

Background Discussion and Business Proposal

1.Introduction

Nowadays, more and more people are immigrating to Canada from all over the world, and Toronto, as the economic and transportation center of Canada, has become the first choice for many people to settle in.

This project is aim to help people explore better facilities in Scarborough, help people understand the area and nearby areas before moving to a new city, state, country or region to work or start a new life. It will help people make a wise decision to choose the best neighborhood in Scarborough.

This project will analyze the characteristics of different neighborhoods, including median housing price, better school according to ratings, crime rates of particular area, road connectivity, weather conditions, good management for emergency, water resources both fresh and waste and so on. So they could have easy access to Café, better schools, shopping malls, and like-minded people.

2.Data

Data Link:https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

We will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes. To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods i. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

We will need the information retrieved from Foursquare as follow: Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue, Name of the venue, Venue Latitude, Venue Longitude, Venue Category. While analysis, we will combine these data with which we scrap from wikipedia.

3.Methodology

libraries

Foursquare API: perform location search, sharing and obtain details about business based on its database of millions of places.

Numpy and Pandas: basic libraries to manipulate series and dataframe

Matplotlib: data visualization

BeautifulSoup and requests: scrap data from webpages

Sklearn: perform machine learning, concretely, the k mean cluster

Folium: construct an interactive map

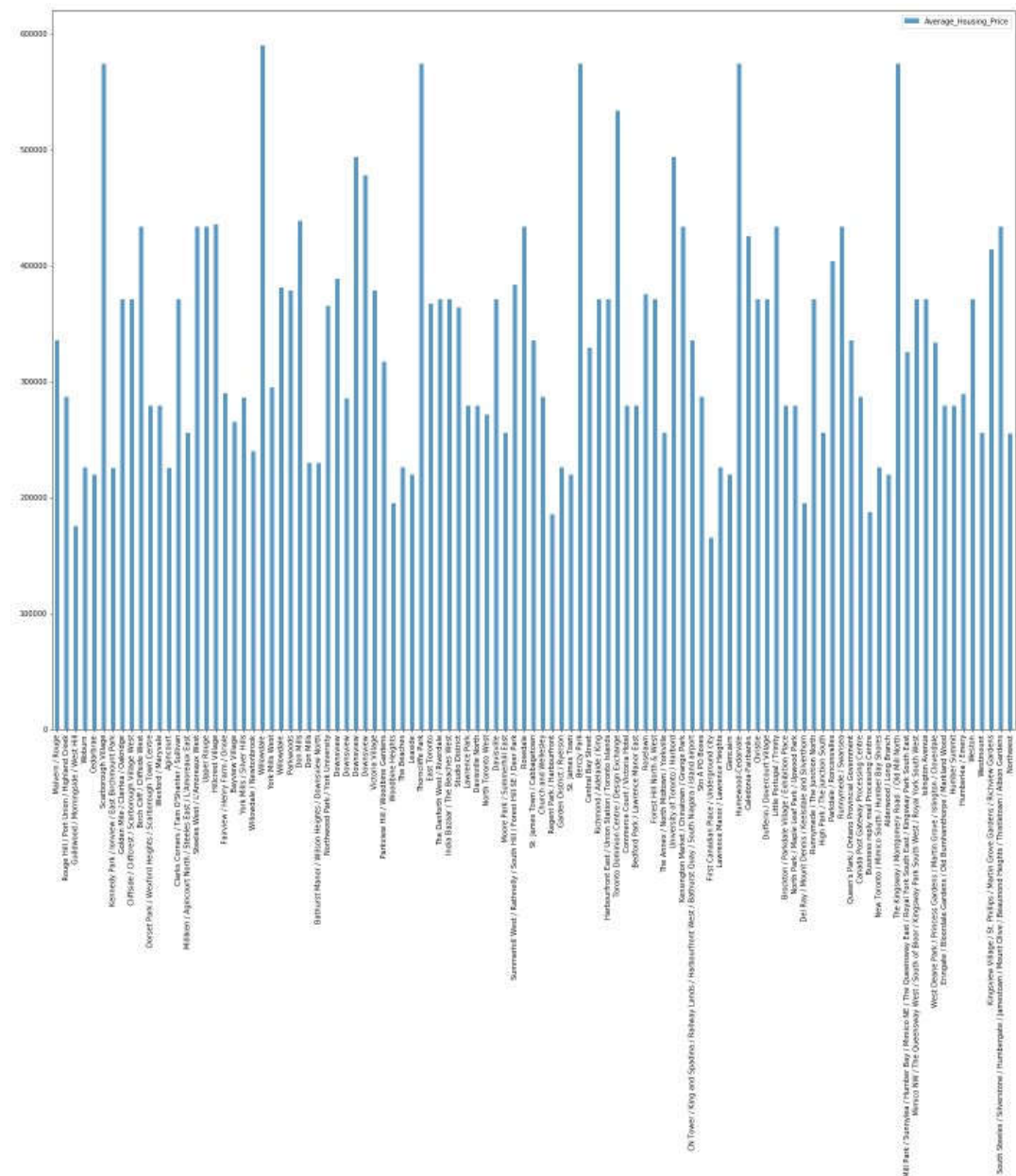
K-Means Cluster

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

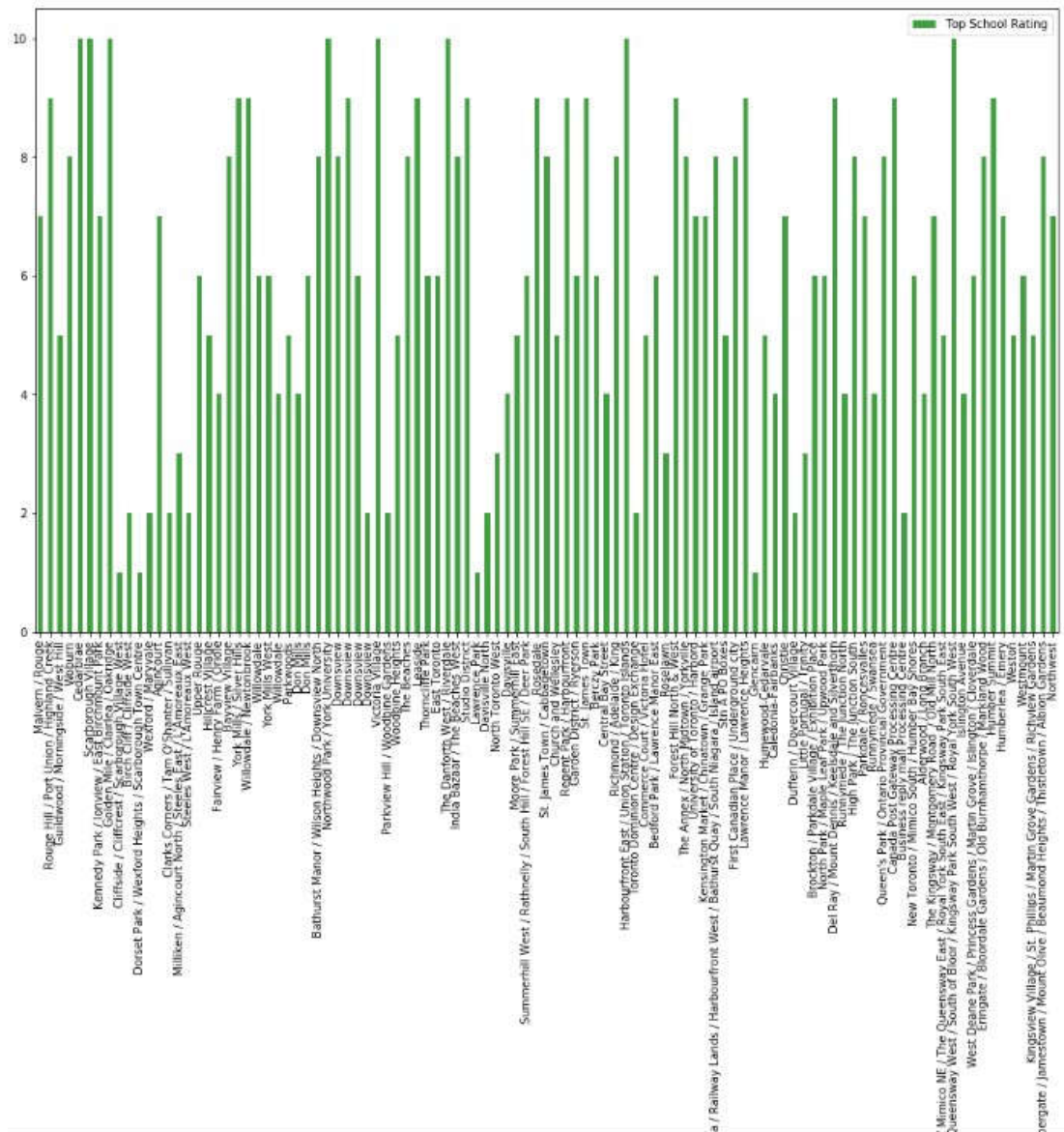
4.Result

Map of Scarborough

Average Housing Price by Clusters in Scarborough



School Rating



5.Discussion

The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shifting there. Social presence in society in terms of like minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

For those who are looking for school for their children, they could reach a balance between house price and school rating

6.Conclusion

In this project, k-means cluster was performed to classify the Scarborough area into ten different clusters and they have very similar neighborhood around them. A particular neighborhood based on average house prices and school rating has presente