίνακες αξιολόγησης απόδοσης Σεναρίου 1 και 2 (Confusion Matrix)

Naive Bayes	Σενάριο 1	Σενάριο 2
Ορθότητα	0.767	0.916
Ευαισθησία	0.323	0.486
Ειδικότητα	0.938	0.951

Πίνακας 5.20: Αποτελέσματα βασικών στατιστικών δεικτών

Η **ορθότητα** (Accuracy) του μοντέλου είναι 76,74% στο πρώτο σενάριο, υποδεικνύοντας ότι ταξινομεί σωστά περισσότερα από τα τρία τέταρτα των περιπτώσεων. Στο δεύτερο σενάριο αυξάνεται αρκετά στο 91,58% και αυτό δείχνει μια καλύτερη συνολική απόδοση και βελτίωση.

Η **ευαισθησία** (Sensitivity) ανέρχεται αρχικά στο 32,31% και για ακόμη μία φορά η χαμηλή αυτή τιμή δείχνει αδυναμία στην ανίχνευση των θετικών περιστατικών. Στη συνέχεια στο Σενάριο 2 για πρώτη φορά ως τώρα, φτάνει στο ποσοστό 48,59% και φανερώνει σαφώς βελτιωμένη ανίχνευση στο μισό σχεδόν των περιπτώσεων.

Η **ειδικότητα** (Specificity), είναι πολύ υψηλή και στα δύο σενάρια, γύρω στο 95% και συνεπώς συνεχίζει να διατηρεί την υψηλή απόδοση στις αρνητικές περιπτώσεις εντοπίζοντας 'τες σχεδόν όλες στο σύνολο.

Επομένως, **το Σενάριο 2 υπερέχει ξεκάθαρα ως προς όλες τις παραμέτρους και αποτελεί ένα πιο ισορροπημένο μοντέλο.** Όπως και προηγουμένως, και αυτό το μοντέλο είναι πολύ καλό στον εντοπισμό των περιπτώσεων χωρίς υπέρβαση ταχύτητας. Ωστόσο, η χαμηλή ευαισθησία τονίζει ότι το μοντέλο εξακολουθεί να δυσκολεύεται να εντοπίσει με ακρίβεια όλες τις περιπτώσεις υπέρβασης ταχύτητας και ενδέχεται να χρειάζεται περαιτέρω βελτιστοποίηση. Βέβαια η διαφορά στην ευαισθησία του δεύτερου σεναρίου είναι αξιοσημείωτη καθώς για πρώτη φορά είναι σε ένα ποσοστό τόσο υψηλό. Αυτό επιβεβαιώνεται εξίσου και με την **τιμή Kappa**, η οποία **δείχνει μέτρια συμφωνία** μεταξύ των προβλέψεων του μοντέλου και των πραγματικών τιμών, ενώ η **τιμή στο test McNemar, βρέθηκε πολύ χαμηλή** και στα δύο σενάρια και υποδηλώνει την στατιστική διαφορά μεταξύ των πραγματικών και των

προβλεπόμενων τιμών. Οι χαμηλές αυτές τιμές καθορίζουν την απαίτηση για βελτίωση.

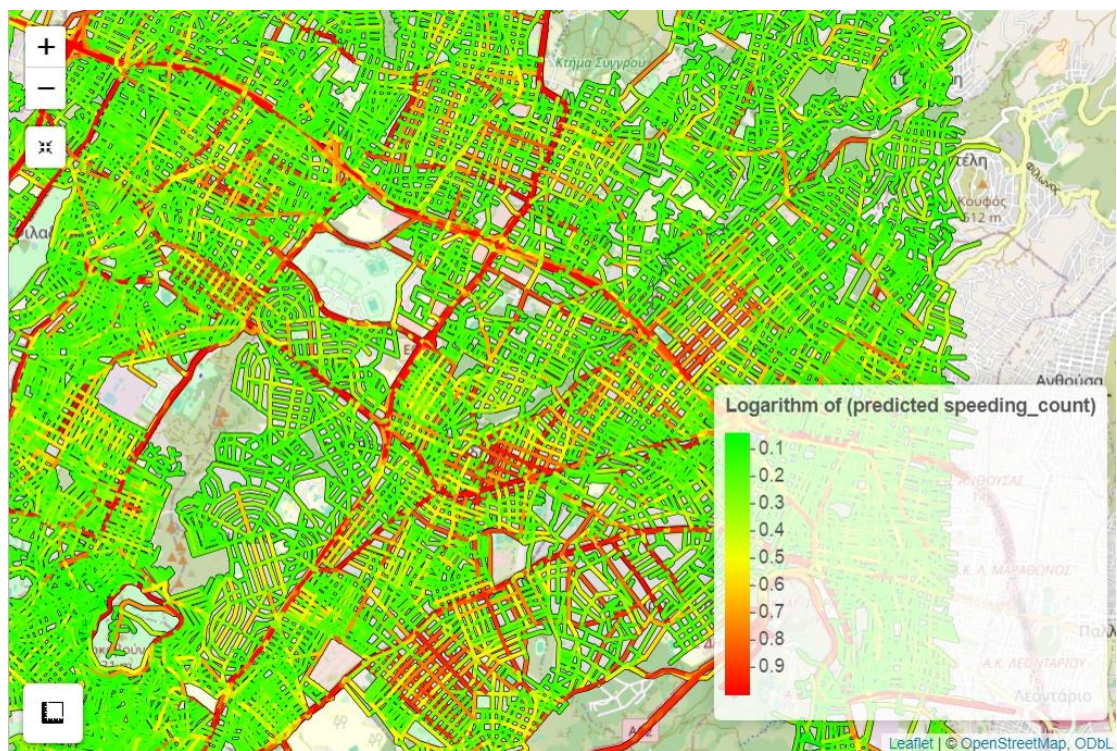
5.4 Απεικόνιση Χάρτη

Ως τελευταίο βήμα υλοποιείται μια **διαδραστική χαρτογραφική απεικόνιση των αποτελεσμάτων** του στατιστικού μοντέλου που επιλέχθηκε από τα προηγούμενα υποκεφάλαια. Η απεικόνιση αυτή στοχεύει στην γεωγραφική αποτύπωση της πιθανότητας εμφάνισης παραβάσεων υπέρβασης των ορίων ταχύτητας στο σύνολο του οδικού δικτύου της Αθήνας.

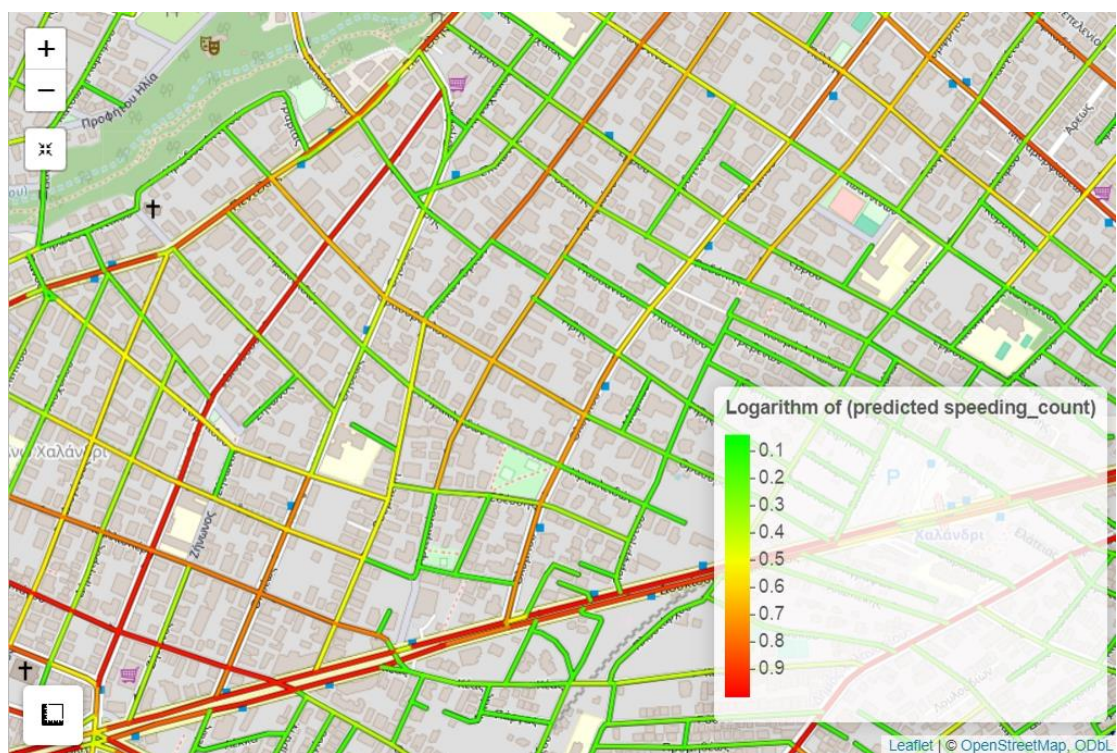
Αρχικά, έγινε η εισαγωγή και επεξεργασία των χωρικών δεδομένων (*GeoJSON*) και η μετατροπή τους σε αντικείμενα *sf*, ώστε να επιτρέπεται η χωρική επεξεργασία και απεικόνισή τους, μέσω της βιβλιοθήκης *leaflet*. Για την απεικόνιση **επιλέχθηκε η μέθοδος Δυαδικής Λογιστικής Παλινδρόμησης σε συνδυασμό με το Σενάριο 1**, καθώς κρίθηκε ότι προσφέρει τον βέλτιστο συνδυασμό ερμηνευσιμότητας και πρακτικής αξίας. Η συγκεκριμένη μέθοδος επιτρέπει την εκτίμηση των πιθανοτήτων **με βάση γραμμικούς συνδυασμούς μεταβλητών**, έτσι ώστε να υπάρχει μια σαφής εικόνα για την συνεισφορά του κάθε παράγοντα στον κίνδυνο εμφάνισης παραβάσεων. Άλλα μοντέλα, όπως τα Τυχαία Δάση και η SVM, εμφάνισαν καλύτερες επιδόσεις. Όμως η **ευκολία ενσωμάτωσης των αποτελεσμάτων** στην χαρτογραφική απεικόνιση που προσφέρει η Δυαδική Λογιστική Παλινδρόμηση κρίθηκε σημαντικότερη. Επιπλέον, συγκριτικά με τα δύο σενάρια παρατηρείται ότι το πρώτο σενάριο επιτρέπει την αναγνώριση όλων των σημείων, στα οποία έχει καταγραφεί οποιασδήποτε διάρκειας υπέρβαση των ορίων ταχύτητας. Ενώ, **το δεύτερο σενάριο** περιορίζει σημαντικά τον αριθμό των καταγεγραμμένων περιπτώσεων και πιθανώς **αποκρύπτει κρίσιμα σημεία κινδύνου**. Συνεπώς, το Σενάριο 1 προσφέρει μια πληρέστερη εικόνα της χωρικής κατανομής του φαινομένου στο οδικό δίκτυο, ώστε να ληφθούν τελικά στοχευμένα μέτρα οδικής ασφάλειας.

Στην συνέχεια γίνεται η αποτύπωση του τελικού χάρτη, με την **χρήση χρωματικής κλίμακας από πράσινο**, το οποίο φανερώνει χαμηλή πιθανότητα, **έως κόκκινο**, το οποίο φανερώνει υψηλή πιθανότητα, καθιστώντας εφικτό τον εντοπισμό των πιο επικίνδυνων οδικών περιοχών. Παράλληλα, παρέχονται ενσωματωμένες δυνατότητες εμφάνισης πληροφοριών ανά οδικό τμήμα, μέσω αναδυόμενων παραθύρων (*popups*), καθώς και εργαλεία πλοήγησης και μέτρησης αποστάσεων. Η προσέγγιση αυτή καθιστά τα αποτελέσματα προσβάσιμα με οπτικοποιημένο και κατανοητό τρόπο και μπορεί να αξιοποιηθεί τόσο για ερευνητικούς σκοπούς όσο και για χάραξη στοχευμένων πολιτικών οδικής ασφάλειας.

Παρακάτω παρατίθενται ενδεικτικές εικόνες από τον διαδραστικό χάρτη που δημιουργήθηκε στο πλαίσιο της παρούσας ανάλυσης:



Εικόνα 5.2: Απεικόνιση πιθανοτήτων υπέρβασης των ορίων ταχύτητας



Εικόνα 5.3: Απεικόνιση πιθανοτήτων υπέρβασης των ορίων ταχύτητας

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1 Σύνοψη Αποτελεσμάτων

Στόχος της παρούσας Διπλωματικής Εργασίας είναι η **εξέταση των παραγόντων επιρροής της υπέρβασης των ορίων ταχύτητας στο οδικό δίκτυο της Αθήνας**. Η έρευνα αξιοποιεί δεδομένα που συλλέχθηκαν από την εταιρία OSeven Telematics και το OpenStreetMap σχετικά με δείκτες μορφολογικών χαρακτηριστικών οδών, δείκτες μέτρων ασφάλειας αλλά και συμπεριφοράς των οδηγών. Πιο συγκεκριμένα, επιδιώκεται να βελτιωθεί η οδική ασφάλεια, μέσα από την δημιουργία ενός εργαλείου που θα αναγνωρίζει τους βασικούς παράγοντες που επηρεάζουν την υπέρβαση των ορίων ταχύτητας και κατ' επέκταση την πρόκληση οδικών ατυχημάτων και θα επιτρέπει την λήψη στοχευμένων μέτρων.

Στα προηγούμενα κεφάλαια, πραγματοποιήθηκε μια συστηματική ανάλυση για την επίτευξη του στόχου αυτού. Αρχικά, έγινε η **βιβλιογραφική ανασκόπηση** (Κεφάλαιο 2), η οποία εστίασε στους παράγοντες που επηρεάζουν την υπέρβαση των ορίων ταχύτητας και την πιθανότητα οδικού ατυχήματος με βάση σύγχρονες ερευνητικές προσεγγίσεις.

Στη συνέχεια, περιεγράφηκαν τα **στατιστικά εργαλεία και οι μέθοδοι ανάλυσης** (Κεφάλαιο 3) που χρησιμοποιήθηκαν, καθώς και τα κριτήρια αξιολόγησης των μοντέλων. Ειδικότερα, παρουσιάστηκαν οι βασικές αρχές της στατιστικής, όπως είναι η ανάλυση της παλινδρόμησης, οι οποίες αξιοποιήθηκαν για τη μοντελοποίηση των δεδομένων, αλλά και οι διαδικασίες αξιολόγησης των μεταβλητών για τον έλεγχο της αξιοπιστίας και της εγκυρότητας των αποτελεσμάτων. Επιπλέον, παρουσιάστηκε και το θεωρητικό υπόβαθρο των αλγορίθμων μηχανικής μάθησης που αξιοποιήθηκαν.

Έπειτα, ακολούθησε η **συλλογή και επεξεργασία των δεδομένων** (Κεφάλαιο 4), όπου χρησιμοποιήθηκαν πραγματικά δεδομένα οδηγικής συμπεριφοράς, τα οποία συλλέχθηκαν μέσω της εταιρίας OSeven Telematics. Τα δεδομένα αυτά, αντλήθηκαν από αισθητήρες κινητών τηλεφώνων και περιλάμβαναν κρίσιμες μεταβλητές, οι οποίες στην συνέχεια φιλτραρίστηκαν και υποβλήθηκαν σε επεξεργασία για την εξαγωγή των κρίσιμων δεικτών οδικής ασφάλειας. Επιπλέον, συλλέχθηκαν και επεξεργάστηκαν δεδομένα σχετικά με τα γεωμετρικά χαρακτηριστικά των οδών του εξεταζόμενου οδικού δικτύου αντλώντας πληροφορίες μέσω του OpenStreetMaps.

Τέλος, στο κεφάλαιο της ανάλυσης των δεδομένων (Κεφάλαιο 5), εφαρμόστηκαν **στατιστικές μέθοδοι για την διερεύνηση των σχέσεων** μεταξύ των παραμέτρων και αναπτύχθηκαν τα τελικά μοντέλα. Πρώτα, έγινε η **επιλογή των βασικών μεταβλητών** μέσα από τον έλεγχο συσχέτισης και αμέσως πραγματοποιήθηκε περιγραφική στατιστική ανάλυση των βασικών δεικτών. Μετά από κατάλληλη επεξεργασία, εφαρμόστηκαν οι επιλεγμένες στατιστικές μέθοδοι παλινδρόμησης και μηχανικής μάθησης για να διερευνηθεί η επίδραση κάθε παραμέτρου στην επιρροή της υπέρβασης των ορίων ταχύτητας από τους οδηγούς, ενώ παράλληλα δοκιμάστηκαν διαφορετικοί συνδυασμοί μεταβλητών για την ανάπτυξη των μοντέλων. **Τα μοντέλα που επιλέχθηκαν ήταν πέντε και χωρίστηκαν σε δύο σενάρια ελέγχου**. Το πρώτο επικεντρώθηκε στον έλεγχο παρουσίας ή απουσίας της υπέρβασης των ορίων ταχύτητας, ενώ το δεύτερο αφορούσε στην ανάλυση της υπέρβασης με ένα συγκεκριμένο όριο. Η εφαρμογή των μοντέλων σε κάθε σενάριο επέτρεψε την διερεύνηση της επίδρασης διαφορετικών τύπων συμπεριφοράς ταχύτητας στα αποτελέσματα της ανάλυσης και βελτίωση της ακρίβειας πρόβλεψης των μοντέλων.

Παρακάτω δίνεται ο συγκεντρωτικός πίνακας των αποτελεσμάτων των αναλύσεων για τις μεθόδους που χρησιμοποιήθηκαν και για τα δύο σενάρια:

Σενάριο 1					
(0 = δεν καταγράφηκε υπέρβαση των ορίων ταχύτητας, 1 = καταγράφηκε υπέρβαση των ορίων ταχύτητας τουλάχιστον ενός δευτερολέπτου)					
	Binary Logistic	Random Forest	Decision Tree	SVM	Naive Bayes
Ορθότητα	0.778	0.784	0.771	0.781	0.767
Ευαισθησία	0.355	0.409	0.330	0.319	0.323
Ειδικότητα	0.942	0.928	0.941	0.958	0.938

Πίνακας 6.1: Συγκεντρωτικός πίνακας αποτελεσμάτων Σεναρίου 1

Σενάριο 2					
(0 = καταγράφηκε υπέρβαση των ορίων ταχύτητας μικρότερη των 15.55 δευτερολέπτων, 1 = καταγράφηκε υπέρβαση των ορίων ταχύτητας τουλάχιστον 15.55 δευτερολέπτων)					
	Binary Logistic	Random Forest	Decision Tree	SVM	Naive Bayes
Ορθότητα	0.941	0.945	0.938	0.939	0.916
Ευαισθησία	0.355	0.392	0.265	0.275	0.486
Ειδικότητα	0.989	0.989	0.993	0.994	0.951

Πίνακας 6.2: Συγκεντρωτικός πίνακας αποτελεσμάτων Σεναρίου 2

6.2 Συνολικά Συμπεράσματα

Κατά τα διάφορα στάδια εκπόνησης της Διπλωματικής Εργασίας προέκυψε μια **σειρά συμπερασμάτων** που συνδέεται άμεσα με τους αρχικούς στόχους και τα ερωτήματα της. Στο παρόν υποκεφάλαιο επιχειρείται να δοθεί μια απάντηση σε αυτά, με την σύνθεση των αποτελεσμάτων των προηγούμενων κεφαλαίων. Τα σημαντικότερα συμπεράσματα συνοψίζονται ως εξής:

- Η διεθνής βιβλιογραφία ανέδειξε ότι, η **υπέρβαση των ορίων ταχύτητας** αποτελεί αδιαμφισβήτητο **κρίσιμο παράγοντα που συνδέεται άμεσα με την πιθανότητα ατυχήματος και τη σοβαρότητα των τραυματισμών**. Βρέθηκε ακόμη, ότι η αύξηση της ταχύτητας μειώνει τον χρόνο αντίδρασης και αυξάνει τη σφοδρότητα της σύγκρουσης.
- Η υπέρβαση των ορίων ταχύτητας **συνδέεται θετικά με μεταβλητές όπως, ο αριθμός διαδρομών ανά εξεταζόμενο οδικό τμήμα (trip_count) και το μήκος οδικού τμήματος (length)**. Αυτό εξηγείται από το γεγονός ότι τόσο ο αριθμός διαδρομών ανά οδικό τμήμα, όσο και το μήκος του τμήματος μπορούν να θεωρηθούν δείκτες έκθεσης στον κίνδυνο.
- Παρατηρείται **αρνητική συσχέτιση της χρήσης κινητού τηλεφώνου κατά την οδήγηση με την παράμετρο της υπέρβασης των ορίων ταχύτητας**. Επομένως, η χρήση κινητού μπορεί να σχετίζεται με πιο προσεκτική οδήγηση, λόγω της απαίτησης ταυτόχρονων πολλαπλών κινήσεων και της αντιστάθμισης του κινδύνου. Ωστόσο, η απόσπαση της προσοχής ενδέχεται να προκαλεί άλλες επικίνδυνες συμπεριφορές.
- Η **υπέρβαση των ορίων ταχύτητας σχετίζεται θετικά με τις απότομες επιβραδύνσεις και τις απότομες επιταχύνσεις**. Η θετική συσχέτιση μεταξύ τους, οφείλεται στις αυξημένες απαιτήσεις αντίδρασης που προκύπτουν σε

υψηλές ταχύτητες. Αναλυτικά, όταν το όχημα κινείται γρηγορότερα από το επιτρεπτό όριο, ο οδηγός έχει λιγότερο χρόνο να προσαρμοστεί στις συνθήκες της οδού, οδηγώντας συχνά σε απότομες επιβραδύνσεις για την αποφυγή κινδύνων. Παράλληλα, η υπερβολική ταχύτητα συχνά σχετίζεται με επιθετική οδήγηση, που περιλαμβάνει ξαφνικές επιταχύνσεις για τη διατήρηση υψηλών ταχυτήτων ή την προσπέραση.

- Οι απότομες επιταχύνσεις και επιβραδύνσεις συνδέονται με αυξημένο κίνδυνο υπέρβασης ταχύτητας στο πρώτο σενάριο, ενώ στο δεύτερο οι επιταχύνσεις δεν παραμένουν εξίσου σημαντικές. Αυτό φανερώνει ότι **η επιθετική οδήγηση είναι πιο καθοριστική για μικρές υπερβάσεις**. Συγκεκριμένα, η διαφορά στη συσχέτιση μεταξύ των δύο σεναρίων δείχνει ότι οι συμπεριφορές αυτές, συνδέονται κυρίως με σύντομες υπερβάσεις ταχύτητας, καθώς η επιθετική οδήγηση οδηγεί σε στιγμιαίες αλλαγές ταχύτητας. Αντίθετα, στις παρατεταμένες υπερβάσεις, οι επιταχύνσεις χάνουν τη σημασία τους, υποδεικνύοντας ότι αυτές σχετίζονται περισσότερο με σταθερές συμπεριφορές, όπως η συστηματική αδιαφορία για τα όρια ταχύτητας ή η επίδραση εξωτερικών παραγόντων, όπως τα χαρακτηριστικά της οδού.
- Ο ταξινομητής **Random Forest** σύμφωνα με τον δείκτη MeanDecreaseGini, έδειξε ότι και στα δύο σενάρια την **πιο κρίσιμη μεταβλητή αποτελεί ο αριθμός διαδρομών ανά εξεταζόμενο οδικό τμήμα** και σταθερά σημαντικές εξίσου παραμένουν το μήκος οδικού τμήματος αλλά και η χρήση κινητού τηλεφώνου κατά την οδήγηση. Τον χαμηλότερο ρόλο διατηρεί η μεταβλητή της κατηγορίας κλίσης της οδού.
- Ο ταξινομητής **Random Forest** παρουσίασε τις καλύτερες επιδόσεις ως προς την πρόβλεψη της υπέρβασης των ορίων ταχύτητας και στα δύο σενάρια που εξετάστηκαν.
- Παρατηρείται σταθερά χαμηλή ευαισθησία σε όλες τις μεθόδους και τα σενάρια. Γεγονός το οποίο, υποδηλώνει **δυσκολία στον εντοπισμό των θετικών περιπτώσεων**, πιθανώς λόγω ανισορροπίας στα δεδομένα, ή περιορισμένων χαρακτηριστικών που διαφοροποιούν το δείγμα.
- Συγκριτικά, **στο Σενάριο 2 όλες οι μέθοδοι έχουν υψηλότερη ακρίβεια σε σχέση με το Σενάριο 1**. Επομένως, το σενάριο αυτό είναι πιο αποτελεσματικό, αφού οδηγεί σε καλύτερη συνολική απόδοση για την εκτίμηση του αρχικού προβλήματος.

6.3 Προτάσεις Βελτίωσης

Σύμφωνα με τα αποτελέσματα και τα συνολικά συμπεράσματα που εξήχθησαν κατά την εκπόνηση της Εργασίας αυτής, ακολουθείται παρακάτω μια **σειρά προτάσεων**, που στοχεύουν στην αξιοποίηση των αποτελεσμάτων και στη βελτίωση του επιπέδου της οδικής ασφάλειας.

- **Αυστηρότερα μέτρα επιτήρησης της ταχύτητας:** Η ενίσχυση των ελέγχων για τις περιπτώσεις υπέρβασης του ορίου ταχύτητας, ιδιαίτερα σε αυτοκινητόδρομους και κύριους δρόμους, και η αναθεώρηση των υπάρχοντων ορίων σε περιοχές με υψηλά ποσοστά ατυχημάτων, θα βοηθήσει στην μείωση

αυτών. Τέτοιες τεχνολογίες θα μπορούσαν να είναι, αυτόματοι σταθμοί μέτρησης ταχύτητας αλλά και κάμερες ασφαλείας.

- **Εκπαίδευση για ασφαλή οδήγηση:** Η υλοποίηση εκπαιδευτικών προγραμμάτων σε κατάλληλες ηλικίες που επικεντρώνονται στην αποφυγή της επιθετικής οδήγησης, μπορεί να βελτιώσουν την συνολική οδηγική συμπεριφορά.
- **Στοχευμένη παρέμβαση στις περιοχές υψηλού κινδύνου:** Η ανάπτυξη υποδομών σε δρόμους ταχείας κυκλοφορίας και η προσθήκη απαραίτητων σημάνσεων και φυσικών εμποδίων, θα συμβάλει στη μείωση της ταχύτητας και στην αύξηση της ασφαλείας.
- **Εφαρμογή τεχνολογιών αιχμής:** Η αξιοποίηση νέων τεχνολογιών, όπως τα έξυπνα συστήματα βοηθητικής οδήγησης (ADAS) και η χρήση τεχνητής νοημοσύνης για την πρόληψη των ατυχημάτων, μπορεί να βελτιώσει την ασφάλεια στους δρόμους. Τέτοια συστήματα παρέχουν προειδοποιήσεις και λαμβάνουν αυτόματα μέτρα για την αποφυγή σύγκρουσης.
- **Προγράμματα επιβράβευσης:** Η αξιοποίηση και ένταξη εφαρμογών (app) για την παρακολούθηση της συμπεριφοράς του οδηγού και τη συγκέντρωση βαθμολογιών (score), μπορεί να δώσει κίνητρο στους οδηγούς για ασφαλέστερη οδήγηση μέσω εκπνώσεων σε ασφαλιστικά προγράμματα. Έτσι θα ενισχυθεί η υπευθυνότητα των οδηγών.

6.4 Προτάσεις για Περαιτέρω Έρευνα

Για την επέκταση της παρούσας Διπλωματικής μου Εργασίας και την περαιτέρω διερεύνηση του αντικειμένου της, ενδιαφέρον παρουσιάζουν οι ακόλουθες προτάσεις:

- **Επέκταση του συνόλου δεδομένων:** Να συμπεριληφθεί ένα ευρύτερο φάσμα πληθυσμού, ώστε να ληφθούν υπόψιν γεωγραφικές και κοινωνικές ιδιαιτεροτήτων αλλά και να βρεθούν πιθανές συνδέσεις.
- **Διερεύνηση σε άλλες πόλεις:** Να πραγματοποιηθεί παρόμοια έρευνα σε διαφορετικές ελληνικές πόλεις, προκειμένου να εξεταστούν οι ιδιαιτερότητες στις συγκοινωνιακές υποδομές και η νοοτροπία σε σύγκριση με την περιοχή της Αθήνας.
- **Ανάλυση συμπεριφοράς σε πραγματικό χρόνο:** Να γίνει εφαρμογή συστημάτων που παρακολουθούν την οδηγική συμπεριφορά σε πραγματικό χρόνο, ώστε να αξιολογηθεί η επίδρασή τους στη μείωση των ατυχημάτων και στη βελτίωση της οδικής ασφαλείας.
- **Μελέτη οδηγών μοτοσυκλετών:** Να εξεταστούν οι παράγοντες που σχετίζονται με την υπέρβαση του ορίου ταχύτητας, όσον αφορά τους ευάλωτους χρήστες της οδού, δηλαδή τους οδηγούς μοτοσυκλετών. Η ανάλυση θα μπορούσε να εστιάσει σε παρόμοιες μεταβλητές με την παρούσα έρευνα αλλά και σε επικίνδυνους ελιγμούς.

- **Διερεύνηση της επίδρασης άλλων παραγόντων:** Να μελετηθούν επιπλέον παράμετροι που επηρεάζουν την υπέρβαση του ορίου ταχύτητας και την αύξηση των ατυχημάτων, όπως οι καιρικές συνθήκες, οι διάφοροι τύποι οχημάτων, οι συνθήκες οδήγησης (π.χ. πρωινές βραδινές ώρες) κ.ά.
- **Επίδραση της χρήσης κινητού τηλεφώνου κατά την οδήγηση:** Να αναλυθεί λεπτομερώς η αρνητική συσχέτιση μεταξύ της χρήσης κινητού και της παραμέτρου speeding. Θα μπορούσε να πραγματοποιηθεί μια μελέτη για την επίδραση διαφορετικών ειδών χρήσης (π.χ. κλήσεις, γραπτά μηνύματα), στην συμπεριφορά των οδηγών, καθώς και για τις πιθανές επιπτώσεις στην διάσπαση προσοχής και σε επικίνδυνες ενέργειες.

7. ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Australian Bureau of Statistics, (n.d). Types of Error: Sampling and Non-Sampling Errors.
2. Beheshti, N. (2022). Random Forest Regression: A basic explanation. Towards Data Science.
3. Brian Beers, (2024). P-Value: What it is, How to calculate it and Examples. Investopedia.
4. Charles M. Farmer, Sheila G. Klauer, Julie A. McClafferty, Feng Guo, (2014). Relationship of Near-Crash/Crash Risk to Time Spent on a Cell Phone While Driving.
5. Christina Harvey, (2023). Accuracy and Precision in Data: Definition, Formula and Examples.
6. Chris Jurewicz, Amir Sobhani, Jeremy Woolley, Jeff Dutschke, Bruce Corbe, (2016). Driver speed selection and crash risk: Insights from the naturalistic driving study.
7. Davis, G. A., et al. (in press, 2006). Speed as a Risk Factor in Serious Run-Off-Road Crashes: Bayesian Case-Control Analysis with Case Speed Uncertainty. Accident Analysis & Prevention.
8. Department for Transport, (2019). Official Statistics: Drugs in reported road fatalities in Great Britain, data to 2022.
9. Elvik, R., (2013). Speed and Road Safety: Synthesis of Evidence from Evaluation Studies. Transportation Research Record: Journal of the Transportation Research Board, 2402(1), 59-69.
10. European Commission, (2022). Road safety in the EU: fatalities in 2021 remain well below pre-pandemic level.
11. European Commission, (2022). Road safety: European Commission rewards effective initiatives and publishes 2021 figures on road fatalities.
12. European Transport Safety Council, (2019). Educating Speeding In Europe PIN Flash Report 36.
13. Exploration of Vehicle Impact Speed – Injury Severity Relationships for Application in Safer Road Design.
14. Hany M. Hassan, Mohamed Shawky, Mohammad Kishta, Atef Garib, Hussain A. Al-Harthei, (2016). Road safety evidence review: Understanding the role of speeding and speed in serious crash trauma: A case study of New Zealand.
15. Harlene Labrum, (2022). Sudden Unintended Acceleration and Car Accidents.
16. Iraj Goli Khatir, Masoud Shayesteh Azar, Zakaria Zakariaei, Hamid Reza Rezaee Rad, Kimia Rasouli, (2022). The prevalence of substance use among drivers with traffic injuries in Mazandaran Province.
17. Jenine K Harris, (2021). Primer on binary logistic regression. National Center for Biotechnology Information (NCBI).
18. Jocelyn Mackie, (2023). 7 Types of Car Accidents and how to avoid them. Lewis and Clark Law School.
19. John, G. H., & Langley, P., (1995). Naive Bayes for Regression.
20. Juan Diego Febres, Susana García-Herrero, Sixto Herrera, J. M. Gutiérrez, J. R. López-García & Miguel A. Mariscal, (2020). Influence of seat-belt use on the severity of injury in traffic accidents. European Transport Research Review.
21. Karen Steward, (2023). Sensitivity vs. Specificity. Technology Networks.
22. Letty Aarts, Ingrid van Schagen, (2005). Driving speed and the risk of road crashes: A review.

23. Liu, C., Chen, C, (2009). An Analysis of Speeding-Related Crashes: Definitions and the Effects of Road Environments. National Highway Traffic Safety Administration.
24. McEvoy S.P. Stevenson M.R., McCartt A.T., Woodward M., Haworth C., Palamara P., Cercarelli, R. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study.
25. MacEvoy, S.P., Stevenson, M.R., Woodward, M. (2007). The prevalence of, and factors associated with, serious crashes involving a distracting activity. *Accident Analysis and Prevention* 39, 475-482.
26. MathWorks, (2019). Understanding Support Vector Machine Regression: Mathematical formulation of SVM.
27. Matthews, R., Legg, S., & Charlton, S., 2003. The effect of cell phone type on drivers' subjective workload during concurrent driving and conversing. *Accident Analysis and Prevention* 35, 441-450.
28. Miguel Perez, Edie Sears, Jacob Valente, Wenyan Huang, Jeremy Sudweeks, (2021). Factors modifying the likelihood of speeding behaviors based on naturalistic driving data.
29. Mohammad Manouchehrifar, Roghaieh Ghalandari, Hamidreza Hatamabadi, (2019). The Impact of Safety Equipment on Traffic Accidents: A Hospital Based Study. *Trauma Monthly Magazine*.
30. Munden, J. M., (1967). The Relation Between a Driver's Speed and His Accident Rate. Road Research Laboratory Report No. 20.
31. National Technical University of Athens, Road Safety Observatory, (2024). Slight decrease in road fatalities in 2023, Greece.
32. Nilsson, G. (1982). The Effects of Speed Limits on Traffic Accidents in Sweden. Proceedings of the International Symposium on the Effects of Speed Limits on Traffic Accidents and Transport Energy Use.
33. Nilsson, G., (2004). Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety. Lund Institute of Technology, Department of Technology and Society, Traffic Engineering.
34. OSeven Telematics, (2024). Data retrieved from the official website of OSeven Telematics. Retrieved from <https://oseven.io/>.
35. Prasad, A., (2021). Regression Trees, Decision Tree for Regression, Machine Learning: How can Regression Trees be used for solving regression problems?. Analytics Vidhya.
36. Raha Hamzeie, Peter T. Savolainen, Timothy J. Gates, (2017). Investigation of drivers' behavior towards speeds using crash data and self-reported questionnaire.
37. Rasifaghihi, N., (2023). From Theory to Practice: Implementing Support Vector Regression for Predictions in Python. Towards Data Science.
38. RF Soames Job, Colin Brodie, (2022). Addressing endogeneity in modeling speed enforcement, crash risk and crash severity simultaneously.
39. Science Direct, (2024). Binary Logistic Regression.
40. Shamsunnahar Yasmin, Naveen Eluru, Md. Mazharul Haque, (2022). An investigation of the speeding-related crash designation through crash narrative reviews sampled via logistic regression.
41. Solomon, D, (1964). Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle. U.S. Department of Commerce/Bureau of Public Roads.
42. Stephen L Brown, Amy Cotton, (2003). Risk-mitigating beliefs, risk estimates, and self-reported speeding in a sample of Australian drivers.
43. Steve G. Gabany, Portia Plummer, Pat Grigg, (1998). Why drivers speed: The speeding perception inventory.
44. Sun, T.-J., Liu, S.-J., et.al. (2020). Influence of road types on road traffic accidents in northern Guizhou Province, China.

45. Tatsuaki Osafune, Toshimitsu Takahashi, Noboru Kiyama, Tsuneo Sobue, Hirozumi Yamaguchi & Teruo Higashino, (2016). Analysis of Accident Risks from Driving Behaviors.
46. Taylor, M. C., Lynam, D. A., Baruya, A. (2002). The Effects of Drivers' Speed on the Frequency of Road Accidents. TRL Report 421.
47. The Editors of Encyclopedia Britannica, (n.d.). Deceleration Injury.
48. Tian-Jing Sun, Si-Jia Liu, et.al. (2020). Influence of road types on road traffic accidents in northern Guizhou Province, China.
49. World Health Organization, (2023). Global status report on road safety 2023.
50. Yang Ding, Xiaohua Zhao, Yiping Wu, et.al, (2023). How psychological factors affect speeding behavior: Analysis based on an extended theory of planned behavior in a Chinese sample.
51. Yannis George, (2014). Distracted driving and mobile phone use: Overview of impacts and countermeasures.
52. Zijun Liang, Yun Xiao, (2020). Analysis of factors influencing expressway speeding behavior in China.
53. Ελληνική Αστυνομία Υπουργείο Προστασίας του Πολίτη, (2023). Στατιστική Επετηρίδα Έτους 2023.
54. Ινστιτούτο Οδικής Ασφάλειας (Ι.Ο.ΑΣ.) «Πάνος Μυλωνάς», (2017). Ενημερωτική καμπάνια για την υπερβολική ταχύτητα και την Αντιπροσωπεία της Ευρωπαϊκής Επιτροπής στην Ελλάδα.