Toronto Neighborhoods & Venues

by Yamar Ba

Intro.

Suppose you have a friend who plans to move to Toronto, Canada but they want to live in an area similar to their neighborhood in downtown manhattan where there are many cafes, restaurants and shops.

Problem

Suppose your friend enlists you to help them find a neighborhood that suits their needs. Which Toronto neighborhood should they live in?

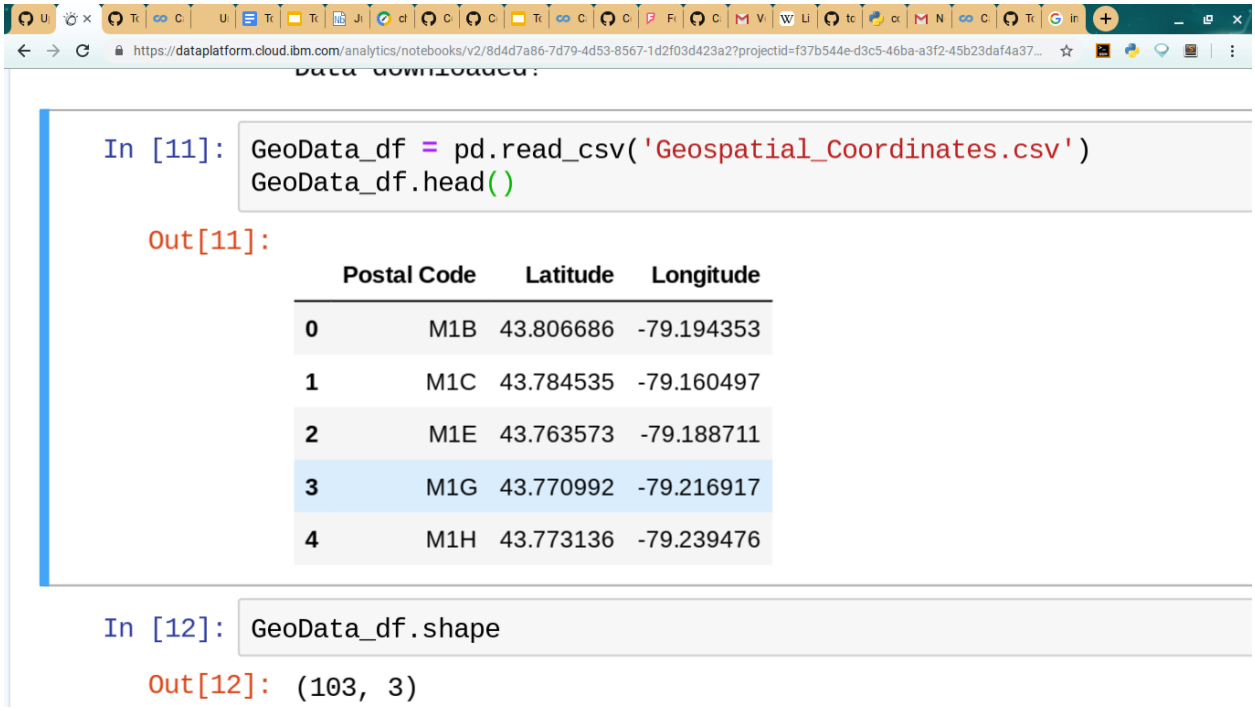
The first approach to this problem is to collect all of the data and tools to make the best possible decision.

Data/ Resources:

1) The following wikipedia page with Toronto neighborhoods categorized: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2) Geospatial coordinates CSV file, with longitude and latitude coordinates of neighborhoods categorized by zip code.
3) Foursquare software will be utilized. local search-and-discovery mobile app which provides search results for its users. The app provides personalized recommendations of places to go near a user's current location based on users' previous browsing history and check-in history.

Methodology

The data from the wikipedia page will be extracted using the software BeautifulSoup. BeautifulSoup is used to extract tables from websites. Once the table is extracted, it will be imported into our dataframe. Similarly, the geospatial coordinates CSV file along will also be added into our dataframe. The two tables have a common feature and that is they both have a column for zip codes. The two tables will be joined by zip code.



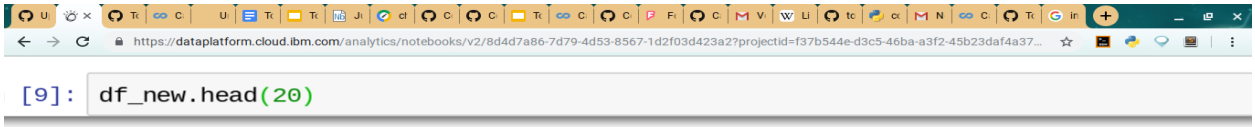
```
In [11]: GeoData_df = pd.read_csv('Geospatial_Coordinates.csv')
         GeoData_df.head()
```

Out[11]:

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

```
In [12]: GeoData_df.shape
```

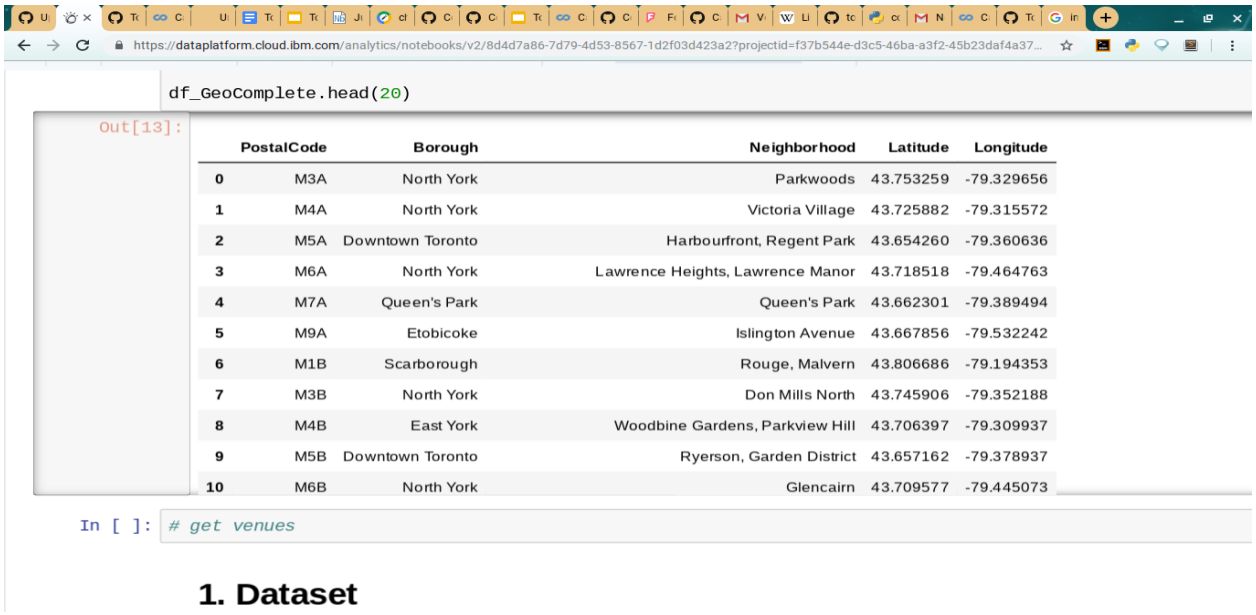
Out[12]: (103, 3)



```
[9]: df_new.head(20)
```

Out[9]:

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront, Regent Park
3	M6A	North York	Lawrence Heights, Lawrence Manor
4	M7A	Queen's Park	Queen's Park
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Rouge, Malvern
7	M3B	North York	Don Mills North
8	M4B	East York	Woodbine Gardens, Parkview Hill
9	M5B	Downtown Toronto	Ryerson, Garden District



```
df_GeoComplete.head(20)
```

Out[13]:

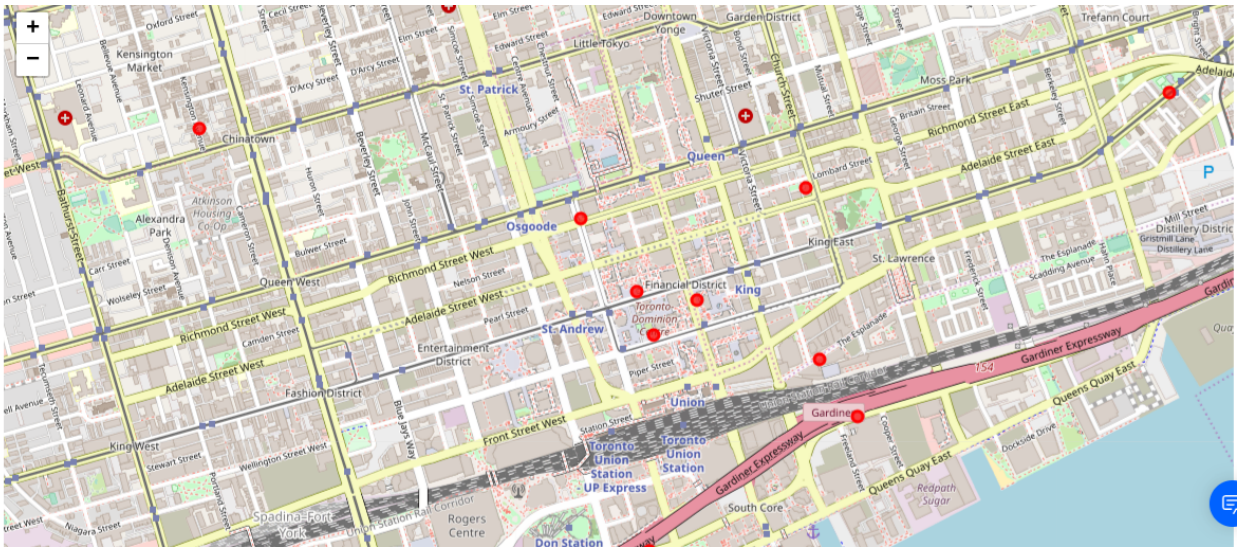
	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront, Regent Park	43.654260	-79.360636
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
6	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
7	M3B	North York	Don Mills North	43.745906	-79.352188
8	M4B	East York	Woodbine Gardens, Parkview Hill	43.706397	-79.309937
9	M5B	Downtown Toronto	Ryerson, Garden District	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073

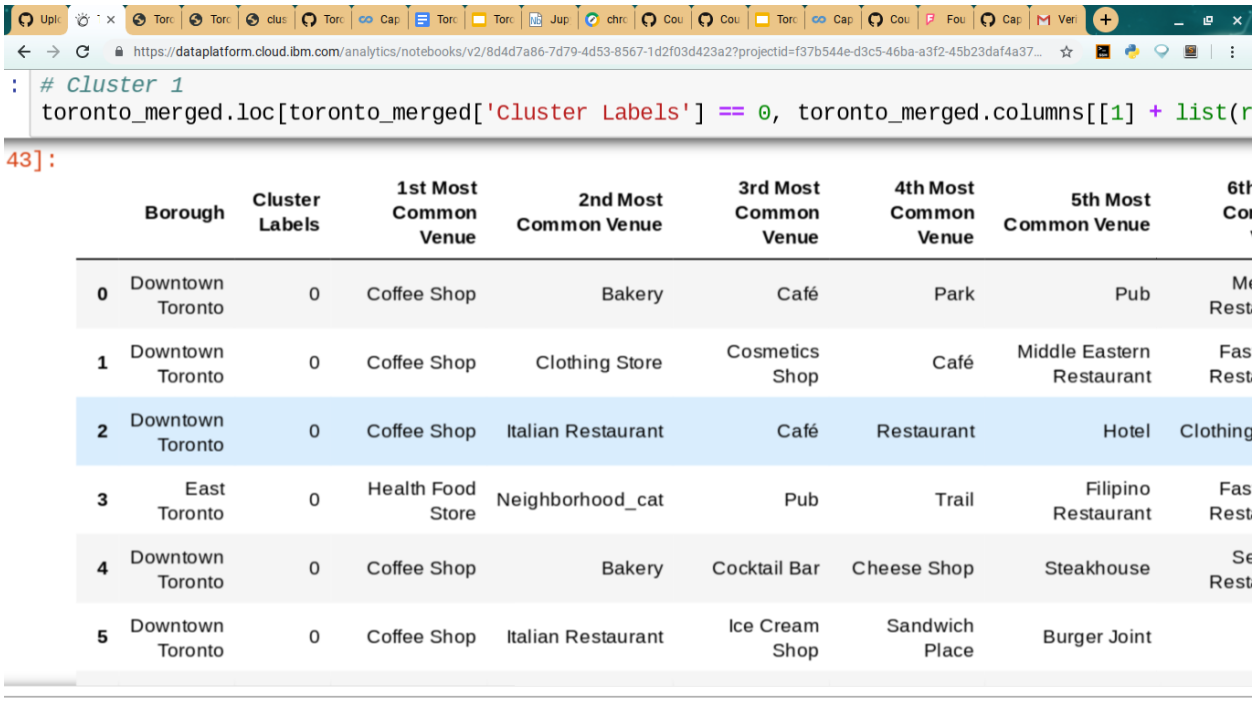
```
In [ ]: # get venues
```

1. Dataset

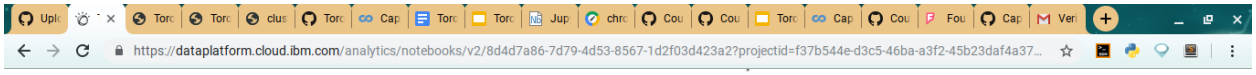
Once are dataframe is set. We then utilize Foursquare to explore Toronto neighborhoods. We want to see all the venues the Toronto area provides for our consideration. Specifically, we are looking for neighborhoods with cafes, restaurants and shops.

The dataframe is then clustered. Four clusters are created. Each cluster has all the popular venues we need in our decision making. Each cluster represents a top neighborhood.Once clustered in the dataframe, the dataframe is them visualized.





	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Co
0	Downtown Toronto	0	Coffee Shop	Bakery	Café	Park	Pub	Me Rest
1	Downtown Toronto	0	Coffee Shop	Clothing Store	Cosmetics Shop	Café	Middle Eastern Restaurant	Fas Rest
2	Downtown Toronto	0	Coffee Shop	Italian Restaurant	Café	Restaurant	Hotel	Clothing
3	East Toronto	0	Health Food Store	Neighborhood_cat	Pub	Trail	Filipino Restaurant	Fas Rest
4	Downtown Toronto	0	Coffee Shop	Bakery	Cocktail Bar	Cheese Shop	Steakhouse	Se Rest
5	Downtown Toronto	0	Coffee Shop	Italian Restaurant	Ice Cream Shop	Sandwich Place	Burger Joint	



```
# Cluster 2
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] + list(range(5,
```

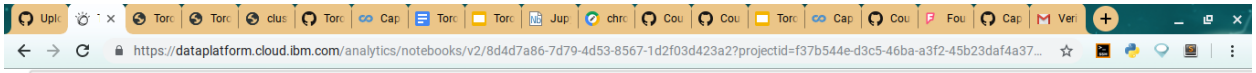
44]:

	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	
32	Downtown Toronto	1	Park	Playground	Trail	Building	Dive Bar	Fast Food Restaurant	Far

```
# Cluster 3
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5,
```

45]:

	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	
28	Central Toronto	2	Summer Camp	Yoga Studio	Flea Market	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant	Far



```
# Cluster 3
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(r
```

45]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Com V
28	Central Toronto	2	Summer Camp	Yoga Studio	Flea Market	Fish & Chips Shop	Filipino Restaurant	Fast Resta

```
# Cluster 4
toronto_merged.loc[toronto_merged['Cluster Labels'] == 3, toronto_merged.columns[[1] + list(r
```

46]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Com V
18	Central Toronto	3	Garden	Yoga Studio	Flea Market	Fish & Chips Shop	Filipino Restaurant	Fast Resta

Conclusion

The cluster with that best fits our model will represent the best neighborhood for our friend. The neighborhood that we would recommend our friend to move to in Toronto, is Downtown Toronto.