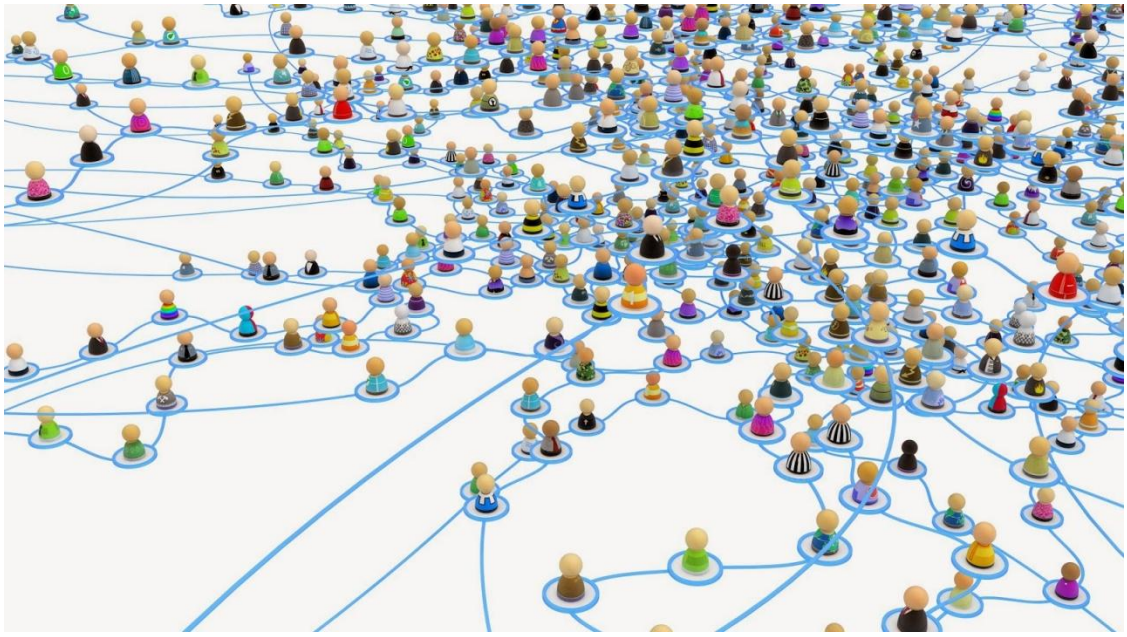
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΚΑΙ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΓΕΩΓΡΑΦΙΑΣ ΚΑΙ ΠΕΡΙΦΕΡΕΙΑΚΟΥ ΣΧΕΔΙΑΣΜΟΥ

Διπλωματική Εργασία

Πρότυπη διαδικτυακή εφαρμογή ανάκτησης πληροφοριών μέσω κοινωνικών δικτύων για τη χωροχρονική ανάλυση κυκλοφοριακών δεδομένων αστικών περιοχών σε περιβάλλον GIS.



Χανιωτάκης Πέτρος

Επιβλέπων: Γεώργιος Ν. Φώτης, Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2016

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΡΕΤΗΡΙΟ ΓΡΑΦΗΜΑΤΩΝ IV
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ IV
ΕΥΡΕΤΗΡΙΟ ΧΑΡΤΩΝ V
ΕΥΧΑΡΙΣΤΙΕΣ VI
ΠΕΡΙΛΗΨΗ VII
ABSTRACT VIII
ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ 9
1.1 Στόχευση της εργασίας 10
1.2 Προσέγγιση του προβλήματος 10
ΚΕΦΑΛΑΙΟ 2: ΓΕΝΙΚΑ ΣΤΟΙΧΕΙΑ 12
2.1 Κυκλοφοριακή συμφόρηση ένα σύγχρονο μεγάλο πρόβλημα 12
2.2 Συλλογή κυκλοφοριακών δεδομένων μέχρι σήμερα 15
2.3 Κυκλοφοριακά δεδομένα σε πραγματικό χρόνο 17
2.4 Twitter 18
2.5 Κοινωνικά δίκτυα και Crowdsourcing εφαρμογές 20
2.6 Προηγούμενες μελέτες σχετικές με τα κοινωνικά δίκτυα 24
2.7 Προηγούμενες μελέτες σχετικές με τα κοινωνικά δίκτυα και την κυκλοφοριακή συμφόρηση 25
2.8 Διαφορετικές αναλύσεις των δεδομένων 27
2.9 Χρησιμότητα της παρούσας εργασίας 29
ΚΕΦΑΛΑΙΟ 3: ΔΙΑΔΙΚΤΥΑΚΗ ΕΦΑΡΜΟΓΗ 30
3.1 Δεδομένα 30
3.1.1 Ανάκτηση των δεδομένων 30
3.1.2 Πηγή των δεδομένων 32
3.2 Διαδικτυακή Εφαρμογή 34
3.2.1 Διαδικασία συλλογής μηνυμάτων 34
3.2.2 Αρχική βάση δεδομένων 36
3.2.3 Χαρτογράφηση σε πραγματικό χρόνο 41
3.2.4 Αξιολόγηση των δεδομένων σε πραγματικό χρόνο 45
3.2.5 Εκκαθάριση των δεδομένων 45

ΚΕΦΑΛΑΙΟ 4: ΕΚΤΟΣ ΣΥΝΔΕΣΗΣ ΑΝΑΛΥΣΗ 47
4.1 Διαχωρισμός των δεδομένων 50
4.2 Χρονική Ανάλυση 51
4.2.1 Επίπεδο συμφόρησης σε σχέση με τον χρόνο, εξαγωγή ώρας αιχμής 51
4.2.2 Υπογραφή της συμφόρησης 65
4.3 Χωρική ανάλυση 75
4.3.1 Αποτύπωση της συμφόρησης σε επίπεδο οδών 75
4.3.2 Υπολογισμός Χωρικής Αυτοσυσχέτισης 86
4.3.3 Αποτύπωση τροχαίων ατυχημάτων 98
ΚΕΦΑΛΑΙΟ 5: ΣΥΝΟΨΗ 105
5.1 Συμπεράσματα 105
5.2 Λογισμικά που χρησιμοποιήθηκαν 106
5.3 Προτάσεις για περαιτέρω μελέτη 107
Βιβλιογραφία-Παραπομπές 108

ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

Εικόνα 1: Έκθεση εταιρίας INRIX 14
Εικόνα 2: Μελλοντικά οικονομικά κόστη της κυκλοφοριακής συμφόρησης σύμφωνα με την εταιρία INRIX 14
Εικόνα 3: Πλήθος ενεργών χρηστών του Twitter ανά χρονολογία 18
Εικόνα 4: Κόμβος συστήματος ασύρματων αισθητήρων 23
Εικόνα 5: Εκτίμηση εμφάνισης της λέξης κλειδί "traffic" καθημερινά στο Twitter 34
Εικόνα 6 : Η βάση δεομένων που δημιουργήθηκε για τις ανάγκες της εργασίας 37
Εικόνα 7: Φωτογραφία χρήστη 40
Εικόνα 8: Φωτογραφία χρήστη 40
Εικόνα 9: Περιβάλλον διαδικτυακής εφαρμογής με αποτυπωμένα 9000 μηνύματα 42
Εικόνα 10: Περιβάλλον διαδικτυακής εφαρμογής με απεικόνιση μηνυμάτων σε μεγάλη κλίμακα 43
Εικόνα 11: Η χωρική ανάλυση των πληροφοριών φτάνει σε επίπεδο λωρίδας 44
Εικόνα 12: Εκ των υστέρων επεξεργασία για την βέλτιστη απόδοση της συμφόρησης 77
Εικόνα 13: Αποτύπωση τροχαίων ατυχημάτων, Νέα Υόρκη 99

ΕΥΡΕΤΗΡΙΟ ΓΡΑΦΗΜΑΤΩΝ

Γράφημα 1: Κατανομή συμφόρησης ανάλογα την ώρα, Dallas 54
Γράφημα 2: Κατανομή συμφόρησης ανάλογα την ημέρα, Dallas 54
Γράφημα 3: Κατανομή συμφόρησης ανάλογα την ώρα, Houston 55
Γράφημα 4: Κατανομή συμφόρησης ανάλογα την ημέρα, Houston 55
Γράφημα 5: Κατανομή συμφόρησης ανάλογα την ώρα, New York 56
Γράφημα 6: Κατανομή συμφόρησης ανάλογα την ημέρα, New York 56
Γράφημα 7: Κατανομή συμφόρησης ανάλογα την ώρα, Birmingham 57
Γράφημα 8: Κατανομή συμφόρησης ανάλογα την ημέρα, Birmingham 57
Γράφημα 9: Κατανομή συμφόρησης ανάλογα την ώρα, Philadelphia 58
Γράφημα 10: Κατανομή συμφόρησης ανάλογα την ημέρα, Philadelphia 58
Γράφημα 11: Κατανομή συμφόρησης ανάλογα την ώρα, Nashville 59
Γράφημα 12: Κατανομή συμφόρησης ανάλογα την ημέρα, Nashville 59
Γράφημα 13: Κατανομή συμφόρησης ανάλογα την ώρα, Indianapolis 60
Γράφημα 14: Κατανομή συμφόρησης ανάλογα την ημέρα, Indianapolis 60
Γράφημα 15: Κατανομή συμφόρησης ανάλογα την ώρα, El Paso 61
Γράφημα 16: Κατανομή συμφόρησης ανάλογα την ημέρα, El Paso 61
Γράφημα 17: Κατανομή συμφόρησης ανάλογα την ώρα, Columbus 62
Γράφημα 18: Κατανομή συμφόρησης ανάλογα την ημέρα, Columbus 62
Γράφημα 19: Κατανομή συμφόρησης ανάλογα την ώρα, Sacramento 63
Γράφημα 20: Κατανομή συμφόρησης ανάλογα την ημέρα, Sacramento 63

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας 1: Κόστη εγκατάστασης και συντήρησης συμβατικών αισθητήρων Πηγή: Road Traffic Data, collection methods and applications, Guillaume Leduc, 2008 16
Πίνακας 2: Μηνύματα από ιδιωτικούς λογαριασμούς 38
Πίνακας 3: Μηνύματα από επαγγελματικούς λογαριασμούς 39
Πίνακας 4:Αξιολόγηση των πληροφοριών σύμφωνα με το Google Traffic 45
Πίνακας 5: Εύρεση μηνυμάτων που δεν είναι χρήσιμα στην παρούσα εργασία 46
Πίνακας 6: Σύνθετες λέξεις που δεν αναφέρονται σε κυκλοφοριακή συμφόρηση 47
Πίνακας 7: Πλήθος μηνυμάτων σε κάθε πόλη 51
Πίνακας 8: Σύγκριση ώρας αιχμής 20 πόλεων με στοιχεία από την εταιρία Tomtom 64
Πίνακας 9: Οδοί με το μεγαλύτερο πρόβλημα συμφόρησης στην Νέα Υόρκη, σύμφωνα με την εταιρία INRIX Πηγή: inrix.com 81
Πίνακας 10: Οι οδοί με την μεγαλύτερη συμφόρηση σύμφωνα με τον συντελεστή που υπολογίστηκε 82
Πίνακας 11: Ολικός δείκτης Moran's I 90
Πίνακας 12: Κατάταξη μητροπολιτικών περιοχών ανάλογα με το επίπεδο συμφόρησης 97
Πίνακας 13: Οδοί με τα περισσότερα ατυχήματα στην Νέα Υόρκη 105

ΕΥΡΕΤΗΡΙΟ ΧΑΡΤΩΝ

Χάρτης 1: Στιγμιότυπο συμφόρησης 12 am, Houston....66
Χάρτης 2 :Στιγμιότυπο συμφόρησης 6 am, Houston....67
Χάρτης 3: Στιγμιότυπο συμφόρησης 9 am, Houston....68
Χάρτης 4: Στιγμιότυπο συμφόρησης 12 pm, Houston 69
Χάρτης 5: Στιγμιότυπο συμφόρησης 2 pm, Houston 70
Χάρτης 6: Στιγμιότυπο συμφόρησης 4 pm, Houston 71
Χάρτης 7: Στιγμιότυπο συμφόρησης 6 pm, Houston 72
Χάρτης 8 :Στιγμιότυπο συμφόρησης 8 pm, Houston 73
Χάρτης 9: Στιγμιότυπο συμφόρησης 10pm, Houston 74
Χάρτης 10: Απεικόνιση συντελεστή κυκλοφοριακής συμφόρησης, Νέα Υόρκη 79
Χάρτης 11: Αποτύπωση των οδών με το μεγαλύτερο πρόβλημα συμφόρησης σύμφωνα με των εταιρία INRIX, Νέα Υόρκη, 2015 80
Χάρτης 12: Μέση συμφόρηση σύμφωνα με το Google Traffic, 2016 Πηγή: maps.google.com 84
Χάρτης 13: Μέση Ημερήσια Κυκλοφορία, Νέα Υόρκη, 2014 Πηγή: dot.ny.gov/tdv 85
Χάρτης 14: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Philadelphia 92
Χάρτης 15: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Dallas....92
Χάρτης 16: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Houston 93
Χάρτης 17: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Seattle 93
Χάρτης 18: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, New York .94
Χάρτης 19: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Columbus..94
Χάρτης 20: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Detroit....95
Χάρτης 21: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Birmingham 95
Χάρτης 22: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Indianapolis 96
Χάρτης 23: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Nashville...96
Χάρτης 24: Απεικόνιση τροχαίων ατυχημάτων, New York-Newark....100
Χάρτης 25: Απεικόνιση τροχαίων ατυχημάτων Dallas-Fort Worth 101
Χάρτης 26: Αποτύπωση οδών στους οποίους εμφανίζονται τα περισσότερα ατυχήματα, Νέα Υόρκη 103

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω όλους όσους συνέβαλλαν στην επιτυχή ολοκλήρωση της παρούσας διπλωματικής εργασίας. Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Γιώργο Φώτη για την καθοδήγηση που μου παρείχε σε όλη τη διάρκεια εκπόνησης της εργασίας καθώς και για τις καίριες παρατηρήσεις και επισημάνσεις του σε όλα τα στάδια εκπόνησης της , όπως επίσης και τους κ. Ευθύμιο Μπακογιάννη και κ. Θάνο Βλαστό, μέλη της τριμελούς επιτροπής.

Τέλος, θα ήθελα να ευχαριστήσω όλους τους φίλους μου και την οικογένειά μου για την αμέριστη συμπαράσταση τους κατά την διάρκεια της φοίτησής μου.

ΠΕΡΙΛΗΨΗ

Στην παρούσα εργασία προτείνεται η αξιοποίηση των κοινωνικών δικτύων ως εναλλακτική πηγή χωρικών δεδομένων πραγματικού χρόνου που σχετίζονται με την κυκλοφοριακή συμφόρηση. Αναπτύσσεται επιχειρηματολογία για το πόσο χρήσιμες είναι οι πληροφορίες από τα κοινωνικά δίκτυα για την λύση του κυκλοφοριακού προβλήματος τόσο για την επεξεργασία τους σε πραγματικό χρόνο, όσο και για εκ των υστέρων ανάλυση. Παρουσιάζεται διαδικτυακή εφαρμογή η οποία αφενός, συλλέγει μηνύματα από το κοινωνικό δίκτυο Twitter που αναφέρονται σε περιστατικά συμφόρησης και αφετέρου τα αποτυπώνει σε διαδικτυακό χάρτη, σε ιστοσελίδα που δημιουργήθηκε για τις ανάγκες της εργασίας. Η εφαρμογή που αναπτύχτηκε πιστεύεται ότι αποτελεί ένα χρήσιμο εργαλείο ενημέρωσης των πολιτών για την κίνηση στους δρόμους, παράμετρος βασική για την επίλυση του προβλήματος. Εκτός από την ανάλυση σε πραγματικό χρόνο τα μηνύματα αποθηκεύονταν σε βάση δεδομένων για περαιτέρω ανάλυση με σκοπό την ανάδειξη χωρικών και χρονικών παραμέτρων της συμφόρησης. Εντός περιόδου 70 ημερών ανακτήθηκαν 130.000 μηνύματα τα οποία προέρχονταν από την χώρα των Ηνωμένων Πολιτειών Αμερικής και στα οποία διενεργήθηκαν διαφόρων ειδών χωροχρονικές αναλύσεις σε περιβάλλον GIS. Τέλος, τα αποτελέσματα των αναλύσεων συγκρίνονται με επίσημα δεδομένα για να αναδειχθεί η αξιοπιστία τους και κατά επέκταση η καταλληλότητα των κοινωνικών δικτύων ως έγκυρη πηγή κυκλοφοριακών πληροφοριών.

Λέξεις κλειδιά: Twitter, Streaming API, κυκλοφοριακή συμφόρηση, κοινωνικά δίκτυα, Γεωγραφικά Συστήματα Πληροφοριών, συμμετοχική ανίχνευση, χρήστες ως αισθητήρες, ΗΠΑ

ABSTRACT

The following diploma thesis advocates the utilization of social networks as an alternative source of real time spatial data regarding congestion. The usefulness of social media information is argued, in regards to relieving congestion problems as well as real time processing and later analysis. A web application is presented that collects data from the social network Twitter, referring to traffic incidents and imprints them on a web map hosted in the page created for the needs of the thesis. The aforementioned application is considered to be a useful tool in the hands of the public and a vital factor in resolving the problem. In addition, the data collected are stored for further analysis aiming to point out the spatial and temporal aspects of congestion. Within 70 days, 130.000 Tweets from USA were retrieved and analyzed using various spatial and temporal procedures in GIS environment. In conclusion, the product of the analysis was compared to official data in order to test the reliability and suitability of social networks as a valid source of traffic information.

Key words: Twitter, Streaming API, traffic, congestion, social media, crowd sourcing, citizens as sensors, USA

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Το συνεχώς αυξανόμενο πρόβλημα της κυκλοφοριακής συμφόρησης στους δρόμους των μεγάλων πόλεων έχει τεράστια κόστη στις κοινωνίες τόσο στα ανεπτυγμένα όσο και στα αναπτυσσόμενα κράτη. Εκτός από τα οικονομικά κόστη το κυκλοφοριακό πρόβλημα επιδρά και στην ψυχολογία και στην παραγωγικότητα των πολιτών. Οι οδηγοί λόγο της αυξημένης κίνησης στους δρόμους δεν μπορούν να εκτιμήσουν το χρόνο ταξιδιού τους με αποτέλεσμα να μην μπορούν να προγραμματίσουν τις καθημερινές τους εργασίες. Ενδεικτικά, ο κάθε οδηγός στις ΗΠΑ το 2011 είχε να αντιμετωπίσει 38 ώρες καθυστέρησης στους δρόμους , αύξηση 138% σε σχέση με τις αντίστοιχες μετρήσεις του 1982 (TTI Urban Mobility Report) [1]. Σημαντικές επιπτώσεις σημειώνονται και στο φυσικό περιβάλλον καθώς σε συνθήκες μποτιλιαρίσματος καταναλώνονται άσκοπα τεράστιες ποσότητες καυσίμου.

Καθώς λοιπόν παρατηρείται μια συνεχόμενη αύξηση στους χρόνους παραμονής των οδηγών στους δρόμους οδηγούμαστε στο συμπέρασμα ότι η μέχρι τώρα διαχείριση του προβλήματος δεν επαρκεί και έτσι θα πρέπει να βρεθούν καινούργιες, εναλλακτικές λύσεις. Έχοντας ως δεδομένο ότι η επέκταση των υπαρχόντων υποδομών είναι πολυδάπανη διαδικασία η εναλλακτική επιλογή είναι η καλύτερη διαχείριση της κίνησης που ίσως αποδειχτεί εξίσου αποτελεσματική .Το πρώτο βήμα για την διαχείριση της κίνησης είναι η καταγραφή σε πραγματικό χρόνο της κατάστασης στους δρόμους, η οποία γίνεται σήμερα με την χρήση διαφόρων τεχνολογιών και εργαλείων όπως κάμερες κυκλοφορίας ή διαφόρων ειδών αισθητήρων. Αυτές οι συμβατικές μέθοδοι παρέχουν δεδομένα μεγάλης ακρίβειας τις περισσότερες φορές αλλά έχουν πολλούς περιορισμούς, όπως το κόστος εγκατάστασης και λειτουργίας αλλά το βασικό είναι ότι έχουν , εκ φύσεως , περιορισμένη κάλυψη του οδικού δικτύου. Άρα το πρόβλημα ανάγεται στο ότι πρέπει να βρεθούν νέες πηγές δεδομένων πραγματικού χρόνου ώστε να αναβαθμιστούν τα συστήματα διαχείρισης της κυκλοφορίας.

Συστήματα τα οποία θα μας παρέχουν σε πραγματικό χρόνο μεγάλο όγκο δεδομένων για την κίνηση θα είναι πολύ χρήσιμα τόσο για τις αρχές όσο και για τους πολίτες. Οι πολίτες για παράδειγμα θα μπορούν να ενημερώνονται άμεσα για την κίνηση , για το που υπάρχει μποτιλιάρισμα ή κάποιο τροχαίο δυστύχημα και να επιλέγουν αναλόγως την διαδρομή που θα ακολουθήσουν. Από την άλλη μεριά οι αρχές θα μπορούν να καταγράφουν συγκεντρωτικά σε πραγματικό χρόνο που υπάρχουν προβλήματα στους δρόμους και να επεμβαίνουν σε λιγότερο χρόνο από ότι γίνεται μέχρι σήμερα. Επίσης τα δεδομένα αυτά αν αποθηκευτούν μπορεί να γίνει στατιστική επεξεργασία ώστε να αποκρυπτογραφηθεί η συμπεριφορά των οχημάτων στο οδικό δίκτυο και εν τέλει να αποφασιστούν ποιες παρεμβάσεις , μεγάλης κλίμακας, πρέπει να γίνουν για να υπάρξει ομαλοποίηση στην κυκλοφορία.

1.1 Στόχευση της εργασίας

Στην παρούσα εργασία γίνεται προσπάθεια να προσεγγιστεί το πρόβλημα της διαχείρισης της κυκλοφοριακής συμφόρησης με ένα καινοτόμο τρόπο ο οποίος μπορεί να αποδειχτεί πολύ χρήσιμος και να διαδραματίσει καταλυτικό ρόλο στην επίλυση του προβλήματος . Στην εργασία χρησιμοποιούνται ως δεδομένα, μηνύματα τα οποία έχουν ανακτηθεί από το κοινωνικό δίκτυο Twitter σε πραγματικό χρόνο και αναφέρονται σε περιστατικά κυκλοφοριακής συμφόρησης . Μετά την ανάκτηση των δεδομένων προτείνονται διάφοροι τρόποι επεξεργασίας τους με αποτέλεσμα την παραγωγή χρήσιμων συμπερασμάτων τα οποία πιστεύεται ότι θα βοηθήσουν στην διαχείριση του προβλήματος.

Στόχος της εργασίας είναι η ανάδειξη των κοινωνικών δικτύων ως μια εναλλακτική πηγή χωρικών δεδομένων για την διαχείριση του κυκλοφοριακού προβλήματος, αλλά και την ανάδειξή τους γενικώς ως μια ανεξάντλητη, έγκυρη πηγή χωρικών δεδομένων. Έτσι εκτός από την παρουσίαση για το πως ανακτούνται και επεξεργάζονται τα δεδομένα στο συγκεκριμένο πρόβλημα γίνεται και επαλήθευση τους μέσω της σύγκρισής τους με επίσημα στοιχεία ώστε να διευκρινιστεί κατά πόσο ανταποκρίνονται στην πραγματικότητα και άρα κατά πόσο είναι αξιόπιστα. Για να το πετύχουμε αυτό χρησιμοποιούνται μόνο δεδομένα τα οποία αναφέρονται σε περιοχές που έχουμε επίσημα στοιχεία ώστε να μπορεί να είναι δυνατή η σύγκριση.

1.2 Προσέγγιση του προβλήματος

Καθημερινά αποστέλλονται 600 εκατομμύρια μηνύματα μέσω της υπηρεσίας του Twitter, από εδώ και στο εξής θα αναφέρονται ως tweets. Ένα ποσοστό από τα tweets εμπεριέχουν την πληροφορία της τοποθεσίας καθώς οι χρήστες πριν κοινοποιήσουν το μήνυμα ενεργοποιούν την υπηρεσία εύρεσης θέσης και έτσι κοινοποιούν εκτός από το μήνυμα και τις συντεταγμένες από όπου στάλθηκε. Άρα γίνεται αντιληπτό ότι καθημερινά κοινοποιείται μέσω της υπηρεσίας του Twitter ένας τεράστιος όγκος χωρικών δεδομένων τα οποία με την σωστή διαχείριση και επεξεργασία μπορούν να χρησιμεύσουν σε πολλές εφαρμογές και έρευνες. Τέτοιου είδους μηνύματα χρησιμοποιούνται και στην παρούσα εργασία καθώς έχει παρατηρηθεί και αποδεικνύεται παρακάτω ότι οι χρήστες όταν βρίσκονται σε συνθήκες έντονης κίνησης στους δρόμους ή σε μποτιλιάρισμα έχουν την τάση να μοιράζονται αυτή την δυσάρεστη τους εμπειρία με τους υπόλοιπους χρήστες της υπηρεσίας πρόθεσή τους είναι αφενός να εκφράσουν την δυσαρέσκειά τους , μέσω του γραπτού μηνύματος, και αφετέρου να προειδοποιήσουν τους υπόλοιπους χρήστες της κοινότητας ότι στην συγκεκριμένη θέση που βρίσκονται υπάρχει πολλή κίνηση στον δρόμο. Το δεύτερο επιτυγχάνεται με την ενεργοποίηση της υπηρεσίας εύρεσης θέσης με αποτέλεσμα το «καρφίτσωμα» του μηνύματος πάνω στον χάρτη.

Τα tweets που επιλέχθηκαν για την εργασία κατέβηκαν μέσω του Streaming API του Twitter. Η λέξη κλειδί που χρησιμοποιήθηκε για την εύρεση των μηνυμάτων είναι η λέξη «traffic». Επίσης έγινε επιλογή μόνο όσων μηνυμάτων ήταν «geotagged» δηλαδή εμπεριείχαν και χωρική πληροφορία. Το βασικό χαρακτηριστικό του Streaming API είναι ότι τα μηνύματα κατεβαίνουν σε πραγματικό χρόνο (real-time). Αυτό σημαίνει ότι ο χρόνος που μεσολαβεί από την στιγμή που κοινοποιείται ένα tweet έως την στιγμή που αποθηκεύεται στην βάση δεδομένων, που δημιουργήθηκε για τους σκοπούς της εργασίας, είναι της τάξης των λίγων δευτερολέπτων. Κατεβάζοντας λοιπόν tweets σε πραγματικό χρόνο μπορούμε με την κατάλληλη επεξεργασία, να έχουμε μια άποψη της κυκλοφορίας στους δρόμους αυτή τη στιγμή.

Στην παρούσα εργασία γίνονται πολλών ειδών αναλύσεις των δεδομένων που ανακτήθηκαν από το Twitter. Αρχικά παρουσιάζεται μια πλατφόρμα η οποία τοποθετεί αυτόματα τα tweets σε ένα παγκόσμιο χάρτη, ο χάρτης που έχει επιλεγεί είναι από μια υπηρεσία της Google η οποία μας δείχνει σε πραγματικό χρόνο την κίνηση στους δρόμους, με καλή ακρίβεια. Έχοντας ως υπόβαθρο την κίνηση όπως την δίνει η Google μπορούμε να κάνουμε οπτικά μια πρώτη σύγκριση για να δούμε κατά πόσο συνάδουν τα δεδομένα από το Twitter με αυτά της Google.

Η διαδικασία ανάκτησης των δεδομένων διήρκησε τρείς μήνες και στο διάστημα αυτό κατέβηκαν 150,000 tweets τα οποία εμπεριέχουν την λέξη κλειδί «traffic» και χωρική πληροφορία. Τα μηνύματα αποθηκεύτηκαν και στην συνέχεια έγινε διαχωρισμός τους σε αυτά που αναφέρονται σε κυκλοφοριακή κίνηση και σε αυτά που χρησιμοποιούν την λέξη «traffic» για να αναφερθούν σε κάτι διαφορετικό, όπως κίνηση χρηστών σε μια ιστοσελίδα. Έπειτα παρατηρήθηκε ότι τα μηνύματα κατά 80% προέρχονταν από τις Ηνωμένες Πολίτες Αμερικής, έτσι η έρευνα επικεντρώθηκε μόνο σε αυτή την χώρα για να γίνουν οι διάφορες αναλύσεις. Αφού επιλέχθηκαν τα δεδομένα που αφορούσαν μόνο τις ΗΠΑ έγινε διαχωρισμός των δεδομένων ανάλογα με την φύση των μηνυμάτων καθώς παρατηρήθηκε ότι πολλά μηνύματα αναφέρονταν σε κίνηση που δημιουργήθηκε από κάποιο αυτοκινητιστικό δυστύχημα. Έτσι απομονώθηκαν τα tweets που αναφέρονταν σε δυστύχημα ώστε από αυτή την βάση δεδομένων να προκύψουν στατιστικές πληροφορίες για τα δυστυχήματα στους δρόμους των διαφόρων πόλεων των ΗΠΑ.

Οι αναλύσεις που πραγματοποιήθηκαν χωρίζονται σε δύο κατηγορίες, τις χωρικές και τις χρονικές. Όσον αφορά τις χωρικές αναλύσεις:

- Αποτύπωση της κίνησης σύμφωνα με τα δεδομένα, στην πόλη της Νέας Υόρκης, δημιουργία χάρτη με τους δρόμους που εμφανίζουν την περισσότερη συμφόρηση, σύγκριση των δεδομένων από επίσημα στοιχεία
- Αποτύπωση των τροχαίων ατυχημάτων σε δύο μεγάλες πόλεις των ΗΠΑ, δημιουργία χάρτη με τους πιο επικίνδυνους δρόμους της Νέας Υόρκης και σύγκριση με τα επίσημα στοιχεία από τον δήμο της πόλης
- Μέτρηση της χωρικής αυτοσυσχέτισης, χρησιμοποιώντας τον δείκτη Moran I ,ανάμεσα σε περιοχές μεγάλων πόλεων που εμφανίζουν συμφόρηση και

εξαγωγή των ευρύτερων περιοχών όπου η συμφόρηση γίνεται εντονότερη. Κατάταξη 10 μεγάλων πόλεων ανάλογα με το πλήθος των παραπάνω ευρύτερων περιοχών και έλεγχος των αποτελεσμάτων από επίσημα στοιχεία.

Όσον αφορά τις χρονικές αναλύσεις έχουμε τα παρακάτω:

- Εξαγωγή διαγραμμάτων που υποδεικνύουν τη συμπεριφορά της κίνησης εντός 24ώρου και προσδιορισμός της ώρας αιχμής για 20 μητροπολιτικές περιοχές που διαθέτουν τα περισσότερα δεδομένα, σύγκριση μεταξύ τους και εξαγωγή συμπερασμάτων
- Δημιουργία χαρτών με τα σημεία όπου εντοπίζεται το μεγαλύτερο πρόβλημα συμφόρησης ανά ώρα στην πόλη του Houston με σκοπό την αποτύπωση της «υπογραφής» της συμφόρησης.

ΚΕΦΑΛΑΙΟ 2: ΓΕΝΙΚΑ ΣΤΟΙΧΕΙΑ

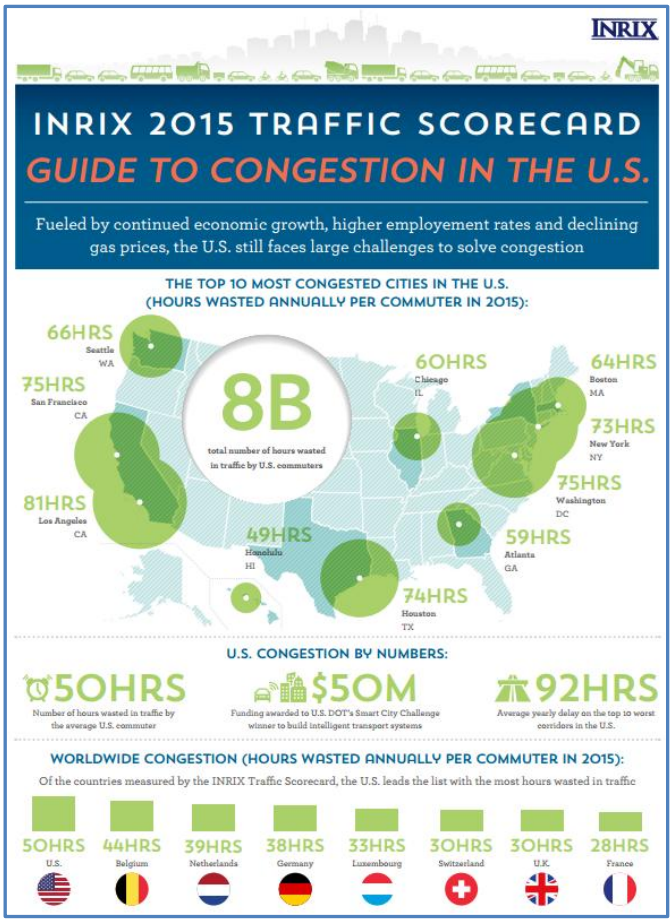
2.1 Κυκλοφοριακή συμφόρηση ένα σύγχρονο μεγάλο πρόβλημα

Η συμφόρηση είναι μια κυκλοφοριακή κατάσταση κατά την οποία η χρήση της οδού αυξάνεται τόσο ώστε τα διερχόμενα οχήματα να κινούνται με χαμηλές ταχύτητες. Αποτέλεσμα της συμφόρησης είναι η αύξηση των χρόνων διέλευσης από την οδό και σε πολλές περιπτώσεις η δημιουργία «ουρών» οχημάτων φαινόμενο της συμφόρησης εμφανίζεται όταν η ζήτηση της οδού αυξάνεται σε σημείο ώστε τα οχήματα να μην μπορούν να κινηθούν απρόσκοπτα οπότε αρχίζουν να αλληλεπιδρούν μεταξύ τους. Στην περίπτωση που ο αριθμός των οχημάτων υπερβεί την χωρητικότητα της οδού παρουσιάζεται το φαινόμενο γνωστό ως «μποτιλιάρισμα» και τα οχήματα παραμένουν ακινητοποιημένα για ορισμένο χρονικό διάστημα αύξηση της ζήτησης της οδού παρατηρείται συνήθως σε ώρες αιχμής οπότε γίνονται μαζικές μετακινήσεις ίσης κυκλοφοριακή συμφόρηση εμφανίζεται σε περιπτώσεις που μειωθεί η χωρητικότητα της οδού ,αποτέλεσμα μη περιοδικών περιστατικών όπως είναι οι εργασίες στον δρόμο, τροχαία ατυχήματα και έντονα καιρικά φαινόμενα (Ιμπριάλου, 2011) [2].

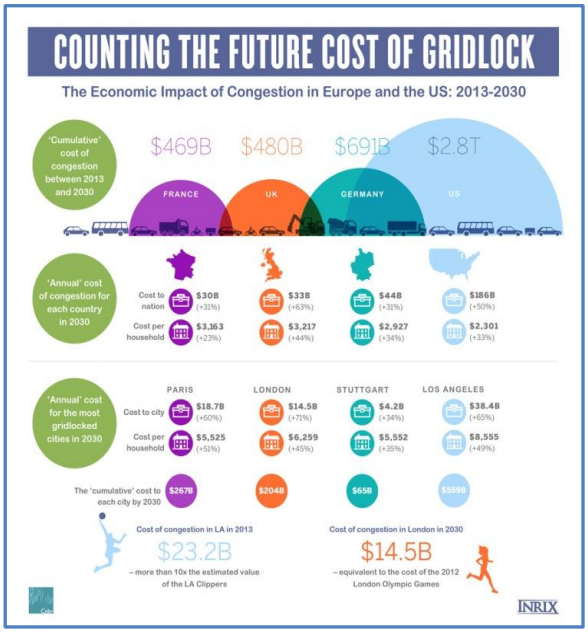
Η μη αποτελεσματική αντιμετώπιση της συμφόρησης στους δρόμους έχει μεγάλες επιπτώσεις στους πολίτες των μεγάλων πόλεων. Η πρωταρχική συνέπεια που επηρεάζει καθοριστικά την καθημερινότητα που κάθε πολίτη είναι ο μεγάλος χρόνος που καταναλώνει σε συνθήκες συμφόρησης στους δρόμους. Οι καθυστερήσεις στο οδικό δίκτυο είναι και ο τρόπος με τον οποίο ποσοτικοποιήται η κυκλοφοριακή συμφόρηση, η διαφορά χρόνου δηλαδή ανάμεσα στο χρόνο που διαρκεί μια συγκεκριμένη διαδρομή και στο χρόνο που θα διαρκούσε αν η ροή των οχημάτων ήταν ομαλή (Aftabuzzaman, 2007) [3]. Παρακάτω παρατίθεται ένα γράφημα (Εικόνα 1) που αναφέρεται στις χαμένες ώρες που έχει δαπανήσει ο κάθε πολίτης των ΗΠΑ σε συνθήκες κυκλοφοριακής συμφόρησης. Το γράφημα προέρχεται από την εταιρία INRIX [4] για την οποία θα γίνει αναφορά παρακάτω, και αναδεικνύει ότι ο κάθε

πολίτης των ΗΠΑ βρίσκεται στο δρόμο 50 ώρες παραπάνω, το χρόνο, λόγο της κυκλοφοριακής συμφόρησης που βιώνει σε καθημερινή βάση. Επίσης όπως προκύπτει από την συγκεκριμένη έρευνα η «πρωταθλήτρια» πόλη στις ΗΠΑ στην συμφόρηση είναι το Λος Αντζελες με 81 ώρες καθυστέρησης το χρόνο και το Σαν Φρανσίσκο και η Ουασιγινκτον με 75 ώρες. Τέλος να σημειωθεί ότι η έρευνα της INRIX η οποία πραγματοποιήθηκε με στοιχεία του 2015, κατατάσσει τις ΗΠΑ ως την χώρα με το μεγαλύτερο πρόβλημα συμφόρησης μεταξύ όλων των χωρών της Δύσης.

Το επόμενο μεγάλο πρόβλημα είναι οι οικονομικές συνέπειες που επιφέρει η συμφόρηση στους δρόμους. Λόγο της φύσης του προβλήματος το οποίο είναι πολυπαραγωντικό οι έρευνες δεν έχουν αποσαφηνίσει ακριβώς με ποιους μηχανισμούς η κακή κατάσταση στους δρόμους των πόλεων επηρεάζει αρνητικά την οικονομία αλλά αυτό που είναι σίγουρο είναι ότι χάνονται πολλά εκατομμύρια ευρώ στις μεγαλουπόλεις λόγω της συμφόρησης. Αρχικά, οι πολίτες όταν έχουν να αντιμετωπίσουν καθημερινά δυσμενείς συνθήκες στην διαδρομή για την δουλειά τους, γίνονται αντιπαραγωγικοί αφενός γιατί δεν μπορούν να προγραμματίσουν το ταξίδι τους και αφετέρου γιατί η όλη διαδικασία τους δημιουργεί άγχος, θυμό και γενικά αρνητικά συναισθήματα. Ένα μεγάλο μέρος των χρημάτων που δαπανούνται έχει να κάνει με τα καύσιμα που καταναλώνονται από τα εκατομμύρια οχήματα που βρίσκονται κολλημένα στην κίνηση. Ενδεικτικά, στις ΗΠΑ το 1994 το ετήσιο κόστος της κυκλοφοριακής συμφόρησης προσδιοριζόταν στα 48 δις δολάρια, από πρόσφατη έρευνα όμως της εταιρίας INRIX προκύπτει ότι το αντίστοιχο ποσό για το 2013 ανέρχεται σε 124 δις ενώ προβλέπει ότι το 2030 οι ΗΠΑ θα δαπανούν κάθε χρόνο 186 δις δολάρια, δηλαδή περίπου το 1% του ΑΕΠ της χώρας. Εάν η κατάσταση δεν αλλάξει και οι προβλέψεις επιβεβαιωθούν οι ΗΠΑ μέχρι το 2030 θα έχει δαπανήσει σωρευτικά 2,8 δις δολάρια όπως παρουσιάζεται σε παρακάτω γράφημα (Εικόνα 2). Τέλος ένα από τα προβλήματα τα οποία δε λαμβάνονται υπόψη με σοβαρότητα από τις αρχές των χωρών είναι η μεγάλη ζημια που γίνεται στο φυσικό περιβάλλον. Τα οχήματα καταναλώνουν ασκόπως τεράστιες ποσότητες καυσίμου σε συνθήκες συμφόρησης με αποτέλεσμα να εκλύεται μεγάλος όγκος καυσαερίων που έχουν σοβαρές επιπτώσεις στο περιβάλλον.



Εικόνα 1: Έκθεση εταιρίας INRIX
Πηγή: inrix.com, 2015



Εικόνα 2: Μελλοντικά οικονομικά κόστη της κυκλοφοριακής συμφόρησης σύμφωνα με την εταιρία INRIX
Πηγή: inrix.com, 2015

Όλες αυτές οι δυσοίωνες προβλέψεις αναγκάζουν τις σύγχρονες κοινωνίες να στραφούν σε νέα μοντέλα διαχείρισης του προβλήματος της κίνησης τα οποία θα βασίζονται σε καινοτόμες ιδέες, αναγνωρίζοντας ότι τα μέχρι τώρα μέσα και μέτρα που έχουν χρησιμοποιηθεί κατά καιρούς δεν μπορούν να ανταπεξέλθουν ικανοποιητικά.

2.2 Συλλογή κυκλοφοριακών δεδομένων μέχρι σήμερα

Η συμφόρηση μπορεί να αντιμετωπιστεί αποτελεσματικά με δύο τρόπους, όταν θεωρούμε δεδομένο ότι η κίνηση γίνεται με ιδιωτικά μέσα και όχι με Μέσα Μαζικής Μεταφοράς. Είτε βελτιώνοντας τις υποδομές του οδικού δικτύου άρα και την χωρητικότητα του, είτε μειώνοντας τις ροές των οχημάτων μέσω της σωστής διαχείρισης της κυκλοφορίας και ενθάρρυνσης των οδηγών να επιλέγουν διαφορετικές διαδρομές. Για να υλοποιηθούν και οι δύο παραπάνω περιπτώσεις χρειάζεται να μελετηθεί το οδικό δίκτυο ως προς την ικανότητα του να ανταπεξέλθει στις ανάγκες των οδηγών. Η μελέτη του οδικού δικτύου προϋποθέτει ότι υπάρχουν αξιόπιστα και ικανοποιητικά σε όγκο δεδομένα της κίνησης στους δρόμους σε πραγματικό χρόνο. Τα δεδομένα αυτά , με την προϋπόθεση πάντα ότι είναι αξιόπιστα μπορούν να παροτρύνουν τους οδηγούς να επιλέξουν εναλλακτικές διαδρομές από αυτές που συνηθίζουν μειώνοντας έτσι την κυκλοφοριακή συμφόρηση αποτελεσματικά. Επίσης με την κατάλληλη επεξεργασία δίνουν στις αρχές πολύτιμες πληροφορίες για την συμπεριφορά της κίνησης στο οδικό δίκτυο η οποία αν συγκριθεί με τις συνθήκες που θα έπρεπε να επικρατούν, μπορεί να εξαχθεί πόρισμα για τις παρεμβάσεις που θα πρέπει να γίνουν. Η στατιστική επεξεργασία των δεδομένων μπορεί να υποδείξει την συμπεριφορά της ροής των οχημάτων σε συνάρτηση με το χρόνο, δηλαδή να εντοπίσει σε ποιον δρόμο ή κόμβο υπάρχει συμφόρηση ανάλογα με την ώρα εντός ενός εικοσιτετραώρου.

Για πολλά χρόνια λόγο της αναγκαιότητας για καλύτερη διαχείριση της κυκλοφοριακής κίνησης είχαν αναπτυχθεί διάφορα εργαλεία συλλογής κυκλοφοριακών δεδομένων, πολλά από τα οποία χρησιμοποιούνται ως σήμερα. Παρακάτω αναφέρονται μερικοί από τους παραδοσιακούς τρόπους συλλογής δεδομένων για την κίνηση.

- Πνευματικοί σωλήνες (Pneumatic road tubes)
- Πιεζοηλεκτρικοί αισθητήρες (Piezoelectric sensors)
- Μαγνητικοί βρογχοι(Magnetic loops)

Τα παραπάνω εργαλεία συλλογής δεδομένων τοποθετούνται πάνω στην οδό και ανιχνεύουν τα οχήματα καθώς διέρχονται από πάνω τους. Συλλέγουν διάφορα στοιχεία ανάλογα με το είδος τους ,όπως τον αριθμό των οχημάτων που εχει περάσει την ταχύτητα τους και το βάρος τους.

Επίσης τα τελευταία χρόνια έχουν αναπτυχθεί συστήματα για ασύρματο εντοπισμό των οχημάτων όπως είναι οι:

- Μαγνητικοί αισθητήρες (Passive magnetic)
- Ραντάρ μικροκυμάτων (Microwave radar)
- Ακουστικοί αισθητήρες (Ultrasonic and passive acoustic)
- Εικόνες και βιντεο από κάμερες κυκλοφορίας

Οι παραπάνω παραδοσιακοί τρόποι συλλογής δεδομένων είναι απαραίτητοι γιατί δίνουν πολύ αξιόπιστα αποτελέσματα αλλά έχουν πολλά μειονεκτήματα. Αρχικά λόγο της φύσης των αισθητήρων αυτών δεν έχουν μεγάλη κάλυψη του οδικού δικτύου με αποτέλεσμα τα δεδομένα που λαμβάνουμε να μην αρκούν ώστε να μελετηθεί η κυκλοφοριακή κίνηση. Επίσης η εγκατάσταση και η συντήρηση έχουν αυξημένα έξοδα όπως φαίνεται στον παρακάτω πίνακα. Τέλος όπως όλες οι συσκευές έτσι και τα συστήματα αυτά έχουν συγκεκριμένο χρόνο ζωής με μέσο όρο τα 10 χρόνια (Πίνακας 1).

Unit Cost Element	Lifetime (years)	Capital Cost ($1000)	Cost Date	O&M Cost ($1000)	Cost Date
Inductive Loop Surveillance on Corridor	5	3-8	2001	0.4-0.6	2005
Inductive Loop Surveillance at Intersection	5	8.6-15.3	2005	0.9-1.4	2005
Machine Vision Sensor on Corridor	10	21.7-29	2003	0.2-0.4	2003
Machine Vision Sensor at Intersection	10	16-25.5	2005	0.2-1	2005
Passive Acoustic Sensor on Corridor		3.7-8	2002	0.2-0.4	1998
Passive Acoustic Sensor at Intersection		5-15	2001	0.2-0.4	2002
Remote Traffic Microwave Sensor on Corridor	10	9-13	2005	0.1-0.58	2005
Remote Traffic Microwave Sensor at Intersection	10	18	2001	0.1	2001
Infrared Sensor Active		6-7.5	2000		
Infrared Sensor Passive		0.7-12	2002		
CCTV Video Camera	10	9-19	2005	1-2.3	2004
CCTV Video Camera Tower	20	4-12	2005		

Πίνακας 1: Κόστη εγκατάστασης και συντήρησης συμβατικών αισθητήρων
Πηγή: Road Traffic Data, collection methods and applications, Guillaume Leduc, 2008

2.3 Κυκλοφοριακά δεδομένα σε πραγματικό χρόνο

Τα σύγχρονα οδικά δίκτυα, ακολουθώντας τους έντονους ρυθμούς της καθημερινότητας των πολιτών, επιβάλλεται πλέον να διαθέτουν συστήματα διαχείρισης της κυκλοφορίας τα οποία επεξεργάζονται δεδομένα πραγματικού χρόνου. Ένα καλό παράδειγμα είναι η πόλη της Bangkok. Κατά την διάρκεια της ώρας αιχμής οι αρχές προσπαθούν να διαχειριστούν την κυκλοφορία των οχημάτων χρησιμοποιώντας αστυνομικούς οι οποίοι έχουν εκπαιδευτεί στην διαχείριση της κίνησης. Οι τροχονόμοι τοποθετούνται σε σημεία που υπάρχει κυκλοφοριακό πρόβλημα και χειροκίνητα χειρίζονται τους φωτεινούς σηματοδότες ανάλογα με την συμφόρηση της κάθε οδού χρησιμοποιώντας στην ουσία τα «δεδομένα» πραγματικού χρόνου που λαμβάνουν ως αυτόπτες μάρτυρες. Επίσης επικοινωνούν μεταξύ τους μέσω ασυρμάτου ώστε η διαχείριση της κίνησης να γίνεται με τον καλύτερο δυνατό τρόπο(Advances in Wireless Sensors and Sensor Networks) [5].Όπως γίνεται αντιληπτό από το παραπάνω παράδειγμα είναι αναγκαία η ύπαρξη πηγής δεδομένων για την κίνηση στους δρόμους πραγματικού χρόνου για την βέλτιστη διαχείριση κυκλοφοριακών προβλημάτων . Τα δεδομένα πρέπει να είναι αξιόπιστα και να συλλέγονται σε ένα σύστημα ώστε το πρόβλημα να επιλύεται συγκεντρωτικά σε όλο το οδικό δίκτυο και όχι σπασμωδικά σε κάθε οδό με κίνδυνο απλώς την μετατόπιση του προβλήματος στις παρακείμενες οδούς.

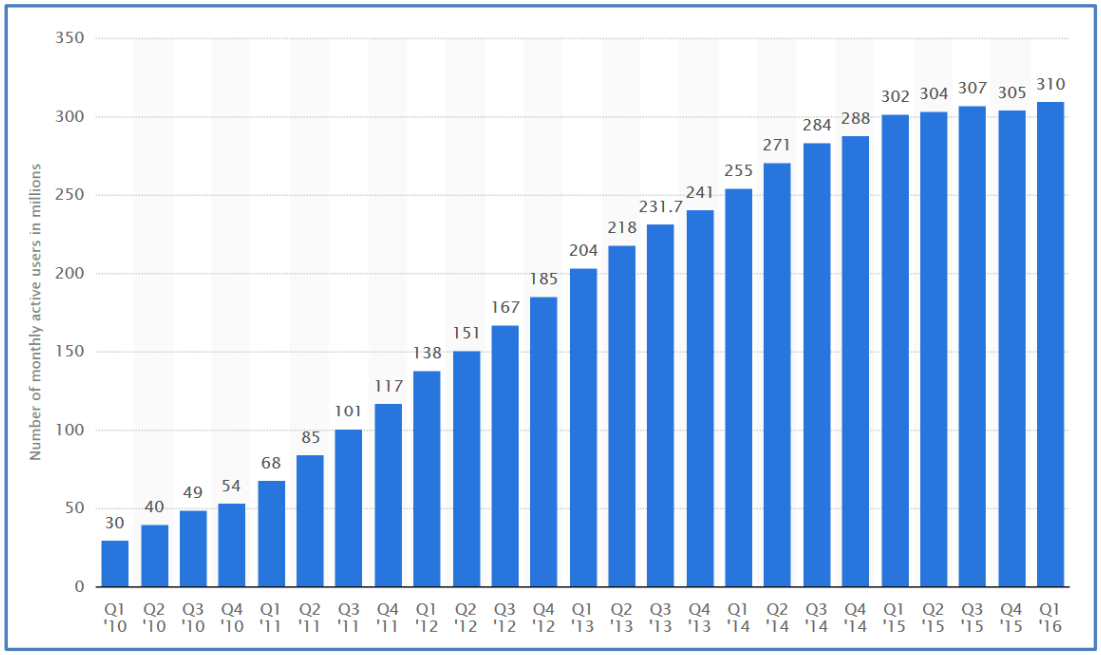
Ένα σύστημα που εκμεταλλεύεται πραγματικού χρόνου δεδομένα δεν θα ήταν μόνο χρήσιμο για την αρχές ώστε να ανταποκριθούν στο πρόβλημα της κυκλοφοριακής συμφόρησης. Εάν οι απλοί πολίτες έχουν πρόσβαση σε ένα αντίστοιχο σύστημα το οποίο θα εμφανίζει σε πραγματικό χρόνο που υπάρχει συμφόρηση θα μπορούν να επιλέγουν από μόνοι τους εναλλακτικές διαδρομές για το ταξίδι τους αποσυμφορίζοντας έτσι τις οδούς. Φυσικά ένα εξελιγμένο σύστημα θα μπορούσε να ενημερώνει τους οδηγούς εκτός για το που υπάρχει συμφόρηση, τον χρόνο που μπορεί να διαρκέσει ή την πιθανότητα να υπάρξει συμφόρηση στον κάθε δρόμο.Για να επιτευχθεί κάτι τέτοιο είναι απαραίτητη η στατιστική ανάλυση δεδομένων τα οποία πρέπει να συλλεχθούν για μεγάλο χρονικό διάστημα ώστε να είναι αξιόπιστα τα αποτελέσματα.

Όπως αναφέρθηκε παραπάνω οι συμβατικοί τρόποι συλλογής δεδομένων παρέχουν ακριβή στοιχεία για την κίνηση αλλά υστερούν σε όγκο δεδομένων και σε κόστος. Τα τελευταία χρόνια με την επανάσταση των έξυπνων τηλεφώνων έχουν αναπτυχθεί πολλές εφαρμογές οι οποίες προσφέρουν σε πραγματικό χρόνο στοιχεία για την κίνηση στους δρόμους χρησιμοποιώντας προφανώς δεδομένα πραγματικού χρόνου. Η εταιρία Tomtom μέσω της υπηρεσίας της HD Traffic παρέχει σε πραγματικό χρόνο στοιχεία για την κίνηση στους δρόμους στους πελάτες της. Τα δεδομένα προέρχονται από 80 εκατομμύρια ανώνυμες συσκευές τηλεφώνων και από ένα εκατομμύριο συσκευές της εταιρίας. Το σύστημα που έχει αναπτύξει η Tomtom εκμεταλλεύεται το

GPS των συσκευών για να προσδιορίσει την ταχύτητα των οχημάτων και την κατεύθυνση τους, με τον τρόπο αυτό υπολογίζει που υπάρχουν αδικαιολόγητες χαμηλές ταχύτητες άρα και καθυστερήσεις [6]. Τέλος σε αυτή την κατηγορία υπάρχει και η εταιρία INRIX [4] η οποία με παρόμοιο τρόπο προσφέρει στους χρήστες της διάφορα δεδομένα για την κίνηση ενώ κάθε χρόνο δημοσιοποιεί την «Traffic Scorecard» [7], μια έκθεση με στατιστικά στοιχεία για πολλές πόλεις του κόσμου τις οποίες της κατατάσσει ανάλογα με την μέση καθυστέρηση που βιώνουν οι πολίτες στους δρόμους. Σε πολλά σημεία της παρούσας εργασίας επικαλείται η έρευνα της INRIX για το 2015 ώστε να αξιολογηθούν τα συμπεράσματα της δικής μας έρευνας.

2.4 Twitter

Το Twitter (Τουίτερ) [8] είναι ένας δημοφιλής ιστοχώρος κοινωνικής δικτύωσης που επιτρέπει στους χρήστες του να στέλνουν και να διαβάζουν σύντομα μηνύματα (tweets) . Κάθε μήνυμα μπορεί να περιέχει έως 140 χαρακτήρες και ο κάθε χρήστης επιλέγει ποιους λογαριασμούς θέλει να ακολουθήσει (follow) ώστε να μπορεί να ενημερώνεται για τα νέα μηνύματα που ανεβαίνουν από τον συγκεκριμένο χρήστη. Η εφαρμογή συνδέει καθημερινά εκατομμύρια ανθρώπους και είναι προσβάσιμη και από υπολογιστή και από κινητό τηλέφωνο. Το Twitter έκανε την εμφάνισή του τον Ιούλιο του 2006 και από τότε ο αριθμός των χρηστών του αυξάνεται με γοργούς ρυθμούς. Όπως παρουσιάζεται και στο παρακάτω διάγραμμα (Εικόνα 3), ο αριθμός των ενεργών χρηστών κάθε μήνα στις αρχές του 2010 ήταν 30 εκατομμύρια ενώ ο αντίστοιχος το 2016 είναι 310 εκατομμύρια.



Εικόνα 3: Πλήθος ενεργών χρηστών του Twitter ανά χρονολογία
Πηγή: statista.com, 2016

Το Twitter, όπως αναφέρθηκε ήδη, συγκαταλέγεται στις εφαρμογές ανταλλαγής μικρών μηνυμάτων (micro-blogging services) [9].Υπάρχουν πολλές εφαρμογές σε αυτή την κατηγορία όπως το Tumblr [10], το Weibo [11], το Identi.ca [12] και πολλά

άλλα, μπορεί το κάθε ένα να έχει τα δικά του ιδιαίτερα χαρακτηριστικά, αλλά όλες αυτές οι εφαρμογές φτιάχτηκαν για να ανταλλάσουν μικρά μυνήματα οι χρήστες τα οποία μπορεί να είναι μυνήματα κειμένου,φωτογραφίες ή αρχεία βίντεο.Στην παρούσα εργασία επιλέχθηκε το Twitter επειδή είναι το ποιό δημοφιλές αλλά θα μπορούσε να εφαρμοστεί σε οποιαδήποτε αντίστοιχη υπηρεσία ή σε συνδιασμό αυτών.

Ένα βασικό χαρακτηριστικό που έχουν οι προαναφερθείσες εφαρμογές είναι ότι οι χρήστες τις χρησιμοποιούν για να περιγράψουν καταστάσεις που συμβαίνουν σε πραγματικό χρόνο(real-time).Σε άλλες εφαρμογές κοινωνικής δικτύωσης όπως το Facebook ή το Flickr οι χρήστες ανανεώνουν τις πληροφορίες του προφίλ τους το πολύ μια φορά την ημέρα εν αντιθέσει με το Twitter που η ενημέρωση των πληροφοριών μέσω των σύντομων μηνυμάτων είναι σε συνεχή βάση.Οι χρήστες του Twitter κοινοποιούν ,ανάλογα την χώρα,έως και 4 μυνήματα την ημέρα με αποτέλεσμα οι χρήστες που τους ακολουθούν να ξέρουν σε τακτά χρονικά διαστήματα τι σκέφτονται πως νιώθουν, με τι ασχολούνται ή τι βλέπουν γύρω τους. Εκ φύσεως λοιπόν η εφαρμογή του Twitter έχει την έννοια της αμεσότητας στην ενημέρωση του κοινού του κάθε χρήστη.

Ένα βασικό χαρακτηριστικό της υπηρεσίας του Twitter είναι η δυνατότητα που δίνει στους χρήστες να προσθέσουν στα μηνύματα που αναρτούν και την τοποθεσία τους. Η διαδικασία αυτή αναφέρεται ως geo-tagging και χρησιμοποιείται από μια πληθώρα εφαρμογών οι οποίες εκμεταλλεύονται τις υπηρεσίες εντοπισμού θέσης των έξυπνων κινητών τηλεφώνων των χρηστών. Το Flickr [13] για παράδειγμα δίνει την επιλογή στους χρήστες του να συμπεριλάβουν στις φωτογραφίες που ανεβάζουν στην υπηρεσία και τα στοιχεία της θέσης από όπου τραβήχτηκε η κάθε φωτογραφία σε μορφή συντεταγμένων οι οποίες εκτίθενται στα μετά-δεδομένα της κάθε εικόνας. Αντιστοίχως στο Twitter ο κάθε χρήστης μπορεί ,εάν το επιθυμεί, να ενεργοποιήσει την υπηρεσία εύρεσης θέσης του κινητού του τηλεφώνου και να συμπεριλάβει την τοποθεσία του στα tweets που κοινοποιεί. Εκτός από το GPS η εφαρμογή χρησιμοποιεί και άλλους τρόπους για να υπολογίσει την θέση του κάθε χρήστη. Το δίκτυο τηλεφωνίας είναι ένα από αυτούς καθώς γίνεται αυτόματος υπολογισμός της απόστασης του κάθε χρήστη από σταθερά σημεία εκπομπής σήματος, όπως είναι οι πυλώνες του δικτύου. Επίσης, η τοποθεσία μπορεί να υπολογιστεί από την απόσταση του κάθε χρήστη με σταθερά σημεία ασύρματης παροχής ίντερνετ (support.twitter.com) [14]. Από τους παραπάνω τρόπους εύρεσης της θέσης των συσκευών των χρηστών το GPS είναι το ποιο αξιόπιστο με ακρίβεια της τάξης των 10 μέτρων εάν η συσκευή έχει εντοπίσει τον κατάλληλο αριθμό δορυφόρων. Εντός των πόλεων όπου υπάρχουν πολλά «σκοτεινά» σημεία για τους δορυφόρους χρησιμοποιείται αυτόματα ο τριγωνισμός από σταθερά σημεία εκπομπής σήματος ή συνδυασμός των παραπάνω δύο μεθόδων. Ο τριγωνισμός από wireless access points είναι ο λιγότερο αξιόπιστος και πολλές φορές η ακρίβεια που δίνει είναι της τάξης των 500 μέτρων. Το Twitter δίνει την δυνατότητα στους χρήστες να χρησιμοποιούν και την ευρύτερη τοποθεσία τους όταν κοινοποιούν ένα μήνυμα δηλαδή το όνομα της

πόλης ή της γειτονιάς τους. Η παρούσα εργασία εκμεταλλεύεται τα geotagged tweets ώστε να συλλέξει χωρικές πληροφορίες. Τα μηνύματα που ανακτήθηκαν προέρχονταν μόνο από συσκευές Android ή ios γιατί τα συγκεκριμένα λειτουργικά συστήματα είναι εγκατεστημένα κατά κύριο λόγο σε φορητές συσκευές (έξυπνα τηλέφωνα), και έτσι διαχωρίστηκαν εύκολα τα μηνύματα που εμπεριείχαν τοποθεσία αλλά είχαν σταλεί από σταθερούς υπολογιστές. Επίσης χρησιμοποιήθηκαν μόνο τα tweets που εμπεριείχαν ακριβές ζεύγος συντεταγμένων και απορρίφθηκαν όσα μηνύματα αναφέρονταν στην ευρύτερη τοποθεσία.

2.5 Κοινωνικά δίκτυα και Crowdsourcing εφαρμογές

Τα κοινωνικά δίκτυα έχουν αλλάξει δραστικά την έννοια της ανταλλαγής και της διάδοσης των πληροφοριών (Kaplan και Haenlein 2010) [15]. Με την ενθάρρυνση των πολιτών να κοινοποιούν και να συνεισφέρουν δεδομένα για τους ίδιους ή για το τι γίνεται γύρω τους τους καθιστά μια πολύτιμη πηγή πληροφοριών που στο μέλλον ίσως αντικαταστήσει τις παραδοσιακές πηγές όπως είναι για παράδειγμα τα ειδησεογραφικά πρακτορεία.Υπό αυτή την έννοια τα κοινωνικά δίκτυα καθιστούν κάθε πολίτη δυνητικό λήπτη και συνεισφέρων πληροφοριών.

Η δύναμη των κοινωνικών δικτύων στην διάδοση σημαντικών πληροφοριών για την κοινή γνώμη έχει γίνει αντιληπτή τα τελευταία χρόνια στον μέγιστο βαθμό. Τον Ιούνιο του 2009 Ιρανοί διαδηλωτές χρησιμοποίησαν το Facebook το Twitter και το Youtube, σε συνδυασμό με το γεωγραφικό τους στίγμα, για να ενημερώσουν όλη την υφήλιο για τις δράσεις και τις κινητοποιήσεις τους , παρακάμπτοντας τους περιορισμούς της κυβέρνησης στην διάδοση των πληροφοριών. . Μέσω του Facebook δημιουργήθηκε το κίνημα της 6ης Απριλίου 2008 στην Αίγυπτο για να παροτρύνει την κοινή γνώμη σε απεργία για την βελτίωση του πολιτικού σκηνικού. Η συμμετοχή στο Facebook ήταν μεγάλη, έφθασε η ιδέα σε 1.000.000 ανθρώπους. Έτσι, η απεργία κατέληξε σε μαζική αντίδραση των πολιτών στο πολιτικό σκηνικό μέχρι ανατροπής του. Το Twitter έπαιξε σημαντικό ρόλο στην ενημέρωση. Η μετάδοση απευθείας από το δρόμο με τη χρήση κινητού τηλεφώνου προς τα κοινωνικά δίκτυα ήταν και ο διαμεσολαβητής με τα διεθνή ειδησεογραφικά πρακτορεία. Η σύνδεση του κινητού τηλεφώνου με το διαδίκτυο έδωσε σε μηδενικό χρόνο είδηση και εικόνα. Το επαναστατικό αυτό κύμα διαδηλώσεων και διαμαρτυριών στην Μέση Ανατολή και στη Βόρειο Αφρική ονομάστηκε Αραβική Άνοιξη (Μελίτα, 2012) [16].Τα παραδείγματα αυτά μας δείχνουν ότι τα κοινωνικά δίκτυα επιτρέπουν στους πολίτες να κοινοποιούν πληροφορίες σε κλίμακα που καμία άλλη τεχνολογία μέχρι σήμερα δεν έχει επιτύχει.

Τα τελευταία χρόνια γίνεται μεγάλη χρήση εφαρμογών που στηρίζονται σε δεδομένα που παραχωρούν εθελοντικά οι χρήστες του διαδικτύου (crowd sourcing).Τέτοιες εφαρμογές χρησιμοποιούνται ευρέως και ένα απλό παράδειγμα είναι η ιστοσελίδα downrightnow.com [17] στην οποία ο χρήστης δηλώνει ότι έχει προβλημα προσβασιμότητας σε κάποια μεγάλη ιστοσελίδα ώστε να ενημερώσει τους υπόλοιπους χρήστες.

Υπάρχουν όμως και πολλές εφαρμογές όπου οι χρήστες κοινοποιούν εθελοντικά γεωγραφικές πληροφορίες. Τέτοιου είδους υπηρεσία είναι το Waze στο οποίο οι χρήστες ενημερώνουν την κοινότητα χρηστών του Waze για το πού υπάρχει κλειστός δρόμος λόγω έργων ή για το πού υπάρχει αυτοκινητιστικό δυστύχημα άρα η ροή των οχημάτων είναι χαμηλή.Η ποιότητα των εξαγόμενων πληροφοριών έχει άμεση σχέση με τον αριθμό των χρηστών που λαμβάνουν μέρος στην όλη διαδικασία.Ο αριθμός των χρηστών που χρησιμοποιούν τέτοιες εφαρμογές ποικίλει από χώρα σε χώρα και από πόλη σε πόλη. Σύμφωνα με τα επίσημα στατιστικά στοιχεία του Waze, στο Παρίσι οι ενεργοί χρήστες τις εφαρμογής είναι περίπου 45,000 ενώ οι αντίστοιχοι στην Αθήνα δεν ξεπερνάνε τους 200. Είναι επόμενο λοιπόν η ποιότητα των δεδομένων στην Αθήνα να υπολείπεται κατά πολύ από την αντίστοιχη για την πόλη του Παρισιού.

Εκτός των παραπάνω εφαρμογών οι πολίτες καλούνται να συνεισφέρουν διάφορες άλλες πληροφορίες στην κοινότητα των χρηστών. Στην ιστοσελίδα FixMyStreet [18] οι χρήστες καλούνται να αναφέρουν προβλήματα στο οδικό δίκτυο όπως σημεία με κατεστραμμένο οδόστρωμα χρησιμοποιώντας τα έξυπνα κινητά τους για να βγάλουν φωτογραφία το περιστατικό και ενεργοποιώντας το GPS να δηλώσουν την γεωγραφική του θέση.

Μια άλλη έκφανση των crowd sourcing εφαρμογών είναι οι λεγόμενες «συμμετοχικής ανίχνευσης» εφαρμογές όπου ο χρήστης καλείται να ενεργοποιήσει την εφαρμογή η οποία με την σειρά της θα ανιχνεύει δεδομένα για τα οποία ενδιαφερόμαστε. Ένα καλό παράδειγμα αυτού του είδους είναι η υπηρεσία της Google «Traffic». Στην συγκεκριμένη υπηρεσία ο χρήστης ανοίγει την εφαρμογή Google Maps και ενεργοποιεί το GPS της συσκευής, αφού ερωτηθεί εάν αποδέχεται να αρχίσουν να στέλνονται στοιχεία απο το κινητό του ανώνυμα στην Google, η εφαρμογή αρχίζει να συλλέγει πληροφορίες σχετικά με την θέση τοιυ κινητού.Με τον τρόπο αυτό η υπηρεσία στην ουσία ανιχνεύει την τοποθεσία των κινητών τηλεφώνων των χρηστών μέσω του GPS ή του δικτύου τηλεφωνίας και συνδυάζοντας την με τον χρόνο μπορεί να υπολογίσει τις ταχύτητες εκατομμυρίων χρηστών. Οι χρήστες μπορούν να ενημερωθούν για την κίνηση μέσω της εφαρμογής η οποία απεικονίζει ένα χάρτη όπου έχουν τονιστεί διαφορετικά οι οδοί ανάλογα με το επίπεδο της συμφόρησης, από το πράσινο που είναι η ελεύθερη ροή μέχρι το βαθύ κόκκινο που είναι το «μποτιλιάρισμα». Επίσης η εφαρμογή διαθέτει και πληροφορίες για το ποιες οδοί είναι κλειστοί λόγω έργων, εκδηλώσεων κτλ. Στην παρούσα εργασία ο χάρτης της Google Traffic χρησιμοποιείται για να διαπιστωθεί κατά πόσο τα δεδομένα της εργασίας είναι έγκυρα.

Όλες οι παραπάνω υπηρεσίες crowd sourcing βασίζονται σε γεωγραφικά δεδομένα που έχουν σταλεί εθελοντικά από τους χρήστες (Volunteered Geographic Information, Goodchild 2007) [19] και βασική αρχή τους είναι ότι προτρέπουν τους χρήστες να μοιραστούν συγκεκριμένου είδους γεωγραφικές πληροφορίες που αφορούν την εκάστοτε εφαρμογή.

Ανάλογα με τις crowd sourcing εφαρμογές οι χρήστες των κοινωνικών δικτύων μοιράζονται πολλές φορές γεωγραφικές πληροφορίες. Οι πληροφορίες αυτές εμπεριέχονται σε μηνύματα στα οποία περιγράφουν γεγονότα που συμβαίνουν γύρω από την θέση στην οποία βρίσκονται δημιουργώντας έτσι ένα σημείο ενδιαφέροντος στον ψηφιακό αλλά και στον γεωγραφικό χώρο. Χαρακτηριστικό παράδειγμα είναι η περίπτωση ενός σεισμού. Μόλις λάβει χώρα ένας σεισμός έχει παρατηρηθεί ότι οι χρήστες σπεύδουν στα κοινωνικά δίκτυα για να μοιραστούν στο Twitter την εμπειρία που μόλις βίωσαν με μηνύματα τα εμπεριέχουν γεωγραφική πληροφορία ώστε να ενημερώσουν τους διαδικτυακούς τους φίλους για την ένταση και την τοποθεσία του γεγονότος. Ένα επίσης πολύ συχνό φαινόμενο με το οποίο καταπιάνεται η παρούσα εργασία είναι η περίπτωση της κυκλοφοριακής συμφόρησης, οι χρήστες των κοινωνικών δικτύων όταν βρεθούν σε συνθήκες μποτιλιαρίσματος στους δρόμους τείνουν να μοιραστούν την πληροφορία αυτή με τους υπόλοιπους χρήστες αποσκοπώντας τις περισσότερες φορές είτε να εκφράσουν την δυσαρέσκεια τους είτε απλά για να ενημερώσουν το κοινό για τις δυσμενείς κυκλοφοριακές συνθήκες.

Οι γεωγραφικές πληροφορίες που κοινοποιούνται μέσω των κοινωνικών δικτύων διαφέρουν από αυτές των crowd sourcing εφαρμογών. Η διαφορά έγκειται στο ότι οι γεωγραφικές πληροφορίες που εμπεριέχονται στα μηνύματα των χρηστών των κοινωνικών δικτύων δεν κοινοποιούνται για κάποιο συγκεκριμένο σκοπό όπως γίνεται με τις εφαρμογές crowd sourcing. Έτσι, οι συγκεκριμένες γεωγραφικές πληροφορίες πρέπει να εξορυχτούν από τον τεράστιο όγκο πληροφοριών που διαδίδονται στη ροή των κοινωνικών δικτύων κάθε δευτερόλεπτο και να υποστούν επεξεργασία ώστε να αποτελέσουν χρήσιμα και ποιοτικά δεδομένα για την παραγωγή οποιουδήποτε συμπεράσματος.

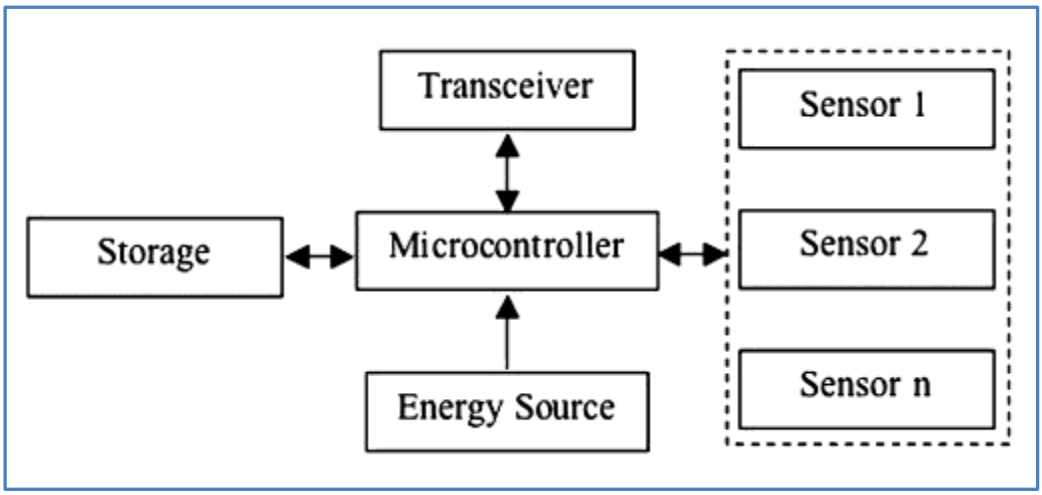
Η συνεχώς αυξανόμενη χρήση των κοινωνικών δικτύων σε συνδυασμό με την τάση των χρηστών να μοιράζονται γεωγραφικές πληροφορίες μέσω των μηνυμάτων τους καθιστά τα συστήματα αυτά μια εναλλακτική πηγή μεγάλου όγκου γεωγραφικών πληροφοριών οι οποίες με την κατάλληλη επεξεργασία μπορούν να χρησιμοποιηθούν σε πολλές εφαρμογές και να δώσουν λύση σε πολλά προβλήματα που αντιμετωπίζουν οι σύγχρονες κοινωνίες.

Οι περισσότερες εργασίες που έχουν γίνει έως τώρα οι οποίες αντιμετωπίζουν τα κοινωνικά δίκτυα ως μέσα ανίχνευσης γεωγραφικών δεδομένων απλώς χρησιμοποιούσαν τα κοινωνικά δίκτυα ως crowdsourcing εφαρμογές. Για παράδειγμα, στην Μεγάλη Βρετανία υπάρχει η ιστοσελίδα uksnowmap.com η οποία προτρέπει τους χρήστες του Twitter να δηλώνουν με συγκεκριμένης μορφής μηνύματα (συγκεκριμένο hashtag και χρήση της τοποθεσίας) που υπάρχει χιόνι. Η αντιμετώπισή όμως αυτή όπως αναφέρθηκε είναι ανάλογη των crowdsourcing.

Η παρούσα εργασία ακολουθεί την ακριβώς αντίθετη αντιμετώπιση .Έστω ότι ο κάθε χρήστης του Twitter λειτουργεί ως ανιχνευτής δεδομένων. Οτιδήποτε παρατηρεί γύρω του και προκαλεί το ενδιαφέρον του το καταγράφει στο προσωπικό του προφίλ και έτσι το μοιράζεται με την κοινότητα των χρηστών του δικτύου. Καθώς

επικοινωνούν οι χρήστες με αυτό τον τρόπο το Twitter γίνεται αυτόματα ένα σύστημα αισθητήρων που έχει εκατομμύρια «δέκτες» σε παγκόσμιο επίπεδο. Στην εργασία «*#Earthquake: Twitter as a Distributed Sensor System» (Stefanidis, 2013) [20]* παρουσιάζεται η αναλογία ανάμεσα σε ένα κοινωνικό δίκτυο όπως το Twitter και σε ένα σύστημα ασύρματων αισθητήρων (Εικόνα 4). Ένας κόμβος ενός συστήματος αισθητήρων, όπως παρουσιάζεται στο παρακάτω διαγραμμα, έχει τα εξής δομικά στοιχεία:

- Ένα ή περισσότερους αισθητήρες που συλλέγουν συγκεκριμένου τύπου και φύσης δεδομένα (sensor)
- Ένα επεξεργαστή για την άμεση επεξεργασία των δεδομένων από τους αισθητήρες (microcontroller)
- Ένα πομποδέκτη για να επικοινωνεί ο κόμβος με το υπόλοιπο σύστημα (transceiver)
- Μια συσκευή αποθήκευσης των δεδομένων (storage)
- Μια πηγή ενέργειας (energy source)



Εικόνα 4: Κόμβος συστήματος ασύρματων αισθητήρων
Πηγή: #Earthquake: Twitter as a Distributed sensor System, Andrew Crooks, 2012

Στην περίπτωσή μας ο κάθε χρήστης ενός κοινωνικού δικτύου αποτελεί τον αισθητήρα του κόμβου καθώς επίσης και τον μικροεπεξεργαστή καθώς συλλέγει τις πληροφορίες που τον ενδιαφέρουν, τις επεξεργάζεται και τις κοινοποιεί. Ο πομποδέκτης είναι το φυσικό ανάλογο του κοινωνικού δικτύου, στην προκειμένη περίπτωση το Twitter, το οποίο αναμεταδίδει τις πληροφορίες στους υπόλοιπους κόμβους του δικτύου. Τέλος η συσκευή αποθήκευσης είναι επίσης το ίδιο το δίκτυο καθώς οι πληροφορίες από τους χρήστες αποθηκεύονται και είναι διαθέσιμες προς επεξεργασία σε μεταγενέστερο χρόνο από την στιγμή που θα κοινοποιηθούν στο δίκτυο. Όπως έχει αναφερθεί το Twitter έχει εκατομμύρια ενεργούς χρήστες σε όλο τον κόσμο άρα γίνεται αντιληπτό ότι αυτό το σύστημα αισθητήρων έχει παγκόσμια κάλυψη με αποτέλεσμα να το καθιστά ασυναγώνιστο στον όγκο των πληροφοριών που συλλέγει κάθε στιγμή.

Οι πληροφορίες αυτές , τις οποίες πρέπει να εξορύξουμε από τον παγκόσμιο ιστό, δεν έχουν κατά κύριο λόγο μια ενιαία μορφή , πολλές φορές μια συγκεκριμένη πληροφορία αποδίδεται από τον κάθε χρήστη διαφορετικά στον γραπτό λόγο. Κάποιος χρήστης θα μπορούσε να γράψει στο Twitter ότι «βρέχει πολύ» κάποιος άλλος ότι «ρίχνει καρεκλοπόδαρα» και οι δύο θέλουν να περιγράψουν μια συγκεκριμένη κατάσταση αλλά την αποδίδουν με διαφορετική φράση. Με το παραπάνω παράδειγμα γίνεται αντιληπτό ότι πολλές φορές δεν είναι εύκολο να εντοπιστούν τα μηνύματα που μας ενδιαφέρουν λόγο της διαφορετικής γραπτής απόδοσης από τους χρήστες και έτσι θα πρέπει να χρησιμοποιηθούν διάφοροι τρόποι εντοπισμού των δεδομένων αυτών με αλγόριθμους Machine Learning αντικείμενο με το οποίο δεν καταπιάνεται η παρούσα εργασία.

2.6 Προηγούμενες μελέτες σχετικές με τα κοινωνικά δίκτυα

Η τεράστια απήχηση που απολαμβάνουν τα κοινωνικά δίκτυα τα τελευταία χρόνια έχει προσελκύσει πολλούς ερευνητές με σκοπό να αποκωδικοποιήσουν τις νέες ηλεκτρονικές κοινωνίες που δημιουργούνται. Έχουν αναπτυχθεί πολλές εργασίες που αφορούν την συλλογή και επεξεργασία δεδομένων από κοινωνικά δίκτυα και μέσω αυτών μπορεί να γίνει αντιληπτό το ευρύ φάσμα εφαρμογών που μπορούν να χρησιμοποιήσουν τα δεδομένα αυτά. Ένα καλό παράδειγμα είναι η εργασία των Adam Tsakalidis, Symeon Papadopoulos , Alexandra Cristea και Yianni Kompatsiari «*Predicting Elections for Multiple Countries Using Twitter and Polls*» [21] στην οποία , το 2014, χρησιμοποίησαν μηνύματα από το Twitter για να προβλέψουν το αποτέλεσμα των ευρωεκλογών της ίδιας χρονιάς σε πολλές χώρες της Ευρωπαϊκής Ένωσης. Οι ερευνητές χρησιμοποίησαν αλγορίθμους Machine Learning ώστε να αποκωδικοποιήσουν τα μηνύματα που ανακτούσαν από το κοινωνικό δίκτυο και μετά από στατιστική επεξεργασία προσπάθησαν να βρουν τις εκλογικές τάσεις που κρύβονταν στα μηνύματα αυτά. Το αποτέλεσμα ήταν υπέρ ικανοποιητικό καθώς οι προβλέψεις τους ήταν κοντά σε σχέση με τα πραγματικά ποσοστά των κομμάτων και με αυτό τον τρόπο ανέδειξαν τα κοινωνικά δίκτυα ως ένα μέσο άντλησης έγκυρων δεδομένων για έρευνες τέτοιου είδους.

Ειδικότερα, στο πεδίο των χωρικών πληροφοριών που συλλέγονται από κοινωνικά δίκτυα έχουν γίνει επίσης διάφορες αξιοσημείωτες εργασίες. Το 2010 οι Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo με την εργασία «*Earthquake Shakes Twitter Users*» προσπάθησαν, με πολύ καλά αποτελέσματα να χαρτογραφήσουν τις αντιδράσεις των πολιτών στο Twitter μετά από ένα σεισμό με στόχο να βρούν σε ποια περιοχή έδρασε ο σεισμός και μέχρι που έγινε αισθητός. Αντίστοιχη εργασία έγινε και το 2012 από τους Andrew Crooks Arie Croitoru Anthony Stefanidis και Jacek Radzikowski με τίτλο «*#Earthquake: Twitter as a Distributed Sensor System*» [20] στην οποία προσπάθησαν να αποτυπώσουν την εστία και και το εύρος ενός από τους περισσότερο αισθητούς σεισμούς στις ΗΠΑ τα τελευταία χρόνια. Τα δεδομένα που χρησιμοποίησαν ήταν μηνύματα χρηστών του Twitter τα οποία ήταν geotagged και αναφέρονταν στο γεγονός του σεισμού. Το παράγωγο της έρευνας αυτής ήταν ένας

χάρτης όπου αποτυπώνονταν τα tweets των χρηστών ο οποίος ήρθε σε αντιπαραβολή με αντίστοιχους χάρτες που έδωσαν οι αρχές των ΗΠΑ που χρησιμοποιούσαν τα επίσημα δεδομένα από σεισμογράφους. Μετά την αντιπαραβολή οι ερευνητές κατέληξαν στο συμπέρασμα ότι το Twitter είναι ένα πολύ καλό και έγκυρο εργαλείο ανάκτησης χωρικών δεδομένων τα οποία θα χρησιμοποιηθούν στην εύρεση σημείων ενδιαφέροντος όπως ένας σεισμός. Μετά από τα ευρήματα τους οι ερευνητές επιχειρηματολόγησαν για το ότι το σύστημα χρήστες-κοινωνικό δίκτυο δουλευει σαν ένα σύστημα ασύρματων αισθητήρων, όπως περιγράφηκε σε παραπάνω ενότητα.

Στην εύρεση γεωγραφικών σημείων ενδιαφέροντος έχουν συμβάλει και άλλες εργασίες όπως η «*From Tags to Trends: A First Glance at Social Media Content Dynamics»* από τους Evaggelos Spyrou και Phivos Mylonas [22] οι οποίοι προσπάθησαν να εντοπίσουν σημεία ενδιαφέροντος από φωτογραφίες του κοινωνικού δικτύου Flickr οι οποίες περιελάμβαναν χωρική πληροφορία στα μεταδεδομένα τους. Αντίστοιχη εργασία που το τελικό της παράγωγο ήταν μια εφαρμογή δημοσίευσε η Ombretta Gaggi το 2013 με τίτλο «*Discovering Local Attractions from Geo-Tagged Photos*» [23] όπου παρουσιάζεται το λογισμικό *Phototrip* το οποίο συλλέγει geotagged φωτογραφίες από το Flickr, εντοπίζει τα σημεία ενδιαφέροντος και κάνει μια ξενάγηση στον χρήστη. Ένα βήμα παραπέρα προς αυτή την κατεύθυνση προχώρησαν οι Abdul Majida , Ling Chena, Gencai Chena , Hamid Turab Mirzaa , Ibrar Hussaina και John Woodwardb το 2012 οι με το άρθρο τους «*A context-aware personalized travelrecommendation system based on geotagged social media data mining*» [24]. Οι ερευνητές χρησιμοποίησαν επίσης geotagged φωτογραφίες χρηστών του Flickr για να δημιουργήσουν ένα προφίλ του κάθε χρήστη με τα ενδιαφέροντα και τα μέρη που του αρέσει να ταξιδεύει. Στην συνέχεια δημιουργησαν ένα σύστημα το οποίο συλλέγει αυτές τις πληροφορίες του κάθε χρήστη και του προτείνει προορισμούς ανα τον κόσμο που ικανοποιούν τις απαιτήσεις του.

2.7 Προηγούμενες μελέτες σχετικές με τα κοινωνικά δίκτυα και την κυκλοφοριακή συμφόρηση

Εκτός από την γενική χρήση των κοινωνικών δικτύων ως πηγή χωρικών δεδομένων έχουν παρουσιαστεί και διάφορες εργασίες που προσεγγίζουν το θέμα της παρούσας εργασίας, την ανάκτηση μηνυμάτων από το Twitter που αναφέρονται στην κυκλοφοριακή συμφόρηση και την επεξεργασία τους. Μια από τις πρώτες εργασίες που έχουν γίνει γύρω από το θέμα είναι των Sri Krisna Endarnoto, Sonny Pradipta, Anto Satriyo Nugroho και James Purnama με τίτλο «*Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application*» [25] το 2011 όπου οι συντάκτες παρουσιάζουν μια εφαρμογή για κινητά τηλέφωνα η οποία ενημερώνει τους χρήστες για το που υπάρχει κυκλοφοριακή συμφόρηση για την πόλη της Τζακαρτα. Τα δεδομένα για την κίνηση στους δρόμους προέρχονται από το λογαριασμό του κέντρου διαχειρησης κυκλοφορίας της Τζακάρτα στο Twitter. Τα δεδομένα αφού ανακτούνται παρουσιάζονται σε ένα χάρτη της πόλης σε πραγματικό χρόνο και έτσι οι χρήστες μπορούν ανά πάσα στιγμή να είναι ενήμεροι για την κίνηση στους δρόμους. Η εργασία αυτή δεν ασχολείται τόσο

με την ανάκτηση των δεδομένων μιας και απλώς κατεβάζει δεδομένα από ένα συγκεκριμένο χρήστη και στην ουσία ασχολείται με την οπτικοποίηση τους και την προβολή τους πάνω σε ένα χάρτη. Την επόμενη χρονιά το 2012, οι Raymondus Kosala και Erwin Adi παρουσίασαν το άρθρο τους με τίτλο «*Harvesting Real Time Traffic Information from Twitter*» [26] στο οποίο και πάλι εξάγονταν από το Twitter δεδομένα για την κυκλοφοριακή κίνηση στην Τζακαρτα και τα χαρτογραφούσαν σε πραγματικό χρόνο αλλά δεν εφαρμόστηκαν κάποιες στατιστικές αναλύσεις στα δεδομένα. Το 2013 οι Schulz και Ristoski με την εργασία τους «*The Car that Hit The Burning House: Understanding Small Scale Incident Related Information in Microblogs*» [27] ασχολήθηκαν με την εξαγωγή πληροφοριών από το Twitter οι οποίες αναφέρονταν σε ατυχήματα στους δρόμους. Η εργασία χρησιμοποιούσε τεχνικές και αλγορίθμους machine learning ώστε να μπορεί ο υπολογιστής να αντιληφθεί εάν το μήνυμα που έχει ανακτηθεί σχετίζεται με ατυχήματα στους δρόμους ή όχι.

Το 2013 παρουσιάστηκε μια παρόμοια εργασία από την Fatma Amin Elsafoury με τίτλο «*Monitoring urban traffic status using twitter messages*» [28] στην οποία αναλύεται ο τρόπος με τον οποίο ανακτούνται τα μυνήματα από το Twitter και διαχωρίζει αυτά που αναφέρονται σε κίνηση στους δρόμους με αλγόριθμο Machine Learning. Τα δεδομένα που χρησιμοποιεί η συγκεκριμένη εργασία αναφέρονται μόνο για την περιοχή του Λονδίνου και πολλά από αυτά είναι ανακτημένα από ένα συγκεκριμένο λογαριασμό στο Twitter τον οποίο διαχειρίζεται το γραφείο διαχείρησης της κυκλοφορίας του Λονδίνου. Επίσης γίνεται ελάχιστη αξιολόγηση των δεδομένων καθώς παρουσιάζεται η αντιπαραβολή τους μόνο με τα δεδομένα του Google Traffic.

Τέλος το 2015 παρουσιάστηκε η εργασία « *REAL-TIME ROAD TRAFFIC INFORMATION DETECTION THROUGH SOCIAL MEDIA*» [29] από τον Chandra P. Khatri στην οποία σε διάστημα 5 μηνών ανακτήθηκαν πολλά εκατομμύρια tweets και χρησιμοποιήθηκε αλγόριθμος Machine Learning για να διαχωριστούν τα μηνύματα που αναφέρονταν σε κυκλοφοριακή συμφόρηση. Η εργασία παρουσιάζει δύο αναλύσεις, πραγματικού χρόνου και ανάλυση αφού έχουν αποθηκευτεί τα δεδομένα. Στην ανάλυση πραγματικού χρόνου στην ουσία διαχωρίζονται τα δεδομένα που έχουν να κανουν με την έρευνα και γίνεται οπτικοποίηση τους ενώ διενεργούνται διάφορες αναλύσεις χωρικές και χρονικές.. Βασικό μειονέκτημα της μεθοδολογίας της εργασίας είναι η ανάκτηση των δεδομένων καθώς αρχικά έγινε ανάκτηση όλων των tweets από την χώρα των ΗΠΑ επί 5 μήνες και μετά έγινε ο διαχωρισμός τους σε αυτά που ήταν χρήσιμα για την έρευνα. Σε αυτό το διάστημα ανακτήθηκαν περίπου 75 εκατομμύρια tweets και εν τέλει αυτά που ήταν σχετικά με την έρευνα ήταν μόνο τα 120.000. Στη παρούσα εργασία αναφέρεται ο τρόπος ώστε να ανακτηθούν πολύ πιο εύκολα τα μηνύματα που μας ενδιαφέρουν από το Twitter API καθώς η αναζήτηση μας είναι επικεντρωμένη μόνο σε αυτά. Με αυτόν τον τρόπο δεν χρειάζεται να καταφύγουμε σε μεθόδους machine learning για να διαχωρίσουμε τα δεδομένα μας καθώς το 96% αυτών είναι μηνύματα τα οποία σχετίζονται με την

έρευνα μας. Επίσης στην εργασία αυτή γίνεται μικρή ανάλυση όσον αφορά στο επίπεδο πόλης, στην εξαγωγή στατιστικών στοιχείων για τους διάφορους δρόμους και στην εύρεση των ωρών αιχμής της κάθε πόλης.

2.8 Διαφορετικές αναλύσεις των δεδομένων

Τα μηνύματα που χρησιμοποιεί η παρούσα εργασία ανακτούνται σε πραγματικό χρόνο μέσω του Streaming API του Twitter. Μόλις κάποιος χρήστης κοινοποιήσει ένα μύνημα το οποίο ανταποκρίνεται στα κριτήρια που έχουμε θέσει ανακτάται από το σύστημα και αποθηκεύεται σε μια βάση δεδομένων. Παρακάτω παρατίθενται οι διαφορετικές αναλύσεις που λαμβάνουν χώρα στην εργασία αρχικά σε πραγματικό χρόνο και μετά αφού έχουν αποθηκευτεί τα δεδομένα:

- **Αποτύπωση των μηνυμάτων σε χάρτη σε πραγματικό χρόνο και έλεγχος των δεδομένων:** Κάθε εγγραφή στην βάση δεδομένων εκτός των άλλων περιέχει τις συντεταγμένες από όπου κοινοποιήθηκε το tweet, σύμφωνα με τις συντεταγμένες αυτές αυτομάτως γίνεται προβολή του μηνύματος σε ένα διαδικτυακό χάρτη ο οποίος παρέχεται στο σύστημα μέσω του API του Google Maps. Η εφαρμογή αυτή από μόνη της αποτελεί ένα πολύ καλό εργαλείο για την ενημέρωση των πολιτών καθώς μπορούν σε πραγματικό χρόνο να βλέπουν που υπάρχει κυκλοφοριακή συμφόρηση και να επιλέγουν εναλλακτικές διαδρομές για το ταξίδι τους.
- **Έλεγχος των δεδομένων μέσω της υπηρεσίας Google Traffic:** Ο χάρτης που έχει επιλεγεί για την προβολή των μηνυμάτων έχει ως επίπεδο την υπηρεσία Google Traffic και ανανεώνεται ανά λίγα δευτερόλεπτα, σύμφωνα με την συγκεκριμένη εφαρμογή γίνεται μια πρώτη αξιολόγηση των μηνυμάτων, κατά πόσο δηλαδή οι πληροφορίες που παρέχουν ανταποκρίνονται στην πραγματικότητα.

Μετά το πέρας των δύο μηνών που διήρκησε η ανάκτηση των 150.000 tweets έγιναν διάφορες αναλύσεις με σκοπό την ανάδειξη τους ως χρήσιμες πρακτικές, χαμηλού κόστους, που μπορούν να συμβάλλουν στην λύση του κυκλοφοριακού προβλήματος:

- **Εντοπισμός χρήσιμων δεδομένων**: Αρχικά έγινε ένας έλεγχος κατα πόσο τα tweets που έχουν ανακτηθεί ανταποκρίνονται στις απαιτήσεις τις εργασίας, πριν την επεξεργασία των δεδομένων διαχωρίστηκαν τα μηνύματα που αφορούσαν κίνηση στους δρόμους με εκείνα που χρησιμοποιούσαν τον λέξη κλειδί “traffic” για να εκφράσουν κάτι διαφορετικό όπως π.χ “traffic light”.

- **Διαχωρισμός δεδομένων**: Μετά την αποθήκευση και την εκκαθάριση των μηνυμάτων δημιουργήθηκαν δύο βάσεις δεδομένων. Η μια περιέχει μηνύματα που αναφέρονται απλώς σε κυκλοφοριακή συμφόρηση και η δεύτερη περιέχει μηνύματα που αναφέρονται σε κυκλοφοριακή συμφόρηση που έχει προκληθεί από κάποιο αυτοκινητιστικό ατύχημα.
- **Εξαγωγή των οδών με την περισσότερη συμφόρηση:** Μετά από τον διαχωρισμό των δεδομένων ανάλογα με την πόλη που στάλθηκαν γίνεται επεξεργασία τους σε περιβάλλον GIS για να προσδιοριστούν οι δρόμοι που, σύμφωνα με τα δεδομένα, παρουσιάζουν μεγαλύτερη πιθανότητα εμφάνισης κυκλοφοριακής συμφόρησης στοιχείο βασικό για την διαχείριση του προβλήματος. Η ανάλυση έγινε στην πόλη της Νέας Υόρκης.
- **Εξαγωγή των οδών με τα περισσότερα ατυχήματα:** Τα δεδομένα επεξεργάζονται για να εξαχθούν οι περισσότερο επικίνδυνοι δρόμοι στην πόλη της Νέας Υόρκης.
- **Εξαγωγή των ωρών αιχμής κάθε πόλης**: Κατά τις ώρες αιχμής σύμφωνα με μετρήσεις, εμφανίζεται το 60% των περιπτώσεων συμφόρησης στους δρόμους έτσι είναι επιβεβλημένο να είναι γνωστό πότε εμφανίζονται οι ώρες αιχμής σε κάθε πόλη. Στην βάση δεδομένων υπάρχει εγγραφή χρόνου για το κάθε tweet και σύμφωνα με την συγκεκριμένη πληροφορία γίνεται ανάλυση για να βρεθούν οι πρωινές και βραδινές ώρες αιχμής για είκοσι μεγάλες πόλεις των ΗΠΑ.
- **Αποτύπωση της ροής των οχημάτων ανάλογα με την ώρα σε ένα 24ωρο**: Κάθε πόλη έχει την δική της «υπογραφή» όσον αφορά την κίνηση των οχημάτων. Η κυκλοφοριακή «υπογραφή» προσδιορίζεται με τον εντοπισμό «σημείων ενδιαφέροντος» στην ροή των οχημάτων ανάλογα με την ώρα εντός εικοσιτετραώρου. Έτσι έχουν παραχθεί χάρτες πυκνότητας σημείων σύμφωνα με τα δεδομένα για την πόλη του Houston ώστε να καταστεί ευκρινές που εντοπίζονται τα σημεία ενδιαφέροντος ανάλογα την ώρα της ημέρας.
- **Εντοπισμός της χωρικής αυτοσυσχέτισης** : Με σκοπό να αναδειχτούν σε κάθε πόλη οι περιοχές όπου η συμφόρηση είναι εντονότερη, μετρήθηκε η αυτοσυσχέτιση υποπεριοχών στις οποίες εμφανίζεται το φαινόμενο σε 10 μεγάλες πόλεις των ΗΠΑ. Η επεξεργασία που διενεργήθηκε περιλάμβανε τον υπολογισμό του ολικού και τοπικού δείκτη Moran I.
- **Κατάταξη των πόλεων ανάλογα το επίπεδο συμφόρησης**: Μεγάλοι οργανισμοί οι οποίοι ασχολούνται με την κίνηση στους δρόμους όπως η INRIX παρουσιάζουν κάθε χρόνο εκθέσεις με στοιχεία για την κίνηση στις μεγάλες πόλεις ανά τον κόσμο. Στις εκθέσεις αυτές υπάρχει ένα βασικό στοιχείο, η καθυστέρηση που βιώνουν οι πολίτες στους δρόμους σε κάθε πόλη, με αυτό το στοιχείο κατατάσσουν τις πόλεις για αναδείξουν αυτές που έχουν την περισσότερη κίνηση. Αντίστοιχη κατάταξη

επιχειρείται στην παρούσα εργασία σύμφωνα με τα αποτελέσματα που προέκυψαν από τον υπολογισμό της χωρικής αυτοσυσχέτισης.

- **Αξιολόγηση των ευρημάτων:** Τα ευρήματα τα οποία προέκυψαν από τις παραπάνω αναλύσεις συγκρίνονται με τα αντίστοιχα των εταιριών INRIX και Tomtom με σκοπό να αξιολογηθούν και να προκύψουν τα αντίστοιχα συμπεράσματα για την αξιοπιστία των δεδομένων. Η αξιολόγηση είναι πολύ σημαντική διαδικασία για να πιστοποιηθεί αν τα δεδομένα που συλλέγονται από το Twitter μπορούν να ανταποκριθούν στις απαιτήσεις της έρευνας. Εκτός από τις παραπάνω εταιρίες επίσημα δεδομένα λαμβάνονται και από τις τοπικές αρχές της κάθε πόλης.

2.9 Χρησιμότητα της παρούσας εργασίας

Το κυκλοφοριακό πρόβλημα απασχολεί καθημερινά σχεδόν όλες τις αρχές των μεγάλων πόλεων και καταπονεί τους πολίτες τους. Παρότι έχουν γίνει πολλές προσπάθειες να λυθεί, η συμφόρηση είναι ένα πρόβλημα που συνεχώς επεκτείνεται με αποτέλεσμα να δαπανούνται τεράστια ποσά για την διαχείριση του ή για καύσιμα τα οποία καταναλώνονται ασκόπως, να επιβαρύνεται το περιβάλλον και να επιβαρύνονται ψυχολογικά οι πολίτες καθημερινά. Η καθημερινή αποτελεσματική διαχείριση της ροής των οχημάτων είναι μια από τις περισσότερο βιώσιμες λύσεις και μάλλον και η πιο αποδοτική. Για την σωστή διαχείριση χρειάζονται δεδομένα που θα αναφέρονται στην κίνηση σε πραγματικό χρόνο, μόνο με δεδομένα πραγματικού χρόνου τα οποία θα είναι αξιόπιστα και θα έχουν τον απαραίτητο όγκο θα μπορεί να γίνει αποτελεσματική επέμβαση από τις αρχές και σε επίπεδο καθημερινότητας αλλά και σε επίπεδο σχεδιασμού νέων υποδομών. Μέχρι τώρα τα δεδομένα που συλλέγονται είναι από συσκευές που κοστίζουν πολύ ,δεν παρέχουν τον απαιτούμενο όγκο δεδομένων και δεν έχουν κάλυψη σε όλο το μήκος του οδικού δικτύου. Μια καλή λύση είναι οι εφαρμογές «συμμετοχικής ανίχνευσης» που παρουσιάστηκαν παραπάνω αλλά και αυτές είναι σε πρώιμο στάδιο και θα πρέπει να περάσουν αρκετά χρόνια οπότε κάθε αυτοκίνητο θα είναι εφοδιασμένο με ένα σύστημα εντοπισμού και θα στέλνει πληροφορίες σε μια κεντρική βάση και με αυτό τον τρόπο θα εξάγονται διάφορα συμπεράσματα.

Η παρούσα εργασία προσπαθεί να αντιμετωπίσει το πρόβλημα της εύρεσης των δεδομένων για την κίνηση με ένα καινοτόμο τρόπο. Τα κοικωνικά δίκτυα λόγο της μαζικής χρήσης τους από τους πολίτες σε όλο τον κόσμο και λόγο της εξάπλωσης των έξυπνων κινητών τηλεφώνων που διαθέτουν συστήματα εντοπισμού, είναι φορείς τεραστίου όγκου χωρικών πληροφοριών οι οποίες διακινούνται καθημερινά στον κυβερνοχώρο. Όπως παρουσιάστηκε παραπάνω τα κοινωνικά δίκτυα μπορούν να παρομιαστούν με ένα σύστημα ανίχνευσης δεδομένων καθώς οι πολίτες-χρήστες των κοινωνικών δικτύων δρουν σαν αισθητήρες με παγκόσμια κάλυψη. Ένα μέρος από τις χωρικές πληροφορίες που παράγονται και κοινοποιούνται στα κοινωνικά δίκτυα σχετίζονται και με την κυκλοφοριακή συμφόρηση. Τα συγκεκριμένα δεδομένα προσπαθεί να αναδείξει η παρούσα εργασία περιγράφοντας τον τρόπο που ανακτούνται, επεξεργάζονται και αξιολογούνται με σκοπό την παροχή αξιόπιστων

πληροφοριών για το πως μπορούν να συμβάλλουν στην αντιμετώπιση του μείζονος προβλήματος της συμφόρησης.

Η εργασία αυτή ίσως να χρειαζόταν ακόμα περισσότερα δεδομένα ώστε να παραχθούν καλύτερα συμπεράσματα αλλά λόγο του περιορισμένου χρόνου και των πολλών περιορισμών που θέτει το Twitter στην δωρεάν ανάκτηση μηνυμάτων μέσω του API του αυτό δεν ήταν εφικτό. Παρόλα αυτά όμως μια πιθανή μεταγενέστερη εργασία στην οποία οι ερευνητές θα είναι σε θέση να καταβάλλουν χρηματικό αντίτιμο την εταιρία για να τους παράσχει περισσότερα δεδομένα ή ακόμα να συμπεριληφθούν περισσότερα κοινωνικά δίκτυα με αποτέλεσμα να μεγαλώσει κατά πολύ ο όγκος των δεδομένων, θα μπορεί να θέσει τις βάσεις για την ανάπτυξη πολύ αποτελεσματικών εφαρμογών διαχείρισης της κυκλοφοριακής συμφόρησης.

ΚΕΦΑΛΑΙΟ 3: ΔΙΑΔΙΚΤΥΑΚΗ ΕΦΑΡΜΟΓΗ

Η παρούσα εργασία και οι αναλύσεις που ακολουθούν βασίζονται στα δεδομένα που λαμβάνονται από το Twitter. Για να εξαχθούν σωστά αποτελέσματα θα πρέπει να γίνει σωστή ανάκτηση των δεδομένων με αποτελεσματικό τρόπο και φυσικά να γίνει και επεξεργασία τους ώστε να αποκλειστούν τα μηνύματα τα οποία δεν μας ενδιαφέρουν και αλλοιώνουν την ποιότητα του δείγματος. Αφού παρουσιαστεί η φύση και η επεξεργασία των δεδομένων παρουσιάζονται οι διάφορες αναλύσεις που έχουν ήδη αναφερθεί.

3.1 Δεδομένα

3.1.1 Ανάκτηση των δεδομένων

Το Twitter δίνει την δυνατότητα στους χρήστες να κατεβάσουν μεγάλο αριθμό από τα μηνύματα που στέλνονται μέσω της υπηρεσίας του χρησιμοποιώντας τα διάφορα APIs που διαθέτει. Το API (Application Programming Interface) είναι η διεπαφή των προγραμματιστικών διαδικασιών που παρέχει ένα λειτουργικό σύστημα, βιβλιοθήκη ή εφαρμογή προκειμένου να επιτρέπει να γίνονται προς αυτά αιτήσεις από άλλα προγράμματα ή/και ανταλλαγή δεδομένων [30]. Στην ουσία είναι ένα κανάλι επικοινωνίας και αλληλεπίδρασης ενός συστήματος , στην περίπτωσή μας το Twitter με μια εξωτερική πηγή (πχ η ιστοσελίδα ή η web εφαρμογή) μέσω συγκεκριμένων προγραμματιστικών εντολών που ορίζονται από έναν οδηγό εντολών και χρήσης του API. Στην εργασία έχει αναπτυχθεί μια εφαρμογή η οποία χρησιμοποιεί το API του Twitter για να κατεβάζει τα tweets και να τα αποθηκεύει σε μία βάση δεδομένων. Υπάρχουν πολλά είδη APIs που παρέχει το Twitter τα οποία χωρίζονται σε δύο κατηγορίες τα Streaming APIs και τα REST APIs. Ένα από τα σημαντικότερα APIs της ομάδας του REST είναι το Search API από το οποίο η εκάστοτε εφαρμογή μπορεί να ανακτήσει μηνύματα , σύμφωνα με τα κριτήρια που ορίζει ο χρήστης, τα οποία

έχουν δημοσιευτεί τις τελευταίες 7 ημέρες. Στην εργασία όμως χρησιμοποιείται ένα από τα Streaming APIs καθώς με αυτό γίνεται ανάκτηση των μηνυμάτων σε πραγματικό χρόνο. Συγκεκριμένα χρησιμοποιήθηκε η Public streams έκδοση του Streaming API καθώς αυτή μπορεί να αναζητήσει και τελικά να αποστείλει στην εφαρμογή που έχει δημιουργήσει ο χρήστης όλα τα δεδομένα που αναρτώνται δημόσια στο Twitter, ανάλογα πάντα με τα κριτήρια που έχουν καθοριστεί εξ αρχής [31].

Το Twitter όμως έχει θέσει πολλούς περιορισμούς στο κατέβασμα των δεδομένων του μέσω των APIs που διαθέτει για οικονομικούς λόγους. Όπως αναφέρθηκε από το REST API μπορεί να ανακτήσει μηνύματα μόνο για τις τελευταίες 7 ημέρες, εάν κάποιος χρήστης ή ερευνητής χρειάζεται περισσότερα δεδομένα θα πρέπει να καταφύγει σε εταιρίες διαχείρισης δεδομένων με τις οποίες συνεργάζεται το Twitter όπως η Gnip [32]. Οι εταιρίες αυτές έχουν βάσεις δεδομένων με όλα τα ιστορικά αρχεία του Twitter στα οποία οι χρήστες μπορούν να έχουν πρόσβαση με την καταβολή του αντίστοιχου ποσού ανάλογα με τον όγκο των δεδομένων που χρειάζονται. Αντίστοιχα με τα ιστορικά δεδομένα έτσι και για τα δεδομένα πραγματικού χρόνου τα οποία διατίθενται στους χρήστες μέσω του Streaming API υπάρχουν περιορισμοί στο κατέβασμα οι οποίοι δυσχεραίνουν το έργο της έρευνας. Λόγο των διαφόρων περιορισμών έχει υπολογιστεί οτι από το Streaming API μπορούν να ανακτηθούν μόνο το 1% των μηνυμάτων που αντιστοιχούν στα κριτήρια που έχουμε θέσει [33]. Στην περίπτωσή μας λοιπόν, έχοντας θέσει ως κρητήριο ότι θέλουμε tweets τα οποία περιέχουν την λέξη “traffic” το API θα μας στείλει μόνο το 1% των tweets που κοινοποιούνται εκείνη την στιγμή με το συγκεκριμένο κρητήριο, γεγονός που περιορίζει κατά πολύ την έρευνα. Για την ανάκτηση περισσότερων δεδομένων θα πρέπει και πάλι να απευθυνθεί ο χρήστης στις εταιρίες που αναφέρθηκαν προηγουμένως και να καταβάλει ένα αξιοσημείωτο χρηματικό ποσό. Όπως αναφέρεται στις οδηγίες χρήσης του Streaming API η επιλογή του 1% των δεδομένων είναι τυχαία γεγονός που είναι σημαντικό γιατί σε διαφορετική περίπτωση το δείγμα που θα λαμβάναμε δεν θα ήταν αξιόπιστο.

Υπάρχουν δύο τρόποι για να ανακτηθούν δεδομένα από το Twitter Streaming API όταν μας ενδιαφέρει η τοποθεσία από όπου στέλνονται τα tweets. Ο πρώτος τρόπος είναι να ορίσουμε στο API ένα χώρο στο χάρτη και να ζητήσουμε να ανακτηθούν όλα τα μυνήματα που στέλνονται από τον συγκεκριμένο χώρο. Η διαδικασία αυτή περιλαμβάνει την εισαγωγή των συντεταγμένων των κορυφών ενός τετραγώνου το οποίο περικλείει την περιοχή που μας ενδιαφέρει. Όταν συλλεχθούν όλα τα μηνύματα από την συγκεκριμένη περιοχή γίνεται εκκαθάριση με διάφορες μεθόδους ώστε να απομονωθούν τα tweets που εμπεριέχουν την λέξη κλειδί που μας ενδιαφέρει. Ο δεύτερος τρόπος ο οποίος και χρησιμοποιήθηκε στην παρούσα εργασία ακολουθεί την αντίθετη διαδρομή. Αρχικά δηλώνεται η λέξη κλειδί (traffic) που μας ενδιαφέρει και ξεκινάει η συλλογή των δεδομένων χωρίς να μας ενδιαφέρει η προέλευσή τους. Μόλις ολοκληρωθεί η συλλογή των δεδομένων απομονώνονται τα μηνύματα τα οποία προέρχονται από την περιοχή που μας ενδιαφέρει. Ο τρόπος αυτός επιλέχθηκε

μετά από πειραματισμό καθώς παρατηρήθηκε ότι 80% των μηνυμάτων που ανακτούσε το σύστημα προέρχονταν από τις Ηνωμένες Πολιτείες Αμερικής , έτσι αποφασίστηκε η εργασία να γίνει για την συγκεκριμένη χώρα αφού υπάρχει πληθώρα δεδομένων. Η επιλογή ενός από του δύο παραπάνω τρόπους ανάκτησης των μηνυμάτων από το API διαδραματίζει σημαντικό ρόλο στην διεξαγωγή της έρευνας. Εάν στην έρευνα που διεξάγεται χρειάζονται μόνο τα μηνύματα που περιέχουν μια συγκεκριμένη λέξη κλειδί όπως στην παρούσα εργασία ο δεύτερος τρόπος είναι μονόδρομος. Αρκετές εργασίες έχουν στόχο να ανακτηθούν μηνύματα με συγκεκριμένη λέξη κλειδί αλλά χρησιμοποιούν τον πρώτο τρόπο, το αποτέλεσμα είναι να πρέπει να κατεβάσουν εκατομμύρια tweets τα οποία επεξεργάζονται με μεθόδους Machine Learning ώστε να απομονώσουν τα tweets που τους ενδιαφέρουν. Είναι αντιληπτό ότι η μέθοδος αυτή είναι χρονοβόρα και μη αποτελεσματική.

3.1.2 Πηγή των δεδομένων

Όπως έχει αναφερθεί ήδη μέσω της υπηρεσίας του Twitter κοινοποιούνται περί των 600 εκατομμυρίων tweets κάθε μέρα σε παγκόσμια κλίμακα. Σε κάποια από αυτά οι χρήστες αναφέρονται στην κυκλοφοριακή συμφόρηση καθημερινά καθώς το πρόβλημα εμφανίζεται επί καθημερινής βάσης. Τα μηνύματα αυτά επιχειρεί να εκμεταλλευτεί η παρούσα εργασία για να συμβάλλει στην λύση του δύσκολου προβλήματος της κίνησης στους δρόμους. Τα tweets που χρησιμοποιούνται για την έρευνα έχουν κατεβεί με την βοήθεια των APIs του Twitter για τα οποία γίνεται αναφορά στην προηγούμενη παράγραφο. Ενδιαφέρον παρουσιάζουν κάποια δημογραφικά στοιχεία για το Twitter όπως παρουσιάζονται στην έρευνα της ιστοσελίδας www.brittonmdg.com [34] η οποία ειδικεύεται στην ανάλυση στατιστικών στοιχείων διαφόρων κοινωνικών δικτύων. Σύμφωνα με την έρευνα το Twitter είναι πολύ δημοφιλές κοινωνικό δίκτυο σε νέους 18-29 ετών (31%) και γενικότερα παρατηρείται κυρίως η χρήση του από ενήλικα άτομα γεγονός που είναι θετικό για την έρευνά μας καθώς αναζητάμε κυρίως μηνύματα οδηγών που εκ των πραγμάτων είναι άτομα άνω των 18 ετών. Σημαντικό στοιχείο της έρευνας είναι επίσης η περίπου ίση κατανομή των χρηστών στις πόλεις με 18% , στα προάστια των πόλεων 19% και εκτός πόλεων 11% δεδομένο που είναι χρήσιμο για την στατιστική ανάλυση. Να σημειωθεί ότι τα ποσοστά αναφέρονται στο σύνολο των χρηστών που χρησιμοποιούν το διαδίκτυο και βασίζονται σε δεδομένα του 2014.

Από τα 600 εκατομμύρια tweets που κοινοποιούνται καθημερινά σύμφωνα με έρευνα του PewSearch Center [35] το 24,4 % προέρχονται από τις ΗΠΑ. Όπως αναλύθηκε παραπάνω από το API μπορούμε να συλλέξουμε μόνο το 1% των tweets, επίσης όπως αναφέρει η σελίδα fastcompany.com [36] μόνο το 20% των tweets που κοινοποιούνται περιέχουν στοιχεία τοποθεσίας. Με ένα απλό υπολογισμό (600.000.000x0.244x0.01x0.2) μπορούμε να ισχυριστούμε ότι υπάρχουν διαθέσιμα περίπου 292,000 tweets μόνο από τις ΗΠΑ. Σε κάποιες εργασίες αναφέρεται ότι

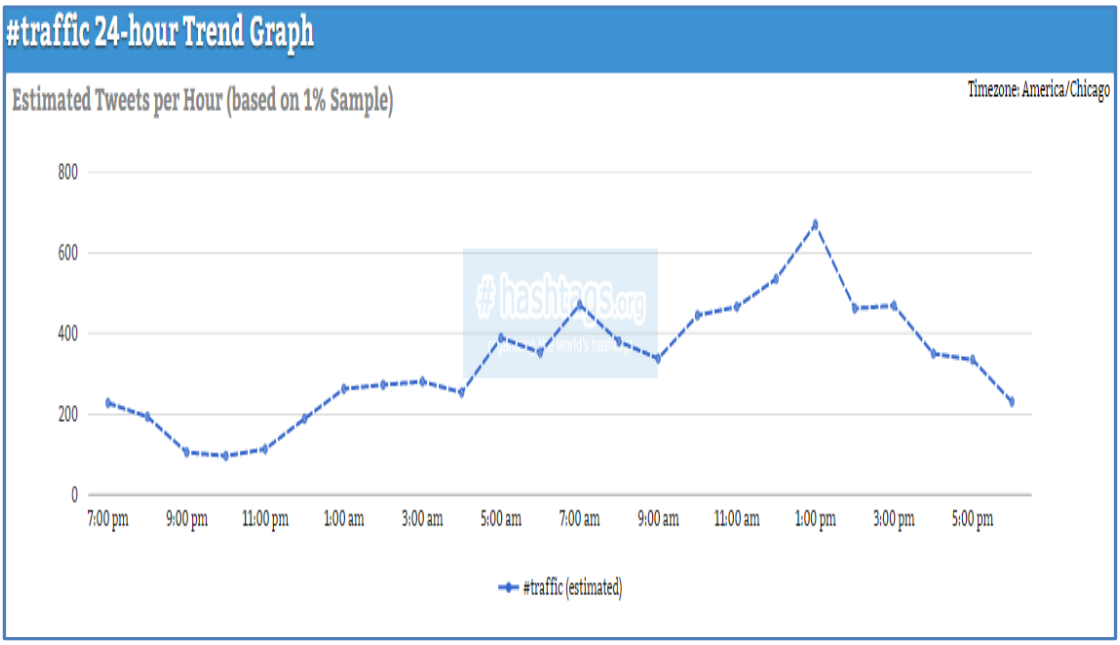
υπάρχει τρόπος να παρακαμφθεί ο περιορισμός του 1% και να υπάρξει η δυνατότητα να ανακτηθούν περισσότερα δεδομένα, επίσης άλλες εργασίες τοποθετούν τα geo-tagged tweets στο 5% του συνολικού αριθμού.

Στην παρούσα εργασία η ανάκτηση των tweets διήρκησε περίπου τρείς μήνες, για την ακρίβεια 70 ημέρες καθώς στην πορεία προέκυψαν διάφορα τεχνικά προβλήματα τα οποία διέκοπταν την απρόσκοπτη λήψη των μηνυμάτων. Κατα την διάρκεια των 70 ημερών η εφαρμογή που συνέλλεγε και αποθήκευε τα μυνήματα λειτουργούσε σε εικοσιτετράωρη βάση. Η αδιάλειπτη ανάκτηση των μηνυμάτων είναι πολύ σημαντική καθώς έπρεπε να συλλεχθούν δεδομένα από όλες της ώρες τις ημέρας ώστε να είναι χρήσιμα για την χρονική ανάλυση που θα παρουσιαστεί παρακάτω. Το κατέβασμα των μηνυμάτων έγινε τους μήνες Μάιο Ιούνιο και Ιούλιο του 2016 και ανακτήθηκαν συνολικά 150,000 tweets. Κατα μέσο όρο κατέβαιναν 2100 μηνύματα το 24ωρο τα οποία στο σύνολο τους περιείχαν την λέξη “traffic” και είχαν πληροφορία τοποθεσίας. Θεωρητικά εάν υπήρχε διαθέσιμο το 100% των μυνημάτων που εμπεριέχουν την λέξη traffic θα μπορούσαμε να συλλέξουμε 210,000 μηνύματα που να αναφέρονταν σε κίνηση στους δρόμους σε παγκόσμια κλίμακα.

Για να γίνει έλεγχος κατά πόσο το σύστημα που χρησιμοποιήθηκε είναι αποτελεσματικό χρησιμοποιήθηκαν μετρήσεις από διαδικτυακές εφαρμογές που δείχνουν πόσες φορές , κατά μέσο όρο, εμφανίζεται κάποια λέξη ή hashtag την ημέρα στα μηνύματα στο Twitter. Σύμφωνα με την σελίδα hashtracking.com [37], η οποία είναι μηχανή αναζήτησης για λέξεις κλειδιά που χρησιμοποιούνται στο Twitter, η λέξη “traffic” είτε ως λέξη μέσα στο κείμενο των μηνυμάτων είτε ως hashtag (#traffic) εμφανιζόταν 290,000 φορές μέσα σε 30 ημέρες. Η συγκεκριμένη εφαρμογή χρησιμοποιεί το API του Twitter χωρίς να πληρώνει όπως δηλώνεται στην σελίδα , άρα μπορεί να ανακτήσει μόνο το 1% των tweets. Άρα την ημέρα η συγκεκριμένη εφαρμογή μπορεί να ανακτήσει περίπου 9700 στα οποία συμπεριλαμβάνονται και τα tweets που διαθέτουν πληροφορία τοποθεσίας και όλα τα υπόλοιπα τα οποία δεν διαθέτουν. Αν θεωρήσουμε ότι οι μετρήσεις είναι σωστές και το 20% περίπου των μηνυμάτων που στέλνονται στο Twitter είναι geo-tagged στην ουσία το hashtracking.com την ημέρα ανακτά περίπου 1950 tweets την ημέρα σε παγκόσμιο επίπεδο. Μια άλλη εφαρμογή το hashtags.com [38] όπως διακρίνεται και στο παρακάτω γράφημα (Εικόνα 5), αναφέρει ότι η λέξη κλειδί “traffic” αναφέρεται ανά τον κόσμο 400 φορές την ώρα κατά μέσο όρο άρα με την αναγωγή που έγινε και προηγουμένως, 1920 tweets την ημέρα τα οποία περιέχουν χωρική πληροφορία. Όπως είναι αντιληπτό από τα προηγούμενα η εφαρμογή που χρησιμοποιήθηκε στην εργασία ήταν αποτελεσματική καθώς ο ρυθμός ανάκτησης των μηνυμάτων, 2100 ανά ημέρα, είναι κοντά αν όχι μεγαλύτερος από τον αντίστοιχο των επαγγελματικών εφαρμογών που παρουσιάστηκαν παραπάνω.

Σε κάποιες εργασίες έχει γίνει η σύγκριση ανάμεσα σε δύο εφαρμογές κατά πόσο τα δεδομένα που λαμβάνουν έχουν κοινά στοιχεία. Τα αποτελέσματα έδειξαν ότι από το Streaming API εάν αρχίζουν να κατεβάζουν παράλληλα δύο εφαρμογές τα tweets που θα λαμβάνονται είναι ακριβώς τα ίδια, έτσι αν ο χρήστης χρειάζεται περισσότερα

δεδομένα από αυτά που του παρέχει το API δεν μπορεί να χρησιμοποιήσει δύο εφαρμογές που θα κατεβάζουν ταυτόχρονα μηνύματα. Ο μοναδικός τρόπος όπως αναφέρθηκε και προυγουμένως είναι να αγοράσει τα δεδομένα που χρειάζεται.



Εικόνα 5: Εκτίμηση εμφάνισης της λέξης κλειδί "traffic" καθημερινά στο Twitter
Πηγή: hashtags.com

3.2 Διαδικτυακή Εφαρμογή

3.2.1 Διαδικασία συλλογής μηνυμάτων

Η όλη διαδικασία ανάκτησης και επεξεργασίας των μηνυμάτων από το Twitter έχει βασικό στόχο την δημιουργία μιας βάσης δεδομένων η οποία θα περιέχει αφενός ένα ικανό όγκο μηνυμάτων ο οποίος θα ανταποκρίνεται στις ανάγκες της έρευνας και αφετέρου μηνύματα τα οποία σχετίζονται αποκλειστικά με την κυκλοφοριακή συμφόρηση αυξάνοντας έτσι την αξιοπιστία των αποτελεσμάτων της ανάλυσης. Η διαδικασία για την δημιουργία της βάσης αυτής αποτελείται από διάφορα βήματα τα οποία θα αναλυθούν παρακάτω. Όπως έχει αναφερθεί ήδη, χρησιμοποιήθηκε το κατάλληλο λογισμικό το ποίο ήταν ικανό να επικοινωνεί με το Streaming API του Twitter και με αυτό τον τρόπο γινόταν ανάκτηση των μηνυμάτων σε πραγματικό χρόνο. Το λογισμικό που χρησιμοποιήθηκε είναι ανοικτού τύπου και ονομάζεται Phirehose [39] με το οποίο στην ουσία παρέχεται ένα εύχρηστο περιβάλλον επικοινωνίας με το Twitter Streaming API. Μέσω του λογισμικού δόθηκαν στο API κάποια κριτήρια που θα έπρεπε να πληρούν τα μηνύματα ώστε να αποθηκεύονται στην βάση δεδομένων. Αρχικά, όπως είναι προφανές, θα έπρεπε να εμπεριέχουν την λέξη κλειδί “traffic”. Κατά την δημιουργία του λογισμικού έγινε προσπάθεια να χρησιμοποιηθούν και άλλες λέξεις κλειδιά ώστε τα αποτελέσματα να είναι περισσότερα. Έτσι δοκιμάστηκαν διάφορες λέξεις της αγγλικής γλώσσας με τις οποίες μπορεί να περιγραφεί η κατάσταση της κυκλοφοριακής συμφόρησης όπως

“blockage”, “bottleneck”, “congestion”, “gridlock” ή “jam”. Μετά από πειραματισμό με τις παραπάνω λέξεις εξάχθηκε το συμπέρασμα ότι δεν είναι σκόπιμο να χρησιμοποιηθούν ως λέξεις κλειδιά καθώς αφενός τα μηνύματα που ανακτούνταν ήταν ελάχιστα σε σχέση με τα μηνύματα με την λέξη “traffic” και αφετέρου πολλές φορές χρησιμοποιούνταν οι λέξεις αυτές για να δηλώσουν κάτι τελείως διαφορετικό από την κυκλοφοριακή κίνηση. Για παράδειγμα η λέξη “jam” σημαίνει μαρμελάδα αλλά στην καθομιλουμένη χρησιμοποιείται για να περιγράψει την κατάσταση μποτιλιαρίσματος. Έτσι αποφασίστηκε να χρησιμοποιηθεί μόνο η λέξη “traffic” η οποία βεβαίως έχει και άλλες χρήσεις από αυτήν της κίνησης στους δρόμους αλλά τα μηνύματα αυτά αφαιρέθηκαν στην μετέπειτα επεξεργασία των δεδομένων.

Το επόμενο βασικό κριτήριο για την συλλογή των μηνυμάτων ήταν φυσικά να περιέχουν χωρική πληροφορία δηλαδή, όπως επίσης έχει αναφερθεί, ο χρήστης θα πρέπει να έχει ενεργοποιήσει τις υπηρεσίες εντοπισμού θέσης πριν κοινοποιήσει το μήνυμα στο μέσο κοινωνικής δικτύωσης. Βασική λεπτομέρεια σε αυτό το σημείο είναι ότι το κάθε μήνυμα πρέπει να περιέχει ζεύγος συντεταγμένων καθώς το Twitter δίνει και την δυνατότητα στους χρήστες να συμπεριλάβουν στα μηνύματα τους την ευρύτερη τοποθεσία τους όπως το όνομα της πόλης τους χωρίς να κοινοποιούν την ακριβή τους θέση. Τα μηνύματα τέτοιου τύπου δεν είναι χρήσιμα για την ανάλυση της παρούσας εργασίας και για αυτό τον λόγο αποκλείστηκαν από την διαδικασία ανάκτησης.

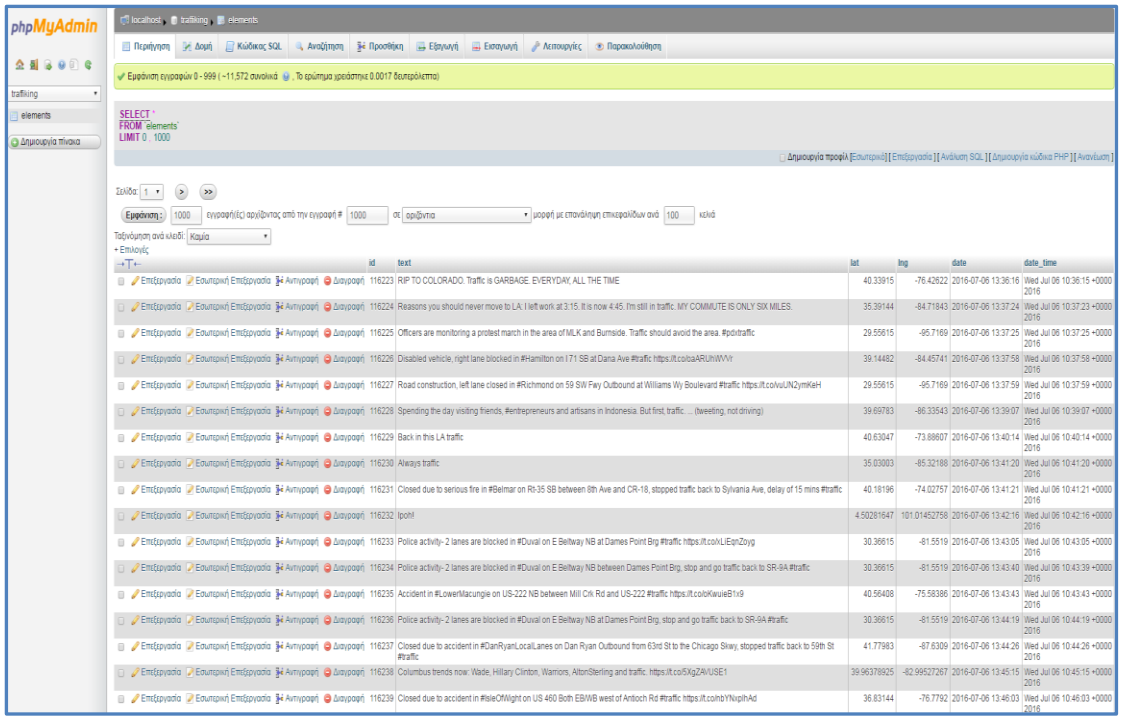
Το Twitter API διαθέτει διάφορες πληροφορίες για το κάθε tweet και μια εξ αυτών είναι το λειτουργικό σύστημα της συσκευής η οποία έχει χρησιμοποιηθεί για να σταλεί το μήνυμα από τον χρήστη. Αυτή η πληροφορία είναι πολύ σημαντική για την έρευνά μας γιατί αναπτύχθηκε η εφαρμογή με τέτοιο τρόπο ώστε να αποθηκεύει μηνύματα τα οποία προέρχονταν μόνο από συσκευές κινητών τηλεφώνων, δηλαδή μηνύματα που έχουν σταλεί από συσκευή που χρησιμοποιεί Android ή IOS ή Windows Mobile, τα τρία ποιο διαδεδομένα λειτουργικά συστήματα για κινητά τηλέφωνα. Με αυτό τον τρόπο αποκλείστηκαν τα μηνύματα τα οποία εμπεριείχαν την λέξη κλειδί “traffic” αλλά προέρχονταν από προσωπικούς σταθερούς υπολογιστές. Η διάκριση αυτή έγινε καθώς το πιο πιθανό είναι τα μηνύματα από προσωπικούς υπολογιστές να χρησιμοποιούσαν την λέξη “traffic” με διαφορετική έννοια από αυτή της κίνησης στους δρόμους, όπως για παράδειγμα Internet traffic. Από την άλλη μεριά τα μηνύματα από τα κινητά τηλέφωνα είναι λογικό το μεγαλύτερο μέρος τους να αναφέρονται σε κυκλοφοριακή κίνηση λόγω της φορητότητας τους, αν και αυτό προφανώς δεν είναι δεδομένο και για αυτό έγινε εκκαθάριση των δεδομένων πριν από την επεξεργασία τους όπως θα αναλυθεί παρακάτω.

3.2.2 Αρχική βάση δεδομένων

Όταν θα γίνεται αναφορά από εδω και στο εξής στην «αρχική βάση δεδομένων» θα εννοείται η βάση στην οποία αποθηκεύονται τα tweets που συλλέγονται κατευθείαν από το Twitter χωρίς να έχουν υποστεί κάποιας μορφής επεξεργασία. Η αρχική βάση δεδομένων είναι ένα κομμάτι του συστήματος (λογισμικού) που αναπτύχθηκε για την ανάκτηση των μηνυμάτων. Κάθε εγγραφή στην βάση δεδομένων αναφέρεται σε ένα συγκεκριμένο tweet και περιέχει τα εξής συστατικά στοιχεία:

- Ένα αύξων αριθμό ο οποίος άρχισε να μετράει από το πρώτο μήνυμα έως το τελευταίο που αποθηκεύτηκε στην βάση δεδομένων, με τελευταία καταγραφή 150,081
- Το κείμενο του εκάστοτε μηνύματος
- Το γεωγραφικό πλάτος του σημείου στο WGS84, από όπου στάλθηκε το κάθε μήνυμα
- Το γεωγραφικό μήκος του σημείου στο WGS84, από όπου στάλθηκε το κάθε μήνυμα
- Την ημερομηνία και την ακριβή ώρα που στάλθηκε το κάθε μήνυμα σε Συγχρονισμένο Παγκόσμιο Χρόνο (UTC)
- Την ημερομηνία και την ώρα που αποθηκεύτηκε το κάθε μήνυμα στην αρχική βάση δεδομένων, σύμφωνα με την ώρα Ελλάδος

Παρακάτω παρατίθεται ένα στιγμιότυπο (Εικόνα 6) από την αρχική βάση δεδομένων όπως υλοποιήθηκε με την βοήθεια του εργαλείου phpMyAdmin [40].



Εικόνα 6 : Η βάση δεομένων που δημιουργήθηκε για τις ανάγκες της εργασίας

Όπως έχει ήδη αναφερθεί η συλλογή των μηνυμάτων διήρκησε περίπου 70 ημέρες κατά τις οποίες αποθηκεύτηκαν στην αρχική βάση δεδομένων περίπου 150.000 μηνύματα. Κατά την διάρκεια της συλλογής αποθηκεύτηκαν στην βάση μηνύματα δύο κατηγοριών που ο διαχωρισμός τους είναι σημαντικός. Στην πρώτη κατηγορία κατατάσσονται τα μηνύματα τα οποία προσδοκούσε η παρούσα εργασία να ανακτήσει. Μηνύματα από πολίτες οι οποίοι βιώνουν κάποια στιγμή της ημέρας το πρόβλημα της κυκλοφοριακής συμφόρησης και θέλουν με την κοινοποίηση της πληροφορίας αυτής να εκφράσουν την δυσαρέσκεια τους ή απλώς θέλουν να ενημερώσουν τους υπόλοιπους χρήστες ότι στο συγκεκριμένο σημείο που βρίσκονται υπάρχει πρόβλημα με την κίνηση και να τους παροτρύνουν να διαλέξουν μια διαφορετική διαδρομή. Εκτός από αυτά, στην αρχική βάση δεδομένων αποθηκεύτηκαν και μηνύματα τα οποία στέλνονταν από επαγγελματικούς λογαριασμούς οι οποίοι έχουν φτιαχτεί για την ενημέρωση των οδηγών, για την κίνηση στους δρόμους, μέσω Twitter. Τα περισσότερα από τα μηνύματα αυτά προέρχονται από την ιστοσελίδα sigalert.com [41] η οποία διαθέτει εφαρμογή που εντοπίζει τα σημεία στους δρόμους που υπάρχει κυκλοφοριακή συμφόρηση και με αυτοματοποιημένο τρόπο μεταδίδει την πληροφορία αυτή στο Twitter. Πολλοί από τους επαγγελματικούς λογαριασμούς ανήκουν στις τοπικές αρχές της εκάστοτε πόλης και προσπαθούν μέσω αυτών να ενημερώσουν τους οδηγούς για το που υπάρχει πολύ κίνηση με στόχο την καλύτερη διαχείριση της. Μερικοί από τους επαγγελματικούς λογαριασμούς από τους οποίους ελήφθησαν μηνύματα είναι οι @TotalTrafficNO , @GMPWhitefield, @WTOPtraffic, @TotalTrafficBOS, @TotalTrafficDFW, @TotalTrafficLA όπως και πολλοί άλλοι. Τα μηνύματα που λαμβάνονται από τέτοιου είδους λογαριασμούς δεν συνάδουν με την αρχική φιλοσοφία της παρούσας εργασίας καθώς δεν είναι αποτέλεσμα απλής καταγραφής

ενός αξιοσημείωτου φαινομένου από τους χρήστες του κοινωνικού δικτύου αλλά η καταγραφή του φαινομένου από ειδικευμένους παράγοντες. Στην ουσία τα μηνύματα αυτά είναι ένα μέσο πληροφόρησης για την κίνηση σύμφωνα με τις συμβατικές πρακτικές διαχείρισης του προβλήματος και δεν εμπίπτουν στην θεωρία που υιοθετεί η εργασία η οποία θεωρεί τους πολίτες ως αισθητήρες που αντιλαμβάνονται που υπάρχει πρόβλημα και το διαδίδουν. Παρόλαυτα τα μηνύματα από τους επαγγελματικούς λογαριασμούς, των οποίων η εγκυρότητά τους προφανώς και δεν αμφισβητείται, χρησιμοποιήθηκαν σε ορισμένες από τις παρακάτω αναλύσεις καθώς αποτελούν περίπου το 30% του συνολικού αριθμού των μηνυμάτων και αν αποκλείονταν θα υπήρχε πρόβλημα επάρκειας δεδομένων. Στην ανάλυση που δεν έλαβαν μέρος τα μηνύματα αυτά ήταν για την κατάταξη των πόλεων σύμφωνα με την κυκλοφοριακή συμφόρηση που βιώνουν οι πολίτες τους καθώς δεν υπήρχε ίσος καταμερισμός των επαγγελματικών λογαριασμών στις διάφορες πόλεις. Εάν χρησιμοποιούνταν τα μηνύματα αυτά για την συγκεκριμένη ανάλυση είναι αντιληπτό ότι δεν θα υπήρχε σωστό δείγμα από την κάθε πόλη και επομένως σοβαρός κίνδυνος εξαγωγής λανθασμένων αποτελεσμάτων.

Στους παρακάτω πίνακες(Πίνακας 2, Πίνακας 3) παρατίθενται διάφορα μηνύματα από τις δύο κατηγορίες που αναφέρθηκαν παραπάνω.

Μηνύματα από ιδιωτικούς λογαριασμούς
Traffic!!!! ☹ ☹
I'm at Turnpike North Stuck In Traffic https://t.co/vNskwQI1Ml
Just got home heavy traffic
Welcome traffic Jakarta
Morning Traffic in LA. #freeways #copterpilot #lostinla #losangeles maverickhelicopters @ Losâ€¦ https://t.co/SrfsggDbWu
traffic congestion!!
Los Angeles traffic... #epicsurvivors #roadtrip
I am so glad I'm back in #London in my home town but f*** #traffic
Stuck in traffic AGAIN!!!
My Lagos!!! Heat and traffic like damn!!! @ Murtala Muhammed International Airport https://t.co/0UsVJQsRj3
traffic sucks be sure to give yourself plenty of time to get to the show tonight!! @ Reggies https://t.co/HiUHq7FoIu
I'm not sure if traffic is worse here or Madison
Boring!!! ☹ #traffic
Why so traffic????? Jeeezzzzz!!!!
Traffic as far as the eye can see..... @ I-26 South Carolina https://t.co/tFbTSfFiby
Stuck in traffic?? .. Take a #Selfie..
Back to normal #city #traffic
Car on fire.Traffic is terrible. DO NOT TAKE N2 @ Nandi Drive https://t.co/wKXITox0ke

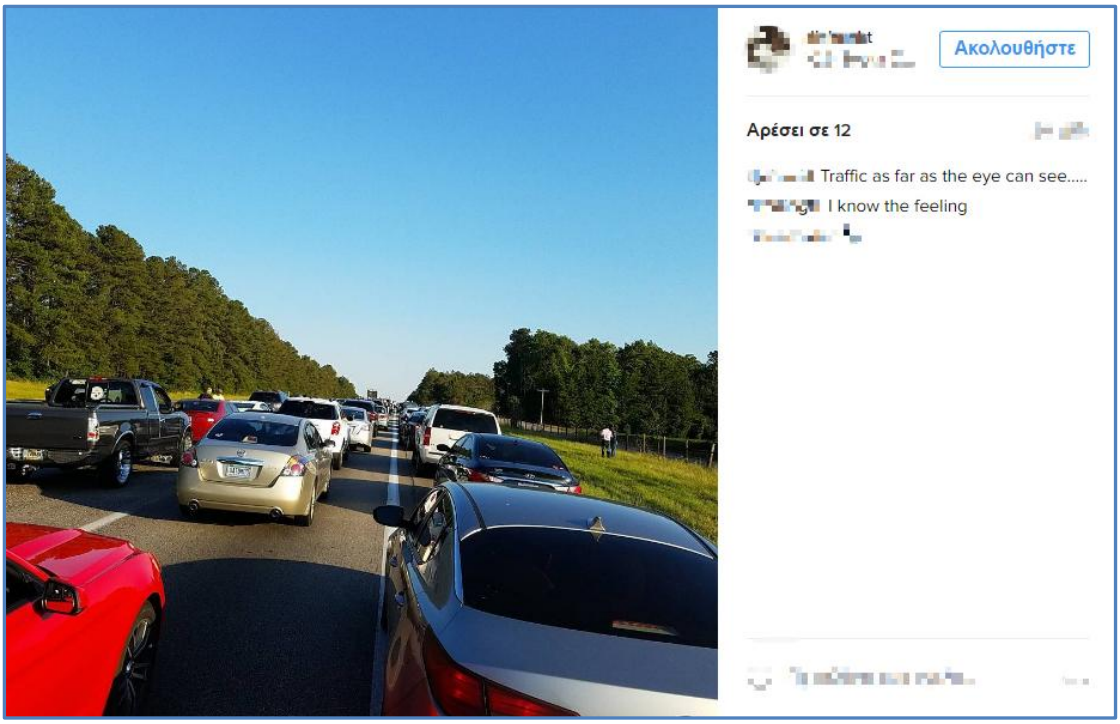
Πίνακας 2: Μηνύματα από ιδιωτικούς λογαριασμούς

Μηνύματα από επαγγελματικούς λογαριασμούς
Stop and go traffic in #Hillsborough on I-275 NB between Ashley Dr and I-4, delay of 2 mins #traffic https://t.co/KJgf5PdTvw
Closed due to accident in #Middletown on Rt-1 SB between PA-213 and Rockhill Dr #traffic https://t.co/BtegSRDq7m
Road construction, two lanes closed in #Hillsborough on I-75 SB at Fletcher Ave #traffic https://t.co/KJgf5PdTvw
Bridge closed in #Providence on Pleasant Vly Pkwy Both NB/SB after Kinsley Ave and before Vly St #traffic https://t.co/8Iu0AzFtbV
Slow moving maintenance vehicle in #Tarrant on N Lp 820 WB between Beach St and Riverside Dr., stopped traffic back to Denton Hwy
Disabled vehicle, caution in #Charleston on I 26 WB after Aviation Ave #traffic https://t.co/dVc3fFdgVj
Two right lanes blocked in #BushAirport on N Sam Houston Tollway EB at John F Kennedy Blvd #traffic https://t.co/vuUN2yElDh

Πίνακας 3: Μηνύματα από επαγγελματικούς λογαριασμούς

Όπως διακρίνεται από τα μηνύματα από τους ιδιωτικούς λογαριασμούς οι πολίτες ανταποκρίνονται ακριβώς όπως είχε προβλεφθεί στην αρχή της έρευνας, βρίσκονται σε κάποιο μποτιλιάρισμα και απλώς κοινοποιούν αυτή την δυσάρεστη εμπερία τους στο Twitter με μηνύματα τύπου «Why so traffic????? Jeeezzzzz!!!!». Επίσης πολλοί προσπαθούν να ενημερώσουν τους υπόλοιπους χρήστες του κοινωνικού δικτύου να αποφύγουν την συγκεκριμένη οδό μέσω μηνυμάτων τύπου «Car on fire.Traffic is terrible. DO NOT TAKE N2», κάποιοι άλλοι πάλι , ίσως λόγο πλήξης κοινοποιούνουν μηνύματα τύπου «Stuck in traffic?? .. Take a #Selfie».

Πολλοί χρήστες θέλοντας να γίνει αντιληπτή ,από τα υπόλοιπα μέλη της κοινότητας , η κατάσταση στην οποία βρίσκεται ο δρόμος που διασχίζουν, επιστρατεύουν τις φωτογραφικές μηχανές των κινητών τους τηλεφώνων για να αποθανατίσουν τις ουρές των αμαξιών που βρίσκονται μπροστά τους. Δύο τέτοια παραδείγματα παρατίθενται παρακάτω (Εικόνα 7, Εικόνα 8), με τους χρήστες να έχουν ανεβάσει την φωτογραφία τους πρωτα στο κοινωνικό δίκτυο Instagram [42] και έπειτα τις αναδημοσιεύουν από το λογαριασμό τους στο Twitter μαζί με το μήνυμα το οποίο εμπεριέχει την λέξη “traffic”.



Εικόνα 7: Φωτογραφία χρήστη
Πηγή: instagram.com

Οποιοσδήποτε πάντως και να είναι ο λόγος για τον οποίο γράφεται το κάθε tweet , ή οποιοσδήποτε και να είναι ο τρόπος που εκφράζεται ο κάθε χρήστης, αυτό που έχει σημασία είναι ότι η πληροφορία της ύπαρξης της κυκλοφοριακής συμφόρησης κοινοποιείται στο μέσο κοινωνικής δικτύωσης σε συνδυασμό πάντα με την τοποθεσία από όπου έχει σταλεί.

Εικόνα 8: Φωτογραφία χρήστη
Πηγή: instagram.com

3.2.3 Χαρτογράφηση σε πραγματικό χρόνο

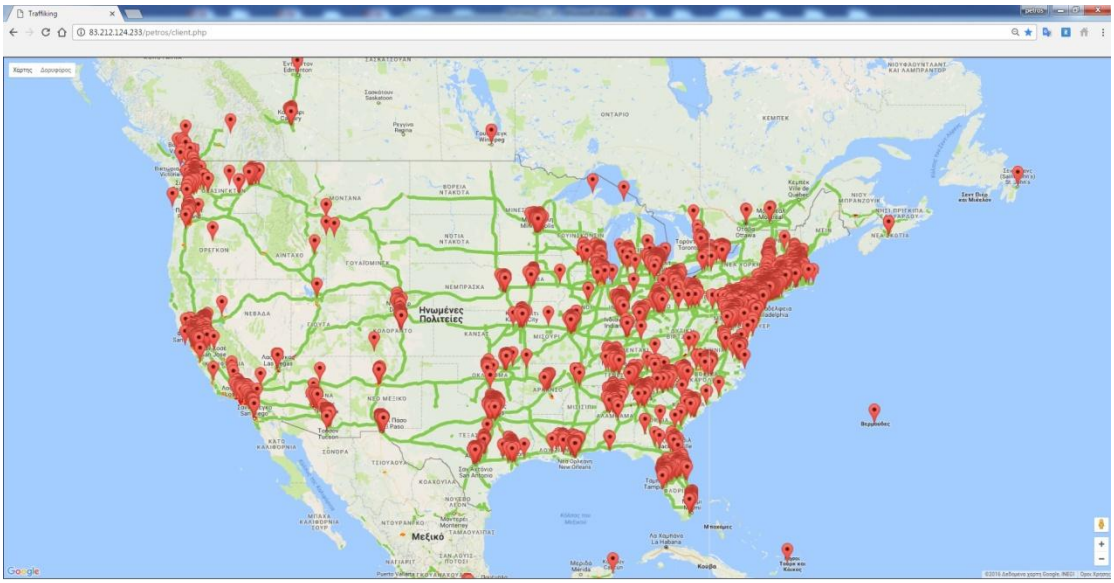
Έως τώρα έχει περιγραφεί η διαδικασία με την οποία τα tweets , σε πραγματικό χρόνο, αποθηκεύονταν στην αρχική βάση δεδομένων. Ένα από τα βασικά μέρη του λογισμικού που δημιουργήθηκε για την κάλυψη των αναγκών της παρούσας εργασίας είχε να κάνει με την χαρτογράφηση των σημείων που αποθηκεύονταν στην βάση δεδομένων επίσης σε πραγματικό χρόνο. Καθώς τα μηνύματα αποθηκεύονταν στην βάση δεδομένων, αυτομάτως μέσω του γεωγραφικού μήκους και πλάτους που τα συνόδευαν, γινόταν η τοποθέτησή τους σε ένα παγκόσμιο διαδικτυακό χάρτη ως σημεία. Στο Twitter όλα τα μηνύματα που περιέχουν ακριβή πληροφορία θέσης συνοδεύονται από γεωγραφικό μήκος και πλάτος στο παγκόσμιο γεωδαιτικό σύστημα αναφοράς(WGS84) καθώς γίνεται χρήση του GPS των έξυπνων τηλεφώνων για την εύρεση θέσης.

Για την χαρτογράφηση των μηνυμάτων δημιουργήθηκε μια ιστοσελίδα η οποία αποτέλεσε ένα διαδικτυακό Γεωγραφικό Σύστημα Πληροφοριών (Web-GIS)προσαρμοσμένο στις ανάγκες τις εργασίας. Τα συστήματα Web-GIS αποτελούν μία πλατφόρμα που έχουν τη δυνατότητα να παρέχουν ευέλικτα εργαλεία στο χειρισμό των διαθέσιμων γεωγραφικών δεδομένων, συντελώντας με αυτό τον τρόπο στην δημιουργία ενός οπτικού και δυναμικού χάρτη στην οθόνη ενός ηλεκτρονικού υπολογιστή(13). Το Web-GIS που δημιουργήθηκε ανακαλούσε σε πραγματικό χρόνο ένα χάρτη από την υπηρεσία Google Maps. Όπως έχει αναφερθεί ήδη η Google παρέχει μια υπηρεσία μέσω της εφαρμογής της Google Maps, το Google Traffic. Στην υπηρεσία αυτή ο χρήστης μπορεί να ενημερωθεί που υπάρχει κίνηση στους δρόμους σε πραγματικό χρόνο. Στην ουσία η υπηρεσία αποτελείται από ένα Web-GIS το οποίο παρέχει ένα παγκόσμιο, δυναμικό χάρτη με αποτυπωμένους όλους τους δρόμους επίσης σε παγκόσμια κλίμακα. Κάθε οδός χρωματίζεται διαφορετικά ανάλογα με το πόση κίνηση υπάρχει πάντα σε πραγματικό χρόνο. Ο υπολογισμός αυτός γίνεται λαμβάνοντας υπόψη την ταχύτητα των οχημάτων και την χωρητικότητα της κάθε οδού. Η χρωματική κατηγοριοποίηση των οδών ανάλογα με τις κυκλοφοριακές συνθήκες στο Google Traffic έχει ως εξής [43]:

- Βαθύ κόκκινο: Υποδηλώνει ότι στην οδό υπάρχει μποτιλιάρισμα
- Κόκκινο: Υποδηλώνει ότι στην οδό υπάρχει κυκλοφοριακή συμφόρηση
- Κίτρινο: Υποδηλώνει ότι τα οχήματα κινούνται με σχετικά χαμηλές ταχύτητες
- Πράσινο: Υποδηλώνει ότι η ροή των οχημάτων είναι ελεύθερη
- Γκρι: Υποδηλώνει ότι δεν υπάρχουν δεδομένα για τις συγκεκριμένες οδούς

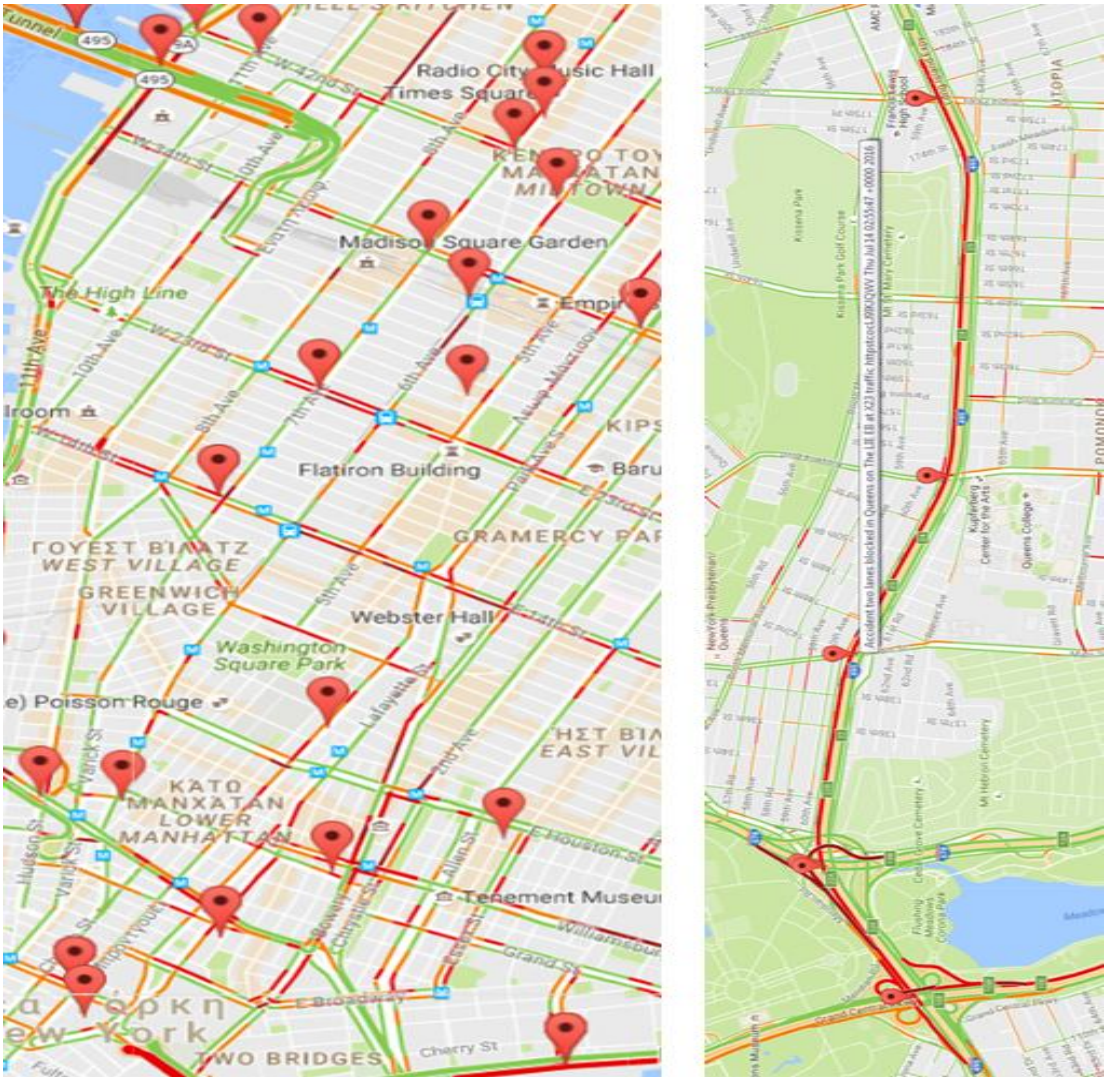
Ο χάρτης της υπηρεσίας Google Traffic λοιπόν χρησιμοποιήθηκε ως βάση για την χαρτογράφηση σε πραγματικό χρόνο των μηνυμάτων που λαμβάνονταν από το Twitter. Η χρήση της συγκεκριμένης εφαρμογής κρίθηκε σκόπιμη καθώς έδινε την δυνατότητα της άμεσης σύγκρισης των δεδομένων του Google Traffic με τα μηνύματα που ανακτούσε το σύστημα. Καθώς τα μηνύματα εμφανίζονταν στον χάρτη με απλή παρατήρηση μπορούσε να εξαχθεί το συμπέρασμα πόσα από αυτά

συνέπιπταν με περιοχές όπου το Google Traffic υποδείκνυε ότι υπάρχει κυκλοφοριακή συμφόρηση. Η σύγκριση αυτή ήταν εφικτή γιατί και οι δύο ομάδες δεδομένων εμφανίζονταν στην ιστοσελίδα σε πραγματικό χρόνο καθώς ο χάρτης από το Google Traffic ανακαλούνταν μέσω του API της Google σε πραγματικό χρόνο και ανανεωνόταν ανά λίγα δευτερόλεπτα, όπως επίσης δυναμική με τον χρόνο ήταν και η εμφάνιση των μηνυμάτων στον χάρτη. Στιγμιότυπα της εφαρμογής με τοποθετημένα τα μηνύματα ως σημεία παρατίθεται παρακάτω. Στο πρώτο εμφανίζεται όλη η περιοχή που καλύπτουν οι Ηνωμένες Πολιτείες Αμερικής ενώ στα επόμενα παρουσιάζονται τμήμα χάρτη από την περιοχή της Νέας Υόρκης σε μεγάλη κλίμακα όπου διακρίνονται οι οδοί με των χρωματικό διαχωρισμό από το Google Traffic.



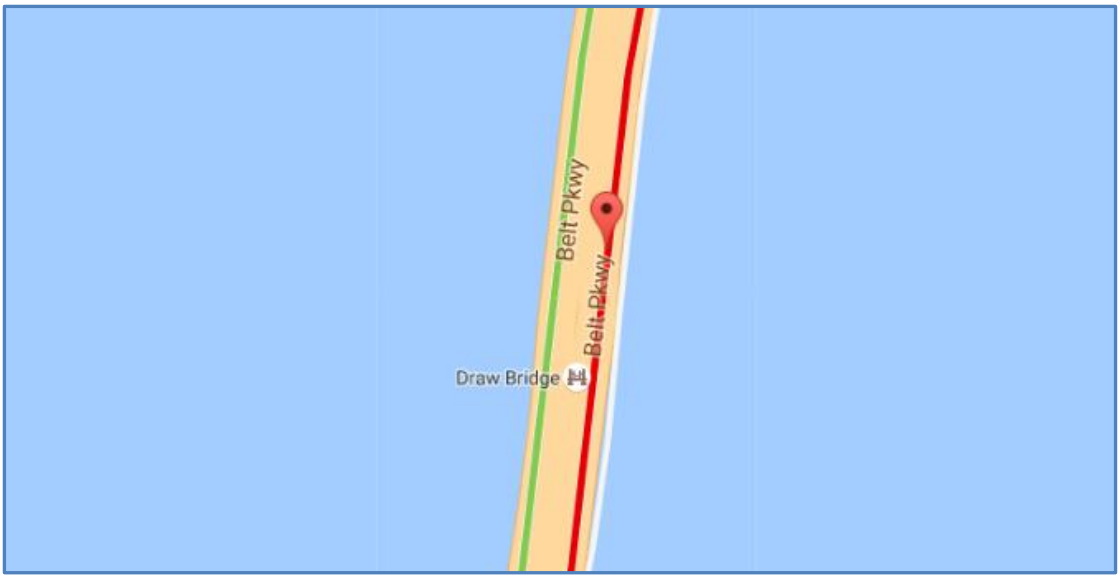
Εικόνα 9: Περιβάλλον διαδικτυακής εφαρμογής με αποτυπωμένα 9000 μηνύματα

Στο πρώτο στιγμιότυπο (Εικόνα 9) τα σημεία απεικονίζονται σωρευτικά καθώς η ανάκτηση τους διήρκησε 3 ημέρες (9000 σημεία περίπου) ενώ στο δεύτερο στιγμιότυπο (Εικόνα 10) η ανάκτηση διήρκησε λίγα λεπτά ώστε να συνάδουν τα δεδομένα από το Google Traffic με την τοποθεσία των μηνυμάτων. Όπως διακρίνεται λοιπόν από το δεύτερο στιγμιότυπο τα μηνύματα τοποθετούνται σε οδούς που είναι χρωματισμένοι με κόκκινο εκεί δηλαδή που αναφέρει το Google Traffic ότι υπάρχει κυκλοφοριακή συμφόρηση. Επίσης διακρίνεται και ένα από τα μηνύματα που εμφανίζονται ως σημεία στον χάρτη καθώς η εφαρμογή έχει σχεδιαστεί με τέτοιο τρόπο ώστε όταν ο χρήστης τοποθετήσει τον κέρσορα πάνω στο σημείο να εμφανίζεται το κείμενο του μηνύματος.



Εικόνα 10: Περιβάλλον διαδικτυακής εφαρμογής με απεικόνιση μηνυμάτων σε μεγάλη κλίμακα

Η λειτουργία GPS που διαθέτουν οι φορητές συσκευές των χρηστών και μέσω αυτού εντοπίζεται η θέση τους, μπορούν να έχουν ακρίβεια μέχρι και 5 μέτρα στην εύρεση της θέσης ανάλογα με τις περιβάλλουσες συνθήκες [44]. Παρατηρήθηκε ότι η τοποθέτηση των tweets στον χάρτη σε πολλές περιπτώσεις, λόγω της ικανοποιητικής ακρίβειας του GPS των συσκευών, επέτρεπε τον διαχωρισμό των δύο λωρίδων στις μεγάλες οδικές αρτηρίες ανάλογα με την κίνηση που επικρατούσε στην κάθε μια. Ένα παράδειγμα του ίδιου δρόμου που στην μια λωρίδα υπάρχει συμφόρηση και στην άλλη όχι παρατίθεται παρακάτω (Εικόνα 11) αναδεικνύοντας την χωρική ανάλυση της πληροφορίας που λαμβάνονται από τα tweets.



Εικόνα 11: Η χωρική ανάλυση των πληροφοριών φτάνει σε επίπεδο λωρίδας

Το λογισμικό που αναπτύχθηκε και παρουσιάστηκε παραπάνω στο σύνολο του αποτελεί ένα χρήσιμο εργαλείο εκμετάλλευσης δεδομένων από το Twitter για την εύρεση θέσεων κυκλοφοριακής συμφόρησης στο οδικό δίκτυο σε πραγματικό χρόνο. Μια τέτοια εφαρμογή θα ήταν πολύ χρήσιμη για τις αρχές τις κάθε πόλης καθώς με πολύ μικρό κόστος, σε σχέση με τις συμβατικές μεθόδους, θα μπορούν να εντοπίζουν σε πραγματικό χρόνο που υπάρχει συμφόρηση και να λαμβάνουν τα ανάλογα μέτρα. Επίσης η εφαρμογή αυτή θα ήταν πολύ χρήσιμη για τους οδηγούς καθώς θα μπορούσαν να ενημερώνονται μέσω αυτής για το που υπάρχει πρόβλημα στους δρόμους και να επιλέξουν αναλόγως την διαδρομή που θα ακολουθήσουν. Το πρόβλημα που υπάρχει ,όπως έχει αναφερθεί ήδη, είναι ότι η υλοποίηση της παραπάνω εφαρμογής έγινε με δεδομένα τα οποία ανακτήθηκαν δωρεάν από το Twitter με αποτέλεσμα να λαμβάνονται μόνο το 1% των μηνυμάτων που θα μπορούσαν να ανακτηθούν. Για να μπορέσει μια τέτοια εφαρμογή να έχει τον απαραίτητο όγκο δεδομένων θα πρέπει να χρησιμοποιεί όλα τα διαθέσιμα μηνύματα στο Twitter τα οποία προσεγγίζονται στις 210,000 παγκοσμίως μέσα σε ένα εικοσιτετράωρο. Επίσης μπορούν να ανακτούνται ταυτόχρονα μηνύματα από πολλές υπηρεσίες micro-blogging ανάλογες με Twitter που σε αυτή την περίπτωση ο όγκος τον δεδομένων θα εκτινασσόταν σε πολλές χιλιάδες μηνυμάτων το λεπτό. Ενδεικτικά, μια από τις πολλές εναλλακτικές εφαρμογές του Twitter που χρησιμοποιείται κυρίως στην Ασία είναι το Weibo το οποίο αριθμεί 600 εκατομμύρια χρήστες. Το παραπάνω πρόβλημα βέβαια αναφέρεται στην χρήση των δεδομένων σε πραγματικό χρόνο καθώς όπως στην παρούσα εργασία, τα δεδομένα αυτά μπορούν να συλλεχθούν για μεγάλο χρονικό διάστημα και να απολέσουν βάση για στατιστική ανάλυση η οποία θα παρουσιαστεί παρακάτω.

3.2.4 Αξιολόγηση των δεδομένων σε πραγματικό χρόνο

Η χρήση την υπηρεσίας Google Traffic ως υπόβαθρο για την χαρτογράφηση των μηνυμάτων έγινε για να διενεργηθεί μια πρώτη αξιολόγηση των δεδομένων που ελήφθησαν από το Twitter. Έτσι, πραγματοποιήθηκε μια χωρική σύγκριση ανάμεσα στις θέσεις των μηνυμάτων που εμφανίζονταν στον χάρτη σε πραγματικό χρόνο και στους δρόμους που υποδείκνυε το Google Traffic ότι υπάρχει συμφόρηση. Η σύγκριση έγινε με την παρατήρηση του χάρτη καθώς τα μηνύματα εμφανίζονταν και πραγματοποιήθηκε για δείγμα 500 μηνυμάτων ο αναλογικός αυτός τρόπος ακολουθήθηκε καθώς η Google δεν διαθέτει τα στοιχεία της κίνησης σε κάποια άλλη μορφή ώστε να μπορούσε να γίνει αυτόματη σύγκριση. Στον παρακάτω πίνακα (Πίνακας 4) παρουσιάζονται πόσα σημεία μηνυμάτων συνέπεσαν στο κάθε χρωματικό επίπεδο του Google Traffic. Όπως διαφαίνεται από τον πίνακα στο δείγμα των 500 σημείων το 89,4% των μηνυμάτων δώσανε την σωστή πληροφορία ότι υπάρχει πρόβλημα με την κίνηση ενώ το 80,6% «ανίχνευσε» ότι υπάρχει μποτιλιάρισμα ή κυκλοφοριακή συμφόρηση πράγμα που αποδεικνύει , πάντα σύμφωνα με τα δεδομένα του Google Traffic, ότι ένα μεγάλο ποσοστό των μηνυμάτων που ανακτήθηκαν είναι αξιόπιστη πηγή πληροφόρησης για την κίνηση στους δρόμους. Αξιοσημείωτο είναι επίσης το γεγονός ότι μόλις το 1,8% του δείγματος ήταν μηνύματα που αναφέρονταν σε κάτι διαφορετικό από την κυκλοφοριακή κίνηση χρησιμοποιώντας την λέξη “traffic”. Αυτό το μικρό ποσοστό των μηνυμάτων αφαιρέθηκε εκ των υστέρων από την βάση δεδομένων πριν γίνει η επεξεργασία τους και θα γίνει αναφορά παρακάτω.

Καταμέτρηση μηνυμάτων		
Επίπεδα	Πλήθος μηνυμάτων	Ποσοστό
Βαθύ κόκκινο (μποτιλιάρισμα)	269	53,8%
Κόκκινο (κυκλοφοριακή συμφόρηση)	135	27%
Κίτρινο (σχετικά χαμηλές ταχύτητες)	43	8,6%
Πράσινο (ελεύθερη ροή)	25	5%
Δεν υπήρχαν δεδομένα από το Google Traffic	19	3,8%

Πίνακας 4:Αξιολόγηση των πληροφοριών σύμφωνα με το Google Traffic

3.2.5 Εκκαθάριση των δεδομένων

Μετά το πέρας της τρίμηνης διαδικασίας ανάκτησης των μηνυμάτων από το Twitter και την συγκρότηση της αρχικής βάσης δεδομένων ακολούθησε το στάδιο του καθαρισμού της βάσης από μηνύματα τα οποία δεν είχαν σχέση με το αντικείμενο της έρευνας. Τα μηνύματα αυτά ,όπως έχει αναφερθεί, περιείχαν την λέξη “traffic” στο κείμενο τους αλλά αναφέρονταν σε κάτι διαφορετικό από κυκλοφοριακή συμφόρηση. Να σημειωθεί στο σημείο αυτό ότι το API του Twitter, όταν τεθεί μια λέξη κλειδί για αναζήτηση, επιστρέφει και αποτελέσματα στα υπάρχουν και σύνθετες λέξεις οι οποίες περιέχουν την λέξη κλειδί. Στην περίπτωση μας ήταν πολύ πιθανό να εμφανιστούν μηνύματα με λέξεις όπως για παράδειγμα “trafficking” τα οποία θα αποτελούσαν θόρυβο για τα δεδομένα μας. Έτσι πριν από την ανάλυση των δεδομένων πραγματοποιήθηκε μια εκκαθάριση της βάσης από τα μηνύματα που

αναφέρονταν σε κάτι διαφορετικό από την κίνηση στους δρόμους είτε χρησιμοποιώντας το “traffic” μεμονωμένα, είτε σε σύνθετη λέξη.

Η αναζήτηση των μηνυμάτων που δεν σχετίζονταν με την κίνηση έγινε με παρατήρηση ενός δείγματος περίπου 1000 μηνυμάτων και εύρεση των λέξεων που συνόδευαν συνήθως των λέξη “traffic” θέλοντας να περιγράψουν μια συγκεκριμένη κατάσταση. Έπειτα πραγματοποιήθηκε αναζήτηση αυτών των λέξεων σε περιβάλλον λογιστικών φύλλων και διαγράφηκαν από την αρχική βάση δεδομένων όλα τα μηνύματα που τις περιείχαν. Στον παρακάτω πίνακα (Πίνακας 5) φαίνονται όλες οι λέξεις που συνόδευαν την λέξη “traffic” και δεν σχετίζονταν με κυκλοφοριακή κίνηση, παραδείγματα και ο αριθμός των μηνυμάτων αυτών που εντοπίστηκαν στο σύνολο της βάσης των 150,081 εγγραφών.

Λέξη	**Παράδειγμα**	**Πλήθος**
Light(s)	I'm at Traffic Light Padang Jawa	369
Internet	7 Figure Traffic Academy. #internetmarketing	12
Blog	How to generate traffic for your blog	97
Trend(s)	Trend Alert: 'Traffic Update'. More trends	137
Website(s)	Do RSS Feeds Detract From Website Traffic Trends?	106
air	I'm at Phuket Air Traffic Control Centre	152
cleared	Accident cleared in #Livingston on Hwy 16 SB between Hwy 64 and before US 190 #traffic	6584

Πίνακας 5: Εύρεση μηνυμάτων που δεν είναι χρήσιμα στην παρούσα εργασία

Όπως διαφαίνεται από τον παραπάνω πίνακα τα περισσότερα μηνύματα που εντοπίστηκαν περιείχαν την λέξη “cleared”. Όλα τα μηνύματα της ομάδας αυτής προέρχονταν από επαγγελματικούς λογαριασμούς και θέλανε να εκφράσουν ότι ο δρόμος είναι ελεύθερος μετά από κάποιο ατύχημα που είχε προηγηθεί με αποτέλεσμα να επέλθει η κανονικότητα στην οδό. Είναι λογικό να αποκλειστούν τέτοιου είδους μηνύματα καθώς αναφέρονταν σε μια κατάσταση που είχε προηγηθεί της ώρας που κοινοποιούνταν το μήνυμα και το γεγονός αυτό θα επηρέαζε την χρονική ανάλυση που θα γίνει παρακάτω.

Εκτός από την παραπάνω διαδικασία αναζητήθηκαν διάφορες λέξεις οι οποίες είχαν ως συνθετικό τη λέξη “traffic”. Οι λέξεις αυτές προήλθαν από ερευνητική εργασία [29] ,για την οποία έγινε αναφορά προηγουμένως, όπου ο ερευνητής αναζητούσε επίσης tweets που να περιείχαν την λέξη “traffic” αλλά από ένα σύνολο πολλών εκατομμυρίων tweet, διαχωρίζοντας αυτά που χρειάζεται με αλγορίθμους Machine Learning. Έτσι, στην εργασία παρατίθενται οι λέξεις που βρέθηκαν αλλά δεν σχετίζονται με την κυκλοφοριακή συμφόρηση οι οποίες και χρησιμοποιήθηκαν και στην παρούσα εργασία για την εκκαθάριση των δεδομένων. Στον παρακάτω πίνακα (Πίνακας 6) εμφανίζονται οι συγκεκριμένες λέξεις και από δίπλα ο αριθμός των tweets που τις περιείχαν.

Λέξη κλειδί	Πλήθος
humantrafficking	14
trafficking	21
traficant	0
sextrafficking	3
trafficked	0
trafficker	0

Πίνακας 6: Σύνθετες λέξεις που δεν αναφέρονται σε κυκλοφοριακή συμφόρηση

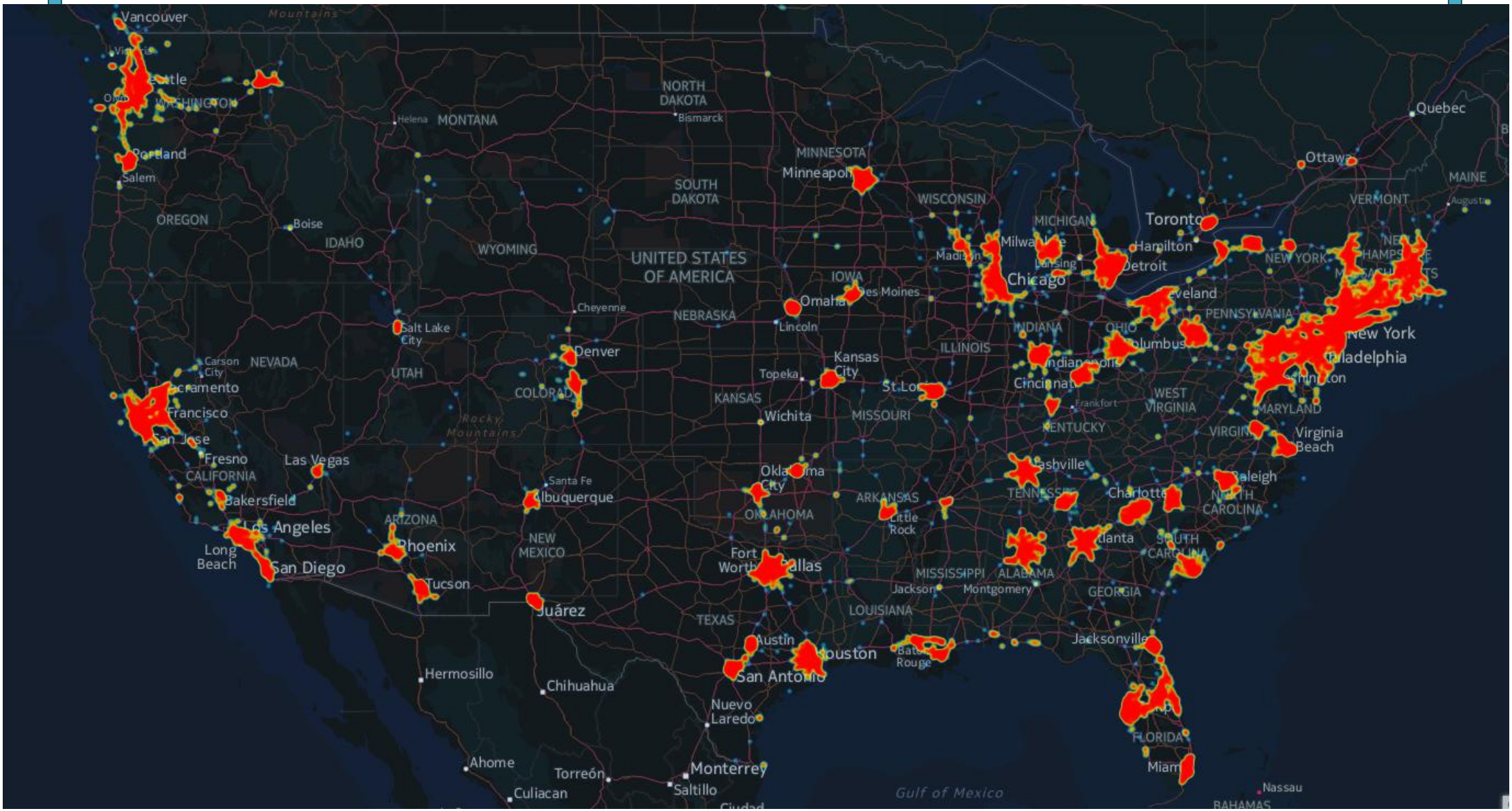
Εάν εξαιρεθούν τα μηνύματα που περιείχαν την λέξη “cleared” τα οποία είναι μια ιδιάζουσα περίπτωση, γίνεται αντιληπτό ότι τα υπόλοιπα μηνύματα που τελικά αποκλείστηκαν από την βάση δεδομένων είναι πολύ λίγα (911) σε σχέση με τον αριθμό των εγγραφών όλης της βάσης, με ποσοστό μόλις 0,61%. Το παραπάνω φαινόμενο μπορεί να εξηγηθεί λόγω της συνθήκης που έχουμε θέσει στο Twitter API ότι χρειαζόμαστε μόνο δεδομένα τα οποία περιέχουν πληροφορία θέσης. Οι χρήστες ενεργοποιούν συνήθως τις υπηρεσίες εύρεσης θέσης για να δηλώσουν την τοποθεσία των μηνυμάτων τους μόνο όταν υπάρχει κάποιος λόγος να κοινοποιηθεί η θέση τους στο κοινωνικό δίκτυο. Κατά αυτό τον τρόπο, δεν είναι σύνηθες να κοινοποιηθεί, για παράδειγμα, η τοποθεσία σε ένα μήνυμα το οποίο αναφέρεται στην διαδικτυακή κινητικότητα μιας ιστοσελίδας (Website traffic). Από την άλλη μεριά στα μηνύματα που κοινοποιούνται για να αναδείξουν κάποιο πρόβλημα στους δρόμους, έχει λογική κάποιοι χρήστες να μοιραστούν την θέση τους με τους άλλους χρήστες ώστε να τους ενημερώσουν για την κατάσταση με σκοπό να αλλάξουν κατεύθυνση και να μην πλησιάσουν στο σημείο. Αυτή ήταν άλλωστε η αρχική λογική που έδωσε το έναυσμα για την πραγματοποίηση της παρούσας εργασίας η οποία επαληθεύεται εκ του αποτελέσματος.

ΚΕΦΑΛΑΙΟ 4: ΕΚΤΟΣ ΣΥΝΔΕΣΗΣ ΑΝΑΛΥΣΗ

Όπως παρουσιάστηκε παραπάνω πραγματοποιήθηκε εκκαθάριση των δεδομένων από τα μηνύματα που ανακτήθηκαν και δεν αναφέρονταν στο αντικείμενο που μελετά η εργασία. Η νέα βάση δεδομένων περιείχε 142,586 εγγραφές από 150,081 που υπήρχαν στην αρχική. Καθώς είχε αποφασιστεί ότι η εργασία θα ασχοληθεί μόνο με την χώρα των ΗΠΑ, απομονώθηκαν με γραφικές μεθόδους μόνο τα σημεία που βρίσκονταν στις ΗΠΑ. Τελικά δημιουργήθηκε μια βάση δεδομένων η οποία περιείχε 126,659 εγγραφές η οποία αποτέλεσε βάση για τις διάφορες εκτός σύνδεσης αναλύσεις και επεξεργασίες που θα παρουσιαστούν παρακάτω. Στον παρακάτω χάρτη απεικονίζονται όλα τα 126,659 σημεία όπως προέκυψε από την πρώτη χαρτογράφηση τους.

Όπως έχει αναφερθεί ήδη οι εκτός σύνδεσης επεξεργασίες χωρίζονται σε δυο κατηγορίες, ,της χωρικής ανάλυσης των δεδομένων και της χρονικής ανάλυσης. Αρχικά θα παρουσιαστεί η χρονική ανάλυση η οποία περιλαμβάνει την εξαγωγή από τα δεδομένα την ώρα αιχμής για 20 πόλεις των ΗΠΑ για τις οποίες ανακτήθηκαν τα περισσότερα δεδομένα και σύγκριση των αποτελεσμάτων με τα αντίστοιχα στοιχεία της εταιρίας Tomtom. Έπειτα επιχειρείται να δημιουργηθεί μια σειρά από χάρτες οι οποίοι θα αναδεικνύουν τα σημεία εντός της πόλης του Houston όπου επικεντρώνεται η περισσότερη ροή των οχημάτων σε σχέση με την ώρα μέσα σε εικοσιτετράωρο.

Κατά την χωρική ανάλυση δημιουργήθηκε ένας χάρτης της Νέας Υόρκης ο οποίος συνδυάζει τα σημεία που υποδεικνύουν που υπάρχει κίνηση με τους αντίστοιχους δρόμους. Τελικά ο χάρτης απεικονίζει σε ποια τμήματα δρόμων της πόλης υπάρχει μεγαλύτερη πιθανότητα να εμφανιστεί κίνηση και το αποτέλεσμα συγκρίνεται με τα αντίστοιχα δεδομένα της εφαρμογής Google Traffic και της εταιρίας INRIX. Ανάλογη ανάλυση γίνεται και με τα μηνύματα που αναφέρονται σε κάποιο τροχαίο ατύχημα με αποτέλεσμα να εξαχθούν τα περισσότερο επικίνδυνα τμήματα δρόμων στην πόλη της νέας Υόρκης. Τέλος επιχειρείται η κατάταξη 10 πόλεων των ΗΠΑ ανάλογα με την συμφόρηση που βιώνουν οι πολίτες στην κάθε μια σύμφωνα με την χωρική αυτοσυσχέτιση που υπολογίσθηκε για τις διάφορες περιοχές των πόλεων. Τα αποτελέσματα συγκρίνονται με τα αντίστοιχα της εταιρίας Tomtom.



Απεικόνιση 125.000 μηνυμάτων από τις Ηνωμένες Πολιτείες Αμερικής

4.1 Διαχωρισμός των δεδομένων

Για όλες τις παρακάτω επεξεργασίες αρχικό βήμα ήταν ο διαχωρισμός των δεδομένων ανάλογα με την πόλη στην οποία αναφέρονται. Ένα βασικό πρόβλημα ήταν πως θα καθοριστούν τα όρια της κάθε πόλης ώστε να εισαχθούν σε περιβάλλον GIS και να γίνει η συσχέτιση των σημείων με το πολύγονο που ορίζει την κάθε πόλη. Το θέμα αυτό ήταν μεγάλης σημασίας καθώς όλες οι εταιρίες τα δεδομένα των οποίων χρησιμοποιούνται για να ελενχθούν τα αντίστοιχα της παρούσας εργασίας δεν αναφέρουν πως ορίζεται η κάθε πόλη άρα θα έπρεπε να επιλεχθούν τα όρια της κάθε πόλης με τον περισσότερο αντικειμενικό τρόπο. Τελικά η λύση βρέθηκε από την επίσημη αρχή των ΗΠΑ που ευθύνεται για τις απογραφές στην χώρα το Bureau of the Census [45]. Στην ιστοσελίδα της υπηρεσίας υπήρχαν δεδομένα σε μορφή κατάλληλη για επεξεργασία σε περιβάλλον GIS, που καθορίζουν διάφορες στατιστικές περιοχές μικρής και μεγάλης κλίμακας οι οποίες έχουν καθοριστεί επίσημα από την κυβέρνηση των ΗΠΑ. Οι περιοχές αυτές (census-designated places) είναι επίσημα καθορισμένες και προορίζονται για στατιστική ανάλυση χωρίς να ανταποκρίνονται σε άλλα κριτήρια όπως τα διοικητικά. Μετά από τον συνδυασμό μικρών και μεγάλων στατιστικών περιοχών επιτευχθηκε τελικά ο ορισμός της κάθε πόλης που μας ενδιέφερε, διαδικασία που αποδείχθηκε απαιτιτική καθώς πολλές φορές η κάθε πόλη συμπεριλάμβανε πάρα πολλές μικρές στατιστικές περιοχές με αποτέλεσμα η συνένωση τους να είναι χρονοβόρα.

Μετά την δημιουργία των πολυγόνων που ορίζουν την κάθε πόλη που μας ενδιέφερε έγινε διαχωρισμός των δεδομένων και δημιουργήθηκαν 20 καινούργιες βάσεις δεδομένων μια για την κάθε πόλη. Οι 20 βάσεις των πόλεων αριθμούν συνολικά 49000 περίπου εγγραφές αριθμός σαφώς πολύ μικρότερος από την αρχική βάση (126000) γεγονός που οφείλεται στο ότι αποφασίσθηκε να αναλυθούν μόνο 20 πόλεις (με τα περισσότερα μηνύματα) και επίσης στο ότι τα όρια των πόλεων που δημιουργήθηκαν απέκλισαν πολλά μηνύματα που ήταν διάσπαρτα στα περίχωρα των πόλεων ή στις συνδετήριες οδούς. Στον παρακάτω πίνακα (Πίνακας 7) παρουσιάζονται τα σημεία που αντιστοιχούν στην κάθε πόλη τα οποία θα χρησιμοποιηθούν για τις χρονικές και τις χωρικές αναλύσεις.

ΔΙΑΧΩΡΙΣΜΟΣ ΜΗΝΥΜΑΤΩΝ ΑΝΑ ΠΟΛΗ		
ΠΟΛΗ		**ΠΛΗΘΟΣ ΜΗΝΥΜΑΤΩΝ**
1	**Dallas**	**5551**
2	**Houston**	**4586**
3	**New York**	**4505**
4	**Birmingham**	**3590**
5	**Detroit**	**3178**
6	**Philadelphia**	**2715**
7	**Nashville**	**2667**
8	**Indianapolis**	**2622**
9	**El Paso**	**2568**
10	**Columbus**	**2476**
11	**Sacramento**	**2135**
12	**Minneapolis**	**1904**
13	**Newark**	**1901**
14	**Cleveland**	**1716**
15	**Tampa**	**1677**
16	**Miami**	**1605**
17	**Atlanta**	**1383**
18	**Pittsburg**	**1302**
19	**Seattle**	**1270**
20	**San Antonio**	**1236**

Πίνακας 7: Πλήθος μηνυμάτων σε κάθε πόλη

4.2 Χρονική Ανάλυση

4.2.1 Επίπεδο συμφόρησης σε σχέση με τον χρόνο, εξαγωγή ώρας αιχμής

Η ώρα αιχμής (peak hour) ορίζεται ως το χρονικό διάστημα εντός μιας ημέρας κατά την οποία οι πολίτες χρησιμοποιούν περισσότερο το οδικό δίκτυο για να μετακινηθούν. Συνήθως εντοπίζονται δύο χρονικά διαστήματα την ημέρα κατά τα οποία η κίνηση στους δρόμους φτάνει στο υψηλότερο επίπεδο, μία το πρωί και μια το απόγευμα. Πολλές φορές οι τιμές των ωρών αιχμής αλλάζουν από πόλη σε πόλη γεγονός που οφείλεται στις διαφορετικές συνήθειες των κατοίκων της κάθε πόλης ή στις διαφορετικές δραστηριότητες που διαθέτει η κάθε πόλη για τους πολίτες. Για παράδειγμα, μια πόλη η οποία διαθέτει πολλές επιλογές για νυκτερινή ζωή ίσως η απογευματινή ώρα αιχμής να είναι μετατοπισμένη προς το βράδυ σε σχέση με μια πόλη στην ύπαιθρο όπου οι εργασίες είναι κατά κύριο λόγο αγροτικές. Από μετρήσεις έχει βρεθεί ότι κατά την διάρκεια των ωρών αιχμής παρουσιάζονται περί του 60% των συνολικών περιστατικών κυκλοφοριακής συμφόρησης. Για να εξαχθεί τα στατιστικό στοιχείο της ώρας αιχμής χρησιμοποιούνται μετρητές στο οδικό δίκτυο οι

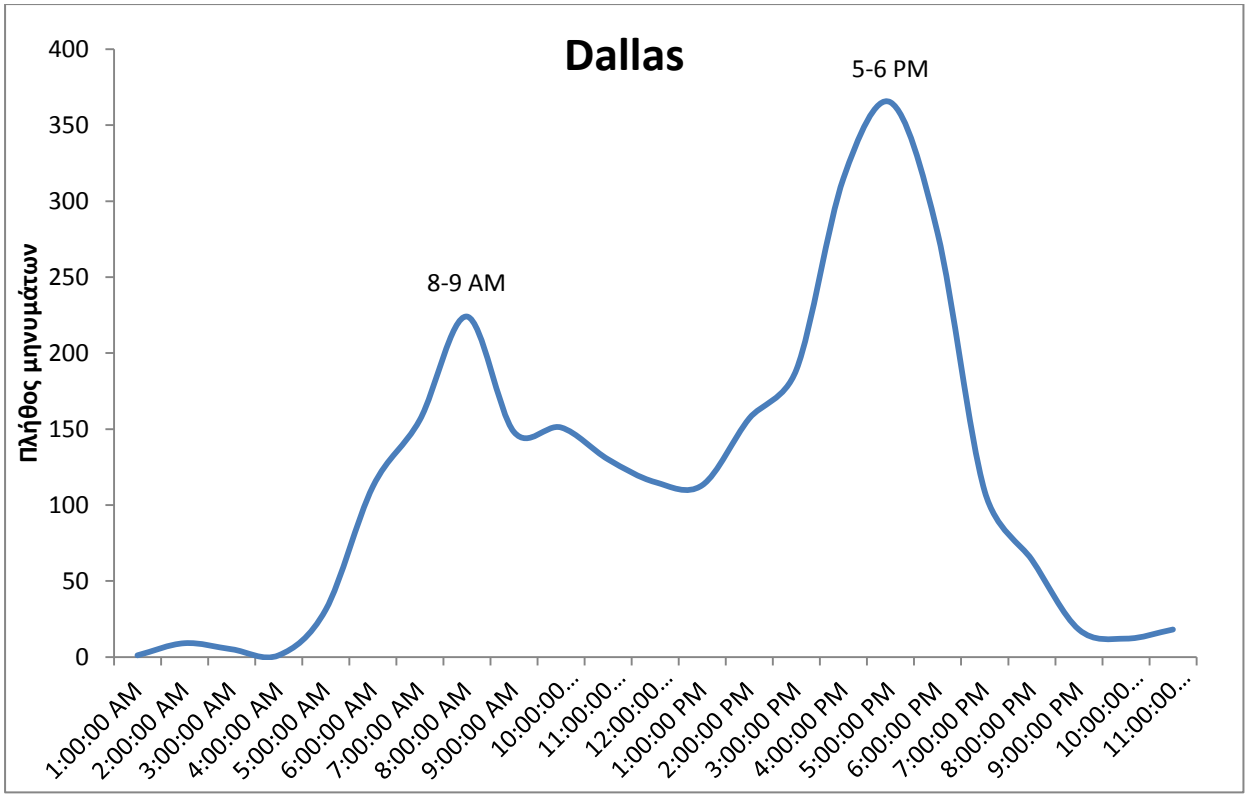
οποίοι υπολογίζουν τους φόρτους των οδών και μετά από επεξεργασία αναδεικνύονται τα χρονικά διαστήματα όπου ο φόρτος είναι μεγαλύτερος. Η ανάδειξη της ώρας αιχμής για μια οδό είναι σημαντική για την καλύτερη διαχείριση της κυκλοφορίας, συνήθως μετά από εξαγωγή της ώρας αιχμής για την κάθε οδό γίνεται επεξεργασία για την εξαγωγή συγκεντρωτικά, των ωρών αιχμής σε επίπεδο πόλης.

Ακολουθώντας την λογική που υπαγορεύει ότι το χρονικό διάστημα που εντοπίζονται οι περισσότερες μαρτυρίες για κυκλοφοριακή συμφόρηση πιθανόν να υπάρχει και ο περισσότερος φόρτος στο οδικό δίκτυο και έχοντας το δεδομένο που αναφέρθηκε παραπάνω ότι 60% των περιπτώσεων συμφόρησης εντοπίζονται στις ώρες αιχμής, έγινε προσπάθεια να υπολογιστούν οι ώρες αιχμής για τις 20 πόλεις στις οποίες ανακτήθηκαν τα περισσότερα δεδομένα. Η κάθε εγγραφή στην βάση δεδομένων περιέχει την πληροφορία της ώρας που στάλθηκε το κάθε μήνυμα έτσι είναι εφικτή η επεξεργασία των δεδομένων για να αναδειχθούν τα χρονικά διαστήματα εντός εικοσιτετραώρου κατά τα οποία λήφθηκαν τα περισσότερα μηνύματα. Για τον σκοπό αυτό δημιουργήθηκαν συγκεντρωτικά διαγράμματα για την κάθε πόλη στα οποία εμφανίζεται το πλήθος των μηνυμάτων που ελήφθησαν ανά ώρα. Τα δεδομένα που χρησιμοποιούνται στο κάθε διάγραμμα λαμβάνονται από την βάση δεδομένων της κάθε πόλης άρα όπως είναι αντιληπτό η κάθε πόλη αντιμετωπίζεται ως σύνολο, γεγονός που ίσως δεν είναι τόσο χρήσιμο όσο η εξαγωγή της ώρας αιχμής κάθε οδού αλλά μια τέτοια ανάλυση δεν ήταν δυνατόν να παρουσιαστεί στην συγκεκριμένη εργασία λόγω έλλειψης δεδομένων. Εκτός από την πληροφορία της ώρας αιχμής τα διαγράμματα των πόλεων που παρουσιάζονται παρακάτω αναδεικνύουν και ποιες άλλες ώρες της ημέρας η συμφόρηση είναι αυξημένη, στοιχείο εξίσου σημαντικό για την καλύτερη κατανόηση της ροής των οχημάτων μέσα στην ημέρα. Κάθε μήνυμα που ανακτήθηκε από το Twitter διέθετε την ακριβή ώρα που κοινοποιήθηκε σύμφωνα με το σύστημα UTC (Coordinated Universal Time) με αποτέλεσμα να χρειαστεί να γίνει μετατροπή στην τοπική ώρα της κάθε πόλης. Η επεξεργασία αυτή αποδείχθηκε απαιτητική δεδομένου ότι οι ΗΠΑ έχουν 9 διαφορετικές χρονικές ζώνες λόγω της μεγάλης έκτασης της χώρας.

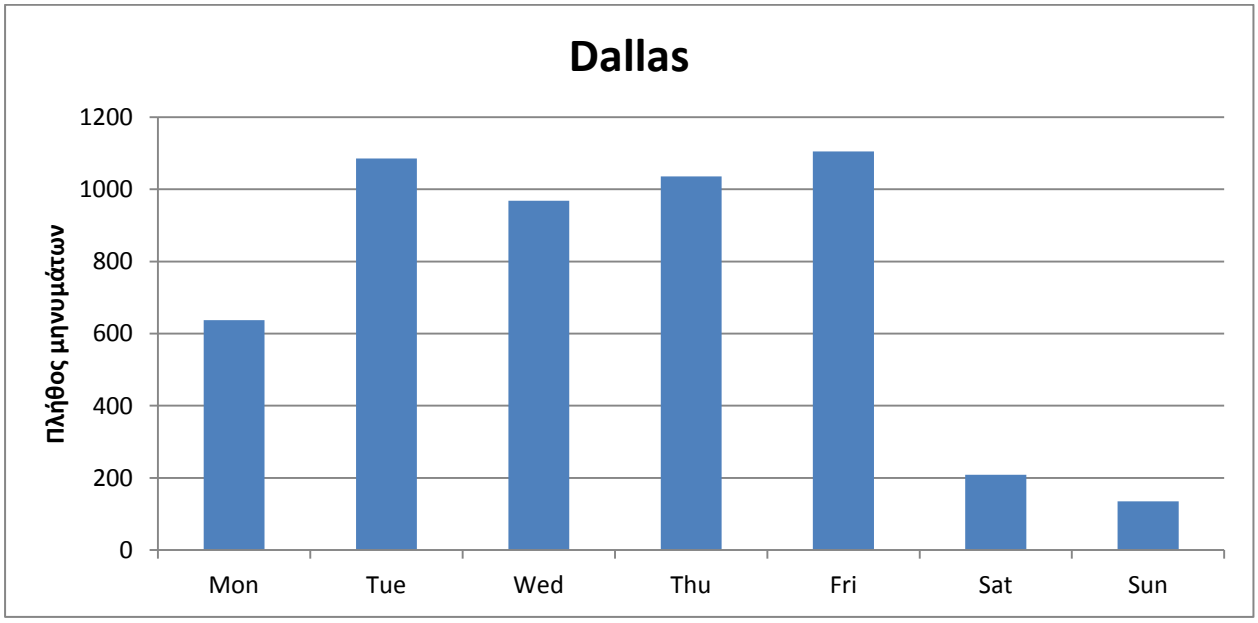
Εκτός από την ώρα αιχμής παρουσιάζεται και μια δεύτερη ανάλυση η οποία έχει να κάνει με την συμφόρηση της κάθε πόλης ανάλογα με την ημέρα εντός της εβδομάδας. Από τα αποτελέσματα είναι ευκρινές ότι στις περισσότερες πόλεις η συμφόρηση είναι σε πολύ χαμηλά επίπεδα τα σαββατοκύριακα χωρίς όμως να υπάρχουν και εξαιρέσεις όπως στην πόλη της Νέας Υόρκης όπου η συμφόρηση του Σαββατοκύριακου κυμαίνεται στα ίδια επίπεδα με την ημέρα της Δευτέρας. Σε πολλές πόλεις διακρίνονται διακυμάνσεις κατά την διάρκεια της εβδομάδας στατιστικό στοιχείο χρήσιμο για την οργάνωση αποτελεσματικού σχεδίου για την διαχείριση της κυκλοφορίας.

Στις παρακάτω σελίδες παρουσιάζονται τα προφίλ των πόλεων σύμφωνα με την χρονική ανάλυση που διεξήχθη. Κατά την διάρκεια της επεξεργασίας για την αποτύπωση της συμφόρησης σε κάθε πόλη ανά ώρα, δημιουργήθηκαν 48 κλάσεις,

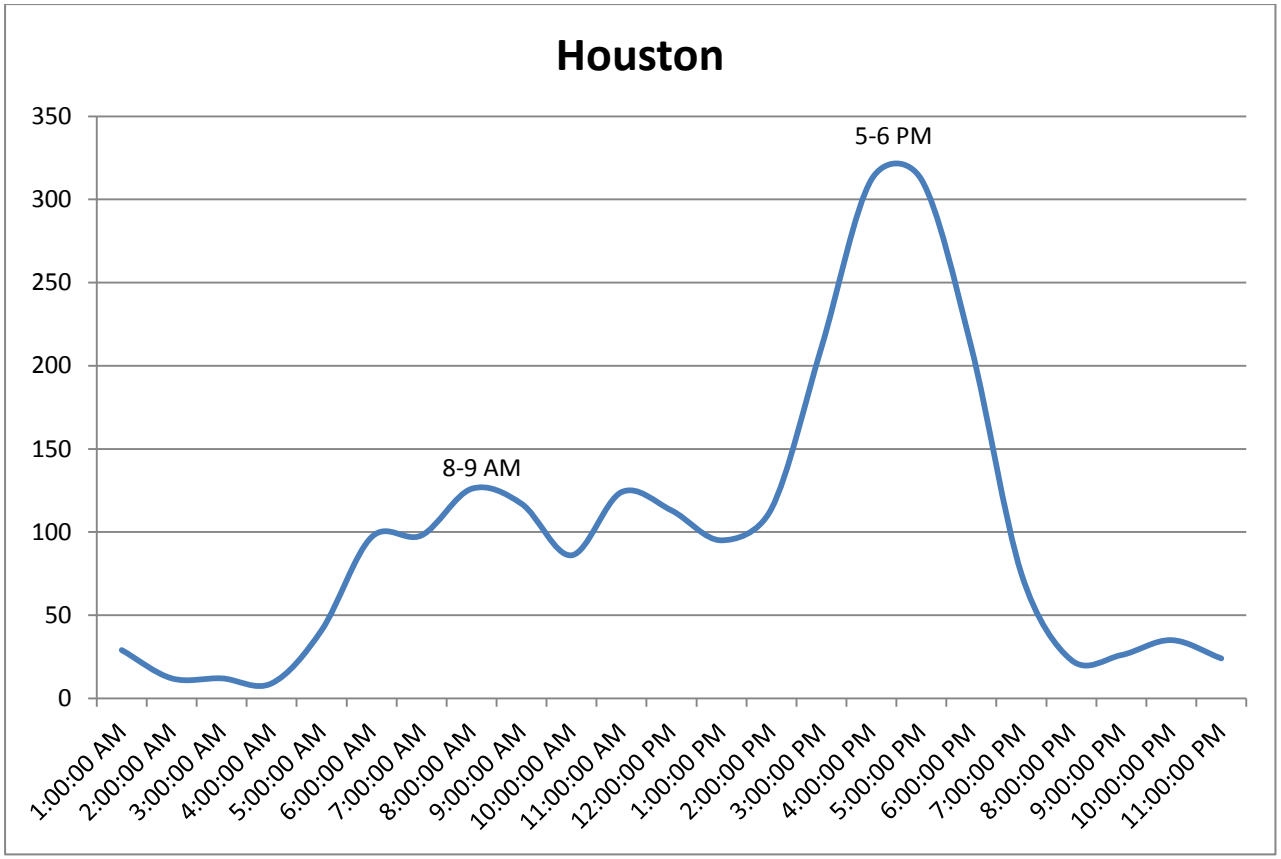
μια για κάθε μισή ώρα της ημέρας, όπου και αντιστοιχήθηκαν όλα τα μηνύματα. Ανάλογη επεξεργασία πραγματοποιήθηκε για την αποτύπωση της συμφόρησης ανάλογα την ημέρα. Σε κάθε προφίλ παρουσιάζεται το διάγραμμα με την συμφόρηση ανά ώρα από όπου εξάγεται και η ώρα αιχμής και το διάγραμμα με την συμφόρηση ανά ημέρα. Επιλέχθηκε να παρατεθούν τα διαγράμματα μόνο 10 πόλεων ενώ συγκεντρωτικά τα αποτελέσματα παρατίθενται στον πίνακα που ακολουθεί στον οποίο συμπεριλαμβάνονται και οι ώρες αιχμής όπως παρουσιάστηκαν στην ετήσια έκθεση της εταιρίας Tomtom για αξιολόγηση των αποτελεσμάτων.



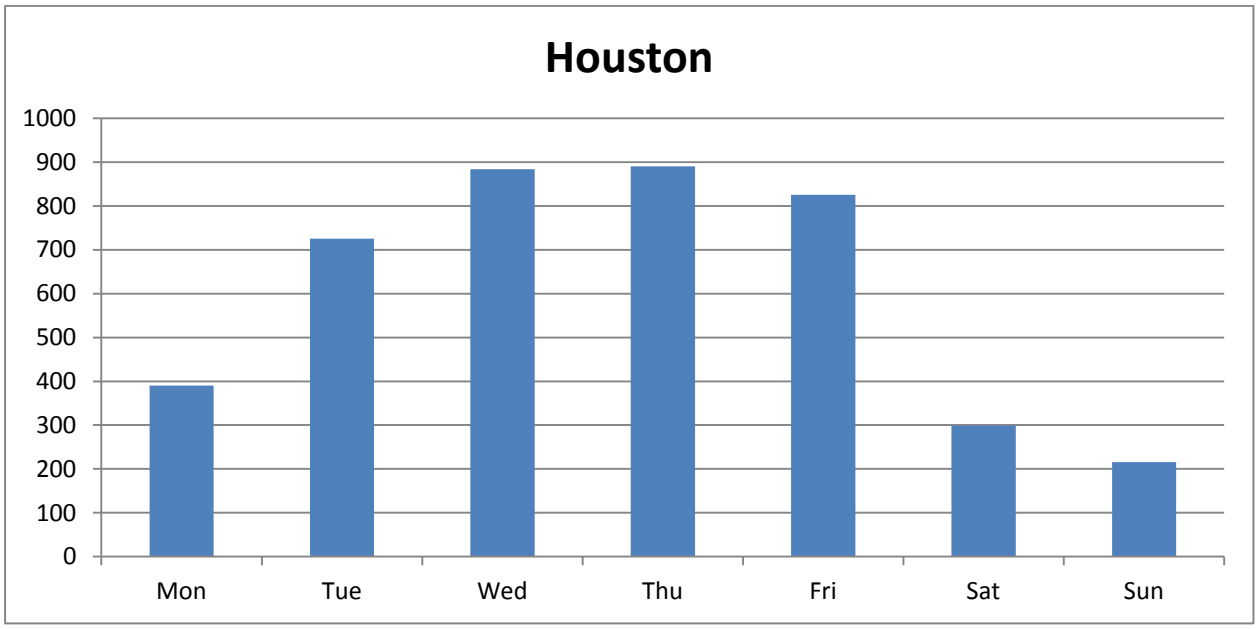
Γράφημα 1: Κατανομή συμφόρησης ανάλογα την ώρα, Dallas



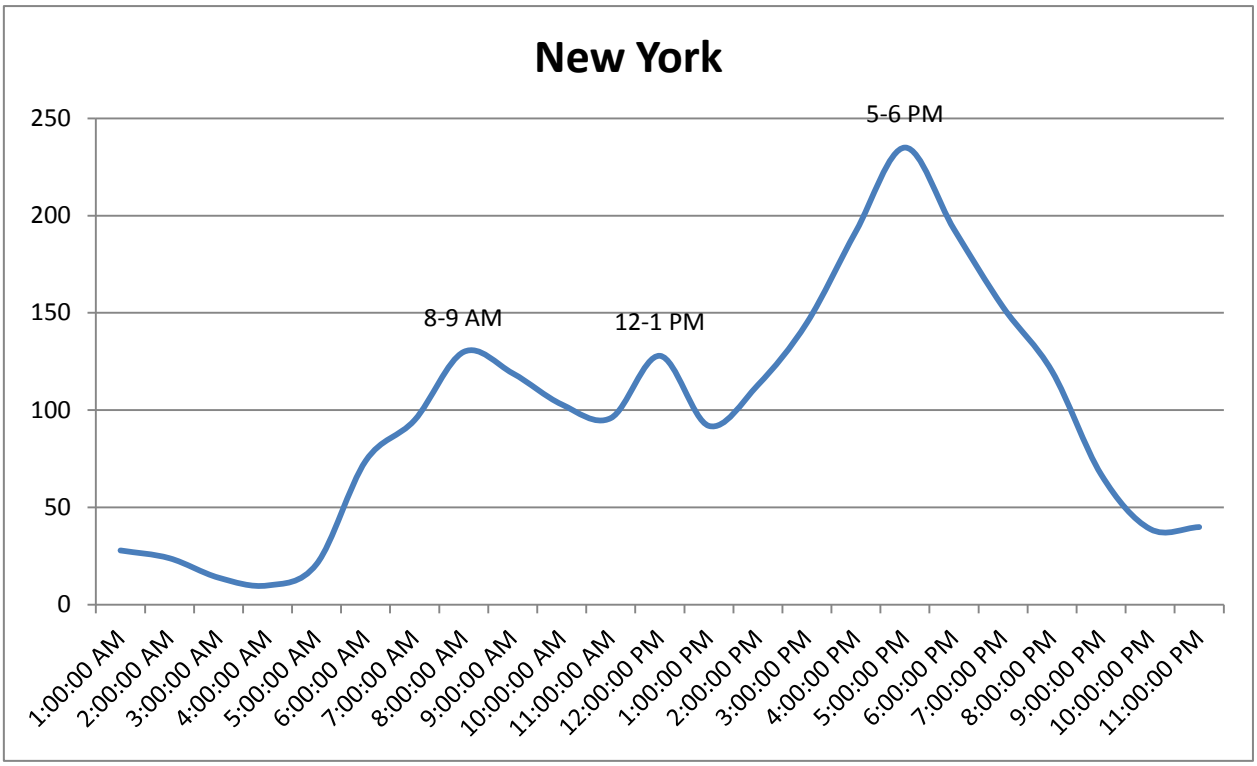
Γράφημα 2: Κατανομή συμφόρησης ανάλογα την ημέρα, Dallas



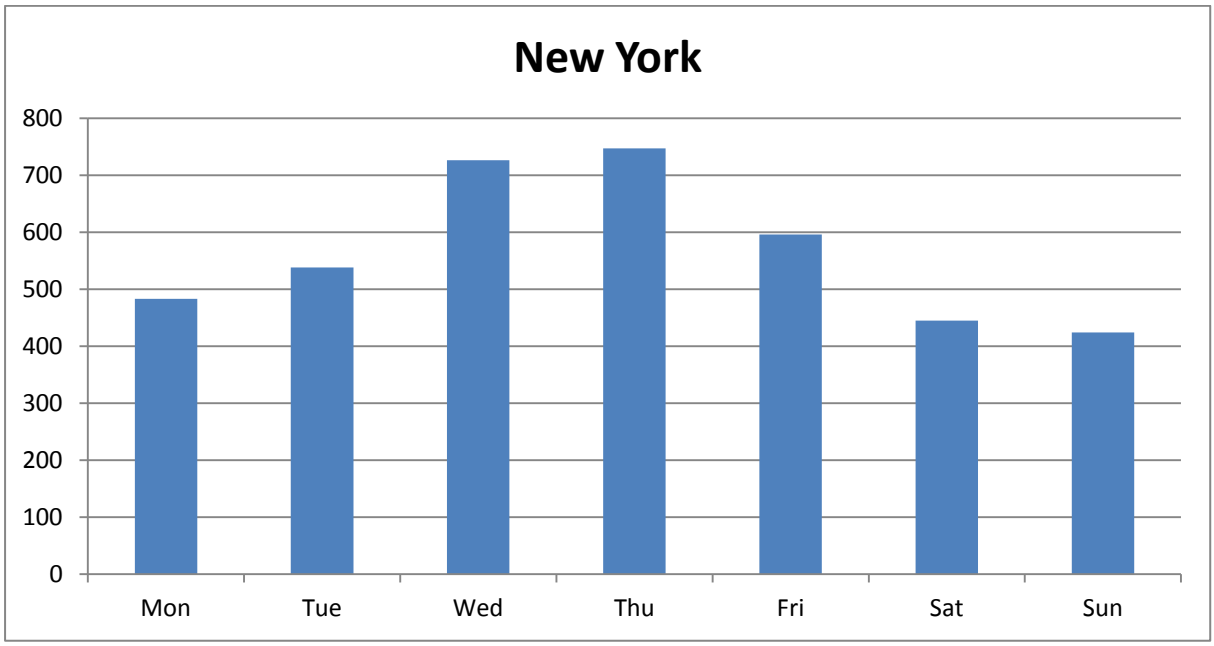
Γράφημα 3: Κατανομή συμφόρησης ανάλογα την ώρα, Houston



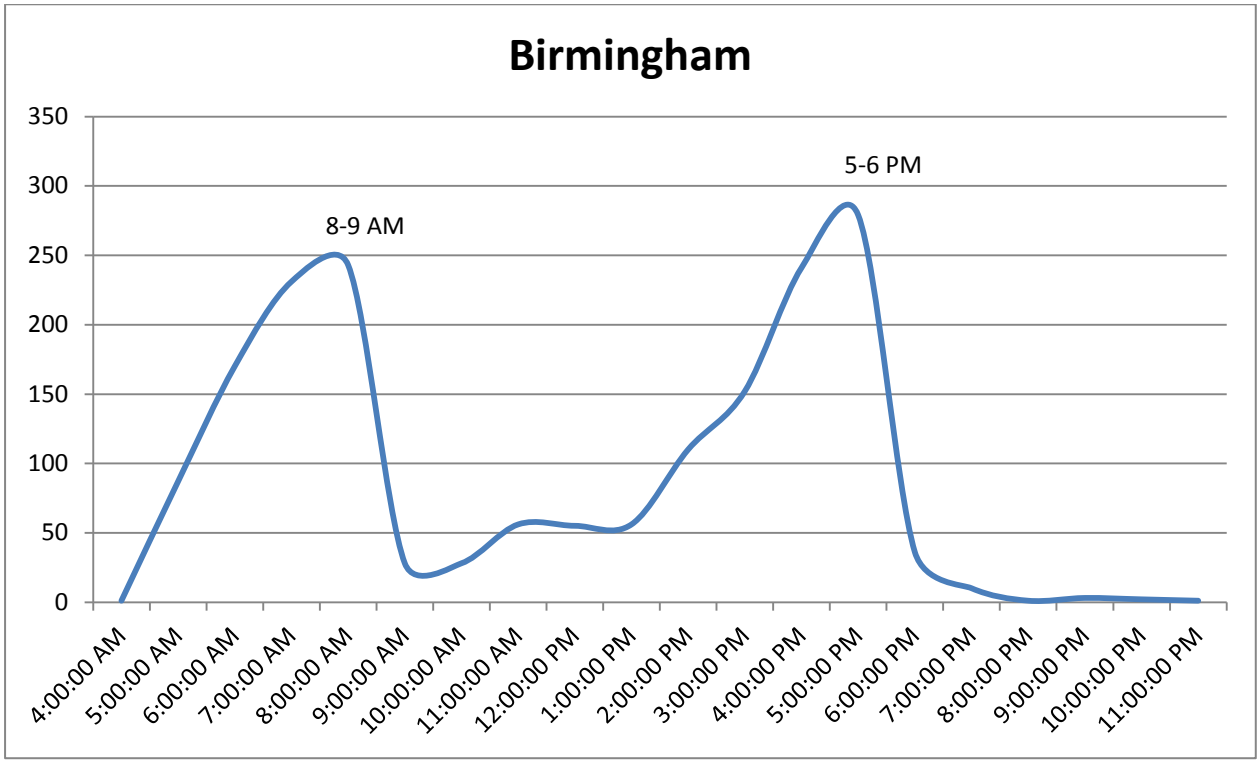
Γράφημα 4: Κατανομή συμφόρησης ανάλογα την ημέρα, Houston



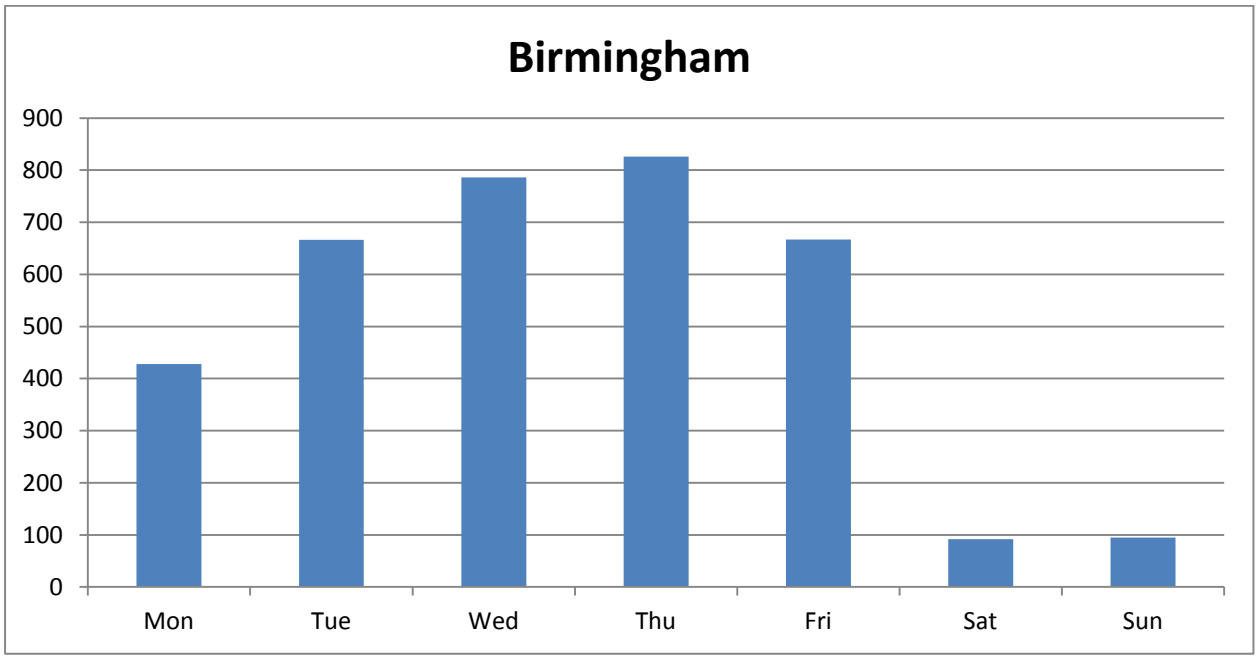
Γράφημα 5: Κατανομή συμφόρησης ανάλογα την ώρα, New York



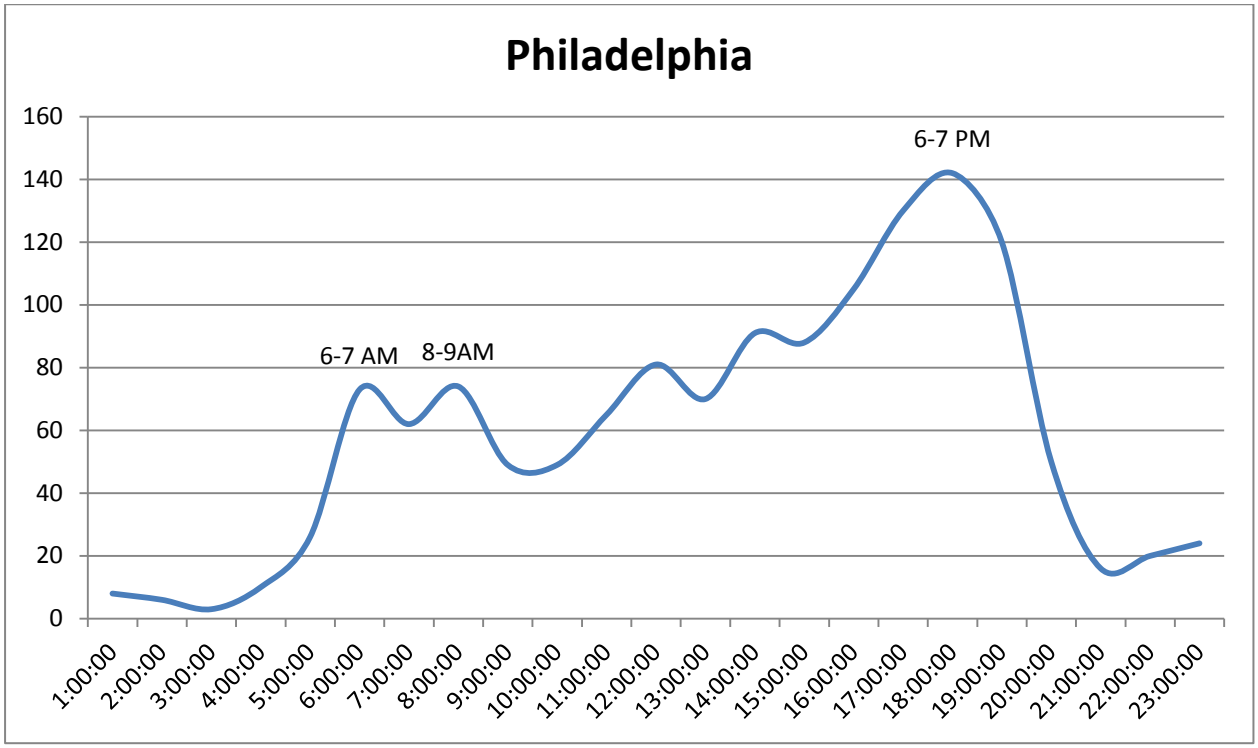
Γράφημα 6: Κατανομή συμφόρησης ανάλογα την ημέρα, New York



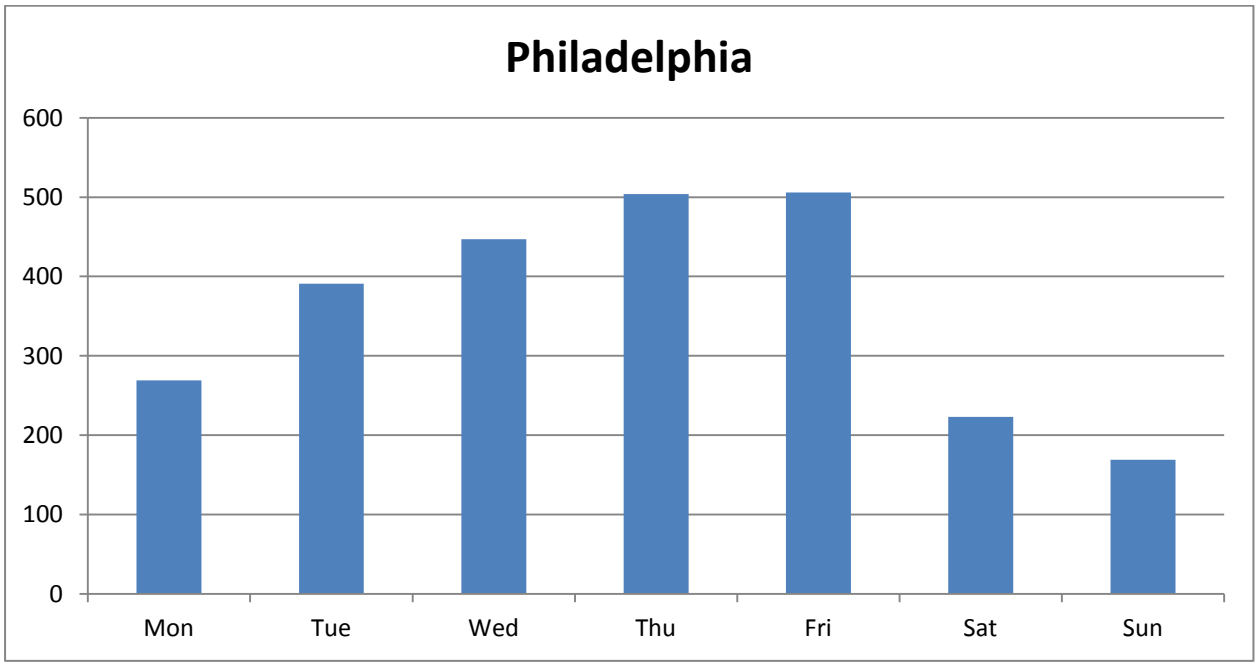
Γράφημα 7: Κατανομή συμφόρησης ανάλογα την ώρα, Birmingham



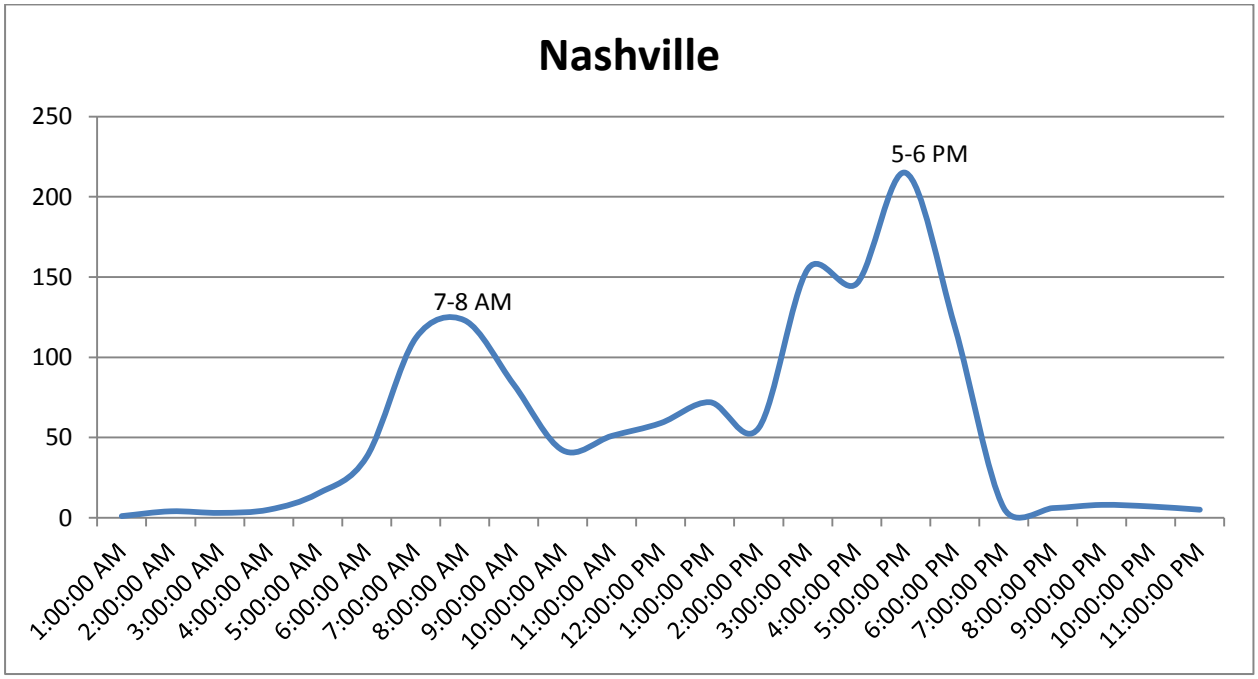
Γράφημα 8: Κατανομή συμφόρησης ανάλογα την ημέρα, Birmingham



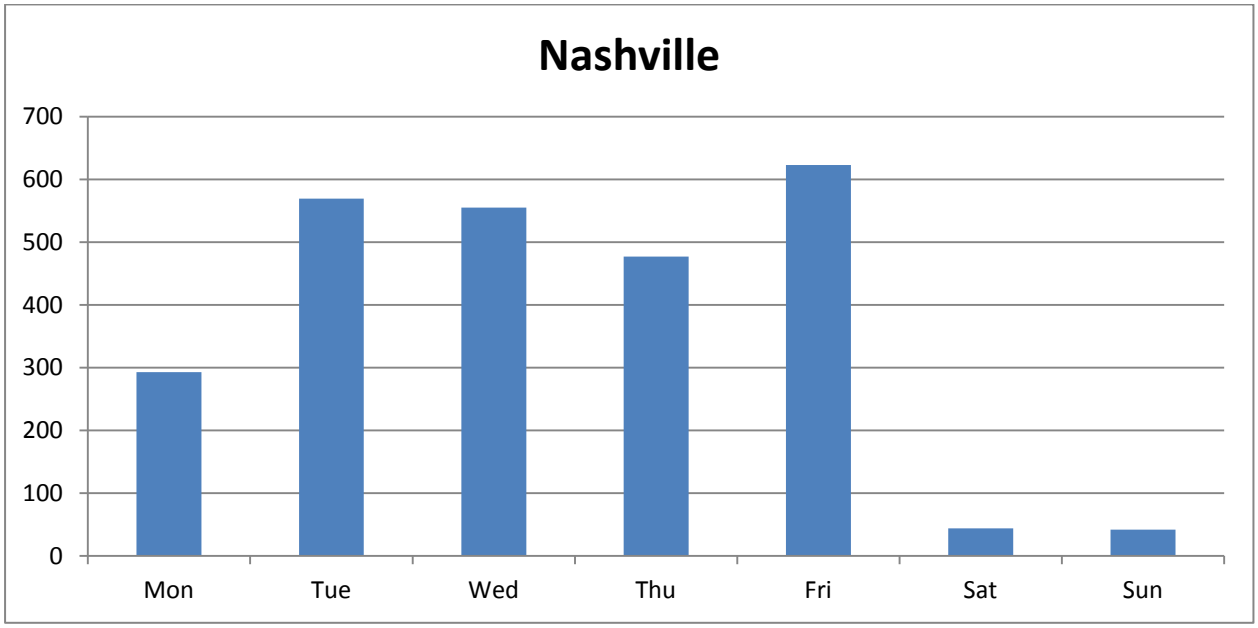
Γράφημα 9: Κατανομή συμφόρησης ανάλογα την ώρα, Philadelphia



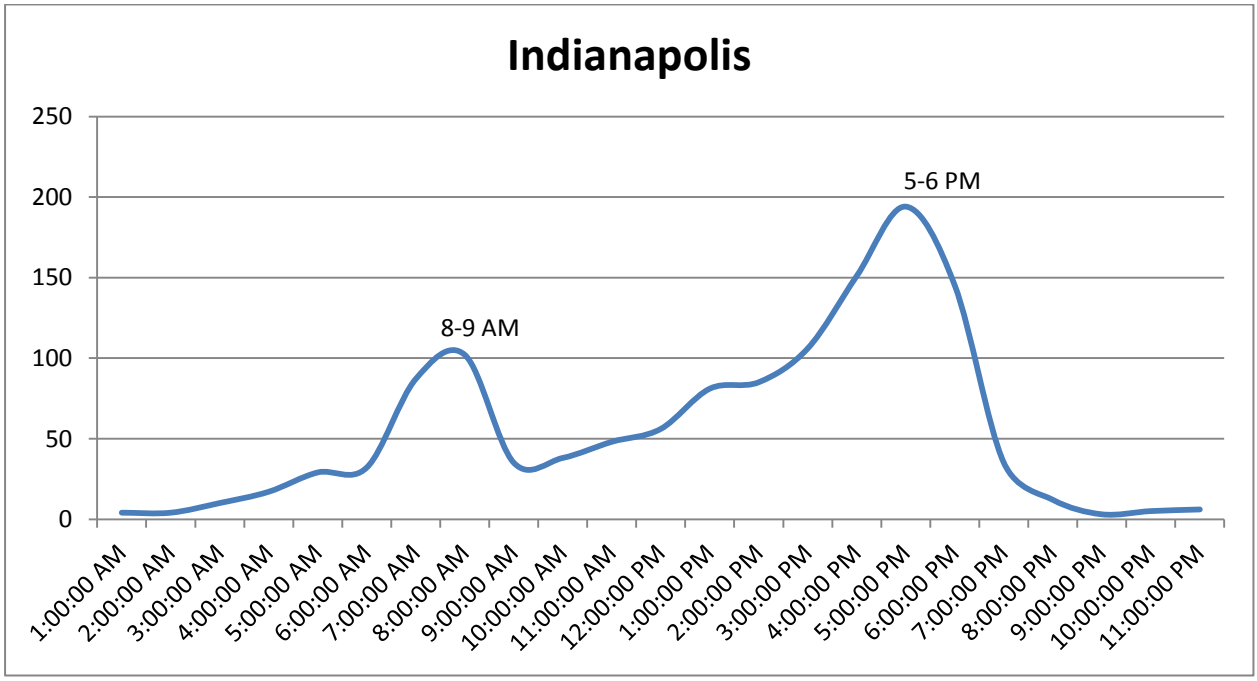
Γράφημα 10: Κατανομή συμφόρησης ανάλογα την ημέρα, Philadelphia



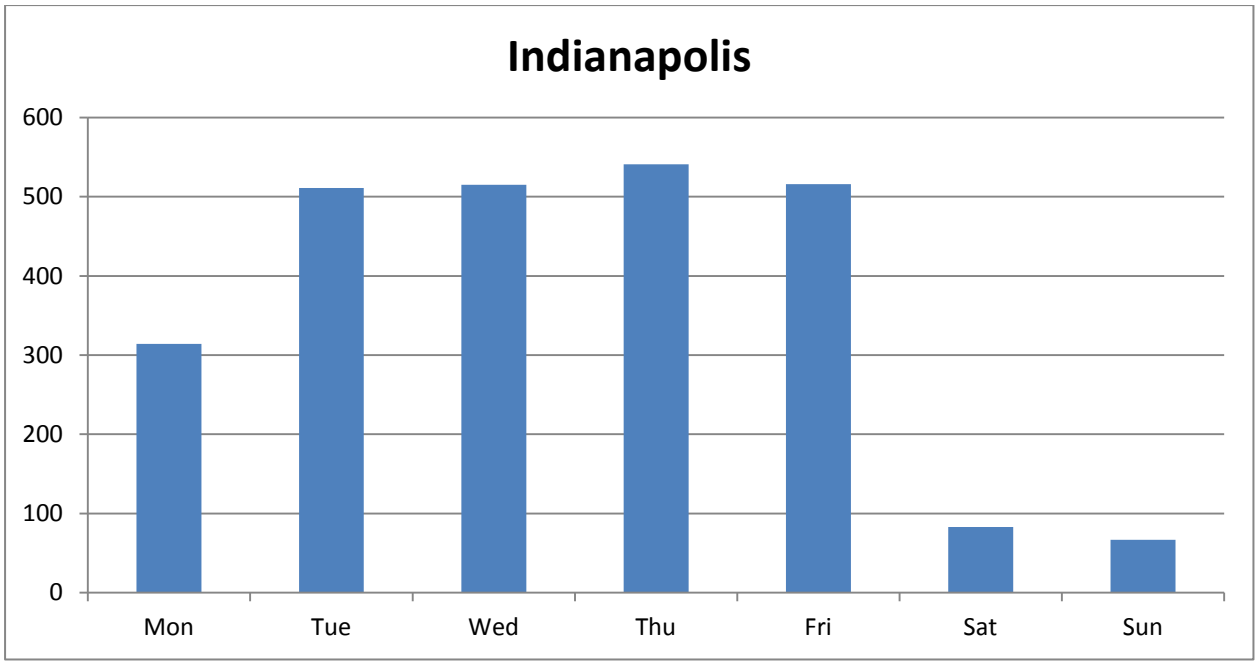
Γράφημα 11: Κατανομή συμφόρησης ανάλογα την ώρα, Nashville



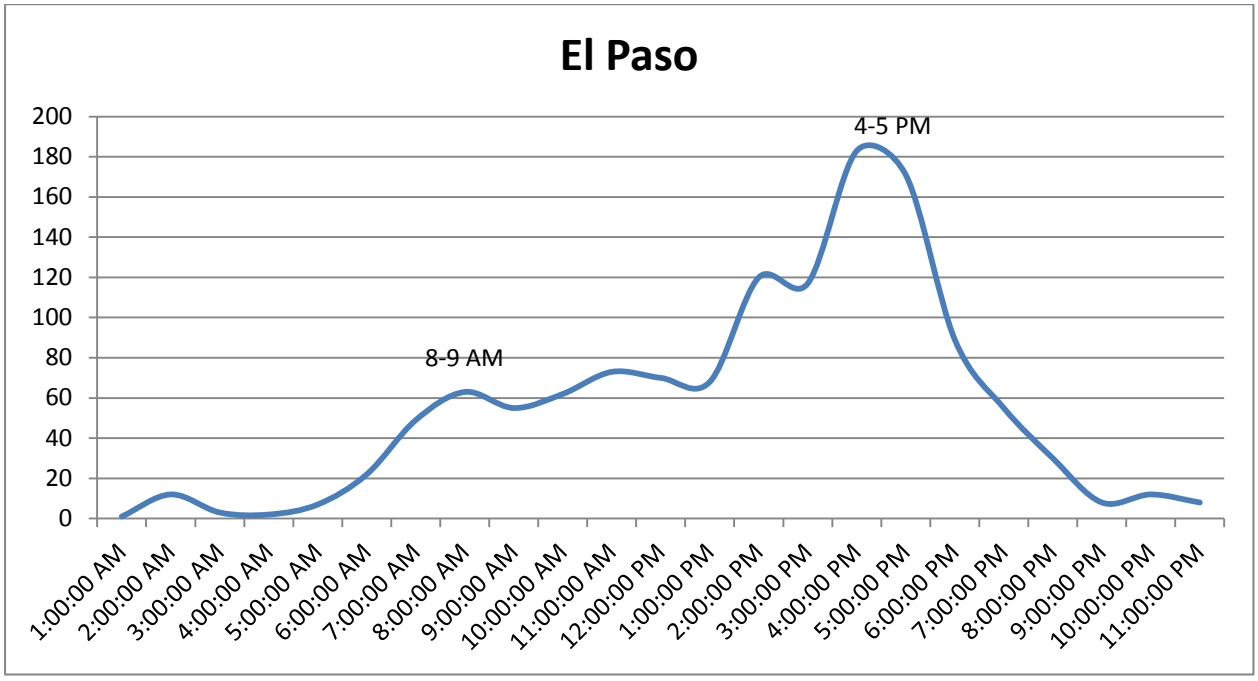
Γράφημα 12: Κατανομή συμφόρησης ανάλογα την ημέρα, Nashville



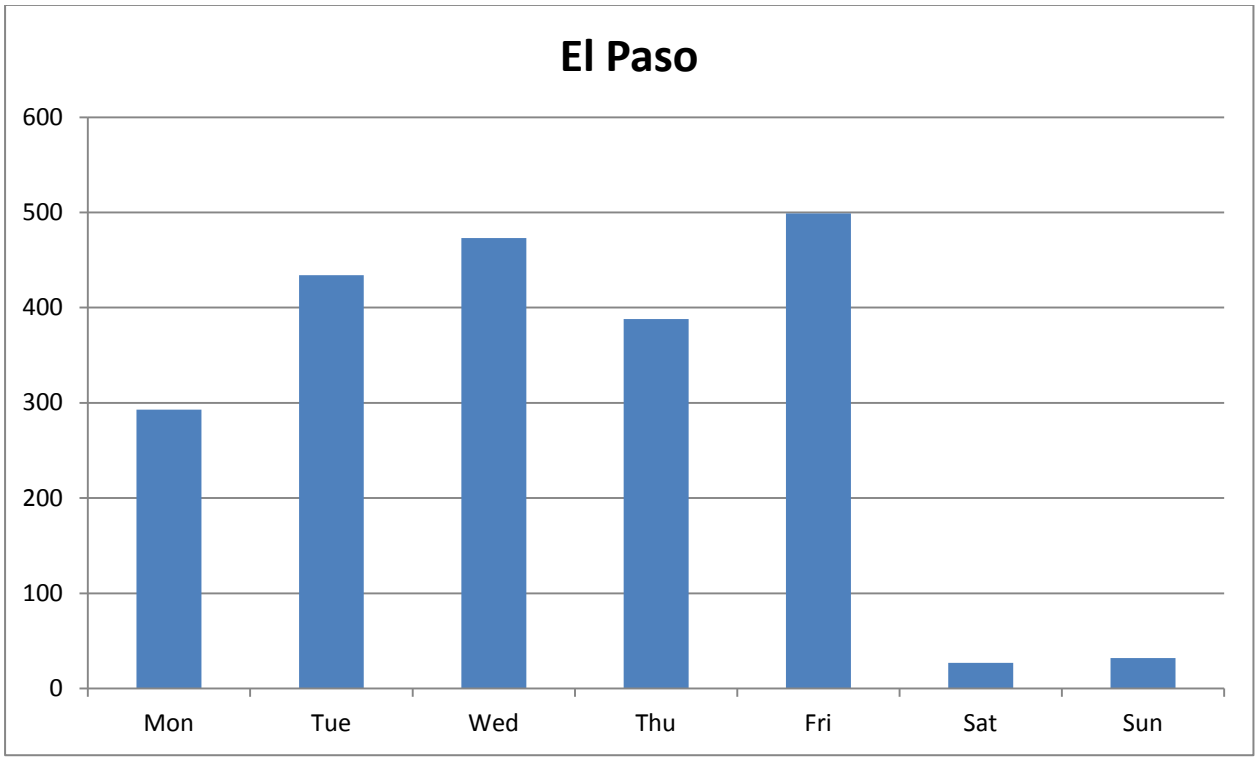
Γράφημα 13: Κατανομή συμφόρησης ανάλογα την ώρα, Indianapolis



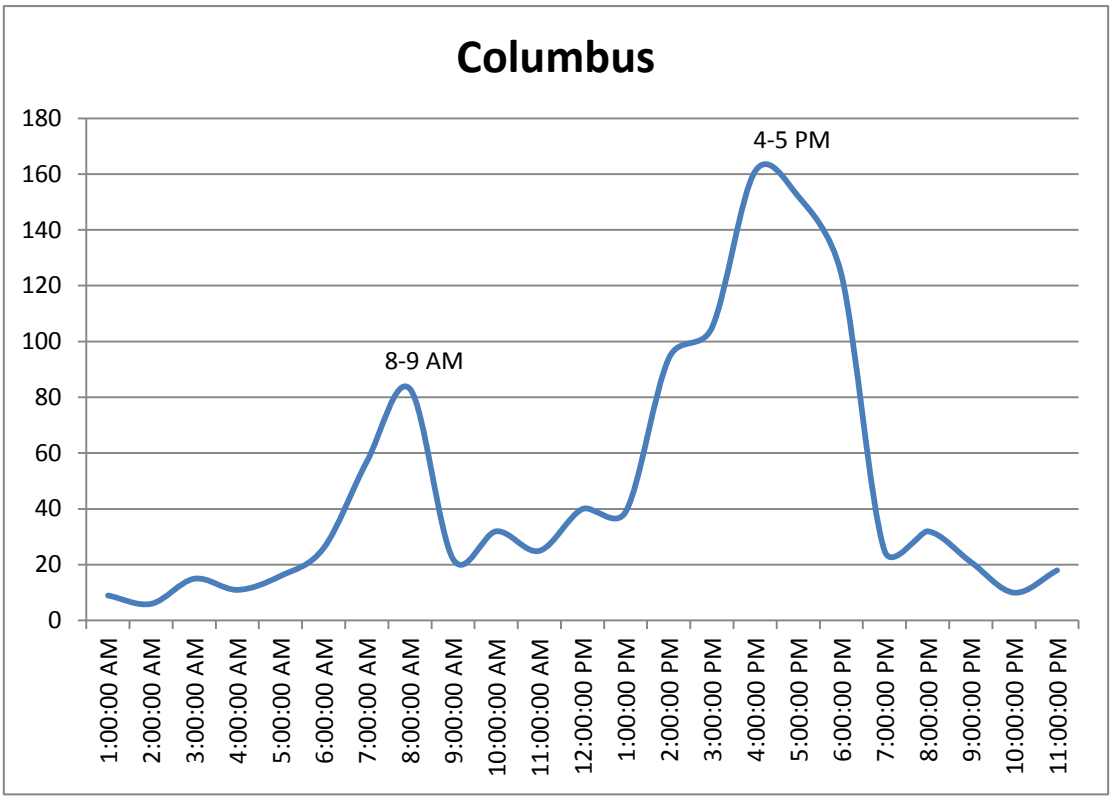
Γράφημα 14: Κατανομή συμφόρησης ανάλογα την ημέρα, Indianapolis



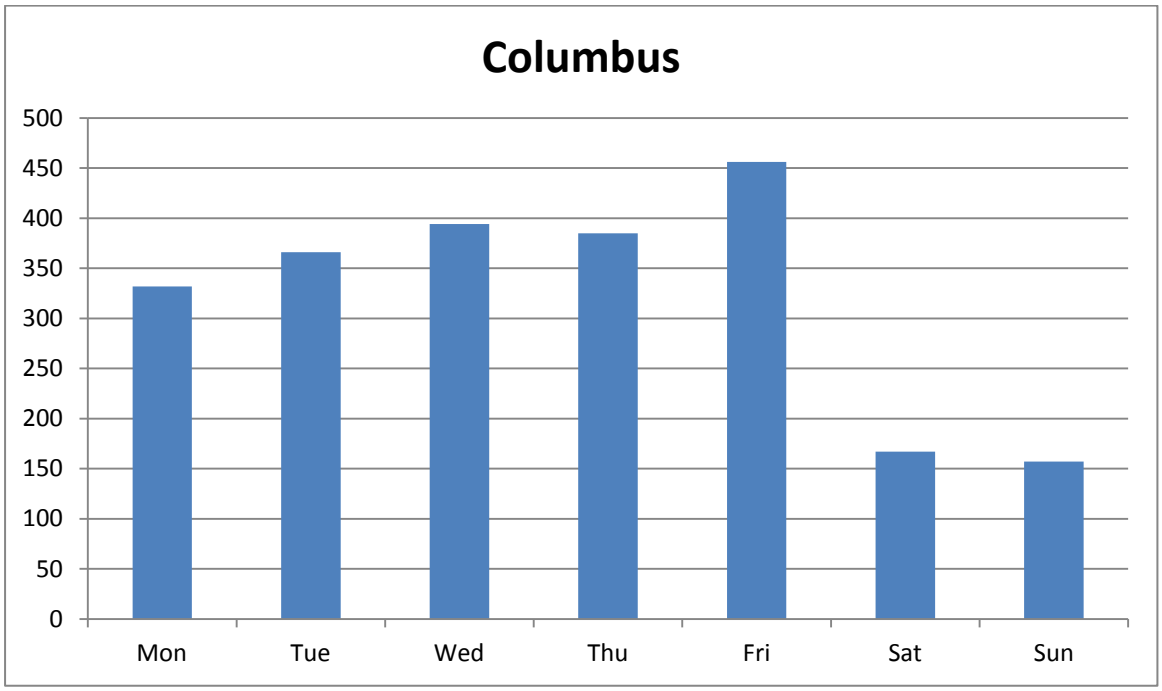
Γράφημα 15: Κατανομή συμφόρησης ανάλογα την ώρα, El Paso



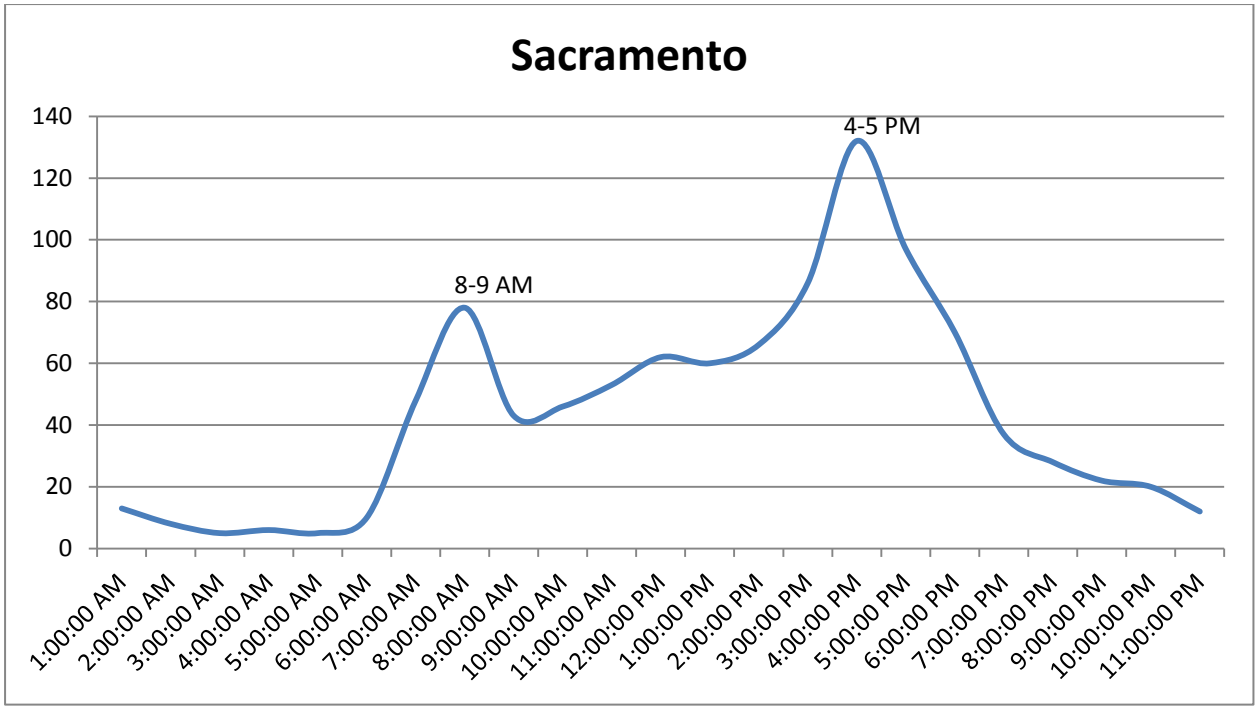
Γράφημα 16: Κατανομή συμφόρησης ανάλογα την ημέρα, El Paso



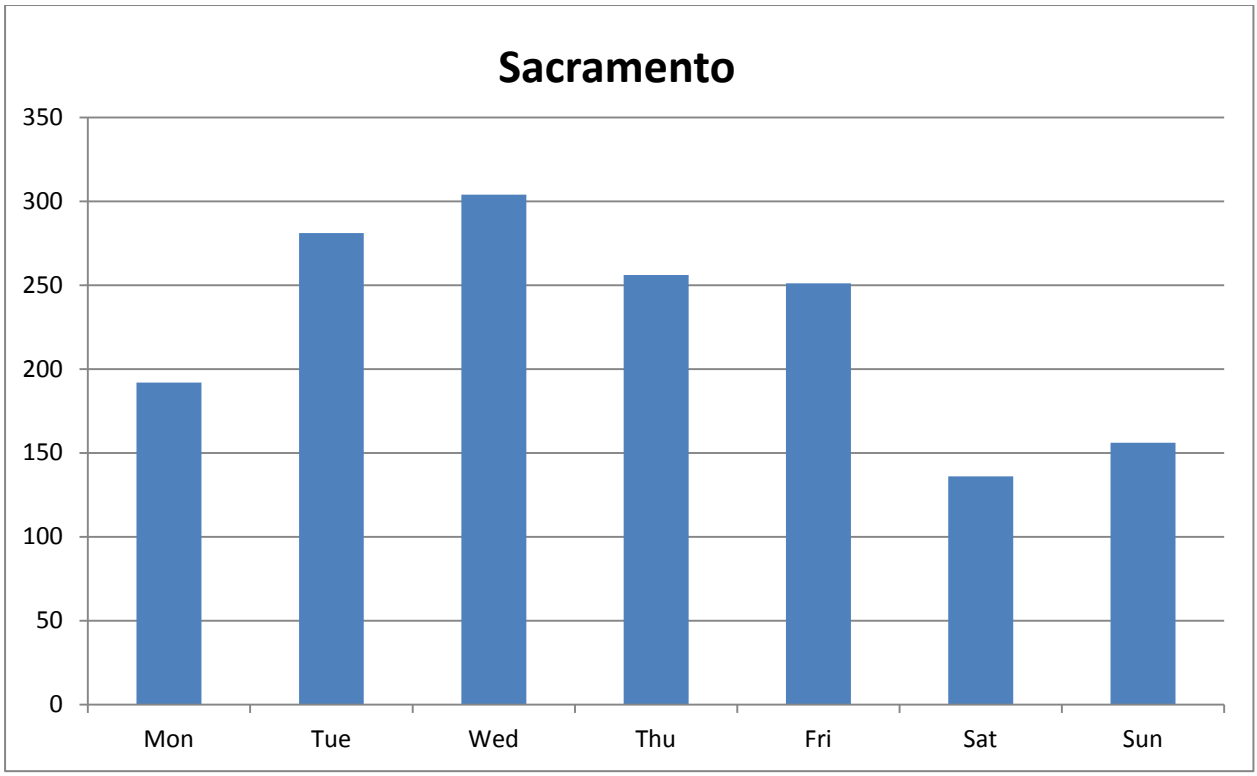
Γράφημα 17: Κατανομή συμφόρησης ανάλογα την ώρα, Columbus



Γράφημα 18: Κατανομή συμφόρησης ανάλογα την ημέρα, Columbus



Γράφημα 19: Κατανομή συμφόρησης ανάλογα την ώρα, Sacramento



Γράφημα 20: Κατανομή συμφόρησης ανάλογα την ημέρα, Sacramento

Πόλη	Ώρα αιχμής		Ώρα αιχμής Tomtom	
	Πρωί	Απόγευμα	Πρωί	Απόγευμα
Dallas	8-9 AM	5-6 PM	7-8 AM	5-6 PM
Houston	-	5-6 PM	7-8 AM	5-6 PM
New York	-	5-6 PM	8-9 AM	5-6 PM
Birmingham	8-9 AM	5-6 PM	7-8 AM	5-6 PM
Detroit	9-10 AM	1-2 PM	8-9 AM	5-6 PM
Philadelphia	-	6-7 PM	8-9 AM	5-6 PM
Nashville	7-8 AM	5-6 PM	7-8 AM	5-6 PM
Indianapolis	8-9 AM	5-6 PM	7-8 AM	5-6 PM
El Paso	8-9 AM	4-5 PM	7-8 AM	5-6 PM
Columbus	8-9 AM	4-5 PM	8-9 AM	5-6 PM
Sacramento	7-8 AM	4-5 PM	7-8 AM	5-6 PM
Minneapolis	8-9 AM	5-6 PM	7-8 AM	5-6 PM
Newark	7-8AM	5-6 PM	-	-
Cleveland	7-8 AM	5-6 PM	7-8 AM	5-6 PM
Tampa	7-8 AM	5-6 PM	7-8 AM	5-6 PM
Miami	8-9 AM	5-6 PM	8-9 AM	5-6 PM
Atlanta	7-8 AM	5-6 PM	7-8 AM	5-6 PM
Pittsburg	8-9 AM	5-6 PM	8-9 AM	5-6 PM
Seattle	8-9 AM	4-5 PM	8-9 AM	5-6 PM
San Antonio	8-9 AM	5-6 PM	7-8 AM	5-6 PM

Πίνακας 8: Σύγκριση ώρας αιχμής 20 πόλεων με στοιχεία από την εταιρία Tomtom

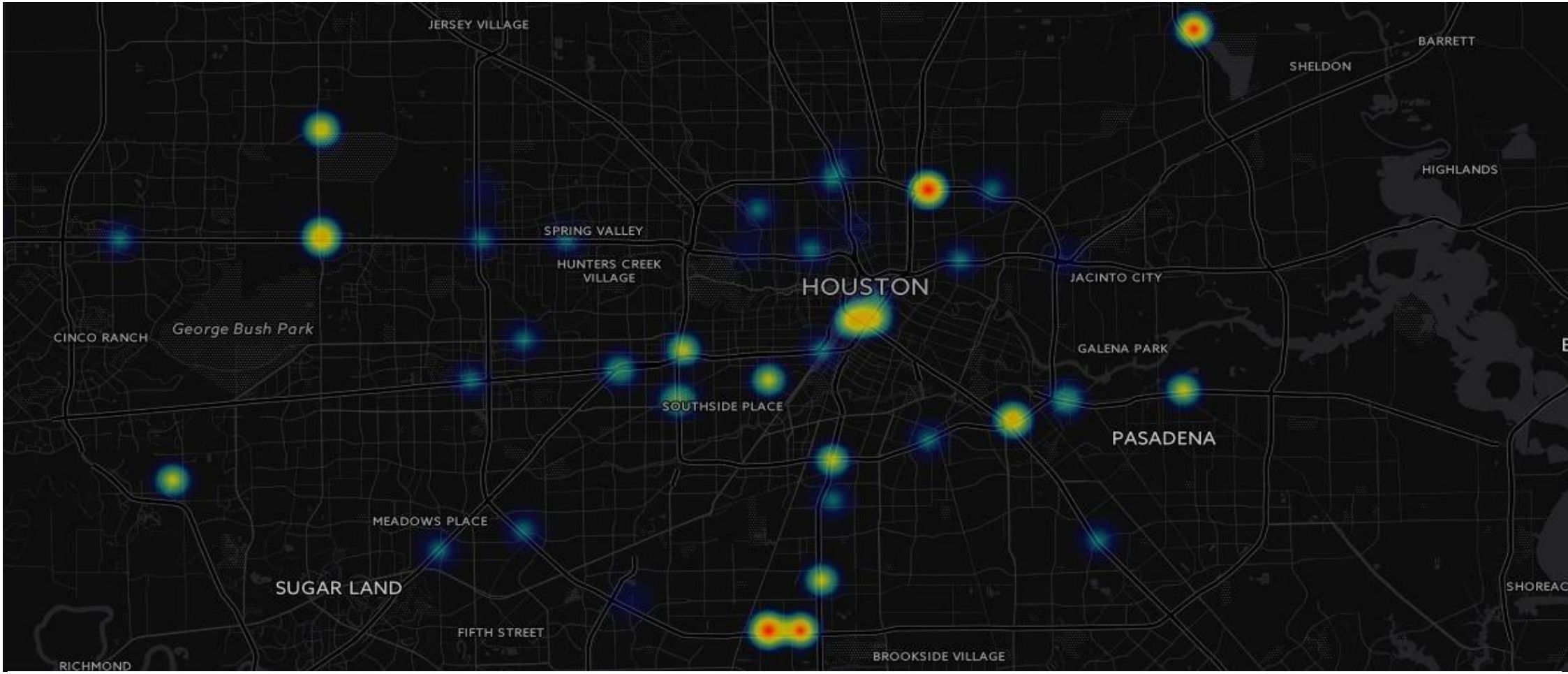
Όπως προκύπτει από τον παραπάνω πίνακα (Πίνακας 8), η ανάλυση που διεξήχθη μπόρεσε επιτυχώς να αναδείξει την ώρα αιχμής της κάθε πόλης. με μικρές αποκλίσεις από τις τιμές που δίνει η έκθεση της εταιρίας Tomtom. Εξαίρεση αποτελεί η πόλη του Detroit όπου η διαφορά με τα επίσημα δεδομένα είναι σημαντική. Σε τρείς περιπτώσεις δεν ήταν εφικτό να επιλεγεί κάποια τιμή ως ώρα αιχμής το πρωί και για τον λόγο αυτό δεν υπάρχει τιμή στην αντίστοιχη θέση στον πίνακα. Στις περιπτώσεις αυτές , όπως η πόλη του Houston, η συμφόρηση κυμαίνεται στα ίδια επίπεδα κατα το χρονικό διάστημα 6AM-12PM όπως φαίνεται στα διαγράμματα. Στην πόλη του Newark η εταιρία δεν διέθετε δεδομένα και για αυτό και δεν υπάρχει τιμή στην αντίστοιχη θέση.

Παρατηρώντας τα διαγράμματα εξέλιξης της συμφόρησης εντός της ημέρας γίνεται άμεσα αντιληπτό ότι σχεδόν σε όλες τις πόλεις, οι περιπτώσεις συμφόρησης στην απογευματινή ώρα αιχμής είναι σχεδόν διπλάσιες από το πρωί, γεγονός που επιβεβαιώνεται σε διάφορες εργασίες αλλά και από την εταιρία Tomtom. Όσον αφορά την εξέλιξη ανά ημέρα είναι ευκρινές ότι, εκτός τα σαββατοκύριακα, η μέρα που σε όλες τις πόλεις έχει την λιγότερη συμφόρηση είναι η Δευτέρα ενώ παρατηρείται ότι σε πολλές πόλεις υπάρχει σημαντική διακύμανση κατά την διάρκεια της εβδομάδας.

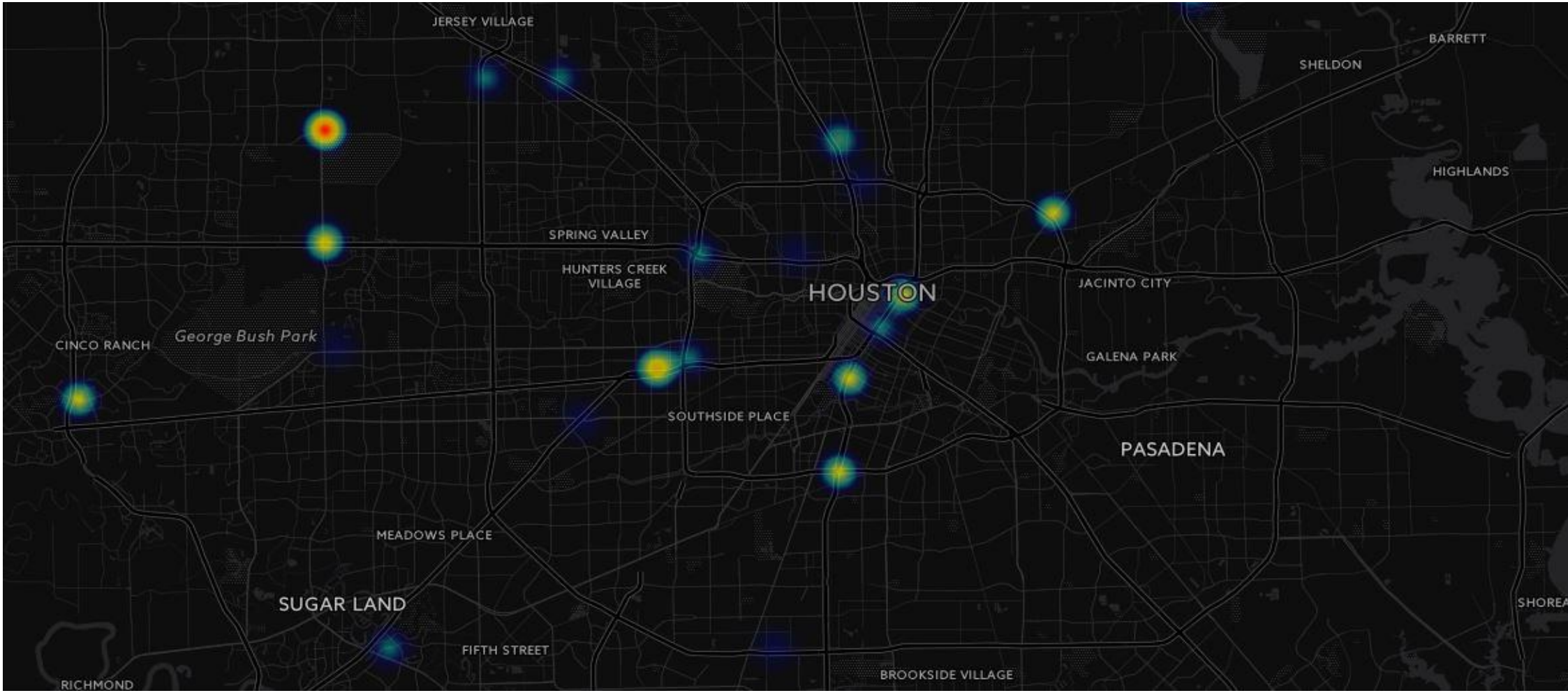
4.2.2 Υπογραφή της συμφόρησης

Κάθε πόλη διαθέτει ένα μοναδικό τρόπο μετάδωσης της κίνησης και κατά επέκταση της συμφόρησης από περιοχή σε περιοχή κατά την διάρκεια της ημέρας. Έτσι σε κάθε πόλη η ροή των οχημάτων δημιουργεί κάποια σταθερά πρότυπα που σχετίζονται με τις περιοχές που επιλέγουν οι περισσότεροι οδηγοί ως προορισμό κάθε ώρα της ημέρας. Η ανάδειξη αυτών των περιοχών «ενδιαφέροντος» για τους περισσότερους οδηγούς κατά την διάρκεια της ημέρας αποτελεί ένα σημαντικό στοιχείο για την λύση τους κυκλοφοριακού προβλήματος καθώς με αυτόν τον τρόπο μπορεί να γίνει περισσότερο κατανοητός ο τρόπος που κινούνται οι πολίτες στο οδικό δίκτυο. Στην παρούσα εργασία γίνεται προσπάθεια να αποτυπωθεί το «κυκλοφοριακό πρότυπου» της πόλης του Houston σύμφωνα με τα δεδομένα που έχουν συλλεχθεί από το Twitter. Στην ουσία η απεικόνιση που ακολουθεί αποτελεί μια μακροσκοπική παρατήρηση των «εστιών» κυκλοφοριακής συμφόρησης στην ευρύτερη περιοχή της πόλης του Houston αναδεικνύοντας τις επιμέρους περιοχές που εμφανίζεται περισσότερο το φαινόμενο κατά την διάρκεια της ημέρας.

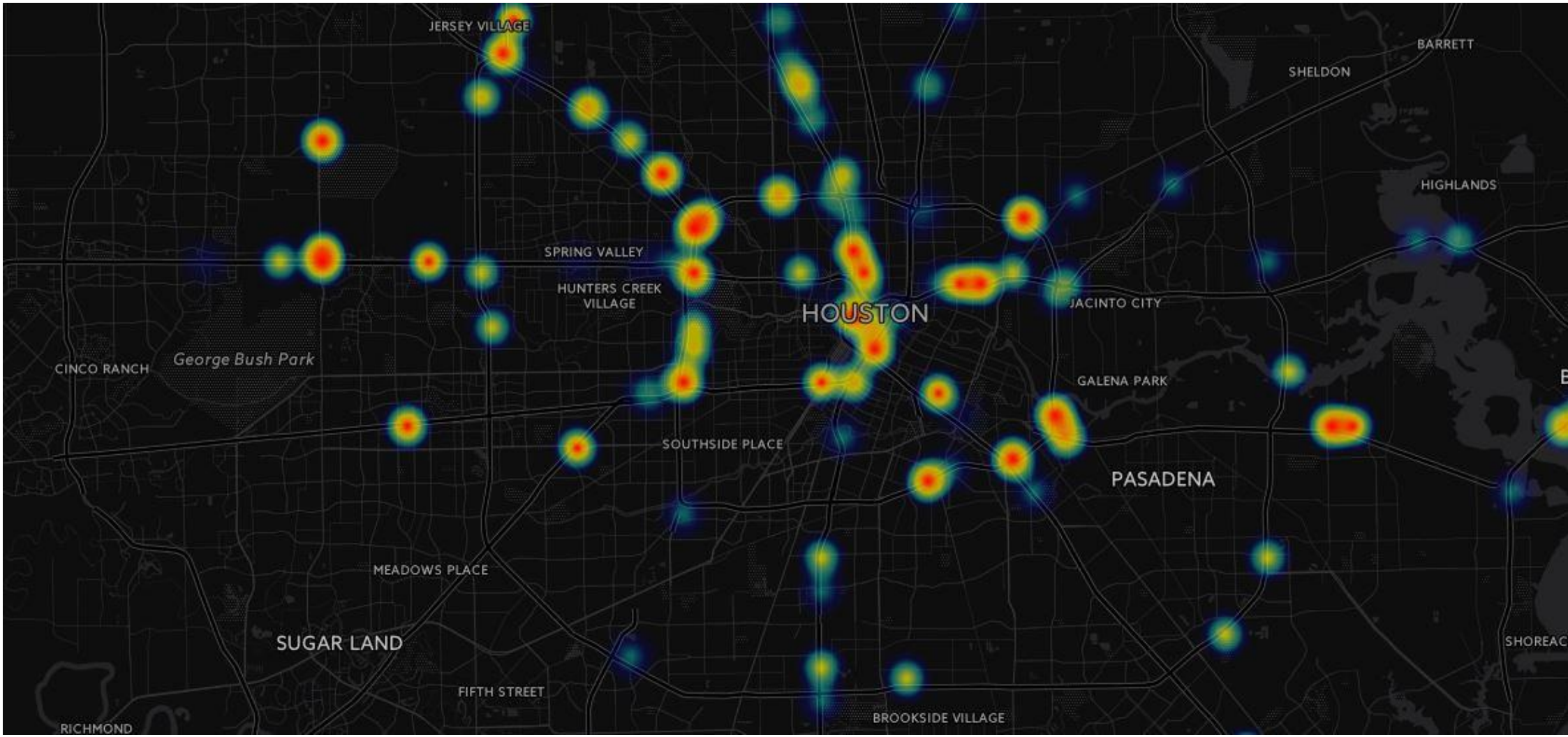
Η γραφική αναπαράσταση των δεδομένων που επιλέχθηκε για την εύρεση του κυκλοφοριακού προτύπου της πόλης του Houston είναι μια σειρά από χάρτες πυκνότητας σημείων όπου με κόκκινο χρωματίζεται η περιοχή του χάρτη που βρίσκονται τα περισσότερα σημεία συμφωνα με την χαρτογράφηση τους που προηγήθηκε. Τα σημεία που χρησιμοποιήθηκαν δεν προέρχονται από την βάση δεδομένων της πόλης του Houston που αναφέρθηκε προυγουμένως καθώς θεωρήθηκε σκόπιμο να εμφανιστούν και οι γύρω περιοχές της πόλης με αποτέλεσμα να χρησιμοποιηθούν περίπου 7000 σημεία.



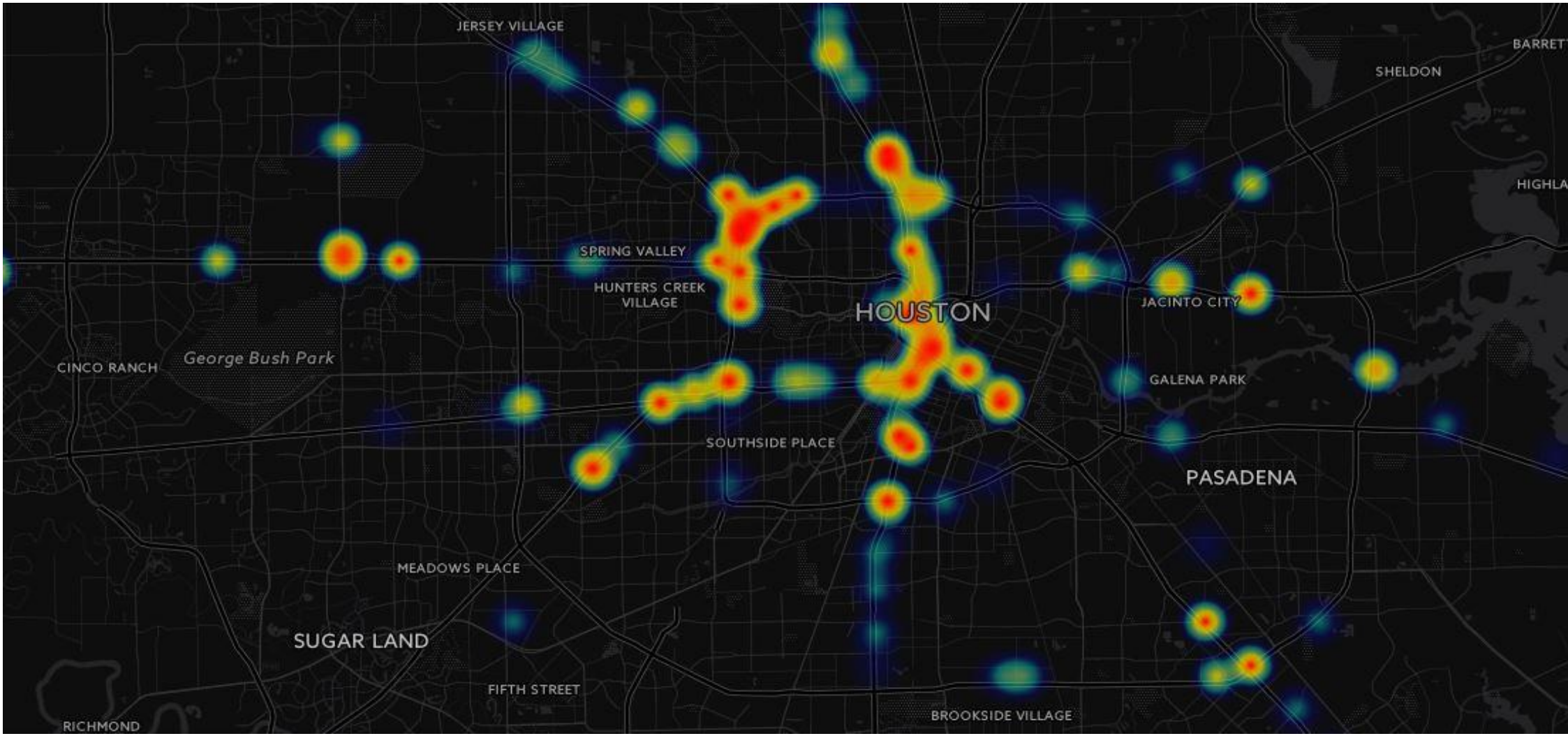
Χάρτης 1: Στιγμιότυπο συμφόρησης 12 am, Houston



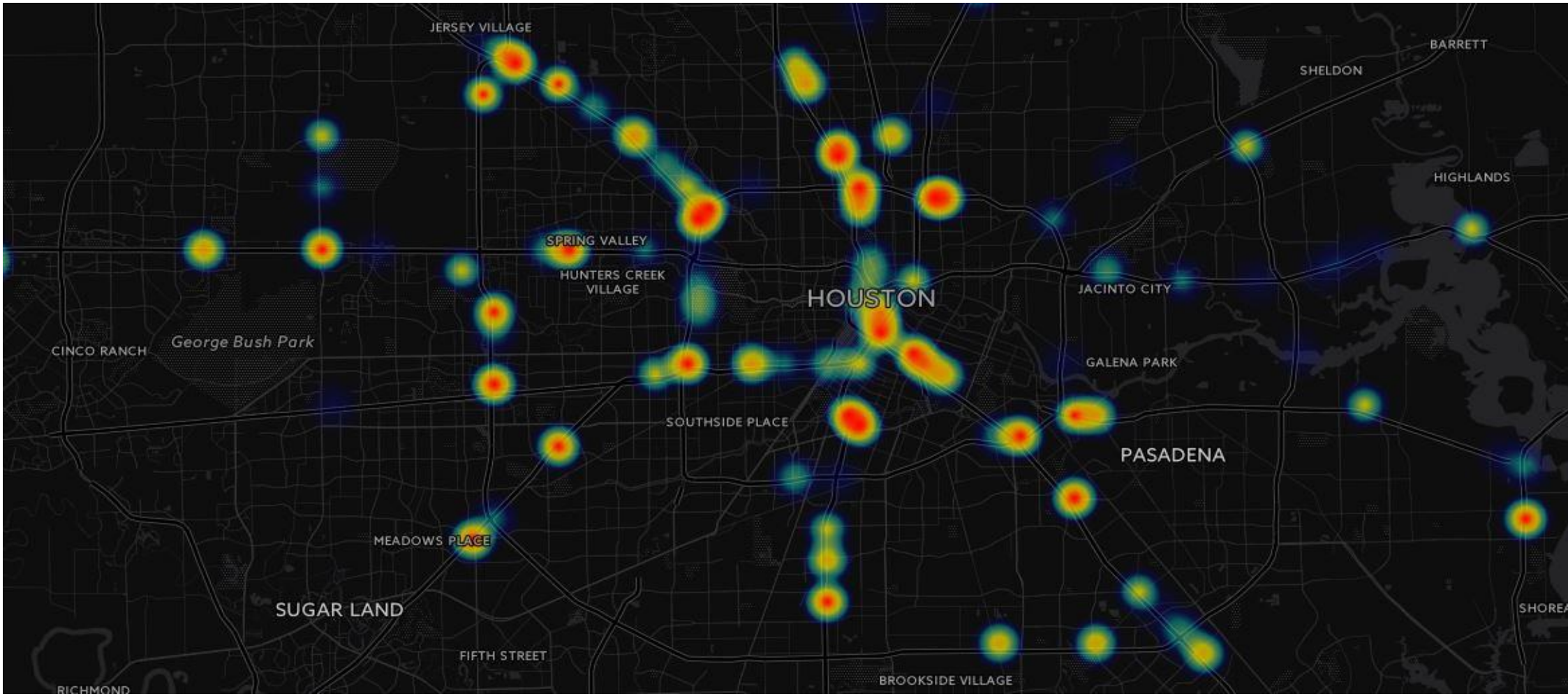
Χάρτης 2 :Στιγμιότυπο συμφόρησης 6 am, Houston



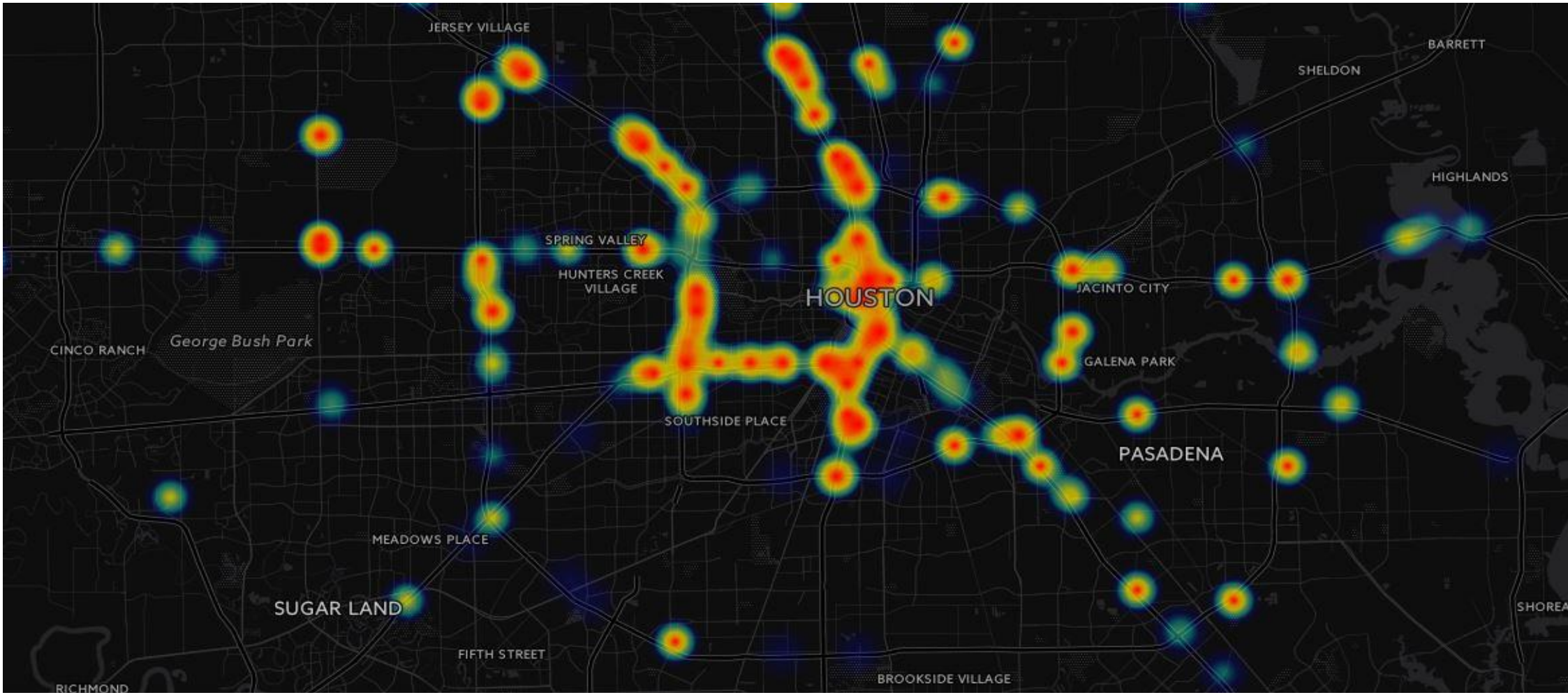
Χάρτης 3: Στιγμιότυπο συμφόρησης 9 am, Houston



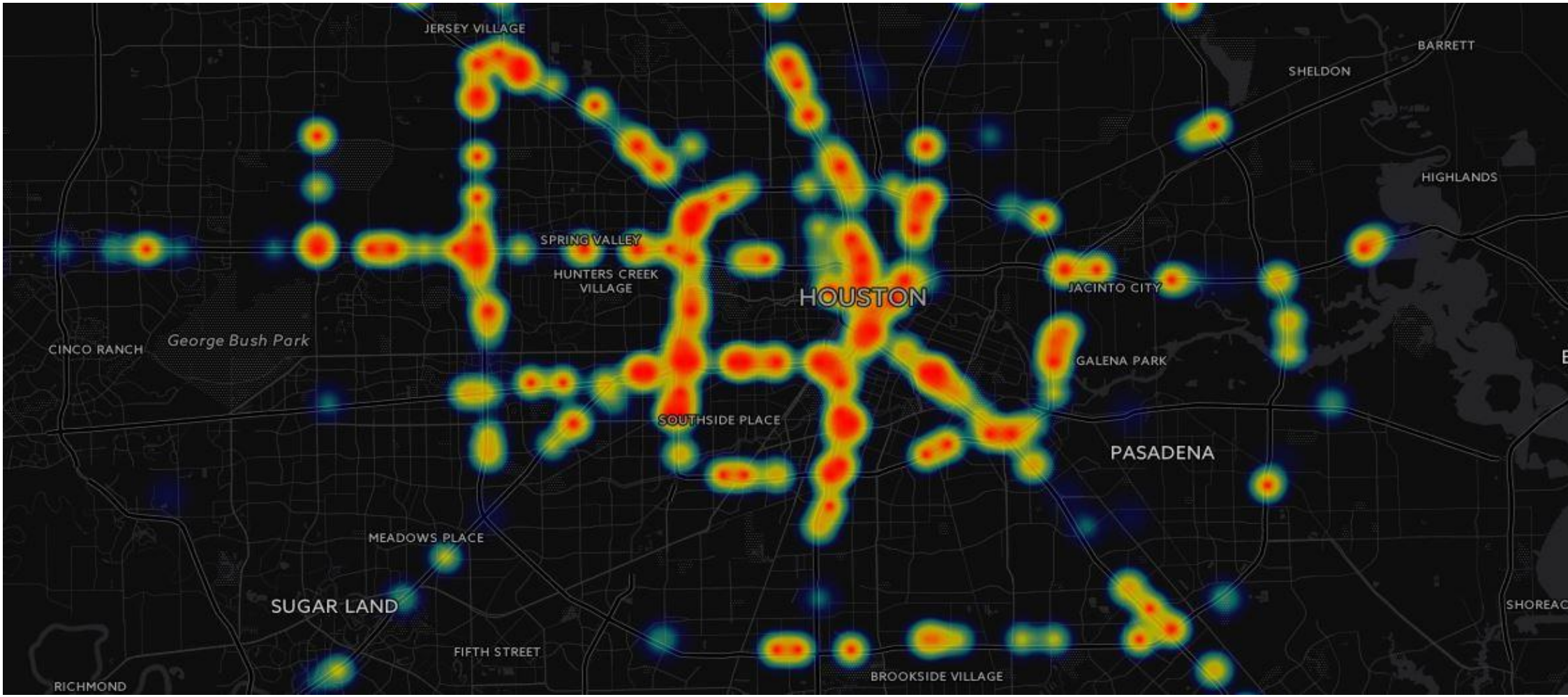
Χάρτης 4: Στιγμιότυπο συμφόρησης 12 pm, Houston



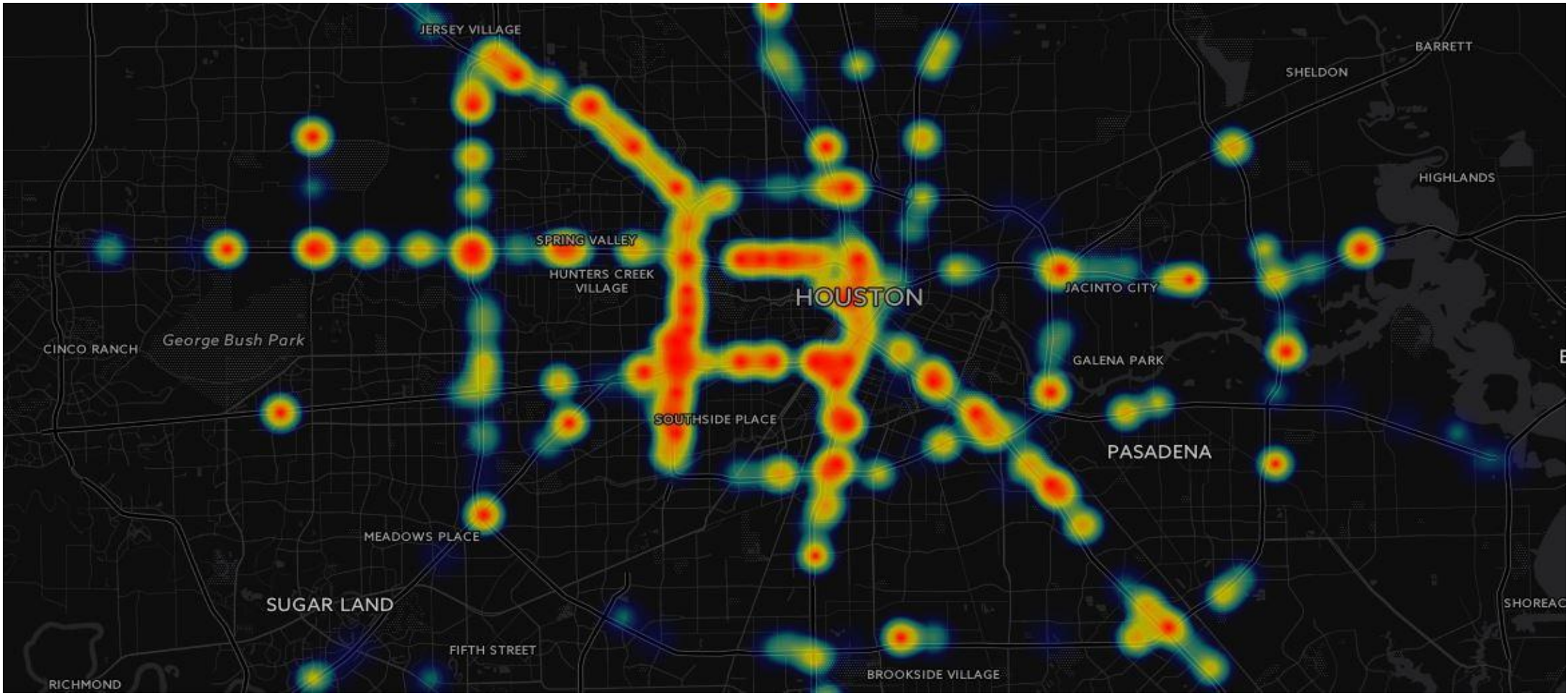
Χάρτης 5: Στιγμιότυπο συμφόρησης 2 pm, Houston



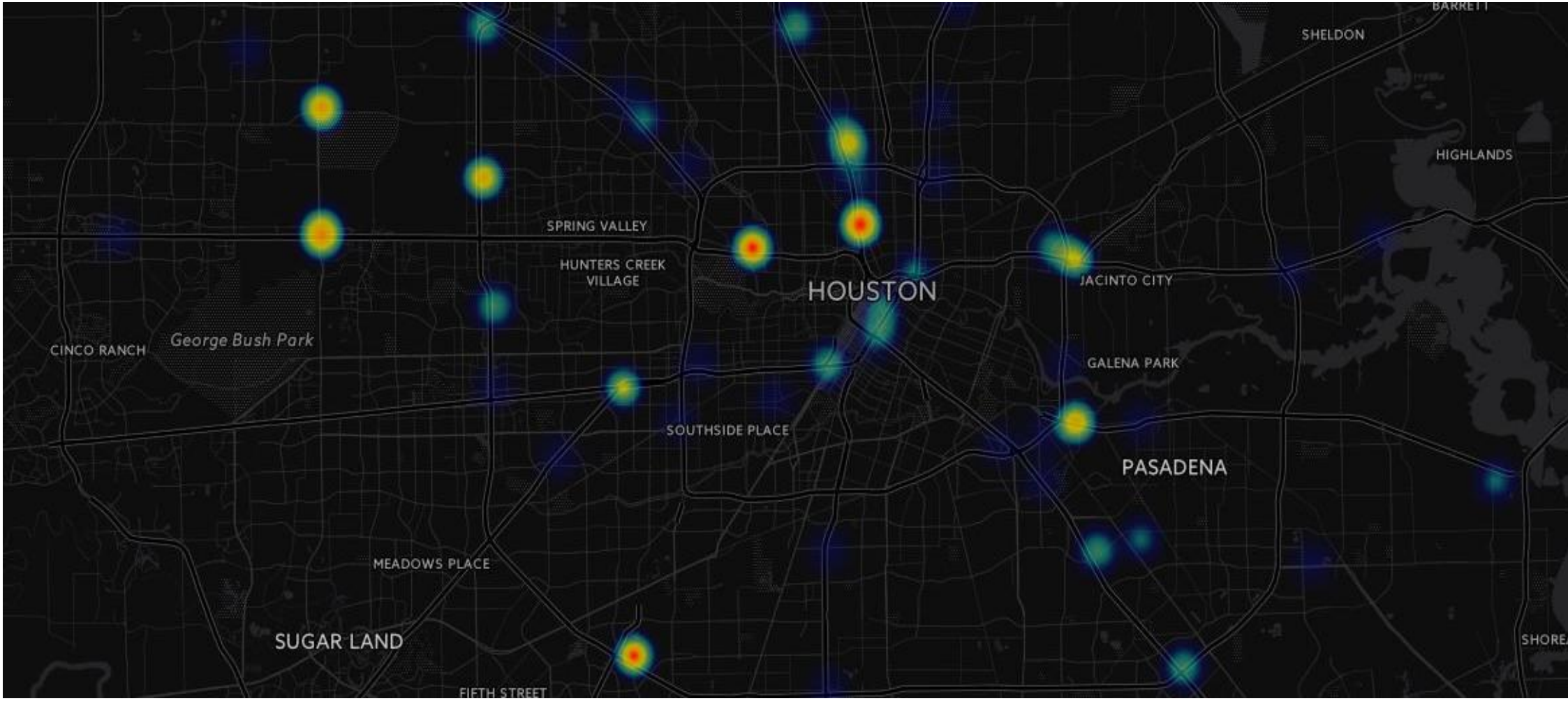
Χάρτης 6: Στιγμιότυπο συμφόρησης 4 pm, Houston



Χάρτης 7: Στιγμιότυπο συμφόρησης 6 pm, Houston



Χάρτης 8 :Στιγμιότυπο συμφόρησης 8 pm, Houston



Χάρτης 9: Στιγμιότυπο συμφόρησης 10pm, Houston

4.3 Χωρική ανάλυση

Κατά την χωρική ανάλυση των δεδομένων που περιγράφεται παρακάτω γίνεται προσπάθεια να εξαχθούν τριών ειδών συμπεράσματα. Αρχικά γίνεται εύρεση των οδών που σύμφωνα με τα δεδομένα έχουν μεγαλύτερη πιθανότητα εμφάνισης συμφόρησης. Έπειτα υπολογίζεται η χωρική αυτοσυσχέτιση σε περιοχές διαφόρων πόλεων με στόχο την εύρεση των περιοχών με το μεγαλύτερο κυκλοφοριακό πρόβλημα. Τέλος εξάγονται οι περιοχές και οι οδοί όπου εμφανίζονται τα περισσότερα τροχαία ατυχήματα. Τα παράγωγα από τις παραπάνω αναλύσεις είναι χάρτες που στην πρώτη περίπτωση είναι σε κλίμακα πόλης ενώ στην δεύτερη και στην τρίτη σε κλίμακα μητροπολιτικής περιοχής. Επίσης παρατίθενται χάρτες στους οποίους απεικονίζονται τα επίσημα συμπεράσματα αντίστοιχων αναλύσεων για αξιολόγηση των αποτελεσμάτων.

4.3.1 Αποτύπωση της συμφόρησης σε επίπεδο οδών

Μια από τις βασικότερες αναλύσεις που παρουσιάζεται στην παρούσα εργασία είναι η αποτύπωση της συμφόρησης σε επίπεδο πόλης και της εξαγωγής των οδών που έχουν την μεγαλύτερη πιθανότητα εμφάνισης του φαινομένου. Η ανάλυση αυτή γίνεται στην πόλη της Νέα Υόρκης καθώς η συγκεκριμένη περιοχή διαθέτει τον απαραίτητο όγκο δεδομένων. Η μεθοδολογία και η εφαρμογή παρουσιάζεται παρακάτω.

Αρχικά, χρησιμοποιώντας την βάση δεδομένων που αντιστοιχούσε στην πόλη της Νέα Υόρκης για την οποία έχει γίνει αναφορά προηγουμένως, έγινε η αποτύπωση των σημείων σύμφωνα με τις συντεταγμένες που συνόδευε το κάθε ένα στο λογισμικό Arcmap . Έπειτα ανακτήθηκε το οδικό δίκτυο της Νέας Υόρκης από την βάση δεδομένων του Openstreetmaps [46] και αποτυπώθηκε επίσης στον χάρτη. Επιλέχθηκαν να αποτυπωθούν μόνο οι πέντε κυριότεροι τύποι οδών οι οποίοι χαρακτηρίζονται ως “motorway”, “trunk”, “primary”, “secondary” και “tertiary” στην βάση δεδομένων του Openstreetmaps και αποτελούν το σύνολο των μεγάλων δρόμων των ΗΠΑ. Η επιλογή αυτή έγινε καθώς σε αυτούς τους τύπους οδών παρατηρήθηκε η μεγαλύτερη συγκέντρωση σημείων με ποσοστό άνω του 90%. Μετά την αποτύπωση των οδών έγινε εκκαθάριση του οδικού δικτύου από τις γραμμές που ορίζανε τις πολλαπλές λωρίδες του κάθε οδικού άξονα με αποτέλεσμα η κάθε οδός να αναπαριστάται από μια και μόνο γραμμή. Η παραπάνω επεξεργασία έγινε καθώς το πλήθος των σημείων δεν επαρκούσε για να εξαχθεί αποτέλεσμα σε επίπεδο λωρίδας. Έπειτα με το κατάλληλο εργαλείο συνδυάστηκαν οι γραμμές του οδικού άξονα με τα σημεία των μηνυμάτων θέτοντας ως προϋπόθεση κατά την επεξεργασία το κάθε σημείο να απέχει έως 10 μέτρα από την εκάστοτε γραμμή. Η επιλογή της τιμής των 10 μέτρων έγινε καθώς μια τυπική λωρίδα στους δρόμους των ΗΠΑ έχει πλάτος

περίπου 3 μέτρα (10 feet) ενώ οι δρόμοι που υπάρχουν στο δίκτυο της Νέας Υόρκης διαθέτουν έως τρείς λωρίδες. Με την παραπάνω διαδικασία επιτεύχθηκε ο συνδυασμός της κάθε οδού με τα σημεία που είναι κοντά με αποτέλεσμα να είναι γνωστός ο αριθμός των σημείων που αντιστοιχεί σε κάθε γραμμή. Στο σημείο αυτό είχε επιτευχθεί η κάθε γραμμή στο Arcgis να διαθέτει στον πίνακα των στοιχείων της (attribute table) τον μοναδικό κωδικό της, το όνομα της οδού, τον τύπο της οδού, το μήκος της και τον αριθμό των σημείων, τα οποία αντιπροσωπεύουν τα μηνύματα, που αντιστοιχούν σε αυτή.

Ένα από τα βασικά προβλήματα στην διαδικασία είναι ο ορισμός της κάθε οδού ώστε να γίνει η κατάταξη τους σύμφωνα με την κυκλοφοριακή συμφόρηση που στατιστικά εμφανίζεται σε αυτές. Εάν ο ορισμός γίνει σύμφωνα με το όνομα που έχει η κάθε οδός τότε δεν είναι εφικτή η σύγκρισή τους καθώς το κάθε τμήμα του οδικού δικτύου θα έχει διαφορετικό μήκος με αποτέλεσμα να υπάρχει ο κίνδυνος να χαρακτηριστεί μια οδός κυκλοφοριακά επιβαρυμένη επειδή έχει μεγάλο μήκος άρα και περισσότερα σημεία σε όλη την έκταση της. Το αρχείο (shapefile) του οδικού δικτύου που ανακτήθηκε από το Openstreetmaps, απεικόνιζε τον κάθε δρόμο σε τμήματα με διαφορετικό μήκος το κάθε ένα. Εάν εξαγόταν συμπέρασμα για την κίνηση ανάλογα με τον αριθμό των σημείων που αντιστοιχούσε στο κάθε τμήμα δρόμου τότε και πάλι θα υπήρχε το ίδιο πρόβλημα στην κατάταξη όπως και προηγουμένως με το όνομα της κάθε οδού. Το πρόβλημα αυτό αντιμετωπίστηκε με την αναγωγή του πλήθους των σημείων στο μήκος του εκάστοτε τμήματος δρόμου. Η αναγωγή έγινε με απλή διαίρεση του πλήθους των σημείων που αντιστοιχούν στο κάθε τμήμα γραμμής προς το μήκος (μέτρα) της γραμμής. Επειδή ο συνολικός αριθμός των σημείων δεν ήταν πολύ μεγάλος από την αναγωγή προέκυψαν μικροί αριθμοί και για αυτό έγινε πολλαπλασιασμός του συντελεστή επί 1000, δηλαδή στην ουσία έγινε αναγωγή των σημείων στο χιλιόμετρο.

Μετά από την παραπάνω διαδικασία δημιουργήθηκαν χρωματικές κλάσεις στις οποίες έγινε αντιστοίχιση της κάθε γραμμής ανάλογα με τον συντελεστή που υπολογίσθηκε προηγουμένως. Επειδή στο αρχείο του οδικού δικτύου υπήρχαν τμήματα των οδών τα οποία είχαν πολύ μικρό μήκος, σε κάποιες περιπτώσεις υπήρχε μεγάλη διακύμανση της τιμής του συντελεστή σε αλληλουχίες τμημάτων των οδών που το συνολικό τους μήκος ήταν λίγες δεκάδες μέτρα. Το φαινόμενο αυτό δημιουργήθηκε από την αντιστοιχία πολλών σημείων σε ένα πολύ μικρό τμήμα της οδού. Καθώς τα σημεία που αντιπροσωπεύουν τα μηνύματα που ανακτήθηκαν, αναφέρονται σε μια κατάσταση του δεν εκτίνεται αυστηρώς τοπικά, με την έννοια ότι μπορεί ο χρήστης να συμπεριλαμβάνει την θέση του μηνύματος του αλλά η συμφόρηση τις περισσότερες φορές εκτείνεται σε μεγάλο μήκος της οδού, θεωρήθηκε σκόπιμο να απαλειφθούν οι γραμμές που εμπίπτουν στην περίπτωση που αναφέρθηκε προηγουμένως. Έτσι όπου υπήρχε μεγάλη διακύμανση στο οδικό δίκτυο και σε αυτό συνέβαλε τμήμα οδού που είχε μήκος κάτω από 100 μέτρα, έγινε επαναυπολογισμός του συντελεστή θεωρώντας ένα τμήμα, το μικρό τμήμα και τα γειτονικά του. Για να γίνει περισσότερο αντιληπτό παρατίθενται παρακάτω μια τέτοια περίπτωση (Εικόνα

12) όπου αρχικά το τμήμα της οδού με μήκος περίπου 100 μέτρα είχε συντελεστή 150 ενώ τα γειτονικά συντελεστή 0. Μετά την επεξεργασία θεωρήθηκαν και τα τρία τμήματα ως ένα με μήκος 400 μέτρα με αποτέλεσμα ο συντελεστής να υπολογιστεί 37 και να λάβει το αντίστοιχο χρωματισμό.

Εικόνα 12: Εκ των υστέρων επεξεργασία για την βέλτιστη απόδοση της συμφόρησης

Το τελικό προϊόν της παραπάνω ανάλυσης είναι ο χάρτης που εμφανίζεται στην επόμενη σελίδα (Χάρτης 10) ο οποίος καλύπτει όλη την περιοχή της πόλης της Νέας Υόρκης καθώς επίσης και ένας πίνακας (Πίνακας 10) με τα τμήματα δρόμων που σύμφωνα με την ανάλυση εμφανίζουν περισσότερες πιθανότητες δημιουργίας κυκλοφοριακής συμφόρησης. Φυσικά προτιμητέο θα ήταν η παραπάνω ανάλυση να σχετίζεται και με τον χρόνο δημιουργώντας χάρτες για κάθε ώρα της ημέρας αλλά κάτι τέτοιο δεν ήταν εφικτό λόγο περιορισμένου όγκου δεδομένων. Οι κλάσεις που εμφανίζονται στον χάρτη διαρθρώνονται ως εξής:

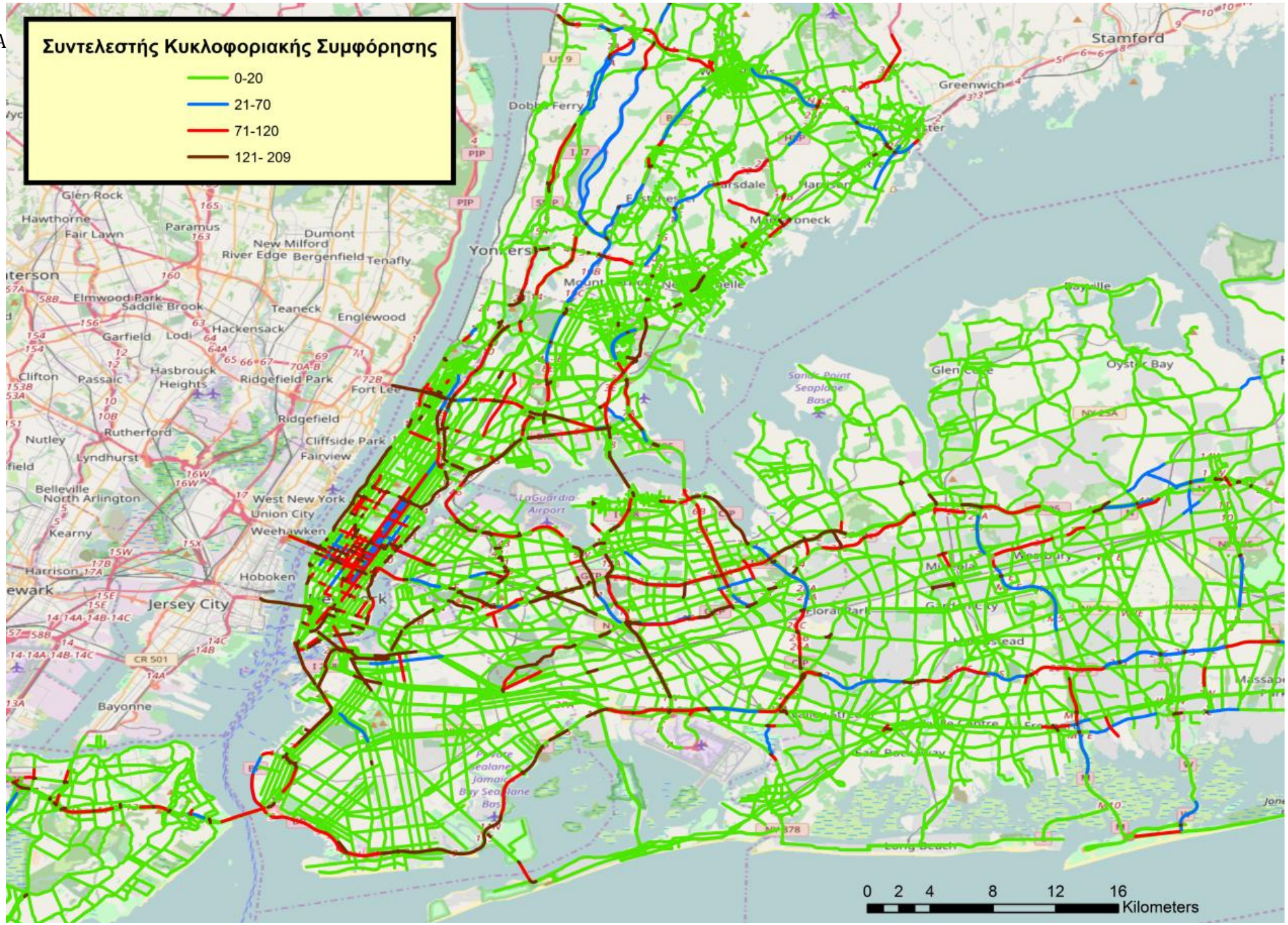
- Πράσινο: 0-20 σημεία ανά χιλιόμετρο
- Μπλέ: 21-70 σημεία ανά χιλιόμετρο
- Κόκκινο: 71-120 σημεία ανά χιλιόμετρο
- Καφέ: 121-209 σημεία ανά χιλιόμετρο

Για την αξιολόγηση των αποτελεσμάτων χρησιμοποιήθηκαν τα στοιχεία από την επίσημη έκθεση της εταιρίας INRIX στην οποία παρατίθεται μια λίστα (Πίνακας 9) με τους δρόμους της Νέας Υόρκης που εμφανίζουν την μεγαλύτερη καθυστέρηση στις διαδρομές των οδηγών. Τα τμήματα που αναφέρει η συγκεκριμένη λίστα, η οποία επίσης παρατίθεται παρακάτω, αποτυπώθηκαν στον χάρτη της Νέας Υόρκης για άμεση σύγκριση των αποτελεσμάτων (Χάρτης 11). Να σημειωθεί στο χάρτη από τα στοιχεία της INRIX δεν υπάρχει ομαδοποίηση των στοιχείων καθώς τα τμήματα αυτά είναι εκείνα με την περισσότερη συμφόρηση και οι διακυμάνσεις είναι μικρές. Στον πίνακα της INRIX η πρώτη στήλη υποδεικνύει την γενική κατάταξη του δρόμου , σε σχέση με την συμφόρηση, σε επίπεδο όλης της χώρας των ΗΠΑ. Όπως είναι φανερό , από την παρατήρηση των δύο χαρτών, η επεξεργασία των δεδομένων που ανακτήθηκαν από το Twitter μπόρεσε να αναδείξει όλους τους δρόμους που αναφέρει η έκθεση της INRIX ως περισσότερο προβληματικούς (worst corridors).

Φυσικά το ίδιο προκύπτει και από ενδελεχή σύγκριση των αποτελεσμάτων των δύο πινάκων που ακολουθούν η οποία θα σχολιαστεί παρακάτω.

Στην συνέχεια παρατίθεται ένα απόσπασμα χάρτη (Χάρτης 12) από την υπηρεσία Google Traffic όπου εμφανίζεται η κατανομή της κινησης στην Νέα Υόρκη σύμφωνα με τα δεδομένα της εταιρίας, για περεταίρω αξιολόγηση των αποτελεσμάτων. Η εφαρμογή της Google είναι υλοποιημένη με τέτοιο τρόπο ώστε να μην μπορεί ο εκάστοτε χρήστης να εμφανίσει στο χάρτη τα συγκεντρωτικά στατιστικά στοιχεία για την κίνηση αλλά να πρέπει να επιλέξει μια συγκεκριμένη μέρα της εβδομάδας και μια συγκεκριμένη ώρα. Η ανάλυση που παρουσιάστηκε παραπάνω βασίστηκε σε όλο τον όγκο των δεδομένων που αναφέρονταν στην Νέα Υόρκη χωρίς να γίνει ανάλυση ημέρας ή ώρας. Για να γίνει λοιπόν σύγκριση με τον χάρτη της Google έπρεπε να βρεθεί μια ημέρα και μια ώρα όπου η συμφόρηση κινείται σε κανονικά επίπεδα. Για την επιλογή αυτή βοήθησε το διάγραμμα που είχα δημιουργηθεί για την Νέα Υόρκη κατά την χρονική ανάλυση και τελικά επιλέχθηκε η ημέρα Παρασκευή και η ώρα 3 το απόγευμα. Όπως παρατηρείται από την σύγκριση των χαρτών τα τμήματα των οδών που υπάρχει πρόβλημα μπόρεσαν να αποδοθούν ικανοποιητικά με την παραπάνω ανάλυση. Φυσικά πολλά τμήματα δεν κατατάσσονται το ίδιο και στους δύο χάρτες το οποίοι είναι θέμα των κλάσεων που θέτει η κάθε ανάλυση και επίσης του όγκου των δεδομένων που έχει η κάθε έρευνα στην διάθεση της. Εάν υπήρχε μεγαλύτερος όγκος μηνυμάτων το αποτέλεσμα θα ήταν αισθητά καλύτερο.

Τέλος παρατίθεται ακόμα ένας χάρτης (Χάρτης 13) ο οποίος δείχνει την μέση ημερίσια κυκλοφορία (Average Annual Daily Traffic) των οδών της πόλης βασισμένος σε μετρήσεις του 2014 και προέρχεται από τις τοπικές αρχές της Νέας Υόρκης [47]. Η μέση ημερήσια κυκλοφορία είναι ο αριθμός των οχημάτων που έχουν διέλθει από την οδό επί ένα χρόνο διαιρεμένος με το 365. Η συμφόρηση σε μια οδό μπορεί να μετρηθεί με τον υπολογισμό του πηλίκου φόρτος προς την χωρητικότητα της οδού (volume to capacity ratio) άρα η μέση ημερήσια κυκλοφορία από μόνη της δεν μπορεί να καταδείξει αν η οδός πάσχει από το πρόβλημα της κυκλοφοριακής συμφόρησης. Παρά ταύτα θεωρήθηκε σκόπιμο να συμπεριληφθεί ο χάρτης με την μέση ημερήσια κυκλοφορία καθώς η Νεα Υόρκη αποδεδειγμένα είναι μια από τις περισσότερο προβληματικές πόλεις στις ΗΠΑ σχετικά με την κυκλοφοριακή συμφόρηση με αποτέλεσμα οι δρόμοι όπου εμφανίζεται περισσότερη κυκλοφορία να εμφανίζονται και περισσότερα περιστατικά συμφόρησης. Την ορθότητα του παραπάνω συλλογισμού την αποδεικνύει η ταύτιση σε πολλά σημεία του χάρτη που προέκυψε από την προηγούμενη ανάλυση και του χάρτη με την Μέση Ημερήσια Κυκλοφορία.



Χάρτης 10: Απεικόνιση συντελεστή κυκλοφοριακής συμφόρησης, Νέα Υόρκη

Χάρτης 11: Αποτύπωση των οδών με το μεγαλύτερο πρόβλημα συμφόρησης σύμφωνα με των εταιρία INRIX, Νέα Υόρκη, 2015

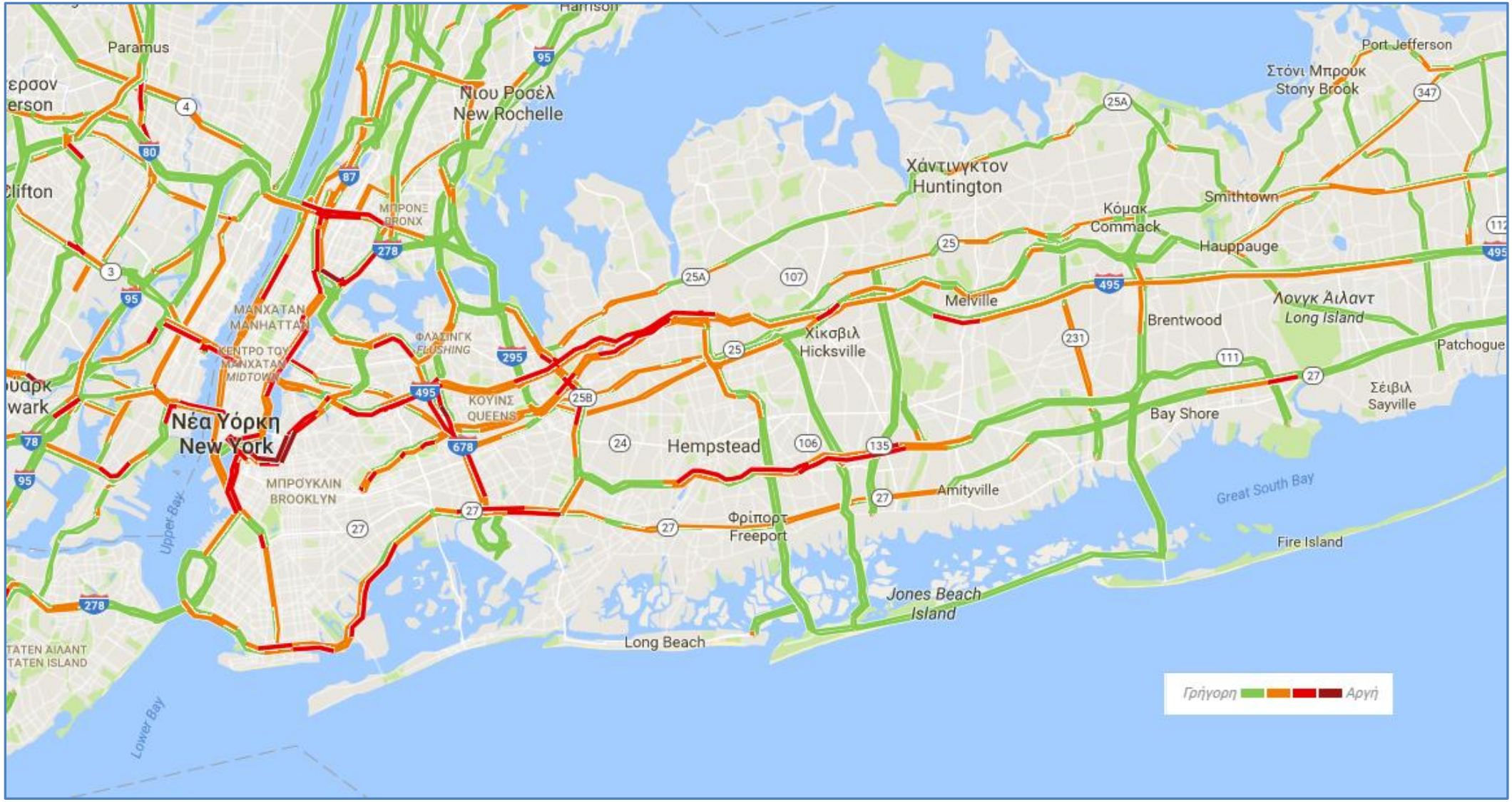
	Rank	Road(s)	Distance (miles)	Freeflow Travel Time (min)	Peak Travel Time (min)	Total Delay Per Year (hrs)	Worst Day/Hour Travel Time (min)
1	5	I-678	3.11	4	16	50	22
2	10	I-495	16.81	18	40	89	67
3	14	I-278	11.05	15	32	70	42
4	21	I-95	9.48	12	22	39	39
5	22	I-495	14.82	16	32	65	62
6	33	I-278	10.34	13	28	60	38
7	39	I-678	5.75	8	17	39	29
8	48	Southern State Pkwy	10.15	10	20	40	33
9	53	Leif Ericson Dr/Belt Pkwy/Shore Pkwy	3.55	4	9	21	17
10	59	I-95	6.70	8	12	18	20
11	60	I-278	3.74	4	9	18	19
12	62	Leif Ericson Dr/Belt Pkwy/Shore Pkwy	4.89	6	12	26	24
13	98	Cross Island Pkwy	7.45	8	15	27	25
14	99	FDR Dr	4.57	6	11	21	17
15	101	Grand Central Pkwy	10.65	12	21	36	35
16	107	I-278	3.04	4	7	15	17
17	118	I-87	3.89	5	9	18	18
18	138	I-95	1.26	2	3	6	6
19	144	Harlem River Dr	2.81	3	6	9	14
20	156	Grand Central Pkwy	4.63	5	9	17	19
21	189	NY-9A	5.53	7	10	15	23
22	207	Southern State Pkwy	10.93	11	17	26	28
23	235	Northern State Pkwy	3.41	3	6	9	10
24	249	I-495	3.19	3	5	7	10
25	255	Cross Island Pkwy	4.64	5	7	7	11
26	267	Hutchinson River Pkwy	4.47	5	7	8	11
27	306	I-87	3.52	4	6	8	9
28	310	Leif Ericson Dr/Belt Pkwy/Shore Pkwy	7.46	8	11	11	18
29	343	Leif Ericson Dr/Belt Pkwy/Shore Pkwy	3.51	4	5	4	6

Πίνακας 9: Οδοί με το μεγαλύτερο πρόβλημα συμφόρησης στην Νέα Υόρκη, σύμφωνα με την εταιρία INRIX
Πηγή: inrix.com

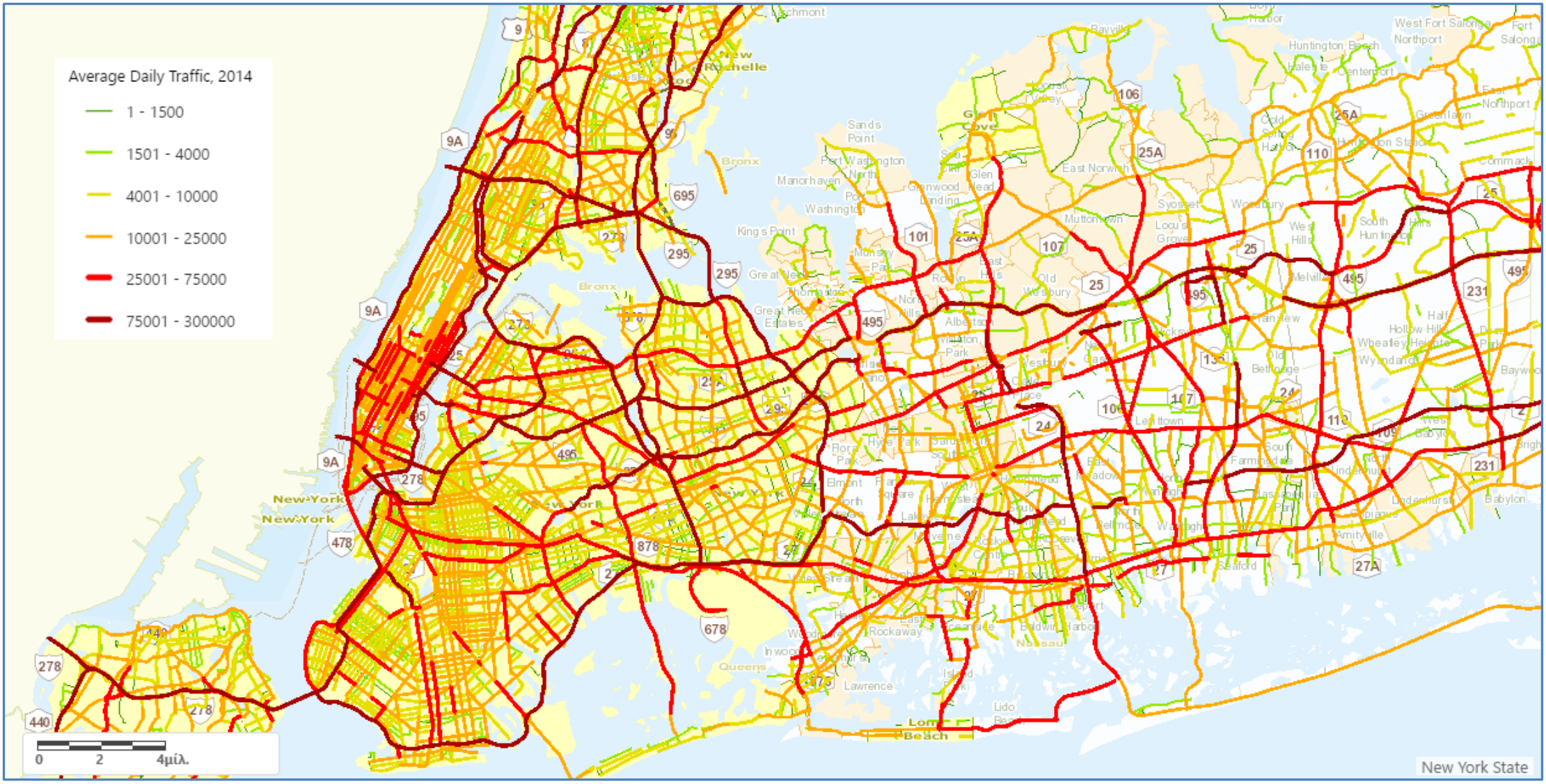
	Συντελεστής	Οδός	Τύπος	Αρίθμηση στοιχείων του πίνακα της INRIX
1	209	Cross Bronx Expressway	motorway	4
2	209	Major Deegan Expressway	motorway	17
3	208	Brooklyn-Queens Expressway	primary	16
4	204	Cross Bronx expressway	secondary	4
5	203	Brooklyn-Queens Expressway	motorway	16
6	197	Cross Bronx Expressway	motorway	18
7	196	Brooklyn-Queens Expressway	motorway	-
8	196	FDR Drive	motorway	14
9	191	FDR Drive	secondary	14
10	189	FDR Drive	secondary	14
11	183	Gowanus Expressway	motorway	11
12	183	FDR Drive	motorway	14
13	179	Cross Bronx Expressway	motorway	18
14	178	Cross Bronx Expressway	secondary	4
15	177	Cross Bronx Expressway	motorway	18
16	177	56th Road	secondary	-
17	173	Hutchinson River Parkway	motorway	-
18	168	Van Wyck Expressway	motorway	1
19	168	Belt Parkway	motorway	12
20	160	Cross Island Parkway	motorway	13
21	157	Brooklyn-Queens Expressway	motorway	3
22	156	Brooklyn-Queens Expressway	secondary	6
23	150	Henry Hudson Parkway	motorway	21
24	147	Belt Parkway	motorway	28
25	146	Goethals Bridge	motorway	16
26	143	Cross Bronx Expressway	motorway	4
27	142	Bronx River Parkway	motorway	10
28	138	Grand Central Parkway	motorway	-
29	137	Belt Parkway	motorway	29
30	134	Belt Parkway	motorway	28
31	133	Northern Boulevard	primary	-
32	133	Northern Boulevard	primary	-
33	131	Bronx River Parkway	primary	10
34	130	Humboldt Street	primary	-
35	130	Long Island Expressway	motorway	2
36	125	Trans-Manhattan Expressway	motorway	4
37	125	Long Island Expressway	trunk	2
38	125	Northern Boulevard	primary	-
39	124	Northern Boulevard	primary	-
40	122	Long Island Expressway	motorway	2

Πίνακας 10: Οι οδοί με την μεγαλύτερη συμφόρηση σύμφωνα με τον συντελεστή που υπολογίστηκε

Οι δύο παραπάνω πίνακες όπως έχει αναφερθεί ήδη αναφέρονται στα τμήματα οδών της Νέας Υόρκης με την μεγαλύτερη κυκλοφοριακή συμφόρηση σύμφωνα με την εταιρία INRIX και σύμφωνα με την χωρική ανάλυση που παρουσιάστηκε. Η κάθε εγγραφή του Πίνακα 9 έχει ένα αύξων αριθμό ο οποίος χρησιμοποιείται στον Πίνακα 10 για να γίνει η αντιστοιχία ανάμεσα στα δεδομένα της INRIX και της ανάλυσης της εργασίας. Στον Πίνακα 10 παρουσιάζονται τα 40 τμήματα οδών που σύμφωνα με την ανάλυση που διεξήχθη διαθέτουν τον μεγαλύτερο συντελεστή κυκλοφοριακής συμφόρησης (αριθμός σημείων προς μήκος τμήματος δρόμου, επί 1000). Μετά από σύγκριση των δύο πινάκων εξάγεται το συμπέρασμα ότι σχεδόν όλα τα τμήματα του Πίνακα 10 αντιστοιχούν σε κάποια εγγραφή του πίνακα της INRIX. Επειδή το οδικό δίκτυο στο οποίο έγινε η επεξεργασία αποτελούνταν από τμήματα γραμμών με μήκος κατά μέσο όρο ενός χιλιομέτρου, τα τμήματα οδών του Πίνακα 10 εμπεριέχονται στα τμήματα οδών του Πίνακα 9 και για αυτό το λόγο σε κάποιες περιπτώσεις γίνεται αντιστοιχία παραπάνω από μιας εγγραφής του Πίνακα 10 με μια εγγραφή του πίνακα της INRIX. Σε πολλές περιπτώσεις οι ονομασίες των οδών δεν είναι ίδιες και στους δύο πίνακες παρόλο που γίνεται αναφορά στον ίδιο δρόμο. Αυτό συμβαίνει καθώς τα περισσότερα τμήματα που αναφέρονται στους πίνακες είναι τμήματα μεγάλων συνδετήριων οδών τα ονόματα των οποίων χρησιμοποιεί η εταιρία INRIX στην έκθεση της, εν αντιθέσει με το Openstreetmaps (και κατ’ επέκταση ο Πίνακας 10) όπου χρησιμοποιούνται οι τοπικές ονομασίες των οδών.



Χάρτης 12: Μέση συμφόρηση σύμφωνα με το Google Traffic, 2016
Πηγή: maps.google.com



Χάρτης 13: Μέση Ημερήσια Κυκλοφορία, Νέα Υόρκη, 2014
Πηγή: dot.ny.gov/tdv

4.3.2 Υπολογισμός Χωρικής Αυτοσυσχέτισης

Μια από τις σημαντικότερες αναλύσεις που μπορεί να διεξαχθεί με βάση τα δεδομενα από το Twitter είναι η εξαγωγή των περιοχών που εμφανίζουν το μεγαλύτερο πρόβλημα κυκλοφοριακής συμφόρησης. Το συγκεκριμένο παράγωγο είναι πολύ σημαντικό για την χάραξη στατηγικής κατά την επεξεργασία για την διευθέτηση του κυκλοφοριακού προβλήματος ,όταν μας ενδιαφέρει να αντιμετωπίσουμε το πρόβλημα σε μια ευρύτερη περιοχή και όχι σε επίπεδο μεμονωμένων οδών.

Ο εντοπισμός των περιοχών με το μεγαλύτερο κυκλοφοριακό πρόβλημα διεξήχθει σε 10 μητροπολιτικές περιοχές των ΗΠΑ. Η κάθε μητροπολιτική περιοχή έχει καθοριστεί από το γραφείο απογραφής των ΗΠΑ(Census berau) από όπου ανακτήθηκε το αρχείο με το όρια τους σε όλη την επικράτεια των ΗΠΑ και αποτελεί στατιστική μονάδα. Στις περισσότερες περιπτώσεις η κάθε μητροπολιτική περιοχή αποτελείται από μια μεγάλη πόλη και τις γύρω κατοικημένες περιοχές. Στο τέλος της ανάλυσης επιχηρήται να καταταχθούν οι μητροπολιτικές περιοχές ανάλογα το πλήθος των περιοχών με μεγάλο κυκλοφοριακό πρόβλημα που διαθέτουν.Τα αποτελέσματα θα συγκριθούν με τα στοιχεία που διαθέτει η εταιρία Tomtom για τις συγκεκριμένες περιοχές για να ελεχθεί η αξιοπιστία τους.

Για να εξαχούν οι περιοχές που μας ενδιαφέρουν χρησιμοποιήθηκε η μέθοδος της χωρικής αυτοσυσχέτησης. Η χωρική αυτοσυσχέτιση μπορεί να οριστεί ως η σχέση μεταξύ των τιμών μιας ενιαίας μεταβλητής που προέρχεται από τη γεωγραφική ρύθμιση των περιοχών στις οποίες εμφανίζονται αυτές οι τιμές. Μετρά την ομοιότητα των δεδομένων μέσα σε μια περιοχή, τον βαθμό δηλαδή στον οποίο ένα χωρικό φαινόμενο αυτοσυσχετίζεται στην περιοχή αυτή ,το επίπεδο αλληλεξάρτησης μεταξύ των μεταβλητών, τη φύση και τη δύναμη της αλληλεξάρτησης αυτής. Έτσι, η χωρική αυτοσυσχέτιση αποτελεί μια αξιολόγηση του συσχετισμού μιας μεταβλητής αναφορικά με τη χωρική της θέση και ακόμη αξιολογεί εάν οι τιμές είναι αλληλένδετες ή εάν υπάρχει ένα χωρικό σχέδιο στη σχέση αυτή. Τα εργαλεία χωρικής αυτοσυσχέτισης εξετάζουν αν η παρατηρηθείσα τιμή μιας μεταβλητής σε μια περιοχή είναι ανεξάρτητη από τις τιμές της μεταβλητής στις γειτονικές τοποθεσίες. Η χωρική αυτοσυσχέτιση μπορεί να είναι θετική ή αρνητική. Όταν είναι θετική, όλες οι παρόμοιες τιμές εμφανίζονται συγκεντρωμένες μαζί ενώ όταν είναι αρνητική οι ανόμοιες τιμές είναι αυτές που εμφανίζονται σε κοντινή απόσταση. Μια θετική χωρική αυτοσυσχέτιση αναφέρεται σε ένα χωρικό πρότυπο όπου τα γεωγραφικά χαρακτηριστικά των παρόμοιων τιμών τείνουν να ομαδοποιηθούν σε ένα χάρτη, ενώ μια αρνητική χωρική αυτοσυσχέτιση δείχνει ένα χωρικό πρότυπο στο οποίο οι γεωγραφικές οντότητες των παρόμοιων τιμών διασκορπίζονται σε ολόκληρο το χάρτη. Όταν η χωρική αυτοσυσχέτιση δεν έχει στατιστική σημαντικότητα τότε το χωρικό πρότυπο κατανομής χαρακτηρίζεται ως τυχαίο. Η γενική μέθοδος στην περιγραφή της αυτοσυσχέτισης για μια μεταβλητή είναι να υπολογιστεί κάποιος δείκτης συνδιακύμανσης για μια σειρά αποστάσεων (ή κατηγοριών απόστασης) από κάθε σημείο. Έτσι, προκύπτει το διάγραμμα συσχετίσεων που επεξηγεί την

αυτοσυσχέτιση σε κάθε συνδυασμό αποστάσεων. Η συμμετοχή σε μια δεδομένη ομάδα καθορίζεται με την ανάθεση ενός βάρους σε κάθε ζευγάρι σημείων. Αυτό το ειδικό βάρος είναι ένας απλός δείκτης που παίρνει την τιμή 1 αν είναι μέσα στην ομάδα ή την τιμή 0. Όταν υπάρχει συστηματική χωρική κατανομή στις τιμές μιας δεδομένης μεταβλητής υπάρχει και χωρική αυτοσυσχέτιση. Η κατανομή αυτή μπορεί να έχει θετική ή αρνητική χωρική αυτοσυσχέτιση. Στη θετική περίπτωση, η τιμή μιας μεταβλητής σε μια δεδομένη θέση τείνει να είναι παρόμοια με τις τιμές της μεταβλητής στις κοντινές θέσεις. Αυτό σημαίνει ότι αν η αξία κάποιας μεταβλητής είναι χαμηλή σε μια δεδομένη θέση, η παρουσία θετικής χωρικής αυτοσυσχέτισης δείχνει ότι οι κοντινές τιμές είναι επίσης χαμηλές. Αντίθετα, η αρνητική χωρική αυτοσυσχέτιση χαρακτηρίζεται από ανόμοιες τιμές σε κοντινές θέσεις και επομένως μια χαμηλή τιμή μπορεί να περιβάλλεται από τις υψηλές αξίες στις κοντινές θέσεις. Ο εντοπισμός της θετικής αυτοσυσχέτισης βοηθά στην ερμηνεία της ευρύτερης περιοχής, ότι εκεί δηλαδή οι οντότητες λειτουργούν με την ίδια ένταση ενώ με τον εντοπισμό της αρνητικής αυτοσυσχέτισης εντοπίζονται οι περιοχές που διακόπτουν τη συνέχεια ενός φαινομένου και επομένως, ανάλογα με την ένταση των τιμών, οι περιοχές αυτές υστερούν ή υπερτερούν σε σχέση με την ευρύτερη περιοχή. Και στις δύο περιπτώσεις, τα συμπεράσματα σχετίζονται με τον χαρακτηρισμό αναγκαιότητας επέμβασης στην περιοχή. Το χωρικό πρότυπο μιας διανομής καθορίζεται από τη ρύθμιση των μεμονωμένων οντοτήτων στο διάστημα και τις γεωγραφικές σχέσεις μεταξύ τους. Η ικανότητα του καθορισμού των χωρικών προτύπων είναι μια προϋπόθεση στην κατανόηση των περίπλοκων χωρικών διαδικασιών που κρύβονται κάτω από την κατανομή ενός φαινομένου. Η χωρική αυτοσυσχέτιση δείχνει το βαθμό στον οποίο το περιστατικό ενός χαρακτηριστικού γνωρίσματος επηρεάζεται από τα παρόμοια χαρακτηριστικά γνωρίσματα στην παρακείμενη περιοχή. Έτσι, οι στατιστικές της χωρικής αυτοσυσχέτισης παρέχουν χρήσιμους δείκτες για τα πρότυπα αυτά.Η μέτρηση της συσχέτισης μεταξύ των γειτονικών παρατηρήσεων σε ένα πρότυπο γίνεται χρησιμοποιώντας διάφορους δείκτες.Στην παρούσα εργασία χρησιμοποιείται ο δείκτης Moran I και στις δυο εκδοχές του, τον ολικό δείκτη και τον τοπικό δείκτη Moran I.

Ολικός δείκτης Moran's I

Ο πρώτος ορισμός του δείκτη Moran's I για μια μεταβλητή *X* με τιμές *xi* για κάθε χωρική οντότητα *i*, ο οποίος προτάθηκε από τον Moran είναι:

$$I = \frac{n}{2A} \frac{\sum_i^n \sum_j^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i^n (x_i - \bar{x})^2}$$

όπου *n* είναι ο αριθμός των χωρικών οντοτήτων (παρατηρήσεων), είναι ο μέσος των τιμών *xi*, *A* είναι ο συνολικός αριθμός συνδέσεων στο σύστημα με βάση τη γειτνίαση και *wij* είναι τα βάρη που ορίζονται με βάση τη χωρική εγγύτητα μεταξύ των

παρατηρήσεων. Ο μαθηματικός τύπος των Cliff και Ord (1973, 1981) που χρησιμοποιείται σήμερα για τον υπολογισμό του δείκτη Moran's I είναι:

$$I = \frac{n\sum_i^n \sum_j^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_i^n \sum_j^n w_{ij})\sum_i^n (x_i - \bar{x})^2}$$

όπου *n* είναι ο αριθμός των χωρικών οντοτήτων (παρατηρήσεων) και *wij* είναι τα βάρη που ορίζονται με βάση τη χωρική εγγύτητα μεταξύ των παρατηρήσεων. Ο ολικός δείκτης Moran's I παίρνει τιμές από –1 ως +1 και η ερμηνεία του είναι παρόμοια με αυτή του συντελεστή συσχέτισης:

1. Τιμές κοντά στο +1 υποδηλώνουν ισχυρή θετική χωρική αυτοσυσχέτιση (αναμένονται χωρικά πρότυπα στα οποία γειτονικές παρατηρήσεις τείνουν να έχουν παρόμοια υψηλές ή χαμηλές τιμές μιας μεταβλητής),
2. Τιμές κοντά στο –1 υποδηλώνουν ισχυρή αρνητική χωρική αυτοσυσχέτιση (π.χ. υψηλές τιμές μιας μεταβλητής τείνουν να βρίσκονται κοντά σε χαμηλές τιμές)
3. Τιμές κοντά στο 0 υποδηλώνουν απουσία χωρικής αυτοσυσχέτισης και επομένως χωρικών προτύπων.

Αξίζει να σημειωθεί ότι γενικά τα χωρικά δεδομένα τείνουν να έχουν θετική χωρική αυτοσυσχέτιση - ως απόρροια του πρώτου νόμου της γεωγραφίας- ενώ η αρνητική χωρική αυτοσυσχέτιση είναι σπάνια [48].Για να αξιολογηθεί αν ο δείκτης χωρικής αυτοσυσχέτισης είναι στατιστικά σημαντικός ή όχι έχει αναπτυχθεί ένα θεωρητικό πλαίσιο που επιτρέπει να διαφανεί αν ο δείκτης Moran's I υποδηλώνει μια συγκεκριμένη χωρική κατανομή των χωρικών δεδομένων που δεν είναι τυχαία. Σύμφωνα με τον Rogerson (2010), αν ο αριθμός των παρατηρήσεων είναι αρκετά μεγάλος, η κατανομή των δειγματικών *I*, υπό την υπόθεση ότι τα χωρικά δεδομένα είναι τυχαία κατανεμημένα στο χώρο, προσεγγίζει την κανονική κατανομή, και ο μέσος και η διακύμανση του *I* μπορούν να χρησιμοποιηθούν στον ορισμό ενός στατιστικού μέτρου Z ως εξής:

$$Z = \frac{I - E[I]}{\sqrt{V[I]}}$$

όπου τα *E*[*I*] και *V*[*I*] είναι αντίστοιχα η αναμενόμενη μέση τιμή και διακύμανση του *I* όταν η μηδενική υπόθεση τυχαίας κατανομής των χωρικών οντοτήτων είναι αληθής. Ουσιαστικά γίνεται ένας έλεγχος υποθέσεων με μηδενική υπόθεση H0=0 ότι ο δείκτης Moran's I είναι 0, που σημαίνει ότι δεν υπάρχει κάποιο χωρικό πρότυπο στα δεδομένα και άρα αυτά είναι τυχαία κατανεμημένα στο χώρο, έναντι της αμφίδρομης

εναλλακτικής υπόθεσης H1≠0 όπου ο δείκτης Moran's I δεν είναι 0, γεγονός που καταδεικνύει ότι τα δεδομένα δεν είναι τυχαία κατανεμημένα στο χώρο και εμφανίζουν συγκριμένα χωρικά πρότυπα ή έχουν συγκεκριμένη χωρική δομή και κατανομή. Αν το στατιστικό μέτρο Z που θα υπολογιστεί είναι μεγαλύτερο του +1,96 ή μικρότερο του -1,96 σε επίπεδο σημαντικότητας α=0.05 (95%), τότε απορρίπτεται η H0, επομένως ο σχετικός δείκτης Moran's I μπορεί να θεωρηθεί στατιστικά σημαντικός.

Τοπικός δείκτης Moran's I

Ο Anselin (1995) προτείνει τη χωρική αποδόμηση των ολικών δεικτών χωρικής αυτοσυσχέτισης και τον ορισμό *τοπικών δεικτών χωρικής σχέσης* (*local indicators of spatial association – LISA*) όπως ο τοπικός δείκτης του Moran *I*. Ο τοπικός δείκτης *Ii* μιας μεταβλητής *X,* με τιμές *xi* για κάθε χωρική οντότητα *i* και μέσο ορίζεται ως εξής (Anselin 1995):

$$I_i = \frac{x_i - \bar{x}}{m_2} \sum_{j=1}^{k} w_{ij}(x_j - \bar{x}),\ j \neq i$$

Όπου $m_2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / n$

Η ερμηνεία του τοπικού δείκτη Moran *Ii* είναι ουσιαστικά η ίδια με αυτή του ολικού δείκτη *I*. Η θετική τιμή του τοπικού *Ii* υποδεικνύει χωρική συγκέντρωση παρόμοιων τιμών (χαμηλών ή υψηλών) ενώ η αρνητική τιμή του *Ii* υποδεικνύει χωρική συγκέντρωση ανόμοιων τιμών, για παράδειγμα μια τοποθεσία με υψηλή τιμή που περιβάλλεται από γείτονες με χαμηλές τιμές (Anselin 1995).

Στην περίπτωση του τοπικού δείκτη Moran's I δεν ισχύει το όριο –1 ως +1 στην τιμή του, αλλά μπορεί να έχει τιμές μεγαλύτερες από +1 και μικρότερες από –1. Με βάση το συνδυασμό των προσήμων των τιμών των ζευγών αυτών και το επίπεδο σημαντικότητας των τοπικών δεικτών Moran's I, είναι δυνατή η δημιουργία του χάρτη χωρικών προτύπων. Πρόκειται για ένα θεματικό χάρτη όπου κάθε χωρική οντότητα έχει ταξινομηθεί σε μία από τις παρακάτω πέντε κατηγορίες:

1. **Υψηλή – Υψηλή (High – High) που αφορά σε χωρικές οντότητες με υψηλή τιμή που συνορεύουν με οντότητες με επίσης υψηλές τιμές της υπό μελέτη μεταβλητής**
2. **Χαμηλή – Χαμηλή (Low – Low) που αφορά σε χωρικές οντότητες με χαμηλή τιμή που συνορεύουν με οντότητες με επίσης χαμηλές τιμές της υπό μελέτη μεταβλητής**
3. **Χαμηλή – Υψηλή (Low – High) που αφορά σε χωρικές οντότητες με χαμηλή τιμή που συνορεύουν με οντότητες με υψηλές τιμές της υπό μελέτη μεταβλητής**

4. **Υψηλή – Χαμηλή (High – Low) που αφορά σε χωρικές οντότητες με υψηλή τιμή που συνορεύουν με οντότητες με χαμηλές τιμές της υπό μελέτη μεταβλητής**
5. **Μη στατιστικά σημαντικό τοπικό δείκτη Moran's I**

Κατά την επεξεργασία αποτυπώθηκαν όλα τα σημεία στον χάρτη των ΗΠΑ όπως επίσης και τα όρια των μητροπολιτικών περιοχών, στην συνέχεια αντιστοιχήθηκαν τα σημεία με τα πολύγωνα που ορίζουν την κάθε περιοχή και απομωνόθηκαν οι 10 περιοχές που διέθεταν τα περρισότερα σημεία. Ο δείκτης Moran και στις δύο του εκδοχές όπως αναφέρθηκε παραπάνω μετράει την χωρική αυτοσυσχέτιση κάποιας μεταβλητής, επειδή τα δεδομένα που έχει στην διάθεση της η εργασία είναι απλά σημεία που αντιπροσωπεύουν τα μηνύματα από το Twitter και δεν διέθεταν κάποια μεταβλητή έπρεπε να δημιουργηθούν οντότητες από τα σημεία τα οποία να σχετίζονται με κάποια αντιπροσωπευτική μεταβλητή. Για τον λόγο αυτό δημιουργήθηκε ένας κάνναβος πάνω από κάθε περιοχή ο οποίος αποτελούνταν από τετράγωνα με ορισμένο μέγεθος(500*500 μέτρα). Σε κάθε τετράγωνο αντιστοιχήκαν τα σημεία που βρίσκονταν εντός αυτού και έτσι δημιουργήθηκαν οι ζητούμενες οντότητες με μεταβλητή το πλήθος των σημείων που τους αντιστοιχούν. Εν τέλει υπολογίσθηκε ο ολικός δείκτης Moran I για κάθε μια από τις δέκα μητροπολιτικές περιοχές ο οποίος παρατίθεται στον παρακάτω πίνακα (Πίνακας 11).

Μητροπολιτική περιοχή	Ολικός δείκτης Moran I	z-scores
Philadelphia	**0.0461**	**9.165**
Dallas	**0.0295**	**9.129**
Houston	**0.0856**	**13.725**
Seattle	**0.0383**	**5.675**
New York	**0.1193**	**38.763**
Columbus	**0.0045**	**2.061**
Detroit	**0.0391**	**3.603**
Birmingham	**0.0489**	**4.400**
Indianapolis	**0.0376**	**8.802**
Nashville	**0.0101**	**2.536**

Πίνακας 11: Ολικός δείκτης Moran's I

Όπως διαφαίνεται από τον παραπάνω πίνακα για όλες τις μητροπολιτικές περιοχές ο ολικός δείκτης Moran I παίρνει θετικές τιμές οι γεγονός που αναδεικνύει ότι δημιουργούνται χωρικά πρότυπα στα οποία γειτονικές παρατηρήσεις τείνουν να έχουν παρόμοια υψηλές ή χαμηλές τιμές της μεταβλητής. Επίσης όπως ήταν αναμενόμενο σε όλες τις περιπτώσεις τα z-scores είναι θετικά και πάνω από το όριο του +1.96. Έτσι ο δείκτης Moran κρίνεται στατιστικά σημαντικός και η πιθανότητα τα στοιχεία να μην είναι κατανεμημένα στο χώρο τυχαία κυμαίνεται σε όλες τις περιοχές άνω του 99%. Τα αποτελέσματα του δείκτη Moran I προφανώς ήταν αναμενόμενα καθώς από την καθημερινή εμπειρία στα οδικά δίκτυα εξάγεται εύκολα το συμπέρασμα ότι κάποιες περιοχές τείνουν τα συγκενρώνουν περισσότερες πιθανότητες να εμφανιστεί κυκλοφοριακή συμφόρηση από κάποιες άλλες.

Καθώς πιστοποιήθηκε , με την χρήση του ολικού δείκτη Moran I, ότι τα δεδομενα σε όλες τις περιοχές δημιουργούν χωρικά πρότυπα, δηλαδή ομαδοποιούνται, έγινε χρήση του τοπικού δείκτη Moran I σε κάθε περιοχή ώστε να γίνει ανάδειξη των περιοχών που εμπεριέχουν πολλά σημεία και γειτνιάζουν με περιοχές με επίσης μεγάλο πλήθος σημείων. Προφανώς χρησιμοποιήθηκαν τα ίδια πολύγωνα που δημιουργήθηκαν κατά την επεξεργασία για τον υπολογισμό του ολικού Moran I. Στις παρακάτω σελίδες παρατίθενται οι θεματικοί χάρτες και των δέκα μητροπολιτικών περιοχών όπου η κάθε περιοχή μία από τις παρακάτω πέντε κατηγορίες:

1. **Κόκκινο: Υψηλή – Υψηλή (High – High) που αφορά σε περιοχές με υψηλή τιμή που συνορεύουν με περιοχές με επίσης υψηλές τιμές**

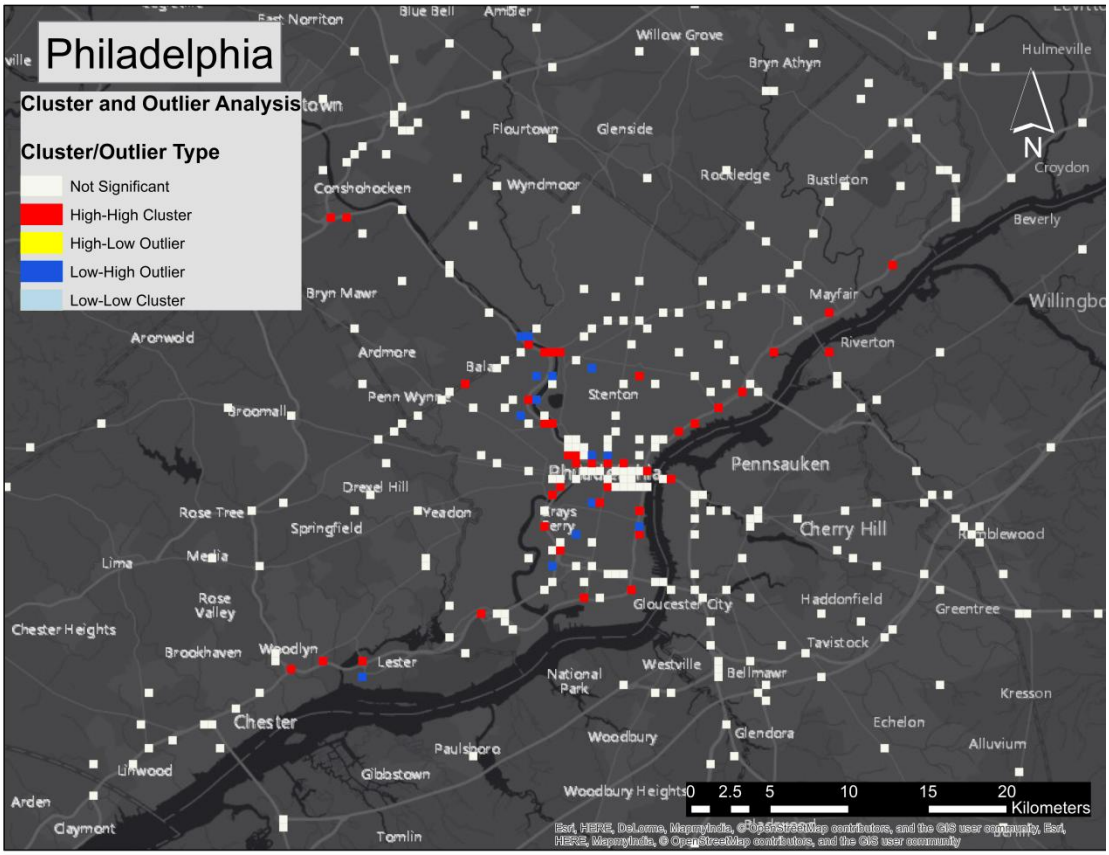
2. **Γαλάζιο: Χαμηλή – Χαμηλή (Low – Low) που αφορά σε περιοχές με χαμηλή τιμή που συνορεύουν με περιοχές με επίσης χαμηλές τιμές**

3. **Μπλε: Χαμηλή – Υψηλή (Low – High) που αφορά σε περιοχές με χαμηλή τιμή που συνορεύουν με περιοχές με υψηλές τιμές**

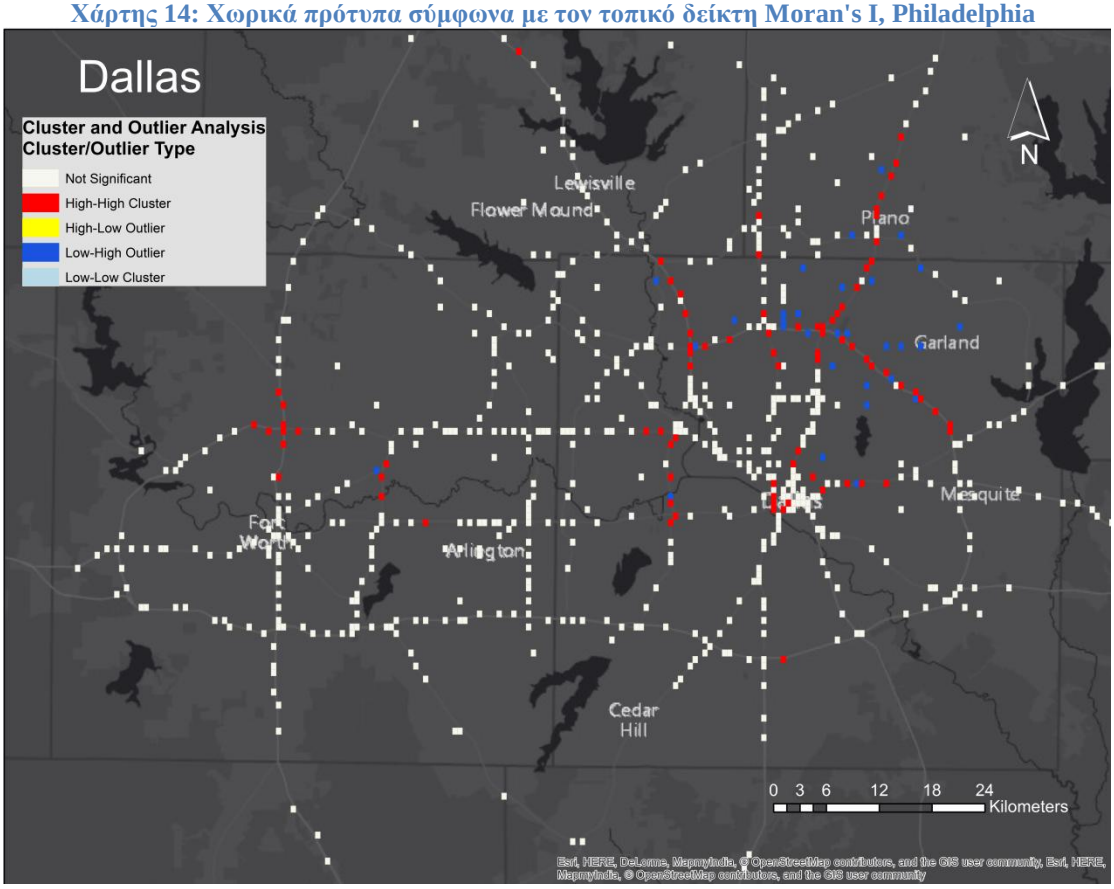
4. **Κίτρινο: Υψηλή – Χαμηλή (High – Low) που αφορά σε περιοχές με υψηλή τιμή που συνορεύουν με περιοχές με χαμηλές τιμές**

5. **Λευκό: Μη στατιστικά σημαντικό τοπικό δείκτη Moran's I**

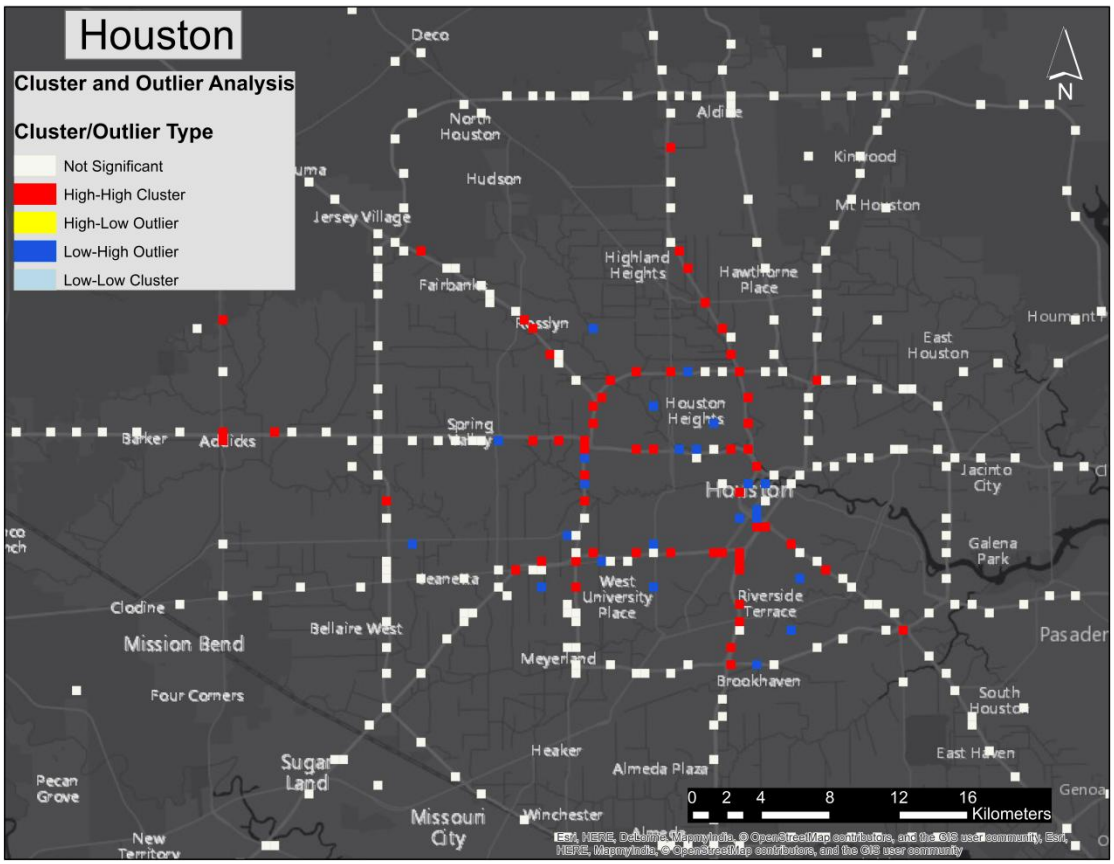
Στην παρούσα ανάλυση μας ενδιαφέρουν οι περιοχές με κόκκινο χρώμα οι οποίες εμφανίζουν μεγάλες τιμές μεταβλητής (πλήθος σημείων) και γύρω τους υπάρχουν περιοχές με εξίσου μεγάλες τιμές μεταβλητής. Οι περιοχές αυτές είναι οι ζητούμενες καθώς μαρτυρούν ότι στις ίδιες και γύρω από αυτές υπάρχει συσσώρευση περιστατικών κυκλοφοριακής συμφόρησης σύμφωνα με τα δεδομένα που διαθέτουμε.



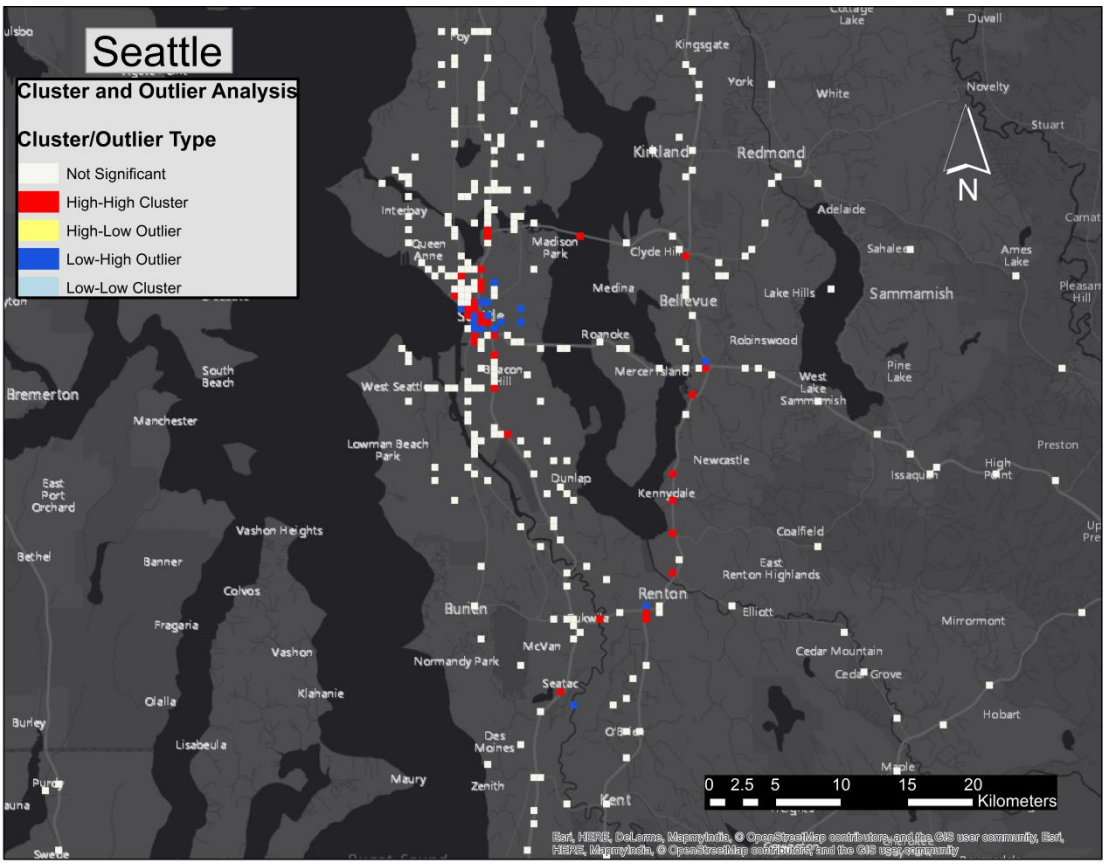
Χάρτης 14: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Philadelphia



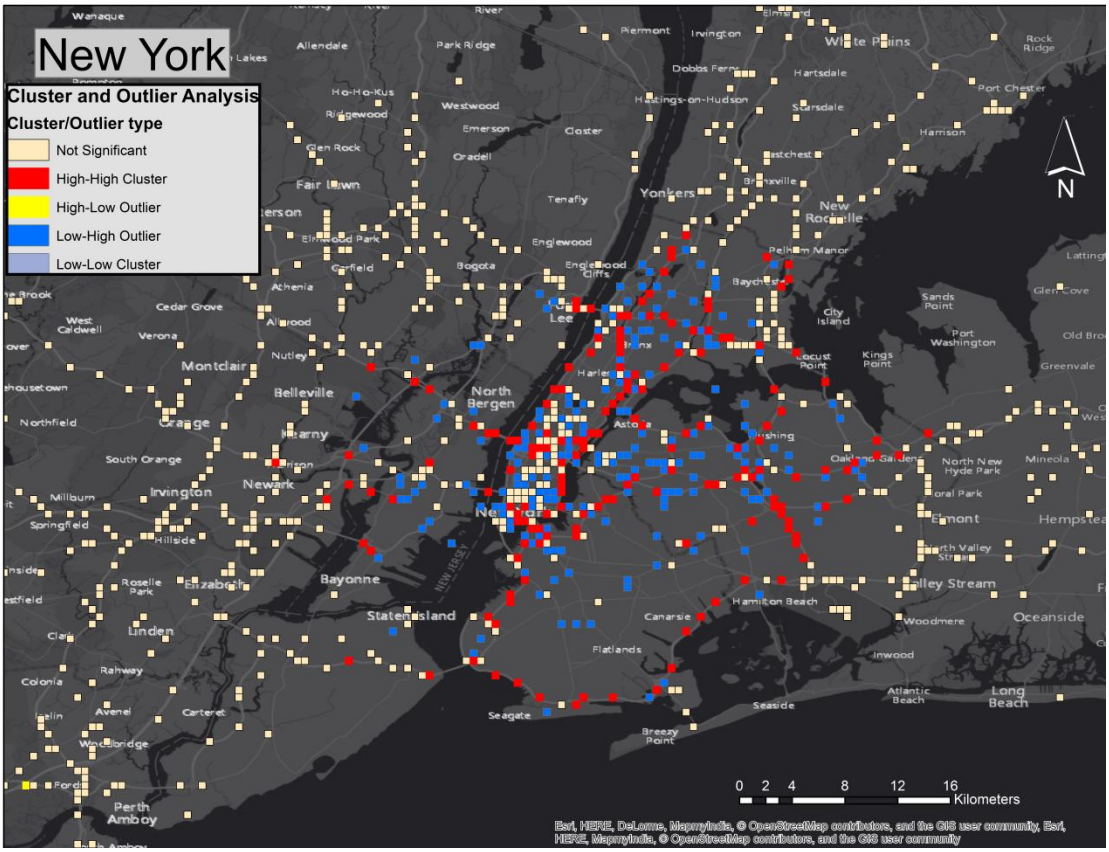
Χάρτης 15: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Dallas



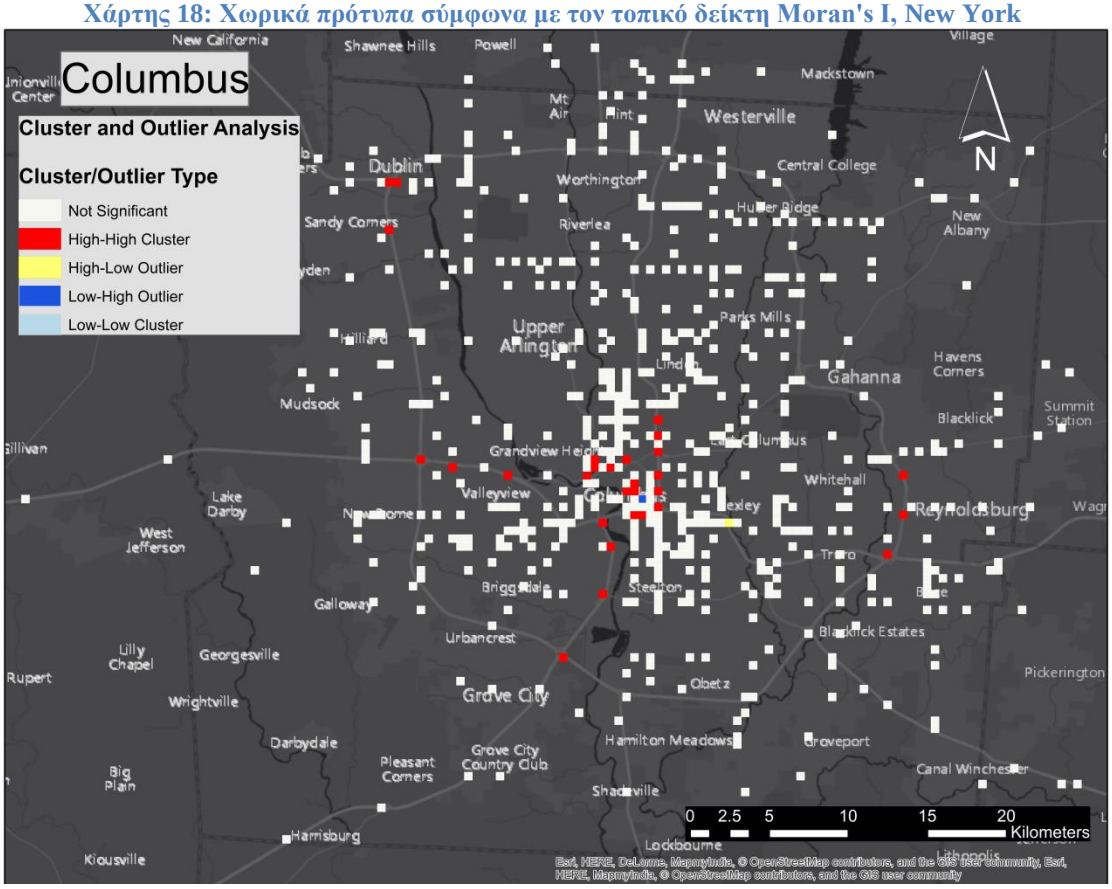
Χάρτης 16: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Houston



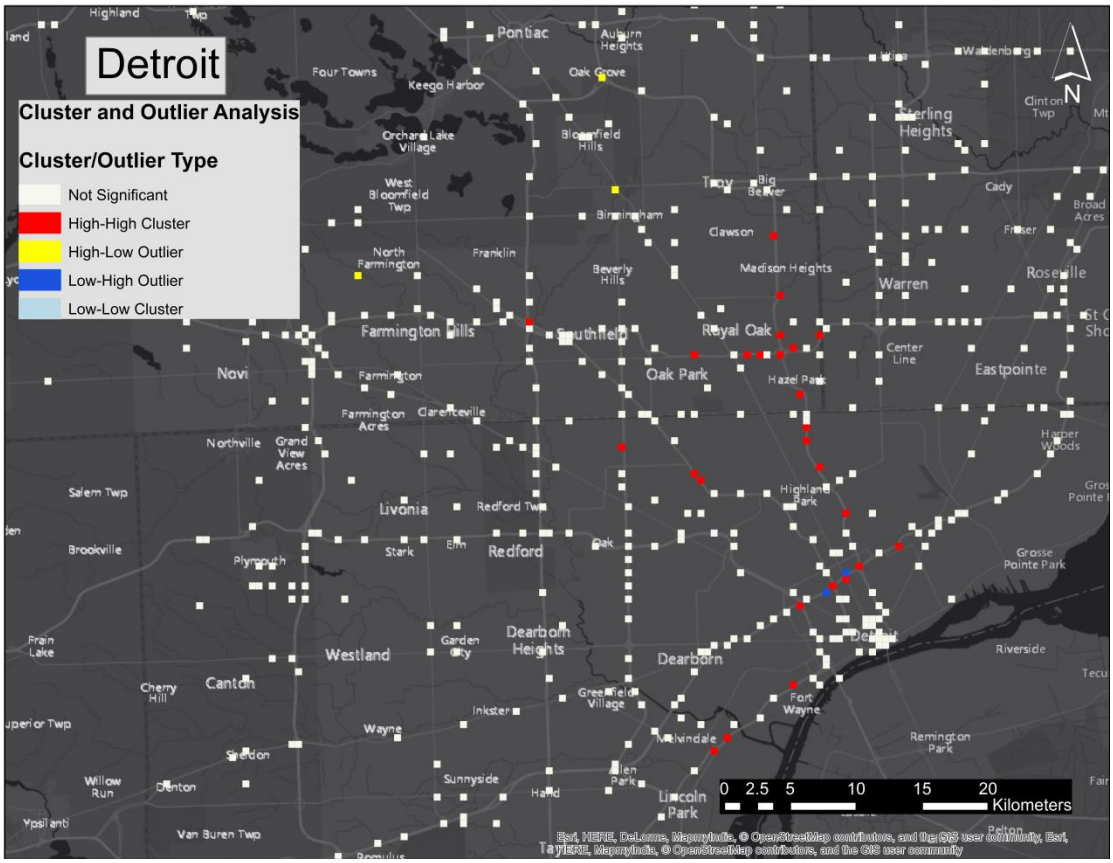
Χάρτης 17: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Seattle



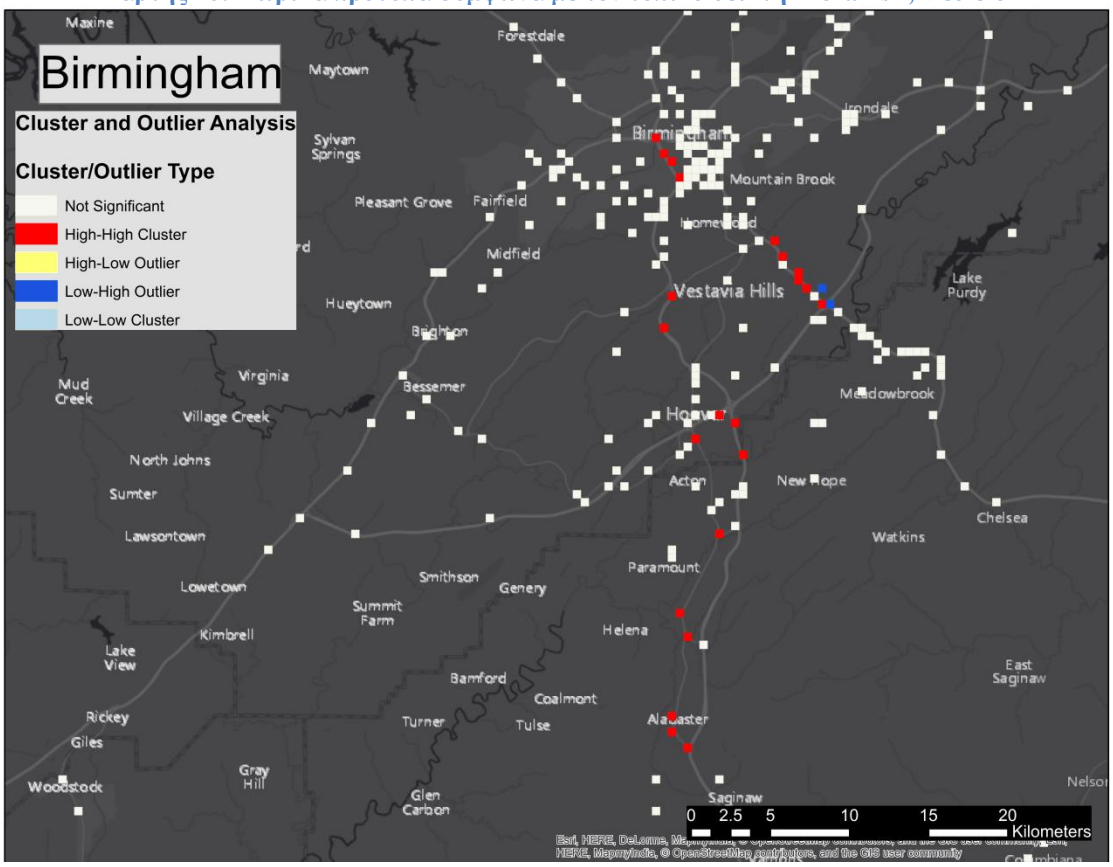
Χάρτης 18: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, New York



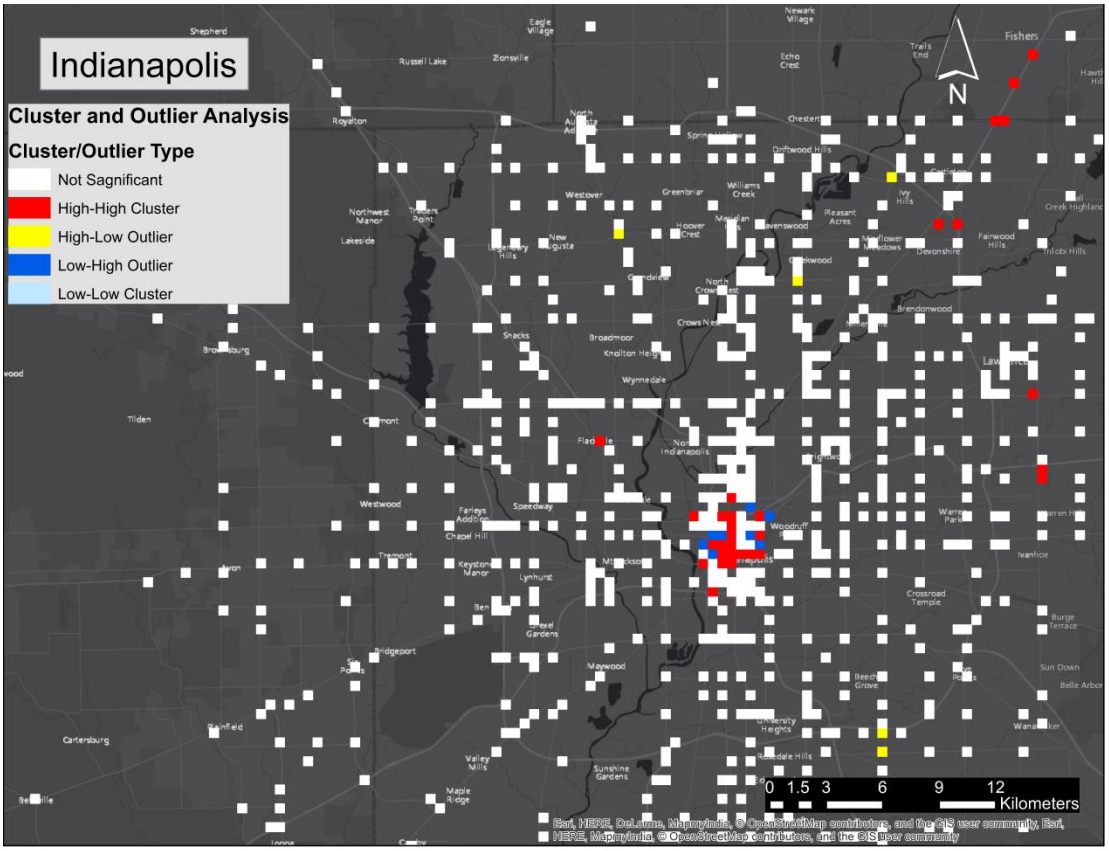
Χάρτης 19: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Columbus



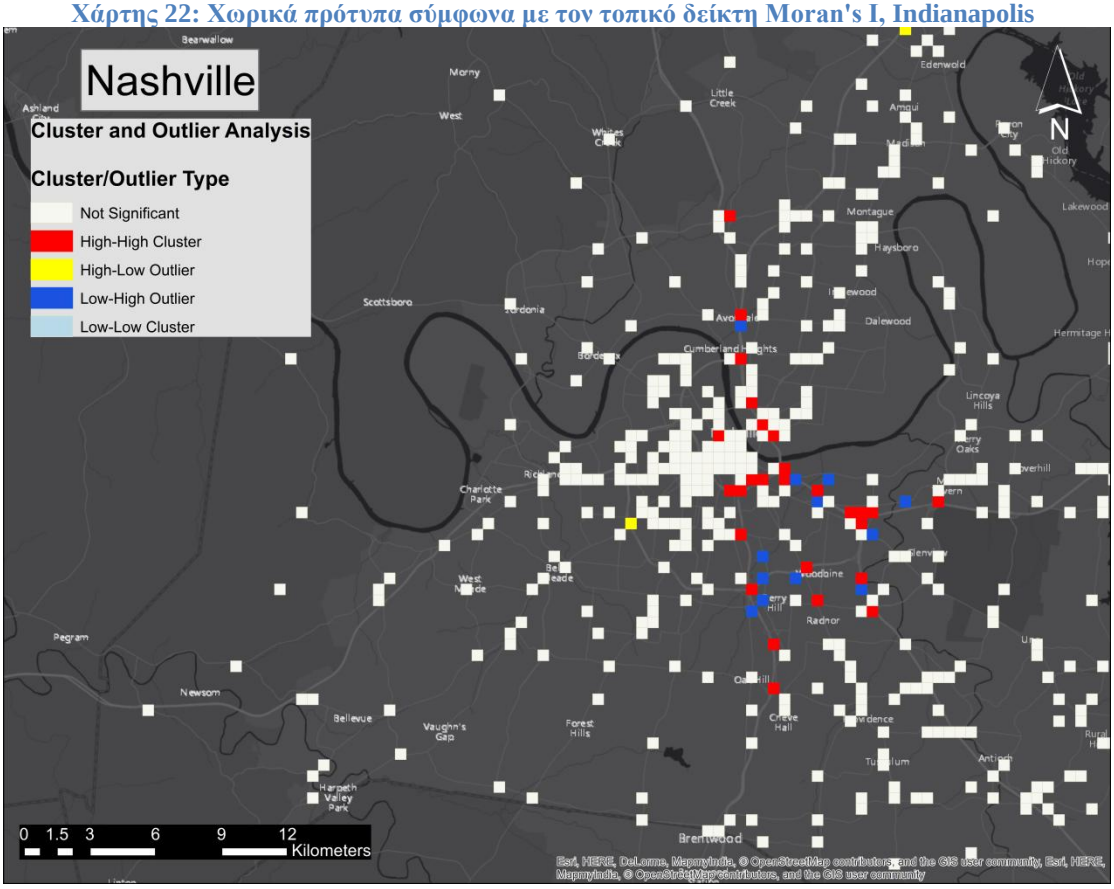
Χάρτης 20: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Detroit



Χάρτης 21: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Birmingham



Χάρτης 22: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Indianapolis



Χάρτης 23: Χωρικά πρότυπα σύμφωνα με τον τοπικό δείκτη Moran's I, Nashville

Με την παραπάνω ανάλυση επιτεύχθηκε η ανάδειξη των υποπεριοχών σε κάθε πόλη οι οποίες εντάσσονται σε μια γενικότερη περιοχή που το πρόβλημα της συμφόρησης είναι περισσότερο έντονο. Το πλήθος των υποπεριοχών αυτών μπορεί να αποτελέσει ένα δείκτη για το πόσο μεγάλο πρόβλημα υπάρχει στην κάθε πόλη. Στον παρακάτω πίνακα (Πίνακας 12) παρουσιάζεται το πλήθος των παραπάνω υποπεριοχών για τις 10 πόλεις που υπολογίσθηκε ο τοπικός δείκτης Moran I και η κατάταξη τους ανάλογα με την καθυστέρηση που βιώνουν κάθε χρόνο οι πολίτες σύμφωνα με την εταιρία Tomtom. Όπως είναι φανερό η κατάταξη που προήλθε από την παραπάνω επεξεργασία μπόρεσε να αναδείξει επιτυχώς ποιες πόλεις έχουν μεγαλύτερο πρόβλημα από τις υπόλοιπες. Να σημειωθεί ότι η αρίθμηση κατά Tomtom αφορά την κατάταξη γενικά των πόλεων σε επίπεδο χώρας. Οι μητροπολιτικές περιοχές οι οποίες τοποθετήθηκαν σε λάθος θέση στην κατάταξη που προήλθε από την παραπάνω επεξεργασία είναι η Columbus και το Nashville γεγονός που θα μπορούσε να αποφευχθεί αν υπήρχε μεγαλύτερος όγκος δεδομένων στην διάθεση της εργασίας.

Μητροπολιτική περιοχή	Κατάταξη	Κατάταξη Tomtom	Congestion Level
New York	165	3	33%
Seattle	81	4	31%
Houston	58	11	25%
Philadelphia	43	17	23%
Dallas	32	34	17%
Columbus	29	50	15%
Nashville	26	19	22%
Detroit	26	49	15%
Birmingham	22	59	13%
Indianapolis	19	68	10%

Πίνακας 12: Κατάταξη μητροπολιτικών περιοχών ανάλογα με το επίπεδο συμφόρησης

4.3.3 Αποτύπωση τροχαίων ατυχημάτων

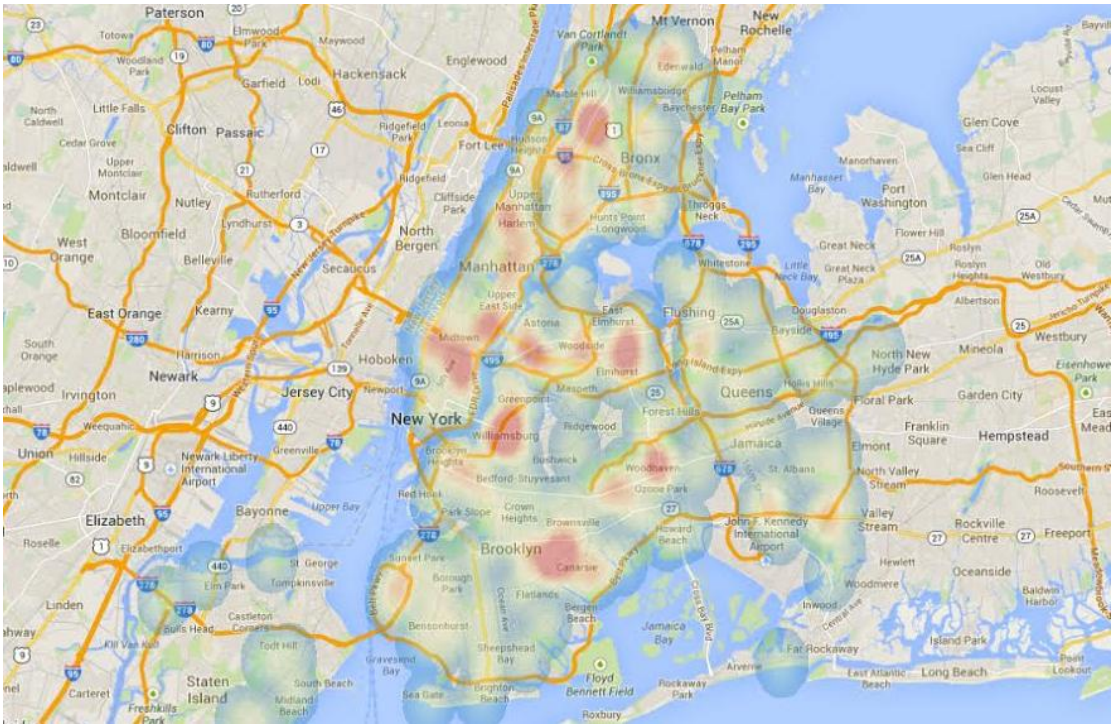
Η κυκλοφοριακή συμφόρηση μπορεί να προκληθεί είτε από αύξηση της ζήτησης της οδού από τα οχήματα είτε από μείωση της χωρητικότητας της οδού όπως η περίπτωση της εκτέλεσης έργων. Στην δεύτερη περίπτωση εντάσσονται και οι περιπτώσεις ύπαρξης τροχαίου ατυχήματος οπότε η οδός κλείνει μερικώς ή ολικώς σε κάποιο σημείο και δυσχεραίνεται η κίνηση των οχημάτων. Μετά από την ανάκτηση των μηνυμάτων και την επεξεργασία με αποτέλεσμα να προκύψει η βάση δεδομένων που χρησιμοποιήθηκε για τις διάφορες αναλύσεις, παρατηρήθηκε ότι ένα μεγάλο μέρος από τα μηνύματα αναφέρονταν σε κάποιο τροχαίο ατύχημα. Πολλά από τα μηνύματα αυτά προέρχονταν από τους επαγγελματικούς λογαριασμούς για τους οποίους έγινε αναφορά παραπάνω και είχαν την μορφή: «***Accident, two lanes blocked in #Portland on I5 SB near Broadway, stop and go traffic***» όπου το μήνυμα ειδοποιεί τους οδηγούς ότι στον δρόμο που εμφανίζεται η τοποθεσία του tweet έχει προκληθεί τροχαίο ατύχημα και υπάρχει μποτιλιάρισμα. Επίσης υπήρχαν πολλά μηνύματα από απλούς χρήστες οι οποίοι θέλοντας να διευκρινίσουν στους υπόλοιπους οδηγούς την κατάσταση που επικρατεί στην οδό που βρίσκονται, ανέφεραν τον λόγο που υπάρχει η συμφόρηση, όπως εκτέλεση έργων ή επικράτηση δυσμενών καιρικών συνθηκών, με την πλειοψηφία των μηνυμάτων αυτών να αναφέρονται στην ύπαρξη τροχαίου ατυχήματος. Καταμετρώντας τα μηνύματα που περιείχαν την λέξη «accident», βρέθηκε ότι αποτελούσαν περίπου το 30% της αρχικής βάσης δεδομένων δηλαδή περίπου 45000 σημεία σε όλη την έκταση των ΗΠΑ. Το ποσοστό είναι αρκετά μεγάλο ώστε να γίνει ο ισχυρισμός ότι τα τροχαία ατυχήματα είναι από τις κύριες αιτίες εμφάνισης έκτακτης κυκλοφοριακής συμφόρησης δεδομένου ότι και στα υπόλοιπα μηνύματα που ελήφθησαν πιθανότατα ένα μεγάλο ποσοστό να αναφέρονται σε συμφόρηση που προκλήθηκε από ατύχημα αλλά δεν περιγράφεται ρητά στον γραπτό λόγο.

Τα τροχαία ατυχήματα είναι εντός δεκάδας με τις επικρατέστερες αιτίες θνησιμότητας στις ΗΠΑ [49]. Έχοντας στην διάθεση της εργασίας τα δεδομένα που αναφέρθηκαν παραπάνω είναι δυνατόν να εξαχθούν οι περιοχές στις οποίες εμφανίζονται τα περισσότερα τροχαία ατυχήματα. Η ανάλυση αυτή είναι πολλή χρήσιμη καθώς με την ανέξοδη και γρήγορη διαδικασία που περιγράφει η εργασία μπορούν να εξαχθούν χρήσιμα συμπεράσματα για το ποιες περιοχές έχουν το μεγαλύτερο πρόβλημα εμφάνισης ατυχημάτων. Τα συμπεράσματα της χωρικής ανάλυσης των ατυχημάτων μπορούν να αποτελέσουν χρήσιμα στοιχεία για τις αρχές για την αντιμετώπιση του φαινομένου καθώς έχοντας τις περιοχές με το μεγαλύτερο πρόβλημα μπορούν να προχωρήσουν σε δράσεις όπως βελτίωση του οδικού δικτύου, τοποθέτηση φωτεινών σηματοδοτών ή μείωση των ορίων ταχύτητας.

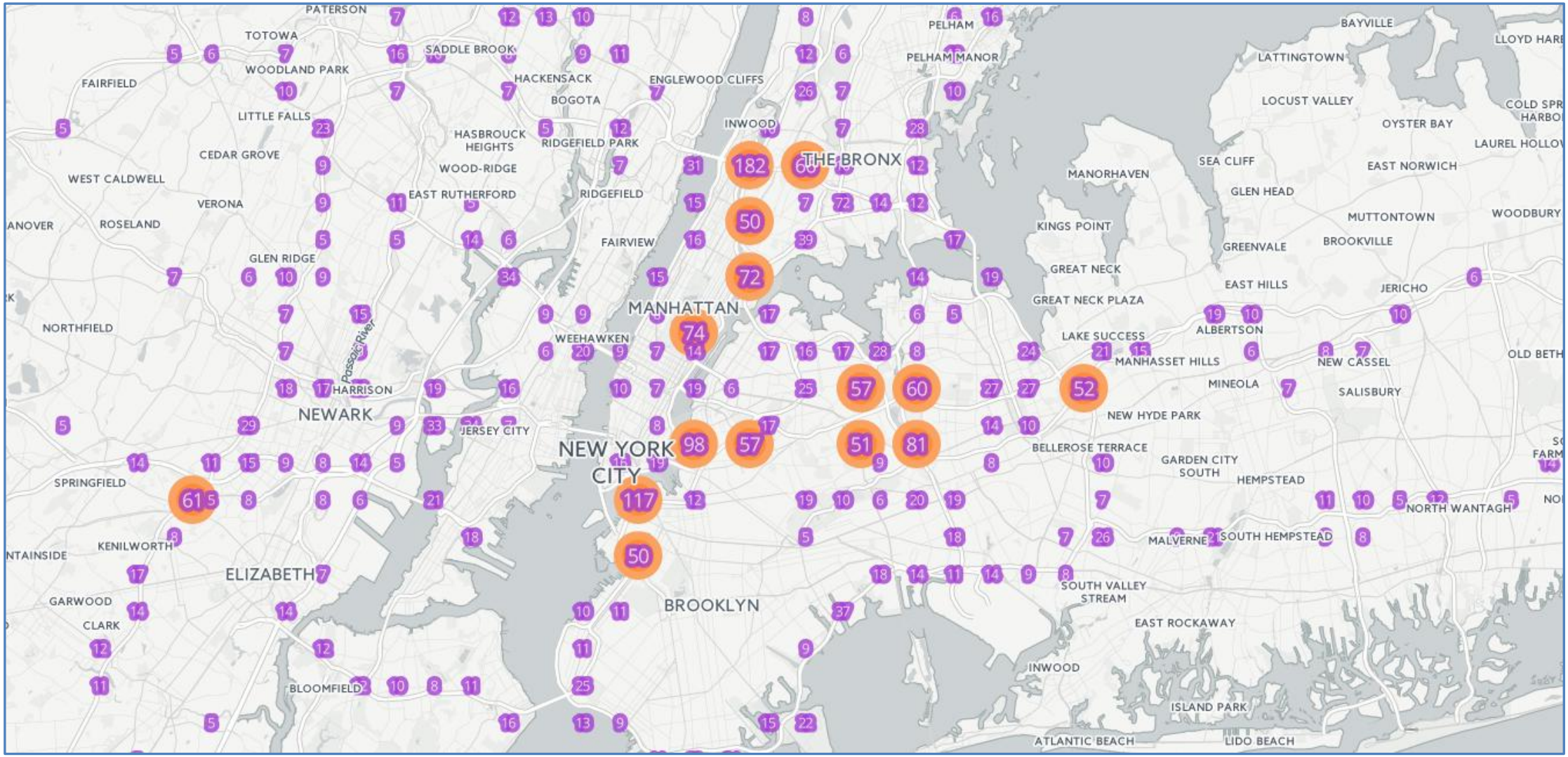
Η ανάδειξη των περιοχών με τα περισσότερα ατυχήματα επιλέχθηκε να υλοποιηθεί σε δύο μητροπολιτικές περιοχές, στην περιοχή του Dallas και Fort-Worth και στην περιοχή της Νέας Υόρκης και Newark. Για κάθε μια από τις παραπάνω μητροπολιτικές περιοχές παρατίθεται παρακάτω ένας χάρτης όπου απεικονίζεται το πλήθος των ατυχημάτων που εμφανίστηκαν στην κάθε υποπεριοχή. Ο χάρτης της

Νέας Υόρκης δημιουργήθηκε σε μεγαλύτερη κλίμακα και με αυτό τον τρόπο είναι εφικτό να εντοπιστούν οι γειτονιές όπου εμφανίζονται τα περισσότερα ατυχήματα ενώ ο χάρτης του Dallas αναδεικνύει τις ευρύτερες περιοχές όπου λαμβάνουν χώρα τα περισσότερα ατυχήματα. Όπως είναι φανερό , στην περίπτωση της Νέας Υόρκης, τα περισσότερα ατυχήματα εμφανίζονται στο κέντρο της πόλης κυρίως στο νησί του Manhattan και στην περιοχή του Brooklyn. Όπως είναι επίσης φανερό στην περίπτωση των Dallas-Fort Worth στο κέντρο των πόλεων εμφανίζονται πολλά ατυχήματα αλλά εξίσου πολλά και στις γύρο μεγάλες οδικές αρτηρίες.

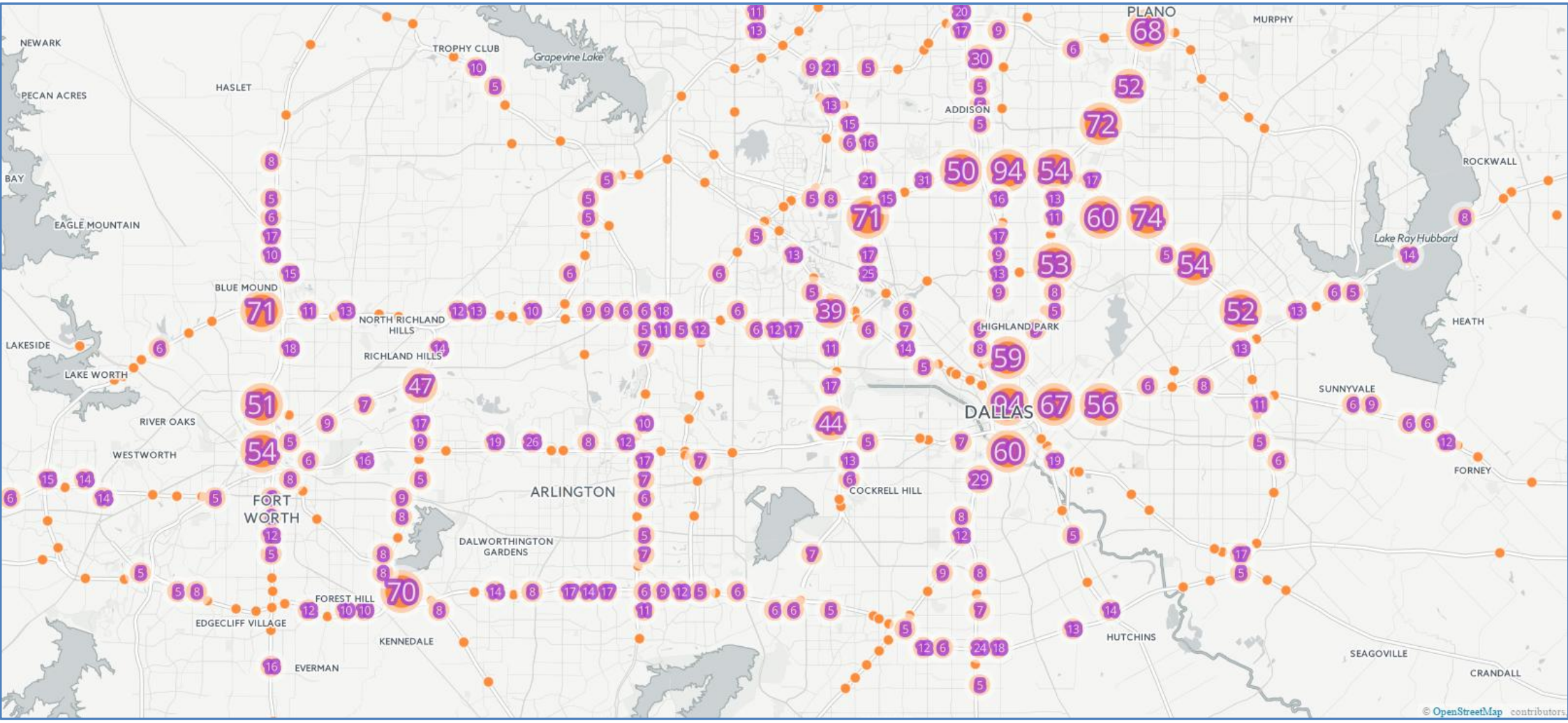
Για την αξιολόγηση των αποτελεσμάτων εμφανίζεται παρακάτω ο χάρτης της Νέας Υόρκης (Εικόνα 13) που παρουσιάζει τις περιοχές όπου έγιναν τα περισσότερα θανάσιμα ατυχήματα το 2013 σύμφωνα με την αστυνομική αρχή, αναφέροντας ότι το 25% των θανάσιμων ατυχημάτων λαμβάνει χώρα μόνο στο 5% της έκτασης της πόλης [50]. Από την σύγκριση του παρακάτω χάρτη και του αντίστοιχου χάρτη που προέκυψε από την ανάλυση διαπιστώνεται ότι έχουν αναδειχθεί επιτυχώς οι περιοχές που έχουν το μεγαλύτερο πρόβλημα ατυχημάτων, εκτός από την περιοχή του Ανατολικού Brooklyn.



Εικόνα 13: Αποτύπωση τροχαίων ατυχημάτων, Νέα Υόρκη
Πηγή data.cityofnewyork.us/



Χάρτης 24: Απεικόνιση τροχαίων ατυχημάτων, New York-Newark

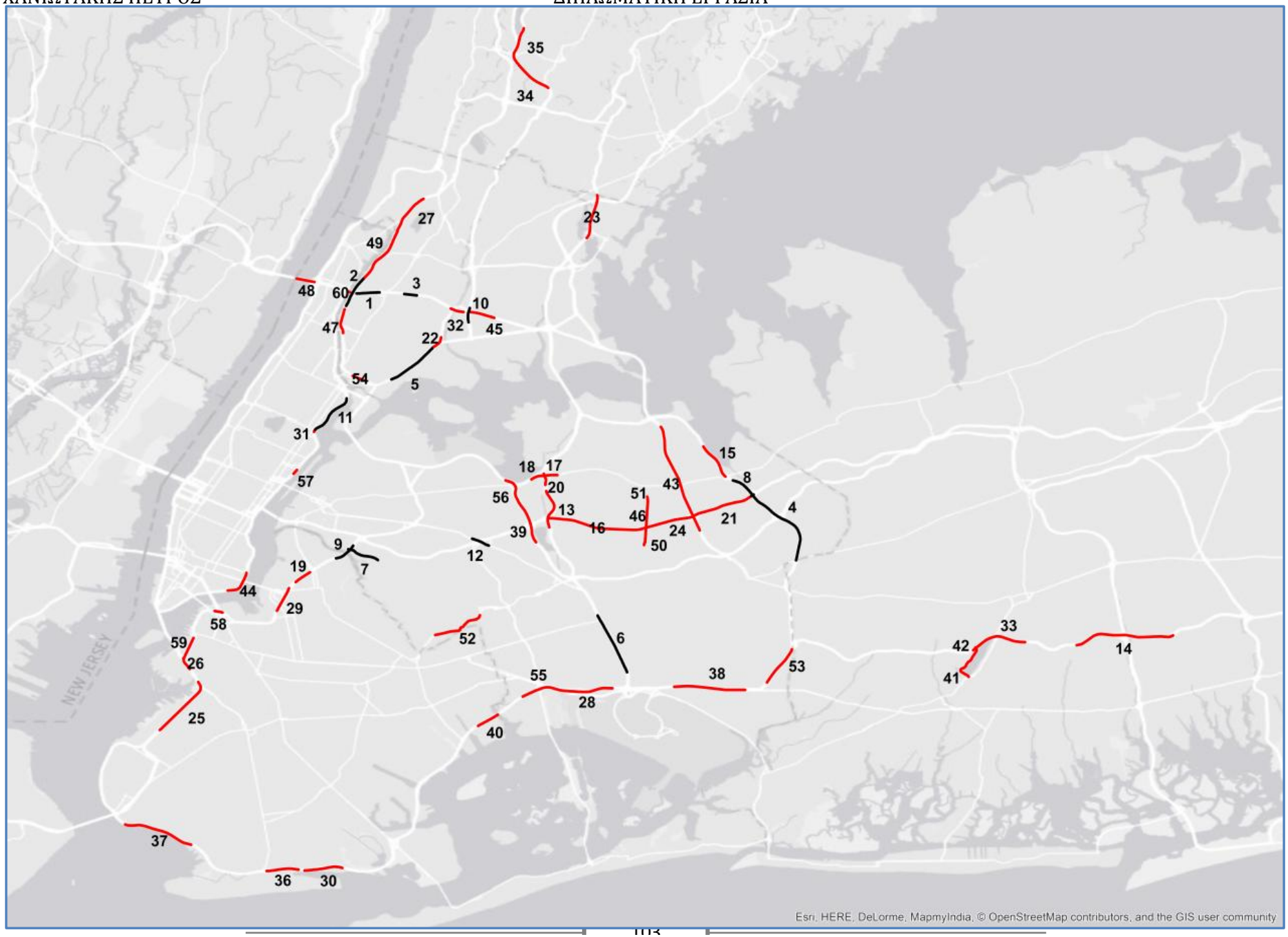


Χάρτης 25: Απεικόνιση τροχαίων ατυχημάτων Dallas-Fort Worth

Εκτός από την ανάλυση για την ανάδειξη των περιοχών με τα περισσότερα ατυχήματα που αντιμετωπίζει το πρόβλημα με μια μακροσκοπική σκοπιά, έγινε προσπάθεια ανάλογης ανάλυσης σε επίπεδο οδών. Σε αυτή την ανάλυση έγινε συσχετισμός των οδών της Νέας Υόρκης με τα σημεία που αποτελούν τις θέσεις των μηνυμάτων στα οποία αναφέρεται η λέξη «accident». Το αποτέλεσμα είναι ένας χάρτης όπου εμφανίζονται τα τμήματα δρόμων για το οποία υπάρχουν περισσότερες μαρτυρίες ότι συνέβη κάποιο τροχαίο δυστύχημα. Η προσέγγιση αυτή αποτελεί μια ακριβέστερη και αποτελεσματικότερη αντιμετώπιση του προβλήματος και τα συμπεράσματά της μπορούν να χρησιμοποιηθούν, όπως και πριν, για τον περιορισμό του φαινομένου, επιλέγοντας τις κατάλληλες παρεμβάσεις.

Κατά την επεξεργασία, τα σημεία που αναφέρονται σε ατύχημα αντιστοιχήθηκαν με τις γραμμές που ορίζουν τις οδούς με παρόμοιο τρόπο όπως και στην ανάλυση για την ανάδειξη των δρόμων με την περισσότερη συμφόρηση με την διαφορά ότι στην συγκεκριμένη περίπτωση δεν έγινε αναγωγή στο μήκος των τμημάτων των οδών. Για να δοθούν σωστά αποτελέσματα κατά την ταξινόμηση όλα τα τμήματα των οδών στα οποία βρέθηκαν σημεία με ατυχήματα χωρίστηκαν σε επιμέρους ώστε κανένα τμήμα να μην υπερβαίνει τα 4 χιλιόμετρα. Επίσης κάθε γραμμή η οποία ήταν κάτω από 100 μέτρα ενοποιήθηκε με τις γειτονικές της . Μετά την επεξεργασία στο δίκτυο υπήρχαν γραμμές που είχαν μήκος από 200 μέτρα έως 4 χιλιόμετρα και με τον τρόπο αυτό περιορίστηκε το φαινόμενο ταξινόμησης μιας οδού ως «πολύ επικίνδυνη» επειδή βρέθηκαν κατά μήκος της πολλά σημεία λόγω της μεγάλης έκτασής της. Η πρακτική αυτή, της μη αναγωγής στο μήκος, ακολουθήθηκε καθώς θεωρήθηκε σκόπιμο η κατάταξη να γίνει σύμφωνα με τον απόλυτο αριθμό περιστατικών για να είναι περισσότερο κατανοητά τα αποτελέσματα.

Στον παρακάτω χάρτη (Χάρτης 26) απεικονίζονται τα 60 τμήματα των δρόμων της πόλης της Νέας Υόρκης στους οποίους έλαβαν χώρα τα περισσότερα ατυχήματα. Με μαύρο χρώμα απεικονίζονται οι οδοί στους οποίους έγιναν 31-71 ατυχήματα ενώ με κόκκινο οι οδοί στους οποίους ανιχνευτήκαν 12-30 ατυχήματα. Δίπλα από κάθε γραμμή εμφανίζεται ένας αύξων αριθμός ο οποίος αναφέρεται στην στήλη « Κατάταξη» του πίνακα που ακολουθεί (Πίνακας 13). Στον πίνακα παρουσιάζονται αναλυτικά τα αποτελέσματα της ανάλυσης με τα ονόματα των οδών και το πλήθος των ατυχημάτων που συνέβησαν σε αυτούς.



Χάρτης 26: Αποτύπωση οδών στους οποίους εμφανίζονται τα περισσότερα ατυχήματα, Νέα Υόρκη

Κατάταξη	Όνομα οδού	Πλήθος ατυχημάτων
1	Cross Bronx Expressway	71
2	Major Deegan Expressway	53
3	Cross Bronx Expressway	43
4	Cross Island Parkway	39
5	Bruckner Expressway	38
6	Van Wyck Expressway	37
7	56th Road	34
8	Long Island Expressway	34
9	Brooklyn-Queens Expressway	34
10	Bronx River Parkway	33
11	FDR Drive	31
12	Queens Boulevard	31
13	Van Wyck Expressway	30
14	Southern State Parkway	30
15	Cross Island Parkway	30
16	Long Island Expressway	29
17	Northern Boulevard	28
18	Northern Boulevard	28
19	Brooklyn-Queens Expressway	28
20	Van Wyck Expressway	28
21	Long Island Expressway	27
22	Sheridan Expressway	27
23	Hutchinson River Parkway	27
24	Long Island Expressway	27
25	Gowanus Expressway	27
26	Brooklyn-Queens Expressway	26
27	Major Deegan Expressway	26
28	Belt Parkway	26
29	Brooklyn-Queens Expressway	24
30	Shore Parkway	24
31	FDR Drive	23
32	Cross Bronx Expressway	23
33	Southern State Parkway	23
34	Sprain Brook Parkway	23
35	Sprain Brook Parkway	23
36	Belt Parkway	22
37	Belt Parkway	22
38	Belt Parkway	22
39	Grand Central Parkway	22
40	Belt Parkway	22
41	Lakeside Drive	21
42	Lakeside Drive	21
43	Clearview Expressway	21
44	FDR Drive	19
45	Cross Bronx Expressway	19

46	Long Island Expressway	19
47	Major Deegan Expressway	19
48	George Washington Bridge	19
49	Major Deegan Expressway	19
50	Utopia Parkway	19
51	Utopia Parkway	19
52	Jackie Robinson Parkway	18
53	Belt Parkway	18
54	Major Deegan Expressway	18
55	Belt Parkway	18
56	Grand Central Parkway	18
57	FDR Drive	15
58	Brooklyn-Queens Expressway	13
59	Hicks Street	12
60	Cross Bronx Expressway	12

Πίνακας 13: Οδοί με τα περισσότερα ατυχήματα στην Νέα Υόρκη

ΚΕΦΑΛΑΙΟ 5: ΣΥΝΟΨΗ

5.1 Συμπεράσματα

Στη παρούσα εργασία αναλύθηκαν οι μέθοδοι οι οποίες μπορούν να ακολουθηθούν με σκοπό την άντληση πληροφοριών από τα κοινωνικά δίκτυα σχετικών με την κυκλοφοριακή συμφόρηση και την επεξεργασία τους σε πραγματικό χρόνο και σε περιβάλλον εκτός σύνδεσης. Ο αρχικός στόχος της εργασίας, δηλαδή την ανάδειξη των κοινωνικών δικτύων ως μια ανεξάντλητη πηγή αξιόπιστων χωρικών πληροφοριών, χρήσιμων για την διαχείριση του κυκλοφοριακού προβλήματος, πιστεύεται ότι επιτεύχθηκε. Στην επιτυχία αυτή διαδραμάτισε σημαντικό ρόλο η σύγκριση των αποτελεσμάτων των διαφόρων αναλύσεων με επίσημα δεδομένα καθώς αναδείχθηκε η αξιοπιστία τους και κατά επέκταση η καταλληλότητα των πληροφοριών από κοινωνικά δίκτυα για χρήση τους σε συστήματα διαχείρισης της κυκλοφορίας.

Βασικό σημείο της εργασίας αποτέλεσε η διαδικτυακή εφαρμογή που υλοποιήθηκε για την ανάκτηση και οπτικοποίηση των μηνυμάτων από το Twitter. Όπως έχει αναφερθεί η συγκεκριμένη εφαρμογή μπορεί να αποτελέσει πηγή ενημέρωσης για τους οδηγούς καθώς μέσω αυτής μπορούν σε πραγματικό χρόνο να εντοπίζουν τα περιστατικά συμφόρησης και να προγραμματίζουν ανάλογα την διαδρομή που θα ακολουθήσουν. Μετά από πειραματισμό έγινε σαφές ότι η συγκεκριμένη εφαρμογή μπορεί να αναδείξει περιστατικά συμφόρησης τα οποία δεν είναι ανιχνεύσιμα εγκαίρως από άλλες εφαρμογές, όπως το Google Traffic. Επίσης σε πολλές περιοχές οι διάφορες υπηρεσίες τύπου Google Traffic δεν λειτουργούν λόγο έλλειψης δεδομένων εν αντιθέσει με την εφαρμογή που παρουσιάστηκε η οποία μπορεί να ανακτήσει πληροφορίες σε παγκόσμιο επίπεδο. Φυσικά για να καταστεί περισσότερο

αποτελεσματική η εφαρμογή χρειάζεται μεγαλύτερος όγκος δεδομένων, για τον λόγο αυτό στην επόμενη ενότητα παρουσιάζονται τρόποι εξόρυξης περισσότερων πληροφοριών.

Οι χρονικές και χωρικές αναλύσεις των μηνυμάτων που παρουσιάστηκαν πιστεύεται ότι ανέδειξαν επιτυχώς τις μεθοδολογίες σύμφωνα με τις οποίες είναι δυνατό να γίνει διαχείριση και επεξεργασία του συγκεκριμένου τύπου δεδομένων. Τα παράγωγα των διαφόρων αναλύσεων αποτελούν χρήσιμες πληροφορίες για την αποκρυπτογράφηση της ροής των οχημάτων στις διάφορες πόλεις και μπορούν να ενταχθούν θεωρητικά ως δεδομένα σε μια μελέτη για την βελτιστοποίηση του οδικού δικτύου ή σε ένα σχέδιο καλύτερης διαχείρισης της κυκλοφορίας. Είναι προφανές ότι τα κοινωνικά δίκτυα εκ φύσεως δεν μπορούν να παράξουν δεδομένα τα οποία είναι σε θέση να ανταποκριθούν στην ακρίβεια των κυκλοφοριακών δεδομένων που παράγονται από άλλες μεθόδους (συμβατικοί ανιχνευτές, συστήματα τύπου Google Traffic) αλλά μπορούν να χρησιμοποιηθούν επικουρικά για την εξαγωγή του βέλτιστου αποτελέσματος λόγο της μεγάλης κάλυψης που παρέχουν, του μικρού κόστους λήψης τους και της φύσης πραγματικού χρόνου που έχουν.

5.2 Λογισμικά που χρησιμοποιήθηκαν

Κατά την διάρκεια της εργασίας χρησιμοποιήθηκαν διάφορα λογισμικά-εργαλεία για την περαίωση των επεξεργασιών που παρουσιάστηκαν. Αρχικά για την ευκολότερη διαχείριση και επικοινωνία με το Twitter streaming API χρησιμοποιήθηκε το λογισμικό ανοικτού τύπου Phirehose [39] το οποίο έχει υλοποιηθεί σε PHP. Η εφαρμογή που δημιουργήθηκε είναι διαθέσιμη στον παρακάτω σύνδεσμο http://83.212.124.233/petros/client.php . Για την χωρική επεξεργασία χρησιμοποιήθηκε το Arcgis της Esri ενώ για την οπτικοποίηση των τροχαίων ατυχημάτων και της υπογραφής της κίνησης το διαδικτυακό εργαλείο https://carto.com/. Στη σελίδα https://petethegreek.carto.com/builder/7f0367f2-5996-11e6-bfa6-0e8c56e2ffdb/embed είναι διαθέσιμα όλα τα δεδομένα που ανακτήθηκαν από όλες τις πόλεις των ΗΠΑ τα οποία λόγο του όγκου τους δεν μπορούσαν να συμπεριληφθούν στο σύνολό τους στην παρούσα εργασία. Η εκκαθάριση των βάσεων δεδομένων και οι χρονικές αναλύσεις πραγματοποιήθηκαν σε περιβάλλον λογιστικών φύλων και συγκεκριμένα στο Microsoft Excel.

5.3 Προτάσεις για περαιτέρω μελέτη

Ένα από τα βασικά προβλήματα στην παρούσα εργασία ήταν η έλλειψη μεγάλου όγκου δεδομένων. Το συγκεκριμένο θέμα μπορεί να αντιμετωπιστεί, όπως έχει αναφερθεί ήδη, με την λήψη όλων των μηνυμάτων που κοινοποιούνται στο Twitter μέσω εταιριών που είναι υπεύθυνες για την διαχείριση τους. Οι μελλοντικοί ερευνητές έχουν την δυνατότητα να ανακτήσουν όλα τα δεδομένα από τις συγκεκριμένες πηγές αν είναι σε θέση να καταβάλουν κάποιο χρηματικό αντίτιμο. Η αύξηση του όγκου των δεδομένων μπορεί να επιτευχθεί και με την χρησιμοποίηση άλλων κοινωνικών δικτύων. Με τον τρόπο αυτό, μπορεί να υπάρξει μια βάση στην οποία θα αποθηκεύονται ταυτόχρονα δεδομένα από Twitter, Facebook, Instagram, Weibo κ.α . Τα παραπάνω κοινωνικά δίκτυα έχουν όπως και το Twitter πολλά εκατομμύρια χρήστες άρα γίνεται αντιληπτό ότι ο συνδυασμός τους θα εκτινάξει τον όγκο των δεδομένων που θα λαμβάνονται. Μια άλλη μέθοδος για αύξηση του όγκου των πληροφοριών είναι η χρήση λέξεων κλειδιών που αναφέρονται σε κυκλοφοριακή συμφόρηση σε διαφορετικές γλώσσες από την Αγγλική. Με τον τρόπο αυτό, το πεδίο δράσης της εφαρμογής θα διευρυνθεί πέρα από τα πλαίσια των αγγλόφωνων περιοχών με στόχο την παγκόσμια κάλυψη.

Εάν ο όγκος των πληροφοριών αυξηθεί θα είναι εφικτό να γίνουν εκ των υστέρων χωρικές και χρονικές αναλύσεις των οποίων τα συμπεράσματα θα είναι περισσότερο αξιόπιστα. Επίσης θα μπορούν να πραγματοποιηθούν αναλύσεις οι οποίες δεν ήταν δυνατόν να γίνουν στην παρούσα εργασία. Ένα παράδειγμα είναι η εξαγωγή της ώρας αιχμής της κάθε οδού για την οποία είναι αντιληπτό ότι χρειάζονται πολλές περισσότερες πληροφορίες.

Όσον αφορά την εφαρμογή, πρέπει να βελτιστοποιηθεί σε πολλά σημεία ώστε να αποτελέσει χρήσιμο εργαλείο για την ενημέρωση των οδηγών. Το βασικό είναι να απαλείφονται τα μηνύματα τα οποία φέρουν πληροφορίες που έχουν παρέλθει χρονικά με αυτόματο τρόπο. Η συγκεκριμένη επεξεργασία δεν θα είναι εύκολη καθώς δεν υπάρχουν δεδομένα για το πότε λαμβάνει τέλος ένα περιστατικό συμφόρησης, μπορούν όμως να υπολογιστούν οι μέσοι όροι της χρονικής διάρκειας της συμφόρησης σε κάθε περιοχή από άλλες πηγές και να εισαχθούν στο σύστημα. Τέλος μια καλή ιδέα είναι να εμφανίζονται στο χάρτη, εκτός από το κείμενο του κάθε μηνύματος, και φωτογραφίες που έχουν κοινοποιηθεί από τους χρήστες, αν είναι διαθέσιμες, ώστε ο κάθε οδηγός να κρίνει σε ποιο επίπεδο είναι η συμφόρηση και να λαμβάνει τις ανάλογες αποφάσεις προγραμματισμού της διαδρομής του.

Βιβλιογραφία-Παραπομπές

1. TTI Urban Mobility Report https://mobility.tamu.edu/ums/
2. Μαρία-Ιωάννα Μ. Ιμπριάλου, "ΕΝΤΟΠΙΣΜΟΣ ΔΕΥΤΕΡΟΓΕΝΩΝ ΣΥΜΒΑΝΤΩΝ ΣΤΟ ΕΣΩΤΕΡΙΚΟ ΠΕΡΙΟΧΩΝ ΣΥΜΦΟΡΗΣΗΣ ΕΛΕΥΘΕΡΗΣ ΛΕΩΦΟΡΟΥ ΜΕ ΧΡΗΣΗ ΑΝΑΛΥΤΙΚΩΝ ΜΕΘΟΔΩΝ", 2011
3. Aftabuzzaman Md, Currie Graham, Sarvi Majid "Evaluating the Congestion Relief Impacts of Public Transport in Monetary Terms", 2010
4. http://inrix.com/
5. Mukhopadhyay, Subhas Chandra, Leung, Henry, "Advances in Wireless Sensors and Sensor Networks", 2010
6. http://www.tomtom.com
7. Ετήσια έκθεση της εταιρίας INRIX http://inrix.com/wp-content/uploads/2016/03/INRIX_2015_US_Scorecard_Infographic.pdf, 2015
8. https://twitter.com
9. Microblogging Wikipedia https://en.wikipedia.org/wiki/Microblogging
10. https://www.tumblr.com/
11. www.weibo.com
12. https://identi.ca/
13. https://www.flickr.com/
14. Twitter Help Center https://support.twitter.com/
15. Andreas M. Kaplan*, Michael Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media", 2010
16. Αικατερίνη-Ειρήνη Μελίτα, "Ο ΡΟΛΟΣ ΤΗΣ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ ΣΕ ΣΧΕΣΗ ΜΕ ΤΗΝ ΠΡΟΣΦΑΤΗ ΑΡΑΒΙΚΗ ΑΝΟΙΞΗ", 2012
17. http://downrightnow.com/
18. https://www.fixmystreet.com/
19. MICHAEL F. GOODCHILD, "CITIZENS AS SENSORS: WEB 2.0 AND THE VOLUNTEERING OF GEOGRAPHIC INFORMATION", 2007
20. Andrew Crooks, Arie Croitoru, Anthony Stefanidis, Jacek Radzikowski, "#Earthquake:Twitter as a Distributed Sensor System", 2013
21. Adam Tsakalidis, Symeon Papadopoulos , Alexandra Cristea και Yianni Kompatsiari, "Predicting Elections for Multiple Countries Using Twitter and Polls", 2014
22. Evaggelos Spyrou, Phivos Mylonas, "From Tags to Trends: A First Glance at Social Media Content Dynamics", 2012
23. Ombretta Gaggi, "Discovering Local Attractions from Geo-Tagged Photos", 2013
24. Majid, Abdul ,Chen, Ling, Chen, Gencai , Mirza, Hamid Turab, Hussain, Ibrar ,Woodward, John, "A context-aware personalized travel recommendation system based on geotagged social media data mining", 2013

25. Sri Krisna Endarnoto, Sonny Pradipta, Anto Satriyo Nugroho, James Purnama, "Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application", 2011
26. Raymondus Kosala, Erwin Adi, "Harvesting Real Time Traffic Information from Twitter", 2012
27. Axel Schulz, Petar Ristoski, "The Car that Hit The Burning House:Understanding Small Scale Incident Related Information in Microblogs", 2011
28. Elsa Foury, "Monitoring Urban Traffic Status. Using Twitter Messages", 2013
29. Khatri, Chandra P., "Real-time road traffic information detection through social media", 2015
30. Διεπαφή προγραμματισμού εφαρμογών Wikipedia https://el.wikipedia.org/wiki/Διεπαφή_προγραμματισμού_εφαρμογών
31. Twitter Developer Documentation, https://dev.twitter.com/streaming/overview
32. Εταιρία διαχείρισης δεδομένων Twitter https://gnip.com/
33. Περιορισμοί στην ανάκτηση των μηνυμάτων από το Twitter https://twittercommunity.com/t/is-there-a-limit-to-the-amount-of-data-the-streaming-api-will-send-out/8482
34. www.brittonmdg.com
35. Κέντρο έρευνας για τα κοινωνικά δίκτυα http://www.pewresearch.org/
36. Geotagged tweets https://www.fastcompany.com/3016802/fast-feed/one-in-five-tweets-is-geotagged-with-your-location-this-website-shows-you-what-you
37. https://www.hashtracking.com/
38. https://www.hashtags.com
39. Το λογισμικό που χρησιμοποιήθηκε για την ανάκτηση των μηνυμάτων https://github.com/fennb/phirehose
40. Εργαλείο δημιουργίας βάσης δεδομένων https://www.phpmyadmin.net/
41. Ιστοσελίδα ενημέρωσης για την συμφόρηση μέσω του Twitter https://www.sigalert.com
42. https://www.instagram.com
43. Η υπηρεσία της Google, Google Traffic Wikipedia https://en.wikipedia.org/wiki/Google_Traffic
44. Ακρίβεια GPS στα έξυπνα τηλέφωνα https://communityhealthmaps.nlm.nih.gov/2014/07/07/how-accurate-is-the-gps-on-my-smart-phone-part-2/
45. Γραφείο απογραφής ΗΠΑ, https://www.census.gov/
46. Πηγή αρχείου οδών (shapefile) http://download.geofabrik.de/north-america/us/new-york.html
47. New York Traffic Viewer, http://gis3.dot.ny.gov/html5viewer/?viewer=tdv
48. Χωρική αυτοσυσχέτιση θεωρία, https://repository.kallipos.gr/bitstream/11419/5032/1/02_chapter_3.pdf
49. Στατιστικά στοιχεία για τα τροχαία ατυχήματα στις ΗΠΑ http://www.who.int/mediacentre/factsheets/fs310/en/

50. Επίσημα δεδομένα για τα τροχαία ατυχήματα από τις αρχές της Νέας Υόρκης https://data.cityofnewyork.us/