



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ  
ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ  
ΥΠΟΔΟΜΗΣ

Διπλωματική Εργασία

## «Ανίχνευση γενικευμένων κυκλοφοριακών μοτίβων σε αστικά δίκτυα»



Μπεφάνη Καλλιόπη

Επιβλέπουσα Καθηγήτρια: Βλαχογιάννη Ελένη, Καθηγήτρια Ε.Μ.Π

Αθήνα, Οκτώβριος 2024

## **ΕΥΧΑΡΙΣΤΙΕΣ**

Ολοκληρώνοντας τις σπουδές μου στο Εθνικό Μετσόβιο Πολυτεχνείο θα ήθελα αρχικά να ευχαριστήσω την κυρία Ελένη Βλαχογιάννη, Καθηγήτρια της Σχολής Πολιτικών Μηχανικών Ε.Μ.Π, αρχικά γιατί μέσω της διδασκαλίας της με έκανε να αγαπήσω την κατεύθυνση του Συγκοινωνιολόγου και έπειτα για την ανάθεση και καθοδήγησή της κατά την συγγραφή της παρούσας Διπλωματικής Εργασίας.

Ευχαριστώ, επίσης, θερμά τον κύριο Παναγιώτη Φαφουτέλλη, Διδάκτορα και Ερευνητή του Ε.Μ.Π, για όλες του τις συμβουλές και την πολύτιμη βοήθειά του καθ' όλη τη διάρκεια της εκπόνησης αυτής της εργασίας.

Τέλος, ευχαριστώ πολύ την οικογένειά μου, γιατί δεν θα τα είχα καταφέρει χωρίς εκείνους, και τους φίλους μου, που έγιναν οικογένεια μέσα από αυτό το ταξίδι.

Μπεφάνη Καλλιόπη

Αθήνα, Οκτώβριος 2024

Τίτλος: «Ανίχνευση γενικευμένων κυκλοφοριακών μοτίβων σε αστικά δίκτυα»

Μπεφάνη Καλλιόπη

Επιβλέπουσα Καθηγήτρια: Ελένη Βλαχογιάννη

## Σύνοψη

Η κυκλοφοριακή συμφόρηση αποτελεί ένα από τα κυριότερα ζητήματα της λειτουργίας των αυτοκινητοδρόμων. Σκοπός της παρούσας διπλωματικής εργασίας αποτελεί ο εντοπισμός και η ανάλυση χωροχρονικών μοτίβων κορεσμού σε αστικά οδικά δίκτυα με τη μέθοδο Dynamic Time Warping (DTW) που ανήκει στην κατηγορία της μη επιβλεπόμενης μάθησης και είναι εξαιρετικά αποτελεσματική ως μέτρο ομοιότητας χρονοσειρών. Η προτεινόμενη προσέγγιση εφαρμόζεται σε κυκλοφοριακά δεδομένα από το δίκτυο φωρατών της Αθήνας για την περίοδο Φεβρουαρίου-Μαρτίου 2023. Από την ανάλυση προκύπτουν 3 διαφορετικά μοτίβα με διαφορετική διακύμανση ωριαίων φόρτων ανά ημέρα (μικρή, αναμενόμενη και μεγάλη ημερήσια διακύμανση). Τα αποτελέσματα δείχνουν ότι στο Κέντρο της Αθήνας και στα Νότια Προάστια εξαιτίας της λειτουργίας του δικτύου στην ικανότητα αρκετές ώρες την ημέρα, τα μοτίβα κυκλοφορίας που εντοπίζονται είναι συνεχώς πάνω από το μέσο μοτίβο ημερήσιας χρονικής εξέλιξης ωριαίου φόρτου και εμφανίζουν μεγαλύτερη διακύμανση. Στις υπόλοιπες περιοχές οι τιμές που καταγράφηκαν ήταν οι αναμενόμενες, δηλαδή κοντά στις φυσιολογικές για τις ώρες εντός και εκτός αιχμής. Είναι προφανές ότι η σε βάθος ανάλυση των μοτίβων σε συγκεκριμένες περιοχές θα μπορούσε να χρησιμοποιηθεί ως εργαλείο διερεύνησης της έντασης και διάρκειας του κορεσμού, καθώς και για μελλοντική ανάλυση της χωροχρονικής εξέλιξης της κυκλοφορίας.

Title: “Detection of generalized traffic patterns in urban networks”

Befani Kalliopi

Thesis Supervising Professor: Eleni Vlahogianni

## **ABSTRACT**

Congestion is one of the main issues in the operation of motorways. The aim of this thesis is to identify and analyze spatio-temporal saturation patterns in urban road networks using the Dynamic Time Warping (DTW) method, which belongs to the category of unsupervised learning and is highly effective as a measure of time series similarity. The proposed approach is applied to traffic data from the Athens traffic light network for the period February-March 2023. The analysis results in 3 different patterns with different variation of hourly traffic volumes per day (small, expected and large daily variation). The results show that in the center of Athens and in the southern suburbs, due to the operation of the network at capacity for several hours per day, the traffic patterns detected are consistently above the average pattern of daily hourly load evolution and show a higher variation. In the remaining areas the values recorded were as expected, meaning close to normal for on- and off-peak hours. It is apparent that an in-depth analysis of the patterns in specific areas could be used as a tool to investigate the intensity and duration of saturation, as well as for future analysis of the spatio-temporal evolution of traffic.

## Περιεχόμενα

1.	Εισαγωγή.....	5
1.1	Τι είναι η κυκλοφοριακή συμφόρηση.....	5
1.2	Το πρόβλημα της κυκλοφοριακής συμφόρησης στις μεγάλες πόλεις .....	6
1.3	Το πρόβλημα της κυκλοφοριακής συμφόρησης στην Αθήνα.....	7
1.4	Σκοπός της Διπλωματικής Εργασίας .....	9
1.5	Διάρθρωση της Διπλωματικής Εργασίας.....	10
2.	Βιβλιογραφική Ανασκόπηση .....	11
2.1	Ανάλυση και Πρόβλεψη της Κυκλοφορίας .....	11
2.2	Συμπεράσματα Βιβλιογραφικής Ανασκόπησης.....	17
3.	Μεθοδολογική Προσέγγιση.....	18
3.1	Εισαγωγή .....	18
3.2	Ο αλγόριθμος Dynamic Time Warping .....	19
3.3	Σύγκριση μεθόδων Dynamic Time Warping και Euclidean.....	23
3.4	Ομαδοποίηση χρονοσειρών με DTW και k-medoids .....	24
3.5	Αξιολόγηση Ομαδοποιήσεων .....	25
4.	Ανάλυση Δεδομένων και Αποτελέσματα.....	26
4.1	Συλλογή Δεδομένων .....	26
4.2	Επεξεργασία δεδομένων .....	27
4.3	Ανάλυση Αποτελεσμάτων.....	29
4.3.1	Παρουσίαση Αποτελεσμάτων της μεθόδου DTW.....	30
4.4	Οπτικοποίηση Αποτελεσμάτων σε Χάρτες .....	35
4.4.1	Κέντρο Αθηνών .....	36
4.4.2	Ανατολικά και Βόρεια Προάστια (Λ. Κηφισίας & Λ. Μεσογείων).....	37
4.4.3	Νότια Προάστια (Λ. Βουλιαγμένης & Λ. Ποσειδώνος).....	38
4.4.4	Δυτικά Προάστια (Λ. Κηφισού & Πειραιάς) .....	39
5.	Συμπεράσματα και Προτάσεις.....	40
5.1	Βασικά Συμπεράσματα.....	40
5.2	Περιορισμοί Έρευνας .....	41
5.3	Προτάσεις για περαιτέρω έρευνα .....	41
	ΒΙΒΛΙΟΓΡΑΦΙΑ .....	42
	ΠΑΡΑΡΤΗΜΑ .....	44

# 1. Εισαγωγή

## 1.1 Τι είναι η κυκλοφοριακή συμφόρηση

Ένα από τα κύρια ζητήματα της λειτουργίας των αυτοκινητοδρόμων είναι η κυκλοφοριακή συμφόρηση (traffic congestion). Οι χαμηλές ταχύτητες, οι μεγαλύτεροι χρόνοι ταξιδιού και η αύξηση των φαινομένων ουράς οχημάτων είναι χαρακτηριστικά της συμφόρησης, η οποία είναι ένα κυκλοφοριακό σενάριο που προκαλείται είτε λόγω αυξημένης ζήτησης είτε λόγω ειδικών γεγονότων (π.χ. ατυχήματα, καιρικές συνθήκες). Συγκεκριμένα, συμβαίνει όταν ο όγκος της κυκλοφορίας αυξάνεται σε σημείο που τα αυτοκίνητα εμπλέκονται μεταξύ τους και επιβραδύνουν τη ροή της κυκλοφορίας. Η ζήτηση πλησιάζει και συχνά ξεπερνά τη χωρητικότητα του οδοστρώματος καθώς αυξάνεται συνεχώς. Αυτή η κατάσταση, που συνήθως αναφέρεται ως "μποτιλιάρισμα", έχει συχνά ως αποτέλεσμα την πολύ αργή κίνηση έως και την πλήρη ακινητοποίηση των αυτοκινήτων για μεγάλα χρονικά διαστήματα.

Ορισμένοι παράγοντες μπορούν είτε να προκαλέσουν είτε να επιδεινώσουν τη συμφόρηση. Οι περισσότεροι από αυτούς είτε αυξάνουν τη ζήτηση, δηλαδή τον αριθμό των οχημάτων που απαιτούνται για τη μεταφορά μιας συγκεκριμένης ποσότητας επιβατών ή προϊόντων, είτε μειώνουν τη χωρητικότητα του οδικού δικτύου σε ένα συγκεκριμένο σημείο ή κατά μήκος ενός συγκεκριμένου μήκους του οδικού δικτύου. Οι καιρικές συνθήκες, τα έργα οδοποιίας σε όλο το δίκτυο και τα τροχαία ατυχήματα αποτελούν άλλες αιτίες συμφόρησης. Η έρευνα για τη κυκλοφοριακή συμφόρηση δεν δύναται να εντοπίσει εκ των προτέρων όλες τις συνθήκες που οδηγούν σε μποτιλιάρισμα. Στην πραγματικότητα έχει διαπιστωθεί ότι μεμονωμένα γεγονότα, όπως τα τροχαία ατυχήματα, μπορούν να έχουν αναπόφευκτες συνέπειες και να εξαπλωθούν, με αποτέλεσμα την επίμονη συμφόρηση σε σημεία και τμήματα όπου η κανονική κυκλοφορία θα μπορούσε να διατηρηθεί για μεγαλύτερο χρονικό διάστημα (Victoria Transport Policy Institute, 2012).

Οι επιπτώσεις της κυκλοφοριακής συμφόρησης είναι πολυάριθμες. Ως προς το χρήστη, παρατηρείται μειωμένη ψυχική ικανοποίηση και άγχος λόγω των απρόβλεπτων χρόνων ταξιδιού. Όταν μάλιστα πρόκειται να μεταβούν στην εργασία τους, πιθανή καθυστέρηση συνεπάγεται και σε μειωμένη παραγωγικότητα. Ο κίνδυνος ασφάλειας και τα οδικά ατυχήματα είναι άλλη μία σημαντική συνέπεια της συμφόρησης. Ως προς το περιβάλλον και την οικονομία, παρατηρείται αύξηση της ηχορύπανσης, της

κατανάλωσης καυσίμου και άρα των ατμοσφαιρικών ρύπων, αλλά και του κόστους του ταξιδιού. Η κυκλοφοριακή συμφόρηση και οι κοινωνικοοικονομικές και περιβαλλοντικές επιπτώσεις της έχουν αυξηθεί σε αστικές και αγροτικές περιοχές σε όλο τον κόσμο, ανεπτυγμένες ή μη, και θα συνεχίσει να επιδεινώνεται, θέτοντας αναμφισβήτητη απειλή για την ποιότητα της αστικής ζωής.



Εικόνα 1.1- Κυκλοφοριακή συμφόρηση στην Αθήνα

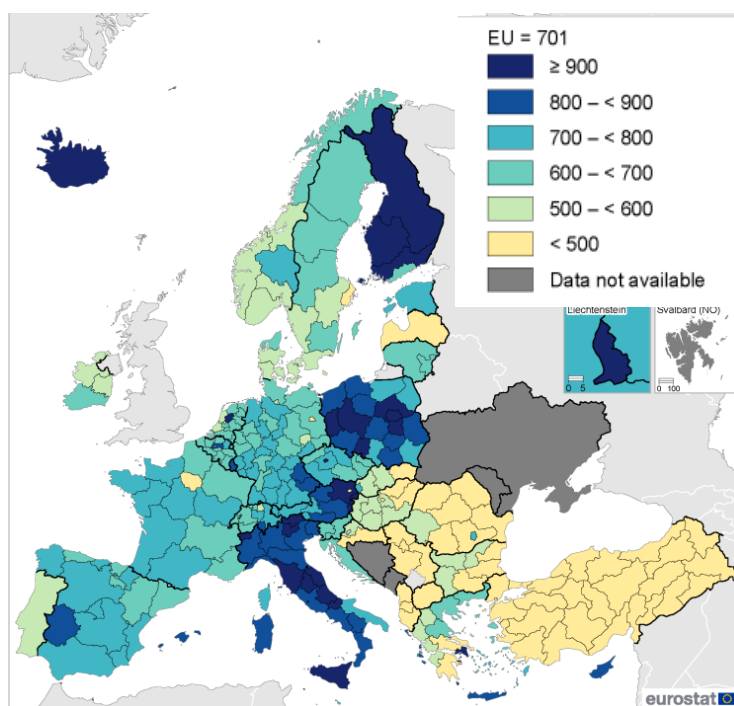
Πηγή: Eurokinissi/Yiannis Panagopoulos

## 1.2 Το πρόβλημα της κυκλοφοριακής συμφόρησης στις μεγάλες πόλεις

Η κυκλοφοριακή συμφόρηση στις μεγάλες πόλεις αποτελεί ένα από τα πιο σοβαρά και διαρθρωτικά προβλήματα που αντιμετωπίζει η σύγχρονη κοινωνία. Έναν από τους σημαντικότερους παράγοντες του προβλήματος αποτελεί η αύξηση του αριθμού των οχημάτων, που οφείλεται στην αύξηση του πληθυσμού και συνεπώς των αναγκών των χρηστών. Αρνητικά επηρεάζει και η ελλιπής δημόσια συγκοινωνία: οι πολίτες προτιμούν να χρησιμοποιήσουν το αυτοκίνητό τους αντί για ένα μη αξιόπιστο και μη αποτελεσματικό μέσο μαζικής μεταφοράς, κάτι που οδηγεί σε αυξημένη ζήτηση για μετακινήσεις στους δρόμους. Τέλος, πολλές πόλεις δεν έχουν σχεδιαστεί με γνώμονα την αυξανόμενη κυκλοφορία, οδηγώντας έτσι τα οδικά δίκτυα να μην μπορούν να ανταποκριθούν αποτελεσματικά στον αυξανόμενο αριθμό οχημάτων.

Υψηλότερα ποσοστά χρήσης Ι.Χ. αυτοκινήτου παρατηρούνται συχνά σε προαστιακές, αγροτικές και περιφερειακές περιοχές, αντανακλώντας, πιθανώς, την έλλειψη εναλλακτικών τρόπων εσωτερικής μεταφοράς επιβατών. Τα υψηλότερα ποσοστά ιδιοκτησίας αυτοκινήτων στην ΕΕ — τουλάχιστον 900 μηχανοκίνητα οχήματα ανά 1

000 κατοίκους το 2021 — απεικονίζονται στην *Εικόνα 2* με την πιο σκούρα απόχρωση του μπλε. Οι 23 περιοχές αυτής της ομάδας συγκεντρώθηκαν κυρίως στην Ιταλία (εννέα περιφέρειες), στην Πολωνία (πέντε περιφέρειες) και τη Φινλανδία (τέσσερις περιφέρειες). Περιλάμβανε επίσης δύο περιφέρειες από την Αυστρία, καθώς και την Αττική και την Πράγα, τις πρωτεύουσες περιοχές της Ελλάδας και της Τσεχίας, και τη Φλέβολαντ στην Ολλανδία.



Εικόνα 1.2- Ποσοστό Μηχανοκίνησης, 2021 (αριθμός οχημάτων ανά 1000 κατοίκους)

Πηγή: Eurostat ([tran\\_r\\_vehst](#)), ([demo\\_pjan](#)) and ([demo\\_r\\_d2jan](#))

### 1.3 Το πρόβλημα της κυκλοφοριακής συμφόρησης στην Αθήνα

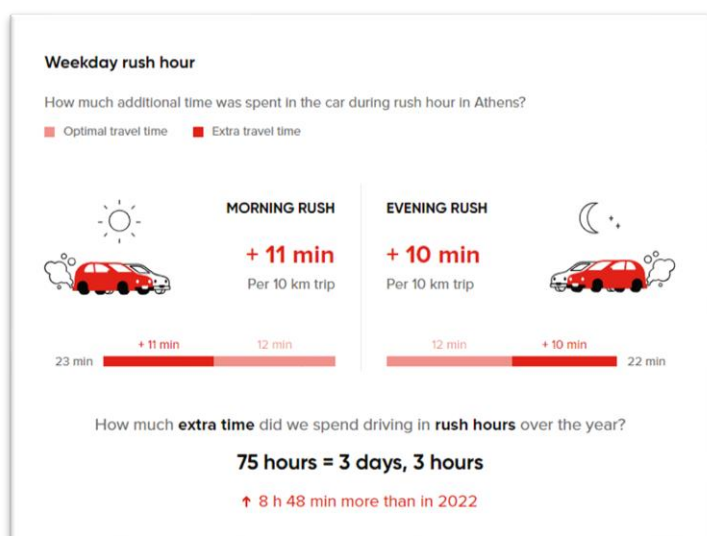
Η παρούσα Διπλωματική Εργασία επικεντρώνεται και πρόκειται να αναλύσει δεδομένα που αφορούν την περιοχή της Αττικής. Η ευρύτερη περιοχή της Αθήνας αποτελείται από σχεδόν 3,8 εκατομμύρια κατοίκους. Το οδικό δίκτυο του Δήμου Αθηναίων έχει συνολικό μήκος περίπου 1000 χιλιομέτρων και καλύπτει περίπου 39 τετραγωνικών χιλιομέτρων. Ο συνολικός αριθμός των αυτοκινήτων που κυκλοφορούν καθημερινά στην Αττική το 2020 εκτιμάται σε περίπου 3 εκατομμύρια . Από τη δεκαετία του 1960, όταν ξεκίνησε η μηχανοκίνηση, η κυκλοφορία στην Αθήνα αυξανόταν σταδιακά μέχρι το 2008 και την παγκόσμια οικονομική κρίση. Κατά τα



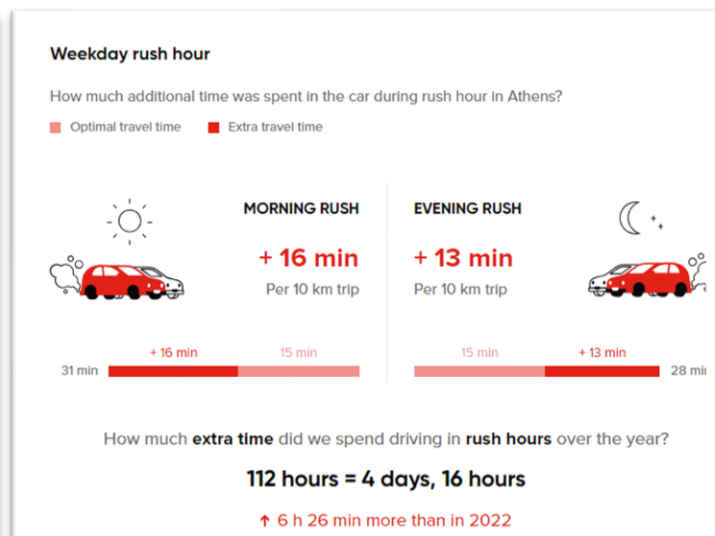
πρώτα χρόνια της ύφεσης, υπήρξε αξιοσημείωτη μείωση της χρήσης των ιδιωτικών αυτοκινήτων. Στη συνέχεια, υπήρξε μια ανάκαμψη της κυκλοφορίας το 2017 και μέχρι το 2019, η κυκλοφορία ήταν σχεδόν ακριβώς όπως ήταν πριν από το 2009. Υπήρξαν περαιτέρω μεταβολές στην κυκλοφορία των αυτοκινήτων ως αποτέλεσμα της επιδημίας του κορονοϊού και των αποκλεισμών που ακολούθησαν, μαζί με μια αξιοσημείωτη μείωση της χρήσης των δημόσιων μέσων μεταφοράς. Μετά από μια αξιοσημείωτη έξαρση της κυκλοφοριακής συμφόρησης τον Οκτώβριο του 2021, οι αρχές της Αθήνας αποφάσισαν να επαναφέρουν τους κυκλοφοριακούς περιορισμούς που ίσχυαν από τον Μάρτιο του 2020 ως αποτέλεσμα της πανδημίας.

Το 2023, ο μέσος χρόνος οδήγησης 10 χιλιομέτρων εντός της μητροπολιτικής περιοχής της Αθήνας ήταν 18 λεπτά και 20 δευτερόλεπτα, κατατάσσοντας την πόλη στην 43η θέση στον κόσμο, σταθερή σε σχέση με το 2022. Στο κέντρο της Αθήνας, ο χρόνος αυτός, συγκριτικά με το 2022, αυξήθηκε κατά 10 δευτερόλεπτα, στα 23 λεπτά και 10 δευτερόλεπτα. Η Πέμπτη 2 Μαρτίου, ήταν η χειρότερη ημέρα για να ταξιδέψει κανείς στην Αθήνα, με τα αυτοκίνητα να καλύπτουν τα 10 χιλιόμετρα, κατά μέσο όρο, σε 32 λεπτά και 10 δευτερόλεπτα. Ο μέσος οδηγός στην περιοχή της Αθήνας, το 2023, ξόδεψε 174 ώρες ταξιδεύοντας κατά τις ώρες αιχμής, εκ των οποίων οι 75 οφείλονταν σε κυκλοφοριακή συμφόρηση, για μια διαδρομή 10 χιλιομέτρων. Αυτό ισοδυναμεί με το χρόνο που απαιτείται για να διαβάσει κανείς 34 βιβλία. Ο χρόνος αυτός αυξάνεται σε 228 ώρες για το κέντρο της Αθήνας, εκ των οποίων 112 ώρες χάθηκαν λόγω της συμφόρησης και θα μπορούσαν να αντικατασταθούν με την ανάγνωση 45 βιβλίων. Οι ετήσιες εκπομπές CO<sub>2</sub> το 2023, στην περιοχή της Αθήνας, ήταν 905 kg (208 kg λόγω κυκλοφοριακής συμφόρησης), ενώ στο κέντρο της πόλης ήταν 992 kg (272 kg λόγω κυκλοφοριακής συμφόρησης). Για να γίνουν τα στοιχεία αυτά πιο κατανοητά στον αναγνώστη, θα χρειαζόταν η ετήσια ανάπτυξη 90 και 99 δέντρων, αντίστοιχα, για την απορρόφηση αυτών των εκπομπών. Τέλος, όσον αφορά το κόστος των καυσίμων, ο μέσος οδηγός χρησιμοποίησε βενζίνη για μια διαδρομή 10 χιλιομέτρων σε ώρα αιχμής και πλήρωσε 597€, εκ των οποίων τα 115€ χάθηκαν λόγω της συμφόρησης (στην περιοχή της Αθήνας) και 615€, εκ των οποίων τα 153€ χάθηκαν λόγω της συμφόρησης στο κέντρο της Αθήνας, αριθμοί που ισοδυναμούν με περίπου 7 δεξαμενές βενζίνης μέσων διαστάσεων.

Στις Εικόνα 3 και Εικόνα 4 παρατηρείται, συγκριτικά, ο επιπλέον χρόνος που δαπανήθηκε κατά τη διάρκεια της πρωινής και απογευματινής ώρα αιχμής στην Αθήνα (ευρύτερη περιοχή και κέντρο).



Εικόνα 1.3- Επιπλέον χρόνοι που δαπανήθηκαν κατά την πρωινή & απογευματινή ώρα αιχμής στην ευρύτερη περιοχή της Αθήνας  
Πηγή: <https://www.tomtom.com/traffic-index/athens-traffic/>



Εικόνα 1.4- Επιπλέον χρόνοι που δαπανήθηκαν κατά την πρωινή & απογευματινή ώρα αιχμής στο κέντρο της Αθήνας  
Πηγή: <https://www.tomtom.com/traffic-index/athens-traffic/>

## 1.4 Σκοπός της Διπλωματικής Εργασίας

Όπως γίνεται αντιληπτό από τις προηγούμενες ενότητες, καθίσταται επιτακτική ανάγκη η παρακολούθηση και ανάλυση της κυκλοφορίας, σε ολόκληρη την έκταση του οδικού δικτύου, καθώς και ο εντοπισμός χωροχρονικών μοτίβων. Αυτό διευκολύνει τους ερευνητές να προβλέψουν πιθανή εμφάνιση κορεσμού και να εφαρμόσουν, τελικά, στρατηγικές βελτίωσης της ροής κυκλοφορίας, ώστε να μειωθούν η ρύπανση και τα ατυχήματα και να βελτιωθεί η εμπειρία οδήγησης για τους οδηγούς.

Σκοπός της παρούσας Διπλωματικής Εργασίας αποτελεί η ανάπτυξη μιας μεθοδολογίας για τον εντοπισμό και την ανάλυση χωροχρονικών κυκλοφοριακών μοτίβων στην ευρύτερη περιοχή της Αθήνας με χρησιμοποιώντας την μέθοδο DTW (Dynamic Time Warping). Πιο συγκεκριμένα, θα συλλεχθούν στοιχεία διελεύσεων οχημάτων από φωρατές σε 389 σημεία της Αττικής για την περίοδο Φεβρουαρίου-Μαρτίου 2023 τις ώρες 07:00 το πρωί έως 01:00 το βράδυ. Τα στοιχεία αυτά, στη συνέχεια, θα ομαδοποιηθούν ανά μία ώρα (18 χρονοσειρές) για όλα τα σημεία των φωρατών και με

τη μέθοδο DTW θα παρατηρηθεί πώς αυτά ομαδοποιούνται κατά τη διάρκεια της ημέρας. Θα πραγματοποιηθούν 3 παραδείγματα-ομαδοποιήσεις (clusters) για την κάθε ώρα της ημέρας, για 3, 4 και 6 clusters και στη συνέχεια θα αξιολογηθούν τα αποτελέσματα βάσει του δείκτη Silhouette (Silhouette Index). Τέλος, τα αποτελέσματα των βέλτιστων ομαδοποιήσεων (clusters)- βάσει του δείκτη Silhouette- θα αποτυπωθούν σε χάρτες όπου για κάθε ώρα θα αποτυπώνεται για τον κάθε φωρατή ένα χρώμα (κίτρινο, πορτοκαλί ή κόκκινο), ανάλογα με την ομάδα που βρίσκεται, για την καλύτερη κατανόηση των αποτελεσμάτων. Ο σχολιασμός και η παρουσίαση των αποτελεσμάτων θα γίνει για 4 περιοχές μελέτης: Κέντρο, Ανατολικά & Βόρεια Προάστια (Λεωφ. Κηφισίας & Λεωφ. Μεσογείων), Νότια Προάστια (Λεωφ. Βουλιαγμένης & Λεωφ. Ποσειδώνος) και Δυτικά Προάστια (Λεωφ. Κηφισού και Πειραιάς), όπου θα γίνει περιγραφή της κίνησης κατά τη διάρκεια της ημέρας και οι ώρες που παρατηρείται συμφόρηση, ή μη, της κυκλοφορίας.

## 1.5 Διάρθρωση της Διπλωματικής Εργασίας

Η Διπλωματική Εργασία χωρίζεται στα παρακάτω κεφάλαια:

Στο Κεφάλαιο 1<sup>ο</sup> γίνεται περιγραφή του φαινομένου της κυκλοφοριακής συμφόρησης σε μεγάλες πόλεις και στην Αθήνα.

Στο Κεφάλαιο 2<sup>ο</sup> γίνεται η βιβλιογραφική ανασκόπηση σε παλαιότερες έρευνες που έχουν γίνει πάνω στην ανίχνευση περιοχών κυκλοφοριακής συμφόρησης.

Στο Κεφάλαιο 3<sup>ο</sup> περιγράφεται η μεθοδολογία που ακολουθήθηκε για τον εντοπισμό και την ανάλυση χωροχρονικών μοτίβων στην περιοχή της Αθήνας.

Στο Κεφάλαιο 4<sup>ο</sup> πραγματοποιείται η ανάλυση των δεδομένων, παρουσιάζονται τα αποτελέσματα, γίνεται η οπτικοποίηση των αποτελεσμάτων σε χάρτες και ακολουθεί ο σχολιασμός τους.

Στο Κεφάλαιο 5<sup>ο</sup> παρατίθενται τα συμπεράσματα που εξάχθηκαν από την εκπόνηση της εργασίας, οι περιορισμοί που εντοπίστηκαν καθώς και προτάσεις για περαιτέρω έρευνα.

## 2. Βιβλιογραφική Ανασκόπηση

### 2.1 Ανάλυση και Πρόβλεψη της Κυκλοφορίας

Τα Ευφυή Συστήματα Μεταφορών (Intelligent Transportation Systems, ITS) και η σύγχρονη διαχείριση της κυκλοφορίας των πόλεων, εξαρτώνται σε μεγάλο βαθμό από την πρόβλεψη της κυκλοφορίας, η οποία επιχειρεί να προβλέψει τις μελλοντικές συνθήκες κυκλοφορίας σε ένα δίκτυο μεταφορών με βάση την ιστορική συμπεριφορά του (Vlahogianni et al. 2004). Η άνευ προηγουμένου πρόοδος στη βαθιά μάθηση κατέστησε δυνατή τη χρήση δεδομένων κυκλοφορίας που ανιχνεύονται από διαφορετικά τμήματα του οδικού δικτύου, αύξησε όμως επίσης σημαντικά την πολυπλοκότητα και τις απαιτήσεις δεδομένων και επεξεργασίας των μοντέλων, καθιστώντας τα λιγότερο χρήσιμα. Κατά συνέπεια, ένα από τα κύρια εμπόδια για την αύξηση της αποτελεσματικότητας, της ακρίβειας και της αξιοπιστίας των μοντέλων πρόβλεψης έχει εντοπιστεί ως η ουσιαστική αναπαράσταση των δεδομένων ροής κυκλοφορίας και του οδικού δικτύου (Fafoutellis, et al. 2023)

Η πρόβλεψη της κυκλοφορίας, έχει χρησιμοποιηθεί εκτενώς στο σχεδιασμό διαδρομών (Wan, et al. 2016), στον έλεγχο ροής οχημάτων (Lv, et al. 2015) και σε άλλες κρίσιμες εφαρμογές. Οι ταξιδιώτες θα μπορούν να σχεδιάσουν τις διαδρομές τους εκ των προτέρων και να γνωρίζουν τον κατά προσέγγιση χρόνο άφιξης, εάν οι συνθήκες κυκλοφορίας (όπως η ροή, η ταχύτητα και η συμφόρηση) μπορούσαν να προβλεφθούν με ακρίβεια. Επιπλέον, οι διαχειριστές των πόλεων θα μπορούσαν να βελτιώσουν την αποδοτικότητα της κυκλοφορίας ελέγχοντας την κυκλοφορία σε κρίσιμες περιοχές.

Τα αστικά δίκτυα μεταφορών είναι συνήθως πολυτροπικά, αποτελούμενα από έναν αριθμό διασυνδεδεμένων υποσυστημάτων που αντιπροσωπεύουν διαφορετικούς τρόπους μεταφοράς, όπως αυτοκίνητα, λεωφορεία και τρένα. Παρέχουν στους κατοίκους των πόλεων πολλαπλές επιλογές μετακίνησης σε περίπτωση διακοπής των υπηρεσιών και είναι σχεδιασμένα για να καλύπτουν ένα ευρύ φάσμα ταξιδιωτικών αναγκών. Ο αυξανόμενος αριθμός των ανθρώπων που ζουν στις πόλεις, η περιορισμένη χωρητικότητα των υποδομών μεταφορών και οι αυξανόμενες ανησυχίες σχετικά με την ανθεκτικότητα των πόλεων καθιστούν πιο σημαντικό από ποτέ το σχεδιασμό, τη διαχείριση και τη λειτουργία πολυτροπικών δικτύων μεταφορών με ολοκληρωμένο τρόπο. Για παράδειγμα, οι υπηρεσίες μετακίνησης με ταξί (ride-hailing) μπορούν να χρησιμοποιηθούν στρατηγικά για να βελτιώσουν την πρόσβαση των ανθρώπων στα

μέσα μαζικής μεταφοράς ή για να υποκαταστήσουν ορισμένες μετακινήσεις με τα μέσα μεταφοράς όταν το σύστημα καθυστερεί ή είναι υπερπλήρες.

Τις προηγούμενες δεκαετίες, η πρόβλεψη της κυκλοφορίας αντιμετωπίζεται ως πρόβλημα χρονοσειράς. Συγχρόνως, εφαρμόζονται παραδοσιακές μέθοδοι μηχανικής μάθησης για τη βελτίωση της ακρίβειας της πρόβλεψης της κυκλοφορίας. Τα τελευταία χρόνια, οι μέθοδοι που βασίζονται στη βαθιά μάθηση παρουσιάζουν ισχυρή ικανότητα στην εξαγωγή χωροχρονικών χαρακτηριστικών. Μοντελοποιούν τη χρονική συσχέτιση με επαναλαμβανόμενα νευρωνικά δίκτυα ή χρονική συνέλιξη και συνήθως χρησιμοποιούν μεθόδους βασισμένες στο Graph Convolution Network (GCN) για την εκμάθηση της χωρικής συσχέτισης. Επιπλέον, ο μηχανισμός αυτοπροσοχής (self-attention mechanism) χρησιμοποιείται ευρέως για τη μοντελοποίηση της δυναμικής χωρικής σχέσης στη δυναμική της κυκλοφορίας και βελτιώνει αποτελεσματικά την ακρίβεια της πρόβλεψης της ροής της κυκλοφορίας.

Πολλές από τις μελετηθέντες εργασίες επικεντρώνονται στην πρόβλεψη ζήτησης ενός τρόπου λειτουργίας και εξετάζουν μόνο τις ενδοτροπικές χωροχρονικές συσχετίσεις, αγνοώντας την πιθανή αλληλεπίδρασή της με άλλους τρόπους. Οι διατροπικές σχέσεις υπάρχουν λόγω της πολύπλοκης ταξιδιωτικής συμπεριφοράς και μπορεί να ποικίλουν στο χώρο και στο χρόνο, καθιστώντας δύσκολη τη μοντελοποίησή τους. Οι υπάρχουσες προσεγγίσεις για την πρόβλεψη πολυτροπικής ζήτησης απαιτούν συνήθως τη συγκέντρωση δεδομένων πολυτροπικής ζήτησης με βάση την ίδια κατάτμηση ζώνης για να καταστεί δυνατή η κοινή χρήση εκμάθησης χαρακτηριστικών. Αυτές οι μέθοδοι δεν είναι σε θέση να συλλάβουν τις ετερογενείς χωρικές συσχετίσεις μεταξύ των τρόπων λειτουργίας σε γενικά πολυτροπικά συστήματα με δίκτυα πολυπλεξίας και ποικίλες χωρικές μονάδες.

Οι Liang et al. (2022) σε έρευνά τους για την πρόβλεψη κοινής ζήτησης σε πολυτροπικά συστήματα μεταφορών (μετρό και διαδρομές με ταξί), βασιζόμενοι με τη σειρά τους σε προηγούμενες έρευνες, εισήγαγαν ένα νέο χωροχρονικό πλαίσιο μοντελοποίησης για την πρόβλεψη της πολυτροπικής ζήτησης, το οποίο ενσωμάτωνε νευρωνικά δίκτυα πολλαπλών συσχετίσεων (multi-relational graph neural networks, MRGNN) με χρονικά δίκτυα συνέλιξης (temporal convolution networks, TCNs) για την από κοινού μοντελοποίηση χωροχρονικών συσχετίσεων σε πολλαπλά μέσα μεταφοράς.

Γενικά, η έρευνα τους παρείχε πληροφορίες σχετικά με τον τρόπο προσαρμογής των νευρωνικών δικτύων γράφων σε ετερογενή δίκτυα κινητικότητας για την πρόβλεψη πολυτροπικής ζήτησης. Προβλήματα εντοπίστηκαν όταν ο αριθμός των σχέσεων είναι μεγάλος, καθώς οι περισσότερες τεχνικές ενσωμάτωσης ετερογενών γραφημάτων (heterogeneous graph embedding) μπορεί να υφίστανται υπερβολική παραμετροποίηση. Μελλοντική έρευνα μπορεί να διερευνήσει τρόπους άμβλυνσης του προβλήματος και βελτίωσης της επεκτασιμότητας του μοντέλου χρησιμοποιώντας πιο προηγμένες τεχνικές κανονικοποίησης (advanced regularization techniques) (Ke et al., 2021). Ακόμη, προτείνεται η διερεύνηση τρόπων γενίκευσης των νευρωνικών δικτύων πολλαπλών συσχετίσεων σε τρόπους μεταφοράς με διαφορετικές χρονικές αναλύσεις. Μια πιθανή μέθοδος είναι η ανάπτυξη ενός χωροχρονικού σύγχρονου νευρωνικού δικτύου γραφημάτων (spatiotemporal synchronous graph neural network) (Song et al., 2020) και η θεώρηση, όχι μόνο ετερογενών χωρικών μονάδων, αλλά και ετερογενών χρονικών μονάδων ως κόμβους στο γράφημα.

Οι Zhang et al (2022) στη δική τους έρευνα προτείνουν μία νέα στρατηγική ανάλυσης χωρικών σχέσεων, η οποία ασχολείται με δεδομένα που λείπουν στη ροή της κυκλοφορίας δημιουργώντας μια ισχυρή χαρακτηριστική συσχέτιση τοπικής περιόδου για κάθε ζεύγος ακολουθιών. Επιπλέον, παρουσιάζουν μία μέθοδο για τον υπολογισμό της υβριδικής σχέσης μεταξύ δύο ακολουθιών, η οποία όχι μόνο αναλύει την ομοιότητα κατάστασης κυκλοφορίας κατασκευάζοντας έναν πίνακα που βασίζεται σε σημείο, αλλά επίσης εξετάζει την ομοιότητα τάσης-κατάστασης κυκλοφορίας κατασκευάζοντας έναν πίνακα που βασίζεται σε παλινδρόμηση. Αντικατοπτρίζουν από κοινού τις ιεραρχικές χωρικές σχέσεις δύο επιπέδων του συστήματος μεταφορών συνολικά. Τα πειράματα διεξάγονται σε τέσσερα σύνολα δεδομένων πραγματικού κόσμου και τα αποτελέσματα δείχνουν ότι η προτεινόμενη μέθοδος υπερέχει των υπαρχουσών μεθόδων.

Πιο συγκεκριμένα, για να μειωθούν οι επιπτώσεις των ελλιπών δεδομένων, αντικαθιστούν τα ελλιπή δεδομένα με τον μέσο όρο άλλων δεδομένων και εξετάζουν τα δεδομένα από πολλαπλά χρονικά βήματα. Κατά τον υπολογισμό της ομοιότητας κατάστασης (state similarity), λαμβάνουν υπόψη τις πληροφορίες πολλαπλών χρονικών βημάτων. Κατά τον υπολογισμό της ομοιότητας τάσης (trend similarity), χρησιμοποιούν τη μέθοδο γραμμικής παλινδρόμησης για την προσαρμογή των δεδομένων κυκλοφορίας. Η μέθοδος γραμμικής παλινδρόμησης λαμβάνει επίσης υπόψη τα πολλαπλά χρονικά βήματα και εξαλείφει την επίδραση των ελλιπών

δεδομένων. Για να μοντελοποιήσουν τις ιεραρχικές χωρικές σχέσεις στα δεδομένα κυκλοφορίας, προτείνουν πίνακες καταστάσεων και τάσεων. Συγκεκριμένα, ο πίνακας κατάστασης υπολογίζεται από την κατάσταση της κυκλοφορίας. Για τον πίνακα τάσεων, εφαρμόζουν τη μέθοδο γραμμικής παλινδρόμησης για τον υπολογισμό των σχέσεων μεταξύ των τάσεων της κυκλοφορίας. Με άλλα λόγια, πρώτα προσαρμόζουν τα δεδομένα κυκλοφορίας με γραμμική παλινδρόμηση για να υπολογίσουν την κλίση της. Στη συνέχεια χρησιμοποιούν τις σχέσεις μεταξύ των κλίσεων για να αναπαραστήσουν τις σχέσεις μεταξύ των τάσεων της κυκλοφορίας. Στη δυναμική της κυκλοφορίας, η σχέση μεταξύ των οδικών κόμβων είναι ιεραρχική.

Υπάρχουν σημαντικές χωροχρονικές συσχετίσεις μεταξύ των ροών κυκλοφορίας γειτονικών οδικών τμημάτων στο οδικό δίκτυο. Ο σωστός προσδιορισμός τέτοιων συσχετίσεων συμβάλλει ουσιαστικά στην βελτίωση της ακρίβειας της πρόβλεψης της κυκλοφοριακής ροής. Έχουν γίνει πολλές προσπάθειες από αρκετούς ερευνητές για την επίλυση αυτού του ζητήματος, αλλά υποθέτουν ότι οι χωροχρονικές συσχετίσεις μεταξύ των κυκλοφοριακών ροών είναι σταθερές τόσο στον χρόνο όσο και στον χώρο, δηλαδή οι βαθμοί στους οποίους οι κυκλοφοριακές ροές επηρεάζουν η μία την άλλη είναι καθορισμένοι. Στη έρευνα των Unsok Ryu et al. (2011) προτείνεται μία μέθοδος πρόβλεψης της κυκλοφορίας με βάση την ομαδοποίηση που λαμβάνει υπόψη τη δυναμική φύση των χωροχρονικών συσχετίσεων.

Προκειμένου να εκφραστεί η βραχυπρόθεσμη εξάρτηση μεταξύ του στοχευμένου οδικού τμήματος και των γειτονικών, εισάγονται οι πίνακες χωροχρονικής συσχέτισης. Τα ιστορικά δεδομένα κυκλοφορίας χωρίζονται σε πολλές ομάδες ανάλογα με την ομοιότητα μεταξύ των πινάκων χωροχρονικής συσχέτισης. Η χωροχρονική ανάλυση συσχέτισης και η επιλογή του προγνωστικού παράγοντα με βάση τις αμοιβαίες πληροφορίες εκτελούνται σε κάθε ομάδα, και τα πολλαπλά μοντέλα πρόβλεψης εκπαιδεύονται ξεχωριστά. Ένα μοντέλο πρόβλεψης, που αντιστοιχεί στο σύμπλεγμα στο οποίο ανήκει το τρέχον μοτίβο κυκλοφορίας, επιλέγεται για να εξάγει το αποτέλεσμα πρόβλεψης. Πειραματικά αποτελέσματα σε δεδομένα πραγματικής κυκλοφορίας δείχνουν ότι η προτεινόμενη μέθοδος επιτυγχάνει καλή ακρίβεια πρόβλεψης διακρίνοντας την ετερογένεια των χωροχρονικών συσχετισμών μεταξύ των ροών κυκλοφορίας.

Αν και έχει σημειωθεί αξιοσημείωτη πρόοδος στην ανάπτυξη μοντέλων πρόβλεψης κυκλοφορίας τις τελευταίες δεκαετίες, εξακολουθεί να είναι ένα ανοιχτό ζήτημα η

επίτευξη ικανοποιητικής απόδοσης πρόβλεψης λόγω της εξαιρετικά μη γραμμικής, στοχαστικής και μη στάσιμης φύσης των ροών κυκλοφορίας. Μια σημαντική τάση που εμφανίστηκε πρόσφατα στον τομέα της έρευνας πρόβλεψης κυκλοφορίας είναι η βελτίωση της απόδοσης πρόβλεψης με την ενσωμάτωση της χωροχρονικής σχέσης μεταξύ γειτονικών οδικών τμημάτων στο δίκτυο κυκλοφορίας στα υπάρχοντα μοντέλα πρόβλεψης. Οι περισσότερες από τις πρώιμες μελέτες πρόβλεψης κυκλοφορίας έχουν λάβει υπόψη μόνο τα δεδομένα κυκλοφορίας από ένα οδικό τμήμα-στόχο, από το οποίο θα πρέπει να προβλεφθεί η μελλοντική κυκλοφορία. Στο πραγματικό δίκτυο κυκλοφορίας, ωστόσο, όλα τα οδικά τμήματα είναι διασυνδεδεμένα επομένως η κυκλοφορία ενός οδικού τμήματος επηρεάζεται με διαφορετικό βαθμό από τις προηγούμενες ροές των άλλων οδικών τμημάτων που είναι γειτονικά με αυτό (Chandra, et al. 2008).

Για παράδειγμα, η κυκλοφοριακή συμφόρηση που εκδηλώθηκε σε ένα οδικό τμήμα μπορεί να εξαπλωθεί σε άλλα τμήματα μετά από αρκετά λεπτά, με αποτέλεσμα την περιφερειακή συμφόρηση. Από την άλλη πλευρά, η συσσώρευση και η διάχυση των οχημάτων σε έναν οδικό κόμβο επηρεάζει τις κυκλοφοριακές ροές στους παρακείμενους κόμβους (Vlahogianni, et al. 2007b). Αυτό υποδηλώνει ότι η συνεκτίμηση μόνο της χρονικής πληροφορίας δεν επαρκεί για την εξήγηση της δυναμικής των κυκλοφοριακών ροών και μπορεί να οδηγήσει σε κακό αποτέλεσμα πρόβλεψης. Ως εκ τούτου, ο ακριβής προσδιορισμός των χωροχρονικών συσχετίσεων μεταξύ των κυκλοφοριακών ροών αποτελεί σημαντικό έργο στην έρευνα για την πρόβλεψη της κυκλοφορίας (Vlahogianni, et al. 2014).

Το πρόβλημα που εντοπίστηκε στη μέθοδο αυτή είναι ότι τα καλύτερα αποτελέσματα πρόβλεψης επιτυγχάνονται όταν τα μοτίβα κυκλοφορίας χωρίζονται σε δύο ομάδες (δηλ.  $K = 2$ ). Σύμφωνα με τη γενική προσδοκία της ποικιλομορφίας της κατανομής της ροής κυκλοφορίας, η δυναμική της χωροχρονικής συσχέτισης είναι πιθανώς πιο περίπλοκη. Φαίνεται παράλογο να γίνεται διάκριση μεταξύ περίπλοκων καταστάσεων κίνησης σε δύο μόνο καθεστώτα κυκλοφορίας. Προτείνεται, σε μελλοντικές εργασίες, η ανάπτυξη μιας πιο επεξεργασμένης μεθόδου ομαδοποίησης, με στόχο την καταγραφή της ετερογένειας της διαστρωματικής συσχέτισης και του μέτρου επικύρωσης.

Η κατανόηση της αστικής κυκλοφορίας σε μεγάλες πόλεις είναι κρίσιμη, δεδομένου ότι τα γεγονότα συμφόρησης συμβαίνουν συχνά και επηρεάζουν τη λειτουργική αξιοπιστία τόσο στις τοπικές περιοχές όσο και σε ολόκληρο το σύστημα. Ως εκ τούτου,



είναι σημαντικό να κατανοήσουμε την επιρροή των δρόμων σε άλλους δρόμους, δηλαδή πώς διαδίδεται η αλλαγή της κατάστασης της κυκλοφορίας. Αυτό θα μπορούσε να παράσχει χρήσιμη καθοδήγηση για τον σχεδιασμό σημάτων έγκαιρης προειδοποίησης για τον σχηματισμό μεγάλων συμφορήσεων. Προς αυτή την κατεύθυνση, ένα σημαντικό βήμα είναι η αποκάλυψη των υφιστάμενων σχέσεων εξάρτησης μεταξύ διαφορετικών οδικών τμημάτων στην πόλη και η αξιολόγηση της κρισιμότητάς τους σύμφωνα με τις καταγεγραμμένες εξαρτήσεις. Κατ' αρχήν, οι εξαρτήσεις αυτές προκαλούνται από τη διάδοση της ροής της κυκλοφορίας. Η εύρεση αυτών των εξαρτήσεων μπορεί να μας πει ποια οδικά τμήματα πρέπει να προσαρμόσουμε όταν θέλουμε να επηρεάσουμε τους άλλους δρόμους (Guo, et al., 2019).

Τα εντοπισμένα οδικά τμήματα και οι συσχετίσεις είναι πολύ σημαντικά για την πρόβλεψη της τοπικής συμφόρησης και των αποτελεσμάτων της. Για παράδειγμα, αφού συμβεί ένα κυκλοφοριακό συμβάν σε ένα οδικό τμήμα, είναι απαραίτητο να προβλεφθεί το εύρος που μπορεί να εξαπλωθεί η συμφόρηση και να εκτιμηθούν οι αντίστοιχοι κίνδυνοι. Αναμένεται ότι το δίκτυο συσχετίσεων μπορεί να παρέχει υποστήριξη για την πρόβλεψη της έναρξης της κυκλοφοριακής συμφόρησης. Εάν υπάρχουν συσχετίσεις μεταξύ πολλών οδικών τμημάτων, θεωρείται ότι η συμφόρηση μπορεί να διαδοθεί μέσω αυτών των οδικών τμημάτων και όχι μέσω της περιοχής ασθενούς συσχέτισης. Σε ορισμένες μελέτες, έχει αποδειχθεί ότι οι χρονικές/χωρικές συσχετίσεις μεταξύ των στοιχείων του συστήματος κυκλοφορίας μπορούν να βοηθήσουν στη βελτίωση της πρόβλεψης στα συστήματα κυκλοφορίας.

Για παράδειγμα, οι Min και Wynter πρότειναν το «Multivariate Spatial-Temporal Autoregressive Moving Average (MSTARMA)» μοντέλο, όπου έχουν ληφθεί υπόψη οι χωρο-χρονικές συσχετίσεις, και πραγματοποίησαν πρόβλεψη της κυκλοφορίας σε πραγματικό χρόνο έως και 1 ώρα νωρίτερα. Οι Pan et al. εισήγαγαν χωρο-χρονικές συσχετίσεις στο Στοχαστικό Μοντέλο Μετάδοσης Κυττάρων (Stochastic Cell Transmission Model, SCTM) και εφάρμοσαν το μοντέλο τους για βραχυπρόθεσμη πρόβλεψη της κατάστασης της κυκλοφορίας. Σε αυτές τις εργασίες, οι χωρικές συσχετίσεις είναι συσχετίσεις μεταξύ διαφορετικών κόμβων σε ένα σταθερό χρονικό σημείο κατά τη διάρκεια διαφορετικών ημερών και οι χρονικές συσχετίσεις είναι συσχετίσεις μεταξύ δύο χρονικών σημείων για τον ίδιο κόμβο κατά τη διάρκεια διαφορετικών ημερών. Σε σύγκριση με αυτά τα προηγούμενα μέτρα, η προσέγγιση των

Guo, et al. λαμβάνει υπόψη την ισχύ συσχέτισης μεταξύ δύο οδικών τμημάτων με την πάροδο του χρόνου και τη χρονική τους υστέρηση, συνολικά. Ειδικότερα, διερευνούν τη σχέση συσχέτισης ως δίκτυο και υπολογίζουν ορισμένες δομικές ιδιότητες του δικτύου συσχέτισης. Τα ευρήματα της έρευνάς αυτής μπορούν να παράσχουν πληροφορίες για την καλύτερη κατανόηση των πολύπλοκων συστημάτων αστικών μεταφορών και θα μπορούσαν να διευκολύνουν μελλοντικές μελέτες για την πρόβλεψη και τον έλεγχο της διάδοσης ακραίων εμπλοκών.

## 2.2 Συμπεράσματα Βιβλιογραφικής Ανασκόπησης

Το φαινόμενο της κυκλοφοριακής συμφόρησης είναι πολύπλευρο. Μπορεί να οριστεί και να προσδιοριστεί με διάφορους τρόπους. Η χρήση αυτών των τεχνικών επιταχύνει τη μελέτη για την εκτίμηση του χρόνου που θα χρειαστεί. Τα μακροχρόνια φαινόμενα κυκλοφοριακής συμφόρησης δημιουργούν σοβαρά ζητήματα για το περιβάλλον, την οικονομία και τη ζωή των κατοίκων, καθιστώντας το θέμα αυτό επιτακτικό για διερεύνηση.

Η δυνατότητα συλλογής δεδομένων για παρατεταμένες χρονικές περιόδους και η εξέλιξη της υπολογιστικής ισχύος έχουν οδηγήσει στη συχνή αναπαράσταση των δεδομένων ως χρονικών σημείων. Με αυτόν τον τρόπο, μια παρατήρηση μπορεί πραγματικά να είναι ένα σύνολο τιμών, γνωστό ως μονομεταβλητή χρονοσειρά. Αυτού του είδους οι δομές δεδομένων χρησιμοποιούνται ευρέως σε πολλούς διαφορετικούς κλάδους. Για παράδειγμα, οι περισσότεροι χρηματοοικονομικοί (τιμές μετοχών, συναλλαγματικές ισοτιμίες), περιβαλλοντικοί (ρύπανση, κατανάλωση ενέργειας) και κοινωνικοοικονομικοί (ΑΕΠ, εργασία, γεννητικότητα, εκπαίδευση) δείκτες συγκεντρώνονται με την πάροδο του χρόνου. Τα δεδομένα χρονοσειρών χρησιμοποιούνται επίσης ευρέως στη μετεωρολογία (καιρικές συνθήκες, σεισμική έρευνα), στο IoT (παρακολούθηση της σωματικής δραστηριότητας, διαχείριση ενέργειας) και στην ιατρική (αρτηριακή πίεση, ηλεκτροκαρδιογράφημα). Στην επιστήμη των δεδομένων, αυτό το είδος ανάλυσης δεδομένων είναι ζωτικής σημασίας.

## 3. Μεθοδολογική Προσέγγιση

### 3.1 Εισαγωγή

Στην ομαδοποίηση δεδομένων χρονοσειρών έχουν οριστεί τρεις βασικοί τύποι προσεγγίσεων (Aghabozorgi et al. 2015, Guijo-Rubio et al. 2018). Ο πρώτος είναι οι ολόκληρες χρονοσειρές, όπου αντιμετωπίζεται μια χρονοσειρά ως μια παρατήρηση. Ως παράδειγμα μπορεί κανείς να πάρει οικονομικές βάσεις δεδομένων όπως η Eurostat. Οι παρατηρήσεις είναι συνήθως χώρες και σε κάθε στήλη υπάρχει μια τιμή του επιλεγμένου δείκτη σε μια συγκεκριμένη χρονική στιγμή. Προκειμένου να ομαδοποιηθούν αυτές οι χώρες με βάση την ομοιότητα των χρονικών σημείων δεδομένων θα πρέπει να υπολογιστεί η απόσταση μεταξύ όλων των σημείων.

Η δεύτερη καθορισμένη προσέγγιση ονομάζεται υπακολουθία (subsequence). Λαμβάνει υπόψη μόνο μία χρονοσειρά και τη χωρίζει σε τμήματα ομαδοποιώντας τα με βάση την ομοιότητα. Μπορεί να χρησιμοποιηθεί σε δεδομένα που συλλέγονται από γυροσκόπια, για παράδειγμα σε έξυπνα ρολόγια, για τον προσδιορισμό των τύπων δραστηριοτήτων που εκτελέστηκαν εκείνη τη χρονική στιγμή. Με αυτόν τον τρόπο μπορούμε να βρούμε στιγμές κατά τις οποίες ένα άτομο καθόταν, στεκόταν κ.λπ.

Η τελευταία προσέγγιση είναι η προσέγγιση χρονικών σημείων και εφαρμόζεται επίσης σε μεμονωμένες χρονοσειρές, αλλά με ομαδοποίηση μεμονωμένων σημείων. Αντιμετωπίζει τη χρονοσειρά με τον ίδιο τρόπο σαν τα σημεία δεδομένων να μην ήταν διατεταγμένα, αλλά ανεξάρτητες τιμές. Παρόλο που και οι τρεις μέθοδοι έχουν σημαντικές εφαρμογές, η παρούσα Διπλωματική Εργασία θα επικεντρωθεί μόνο στην ανάλυση ολόκληρων χρονοσειρών.

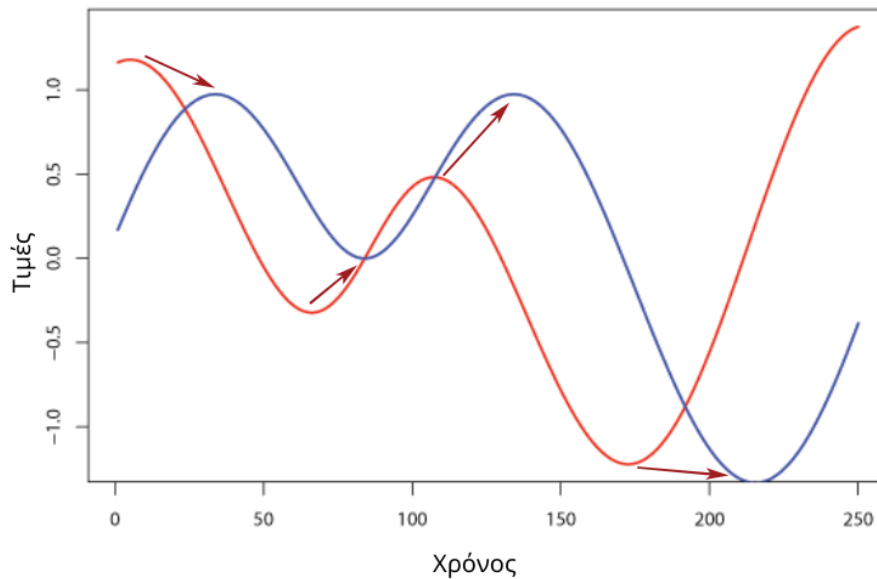
Συλλογή Στοιχείων	Δεδομένα κυκλοφοριακού φόρτου από 390 φωρατές στην Αθήνα
Επεξεργασία στοιχείων	Αφαίρεση φωρατών με λάθος τιμές, δημιουργία αρχείων με ωριαίους φόρτους
Εύρεση Ομοιοτήτων μεταξύ Χρονοσειρών	Με τον αλγόριθμο DTW
Ομαδοποίηση Χρονοσειρών	Με τη μέθοδο k-medoids και τον αλγόριθμο PAM
Εξαγωγή Αποτελεσμάτων	Αξιολόγηση βάσει του δείκτη Silhouette
Οπτικοποίηση Αποτελεσμάτων σε Χάρτες	Αποτύπωση αποτελεσμάτων σε χάρτες τεσσάρων περιοχών της Αθήνας & σχολιασμός

Διάγραμμα 3.1 – Διάγραμμα Ροής Μεθοδολογικής Προσέγγισης

### 3.2 Ο αλγόριθμος Dynamic Time Warping

Ο αλγόριθμος Dynamic Time Warping (DTW) έχει κερδίσει τη δημοτικότητά του επειδή είναι εξαιρετικά αποτελεσματικός ως μέτρο ομοιότητας χρονοσειρών, το οποίο ελαχιστοποιεί τις επιπτώσεις της μετατόπισης και της παραμόρφωσης στο χρόνο, επιτρέποντας τον "ελαστικό" μετασχηματισμό των χρονοσειρών προκειμένου να ανιχνεύονται παρόμοια σχήματα με διαφορετικές φάσεις. Δεδομένων δύο χρονοσειρών  $X = (x_1, x_2, \dots, x_N)$ ,  $N \in \mathbb{N}$  και  $Y = (y_1, y_2, \dots, y_M)$ ,  $M \in \mathbb{N}$  που αναπαρίστανται από τις ακολουθίες τιμών (ή καμπύλες που αναπαρίστανται από τις ακολουθίες κορυφών), η DTW δίνει βέλτιστη λύση σε χρόνο  $O(MN)$ , ο οποίος θα μπορούσε να βελτιωθεί περαιτέρω μέσω διαφόρων τεχνικών, όπως η πολλαπλή κλιμάκωση (multi-scaling). Ο μόνος περιορισμός που τίθεται στις ακολουθίες δεδομένων είναι ότι θα πρέπει να δειγματοληπτούνται σε ισαπέχοντα χρονικά σημεία, το οποίο μπορεί να επιλυθεί με επαναληπτική δειγματοληψία.

### Χρονοσειρά πριν την ευθυγράμμιση



Εικόνα 3.1- Αρχικές χρονοσειρές, τα βέλη δείχνουν τα επιθυμητά σημεία ευθυγράμμισης

Εάν οι ακολουθίες λαμβάνουν τιμές από κάποιο χώρο χαρακτηριστικών  $\Phi$ , τότε για να συγκρίνουμε δύο διαφορετικές ακολουθίες  $X$  και  $Y \in \Phi$ , πρέπει να χρησιμοποιήσουμε το μέτρο τοπικής απόστασης, το οποίο ορίζεται ως μια συνάρτηση:  $d$  :

$$\Phi \times \Phi \rightarrow \mathbb{R} \geq 0 \quad (1)$$

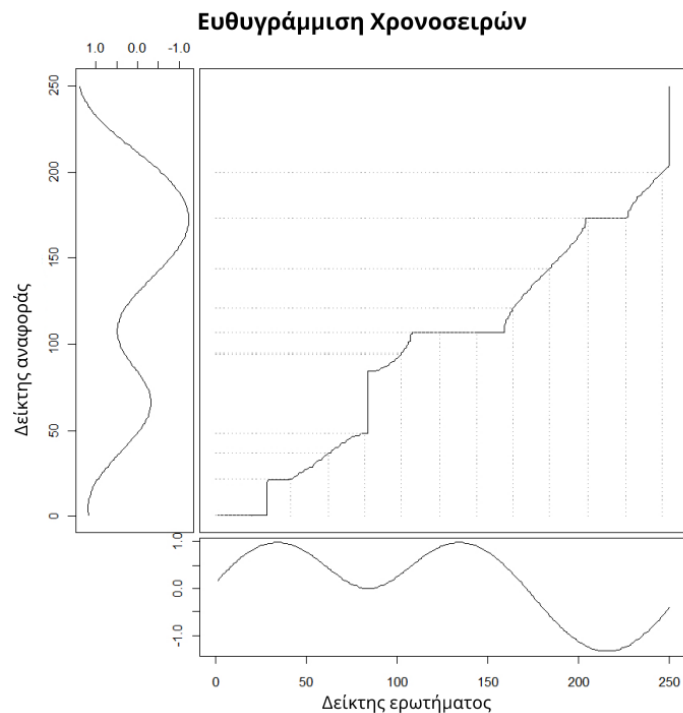
Διαισθητικά, το  $d$  έχει μικρή τιμή όταν οι ακολουθίες είναι παρόμοιες και μεγάλη τιμή αν είναι διαφορετικές. Δεδομένου ότι ο αλγόριθμος δυναμικού προγραμματισμού βρίσκεται στον πυρήνα του DTW, συνηθίζεται να αποκαλείται αυτή η συνάρτηση απόστασης "συνάρτηση κόστους" και το ζητούμενο της βέλτιστης ευθυγράμμισης των ακολουθιών γίνεται το ζητούμενο της διάταξης όλων των σημείων των ακολουθιών με ελαχιστοποίηση της συνάρτησης κόστους (ή της απόστασης).

Ο αλγόριθμος ξεκινά με τη δημιουργία του πίνακα αποστάσεων  $C \in \mathbb{R}^{N \times M}$  που αντιπροσωπεύει όλες τις αποστάσεις ανά ζεύγη μεταξύ των  $X$  και  $Y$ . Αυτός ο πίνακας αποστάσεων ονομάζεται **πίνακας τοπικού κόστους** για την ευθυγράμμιση δύο ακολουθιών  $X$  και  $Y$  :

$$C_i \in \mathbb{R}^{N \times M} : c_{i,j} = ||x_i - y_j||, i \in [1 : N], j \in [1 : M] \quad (2)$$

Αφού κατασκευαστεί ο τοπικός πίνακας κόστους, ο αλγόριθμος βρίσκει το μοτίβο ευθυγράμμισης που διέρχεται από τις περιοχές χαμηλού κόστους - "κοιλιάδες" στον πίνακα κόστους. Αυτό το μοτίβο ευθυγράμμισης (ή μοτίβο στρέβλωσης ή συνάρτηση

στρέβλωσης) ορίζει την αντιστοιχία ενός στοιχείου  $x_i \in X$  με  $y_j \in Y$  ακολουθώντας την οριακή συνθήκη που αντιστοιχίζει το πρώτο και το τελευταίο στοιχείο των  $X$  και  $Y$  μεταξύ τους.



Εικόνα 3.2- Το βέλτιστο μονοπάτι στρέβλωσης που ευθυγραμμίζει τις χρονοσειρές από την Εικόνα 3.1

Το μοτίβο ευθυγράμμισης που κατασκευάζεται από την DTW είναι μια ακολουθία σημείων  $p = (p_1, p_2, \dots, p_K)$  με  $p_l = (p_l, p_j) \in [1 : N] \times [1 : M]$  για  $l \in [1 : K]$  που πρέπει να ικανοποιεί τα ακόλουθα κριτήρια:

- 1) **Οριακή συνθήκη:**  $p_1 = (1, 1)$  και  $p_K = (N, M)$ . Τα σημεία έναρξης και λήξης της διαδρομής στρέβλωσης πρέπει να είναι το πρώτο και το τελευταίο σημείο ευθυγραμμισμένων ακολουθιών
- 2) **Συνθήκη μονοτονίας:**  $n_1 \leq n_2 \leq \dots \leq n_K$  και  $m_1 \leq m_2 \leq \dots \leq m_K$ . Αυτή η συνθήκη διατηρεί τη χρονική διάταξη των σημείων.
- 3) **Συνθήκη μεγέθους βήματος:** αυτό το κριτήριο περιορίζει το μονοπάτι στρέβλωσης από μεγάλα άλματα (μετατοπίσεις στο χρόνο) κατά την ευθυγράμμιση των ακολουθιών. Η βασική συνθήκη μεγέθους βήματος διατυπώνεται ως  $p_{l+1} - p_l \in \{(1, 1), (1, 0), (0, 1)\}$ .

Η συνάρτηση κόστους που σχετίζεται με ένα μοτίβο στρέβλωσης και υπολογίζεται σε σχέση με τον τοπικό πίνακα κόστους (ο οποίος αντιπροσωπεύει όλες τις αποστάσεις ανά ζεύγη) θα είναι:

$$c_p(X, Y) = \sum_{l=1}^L c(x_{nl}, y_{ml}) \quad (3)$$

Το μοτίβο στρέβλωσης που έχει ελάχιστο κόστος σχετιζόμενο με την ευθυγράμμιση ονομάζεται βέλτιστο μοτίβο στρέβλωσης. Θα ονομάζουμε αυτό το μονοπάτι  $P^*$ . Ακολουθώντας τον ορισμό του βέλτιστου μοτίβου στρέβλωσης για να βρούμε ένα τέτοιο μοτίβο, πρέπει να δοκιμάσουμε κάθε πιθανό μοτίβο στρέβλωσης μεταξύ των  $X$  και  $Y$ , το οποίο θα μπορούσε να είναι υπολογιστικά δύσκολο λόγω της εκθετικής αύξησης του αριθμού των βέλτιστων μοτίβων καθώς τα μήκη των  $X$  και  $Y$  αυξάνονται γραμμικά. Για να ξεπεραστεί αυτή η πρόκληση, ο DTW χρησιμοποιεί τον αλγόριθμο δυναμικού προγραμματισμού με πολυπλοκότητα μόνο  $O(MN)$ .

Το μέρος του δυναμικού προγραμματισμού του αλγορίθμου DTW χρησιμοποιεί τη συνάρτηση απόστασης DTW

$$DTW(X, Y) = c_p * (X, Y) = \min\{c_p(X, Y), p \in P^{N \times M}\} \quad (4)$$

όπου  $P^{N \times M}$  είναι το σύνολο όλων των πιθανών μονοπατιών (path) στρέβλωσης και κατασκευάζει τον συσσωρευμένο πίνακα κόστους ή παγκόσμιο πίνακα κόστους  $D$  που ορίζεται ως εξής:

1. Πρώτη γραμμή:  $D(1, j) = \sum_{k=1}^j c(x_1, y_k), j \in [1, M]$
2. Πρώτη στήλη:  $D(i, 1) = \sum_{k=1}^i c(x_k, y_1), i \in [1, N]$
3. Όλα τα άλλα στοιχεία:  $D(i, j) = \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\} + c(x_i, y_j), i \in [1, N], j \in [1, M]$ .

Το χρονικό κόστος για την κατασκευή αυτού του πίνακα είναι  $O(NM)$  το οποίο ισούται με το κόστος του ακόλουθου αλγορίθμου, όπου  $X$  και  $Y$  είναι οι χρονοσειρές εισόδου και  $C$  είναι ο τοπικός πίνακας κόστους που αντιπροσωπεύει όλες τις αποστάσεις ανά ζεύγη μεταξύ των  $X$  και  $Y$ .

### 3.3 Σύγκριση μεθόδων Dynamic Time Warping και Euclidean

Κατά τη βιβλιογραφική αναζήτηση εντοπίστηκε και η Euclidean μέθοδος η οποία ενώ μοιάζει με την DTW, παρουσιάζει αρκετά μειονεκτήματα που καθιστούν ακατάλληλη τη χρήση της σε ορισμένες εφαρμογές:

α) Συγκρίνει μόνο χρονοσειρές ίδιου μήκους

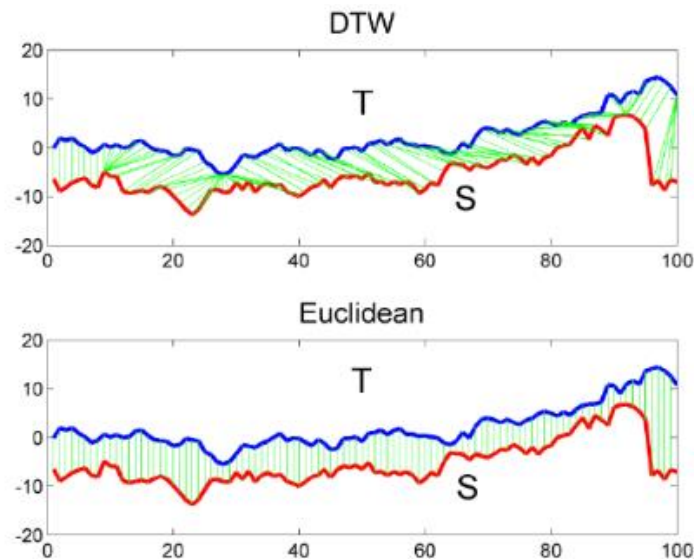
β) Δεν χειρίζεται τις ακραίες τιμές

γ) Είναι πολύ ευαίσθητη σε σχέση με έξι μετασχηματισμούς σχήματος: μετατόπιση (shifting), ομοιόμορφη κλιμάκωση πλάτους (uniform amplitude scaling), ομοιόμορφη κλιμάκωση χρόνου (uniform time scaling), ομοιόμορφη διπλή κλιμάκωση (uniform bi-scaling), χρονική στρέβλωση (time warping) και μη ομοιόμορφη κλιμάκωση πλάτους (non-uniform amplitude-scaling).

Η DTW παρέχει μεγαλύτερη ανθεκτικότητα στον υπολογισμό της ομοιότητας. Με τη μέθοδο αυτή, μπορούν να συγκριθούν και χρονοσειρές διαφορετικού μήκους, επειδή αντικαθιστά τη σύγκριση ενός προς ένα σημείο, που χρησιμοποιείται στην Ευκλείδεια απόσταση, με μια σύγκριση πολλών προς ένα (και αντίστροφα). Το κύριο χαρακτηριστικό αυτού του μέτρου απόστασης είναι ότι επιτρέπει την αναγνώριση παρόμοιων σχημάτων, ακόμη και αν παρουσιάζουν μετασχηματισμούς σχήματος, όπως μετατόπιση ή/και κλιμάκωση. Επίσης, υπολογίζει τη μικρότερη απόσταση μεταξύ όλων των σημείων, το οποίο επιτρέπει ένα- προς πολλά ταιριαρίσματα.

Η Ευκλείδεια μέθοδος είναι η πιο αποτελεσματική, αλλά αναγκάζει και τις δύο χρονοσειρές να έχουν το ίδιο μήκος (Guijo-Rubio et al. (2018), Wang et al. (2013)), αλλά η DTW είναι η πιο ακριβής (Wang et al. (2013)). Στην Εικόνα 3.3 απεικονίζονται οι διαφορές των δύο μεθόδων.





Εικόνα 3.3 - Σύγκριση μεταξύ της απόστασης DTW και της Ευκλείδειας απόστασης (οι πράσινες γραμμές αντιπροσωπεύουν την αντιστοίχιση μεταξύ των σημείων των χρονοσειρών T και S). Η πρώτη επιτρέπει συγκρίσεις πολλών προς ένα σημείων, ενώ η ευκλείδεια απόσταση σημείο προς σημείο.

### 3.4 Ομαδοποίηση χρονοσειρών με DTW και k-medoids

Η ομαδοποίηση (clustering) είναι η διαδικασία διάσπασης μιας αφηρημένης ομάδας σημείων/αντικειμένων δεδομένων σε κλάσεις παρόμοιων αντικειμένων, έτσι ώστε όλα τα αντικείμενα σε μια ομάδα να έχουν παρόμοια χαρακτηριστικά. Μία ομάδα  $n$  αντικειμένων αναλύεται σε  $k$  αριθμό συστάδων/ομάδων με βάση τις ομοιότητές τους.

Τα K-medoids είναι μια μέθοδος, χωρίς επίβλεψη, με μη επισημειωμένα δεδομένα προς ομαδοποίηση. Πρόκειται για μια αυτοσχέδια έκδοση του αλγορίθμου K-Means, η οποία έχει σχεδιαστεί κυρίως για την αντιμετώπιση της ευαισθησίας των ακραίων δεδομένων. Σε σύγκριση με άλλους αλγορίθμους κατάτμησης, ο αλγόριθμος είναι απλός, γρήγορος και εύκολος στην υλοποίηση. Το Medoid είναι ένα σημείο στη συστάδα από το οποίο το άθροισμα των αποστάσεων από άλλα σημεία δεδομένων είναι ελάχιστο. Υπάρχουν τρεις τύποι αλγορίθμων για την ομαδοποίηση K-Medoids:

- PAM (Partitioning Around Clustering)
- CLARA (Clustering Large Applications)
- CLARANS (Randomized Clustering Large Applications)

Ο PAM είναι ο πιο ισχυρός αλγόριθμος από τους τρεις αλγορίθμους, αλλά έχει το μειονέκτημα της χρονικής πολυπλοκότητας. Αυτός ο αλγόριθμος θα χρησιμοποιηθεί στην παρούσα Διπλωματική Εργασία.

Τα στάδια που ακολουθούνται στον αλγόριθμο PAM είναι τα ακόλουθα:

1. Επιλέγεται  $k$  αριθμός τυχαίων σημείων από τα δεδομένα και αναθέτονται αυτά τα  $k$  σημεία σε  $k$  αριθμό συστάδων. Αυτά είναι τα αρχικά medoids.
2. Για όλα τα υπόλοιπα σημεία δεδομένων, υπολογίζεται η απόσταση από κάθε medoid και ανατίθεται στη συστάδα με το πλησιέστερο medoid.
3. Υπολογίζεται το συνολικό κόστος (άθροισμα όλων των αποστάσεων από όλα τα σημεία δεδομένων προς τα medoids).
4. Επιλέγεται ένα τυχαίο σημείο ως νέο medoid και ανταλλάσσεται με το προηγούμενο medoid. Επαναλαμβάνονται τα βήματα 2 και 3.
5. Εάν το συνολικό κόστος του νέου medoid είναι μικρότερο από εκείνο του προηγούμενου medoid, καθίσταται το νέο medoid μόνιμο και επαναλαμβάνεται το βήμα 4.
6. Εάν το συνολικό κόστος του νέου medoid είναι μεγαλύτερο από το κόστος του προηγούμενου medoid, αναιρείται η ανταλλαγή και επαναλαμβάνεται το βήμα 4.
7. Οι επαναλήψεις συνεχίζονται έως ότου δεν προκύψει καμία αλλαγή με νέα medoids για την ταξινόμηση των σημείων δεδομένων.

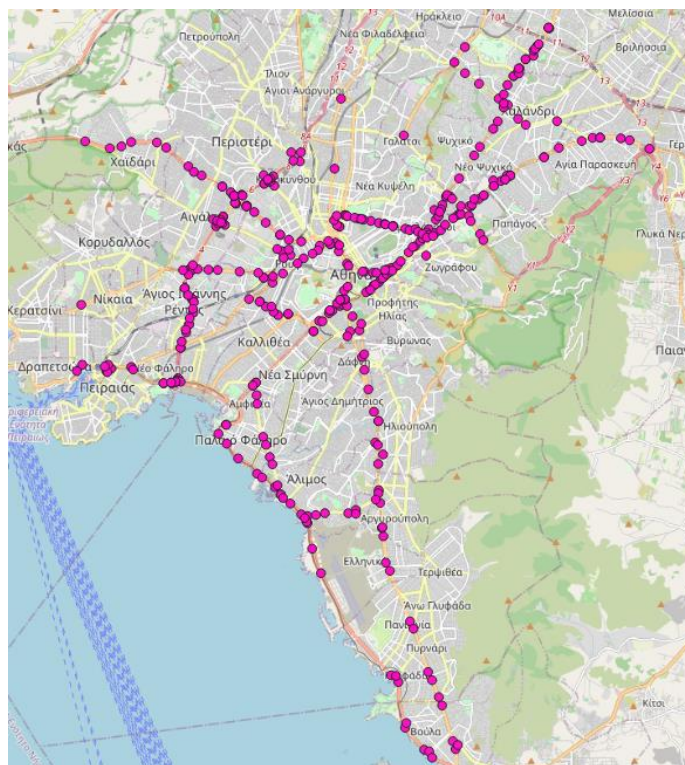
### 3.5 Αξιολόγηση Ομαδοποιήσεων

Η αξιολόγηση των αποτελεσμάτων της μεθόδου DTW, στην παρούσα Διπλωματική Εργασία, θα γίνει με τον δείκτη Silhouette (Sil). Ο δείκτης Silhouette κυμαίνεται από -1 έως 1 και μετρά την ομοιότητα μεταξύ σημείων δεδομένων εντός μιας ομάδας και την ανομοιότητα μεταξύ σημείων δεδομένων σε διαφορετικές ομάδες. Μια υψηλότερη τιμή υποδηλώνει καλύτερη ποιότητα ομαδοποίησης. Μια υψηλή τιμή υποδεικνύει ότι το αντικείμενο ταιριάζει καλά στη δική του ομάδα και ελάχιστα στις γειτονικές ομάδες.

## 4. Ανάλυση Δεδομένων και Αποτελέσματα

### 4.1 Συλλογή Δεδομένων

Στο πλαίσιο της διπλωματικής, συλλέχθηκαν δεδομένα από φωρατές στην Αθήνα. Οι εν λόγω φωρατές (detectors) είναι συσκευές που χρησιμοποιούνται για την παρακολούθηση και τη συλλογή δεδομένων σχετικά με τη κυκλοφοριακή ροή, τον κυκλοφοριακό όγκο και την ταχύτητα της κυκλοφορίας στους δρόμους. Τα δεδομένα τα οποία χρησιμοποιήθηκαν στην έρευνα, αφορούν σε μετρήσεις την περίοδο Φεβρουαρίου-Μαρτίου 2023. Οι μετρήσεις αυτές έγιναν από 390 φωρατές στην Αθήνα οι οποίοι κατέγραψαν, για τις ώρες 07:00 το πρωί έως 01:00 το βράδυ, τον κυκλοφοριακό φόρτο στο οδικό δίκτυο της περιοχής.



Εικόνα 4.1 - Σημεία μέτρησης κυκλοφοριακών μεγεθών στην Αθήνα

Τα δεδομένα που αναλύθηκαν έχουν ομαλοποιηθεί και επεξεργαστεί όπως θα δούμε στην επόμενη ενότητα. Η βάση δεδομένων έχει την μορφή του Πίνακα 4.1.

Πίνακας 4.1- Μορφή βάσης δεδομένων μετρήσεων

ΗΜΕΡΟΜΗΝΙΑ	ΩΡΑ	MS106	MS107	MS108	MS109
1/2/2023	00:00-01:00	986	77	810	62
1/2/2023	07:00-08:00	4271	449	4718	706
1/2/2023	08:00-09:00	4783	452	5409	669
1/2/2023	09:00-10:00	3994	425	5178	661
1/2/2023	10:00-11:00	3936	374	4618	648
1/2/2023	11:00-12:00	3988	425	4189	688
1/2/2023	12:00-13:00	4085	467	4139	672
1/2/2023	13:00-14:00	4421	548	4169	717
1/2/2023	14:00-15:00	4626	570	4620	764
1/2/2023	15:00-16:00	4930	557	4859	754

## 4.2 Επεξεργασία δεδομένων

Αρχικά, η βάση δεδομένων έπρεπε να απαλλαχθεί από λάθη. Παρατηρήθηκε μεγάλος όγκος δεδομένων όπου οι φωρατές κατέγραψαν είτε μηδενικές μετρήσεις, είτε λανθασμένες. Οι φωρατές αυτοί αφαιρέθηκαν από τη βάση δεδομένων. Πιο συγκεκριμένα αφαιρέθηκαν οι MS: 111, 113, 115, 117, 121, 123, 125, 134, 140, 143, 145, 204, 340, 341, 362, 364, 365, 470, 471, 472, 473, 485, 487, 489, 502, 690, 691, 730, 738, 747, 748, 749, 757.

Στη συνέχεια, η βάση δεδομένων χωρίστηκε σε 18 αρχεία για κάθε ώρα της ημέρας με την ονομασία «traffic\_feb\_mar\_23\_i», όπου  $i=1,2,3$  οι αριθμοί που αντιστοιχίζονται σε κάθε εύρος ώρας. Καθένα από αυτά τα αρχεία, δηλαδή, αφορά 1 ώρα της ημέρας, για παράδειγμα 08:00-09:00, και περιέχει όλα τα δεδομένα των φωρατών για τον Φεβρουάριο και το Μάρτιο του 2023.

Πρώτο βήμα για την επεξεργασία δεδομένων αποτέλεσε η εγκατάσταση βιβλιοθηκών στο περιβάλλον R, οι οποίες είναι απαραίτητες για όλες τις διαδικασίες ανάγνωσης και τροποποίησης των δεδομένων, καθώς και την αξιολόγησή τους. Ορισμένες από αυτές τις βιβλιοθήκες συνοψίζονται παρακάτω:

- ggplot2: είναι ένα σύστημα για τη δημιουργία γραφημάτων, βασισμένο στη Γραμματική των Γραφημάτων (The Grammar of Graphics). Δίνεις τα δεδομένα, λες στο ggplot2 πώς να αντιστοιχίσει τις μεταβλητές οπτικά, ποια γραφικά μέσα να χρησιμοποιήσει, και αυτό αναλαμβάνει τις λεπτομέρειες.
- tsclust: αποσκοπεί στην εφαρμογή ενός μεγάλου συνόλου καθιερωμένων μέτρων ανομοιότητας χρονοσειρών που έχουν αξιολογηθεί από ομοτίμους,

συμπεριλαμβανομένων μέτρων που βασίζονται σε ακατέργαστα δεδομένα, εξαγόμενα χαρακτηριστικά, υποκείμενα παραμετρικά μοντέλα, επίπεδα πολυπλοκότητας και συμπεριφορές πρόβλεψης.

- `dtwclust`: αυτή είναι η κύρια λειτουργία για την ομαδοποίηση χρονοσειρών. Υποστηρίζει την ομαδοποίηση κατά τμήματα, την ιεραρχική, την ασαφή, την `k-Shape` και την `TADPole`.
- `tidyr`: τα `tidy data` περιγράφουν έναν τυποποιημένο τρόπο αποθήκευσης δεδομένων που χρησιμοποιείται όπου είναι δυνατόν σε όλο το `tidyverse`.
- `dplyr`: είναι μια γραμματική χειρισμού δεδομένων, η οποία παρέχει ένα συνεπές σύνολο ρημάτων που βοηθούν να επιλυθούν οι πιο συνηθισμένες προκλήσεις χειρισμού δεδομένων.
- `dtw`: είναι μια οικογένεια αλγορίθμων οι οποίοι υπολογίζουν την τοπική επιμήκυνση ή συμπίεση που πρέπει να εφαρμοστεί στους χρονικούς άξονες δύο χρονοσειρών προκειμένου να αντιστοιχιστεί βέλτιστα η μία (`query`) στην άλλη (`reference`).
- `BBmisc`: Κανονικοποιεί τα αριθμητικά δεδομένα σε μια δεδομένη κλίμακα.
- `TSdist`: Ένα σύνολο από συνήθως χρησιμοποιούμενα μέτρα απόστασης και ορισμένες πρόσθετες συναρτήσεις που, αν και αρχικά δεν είχαν σχεδιαστεί για τον σκοπό αυτό, μπορούν να χρησιμοποιηθούν για τη μέτρηση της ανομοιότητας μεταξύ χρονοσειρών. Τα μέτρα αυτά μπορούν να χρησιμοποιηθούν για την εκτέλεση ομαδοποίησης, ταξινόμησης ή άλλων εργασιών εξόρυξης δεδομένων που απαιτούν τον ορισμό ενός μέτρου απόστασης μεταξύ χρονοσειρών (U. Mori, A. Mendiburu και J.A. Lozano (2016)).

Δεύτερο βήμα, αποτέλεσε η εισαγωγή των δεδομένων στο περιβάλλον της R. Πιο συγκεκριμένα κάθε αρχείο “`csv`” «`traffic_feb_mar_23_i`» ( $i=1,2,3,..$ ) μετονομάστηκε για λόγους απλούστευσης σε «`datai`», όπου  $i=1,2,3,..$ , και έπειτα αφαιρέθηκαν οι στήλες της ημερομηνίας και της ώρας. Αμέσως μετά δημιουργήθηκε η γραφική παράσταση της χρονοσειράς για κάθε ένα από τα 18 αρχεία. Τέλος, αφαιρέθηκε και η στήλη `count` και σε κάθε αρχείο έγινε η μετατροπή των σειρών σε στήλες, ώστε να έχουμε σε κάθε σειρά τους φωρατές και σε κάθε στήλη την κάθε ημέρα της χρονικής περιόδου (Πίνακας 4.2).

Πίνακας 4.2-Τελική μορφή δεδομένων (παράδειγμα 00:00-01:00)

	1	2	3	4	5
<b>MS106</b>	986	1039	1076	450	510
<b>MS107</b>	77	71	74	21	26
<b>MS108</b>	810	905	999	457	483
<b>MS109</b>	62	69	89	32	47
<b>MS110</b>	1011	1021	1052	444	516
<b>MS112</b>	821	940	984	469	483
<b>MS114</b>	990	1023	1010	422	504
<b>MS116</b>	522	618	664	307	342
<b>MS118</b>	858	892	854	371	449
<b>MS119</b>	129	114	127	65	71
<b>MS120</b>	478	632	683	311	311

Όσον αφορά στον κώδικα και την εφαρμογή της μεθόδου «Dynamic Time Warping», ακολουθήθηκε η ίδια διαδικασία για τα 18 αρχεία δεδομένων:

1. Ομαλοποίηση των επιλεγμένων δεδομένων
2. Ομαδοποίηση σε 6 clusters,
  - a. παρουσίαση λίστας των σημείων και σε ποιο cluster ανήκουν
  - b. μετατροπή αποτελεσμάτων σε αρχείο “csv”
  - c. αξιολόγηση βάσει του δείκτη Silhouette (>0,5)
3. Ομαδοποίηση σε 4 clusters,
  - a. παρουσίαση λίστας των σημείων και σε ποιο cluster ανήκουν
  - b. μετατροπή αποτελεσμάτων σε αρχείο “csv”
  - c. αξιολόγηση βάσει του δείκτη Silhouette (>0,5)
4. Ομαδοποίηση σε 3 clusters,
  - a. παρουσίαση λίστας των σημείων και σε ποιο cluster ανήκουν
  - b. μετατροπή αποτελεσμάτων σε αρχείο “csv”
  - c. αξιολόγηση βάσει του δείκτη Silhouette (>0,5)

### 4.3 Ανάλυση Αποτελεσμάτων

Τα αποτελέσματα της μεθόδου «Dynamic Time Warping» που θα παρουσιαστούν και θα σχολιαστούν, θα αφορούν στις ώρες της **πρωινής αιχμής 8:00-09:00**, της **μεσημεριανής αιχμής, 14:00-15:00 & 15:00-16:00** και δύο ώρες **εκτός αιχμής, 11:00-**

**12:00** το πρωί και **23:00-00:00** το βράδυ. Για λόγους πληρότητας, όσα αποτελέσματα δεν σχολιαστούν θα βρίσκονται στο **ΠΑΡΑΡΤΗΜΑ**.

Όπως αναφέρθηκε και στην προηγούμενη ενότητα, τα συλλεχθέντα δεδομένα ομαδοποιήθηκαν ανά μία ώρα (18 χρονοσειρές) για όλα τα σημεία των φωρατών και με την μέθοδο DTW παρατηρήθηκαν πώς αυτά ομαδοποιούνται κατά τη διάρκεια της ημέρας. Πραγματοποιήθηκαν 3 παραδείγματα-ομαδοποιήσεις (clusters) για την κάθε ώρα της ημέρας, για 3, 4 και 6 clusters και στη συνέχεια αξιολογήθηκαν τα αποτελέσματα με βάση τον δείκτη Silhouette (Silhouette Index). Μία υψηλή τιμή του δείκτη ( $>0,5$ ) υποδηλώνει καλύτερη ποιότητα ομαδοποίησης. Στον Πίνακα 4.3 παρουσιάζονται τα αποτελέσματα της αξιολόγησης βάσει του δείκτη Silhouette. Για κάθε χρονοσειρά, παρατηρούνται με πράσινο χρώμα τα αποτελέσματα της αξιολόγησης και άρα ο αποδεκτός αριθμός cluster. Για παράδειγμα, στις 00:00-01:00 το βράδυ παρατηρήθηκε καλύτερη ομαδοποίηση σε 3 ομάδες (clusters), ενώ στις 07:00-08:00 το πρωί παρατηρήθηκε καλύτερη ομαδοποίηση σε 4 ομάδες.

Πίνακας 4.3- Αποτελέσματα δείκτη Silhouette

	00:00-01:00	07:00-08:00	08:00-09:00	09:00-10:00	10:00-11:00	11:00-12:00	12:00-13:00	13:00-14:00	14:00-15:00
<b>6 clusters</b>	0.418	0.429	0.406	0.412	0.436	0.425	0.455	0.396	0.472
<b>4 clusters</b>	0.514	<b>0.506</b>	0.448	<b>0.504</b>	0.531	0.424	<b>0.542</b>	0.527	0.502
<b>3 clusters</b>	<b>0.571</b>	0.493	<b>0.514</b>	0.464	<b>0.545</b>	<b>0.576</b>	0.527	<b>0.600</b>	<b>0.587</b>

	15:00-16:00	16:00-17:00	17:00-18:00	18:00-19:00	19:00-20:00	20:00-21:00	21:00-22:00	22:00-23:00	23:00-00:00
<b>6 clusters</b>	0.462	0.447	0.448	0.427	0.419	0.439	0.354	0.388	0.426
<b>4 clusters</b>	0.523	0.485	0.526	0.490	0.484	0.487	0.549	0.568	0.520
<b>3 clusters</b>	<b>0.576</b>	<b>0.572</b>	<b>0.538</b>	<b>0.551</b>	<b>0.567</b>	<b>0.540</b>	<b>0.552</b>	<b>0.578</b>	<b>0.559</b>

#### 4.3.1 Παρουσίαση Αποτελεσμάτων της μεθόδου DTW

Για κάθε ώρα της ημέρας, παρουσιάζονται οι μέσες χρονοσειρές των διαφορετικών ομάδων (clusters) που εντοπίστηκαν. Για το σχολιασμό των διαγραμμάτων με τα αποτελέσματα της μεθόδου DTW χρειάζεται να επεξηγηθούν, αρχικά, οι τιμές του άξονα y. Ο άξονας y (value) αποτελείται από κανονικοποιημένες τιμές, θετικές και αρνητικές, που δηλώνουν τη συμπεριφορά των διάφορων φωρατών σε σχέση με την

μέση τιμή, ώστε αυτές να ακολουθούν την κανονική κατανομή με μέσο όρο το μηδέν. Ορίζοντας ως το μέσο όρο το «μηδέν», παρατηρούμε τιμές να κυμαίνονται είτε κοντά σε αυτόν (τιμές κοντά στο μηδέν), είτε πολύ μακριά από αυτόν (πολύ θετικές, ή πολύ αρνητικές τιμές). Στην ουσία, αυτή η κανονικοποίηση έχει γίνει από τη μέθοδο ώστε οι τιμές των χρονοσειρών να «έρθουν» στην ίδια κλίμακα, καθώς δύο χρονοσειρές μπορεί να παρουσιάζουν ίδιο σχήμα, δηλαδή παρόμοιο κυκλοφοριακό μοτίβο, αλλά διαφορετική ζήτηση. Για παράδειγμα, από ένα φωρατή που βρίσκεται σε κεντρικό σημείο περνάνε πολλά αυτοκίνητα και σε έναν άλλο που βρίσκεται μακριά από το κέντρο περνάνε πολύ λιγότερα. Αυτό το σύνηθες φαινόμενο επιλύεται με τη χρήση του παρακάτω τύπου για την κανονικοποίηση των τιμών:

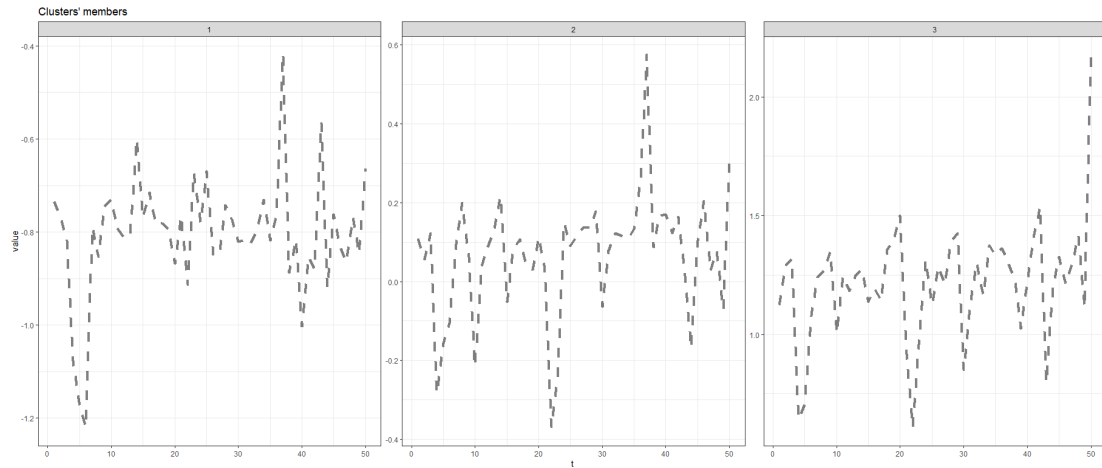
$$X_{norm} = \frac{X_i - \mu}{\sigma}$$

όπου  $\mu$  η μέση τιμή και  $\sigma$  η τυπική απόκλιση.

- **08:00-09:00 –3 ΟΜΑΔΕΣ**

Στο *Διάγραμμα 4.1* παρατηρείται ότι στη 2<sup>η</sup> ομάδα έχουν ομαδοποιηθεί φωρατές οι οποίοι παρουσιάζουν φυσιολογικό φόρτο στην ώρα αιχμής, καθώς βρίσκονται κοντά στη μέση τιμή (0). Οι φωρατές της 1<sup>ης</sup> ομάδας βρίσκονται πιθανώς σε περιοχές μακριά από το κέντρο, στις οποίες παρατηρείται μικρότερος φόρτος από τον μέσο όρο, καθώς παρατηρούνται τιμές 0,8 φορές την τυπική απόκλιση ( $0,8 \cdot \sigma$ ) μικρότερες από τη μέση τιμή. Στην 3<sup>η</sup> ομάδα, παρατηρούνται τιμές 1 έως 2 φορές την τυπική απόκλιση πάνω από τον μέσο όρο και εκεί ανήκουν φωρατές οι οποίοι βρίσκονται σε περιοχές όπως το κέντρο όπου εμφανίζεται μεγαλύτερος φόρτος από τον φυσιολογικό κατά την ώρα αιχμής.

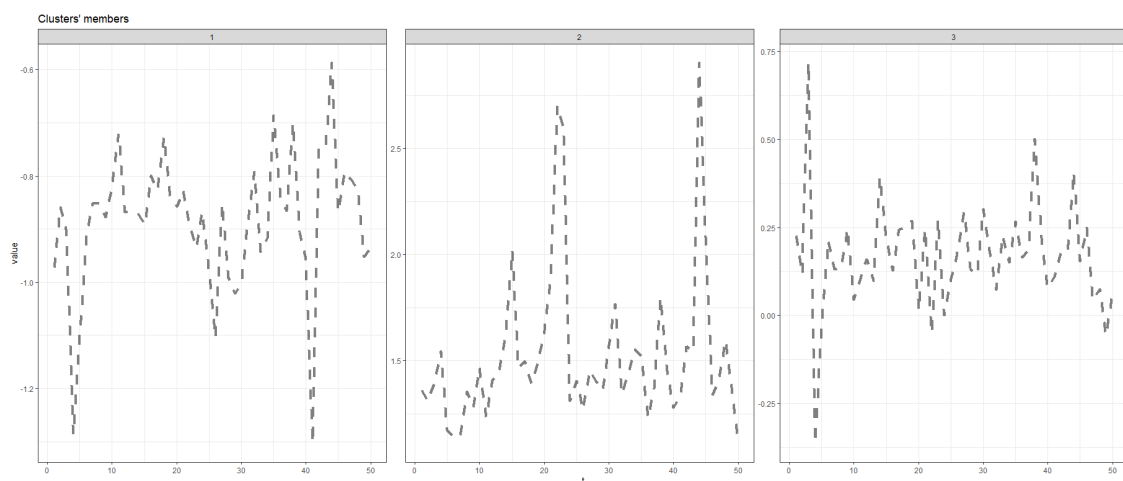




Διάγραμμα 4.1- Αποτελέσματα 3 ομάδων για τις 08:00-09:00

- **11:00-12:00 –3 ΟΜΑΔΕΣ**

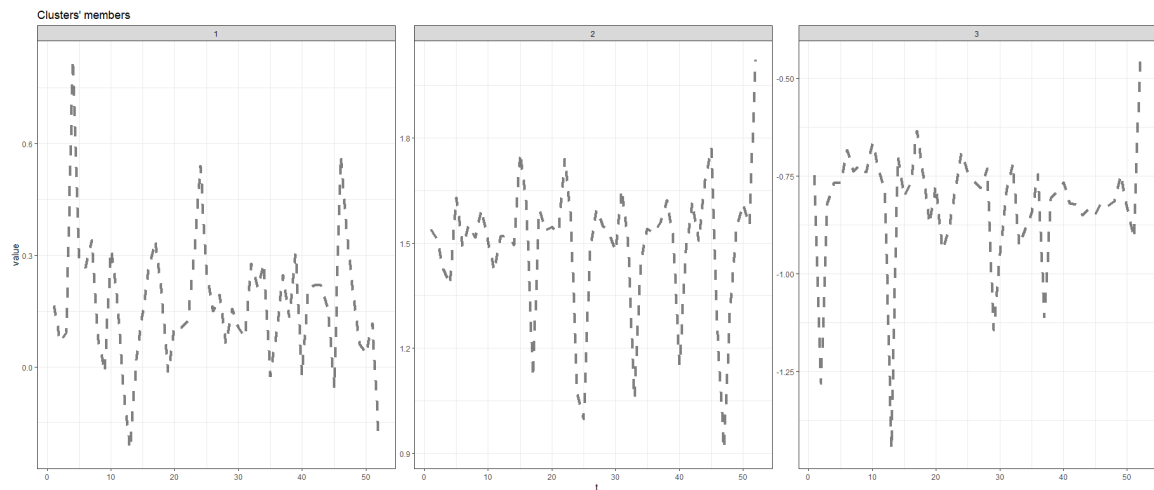
Παρατηρώντας το Διάγραμμα 4.2, διαπιστώνεται ότι στην 3<sup>η</sup> ομάδα έχουν ομαδοποιηθεί φορατές οι οποίοι κατέγραψαν φυσιολογικές τιμές φόρτου, κοντά στο «μηδέν», για την ώρα 11:00-12:00 το πρωί που βρίσκεται εκτός αιχμής. Στην 1<sup>η</sup> ομάδα κατατάχθηκαν φορατές με τιμές φόρτου λίγο μικρότερες της μέσης τιμής, κυρίως κοντά στο -0,8. Ενώ, τέλος, στη 2<sup>η</sup> ομάδα οι φορατές παρουσίασαν τιμές πολύ παραπάνω της μέσης τιμής, κοντά στο +1,5, ενώ κάποιοι φορατές έφτασαν και το +2,5. Αυτό σημαίνει ότι σε κάποιους δρόμους, υπήρξε μεγαλύτερος κυκλοφοριακός φόρτος παρόλο που η ώρα είναι εκτός της «συνηθισμένης» αιχμής.



Διάγραμμα 4.2- Αποτελέσματα 3 ομάδων για τις 11:00-12:00

- **14:00-15:00 –3 ΟΜΑΔΕΣ**

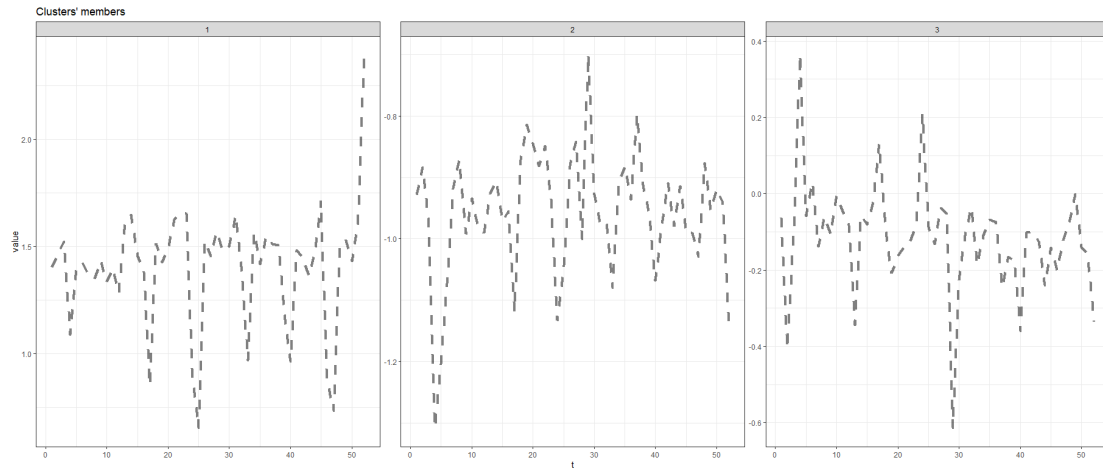
Στο Διάγραμμα 4.3 παρατηρείται ότι οι φωρατές, που κατέγραψαν φυσιολογικές τιμές φόρτου σε ώρα εντός αιχμής, έχουν ομαδοποιηθεί στην 1<sup>η</sup> ομάδα. Στη 2<sup>η</sup> ομάδα κατατάχθηκαν φωρατές με πολύ υψηλότερες από τη μέση τιμή τιμές (+1,2 έως +1,8). Ενώ στην 3<sup>η</sup> ομάδα οι φωρατές κατέγραψαν κυκλοφοριακό φόρτο λίγο κάτω από τον μέσο όρο (-0,75).



Διάγραμμα 4.3-Αποτελέσματα 3 ομάδων για τις 14:00-15:00

- **15:00-16:00 –3 ΟΜΑΔΕΣ**

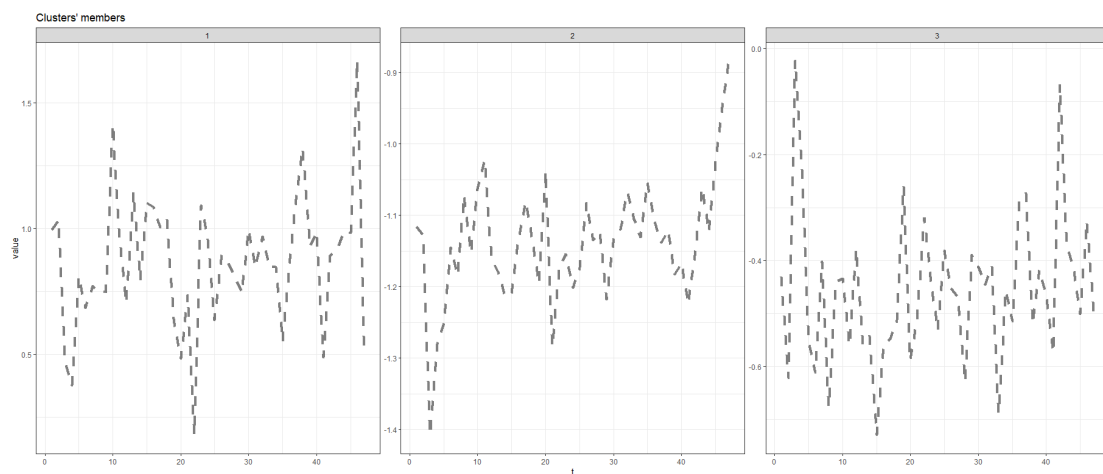
Στο Διάγραμμα 4.4 στην 3<sup>η</sup> ομάδα έχουν ομαδοποιηθεί οι φωρατές με την αναμενόμενη κίνηση κατά την ώρα αιχμής, ενώ στην 1<sup>η</sup> ομάδα έχουν ομαδοποιηθεί οι φωρατές σε οδούς όπου ο φόρτος κατά την ώρα της αιχμής ήταν μεγαλύτερος από το φυσιολογικό. Τέλος, στην 2<sup>η</sup> ομάδα βρίσκονται οι φωρατές σε περιοχές όπου καταγράφηκε λιγότερη κίνηση.



Διάγραμμα 4.4-Αποτελέσματα 3 ομάδων για τις 15:00-16:00

- **23:00-00:00 – 3 ΟΜΑΔΕΣ**

Στο Διάγραμμα 4.5 στο διάστημα 23:00-00:00 το βράδυ, παρατηρείται μειωμένος κυκλοφοριακός φόρτος στους φωρατές των ομάδων 2 και 3, ενώ στην 1<sup>η</sup> ομάδα παρατηρούνται τιμές λίγο μεγαλύτερες από την μέση τιμή.



Διάγραμμα 4.5-Αποτελέσματα 3 ομάδων για τις 23:00-00:00





#### 4.4 Οπτικοποίηση Αποτελεσμάτων σε Χάρτες

Στην παρούσα ενότητα πρόκειται να αποτυπωθούν σε χάρτες τα αποτελέσματα της ανάλυσης, που πραγματοποιήθηκε στο προηγούμενο κεφάλαιο. Πιο συγκεκριμένα, έγινε χρήση του προγράμματος “QGIS”, το οποίο είναι ένα λογισμικό Γεωγραφικών Συστημάτων Πληροφοριών δομημένο σε ανοιχτό και ελεύθερο κώδικα, φιλικό προς τον χρήστη.

Όσον αφορά στους χάρτες, ο σχολιασμός θα γίνει για 4 περιοχές μελέτης:

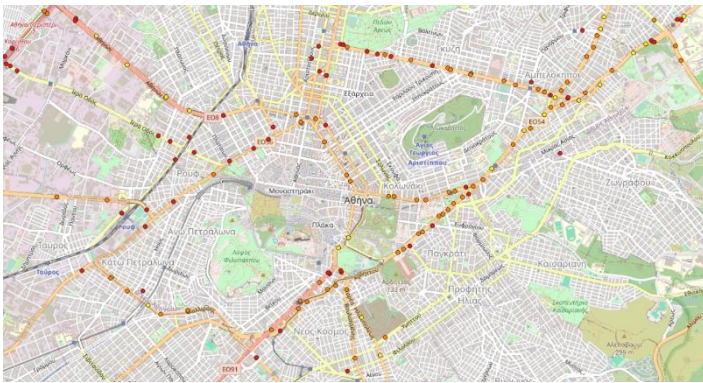
- i. Κέντρο,
- ii. Ανατολικά & Βόρεια Προάστια (Λεωφ. Κηφισίας & Λεωφ. Μεσογείων),
- iii. Νότια Προάστια (Λεωφ. Βουλιαγμένης & Λεωφ. Ποσειδάωνος) και
- iv. Δυτικά Προάστια (Λεωφ. Κηφισού και Πειραιάς)

Για τον κάθε φωρατή, σε κάθε χάρτη, θα αντιστοιχεί ένα χρώμα, ανάλογα με την ομάδα που βρίσκεται. Πιο συγκεκριμένα:

-  Κίτρινο, για ακραίες τιμές, πάνω από τον μέσο όρο
-  Κόκκινο, για ακραίες τιμές, κάτω από τον μέσο όρο
-  Πορτοκαλί, για τιμές κοντά στον μέσο όρο
-  Μουσταρδί, για τιμές λίγο πάνω από τον μέσο όρο (για τις περιπτώσεις που έχουμε 4 ομαδοποιήσεις, αντί για 3).

#### 4.4.1 Κέντρο Αθηνών

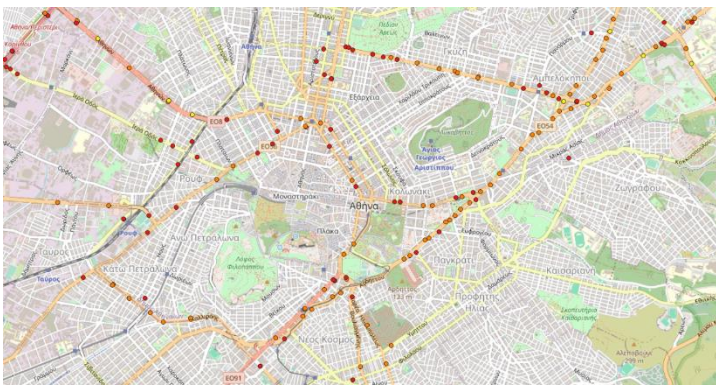
Παρατηρώντας τις παρακάτω εικόνες γίνεται αντιληπτό ότι κατά την ώρα της πρωινής αιχμής στις 08:00-09:00 ο φόρτος ήταν υψηλότερος από τον μέσο στον δακτύλιο, ενώ τις υπόλοιπες ώρες η πληθώρα των φορατών κατέγραψαν τιμές φόρτου κοντά στον μέσο όρο, με εξαίρεση το διάστημα 23:00-00:00 το βράδυ όπου στο τμήμα της Λεωφόρου Βασιλέως Κωνσταντίνου και Βασιλίσσης Σοφίας παρατηρήθηκαν τιμές υψηλότερες από το μέσο όρο.



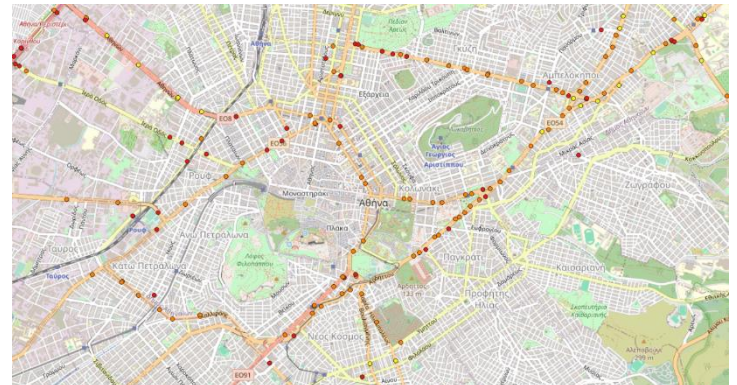
Εικόνα 4.2- Κυκλοφοριακός Φόρτος στο Κέντρο της Αθήνας στις 08:00-09:00



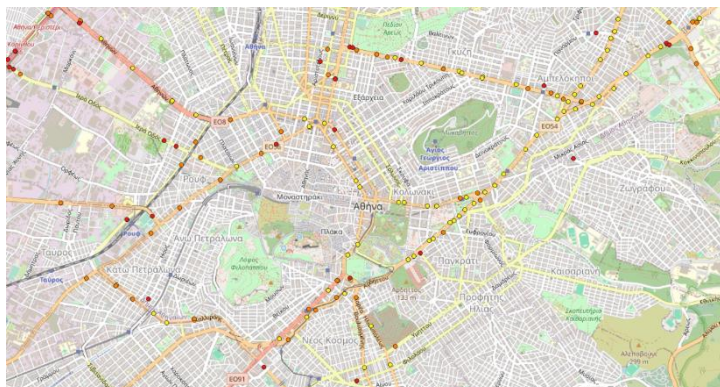
Εικόνα 4.3 - Κυκλοφοριακός Φόρτος στο Κέντρο της Αθήνας στις 11:00-12:00



Εικόνα 4.4- Κυκλοφοριακός Φόρτος στο Κέντρο της Αθήνας στις 14:00-15:00



Εικόνα 4.5- Κυκλοφοριακός Φόρτος στο Κέντρο της Αθήνας στις 15:00-16:00

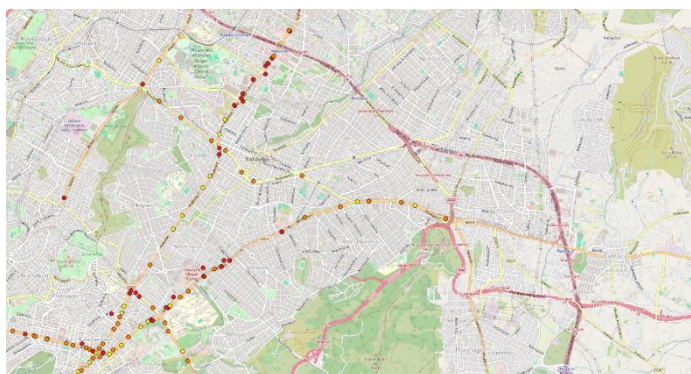


Εικόνα 4.6- Κυκλοφοριακός Φόρτος στο Κέντρο της Αθήνας στις 23:00-00:00

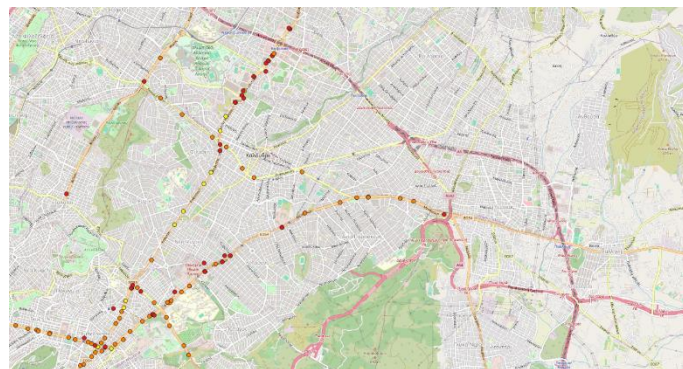


#### 4.4.2 Ανατολικά και Βόρεια Προάστια (Λ. Κηφισίας & Λ. Μεσογείων)

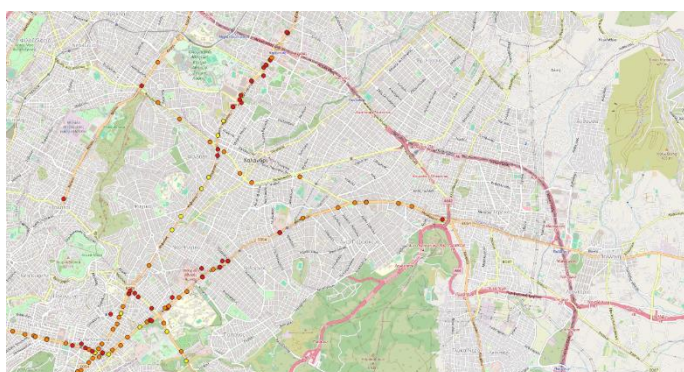
Όσον αφορά τα Ανατολικά και Βόρεια Προάστια, στη Λεωφόρο Κηφισίας παρατηρήθηκαν τιμές κάτω από τον μέσο όρο κυρίως στο βόρειο τμήμα της Λεωφόρου, ενώ στο νότιο τμήμα παρατηρήθηκαν τιμές κυκλοφοριακού φόρτου πάνω από τον μέσο όρο. Στη Λεωφόρο Μεσογείων, οι τιμές φόρτου κυμάνθηκαν κυρίως κοντά στον μέσο όρο με εξαίρεση κάποιων κόμβων όπου παρατηρήθηκαν τιμές πάνω από το μέσο όρο.



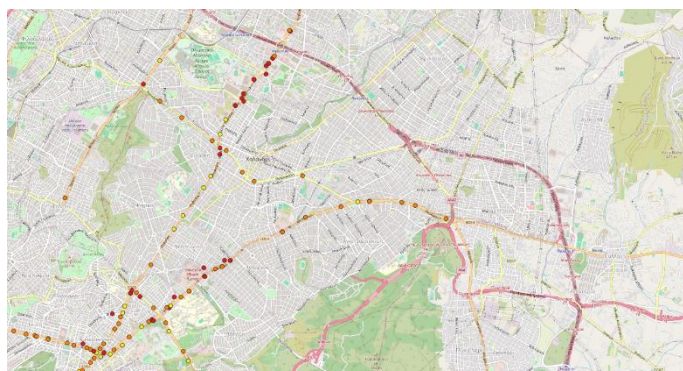
Εικόνα 4.7- Κυκλοφοριακός Φόρτος στα Ανατολικά & Βόρεια Προάστια της Αθήνας στις 08:00-09:00



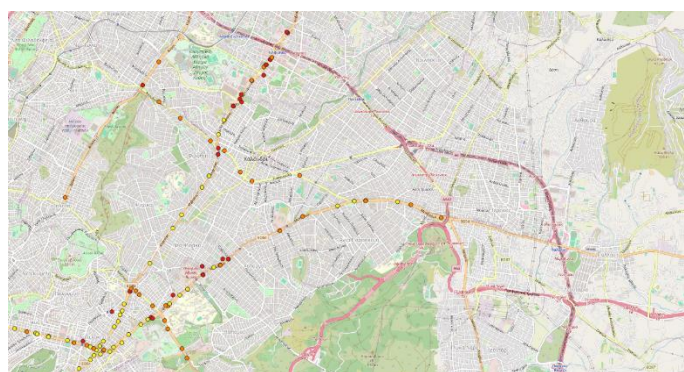
Εικόνα 4.8- Κυκλοφοριακός Φόρτος στα Ανατολικά & Βόρεια Προάστια της Αθήνας στις 11:00-12:00



Εικόνα 4.9- Κυκλοφοριακός Φόρτος στα Ανατολικά & Βόρεια Προάστια της Αθήνας στις 14:00-15:00



Εικόνα 4.10- Κυκλοφοριακός Φόρτος στα Ανατολικά & Βόρεια Προάστια της Αθήνας στις 15:00-16:00

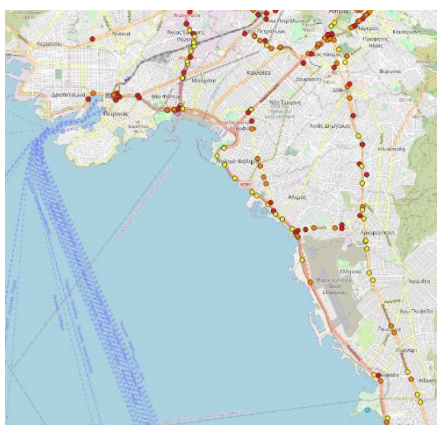


Εικόνα 4.11- Κυκλοφοριακός Φόρτος στα Ανατολικά & Βόρεια Προάστια της Αθήνας στις 23:00-00:00

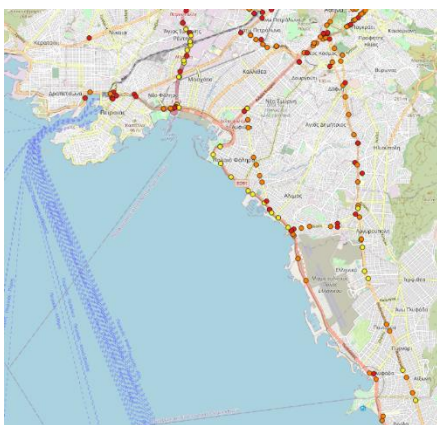


#### 4.4.3 Νότια Προάστια (Λ. Βουλιαγμένης & Λ. Ποσειδώνος)

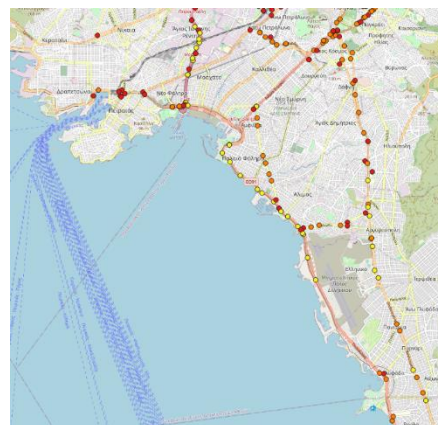
Στα Νότια Προάστια, στη Λεωφόρο Βουλιαγμένης, το διάστημα 08:00-09:00 το πρωί, 15:00-16:00 το απόγευμα και 23:00-00:00 το βράδυ ο φόρτος των διερχόμενων αυτοκινήτων ήταν μεγαλύτερος από τον αναμενόμενο, ενώ τα διαστήματα 11:00-12:00 και 14:00-15:00 το μεσημέρι παρατηρήθηκε κοντά στη μέση τιμή. Στη Λεωφόρο Ποσειδώνος, ο φόρτος παρατηρήθηκε μεγαλύτερος από τον αναμενόμενο καθ' όλη τη διάρκεια της ημέρας.



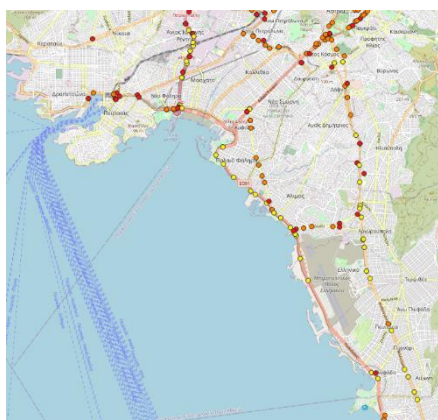
Εικόνα 4.12- Κυκλοφοριακός Φόρτος στα Νότια Προάστια της Αθήνας στις 08:00-09:00



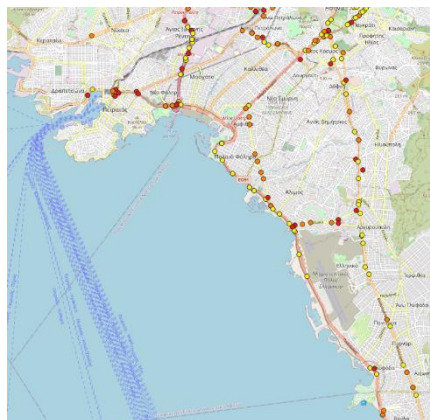
Εικόνα 4.13- Κυκλοφοριακός Φόρτος στα Νότια Προάστια της Αθήνας στις 11:00-12:00



Εικόνα 4.14- Κυκλοφοριακός Φόρτος στα Νότια Προάστια της Αθήνας στις 14:00-15:00



Εικόνα 4.15- Κυκλοφοριακός Φόρτος στα Νότια Προάστια της Αθήνας στις 15:00-16:00

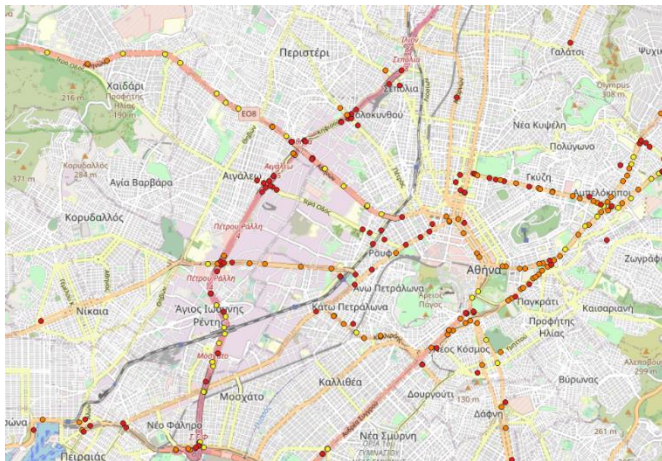


Εικόνα 4.16- Κυκλοφοριακός Φόρτος στα Νότια Προάστια της Αθήνας στις 23:00-00:00

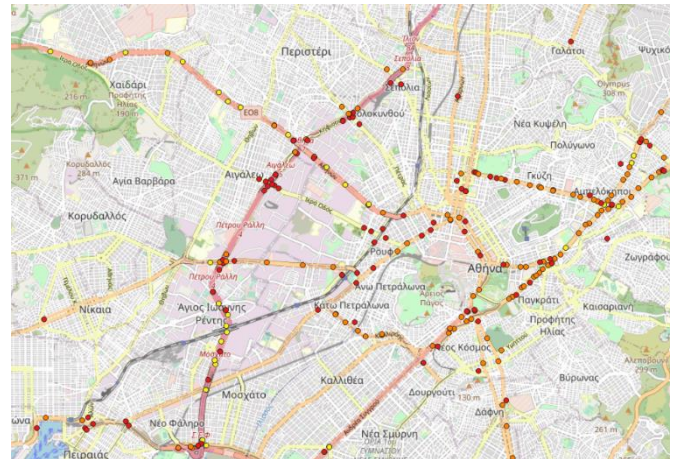


#### 4.4.4 Δυτικά Προάστια (Λ. Κηφισού & Πειραιάς)

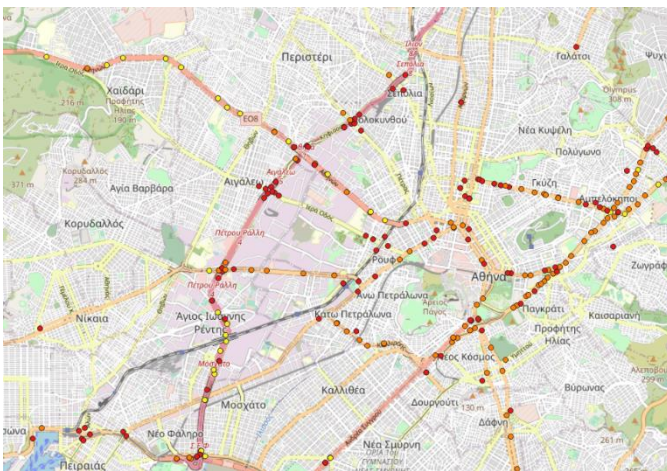
Τέλος, στα Δυτικά Προάστια οι τιμές του φόρτου παρέμειναν κάτω από τον μέσο όρο σε όλα τα εξεταζόμενα χρονικά διαστήματα, με εξαίρεση το σημείο της Λεωφόρου Κηφισού κοντά στον Άγιο Ιωάννη Ρέντη, όπου παρατηρήθηκαν τιμές πάνω από το μέσο όρο.



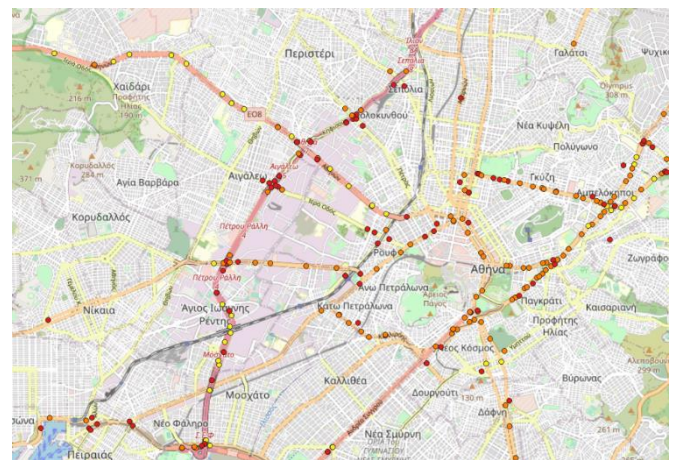
Εικόνα 4.17- Κυκλοφοριακός Φόρτος στα Δυτικά Προάστια της Αθήνας στις 08:00-09:00



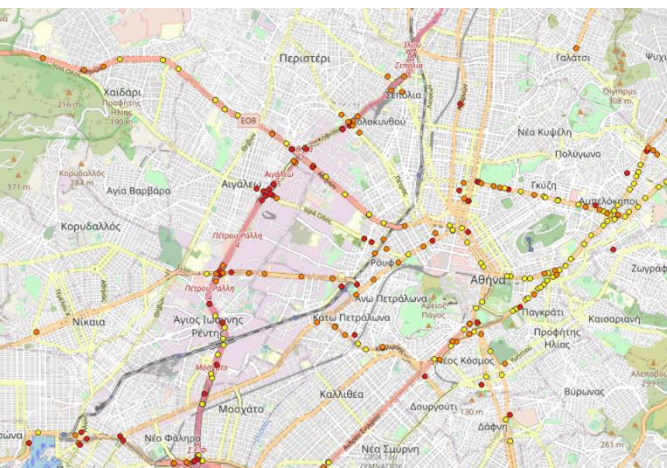
Εικόνα 4.18- Κυκλοφοριακός Φόρτος στα Δυτικά Προάστια της Αθήνας στις 11:00-12:00



Εικόνα 4.19- Κυκλοφοριακός Φόρτος στα Δυτικά Προάστια της Αθήνας στις 14:00-15:00



Εικόνα 4.20- Κυκλοφοριακός Φόρτος στα Δυτικά Προάστια της Αθήνας στις 15:00-16:00



Εικόνα 4.21- Κυκλοφοριακός Φόρτος στα Δυτικά Προάστια της Αθήνας στις 23:00-00:00



## 5. Συμπεράσματα και Προτάσεις

### 5.1 Βασικά Συμπεράσματα

Η παρακολούθηση και ανάλυση του κυκλοφοριακού φόρτου είναι απαραίτητη για την αποτελεσματική λειτουργία των μεταφορικών δικτύων και την αντιμετώπιση προβλημάτων κυκλοφορίας. Μέσω της παρακολούθησης του κυκλοφοριακού φόρτου μπορούν να αναγνωριστούν σημεία συμφόρησης και προβλήματα κίνησης, ώστε να ληφθούν οι απαραίτητες δράσεις για τη βελτίωση της κυκλοφορίας. Επιπλέον, η ανάλυση του κυκλοφοριακού φορτίου μπορεί να παράσχει πολύτιμες πληροφορίες για τον σχεδιασμό και την ανάπτυξη των μεταφορικών υποδομών, με στόχο τη βελτίωση της ασφάλειας και τη μείωση του χρόνου διαδρομής. Η συστηματική παρακολούθηση και ανάλυση του κυκλοφοριακού φορτίου αποτελεί ένα ουσιαστικό εργαλείο για τη βελτίωση της κίνησης και την αποτελεσματική διαχείριση του οδικού δικτύου.

Στην παρούσα Διπλωματική Εργασία εντοπίστηκαν και αναλύθηκαν χωροχρονικά μοτίβα στην ευρύτερη περιοχή της Αθήνας με μη επιβλεπόμενη μάθηση, χρησιμοποιώντας την μέθοδο DTW. Για την ανάλυση χρησιμοποιήθηκαν δεδομένα φόρτων για την περίοδο Φεβρουαρίου-Μαρτίου 2023, τα οποία έχουν ληφθεί από φωρατές ανίχνευσης κυκλοφορίας τοποθετημένους ανά τακτές αποστάσεις σε όλη την περιοχή μελέτης. Τα δεδομένα ομαδοποιήθηκαν ικανοποιητικά σε ομάδες σαφώς διακριτές μεταξύ τους και στη συνέχεια αποτυπώθηκαν σε χάρτες για την καλύτερη κατανόηση της διακύμανσης του κυκλοφοριακού φόρτου.

Από την ανάλυση των χρονοσειρών ημερήσιας μεταβολής του ωριαίου φόρτου όλων των φωρατών προέκυψαν 3 διαφορετικά μοτίβα με διαφορετική διακύμανση φόρτων ανά ημέρα. Αυτά που παρουσιάζουν μικρή ημερήσια διακύμανση, και άλλα που έχουν μεγαλύτερη ημερήσια διακύμανση. Είναι προφανές ότι τα μοτίβα με μικρή διακύμανση είναι και πιο προβλέψιμα.

Τα μοτίβα αυτά εμφανίζονται σε διαφορετικά σημεία του δικτύου ανά ώρα δίνοντας τη δυνατότητα για μελλοντική ανάλυση της χωροχρονικής εξέλιξης της κυκλοφορίας.

Από τη σε βάθος ανάλυση των μοτίβων ανά περιοχή, παρατηρήθηκε ότι στο Κέντρο της Αθήνας και στα Νότια Προάστια εξαιτίας της λειτουργίας του δικτύου στην ικανότητα αρκετές ώρες την ημέρα, τα μοτίβα κυκλοφορίας που εντοπίζονται είναι συνεχώς πάνω από το μέσο μοτίβο ημερήσιας χρονικής εξέλιξης ωριαίου φόρτου και

εμφανίζουν μεγαλύτερη διακύμανση. Στις υπόλοιπες περιοχές οι τιμές που καταγράφηκαν ήταν οι αναμενόμενες, δηλαδή κοντά στις φυσιολογικές για τις ώρες εντός και εκτός αιχμής. Είναι προφανές ότι η σε βάθος ανάλυση των μοτίβων σε συγκεκριμένες περιοχές θα μπορούσε να χρησιμοποιηθεί ως εργαλείο διερεύνησης της έντασης και διάρκειας του κορεσμού.

## 5.2 Περιορισμοί Έρευνας

Κατά την εκπόνηση της παρούσας Διπλωματικής Εργασίας εντοπίστηκαν κάποιοι περιορισμοί. Η βάση δεδομένων ήταν μικρή, αφορούσε τους μήνες Φεβρουάριο και Μάρτιο. Αν τα δεδομένα αντιστοιχούσαν σε ένα μεγαλύτερο χρονικό διάστημα, για παράδειγμα ενός έτους, τότε τα αποτελέσματα θα μπορούσαν να είναι πιο γενικεύσιμα. Επίσης, για τον εντοπισμό πιο βραχυχρόνιων μοτίβων θα ήταν προτιμότερο η ανάλυση να είχε γίνει σε δεδομένα ανά μικρότερο χρονικό διάστημα, για παράδειγμα ανά 5 λεπτά. Λόγω έλλειψης δεδομένων ανά 5 λεπτά, όμως, χρησιμοποιήθηκαν δεδομένα ανά μία ώρα.

## 5.3 Προτάσεις για περαιτέρω έρευνα

Με την ολοκλήρωση της παρούσας Διπλωματικής Εργασίας και βάσει των αποτελεσμάτων και των συμπερασμάτων που προέκυψαν, ακολουθούν ορισμένες προτάσεις για περαιτέρω έρευνα. Αρχικά, τα αποτελέσματα της παρούσας έρευνας θα μπορούσαν να αξιοποιηθούν σε μελλοντικές έρευνες για τη βελτίωση της ακρίβειας μοντέλων πρόβλεψης, καθώς και για μελλοντική ανάλυση της χωροχρονικής εξέλιξης της κυκλοφορίας. Ενδιαφέρον θα είχε, επίσης, να γίνει μια εκτενής έρευνα που θα αφορά συνδυαστικά και τα αυτοκίνητα και τα άλλα MMM, ώστε να διαπιστωθεί πώς η αύξηση του κυκλοφοριακού φόρτου στις οδούς επηρεάζει την κίνηση των υπόλοιπων MMM. Τέλος, η μέθοδος DTW θα μπορούσε να εφαρμοστεί και σε άλλες πόλεις ώστε να διαπιστωθεί αν θα προκύπταν παρόμοια συμπεράσματα, ανάλυση η οποία θα μπορούσε να χρησιμοποιηθεί ως εργαλείο διερεύνησης της έντασης και διάρκειας του κορεσμού.

## BIBΛΙΟΓΡΑΦΙΑ

- [1] Antonis Chaziris, George Yannis, A critical assessment of Athens Traffic Restrictions using multiple data sources, 2022
- [2] Carmelo Cassisi et al., Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining, Advances in Data Mining Knowledge Discovery and Applications, Chapter 3, 2012
- [3] Constantinos Antoniou, Haris N. Koutsopoulos, George Yannis, Dynamic data-driven local traffic state estimation and prediction, Transportation Research Part C: Emerging Technologies, Volume 34, Pages 89-107, 2013
- [4] Eleni I. Vlahogianni , John C. Golias & Matthew G. Karlaftis (2004) Short-term traffic forecasting: Overview of objectives and methods, Transport Reviews, 24:5, 533-557, DOI: 10.1080/0144164042000195072
- [5] Fafoutellis, P., Vlahogianni, E.I. Unlocking the Full Potential of Deep Learning in Traffic Forecasting Through Road Network Representations: A Critical Review. Data Sci. Transp. 5, 23 (2023)
- [6] Fattah, Md. Abdul, Syed Riad Morshed, and Abdulla-Al Kafy. "Insights into the Socio-Economic Impacts of Traffic Congestion in the Port and Industrial Areas of Chittagong City, Bangladesh." Transportation Engineering 9 (2022)
- [7] Guo, S., Zhou, D., Fan, J. et al. Identifying the most influential roads based on traffic correlation networks. EPJ Data Sci. 8, 28 (2019)
- [8] Higgins, Christopher D., Matthias N. Sweet, and Pavlos S. Kanaroglou. "All minutes are not equal: travel time and the effects of congestion on commute satisfaction in Canadian cities." *Transportation* 45 (2018): 1249-1268. )
- [9] J. Wan, J. Liu, Z. Shao, A. V. Vasilakos, M. Imran, and K. Zhou, "Mobile crowd sensing for traffic prediction in internet of vehicles," Sensors, vol. 16, no. 1, p. 88, 2016.
- [10] Khalid Mohammed Almatar, Traffic congestion patterns in the urban road network: (Dammam metropolitan area), Ain Shams Engineering Journal, Volume 14, Issue 3, 2023
- [11] Liu, Haiyang, Chunjiang Zhu, Detian Zhang and Qing Li. "Attention-based Spatial-Temporal Graph Convolutional Recurrent Networks for Traffic Forecasting." *International Conference on Advanced Data Mining and Applications* (2023).

- [12] Pavel Senin, “Dynamic Time Warping Algorithm Review”, 2008
- [13] Ryu, U., Wang, J., Pak, U. et al. A clustering based traffic flow prediction method with dynamic spatiotemporal correlation analysis. *Transportation* 49, 951–988 (2022)
- [14] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, Teh Ying Wah, Time-series clustering – A decade review, *Information Systems*, Volume 53, 2015, Pages 16-38
- [15] W. Zhang, Z. Wu, X. Zhang, G. Song, Y. Wang and J. Chen, "Robust and Hierarchical Spatial Relation Analysis for Traffic Forecasting," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 1, pp. 201-217, Jan. 2023
- [16] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Y. Wang, “Traffic flow prediction with big data: A deep learning approach,” *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Dec. 2015.
- [17] Y. Zhao, Y. Lin, H. Wen, T. Wei, X. Jin and H. Wan, "Spatial-Temporal Position-Aware Graph Convolution Networks for Traffic Flow Forecasting," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 8, pp. 8650-8666, 2023
- [18] Yu Sang Chang, Yong Joo Lee, Sung Sup Brian Choi, Is there more traffic congestion in larger cities? -Scaling analysis of the 101 largest U.S. urban centers-, *Transport Policy*, Volume 59, 2017, Pages 54-63
- [19] Yuebing Liang, Guan Huang, Zhan Zhao, Joint demand prediction for multimodal systems: A multi-task multi-relational spatiotemporal graph neural network approach, *Transportation Research Part C: Emerging Technologies*, Volume 140, 2022
- [20] I. M. Φραντζεσκάκης, I.K. Γκόλιας, M.X. Πίτσιαβα-Λατινοπούλου, 2009. *Κυκλοφοριακή Τεχνική*. Εκδόσεις Παπασωτηρίου.
- [21] Google: [https://en.wikipedia.org/wiki/Traffic\\_congestion](https://en.wikipedia.org/wiki/Traffic_congestion)
- [22] Google: <https://blog.goodvisionlive.com/the-ripple-effect-of-traffic-congestion-in-cities>
- [23] Google: [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Transport\\_statistics\\_at\\_regional\\_level#Road\\_transport\\_and\\_accidents](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Transport_statistics_at_regional_level#Road_transport_and_accidents)
- [24] Google: <https://www.tomtom.com/traffic-index/athens-traffic/>
- [25] Google: [https://data.gov.gr/datasets/road\\_traffic\\_attica/](https://data.gov.gr/datasets/road_traffic_attica/)

## ΠΑΡΑΡΤΗΜΑ

### 1) Κώδικας (ενδεικτικά για το διάστημα 08:00-09:00)

```
#install packages
install.packages('dtwclust')
install.packages("ggplot2")
install.packages("BBmisc")
install.packages("TSclust")
library("ggplot2")
library("TSclust")
library("dtwclust")
library(tidyr)
library(dplyr)
library(dtw)
library(BBmisc)
library(TSdist)

#####
#data processing:
# 1)diagrafh fwratwn me mh swstes methrseis h' me polles mhdenikes: [111, 113, 115, 117, 121, 123, 125, 134, 140, 143, 145, 204, 340, 341,
362, 364, 365, 470, 471, 472, 473, 485, 487, 489, 502, 690, 691, 730, 738, 747, 748, 749, 757]
# 2)dhmiourgia 18 excel arxeiwn me onoma "datai", opou i=1,2,3,.. gia kathe evros wras:
#### 1.00:00-01:00
#### 2.07:00-08:00
#### 3.08:00-09:00
#### 4.09:00-10:00
#### 5.10:00-11:00
#### 6.11:00-12:00
#### 7.12:00-13:00
#### 8.13:00-14:00
#### 9.14:00-15:00
#### 10.15:00-16:00
#### 11.16:00-17:00
#### 12.17:00-18:00
#### 13.18:00-19:00
#### 14.19:00-20:00
#### 15.20:00-21:00
#### 16.21:00-22:00
#### 17.22:00-23:00
#### 18.23:00-00:00

#import data
#DATA1 BETWEEN 00:00-01:00 EVERYDAY
##### CSV ME ROWS TO COLUMNS data1<- read.csv("C:/Users/kelly/Desktop/Befani_Diplomatiki/data1.csv")

data1<- read.csv("C:/Users/kelly/Desktop/Befani_Diplomatiki/sorted_data/traffic_feb_mar_23_1.csv") #date and time as columns

#dropping variables from dataset
data1$DATE<- NULL #dropping column with DATE
data1$TIME<-NULL #dropping column with TIME
#data1

# time series - diagramma BETWEEN 00:00-01:00 EVERYDAY FOR MS106 ONLY
ggplot(data1, aes(x=COUNT, y=MS106)) +
  geom_line() +
  xlab("everyday at 00:00-01:00")+ylab("value of MS106")

#dropping extra variables from dataset to prepare for clustering
data1$COUNT<- NULL #dropping column with COUNT

#turning rows to columns in data1
final_data1 <- as.data.frame(t(data1))
final_data1

#####
#NUMBER OF CLUSTERS: SIX
#####
```

```
#5 k-medoids - Normalising the chosen data and performing clustering using PAM and DTW distance
head(final_data1)
final_data1.norm <-BBmisc::normalize(final_data1, method="standardize")
```

```
#clustering with 6 clusters
clust.pam <- tsclust(final_data1.norm, type="partitional", k=6L, distance="dtw", clustering="pam")
plot(clust.pam, type = "centroids", clus = 1L:6L)
plot(clust.pam, type = "centroids", clus = 1L)
plot(clust.pam, type = "centroids", clus = 2L)
plot(clust.pam, type = "centroids", clus = 3L)
plot(clust.pam, type = "centroids", clus = 4L)
plot(clust.pam, type = "centroids", clus = 5L)
plot(clust.pam, type = "centroids", clus = 6L)
```

```
#Below is a list of points and which cluster they belong to.
data1_6_pam<-t(cbind(final_data1[,0], cluster = clust.pam@cluster))
write.csv(data1_6_pam, "data1_6_pam.csv")
```

```
#plot for each cluster
plot(clust.pam, type = "sc", clus = 1L) #type="sc" is actually a combination of series and centroids
plot(clust.pam, type = "sc", clus = 2L)
plot(clust.pam, type = "sc", clus = 3L)
plot(clust.pam, type = "sc", clus = 4L)
plot(clust.pam, type = "sc", clus = 5L)
plot(clust.pam, type = "sc", clus = 6L)
```

```
plot(clust.pam, type = "series", clus = 1L)
plot(clust.pam, type = "series", clus = 2L)
plot(clust.pam, type = "series", clus = 3L)
plot(clust.pam, type = "series", clus = 4L)
plot(clust.pam, type = "series", clus = 5L)
plot(clust.pam, type = "series", clus = 6L)
```

```
plot(clust.pam, type = "centroids", clus = 1L)
plot(clust.pam, type = "centroids", clus = 2L)
plot(clust.pam, type = "centroids", clus = 3L)
plot(clust.pam, type = "centroids", clus = 4L)
plot(clust.pam, type = "centroids", clus = 5L)
plot(clust.pam, type = "centroids", clus = 6L)
```

```
#EVALUATION
cvi(clust.pam)
```

```
#####
#REDUCING NUMBER OF CLUSTERS-----> 4 CLUSTERS
#####
```

```
#NUMBER OF CLUSTERS: 4
#5 k-medoids - Normalising the chosen data and performing clustering using PAM and DTW distance
head(final_data1)
final_data1.norm <-BBmisc::normalize(final_data1, method="standardize")
```

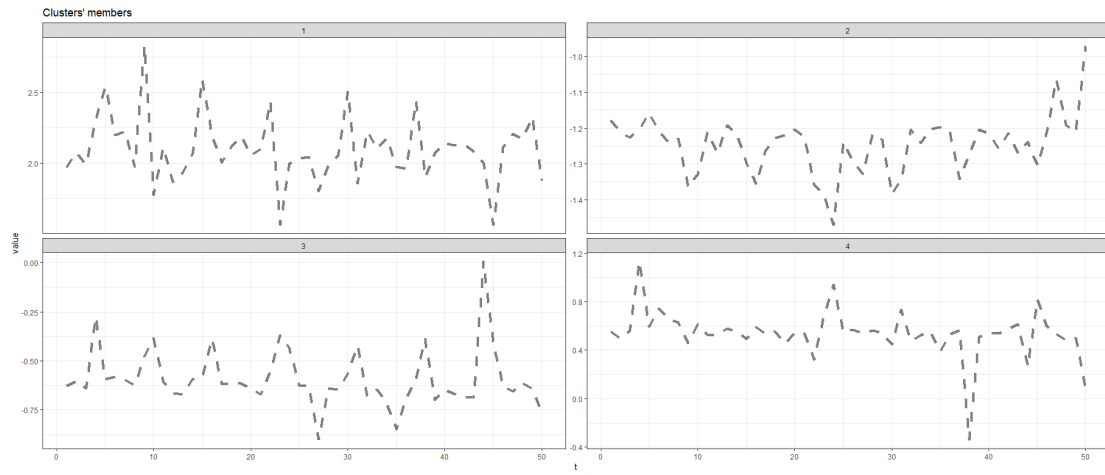
```
#clustering with 4 clusters
clust.pam <- tsclust(final_data1.norm, type="partitional", k=4L, distance="dtw", clustering="pam")
plot(clust.pam, type = "centroids")
plot(clust.pam, type = "centroids", clus = 1L:4L)
plot(clust.pam, type = "centroids", clus = 1L)
plot(clust.pam, type = "centroids", clus = 2L)
plot(clust.pam, type = "centroids", clus = 3L)
plot(clust.pam, type = "centroids", clus = 4L)
```

```
#Below is a list of points and which cluster they belong to.
data1_4_pam<-t(cbind(final_data1[,0], cluster = clust.pam@cluster))
write.csv(data1_4_pam, "data1_4_pam.csv")
```

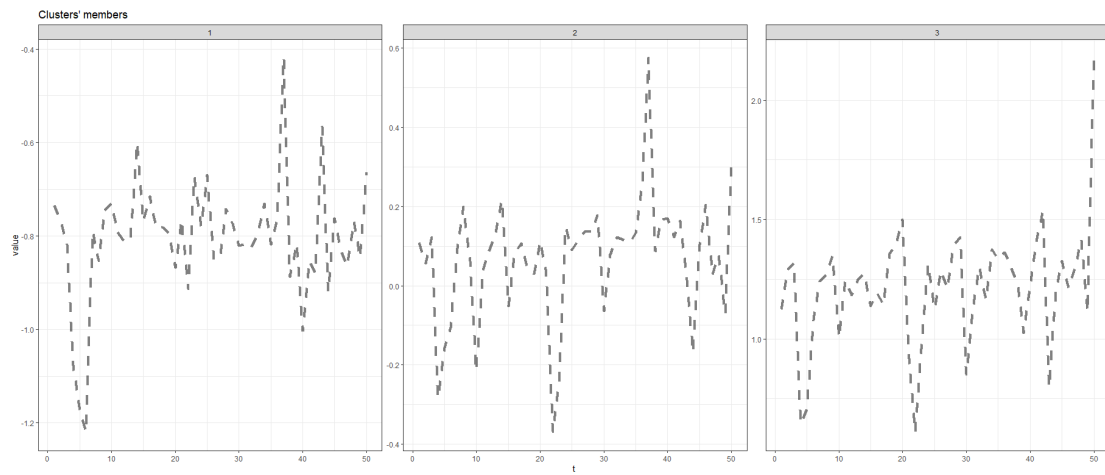
```
#plot for each cluster
```



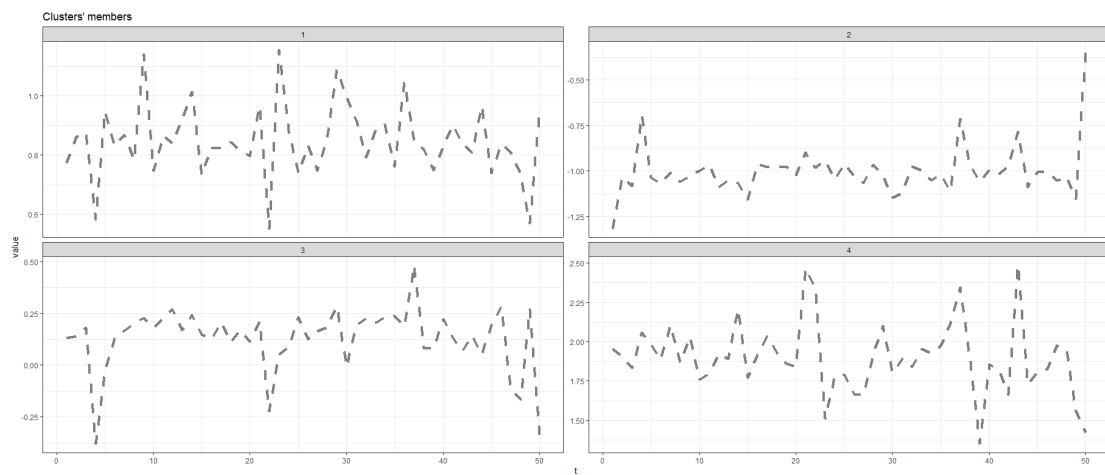
### Data2- 07:00-08:00



### Data3- 08:00-09:00

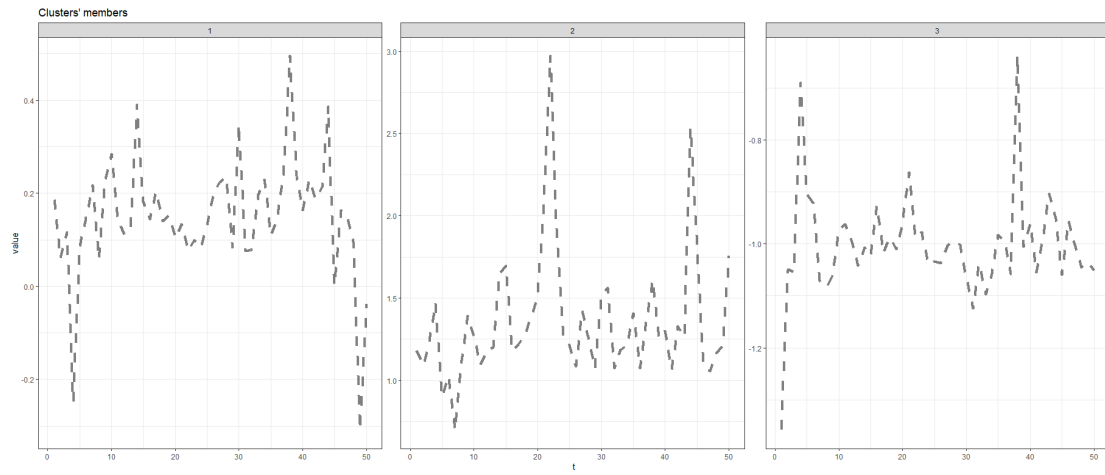


### Data4- 09:00-10:00

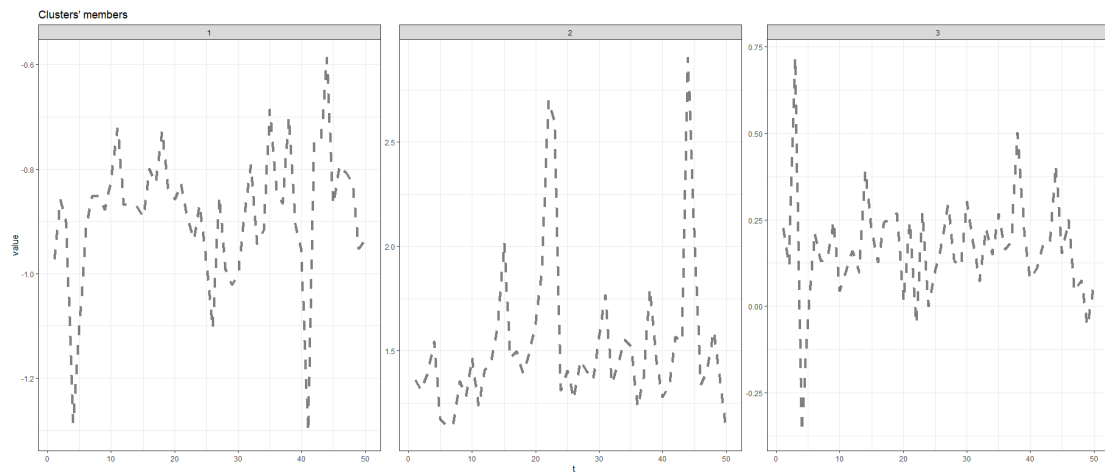




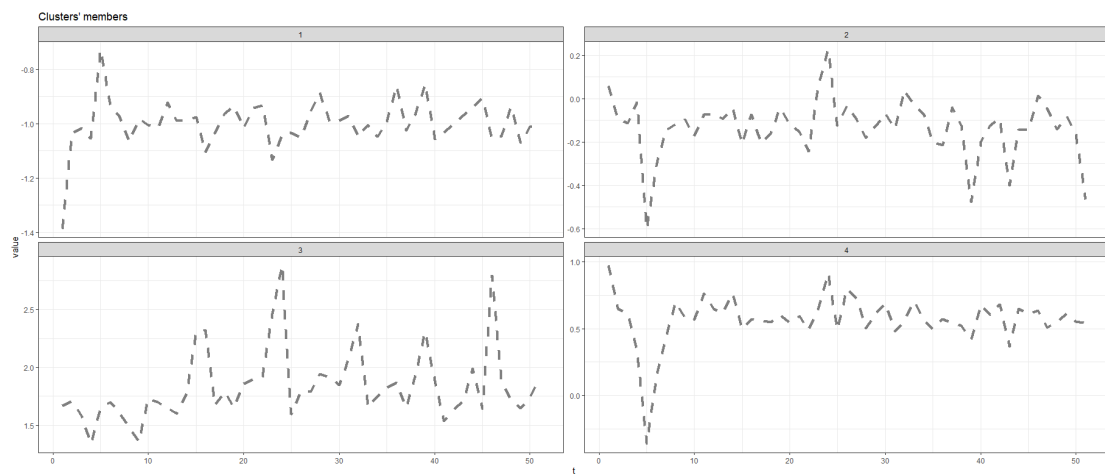
### Data5- 10:00-11:00



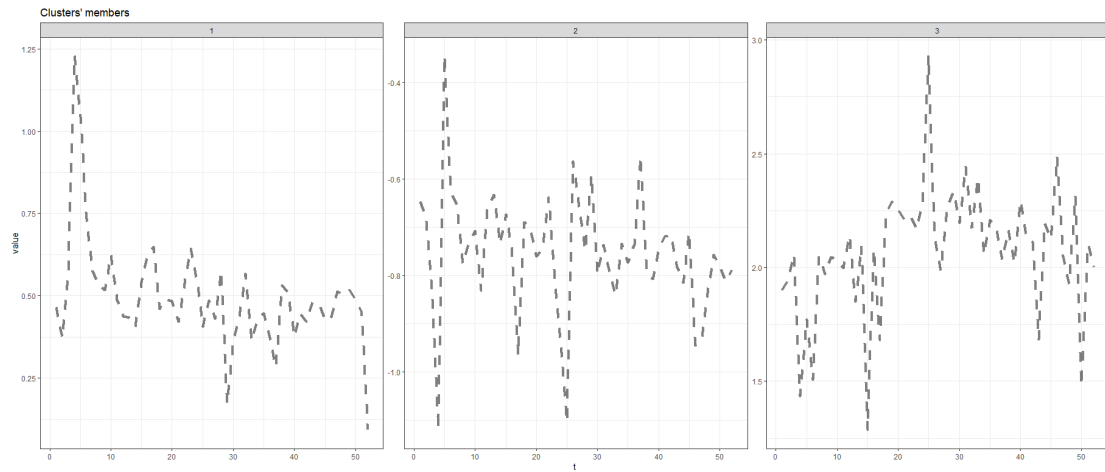
### Data6- 11:00-12:00



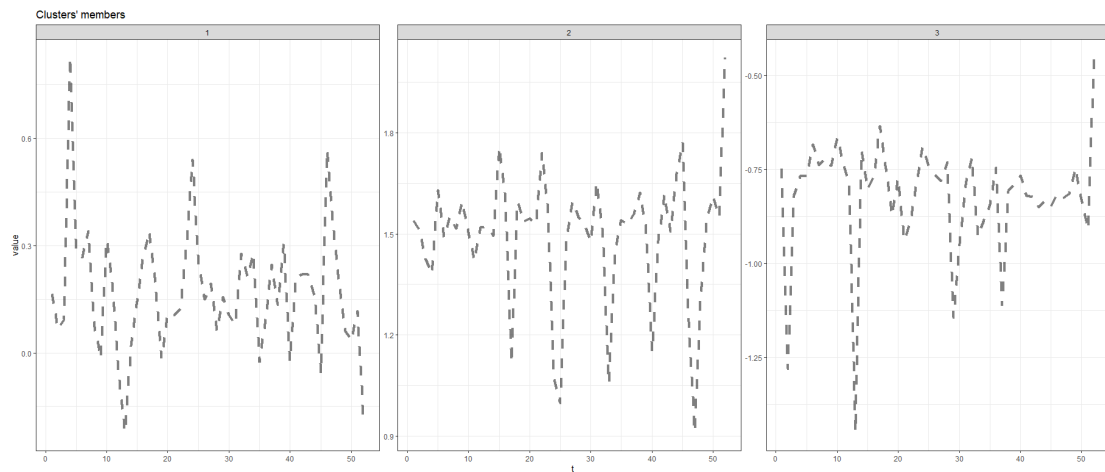
### Data7- 12:00-13:00



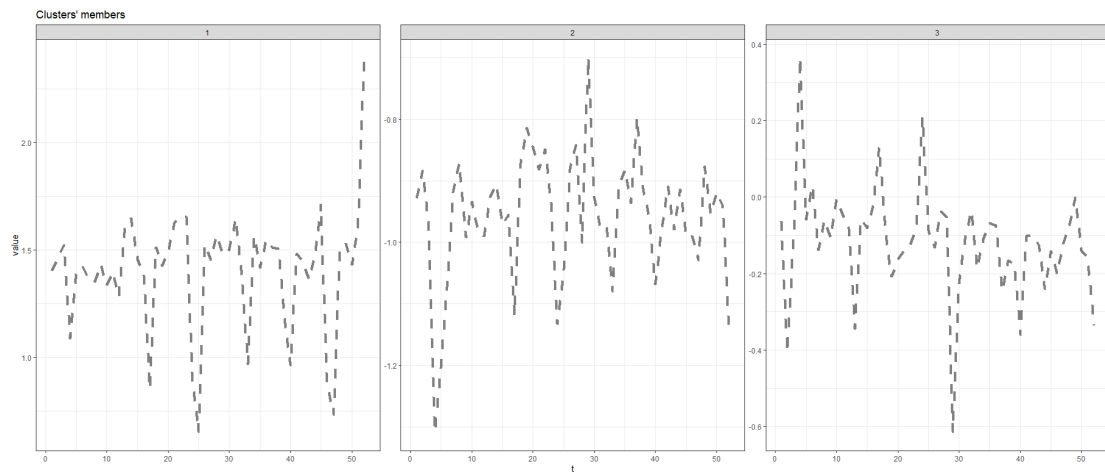
### Data8- 13:00-14:00



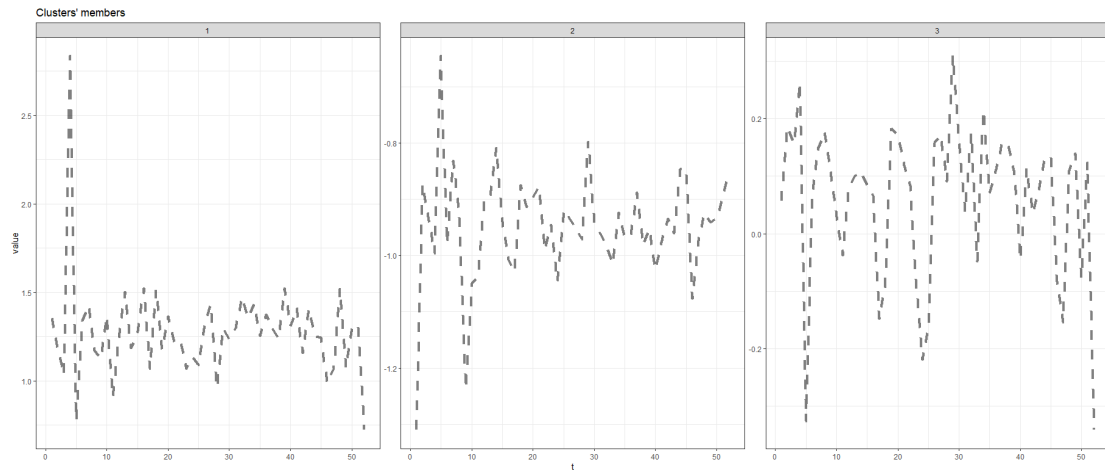
### Data9- 14:00-15:00



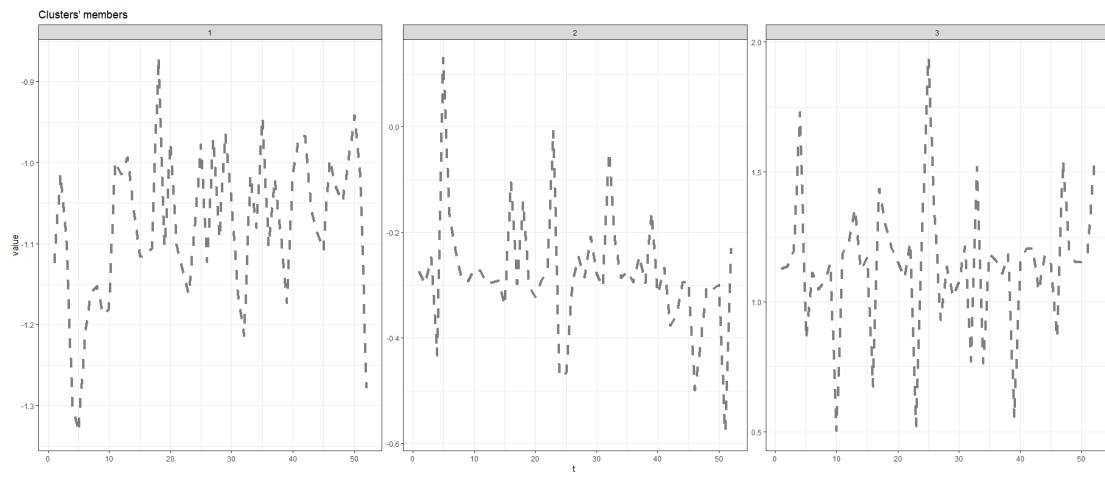
### Data10- 15:00-16:00



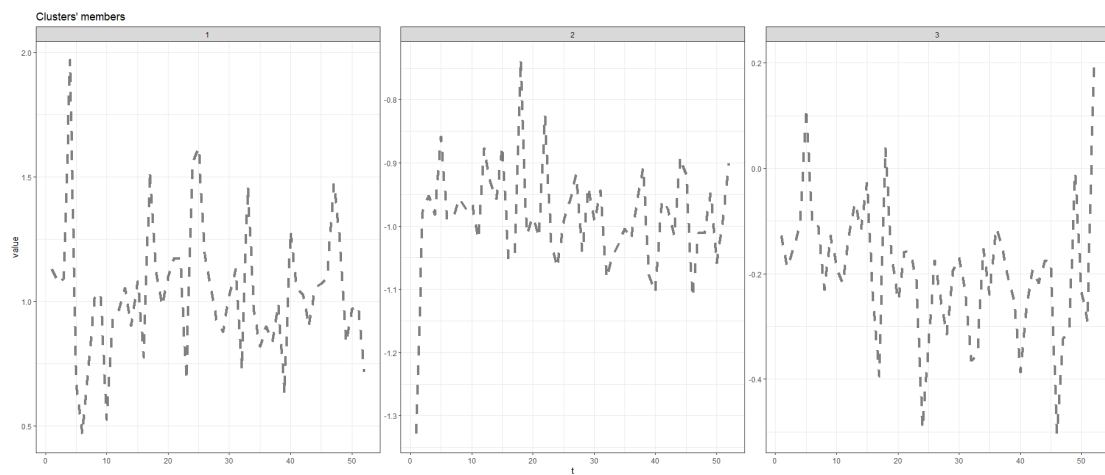
### Data11- 16:00-17:00



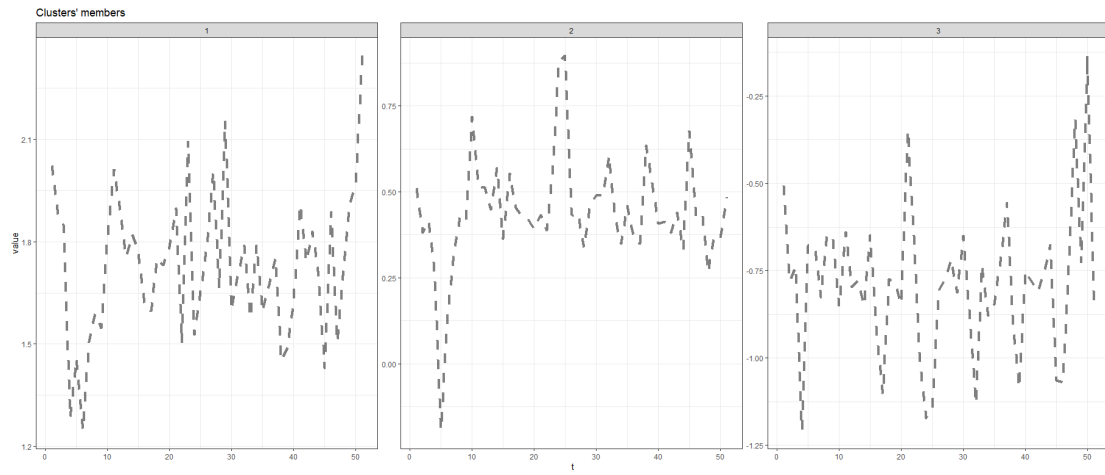
### Data12- 17:00-18:00



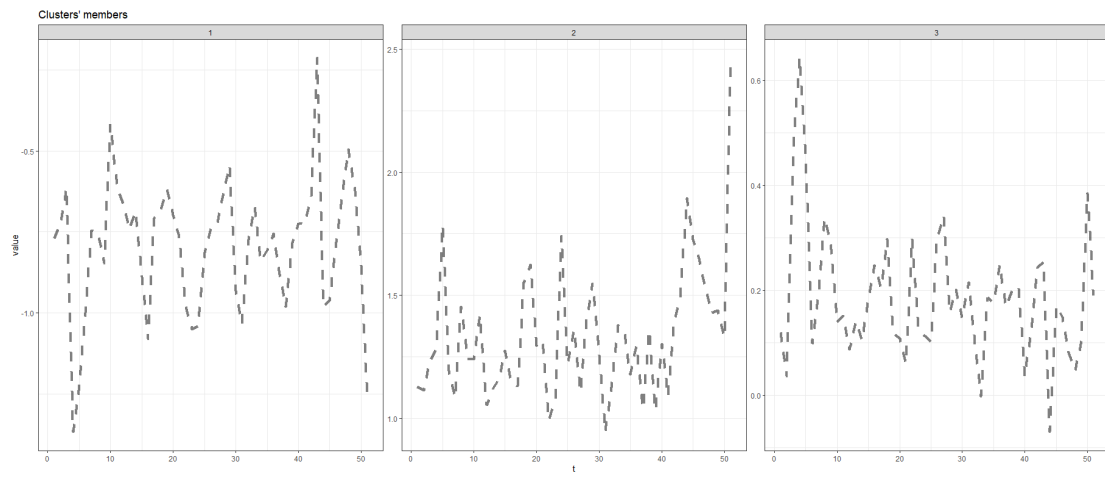
### Data13- 18:00-19:00



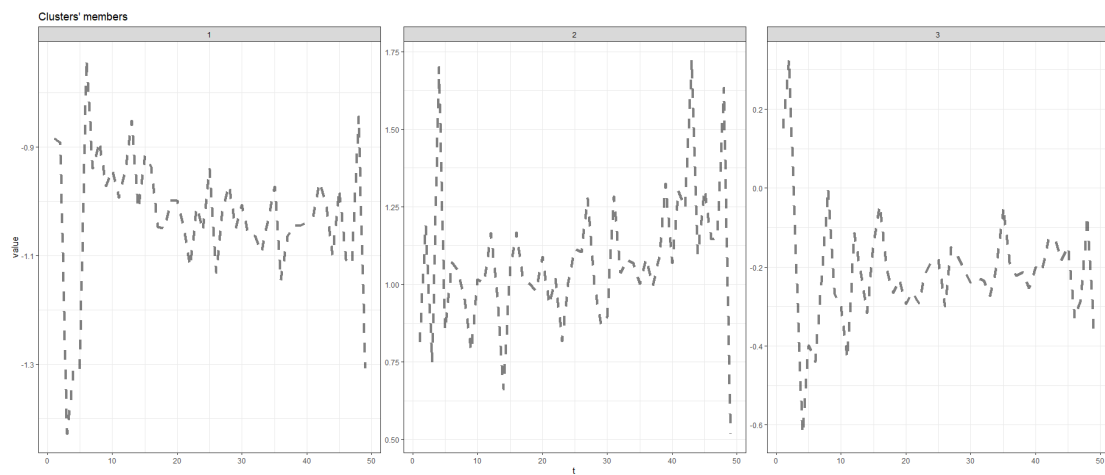
### Data14- 19:00-20:00



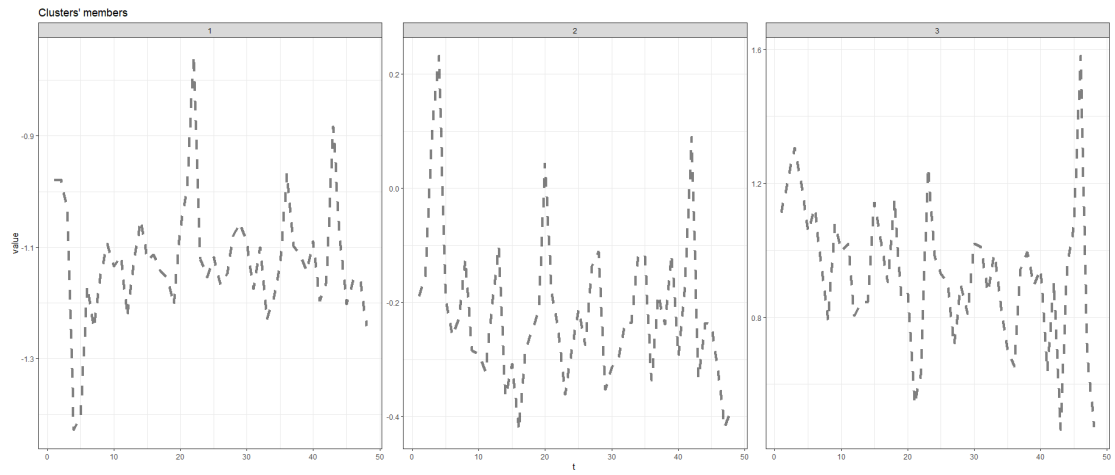
### Data15- 20:00-21:00



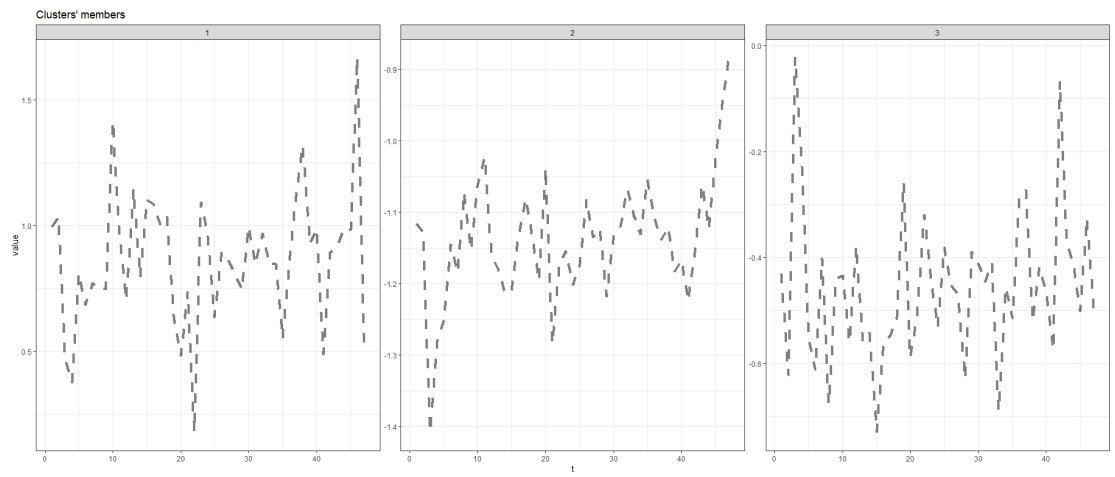
### Data16- 21:00-22:00



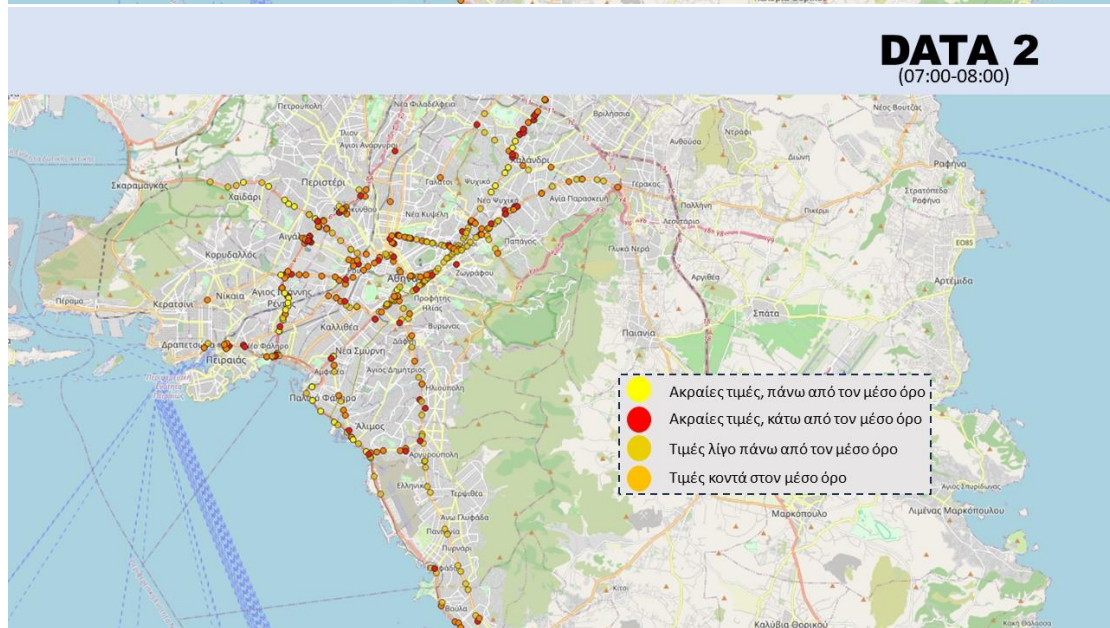
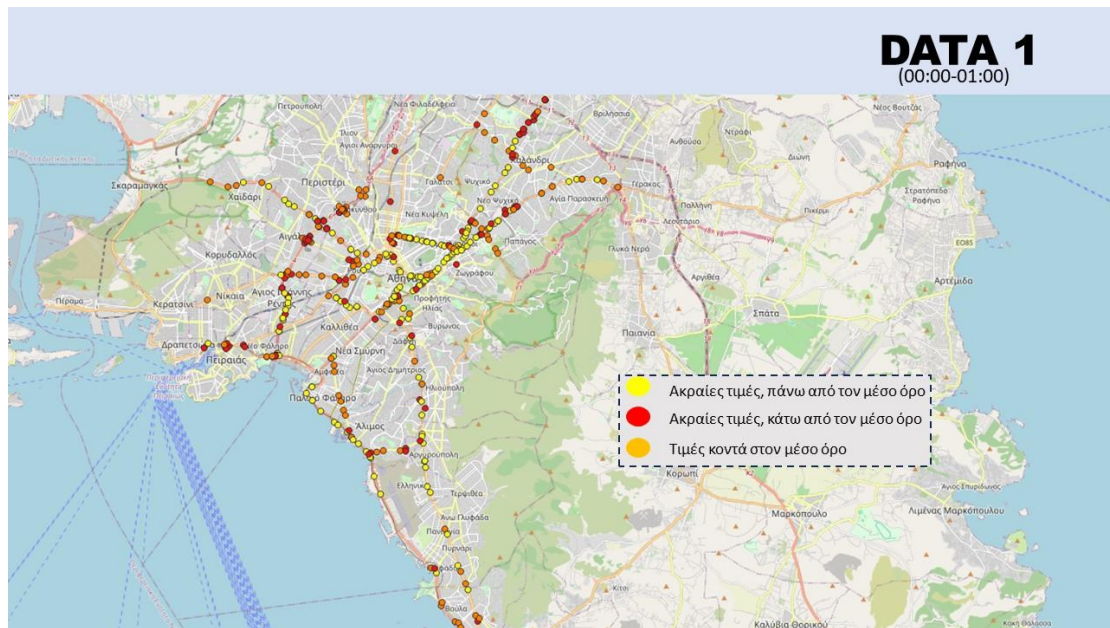
### Data17- 22:00-23:00



### Data18- 23:00-00:00

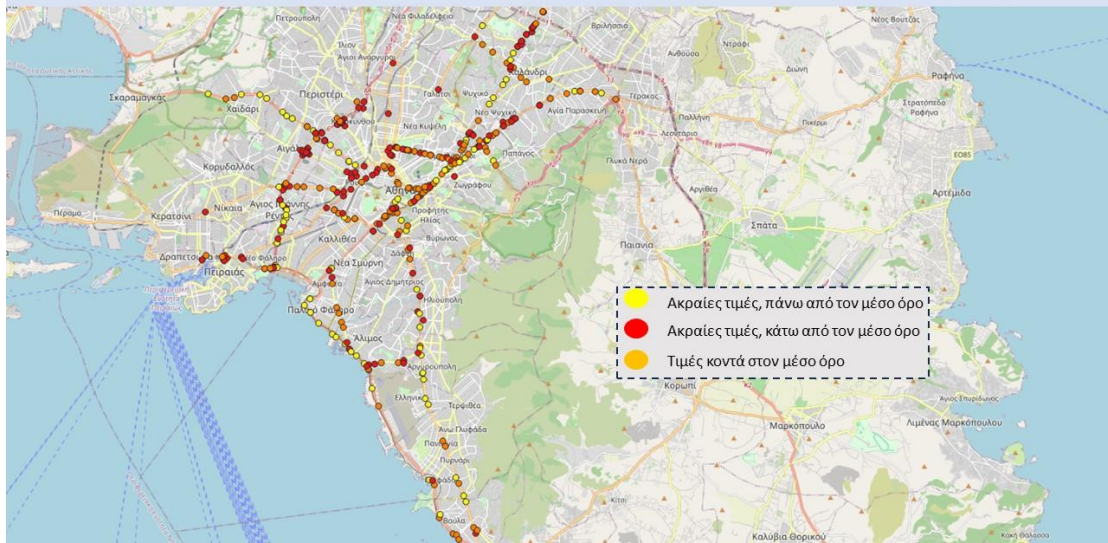


### 3) Χάρτες όλων των σημείων, για όλες τις ώρες της ημέρας

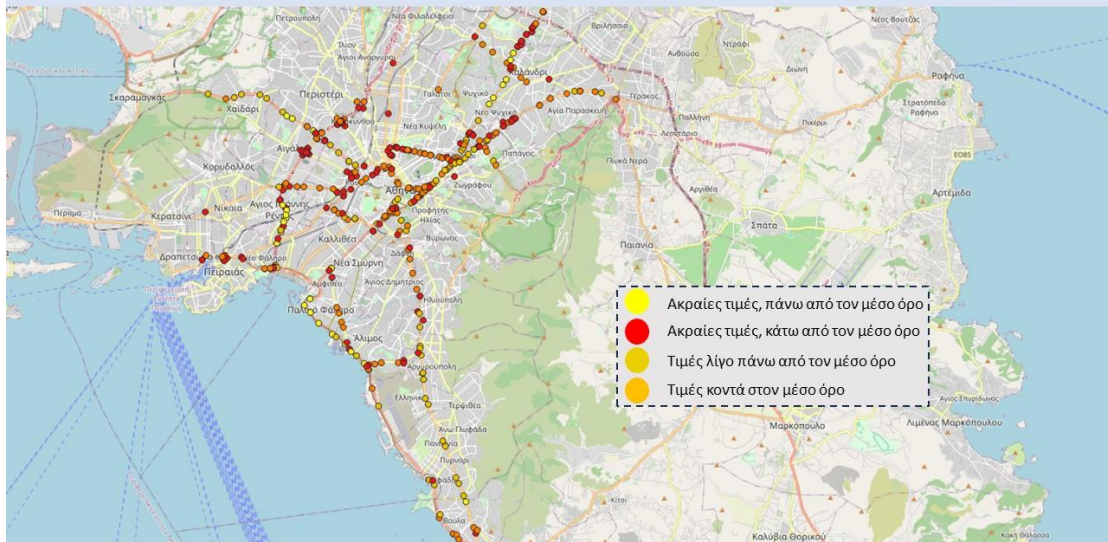




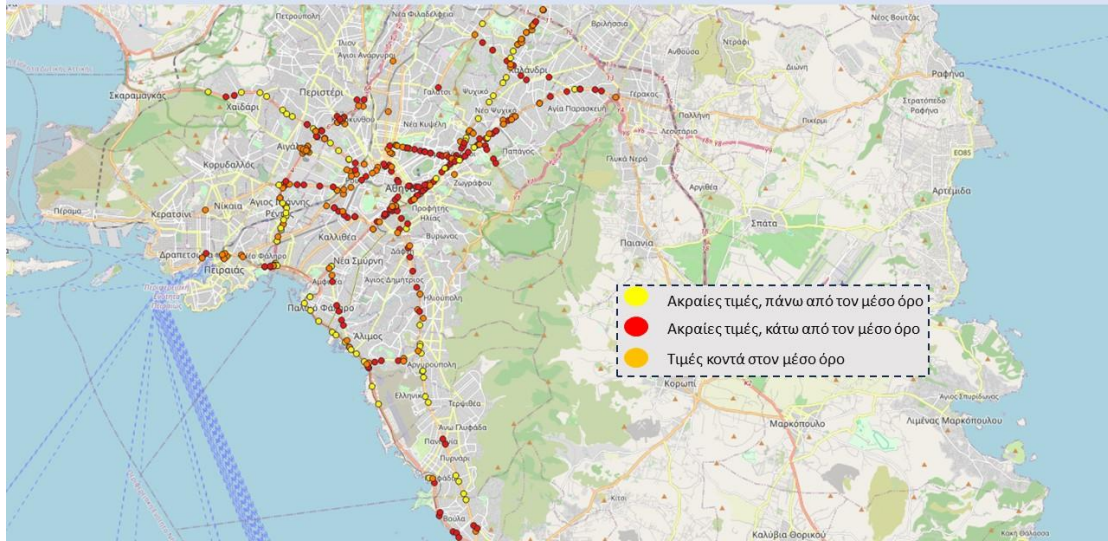
### DATA 3 (08:00-09:00)



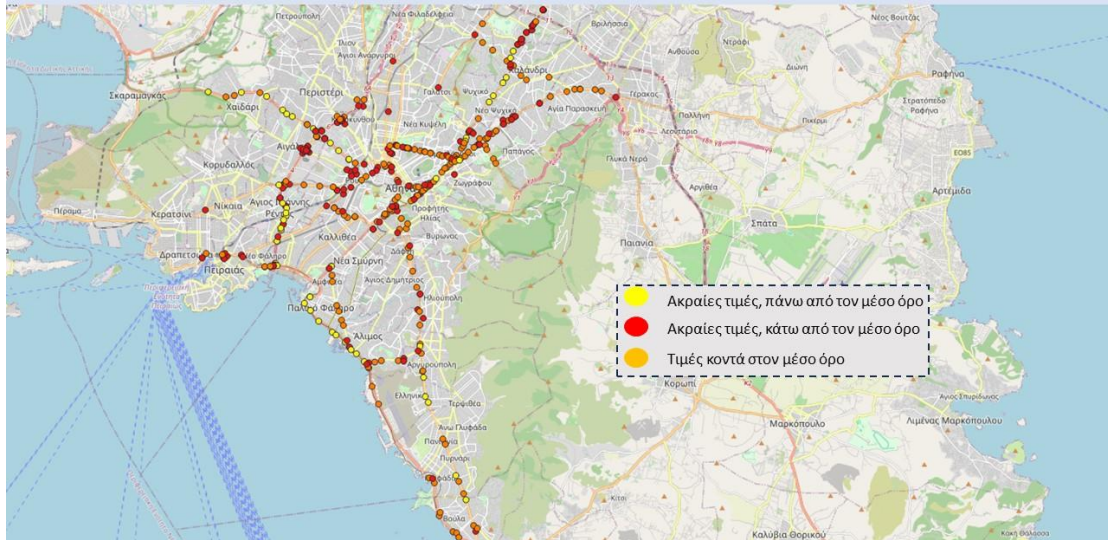
### DATA 4 (09:00-10:00)



## DATA 5 (10:00-11:00)

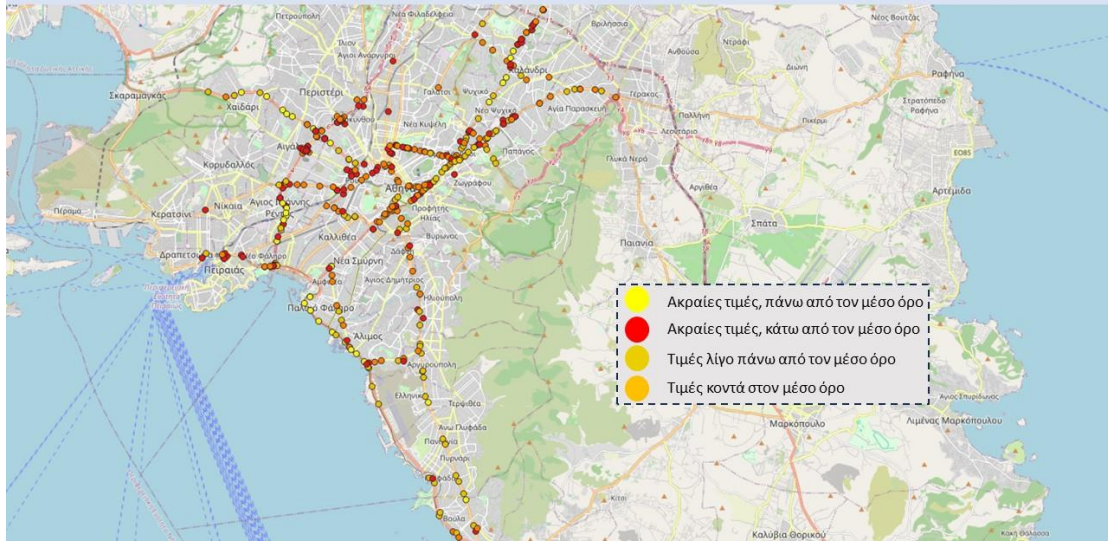


## DATA 6 (11:00-12:00)

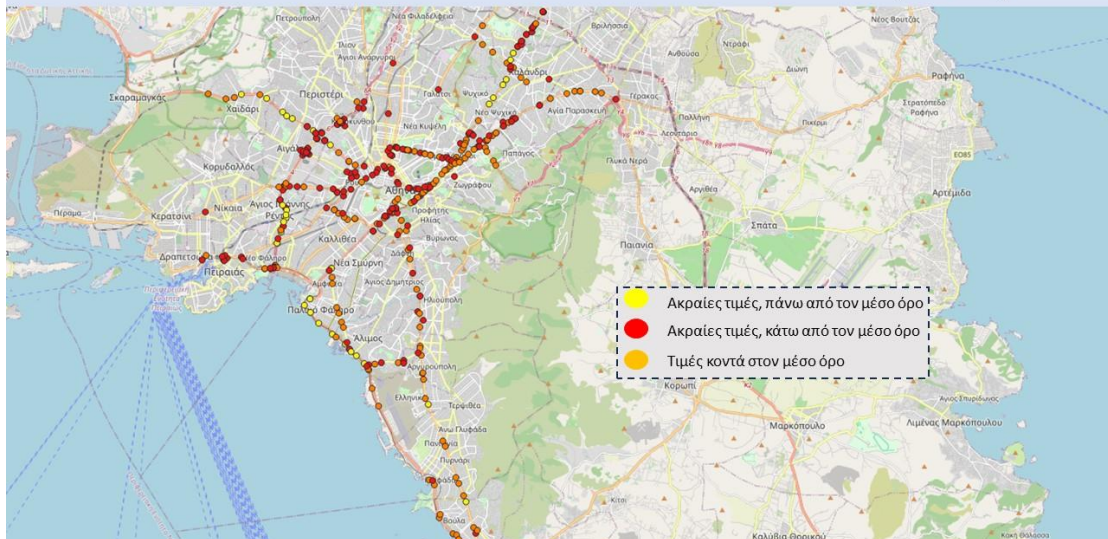




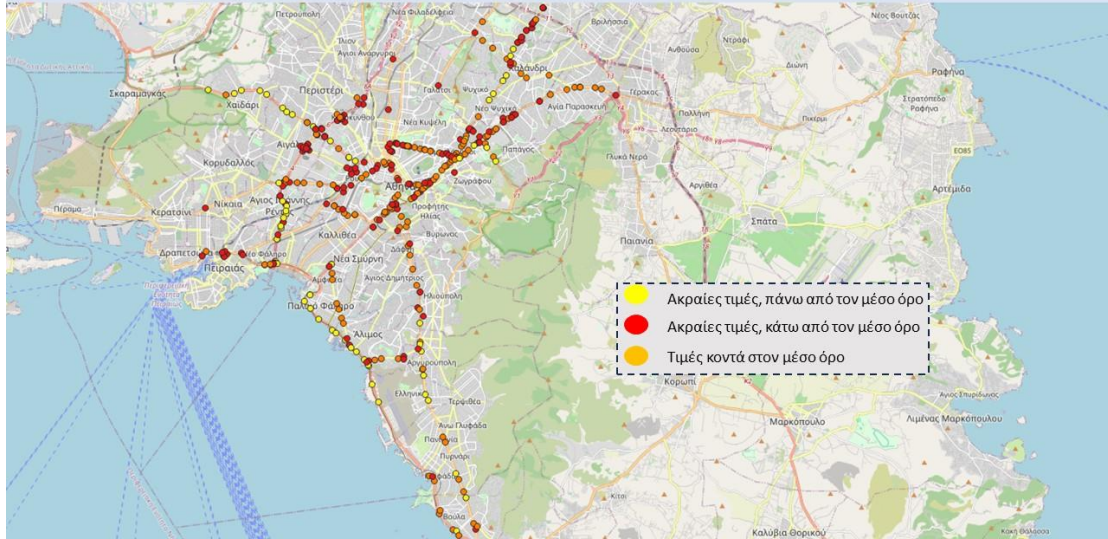
## DATA 7 (12:00-13:00)



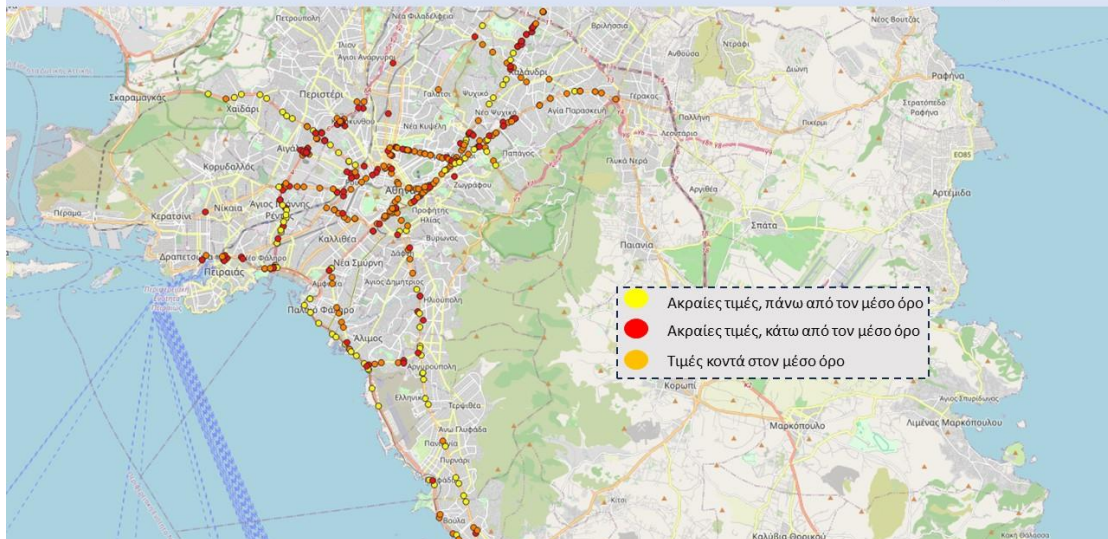
## DATA 8 (13:00-14:00)



## DATA 9 (14:00-15:00)



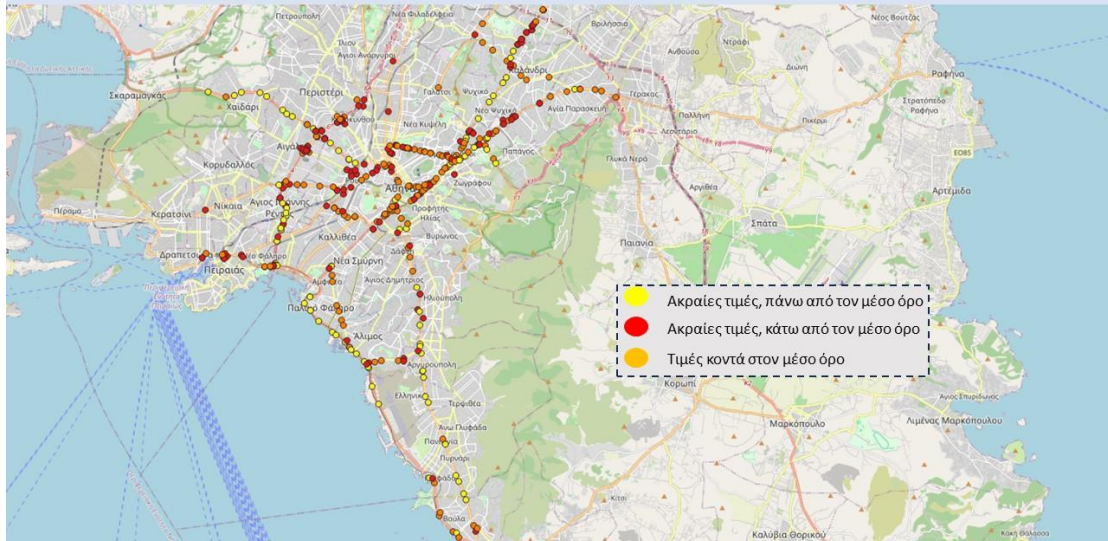
## DATA 10 (15:00-16:00)





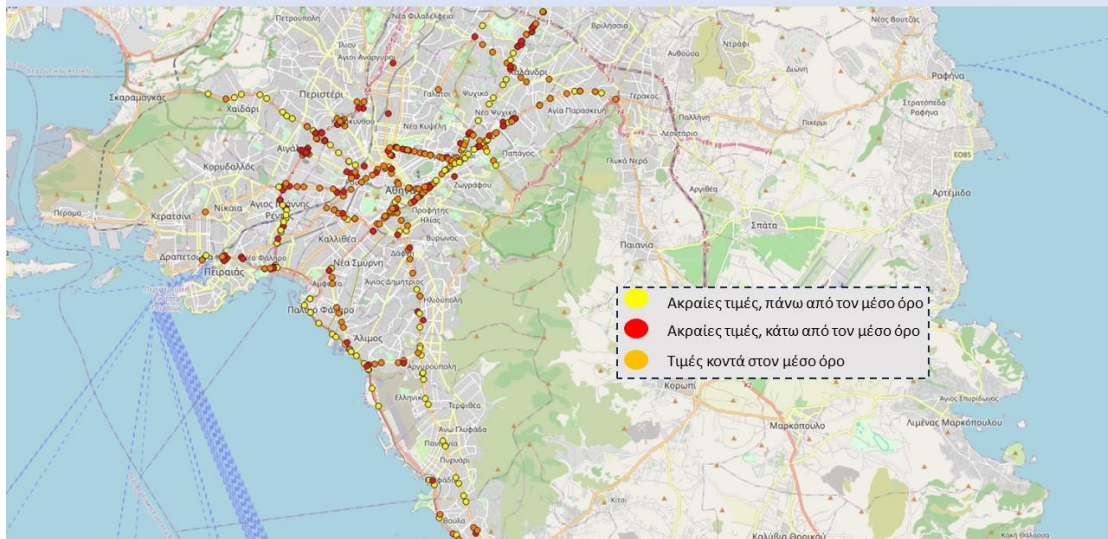
# DATA 11

(16:00-17:00)



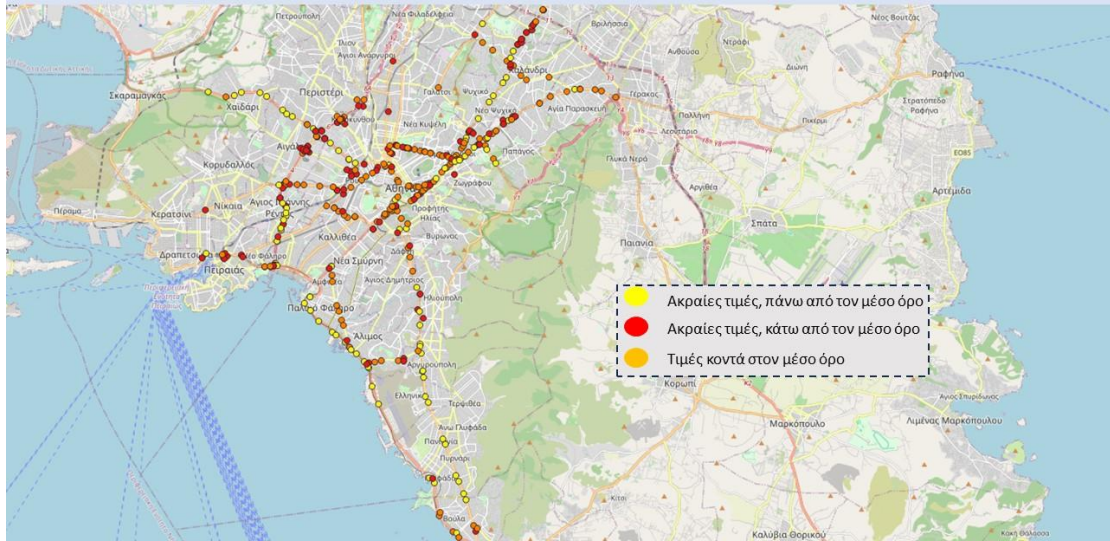
# DATA 12

(17:00-18:00)



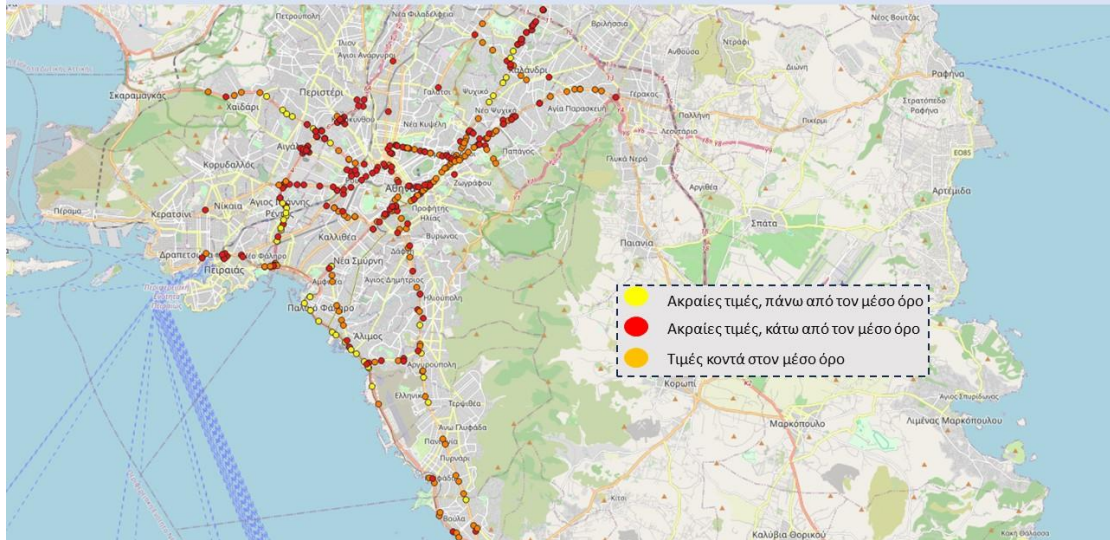
# DATA 13

(18:00-19:00)



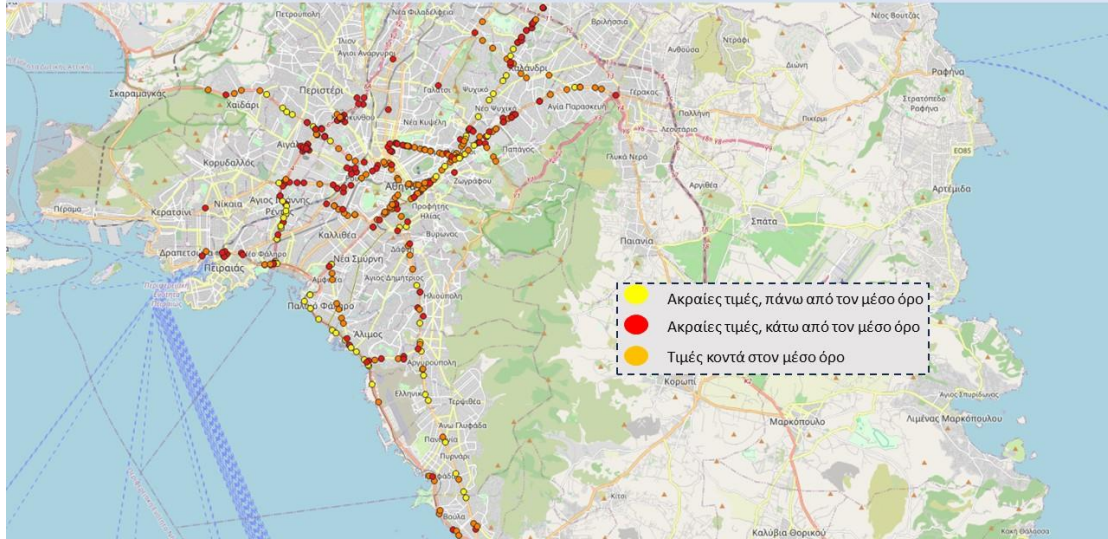
# DATA 14

(19:00-20:00)

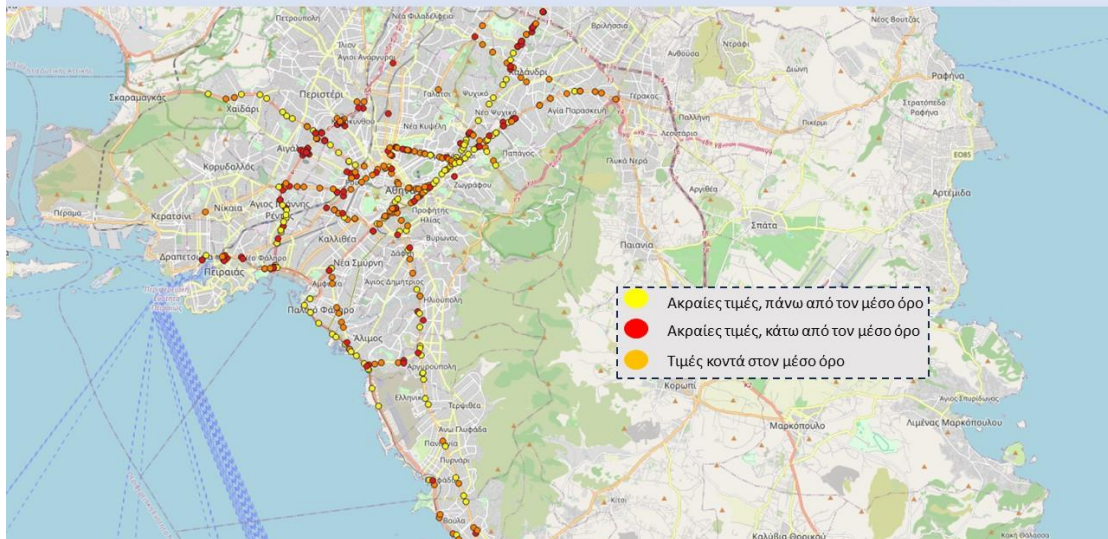




## DATA 15 (20:00-21:00)

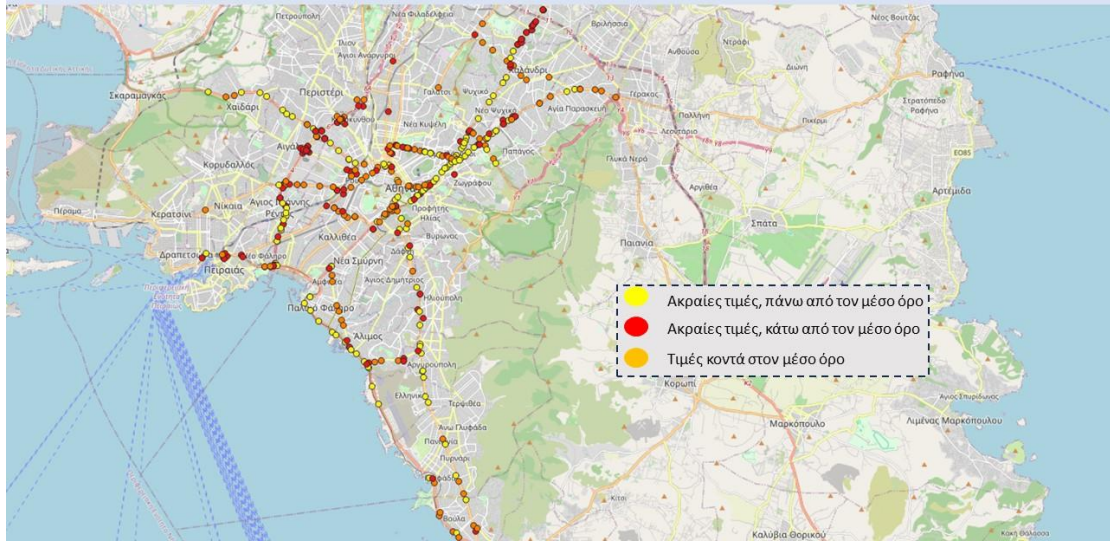


## DATA 16 (21:00-22:00)



# DATA 17

(22:00-23:00)



# DATA 18

(23:00-00:00)

