FINAL CAPSTONE PROJECT REPORT.

Introduction:

The problem solved in this course of this project was extracting data from Toronto open data evolve round using open data available to create a guide list of dedicated routes around various Neighborhood with Toronto as we will demonstrate our findings and share the outcomes.

Data Source: https://ckan0.cf.opendata.inter.prod-toronto.ca/api/3/action/package_show

Method of Data Extraction

Using Python Matplotlib function to extract the data into the dataframe, we are able to commence the process of data cleaning.

Figure 1: Shows the Jpeg sample view of the data formation.

Processing and Cleaning of Data

```
In [3]: df.head()
```

Out[3]:

	ADDRESSNUMBERTEXT	ADDRESSSTREET	ASSETTYPE	BIA	DIRECTION	FROMSTREET	FRONTINGSTREET	ID	LATITUDE	LONGITUDE	OI
0	70	The Pond Rd	Rack	None	None	Seneca Lane	None	BP-11699	43.770770	-79.499834	4
1	8	Assiniboine Rd	Rack	None	None	Nelson Rd	None	BP-11900	43.768166	-79.504212	9
2	8	Kensington Ave	Ring	Kensington Market	None	Kensington Ave	None	BP-03501	43.653058	-79.400127	66
3	21	Canniff St	Ring	None	None	Strachan Ave	None	BP-12883	43.641451	-79.411493	10
4	87	Avenue Rd	Ring	None	North	Elgin Ave	Avenue Rd	BP-15966	43.672000	-79.395336	17

Further Data extrapolation:

At this point more information to establish, needed data is required.

Using NumPy and pandas, we are able to redefine the data…

```
In [6]: df_1.iloc[:10]
```

Out[6]:

	_id	X	Y	ADDRESSSTREET	LONGITUDE	LATITUDE
0	5500530	304813.343	4847867.805	The Pond Rd	-79.499834	43.770770
1	5500531	304460.910	4847578.496	Assiniboine Rd	-79.504212	43.768166
2	5500532	312856.125	4834795.334	Kensington Ave	-79.400127	43.653058
3	5500533	311940.672	4833504.872	Canniff St	-79.411493	43.641451
4	5500534	313239.951	4836900.226	Avenue Rd	-79.395336	43.672000
5	5500535	312876.928	4834998.675	Baldwin St	-79.399866	43.654888
6	5500536	313503.184	4833936.254	Blue Jays Way	-79.392120	43.645318
7	5500537	309323.837	4835314.509	Bloor St W	-79.443913	43.657761
8	5500538	313505.863	4835081.661	Mc Caul St	-79.392068	43.655628
9	5500539	309633.259	4835449.198	Bloor St W	-79.440076	43.658971

```
In [7]: df_2 =df_1[['X', 'Y']]
```

Data Set is Ready foe use:

These datasets will consist of the street address, Serial Id, Longitude and Latitude etc.

Here is some information

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Section 2:

Methodology:

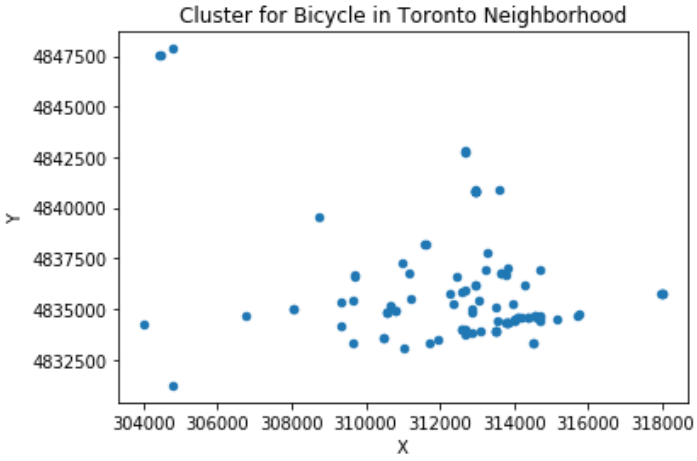
Testing the clusters using Scatter plot graph

Since, the project seeks to establish dedicated routes for Bicycle around the neighborhood which wasn't known but only in data format.

The Various Longitude and Latitude points are tested, we are able to identify the various clusters for bicycle with the Neighborhood in Toronto.

```
In [9]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        df_2 = pd.DataFrame(data, columns=['x', 'y'])
        df.plot.scatter(x='X',y='Y',title = 'Cluster for Bicycle in Toronto Neighborhood')
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7fbff86316a0>
```




```
In [10]: !pip install -U scikit-learn
```

From the display we have an understanding of the pictures on the various routes on the Map,

To employ these locations on the Map, we need to invoke Folium lib using python to display the locations.

MAP:

Maps derived in the course of this project generated using Folium Libs

The importance of the map is to aid visual associated with the cluster points.

As you can observe that the scatter plot graph share similarity with the cluster points found on the Map.

This is another way to acertaining the veracity of the plot and points.

The Map gives a good geographical mapping of the locations and makes it easy for accessibility.

This report has establish the relationship between the cluster points on the map and the Scatter plot graph.

MAP with cluster locations :

LIST accompanying the MAP:

	ADDRESSSTREET	LONGITUDE	LATITUDE
0	The Pond Rd	-79.499834	43.770770
1	Assiniboine Rd	-79.504212	43.768166
2	Kensington Ave	-79.400127	43.653058
3	Canniff St	-79.411493	43.641451
4	Avenue Rd	-79.395336	43.672000
5	Baldwin St	-79.399866	43.654888
6	Blue Jays Way	-79.392120	43.645318
7	Bloor St W	-79.443913	43.657761
8	Mc Caul St	-79.392068	43.655628
9	Bloor St W	-79.440076	43.658971
10	Davenport Rd	-79.423442	43.675403
11	Lombard St	-79.377324	43.651417
12	Yonge St	-79.388269	43.672935
13	Western Battery Rd	-79.414028	43.639938
14	King St E	-79.364787	43.651970
15	Bloor St W	-79.388866	43.669758
16	Dewson St	-79.427117	43.656299
17	Mount Pleasant Rd	-79.391024	43.708137
18	Bathurst St	-79.402566	43.645471
19	Cottingham St	-79.394941	43.680088
20	Victoria St	-79.377482	43.652090
21	Elizabeth St	-79.386163	43.657374
22	Queen St W	-79.440030	43.639981

In Conclusion:

The information extracted from the data can serve as a guide to new visitor or migrant to this neighborhood.

More so, the list accomplishing this data can aid planning on further expansion by the city district council for more dedicated routes.

This report has establish the relationship between the cluster points on the map and the Scatter plot graph.