THE BATTLE OF NEIGHBORHOODS REPORT

TOWARDS IBM DATA SCIENCE
CERTIFICATION ORGANIZED BY COURSERA

BY

OYEKANMI O. OLATUNDE

OCTOBER 22, 2019

By Oyekanmi Olatunde: Towards Data Science Certification

1. Introduction:

1.1. Background

Toronto is the largest city in Canada and an economically viable place in terms of business, finance, entertainment, technology and so on. With a population of about 3 million people, the city is ranked as one of the most livable cities in the world according to the report by the Economist Intelligence Unit (City of Toronto, 2019). Within the city of Toronto, Downtown Toronto can boast of the city's main business areas and the most culturally diverse city in Toronto, Ontario. My friend(Tunde) and I, have lived in a popular neighborhood within Downtown Toronto for more than a decade now and recently, he shared his intention of leaving downtown for another unknown city in Toronto. Tunde wants a neighborhood that is similar to his current borough because he feels that would help him settle easily.

1.2. Interest

In the first part of this project, I will compare two most populated Boroughs in the city of Toronto to seek for clusters or patterns among neighborhoods. In addition to my analysis, I will also analyze and recommend a place in any of the overall neighborhoods for anyone planning to open a restaurant. The location of interest must be a densely populated area with few or no restaurant. I will utilize the data scientist skills to explore Toronto neighborhood datasets and extract the needed information. Advantages of each area will then be clearly expressed so that the best possible final location can be chosen by stakeholders

2. Data Acquisition and Cleaning

2.1. Data Acquisition:

Three different datasets were used, someone of which was scraped from Wikipedia and Canada open data portal. You can view the data through the link shown below ;

- Canada data: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- PopulationInformation: https://www12.statcan.gc.ca/censusrecensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&SR=1&S=22&O=A&RPP=9999&PR=0
- Geocoordinates of Canada by Postal codes: http://cocl.us/Geospatial_data

2.2. Data Cleaning :

Some data downloaded or scraped from the website has some missing data. For example, the data in **Fig. 1** below was an Html data scraped from Wikipedia. After the dataset was read into a data frame, it was realized that some Boroughs and Neighborhoods are not assigned.

```
In [254]: #scrape Canada html data from web
html_doc = "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M"
df_Canada = pd.read_html(html_doc, header=0)
df_Can = df_Canada[0]
df_Can.head()
```

Out[254]:

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

Fig. 1: Canada table read into a dataframe

In the data cleaning process, I first selected the rows with an assigned Borough, appended multiple neighborhoods with the same postal codes and replaced unassigned neighborhoods with their Boroughs. I also read the comma-separated file which had the geocoordinates by postal codes and then performs a left-on merge on the three data frames to get the required dataset. I further sorted the data frame in the order of decreasing population so I can extract the two most populated boroughs, which is our data of interest. The resulting data frame is shown below;

Out[321]:

	Borough	Population
0	Downtown Toronto	271186.0
1	West Toronto	212782.0
2	Central Toronto	164502.0
3	East Toronto	113966.0

Fig. 2 : Dataset of Boroughs in descending order of population

The result above shows that the top two most populated Borough in Toronto is **Downtown T oronto** and **West Toronto** with a population of 27,1186 and 21,2782 people respectively. Sin ce our stakeholder is interested in not only a similar neighborhood but also in a place with a relative population as his current borough, it is not a bad decision to choose the next in line in terms of population, **West Toronto**. The fig. 3 below has the Neighborhoods and geocoordinat es of west and downtown Toronto.

```
In [272]: # Cocatenate the dataframes of West Toronto and Downtown Toronto
West_Downtown_Tor = pd.concat([West_Tor, Downtown_Tor], join='inner').reset_index(drop=True)
West_Downtown_Tor.head()
```

Out[272]:

	Borough	Neighbourhood	Population	Latitude	Longitude
0	West Toronto	Dufferin,Dovercourt Village	44950.0	43.669005	-79.442259
1	West Toronto	Parkdale Village,Exhibition Place,Brockton	40957.0	43.636847	-79.428191
2	West Toronto	High Park,The Junction South	40035.0	43.661608	-79.464763
3	West Toronto	Swansea,Runnymede	34299.0	43.651571	-79.484450
4	West Toronto	Trinity,Little Portugal	32684.0	43.647927	-79.419750

Fig. 3 : Dataframe with West and Donwtown Toronto Dataset

We can visualize the neighborhoods of downtown and west Toronto on the map of Toronto as shown below ;

Fig. 4 : Toronto map with neighborhoods in West and Downtown Toronto

3. Exploratory Data Analysis

Using the Foursquare API, the top 100 venues for each neighborhood were extracted. While some other neighborhood venues are not up to 100, few neighborhoods stood out as one with the highest number of venues. For example, the dataframe below displays neighborhoods with the highest number of venues. This implies these neighborhood areas are somewhat busy areas with lots of traffics.

```
In [345]: #Sort dataframe by venues in descending order
          West_Downtown_most_venues_sorted= West_Downtown_most_venues.sort_values('Venue', ascending=False)
          # Since the max is 100 , we print dataframe with venues equal to the maximum values
          West_Downtown_most_venues_sorted= West_Downtown_most_venues[West_Downtown_most_venues['Venue']==100]
          West_Downtown_most_venues_sorted
```

Out[345]:

Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adelaide,Richmond,King	100	100	100	100	100	100
Chinatown,Kensington Market,Grange Park	100	100	100	100	100	100
Commerce Court,Victoria Hotel	100	100	100	100	100	100
Design Exchange,Toronto Dominion Centre	100	100	100	100	100	100
First Canadian Place,Underground city	100	100	100	100	100	100
Ryerson,Garden District	100	100	100	100	100	100
St. James Town	100	100	100	100	100	100
Union Station,Toronto Islands,Harbourfront East	100	100	100	100	100	100

Fig. 5: The neighborhoods with the highest number of venues

Another great idea is to find all venues that are restaurants in the neighborhoods. Since the project's aim is to also suggest a location for opening a restaurant, it would be nice to know the number of venues that are restaurants in each neighborhood. We can visualize the restaurants on a Toronto map as displayed below ;

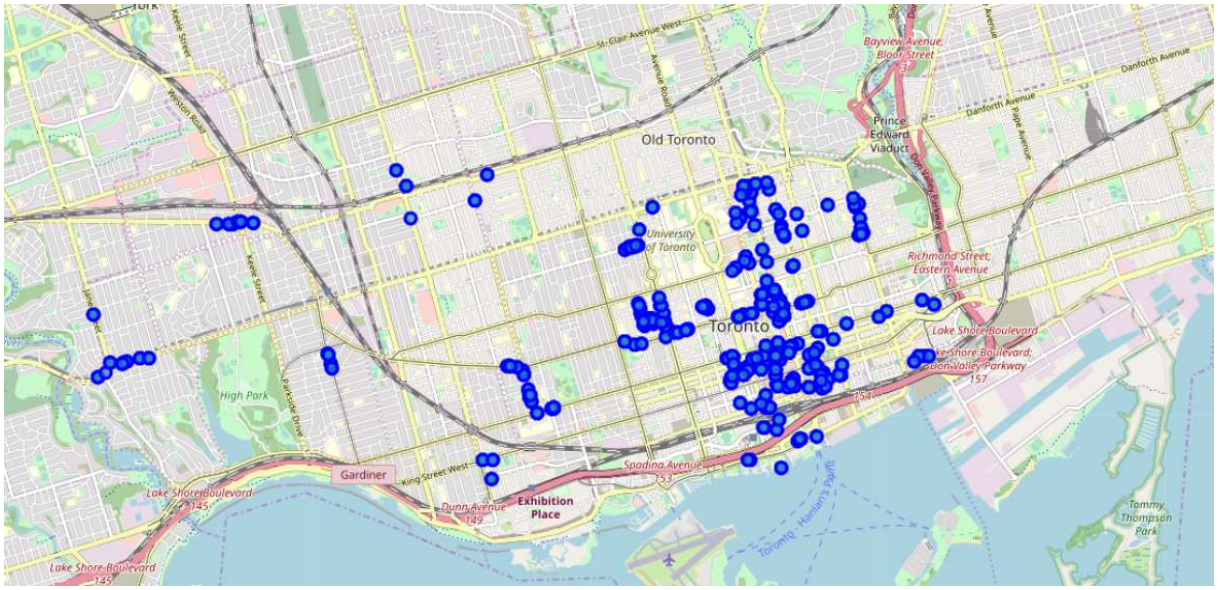

Fig. 6 : Map of Toronto with restaurants in West and Downtown

From the map above, we can see that most of the restaurants are clustered around the downtown part of Toronto(about the eastern side of the map), although we see few clusters on the west. This should be expected since the population of downtown is about 60,000 more than of the west. However, despite the population difference, the restaurant density in downtown is still considerably greater.

One of our interest areas is to look for an area with fewer restaurants, so we can prevent/reduce competition. To do this, I extracted and visualized the top 20 neighborhoods with the highest number of restaurants as shown in the figure below ;

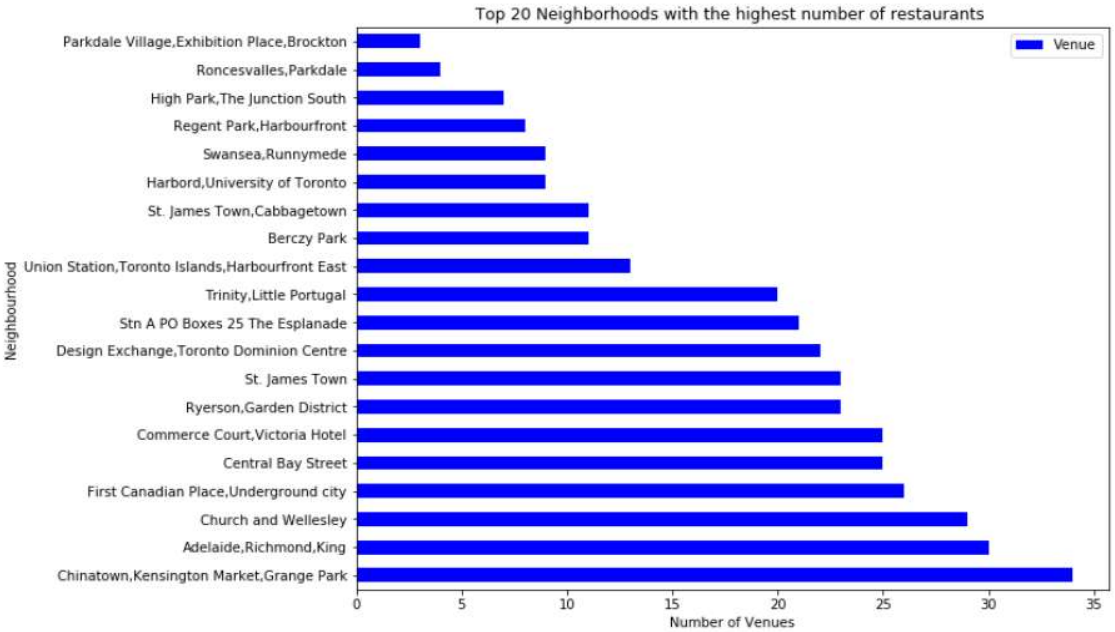

Fig. 7: Top 20 neighborhoods grouped by the number of retaurants

The result shows that Chinatown neighborhoods have the highest number of restaurants with a value of about 34 while Parkdale village neighborhoods are the least with about 4 restaurants. Since population of the neighborhood is also a significant factor to consider before opening a restaurant, it is necessary to take into account areas with considerable population in addition to fewer restaurants. In this analysis, I will choose the Toronto Islands. One reason for this is because it has fewer restaurants and a busy place. I realized that despite having more venues within its neighborhoods, the number of restaurants is fewer. This could mean that the neighborhoods are a busy type and have high viability for business.
To also determine the most common restaurants in the neighborhoods, let's consider the horizontal bar graph below ;

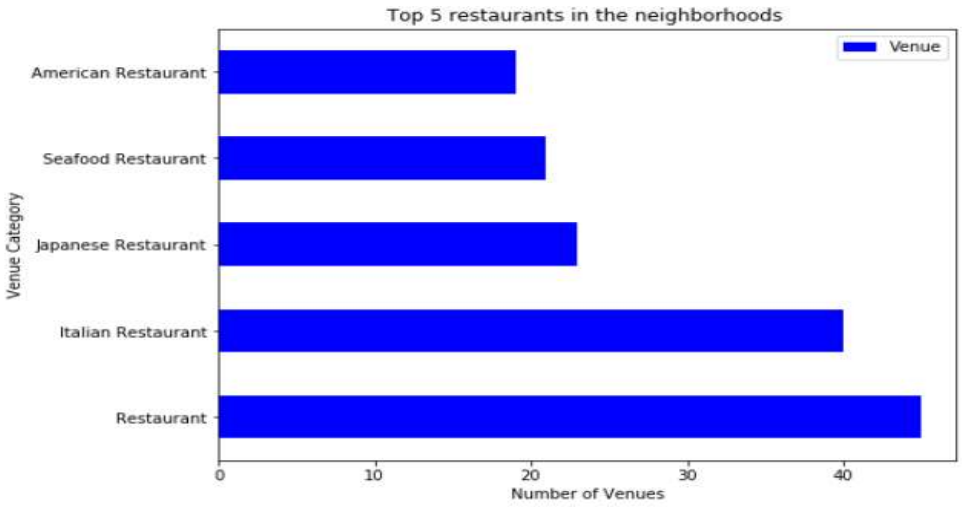

Fig. 8 : Top 5 restaurants in the neighborhood

The chart above shows that Italian restaurants have the greatest number of traffics. Although we saw "Restaurant" with the greatest number of traffic, I will neglect that because our data do not specify the type of restaurant it is i.e. I will consider it as a general-purpose restaurant and not for a niche of customers. We would explore further each neighborhood to determine its top 10 venues. We perform a one-hot encoding on our datasets, group the resulting data frame by Neighborhoods and computing the mean of occurrence for each category. We can then iterate over each neighborhood to get the result as shown below;

Out[314]:

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide,Richmond,King	Coffee Shop	Café	American Restaurant	Thai Restaurant	Bar	Steakhouse	Burger Joint	Hotel	Restaurant	Asian Restaurant
1	Bathurst Quay,CN Tower,Railway Lands,South Nia...	Airport Service	Airport Lounge	Airport Terminal	Boutique	Sculpture Garden	Plane	Boat or Ferry	Bar	Harbor / Marina	Airport Gate
2	Berczy Park	Coffee Shop	Cocktail Bar	Cheese Shop	Bakery	Steakhouse	Beer Bar	Seafood Restaurant	Café	Farmers Market	Jazz Club
3	Central Bay Street	Coffee Shop	Italian Restaurant	Café	Sandwich Place	Burger Joint	Ice Cream Shop	Gym / Fitness Center	Salad Place	Chinese Restaurant	Bar
4	Chinatown,Kensington Market,Grange Park	Café	Chinese Restaurant	Vegetarian / Vegan Restaurant	Bar	Mexican Restaurant	Vietnamese Restaurant	Coffee Shop	Dumpling Restaurant	Bakery	Donut Shop

Fig. 9: Ranking of the top 10 venues in some neighborhoods

For simplicity, we can visualize the top 5 venues for some of the neighborhood as shown below;

```
----Adelaide,Richmond,King----
                venue  freq
0         Coffee Shop  0.07
1                Café  0.05
2     Thai Restaurant  0.04
3                 Bar  0.04
4  American Restaurant  0.04


----Bathurst Quay,CN Tower,Railway Lands,
              venue  freq
0   Airport Service  0.18
1    Airport Lounge  0.12
2  Airport Terminal  0.12
3          Boutique  0.06
4           Airport  0.06


----Trinity,Little Portugal----
               venue  freq
0                Bar  0.10
1        Coffee Shop  0.06
2        Men's Store  0.06
3  Asian Restaurant  0.04
4        Pizza Place  0.03


----Union Station,Toronto Islands,
              venue  freq
0     Coffee Shop  0.13
1           Hotel  0.05
2        Aquarium  0.05
3           Café  0.04
4  Scenic Lookout  0.03
```

```
----Berczy Park----
              venue  freq
0      Coffee Shop  0.07
1     Cocktail Bar  0.05
2  Farmers Market  0.04
3       Steakhouse  0.04
4         Beer Bar  0.04


----Central Bay Street----
                venue  freq
0         Coffee Shop  0.15
1  Italian Restaurant  0.07
2               Café  0.05
3      Ice Cream Shop  0.03
4        Burger Joint  0.03


----Stn A PO Boxes 25 The Esplanade-
          venue  freq
0  Coffee Shop  0.11
1         Café  0.04
2   Restaurant  0.04
3     Beer Bar  0.03
4        Hotel  0.03


----Swansea,Runnymede----
                venue  freq
0               Café  0.09
1         Coffee Shop  0.09
2    Sushi Restaurant  0.06
3         Pizza Place  0.06
4  Italian Restaurant  0.06
```

Fig. 10: Top 5 venues for some restaurants in the neighborhood

From the result, it was obvious that our suspected place for a restaurant, Toronto Islands has no restaurant in its top 5 venues. This supports the fact that the probability for competition will be less if a **new restaurant** is brought to the neighborhood.

Finally, I performed the k-means clustering to segment the neighborhoods into 5 clusters and visualize on the Toronto map as shown in Fig . 11 below

Fig. 11 : Neighborhood Clusters

From the figure above, we can see that some neighborhoods in west Toronto form clusters with some neighborhoods in downtown. This shows that they share some similarities which could also be seen by direct comparison of the top 10 venues from each neighborhood

4. Results and Discussion

Problem1: **Comparing Neighborhoods in Downtown Toronto and West Toronto**

The map in Fig.11 showed that some part of Downtown Toronto has common features to that of West Toronto. We can see that from the above map that neighborhoods with red markers have similar features and so, form a cluster. Hence, from the result,It can be said that our neighborhood of interest, Toronto Islands is in the same clusters as Parkdale village, Trinity, Little Portugal, and others.Perhaps Tunde (the stakeholder) might not want to go that far from where he currently stays, he can choose to move to Parkdale or Trinity which seems to be the closest similar neighborhood in distance.

Problem 2: Recommended location for a restaurant

From Fig. 7, We can see than Chinatown, Kensington Market and Grange Park neighborhood altogether have the highest number of restaurants. I would choose Union Station, Toronto Islands, Harbourfront East as the area I would like to explore for the restaurant opportunities. One of the reasons is that it has fewer restaurants and also a busy place (see Fig. 5 and 7). We could see that despite having more venues within the neighborhoods, restaurants are fewer. This could mean that the neighborhoods is a busy type and have high viability for business. The chart in Fig. 8 also shows that there are more Italian restaurants than others in the neighborhoods. Apart from the first " Restaurant" which has no specified demographics, other restaurants are made to serve delicacies for some specific set of people. This doesn't rule out that some Americans might also go to Italian restaurants and so on.Also, among the top 10 most common venues for each neighborhood, every other neighborhood has more than one restaurant on their list except Toronto Islands which shows that there would be less competition for a new restaurant(See Fig. 9). Hence, within the scope of my report, I have

found Toronto Island neighborhood as a viable place for opening a restaurant.We can further explore the wellbeing, spending power, and perhaps the income of the people in the neighborhood. As written above, we can also get the demographics of the people which can help us to refine our restaurant type, perhaps, it could be a general-purpose restaurant with some Italian, Chinese or American meals available on request.

5. Conclusion

The first part of the project was to identify a similar neighborhood for a friend who intends to leave Toronto Islands in downtown Toronto to seek a similar environment. I was able to identify a neighborhood with a similar feature by k-means clustering. With the stakeholder also considering a little densely populated neighborhood, I considered the next-in-line city after downtown Toronto in terms of population and further did an exploratory analysis of its neighborhoods. Also, the project is intended to recommend an optimal location for someone hoping to open a restaurant in any of the neighborhoods in either West or Downtown Toronto. I considered the top 20 neighborhoods with the highest number of restaurants. Also, I tried to explore further to determine the top 5 restaurants in the neighborhoods. I realized a distinct neighborhood location with an opportunity, because despite having more venues that most neighborhoods, there are fewer restaurants. Based on my result, I suggested that the stakeholder should consider a general-purpose restaurant and not the one made for demographics. I believe that will increase the restaurant's chance of exploiting consumer opportunities. Finally, while it's beyond the scope of my project, I believe having additional information like population demographics, wellbeing, income per household & spending power of the people in the neighborhood would have a great impact on the project

References

Toronto, C. O. (n.d.). *World Rankings for Toronto.* Retrieved October 14, 2019, from www.toronto.ca: https://www.toronto.ca/city-government/data-research-maps/toronto-progress-portal/world-rankings-for-toronto/