Battle of Neighbourhoods

Capstone Project

Vikas Nagar

Agenda

- Introduction
- Data Source
- Methodology
- Results
- Discussion
- Conclusion

Introduction

- ABC Corp. business group wants to open a new venture in Toronto and wants to identify which would be the best business area to step into and in which neighborhood to open that business.
- They want to go big and establish a brand of them selves instead of competing with the already existing business groups in that area.
- Hence they want an analysis to be done to find out 2 things the most fast moving business venture which is very popular and the areas where its not available or not so popular. Their approach is contrary to the regular businesses and they want to get an analysis for this before finalizing the venue.

Data Source

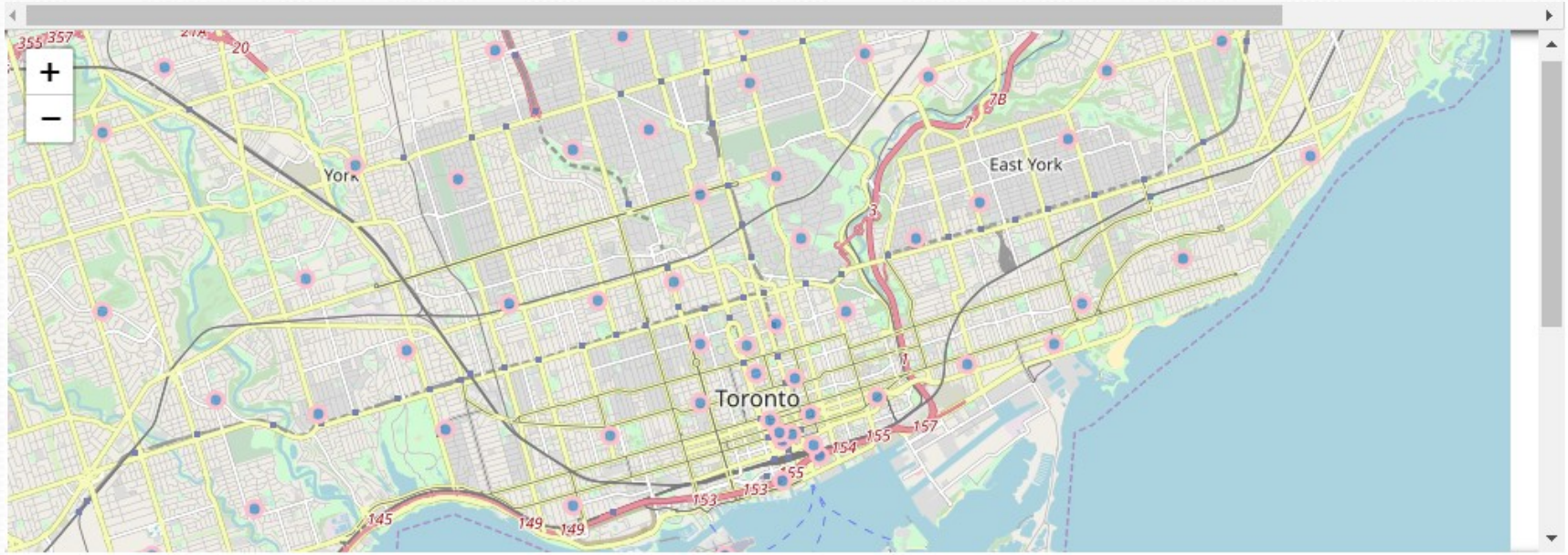
- Looking into the business requirements, we can see that there might be multiple ways of approaching the problem, right from capital required to best areas.
- The approach followed will be as follows:
- Get the Postal codes, Neighborhoods, Borough data from wiki
- Use the Foursquare API for the details of the Neighborhoods and most common / popular venues there to identify the most common venues across Toronto.
- Cluster the data to see the areas already having the most common business and identify the most common business for the ABC Corp. business group. This should give an idea on the most common business ventures which can be suggested to the ABC Corp. group.
- Now identify which are the Neighborhoods where such businesses are not the top venues. This means that if a good option is provided to the end users there is a potential of the business working out in the long run.
- Though there might be other approaches taken after doing this analysis, but for the same of the project we will restrict those here.

Methodology

- The methodology will be as follows:
  - Get all the postal codes with the Borough and Neighbourhood from wikipedia by scraping the data
  - For these use the longitude and latitude data prepare a dataframe . This can be displayed on a map using Folium.
  - Now to get the most common businesses use the Foursquare api (https://developer.foursquare.com) using a developer account.
  - Once we have the data prioritize by the topmost businesses in each neighbourhood. Then identify the Neighbourhoods where these common businesses do not appear in the top 2 business venues.
  - The Neighbourhoods tus identified can be suggested to the ABC Corp. As where they can start the new venture.

Results

- After scraping the data for the postal codes and mapping them to map using Folium we can see the map as follows:



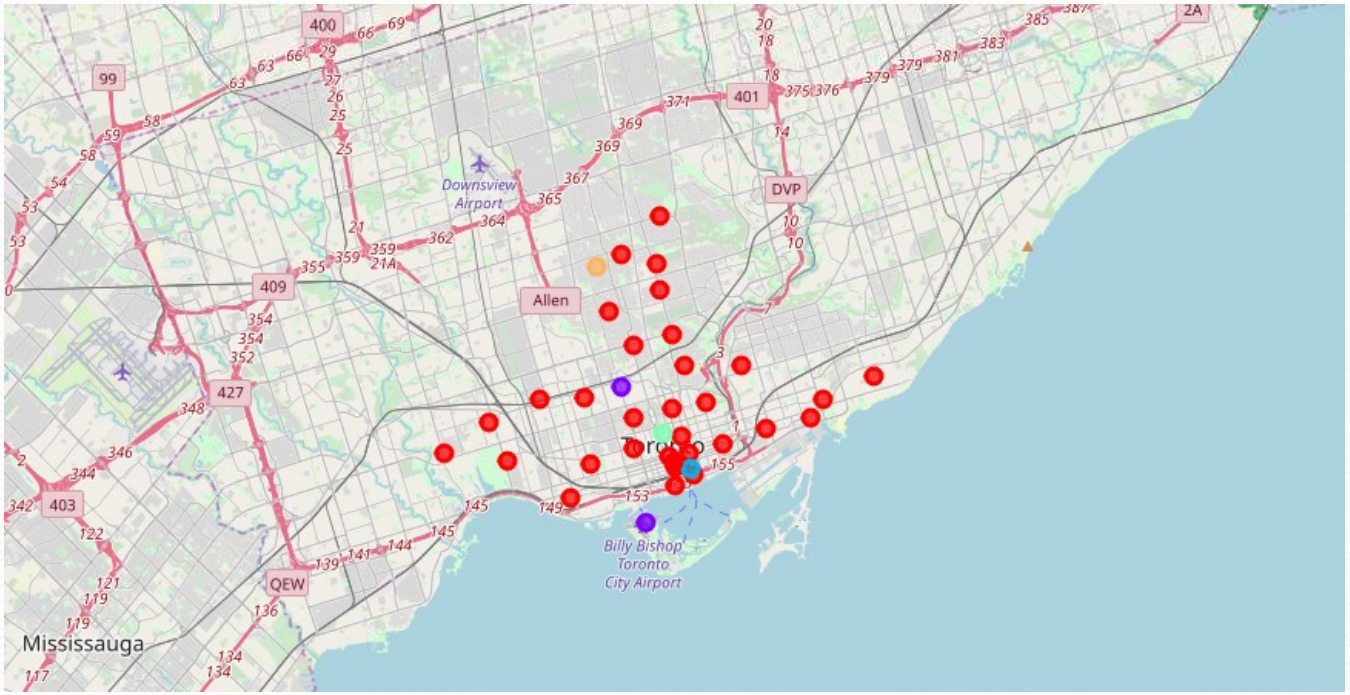
Results Continued...

Data from the Analysis shows very clearly that the most common business is a Coffee Shop or a Cafe

1st Most Common Venue	2nd Most Common Venue	
Coffee Shop	Café	7
Café	Bar	2
Supermarket	Pharmacy	1
Coffee Shop	Pub	1
Bar	Café	1
Breakfast Spot	Gift Shop	1
Bus Line	Trail	1
Café	Bookstore	1
	Coffee Shop	1
	Grocery Store	1
Coffee Shop	Aquarium	1
	Clothing Store	1
	Cocktail Bar	1
	Hotel	1
	Italian Restaurant	1
	Restaurant	1
Sandwich Place	Park	1
Coffee Shop	Sporting Goods Shop	1
	Sushi Restaurant	1
Dessert Shop	Sandwich Place	1
Dim Sum Restaurant	Park	1
Garden	Yoga Studio	1
Greek Restaurant	Ice Cream Shop	1
Gym	Playground	1
Hotel	Breakfast Spot	1
Light Rail Station	Yoga Studio	1
Neighborhood	Grocery Store	1
Park	Playground	1
Pizza Place	Coffee Shop	1
Restaurant	Coffee Shop	1
Airport Service	Airport Terminal	1

Results Continued…

- Now we create the clusters using Kmeans and map the same again on the map using Folium



Results Continued…

- Now we can filter the data to see the neighbourhoods where the 1st or second business is not Coffee Shop or Cafe.

	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
37	East Toronto	The Beaches	43.676357	-79.293031	0	Neighborhood	Grocery Store	Coffee Shop
41	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	0	Greek Restaurant	Ice Cream Shop	Coffee Shop
42	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572	0	Sandwich Place	Park	Burger Joint
44	Central Toronto	Lawrence Park	43.728020	-79.388790	0	Dim Sum Restaurant	Park	Lake
45	Central Toronto	Davisville North	43.712751	-79.390197	0	Hotel	Breakfast Spot	Food & Drink Shop
47	Central Toronto	Davisville	43.704324	-79.388790	0	Dessert Shop	Sandwich Place	Restaurant
48	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160	0	Gym	Playground	Park
50	Downtown Toronto	Rosedale	43.679563	-79.377529	0	Park	Playground	Trail
63	Central Toronto	Roselawn	43.711695	-79.416936	4	Garden	Yoga Studio	Farmers Market
64	Central Toronto	Forest Hill North, Forest Hill West	43.696948	-79.411307	0	Bus Line	Trail	Park
68	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420	1	Airport Service	Airport Terminal	Airport Lounge
76	West Toronto	Dovercourt Village, Dufferin	43.669005	-79.442259	0	Supermarket	Pharmacy	Bakery
83	West Toronto	Parkdale, Roncesvalles	43.648960	-79.456325	0	Breakfast Spot	Gift Shop	Italian Restaurant
87	East Toronto	Business reply mail Processing Centre969 Eastern	43.662744	-79.321558	0	Light Rail Station	Yoga Studio	Recording Studio

Discussion

- Thus we observed that when we clustered the most popular businesses data using the Fourswuare API details for the top 3 items, we got the results as the Coffee Shop or Cafe as the top most popular spots.
- Post that we identified the areas were this was not the case and there was scope of opening these .
- These can hence be suggested to ABC Corp. As an analysis.
- There can be furthest analysis done to identify the suggested areas for offices and maybe open a shop there first so that the ROI is better.

Conclusion

- This concludes the Capstone Project where we demonstrated the use of
- Python we scraping
- Machine learning clustering and
- Map presentation using Folium
- Analytical capabilities of python

Using these we were able to analyze the business problem for ABC Corp. To identify a new business venture and possible venues for opening those.