Analyzing and Clustering Toronto Neighbourhoods

IBM Data Science Capstone Project

Introduction

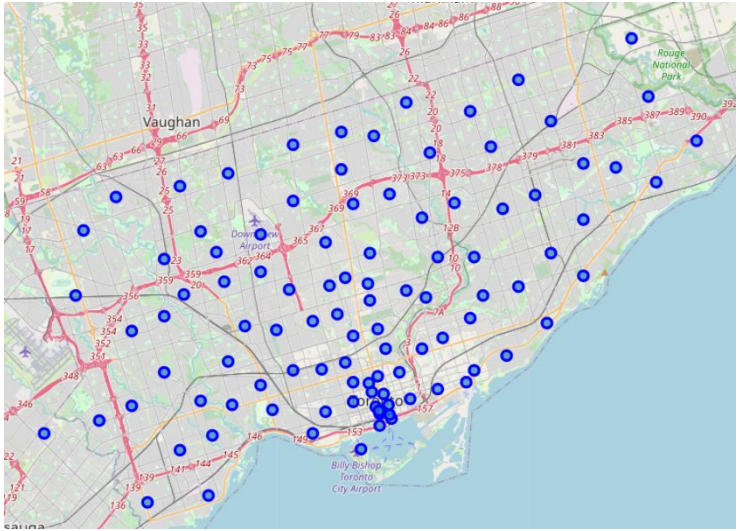
- Toronto is the most cultural city in the world and has many neighbourhoods which have different cultural experiences.
- We have to figure out which neighbourhood is the best to open an Indian restaurant.

Data gathering and cleaning

- Toronto's neighbourhood data web scraped from Wikipedia using BeautifulSoup.
- Toronto's venues data gathered using Foursquare API.
- Geographical coordinates gathered using Geopy Library.
- Cleaned the data by dropping the null values and converting the latitude and longitude to float.

Plotting all the neighbourhoods

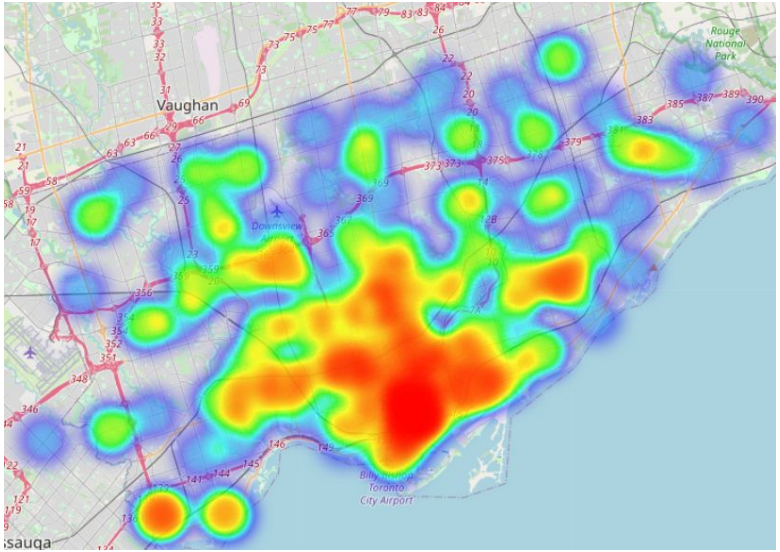
- Plotted all the neighbourhoods in Toronto, ON.
- Map generated using Folium.



Plotting all the Indian Restaurants.

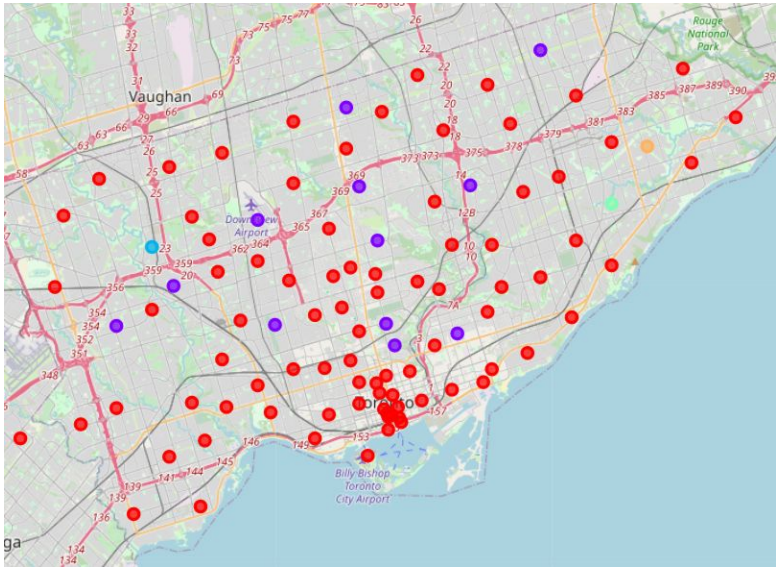
- Not a lot of Indian restaurants in Toronto.
- Therefore we need not worry about the competing Indian restaurants as they are pretty spaced out.

Plotting all the venues in Toronto.



- Density of venues in Toronto plotted using Heatmap.
- Downtown Toronto and North York has the greatest number of venues as compared to other areas.
- Therefore, according to these Heatmap Downtown Toronto and North York are best for an Indian restaurant.

Clustering Neighbourhoods



- Dividing neighbourhoods into 5 clusters based on the top 10 most common venues.
- Cluster 0 (Red) is the most dominant which consists largely of food places, coffee places and parks.
- These are the places where people mostly frequent.
- Therefore, these places are also an Indian Restaurant.

Results and Future directions

- Considering both, Heatmap and the clusters, Downtown Toronto looks like the best place to open an Indian Restaurant.
- There only 3 restaurants which are near to the lake. So a bit of those places would be the perfect place.
- We can also add the property prices to do further analysis and find the place which is least expensive to buy or rent.