2. Data Acquisition and cleaning:

2.1 Data Acquisition

The data acquired for this project is a combination of data from three sources.

The first data source of the project uses a Toronto crime data that shows the crime per Neighborhood in Toronto this data is from the Toronto Polish Department Data http://data.torontopolice.on.ca/datasets/mci-2014-to-2019. I made changes to remove the useless data. The dataset contains the following columns:

- Neighborhood: the common name of Toronto.
- Population : the population of each neighborhood
- AVG : the average of crimes for each category (between 2014 and 2019)
- Total: the total average of crimes of all categories.

The second source of data is scraped from a Wikipedia page that contains the list of Toronto Neighborhoods. The columns are : Borough and Neighborhood.

- Borough: The names of the 10 Boroughs of Toronto.
- Neighborhood: the names of Neighborhood of each Borough.

The third source of data is the one that we used in the last assignment which contains the list of all the Neighborhoods, Boroughs and their Latitude and Longitude.

2.2 Data Cleaning

The data preparation for each of the three sources of data is done separately.

1/- From the Toronto crime data, the crimes during the years (2014-2019) are selected.

	Neighbourhood	Hood_ID	Population	Assault_AVG	AutoTheft_AVG	BreakandEnter_AVG	Homicide_AVG	Robbery_AVG	TheftOver_AVG	total
0	Waterfront Communities-The Island	77	65913	851.8	53.7	247.3	1.0	82.2	56.2	1292.2
1	Bay Street Corridor	76	25797	771.0	32.8	158.7	1.5	121.3	52.3	1137.6
2	Church-Yonge Corridor	75	31340	642.8	37.8	188.5	2.0	135.7	33.8	1040.6
3	West Humber-Clairville	1	33312	301.8	366.7	137.8	1.5	91.8	52.2	951.8
4	Moss Park	73	20506	474.7	30.2	148.5	2.5	125.5	18.8	800.2
5	York University Heights	27	27593	333.2	106.3	113.2	0.8	75.8	36.3	665.6
6	Downsview-Roding-CFB	26	35052	395.8	107.8	78.8	1.3	64.7	15.2	663.6
7	Kensington-Chinatown	78	17945	368.2	27.5	150.8	1.5	64.0	26.7	638.7
8	Woburn	137	53485	384.7	46.0	105.2	1.2	83.5	13.7	634.3
9	West Hill	136	27392	402.0	26.5	82.5	0.8	65.2	6.7	583.7

2/-The second Data contains the list of Boroughs and their Neighborhoods.
It is scraped from Wikipedia page (https://en.wikipedia.org/wiki/List_of_city-designated_neighbourhoods_in_Toronto) using the Beautiful soup library in Python. Using this library, we can extract the data in the table as shown in the webpage.

After the web scraping, string manipulation is required to get the names of boroughs in the correct form. It is important because we will merge the two datasets using the Neighborhood names.

	Borough	Neighbourhood
0	Scarborough	Agincourt
1	Scarborough	Agincourt
2	Etobicoke	Alderwood
3	Old City of Toronto	The Annex
4	North York	Don Mills
5	North York	Bathurst Manor
6	Old City of Toronto	Bay Street
7	North York	Bayview Village
8	North York	Bayview Woods
9	North York	Bedford Park

The two datasets are merged on the Neighborhood names to form a new dataset that combines the necessary information in one dataset.
The purpose is to visualize the crimes rates in each Borough with the average crime number from 2014 to 2019.

	Borough	Population	Assault_AVG	AutoTheft_AVG	BreakandEnter_AVG	Homicide_AVG	Robbery_AVG	TheftOver_AVG	total
0	Scarborough	259450	1635.9	306.2	584.5	5.6	354.4	70.3	2956.9
1	North York	162689	811.6	263.7	315.2	3.1	157.8	63.2	1614.6
2	Etobicoke	89067	550.1	112.1	193.6	2.2	71.9	31.5	961.4
3	Old City of Toronto	82751	527.2	76.5	233.8	2.1	74.5	33.2	947.3
4	York	34803	225.0	40.2	50.5	1.5	48.1	5.2	370.5
5	East York	27699	138.6	12.9	57.9	0.9	23.1	6.9	240.3

After visualizing the crime in each borough we can find the one with the lowest crime number and hence tag that borough as the safest one.

3/-The third source of data is acquired from the list of neighborhoods that we worked with in WEEK03 assignment (Wikipedia).
This dataset is created from scratch, the pandas data frame is created with the names of neighborhood and Boroughs, latitude and longitude.

	Postal Code	Borough	Neighbourhood
0	M1B	Scarborough	Malvern
1	M1B	Scarborough	Rouge
2	M1C	Scarborough	Rouge Hill
3	M1C	Scarborough	Port Union
4	M1C	Scarborough	Highland Creek
5	M1E	Scarborough	Guildwood
6	M1E	Scarborough	Morningside
7	M1E	Scarborough	West Hill
8	M1G	Scarborough	Woburn
9	M1H	Scarborough	Cedarbrae
10	M1J	Scarborough	Scarborough Village

The coordinates of the neighborhoods is to be obtained using **Google Maps API Geocoding** to get the final dataset.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern	43.806686	-79.194353
1	M1B	Scarborough	Rouge	43.806686	-79.194353
2	M1C	Scarborough	Rouge Hill	43.784535	-79.160497
3	M1C	Scarborough	Port Union	43.784535	-79.160497
4	M1C	Scarborough	Highland Creek	43.784535	-79.160497
5	M1E	Scarborough	Guildwood	43.763573	-79.188711
6	M1E	Scarborough	Morningside	43.763573	-79.188711
7	M1E	Scarborough	West Hill	43.763573	-79.188711
8	M1G	Scarborough	Woburn	43.770992	-79.216917
9	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

The new dataset is used to generate the venues for each neighborhood using the Foursquare API.