



Coursera IBM Capstone Project: Opening a new Coffeehouse in Toronto

By: Rishi Mahajan



1.Business Problem

1. Find the most suitable location for the entrepreneur to open a coffeehouse in the Toronto area.
2. Objective: By using some techniques of data science and machine learning methods such as clustering we will guide entrepreneur, on which location he should open a shop to gain maximum profit.
3. Business question: If an entrepreneur wants to open a coffeehouse in toronto,where should he consider to open a coffeehouse?

2.Data

1. List of neighborhoods in Toronto, Canada.
2. Latitude and Longitude of these neighborhoods.
3. Venue data related to Coffeehouse. Venue data will help us find the neighborhoods that are most suitable to open a coffeehouse.

3. Sources of Data

1. Scrapping of Toronto neighborhoods via Wikipedia.
(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
2. Getting Latitude and Longitude data of these neighborhoods via a csv file and merging into the Toronto neighborhood data.
(https://cocl.us/Geospatial_data)
3. Using Foursquare API to get venue data related to these neighborhoods.

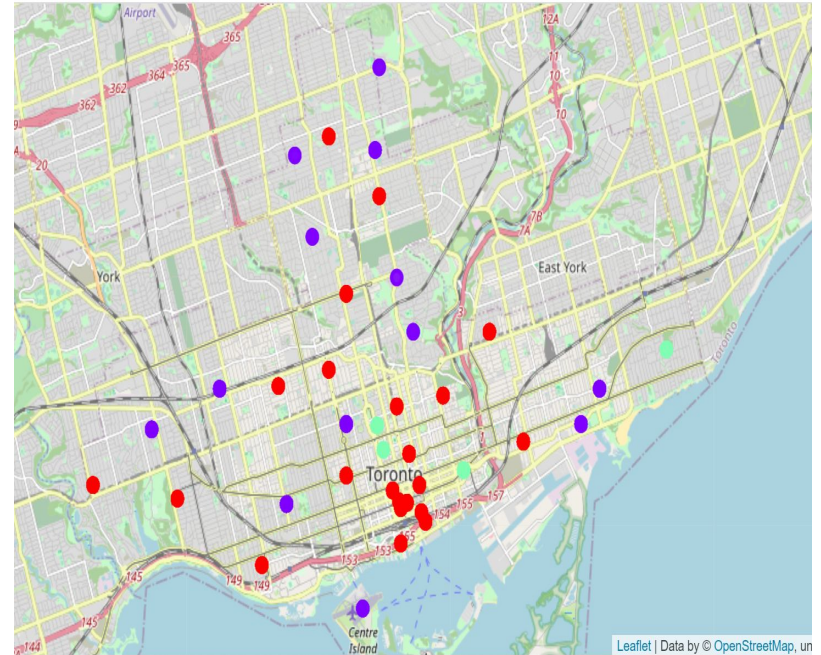
4. Methodology

1. Web scraping Wikipedia page for neighbourhoods list.
2. Get latitude and longitude coordinates using CSV file provided by IBM.
3. Visualised data by using Folium Package.
4. Use Foursquare API to get venue data.
5. Filter venue category by Coffee Shop.
6. Perform clustering on the data by using k-means clustering.
7. Visualize the clusters in a map using Folium.

5. Results

Categorized the neighbourhoods into 3 clusters

1. Cluster 0: Neighborhoods with a high number of Coffee Shops (Red Dots).
2. Cluster 1: Neighborhoods with little more number of coffee shops than cluster 2 (Purple Dot).
3. Cluster 2: Neighborhoods with very less number of Coffee Shops (Light Green dot).



6. Discussion

1. Most of the coffee shops are concentrated in the central area of the city.
2. Highest number in cluster 0 and moderate number in cluster 1.
3. Cluster 2 has very low number coffee shops in the neighbourhoods.
4. Cluster 2 areas are near to Queens Park, Glen Stewart Park and other areas.

7. Recommendations

1. Open new coffee house in neighbourhoods in cluster 2.
2. Opening near 'Glen Stewart Park' will be good option.
3. If entrepreneur wants to open in main city then location near 'Queens Park' will be best suggestion

8. Limitation and future work

1. Should consider parameters like population density, young and corporate people in that area, rent that could influence the decision to open a new shop.
2. Future research can take into consideration these factors.

9. Conclusion

1. The neighbourhoods in cluster 2 are the most preferred locations to open a Coffeehouse.
2. This project will help the relevant entrepreneur to capitalize on the opportunities in high potential locations while avoiding overcrowded areas in their decisions to open a new coffeehouse.

Thank You