



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

Ανίχνευση προφίλ κινητικότητας, αλυσίδας και διάρκειας δραστηριοτήτων με χρήση δεδομένων ευρείας κλίμακας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ



Αθήνα, Μάρτιος 2021

Κωνσταντίνου Χρίστος

Επιβλέπουσα Καθηγήτρια: Ελένη Ι. Βλαχογιάννη

Ευχαριστίες

Η διπλωματική εργασία με τίτλο «Ανίχνευση προφίλ κινητικότητας, αλυσίδας και διάρκειας δραστηριοτήτων με χρήση δεδομένων ευρείας κλίμακας» πραγματοποιήθηκε στα πλαίσια φοίτησης του προγράμματος σπουδών του τμήματος Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου κατά το ακαδημαϊκό έτος 2020-2021.

Η ολοκλήρωση της παρούσας διπλωματικής εργασίας υλοποιήθηκε με την υποστήριξη της κ. Ελένης Ι. Βλαχογιάννης (Αναπληρώτρια Καθηγήτρια της Σχολής Πολιτικών Μηχανικών του Ε.Μ.Π.) και του κ. Παναγιώτη Φαφουτέλλη (υποψήφιο Διδάκτορα και ερευνητή του Ε.Μ.Π.) στους οποίους θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες για την άριστη συνεργασία που είχαμε, την συνεχή καθοδήγησή τους καθ' όλη τη διάρκεια εκπόνησης της μελέτης, καθώς και για την προθυμία τους και τον πολύτιμο χρόνο που μου διέθεσαν.

Ανίχνευση προφίλ κινητικότητας, αλυσίδας και διάρκειας δραστηριοτήτων με χρήση δεδομένων ευρείας κλίμακας

Κωνσταντίνου Χρίστος

Επιβλέπων: Ελένη Ι. Βλαχογιάννη

Σύνοψη

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανίχνευση των μοτίβων κινητικότητας, των καθημερινών αλυσίδων των δραστηριοτήτων και η ποσοτικοποίηση των διαρκειών των διαφορετικών δραστηριοτήτων. Αναλυτικά δεδομένα συλλέχθηκαν για 696 οδηγούς και τα ταξίδια τα οποία πραγματοποίησαν (περισσότερα από 150000 ταξίδια συνολικά) στην Ελλάδα από τη περίοδο του Δεκεμβρίου 2017 έως τον Αύγουστο 2019. Τα συγκεντρωτικά αποτελέσματα βασίστηκαν σε αυτούς με περισσότερα από 200 ταξίδια στη παραπάνω χρονική περίοδο. Μέσω συνεργασίας μη εποπτευόμενων αλγορίθμων μάθησης (Dbscan, k-nearest neighbors) καθώς και χωροχρονικών χαρακτηριστικών, οι προορισμοί κάθε οδηγού κατηγοριοποιήθηκαν. Όλα τα εξατομικευμένα μοτίβα κινητικότητας και οι εξατομικευμένες καθημερινές αλυσίδες δραστηριοτήτων εξάχθηκαν. Επιπλέον, έγινε καθημερινή ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών (Σπίτι, Δουλειά, Προσωπικές Δραστηριότητες, Υπόλοιπες Τοποθεσίες). Μέσω της στατιστικής μεθόδου (Ανάλυση Επιβίωσης) πραγματοποιήθηκε η κανονικοποίηση των διαρκειών των δραστηριοτήτων ανά τύπο και ημέρα. Τα ευρήματα δείχνουν ότι η προτεινόμενη μεθοδολογία είναι σε θέση να ανιχνεύσει τα εξατομικευμένα προφίλ κινητικότητας και τη καθημερινή αλυσίδα δραστηριοτήτων. Η στατιστική ανάλυση των καθημερινών μοτίβων κινητικότητας δείχνουν ότι ανάμεσα στις μέρες Δευτέρα έως Παρασκευή υπάρχει αξιοσημείωτη ομοιότητα και ξεχωριστά κάθε μέρα παρουσιάζει σημαντική κανονικότητα. Επιπλέον δεν υπάρχει σημαντική διαφορά όσον αφορά τα ταξίδια από και προς τη Δουλειά ανάμεσα σε αυτές τις μέρες. Τέλος, οι αλυσίδες και οι διάρκειες δραστηριοτήτων ακολουθούν ένα σχετικά σταθερό μοτίβο για όλες τις μέρες εκτός της Κυριακής.

Λέξεις-κλειδιά: μοτίβο κινητικότητας, αλυσίδες δραστηριοτήτων, διάρκειες δραστηριοτήτων, μη εποπτευόμενη μάθηση, δεδομένα ευρείας κλίμακας

Mining Human Mobility Patterns, Activity Chains and Duration using Big Data

Konstantinou Christos

Supervisor: Eleni I. Vlahogianni

Abstract

The objective of the present Thesis is the detection of human mobility patterns, the daily activity chains and to provide the quantification of different activities' duration. Detailed data gathered for 696 drivers and the trips conducted (more than 150000 total trips) in Greece from December 2017 to August 2019. Results are based on drivers having more than 200 trips between the above period. Through the collaboration of unsupervised learning algorithms such as DbSCAN and k-nearest neighbors and spatiotemporal characteristics, the destinations of every driver were categorized. All the individual mobility patterns and the personal typical daily trip chains were ascertained. Moreover, a daily quantification of trips between primary destinations (home, work, personal activities, other) was derived. Performing the statistical method (survival analysis), the standardization of activity durations per type and day inferred. The results indicate that the proposed methodology can reveal each individual's mobility profile and trip chain. Statistical analysis of the daily mobility patterns indicate that days from Monday to Friday reveal a significant resemblance and each day has a notable regularity. Also there is no significant difference concerning the trips conducted to/from work between these days. Finally individual trip chains and activity durations follow a relatively stable pattern for all days, except for Sunday.

Keywords: mobility patterns, activity chains, activity duration, unsupervised learning, Big data

Περίληψη

Οι βλέψεις των επιστημόνων για τα παρούσα και μελλοντικά αστικά συγκοινωνιακά δίκτυα επικεντρώνονται περισσότερο στις εξατομικευμένες λύσεις κινητικότητας για κάθε χρήστη ξεχωριστά. Αυτό έχει στόχο την βελτίωση της αποδοτικότητας των κυκλοφοριακών συνθηκών. Επίσης, αναδυόμενες υπηρεσίες κινητικότητας και τα αυτόνομα αυτοκίνητα όχι μόνο επιθυμούν, αλλά συχνά έχουν ανάγκη την ιδέα της εξατομικευμένης λύσης κινητικότητας για να επιτευχθεί η σωστή τους λειτουργία.

Συνεπώς η ολοκληρωτική κατανόηση της συμπεριφοράς του κάθε χρήστη στο συγκοινωνιακό δίκτυο είναι ζωτικής σημασίας για την επιτυχία μελλοντικών λύσεων κινητικότητας των ανθρώπων. Για αυτό το λόγο στη παρούσα Διπλωματική Εργασία γίνεται προσπάθεια δημιουργίας μίας δομής μέσω της χρήσης μιας βάσης δεδομένων ευρείας κλίμακας και μη εμποτευόμενων αλγορίθμων μάθησης για να επιτευχθεί αυτό. Πιο συγκεκριμένα η δομή αυτή έχει στόχο να ανιχνεύσει προφίλ κινητικότητας, καθημερινή χαρακτηριστική αλυσίδα για κάθε χρήστη ξεχωριστά και να ποσοτικοποιήσει τις διάρκειες των διαφορετικών δραστηριοτήτων.

Μέσω της βιβλιογραφικής ανασκόπησης τα συμπεράσματα τα οποία βγαίνουν είναι ποικίλα. Αρχικά τα μοτίβα κινητικότητας των ανθρώπων παρουσιάζουν σημαντική κανονικότητα και επαναληπτικότητα. Επιπλέον φαίνεται πως κάθε άνθρωπος χαρακτηρίζεται από μία τυπική καθημερινή αλυσίδα δραστηριοτήτων. Επίσης τα μοτίβα κινητικότητας και οι αλυσίδες δραστηριοτήτων του συνόλου των ανθρώπων έχουν κατηγοριοποιηθεί και έχει αποδειχθεί ότι η πρόβλεψη τους είναι εφικτή. Τα παραπάνω συμπεράσματα αποτέλεσαν την αφορμή και το σκοπό της παρούσας έρευνας τόσο προς επιβεβαίωση αλλά και εμπλούτιση αυτών.

Αναλυτικότερα η αρχική βάση δεδομένων περιείχε περισσότερα από 150000 ταξίδια την περίοδο του Δεκεμβρίου 2017 έως Αύγουστο 2019 στην Ελλάδα. Μέσω της κατηγοριοποίησης των προορισμών που επετεύχθη μέσω αλγορίθμων μάθησης χωρίς επίβλεψη (DBSCAN, K-nearest-neighbors) και χωροχρονικών στοιχείων, αλλά και την ποσοτικοποίηση μεταξύ αυτών ανιχνεύονται τα μοτίβα κινητικότητας και η καθημερινή χαρακτηριστική αλυσίδα του κάθε χρήστη. Επιπλέον μέσω της στατιστικής ανάλυσης του συνόλου των καθημερινών μοτίβων κινητικότητας και των καθημερινών αλυσίδων προκύπτει πώς για τις μέρες μεταξύ Δευτέρας έως Παρασκευή παρουσιάζεται μια σταθερή κανονικότητα και ένα σχετικά σταθερό μοτίβο αντίστοιχα. Επίσης φαίνεται πως τα ταξίδια τα οποία πραγματοποιούνται από/προς τη Δουλειά δεν παρουσιάζουν σημαντική διαφορά ανάμεσα σε αυτές τις μέρες. Τέλος μέσω της προτυποποίησης των διαρκειών με τη μέθοδο της Ανάλυσης Επιβίωσης (Survival Analysis) διαφαίνονται αρκετές ομοιότητες μεταξύ όλων των ημερών πλην της Κυριακής όσον αφορά τις διάρκειες των δραστηριοτήτων. Πιο συγκεκριμένα όσον αφορά τις διάρκειες της δραστηριότητας του Σπιτιού συμπεραίνεται ότι οι μέρες από Δευτέρα έως Πέμπτη ακολουθούν παρόμοιο μοτίβο. Επιπλέον διαφαίνεται ότι οι διάρκειες της δραστηριότητας της Δουλειάς ακολουθούν παρόμοιο μοτίβο για τη πλειοψηφία των ημερών από Δευτέρα έως Παρασκευή καθώς και για τις ημέρες Σάββατο με Κυριακή. Επιπρόσθετα οι διάρκειες των Προσωπικών Δραστηριοτήτων δεν παρουσιάζουν σημαντική διαφορά για τη πλειοψηφία των ημερών από Δευτέρα

έως Σάββατο. Τέλος, οι διάρκειες όλων των Υπόλοιπων Δραστηριοτήτων ακολουθούν παρόμοιο μοτίβο για τις μέρες από Δευτέρα έως Πέμπτη.

Από την έρευνα που πραγματοποιήθηκε στη παρούσα Διπλωματική Εργασία προέκυψαν κάποιες συστάσεις για περαιτέρω έρευνα. Πιο συγκεκριμένα αυτές είναι η δημιουργία αλγορίθμου που προβλέπει την καθημερινή αλυσίδα αλλά και τα μοτίβα κινητικότητας καθώς και η δημιουργία εφαρμογών που εκμεταλλεύονται τις παραπάνω πληροφορίες με ποικίλους τρόπους. Στόχος τέτοιων εφαρμογών αποτελεί η βέλτιστη λειτουργία των συγκοινωνιακών δικτύων αλλά και η αναπροσαρμογή των δραστηριοτήτων.

Περιεχόμενα

1. Εισαγωγή.....	15
1.1 Διαχείριση Κινητικότητας.....	15
1.2 Σκοπός της Εργασίας	16
1.3 Διάρθρωση της Εργασίας	17
2. Βιβλιογραφική Ανασκόπηση	18
2.1 Πρότυπα Ανίχνευσης Κινητικότητας.....	18
2.2 Μοτίβα κινητικότητας των ανθρώπων	18
2.2.1 Επαναληπτικότητα καθημερινών προορισμών	22
2.3 Πρόβλεψη κινητικότητας οχημάτων.....	22
2.4 Συμπεράσματα Βιβλιογραφίας	22
3. Μεθοδολογική Προσέγγιση	24
3.1 Περιγραφή Διαδικασίας.....	24
3.2 Θεωρητικό Υπόβαθρο.....	25
3.2.1 Γλώσσα Προγραμματισμού Python	26
3.2.2 Αλγόριθμος ομαδοποίησης DBSCAN	27
3.2.3 Αλγόριθμος μηχανικής μάθησης K-nearest-neighbors	28
3.2.4 Ανάλυση Επιβίωσης (Survival Analysis).....	29
4. Ανάλυση και Αποτελέσματα.....	33
4.1 Περιγραφή βάσης δεδομένων	33
4.2 Αρχική επεξεργασία βάσης δεδομένων	33
4.3 Οπτικοποίηση δεδομένων ανά οδηγό	34
4.4 Κατηγοριοποίηση των προορισμών ανά οδηγό.....	38
4.5 Εύρεση προφίλ κινητικότητας ανά οδηγό	40
4.6 Εύρεση χωροχρονικού προφίλ	41
4.7 Κατηγοριοποίηση συνόλου των προορισμών.....	44
4.8 Καθημερινή αλυσίδα δραστηριοτήτων των χρηστών και τα στατιστικά τους ...	47
4.9 Προτυποποίηση διαρκειών των δραστηριοτήτων	49
4.9.1 Σπίτι	50
4.9.2 Δουλειά	52
4.9.3 Προσωπικές Δραστηριότητες	55
4.9.4 Υπόλοιπες Τοποθεσίες.....	57
5. Συμπεράσματα και Προτάσεις.....	61
5.1 Σύνοψη Μεθοδολογίας και Αποτελεσμάτων	61
5.2 Βασικά Συμπεράσματα	61

5.3 Συστάσεις για περαιτέρω έρευνα	62
Βιβλιογραφία.....	64
ΠΑΡΑΡΤΗΜΑ Α.....	66
ΠΑΡΑΡΤΗΜΑ Β.....	87
ΠΑΡΑΡΤΗΜΑ Γ.....	91

Ευρετήριο Διαγραμμάτων

Διάγραμμα 1: Αριθμός χρηστών έξυπνων κινητών τηλεφώνων παγκοσμίως (Πηγή: Statista.com)	16
Διάγραμμα 2: Ροή κεφαλαίων εργασίας.....	17
Διάγραμμα 3: Βασικές κατηγορίες μοντέλων κινητικότητας.....	18
Διάγραμμα 4: 8 αντιπροσωπευτικά μοτίβα δραστηριοτήτων (Πηγή: Allahviranloo (2016))	19
Διάγραμμα 5: 12 αντιπροσωπευτικά μοτίβα δραστηριοτήτων (Πηγή: Hafezi et al. (2017))	20
Διάγραμμα 6: Ομαδοποίηση Καθημερινών μοτίβων δραστηριότητας και τα κοινωνικό-δημογραφικά τους χαρακτηριστικά (Πηγή: Jiang et al. (2012))	21
Διάγραμμα 7: Ομαδοποίηση μοτίβων δραστηριότητας τα Σαββατοκύριακα και τα κοινωνικό-δημογραφικά τους χαρακτηριστικά (Πηγή: Jiang et al. (2012)).....	21
Διάγραμμα 8: Διάγραμμα ροής εργασιών	25
Διάγραμμα 9: Θεωρητικό υπόβαθρο εργασίας	26
Διάγραμμα 10: Εργαλεία που χρησιμοποιήθηκαν για την Ανάλυση Επιβίωσης (Survival Analysis).....	30
Διάγραμμα 11: Ώρα άφιξης για όλες τις τοποθεσίες του οδηγού 1	35
Διάγραμμα 12: Χωροχρονικές αφίξεις του οδηγού 1	35
Διάγραμμα 13: Heat map όλων των τοποθεσιών που έχει βρεθεί ο οδηγός 1	36
Διάγραμμα 14: Απεικόνιση όλων των ταξιδιών του οδηγού 1	37
Διάγραμμα 15: Απεικόνιση μοτίβου μετακινήσεων ανά ημέρα για τον οδηγό 1	37
Διάγραμμα 16: Μέθοδος κατηγοριοποίησης προορισμών	38
Διάγραμμα 17: Κατηγοριοποίηση προορισμών του οδηγού 1.....	40
Διάγραμμα 18: Μοτίβο κινητικότητας του οδηγού 1	41
Διάγραμμα 19: Πιθανότητα άφιξης στους βασικούς προορισμούς για κάθε ώρα της ημέρας για τον οδηγό 1	41
Διάγραμμα 20: Πιθανότητα εύρεσης του οδηγού 1 στις βασικές τοποθεσίες για κάθε ώρα της ημέρας.....	42
Διάγραμμα 21: Χαρακτηριστική αλυσίδα δραστηριοτήτων του οδηγού 1	43
Διάγραμμα 22: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών τη Δευτέρα.....	44
Διάγραμμα 23: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών την Τρίτη	45
Διάγραμμα 24: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών την Τετάρτη	45
Διάγραμμα 25: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών την Πέμπτη.....	45
Διάγραμμα 26: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών την Παρασκευή.....	46
Διάγραμμα 27: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών το Σάββατο	46
Διάγραμμα 28: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών την Κυριακή.....	46
Διάγραμμα 29: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Δευτέρας με βάση τα τυπικά προφίλ των οδηγών	49

Διάγραμμα 30: Πιθανότητα διάρκειας παραμονής στη τοποθεσία Σπίτι για κάθε μέρα της εβδομάδας	50
Διάγραμμα 31: Πιθανότητα τερματισμού παραμονής στη τοποθεσία Σπίτι για κάθε μέρα της εβδομάδας.....	51
Διάγραμμα 32: Πιθανότητα διάρκειας παραμονής στην τοποθεσία Δουλειά για κάθε μέρα της εβδομάδας.....	53
Διάγραμμα 33: Πιθανότητα τερματισμού παραμονής στη τοποθεσία Δουλειά για κάθε μέρα της εβδομάδας.....	54
Διάγραμμα 34: Πιθανότητα διάρκειας παραμονής στην τοποθεσία Προσωπικές Δραστηριότητες για κάθε μέρα της εβδομάδας	55
Διάγραμμα 35: Πιθανότητα τερματισμού παραμονής στις Προσωπικές Δραστηριότητες για κάθε μέρα της εβδομάδας	56
Διάγραμμα 36: Πιθανότητα διάρκειας παραμονής στις Υπόλοιπες Τοποθεσίες για κάθε μέρα της εβδομάδας	58
Διάγραμμα 37: Πιθανότητα τερματισμού παραμονής στις Υπόλοιπες Τοποθεσίες για κάθε μέρα της εβδομάδας	59
Διάγραμμα 38: Ώρα άφιξης για όλες τις τοποθεσίες του οδηγού 2	66
Διάγραμμα 39: Χωροχρονικές αφίξεις του οδηγού 2.....	66
Διάγραμμα 40: Heatmap όλων των τοποθεσιών που έχει βρεθεί ο οδηγός 2	67
Διάγραμμα 41: Απεικόνιση όλων των ταξιδιών του οδηγού 2	67
Διάγραμμα 42: Απεικόνιση μοτίβου μετακινήσεων ανά ημέρα για τον οδηγό 2.....	67
Διάγραμμα 43: Κατηγοριοποίηση προορισμών του οδηγού 2.....	68
Διάγραμμα 44: Μοτίβο κινητικότητας του οδηγού 2	68
Διάγραμμα 45: Πιθανότητα άφιξης στους βασικούς προορισμούς για κάθε ώρα της ημέρας για τον οδηγό 2	68
Διάγραμμα 46: Πιθανότητα εύρεσης του οδηγού 2 στις βασικές τοποθεσίες για κάθε ώρα της ημέρας.....	69
Διάγραμμα 47: Χαρακτηριστική αλυσίδα δραστηριοτήτων του οδηγού 2	69
Διάγραμμα 48: Ώρα άφιξης για όλες τις τοποθεσίες του οδηγού 3	70
Διάγραμμα 49: Χωροχρονικές αφίξεις του οδηγού 3.....	70
Διάγραμμα 50: Heat map όλων των τοποθεσιών που έχει βρεθεί ο οδηγός 3	71
Διάγραμμα 51: Απεικόνιση όλων των ταξιδιών του οδηγού 3	71
Διάγραμμα 52: Απεικόνιση μοτίβου μετακινήσεων ανά ημέρα για τον οδηγό 3.....	72
Διάγραμμα 53: Κατηγοριοποίηση προορισμών του οδηγού 3.....	72
Διάγραμμα 54: Μοτίβο κινητικότητας του οδηγού 3	72
Διάγραμμα 55: Πιθανότητα άφιξης στους βασικούς προορισμούς για κάθε ώρα της ημέρας για τον οδηγό 3	73
Διάγραμμα 56: Πιθανότητα εύρεσης του οδηγού 3 στις βασικές τοποθεσίες για κάθε ώρα της ημέρας.....	73
Διάγραμμα 57: Χαρακτηριστική αλυσίδα δραστηριοτήτων του οδηγού 3	74
Διάγραμμα 58: Ώρα άφιξης για όλες τις τοποθεσίες του οδηγού 4	75
Διάγραμμα 59: Χωροχρονικές αφίξεις του οδηγού 4.....	75
Διάγραμμα 60: Heatmap όλων των τοποθεσιών που έχει βρεθεί ο οδηγός 4	76
Διάγραμμα 61: Απεικόνιση όλων των ταξιδιών του οδηγού 4	77
Διάγραμμα 62: Απεικόνιση μοτίβου μετακινήσεων ανά ημέρα για τον οδηγό 4.....	77
Διάγραμμα 63: Κατηγοριοποίηση προορισμών του οδηγού 4.....	78
Διάγραμμα 64: Μοτίβο κινητικότητας του οδηγού 4	79

Διάγραμμα 65: Πιθανότητα άφιξης στους βασικούς προορισμούς για κάθε ώρα της ημέρας για τον οδηγό 4	79
Διάγραμμα 66: Πιθανότητα εύρεσης του οδηγού 4 στις βασικές τοποθεσίες για κάθε ώρα της ημέρας.....	80
Διάγραμμα 67: Χαρακτηριστική αλυσίδα δραστηριοτήτων του οδηγού 4	80
Διάγραμμα 68: Ώρα άφιξης για όλες τις τοποθεσίες του οδηγού 5	81
Διάγραμμα 69: Χωροχρονικές αφίξεις του οδηγού 5	81
Διάγραμμα 70: Heatmap όλων των τοποθεσιών που έχει βρεθεί ο οδηγός 5	82
Διάγραμμα 71: Απεικόνιση όλων των ταξιδιών του οδηγού 5	82
Διάγραμμα 72: Απεικόνιση μοτίβου μετακινήσεων ανά ημέρα για τον οδηγό	83
Διάγραμμα 73: Κατηγοριοποίηση προορισμών του οδηγού 5.....	84
Διάγραμμα 74: Μοτίβο κινητικότητας του οδηγού 5	84
Διάγραμμα 75: Πιθανότητα άφιξης στους βασικούς προορισμούς για κάθε ώρα της ημέρας για τον οδηγό 5	85
Διάγραμμα 76: Πιθανότητα εύρεσης του οδηγού 5 στις βασικές τοποθεσίες για κάθε ώρα της ημέρας.....	85
Διάγραμμα 77: Χαρακτηριστική αλυσίδα δραστηριοτήτων του οδηγού 5	86
Διάγραμμα 78: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Τρίτης με βάση τα τυπικά προφίλ των οδηγών	91
Διάγραμμα 79: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Τετάρτης με βάση τα τυπικά προφίλ των οδηγών.....	91
Διάγραμμα 80: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Πέμπτης με βάση τα τυπικά προφίλ των οδηγών	92
Διάγραμμα 81: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Παρασκευής με βάση τα τυπικά προφίλ των οδηγών	92
Διάγραμμα 82: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα του Σαββάτου με βάση τα τυπικά προφίλ των οδηγών	93
Διάγραμμα 83: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Κυριακής με βάση τα τυπικά προφίλ των οδηγών	93

Ευρετήριο πινάκων

Πίνακας 1: Παράμετρος εισόδου αλγορίθμου K-nearest-neighbors	39
Πίνακας 2: Παράμετροι εισόδου αλγορίθμου DBSCAN.....	39
Πίνακας 3: Στατιστικά ώρας και διάρκειας των βασικών δραστηριοτήτων του οδηγού 1.....	43
Πίνακας 4: Ποσοστά των 5 συνηθέστερων αλυσίδων δραστηριότητας τις Καθημερινές	47
Πίνακας 5: Ποσοστά των 5 συνηθέστερων αλυσίδων δραστηριότητας τα Σαββατοκύριακα.....	48
Πίνακας 6: Σύγκριση ημερών σε σχέση με τη διάρκεια παραμονής στην τοποθεσία Σπίτι	52
Πίνακας 7: Σύγκριση ημερών σε σχέση με τη διάρκεια παραμονής στη τοποθεσία Δουλειά	54
Πίνακας 8: Σύγκριση ημερών σε σχέση με τη διάρκεια παραμονής στις Προσωπικές Δραστηριότητες	57
Πίνακας 9: Σύγκριση ημερών σε σχέση με τη διάρκεια παραμονής στις Υπόλοιπες Τοποθεσίες	59
Πίνακας 10: Στατιστικά ώρας και διάρκειας των βασικών δραστηριοτήτων του οδηγού 2.....	87
Πίνακας 11: Στατιστικά ώρας και διάρκειας των βασικών δραστηριοτήτων του οδηγού 3.....	88
Πίνακας 12: Στατιστικά ώρας και διάρκειας των βασικών δραστηριοτήτων του οδηγού 4.....	89
Πίνακας 13: Στατιστικά ώρας και διάρκειας των βασικών δραστηριοτήτων του οδηγού 5.....	90

Ευρετήριο Εικόνων

Εικόνα 1: Επεξήγηση του χαρακτηρισμού των σημείων του αλγορίθμου DBSCAN (Πηγή: KDnuggets.com)	28
Εικόνα 2: Παράδειγμα λειτουργίας KNN αλγορίθμου (Πηγή: Medium.com)	29
Εικόνα 3: Βασικά διαγράμματα Ανάλυσης Επιβίωσης (Πηγή: ResearchGate.net) ...	30
Εικόνα 4: Επεξήγηση συναρτήσεων Επιβίωσης και Αθροιστικής Πυκνότητας (Πηγή: Wikimedia commons)	31
Εικόνα 5: Αρχική μορφή βάσης δεδομένων	33
Εικόνα 6: Αρχική επεξεργασία βάσης δεδομένων	34

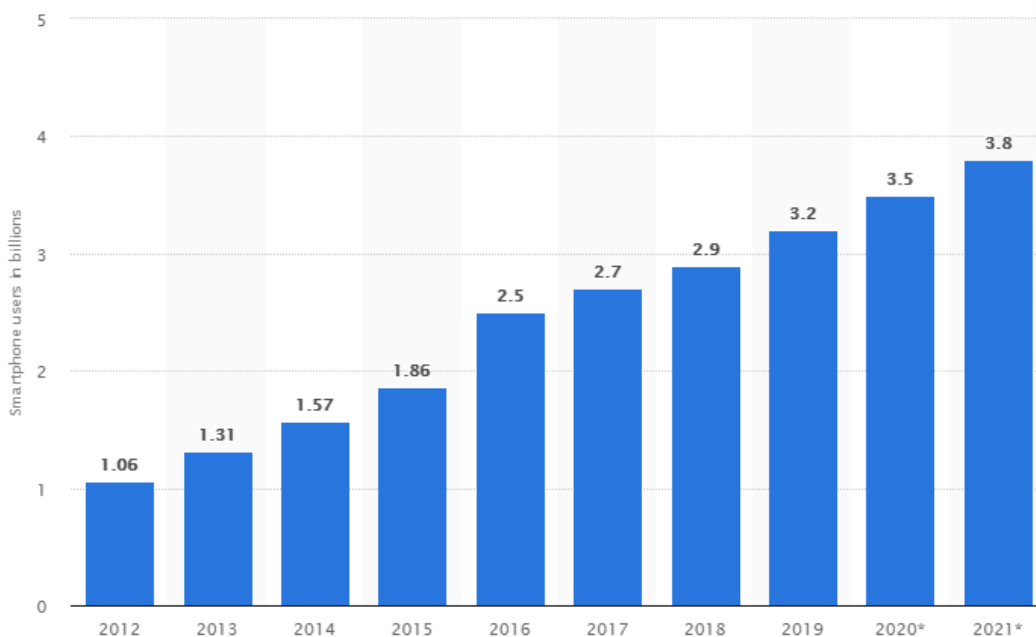
1. Εισαγωγή

1.1 Διαχείριση Κινητικότητας

Η αύξηση της κυκλοφοριακής συμφόρησης στα αστικά κέντρα είναι ένα πολύ σημαντικό πρόβλημα που επηρεάζει σημαντικά και από πολλές απόψεις τη ποιότητα της ζωής σε αυτά (Stathopoulos et al. 2003). Υπηρεσίες κινητικότητας των ανθρώπων όπως Mobility-As-A-Service (MAAS), Mobility-On-Demand (MOD) καθώς και τα αυτόνομα οχήματα έχουν μεγάλη άνθηση. Τέτοιες υπηρεσίες κινητικότητας θεωρείται από τον επιστημονικό κλάδο πως θα βοηθήσουν στην βελτίωση του προβλήματος της κυκλοφοριακής συμφόρησης αλλά και στην ασφάλεια των χρηστών του συγκοινωνιακού δικτύου (Dimitrakopoulos et al. 2010). Αυτές οι υπηρεσίες όμως με σκοπό τη βέλτιστη λειτουργία τους επιζητούν εξατομικευμένες λύσεις κινητικότητας ώστε να επιτευχθεί η σωστή λειτουργία τους. Για το παραπάνω λόγο η ολοκληρωμένη κατανόηση της συμπεριφοράς του κάθε χρήστη στο συγκοινωνιακό δίκτυο αποτελεί προϋπόθεση για την επιτυχία μελλοντικών λύσεων κινητικότητας τόσο στο σχεδιασμό, στη διαχείριση και στη λειτουργία.

Η επιστημονική κοινότητα έχει στρέψει πλέον την προσοχή της στην κατανόηση των επιλογών των ανθρώπων σε σχέση με τις διαφορετικές δραστηριότητες που εκτελούν, τον προγραμματισμό αυτών των δραστηριοτήτων και κατά συνέπεια το αποτύπωμα που αφήνουν κατά τη διάρκεια της ημέρας στο συγκοινωνιακό δίκτυο. Για το λόγο αυτό τα τελευταία χρόνια έχουν αναπτυχθεί πολλά μοντέλα κινητικότητας που διαχωρίζονται σε 3 κατηγορίες (μοντέλα από καταμετρήσεις, μοντέλα βασισμένα σε συντεταγμένες και μαθηματικά μοντέλα). Παρόλα αυτά η δυναμική των οχημάτων στο συγκοινωνιακό δίκτυο και σε ένα σύστημα επικοινωνίας είναι αρκετά περίπλοκη.

Όπως φαίνεται η ανάγκη για δημιουργία και άλλων τέτοιων μοντέλων και δομών είναι μεγάλη και δύσκολη. Σε συνδυασμό όμως με τη ραγδαία αύξηση της τεχνολογίας που καθιστά ευκολότερη και οικονομικότερη τη συλλογή μεγάλου όγκου δεδομένων η δημιουργία αποκτά μια έννοια με λιγότερο κόστος τόσο χρηματικό όσο και χρονικό. Ενδεικτικά και προς επικύρωση των ανωτέρω, στο παρακάτω διάγραμμα διαφαίνεται η αύξηση της χρήσης τηλεφώνων που αποτελούν και κύρια πηγή άντλησης δεδομένων, τα τελευταία χρόνια όσον αφορά στις συγκοινωνιακές έρευνες (Διάγραμμα 1).



Διάγραμμα 1: Αριθμός χρηστών έξυπνων κινητών τηλεφώνων παγκοσμίως (Πηγή: Statista.com)

Με βάση τα παραπάνω είναι αναγκαία η ανάπτυξη αλγοριθμικών δομών για τη πλήρη εξατομικευμένη κατανόηση της κινητικότητας των χρηστών στο συγκοινωνιακό δίκτυο και μέσω του συνόλου των αποτελεσμάτων της εξατομικευμένης ανάλυσης την εξαγωγή συμπερασμάτων που θα βοηθήσουν στον σχεδιασμό, την διαχείριση και την λειτουργία των συστημάτων αστικών μεταφορών. Πιο συγκεκριμένα η δομή που θα παρουσιαστεί είναι ικανή να ανιχνεύσει τα μοτίβα κινητικότητας των ανθρώπων, τις αλυσίδες δραστηριοτήτων και να ποσοτικοποιήσει τις διάρκειες των διαφορετικών δραστηριοτήτων.

1.2 Σκοπός της Εργασίας

Η κατανόηση των επιλογών των ανθρώπων σε σχέση με τις διαφορετικές δραστηριότητες τους, τον προγραμματισμό αυτών και την επίπτωση των επιλογών αυτών στο συγκοινωνιακό δίκτυο είναι το επίκεντρο της συγκοινωνιακής έρευνας για αρκετά χρόνια τώρα. Ουσιαστικά οι επιστήμονες επικεντρώνονται στην ολοκληρωτική κατανόηση της συμπεριφοράς κάθε χρήστη στο συγκοινωνιακό δίκτυο. Πλέον με τη ραγδαία εξέλιξη της τεχνολογίας είναι εύκολο και οικονομικό να σχηματιστούν βάσεις δεδομένων με προσωποποιημένα χαρακτηριστικά των μετακινήσεων των χρηστών.

Συνεπώς, ο σκοπός της παρούσας Διπλωματικής Εργασίας είναι η δημιουργία αλγορίθμων για την κατανόηση του τρόπου μετακίνησης των χρηστών από δεδομένα που συλλέγονται από αισθητήρες κινητών τηλεφώνων. Για το σκοπό αυτό αναπτύσσονται αλγόριθμοι που μέσω μιας βάσης δεδομένων με συντεταγμένες εξάγουν τα προφίλ κινητικότητας, τις καθημερινές χαρακτηριστικές αλυσίδες δραστηριοτήτων των χρηστών και ποσοτικοποιούν τις διάρκειες των διαφορετικών δραστηριοτήτων. Επιπλέον αναπτύσσονται αλγόριθμοι για την ανίχνευση των σημαντικών πληροφοριών μέσω στατιστικών αναλύσεων επί του συνόλου των μοτίβων κινητικότητας, των αλυσίδων-δραστηριοτήτων και τις διάρκειες αυτών που

θα βοηθήσουν στον σχεδιασμό, την διαχείριση και την λειτουργία των συστημάτων αστικών μεταφορών.

Για να επιτευχθούν τα παραπάνω γίνεται οπτικοποίηση των χωροχρονικών χαρακτηριστικών για κάθε χρήστη με βάση τις συντεταγμένες που έχει βρεθεί, κατηγοριοποιούνται και ποσοτικοποιούνται οι βασικοί του προορισμοί με συνεπαγωγή την εύρεση των μοτίβων κινητικότητας και της χαρακτηριστικής αλυσίδας δραστηριοτήτων των χρηστών. Εν συνεχεία μέσω στατιστικών αναλύσεων, προτυποποίησης των διαρκειών ανά τύπο και ημέρα και ποσοτικοποίησης των βασικών προορισμών με βάση όλους τους οδηγούς για κάθε μέρα της εβδομάδας εξάγονται οι παραπάνω προς αναζήτηση πληροφορίες.

1.3 Διάρθρωση της Εργασίας

Η έρευνα η οποία ακολουθεί έχει την εξής δομή :

Το παρόν κεφάλαιο αποτελεί το Κεφάλαιο 1 και είναι η εισαγωγή της παρούσας Διπλωματικής Εργασίας. Πιο συγκεκριμένα, αρχικά γίνεται αναφορά στη σημασία της έρευνας των εξατομικευμένων μοτίβων κινητικότητας των ανθρώπων. Εν συνεχεία παρουσιάζεται ο σκοπός και η δομή της παρούσας Διπλωματικής Εργασίας.

Το Κεφάλαιο 2 αποτελείται από τη βιβλιογραφική ανασκόπηση σε έρευνες που πραγματεύονται παρόμοια αντικείμενα με αυτό της Διπλωματικής Εργασίας και τα συμπεράσματα που εξάχθηκαν από αυτή την ανάλυση της βιβλιογραφίας. Εκ τούτου, υπογραμμίζεται ξανά ο σκοπός της εργασίας.

Το Κεφάλαιο 3 συνίσταται από τη περιγραφή της διαδικασίας της μεθοδολογικής προσέγγισης που ακολουθείται ώστε να επιτευχθεί ο στόχος της διπλωματικής εργασίας καθώς και την αναλυτική επεξήγηση του θεωρητικού υποβάθρου που κατέστησε τη παρούσα έρευνα δυνατή.

Το Κεφάλαιο 4 αποτελείται από την εκτενή ανάλυση και τα αποτελέσματα της μεθοδολογίας που ακολουθήθηκε. Πιο συγκεκριμένα γίνεται η περιγραφή της αρχικής βάσης δεδομένων, ο τρόπος επεξεργασίας της και τα βήματα της μεθοδολογίας που χρησιμοποιήθηκαν παράλληλα και με τα αντίστοιχα αποτελέσματα.

Τέλος το Κεφάλαιο 5 αποτελείται από τα τελικά συμπεράσματα τα οποία προέκυψαν από τις αναλύσεις και γίνονται συστάσεις για περαιτέρω έρευνα στο συγκεκριμένο θέμα. Παρακάτω φαίνονται και σχηματικά τα μέρη από τα οποία απαρτίζεται η εργασία (Διάγραμμα 2).



Διάγραμμα 2: Ροή κεφαλαίων εργασίας

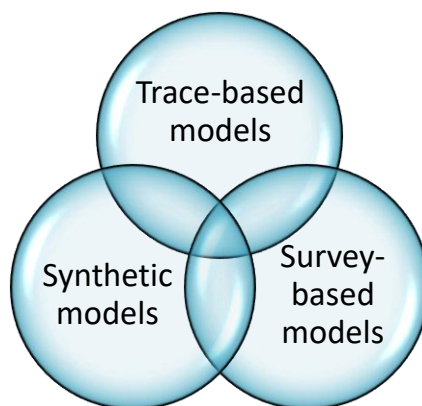
2. Βιβλιογραφική Ανασκόπηση

2.1 Πρότυπα Ανίχνευσης Κινητικότητας

Η ενασχόληση των επιστημόνων όσον αφορά τα παρούσα και μελλοντικά αστικά συγκοινωνιακά συστήματα επικεντρώνεται στην προσωποποιημένη κινητικότητα για κάθε χρήστη ξεχωριστά. Αυτό είναι απαραίτητο για να βελτιωθεί η αποδοτικότητα η βιωσιμότητα και οι κυκλοφοριακές συνθήκες ειδικά λόγω της άνθησης υπηρεσιών μετακίνησης όπως Mobility-As-A-Service (MAAS), Mobility-On-Demand (MOD) και των αυτόνομων οχημάτων που απαιτούν τη παραπάνω γνώση. Συνεπώς η ολοκληρωτική κατανόηση της συμπεριφοράς κάθε χρήστη στο συγκοινωνιακό δίκτυο είναι ζωτικής σημασίας για την επιτυχία μελλοντικών λύσεων κινητικότητας. Παρακάτω θα παρουσιαστούν συνοπτικά οι κυριότερες έρευνες οι οποίες πραγματεύονται την ολοκληρωτική κατανόηση της συμπεριφοράς κάθε χρήστη στο συγκοινωνιακό δίκτυο και την εξαγωγή γενικών συμπερασμάτων μέσω αυτών. Πιο συγκεκριμένα, θα αναλυθεί το πλαίσιο της κάθε έρευνας και τα βασικά αποτελέσματα με έμφαση στα στοιχεία με συνάφεια με το αντικείμενο της διπλωματικής εργασίας.

Οι [Harri et al. \(2009\)](#) κατηγοριοποίησαν τα μοντέλα κινητικότητας με βάση τις επιστημονικές προσπάθειες δημιουργίας αυτών τα τελευταία χρόνια. Πιο συγκεκριμένα ένας μεγάλος αριθμός μοντέλων κινητικότητας είναι διαθέσιμος και αυτά χωρίζονται σε 3 βασικές διαφορετικές κατηγορίες (Διάγραμμα 3):

- Συνθετικά μοντέλα (μέσω μαθηματικών μοντέλων)
- Μοντέλα έρευνας (μέσω καταμετρήσεων)
- Μοντέλα βασισμένα σε αληθινές συντεταγμένες



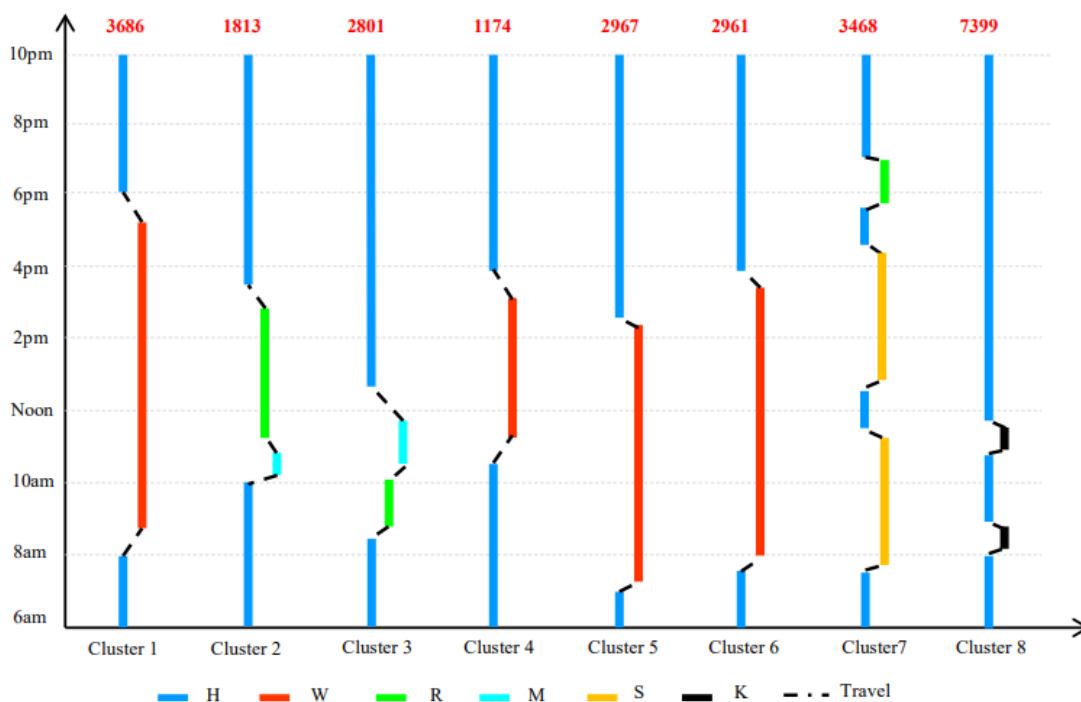
Διάγραμμα 3: Βασικές κατηγορίες μοντέλων κινητικότητας

2.2 Μοτίβα κινητικότητας των ανθρώπων

Οι [Gonzalez et al. \(2008\)](#) πραγματοποίησαν μια έρευνα με στόχο την κατανόηση των ξεχωριστών ανθρωπίνων προτύπων κινητικότητας και την εξαγωγή γενικών συμπερασμάτων με βάση αυτά. Μέσω της μελέτης των συντεταγμένων από 100.000 ανώνυμους χρήστες κινητών τηλεφώνων για μια περίοδο 6 μηνών, διαπιστώθηκε πως οι ανθρώπινες συντεταγμένες παρουσιάζουν μεγάλη κανονικότητα στα

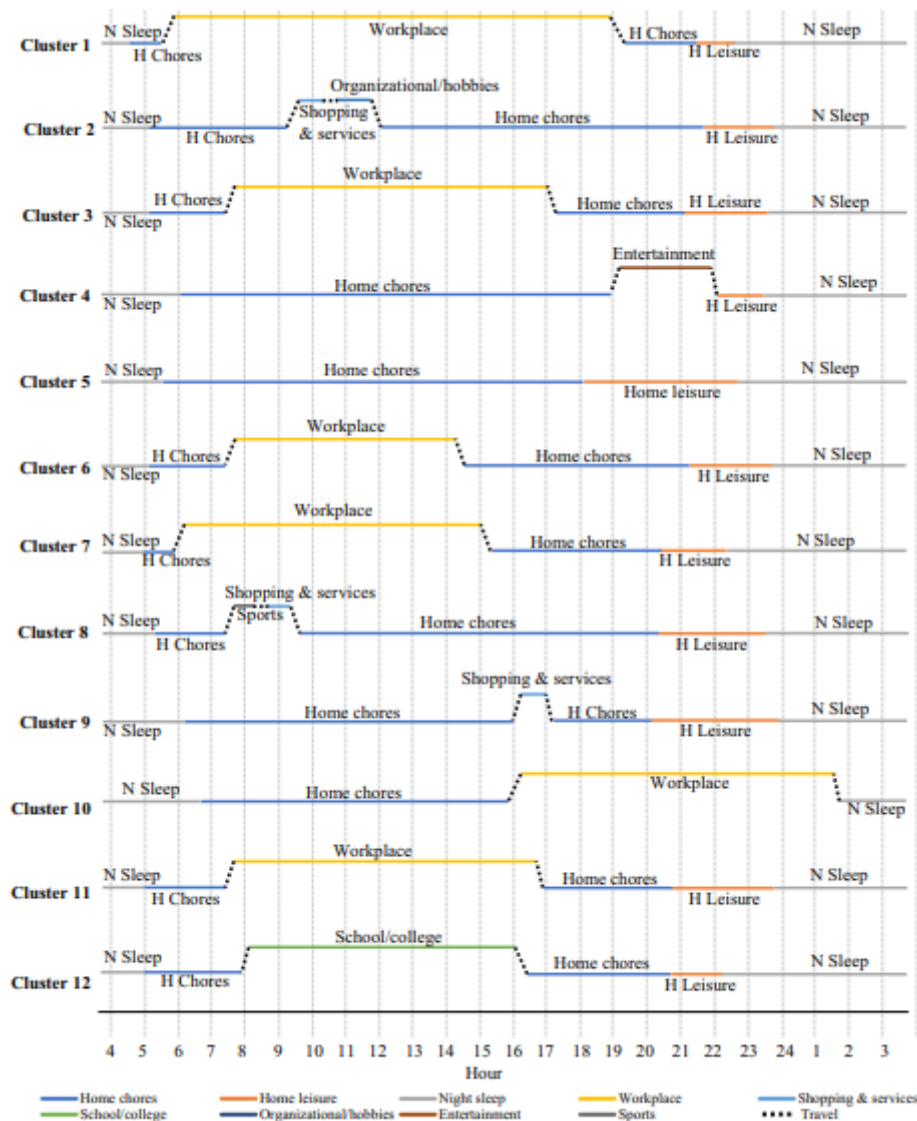
χωροχρονικά τους χαρακτηριστικά. Επιπλέον διαπιστώθηκε ότι κάθε χρήστης χαρακτηρίζεται από μία τυπική ξεχωριστή αλυσίδα δραστηριοτήτων καθώς και από μεγάλη πιθανότητα επιστροφής σε μεγάλης επισκεψιμότητας προορισμούς για τον ίδιο. Εν συνεχεία μέσω της διόρθωσης της ανισοτροπίας της κάθε συντεταγμένης συμπεραίνεται ότι τα ξεχωριστά μοτίβα κινητικότητας συγκεντρώνονται σε μία χωρική πιθανοτική κατανομή που με τη σειρά του αυτό υποδεικνύει πως όλοι οι άνθρωποι ακολουθούν απλά επαναλαμβανόμενα μοτίβα.

Ο Allahviranloo (2016) στην έρευνα την οποία διεξήγαγε με στόχο την αναγνώριση μοτίβων και των εξατομικευμένων συμπεριφορών μετακίνησης χρησιμοποίησε δεδομένα της περιόδου 2000-2001 που αποτελούνταν από τα μοτίβα που ακολουθούν περισσότεροι από 26.000 άνθρωποι. Ουσιαστικά προτάθηκε μια μαθηματική/στατιστική δομή, με στόχο την εξαγωγή της συμπεριφοράς των μετακινήσεων των ατόμων με βάση τα κοινωνικό-δημογραφικά τους προφίλ, με τη χρήση τεχνικών αναγνώρισης μοτίβων. Πιο συγκεκριμένα επιτεύχθηκε η πρόβλεψη με μέση ακρίβεια 80,3% των τύπων των δραστηριοτήτων αυτών των ατόμων καθόλη τη διάρκεια της ημέρας. Επίσης σημειώνεται ότι λόγω ελλείψεων στα trip-based μοντέλα (μοντέλα που ασχολούνται μόνο με τον επόμενο προορισμό) καθώς επικεντρώνονται σε ανάλυση σε συνολικό επίπεδο και χάνουν δραστηριότητες που συνδέονται μεταξύ τους, είναι αναγκαία η εφαρμογή tour-based μοντέλων (μοντέλων που ασχολούνται με το σύνολο των μετακινήσεων του χρήστη και όχι μόνο με τον επόμενο προορισμό). Το συμπέρασμα της μεθόδου ομαδοποίησης ήταν ότι υπάρχουν 8 αντιπροσωπευτικά μοτίβα κατά τη διάρκεια 5π.μ. με 11μ.μ τα οποία φαίνονται παρακάτω στη ξενόγλωσση (Διάγραμμα 4). Όπου H= 'Μέσα στο Σπίτι', W= 'Δουλειά', S= 'Σχολείο', M='Συντήρηση', R= 'Διασκέδαση/Κοινωνικοποίηση', K='Προσωπικά' και Κ= 'Αφήνω-Παίρνω με το αμάξι'.



Διάγραμμα 4: 8 αντιπροσωπευτικά μοτίβα δραστηριοτήτων (Πηγή: Allahviranloo (2016))

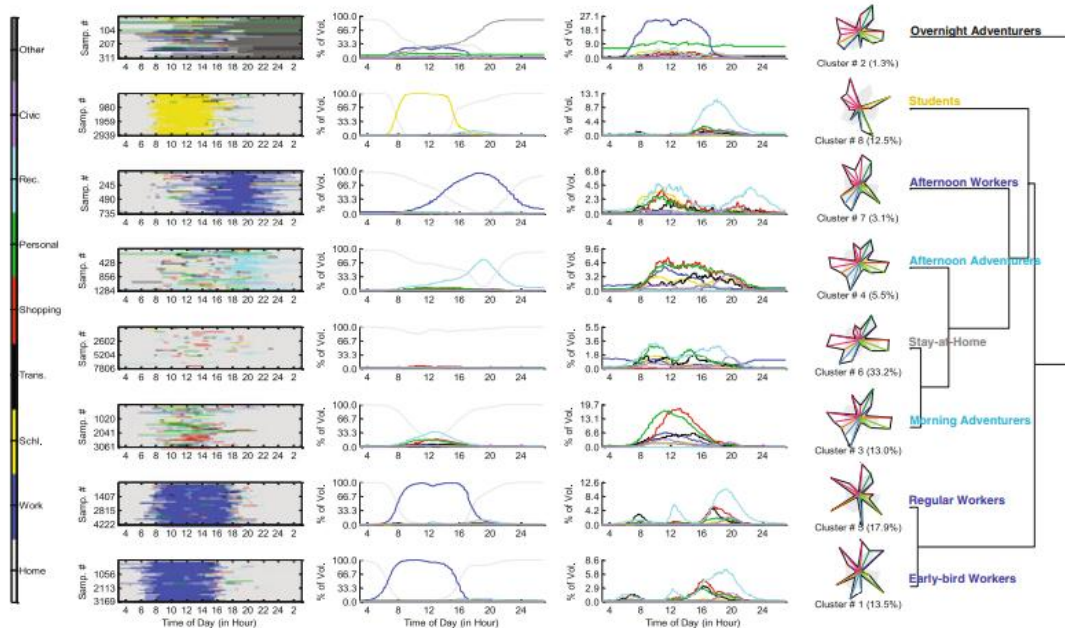
Οι Hafezi et al. (2017) στην έρευνα που διεξήγαγαν χρησιμοποίησαν δεδομένα 1389 ανθρώπων (2 ημέρες ο καθένας). Αντικείμενο αυτής της έρευνας είναι η δημιουργία ενός περιεκτικού μοντέλου αναγνώρισης μοτίβων που εξάγει ομαδοποιήσεις ομοιογενών καθημερινών μοτίβων δραστηριότητας. Πιο συγκεκριμένα μέσω του αλγόριθμου FCM και του δείκτη Dunn αναγνωρίστηκαν 12 διαφορετικές ομαδοποιήσεις ομοιογενών καθημερινών μοτίβων δραστηριοτήτων. Εν συνεχεία μέσω του CART διαχωριστή κατηγοριοποιήθηκαν τα άτομα σε αυτές. Επίσης καταφέρθηκε η πρόβλεψη πληροφοριών που έχουν να κάνουν με την ώρα άφιξης, την διάρκεια παραμονής την απόσταση και τον τρόπο μεταφοράς. Για να επιτευχθούν τα παραπάνω το 24ωρο χωρίστηκε σε χρονικά διαστήματα των 5 λεπτών. Οι ομοιογενείς ομαδοποιήσεις που εξάχθηκαν μέσω του παραπάνω μοντέλου φαίνονται παρακάτω στη ξενόγλωσση (Διάγραμμα 5).



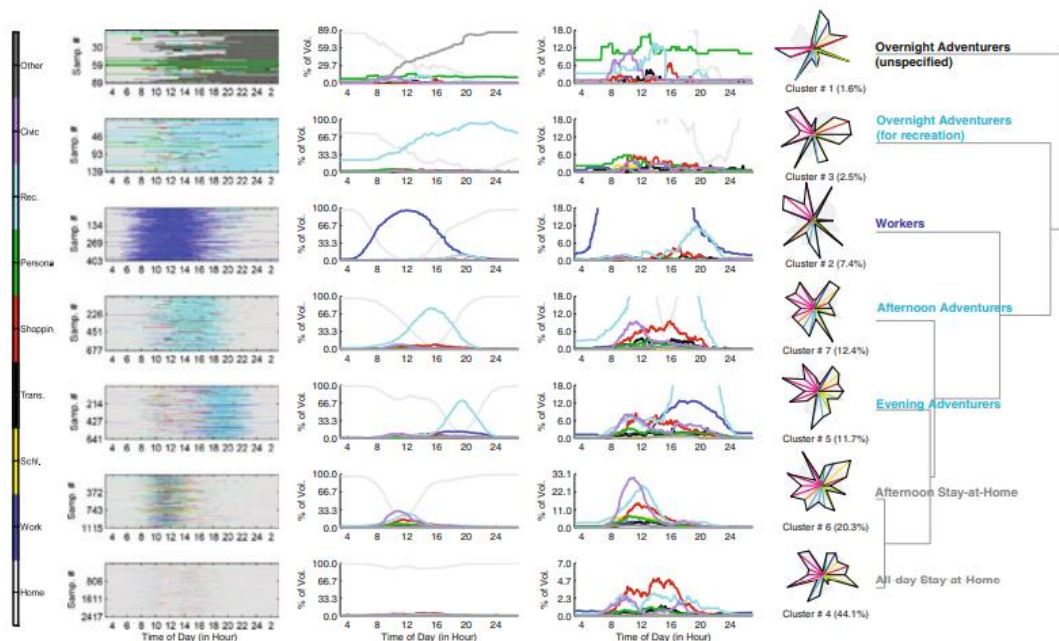
Διάγραμμα 5: 12 αντιπροσωπευτικά μοτίβα δραστηριοτήτων (Πηγή: Hafezi et al. (2017))

Οι Jiang et al. (2012) στην έρευνα τους χρησιμοποίησαν δεδομένα για περισσότερους από 30.000 ανθρώπους για τη περίοδο Ιανουαρίου 2007 έως Φεβρουαρίου 2008. Η ενασχόληση και τα συμπεράσματα της παρούσας έρευνας

χωρίζεται σε Καθημερινές και Σαββατοκύριακα. Πιο συγκεκριμένα αναλύεται η δομή των καθημερινών δραστηριοτήτων των χρηστών, οι παραλλαγές αυτών, η σχέση τους με το χρόνο και τέλος η ομαδοποίηση των καθημερινών συμπεριφορών και η σχέση αυτών με τα κοινωνικό-δημογραφικά τους στοιχεία. Τα αποτελέσματα είναι πως ο πληθυσμός μπορεί να χωριστεί σε 8 και 7 διαφορετικές αντιπροσωπευτικές ομάδες τις καθημερινές και τα Σαββατοκύριακα αντίστοιχα. Τα συνολικά αποτελέσματα της έρευνας ξεχωριστά για τις Καθημερινές (Διάγραμμα 6) και τα Σαββατοκύριακα (Διάγραμμα 7) φαίνονται παρακάτω στη ξενόγλωσση.



Διάγραμμα 6: Ομαδοποίηση Καθημερινών μοτίβων δραστηριότητας και τα κοινωνικό-δημογραφικά τους χαρακτηριστικά (Πηγή: Jiang et al. (2012))



Διάγραμμα 7: Ομαδοποίηση μοτίβων δραστηριότητας τα Σαββατοκύριακα και τα κοινωνικό-δημογραφικά τους χαρακτηριστικά (Πηγή: Jiang et al. (2012))

Οι [Ma et al. \(2013\)](#) στην έρευνα τους προτείνουν μια δομή με στόχο τη μοντελοποίηση των μοτίβων ταξιδιών. Η ενασχόληση τους έγκειται στα ταξίδια που γίνονται μέσω M.M.M. Μέσω των αλγορίθμων DBSCAN, K-means++ και της θεωρίας ακατέργαστων συνόλων εξάχθηκαν τα μοτίβα ταξιδιών μέσω της αλυσίδας των δραστηριοτήτων κάθε χρήστη και εν συνεχεία κατηγοριοποιήθηκαν αυτά τα μοτίβα. Επιπλέον σύγκριναν την παραπάνω δομή αλγορίθμων με άλλους 4 αλγόριθμους κατηγοριοποίησης και τα αποτελέσματα όσον αφορά την ακρίβεια και την απόδοση ήταν καλύτερα. Τα σημαντικά αποτελέσματα αυτής της έρευνας είναι η δημιουργία μιας δομής ικανής να αναγνωρίσει τα μοτίβα των ανθρώπων στις μετακινήσεις τους καθώς και η κατηγοριοποίηση των μοτίβων λόγω της κανονικότητας που ακολουθούν και όλα αυτά για μία μεγάλη βάση δεδομένων.

2.2.1 Επαναληπτικότητα καθημερινών προορισμών

Οι [Susilo et al. \(2014\)](#) στην έρευνα την οποία διεξήγαγαν με αντικείμενο την επαναληπτικότητα των ανθρωπίνων καθημερινών προορισμών (που έχουν να κάνουν με μετακίνηση με μέσο) και ο συνδυασμός αυτών, χρησιμοποίησαν το δείκτη Herfindahl–Hirschman και δεδομένα 6 εβδομάδων. Τα αποτελέσματα δείχνουν ότι η επαναληπτικότητα αυτή επηρεάζεται περισσότερο από τις εξωτερικές δουλειές του Σπιτιού και τις ενδοοικογενειακές καταστάσεις. Πιο συγκεκριμένα υποχρεωτικές δραστηριότητες όπως η Δουλειά το σχολείο και Αφήνω/Παίρνω με το αμάξι δραστηριότητες έχουν μεγαλύτερη επαναληπτικότητα από άλλες δραστηριότητες όπως ταξίδια αναψυχής και επαγγελματικά που παρουσιάζουν μεγαλύτερη ευελιξία. Επιπλέον συμπεραίνεται ότι αυτή η επαναληπτικότητα επηρεάζεται περισσότερο από τις δεσμεύσεις και τις υποχρεώσεις του κάθε ατόμου. Ένα ακόμα ενδιαφέρον εύρημα αυτής της έρευνας είναι ότι οι κάτοικοι αστικών περιοχών τείνουν να έχουν μεγαλύτερη επαναληπτικότητα στα καθημερινά τους μοτίβα. Τέλος φαίνεται από το σύνολο της έρευνας ότι ο δείκτης HHI μπορεί να χρησιμοποιηθεί για τον χαρακτηρισμό της επαναληπτικότητας των μοτίβων των δραστηριοτήτων.

2.3 Πρόβλεψη κινητικότητας οχημάτων

Οι [Yong Li et al. \(2014\)](#) στην έρευνα που πραγματοποίησαν με θέμα τα όρια της προβλεψιμότητας της κινητικότητας οχημάτων χρησιμοποίησαν δύο μεγάλες βάσεις δεδομένων αστικών περιοχών. Μέσω της θεωρίας της εντροπίας αποδεικνύεται ότι υπάρχει κανονικότητα στις χωροχρονικές καθημερινές κινήσεις των οχημάτων, μία πληροφορία που μπορεί να χρησιμοποιηθεί ώστε να αυξηθεί η ακρίβεια της πρόβλεψης των κινήσεων των οχημάτων. Πιο συγκεκριμένα η ακρίβεια της πρόβλεψης του επόμενου προορισμού φτάνει το όριο του 80%-99%. Επιπλέον με προϋπόθεση την χρονική κανονικοποίηση, άνω της 70% ακρίβεια πρόβλεψης μπορεί να επιτευχθεί όσον αφορά τη διάρκεια παραμονής στις δραστηριότητες αντίστοιχα και για τις δύο βάσεις δεδομένων. Συνολικά το σημαντικό εύρημα της παρούσας έρευνας είναι πως η ανάπτυξη μοντέλων και αλγορίθμων πρόβλεψης είναι εφικτή και πλέον με επιστημονική αποδεδειγμένη βάση.

2.4 Συμπεράσματα Βιβλιογραφίας

Με βάση τα παραπάνω συμπεραίνεται ότι οι συντεταγμένες που ακολουθούν οι άνθρωποι παρουσιάζουν μεγάλη κανονικότητα με συνέπεια να εκτελούν επαναλαμβανόμενα μοτίβα και να χαρακτηρίζονται από καθημερινές αλυσίδες.

Σύμφωνα με την υπάρχουσα βιβλιογραφία παρατηρείται ότι επιστήμονες κινούνται στην εξαγωγή των παραπάνω πληροφοριών μέσα από διάφορους τρόπους και κυρίως μέσω μοντέλων έρευνας έχουν εξαχθεί για δείγματα πληθυσμού τα χαρακτηριστικά μοτίβα κινητικότητας που ακολουθούνται από όλους τους χρήστες. Επιπλέον έχουν αναλυθεί και οι λόγοι της παρουσίας επαναληπτικότητας στους προορισμούς που επιλέγει ο κάθε χρήστης. Πλέον επίσης έχει αποδειχτεί ότι η δημιουργία μοντέλων και αλγορίθμων με υψηλή ακρίβεια πρόβλεψης, που έχουν ως δεδομένο αληθινές συντεταγμένες των ανθρώπων, είναι εφικτή με κατάλληλη επεξεργασία των δεδομένων και την εκμετάλλευση της χρονικής κανονικότητας των διαρκειών των δραστηριοτήτων.

Με βάση τα δεδομένα της προηγούμενης παραγράφου και των προηγούμενων υποκεφαλαίων (2.1 – 2.3) παρατηρείται πως η ανάγκη κατασκευής δομών που εξάγουν τα μοτίβα κινητικότητας των ανθρώπων, και της καθημερινής χαρακτηριστικής αλυσίδας αυτών είναι αναγκαία, καθώς καθιστά μια πολύ σημαντική πληροφορία για το σχεδιασμό, την διαχείριση και λειτουργία των αστικών συγκοινωνιακών συστημάτων.

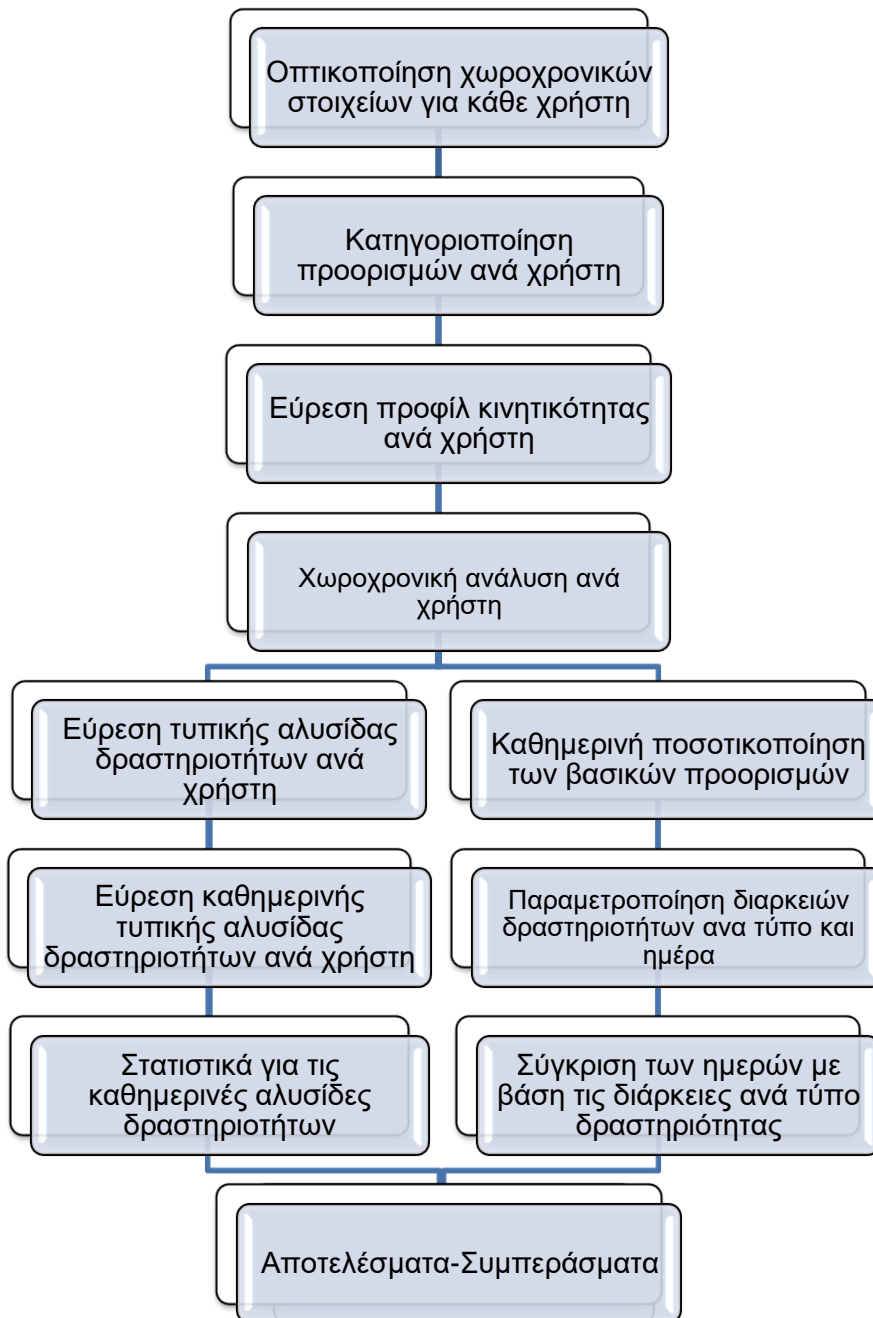
Συνεπώς τα ζητήματα που καλείται να καλύψει η παρούσα Διπλωματική Εργασία, είναι η κατασκευή μιας δομής, διαφορετικής μορφής σε σχέση με τις προϋπάρχουσες που εξάγει μέσω πραγματικών συντεταγμένων τα καθημερινά μοτίβα κινητικότητας των ανθρώπων και την καθημερινή χαρακτηριστική αλυσίδα αυτών. Επιπλέον καλείται να κανονικοποιήσει τις διάρκειες των διαφορετικών δραστηριοτήτων ως μία πληροφορία αύξησης της ακρίβειας της εν δυνάμει πρόβλεψης. Τέλος, καλείται μέσω στατιστικών αναλύσεων να εξάγει πληροφορίες που σχετίζονται με τα παραπάνω με στόχο την ολοκληρωτική κατανόηση της συμπεριφοράς των χρηστών στο συγκοινωνιακό δίκτυο.

3. Μεθοδολογική Προσέγγιση

3.1 Περιγραφή Διαδικασίας

Στόχος της παρούσας Διπλωματικής Εργασίας αποτελεί η εξαγωγή της πληροφορίας των ανθρωπίνων προφίλ κινητικότητας, των αλυσίδων δραστηριοτήτων και των διαρκειών αυτών των δραστηριοτήτων. Οι παραπάνω στόχοι επιτεύχθηκαν μέσω της γλώσσας προγραμματισμού Python στη πλατφόρμα Google Colaboratory. Αρχικά μέσω μιας βάσης δεδομένων ευρείας κλίμακας που εμπεριέχει αναλυτικά στοιχεία των ταξιδιών που πραγματοποιήθηκαν από 696 οδηγούς από τον Δεκέμβριο του 2017 έως τον Αύγουστο του 2019 (η ενασχόληση μας έγκειται σε αυτούς που πραγματοποίησαν περισσότερα από 200 ταξίδια) οπτικοποιήθηκαν τα χωροχρονικά στοιχεία για κάθε χρήστη. Εν συνεχεία για κάθε χρήστη ξεχωριστά κατηγοριοποιήθηκαν οι βασικοί του προορισμοί μέσω των χωροχρονικών χαρακτηριστικών των προορισμών και της συνεργασίας αλγορίθμων ομαδοποίησης και μηχανικής μάθησης και πιο συγκεκριμένα μέσω των αλγορίθμων DBSCAN και K-nearest-neighbors. Επιπλέον ποσοτικοποιήθηκαν για κάθε χρήστη ξεχωριστά τα ταξίδια που πραγματοποιήθηκαν μεταξύ των βασικών του προορισμών με συνεπαγωγή την εξαγωγή του προφίλ κινητικότητας του κάθε χρήστη. Συνεχίζοντας με την ενασχόληση για κάθε χρήστη ξεχωριστά, πραγματοποιήθηκε χωροχρονική ανάλυση και βρέθηκαν για κάθε χρήστη τα τυπικά χωροχρονικά προφίλ (αλυσίδα δραστηριοτήτων). Έχοντας εξάγει πλέον τις βασικές τοποθεσίες για κάθε οδηγό ακολουθεί η ποσοτικοποίηση των προορισμών και των ταξιδιών μεταξύ αυτών για κάθε μέρα της εβδομάδας ξεχωριστά. Στη συνέχεια γίνεται χωροχρονική ανάλυση επί του συνόλου των οδηγών και εξάγονται στατιστικά στοιχεία για τις αλυσίδες δραστηριοτήτων για κάθε μέρα της εβδομάδας ξεχωριστά. Επιπρόσθετα γίνεται προτυποποίηση των διαρκειών των δραστηριοτήτων ανά τύπο και ημέρα της εβδομάδας μέσω της στατιστικής μεθόδου Ανάλυση Επιβίωσης (Survival Analysis) και πιο συγκεκριμένα μέσω του Kaplan Meier estimator και των Καμπύλων Επιβίωσης (της Συνάρτησης Επιβίωσης (Survival Function) και της Συνάρτησης Αθροιστικής Πυκνότητας (Cumulative Distribution Function)). Τέλος συγκρίθηκαν οι στατιστικές διαφορές μεταξύ των ημερών της εβδομάδας, μέσω του Log Rank test, με σκοπιά τις διάρκειες δραστηριοτήτων. Με βάση όλα τα παραπάνω βήματα εξάχθηκαν τα ανάλογα αποτελέσματα και τα εξ αυτών συμπεράσματα.

Στο παρακάτω διάγραμμα ροής παρουσιάζεται με σχηματικό τρόπο η πορεία της μεθοδολογικής προσέγγισης (Διάγραμμα 8).



Διάγραμμα 8: Διάγραμμα ροής εργασιών

3.2 Θεωρητικό Υπόβαθρο

Για την ανάλυση των δεδομένων χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python μέσω της πλατφόρμας Google Colaboratory. Αναλυτικότερα αφού επεξεργάστηκε η βάση δεδομένων και οπτικοποιήθηκαν τα δεδομένα για κάθε χρήστη οδηγό ξεχωριστά στην συνέχεια κατηγοριοποιήθηκαν οι βασικοί τους προορισμοί μέσω της συνεργασίας των αλγορίθμων μη εποπτευόμενης μάθησης DBSCAN και K-nearest-neighbors. Εν συνεχεία μέσω κώδικα στη γλώσσα προγραμματισμού Python ποσοτικοποιήθηκαν οι μετακινήσεις μεταξύ των βασικών προορισμών για κάθε χρήστη (προφίλ κινητικότητας) αλλά και στο σύνολο. Επίσης μέσω κώδικα βρέθηκαν τα χωροχρονικά τυπικά καθημερινά προφίλ (καθημερινή αλυσίδα δραστηριοτήτων) όλων των χρηστών. Τέλος μέσω της στατιστικής μεθόδου

Ανάλυση Επιβίωσης (Survival Analysis) και πιο συγκεκριμένα μέσω της Συνάρτησης Επιβίωσης (Survival Function) και Συνάρτηση Αθροιστικής Πυκνότητας (Cumulative Distribution Function) προτυποποιήθηκαν οι διάρκειες των δραστηριοτήτων ανά τύπο και ημέρα. Τέλος μέσω του Log rank test διαπιστώθηκαν τυχόν στατιστικές διαφορές ανάμεσα στις μέρες της εβδομάδας όσον αφορά τις διάρκειες των δραστηριοτήτων. Στο Διάγραμμα 9 διαφαίνεται το θεωρητικό υπόβαθρο της Διπλωματικής Εργασίας.



Διάγραμμα 9: Θεωρητικό υπόβαθρο εργασίας

3.2.1 Γλώσσα Προγραμματισμού Python

Η Python είναι μία γλώσσα προσακτικού προγραμματισμού και υποστηρίζει τόσο το διαδικαστικό όσο και το αντικειμενοστραφές προγραμματιστικό υπόδειγμα. Για την υλοποίηση των σκοπών της εργασίας χρησιμοποιήθηκαν τα εξής πακέτα στη πλατφόρμα Google Colaboratory :

- files (ανάγνωση αρχείων)
- pandas (πίνακες για χειρισμό και ανάλυση δεδομένων)
- io (διαχείριση λειτουργιών εσόδων και εξόδων που σχετίζονται με το αρχείο)
- matplotlib.pyplot (οπτικοποίηση δεδομένων)
- seaborn (οπτικοποίηση δεδομένων)
- numpy (δημιουργία n-διαστάσεων πινάκων)
- haversine (υπολογισμός αποστάσεων μεταξύ γεωγραφικών μηκών και πλατών)
- folium (οπτικοποίηση δεδομένων σε ρεαλιστικούς χάρτες)
- math (χρησιμοποίηση μαθηματικών συναρτήσεων)
- HeatMap (οπτικοποίηση δεδομένων)
- Axes3D (τριδιάστατη απεικόνιση δεδομένων)
- mplot3d (τριδιάστατη απεικόνιση δεδομένων)
- NearestNeighbors (αλγόριθμος μηχανικής μάθησης)
- DBSCAN (αλγόριθμος μηχανικής μάθησης)
- display (εμφάνιση συμβόλου ποσοστών)
- Digraph (οπτικοποίηση δεδομένων)
- networkx (οπτικοποίηση δεδομένων)

- to_agraph (οπτικοποίηση δεδομένων)
- KaplanMeierFitter (υπολογισμός καμπύλης επιβίωσης και απεικόνιση)
- NelsonAalenFitter (υπολογισμός καμπύλης αθροιστικής πυκνότητας και απεικόνιση)
- logrank_test (σύγκριση στατιστικής διαφοράς μεταξύ δειγμάτων)

3.2.2 Αλγόριθμος ομαδοποίησης DBSCAN

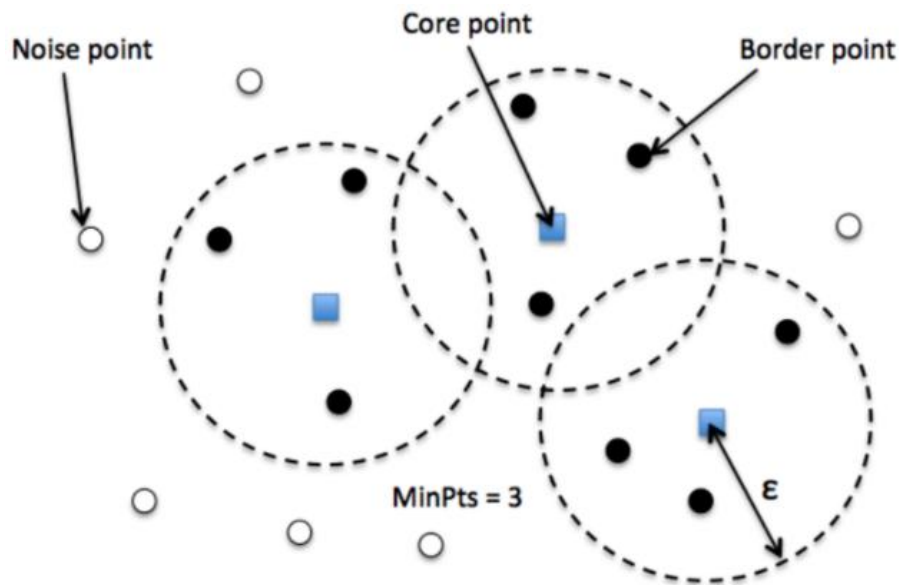
Ο αλγόριθμος ομαδοποίησης DBSCAN (Density based spatial clustering and noise) είναι μία δημοφιλής μέθοδος μη εποπτευόμενης μάθησης που χρησιμοποιείται στη κατασκευή μοντέλων και στους αλγόριθμους μηχανικής μάθησης. Βασίζεται στη πυκνότητα των σημείων και είναι ικανός να βρει ομαδοποιήσεις ανεξαρτήτου σχήματος καθώς επίσης να βρει και ακραίες τιμές (noise) (M.Ester et al.,(1996)). Η κύρια ιδέα πίσω από αυτόν τον αλγόριθμο έγκειται στο ότι ένα σημείο ανήκει σε μια ομαδοποίηση αν βρίσκεται κοντά σε πολλά σημεία από αυτή την ομαδοποίηση. Οι δύο βασικές παράμετροι που χρησιμοποιούνται σε αυτόν τον αλγόριθμο είναι το

- **eps**: Η απόσταση που υποδεικνύει τα γειτονικά σημεία. Δύο σημεία θεωρούνται γειτονικά όταν η απόσταση μεταξύ αυτών είναι μικρότερη η ίση του eps.
- **minpts**: Ο ελάχιστος αριθμός σημείων ώστε να σχηματιστεί ομαδοποίηση.

Βασιζόμενοι σε αυτές τις δύο παραμέτρους ένα σημείο των δεδομένων μπορεί να θεωρηθεί ως σημείο πυρήνα, συνοριακό σημείο ή ακραία τιμή.

- **Σημείο πυρήνα (Core point)**: Ένα σημείο χαρακτηρίζεται έτσι όταν σε ακτίνα eps από αυτό υπάρχουν τουλάχιστον τόσα σημεία (συμπεριλαμβανομένου του ίδιου) όσο ο αριθμός minpts.
- **Συνοριακό σημείο (Border point)**: Ένα σημείο χαρακτηρίζεται έτσι όταν είναι μέσα στην ακτίνα ενός σημείου πυρήνα και στη δικιά του ακτίνα υπάρχουν λιγότερα σημεία από minpts.
- **Ακραία τιμή (Outlier)**: Ένα σημείο χαρακτηρίζεται έτσι όταν δεν είναι σημείο πυρήνα και δεν βρίσκεται στην ακτίνα κανενός σημείου πυρήνα.

Για καλύτερη κατανόηση των ανωτέρω η Εικόνα 1 αποτελεί ένα απλό παράδειγμα προσδιορισμού των σημείων.



Εικόνα 1: Επεξήγηση του χαρακτηρισμού των σημείων του αλγορίθμου DBSCAN (Πηγή: KDnuggets.com)

Ο τρόπος λειτουργίας του αλγορίθμου ακολουθεί την εξής σειρά.

1. Οι παράμετροι $minpts$ και ϵ καθορίζονται.
2. Ένα αρχικό σημείο επιλέγεται στη τύχη. Αν σε ακτίνα ϵ από αυτό το σημείο βρίσκονται ίσα ή περισσότερα σημεία της τιμής $minpts$ τότε αυτό θεωρείται ως σημείο πυρήνα (core point) και η δημιουργία ομαδοποίησης ξεκινά. Αν δεν ισχύει το παραπάνω τότε αυτό το σημείο θεωρείται ως ακραία τιμή (outlier). Όταν έχει ξεκινήσει η δημιουργία ομαδοποίησης, όλα τα σημεία που βρίσκονται εντός της ακτίνας με τιμή ϵ συγκαταλέγονται στην ομαδοποίηση. Αν και αυτά τα σημεία είναι σημεία πυρήνα (core points) τότε τα σημεία με απόσταση μικρότερη του ϵ από αυτά προστίθενται στην παραπάνω ομαδοποίηση. Σημεία που έχουν χαρακτηριστεί ως ακραίες τιμές (outliers) μπορεί να ξαναεπισκεφτούν και να γίνουν μέρος άλλης ομαδοποίησης.
3. Επόμενο βήμα είναι η τυχαία επιλογή από τα σημεία που δεν έχουν επισκεφτεί στα προηγούμενα βήματα και η ίδια διαδικασία επαναλαμβάνεται.
4. Η διαδικασία τελειώνει όταν όλα τα σημεία έχουν επισκεφτεί.

Η μέθοδος μέτρησης της απόστασης μεταξύ των σημείων καθορίζεται από το χρήστη.

Εφαρμόζοντας τα παραπάνω βήματα ο αλγόριθμος DBSCAN είναι ικανός να κατηγοριοποιήσει με τις κατάλληλες παραμέτρους σε μεγάλης και μικρής πυκνότητας περιφέρειες όλα τα σημεία που έχει ως είσοδο.

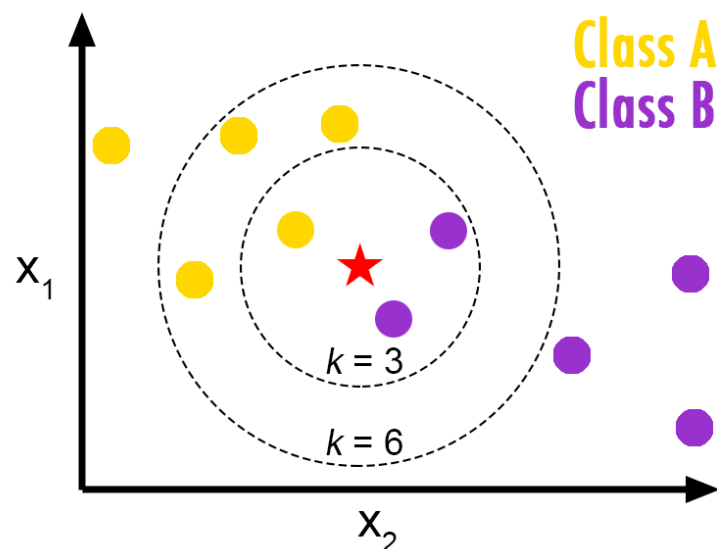
3.2.3 Αλγόριθμος μηχανικής μάθησης K-nearest-neighbors

Ο αλγόριθμος K-nearest-neighbors είναι ένας απλός αλγόριθμος μηχανικής μάθησης που μπορεί να χρησιμοποιηθεί για επίλυση προβλημάτων ομαδοποίησης και παλινδρόμησης. Τα βήματα λειτουργίας που χρησιμοποιήθηκαν στη παρούσα Διπλωματική Εργασία είναι τα εξής :

1. Επιλογή αριθμού K των γειτόνων.

2. Υπολογισμός της απόστασης των K αριθμών των γειτόνων για κάθε σημείο στα δεδομένα.
3. Ανακατάταξη των παραπάνω αποστάσεων σε φθίνουσα σειρά.
4. Εύρεση του 80% της μέγιστης απόστασης ώστε να δεχτεί σαν είσοδο ο αλγόριθμος DBSCAN αυτόν τον αριθμό σαν παράμετρο ϵ . Αυτό υποδεικνύει ότι σε κάθε ομαδοποίηση το 80% θα είναι σημεία πυρήνα (core points).

Για καλύτερη κατανόηση των ανωτέρω παρατίθεται η Εικόνα 2 που παριστά ένα μικρό παράδειγμα λειτουργίας του αλγορίθμου.



Εικόνα 2: Παράδειγμα λειτουργίας KNN αλγορίθμου (Πηγή: Medium.com)

Ουσιαστικά αυτός ο αλγόριθμος χρησιμοποιήθηκε ως προς την εύρεση εκείνης της ελάχιστης ακτίνας, με ορισμένο αριθμό K ίσο με $minpts$ του DBSCAN.

3.2.4 Ανάλυση Επιβίωσης (Survival Analysis)

Η Ανάλυση Επιβίωσης (Survival Analysis) είναι μια στατιστική μέθοδος που ασχολείται με το πόσο χρόνο χρειάζεται μέχρι ένα συγκεκριμένο γεγονός να λάβει χώρα.

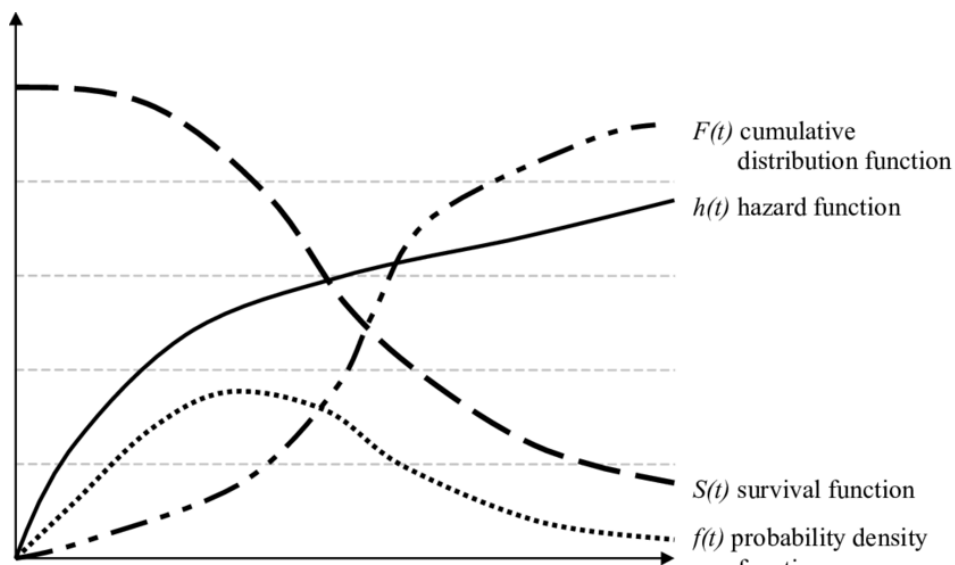
Τα εργαλεία που χρησιμοποιήθηκαν (Διάγραμμα 10) για την επίτευξη αυτής της ανάλυσης ήταν τα εξής:

- **Kaplan Meier Estimator** για τον υπολογισμό της συνάρτησης Επιβίωσης (Survival function) και τη καμπύλη αυτής.
- **Kaplan Meier Estimator** για τον υπολογισμό της συνάρτησης της Αθροιστικής Πυκνότητας (Cumulative distribution function) και τη καμπύλη αυτής.
- **Log Rank Test** για τη σύγκριση καμπύλων επιβίωσης.



Διάγραμμα 10: Εργαλεία που χρησιμοποιήθηκαν για την Ανάλυση Επιβίωσης (Survival Analysis)

Πέρα από τις καμπύλες που χρησιμοποιήθηκαν στην εκπόνηση της Διπλωματικής Εργασίας υπάρχουν και άλλες μορφές καμπύλων (Εικόνα 3) που συμβάλλουν στην Ανάλυση Επιβίωσης (Survival Analysis). Παρακάτω φαίνεται η μορφή των βασικών καμπύλων.



Εικόνα 3: Βασικά διαγράμματα Ανάλυσης Επιβίωσης (Πηγή: ResearchGate.net)

A) Kaplan Meier estimator και Συνάρτηση Επιβίωσης

Ο Kaplan Meier estimator είναι ένας μη παραμετρικός στατιστικός εκτιμητής που χρησιμοποιείται για τον υπολογισμό της Συνάρτησης Επιβίωσης (Survival Function). Ο εκτιμητής της συνάρτησης επιβίωσης $S(t)$ δηλαδή της πιθανότητας ότι η διάρκεια επιβίωσης είναι μεγαλύτερη από t δίνεται από :

$$\diamond S(t) = \prod_{i: t_i \leq t} (1 - d_i / n_i) \quad (1)$$

Όπου t_i είναι ο χρόνος που τουλάχιστον ένα γεγονός συνέβη, d_i ο αριθμός των γεγονότων (στη παρούσα εργασία τερματισμός δραστηριότητας) που συνέβησαν έως τη χρονική στιγμή t_i , και n_i ο αριθμός των γεγονότων που δεν έχουν συμβεί μέχρι τη χρονική στιγμή t_i .

Υπολογίζοντας τον παραπάνω εκτιμητή δημιουργείται η συνάρτηση Επιβίωσης (Survival Function) η οποία είναι η συνάρτηση που δίνει τη πιθανότητα επιβίωσης ενός αντικειμένου ενδιαφέροντος πριν από κάποια ορισμένη χρονική στιγμή.

B) Kaplan Meier estimator και Συνάρτηση Αθροιστικής Πυκνότητας

Έστω ότι T είναι μία συνεχόμενη τυχαία μεταβλητή με συνάρτηση αθροιστικής πυκνότητας $F(t)$ στο διάστημα $[0, \infty)$. Η συνάρτηση Επιβίωσης είναι :

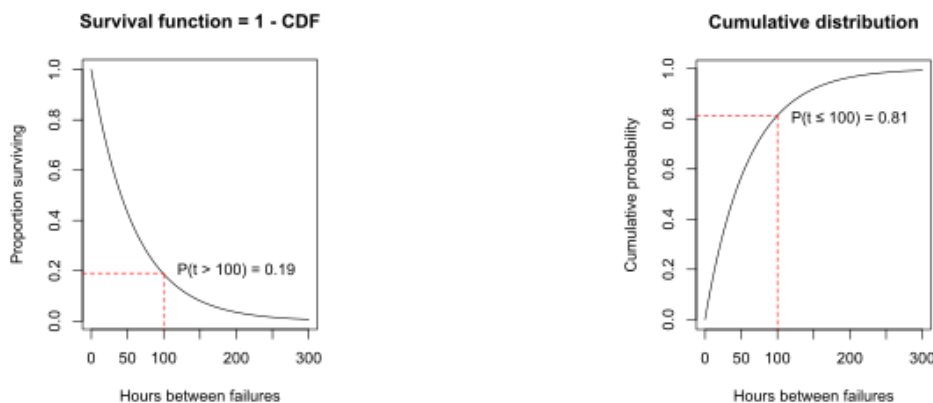
$$\diamond S(t) = P(\{T > t\}) = \int_t^{\infty} f(u) du = 1 - F(t) \quad (2)$$

Όπου η συνάρτηση Αθροιστικής Πυκνότητας (cumulative density) μίας τυχαίας μεταβλητής X είναι η συνάρτηση που ορίζεται από την εξής μαθηματική εξίσωση

$$\diamond F_X = P(X \leq x) \quad (3)$$

Με το δεξιό μέρος της συνάρτησης να συμβολίζει τη πιθανότητα μια τυχαία μεταβλητή X να παίρνει τιμή μικρότερη ή ίση με το x .

Παρακάτω φαίνεται σε μορφή γραφήματος η Συνάρτηση Επιβίωσης καθώς και η Συνάρτηση Αθροιστικής Πυκνότητας με παρατηρήσεις και ενδείξεις στη ξενόγλωσση που επεξηγούν τη μορφή τους (Εικόνα 4).



Εικόνα 4: Επεξήγηση συναρτήσεων Επιβίωσης και Αθροιστικής Πυκνότητας (Πηγή: Wikimedia commons)

Ουσιαστικά οι παραπάνω συναρτήσεις υποδεικνύουν τη πιθανότητα επιβίωσης και τη πιθανότητα τερματισμού επιβίωσης στο πέρασμα μια χρονοσειράς αντίστοιχα.

Γ) Log Rank Test

Το Log Rank Test είναι ένα μη παραμετρικό τεστ υπόθεσης που συγκρίνει τις κατανομές επιβίωσης ανάμεσα σε δύο ή περισσότερα δείγματα. Πιο συγκεκριμένα συγκρίνει τις εκτιμήσεις των συναρτήσεων επιβίωσης για κάθε παρατηρήσιμη χρονική στιγμή. Κατασκευάζεται από τον υπολογισμό του αριθμού των παρατηρήσιμων και αναμενόμενων συμβάντων σε μία από τις δύο ομάδες για κάθε παρατηρήσιμη χρονική στιγμή-συμβάν και στη συνέχεια προσθέτει αυτά ώστε να βγει ένα γενικό συμπέρασμα σε σχέση με όλα τα χρονικά σημεία που πραγματοποιήθηκε ένα γεγονός.

Έστω δύο ομάδες 1,2 προς σύγκριση.

Έστω $1, \dots, J$ οι διακριτές φορές των παρατηρήσιμων γεγονότων για κάθε ομάδα.

Έστω $N_{1,j}$ και $N_{2,j}$ ο αριθμός των γεγονότων που δεν έχουν γίνει ακόμα τη χρονική στιγμή j .

Έστω $O_{1,j}$ και $O_{2,j}$ ο παρατηρήσιμος αριθμός γεγονότων των ομάδων τη χρονική στιγμή j .

Έστω $N_j = N_{1,j} + N_{2,j}$ και $O_j = O_{1,j} + O_{2,j}$

Η μηδενική υπόθεση είναι ότι οι δύο ομάδες προ σύγκριση έχουν πανομοιότυπες συναρτήσεις Επιβίωσης δηλαδή $H_0: h_1(t) = h_2(t)$.

Η υπέρ-γεωμετρική κατανομή πίσω από H_0 για κάθε ομάδα $i = 1, 2$, $O_{i,j}$ με παραμέτρους $N_j, N_{i,j}, O_j$ έχει τιμή $E_{i,j} = N_{i,j} O_j / N_j$ και διακύμανση

$$V_{i,j} = E_{i,j} \left(\frac{N_j - O_j}{N_j} \right) \left(\frac{N_j - N_{i,j}}{N_j - 1} \right) \quad (4)$$

Για όλα τα $j = 1, \dots, J$ το στατιστικό τεστ Log Rank συγκρίνει το $O_{i,j}$ με το αναμενόμενο $E_{i,j}$ κάτω του H_0 και χαρακτηρίζεται από τη παρακάτω μαθηματική εξίσωση

$$Z = \frac{\sum_j^J (O_{i,j} - E_{i,j})}{\sqrt{\sum_j^J V_{i,j}}} \xrightarrow{d} N(0,1) \text{ για } i=1 \text{ ή } 2 \quad (5)$$

Ουσιαστικά αν το αποτέλεσμα αυτού του στατιστικού τεστ και πιο συγκεκριμένα της τιμής του p είναι μικρότερη του 0.05 σημαίνει ότι οι δύο ομάδες που εξετάστηκαν έχουν σημαντική στατιστική διαφορά.

4. Ανάλυση και Αποτελέσματα

4.1 Περιγραφή βάσης δεδομένων

Η αρχική μορφή της βάσης δεδομένων περιέχει 153953 ταξίδια που πραγματοποιήθηκαν στην Ελλάδα τη περίοδο Δεκεμβρίου 2017 έως Αύγουστο 2019. Κάθε χρήστης-οδηγός έχει κάνει από 20 έως 2711 ταξίδια. Κάθε γραμμή αντιστοιχεί σε ένα ταξίδι. Συγκεκριμένα για κάθε ταξίδι έχουμε:

- tripid: ένα μοναδικό id για κάθε ταξίδι
- driverid: id του οδηγού που έκανε το ταξίδι
- tripended: ώρα και ημερομηνία που έφτασε στον προορισμό
- tripstarted: ώρα και ημερομηνία που ξεκίνησε το ταξίδι
- duration: διάρκεια του ταξιδιού σε δευτερόλεπτα
- distance_total: μήκος διαδρομής σε χιλιόμετρα
- longitude_end, latitude_end: συντεταγμένες σημείου προορισμού
- longitude_start, latitude_start: συντεταγμένες σημείου προέλευσης

Για καλύτερη κατανόηση του dataset στην Εικόνα 5 φαίνεται η μορφή του.

	tripid	driverid	tripended	tripstarted	duration	distance_total	longitude_end	latitude_end	longitude_start	latitude_start
0	00000261-A935-44B2-973F-61402522A2F9	986	2019-05-08T16:23:42.000+0300	2019-05-08T15:27:18.396+0300	3385	83.207933	24.388899	40.922304	25.187223	41.125152
1	00004AFC-E870-46F0-BAB7-AAB0139C240C	15	2018-10-12T12:35:10.000+0300	2018-10-12T12:33:27.960+0300	102	0.801772	23.794172	37.987561	23.800403	37.991179
2	00005A13-A0D6-48B0-8751-A60CABBBC3443	1085	2019-05-09T11:08:23.000+0300	2019-05-09T10:29:22.189+0300	2341	15.566860	23.694165	37.950428	23.675677	37.993737
3	00011DCF-4CDC-48E5-8FFA-D977728A20F0	104	2018-04-14T23:42:35.000+0300	2018-04-14T23:41:14.109+0300	81	0.539993	23.834691	38.032177	23.831831	38.035411
4	00014E90-F704-4FDB-B7D8-0AD812C83A13	208	2019-01-23T19:21:31.000+0200	2019-01-23T18:32:27.099+0200	2945	13.888555	23.740153	37.978229	23.854147	38.024992
...
153948	FFFF423C-3D44-49B8-A9AD-DB3B553844B3	788	2019-07-31T07:51:39.000+0300	2019-07-31T07:34:39.741+0300	1019	9.599012	23.713786	37.948559	23.641154	37.979062
153949	FFFF5F59-C9B2-4910-A2F6-88F76B85ECCB	783	2019-07-28T01:07:00.000+0300	2019-07-28T00:54:39.953+0300	741	6.396318	23.740931	37.964331	23.695944	37.937535
153950	FFFF85EB-6B66-4A30-9AD8-40801E9DEAF4	160	2019-05-19T09:36:42.000+0300	2019-05-19T09:17:03.421+0300	1178	17.347190	26.012418	38.214763	26.100801	38.316609
153951	FFFF8AAB-AFCD-4CDD-BC04-31C2E60E8D0E	993	2019-05-19T22:13:26.000+0300	2019-05-19T22:05:24.974+0300	482	1.894998	23.843506	38.032061	23.840942	38.034640
153952	FFFFD309-47B5-45DB-90EA-DCE7D54CAD55	208	2018-02-23T22:25:14.000+0200	2018-02-23T21:53:37.188+0200	1898	29.518685	23.645830	37.939679	23.836343	38.101842

153953 rows x 10 columns

Εικόνα 5: Αρχική μορφή βάσης δεδομένων

4.2 Αρχική επεξεργασία βάσης δεδομένων

Αρχικά πραγματοποιείται ανακατάταξη της βάσης δεδομένων με βάση τον οδηγό και τη χρονολογική σειρά των ταξιδιών του κάθε χρήστη. Οπότε έχουμε ως κατάληξη την παρακάτω δομή των δεδομένων (Εικόνα 6).

	tripid	driverid	tripeded	tripstarted	duration	distance_total	longitude_end	latitude_end	longitude_start	latitude_start
28939	300E76B1-6538-42E0-8227-2C330881F31E	0	2019-03-28T07:31:26.000+0200	2019-03-28T07:26:59.267+0200	268	2.878174	22.801831	37.567184	22.815462	37.584628
21033	22F7EE79-C036-40BA-B67A-AF74A42DFD2B	0	2019-03-28T15:53:59.000+0200	2019-03-28T15:47:12.150+0200	407	4.357984	22.810559	37.596629	22.802942	37.564933
119240	C5E792FE-EE79-4D59-9432-5A5749398BB1	0	2019-03-29T07:30:19.000+0200	2019-03-29T07:26:42.927+0200	217	1.812701	22.801795	37.567168	22.814132	37.575838
43406	480CF238-9276-427A-9F85-1C596C8CAB90	0	2019-03-29T15:15:37.000+0200	2019-03-29T15:06:39.228+0200	539	4.536248	22.810598	37.596629	22.801387	37.565672
118999	C57DDE36-8557-49BD-895E-52E7C85B69B4	0	2019-03-29T17:41:03.000+0200	2019-03-29T17:34:57.128+0200	366	5.521724	22.745447	37.586957	22.799556	37.597037
...
97738	A2787C2A-A3C8-416F-9D8D-872D0E3D9446	1279	2019-07-31T12:14:39.000+0300	2019-07-31T12:02:27.434+0300	732	5.076759	22.695770	40.735761	22.699220	40.731815
53133	5806D2CE-D19C-4DD8-AE9C-DB60156C1945	1279	2019-07-31T13:23:21.000+0300	2019-07-31T13:08:22.145+0300	899	16.242303	22.834721	40.688311	22.700072	40.731481
42914	473353EE-EEDB-4F3B-835E-8A5914F241DC	1279	2019-07-31T14:57:19.000+0300	2019-07-31T14:24:29.476+0300	1970	24.866554	22.695804	40.735799	22.930807	40.642795
47262	4E72FA0C-4FCF-478A-AD0B-4FA5BAF296CC	1279	2019-07-31T22:27:36.000+0300	2019-07-31T22:24:09.328+0300	208	2.065735	22.695721	40.735790	22.695178	40.721914
88360	92FA1931-2130-4F79-BC7E-939E86E28F59	1279	2019-07-31T22:37:41.000+0300	2019-07-31T22:34:16.354+0300	205	1.793653	22.687579	40.722245	22.699373	40.728190

153953 rows x 10 columns

Εικόνα 6: Αρχική επεξεργασία βάσης δεδομένων

Στο σύνολο τους οι οδηγοί στα δεδομένα είναι 696 με τους 263 από αυτούς να έχουν συνολικά πάνω από 200 ταξίδια στη περίοδο του Δεκεμβρίου 2017 έως τον Αύγουστο 2019. Η ενασχόληση και η εξαγωγή συμπερασμάτων αφορά αυτούς με παραπάνω από 200 ταξίδια στο σύνολο. Αυτό γίνεται για σκοπούς εξαγωγής ρεαλιστικών συμπερασμάτων. Επιπλέον τα χρονικά δεδομένα της ώρας άφιξης και αφετηρίας των ταξιδιών ομαδοποιήθηκαν σε χρονικά διαστήματα της 1 ώρας μέσα στην ημέρα (π.χ. 00:18 → 0, 23:52 → 23)

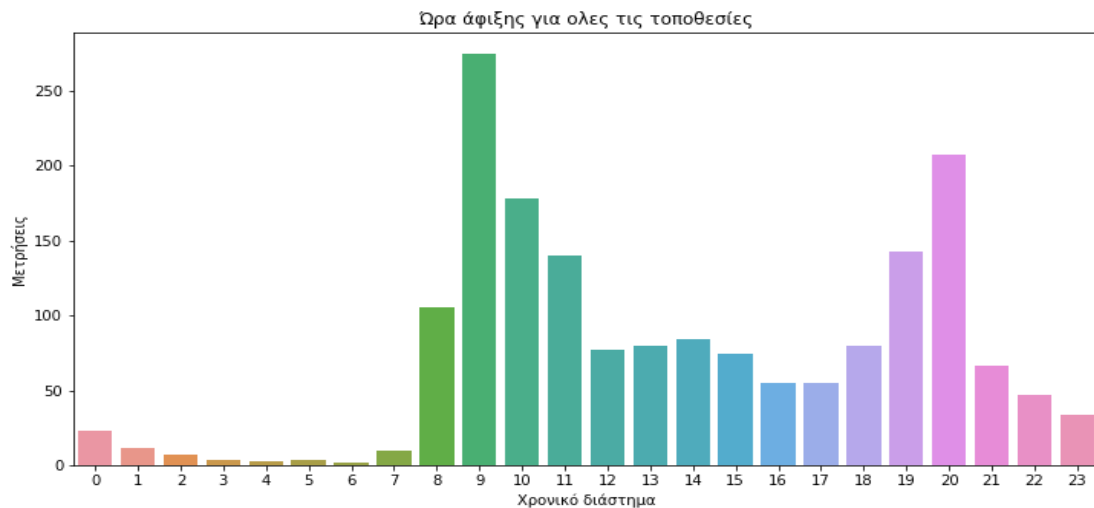
4.3 Οπτικοποίηση δεδομένων ανά οδηγό

Με στόχο τη καλύτερη κατανόηση των δεδομένων ευρείας κλίμακας η οπτικοποίηση αυτών για κάθε οδηγό ξεχωριστά είναι αναγκαία. Στο παρόν κεφάλαιο καθώς και στο παράρτημα Α παρουσιάζονται διαγράμματα που σχετίζονται με γενικές πληροφορίες των τοποθεσιών του προορισμού αλλά και της αφετηρίας των ταξιδιών για 5 χαρακτηριστικούς οδηγούς από χωρική αλλά και από χρονική σκοπιά.

Απώτερος σκοπός της παρούσας μεθοδολογίας αποτελεί η βοήθεια στη κατανόηση των επόμενων μεθοδολογιών και των συμπερασμάτων που θα εξαχθούν.

Στα υποκεφάλαια 4.3 έως 4.6 παρουσιάζεται η ανάλυση για έναν χαρακτηριστικό οδηγό. Τα ανάλογα αποτελέσματα και διαγράμματα, ώστε να κατανοηθεί η εγκυρότητα της μεθοδολογίας, για άλλους 4 χαρακτηριστικούς οδηγούς υπάρχουν στο Παράρτημα Α και Β.

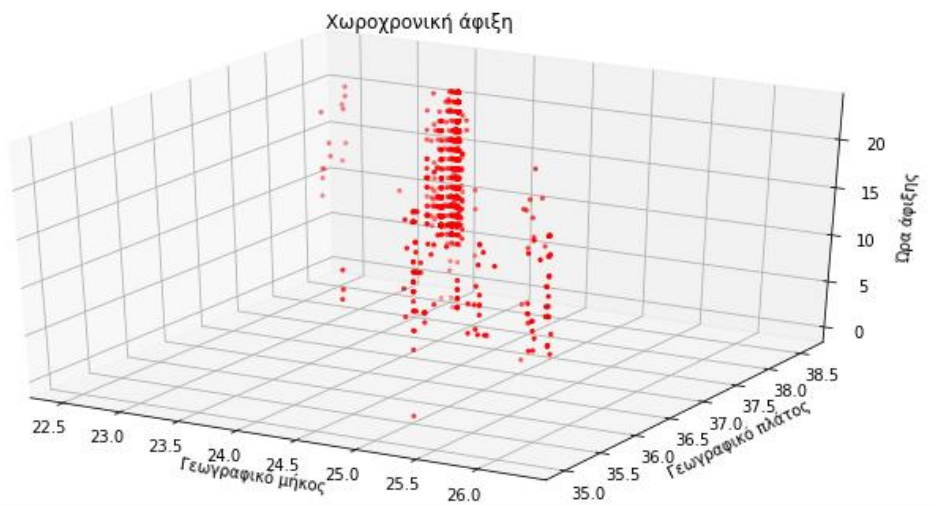
Αρχικά απομονώνονται τα ταξίδια που έχει πραγματοποιήσει ο κάθε οδηγός ώστε να βρούμε πληροφορίες και να εξάγουμε συμπεράσματα εξ αυτών. Αναγκαίος τρόπος στην αρχή για να επιτευχθεί αυτό, ειδικά σε βάσεις δεδομένων ευρείας κλίμακας, είναι η οπτικοποίηση των δεδομένων. Πρώτος στόχος αποτελεί να βρούμε για τον οδηγό αν υπάρχει κάποια επαναληπτικότητα στις αφίξεις χρονικά. Στο Διάγραμμα 11 (count plot) φαίνονται για τον οδηγό 1 οι μετρήσεις των αφίξεων στις τοποθεσίες για κάθε ώρα ξεχωριστά μέσα στην ημέρα.



Διάγραμμα 11: Ωρα άφιξης για όλες τις τοποθεσίες του οδηγού 1

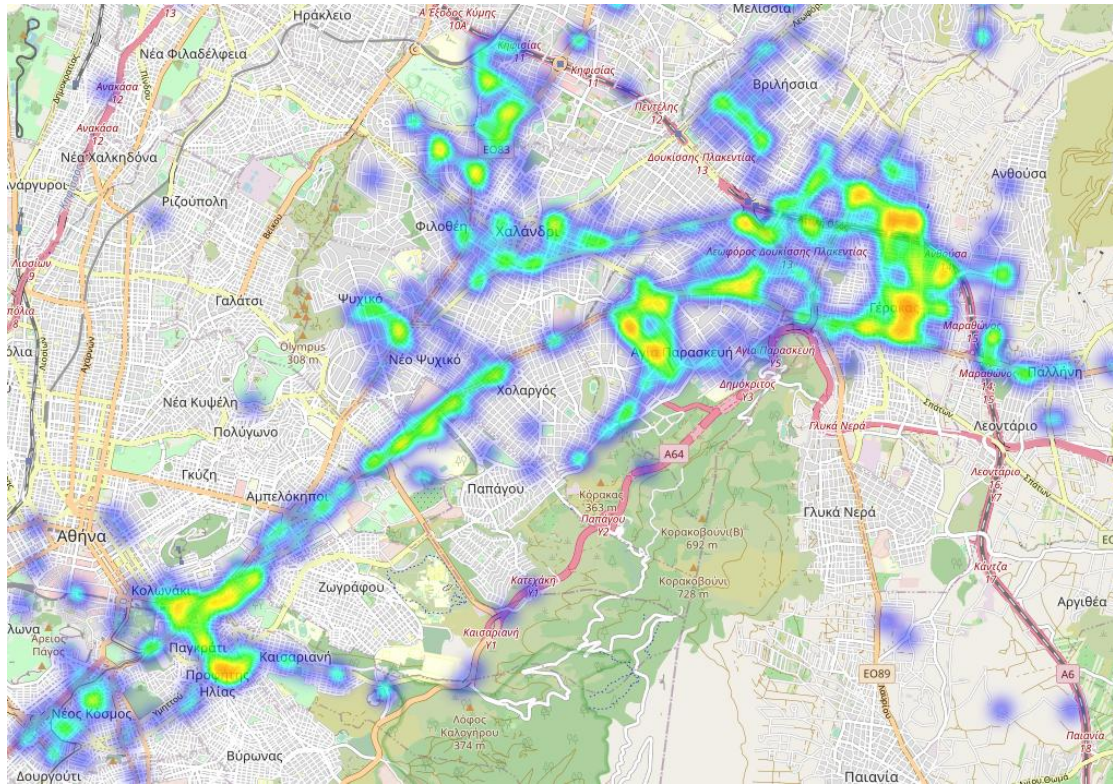
Παρατηρείται πως ο οδηγός 1 πραγματοποιεί τις περισσότερες του μετακινήσεις τις πρωινές (8-9-10-11) και τις βραδινές ώρες (19-20) της ημέρας.

Σε αλληλουχία με το παραπάνω διάγραμμα που αποτελεί μόνο χρονική ανάλυση, κατασκευάζεται ένα χωροχρονικό 3D διάγραμμα για τον οδηγό 1 για τις τοποθεσίες άφιξης και τις αντίστοιχες ώρες ώστε να εξαχθούν συμπεράσματα για συνήθειες τοποθεσίες προορισμού του χρήστη (Διάγραμμα 12).



Διάγραμμα 12: Χωροχρονικές αφίξεις του οδηγού 1

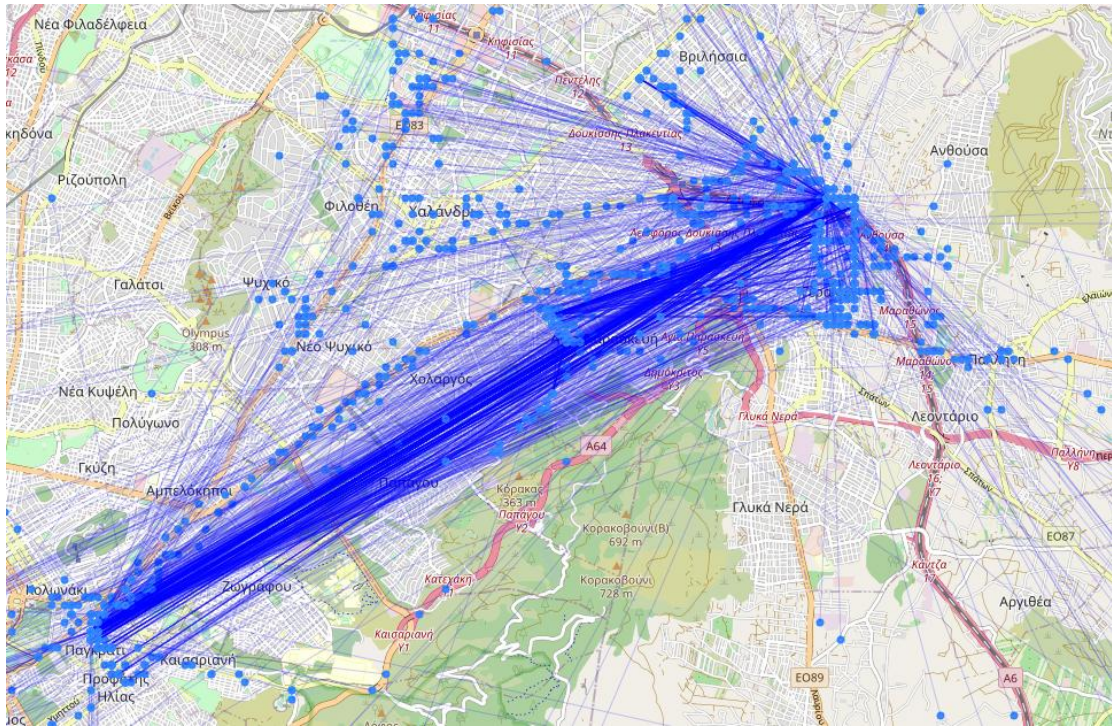
Προκύπτει πως υπάρχουν σημεία που ο χρήστης επισκέπτεται ως προορισμούς πιο συχνά. Με στόχο την εύρεση αυτών των σημείων διαμορφώνεται ένα Heat map (Διάγραμμα 13) όλων των τοποθεσιών (αφετηρίας και προορισμού) που έχει βρεθεί ο οδηγός. Με κέντρο αναφοράς τις πιο πολυσύχναστες τοποθεσίες φαίνεται ο παρακάτω χάρτης.



Διάγραμμα 13: Heat map όλων των τοποθεσιών που έχει βρεθεί ο οδηγός 1

Ουσιαστικά αυτό που φαίνεται σε αυτό το διάγραμμα είναι ότι όσο πιο έντονος χρωματισμός υπάρχει σε ένα σημείο τόσο πιο πολλές φορές ο χρήστης έχει επισκεφτεί αυτή τη τοποθεσία. Με αυτό τον τρόπο προκύπτει μια γενική εικόνα των βασικών τοποθεσιών του χρήστη σε ένα χάρτη.

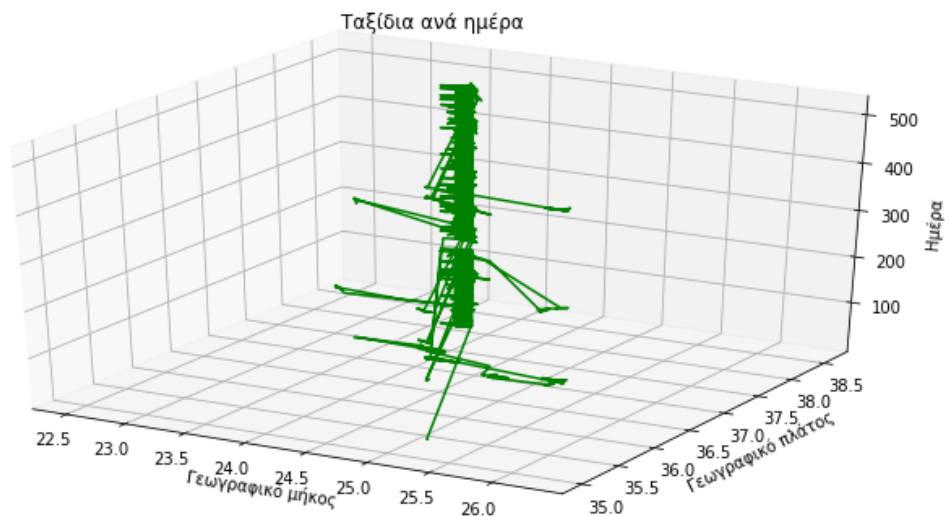
Προς επιβεβαίωση των παραπάνω εμφανών τοποθεσιών με περισσότερη συχνότητα δημιουργείται ένα γράφημα που παρουσιάζει όλες τις μετακινήσεις του οδηγού (αφετηρία-προορισμός) με κέντρο αναφοράς του χάρτη ίδιο με το προηγούμενο (Διάγραμμα 14).



Διάγραμμα 14: Απεικόνιση όλων των ταξιδιών του οδηγού 1

Ουσιαστικά οι μπλε κουκίδες αποτελούν τα σημεία άφιξης και αφετηρίας των ταξιδιών του χρήστη και όσο πιο έντονο χρώμα μπλε έχουν οι γραμμές τόσο πιο συχνά πραγματοποιείται ταξίδι ανάμεσα σε αυτές τις τοποθεσίες.

Τέλος με στόχο την εύρεση καθημερινών μοτίβων αλλά και την αναζήτηση επαναληπτικότητας αυτών των μοτίβων ανά τις ημέρες οπτικοποιούνται σε ένα 3D διάγραμμα όλες οι μετακινήσεις του οδηγού 1 με χρονολογική σειρά ανά ημέρα (Διάγραμμα 15).



Διάγραμμα 15: Απεικόνιση μοτίβου μετακινήσεων ανά ημέρα για τον οδηγό 1

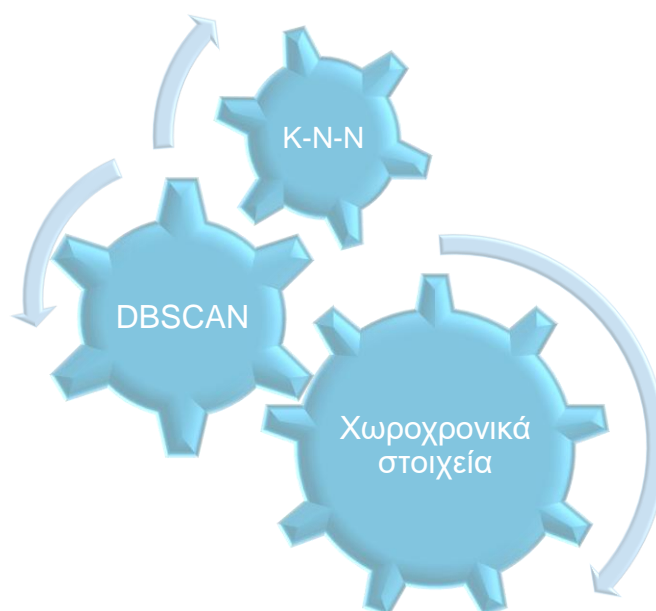
Είναι ευδιάκριτο πως υπάρχουν καθημερινά μοτίβα, τοποθεσίες με μεγαλύτερη συχνότητα και επαναλαμβανόμενα ταξίδια ανά τις ημέρες για τον συγκεκριμένο

οδηγό. Στόχος στην συνέχεια αποτελεί ο προσδιορισμός και η συγκεκριμενοποίηση αυτών.

Με το ίδιο σκεπτικό ακολουθούν τα παραπάνω 5 διαγράμματα και για τους υπόλοιπους 4 χαρακτηριστικούς οδηγούς στο παράρτημα Α.

4.4 Κατηγοριοποίηση των προορισμών ανά οδηγό

Μέσω συνεργασίας των αλγορίθμων ομαδοποίησης, μηχανικής μάθησης (DBSCAN και K-nearest- neighbors) επιτυγχάνεται η ομαδοποίηση των προορισμών του κάθε χρήστη ξεχωριστά με βάση τη πυκνότητα των σημείων. Εν συνεχεία και με βάση τα χωροχρονικά στοιχεία της κάθε ομαδοποίησης διακρίνονται οι βασικές τοποθεσίες του χρήστη ως **Σπίτι**, **Δουλειά**, **Προσωπικές Δραστηριότητες** όπως το γυμναστήριο, οι οικογενειακές υποχρεώσεις και άλλα σημεία που παρουσιάζουν σχετικά μεγάλη συχνότητα, **Υπόλοιπες Τοποθεσίες** όπως οι διακοπές σε διαφορετικά μέρη (Διάγραμμα 16).



Διάγραμμα 16: Μέθοδος κατηγοριοποίησης προορισμών

Πιο συγκεκριμένα, αρχικά μέσω του αλγόριθμου K-nearest-neighbors προκύπτει για κάθε οδηγό ξεχωριστά την ακτίνα του κύκλου η οποία έχει τιμή τέτοια ώστε το 5% των προορισμών που έχει βρεθεί ο χρήστης να εμπεριέχονται μέσα σε αυτό το κύκλο. Από αυτή τη τιμή της ακτίνας ενδιαφέρον έχει το 80% αυτής ώστε να τη χρησιμοποιηθεί ως παράμετρο eps στον αλγόριθμο DBSCAN. Επιπλέον η παράμετρος minpts του αλγόριθμου παίρνει τιμή ίση με το 5% των συνολικών προορισμών. Μέσω αυτής της μεθοδολογίας προσδιορίζεται στον αλγόριθμο ότι το 80% των σημείων κάθε ομαδοποίησης είναι core points (όπως αναλύθηκε στο θεωρητικό υπόβαθρο) θεωρώντας μια αρκετά ρεαλιστική παραδοχή για το σύνολο των οδηγών στο dataset. Τελικά ο αλγόριθμος εξάγει κάποιες ομαδοποιήσεις αυτόματα με βάση τη πυκνότητα των σημείων που πληρούν τις παραπάνω παραδοχές που ορίστηκαν. Αυτή η μεθοδολογία έχει στόχο την βέλτιστη εύρεση των ομαδοποιήσεων καθώς και την αυτοματοποιημένη διαδικασία κατηγοριοποίησης των

προορισμών για όλους τους οδηγούς. Στους παρακάτω πίνακες καταγράφονται οι παράμετροι και αντίστοιχες τιμές τους για τους αλγόριθμους που χρησιμοποιήθηκαν (Πίνακας 1 & 2)

Πίνακας 1: Παράμετρος εισόδου αλγορίθμου K-nearest-neighbors

Παράμετρος	Τιμή
K	5% του συνόλου των προορισμών του χρήστη

Πίνακας 2: Παράμετροι εισόδου αλγορίθμου DBSCAN

Παράμετρος	Τιμή
Eps	80% της ακτίνας που υπολογίζει ο αλγόριθμος K-nearest-neighbors
Minpts	5% του συνόλου των προορισμών του χρήστη

Για τον προσδιορισμό μιας ομαδοποίησης και τη κατηγοριοποίηση αυτών δεν αρκεί μόνο η χωρική ανάλυση που επιτυγχάνεται με το παραπάνω τρόπο, αλλά είναι αναγκαία η χωροχρονική ανάλυση. Συνεπώς για τον προσδιορισμό αυτό θεωρείται ως

- **Σπίτι** την ομαδοποίηση εκείνη η οποία έχει τις περισσότερες παρατηρήσεις ως προορισμός την ημέρα της Κυριακής.

Στη συνέχεια αφαιρώντας όλες τις τοποθεσίες με προορισμό το Σπίτι βρίσκεται η ώρα άφιξης (+- 1 ώρα) με τις περισσότερες παρατηρήσεις και προσδιορίζεται ως

- **Δουλειά** την ομαδοποίηση εκείνη με τις περισσότερες παρατηρήσεις της ώρας εκείνης.

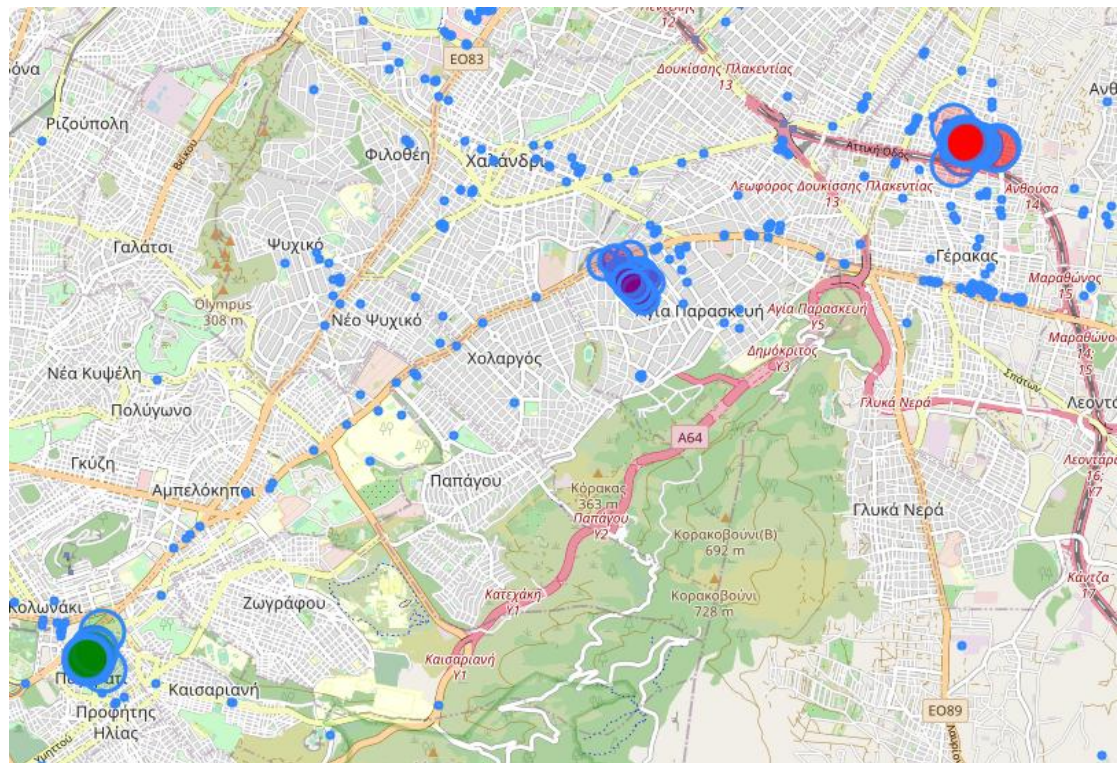
Σε περίπτωση που ο αλγόριθμος έχει βρει και άλλες ομαδοποιήσεις με τα προαναφερθέντα χωρικά χαρακτηριστικά αυτές προσδιορίζονται ως

- **Προσωπικές Δραστηριότητες**

Τέλος, όλες οι υπόλοιπες τοποθεσίες που δεν συγκαταλέγονται σε κάποια ομαδοποίηση (noise of DBSCAN) κατηγοριοποιούνται μαζί ως

- **Υπόλοιπες Τοποθεσίες**

Σε μορφή γραφήματος σε χάρτη και σε αλληλουχία (προς επικύρωση των αποτελεσμάτων) με το προηγούμενο κεφάλαιο της οπτικοποίησης χωροχρονικών δεδομένων φαίνονται οι ομαδοποιήσεις (κατηγοριοποιήσεις προορισμών) για τον οδηγό 1 στο Διάγραμμα 17. Στον παρακάτω χάρτη προσδιορίζονται με **κόκκινο** χρώμα οι τοποθεσίες που έχουν ως προορισμό το **Σπίτι**, με **πράσινο** χρώμα οι τοποθεσίες που έχουν ως προορισμό τη **Δουλειά**, με **μωβ** χρώμα οι τοποθεσίες που έχουν ως προορισμό τις **Προσωπικές Δραστηριότητες** και τέλος με **μπλε** χρώμα όλες οι **Υπόλοιπες Τοποθεσίες**.



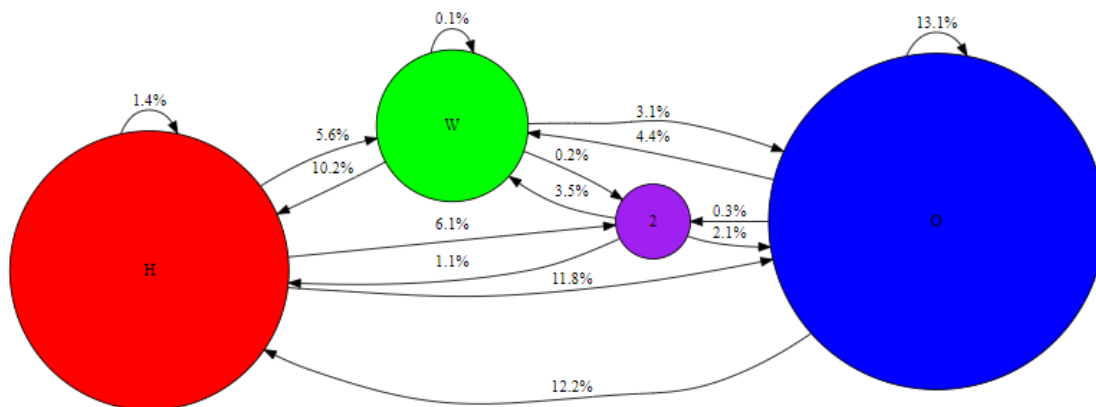
Διάγραμμα 17: Κατηγοριοποίηση προορισμών του οδηγού 1

Αυτή η διαδικασία επαναλαμβάνεται για όλους τους οδηγούς με περισσότερα από 200 ταξίδια. Στο παράρτημα Α παρουσιάζονται οι ομαδοποιήσεις για τους υπόλοιπους 4 χαρακτηριστικούς οδηγούς σε μορφή του παραπάνω γραφήματος.

4.5 Εύρεση προφίλ κινητικότητας ανά οδηγό

Κάθε χρήστης-οδηγός εκτελεί μια αλληλουχία προορισμών. Ο στόχος του παρόντος υποκεφαλαίου είναι η ποσοτικοποίηση των ταξιδιών που γίνονται ανάμεσα στους βασικούς προορισμούς για κάθε χρήστη ξεχωριστά. Αυτό επιτυγχάνεται από τη τοποθέτηση των ταξιδιών σε χρονολογική σειρά όπως έχει ήδη γίνει παραπάνω και την εύρεση των ποσοστών μεταξύ αυτών επί του συνόλου των ταξιδιών.

Στο Διάγραμμα 18 φαίνονται σε μορφή γραφήματος (network chart) αυτά τα ποσοστά για τον οδηγό 1 που αποτελεί και το προφίλ κινητικότητας του όπου με **κόκκινο** χρώμα προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό το **Σπίτι**, με **πράσινο** χρώμα προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό τη **Δουλειά**, με **μωβ** χρώμα προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό τις **Προσωπικές Δραστηριότητες** και τέλος με **μπλε** χρώμα προσδιορίζονται όλες οι **Υπόλοιπες Τοποθεσίες**.

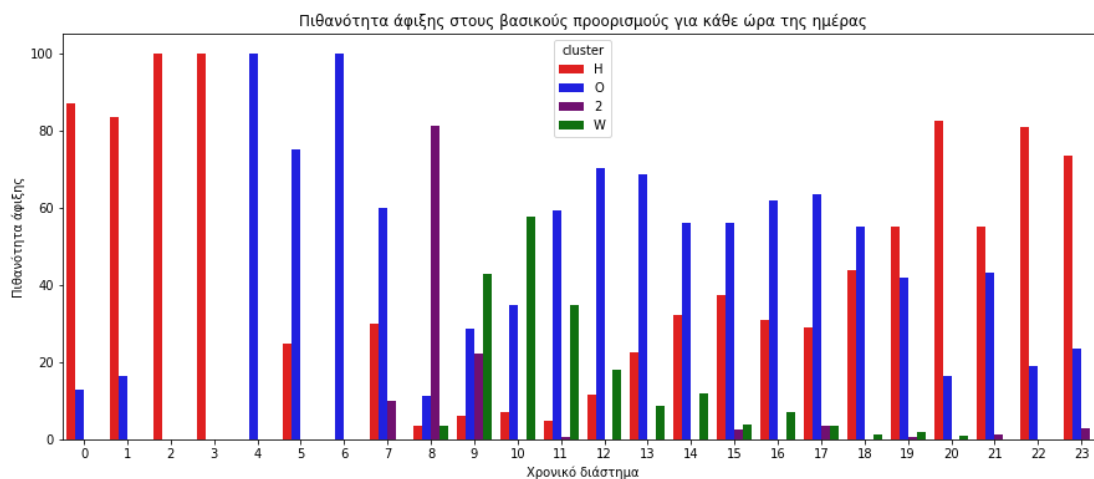


Διάγραμμα 18: Μοτίβο κινητικότητας του οδηγού 1

Ουσιαστικά αυτό το γράφημα αποτελεί το χωρικό προφίλ του οδηγού και κατά συνέπεια το προφίλ κινητικότητας του χρήστη ανάμεσα στους βασικούς προορισμούς. Στο παράρτημα Α παρουσιάζονται τα προφίλ κινητικότητας των υπόλοιπων 4 χαρακτηριστικών οδηγών.

4.6 Εύρεση χωροχρονικού προφίλ

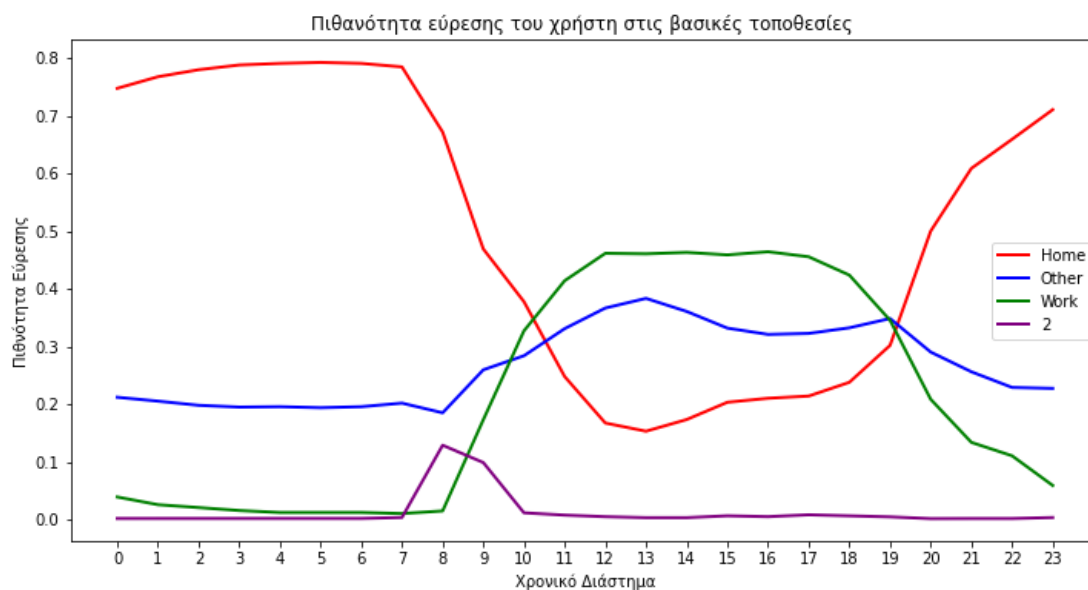
Με στόχο την εύρεση του χωροχρονικού προφίλ του οδηγού οπτικοποιείται (Διάγραμμα 19) η πιθανότητα άφιξης του στους βασικούς προορισμούς για κάθε ώρα της ημέρας όπου με κόκκινο χρώμα προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό το Σπίτι, με πράσινο χρώμα προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό τη Δουλειά, με μωβ χρώμα προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό τις Προσωπικές Δραστηριότητες και τέλος με μπλε χρώμα προσδιορίζονται όλες οι Υπόλοιπες Τοποθεσίες.



Διάγραμμα 19: Πιθανότητα άφιξης στους βασικούς προορισμούς για κάθε ώρα της ημέρας για τον οδηγό 1

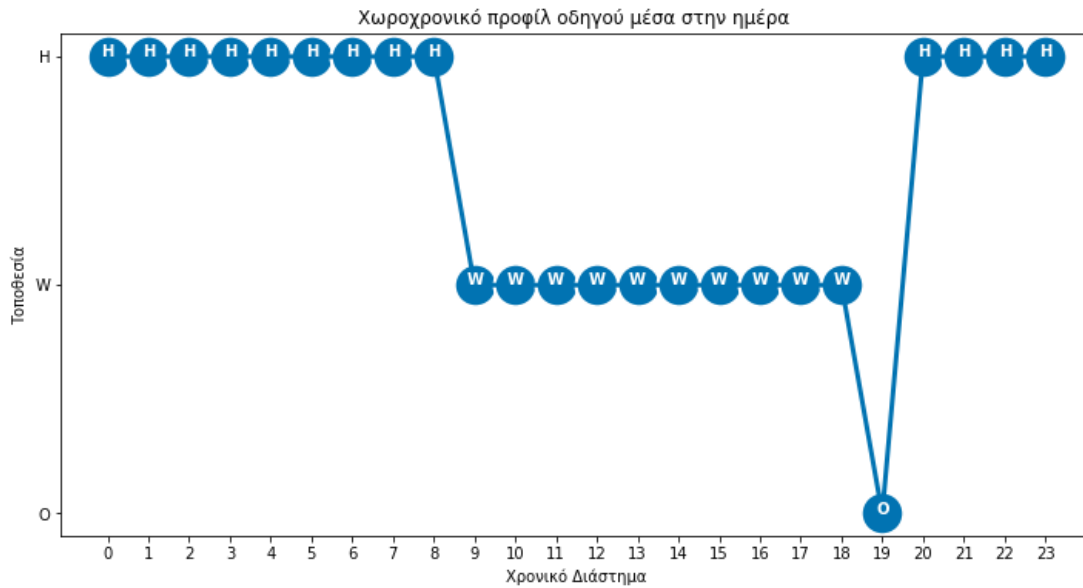
Εν συνεχεία της παραπάνω πληροφορίας για τον οδηγό και τις ώρες άφιξης του στους βασικούς προορισμούς είναι εφικτό με βάση την έναρξη του επόμενου ταξιδιού να προσδιοριστούν τη διάρκεια παραμονής σε κάθε ξεχωριστή δραστηριότητα που πραγματοποίησε ο χρήστης. Με αυτό το τρόπο προκύπτει μέσα στη διάρκεια της ημέρας για κάθε ξεχωριστή ώρα η συχνότητα εύρεσης του χρήστη σε αυτή τη τοποθεσία.

Συνεπώς είναι γνωστή για κάθε χρονικό διάστημα μέσα στην ημέρα με βάση το ιστορικό του χρήστη η πιθανότητα εύρεσης του σε κάθε τοποθεσία. Η παραπάνω πληροφορία αποτυπώνεται στο Διάγραμμα 20 για τον οδηγό 1.



Διάγραμμα 20: Πιθανότητα εύρεσης του οδηγού 1 στις βασικές τοποθεσίες για κάθε ώρα της ημέρας

Με βάση το παραπάνω διάγραμμα (δηλαδή τη μέγιστη πιθανότητα εύρεσης σε κάποια τοποθεσία για κάθε χρονικό διάστημα) και θεωρώντας δεδομένη την παρουσία του χρήστη στη Δουλειά για χρόνο παραμονής ίσο με τη πιο συχνή παρατήρηση της συγκεκριμένης ομαδοποίησης, στο Διάγραμμα 21 παρουσιάζεται το χωροχρονικό προφίλ (χαρακτηριστική αλυσίδα δραστηριοτήτων επί του συνόλου των ημερών) του οδηγού 1. Η παραδοχή αυτή γίνεται γιατί οι Υπόλοιπες Τοποθεσίες αποτελούν πιο συχνή παρατήρηση ως προορισμός συνήθως από τη Δουλειά καθώς έχουν ομαδοποιηθεί στο σύνολο τους ενώ αποτελούν διαφορετικά μέρη μεταξύ τους αλλά και στο σύνολο των ημερών, και με αυτό τον τρόπο διαστρεβλώνεται η πραγματικότητα αν θεωρηθεί ως χωροχρονικό προφίλ η μέγιστη πιθανότητα για κάθε ώρα ξεχωριστά. Στην ενασχόληση για κάθε μέρα της εβδομάδας ξεχωριστά αυτή η παραδοχή δεν λαμβάνεται υπόψη. Το παρακάτω χωροχρονικό προφίλ είναι αυτό με τη μεγαλύτερη πιθανότητα πραγματοποίησης για όλες τις ημέρες της εβδομάδας.



Προς ολοκλήρωση της εξατομικευμένης ανάλυσης στο Πίνακα 3 παρατίθενται στατιστικά που αφορούν την ώρα άφιξης στον προορισμό και τη διάρκεια παραμονής για κάθε χρήστη και τοποθεσία ξεχωριστά, **Hour**: Ώρα άφιξης, **Duration**: Διάρκεια παραμονής, **Count**: Σύνολο παρατηρήσεων, **Mean**: Μέση τιμή, **Std**: Τυπική απόκλιση των τιμών, **Min**: Ελάχιστη τιμή παρατηρήσεων, **25%**: Πρώτο τεταρτημόριο, **50%**: Διάμεσος, **75%**: Τρίτο Τεταρτημόριο, **Max**: Μέγιστη τιμή παρατηρήσεων

Σπίτι

	hour	duration
count	586.000000	586.000000
mean	16.854949	11.737201
std	5.646067	7.989901
min	0.000000	0.000000
25%	15.000000	7.250000
50%	19.000000	12.000000
75%	20.000000	15.000000
max	23.000000	49.000000

Δουλειά

	hour	duration
count	320.000000	320.000000
mean	10.375000	9.259375
std	1.972388	3.522196
min	8.000000	0.000000
25%	9.000000	8.000000
50%	10.000000	9.000000
75%	11.000000	11.000000
max	20.000000	27.000000

Προσωπική Δραστηριότητα '2'

	hour	duration
count	156.000000	156.000000
mean	8.858974	0.365385
std	2.171513	1.665926
min	7.000000	0.000000
25%	8.000000	0.000000
50%	8.000000	0.000000
75%	9.000000	0.000000
max	23.000000	12.000000

Υπόλοιπες Τοποθεσίες

	hour	duration
count	706.000000	706.000000
mean	13.903683	4.732295
std	4.272201	19.225958
min	0.000000	0.000000
25%	11.000000	0.000000
50%	13.000000	2.000000
75%	18.000000	4.000000
max	23.000000	408.000000

Πίνακας 3: Στατιστικά ώρας και διάρκειας των βασικών δραστηριοτήτων του οδηγού 1

Η παραπάνω μεθοδολογία κατηγοριοποίησης και χωροχρονικής ανάλυσης επαναλαμβάνεται για όλους τους οδηγούς με περισσότερα από 200 ταξίδια. Τα σχετικά γραφήματα για τους υπόλοιπους 4 χαρακτηριστικούς οδηγούς βρίσκονται στο παράρτημα Α. Οι πίνακες με τα στατιστικά της ώρας άφιξης και της διάρκειας δραστηριοτήτων των υπόλοιπων 4 οδηγών παρουσιάζονται στο παράρτημα Β.

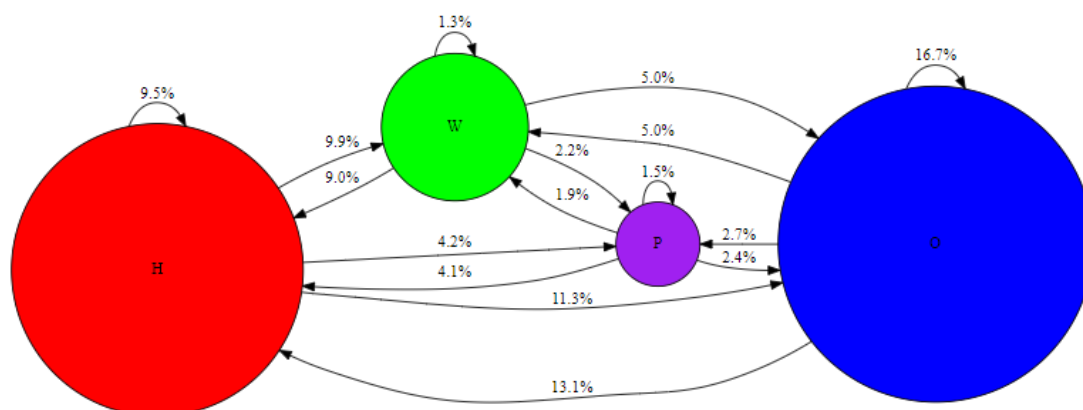
4.7 Κατηγοριοποίηση συνόλου των προορισμών

Σκοπός αυτού του υποκεφαλαίου είναι η εξαγωγή γενικών συμπερασμάτων όσον αφορά τις μετακινήσεις που γίνονται μεταξύ των βασικών προορισμών για κάθε ημέρα της εβδομάδας.

Για να επιτευχθεί αυτό η παραπάνω μεθοδολογία που παρουσιάστηκε στα υποκεφάλαια 3, 4, 5 εφαρμόζεται για το σύνολο των οδηγών με περισσότερα από 200 ταξίδια σε καθημερινό επίπεδο. Τα συμπεράσματα που επιδιώκονται να εξαχθούν αφορούν τις μετακινήσεις ανάμεσα στις βασικές τοποθεσίες και τη ποσοτικοποίηση που υπάρχει στα ταξίδια ανάμεσα σε αυτές ξεχωριστά για κάθε μέρα της εβδομάδας. Εκ τούτου διαχωρίζονται οι προορισμοί των ταξιδιών που καταγράφηκαν για κάθε χρήστη σε μέρες της εβδομάδας. Επιπλέον, καθώς οι Προσωπικές Δραστηριότητες είναι ξεχωριστές και συνήθως περισσότερες από μια για κάθε χρήστη ομαδοποιούνται σαν μία δραστηριότητα (με ονομασία 'P') για όλους τους οδηγούς ώστε να καθίσταται δυνατό να εξαχθούν τα άνωθεν συμπεράσματα.

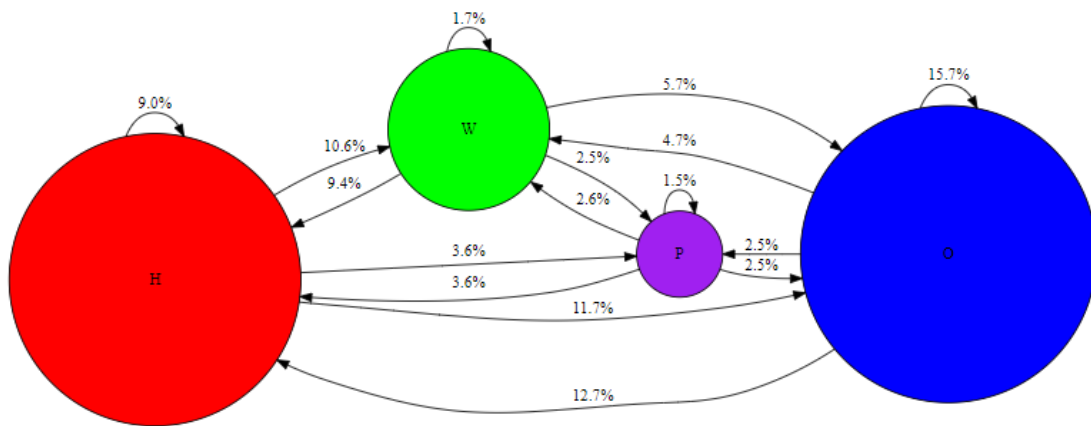
Ακολουθώντας την ίδια διαδικασία σε επίπεδο ημέρας για όλους τους οδηγούς εξάγεται για κάθε μέρα της εβδομάδας ξεχωριστά το χωρικό της προφίλ με βάση τους οδηγούς με περισσότερα από 200 ταξίδια και τα Διαγράμματα 22 έως 28 μας δίνουν αυτή τη πληροφορία όπου με **κόκκινο** χρώμα προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό το **Σπίτι**, με **πράσινο** χρώμα προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό τη **Δουλειά**, με **μωβ** χρώμα προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό τις **Προσωπικές Δραστηριότητες** και τέλος με **μπλε** χρώμα προσδιορίζονται όλες οι **Υπόλοιπες Τοποθεσίες**.

Δευτέρα



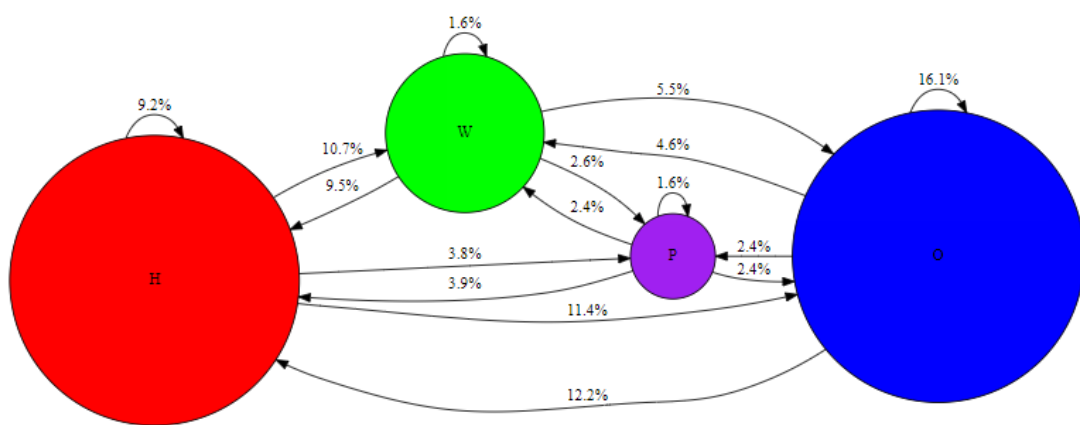
Διάγραμμα 22: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών τη Δευτέρα

Τρίτη



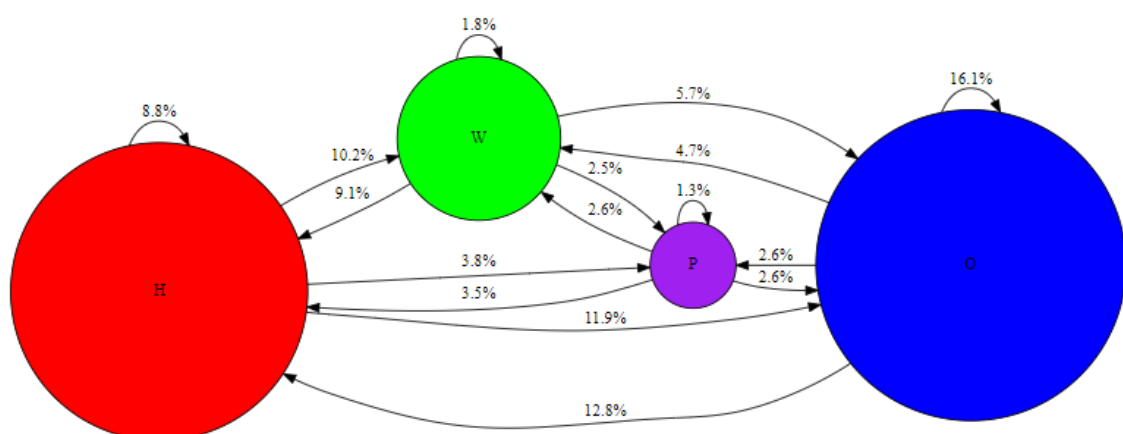
Διάγραμμα 23: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών την Τρίτη

Τετάρτη



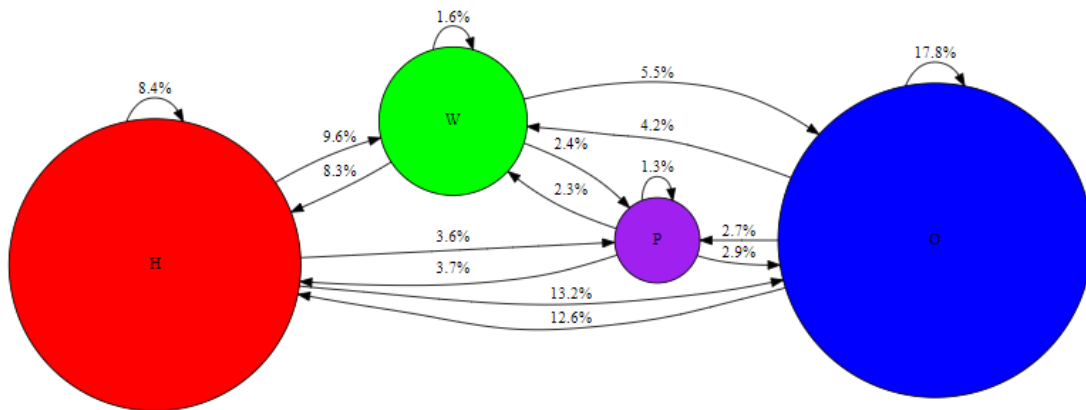
Διάγραμμα 24: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών την Τετάρτη

Πέμπτη



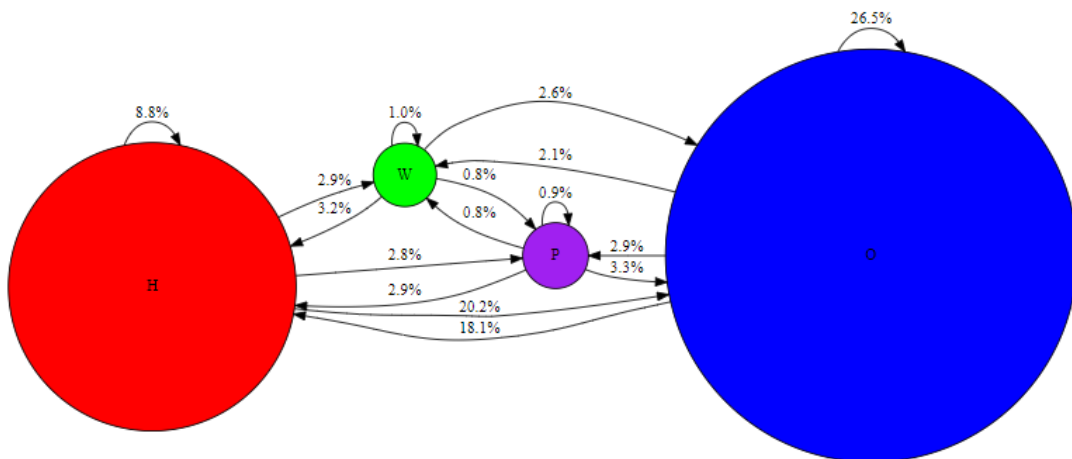
Διάγραμμα 25: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών την Πέμπτη

Παρασκευή



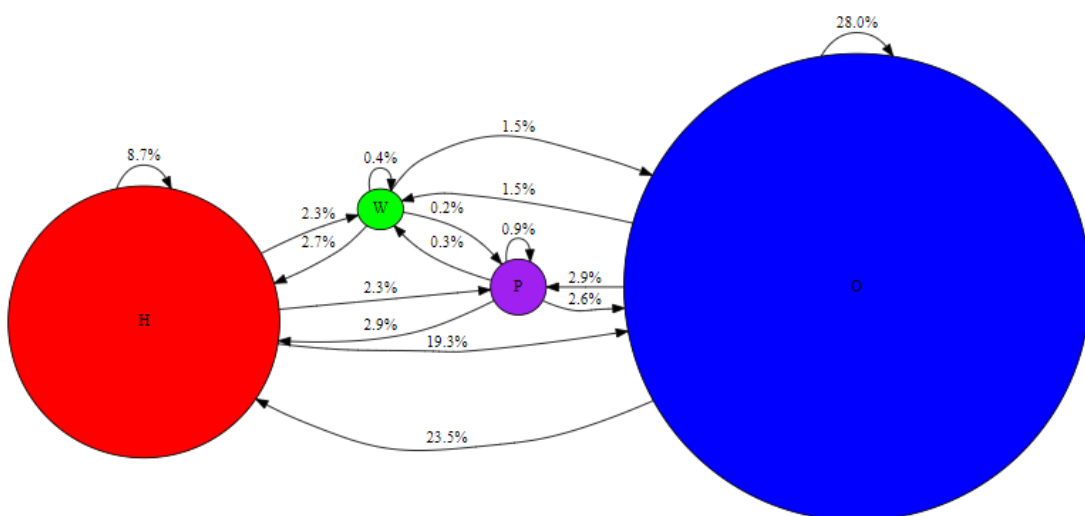
Διάγραμμα 26: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών την Παρασκευή

Σάββατο



Διάγραμμα 27: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών το Σάββατο

Κυριακή



Διάγραμμα 28: Ποσοτικοποίηση των ταξιδιών μεταξύ των βασικών προορισμών την Κυριακή

Είναι προφανής από τα παραπάνω γραφήματα η ομοιότητα που παρουσιάζουν οι μέρες από τη Δευτέρα έως τη Παρασκευή με κριτήριο τη ποσοτικοποίηση των μετακινήσεων μεταξύ των βασικών προορισμών (επί του συνόλου των μοτίβων κινητικότητας των χρηστών).

Πιο συγκεκριμένα διαφαίνεται έντονη ομοιότητα στα ποσοστά των ταξιδιών από και προς τη Δουλειά για τις μέρες Δευτέρα ως Παρασκευή.

Επίσης φαίνεται πως η Κυριακή αποτελεί τη μέρα με τη μεγαλύτερη αβεβαιότητα κατάληξης του οδηγού στους βασικούς προορισμούς αφού είναι αρκετά μεγάλο το ποσοστό ύπαρξης των Υπόλοιπων Τοποθεσιών.

4.8 Καθημερινή αλυσίδα δραστηριοτήτων των χρηστών και τα στατιστικά τους

Στόχος της παρούσας μεθοδολογίας αποτελεί η εύρεση της **τυπικής αλυσίδας δραστηριοτήτων** για κάθε μέρα της εβδομάδας και η εξαγωγή στατιστικών και συμπερασμάτων. Αυτό επιτυγχάνεται εφαρμόζοντας την ίδια μεθοδολογία με το υποκεφάλαιο 4.6, αυτή τη φορά όμως για κάθε ημέρα της εβδομάδας ξεχωριστά. Η παραδοχή που γίνεται στη παρούσα μεθοδολογία είναι η θεώρηση ότι μια δραστηριότητα δεν μπορεί να διαρκέσει πάνω από 96 ώρες. Επιπλέον παύει να υφίσταται η παραδοχή της δεδομένης παρουσίας του οδηγού στη Δουλειά, ενώ πλέον η τυπική καθημερινή αλυσίδα δραστηριοτήτων του χρήστη διαμορφώνεται για κάθε χρονικό διάστημα μέσω της τοποθεσίας με τις περισσότερες παρατηρήσεις παραμονής σε αυτή.

Εφαρμόζοντας την παραπάνω μεθοδολογία προκύπτουν για όλους τους χρήστες-οδηγούς που πραγματοποίησαν περισσότερα από 200 ταξίδια για κάθε μέρα της εβδομάδας την τυπική χωροχρονική αλυσίδα ταξιδιών που πραγματοποιεί κάθε χρήστης και προκύπτουν τα παρακάτω στατιστικά συμπεράσματα για τις Καθημερινές και τα Σαββατοκύριακα αντίστοιχα (Πίνακας 4 & 5).

Πίνακας 4: Ποσοστά των 5 συνηθέστερων αλυσίδων δραστηριότητας τις Καθημερινές

Δευτέρα		Τρίτη		Τετάρτη		Πέμπτη		Παρασκευή	
H-W- H	26.24%	H- W-H	25.48%	H-W- H	27.38%	H-W- H	23.57%	H-W-H	22.43%
H	15.97%	H	14.83%	H	14.45%	H	15.21%	H	12.93%
H-O-H	7.6%	H-O- H	11.41%	H-O-H	9.51%	H-O-H	10.27%	H-O-H	8.75%
H-W- O-H	6.84%	H- W- O-H	7.6%	H-W- O-H	7.6%	H-W- O-H	7.22%	H-W-O-H	7.98%
H-O- H-O-H	3.42%	H-O- H-O- H	3.42%	H-W- H-O-H	4.18%	H-O- H-O-H	4.94%	H-W-H-O- H	4.56%

Με H και χρώμα **κόκκινο** συμβολίζεται η τοποθεσία **Σπίτι**.

Με W και χρώμα **πράσινο** συμβολίζεται η τοποθεσία **Δουλειά**.

Με O και χρώμα **μπλε** συμβολίζονται οι **Υπόλοιπες Τοποθεσίες**.

Πίνακας 5: Ποσοστά των 5 συνηθέστερων αλυσίδων δραστηριότητας τα Σαββατοκύριακα

Σάββατο		Κυριακή	
H-W- H	22.43%	H	20.53%
H	12.93%	H-O- H	17.49%
H-O-H	8.75%	H-O	8.37%
H-W- O-H	7.98%	O-H	7.22%
H-W- H-O-H	4.56%	H-O- H-O- H	6.84%

Με H και χρώμα **κόκκινο** συμβολίζεται η τοποθεσία **Σπίτι**.
 Με W και χρώμα **πράσινο** συμβολίζεται η τοποθεσία **Δουλειά**.
 Με O και χρώμα **μπλε** συμβολίζονται οι **Υπόλοιπες Τοποθεσίες**.

Μέσω των παραπάνω πινάκων φαίνεται η επαναληπτικότητα και η ομοιότητα που παρουσιάζουν οι αλυσίδες ταξιδιών ανάμεσα στις μέρες από Δευτέρα έως Παρασκευή. Επίσης μέσω των παραπάνω ποσοστών διαφαίνεται η κανονικότητα που παρουσιάζουν οι ημέρες της εβδομάδας με βάση τις αλυσίδες ταξιδιών.

Επίσης ενδιαφέρον παρουσιάζει το ποσοστό των οδηγών που παρόλο που έχουν διαφορετική τυπική χωροχρονική αλυσίδα δραστηριοτήτων από την μέγιστη παρατήρηση των αντίστοιχων ημερών, η μέγιστη παρατήρηση εμπεριέχεται στην ευρύτερη προσωπική τους τυπική αλυσίδα (πχ. το H-W-H εμπεριέχεται στο H-W-H-O-H). Για την ημέρα της Κυριακής η ενασχόληση μας έγκειται στη δεύτερη μέγιστη παρατήρηση.

Από τους 263 οδηγούς οι :

- 100 οδηγοί εμφανίζουν την αλυσίδα **H-W-H** την Δευτέρα
- 101 οδηγοί εμφανίζουν την αλυσίδα **H-W-H** την Τρίτη
- 109 οδηγοί εμφανίζουν την αλυσίδα **H-W-H** την Τετάρτη
- 103 οδηγοί εμφανίζουν την αλυσίδα **H-W-H** την Πέμπτη
- 103 οδηγοί εμφανίζουν την αλυσίδα **H-W-H** την Παρασκευή
- 151 οδηγοί εμφανίζουν την αλυσίδα **H-O-H** το Σάββατο
- 137 οδηγοί εμφανίζουν την αλυσίδα **H-O-H** την Κυριακή

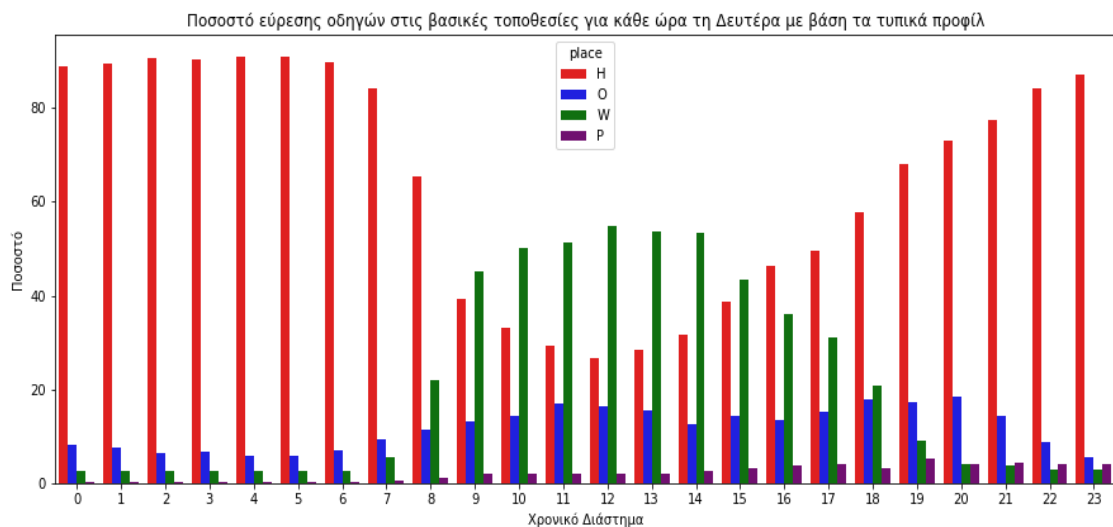
Επίσης ένα στατιστικό στοιχείο που παρουσιάζει ενδιαφέρον είναι το σύνολο των διαφορετικών τυπικών χωροχρονικών αλυσίδων δραστηριότητας που παρατηρήθηκαν ξεχωριστά για κάθε ημέρα. Πιο συγκεκριμένα από 263 οδηγούς την ημέρα:

- **Δευτέρα** παρατηρήθηκαν 68 διαφορετικές αλυσίδες ταξιδιών
- **Τρίτη** παρατηρήθηκαν 64 διαφορετικές αλυσίδες ταξιδιών
- **Τετάρτη** παρατηρήθηκαν 61 διαφορετικές αλυσίδες ταξιδιών
- **Πέμπτη** παρατηρήθηκαν 61 διαφορετικές αλυσίδες ταξιδιών
- **Παρασκευή** παρατηρήθηκαν 70 διαφορετικές αλυσίδες ταξιδιών
- **Σάββατο** παρατηρήθηκαν 58 διαφορετικές αλυσίδες ταξιδιών

- **Κυριακή** παρατηρήθηκαν 44 διαφορετικές αλυσίδες ταξιδιών

Τέλος στο παράρτημα Γ παρατίθενται συγκεντρωτικά διαγράμματα (bar plots) για κάθε μέρα της εβδομάδας ξεχωριστά, που υποδεικνύουν σε μορφή ποσοστού για κάθε ώρα της ημέρας τη τυπική χωρική εύρεση του συνόλου των χρηστών.

Ενδεικτικά στο Διάγραμμα 29 φαίνεται αυτό της Δευτέρας όπου με **κόκκινο** χρώμα και H προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό το **Σπίτι**, με **πράσινο** χρώμα και W προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό τη **Δουλειά**, με **μωβ** χρώμα και P προσδιορίζονται οι τοποθεσίες που έχουν ως προορισμό τις **Προσωπικές Δραστηριότητες** και με **μπλε** χρώμα και O προσδιορίζονται όλες οι **Υπόλοιπες Τοποθεσίες**.



Διάγραμμα 29: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Δευτέρας με βάση τα τυπικά προφίλ των οδηγών

Μέσω των παραπάνω πινάκων, στατιστικών στοιχείων και των διαγραμμάτων που παρατίθενται στο παρόν υποκεφάλαιο και στο παράρτημα Γ φαίνεται η επαναληπτικότητα και η ομοιότητα που παρουσιάζουν οι αλυσίδες δραστηριοτήτων ανάμεσα στις μέρες από Δευτέρα έως Παρασκευή. Επίσης μέσω των παραπάνω ποσοστών και των διαγραμμάτων στο παράρτημα Γ διαφαίνεται η κανονικότητα που παρουσιάζουν οι ημέρες της εβδομάδας με βάση τις αλυσίδες ταξιδιών. Οι παραπάνω πληροφορίες είναι πολύ σημαντικές για την αναγνώριση των καθημερινών μοτίβων των χρηστών.

4.9 Προτυποποίηση διάρκειών των δραστηριοτήτων

Μετά τη καθημερινή κατηγοριοποίηση και ποσοτικοποίηση των βασικών προορισμών καθώς και τη καθημερινή χωροχρονική ανάλυση και εξαγωγή στατιστικών συμπερασμάτων εξ αυτών, σειρά στη μεθοδολογία που ακολουθείται έχει η προτυποποίηση των διάρκειών των βασικών προορισμών.

Στόχος είναι η εύρεση στατιστικών συμπερασμάτων για τις διάρκειες των δραστηριοτήτων γιατί σε συνδυασμό με τις παραπάνω πληροφορίες δημιουργείται

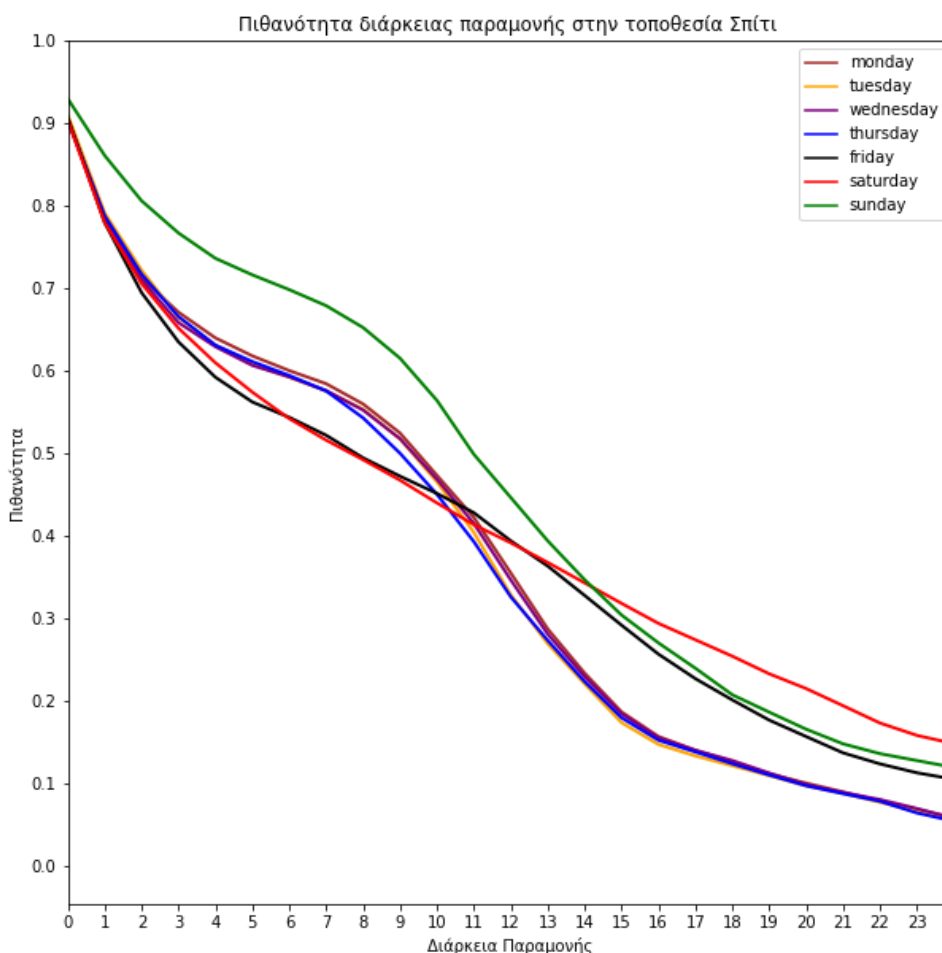
ένα σύνολο πληροφοριών με μεγάλη δυναμική και ζωτικής σημασίας καθώς δύναται να βελτιστοποιεί την πρόβλεψη των καθημερινών αλυσίδων ταξιδιών όλων των χρηστών.

Για να επιτευχθεί η παραπάνω μεθοδολογία χρησιμοποιείται η στατιστική μέθοδος της Ανάλυσης Επιβίωσης (Survival Analysis) η οποία έχει αναλυθεί στο υποκεφάλαιο του θεωρητικού υποβάθρου. Η ανάλυση αφορά τους βασικούς προορισμούς που ορίστηκαν για κάθε ημέρα της εβδομάδας ξεχωριστά.

Πρώτα κατηγοριοποιούνται οι διάρκειες των δραστηριοτήτων ανά τύπο και ημέρα της εβδομάδας. Οι βασικοί προορισμοί είναι το Σπίτι, η Δουλειά, οι Προσωπικές Δραστηριότητες και οι Υπόλοιπες Τοποθεσίες ενοποιημένες πλέον από όλους τους χρήστες.

4.9.1 Σπίτι

Αρχικά αναλύεται η τοποθεσία Σπίτι και δημιουργείται η Kaplan Meier καμπύλη Επιβίωσης (Διάγραμμα 30) για κάθε μέρα της εβδομάδας όσον αφορά τις διάρκειες.

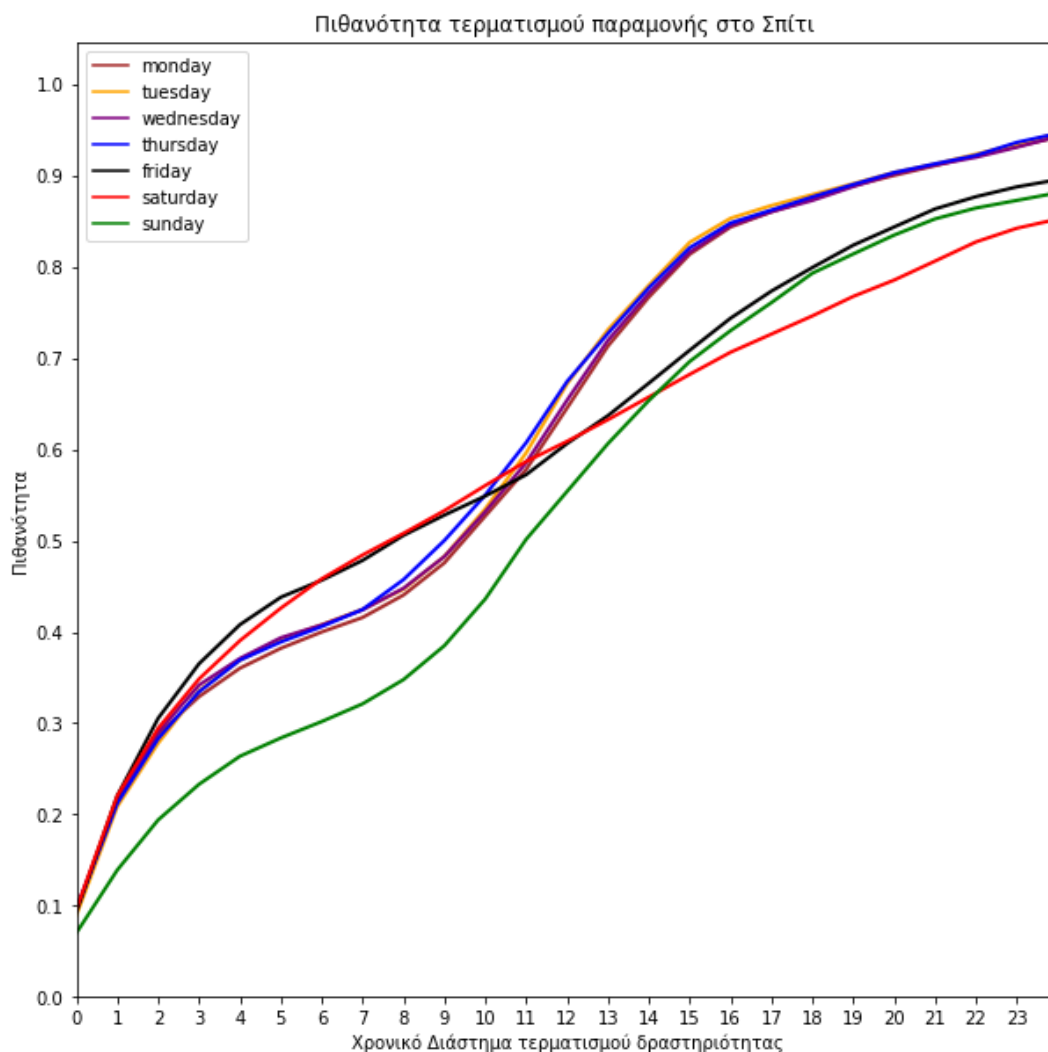


Διάγραμμα 30: Πιθανότητα διάρκειας παραμονής στη τοποθεσία Σπίτι για κάθε μέρα της εβδομάδας

Το Διάγραμμα 30 δείχνει τη πιθανότητα διάρκειας παραμονής οποιουδήποτε χρήστη στη τοποθεσία Σπίτι για κάθε μέρα της εβδομάδας ξεχωριστά. Στον άξονα x του διαγράμματος οι δείκτες συμβολίζουν διαστήματα (πχ το 0 = 0 λεπτά έως 59 λεπτά

κλπ). Για λόγους ευκρίνειας του διαγράμματος παρουσιάζονται διάρκειες παραμονής έως 24 ωρών. Παρατηρείται ότι οι μέρες Δευτέρα έως Πέμπτη παρουσιάζουν χαρακτηριστική ομοιότητα στη πιθανότητα διάρκειας παραμονής στη τοποθεσία Σπίτι και αυτό αποτελεί μια σημαντική πληροφορία στην εν δυνάμει μείωση του υπολογιστικού κόστους πρόβλεψης της διάρκειας παραμονής για τη τοποθεσία Σπίτι.

Σε αλληλουχία με το παραπάνω διάγραμμα χρησιμοποιώντας τη Συνάρτηση Αθροιστικής Πυκνότητας (Cumulative Distribution Function) κατασκευάζεται το παρακάτω διάγραμμα που ορίζει τη πιθανότητα τερματισμού της παραμονής στη τοποθεσία Σπίτι.



Διάγραμμα 31: Πιθανότητα τερματισμού παραμονής στη τοποθεσία Σπίτι για κάθε μέρα της εβδομάδας

Όμοια με το Διάγραμμα 30 και στο Διάγραμμα 31 φαίνεται η ομοιότητα των ημερών Δευτέρα έως Πέμπτη στη πιθανότητα τερματισμού της δραστηριότητας Σπίτι.

Προς επιβεβαίωση των ανώτερων ενδείξεων περί ομοιότητας στις διάρκειες της δραστηριότητας με προορισμό το Σπίτι εκτελούμε τη δοκιμή κατάταξης (Log-rank Test) που αναλύθηκε στο θεωρητικό υπόβαθρο. Ο Πίνακας 6 δείχνει μεταξύ όλων των ημερών την τιμή του p-value. Η τιμή αυτή κυμαίνεται από 0-1 και όσο

μεγαλύτερη είναι αυτή τόσο μικρότερη στατιστική διαφορά υπάρχει ανάμεσα στις μέρες που εξετάζονται. Τιμή μικρότερη του 0.05 υποδεικνύει σημαντική στατιστική διαφορά μεταξύ των δειγμάτων.

P-value	Δευτέρα	Τρίτη	Τετάρτη	Πέμπτη	Παρασκευή	Σάββατο	Κυριακή
Δευτέρα	-	0.15	0.60	0.13	<0.005	<0.005	<0.005
Τρίτη	-	-	0.37	0.89	<0.005	<0.005	<0.005
Τετάρτη	-	-	-	0.32	<0.005	<0.005	<0.005
Πέμπτη	-	-	-	-	<0.005	<0.005	<0.005
Παρασκευή	-	-	-	-	-	0.02	<0.005
Σάββατο	-	-	-	-	-	-	0.01
Κυριακή	-	-	-	-	-	-	-

Πίνακας 6: Σύγκριση ημερών σε σχέση με τη διάρκεια παραμονής στην τοποθεσία Σπίτι

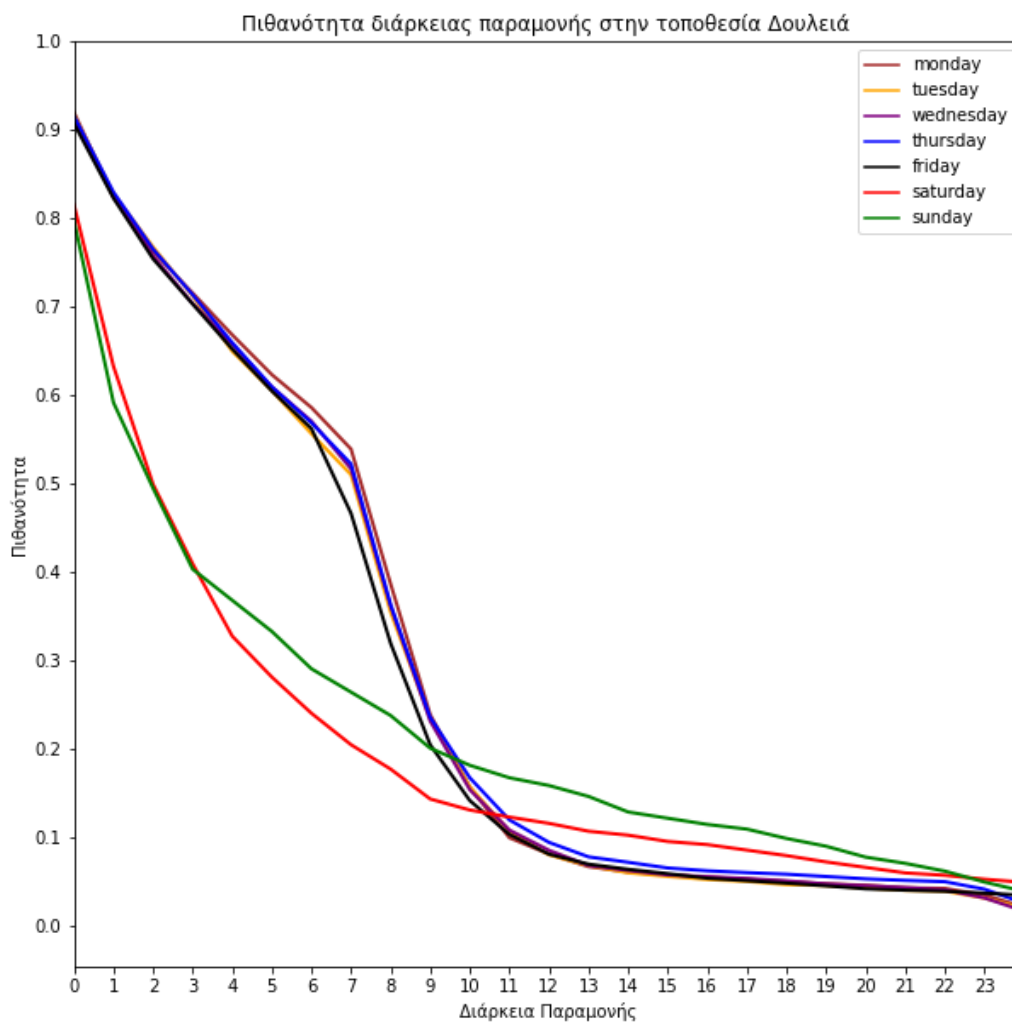
Με κόκκινο σημειώνονται οι τιμές που υποδεικνύουν στατιστική διαφορά.

Τέλος για την τοποθεσία **Σπίτι** οι διάμεσοι της διάρκειας παραμονής για κάθε μέρα ξεχωριστά είναι οι εξής :

- **Δευτέρα** : 10 ώρες
- **Τρίτη** : 10 ώρες
- **Τετάρτη** : 10 ώρες
- **Πέμπτη** : 9 ώρες
- **Παρασκευή** : 8 ώρες
- **Σάββατο** : 8 ώρες
- **Κυριακή** : 11 ώρες

4.9.2 Δουλειά

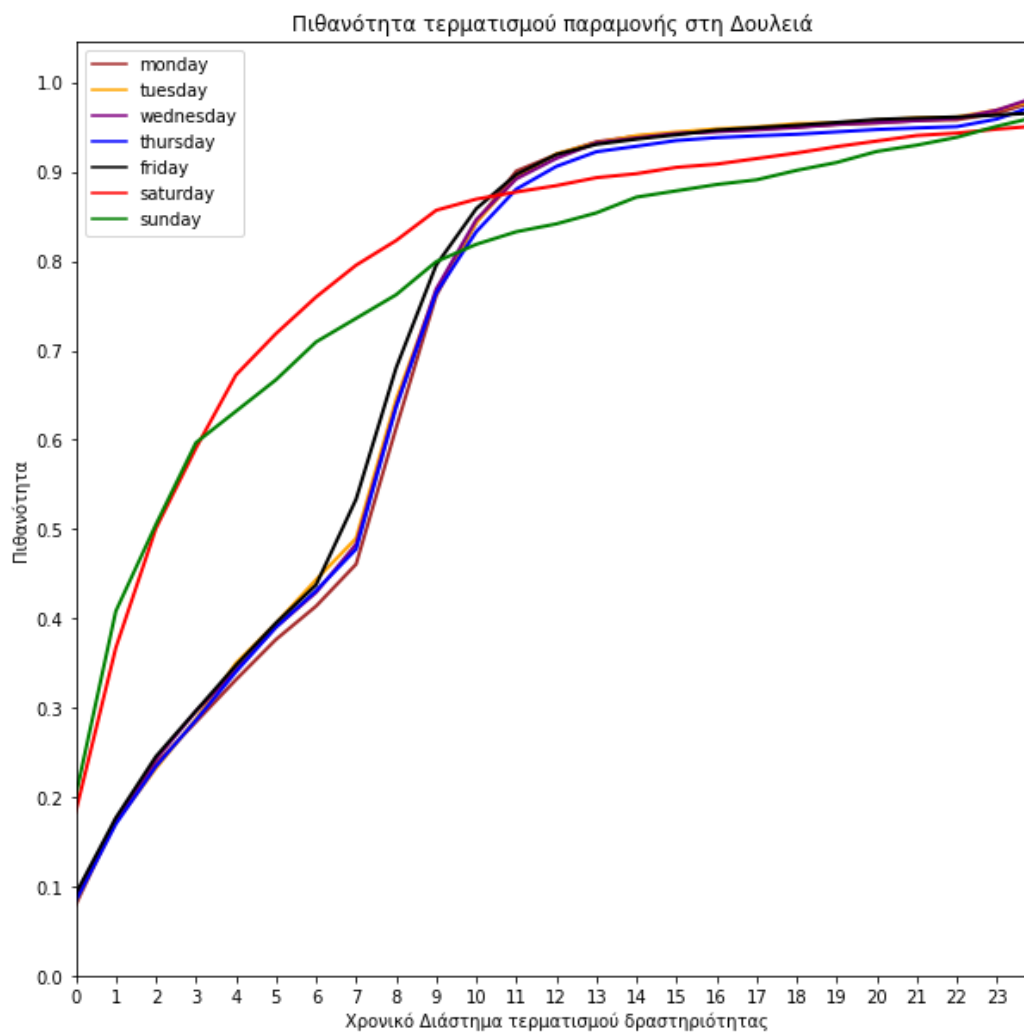
Ακολουθώντας την ίδια διαδικασία με τη τοποθεσία Σπίτι στο Διάγραμμα 32 φαίνεται η καμπύλη Επιβίωσης (Kaplan Meier estimator) για κάθε ημέρα της εβδομάδας για τη τοποθεσία της Δουλειάς.



Διάγραμμα 32: Πιθανότητα διάρκειας παραμονής στην τοποθεσία Δουλειά για κάθε μέρα της εβδομάδας

Σε αντιστοιχία με τις διευκρινίσεις που δόθηκαν στο προηγούμενο υποκεφάλαιο 4.9.1 το Διάγραμμα 32 δείχνει τη πιθανότητα διάρκειας παραμονής οποιουδήποτε χρήστη στη τοποθεσία **Δουλειά** για κάθε μέρα της εβδομάδας ξεχωριστά. Παρατηρείται ότι οι μέρες Δευτέρα έως Πέμπτη παρουσιάζουν χαρακτηριστική ομοιότητα στη πιθανότητα διάρκειας παραμονής για τη τοποθεσία Δουλειά.

Σε αλληλουχία με το Διάγραμμα 32 χρησιμοποιώντας τη Συνάρτηση Αθροιστικής Πυκνότητας (Cumulative Distribution Function) κατασκευάζεται το Διάγραμμα 33 που ορίζει τη πιθανότητα τερματισμού της παραμονής στη τοποθεσία **Δουλειά**.



Διάγραμμα 33: Πιθανότητα τερματισμού παραμονής στη τοποθεσία Δουλειά για κάθε μέρα της εβδομάδας

Όμοια με το διάγραμμα 32 και σε αυτό φαίνεται να υπάρχει ομοιότητα ανάμεσα στις μέρες στη πιθανότητα τερματισμού της δραστηριότητας Δουλειά.

Προς επιβεβαίωση των ανώτερων ενδείξεων περί ομοιότητας και επαναληπτικότητας στις διάρκειες της δραστηριότητας με προορισμό τη Δουλειά εκτελούμε τη δοκιμή κατάταξης (Log-rank Test) που αναλύθηκε στο θεωρητικό υπόβαθρο (Πίνακας 7).

P-value	Δευτέρα	Τρίτη	Τετάρτη	Πέμπτη	Παρασκευή	Σάββατο	Κυριακή
Δευτέρα	-	0.28	0.21	0.66	0.01	<0.005	<0.005
Τρίτη	-	-	0.92	0.12	0.18	<0.005	<0.005
Τετάρτη	-	-	-	0.09	0.24	<0.005	<0.005
Πέμπτη	-	-	-	-	0.01	<0.005	<0.005
Παρασκευή	-	-	-	-	-	<0.005	<0.005
Σάββατο	-	-	-	-	-	-	0.23
Κυριακή	-	-	-	-	-	-	-

Πίνακας 7: Σύγκριση ημερών σε σχέση με τη διάρκεια παραμονής στη τοποθεσία Δουλειά

Εν τέλει από το Πίνακα 7 φαίνεται πως οι μέρες Δευτέρα έως Πέμπτη δεν παρουσιάζουν σημαντική στατιστική διαφορά μεταξύ τους. Επίσης μεταξύ τους οι

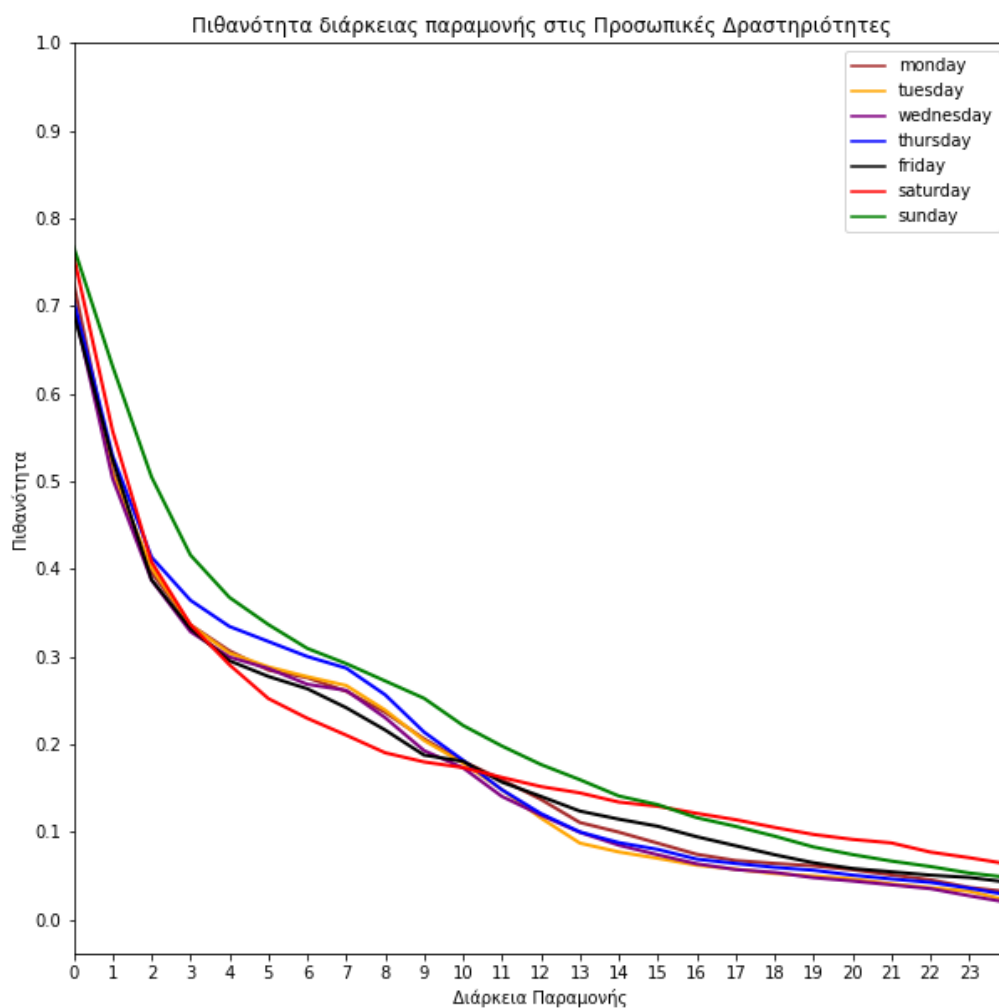
μέρες Σάββατο και Κυριακή δεν παρουσιάζουν σημαντική στατιστική διαφορά. Επίσης η Παρασκευή δεν παρουσιάζει σημαντική διαφορά με τις μέρες Τρίτη και Τετάρτη αν το όριο του p-value θεωρηθεί το 0.05.

Τέλος για τη τοποθεσία **Δουλειά** οι διάμεσοι της διάρκειας παραμονής για κάθε μέρα ξεχωριστά είναι οι εξής :

- **Δευτέρα** : 8 ώρες
- **Τρίτη** : 8 ώρες
- **Τετάρτη** : 8 ώρες
- **Πέμπτη** : 8 ώρες
- **Παρασκευή** : 7 ώρες
- **Σάββατο** : 2 ώρες
- **Κυριακή** : 2 ώρες

4.9.3 Προσωπικές Δραστηριότητες

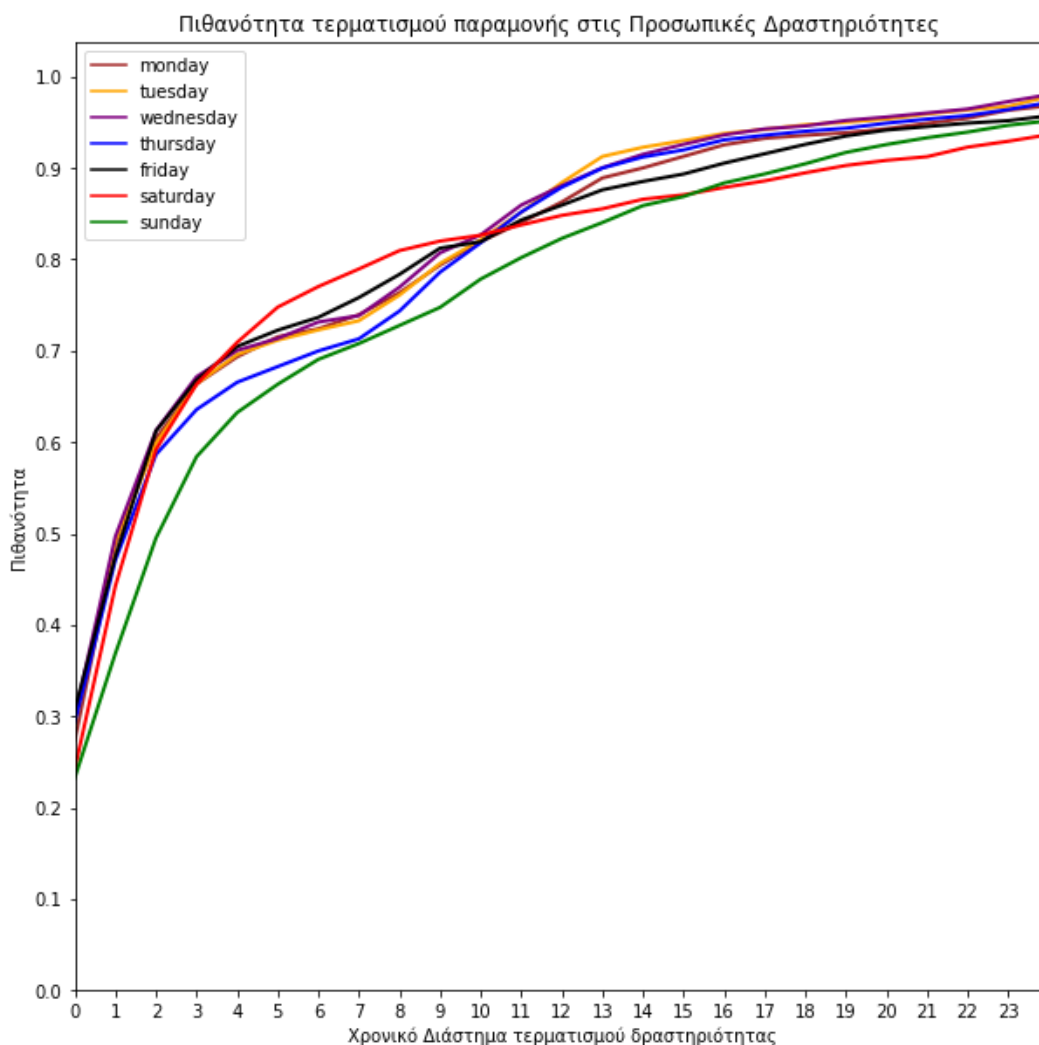
Ακολουθώντας την ίδια διαδικασία στο Διάγραμμα 34 παρουσιάζεται η καμπύλη Επιβίωσης (μέσω του Kaplan Meier estimator) για κάθε μέρα της εβδομάδας των Προσωπικών Δραστηριοτήτων.



Διάγραμμα 34: Πιθανότητα διάρκειας παραμονής στην τοποθεσία Προσωπικές Δραστηριότητες για κάθε μέρα της εβδομάδας

Και σε αυτό το διάγραμμα φαίνεται ομοιότητα στη πιθανότητα, ανάμεσα σε κάποιες μέρες και ιδιαίτερα όταν η διάρκεια παραμονής έχει μικρή τιμή.

Σε αλληλουχία με το Διάγραμμα 34 χρησιμοποιώντας τη Συνάρτηση Αθροιστικής Πυκνότητας (Cumulative Distribution Function) κατασκευάζεται το Διάγραμμα 35 που ορίζει τη πιθανότητα τερματισμού της παραμονής στις Προσωπικές Δραστηριότητες.



Διάγραμμα 35: Πιθανότητα τερματισμού παραμονής στις Προσωπικές Δραστηριότητες για κάθε μέρα της εβδομάδας

Τα συμπεράσματα είναι ίδια με αυτά του προηγούμενου διαγράμματος, οπότε είναι αναγκαία η χρήση του Log Rank Test ώστε να εξαχθούν οι μέρες που μεταξύ τους παρουσιάζουν μικρή στατιστική διαφορά.

Στο Πίνακα 8 παρουσιάζονται τα αποτελέσματα του Log Rank Test όσον αφορά τις Προσωπικές Δραστηριότητες των χρηστών.

P-value	Δευτέρα	Τρίτη	Τετάρτη	Πέμπτη	Παρασκευή	Σάββατο	Κυριακή
Δευτέρα	-	0.22	0.07	0.84	0.83	0.10	<0.005
Τρίτη	-	-	0.54	0.31	0.18	0.01	<0.005
Τετάρτη	-	-	-	0.10	0.04	<0.005	<0.005
Πέμπτη	-	-	-	-	0.76	0.11	<0.005
Παρασκευή	-	-	-	-	-	0.24	<0.005
Σάββατο	-	-	-	-	-	-	0.02
Κυριακή	-	-	-	-	-	-	-

Πίνακας 8: Σύγκριση ημερών σε σχέση με τη διάρκεια παραμονής στις Προσωπικές Δραστηριότητες

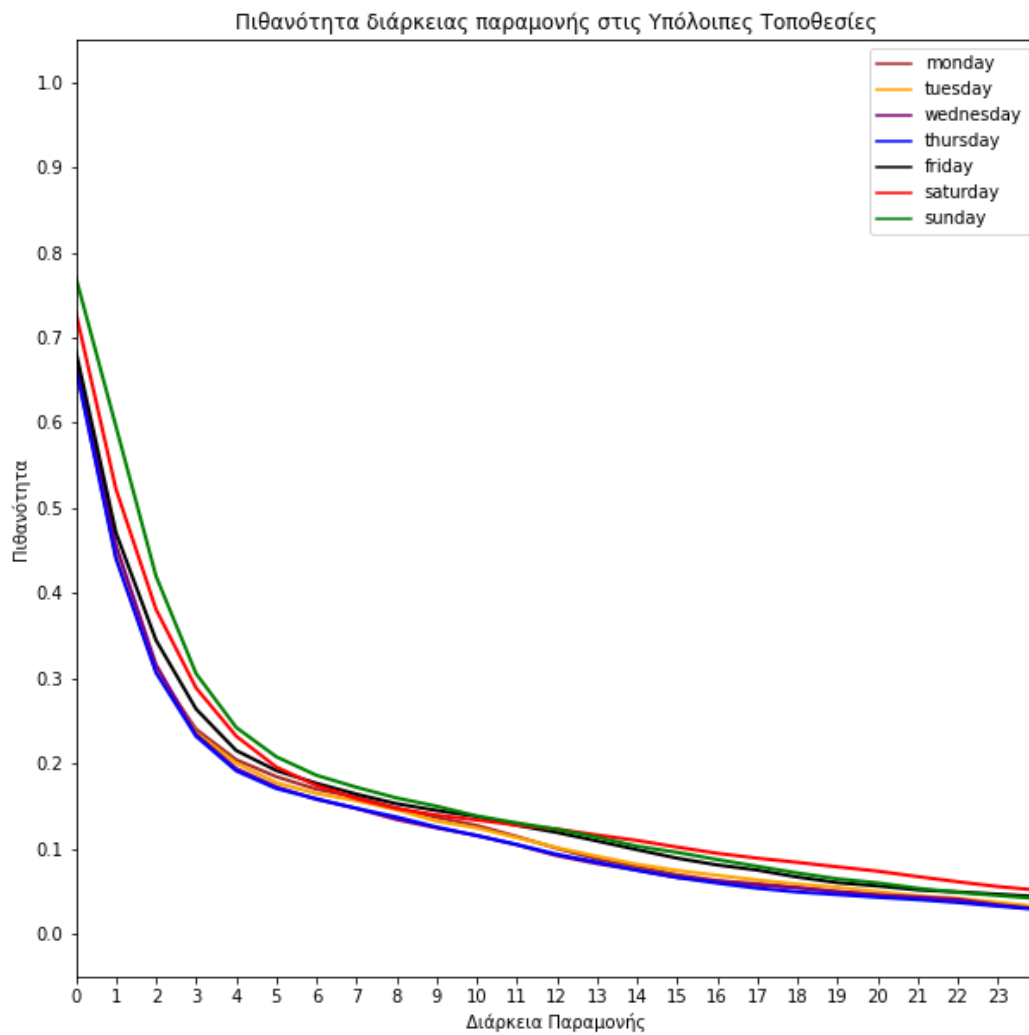
Συμπεραίνεται ότι όσον αφορά τις Προσωπικές Δραστηριότητες η πλειοψηφία των ημερών μεταξύ τους δεν παρουσιάζουν σημαντική στατιστική διαφορά.

Τέλος για τις Προσωπικές Δραστηριότητες οι διάμεσοι της διάρκειας παραμονής για κάθε μέρα ξεχωριστά είναι οι εξής :

- Δευτέρα : 2 ώρες
- Τρίτη : 2 ώρες
- Τετάρτη : 2 ώρες
- Πέμπτη : 2 ώρες
- Παρασκευή : 2 ώρες
- Σάββατο : 2 ώρες
- Κυριακή : 3 ώρες

4.9.4 Υπόλοιπες Τοποθεσίες

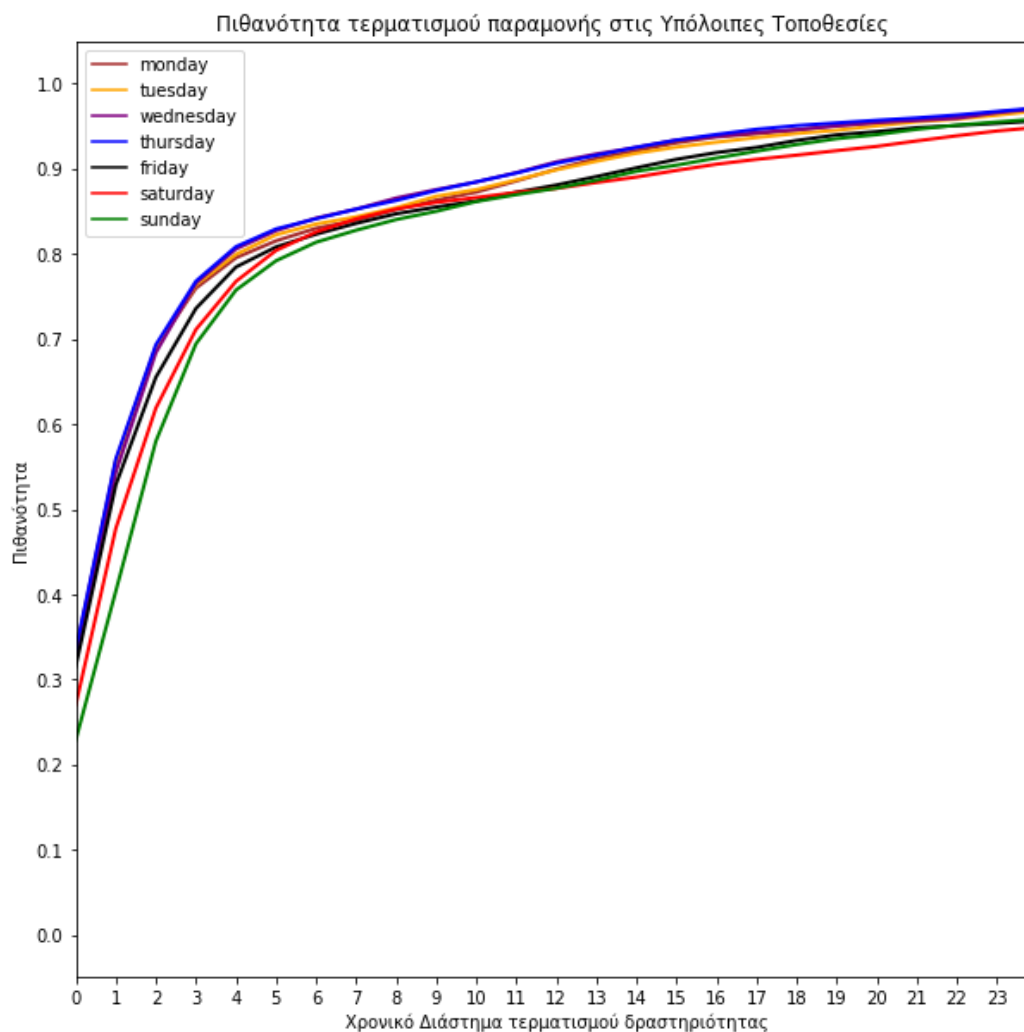
Σε συνέχεια με τις παραπάνω διαδικασίες το Διάγραμμα 36 παρουσιάζεται η καμπύλη Επιβίωσης (μέσω του Kaplan Meier estimator) για τις Υπόλοιπες Τοποθεσίες.



Διάγραμμα 36: Πιθανότητα διάρκειας παραμονής στις Υπόλοιπες Τοποθεσίες για κάθε μέρα της εβδομάδας

Όμοια με τις παραπάνω βασικές τοποθεσίες διακρίνεται πάλι ομοιότητα στη πιθανότητα διάρκειας παραμονής στις Υπόλοιπες Τοποθεσίες ανάμεσα στις μέρες της εβδομάδας.

Σε αλληλουχία με το Διάγραμμα 36 χρησιμοποιώντας τη Συνάρτηση Αθροιστικής Πυκνότητας (Cumulative Distribution Function) κατασκευάζεται το Διάγραμμα 37 που δείχνει τη πιθανότητα τερματισμού της παραμονής στις Υπόλοιπες Τοποθεσίες.



Διάγραμμα 37: Πιθανότητα τερματισμού παραμονής στις Υπόλοιπες Τοποθεσίες για κάθε μέρα της εβδομάδας

Προς επιβεβαίωση των ανώτερων ενδείξεων περί ομοιότητας στις διάρκειες της δραστηριότητας με προορισμό τις Υπόλοιπες Τοποθεσίες εκτελείται η δοκιμή κατάταξης (Log-Rank Test) μεταξύ των ημερών. Στο Πίνακα 9 φαίνονται τα αποτελέσματα αυτής της δοκιμής.

P-value	Δευτέρα	Τρίτη	Τετάρτη	Πέμπτη	Παρασκευή	Σάββατο	Κυριακή
Δευτέρα	-	0.62	0.83	0.35	<0.005	<0.005	<0.005
Τρίτη	-	-	0.75	0.16	<0.005	<0.005	<0.005
Τετάρτη	-	-	-	0.24	<0.005	<0.005	<0.005
Πέμπτη	-	-	-	-	<0.005	<0.005	<0.005
Παρασκευή	-	-	-	-	-	<0.005	<0.005
Σάββατο	-	-	-	-	-	-	0.01
Κυριακή	-	-	-	-	-	-	-

Πίνακας 9: Σύγκριση ημερών σε σχέση με τη διάρκεια παραμονής στις Υπόλοιπες Τοποθεσίες

Τα αποτελέσματα υποδεικνύουν πως δεν υπάρχει σημαντική διαφορά ανάμεσα στις μέρες Δευτέρα έως Πέμπτη όσον αφορά τις διάρκειες των δραστηριοτήτων.

Τέλος για τις Υπόλοιπες Τοποθεσίες οι διάμεσοι της διάρκειας παραμονής για κάθε μέρα ξεχωριστά είναι οι εξής :

- **Δευτέρα** : 1 ώρα
- **Τρίτη** : 1 ώρα
- **Τετάρτη** : 1 ώρα
- **Πέμπτη** : 1 ώρα
- **Παρασκευή** : 1 ώρα
- **Σάββατο** : 2 ώρες
- **Κυριακή** : 2 ώρες

Μέσω των παραπάνω μεθοδολογιών προτυποποιήθηκαν οι διάρκειες δραστηριοτήτων ανά τύπο και ημέρα. Αυτό έχει ως απώτερο σκοπό τη μείωση πλεονασμών στα δεδομένα μας και τη βελτίωση της ακεραιότητας των δεδομένων.

5. Συμπεράσματα και Προτάσεις

5.1 Σύνοψη Μεθοδολογίας και Αποτελεσμάτων

Αντικείμενο της παρούσας εργασίας αποτελεί η ανίχνευση προφίλ κινητικότητας, αλυσίδας και διάρκειας δραστηριοτήτων. Αυτό επιτεύχθηκε μέσω της ανάλυσης μίας βάσης δεδομένων ευρείας κλίμακας που εμπειρεύει αναλυτικά στοιχεία για τα ταξίδια που πραγματοποιήσαν 696 οδηγοί τη περίοδο Δεκεμβρίου 2017 με Αύγουστο 2019.

Πιο συγκεκριμένα αρχικά μέσω της συνεργασίας αλγορίθμων μάθησης χωρίς επίβλεψη και χωροχρονικών στοιχείων των προορισμών κατηγοριοποιήθηκαν οι προορισμοί για κάθε χρήστη ξεχωριστά. Στο κεφάλαιο 4 καθώς και στο παράρτημα, μέσω της οπτικοποίησης των δεδομένων για 5 χρήστες διαφαίνεται ότι η κατηγοριοποίηση των προορισμών έχει μεγάλη ακρίβεια. Εν συνεχεία για κάθε χρήστη ξεχωριστά ποσοτικοποιήθηκαν τα ταξίδια ανάμεσα στους βασικούς του προορισμούς. Επίσης μέσω της χωροχρονικής ανάλυσης ανιχνεύεται η τυπική αλυσίδα δραστηριοτήτων του κάθε χρήστη σε επίπεδο συνόλου των ταξιδιών που πραγματοποιήθηκαν από τον ίδιο. Η μεγάλη ακρίβεια που έχει αυτή η μέθοδος διαφαίνεται μέσω της οπτικοποίησης των χωροχρονικών δεδομένων για 5 χρήστες στο Κεφάλαιο 4 και στο Παράρτημα Α. Τέλος, όσον αφορά την εξατομικευμένη ανάλυση ανιχνεύεται η τυπική καθημερινή αλυσίδα δραστηριοτήτων του κάθε χρήστη-οδηγού.

Η παραπάνω διαδικασία επαναλαμβάνεται για όλους τους χρήστες πλέον σε επίπεδο ημέρας εβδομάδος και εξάγονται στατιστικά συμπεράσματα όσον αφορά τα ταξίδια μεταξύ των βασικών προορισμών (Σπίτι, Δουλεία, Προσωπικές Δραστηριότητες, Υπόλοιπες Τοποθεσίες) και τις καθημερινές αλυσίδες δραστηριοτήτων τα οποία διαφαίνονται με τη μορφή διαγραμμάτων και πινάκων στο Κεφάλαιο 4. Τέλος όσον αφορά τα συγκεντρωτικά στατιστικά συμπεράσματα, μέσω της προτυποποίησης των διαρκειών των δραστηριοτήτων από τη χρήση του Kaplan Meier estimator και του Log Rank Test εξάγονται σημαντικές πληροφορίες τόσο για τις διάρκειες των δραστηριοτήτων όσο και για τις ομοιότητες που υπάρχουν ανάμεσα στις μέρες της εβδομάδας.

5.2 Βασικά Συμπεράσματα

Με βάση την παραπάνω μεθοδολογία και τα εξαγόμενα συγκεντρωτικά αποτελέσματα προκύπτουν τα ποικίλα συμπεράσματα. Πιο συγκεκριμένα προκύπτει ότι η οπτικοποίηση των δεδομένων ευρείας κλίμακας προσφέρει στον αναλυτή καλύτερη εποπτεία και επικύρωση των αποτελεσμάτων οπότε είναι αναγκαίο να αποτελεί μέρος της μεθοδολογίας στην παρούσα αλλά και σε συναφείς έρευνες. Επίσης διαπιστώνεται πως η κατηγοριοποίηση των προορισμών ενός οδηγού είναι εφικτή μέσω της συνεργασίας των μη εποπτευόμενων αλγορίθμων μάθησης (DBSCAN, K-nearest-neighbors) και τα χωροχρονικά χαρακτηριστικά των προορισμών που αναφέρθηκαν στο Κεφάλαιο 4, με ικανοποιητικά αποτελέσματα. Επιπλέον μέσω των αποτελεσμάτων αποδεικνύεται ότι η παραπάνω μεθοδολογία είναι ικανή να ανιχνεύσει το προφίλ κινητικότητας και την καθημερινή χαρακτηριστική αλυσίδα του κάθε χρήστη. Τέλος, μέσω της εξατομικευμένης ανάλυσης προκύπτει ότι

η παραπάνω μεθοδολογία είναι σε θέση να ποσοτικοποιήσει τις μετακινήσεις μεταξύ των βασικών προορισμών (Σπίτι, Δουλειά, Προσωπικές Δραστηριότητες, Υπόλοιπες Τοποθεσίες) για κάθε μέρα της εβδομάδας ξεχωριστά.

Τα αποτελέσματα της στατιστικής έρευνας υποδεικνύουν ότι τα καθημερινά προφίλ κινητικότητας παρουσιάζουν μία σταθερή κανονικότητα για τις μέρες από Δευτέρα έως Παρασκευή ενώ τα ταξίδια που πραγματοποιούνται από/προς τη Δουλειά αυτές τις μέρες δεν παρουσιάζουν σημαντική διαφορά μεταξύ τους. Επίσης για τις μέρες Δευτέρα έως Παρασκευή τα συγκεντρωτικά αποτελέσματα υποδεικνύουν ότι οι καθημερινές αλυσίδες δραστηριοτήτων παρουσιάζουν ένα σχετικά σταθερό μοτίβο. Όσον αφορά τις διάρκειες της δραστηριότητας του Σπιτιού συμπεραίνεται ότι οι μέρες από Δευτέρα έως Πέμπτη ακολουθούν παρόμοιο μοτίβο. Επιπλέον διαφαίνεται ότι οι διάρκειες της δραστηριότητας της Δουλειάς ακολουθούν παρόμοιο μοτίβο για τη πλειοψηφία των ημερών από Δευτέρα έως Παρασκευή καθώς και για τις ημέρες Σάββατο με Κυριακή. Επιπρόσθετα οι διάρκειες των Προσωπικών Δραστηριοτήτων δεν παρουσιάζουν σημαντική διαφορά για τη πλειοψηφία των ημερών από Δευτέρα έως Σάββατο. Τέλος, οι διάρκειες όλων των Υπόλοιπων Δραστηριοτήτων ακολουθούν παρόμοιο μοτίβο για τις μέρες από Δευτέρα έως Πέμπτη.

Το σύνολο των αποτελεσμάτων αποτελούν ένα σύνολο πληροφοριών ζωτικής σημασίας στον μοντελισμό της ζήτησης (demand modeling) καθώς και στο σχεδιασμό, τη διαχείριση και την εφαρμογή αστικών συγκοινωνιακών συστημάτων στο παρόν και μελλοντικό συγκοινωνιακό δίκτυο.

5.3 Συστάσεις για περαιτέρω έρευνα

Μέσω της βιβλιογραφικής ανασκόπησης και του συνόλου της παρούσας Διπλωματικής Εργασίας είναι φανερό ότι υπάρχουν περιθώρια περαιτέρω έρευνας στο συναφές θέμα που πραγματεύεται η εργασία. Μερικές συστάσεις για περαιτέρω διερεύνηση του παρόντος θέματος είναι οι ακόλουθες.

Αρχικά, θα ήταν εφικτό μέσω του συνδυασμού των καθημερινών αλυσίδων των δραστηριοτήτων και της Ανάλυσης Επιβίωσης να δημιουργηθεί ένα μοντέλο πρόβλεψης της αλυσίδας του κάθε χρήστη.

Επιπλέον η παραπάνω έρευνα θα μπορούσε να γίνει συγκεκριμένα σε μια οριοθετημένη περιοχή (πχ. Κρήτη) για ένα μεγάλο σύνολο οδηγών με συμπεράσματα τα οποία θα μπορούσαν να βελτιστοποιήσουν την αξιοποίηση των υπαρχόντων (σηματοδότηση, διπλής κατεύθυνσης δρόμος κάποιες ώρες της ημέρας) αλλά και των εν δυνάμει μελλοντικών συγκοινωνιακών δικτύων (αριθμός λωρίδων, περιοχές σύνδεσης μέσω αυτοκινητόδρομων).

Επίσης θα ήταν εφικτό να δημιουργηθεί μια εφαρμογή στο κινητό που με βάση τη πρόβλεψη σε συνδυασμό με την καθημερινή αλληλεπίδραση του χρήστη με αυτή θα του προσφέρει χρήσιμες πληροφορίες για το επόμενο του ταξίδι. Τέτοιες πληροφορίες θα μπορούσαν να είναι η κυκλοφοριακή συμφόρηση, οι εναλλακτικές διαδρομές που θα μπορούσε να ακολουθήσει, η παρουσίαση μέσω γραφημάτων της κυκλοφοριακής συμφόρησης άλλες ώρες της ημέρας με βάση την πρόβλεψη που έχει προέλθει ώστε να δίνεται η δυνατότητα στο χρήστη για τυχόν αναπρογραμματισμό των δραστηριοτήτων του.

Τέλος, παρατηρώντας τη συνεχή διόγκωση του προβλήματος της κυκλοφοριακής συμφόρησης μια λύση θα ήταν η επιστημονική κοινότητα να στραφεί σε μη συμβατικές λύσεις που επηρεάζουν τη ρίζα του προβλήματος. Μία τέτοια λύση μπορεί να αποτελέσει η δημιουργία μίας εφαρμογής και ανά χρονικά διαστήματα μια δομή που με τη πάροδο του χρόνου και την πληροφορία της καθημερινής αλυσίδας και της διάρκειας των δραστηριοτήτων που δέχεται από ένα μεγάλο σύνολο χρηστών, αναπροσαρμόζει τις δραστηριότητες του κάθε χρήστη. Αυτό έχει ως στόχο τη βέλτιστη χωρική και χρονική κατανομή όλων των χρηστών σε όλο το δίκτυο της περιοχής. Η παραπάνω σκέψη και υλοποίηση της είναι προφανές ότι με την εξέλιξη των μεταφορών και συγκεκριμένα των αυτόνομων οχημάτων θα μπορούσε να έχει βέλτιστα αποτελέσματα καθώς με τον έλεγχο της ταχύτητας του κάθε οχήματος η αναπροσαρμογή των δραστηριοτήτων θα είναι μικρότερης κλίμακας και συνεπώς ευκολότερα επιτεύξιμη.

Βιβλιογραφία

A. Stathopoulos and M. G. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transp. Res. C, Emerging Technol.*, vol. 11, no. 2, pp. 121–135, Apr. 2003.

G. Dimitrakopoulos and P. Demestichas, "Intelligent transportation systems," *IEEE Veh. Technol. Mag.*, vol. 5, no. 1, pp. 77–84, Mar. 2010.

González, M., Hidalgo, C. & Barabási, AL. Understanding individual human mobility patterns. *Nature* 453, 779–782 (2008). <https://doi.org/10.1038/nature06958>

Hafezi, M.H., Liu, L. & Millward, H. A time-use activity-pattern recognition model for activity-based travel demand modeling. *Transportation* 46, 1369–1394 (2019). <https://doi.org/10.1007/s11116-017-9840-9>

J. Harri, F. Filali and C. Bonnet, "Mobility models for vehicular ad hoc networks: a survey and taxonomy," in *IEEE Communications Surveys & Tutorials*, vol. 11, no. 4, pp. 19-41, Fourth Quarter 2009, doi: 10.1109/SURV.2009.090403.

Jiang, S., Ferreira, J. & González, M.C. Clustering daily patterns of human activities in the city. *Data Min Knowl Disc* 25, 478–510 (2012). <https://doi.org/10.1007/s10618-012-0264-z>

M. Allahviranloo, "Pattern Recognition and Personal Travel Behavior," in *Transportation Research Board 95th Annual Meeting*, 2016.

M. Ester, H.Peter Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: *KDD*, 1996, pp. 226–231.

Susilo, Y.O., Axhausen, K.W. Repetitions in individual daily activity–travel–location patterns: a study using the Herfindahl–Hirschman Index. *Transportation* 41, 995–1011 (2014). <https://doi.org/10.1007/s11116-014-9519-4>

W. Tang, D. Pi and Y. He, "A density-based clustering algorithm with sampling for travel behavior analysis", *Proceedings of International Conference on Intelligent Data Engineering and Automated Learning*, pp. 231-239, 2016.

Xiaolei Ma, Yao-Jan Wu, Yinhai Wang, Feng Chen, Jianfeng Liu, Mining smart card data for transit riders' travel patterns, *Transportation Research Part C: Emerging Technologies*, Volume 36, 2013, Pages 1-12, ISSN 0968-090X, <https://doi.org/10.1016/j.trc.2013.07.010>.

Y. Li, D. Jin, P. Hui, Z. Wang and S. Chen, "Limits of Predictability for Large-Scale Urban Vehicular Mobility," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2671-2682, Dec. 2014, doi: 10.1109/TITS.2014.2325395.

Εξώφυλλο: <https://www.thetimes.co.uk/article/roadmap-for-the-future-of-scotland-gqnx36998>

Εφαρμογή Survival Analysis:

<https://www.kdnuggets.com/2020/07/complete-guide-survival-analysis-python-part1.html>

<https://www.kdnuggets.com/2020/07/guide-survival-analysis-python-part-2.html>

<https://www.kdnuggets.com/2020/07/guide-survival-analysis-python-part-3.html>

DBSCAN example graph: <https://www.kdnuggets.com/2020/04/dbscan-clustering-algorithm-machine-learning.html>

K-N-N graph: <https://equipintelligence.medium.com/k-nearest-neighbor-classifier-knn-machine-learning-algorithms-ed62feb86582>

Number of smart-phone users worldwide from 2016 to 2021:

<https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>

Survival analysis: Main distributional representations:

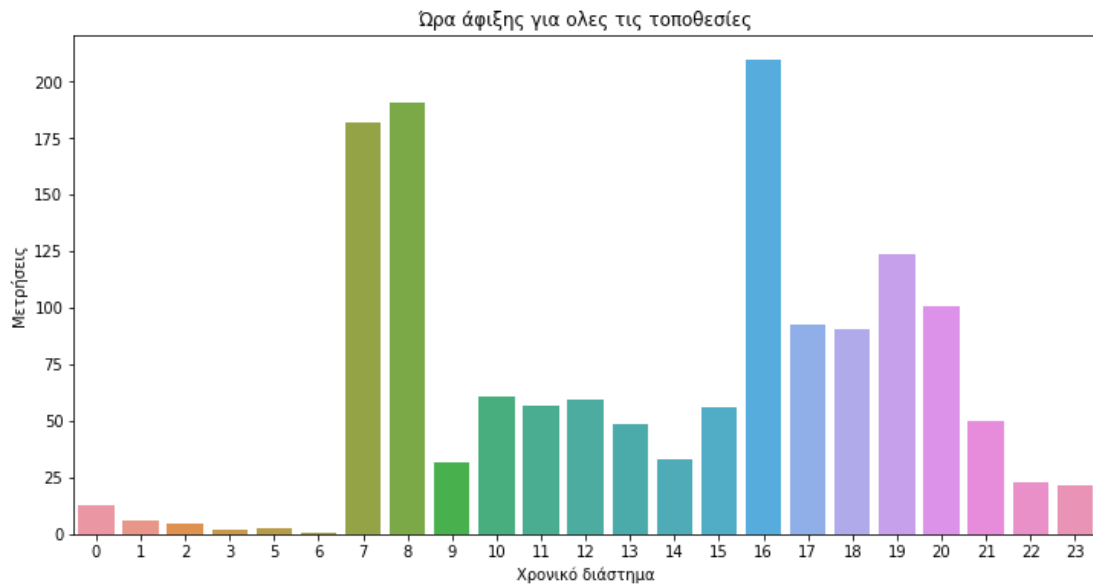
https://www.researchgate.net/figure/Survival-analysis-Main-distributional-representations_fig1_252625034

Survival function and Cumulative distribution example:

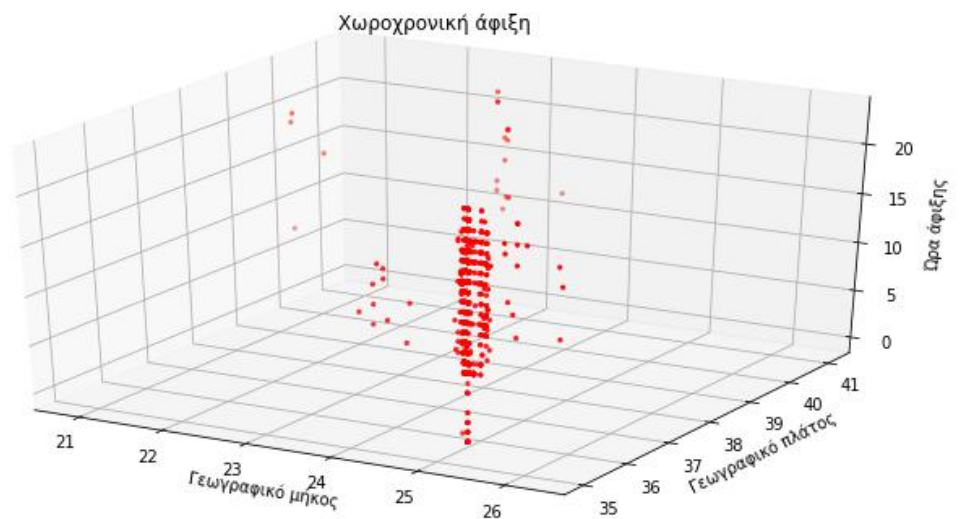
https://commons.wikimedia.org/wiki/File:Survival_function_is_1_-_CDF.svg

ΠΑΡΑΡΤΗΜΑ Α

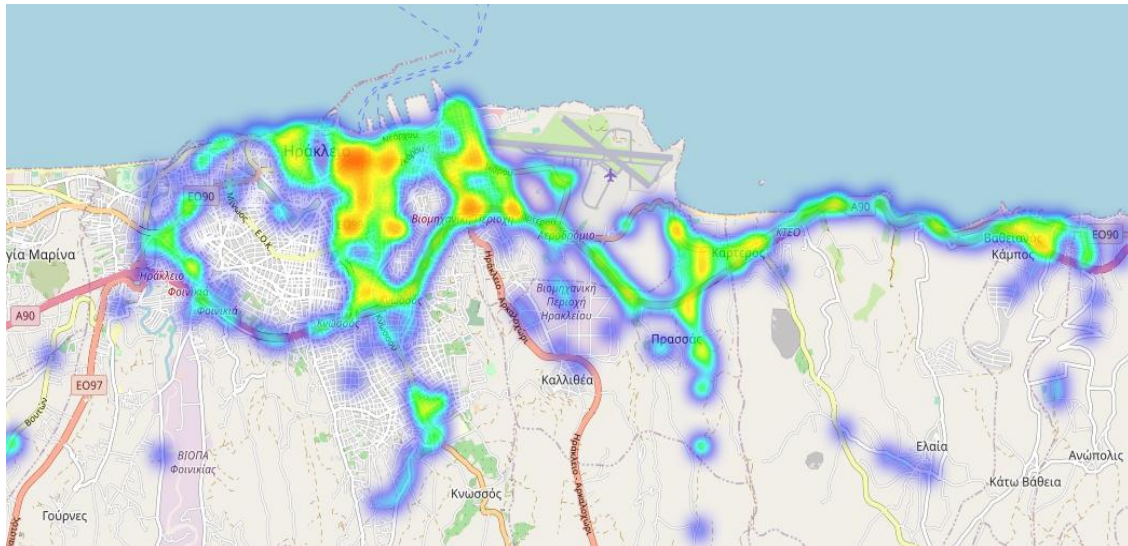
ΟΔΗΓΟΣ 2



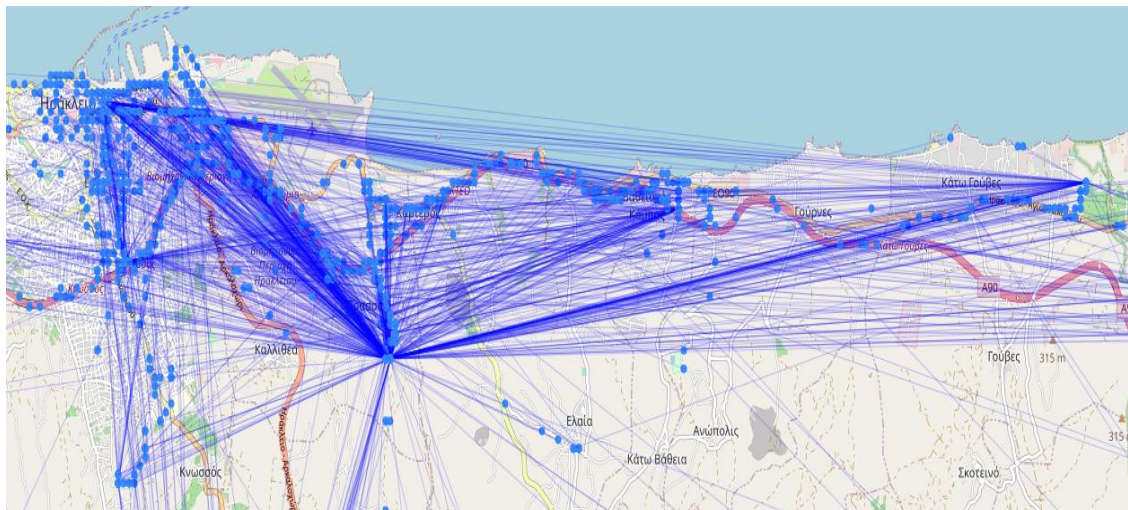
Διάγραμμα 38: Ωρα άφιξης για όλες τις τοποθεσίες του οδηγού 2



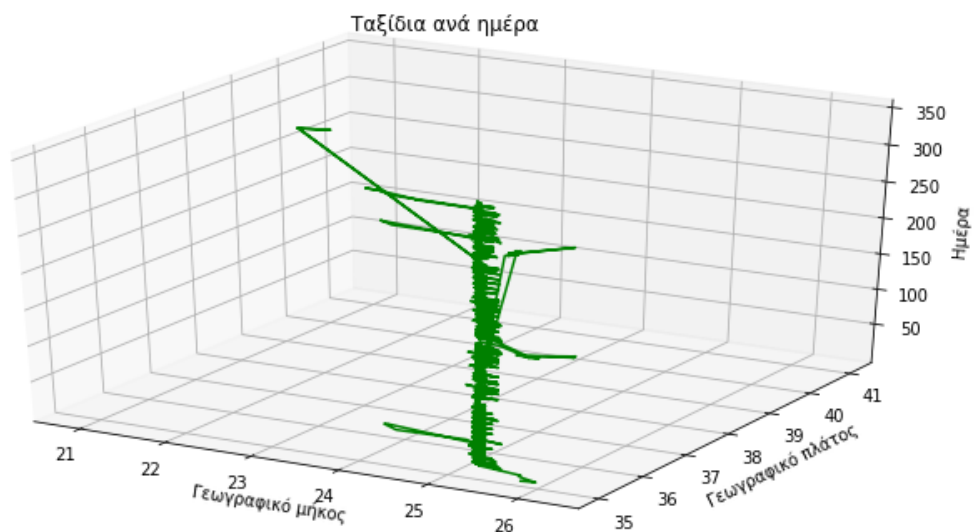
Διάγραμμα 39: Χωροχρονικές αφίξεις του οδηγού 2



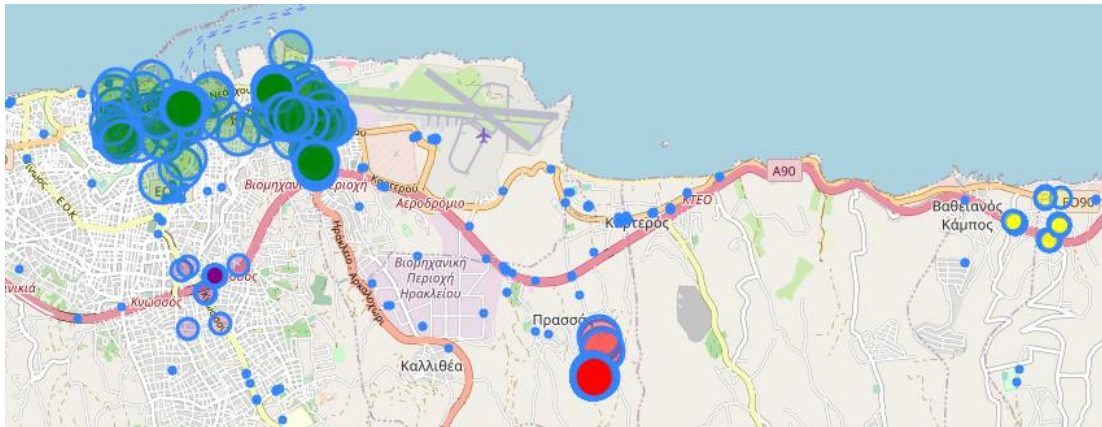
Διάγραμμα 40: Heatmap όλων των τοποθεσιών που έχει βρεθεί ο οδηγός 2



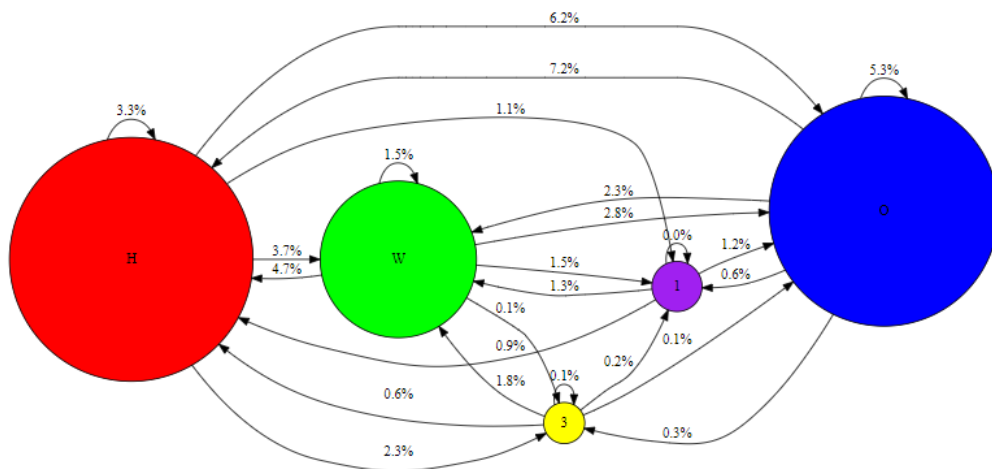
Διάγραμμα 41: Απεικόνιση όλων των ταξιδιών του οδηγού 2



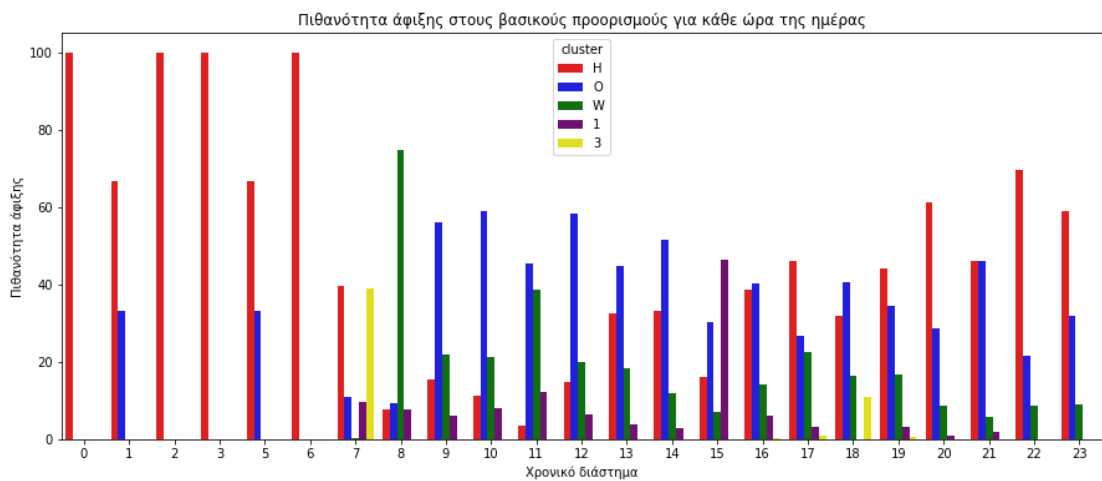
Διάγραμμα 42: Απεικόνιση μοτίβου μετακινήσεων ανά ημέρα για τον οδηγό 2



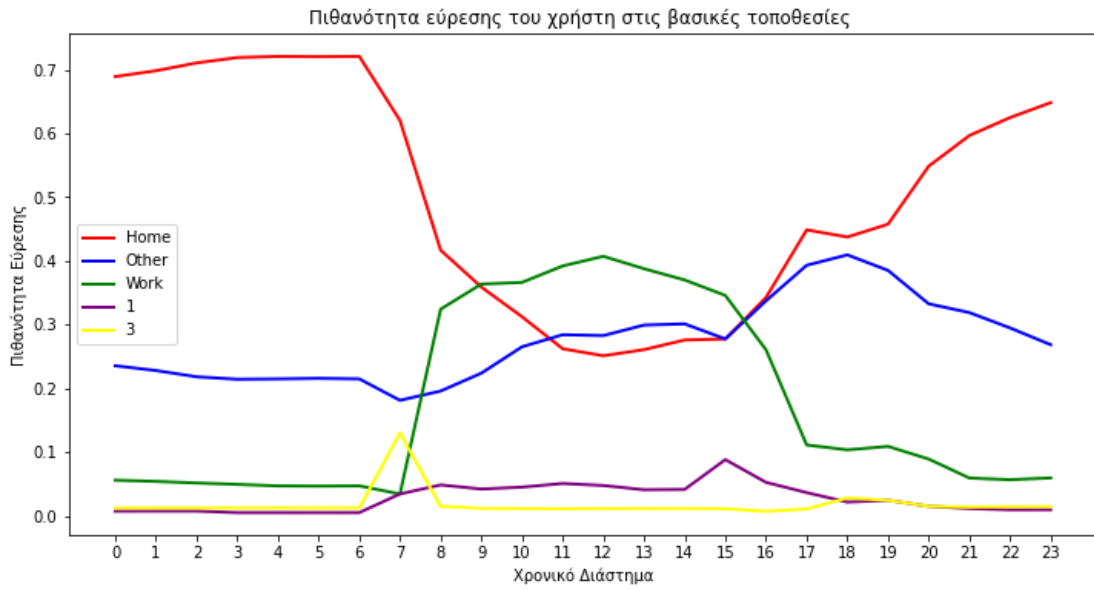
Διάγραμμα 43: Κατηγοριοποίηση προορισμών του οδηγού 2



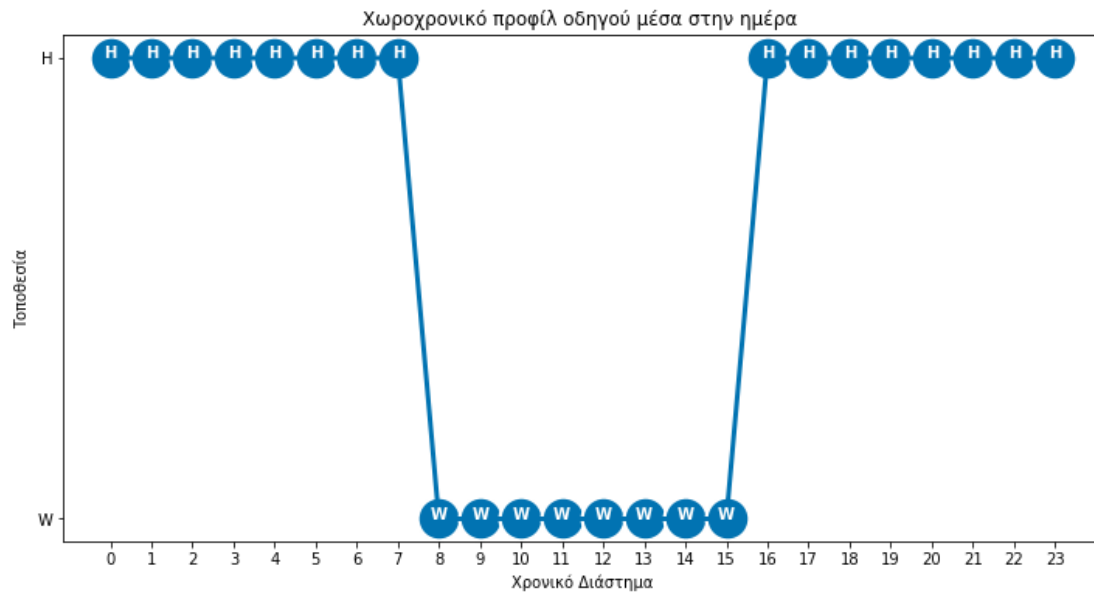
Διάγραμμα 44: Μοτίβο κινητικότητας του οδηγού 2



Διάγραμμα 45: Πιθανότητα άφιξης στους βασικούς προορισμούς για κάθε ώρα της ημέρας για τον οδηγό 2

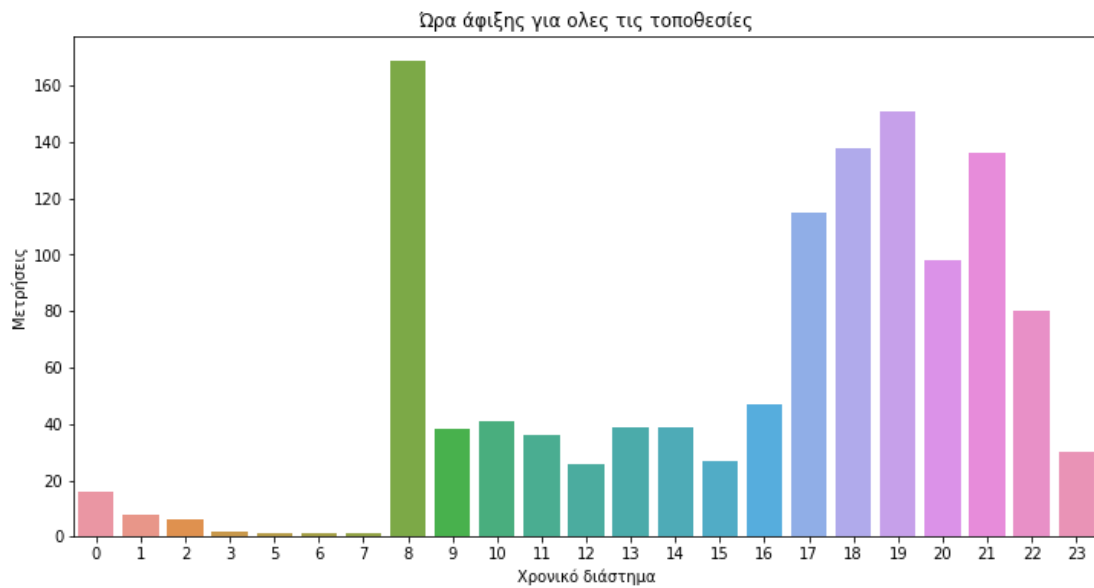


Διάγραμμα 46: Πιθανότητα εύρεσης του οδηγού 2 στις βασικές τοποθεσίες για κάθε ώρα της ημέρας

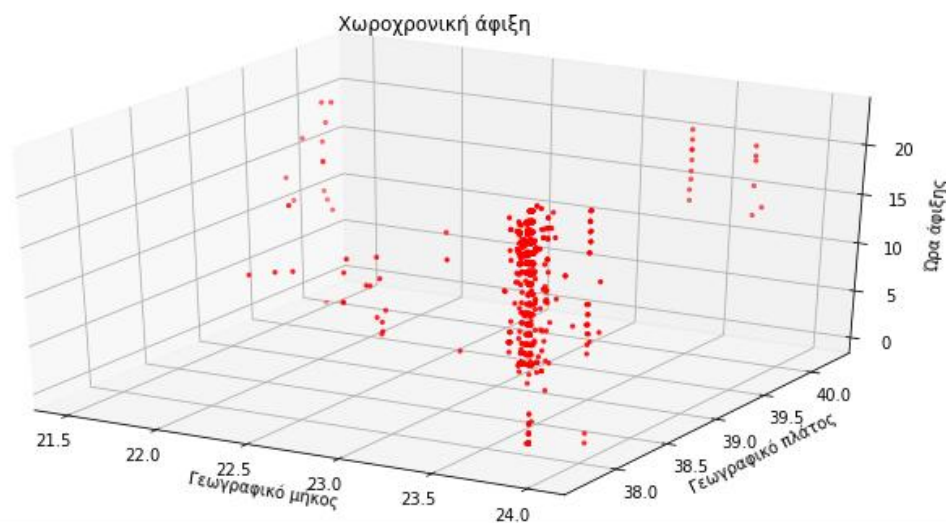


Διάγραμμα 47: Χαρακτηριστική αλυσίδα δραστηριοτήτων του οδηγού 2

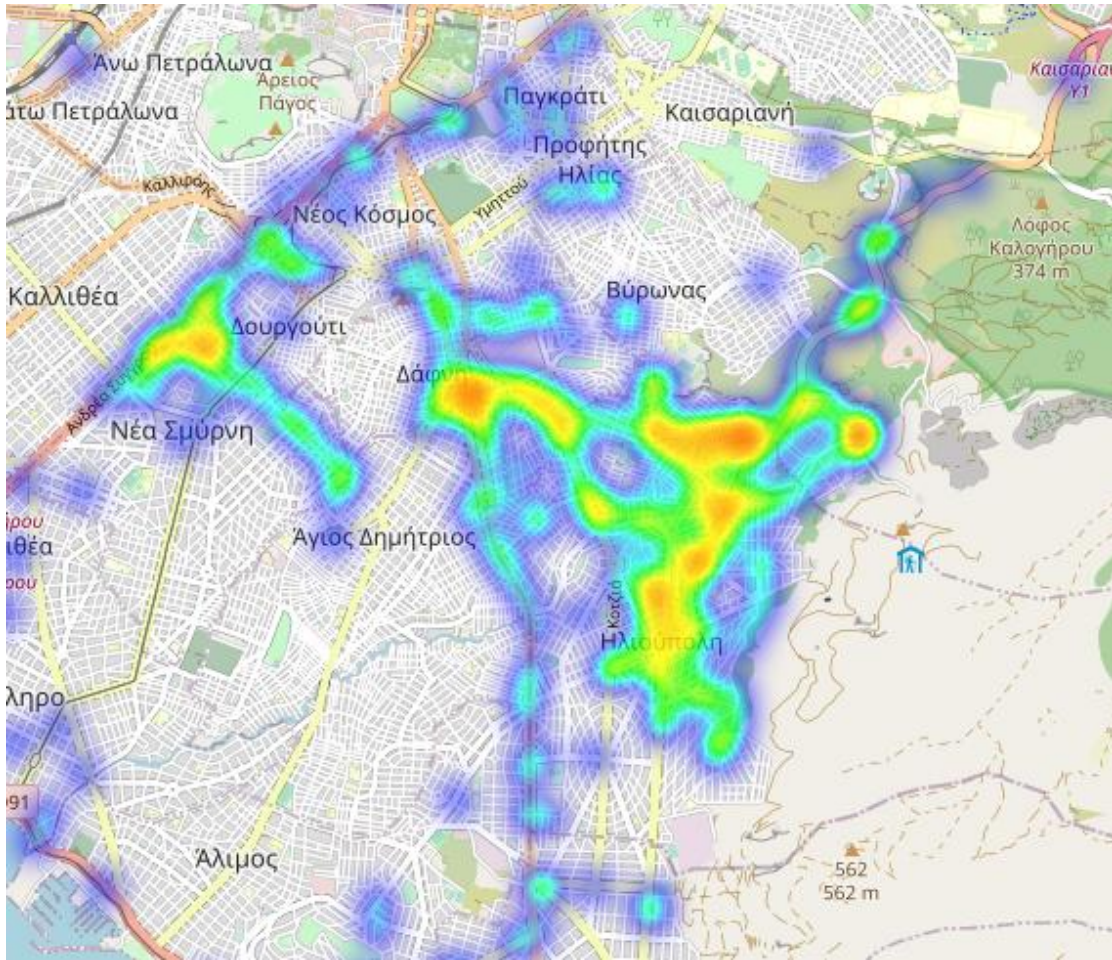
ΟΔΗΓΟΣ 3



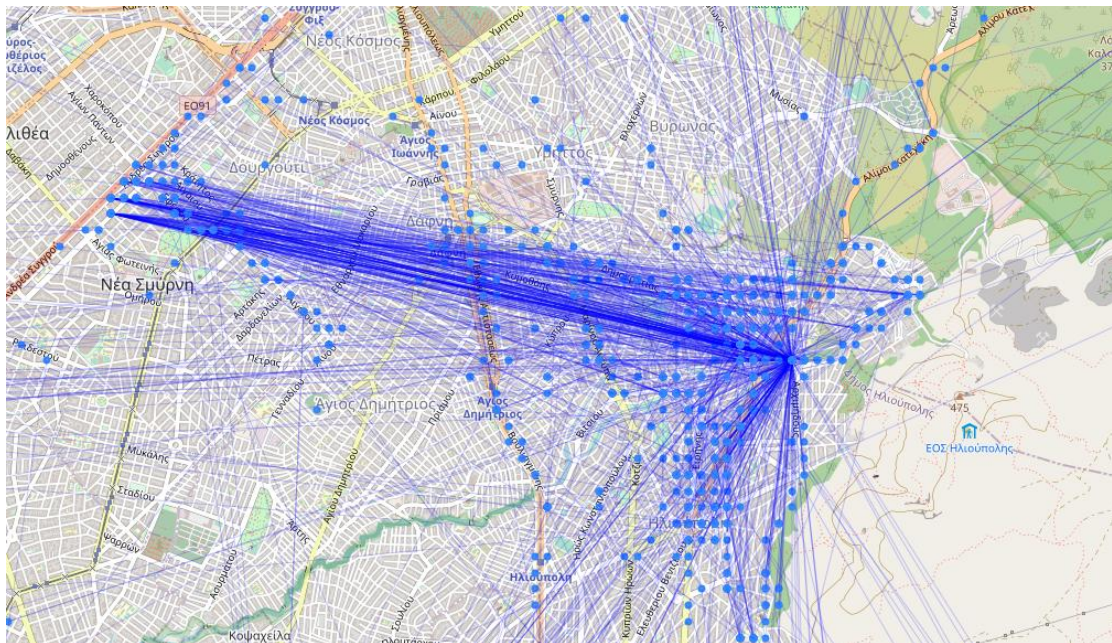
Διάγραμμα 48: Ωρα άφιξης για όλες τις τοποθεσίες του οδηγού 3



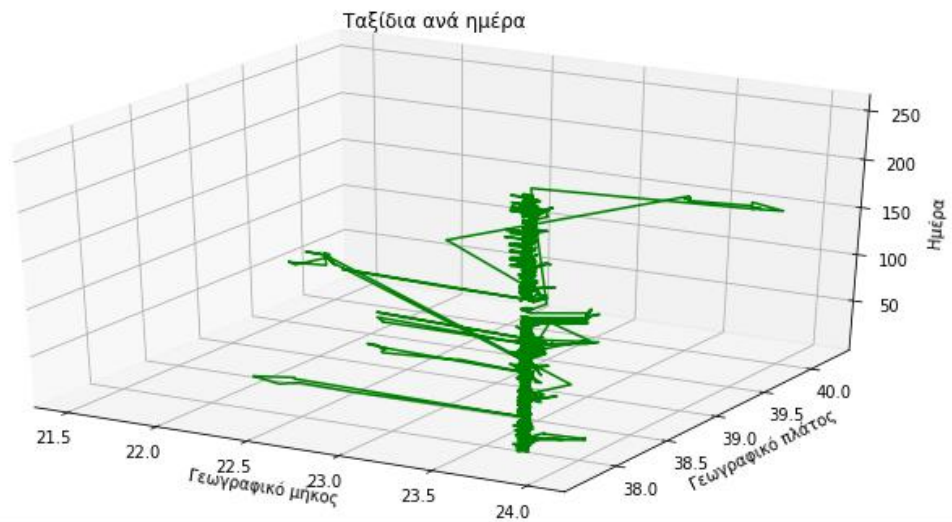
Διάγραμμα 49: Χωροχρονικές αφίξεις του οδηγού 3



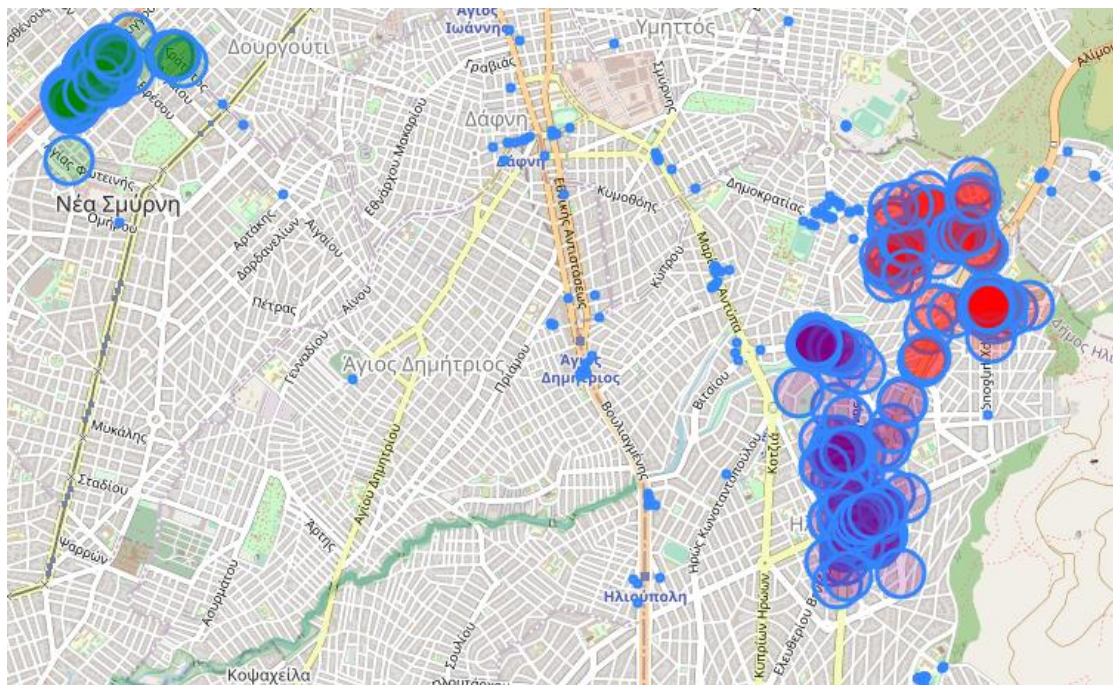
Διάγραμμα 50: Heat map όλων των τοποθεσιών που έχει βρεθεί ο οδηγός 3



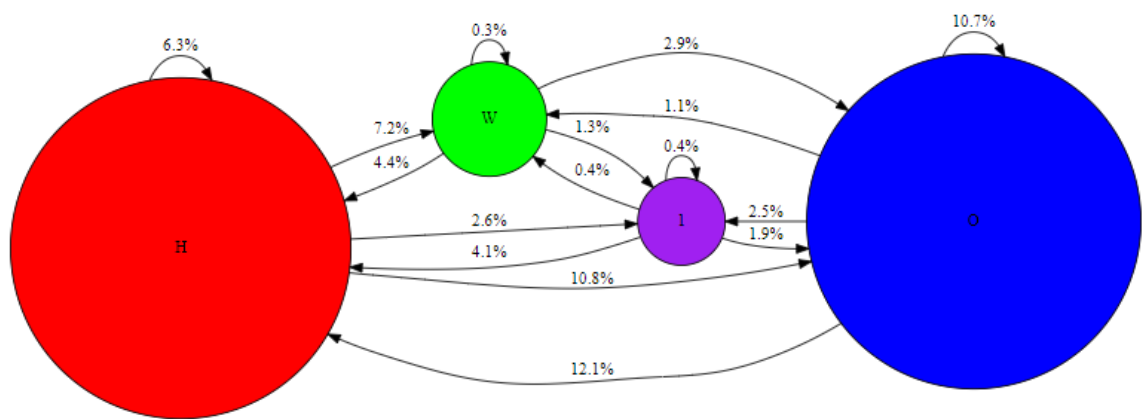
Διάγραμμα 51: Απεικόνιση όλων των ταξιδιών του οδηγού 3



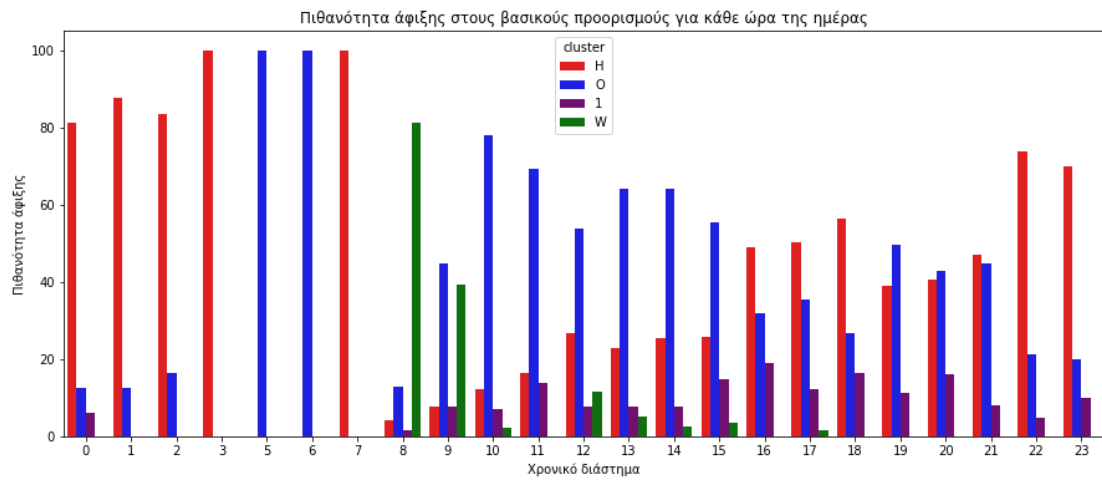
Διάγραμμα 52: Απεικόνιση μοτίβου μετακινήσεων ανά ημέρα για τον οδηγό 3



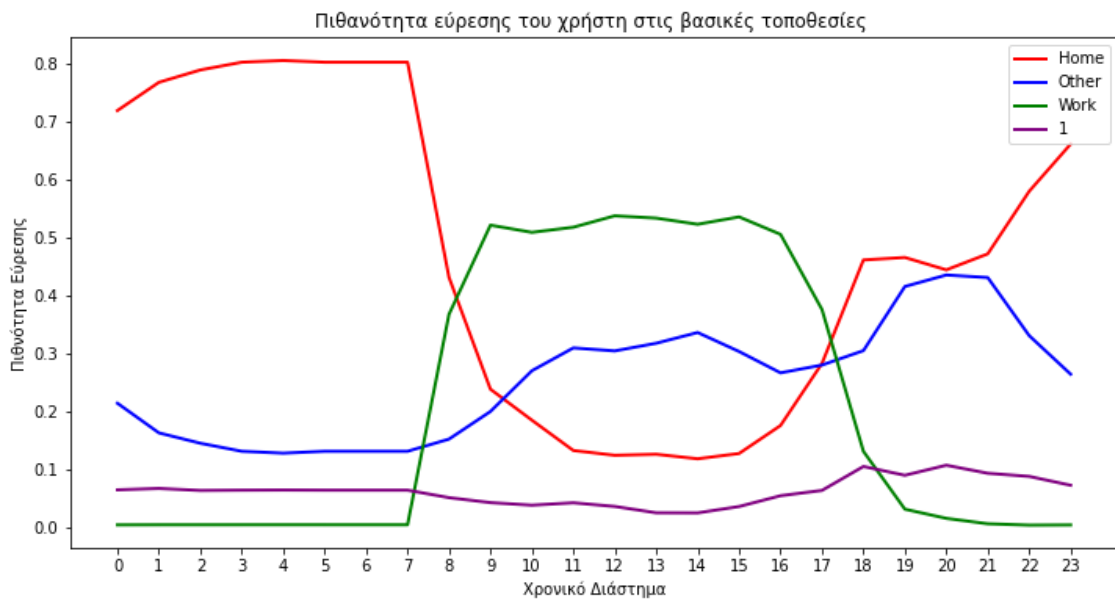
Διάγραμμα 53: Κατηγοριοποίηση προορισμών του οδηγού 3



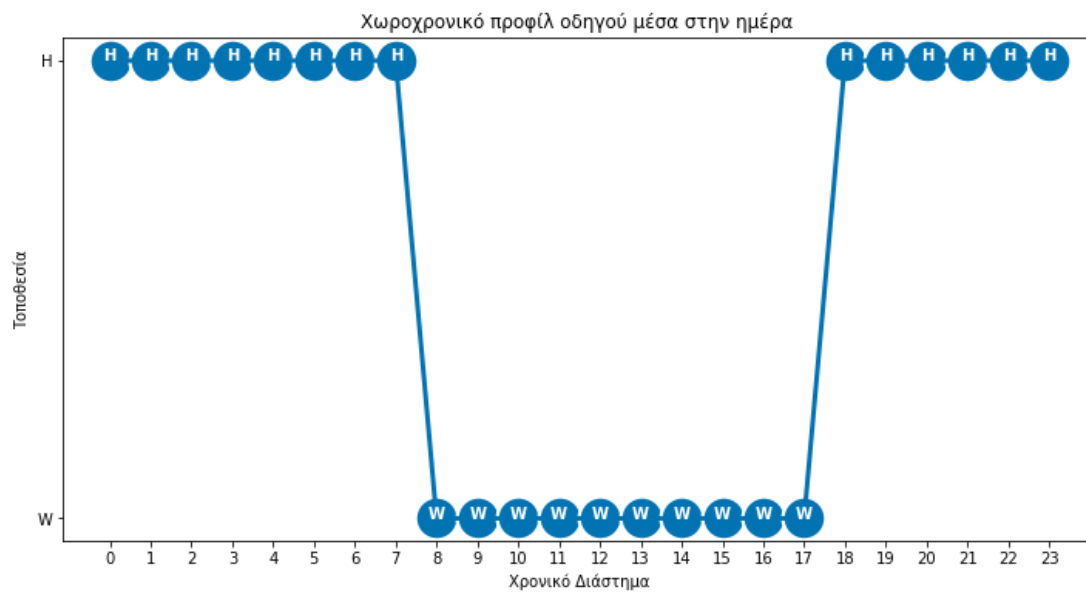
Διάγραμμα 54: Μοτίβο κινητικότητας του οδηγού 3



Διάγραμμα 55: Πιθανότητα άφιξης στους βασικούς προορισμούς για κάθε ώρα της ημέρας για τον οδηγό 3

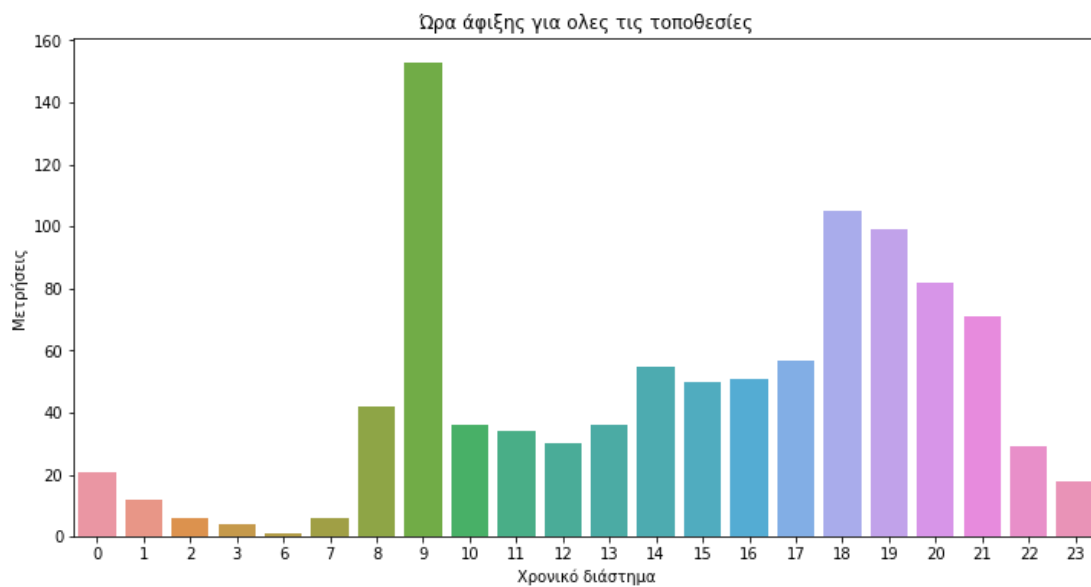


Διάγραμμα 56: Πιθανότητα εύρεσης του οδηγού 3 στις βασικές τοποθεσίες για κάθε ώρα της ημέρας

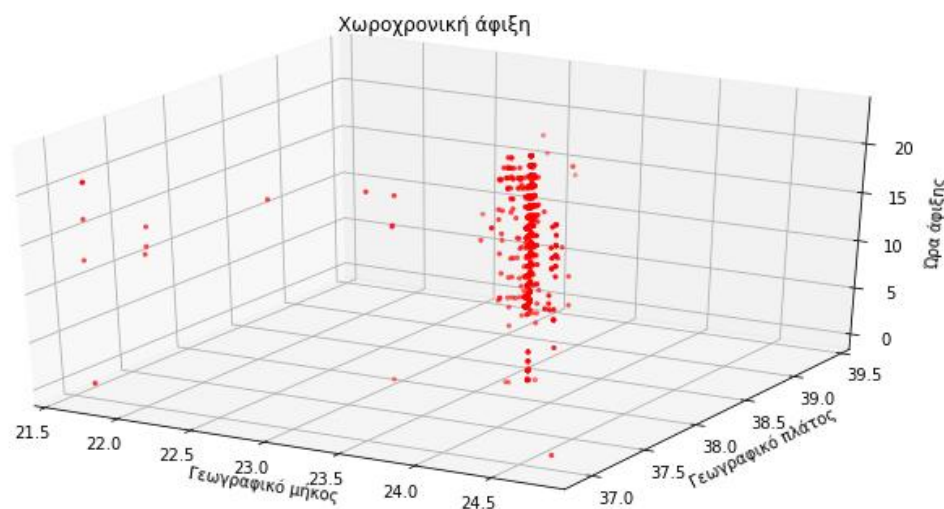


Διάγραμμα 57: Χαρακτηριστική αλυσίδα δραστηριοτήτων του οδηγού 3

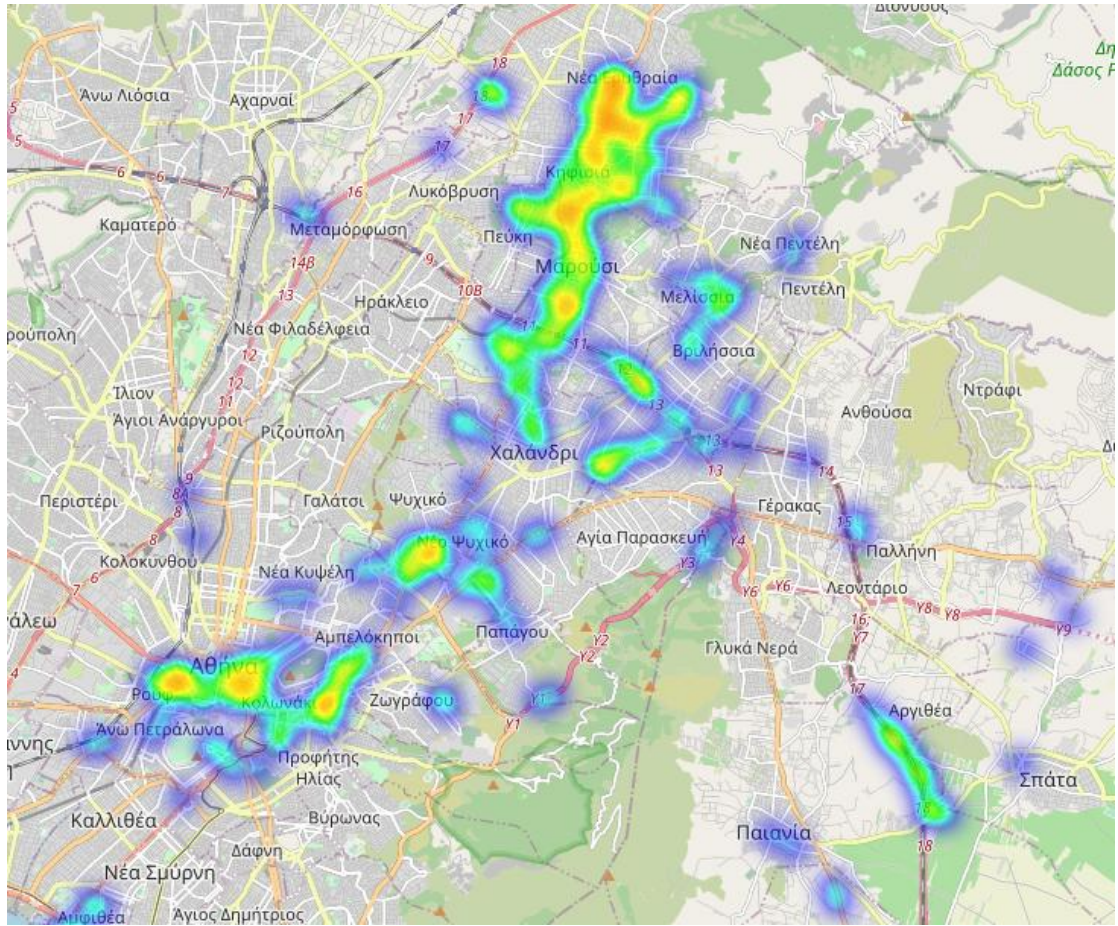
ΟΔΗΓΟΣ 4



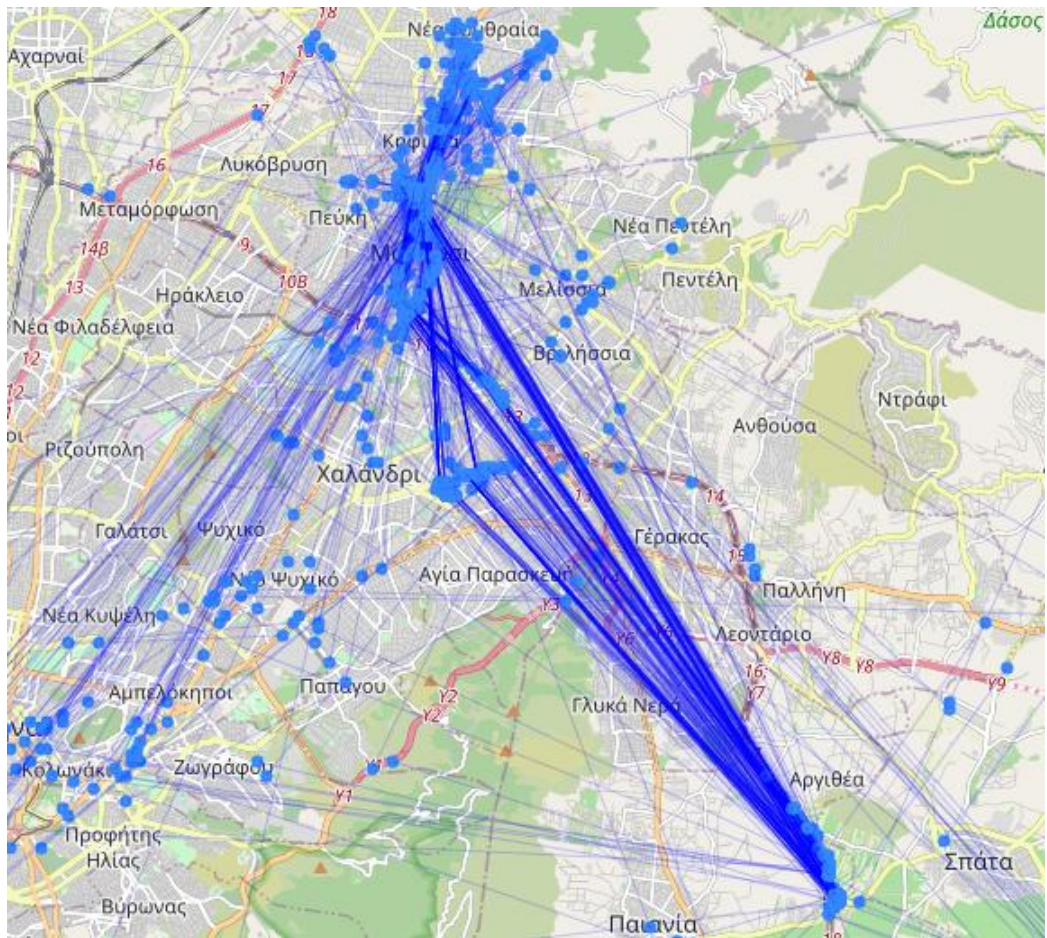
Διάγραμμα 58: Ωρα άφιξης για όλες τις τοποθεσίες του οδηγού 4



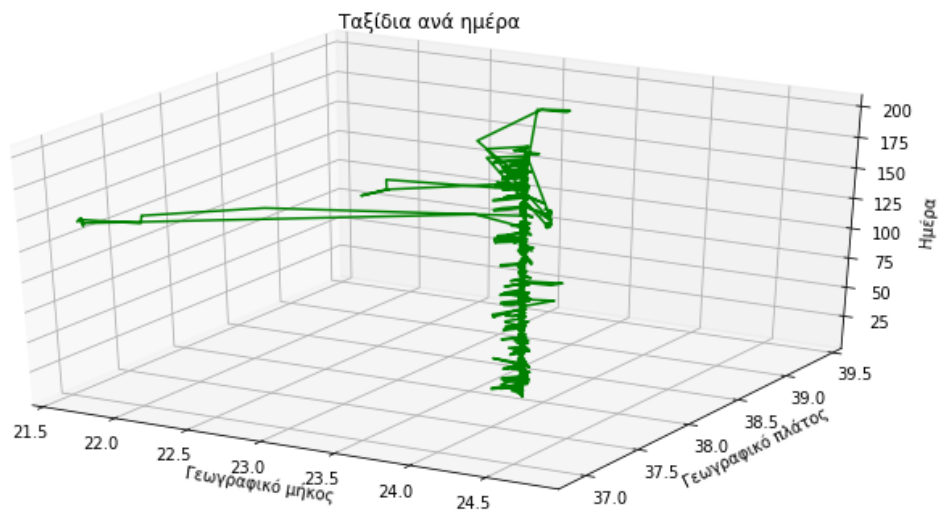
Διάγραμμα 59: Χωροχρονικές αφίξεις του οδηγού 4



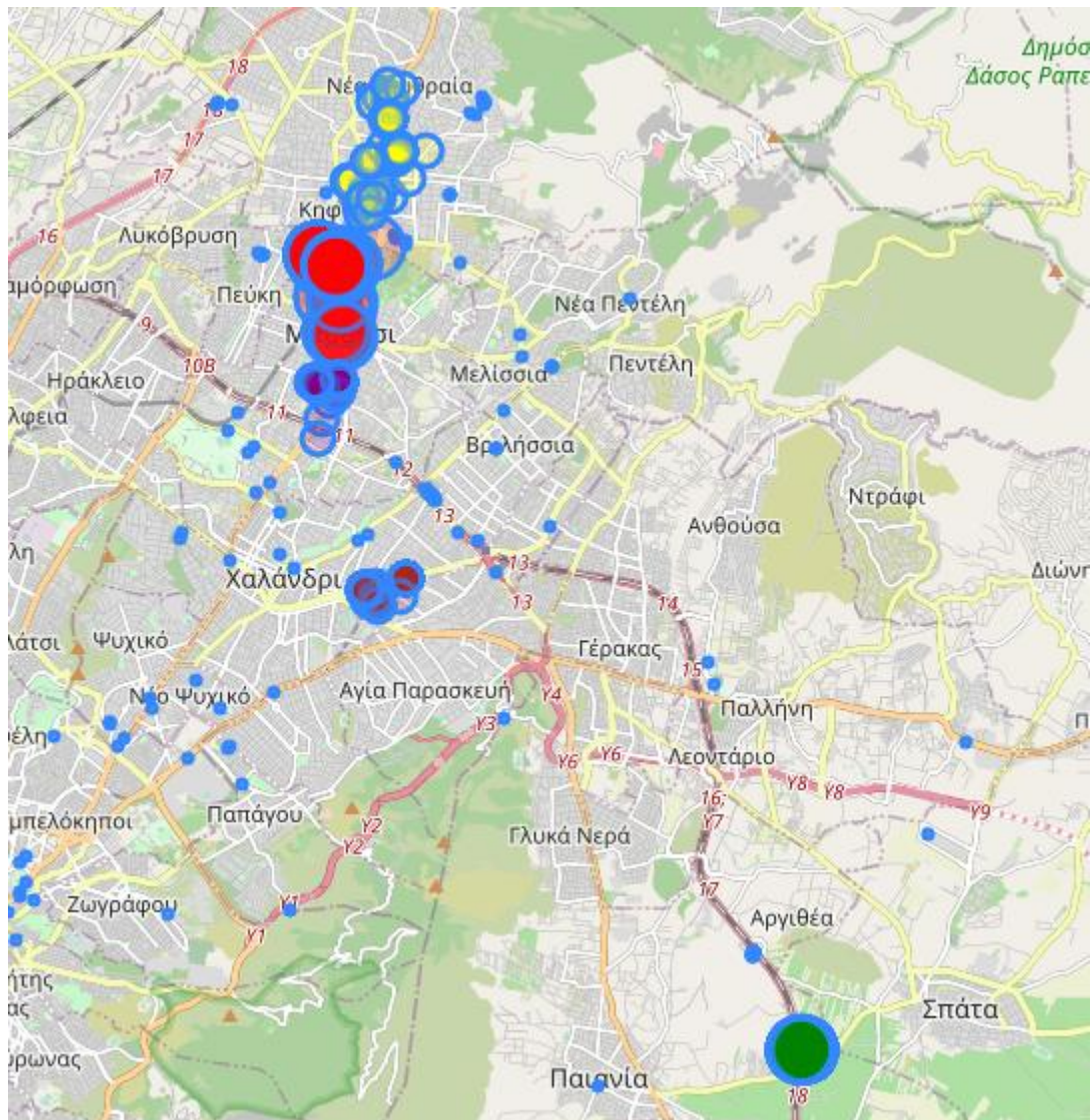
Διάγραμμα 60: Heatmap όλων των τοποθεσιών που έχει βρεθεί ο οδηγός 4



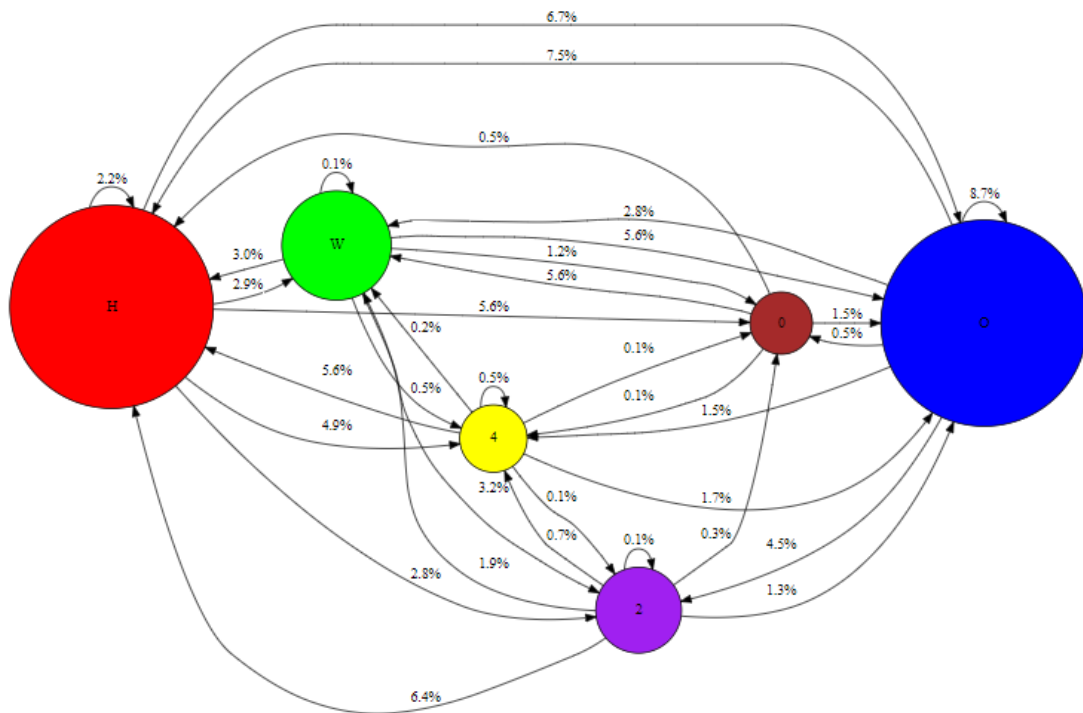
Διάγραμμα 61: Απεικόνιση όλων των ταξιδιών του οδηγού 4



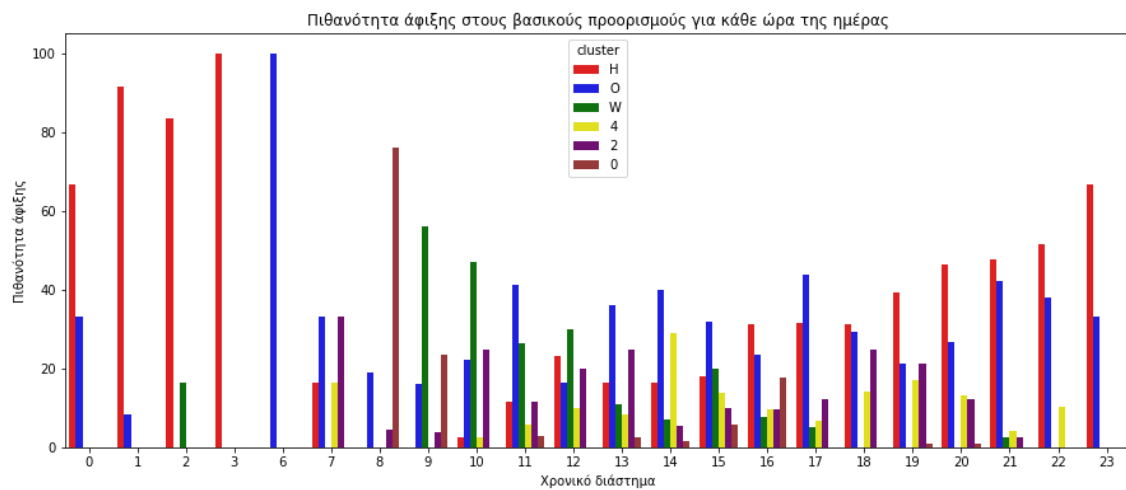
Διάγραμμα 62: Απεικόνιση μοτίβου μετακινήσεων ανά ημέρα για τον οδηγό 4



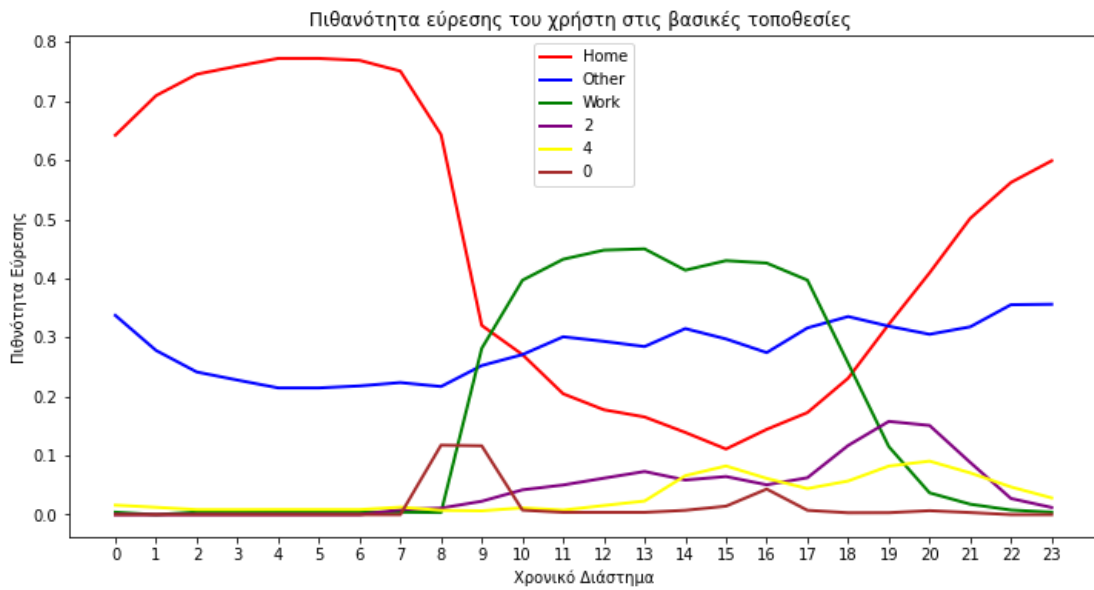
Διάγραμμα 63: Κατηγοριοποίηση προορισμών του οδηγού 4



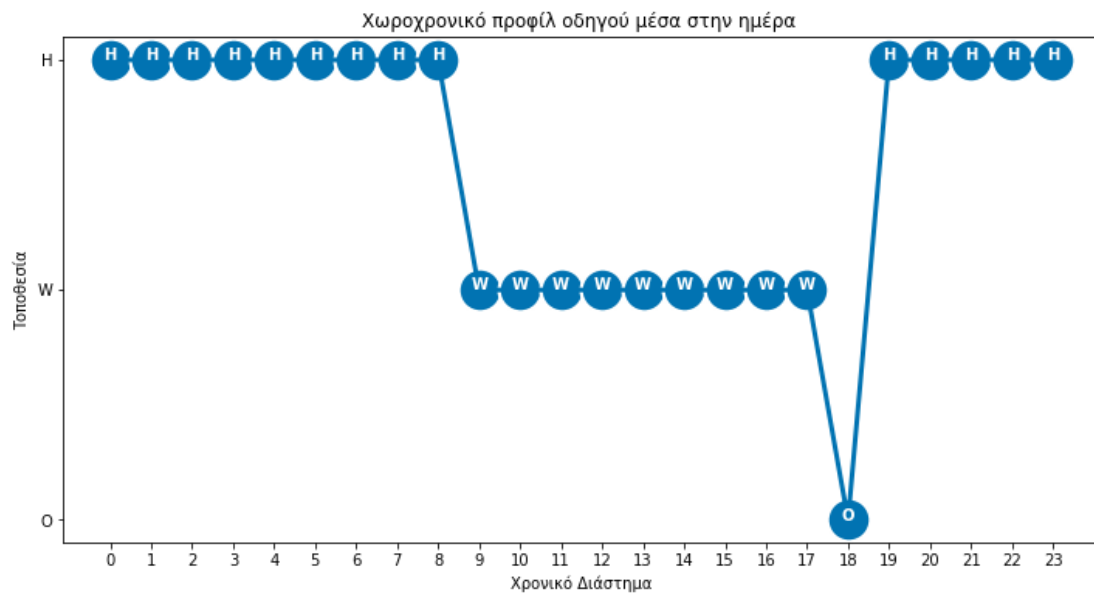
Διάγραμμα 64: Μοτίβο κινητικότητας του οδηγού 4



Διάγραμμα 65: Πιθανότητα άφιξης στους βασικούς προορισμούς για κάθε ώρα της ημέρας για τον οδηγό 4

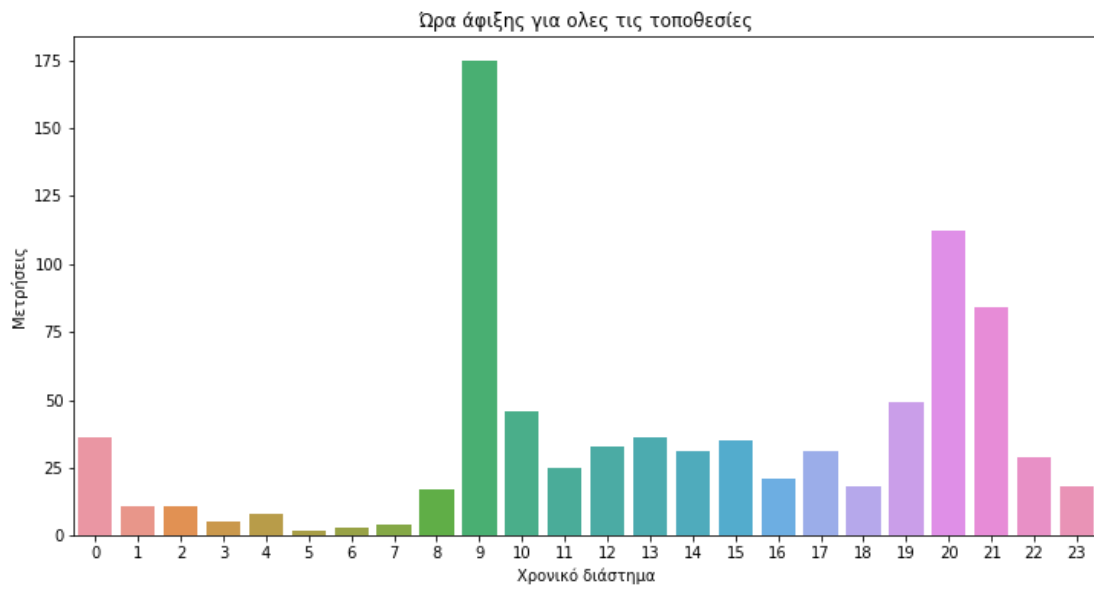


Διάγραμμα 66: Πιθανότητα εύρεσης του οδηγού 4 στις βασικές τοποθεσίες για κάθε ώρα της ημέρας

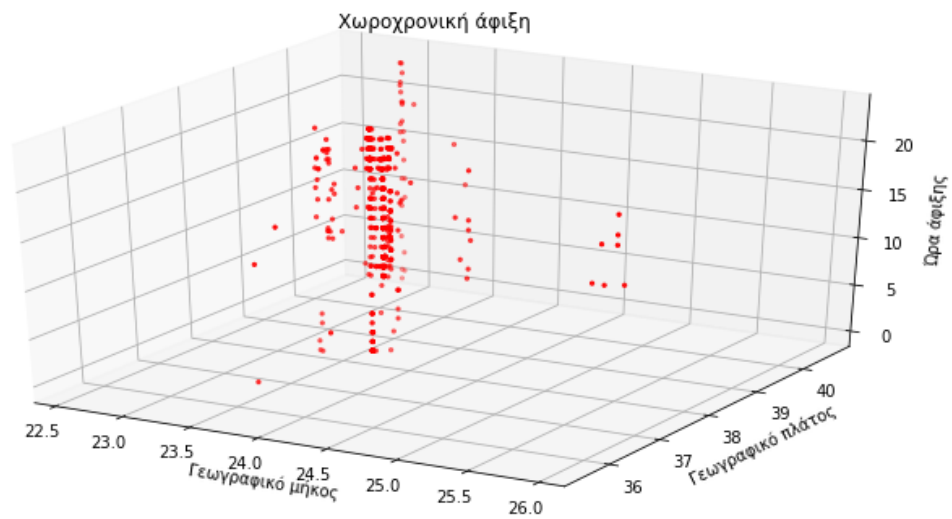


Διάγραμμα 67: Χαρακτηριστική αλυσίδα δραστηριοτήτων του οδηγού 4

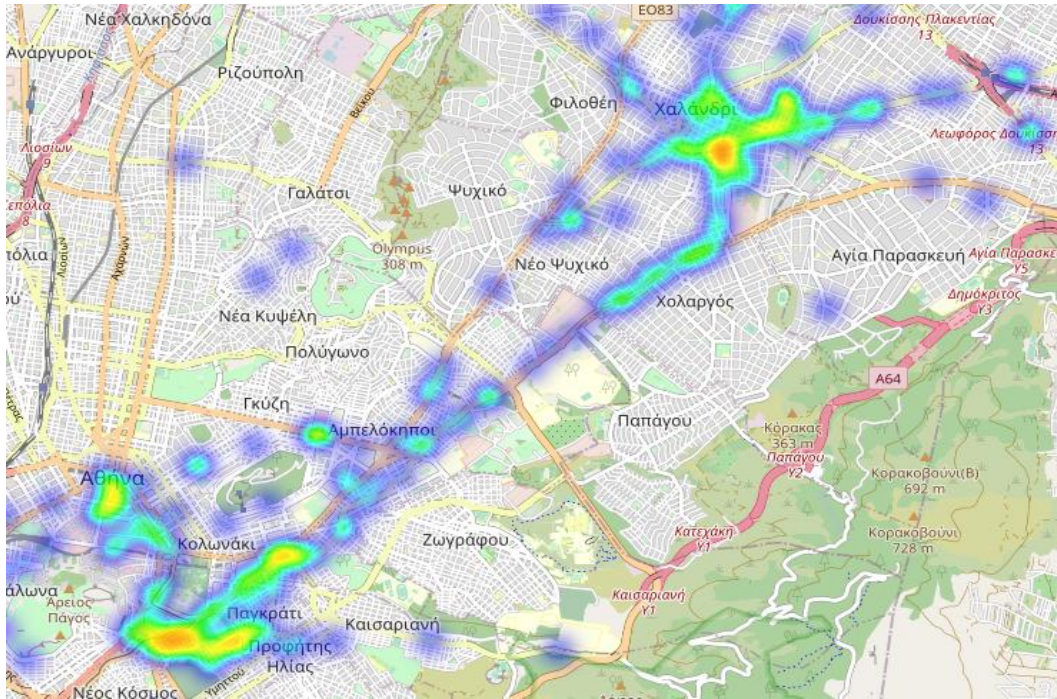
Οδηγός 5



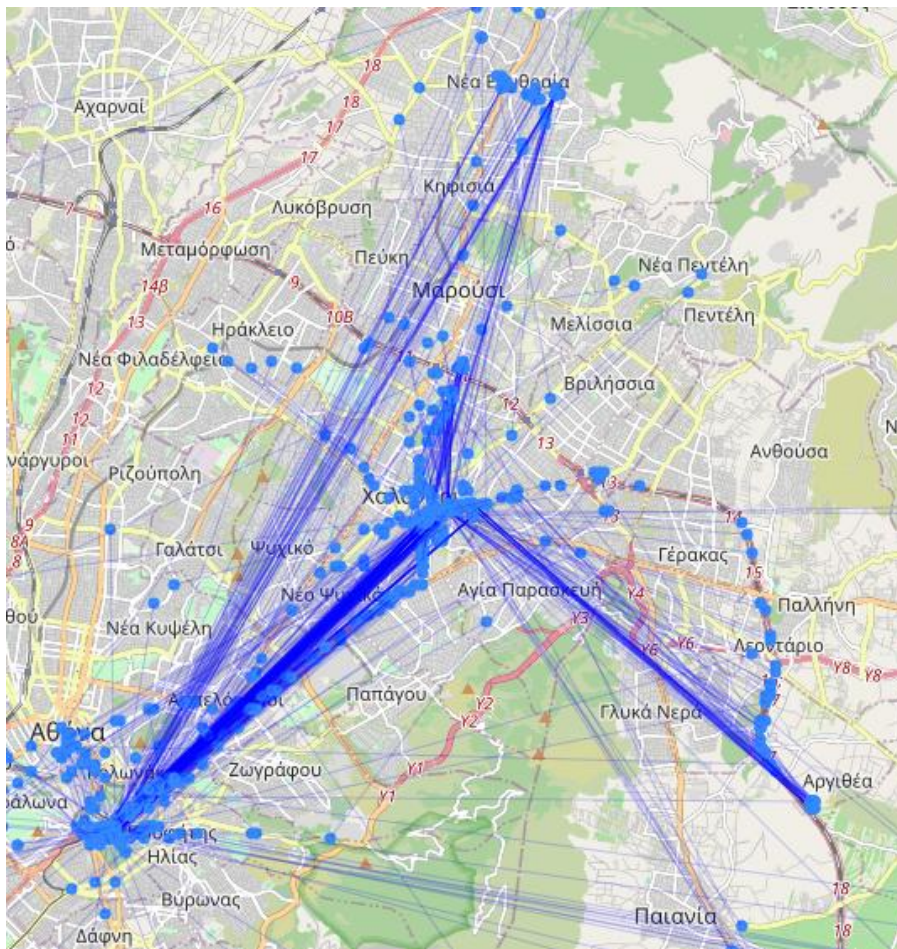
Διάγραμμα 68: Ωρα άφιξης για όλες τις τοποθεσίες του οδηγού 5



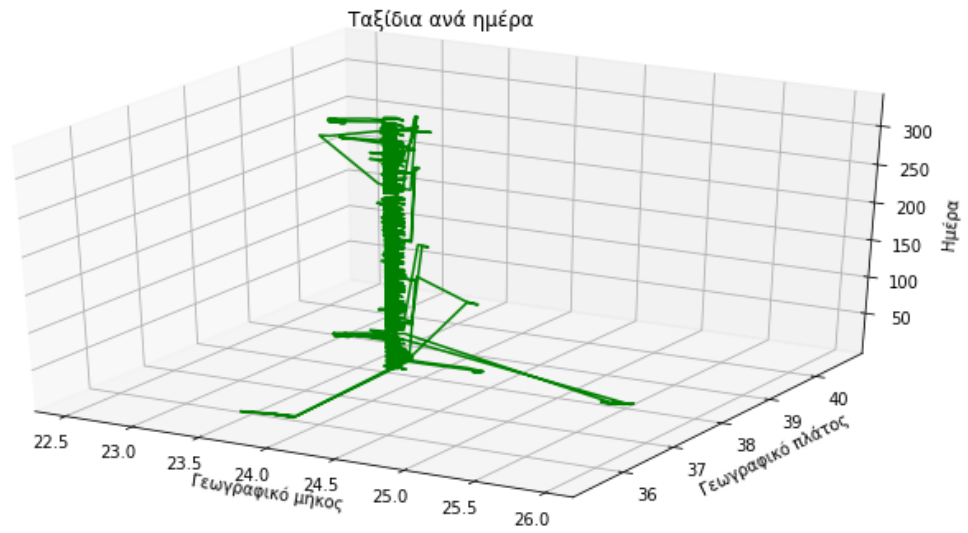
Διάγραμμα 69: Χωροχρονικές αφίξεις του οδηγού 5



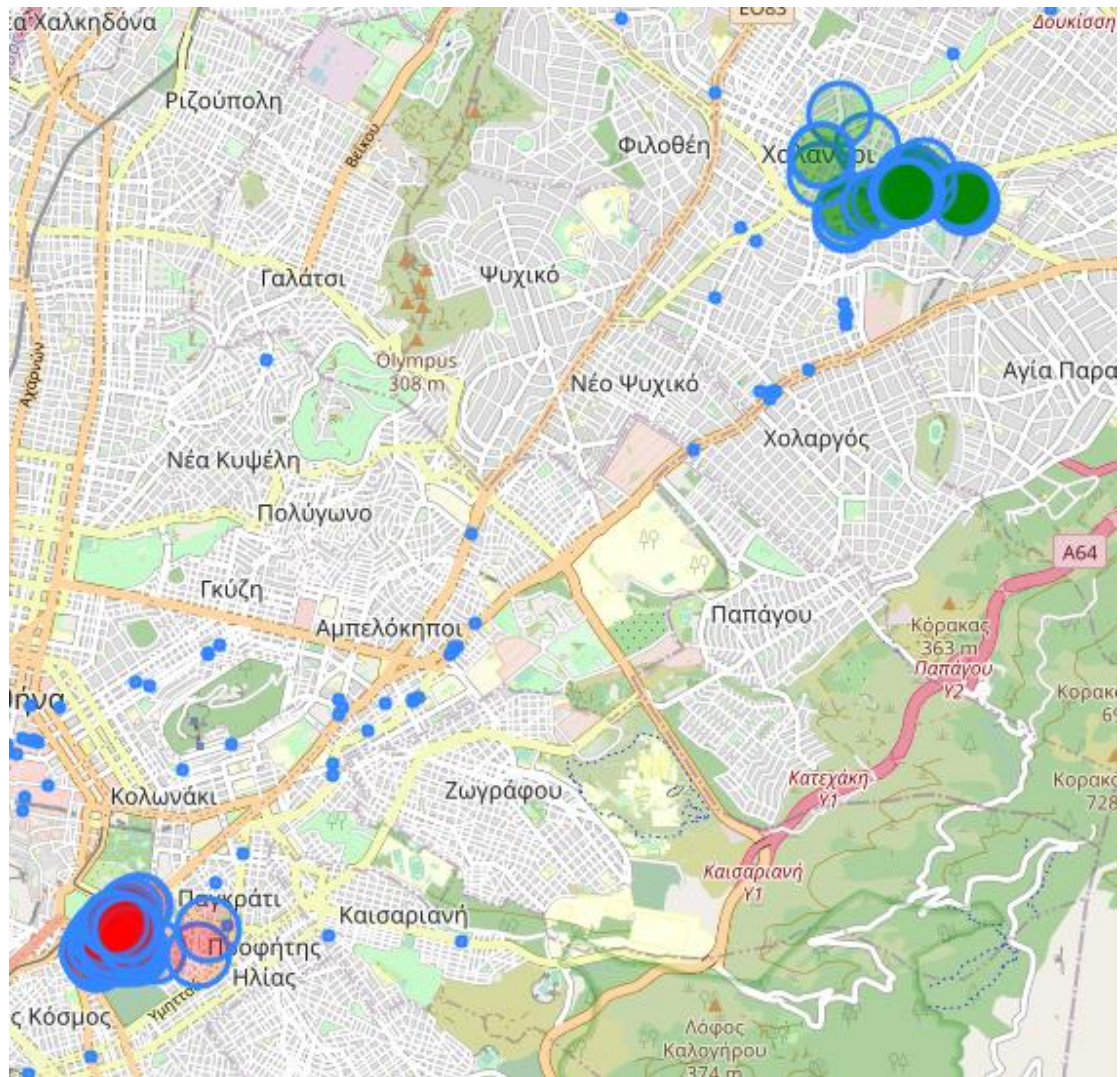
Διάγραμμα 70: Heatmap όλων των τοποθεσιών που έχει βρεθεί ο οδηγός 5



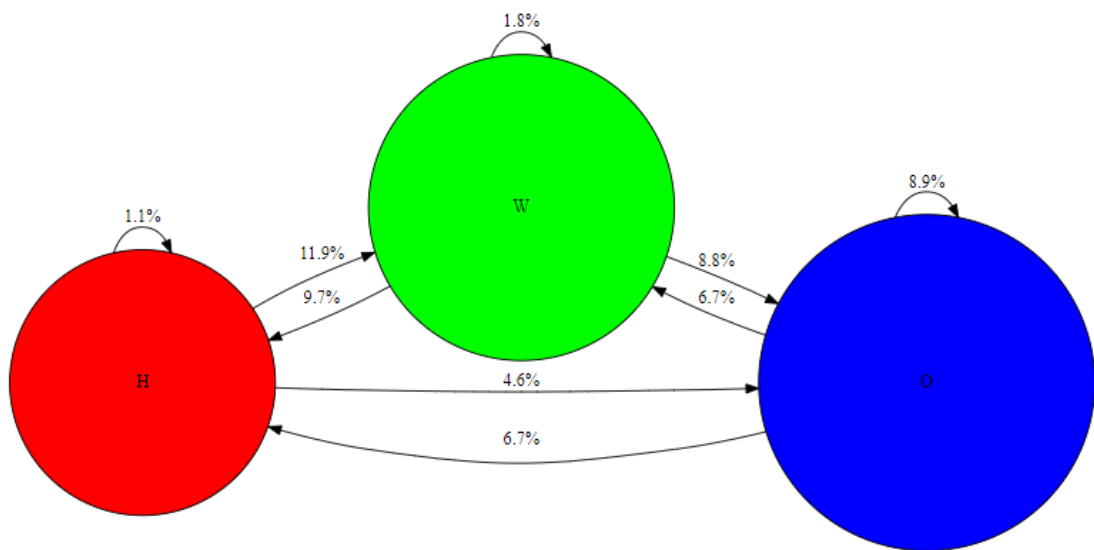
Διάγραμμα 71: Απεικόνιση όλων των ταξιδιών του οδηγού 5



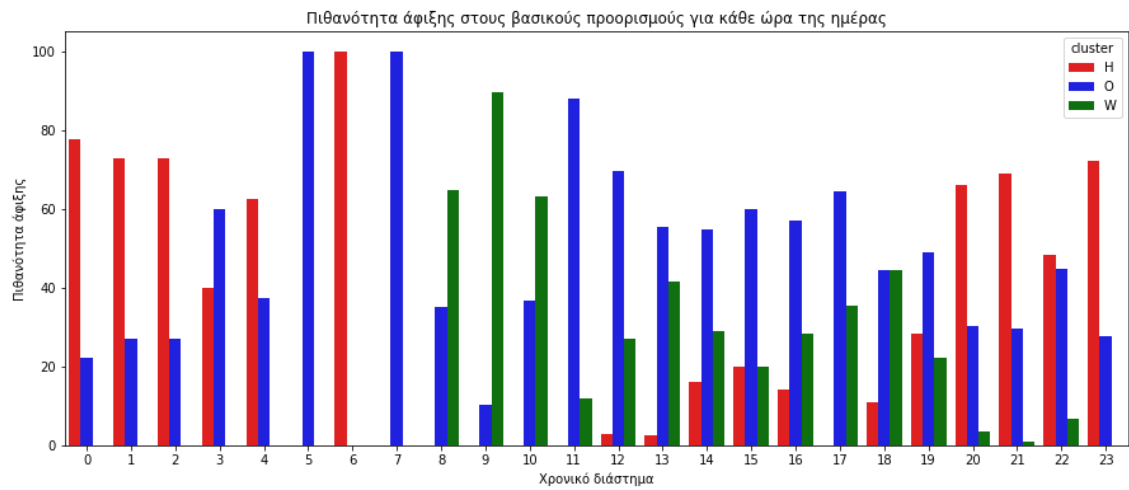
Διάγραμμα 72: Απεικόνιση μοτίβου μετακινήσεων ανά ημέρα για τον οδηγό



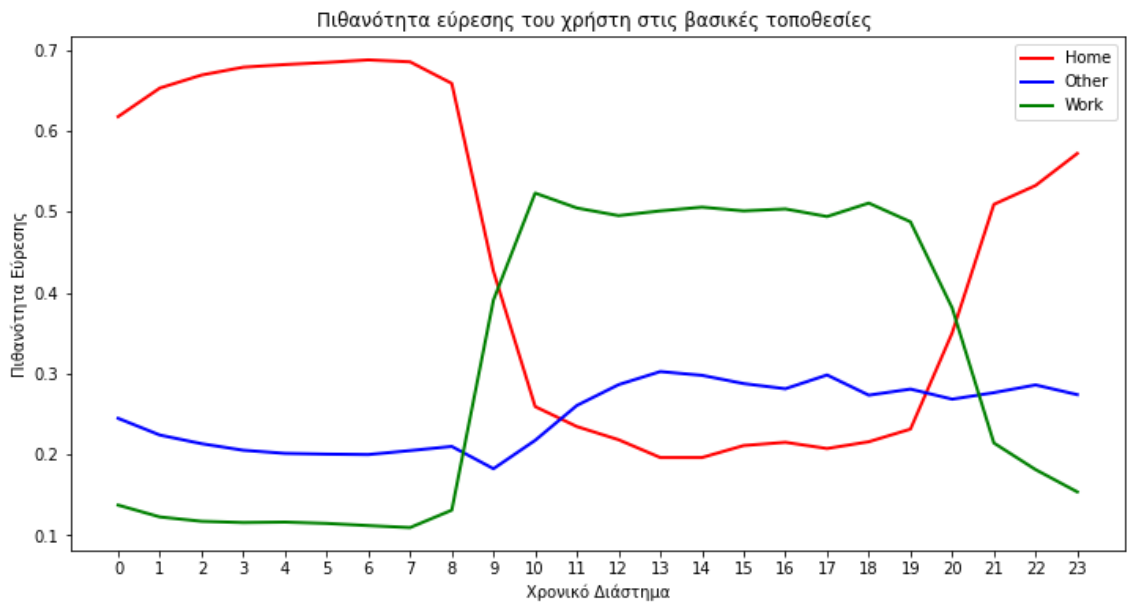
Διάγραμμα 73: Κατηγοριοποίηση προορισμών του οδηγού 5



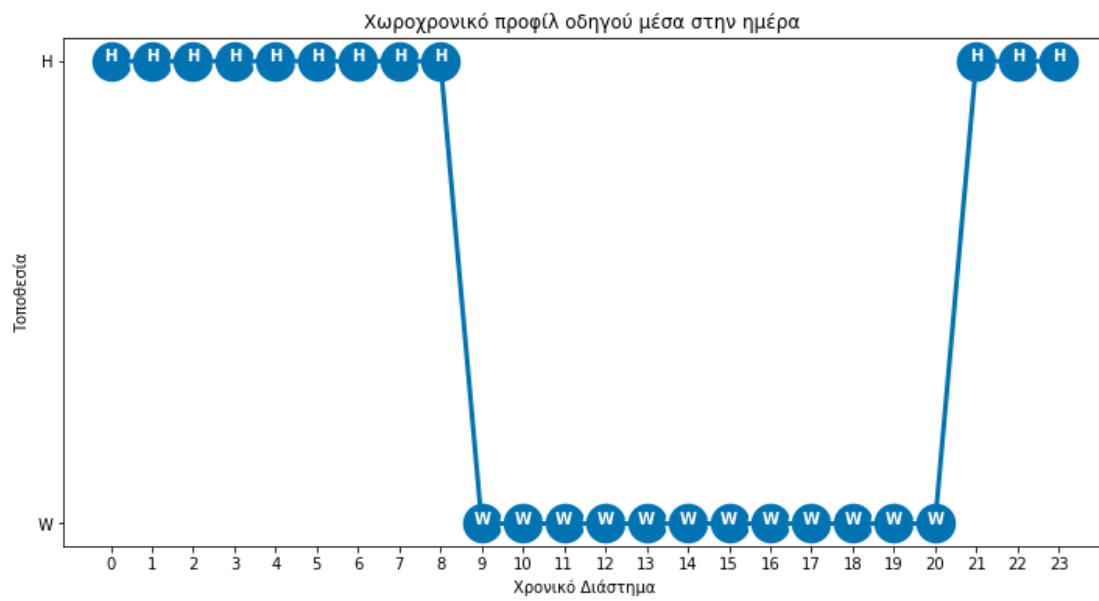
Διάγραμμα 74: Μοτίβο κινητικότητας του οδηγού 5



Διάγραμμα 75: Πιθανότητα άφιξης στους βασικούς προορισμούς για κάθε ώρα της ημέρας για τον οδηγό 5



Διάγραμμα 76: Πιθανότητα εύρεσης του οδηγού 5 στις βασικές τοποθεσίες για κάθε ώρα της ημέρας



Διάγραμμα 77: Χαρακτηριστική αλυσίδα δραστηριοτήτων του οδηγού 5

ΠΑΡΑΡΤΗΜΑ Β

ΟΔΗΓΟΣ 2

Σπίτι

	hour	duration
count	495.000000	495.000000
mean	14.888889	9.846465
std	5.760203	11.780751
min	0.000000	0.000000
25%	10.000000	2.000000
50%	16.000000	10.000000
75%	19.000000	13.000000
max	23.000000	154.000000

Δουλειά

	hour	duration
count	318.000000	318.000000
mean	11.955975	4.984277
std	4.474034	5.924052
min	7.000000	0.000000
25%	8.000000	1.000000
50%	10.000000	4.000000
75%	16.000000	8.000000
max	23.000000	58.000000

Προσωπική Δραστηριότητα '1'

	hour	duration
count	102.000000	102.000000
mean	12.205882	2.009804
std	3.973522	4.419791
min	7.000000	0.000000
25%	8.000000	0.000000
50%	12.500000	0.000000
75%	15.000000	1.750000
max	21.000000	32.000000

Προσωπική Δραστηριότητα '3'

	hour	duration
count	84.000000	84.000000
mean	8.678571	1.523810
std	3.955051	6.388009
min	7.000000	0.000000
25%	7.000000	0.000000
50%	7.000000	0.000000
75%	7.000000	0.000000
max	19.000000	48.000000

Υπόλοιπες Τοποθεσίες

	hour	duration
count	466.000000	466.000000
mean	14.806867	5.388412
std	4.268569	14.086881
min	1.000000	0.000000
25%	11.000000	0.000000
50%	16.000000	2.000000
75%	18.000000	4.000000
max	23.000000	202.000000

Πίνακας 10: Στατιστικά ώρας και διάρκειας των βασικών δραστηριοτήτων του οδηγού 2

ΟΔΗΓΟΣ 3

Σπίτι

	hour	duration
count	484.000000	484.000000
mean	17.518595	5.909091
std	5.116618	5.702057
min	0.000000	0.000000
25%	17.000000	1.000000
50%	19.000000	3.000000
75%	21.000000	10.000000
max	23.000000	34.000000

Δουλειά

	hour	duration
count	162.000000	162.000000
mean	8.432099	8.58642
std	1.439792	2.34228
min	8.000000	0.000000
25%	8.000000	9.000000
50%	8.000000	9.000000
75%	8.000000	9.000000
max	17.000000	24.000000

Προσωπική Δραστηριότητα '1'

	hour	duration
count	124.000000	124.000000
mean	17.161290	2.645161
std	3.843244	5.386212
min	0.000000	0.000000
25%	16.000000	0.000000
50%	18.000000	0.000000
75%	20.000000	2.000000
max	23.000000	32.000000

Υπόλοιπες Τοποθεσίες

	hour	duration
count	475.000000	475.000000
mean	16.141053	2.962105
std	4.522226	5.417451
min	0.000000	0.000000
25%	13.000000	0.000000
50%	18.000000	1.000000
75%	20.000000	3.000000
max	23.000000	44.000000

Πίνακας 11: Στατιστικά ώρας και διάρκειας των βασικών δραστηριοτήτων του οδηγού 3

ΟΔΗΓΟΣ 4

Σπίτι

	hour	duration
count	276.000000	276.000000
mean	16.282609	9.054348
std	6.384059	6.913540
min	0.000000	0.000000
25%	15.000000	2.000000
50%	18.500000	11.000000
75%	20.000000	14.000000
max	23.000000	63.000000

Δουλειά

	hour	duration
count	149.000000	149.000000
mean	10.523490	6.637584
std	2.647242	2.802544
min	2.000000	0.000000
25%	9.000000	4.000000
50%	9.000000	8.000000
75%	11.000000	9.000000
max	21.000000	13.000000

Προσωπική Δραστηριότητα '0'

	hour	duration
count	85.000000	85.000000
mean	9.952941	0.152941
std	3.027280	0.450023
min	8.000000	0.000000
25%	8.000000	0.000000
50%	9.000000	0.000000
75%	9.000000	0.000000
max	20.000000	2.000000

Προσωπική Δραστηριότητα '2'

	hour	duration
count	117.000000	117.000000
mean	15.658120	1.572650
std	3.826321	1.101067
min	7.000000	0.000000
25%	13.000000	0.000000
50%	18.000000	2.000000
75%	19.000000	2.000000
max	21.000000	4.000000

Προσωπική Δραστηριότητα '4'

	hour	duration
count	91.000000	91.000000
mean	16.846154	1.395604
std	3.025532	2.169891
min	7.000000	0.000000
25%	14.000000	0.000000
50%	18.000000	1.000000
75%	19.000000	2.000000
max	22.000000	15.000000

Υπόλοιπες Δραστηριότητες

	hour	duration
count	280.000000	280.000000
mean	15.517857	5.414286
std	5.016071	24.105945
min	0.000000	0.000000
25%	12.000000	0.000000
50%	17.000000	1.000000
75%	19.000000	3.000000
max	23.000000	300.000000

Πίνακας 12: Στατιστικά ώρας και διάρκειας των βασικών δραστηριοτήτων του οδηγού 4

ΟΔΗΓΟΣ 5

Σπίτι

	hour	duration
count	246.000000	246.000000
mean	15.947154	17.382114
std	8.044345	14.216128
min	0.000000	0.000000
25%	15.000000	10.000000
50%	20.000000	12.000000
75%	21.000000	18.000000
max	23.000000	83.000000

Δουλειά

	hour	duration
count	283.000000	283.000000
mean	11.091873	10.434629
std	3.433053	11.505401
min	8.000000	0.000000
25%	9.000000	4.000000
50%	9.000000	10.000000
75%	13.000000	11.000000
max	22.000000	88.000000

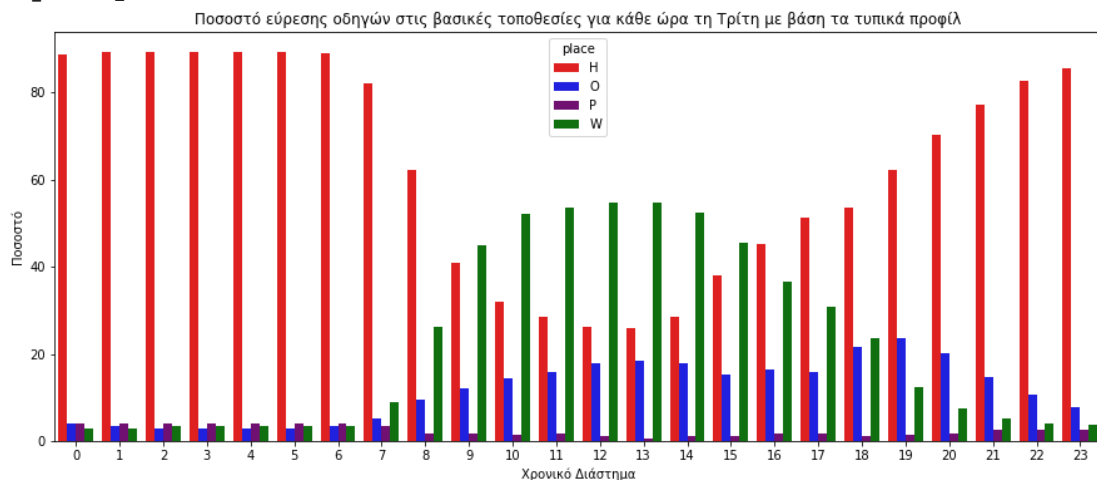
Υπόλοιπες Τοποθεσίες

	hour	duration
count	311.000000	311.000000
mean	14.524116	7.273312
std	5.492978	11.318522
min	0.000000	0.000000
25%	11.000000	2.000000
50%	15.000000	3.000000
75%	19.000000	8.000000
max	23.000000	91.000000

Πίνακας 13: Στατιστικά ώρας και διάρκειας των βασικών δραστηριοτήτων του οδηγού 5

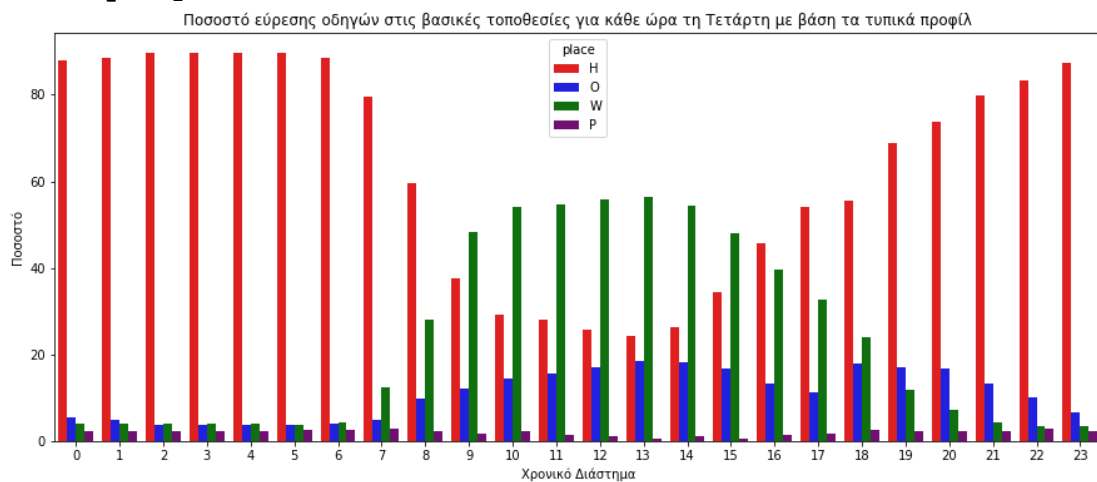
ΠΑΡΑΡΤΗΜΑ Γ

Τρίτη



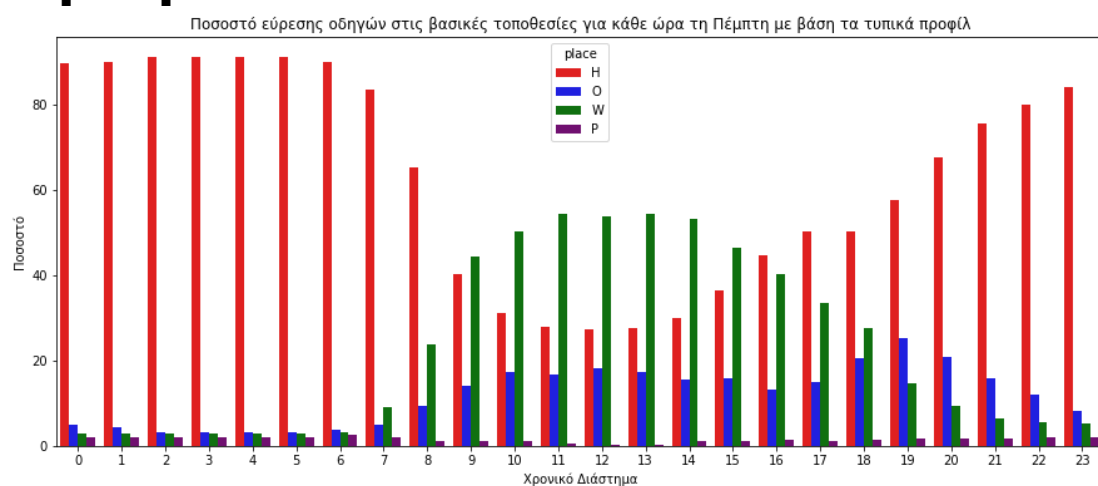
Διάγραμμα 78: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Τρίτης με βάση τα τυπικά προφίλ των οδηγών

Τετάρτη



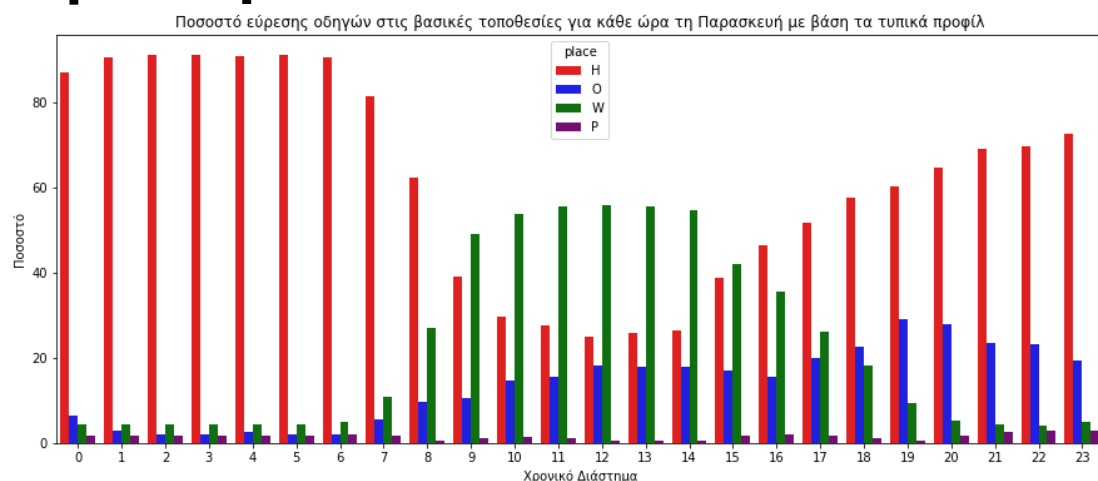
Διάγραμμα 79: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Τετάρτης με βάση τα τυπικά προφίλ των οδηγών

Πέμπτη



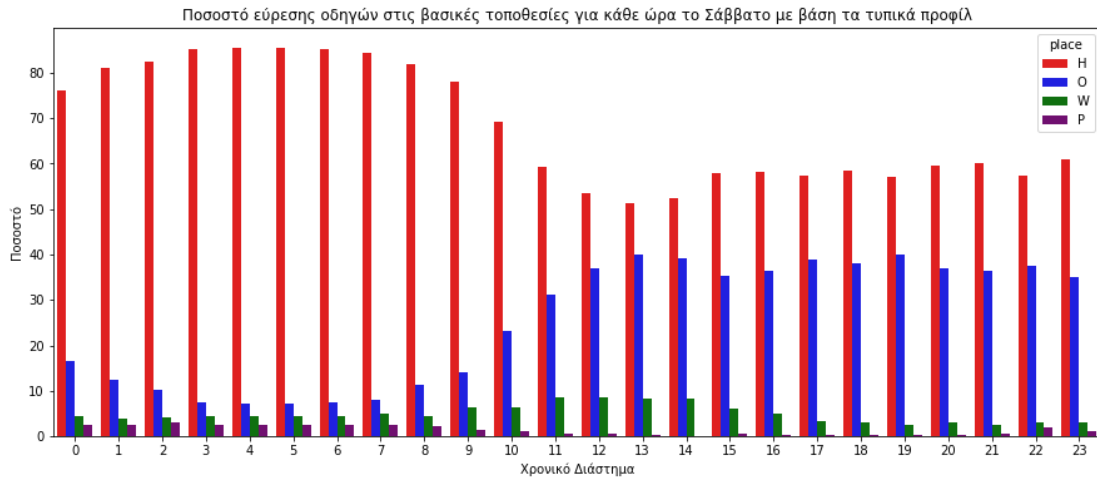
Διάγραμμα 80: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Πέμπτης με βάση τα τυπικά προφίλ των οδηγών

Παρασκευή



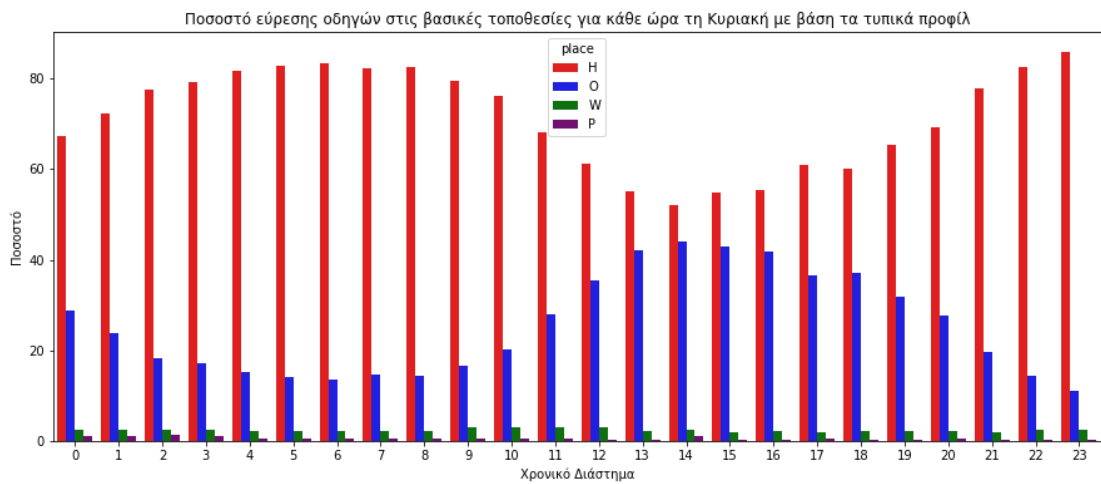
Διάγραμμα 81: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Παρασκευής με βάση τα τυπικά προφίλ των οδηγών

Σάββατο



Διάγραμμα 82: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα του Σαββάτου με βάση τα τυπικά προφίλ των οδηγών

Κυριακή



Διάγραμμα 83: Ποσοστό εύρεσης οδηγών στις βασικές τοποθεσίες για κάθε ώρα της Κυριακής με βάση τα τυπικά προφίλ των οδηγών