Capstone Project- The Battle of Neighborhoods: Restaurants in Toronto

This document is prepared as a partial requirement to complete Coursera's [IBM Data Science Professional Certificate](#) specialization.

1. Introduction

In line with the recent trends of globalization and global awareness, one of the qualities of major cities of the world is their suitability for expats and visiting foreigners. One of the ways that a cities suitability can be measured the diversity of its restaurants. Hence, the problem we will try to address here is the availability and distribution of restaurants serving foreign foods in Toronto, Canada. We will be using location data from Foresquare to give insight to people considering to visit or relocate to Toronto as to how they should choose neighborhoods.

2. Data

For this project, we need the following data:

1. Toronto data that contains list of neighborhoods
    Data source:
    https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
    Description:
    We will Scrap Toronto neighborhoods along with their postal codes from Table on the Wikipedia page
2. Geographical data containing latitude and longitude of postal codes
    Data source:
    http://cocl.us/Geospatial_data
    Description:
    We will read the csv data and import the longitude and latitude corresponding to the postal codes.
3. Restaurants in each neighborhood of Toronto:
    Data source:
    Foursquare API calls
    Description:
    By using this API we will get all the venues in each neighborhood. We can filter these venues to get only restaurants.

3. Methodology

3.1. Data Preparation

The first step of preparing the data is web scraping of Toronto Neighborhoods Table from Wikipedia. For this purpose, the **BeautifulSoup4** library is used to read the Wikipedia table.

```
source = requests.get("https://en.wikipedia.org/wiki/List_of_postal_
codes_of_Canada:_M").text
soup = BeautifulSoup(source, 'lxml')
```

After further processing the data, a data frame with columns containing the postal code along with the neighborhood and borough names is created. The header of the data-frame is shown below.

| PostalCode	Borough	Neighborhood
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

The next step is to obtain the longitude and latitude of these postal codes. This is realized by utilizing the Nominatim library from geopy.geocoder client as follows.

```
geodf=pd.read_csv('http://cocl.us/Geospatial_data')
geodf.rename(columns={'Postal Code':'PostalCode'},inplace=True)
df_toronto= pd.merge(geodf,df3, on='PostalCode')
df_toronto=df_toronto[['PostalCode','Borough','Neighborhood','Latitude'
,'Longitude']] address = 'Toronto, ON, Canada' geolocator =
Nominatim(user_agent="ny_explorer") location =
geolocator.geocode(address) latitude = location.latitude longitude =
location.longitude
Toronto_data=df_toronto[df_toronto['Borough'].str.contains("Toronto")].
reset_index(drop=True) Toronto_data.head()
```

The head of the resulting data frame is shown below.

PostalCode	Borough	Neighborhood	Latitude	Longitude
M4E	East Toronto	The Beaches	43.676357	-79.293031
M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
M4M	East Toronto	Studio District	43.659526	-79.340923
M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

Let us use folium to visualize the geographic data we have so far by creating a map of Toronto with neighborhoods superimposed on top.
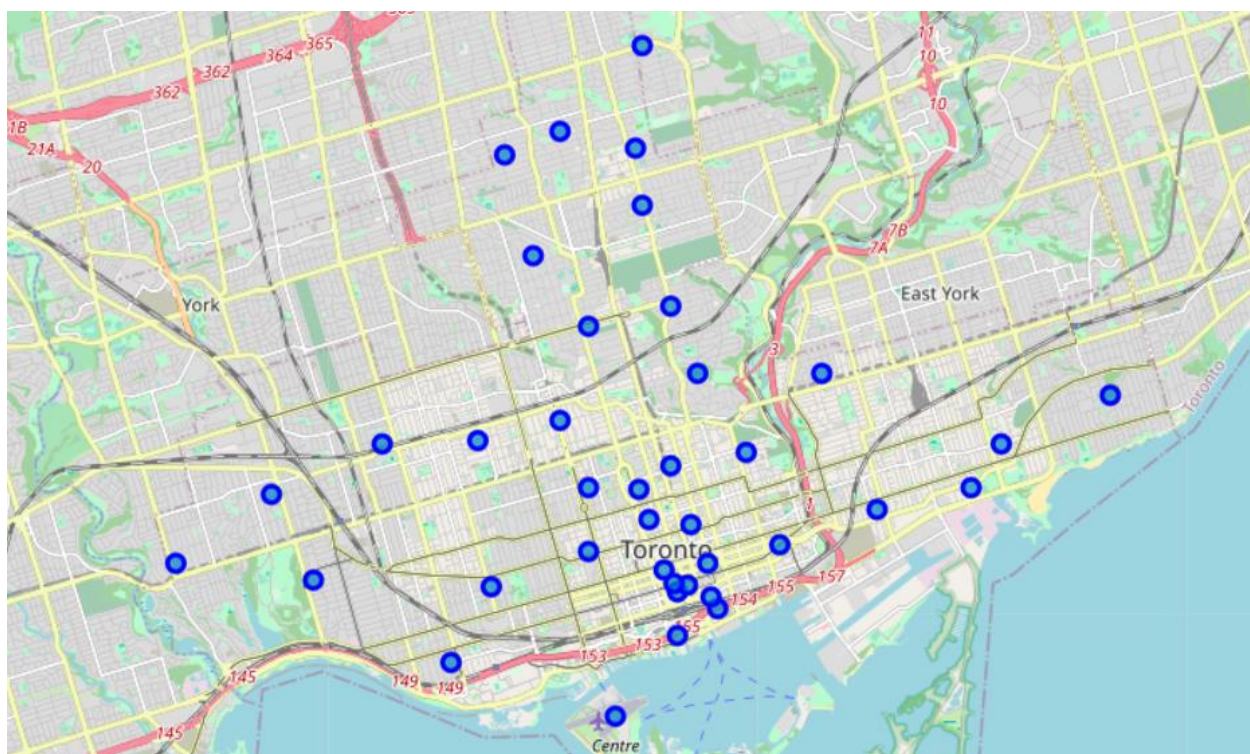
3.2.   Exploratory Data Analysis

Exploratory Data Analysis will help us understand and show the diversity and distribution of restaurants in Toronto. For this purpose we will make an API call to Foresquare server and get the venues near the locations, presumably the centroids of the neighborhoods. After making the call and reshaping the result, the head of the nearby venues are obtained as follows.

name	categories	lat	lng
The Fox Theatre	Indie Movie Theater	43.672801	-79.287272
Kew Gardens	Park	43.669038	-79.298538
Kew-Balmy Beach	Beach	43.667372	-79.295312
Woodbine Beach	Beach	43.663112	-79.306374
Hollandaise Diner	Breakfast Spot	43.686527	-79.308897

However, we are only interested in venues falling under Restaurant category and we need to categories the venues into neighborhood so that we can reach at our understanding of the density and distribution of restaurants among the neighborhoods in Toronto. For this purpose, we will use the pandas one-hot encoding.

After using the one-hot encoding for all venues, we get the following table.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	The Fox Theatre	43.672801	-79.287272	Indie Movie Theater
1	The Beaches	43.676357	-79.293031	Kew Gardens	43.669038	-79.298538	Park
2	The Beaches	43.676357	-79.293031	Kew-Balmy Beach	43.667372	-79.295312	Beach
3	The Beaches	43.676357	-79.293031	Woodbine Beach	43.663112	-79.306374	Beach
4	The Beaches	43.676357	-79.293031	Hollandaise Diner	43.686527	-79.308897	Breakfast Spot
...	...	...	...	...	...	...	...
3895	Business reply mail Processing Centre, South C...	43.662744	-79.321558	Union Pearson Express	43.644362	-79.383199	Train Station
3896	Business reply mail Processing Centre, South C...	43.662744	-79.321558	Steam Whistle Brewing	43.641752	-79.387089	Brewery
3897	Business reply mail Processing Centre, South C...	43.662744	-79.321558	Black Camel	43.677016	-79.389367	BBQ Joint
3898	Business reply mail Processing Centre, South C...	43.662744	-79.321558	Art Gallery of Ontario	43.654003	-79.392922	Art Gallery
3899	Business reply mail Processing Centre, South C...	43.662744	-79.321558	LCBO	43.681497	-79.391261	Liquor Store

Let us see the most common restaurants in Toronto. As it can be seen from the following table Japanese, French and American Restaurant are the most common restaurants in Toronto.

Venue_Category	Frequency
Japanese Restaurant	99
Restaurant	84
French Restaurant	59
American Restaurant	50
Mediterranean Restaurant	48
Asian Restaurant	45
Italian Restaurant	42
Vegetarian / Vegan Restaurant	41
Vietnamese Restaurant	33
Caribbean Restaurant	32

As an Ethiopian, I am also interested to find out the existence on Ethiopian Restaurants in Toronto. I directly visited the Foresquare website and found out many restaurants serving Ethiopian food. However, as it can be seen on the following table only nine Ethiopian restaurants are returned by API call. This confused me a bit and when I thought it further, this could be because of two factors. The first reason may be that Ethiopian Restaurant are not common venues as compared to other popular venues hence they are not returned to the API call. The second reason may be that related to not explicitly categorize the Ethiopian Restaurants in the venue category.

```
                  Venue Category  Neighborhood
          Turkish Restaurant                 1
   Latin American Restaurant                 7
 Eastern European Restaurant                 7
         Ethiopian Restaurant                 9
```

4. Result and Discussion

We have found that Japanese, French and American Restaurant are the most common restaurants in Toronto. Because of this any visitor, preferring Japanese, French or American food would be happy.

We have also found that Christie neighborhood is the best neighborhood in terms of diversity of the restaurants

Neighborhood	
Christie	26
University of Toronto, Harbord	25
The Annex, North Midtown, Yorkville	25
Kensington Market, Chinatown, Grange Park	24
Richmond, Adelaide, King	24

5. Conclusion

In this project, we have tried to acquire location data and merge them to obtain the necessary insight expected. There is definitely room for improvement on the workflows and the data used. However, as far as the capstone project is required, the author believes what is done is enough.