

Starbucks Site Selection – Toronto, Canada

IBM Data Science Professional Certificate

Arjun Manjini

coursera



IBM

Finding new potential Starbucks locations due to change in customer behaviour

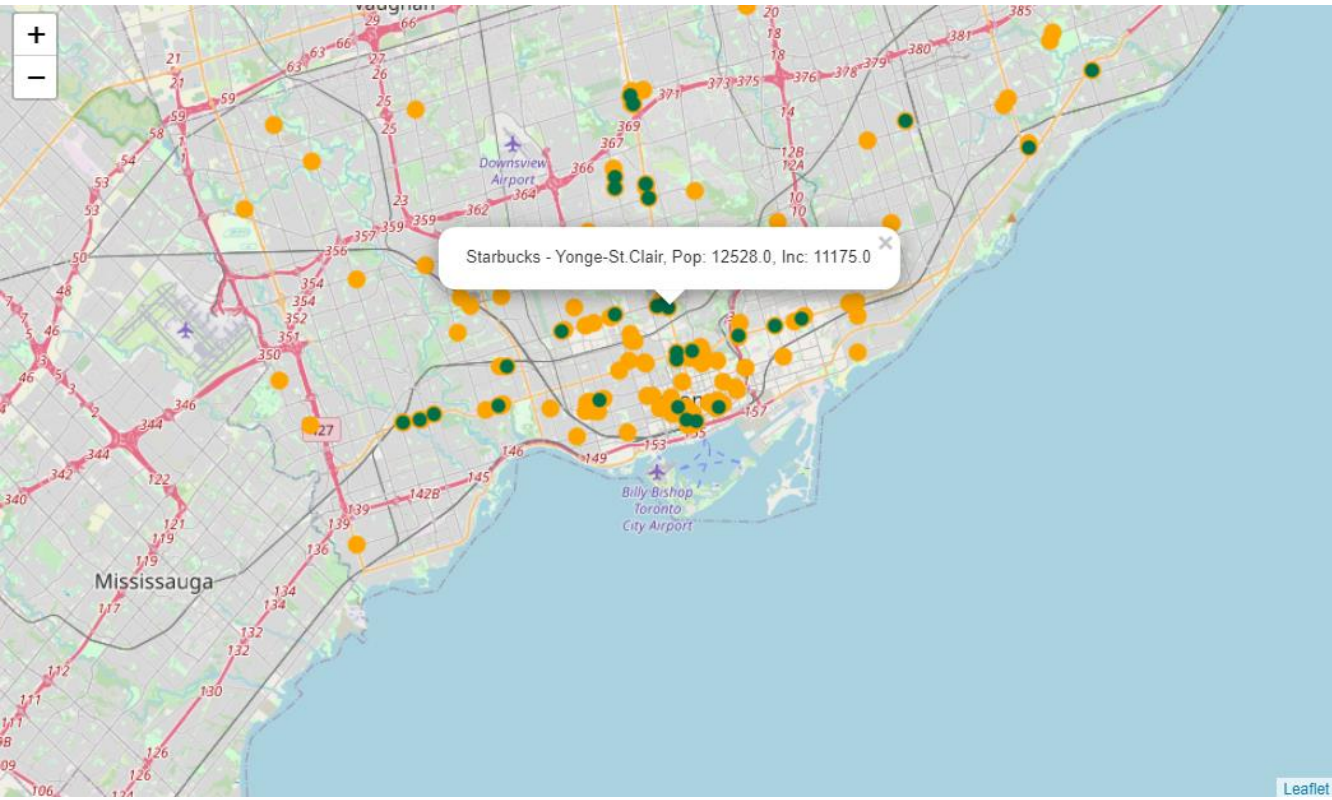


Fig. – Locations of Starbucks (Green) and other Coffee shops (Orange) in Toronto.

- Starbucks global market planning department and Starbucks’ local representatives of Toronto are closing down stores mostly in core Downtown area.
- Toronto’s diverse suburban cityscape provides ample opportunities to open the new-format Starbucks
- Strategy –
 - Build a model using clustering machine learning algorithm to find new potential neighborhoods
 - Look for neighborhoods with high footfalls and low competitors (other coffee shops)



Data acquisition and cleaning

- Population demographics information (total population, income, population per age group) of each neighborhood was obtained from [Open Data Portal – City of Toronto](#).
- Geographical coordinates of the neighborhoods were extracted using the [OpenCage Geocoding API](#).
- Locations of Starbucks and other coffee shops retrieved using [Foursquare API](#).
- After data cleaning, we had -
 - 139 Neighborhoods and 22 features in the Neighborhoods Profile Data frame
 - 73 Neighborhoods have a total of 172 coffee shops, of which, 40 were Starbucks



KMeans algorithm - Clustering analysis

Elbow curve method

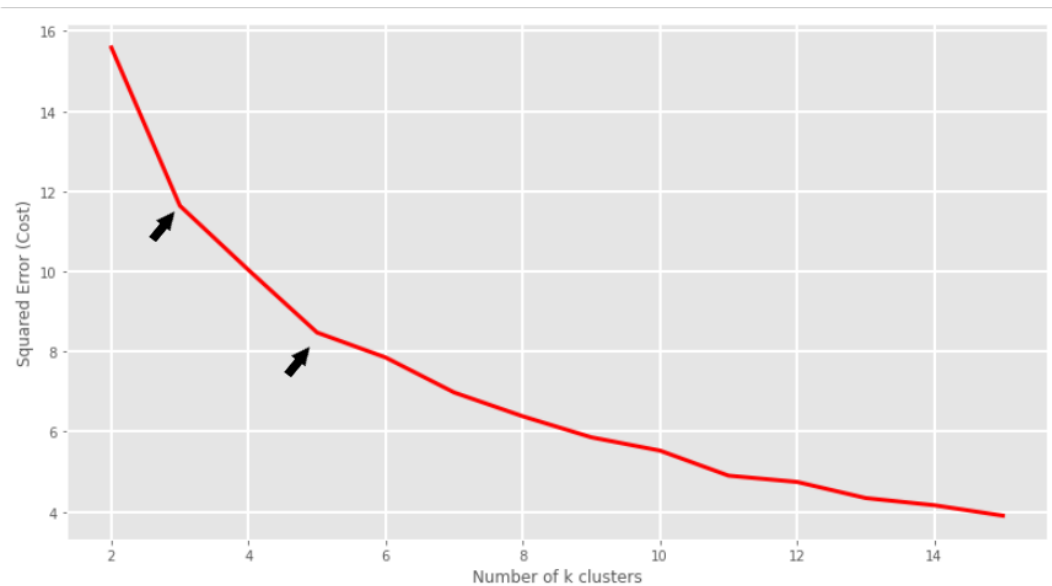


Fig. – Elbow curve method: elbow forms at both K=3 and K= 5.

- K=5 was found to be the optimal number of clusters for our study analysis
 - High silhouette scores, and
 - Even distribution of clusters

Silhouette curve method

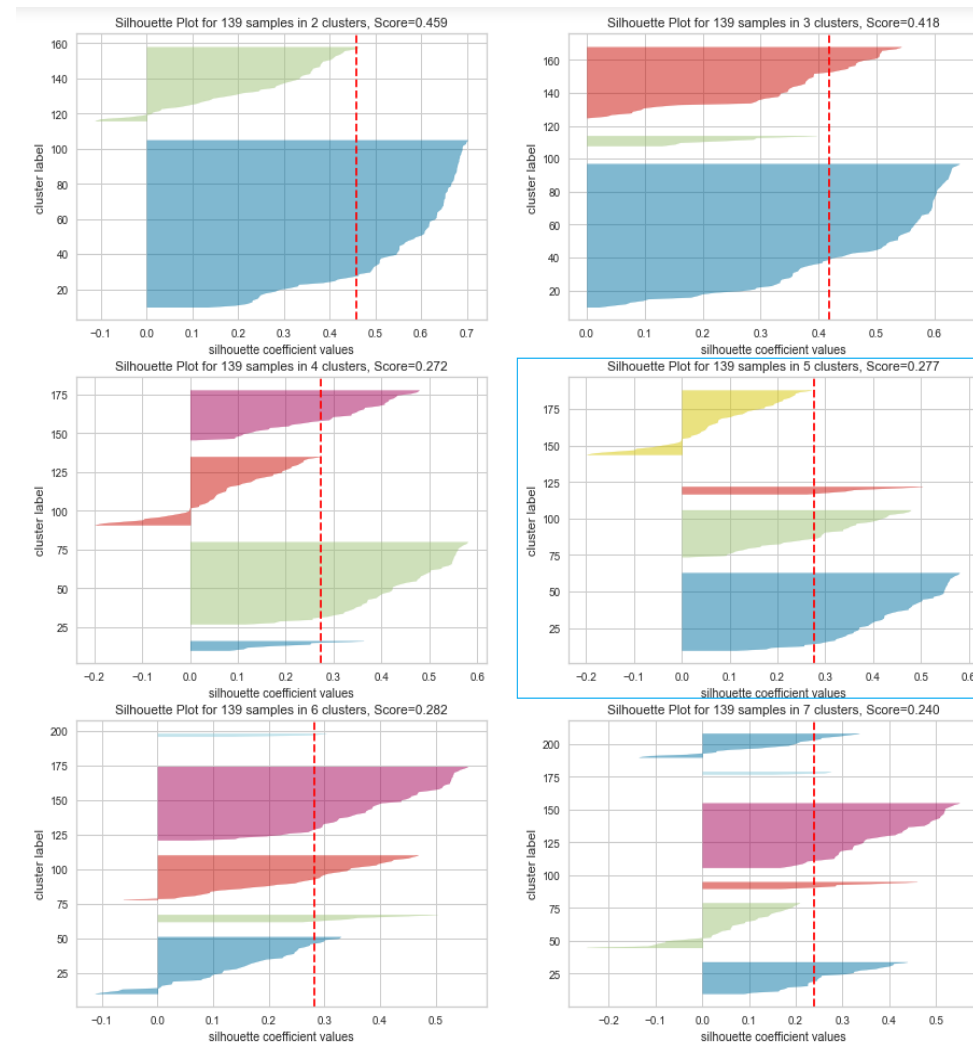


Fig. – Silhouette curve method: The clusters distribution is more uniform in K=5 than K=3.



Clustered neighborhoods

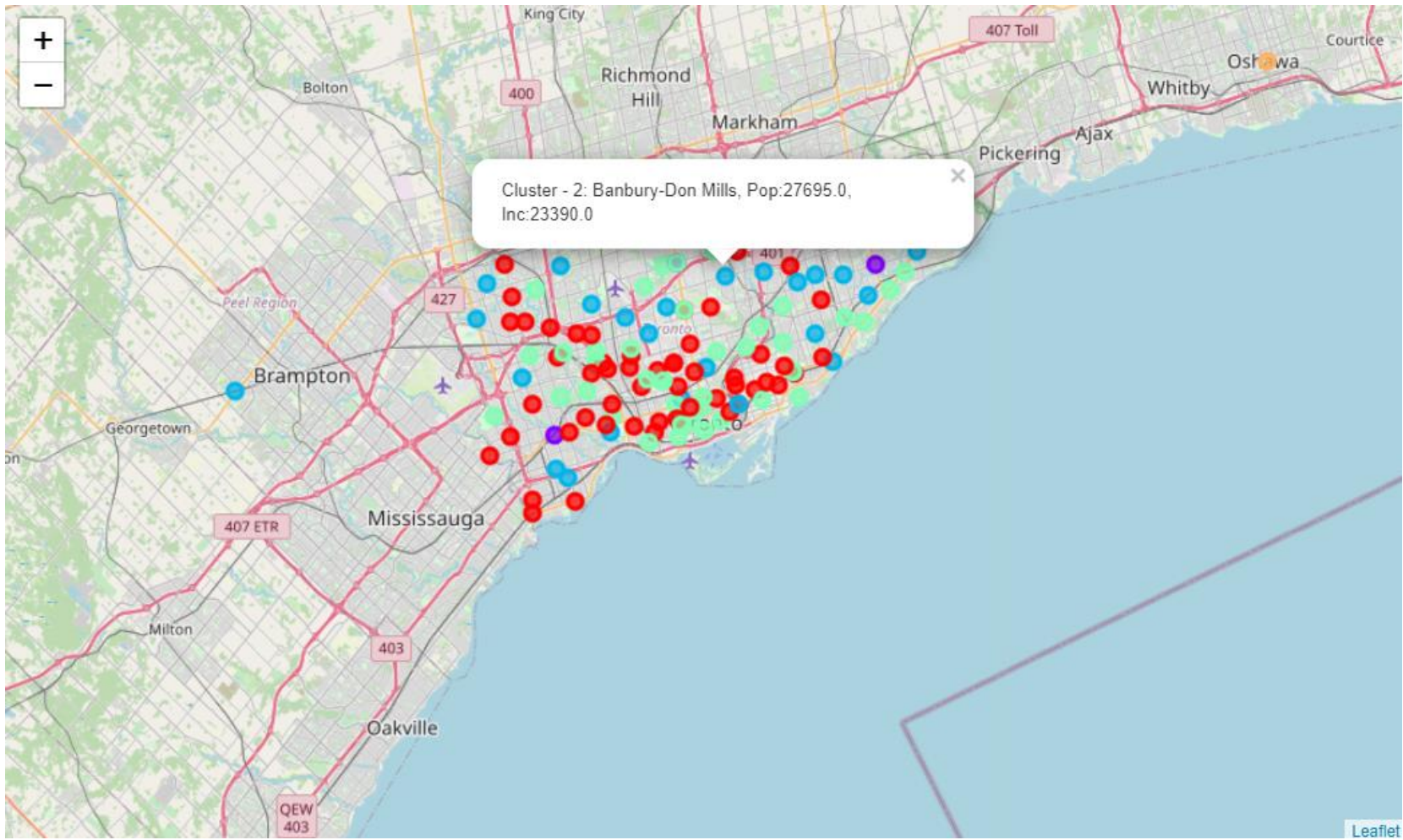


Fig. – Clustered neighborhoods of Toronto based on similarities in features: Cluster-0 (Red), Cluster-1 (Purple), Cluster-2 (Blue), Cluster-3 (Cyan), and Cluster-4 (Orange).



Cluster examination

Population

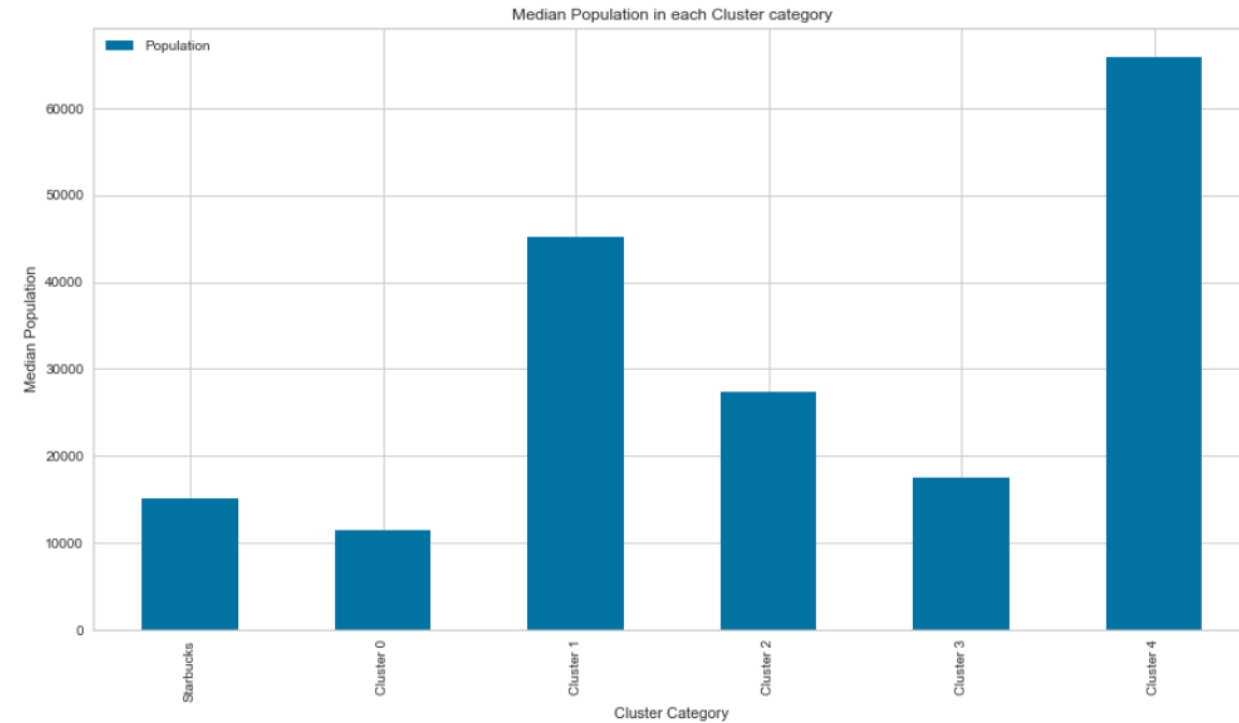


Fig. – Median population of neighborhoods with Starbucks and of each cluster.

- Population in clusters 0, 2, and 3 are closely related to neighborhoods with Starbucks.
- Clusters 4 have the highest median population.

Income

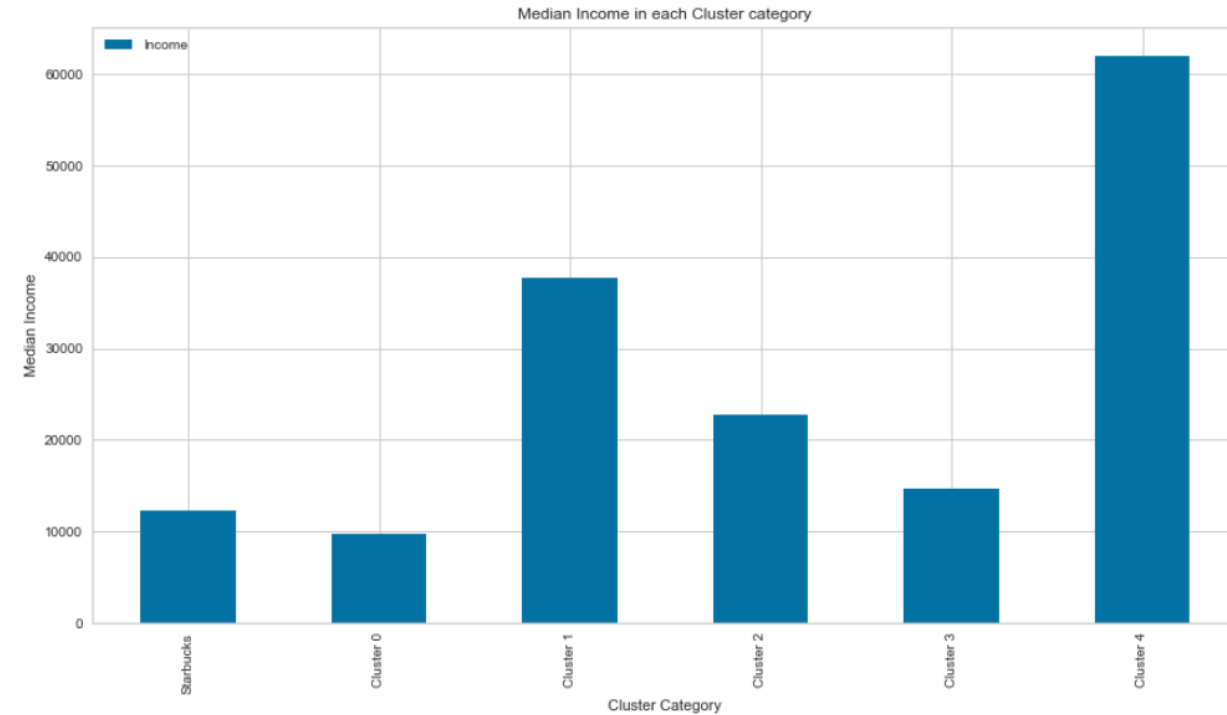


Fig. – Median income of neighborhoods with Starbucks and of each cluster.

- Clusters 0, 2, and 3 are low- and mid-income neighborhoods, similar to the neighborhoods with Starbucks.



Cluster examination

Population grouped by age groups

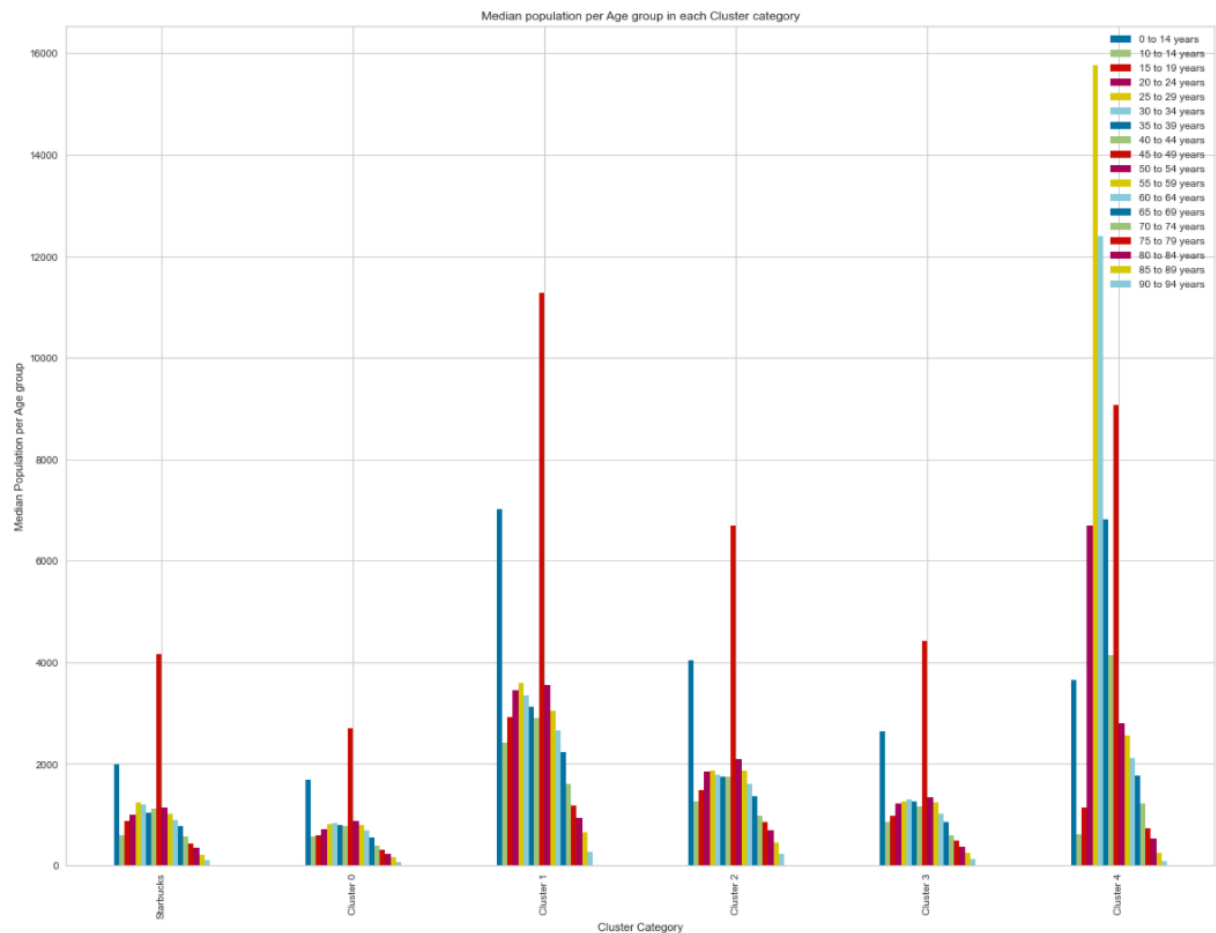


Fig. – Median population grouped by age groups of neighborhoods with Starbucks and of each cluster.

Competitors

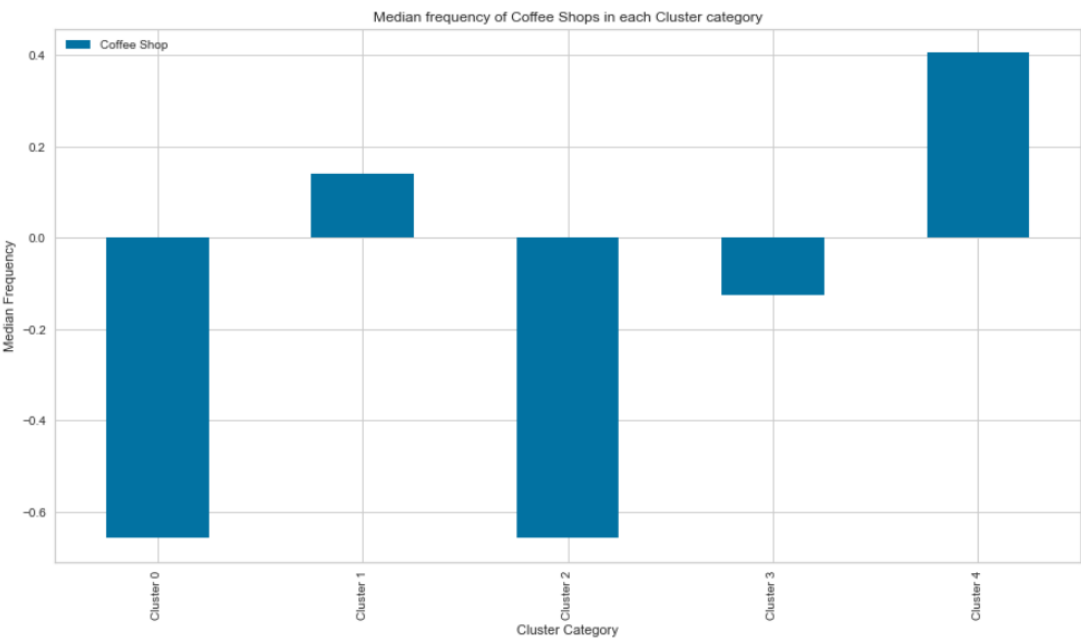
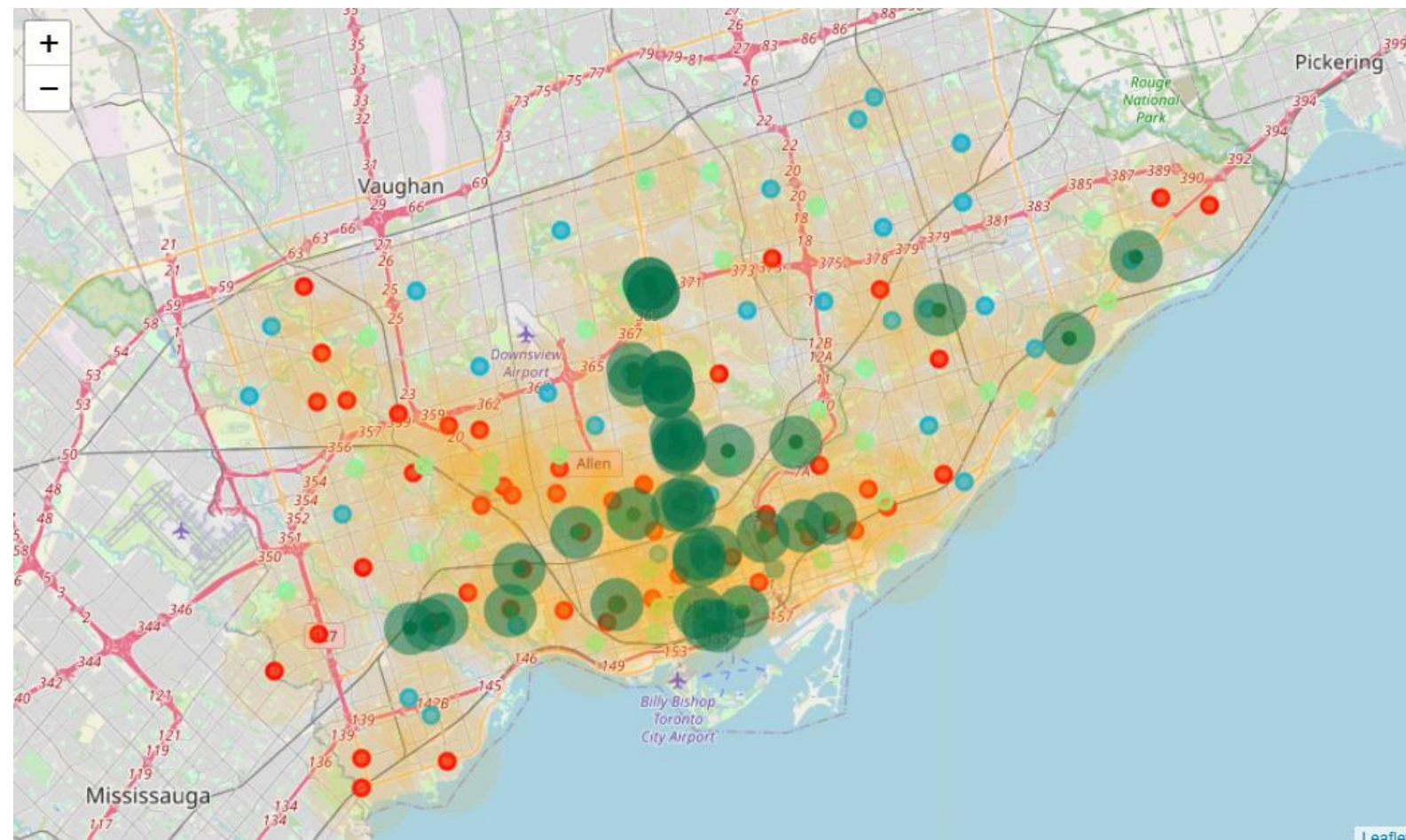


Fig. – Median frequency of coffee shops in each cluster.

- Clusters 0, 2, and 3 have low competitors.
- Clusters 0, 2, and 3 are the best sites to open the new format Starbucks considering all factors.



Potential locations of new Starbucks



- Orange shaded areas (not in the vicinity of current Starbucks locations) are the optimal neighborhoods to open new Starbucks stores.
- Most locations away from core Downtown Toronto.
- Selected neighborhoods have –
 - MID population
 - HIGH Middle-aged adults and MID Young adults
 - MID Spending power (income)
 - LOW competitors

Fig. – Locations of current Starbucks (Green), and potential sites for future expansion (Orange).



Conclusion

- Built a useful model to identify optimal neighborhoods in Toronto for Starbucks expansion.
- Similar strategy can be deployed for other Canadian cities and towns.
- Accuracy of the model has room for improvement.
- Future directions –
 - Include Starbucks customer data for better recommendation performance
 - Include user preferences, proximity to landmarks/public areas, and real estate availability
 - Create a hybrid recommendation engine to suggest specific Starbucks store format for different neighborhoods.

