IDENTIFYING DIFFERENCE BETWEEN MANHATTAN AND TORONTO

BY RAHUL S NAIR

Use Case

- **To establish dissimilarities between Manhattan and Toronto located situated miles apart based on the venues of Neighborhoods**

Feature Engineering

- **Pin Code of Neighborhood is used to generated longitude and latitude details**

- **After the usage of pin code , it is dropped**

- **Categorical Variables are converted in numeric using One Hot Encoding**

- **Similar Data are grouped to form meaningful dataset which is used for clustering**

|	Borough	Neighbourhood	Latitude	Longitude
0	Downtown Toronto	Harbourfront	43.654260	-79.360636
1	Downtown Toronto	Regent Park	43.654260	-79.360636
2	Downtown Toronto	Ryerson	43.657162	-79.378937
3	Downtown Toronto	Garden District	43.657162	-79.378937
4	Downtown Toronto	St. James Town	43.651494	-79.375418

Training Data

- **Column Borough is dropped once feature engineering is done**
- **Data is synthesized to the following format for the machine learning algorithm to learn**

|	Neighbourhood	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	Adelaide	Opera House	Asian Restaurant	Speakeasy	Steakhouse	Hotel	Plaza	Seafood Restaurant	Concert Hall	Greek Restaurant	Vegetarian / Vegan Restaurant
1	Bathurst Quay	Airport Lounge	Airport	Boutique	Harbor / Marina	Plane	Coffee Shop	Airport Food Court	Airport Gate	Airport Terminal	American Restaurant
2	Berczy Park	Beer Bar	Concert Hall	Vegetarian / Vegan Restaurant	Park	Museum	Thai Restaurant	French Restaurant	Farmers Market	Liquor Store	Steakhouse
3	Brockton	Coffee Shop	Breakfast Spot	Furniture / Home Store	Caribbean Restaurant	Café	Bar	Italian Restaurant	Pet Store	Gym	French Restaurant
4	Business Reply Mail Processing Centre 969 Eastern	Pizza Place	Auto Workshop	Restaurant	Burrito Place	Farmers Market	Skate Park	Brewery	Fast Food Restaurant	Comic Shop	Garden Center

Exploratory Data Analysis

- **Following graph shows the Neighborhoods in Toronto**
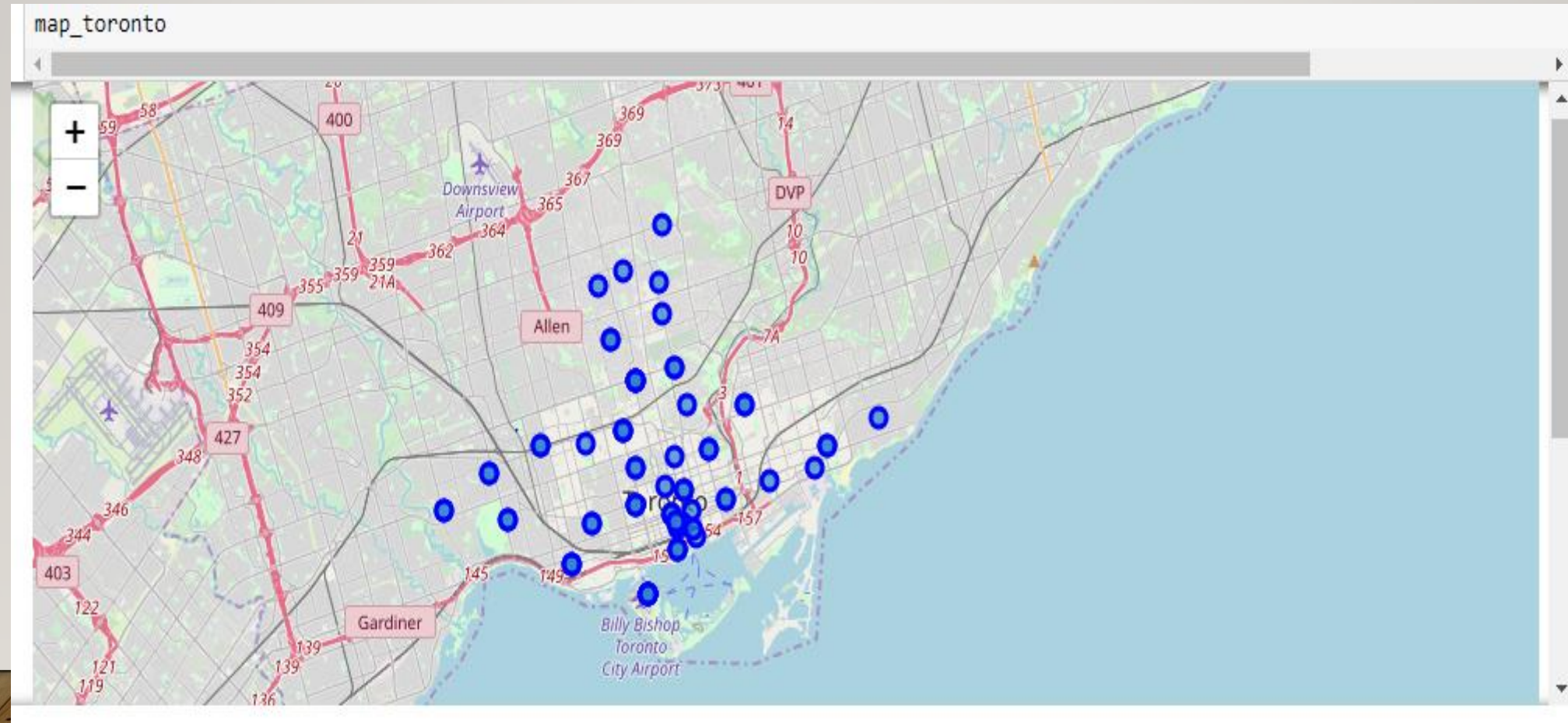
Exploratory Data Analysis
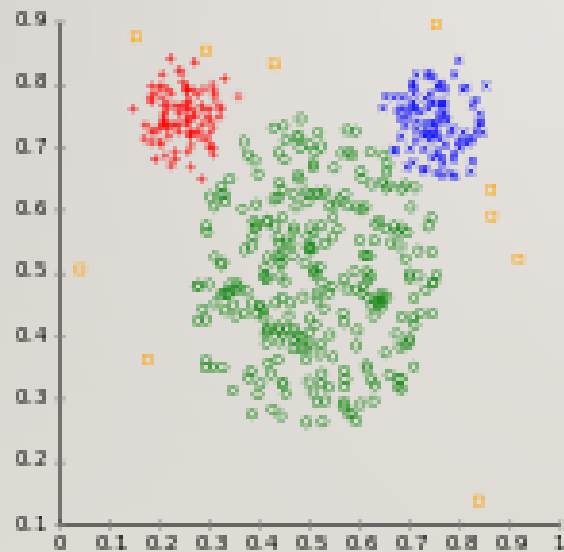
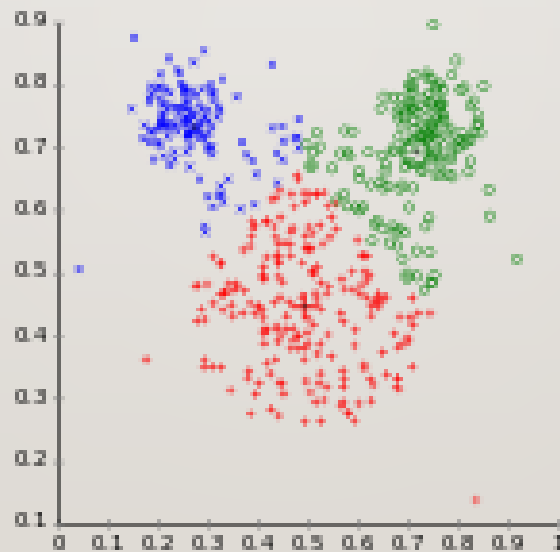- Following graph shows the Neighborhoods in Manhattan
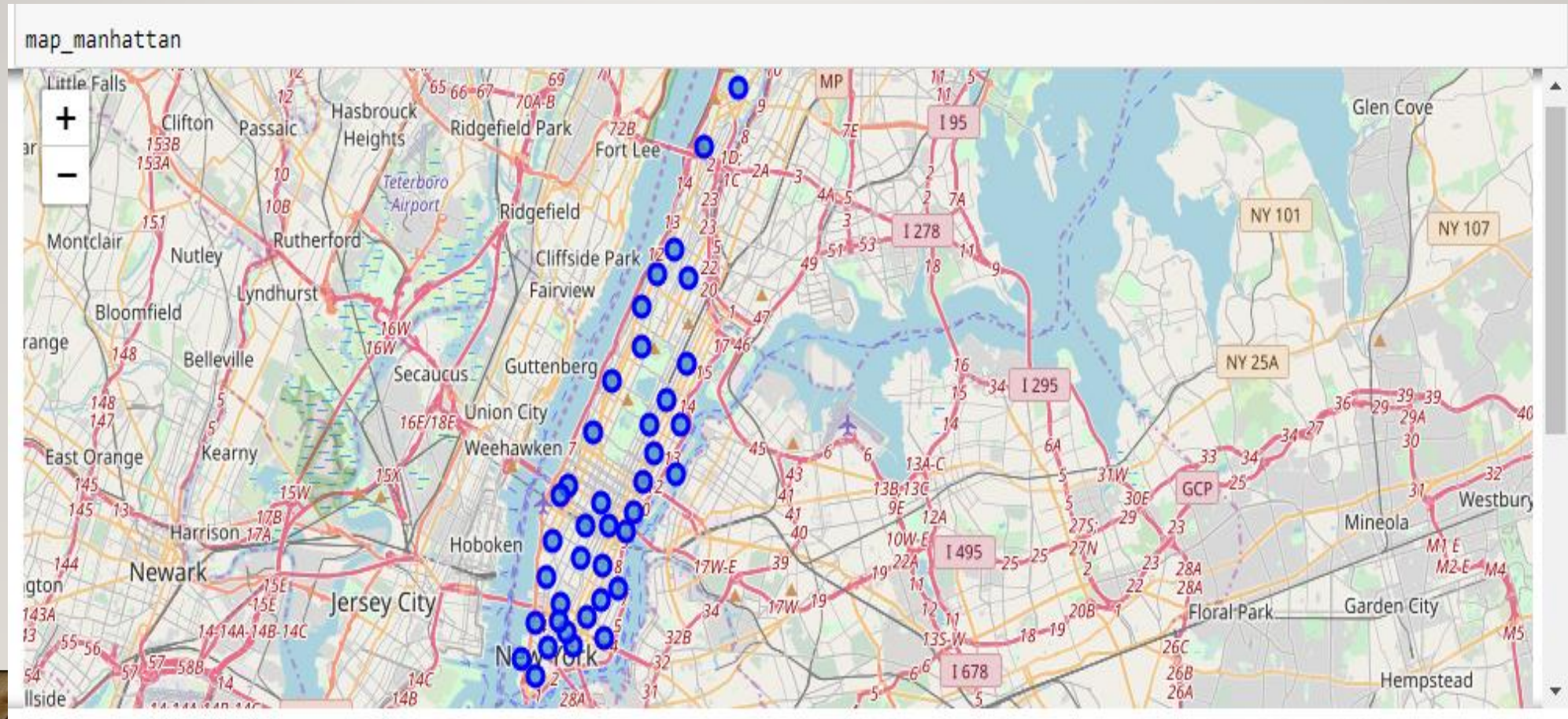
K Means Clustering

- **K Means is a un supervised learning algorithm that helps in clustering data into different clusters. Following image shows an example on how K Means works**

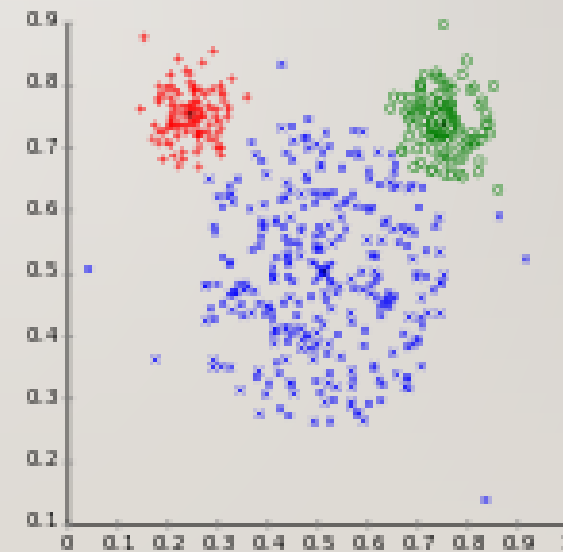Different cluster analysis results on "mouse" data set:
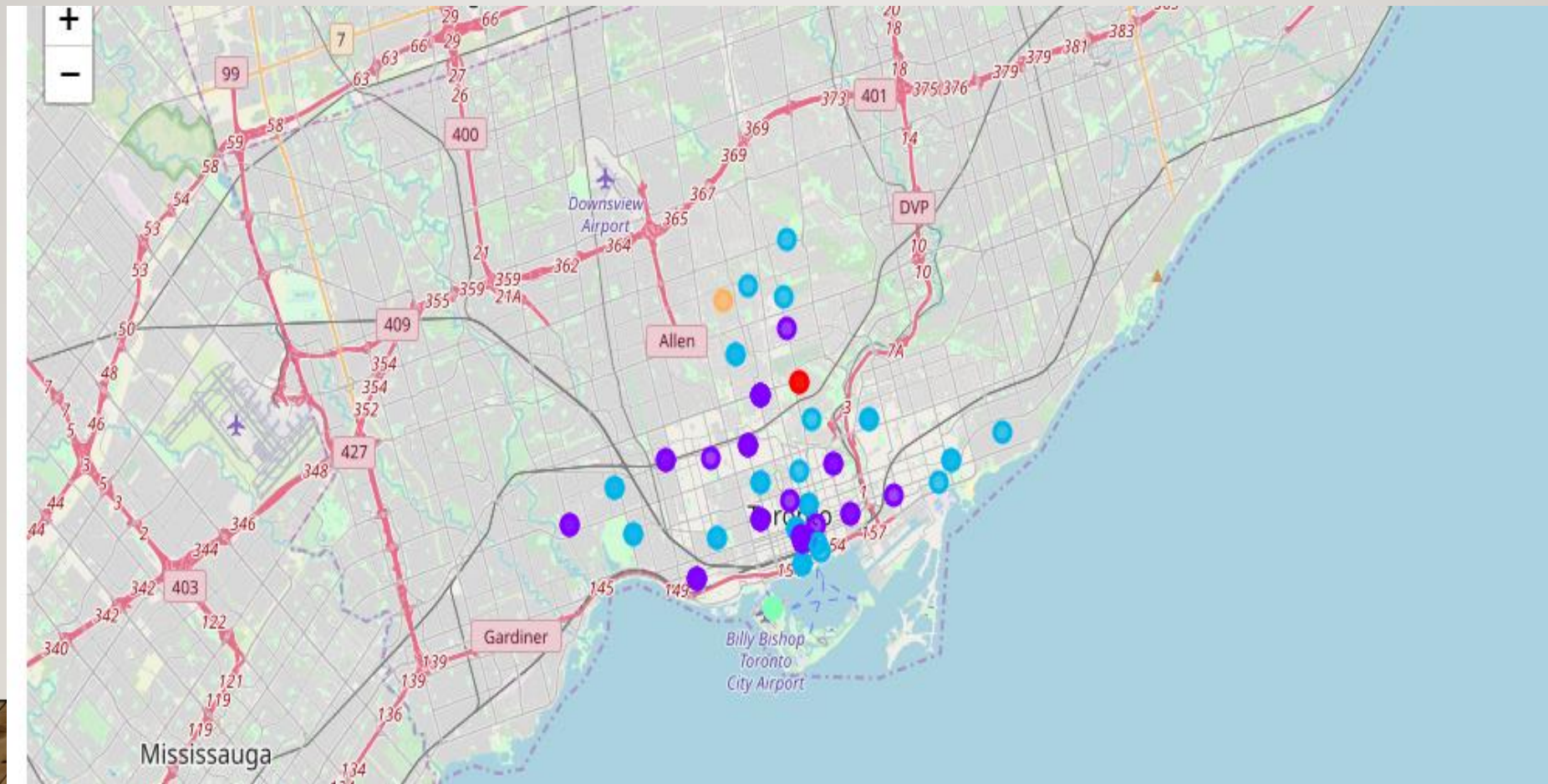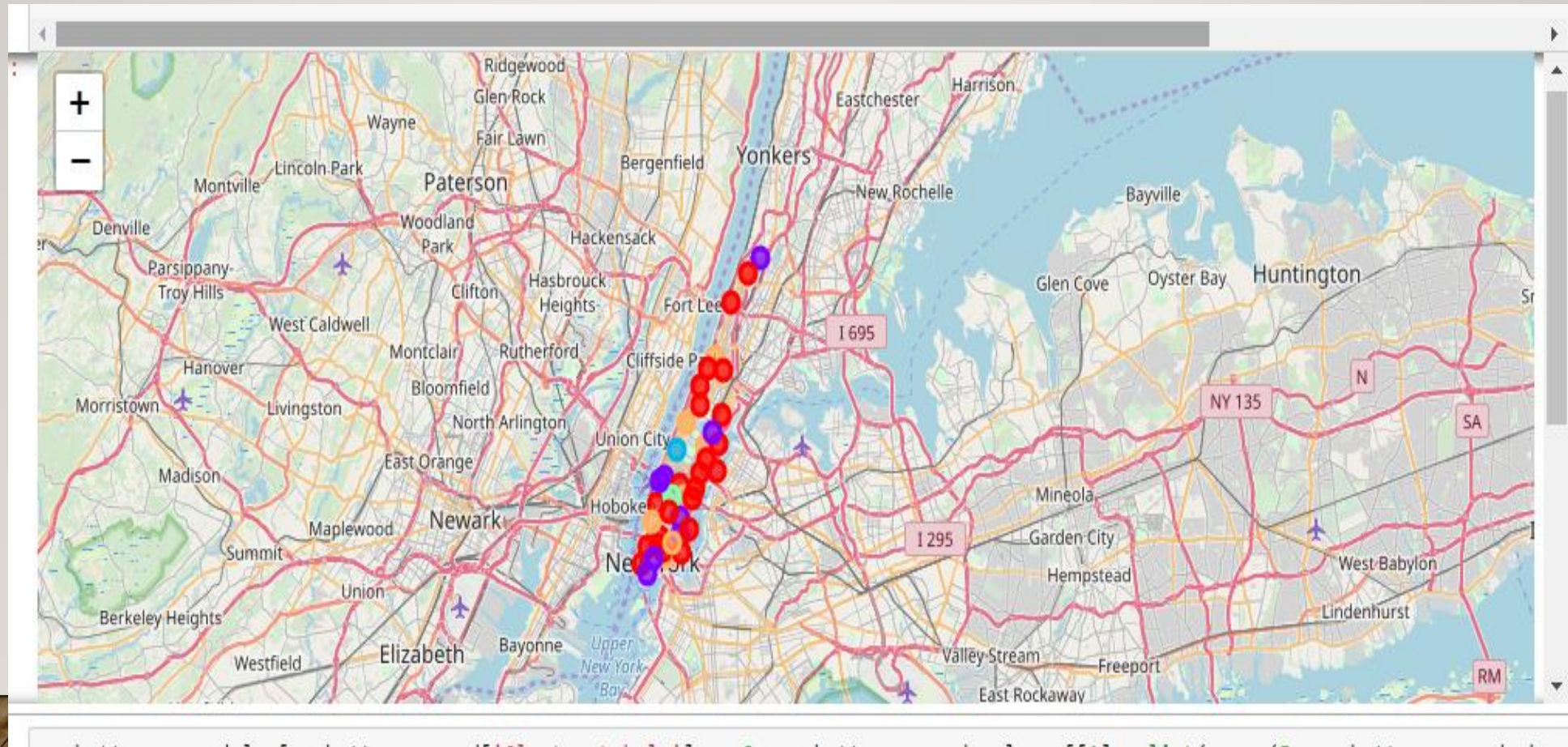Original Data · k-Means Clustering · EM Clustering

K Means on Training Data

- **Following graph shows how K Means Cluster done Neighborhood clustering based on data of Toronto venues**

K Means on Training Data

- **Following graph shows how K Means Cluster done Neighborhood clustering based on data of Manhattan venues**

Conclusion

Following information were obtained after analyzing the cluster data

Toronto

- **More number of coffee shops are present. People tend to drink more coffee**
- **People prefer entertainment venues such as comic book store, parks**
- **Ice cream shops are more in some areas**

Manhattan

- **A Cluster that shows domination on pubs are present.**
- **People prefer entertainment venues such as pubs, parks**
- **Higher number of gym/fitness centers are present**