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 liveable 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 client wants to shift to unknown city in Toronto and require help to find the best neighbourhood to start a restaurant. My client wants a neighbourhood that is like his current borough because he feels that would help him settle easily.

1.2. Target

I will analyse and recommend a place in any of the overall neighbourhoods for my client 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 neighbourhood 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

Population Information: https://www12.statcan.gc.ca/censusrecensement/2016/dppd/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 the below Figure was an Html data scraped from Wikipedia. After the dataset was read into a data frame, it was realized that some Boroughs and Neighbourhoods are not assigned.

```
[2]: #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[2]:

	Postcode	Borough	Neighborhood
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

In the data cleaning process, I first selected the rows with an assigned Borough, appended multiple neighbourhoods with the same postal codes and replaced unassigned neighbourhoods 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;

ıt[14]:

|	Borough	Population
0	Central Toronto	164502.0
1	Downtown Toronto	271196.0
2	East Toronto	113966.0
3	West Toronto	212782.0

The result above shows that the top two most populated Borough in Toronto is Downtown Toronto and West Toronto with a population of 27,1196 and 21,2782 people respectively. Since our stakeholder is interested in not only a similar neighbourhood 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 below figure has the Neighbourhood s and geocoordinate of west and downtown Toronto.

```
In [18]:  # 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[18]:

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

Toronto map with neighbourhood s in West and Downtown Toronto
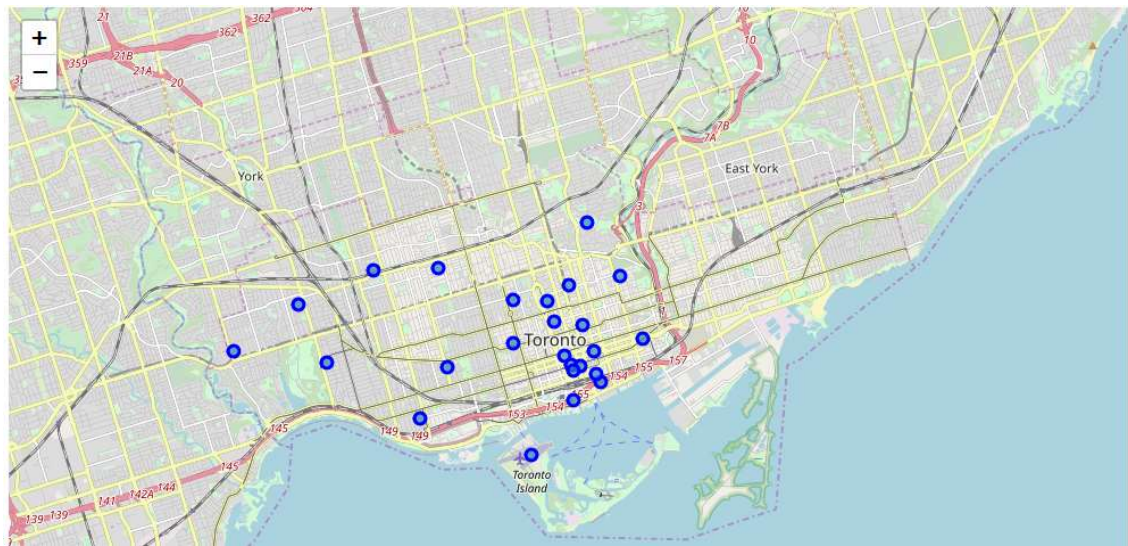
3. Exploratory Data Analysis

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

```
#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
```

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

Another great idea is to find all venues that are restaurants in the neighbourhood s. 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 neighbourhood. We can visualize the restaurants on a Toronto map as displayed below;
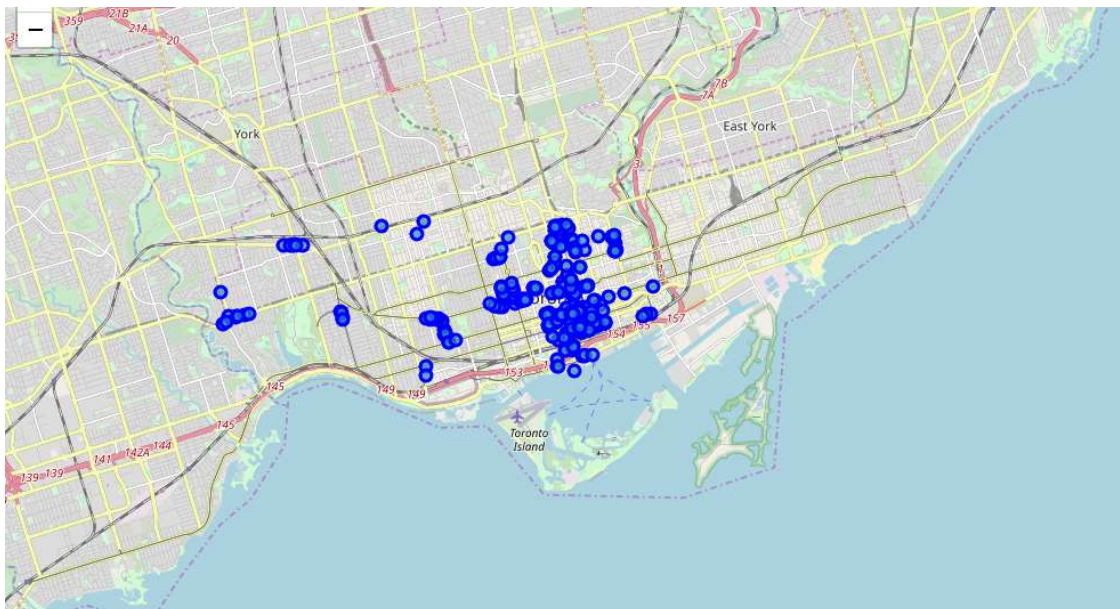

From the map above, we can see that most of the restaurants are clustered around the downtown part of Toronto (about the eastern side of the map), although we see few clusters on the west. This should be expected since the population of downtown is about 60,000 more than of the west. However, despite 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 neighbourhood s with the highest number of restaurants as shown in the figure below;

Top 20 Neighborhoods with the highest number of restaurants
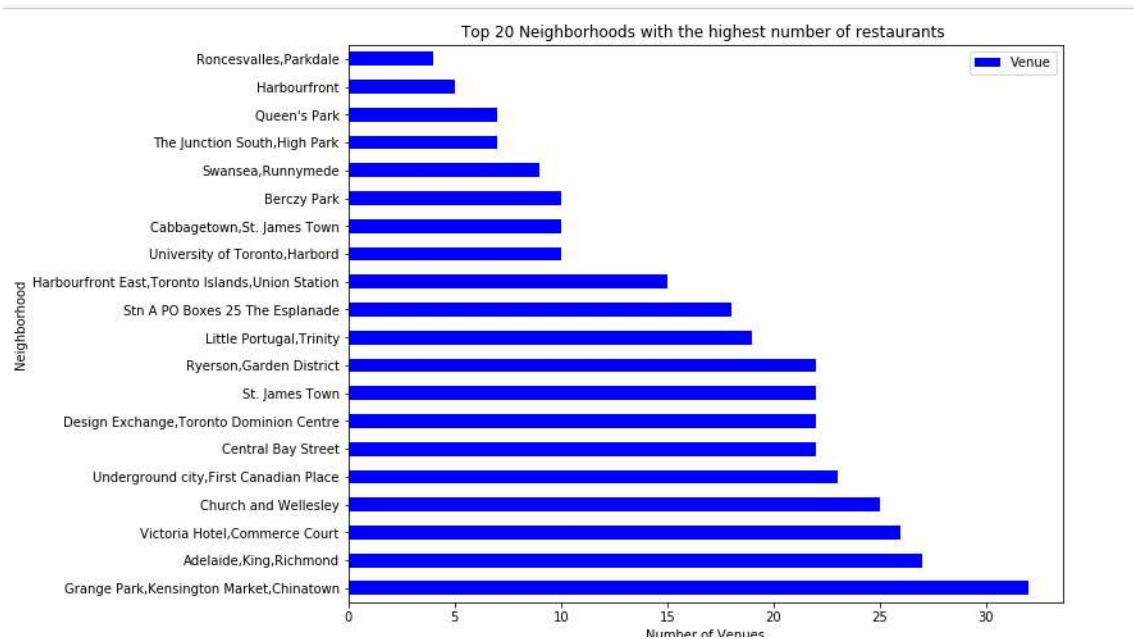
The result shows that Chinatown neighbourhood s have the highest number of restaurants with a value of about 34 while Parkdale village neighbourhood s are the least with about 4 restaurants. Since population of the neighbourhood is also a significant factor to consider before opening a restaurant, it is necessary to consider 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 neighbourhood s, the number of restaurants is fewer. This could mean that the neighbourhood is a busy type and have high viability for business. To also determine the most common restaurants in the neighbourhood s, let's consider the horizontal bar graph below;

Top 5 restaurants in the neighborhoods
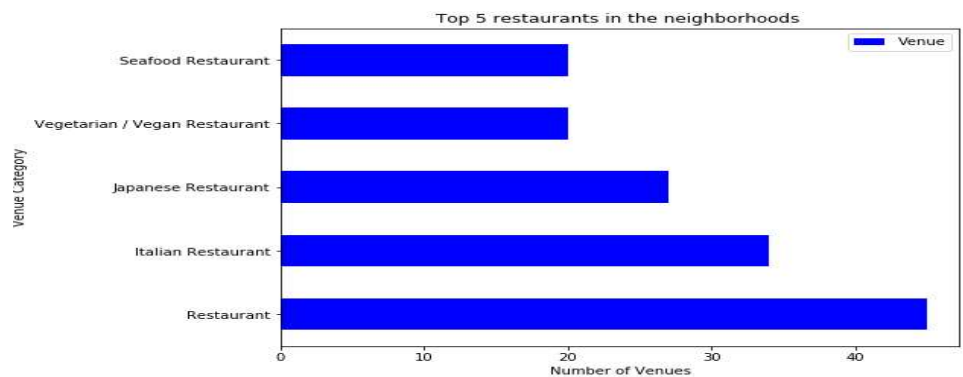
Top 5 restaurants in the neighbourhood. The chart above shows that Italian restaurants have the greatest number of traffics. Although we saw "Restaurant" with the greatest number of traffic, I will neglect that because our data do not specify the type of restaurant it is i.e. I will consider it as a

general-purpose restaurant and not for a niche of customers. We would explore further each neighbourhood to determine its top 10 venues. We perform a one-hot encoding on our datasets, group the resulting data frame by Neighbourhood s and computing the mean of occurrence for each category. We can then iterate over each neighbourhood to get the result as shown below;

```
: # add clustering labels
  neighborhoods_venues_sorted.head(5)
```

	Neighborhood	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,King,Richmond	Coffee Shop	Café	Bar	Steakhouse	Restaurant	Asian Restaurant	Hotel	Sushi Restaurant	Thai Restaurant	Burger Joint
1	Berczy Park	Coffee Shop	Cocktail Bar	Farmers Market	Beer Bar	Seafood Restaurant	Bakery	Steakhouse	Cheese Shop	Café	Butcher
2	Brockton,Parkdale Village,Exhibition Place	Café	Breakfast Spot	Coffee Shop	Gym	Grocery Store	Pet Store	Performing Arts Venue	Nightclub	Italian Restaurant	Intersection
3	Cabbagetown,St. James Town	Coffee Shop	Restaurant	Pub	Park	Pizza Place	Bakery	Café	Italian Restaurant	Sandwich Place	Breakfast Spot
4	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Burger Joint	Japanese Restaurant	Sandwich Place	Ice Cream Shop	Bar	Gym / Fitness Center	Bakery

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

```
----Adelaide,King,Richmond----
          venue  freq
0   Coffee Shop  0.08
1    Steakhouse  0.04
2           Bar  0.04
3          Café  0.04
4         Hotel  0.03


----Berczy Park----
                venue  freq
0         Coffee Shop  0.09
1        Cocktail Bar  0.05
2          Steakhouse  0.04
3  Seafood Restaurant  0.04
4      Farmers Market  0.04


----Brockton,Parkdale Village,Exhibition Place----
             venue  freq
0             Café  0.12
1   Breakfast Spot  0.08
2      Coffee Shop  0.08
3              Bar  0.04
4    Burrito Place  0.04
```

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 neighbourhood. Finally, I performed the k-means clustering to segment the neighbourhood s into 5 clusters and visualize on the Toronto map as shown below

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

4. Results and Discussion

   Recommended location for a restaurant from the above listed figures, we can see than Chinatown, Kensington Market and Grange Park neighbourhood 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 a busy place. We could see that despite having more venues within the neighbourhood s, restaurants are fewer. This could mean that the neighbourhood s is a busy type and have high viability for business. The chart also shows that there are more Italian restaurants than others in the neighbourhoods. 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 neighbourhood, every other neighbourhood has more than one restaurant on their list except Toronto Islands which shows that there would be less competition for a new restaurant. Hence, within the scope of my report, I have

We can further explore the wellbeing, spending power, and perhaps the income of the people in the neighbourhood. 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

I was able to identify a neighbourhood with a similar feature by k-means clustering. With the stakeholder also considering a little densely populated neighbourhood, I considered the next-in-line city after downtown Toronto in terms of population and further did an exploratory analysis of its neighbourhoods. The project is intended to recommend an optimal location for my client hoping to open a restaurant in any of the neighbourhood s in either West or Downtown Toronto. I considered the top 20 neighbourhood s with the highest number of restaurants. Also, I tried to explore further to determine the top 5 restaurants in the neighbourhood s. I realized a distinct neighbourhood location with an opportunity, because despite having more venues the most neighbourhood s, 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 neighbourhood would have a great impact on the project.