Stake Holders interested in opening restaurant in Toronto

1. Introduction:

1.1 Background:

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

Canada Post provides a free postal code look-up tool on its website, via its applications for such smartphones as the iPhone and BlackBerry, and sells hard-copy directories and CD-ROMs. Many vendors also sell validation tools, which allow customers to properly match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries.

1.2 Problem:

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening restaurant in Toronto

Since there are lots of restaurants in Berlin, we will try to detect locations that are not already crowded with restaurants.

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

2. Data acquisition:

Data is imported from the below link:
'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'

We have to do web scraping in order to get the data from the above website.

The final dataset which we will get after scraping the website is as below:

	Postal code\n	Borough\n	Neighborhood\n
0	M1A\n	Not assigned\n	\n
1	M2A\n	Not assigned\n	\n
2	M3A\n	North York\n	Parkwoods\n
3	M4A\n	North York\n	Victoria Village\n
4	M5A\n	Downtown Toronto\n	Regent Park / Harbourfront\n

3. Data Cleaning:

We could see that the data which we got from the website is not clean.

So, we have to clean the data.

First, we have to remove '\n' characters after each word.

Then, we have to rename the column names.

Remove all the rows in the column 'Borough' which is are assigned to the value 'Not Assigned'. And also, we need to check the null values in the data set and we need to manage them. And also, we need to manage the duplicate values in the dataset.

As there are 77 rows with Not Assigned value in Borough column so we will remove all the 77 rows. And also, there are no null and duplicate values in the dataset. And also we need to replace the character '/' with ','.

So, after cleaning the data we could see that there are 103 rows in the dataset.

4. Importing Geospatial Data:

Now that you have built a dataframe of the postal code of each neighborhood along with the borough name and neighborhood name, in order to utilize the Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood.

In an older version of this course, we were leveraging the Google Maps Geocoding API to get the latitude and the longitude coordinates of each neighborhood. However, recently Google started charging for their API: http://geoawesomeness.com/developers-up-in-arms-over-google-maps-api-insane-price-hike/, so we will use the Geocoder Python package instead: https://geocoder.readthedocs.io/index.html.

The problem with this Package is you have to be persistent sometimes in order to get the geographical coordinates of a given postal code. So you can make a call to get the latitude and longitude coordinates of a given postal code and the result would be None, and then make the call again and you would get the coordinates. So, in order to make sure that you get the coordinates for all of our neighborhoods, you can run a while loop for each postal code. Taking postal code M5G as an example, your code would look something like this:

Given that this package can be very unreliable, in case you are not able to get the geographical coordinates of the neighborhoods using the Geocoder package, here is a link to a csv file that has the geographical coordinates of each postal code: http://cocl.us/Geospatial_data

After importing the dataset, the data frame will looks like below:

	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

We have to merge these data frame to the mail data frame and the resulting data frame looks like:

	Postal code	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	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

5. Exploring and Clustering the neighborhoods in Toronto

We have to explore and cluster the data with only Boroughs that contain the word Toronto.
After doing the above analysis the data frame looks like below:

	Postal code	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M4E	East Toronto	The Beaches	43.676357	-79.293031

Now we need to get the geographical coordinates of Toronto.
Using geopy.geocoders from Nominatim we can find the geographical coordinate of Toronto are 43.6534817, -79.3839347.

Using folium, we should create a map of Toronto using longitude and latitude values. We need to define the Foursquare Credentials and Version.

Exploring the first neighborhood in our data frame:

We need to get the latitude and longitude of the first neighborhood Regent Park, Harbourfront are 43.6542599, -79.3606359.

After doing the analysis we will get 48 venues returned by Foursquare. And the data frame will look like below:

	name	categories	lat	lng
0	Roselle Desserts	Bakery	43.653447	-79.362017
1	Tandem Coffee	Coffee Shop	43.653559	-79.361809
2	Morning Glory Cafe	Breakfast Spot	43.653947	-79.361149
3	Cooper Koo Family YMCA	Distribution Center	43.653249	-79.358008
4	Body Blitz Spa East	Spa	43.654735	-79.359874

Exploring each neighborhood:

After analyzing, we will get 231 unique categories. After getting this we need to print each neighborhood along with top 5 most common venues.

Clustering the Neighborhoods:

We need to cluster the Neighborhoods in order to achieve the place where the stakeholders can open the restaurant.

Examining the Clusters:
We got 5 clusters from the above analysis. Now, we need to examine those 5 clusters.

Cluster1:

Cluster 1:

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

	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
4	East Toronto	0	Trail	Health Food Store	Pub	Women's Store	Dance Studio	Electronics Store	Eastern European Restaurant	Donut Shop	Doner Restaurant	Dog Run

Cluster2:

Cluster 2:

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

	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
0	Downtown Toronto	1	Coffee Shop	Park	Bakery	Pub	Breakfast Spot	Café	Theater	Mexican Restaurant	Restaurant	Chocolate Shop
1	Downtown Toronto	1	Coffee Shop	Sushi Restaurant	Diner	Yoga Studio	Burrito Place	Beer Bar	Italian Restaurant	Juice Bar	Sandwich Place	Burger Joint
2	Downtown Toronto	1	Clothing Store	Coffee Shop	Café	Cosmetics Shop	Restaurant	Japanese Restaurant	Italian Restaurant	Bubble Tea Shop	Middle Eastern Restaurant	Ramen Restaurant
3	Downtown Toronto	1	Café	Coffee Shop	Cocktail Bar	American Restaurant	Gastropub	Hotel	Restaurant	Gym	Italian Restaurant	Department Store
5	Downtown Toronto	1	Coffee Shop	Cocktail Bar	Beer Bar	Cheese Shop	Restaurant	Café	Bakery	Seafood Restaurant	Bistro	Jazz Club
6	Downtown Toronto	1	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Ice Cream Shop	Middle Eastern Restaurant	Bar	Thai Restaurant	Burger Joint	Fried Chicken Joint
7	Downtown Toronto	1	Grocery Store	Café	Park	Candy Store	Athletics & Sports	Restaurant	Italian Restaurant	Diner	Nightclub	Coffee Shop
8	Downtown Toronto	1	Coffee Shop	Café	Restaurant	Thai Restaurant	Deli / Bodega	Gym	Clothing Store	Hotel	Bakery	Steakhouse
9	West Toronto	1	Bakery	Pharmacy	Supermarket	Music Venue	Bank	Brewery	Pool	Recording Studio	Café	Middle Eastern Restaurant
10	Downtown Toronto	1	Coffee Shop	Aquarium	Hotel	Café	Restaurant	Brewery	Sporting Goods Shop	Italian Restaurant	Scenic Lookout	Fried Chicken Joint

Cluster 3:

Cluster 3:

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

	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
18	Central Toronto	2	Swim School	Park	Construction & Landscaping	Bus Line	Women's Store	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Donut Shop	Doner Restaurant
21	Central Toronto	2	Park	Jewelry Store	Trail	Bus Line	Sushi Restaurant	Deli / Bodega	Electronics Store	Eastern European Restaurant	Donut Shop	Doner Restaurant
33	Downtown Toronto	2	Park	Playground	Trail	Women's Store	Dance Studio	Electronics Store	Eastern European Restaurant	Donut Shop	Doner Restaurant	Dog Run

Cluster 4:

Cluster 4:

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

	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
19	Central Toronto	3	Garden	Women's Store	Deli / Bodega	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Donut Shop	Doner Restaurant	Dog Run

Cluster 5:

Cluster 5:

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 4, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	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
29	Central Toronto	4	Park	Women's Store	Deli / Bodega	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center

Results and Conclusion:

As per the above results, the stakeholders can open the restaurant in Cluster1, Cluster 4 and Cluster 5.