ASSESSING THE NEIGHBORHOOD WITH LESSER PIZZA VENUE IN TORONTO

By Ebrima Sawaneh

CONTENTS

INTRODUCTION

- The objective of this project is to help the management of Banjul Pizza Ltd decides on possible Pizza restaurants in Toronto Canada.
- The aim is to open the first restaurant in a neighborhood with lesser pizza restaurants thereby a possible lesser competition too.

DATA SOURCES

- Wikipedia - borough, neighborhood and postal code of Toronto.
- Toronto Geospace information – the geo data for all neighborhoods in Toronto with postal code.
- Foursquare location data– to collect venues that are nearby each geo location.

METHODOLOGY

- Use BeautifulSoup package and request to collect data and put into Pandas DataFrame
- The various tables are merge into one

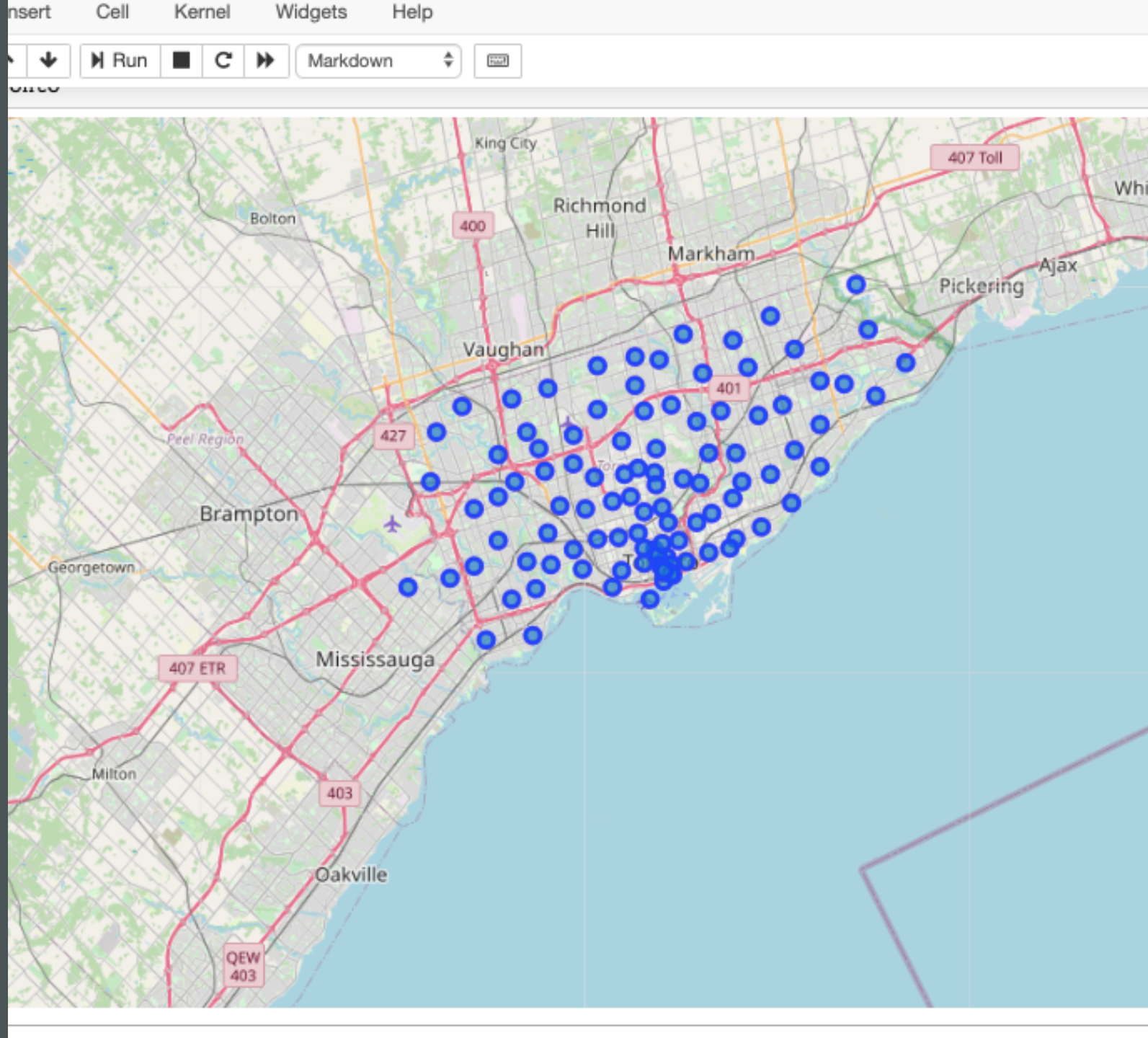
```
In [68]: # Now merging both DataFrames
torontodata = pd.merge(df2,geo,on='Postal Code')
torontodata.head()
```

Out[68]:

	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

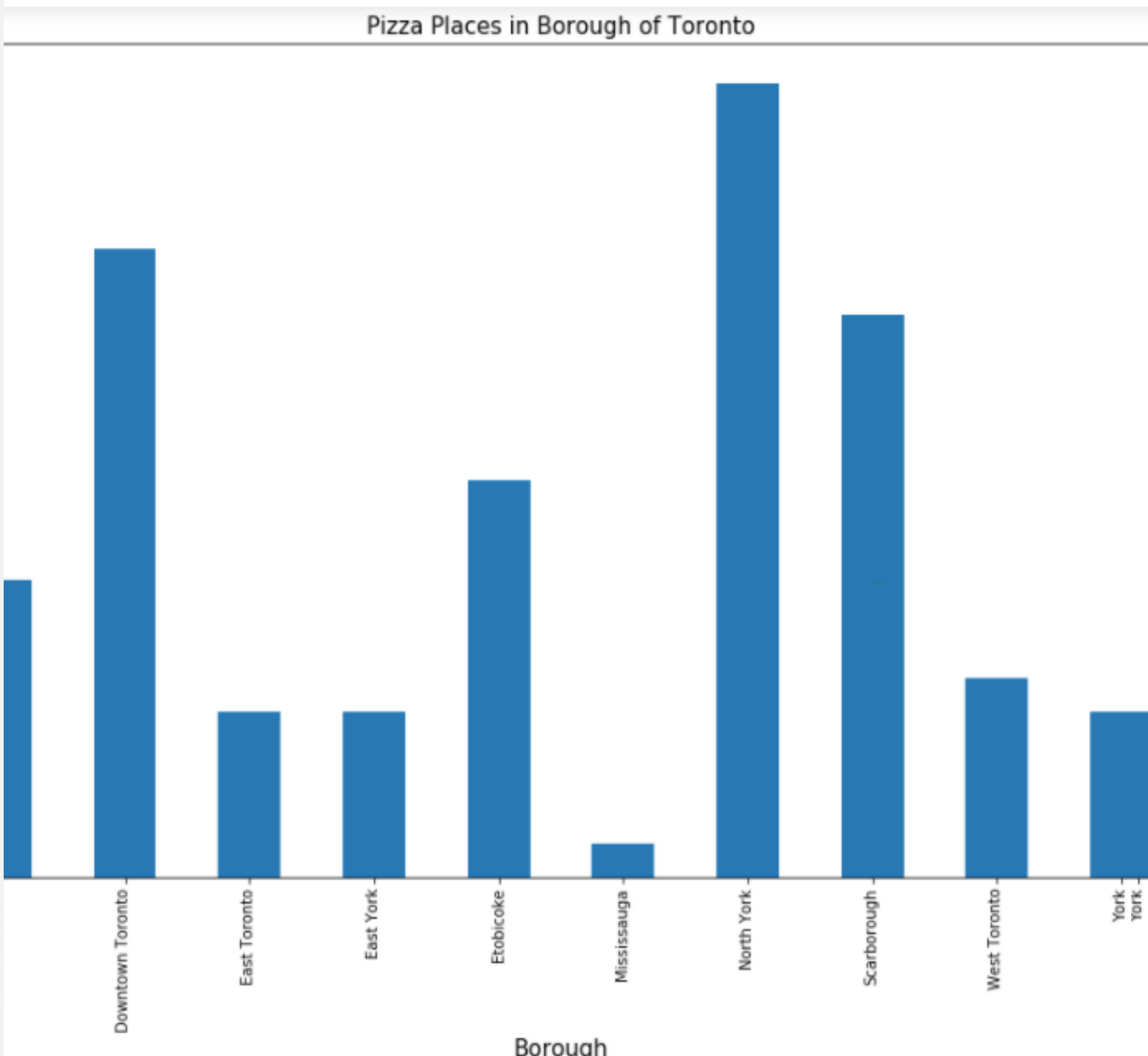
DATA EXPLORATION

- Use Folium to display the neighborhood on the map of Toronto.



DATA EXPLORATION

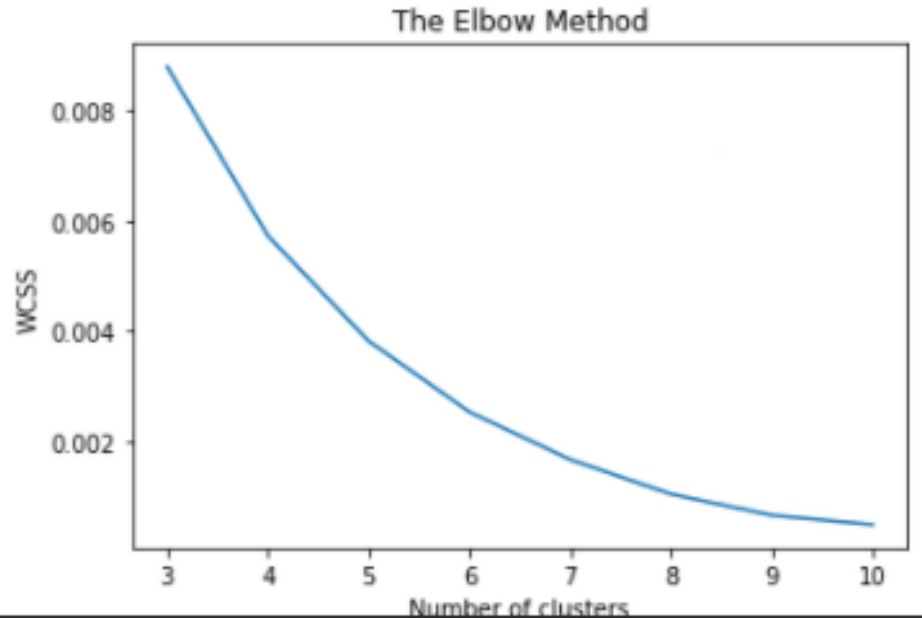
- Use Matplotlib to display pizza venue per borough.



K-MEANS CLUSTERING

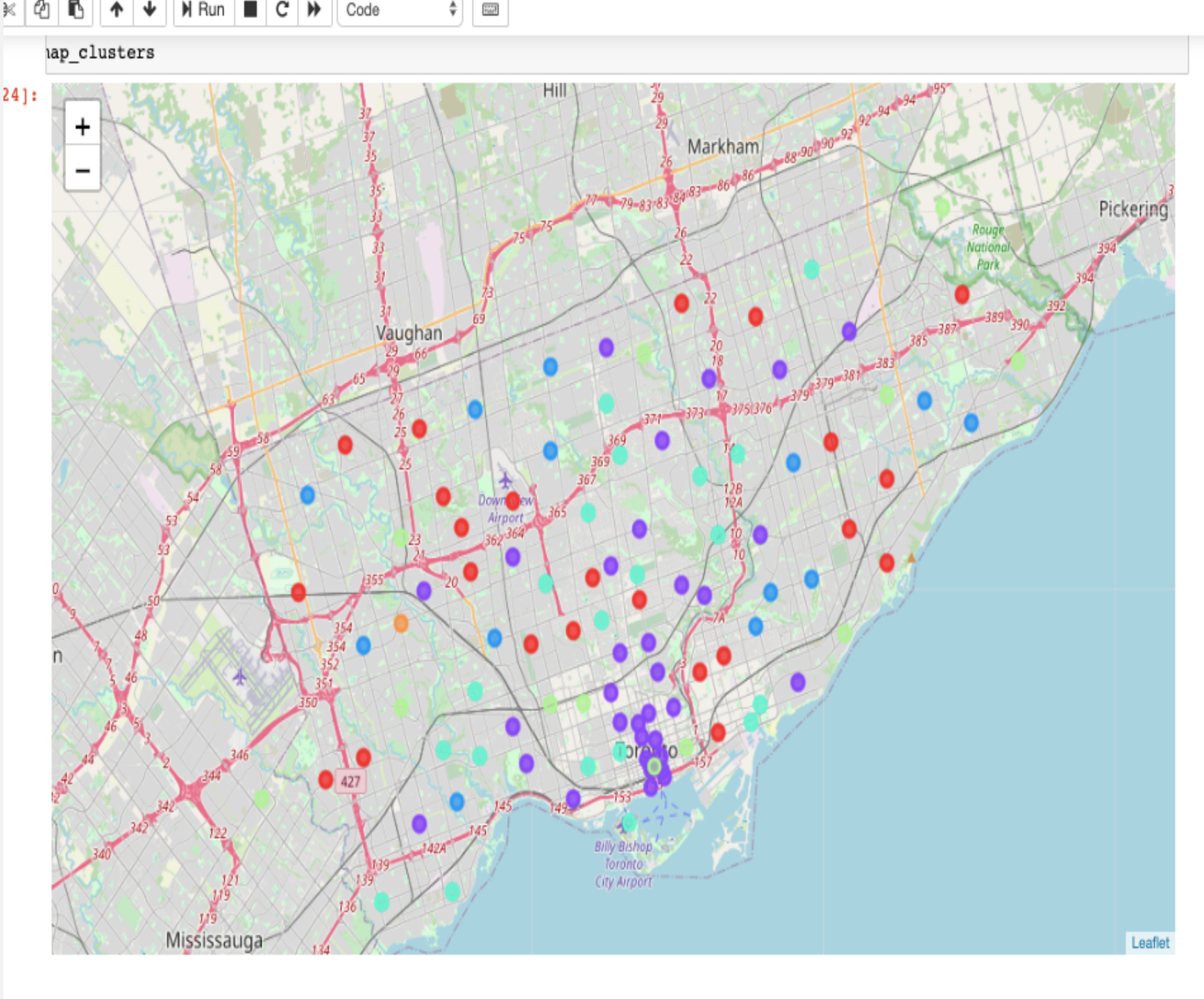
- Use elbow method to assess suitable number of clusters

```
# Matplotlib and associated plotting modules
from matplotlib import pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors

toronto_Pizza_Clustering = toronto_Pizza.drop('Neighborhood', 1)


wcss = []

for i in range(3, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', max_iter= 50)
    kmeans.fit(toronto_Pizza_Clustering)
    wcss.append(kmeans.inertia_)
plt.plot(range(3, 11), wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```



RESULTS

- Visualize clusters on the Map



Let's group the results by summing the Cluster Labels

```
In [120]: toronto_Pizza_Clustering_merged2.groupby(['Cluster Labels'])['Pizza Place'].agg('sum')
```

```
Out[120]: Cluster Labels
          0    1.206009
          1    0.467172
          2    1.020216
          3    0.656773
          4    0.000000
          5    0.121212
          Name: Pizza Place, dtype: float64
```

RESULTS

- Visualize clusters on the Map

DISCUSSION & CONCLUSION

- The K-Mean clustering result shows that Cluster 4 is the cluster which has less Pizza venues while cluster 3 has more pizza venues. I will then conclude that cluster 4 should be the firms first Pizza location.
- Further research about the boroughs within the cluster as there are other variables that may impact sales such as traffic, population, crime rate etc.

THANK YOU