Most common types of restaurants in Toronto Area

Background

- Market research is a crucial step to forge the success of any business.

- Many new restaurants fail due to location (serving a cuisine that is the exact same as several other restaurants in the area).

- Location and customer base are two key components to a restaurant's success.

Solution?

- Having location information with regards to which areas have what kind of restaurants (or lack thereof) will give entrepreneurs an edge to make various informed business decisions.
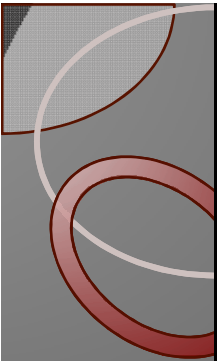
Methodology – Data Sources

- Datasets required to tackle the problem are:

  Toronto's geospatial data *(https://cocl.us/Geospatial_data)*;

  Borough, neighborhood(s), postal code data in the Toronto area *(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)* and

  Restaurants within in each neighborhood (Foursquare API).

Methodology – Data Cleaning

- Web scraping – BeautifulSoup package.

- Remove entries with Borough values 'Not assigned'.

- Replace Neighborhood values 'Not assigned' with respective Borough value.

- Merge entries with same postal code and collate Neighborhood values.

- Merge dataframe with geospatial dataframe.

Methodology – Cluster Analysis

- Use K-means cluster algorithm with target value k of 5.

- Will segment Toronto city data into 5 clusters.

- Use Foursquare API to gather information on restaurant cuisines within a 500m radius from each neighbourhood within a cluster.
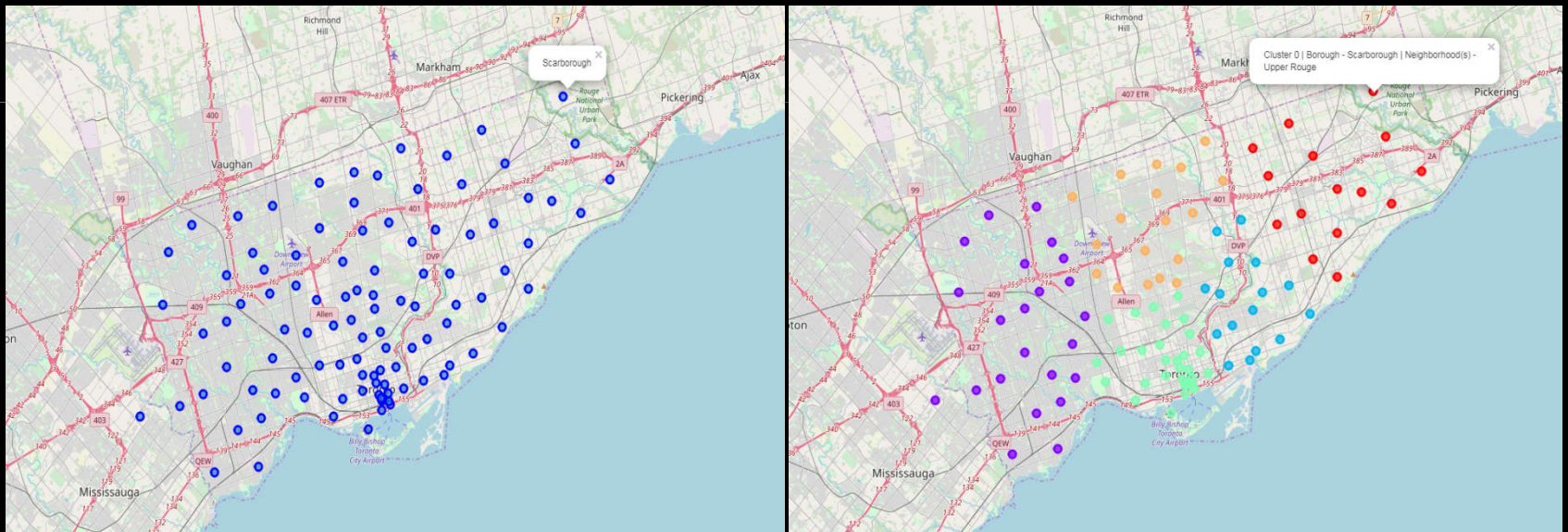
Results – Cleaned dataframe

- The final shape of the dataframe is 103 rows and 5 columns.

Out[8]:

	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Results – Clustering

- Comparison of normal vs clustered map
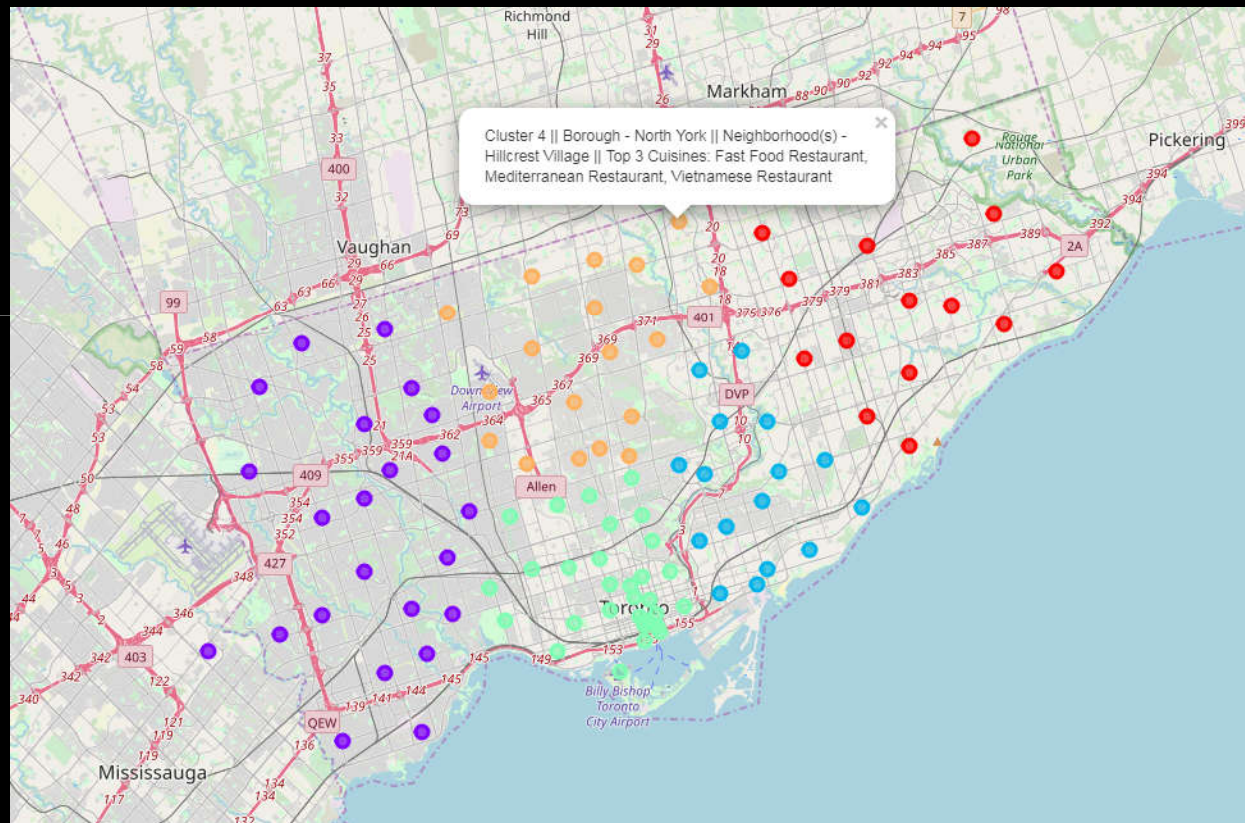
Results – Foursquare API

- The Foursquare API was used to gather information of the most common restaurants cuisines and rank them accordingly. This was done by first one hot encoding, grouping and calculating the frequency.

- Postalcode column was used as the key instead of the neighborhood column.

- This data was expanded to include the top 10 most common restaurant cuisines and merged with the first dataframe.

```
----M5M----
                  venue  freq
0      Italian Restaurant  0.18
1    Fast Food Restaurant  0.18
2        Sushi Restaurant  0.09
3     American Restaurant  0.09
4        Greek Restaurant  0.09


----M5P----
                  venue  freq
0        Sushi Restaurant   1.0
1       Afghan Restaurant   0.0
2  New American Restaurant   0.0
3     Japanese Restaurant   0.0
4       Jewish Restaurant   0.0


----M5R----
                  venue  freq
0          Indian Restaurant  0.25
1          Jewish Restaurant  0.25
2        American Restaurant  0.25
3  Vegetarian / Vegan Restaurant  0.25
4          Theme Restaurant  0.00
```

Results – Final map

Initial observations

- In cluster 0 (red), which is predominantly the Scarborough borough on the Eastern side, the two most common restaurants in the Top 3 are Vietnamese and Doner Restaurants.

- In cluster 1 (purple), which contains boroughs Etobicoke, York and North York on the western side, has far more NaN entries. This could indicate a possible source of development and worth further investigation. This is similar to cluster 4 (yellow), which is on the Northern side and consists mostly of the North York borough.

- In cluster 2 (blue), which is on the Southern Eastern side, has Fast Food and Italian cuisines in 1st place. Similarly in cluster 3 (light green), which is in Downtown Toronto on the Southern side, is predominantly Italian cuisine.

Conclusions

- The study consisted of acquiring location data related to Toronto City and leveraging the Foursquare API to get restaurant cuisines around various neighborhoods in Toronto City.

- Dataset was cleaned and K-mean cluster algorithm was employed.

- Foursquare API was used to get the Top 10 most common restaurant cuisines in every neighborhood(s) in each cluster.

- Embedded into an interactive map using Folium that gives a popup with all relevant data.

- Information can be used by potential entrepreneurs to make informed business decisions on where to setup the restaurant and what type of cuisine to offer.