**Applied Data Science Capstone Project: The Battle of Neighborhoods Better Neighborhood in Toronto**

**Introduction**:

This project was conducted with aims to assist individuals who plan to migrate, settle and or visit Toronto. It analyses and explores its neighborhood. Thus, with its availability and application, individuals can make informed, efficient and data driven decision when choosing best neighborhood to visit in Scarborough, Toronto.

**Background Information**

Globalization has increased net incoming migration especially in countries of the North. Vast number of people migrate to various places, i.e. states of Canada in this scenario. This has triggered a need for information of where these individuals are heading to. This information could be about, job opportunities, vacation spots, or even best neighborhoods for settlement. With this need of info, arise a need for data, analytics, and science. Thanks to Big Data, we can retrieve info about, good housing prices and reputed schools, hospitals, estates, restaurants, shopping centres, parks, and many more other factors that drive migration. This project then, as already stated, is for such individuals who are on the quest for better neighborhoods. This Project aims to create comparative analysis between neighborhoods. Foursquare API will be employed to carry out such analytics. The final reporting will include, housing price and school based on ratings, crime rates, spatial analytics i.e. connectivity, weather, management for emergency, water resources and recreational facilities.

**Problems to Addrress**

This report will address, better neighborhood of Toronto for any individual who aims to settle there. Socio-economic affairs of that particular neighborhood. Connectivity and transport routes. Housing prices in order and Schools, through ratings and online reputation these poses.

**The Location:**

Scarborough one of the most diverse and multicultural areas in the Greater Toronto Area, which is home to various religious groups and places of worship and popular destination for new immigrants in Canada.

### **Methods and Materials**

Foursquare API: Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business. Clustering: To compare the similarities of two cities, the neighborhoods will be explored, segmented, and grouped into clusters to find similar neighborhoods. Thus k-means clustering algorithm will be employed. Libraries Used: Pandas: For creating and manipulating dataframes. Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map. Scikit Learn: For importing k-means clustering. JSON: Library to handle JSON files. XML: To separate data from presentation and XML stores data in plain text format. Geocoder: To retrieve Location Data. Beautiful Soup and Requests: To scrap and library to handle http requests. Matplotlib: Python Plotting Module.

### **Data Description:**

### Data for this project was retrieved from [List of postal codes of Canada: M i.e. Toronto](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). Scarborough dataset which was scrapped from this page in the past or previous project will be used to further this study and observation. Dataset consisting of latitude and longitude, zip codes. The data retrieved contains info of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows: Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venues, Name of the venue e.g. the name of a store or restaurant, Venue Latitude, Venue Longitude, Venue Category

**Results**

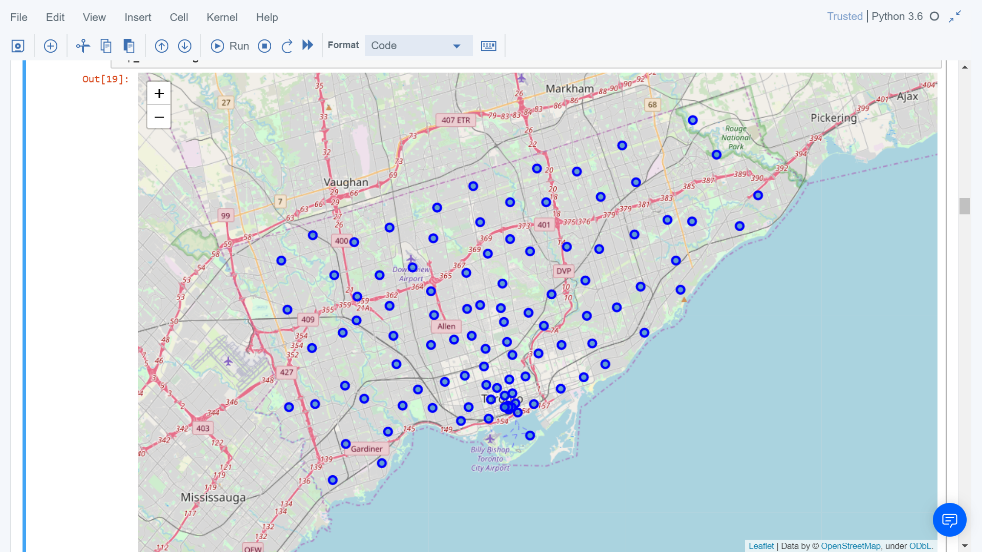
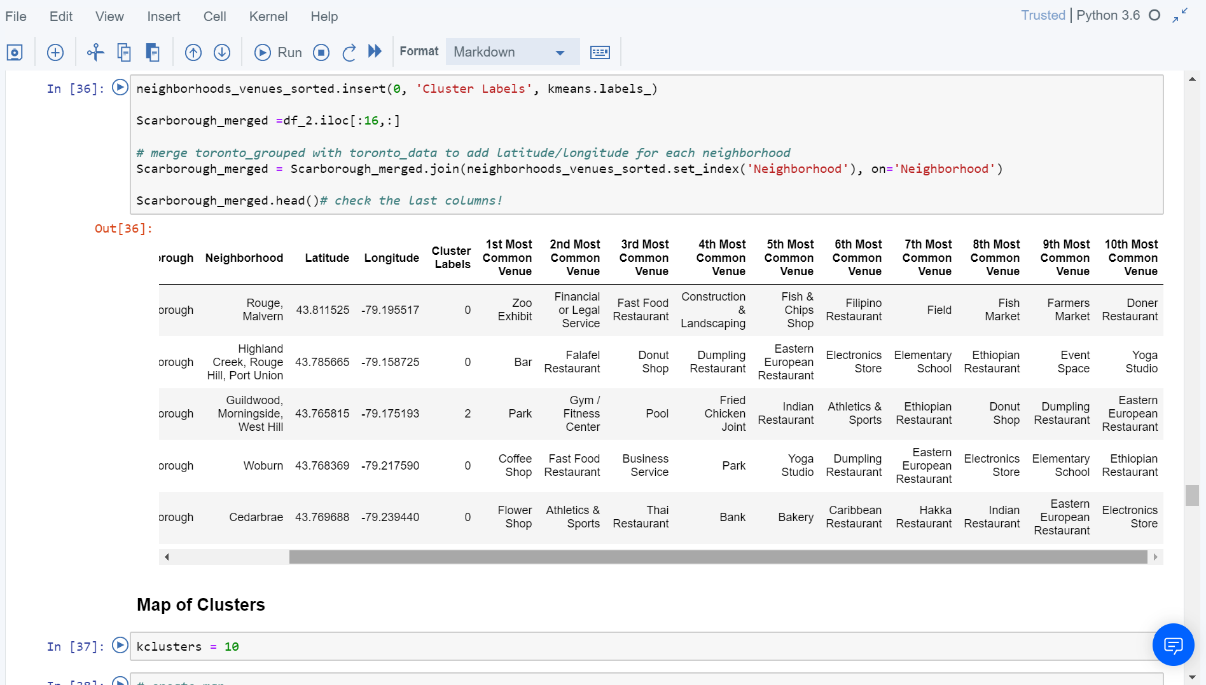
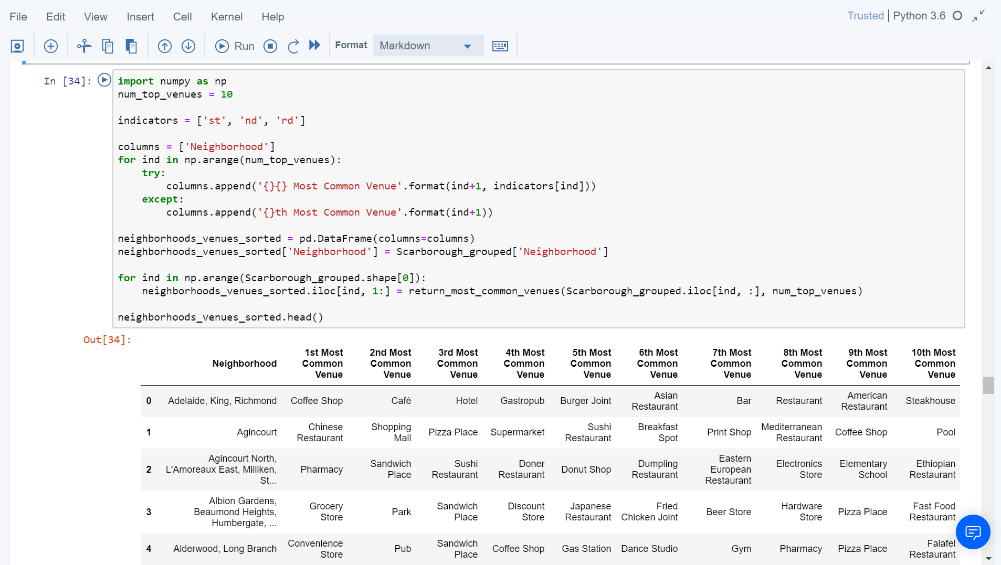


Figure 1. Map of Scarborough

Figure 2. K-Means Clustering

Figure 3. Most Common venues near Neighborhood

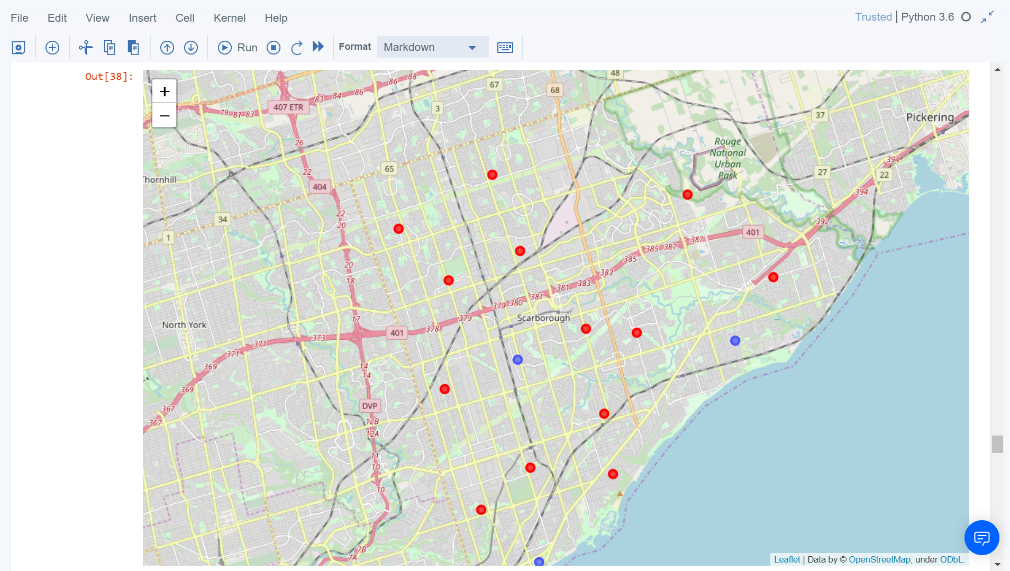


Figure 4. Map of Clusters in Scarborough

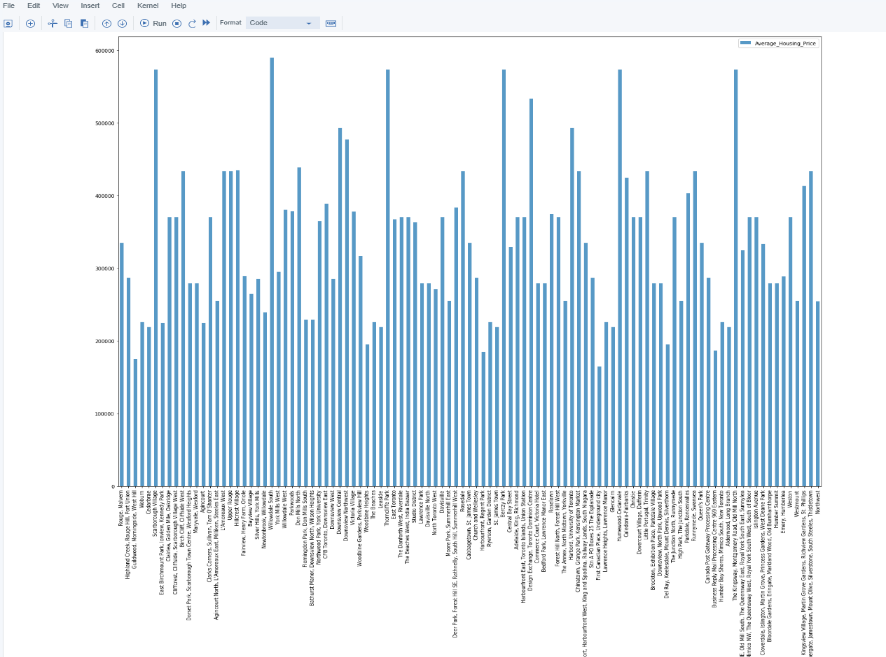
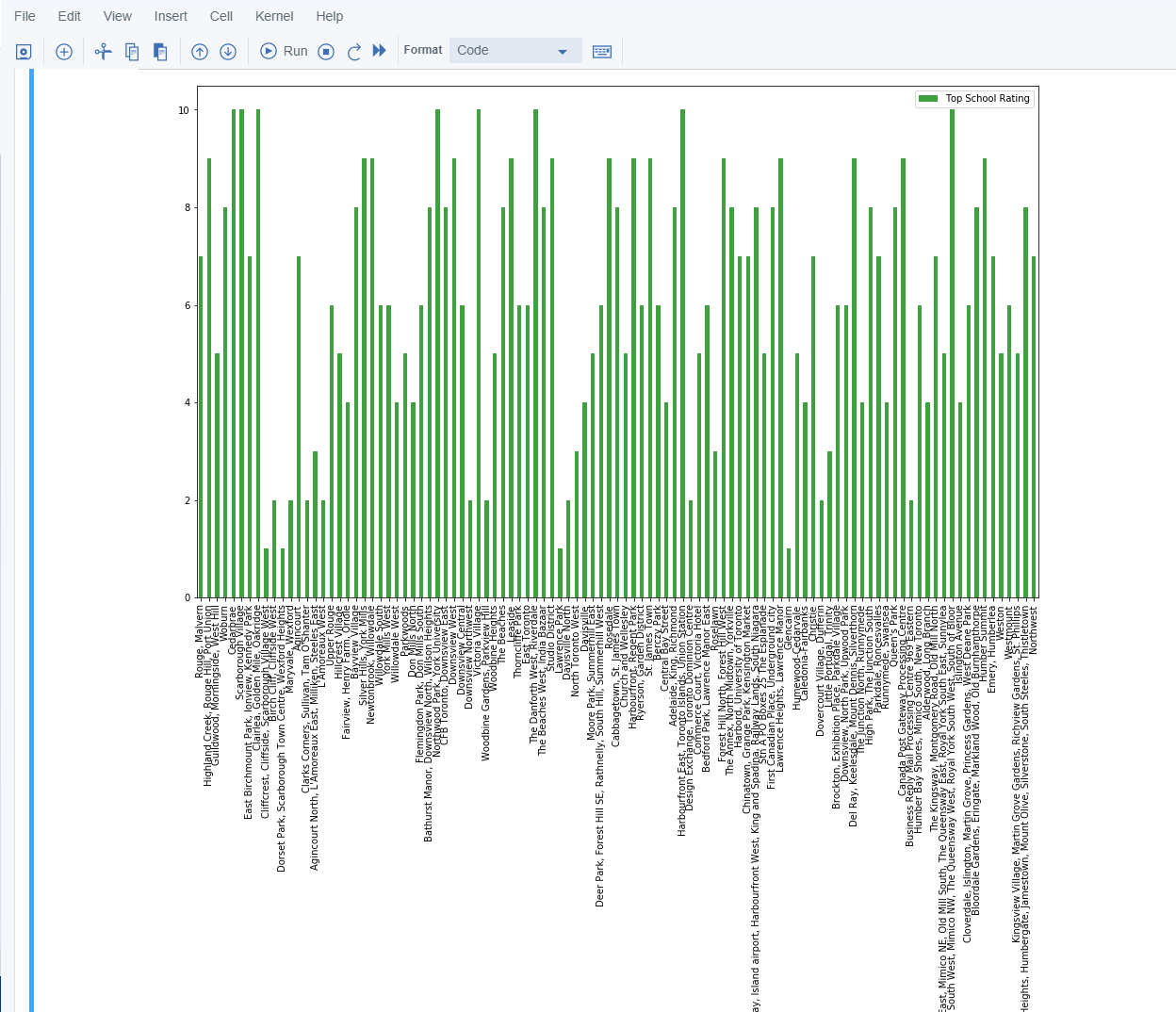


Figure 5. Average Housing Price by Clusters in Scarborough

Figure 6. School Ratings by Clusters in Scarborough

**Discussion**

Here, different categories of venues were explored. Initially, were a sum of 103 unique postal codes has been observed for 10 diffrent Borough. Using geocoder of Geopy library for Toronto. Then the result above were attained through spliting the data into longitude and latitude of the neighborhood. Each neighborhood was analysed and led to one-hot encoding. After one hot encoding and venue exploration, the k-means algorithm for k=5 was applied. These five clusters identified in five different colors. Cluster 0 is having only one neighborhood, cluster 1 is having the maximum number of neighborhoods.