Advertising Strategy for Fashion Company

1.0 Introduction

1.1 Business Problem & Analysis Objective

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.

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

1.2 Data Requirements

In this project, we will take into account of the supply of advertising location and demand from customers to assess the potential of each region. Furthermore, bus stop has decide as main advertising location, the number of nearby bus stop within a 3km radius to a neighborhood is used to determine the advertising density. On the other hand, the fashion demand from a customer is assumed to be driven by purchasing power of each neighborhood's population. The purchasing power is both determined by population size and average household income.

Data Sources:

1. Nearby bus stop- Foursquare
2. Location data - Wikipedia
3. Demographic/ socioeconomic data (population size by neighbourhood, average after tax household income by neighbourhood) - Wellbeing Toronto

1.3 Methodology

1. Various data sets were imported and subsequently combined into a single data frame.
2. Foursquare API is used to determine the number of nearby bus stopwithin a 3km radius to a neighborhood, this is returned as bus stop count and appended to the combined dataframe.

3. To determine the most strategic location, neighbourhoods will be segregated into different regions via k-means clustering.
4. The statistical data on each region (or cluster) is then generated and explored.

2.0 Data Acquisition and Cleaning

There are 4 feature data which requires to perform the data analysis:

1. Toronto neighorhood location data – We were web scrap Toronto neighborhood data from Wikipedia, appending the location data with geocode data, then saved into dataframe using pandas.

	Postcode	Borough	Neighbourhood	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	Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Heights	43.718518	-79.464763
4	M6A	North York	Lawrence Manor	43.718518	-79.464763
5	M7A	Downtown Toronto	Queen's Park	43.662301	-79.389494

2. Toronto population data and average household income by neighborhood – which export the CSV file from wellbeing_toronto website, then loaded the data into dataframe and merge into one.

	Neighbourhood	After-Tax Household Income	Total Population
0	West Humber-Clairville	59703.0	33312.0
1	Mount Olive-Silverstone-Jamestown	46986.0	32954.0
2	Thistletown-Beaumond Heights	57522.0	10360.0
3	Rexdale-Kipling	51194.0	10529.0
4	Elms-Old Rexdale	49425.0	9456.0
5	Kingsview Village-The Westway	50714.0	22000.0

3. Nearby bus stop data – Foursquare was used to fetch the nearby busstop data as json format, then assign the relevant part of json to venues and transform venues into a dataframe and keep only columns that include venue name, and anything that is associated with location, then make a list for bus-stop count.

	name	categories	lat	lng	labeledLatLngs	distance	cc	country	formattedAddress
0	Missisauga Bus Stop #1309	Bus Stop	43.639558	-79.584139	[{'label': 'display', 'lat': 43.639558, 'lng':...	711	CA	Canada	[Canada]
1	Sabine Bus Stop	Bus Line	43.667052	-79.559617	[{'label': 'display', 'lat': 43.66705184632205...	2978	CA	Canada	[Martin Grove, Canada]
2	Mississauga/Toronto bus stop	Bus Line	43.631054	-79.555737	[{'label': 'display', 'lat': 43.63105411644539...	2216	CA	Canada	[Canada]
3	Bus stop #2915	None	43.645113	-79.610573	[{'label': 'display', 'lat': 43.64511315750388...	2694	CA	Canada	[Mississauga ON, Canada]
4	MiWay bus stop - bus #20 east to Islington Sub...	Bus Stop	43.607505	-79.632238	[{'label': 'display', 'lat': 43.607505, 'lng':...	5978	CA	Canada	[Mississauga ON, Canada]
5	The Queensway & Burma Dr WB Bus Stop	Bus Stop	43.628098	-79.494992	[{'label': 'display', 'lat': 43.628098, 'lng':...	6841	CA	Canada	[Toronto ON M8Y 1K1, Canada]

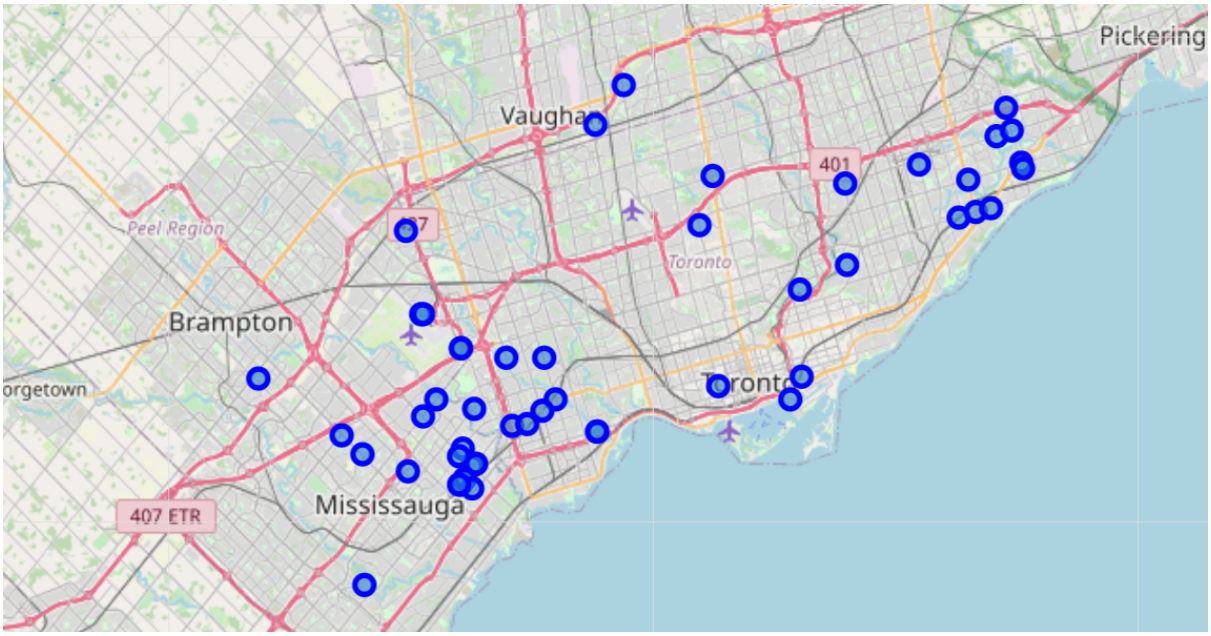
The final step is merged all feature datas into one single dataframe.

	Postcode	Borough	Neighbourhood	Latitude	Longitude	After-Tax Household Income	Total Population	Busstop Count
0	M4A	North York	Victoria Village	43.725882	-79.315572	43743.0	17510.0	50
1	M1B	Scarborough	Rouge	43.806686	-79.194353	72784.0	46496.0	35
2	M1B	Scarborough	Malvern	43.806686	-79.194353	53425.0	43794.0	35
3	M1C	Scarborough	Highland Creek	43.784535	-79.160497	87321.0	12494.0	38
4	M3C	North York	Flemingdon Park	43.725900	-79.340923	43511.0	21933.0	49
5	M6C	York	Humewood-Cedarvale	43.693781	-79.428191	49252.0	14365.0	49

3.0 Data Viualization

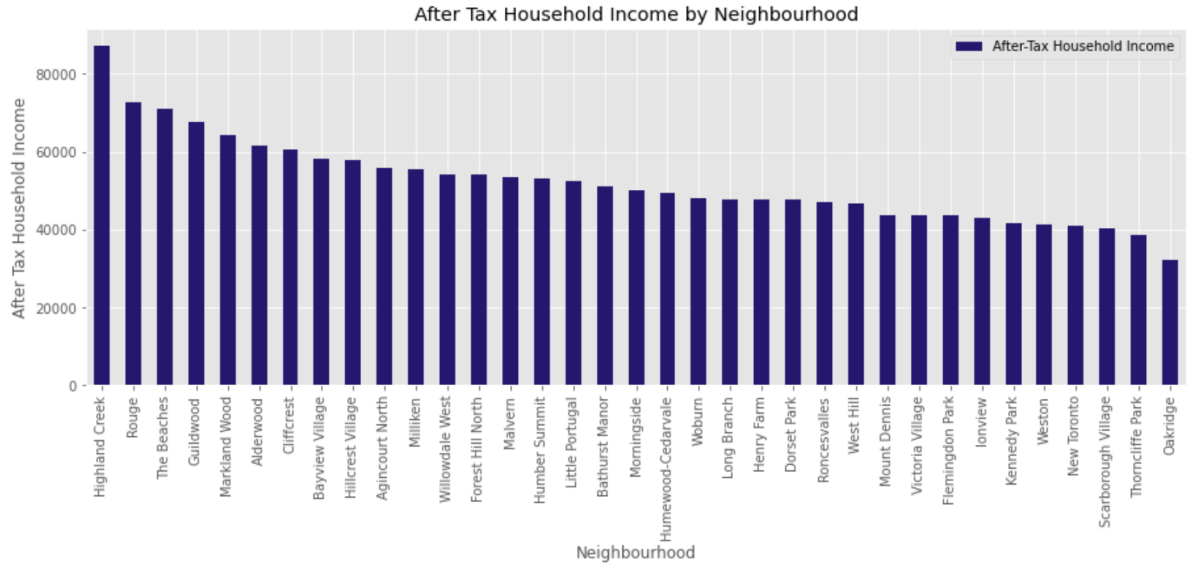
3.1 Folium Map

Visualizing the bus stops on the map as blue circle markers.



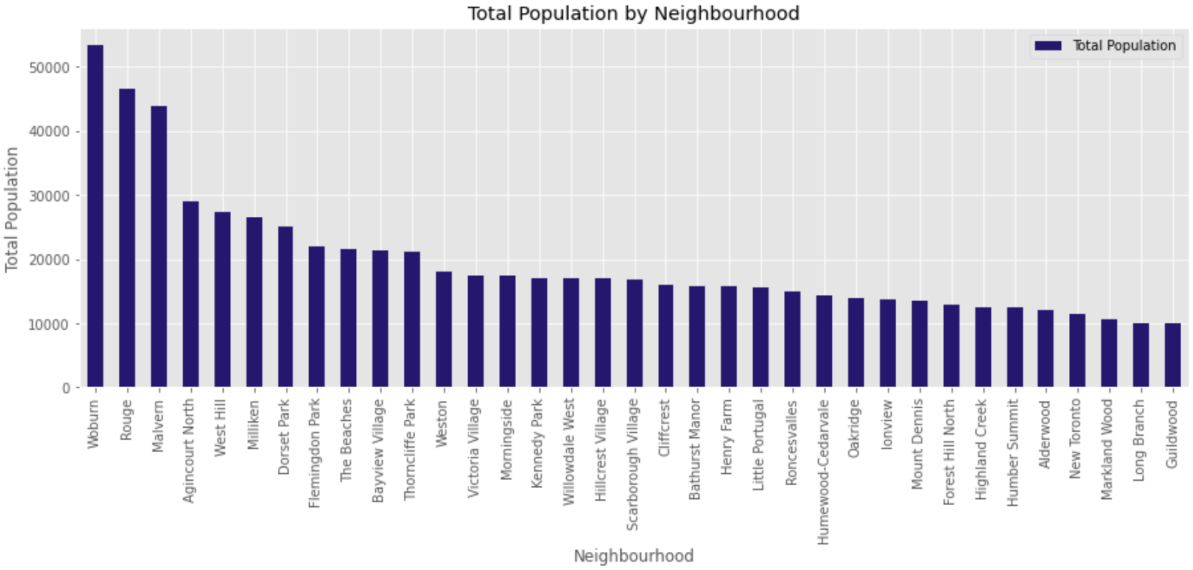
3.2 Visualising Average household income

This chart gives information about the average household income by neighborhoods, except the “Highland creek” region, the rest of regions seems only has small differences compare one with another.



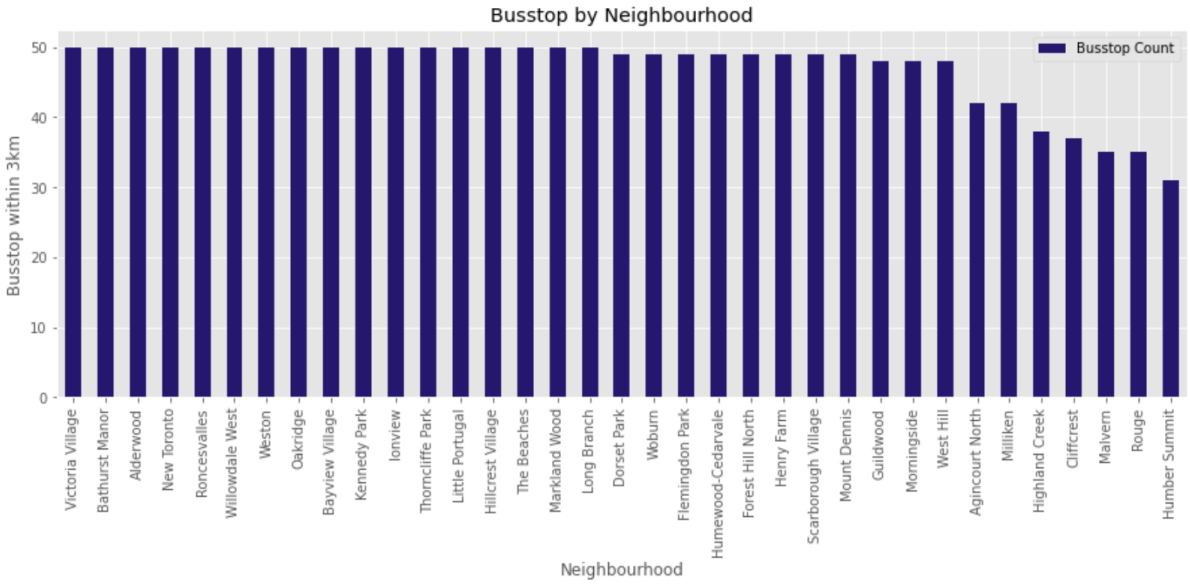
3.3 Visualising the population size by neighborhood

This chart shows the population size in each neighborhood, it can be seen that the number of the population in top 3 regions are relatively larger than others.



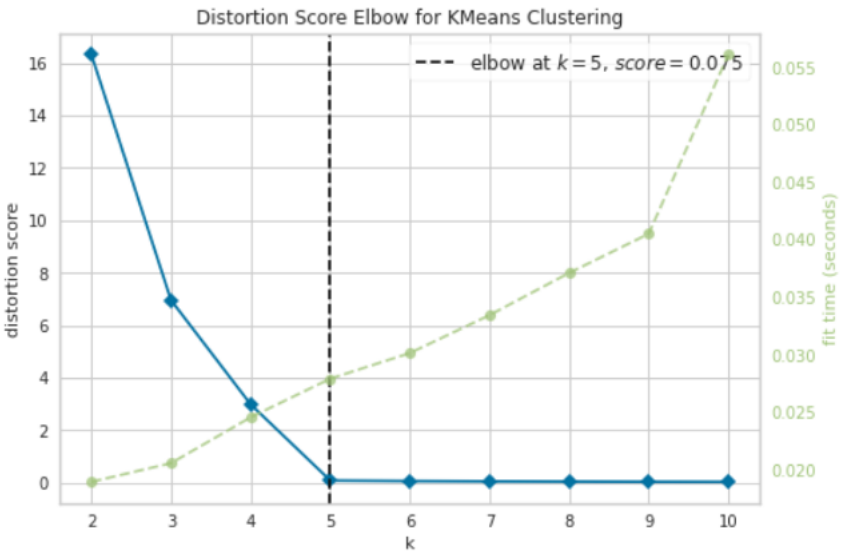
3.4 Visualising the nearby busstops by neighborhood

This chart illustrates the number of bus-stop in each neighborhood, as the maximine response we can got from Foursquare, seems many regions reaches the highest 50.



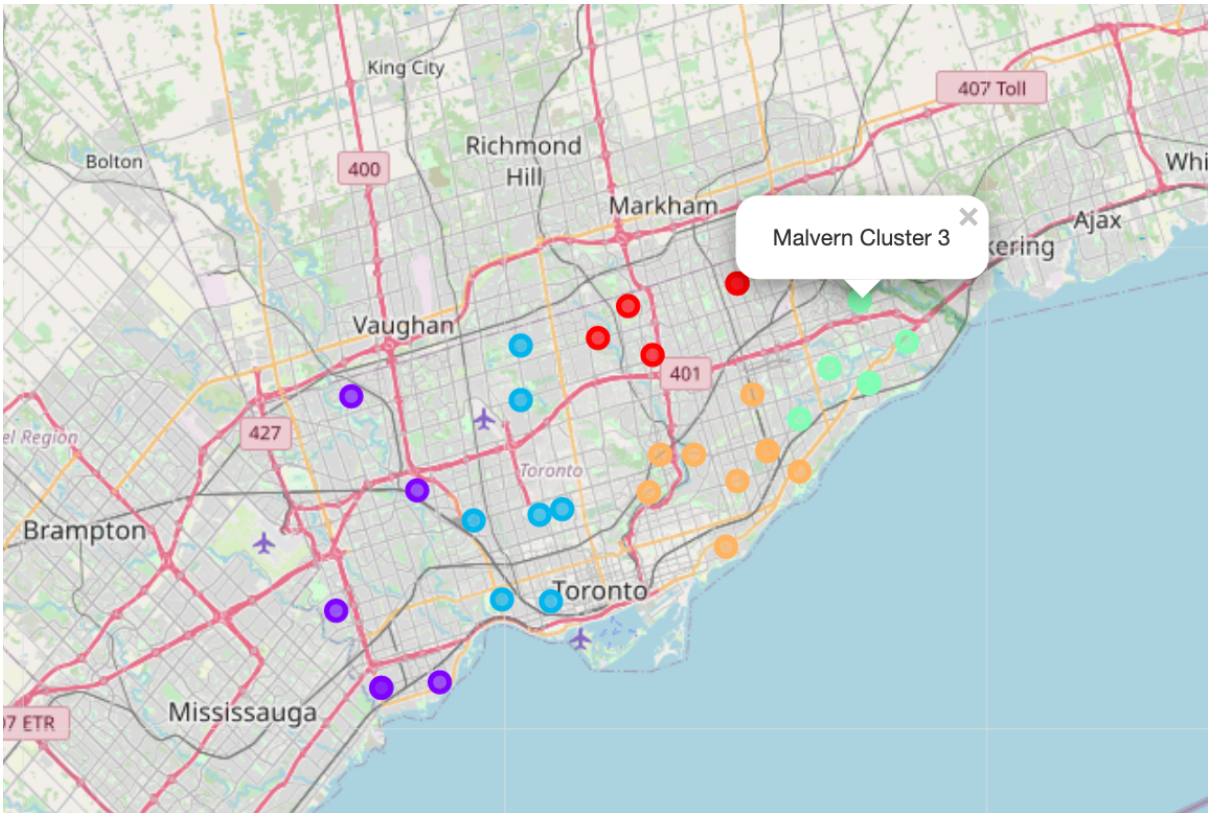
4.0 K-means Clustering

We can now cluster the neighborhoods, working only with required info. First we determined the optimal K for clustering using the Elbow method, the optimal K is 4 and clustering is performed using K=5.

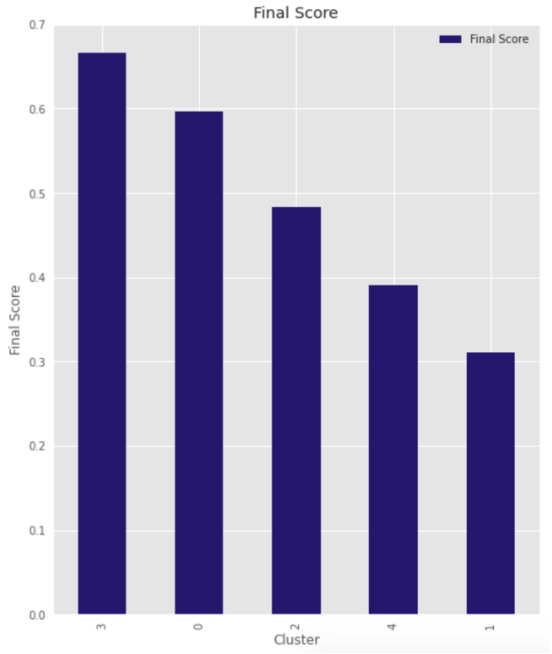


The clustering model then clusters the neighborhoods and provides for each neighborhood which is representative of the cluster it belongs to.

	Cluster Labels	After-Tax Household Income	Total Population	Latitude	Longitude	Busstop Count
0	0.0	54945.200000	21947.600000	43.799946	-79.333027	46.600000
1	1.0	51477.666667	12427.166667	43.652861	-79.541607	46.833333
2	2.0	50246.285714	14872.285714	43.702257	-79.439444	49.571429
3	3.0	58271.125000	28469.625000	43.775544	-79.196466	43.750000
4	4.0	46855.111111	18629.444444	43.719356	-79.291146	48.333333



5.0 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.