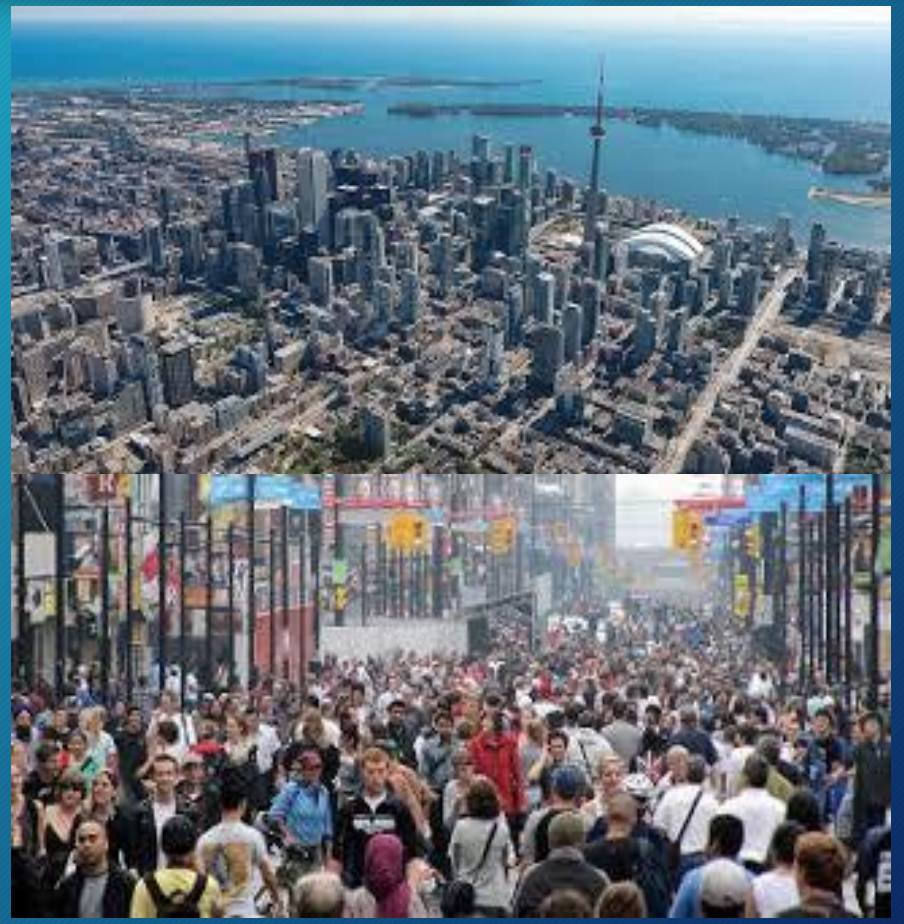
The Battle of the Neighborhoods : Report

Applied Data Science Capstone by IBM/Coursera

Jaminton Moreno

Introduction: Business Problem

- Toronto is the most populous city in Canada, with a population of 2.9 million in 2017. Most are within the Greater Toronto Area (GTA). Toronto is the fastest growing city in North America and is the anchor of an urban agglomeration, known as the Golden Horseshoe in southern Ontario.
- Toronto encompasses a geographic area formerly administered by many separate municipalities. Each of these townships has developed a distinct history and identity over the years, and their names remain in common use among Toronto residents. Former boroughs include East York, Etobicoke, Forest Hill, Mimico, North York, Parkdale, Scarborough, Swansea, Weston, and York. Throughout the city there are hundreds of small neighborhoods and some larger neighborhoods that cover a few square kilometers.



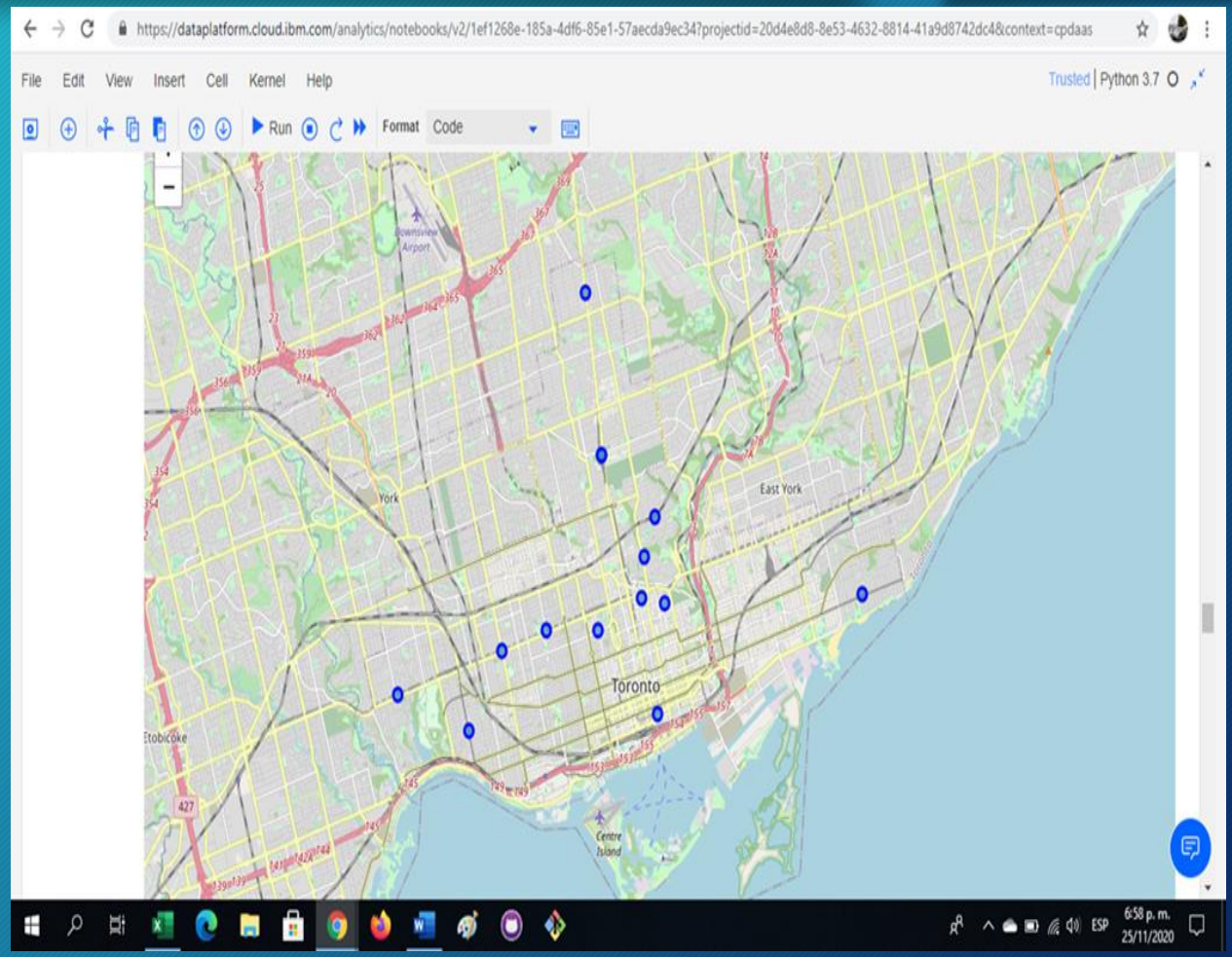
Data Acquisition

- According to the definition of our problem, the factors that will influence our decision are:
- 1. All existing Italian food restaurants.
- 2. Age group of people with income.
- 3. Distance of the neighborhood from the city center

A grid of regularly spaced locations, centered around the city center, will be used to define the neighborhoods.

The necessary data will be collected from the following source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The number of restaurants and their type and location in each neighborhood will be obtained using the Foursquare API and the area centers will be generated algorithmically.

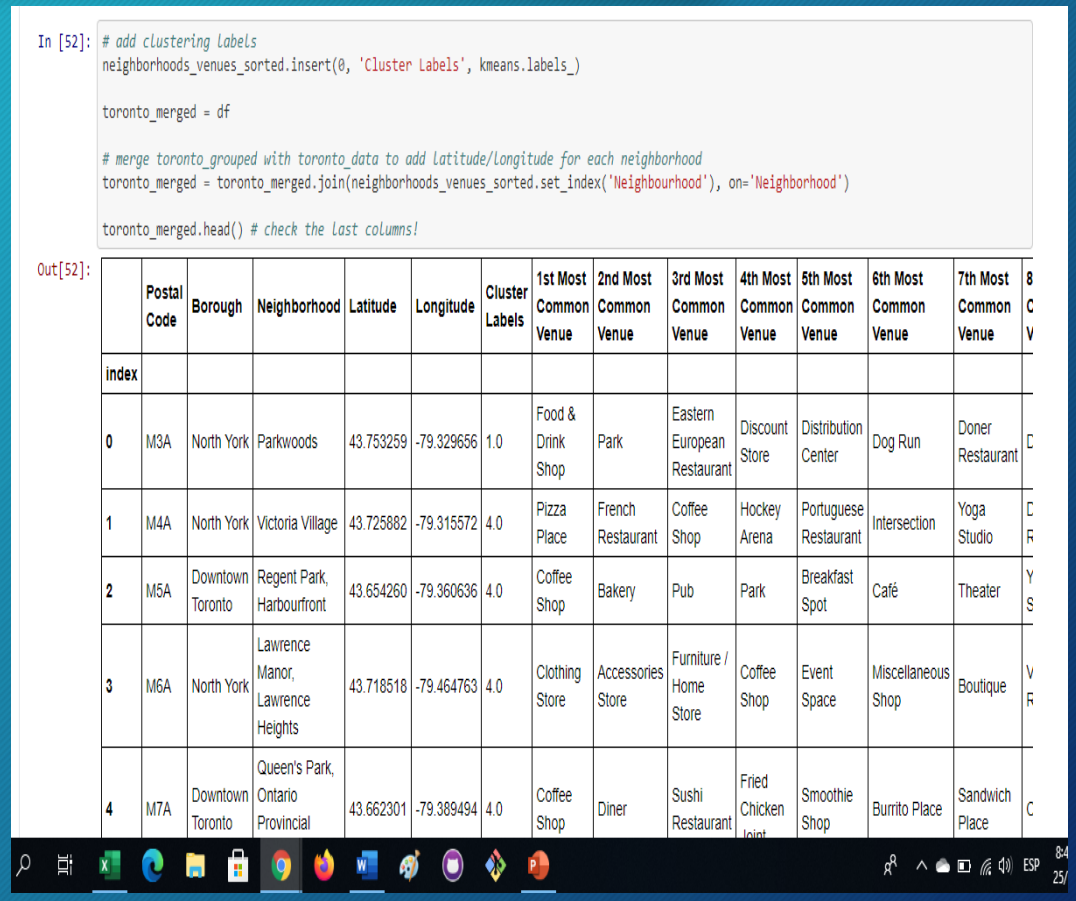


Analysis

- Identification, capture and cleaning of data.

Here we are using wikipedia to get data on Toronto, Canada. Then we remove the redundant values (data cleaning). Then we group the neighborhoods similar to the Bronx. Now the data is clean and ready to use.

After cleaning and grouping the data it is combined with the data of the neighborhood, the postal address and the Latitude - longitude. We will store those values in a separate data frame. The resulting data frame contains details about Zip Code, Brough, Neighborhood, Latitude, and Longitude. It is visualized using a folio map.



In [52]:
```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

toronto_merged = df

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
toronto_merged = toronto_merged.join(neighborhoods_venues_sorted.set_index('Neighbourhood'), on='Neighborhood')

toronto_merged.head() # check the last columns!
```

Out[52]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	8 C V
index														
0	M3A	North York	Parkwoods	43.753259	-79.329656	1.0	Food & Drink Shop	Park	Eastern European Restaurant	Discount Store	Distribution Center	Dog Run	Doner Restaurant	D
1	M4A	North York	Victoria Village	43.725882	-79.315572	4.0	Pizza Place	French Restaurant	Coffee Shop	Hockey Arena	Portuguese Restaurant	Intersection	Yoga Studio	D R
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	4.0	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Café	Theater	Y S
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	4.0	Clothing Store	Accessories Store	Furniture / Home Store	Coffee Shop	Event Space	Miscellaneous Shop	Boutique	V R
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial	43.662301	-79.389494	4.0	Coffee Shop	Diner	Sushi Restaurant	Fried Chicken Joint	Smoothie Shop	Burrito Place	Sandwich Place	C

Analysis

- Explore the neighborhoods of Toronto.

First, we scan all neighborhoods in the city of Toronto, using latitude and longitude data, using the Foursquare API to get the available restaurant locations in Toronto. Explore the unique categories in the neighborhood. Filter the details of the places for all possible 'Restaurants'. Find each neighborhood along with the most common places and identify the top 10 places for each neighborhood.

- Clustering

With an assumption of 5 clusters, use the K-Cluster algorithm to create 5 different clusters in Toronto with a similar set of locations. Explore each cluster and determine the discriminatory place categories that distinguish each cluster. Identify the groups and districts / neighborhoods with the maximum number of restaurants and their types.

Cluster 1

```
In [57]: toronto_merged_nonan.loc[toronto_merged_nonan['Cluster Labels'] == 0, toronto_merged_nonan.columns[[1] + list(range(5, toronto_merged_nonan.shape[1]))]]
```

Out[57]:

	Borough	Cluster Labels	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
index												
32	Scarborough	0.0	Playground	Yoga Studio	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Eastern European Restaurant
85	Scarborough	0.0	Park	Playground	Sculpture Garden	Donut Shop	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Drugstore

Cluster 2

```
In [58]: toronto_merged_nonan.loc[toronto_merged_nonan['Cluster Labels'] == 1, toronto_merged_nonan.columns[[1] + list(range(5, toronto_merged_nonan.shape[1]))]]
```

Out[58]:

	Borough	Cluster Labels	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
index												
0	North York	1.0	Food & Drink Shop	Park	Eastern European Restaurant	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Yoga Studio

Results and Conclusion

- Our analysis shows that although there are a large number of restaurants in Toronto, there are pockets of low-density restaurants fairly close to the city center.
- The cluster with the maximum number of restaurants will have the best chance of having a new restaurant due to the need in the area. According to the resulting groups, Group 1 and Group 5 have a greater number of restaurants than the rest of the groups.
- With these data, a more detailed investigation can be carried out in each of these sectors in order to determine the best place to open the new restaurant.