Capstone Project - The Battle of Neighborhoods

Abhijith Antony

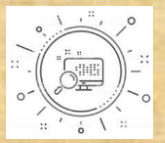
Objective

- Investing in a restaurant can be a very risky venture without a proper plan and strategy.
- Some of the factors to be considered while deciding the location is the competition in a neighborhood and the footfall of people in an area.
- Goal of the thesis is to identify an ideal neighborhood in Toronto where we can invest in, to open a restaurant.
- We will leverage the information available online regarding restaurants and other stores in an area and use K means clustering to divide the region into different areas with different degrees of competition.
- We assume that there is a high footfall of people in an area based on presence of other avenues.

Approach

Data sourcing

Data Cleaning & Manipulation



Exploration of the dataset

K Means Clustering

Data sourcing

Source 1: Information related to postal code and neighborhood names:
Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Source 2: Foursquare API : Information regarding the avenues in a neighborhood.

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Parkwoods	43.752935	-79.335641	Donalda Golf & Country Club	43.752816	-79.342741	Golf Course
1	Parkwoods	43.752935	-79.335641	Brookbanks Park	43.751976	-79.332140	Park
2	Parkwoods	43.752935	-79.335641	TD Canada Trust	43.753253	-79.347851	Bank
3	Parkwoods	43.752935	-79.335641	Variety Store	43.751974	-79.333114	Food & Drink Shop
4	Parkwoods	43.752935	-79.335641	TTC Stop #9075	43.757596	-79.338155	Train Station

Data Cleaning and Manipulation

Data Cleaning:

- Remove all the *'Not assigned'* neighborhoods from the list.
- More than one neighborhood can exist in one postal code area, these different neighborhoods must be combined in a single row, ensuring unique postal codes.
- Filter out avenues with *'Restaurant'* in its name.
- Fetching data from HTML page and converting into a data frame.

Exploration of the dataset

Observations:

Top Avenue: Tim Horton

Top Neighborhood: Downsview and Willowdale (68)

Neighborhood with least number of Avenues: Rouge Hill, Port Union, Highland Creek (1)

Packages Used:

Foursquare API: To fetch location information for the corresponding neighborhoods.

BeautifulSoup: For Web scraping data.

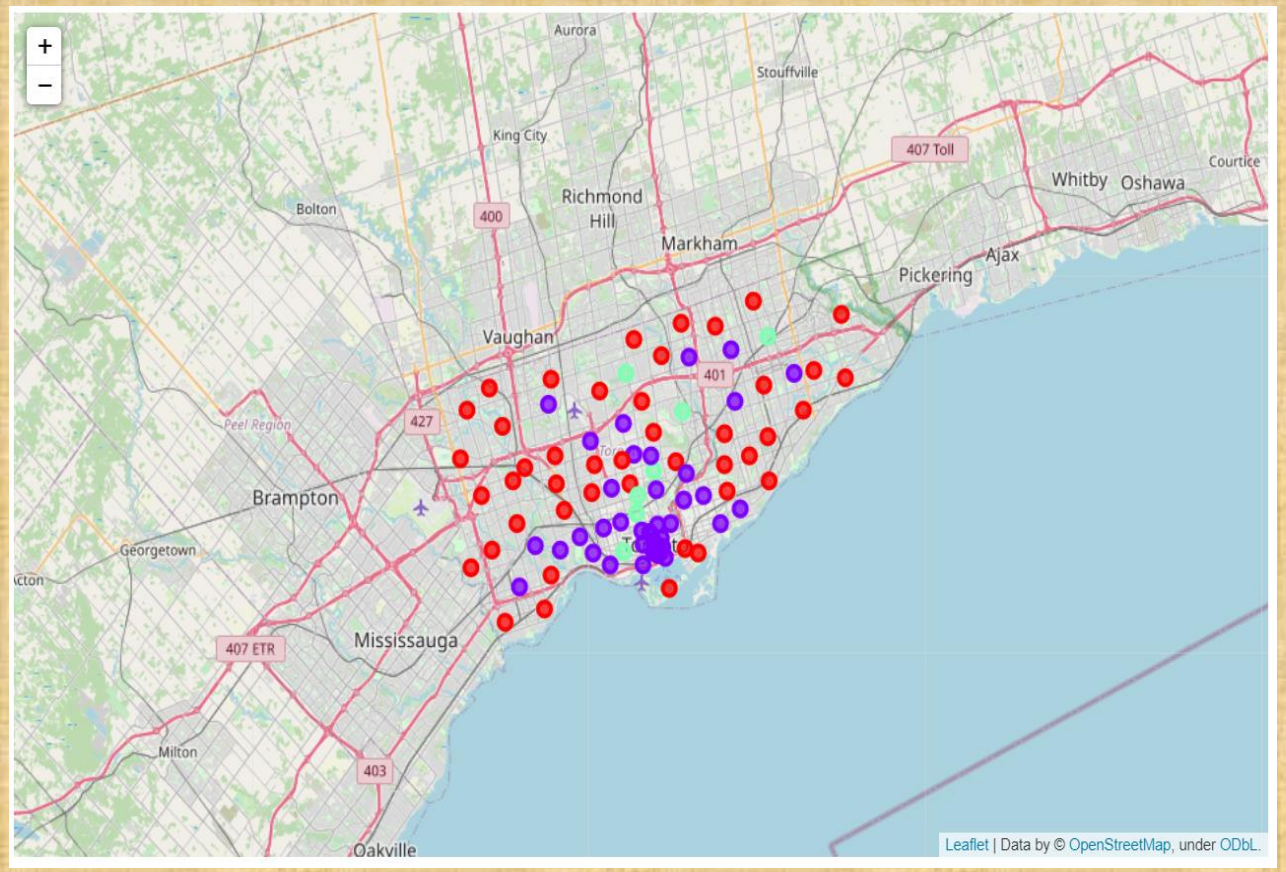
K Means Clustering

Observations:

- Red: Cluster 0: With least number of restaurants in the surroundings.

- Green: Cluster 1: With Average number of restaurants in the surroundings.
-
- Blue: Cluster 2: With Maximum number of restaurants in the surroundings (Will be the most competitive regions).

- We see that the least amount of competition is faced by the red or cluster 0 is mostly on the outskirts of the city. But we do see few neighborhoods closer to city coming in this segment which can be potential investment opportunity.



Recommendations

- We see that Cluster 0 is the least competitive and in cluster 0 neighborhood Woodbine Heights has the maximum footfall and other avenues in the region hence this would be the prime target followed by Regent Park, Harbourfront, Leaside, Runnymede and The junction North neighborhoods.

- In cluster 1 Downsview is most optimum neighborhood to invest in, this has the highest number of avenues in the region and since its in cluster1 there is only average competition which can be leveraged here.

- Cluster 2 is highly competitive with already a very high number of restaurants in the region hence I would not recommend investing here for the time being unless we have any further information which might change the current scenario.