TORONTO
CLUSTERING TO IDENTIFY DISTRIBUTION OF ASIAN RESTAURANTS

- OBJECTIVE :
- The objective of this project is to provide solution to support business needs to use available data to identify an optimum location to open a south Indian cuisine restaurant in Toronto area.

- The datasets referred to in this exercise are:
- - Wikipedia for neighbourhood information on Toronto including latitudes and longitudes.
- - Four square API for details of existing restaurant and other places of interest in Toronto

Due to time constraints the following information that can enhance results was not used.

- o ZOMATO API URL
- o https://developers.zomato.com/api

DATA ACQUISITON

- DATA DESCRIPTION AND ANALYSIS :
- The data used to solve the problem is described below:
- A.) IDENTIFICATION OF TORONTO NEIGHBOURHOODS
- (“https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M”
- This is following earlier exercise where wiki pages are scraped for data and stored in pandas DF.
- IBM team CSV file to identify the longitude and lattitudes of the neighbourhood and appending to data frame.

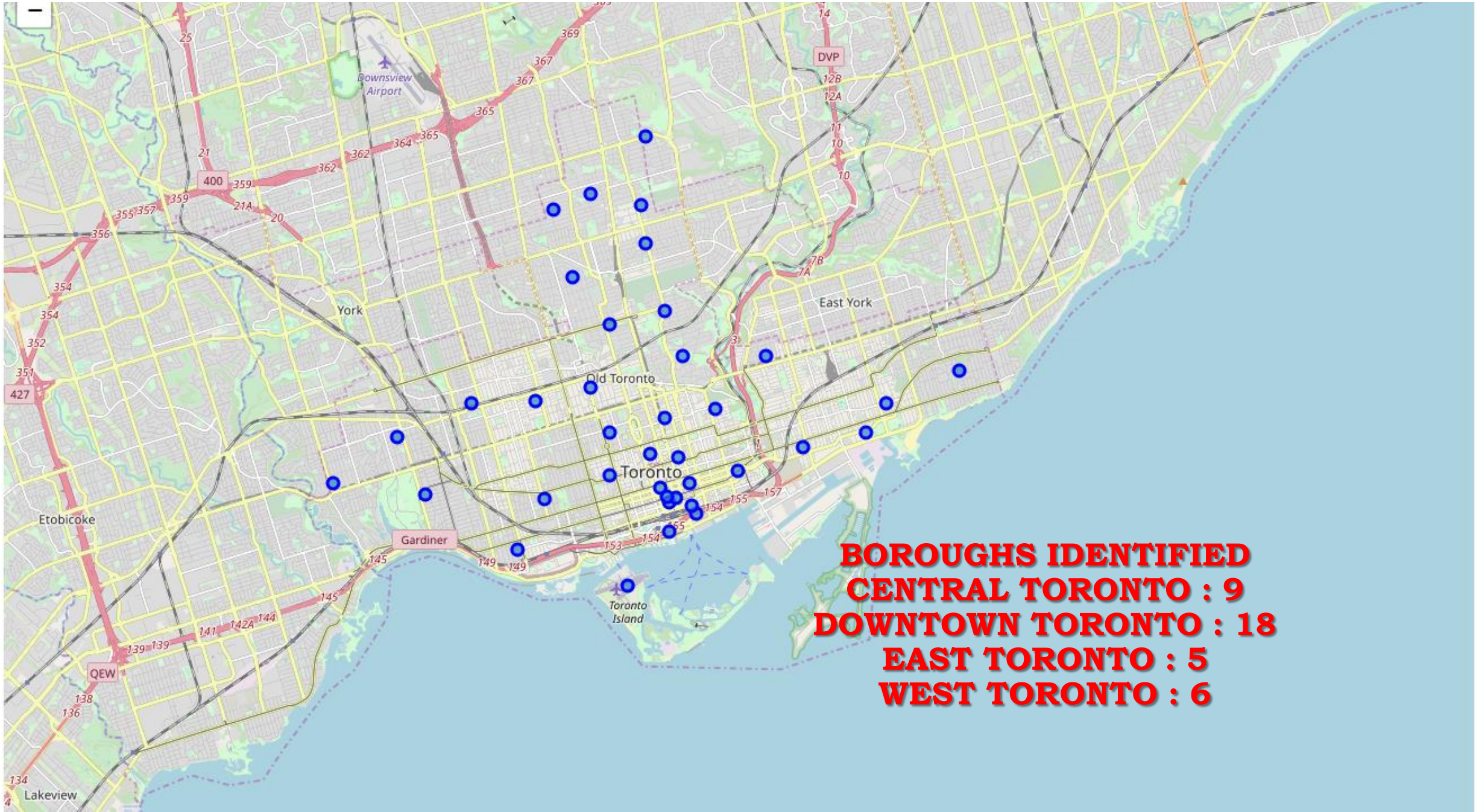
- B.) IDENTIFICATION OF TOP VENUES USING FOURSQUARE API.
- By using foursquare API and using client ID and password, I pull out all details available for the venues and restaurants.

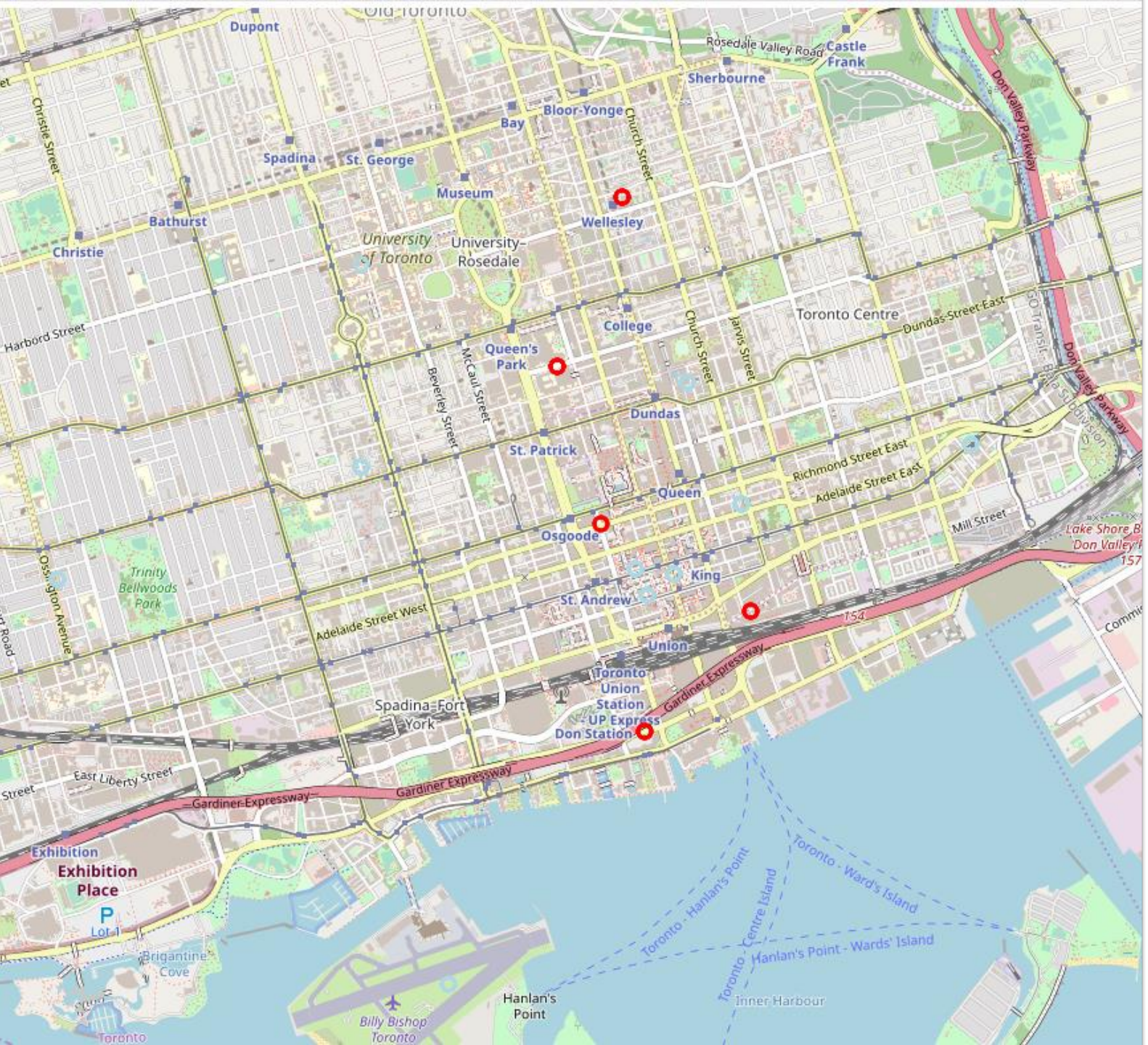
- C.) GROUPING VENUES BY NEIGHBOURHOOD.
- Analyse and group all data by Neighbourhood and clustering.

- D.) IDENTIFICATION OF CUSINE TYPE
- Analyse and filter by cuisine type and map the details by K means clustering algorithm and allocate location to nearest cluster.

RESULTS

- Identification of Boroughs and neighbourhoods in Toronto
- Identification of restaurants using 4 square
- Identification of Asian restaurants and clustering using KMEANS


Downsview Airport
York
East York
Old Toronto
Toronto
Etobicoke
Lakeview
Toronto Island
Gardiner
QEW
400
427
DVP
BOROUGHS IDENTIFIED
CENTRAL TORONTO : 9
DOWNTOWN TORONTO : 18
EAST TORONTO : 5
WEST TORONTO : 6



ASIAN RESTAURANTS

- The result indicates a low count of available Asian restaurants and shows a very high potential for client to further evaluate opening of Asian restaurants in Toronto area.
- There is room for improvement using additional data sets which could include statistics of community and prices of similar cuisine and restaurants in the region.

CONCLUSION

- The data sets used provided an excellent insight into the distribution of restaurants and identification of Asian restaurants.
- The result indicates a low count of available Asian restaurants and shows a very high potential for client to further evaluate opening of Asian restaurants in Toronto area.
- There is room for improvement using additional data sets which could include statistics of community and prices of similar cuisine and restaurants in the region.
- Due to time limit, I was unable to further explore on enhancing the data sets.

- The direction to client is based on the data explored there is huge potential for opening of Asian Restaurant in Toronto and the results can be narrowed down with use of additional datasets and price factors.