DATA SCIENCE CAPSTONE PROJECT

BATTLE OF NEIGHBOURHOODS

DATA FOR PROJECT

The data will be obtained

1. Wikipedia(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

2. Geospatial Data(https://cocl.us/Geospatial_data)

Wikipedia Data

This data is scraped from Wikipedia and looks like:

	Postal Code	Borough	Neighborhood
0	M1A	Not assigned	NaN
1	M2A	Not assigned	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

Geospatial Data

	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

Merging two data to obtain postal codes,neighbourhoods of corresponding boroughs along with location

	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

Visualizing Neighbourhood data
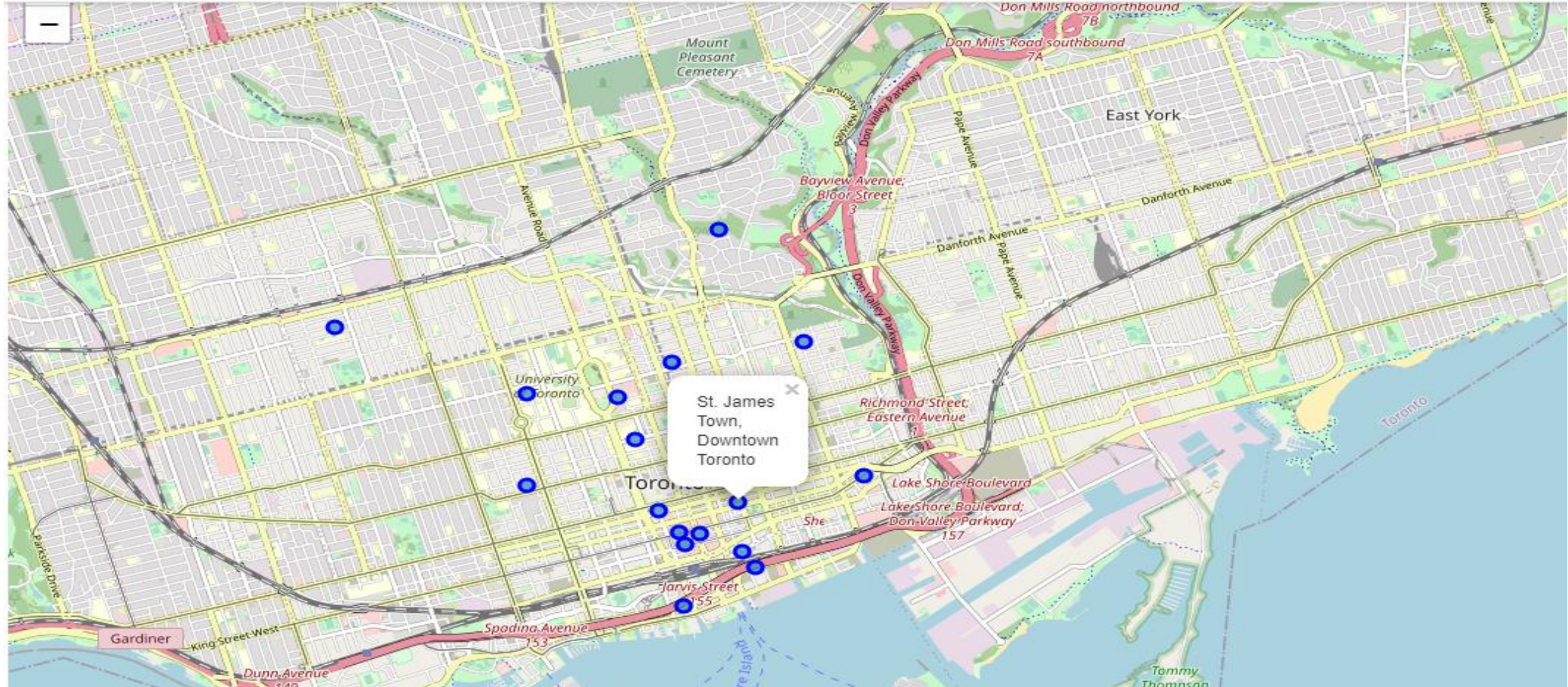
Last steps

- Using Foursquare Api to obtain venues within the radius of 500 metres

- Analysing the data obtained and then converting in the onehot vector form

- Using Clustering algorithm to find clusters which are restaurant rich.