IBM Data Science Capstone Project

Using k-means clustering to find the best area in Toronto to expand the advertising for a fashion company - Boozt.

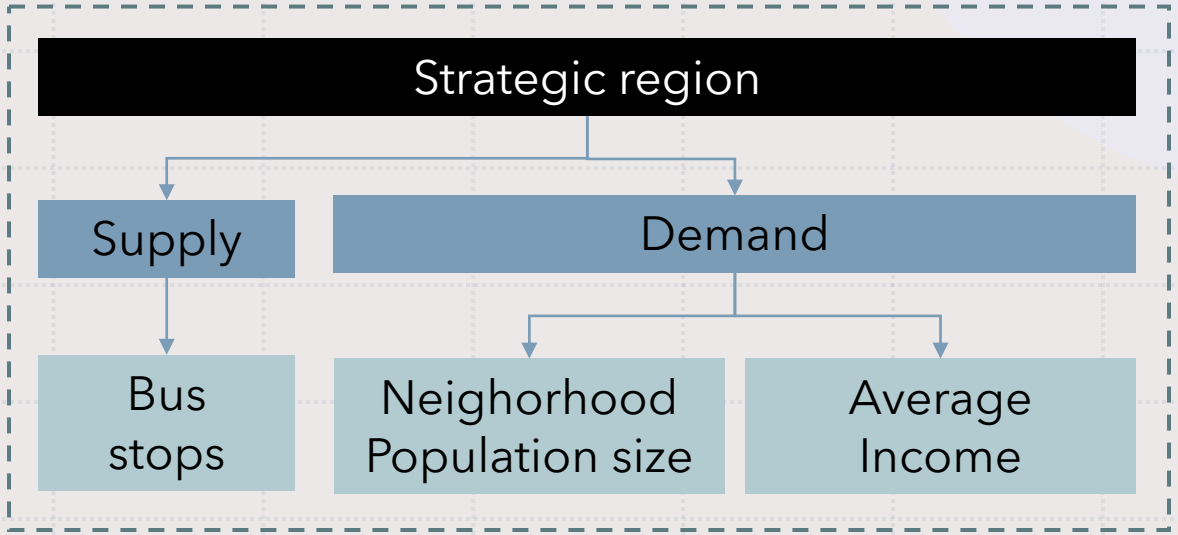
Table of Contents

Fashion Company: Strategic Expansion

Context:

In this project, we will hypothesize that an online shopping fashion company – Boozt, decides to expand their business to the city of Toronto. To ensure that every dollar is strategically spent, this fashion company must selectively choose the most optimal regions for advertising to grab the potential customers

Issue Tree:

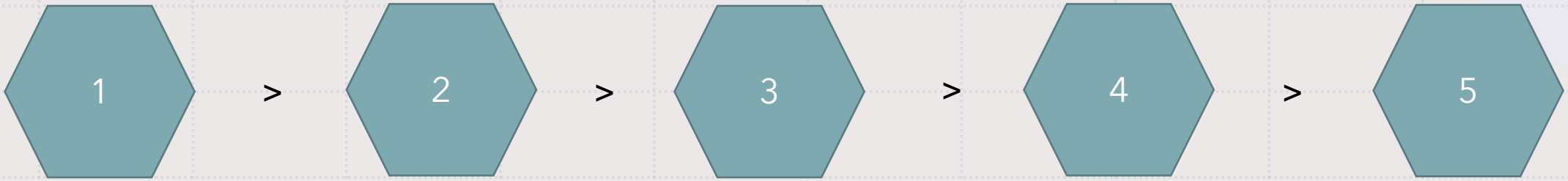


Problem Statement:

This project seeks to determine the most strategic region in Toronto for a fashion company to spread of advertising.

Table of Contents

Data Acquisition and Cleaning



Toronto neighbor-hood data was web scraped from Wikipedia

Household income and neighborhood population size data were extracted from Toronto Wellbeing.

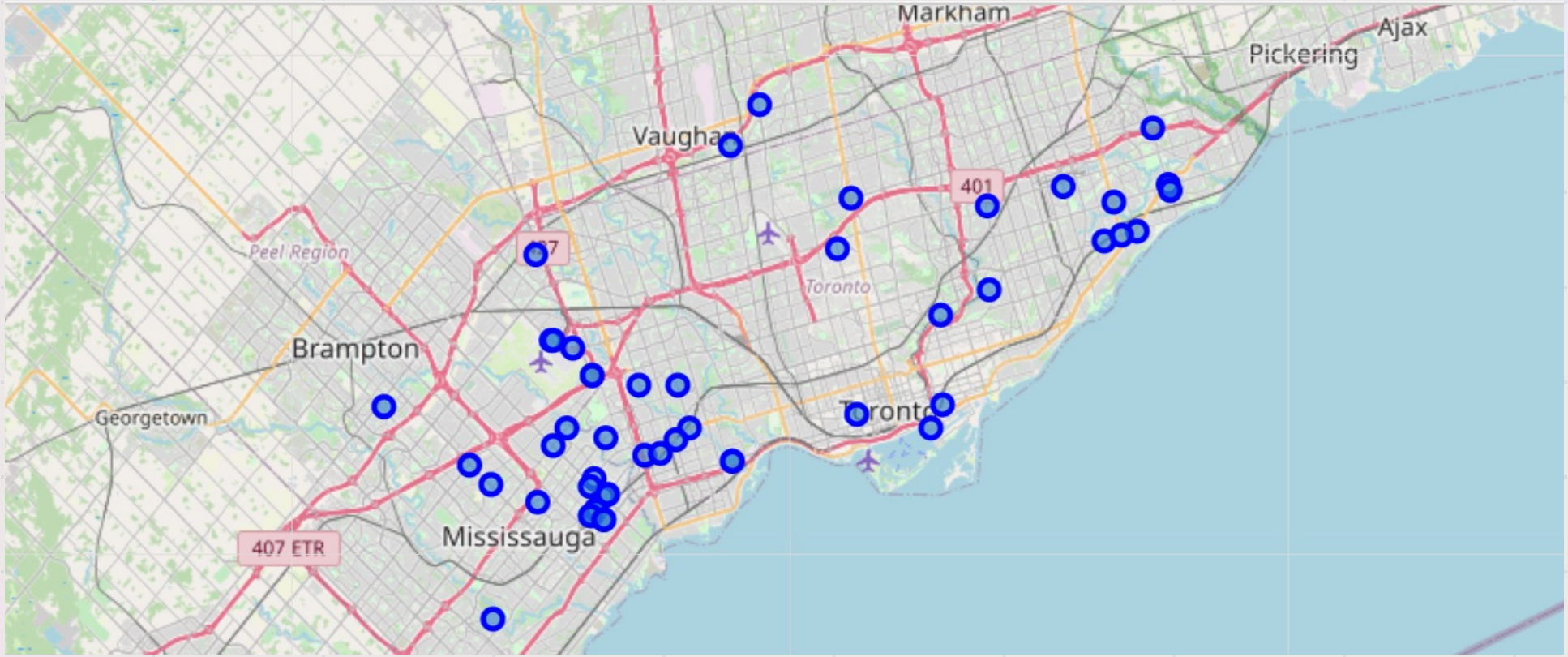
Numbers of nearby busstops from each neighorbood was extracted from Foursquare.

Household income, population size and nearny busstops are all merged into one single data frame.

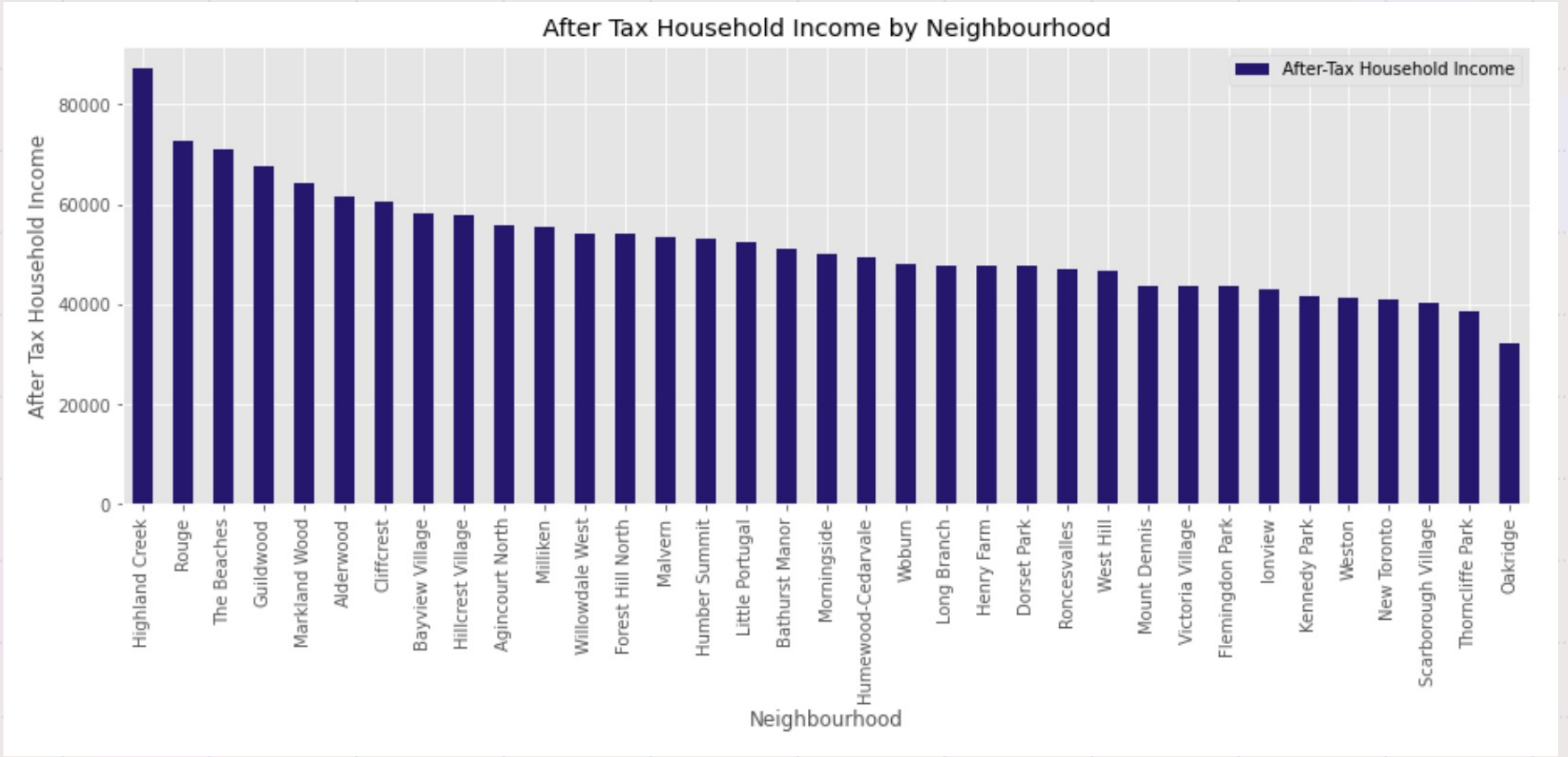
The final clean data contants 3 features (population size, household income, nearby busstops)

Table of Contents

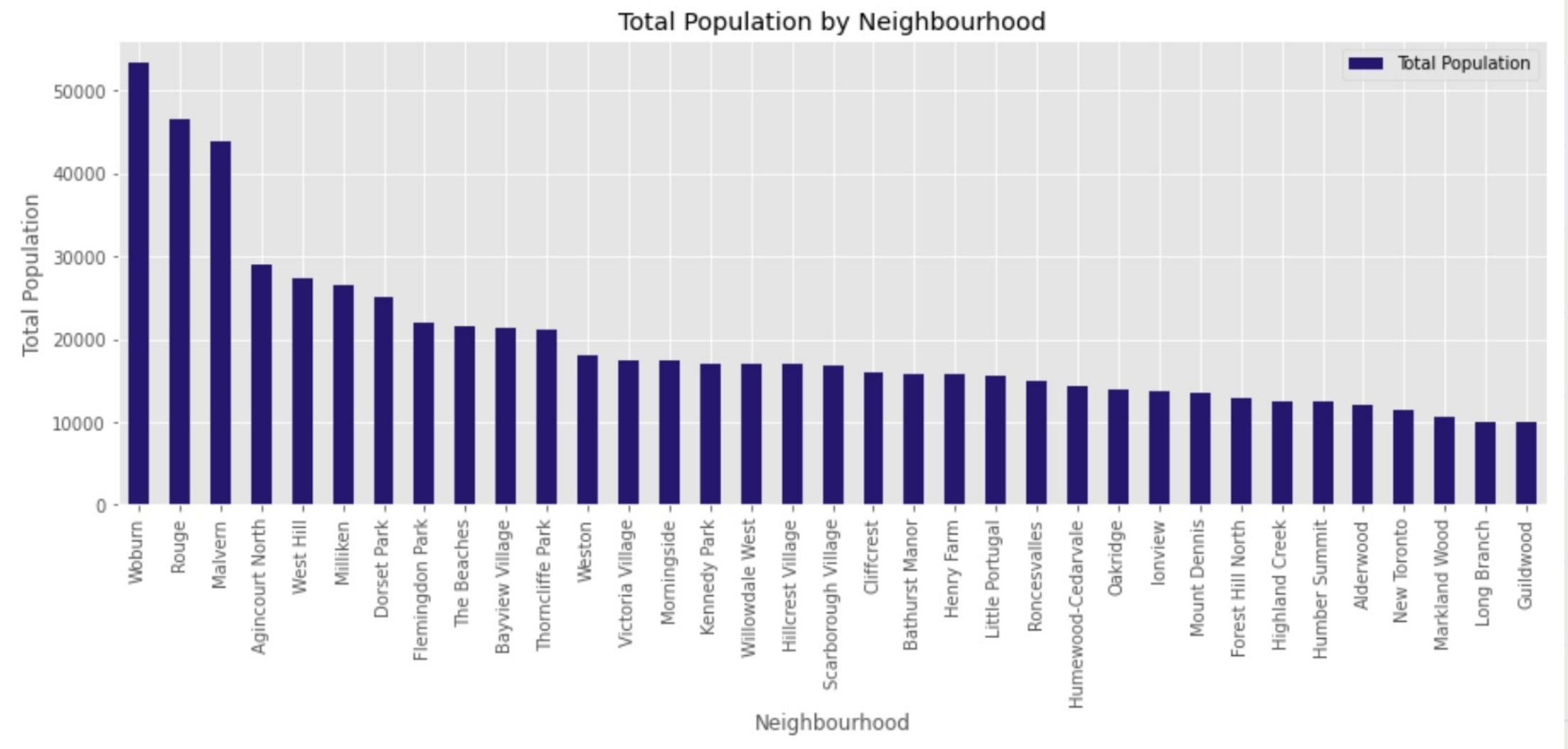
Exploratory Data Analysis: Folium Mapping



Exploratory Data Analysis: Household income



Exploratory Data Analysis: Population size



Exploratory Data Analysis: Bus-stop count

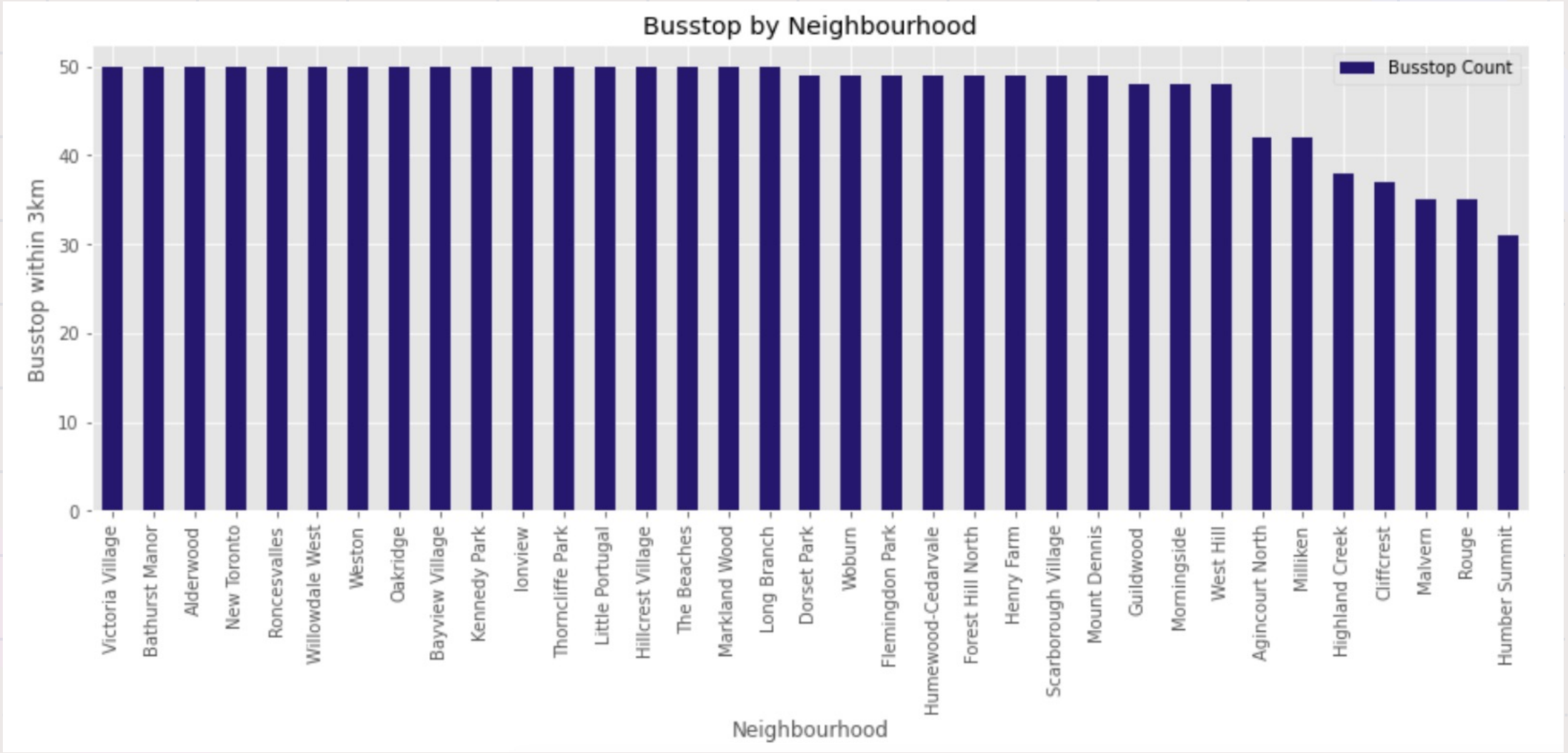
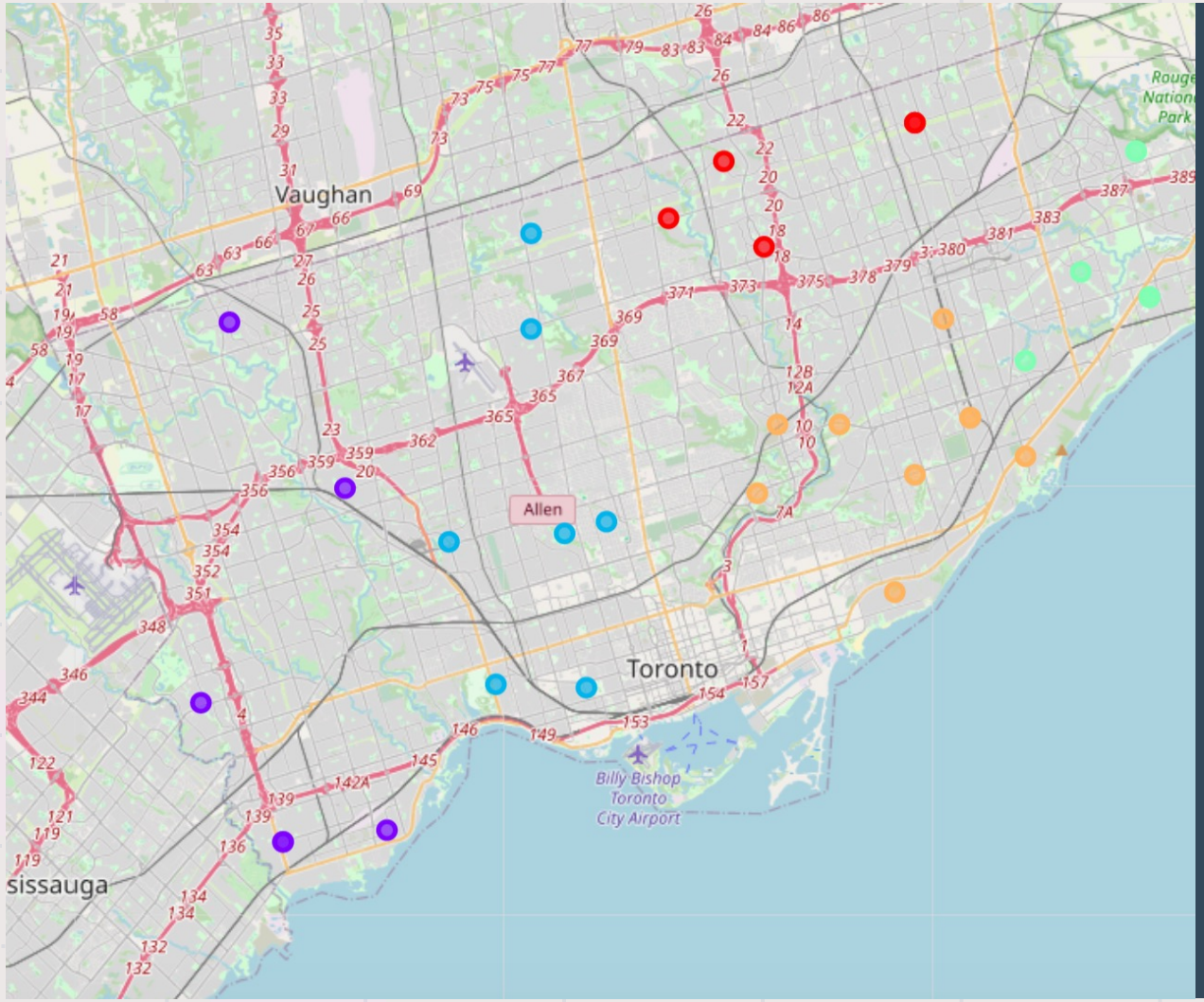


Table of Contents

K-means Clustering



5 Clusters returned based on Elbow method.

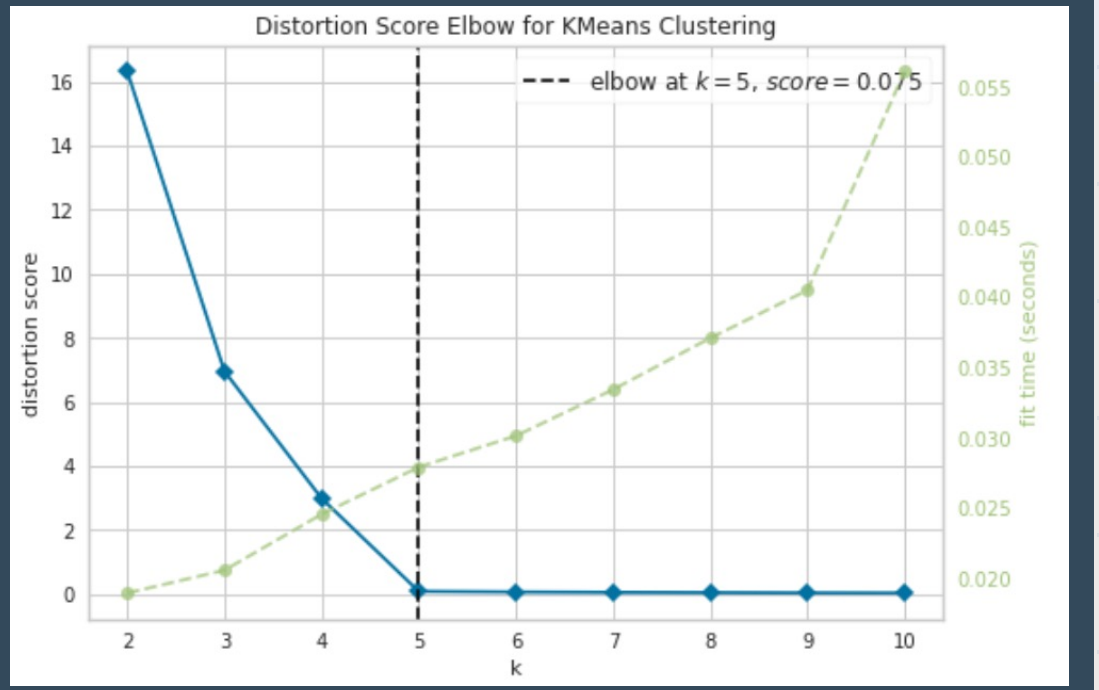
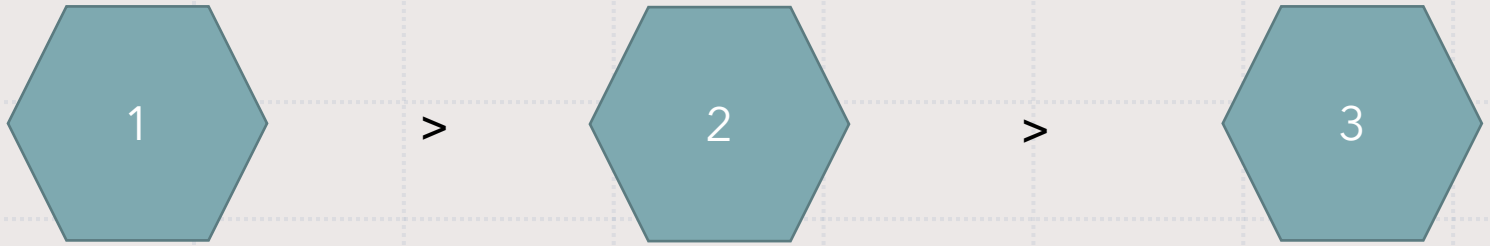


Table of Contents

Analysing Results



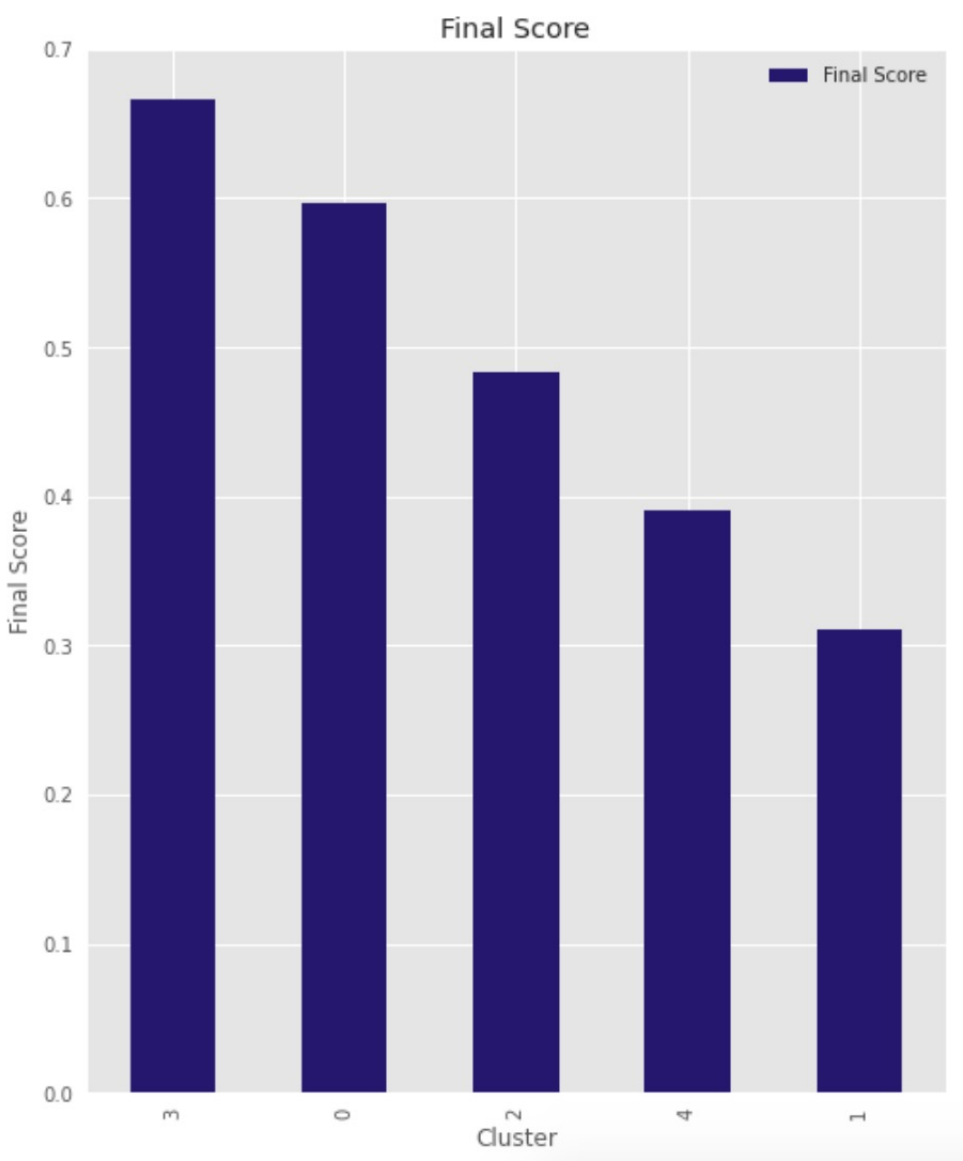
Normalising the feature data (household income, population size, nearby busstop count).

Calculate the mean score of each cluster

Visualising the results with bar chart.

Table of Contents

Conclusion & Future Directions



Results:

Cluster 3 is the best region to expand the advertising to, followed by cluster 0, both clusters has final score of more than 0.5, therefore are the considered most optimal.

Future Directions:

- Population size and average household income alone produres limit results. For realistically speaking, online fashion demand can also driven by age, gender, etc.
- We are only considered making the bus-stop billboard in this project, as there are hunreds way of advertise, more researches can be perform to find the best effective way of reaching potential customers.