

# Τεχνικές Εξόρυξης Δεδομένων

## Εργασία 2

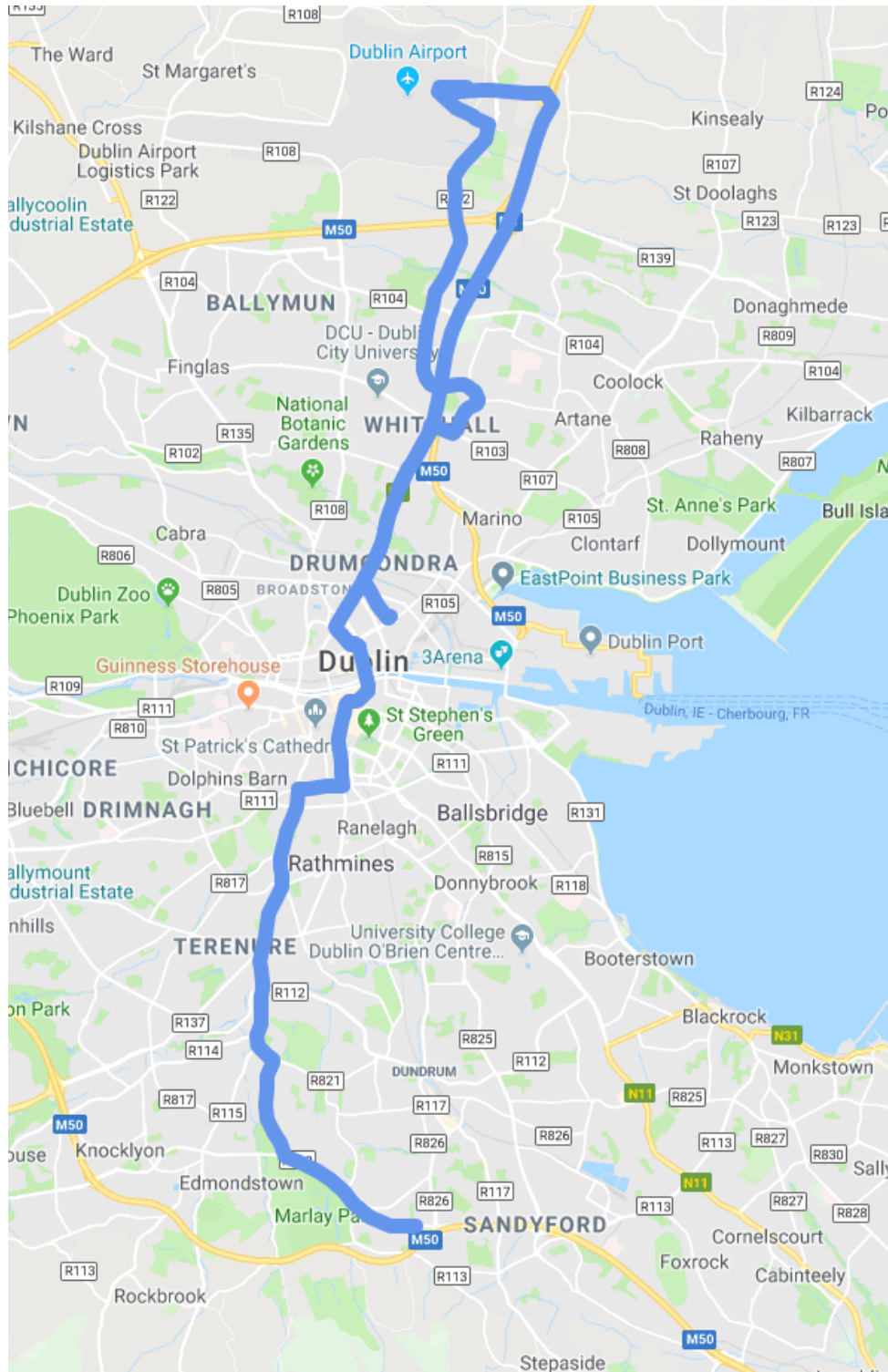
Όνομα	Επώνυμο	ΑΜ
Αναστάσιος	Αντωνόπουλος	1115201400014
Γεώργιος	Τσομπανέλης	1115201400213

### Ερώτημα 1

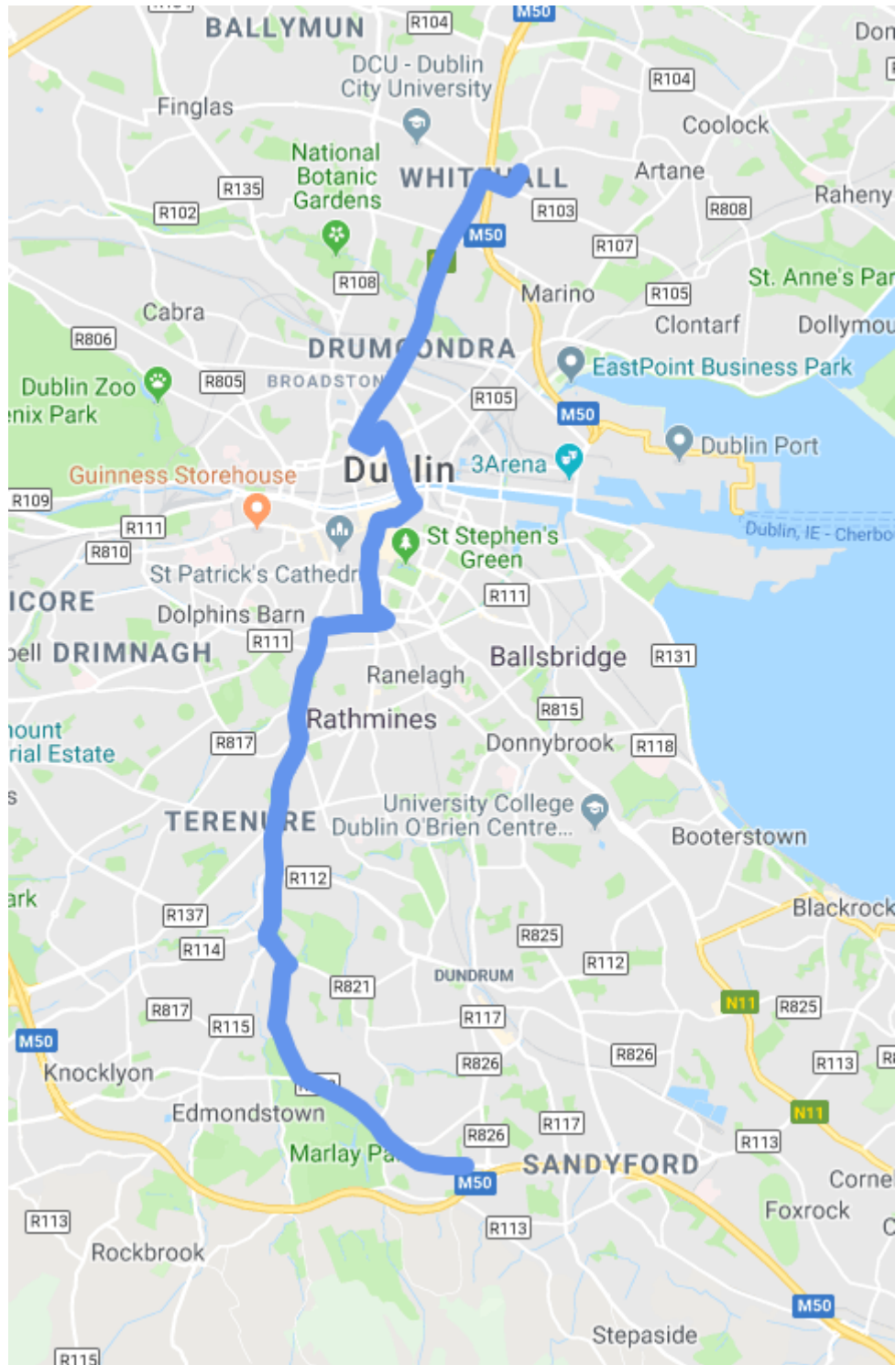
Στο ερώτημα 1 έχουμε επιλέξει τυχαία 5 διαφορετικά Journey\_Pattern\_ID και με βάση αυτά, περνάμε τα σχετικά lats, lons στις αντίστοιχες λίστες τις οποίες παίρνει σαν όρισμα η συνάρτηση gmap.plot. Έτσι δημιουργούνται 5 html(mymap\_1, mymap\_2, mymap\_3, mymap\_4, mymap\_5) αρχεία χάρτη με τις διαφορετικές διαδρομές.

Παρακάτω παρουσιάζονται οι διαδρομές αυτές:

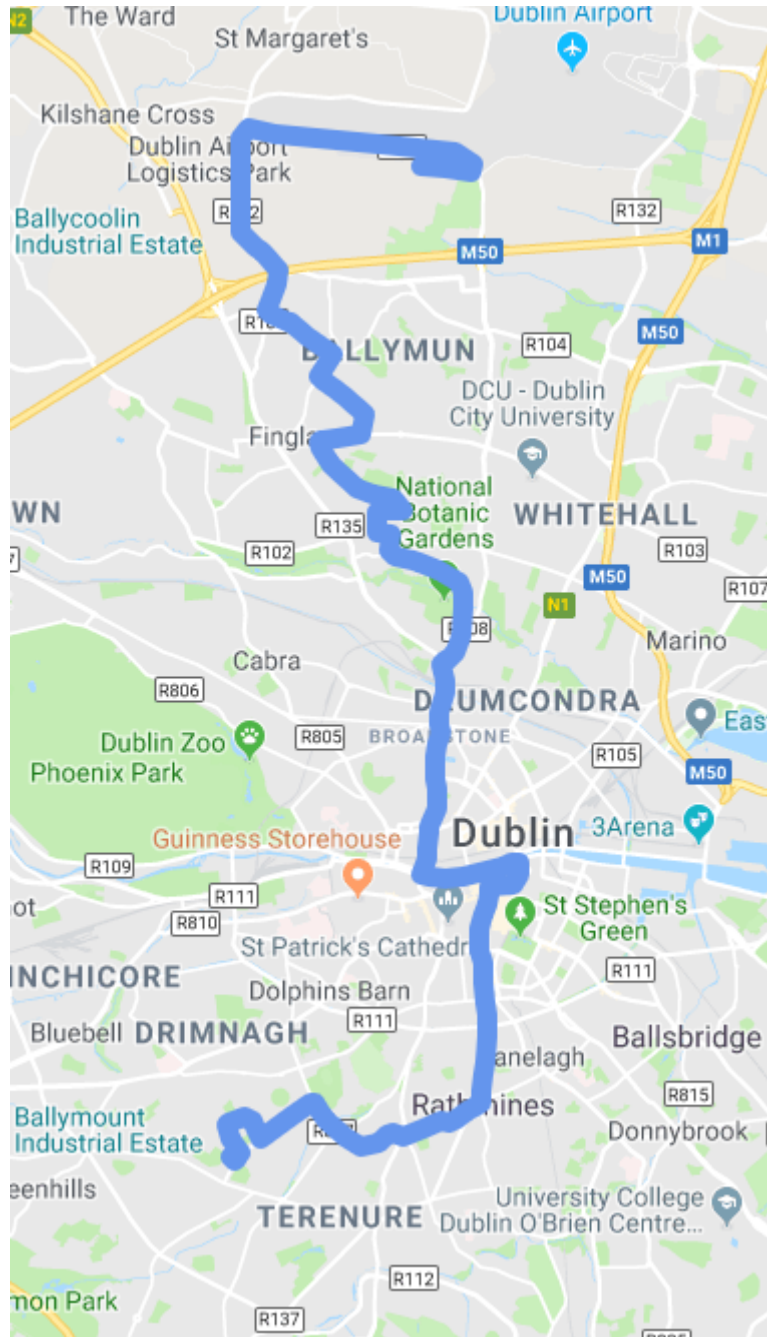
## mymap\_1



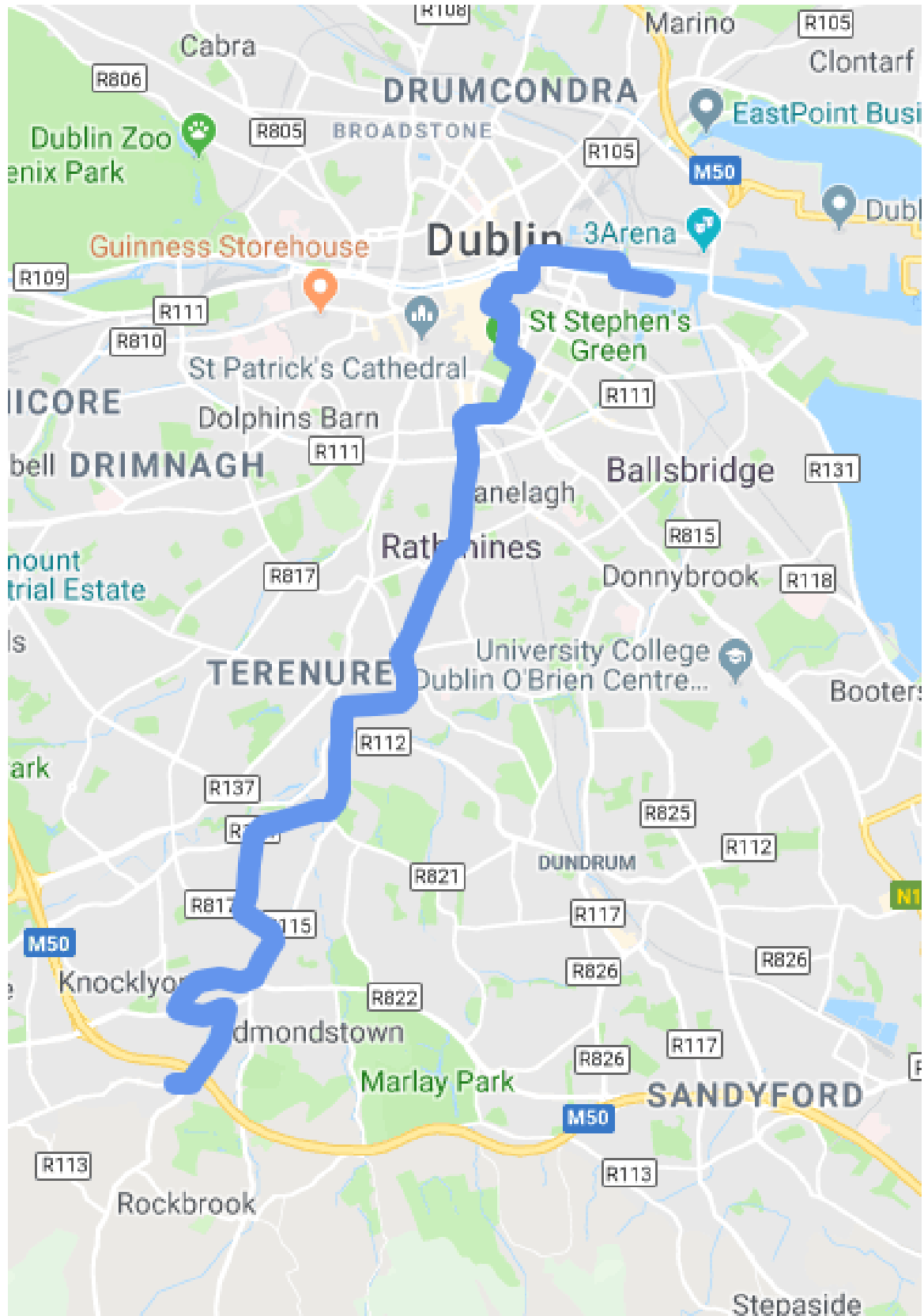
mymap\_2



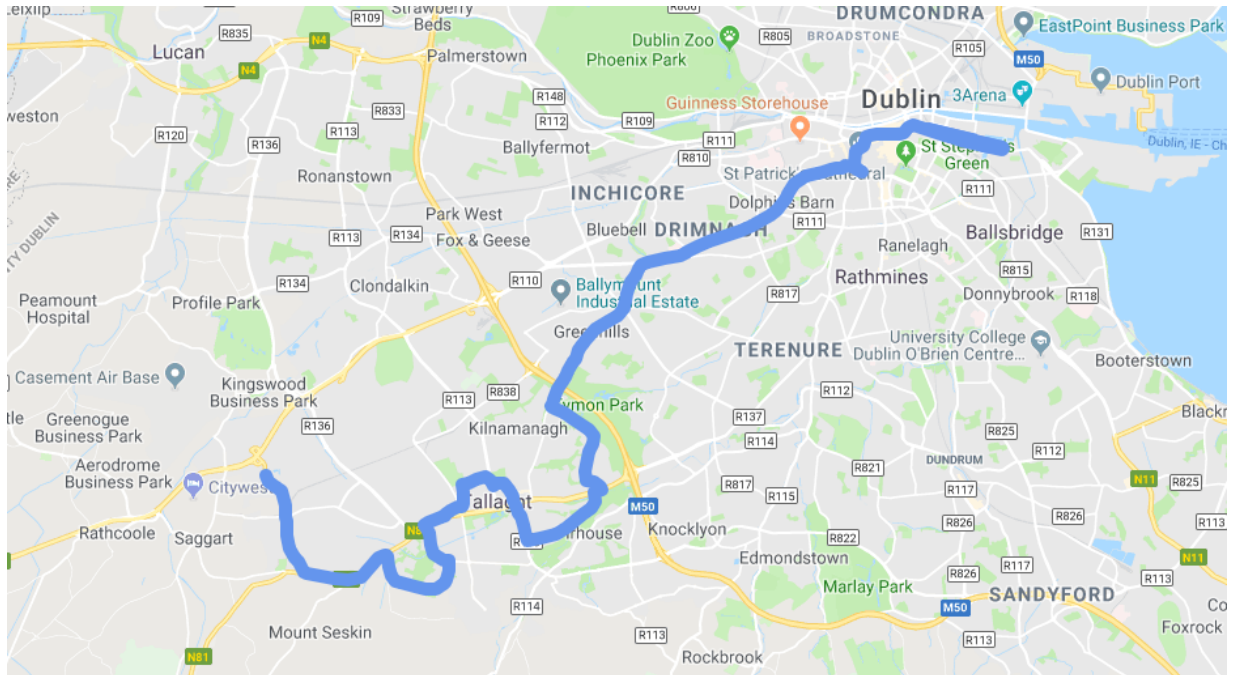
mymap\_3



mymap\_4



mymap\_5



## Ερώτημα 2

A1) Για το ερώτημα 2 A1 ζητείται να βρούμε τους 5 κοντινότερους γείτονες με βάση τη μετρική DTW. Κάνουμε, για αυτό το λόγο, χρήση της συνάρτησης `fastdtw` από τη βιβλιοθήκη `fastdtw`. Επιπλέον, για τον υπολογισμό της απόστασης δύο σημείων χρησιμοποιούμε τη συνάρτηση `haversine` από τη βιβλιοθήκη `haversine`. Παρακάτω παραθέτουμε τα πινακάκια:



## Test\_trip\_1



Test Trip 1  
 $\Delta T = 235$  sec



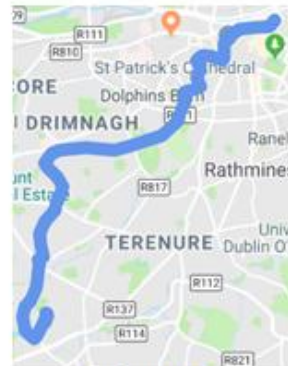
Neighbor 1  
JP\_ID: 01501001  
DTW: 0.0 km



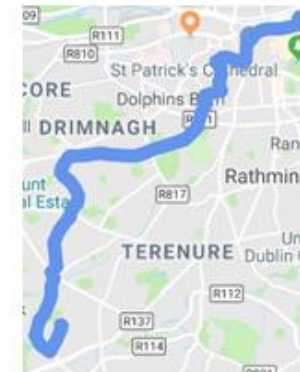
Neighbor 2  
JP\_ID: 01501001  
DTW: 4.75 km



Neighbor 3  
JP\_ID: 01501001  
DTW: 5.06 km



Neighbor 4  
JP\_ID: 01501001  
DTW: 5.28 km



Neighbor 5  
JP\_ID: 01501001  
DTW: 5.41 km

Test\_trip\_2



Test Trip 2  
 $\Delta T = 178$  sec



Neighbor 1  
JP\_ID: 01200001  
DTW: 0.0 km



Neighbor 2  
JP\_ID: 01200001  
DTW: 4.04 km



Neighbor 3  
JP\_ID: 01200001  
DTW: 4.52 km



Neighbor 4  
JP\_ID: 01200001  
DTW: 4.61 km



Neighbor 5  
JP\_ID: 01200001  
DTW: 5.06 km



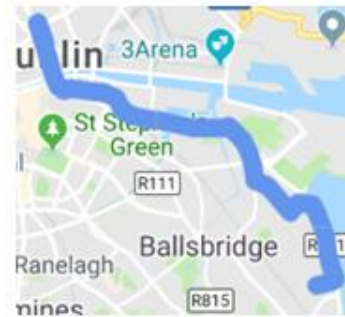
## Test\_trip\_3



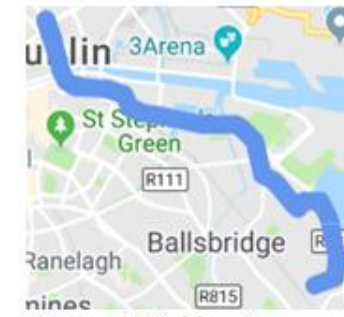
## Test\_trip\_4



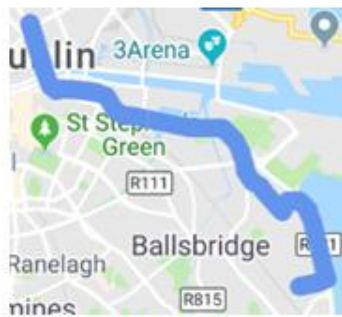
Test Trip 4  
 $\Delta T = 202.55$  sec



Neighbor 1  
JP\_ID: 00010002  
DTW: 0.0 km



Neighbor 2  
JP\_ID: 00010002  
DTW: 3.54 km



Neighbor 3  
JP\_ID: 00010002  
DTW: 3.90 km



Neighbor 4  
JP\_ID: 00010002  
DTW: 4.28 km



Neighbor 5  
JP\_ID: 00010002  
DTW: 4.82 km

## Test\_trip\_5



Test Trip 5  
 $\Delta T = 186.75$  sec



Neighbor 1  
JP\_ID: 01300001  
DTW: 0.0 km



Neighbor 2  
JP\_ID: 01300001  
DTW: 6.24 km



Neighbor 3  
JP\_ID: 01300001  
DTW: 6.49 km



Neighbor 4  
JP\_ID: 01300001  
DTW: 6.58 km



Neighbor 5  
JP\_ID: 01300001  
DTW: 6.61 km

## **Ερώτημα 2**

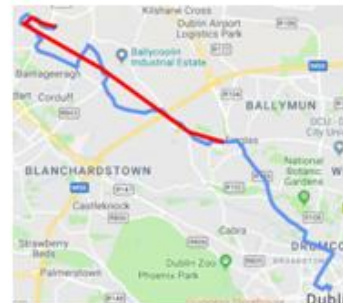
A2) Στο ερώτημα αυτό για την εύρεση των υποδιαδρομών χρησιμοποιούμε τη συνάρτηση lcs την οποία χρησιμοποιήσαμε έτοιμη από το διαδίκτυο. Παρακάτω παρουσιάζονται οι αντίστοιχοι πίνακες για κάθε test trip:



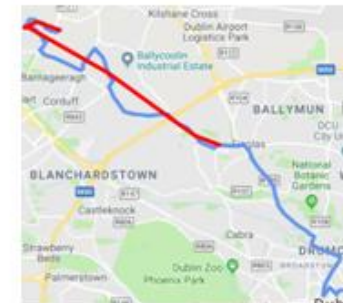
### Test Trip 1



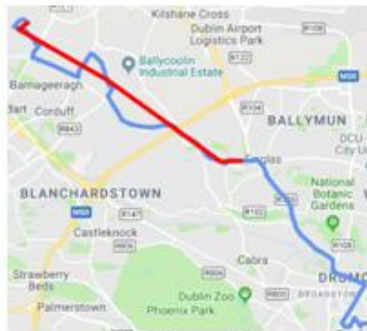
Test Trip 1  
 $\Delta T = 60\text{sec}$



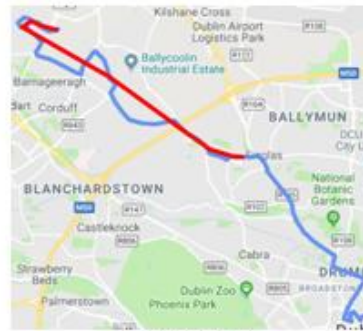
Neighbor 1  
JP\_ID: 040D1002  
#Matching Points: 11



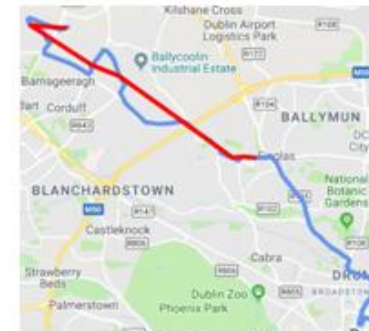
Neighbor 2  
JP\_ID: 040D1002  
#Matching Points: 11



Neighbor 3  
JP\_ID: 040D1002  
#Matching Points: 9



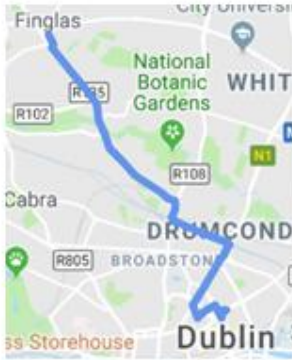
Neighbor 4  
JP\_ID: 040D1002  
#Matching Points: 9



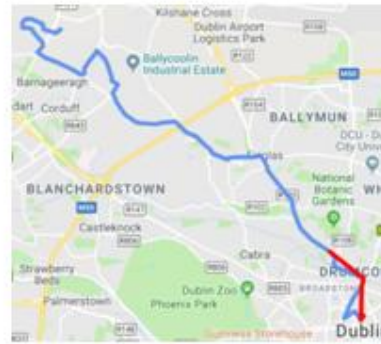
Neighbor 5  
JP\_ID: 040D1002  
#Matching Points: 9



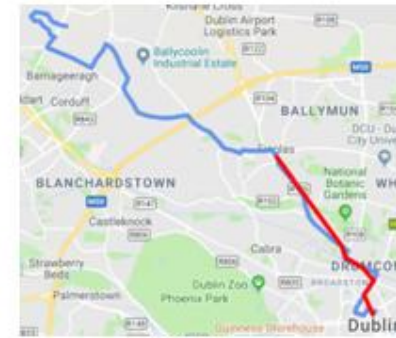
## Test Trip 2



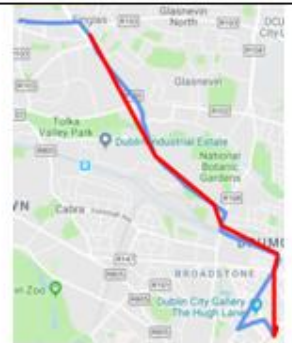
Test Trip 2  
 $\Delta T = 54.4$  sec



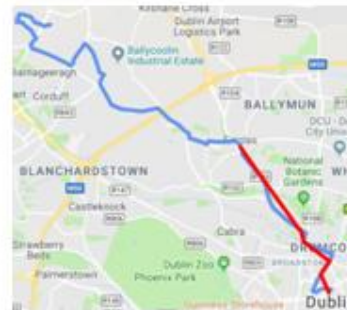
Neighbor 1  
JP\_ID: 040D1002  
#Matching Points: 17



Neighbor 2  
JP\_ID: 040D1002  
#Matching Points: 10



Neighbor 3  
JP\_ID: 040D1002  
#Matching Points: 9



Neighbor 4  
JP\_ID: 040D1002  
#Matching Points: 9



Neighbor 5  
JP\_ID: 040D1002  
#Matching Points: 9

Test Trip 3



Test Trip 3  
 $\Delta T = 29.17$  sec



Neighbor 1  
JP\_ID:00790001  
#Matching Points: 39



Neighbor 2  
JP\_ID:00790001  
#Matching Points: 6



Neighbor 3  
JP\_ID:00790001  
#Matching Points: 6

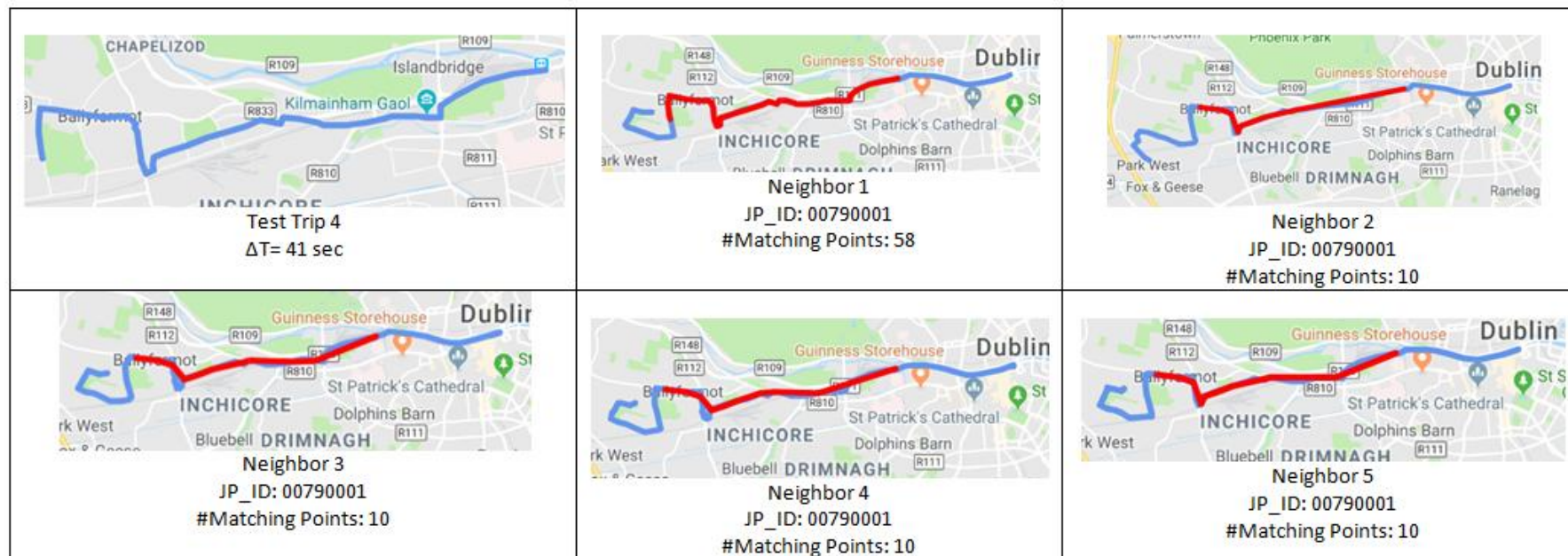


Neighbor 4  
JP\_ID:00790001  
#Matching Points: 6



Neighbor 5  
JP\_ID:00790001  
#Matching Points: 5

# Test Trip 4



# Test Trip 5



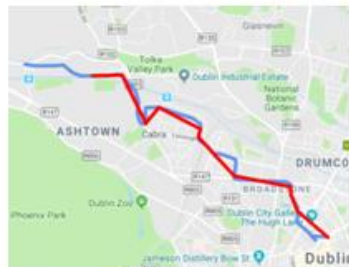
Test Trip 5  
 $\Delta T = 49.8 \text{ sec}$



Neighbor 1  
 JP\_ID: 01200001  
 #Matching Points: 72



Neighbor 2  
 JP\_ID: 01200001  
 #Matching Points: 14



Neighbor 3  
 JP\_ID: 01200001  
 #Matching Points: 13



Neighbor 4  
 JP\_ID: 01200001  
 #Matching Points: 12



Neighbor 5  
 JP\_ID: 01200001  
 #Matching Points: 11



### Ερώτημα 3

Σε αυτό το ερώτημα ζητείται να προβλέψουμε τα journey\_pattern\_ID από πέντε διαφορετικά test trip χρησιμοποιώντας τον knn classifier. Όπως αναφέρθηκε και με διευκρίνιση στο piazza δεν μπορούμε να χρησιμοποιήσουμε τον έτοιμο του sklearn , οπότε χρησιμοποιήσαμε κάποιον έτοιμο από το διαδίκτυο κάνοντάς του κάποιες αλλαγές, ώστε να δουλεύει για τη δική μας μορφή δεδομένων. Παρακάτω παρουσιάζονται οι προβλέψεις:

Test Trip Id	Predicted Journey Pattern Id
0	00400001
1	00070004
2	07470001
3	00791001
4	01201001



