Report: Capstone Project – Battle of Neighborhood

Cheap and best Food **in CANADA**

Topics to be discussed are:

- *Launching Problem Statements*
- *Data Section*
- *working on Problem Statements*
- *Conclusion*

Launching Problem Statement

As many of the people are interested in travelling different parts of the world and to taste different foods and cuisines, it is necessary to spend money wisely and at the same time we should not compromise on the quality of the food.

Keeping this in mind, I have brainstormed some problems statements. They are:

1. Any newbie to the city Scarborough, Canada cannot find the best hotels.
2. In case they are able to find the best hotel, the cost of the hotel is too high.
3. **In order to avoid this situation, we are going to derive insights of the hotels which have best rating and also low cost.**

Data Section

This section deals with which data we are going to use and how to deal with that data.

Data that are going to use is from the Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

In this page we have table contains the details of postal code, borough and neighborhood. We are going to the scrap this data using **Request and lxml.html** packages.

After scrapping the data we will **pandas** library to convert data to Dataframe.

Then we will be using **numpy** for numerical calculations.

Geographical data is taken from: http://cocl.us/Geospatial_data

geopy.geocoders is used for get location details.

Foursquare API is used for making request to get the results in the desired location and using this we can do various operations.

We use **re** package to get matching string.

json package is used for getting the response from the **foursquare API**.

matplotlib library is used for plotting data, used for visualization.

folium is used for plotting geographical data.

sklearn is powerful library which contains built-in models, in this we use **k-means clustering** for clustering data

Working on Problem statements

- **Importing the data**

  To start the work, we want to import the data. Here, instead of importing the data, we are going to scrap the data from Wikipedia. To scrap the data we use *request, lxml.html* libraries. In this first we request the URL (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M), then it is converted to text content.

- **Converting to Dataframe**

  The obtained data is then converted to Dataframe using *pandas* library.

	Postal Code\n	Borough\n	Neighborhood\n
0	M1A\n	Not assigned\n	Not assigned\n
1	M2A\n	Not assigned\n	Not assigned\n
2	M3A\n	North York\n	Parkwoods\n
3	M4A\n	North York\n	Victoria Village\n
4	M5A\n	Downtown Toronto\n	Regent Park, Harbourfront\n

- **Cleaning the data in Dataframe**

  As raw data may contain inconsistent and NaN value. We want to remove this noisy data to get good accuracy. In the given data we are extra character i.e. \n and then we have not assigned value in some columns. Sometimes, the column misplaced in wrong order. So we want correct all kind of inconsistency in the given data. Since some the neighborhood having same postal code we want to merge the neighborhood. The following is the cleaned Dataframe,

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

- **Importing geological data**

  We are going to import geographical data from: http://cocl.us/Geospatial_data

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

- **Merging geological data and scrapped data**

  Since to merge data we want a common column, to make it we can want to change the column name *postal code* to *Postcode.*

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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

- **Plotting the map**

  Using *geopy.geocoders* library we get the location of *Scarborough, Canada* and plot the location using *folium* package. Folium is the best library to plot the geographical and used visualize the data more precisely.

  Following are the places in Scarborough, Canada

- **Using Foursquare API**
  This is one of the best API to provide day to day data related to location and it is one of the easiest way get the location data i.e. hotels, hospital, etc.. Using this data we can get more insights. This API helps to get required data by passing the URI in such a way what data we are looking for.


```
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    lat,
    lng,
    radius,
    LIMIT)
```

- **Making a data to fit for our Problem statement**
  Here we use *re* package to get data which has venue category as hotels or restaurants. By this way we filter the data.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Fairview, Henry Farm, Oriole	43.778517	-79.346556	New York Fries - Fairview Mall	43.778605	-79.343577	Restaurant
1	Fairview, Henry Farm, Oriole	43.778517	-79.346556	Thai Express	43.777990	-79.344091	Restaurant
2	Fairview, Henry Farm, Oriole	43.778517	-79.346556	Taco Bell	43.778611	-79.343186	Restaurant
3	Willowdale, Willowdale East	43.770120	-79.408493	Symposium Cafe Restaurant & Lounge	43.771075	-79.413396	Restaurant
4	Don Mills	43.725900	-79.340923	Swiss Chalet	43.726747	-79.341625	Restaurant

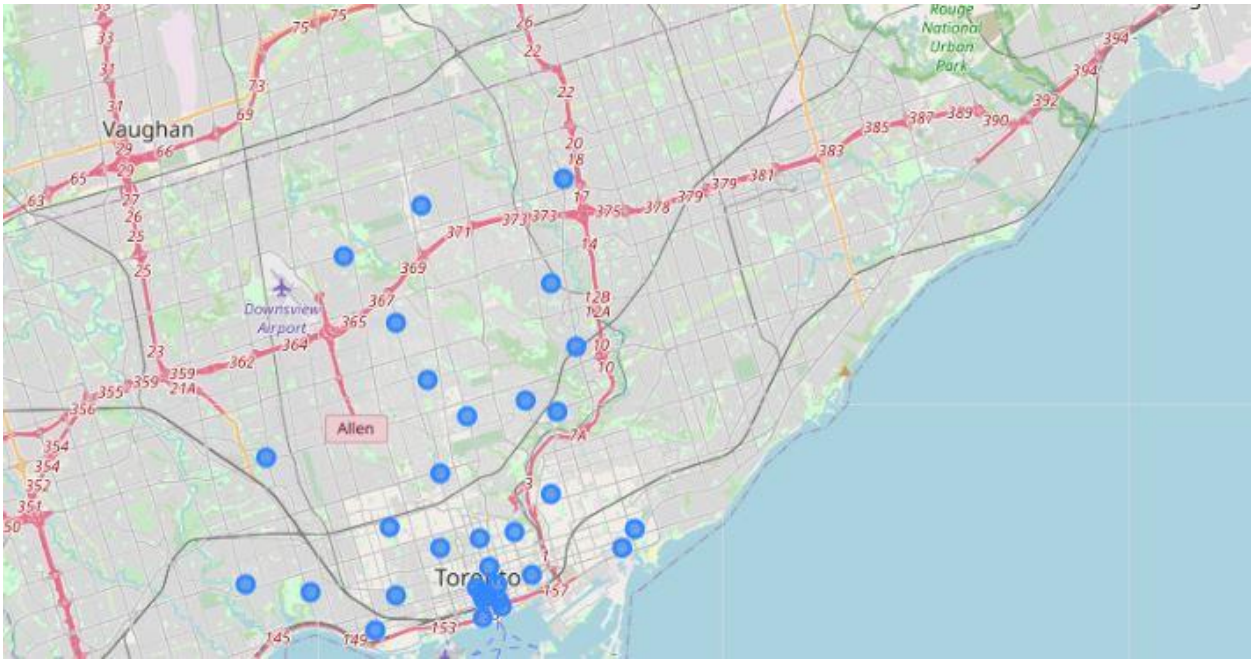
Once we filter data to analyze further we convert the row wise data of venue to column

	Neighborhood	Actinolite	Bagel Plus	Bannock	Bardi's Steak House	Bymark	Canoe	Carrot Commons	Cider House	Darbar Persian Grill	Drake One Fifty	El Nahual	Fabarnak	Founder Restaurant & Bar	GEORGE Restaurant
0	Fairview, Henry Farm, Oriole	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Fairview, Henry Farm, Oriole	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Fairview, Henry Farm, Oriole	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Willowdale, Willowdale East	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Don Mills	0	0	0	0	0	0	0	0	0	0	0	0	0	0

- **Applying cluster model**

  A cluster model is applied using *sklearn* library, which provides set of machine learning and they are predefined. It is enough to pass you dataset as an argument.

  The following diagram shows the filter out location which has only location of hotels.



  From the above diagram it is clearly seen that, location are filtered out. When you compare this with last map, you will know the difference.

  By applying k-means clustering model in sklearn, we will be able cluster the data. Here we apply 4 clusters.

The following are the clustered table

Cluster 1

	Borough	1st Most Common restaurant/hotel	2nd Most Common restaurant/hotel	3rd Most Common restaurant/hotel	Cluster Labels
32	East Toronto	The Green Wood	Wish	El Nahual	0.0

Cluster 2

	Borough	1st Most Common restaurant/hotel	2nd Most Common restaurant/hotel	3rd Most Common restaurant/hotel	Cluster Labels
2	North York	Swiss Chalet	Harvey's	Wish	1.0
3	North York	Swiss Chalet	Harvey's	Wish	1.0
8	East Toronto	Harvey's	Wish	Murgatroid	1.0
22	North York	Harvey's	Darbar Persian Grill	Wish	1.0

Cluster 3

	Borough	1st Most Common restaurant/hotel	2nd Most Common restaurant/hotel	3rd Most Common restaurant/hotel	Cluster Labels
0	North York	New York Fries - Fairview Mall	Thai Express	Taco Bell	2.0
1	North York	Symposium Cafe Restaurant & Lounge	El Nahual	Mary Be Kitchen	2.0
4	North York	Bagel Plus	Wish	Fabamak	2.0
5	East York	The Leaside Pub	Wish	El Nahual	2.0
6	East York	Swiss Chalet	Wish	Mary Be Kitchen	2.0
7	East Toronto	Rikkochez	Carrot Commons	Wish	2.0
9	Central Toronto	Sushi Shop	Wish	Fabamak	2.0
10	Central Toronto	Starving Artist	Wish	Fabamak	2.0
11	Central Toronto	Mary Be Kitchen	Wish	Murgatroid	2.0
12	Downtown Toronto	The Pear Tree	Murgatroid	Wish	2.0
13	Downtown Toronto	Wish	Fabamak	O. Noir	2.0
14	Downtown Toronto	Impact Kitchen	Wish	Murgatroid	2.0
16	Downtown Toronto	Victoria's Restaurant	Bannock	GEORGE Restaurant	2.0
17	Downtown Toronto	The Works Gourmet Burger Bistro	The Keg Steakhouse + Bar - Esplanade	Wish	2.0
18	Downtown Toronto	Bardi's Steak House	Bymark	The Keg Steakhouse + Bar - York Street	2.0
19	Downtown Toronto	Kellys Landing	Steam Whistle's Biergarten	Sushi Shop	2.0
20	Downtown Toronto	Victoria's Restaurant	The Keg Steakhouse + Bar - York Street	Canoe	2.0
21	Downtown Toronto	Victoria's Restaurant	Bardi's Steak House	The Keg Steakhouse + Bar - York Street	2.0
23	Downtown Toronto	Rasa	Harvest Kitchen	Wish	2.0
24	Downtown Toronto	Victoria's Restaurant	The Works Gourmet Burger Bistro	The Keg Steakhouse + Bar - Esplanade	2.0
25	Downtown Toronto	Drake One Fifty	Bardi's Steak House	Bymark	2.0
26	Downtown Toronto	Actinolite	Murgatroid	Mary Be Kitchen	2.0
27	West Toronto	Montgomery's	Founder Restaurant & Bar	Murgatroid	2.0
28	West Toronto	Vogue Supper Club	Fabamak	Montgomery's	2.0
29	York	El Nahual	Murgatroid	Mary Be Kitchen	2.0
30	West Toronto	Cider House	Wish	Fabamak	2.0
31	West Toronto	Supper Solved	Wish	Fabamak	2.0
33	Etobicoke	New Toronto Fish & Chips	Murgatroid	Mary Be Kitchen	2.0

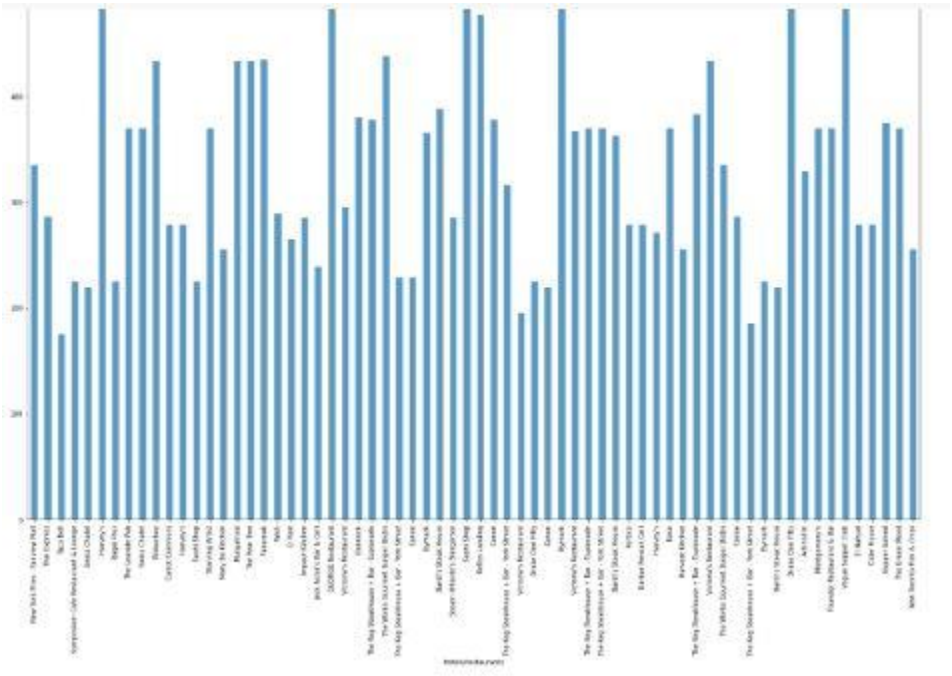
Cluster 4

	Borough	1st Most Common restaurant/hotel	2nd Most Common restaurant/hotel	3rd Most Common restaurant/hotel	Cluster Labels
15	Downtown Toronto	Jack Astor's Bar & Grill	Wish	Murgatroid	3.0

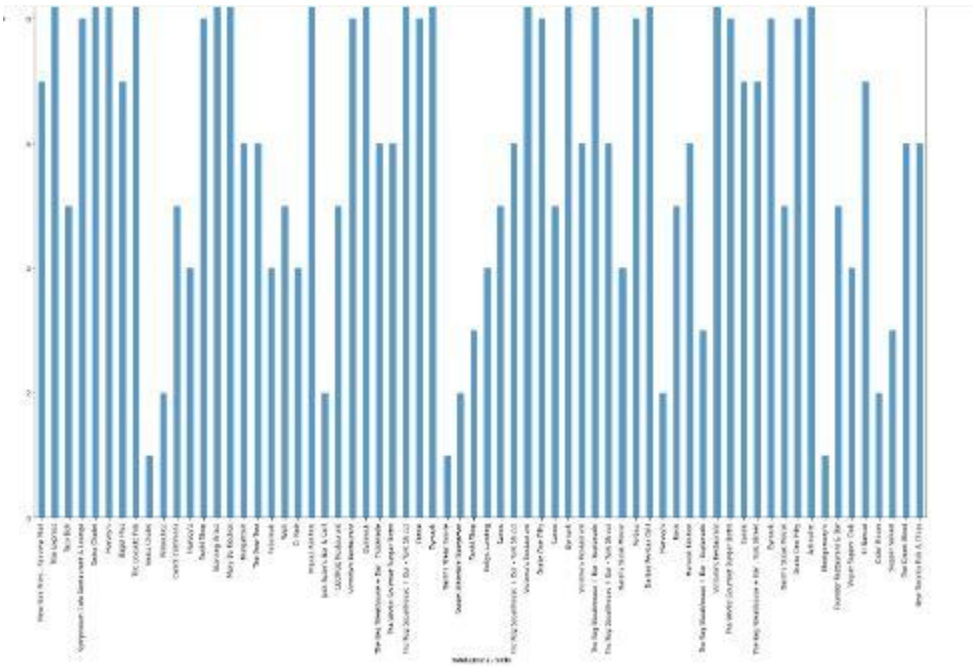
- **Finding best hotel**

To find best hotel, first we have to find the two things, which has average low price and having best rating. Since we do not have specific data, we are going to assign sample values. By applying sample value we build bar graph to see the hikes.

First we can have the data of price and plot it in bar graph. Following is the graph that tells about of price of the hotels/restaurant.



Following is the graph that tells about of rating of the hotels/restaurant.



From this we derive at the table, which tells the top hotel in Scarborough, Canada. Following is the table tells the top hotel in the city,

	hotels/restaurants	Food Rating	Average_food_Price
0	The Keg Steakhouse + Bar - York Street	10.0	185.0
1	The Keg Steakhouse + Bar - York Street	10.0	229.0
2	The Keg Steakhouse + Bar - York Street	10.0	316.0
3	Swiss Chalet	10.0	219.0
4	Harvey's	10.0	271.0
5	Harvey's	10.0	279.0
6	Impact Kitchen	10.0	285.0
7	Darbar Persian Grill	10.0	279.0
8	Bymark	9.0	225.0
9	Bymark	9.0	225.0
10	Bymark	8.0	225.0
11	Victoria's Restaurant	9.0	195.0
12	Victoria's Restaurant	9.0	295.0
13	Victoria's Restaurant	9.0	195.0
14	Victoria's Restaurant	9.0	295.0
15	Victoria's Restaurant	8.0	195.0
16	Victoria's Restaurant	8.0	295.0
17	Actinolite	9.0	329.0
18	Thai Express	9.0	286.0
19	Mary Be Kitchen	9.0	255.0
20	Drake One Fifty	8.0	225.0
21	Drake One Fifty	8.0	225.0
22	Portico	8.0	279.0

Conclusion

From the above working, we are able to get the segregate data related to hotel and we are able to apply k-means clustering on that data. Then we are able to find the best hotel in the city. (Note: price and ratings are assumed data values)

The top three hotels are

- The Keg steakhouse + Bar – York Street.
- Swiss Chalet
- Harvey's