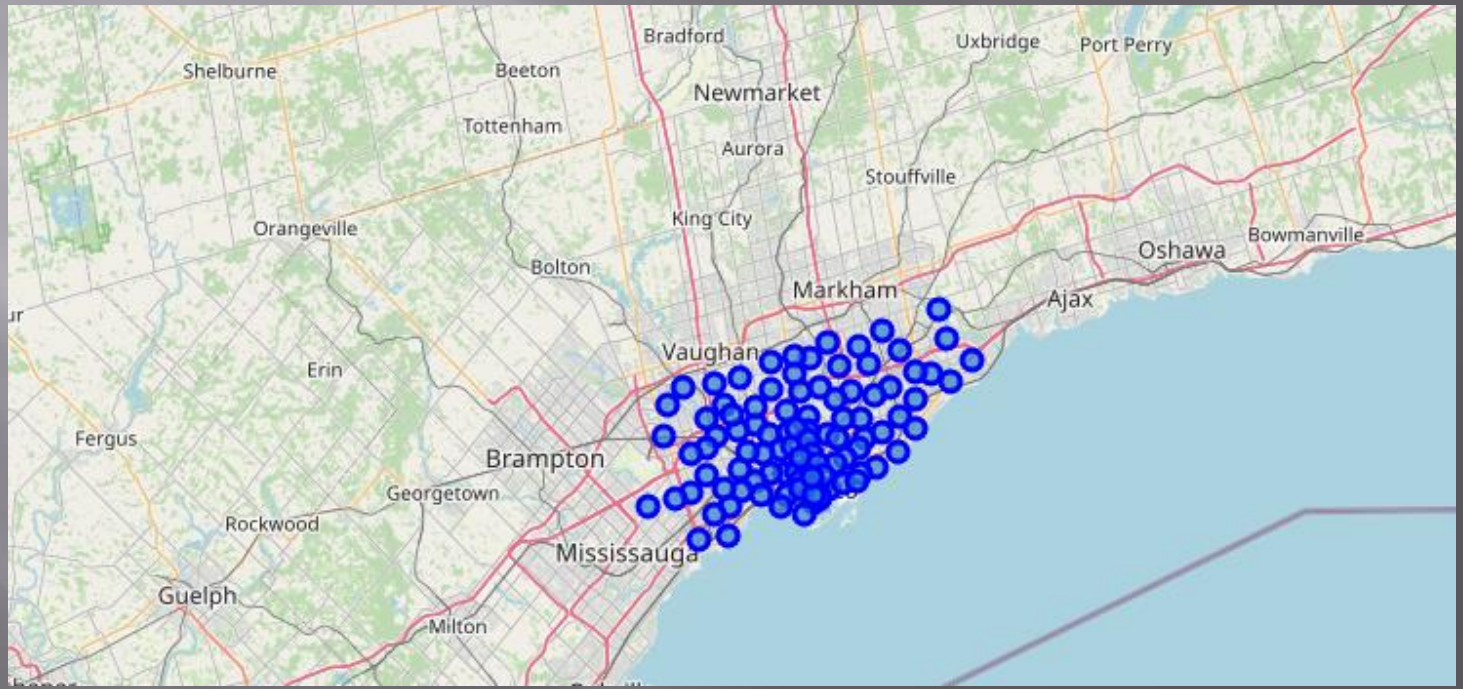
Business Prediction in Neighborhood

Data acquisition

- We have acquired the required data from Wikipedia Toronto city postal code page.
- Url is https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,
- From above page we have downloaded the all borough details and respective neighborhood details along with postal codes
- To acquire data from web page we have used Beautiful Soup object
- For latitude and longitude details, we have used Geospatial_Coordinates excel provided by Coursera in 3rd week assignment.
- To acquire final data set, we have merged both data sets using postal code

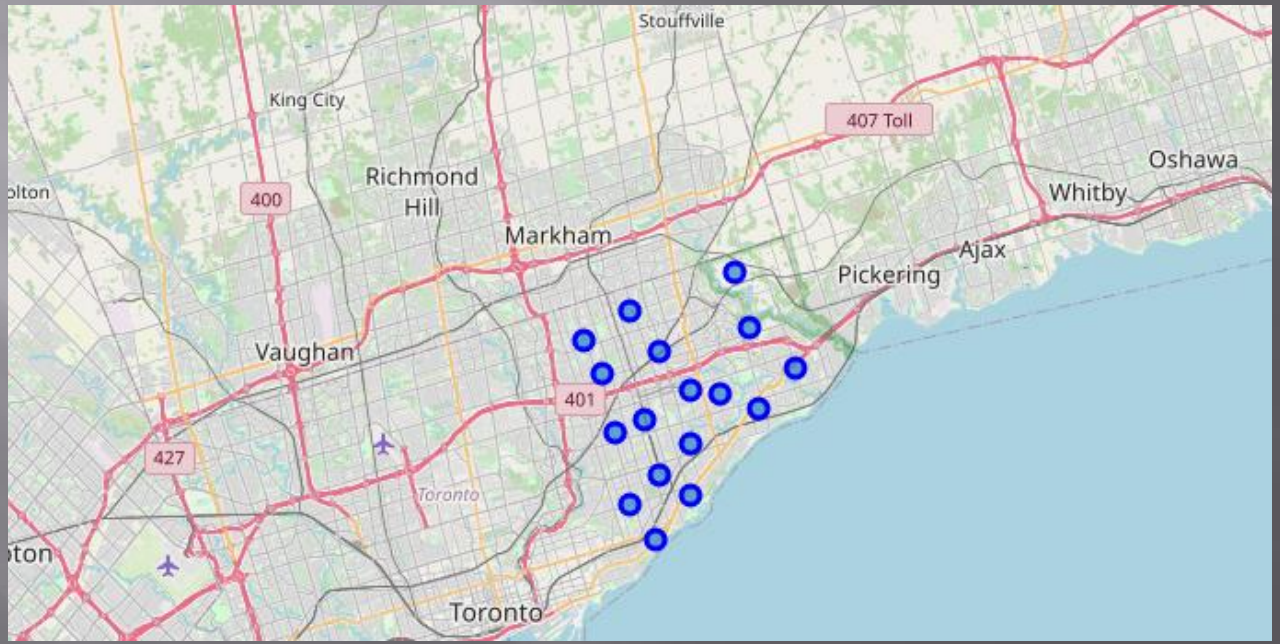
Data Visualization

To visualize data set, we have used Folium. Using Folium, we can zoom in Zoom out the visualizing picture. Please find Toronto picture below.



Scarborough Neighborhoods Visualization

I have chosen Scarborough for business prediction, we can apply same Model on all borough in Toronto data set. Please find Scarborough map below



Feature Engineering and Neighborhood exploration

- To explore the neighborhoods, we have used FourSquare API
- Using Four Square API, we have gathered all the venue details in the respective neighborhoods along with category of venue
- To get all the details about venues in the neighborhood, we have create project in Four Square, and access token
- In feature engineering , we have one hot coded all venue categories.

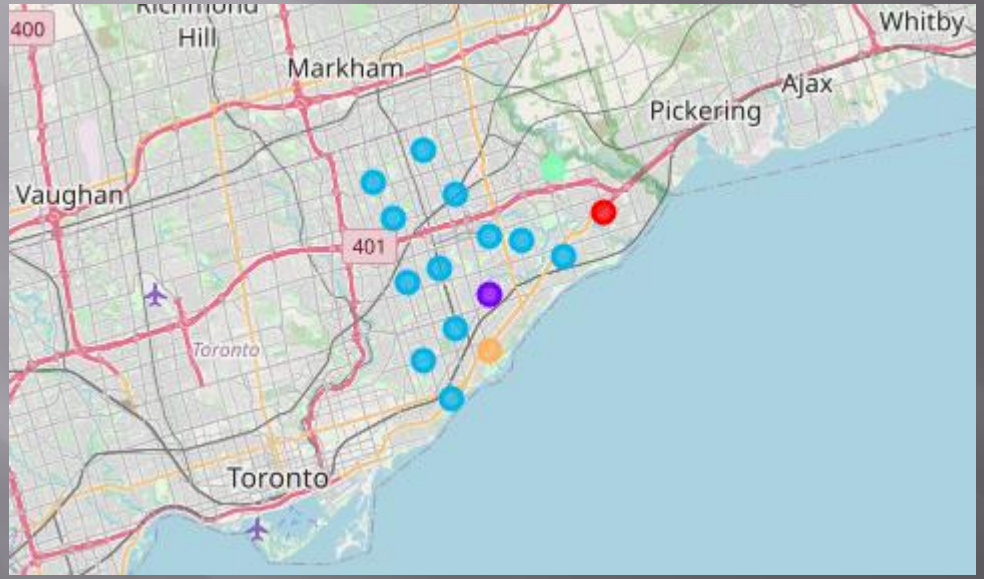
Model building and Results

- We have used K-means clustering ML model for more simplification
- Using K-means clustering, we have divided the neighborhoods into 5 clusters

- Find top 3 venues places from 5 clusters

Cluster 1: Smoke Shop, Playground, Vietnamese Restaurant
Cluster 2: in cluster 2, we have many neighborhoods, it is difficult to dicided on most prefered place by people in the cluster. from the above data we can choose Coffee Shop is most prefered place by people.
Cluster 3: Fast Food Restaurant, Vietnamese Restaurant, Ice Cream Shop
Cluster 4: American Restaurant, Motel, College Stadium
Cluster 5: Home Service, Bar, College Stadium

Clusters Visualization



Conclusion

Finally, using above data business man can decide,
which business he can start in respective cluster, to avaoid loss and again more profits

Thank You