1. **Introduction**

When you buy a house, you aren’t just buying a house. In a way, you’re buying a neighborhood.

This is a resident who is looking for houses or apartments in the boroughs of Toronto. This renter wants to find similar neighborhoods like the old one. This neighborhood may contain many places such as Grocery Store, Bank, Park, Shopping Mall, Restaurant, Pharmacy, Gym and School.

People have various need for neighborhood. A single man may prefer gym or bar. A housewife may consider school and grocery. When a resident moving, he may want to move to a neighborhood most like his old one.

This resident can assess new neighborhoods based on old neighborhood and find those most close options to make final decision.

1. **Data**

We will need geo-locational data about the neighborhoods in borough. Those data included neighborhoods’ name, latitude and longitude. If we know neighborhood names in the borough, we can use geopy library to get the latitude and longitude values of neighborhood. And we can use Folium to visualize those neighborhoods on map.

We will need venues data in neighborhoods. If we have the neighborhood name, latitude and longitude, we can use Foursquare to get venues data in the neighborhood in certain distance. The venues data included venue name, location, latitude, longitude, distance and PostalCode.

1. **Data Analyze**

Import the base neighborhood data with Borough name, latitude and longitude.

A screenshot of a cell phone

Description automatically generated

Figure3.1 Base Data

Use Folium to draw all neighborhood in map with their latitude and longitude.

A close up of a map

Description automatically generated

Figure 3.2 Base Folium Map

In order to get venues data, we need to use Foursquare base on Neighborhood name, latitude and longitude. With the URL established by Foursquare, we can get information of every venue in all neighborhoods. We only choose the name and category of venues.

A screenshot of a cell phone

Description automatically generated

Figure 3.3 Venues Data

With all venues data in each neighborhood, we can find the first 10 most common venue in each neighborhood.

A screenshot of a cell phone

Description automatically generated

Figure 3.4 Most Common Venue

1. **Predictive Modeling**

We choose K-means Clustering Model to cluster all neighborhoods. K-means Modeling can use in neighborhood segmentation. With K-means Modeling, neighborhoods can be clustered to k partitions. Each cluster have most similar features and same label.

A screenshot of a cell phone

Description automatically generated

Figure 4.1 Cluster Labels

With cluster labels with neighborhoods, we can use Folium to draw a clustered map of neighborhoods. Each color represents one cluster. There are 5 clusters since we choose k=5. If the resident has other requirements, we can change the value of k to get other results.

**A close up of a map

Description automatically generated**

Figure 4.2 Labeled Folium Map

1. **Conclusions**

After data analyze and predictive modeling, we have the labeled map. When the resident plan to move, he should know his old neighborhood belongs to which cluster and to find other neighborhoods in the same cluster. The same labeled neighborhood are those have most similar venues compare to the old neighborhood.