THE BATTLE OF NEIGHBORHOODS (WEEK 2)- REPORT

Introduction

In this week's submission, I have selected to use the city of Toronto and find neighborhoods that would be best suited for a restaurant. This is created in a way to allow a person to get a brief knowledge into who will be the target market for a certain area based on the overall population of that area.

Business Problem

Toronto encompasses a geographical area formerly administered by many separate municipalities. These municipalities have each developed a distinct history and identity over the years, and their names remain in common use among Torontonians. Former municipalities include East York, Etobicoke, Forest Hill, Mimico, North York, Parkdale, Scarborough, Swansea, Weston and York. Throughout the city there exist hundreds of small neighbourhoods and some larger neighbourhoods covering a few square kilometres.

Toronto's neighborhood is divided into multiple types ranging from Old Toronto to Suburbs to Industrial areas. Toronto also is the most populous city in Canada and has a really diverse population allowing it to become an important destination to a lot of immigrants arriving to Canada.

This diversity as well as the sheer size of the city itself makes it really difficult for a new business to decide what area would be most suited for their business in such a way that they can maximize their profit without worrying about as less competition as possible.

Thus, that’s the idea behind this project as a way to allow new business with some knowledge on where and when to open a new business in an area.

Data Used

Based on the problem undertaken, multiple components affect the decision to be taken, some of which are:

1. Any existing restaurant in the borough or neighborhood.

2. Age and income of the people visiting these restaurants.

3. Distance to the city centre.

Utilization of **Wikipedia** was used to generate areas centres along with the use of the **Foursquare** API to get the number of restaurant and their types in the areas obtained through the Wikipedia.

```
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_toronto)

map_toronto
```

Out[16]:

```
In [17]: # The code was removed by Watson Studio for sharing.

Foursquare API Client_ID verified
```

Fig.: *Snapshot of utilization of Foursquare API.*

Methodology

As discussed in the Business problem, the main initiative of the project is to find the best possible location for a restaurant to open to minimize competition and maximize profit.

To achieve the same, we require two data sets, one that provides us with boroughs information in Toronto as well as another that provides us with the age group and population of said areas. To solve the problem, K-Means algorithm was used.

K-Means clustering is a method used for clustering analysis. It is a type of unsupervised learning, which is used when there is unlabelled data present, i.e. data without defined categories or groups. The goal of the algorithm is to find groups in the data with the number of groups represented by the variable *K*. The algorithm works iteratively to assign each data point to one of *K*-groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the *K*-means clustering algorithm are:

- The centroids of the *K* clusters, which can be used to label new data
- Labels for the training data (each data point is assigned to a single cluster)

Following steps were taken:

- Gathering of data from different sources based on latitude and longitude as well as the types of restaurant in a locality.
- Utilization of K-Means to identify good locations close to centre with low number of restaurant and their types.
- Creation of clusters of location that meet the basic requirements set by the stake-holders. Radius of 500 metres was created to keep a check on other restaurants.

Methodology

Manipulation of data

Wikipedia was used to get data for Toronto, Canada. This data was cleaned by removing any and all unnecessary value.

	Postal_Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M6A	North York	Lawrence Heights
4	M6A	North York	Lawrence Manor
5	M7A	Downtown Toronto	Queen's Park
6	M9A	Etobicoke	Islington Avenue
7	M1B	Scarborough	Rouge
8	M1B	Scarborough	Malvern
9	M3B	North York	Don Mills North
10	M4B	East York	Woodbine Gardens
11	M4B	East York	Parkview Hill
12	M5B	Downtown Toronto	Ryerson
13	M5B	Downtown Toronto	Garden District
14	M6B	North York	Glencairn
15	M9B	Etobicoke	Cloverdale
16	M9B	Etobicoke	Islington
17	M9B	Etobicoke	Martin Grove
18	M9B	Etobicoke	Princess Gardens
19	M9B	Etobicoke	West Deane Park
20	M1C	Scarborough	Highland Creek
21	M1C	Scarborough	Rouge Hill
22	M1C	Scarborough	Port Union
23	M3C	North York	Flemingdon Park

Fig.: *Snapshot of data after cleaning process.*

Combining different data sources

Datasets of postal address was combined with dataset of latitude and longitude in a separate data frame. This new data frame was visualized using a folium map.

Usage of Foursquare API

Foursquare API was utilized to get the data on restaurant in Toronto and filtered the venue for all possible restaurants. Each neighborhood along with the top ten most common venues was identified.

Fig.: *Following map was created using folium.*

Clustering

K-Means algorithm was used to come up with 5 clusters that were different in Toronto. Identify the clusters & Boroughs/Neighborhoods with Maximum number restaurants and their types.

Fig.: *Following map was created following the clustering process.*

Results & Discussion

Analysis of the data received shows that Toronto has a large number of restaurants but there are some areas with lower number of restaurants in the city near the city centre. There are 4 boroughs and 39 neighborhoods in the geographical co-ordinates **43.6534817, -79.3839347.**

Based on the analysis, Cluster 3 has the highest amounts of restaurants than the rest of them. This makes introduction of a new restaurant in that cluster non-feasible due to the number of restaurants present. Thus, recommended zones should only be taken as a starting point in area selection followed by a detailed analysis to get the full details to get the most out of the area.

Cluster 3

```
In [47]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[47]:

	Borough	Cluster_Labels	Cluster Labels	1st Popular Venues	2nd Popular Venues	3rd Popular Venues	4th Popular Venues	5th Popular Venues	6th Popular Venues	7th Popular Venues	8th Pop Ven
0	East Toronto	2	2	Beach	Pizza Place	Thai Restaurant	Breakfast Spot	Park	Tea Room	Coffee Shop	Bar
1	East Toronto	2	2	Coffee Shop	Café	Vegetarian / Vegan Restaurant	Pizza Place	Cosmetics Shop	Japanese Restaurant	Movie Theater	Sho Stor
6	Downtown Toronto	2	2	Diner	Restaurant	Gastropub	Grocery Store	Pub	Indian Restaurant	Rock Club	Spa
7	Downtown Toronto	2	2	Coffee Shop	Japanese Restaurant	Gay Bar	Burger Joint	Café	Restaurant	Sushi Restaurant	Yog Stu
8	Downtown Toronto	2	2	Coffee Shop	Café	Restaurant	Hotel	Italian Restaurant	Pizza Place	Sporting Goods Shop	Spo
9	Downtown Toronto	2	2	Coffee Shop	Pizza Place	Grocery Store	Café	Food & Drink Shop	Breakfast Spot	Market	Bus
10	Downtown Toronto	2	2	Coffee Shop	Café	Restaurant	Italian Restaurant	Japanese Restaurant	Hotel	Beer Bar	Bak
11	Downtown Toronto	2	2	Coffee Shop	Restaurant	Café	Hotel	American Restaurant	Gym	Gastropub	Sea Res
12	Central Toronto	2	2	Coffee Shop	Restaurant	Italian Restaurant	Fast Food Restaurant	Yoga Studio	Diner	Pizza Place	Par
14	West Toronto	2	2	Café	Coffee Shop	Eastern European Restaurant	Bakery	Indian Restaurant	Beach	Shoe Store	Res
15	West Toronto	2	2	Café	Coffee Shop	Bakery	Bank	Pizza Place	Indian Restaurant	Dessert Shop	Flo Sho

Fig.: *Cluster 3 was found to be the one with most amount of restaurants.*

Conclusion

The purpose of this coursework was to identify areas in Toronto that have a low number of restaurants and help new businesses in choosing an optimal location for a restaurant. This was achieved with the utilization of data from Wikipedia as well as the use of the Foursquare API and zones were created to serve as a starting point for further investigation by the stakeholders themselves.

The decision on the optimal location would depend on the stakeholders themselves and will include additional factors besides the one used in this project, such as proximity of location to prime spots like Parks or offices, level of noise and access to the main roads as well as the affordability of real estate along with the socio-economic dynamics of the neighborhoods. Thus, this concludes the project as a whole.