Introduction

Finding a better place or neighborhood to start a business is a bigger problem than the business itself. In this world, full of rapid growth and submerged in technology finding a neighborhood is not a difficult task.

You can easily get the ratings, locality information, neighbourhood details, nearby stores at your fingertips.

But getting a reliable and safe neighbourhood is a difficult task to handle. There are many vendors who are in search of a better place and can be a tough competitor to the Businesses. Hence targeting those places which have lower supply of what your Business has to offer and lower competitions from the other vendors, is an essential task and a reckless to do from the vendors side.
Hence, this project wishes to explore those places which you should target to grow your Business a nd target the customers.

Business Problem

- Explore the given geographic location and suggest a specific spot in order to maximize the impact of the Business on the targeted audience.
- Compare the Neigbhourhood of the given location with big cities viz. Toronto, New York and cluster them and suggest what could be the similarity between them.
- Also specify what sort of business type will be best suitable for these localities.

Target Audience

The model wishes to answer the specific geographic spot which will turn out to be the maximum profit for their businesses. The model has its target audience as

- Small Vendors
- Housewife's who wish to start a career
- Small startup groups
- Student willing to start something to fun their study
- Large scale retailers, willing to sell in different localities

Data

The data used analysis and recommendation contains the neighborhood details for the New York City and Toronto City.

Data Provider

For the clustering of the New York neighborhood and segmenting them, the data is available here

Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs a nd the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood.

For the Toronto Data Set, the data is provided through:

- Postal Code and Neighborhood: Wikipedia
- Co-Ordinates of the Neighborhood: Geospatial Data
- Localities of Each Neighborhood: Foursquare APIs

Data Acquisition

```
The data is available in the form of the table on the wiki page here.
The data is scraped from the given web page and is transformed from the table
into a data frame which is readily available for segmentation.
The table tag has a class associated with it, as class = “wikitable sortable”
```

Data Set Summary

.

The data set contains details of neighborhood, latitudes, longitudes, postal codes for each row/field. To explore the data around each venue, we utilized the four-square APIs.

Four Square APIs, have several endpoints groups which contains several endpoints for the scripts to send request to. The endpoints groups are venues, explore, search, etc.

The venues API endpoint is used to get all the relevant venues around the given latitude-longitude coordinates.

The parameters which forms the URL are:

- LIMIT
- Radius
- Client ID
- Client Secret
- Latitude
- Longitude

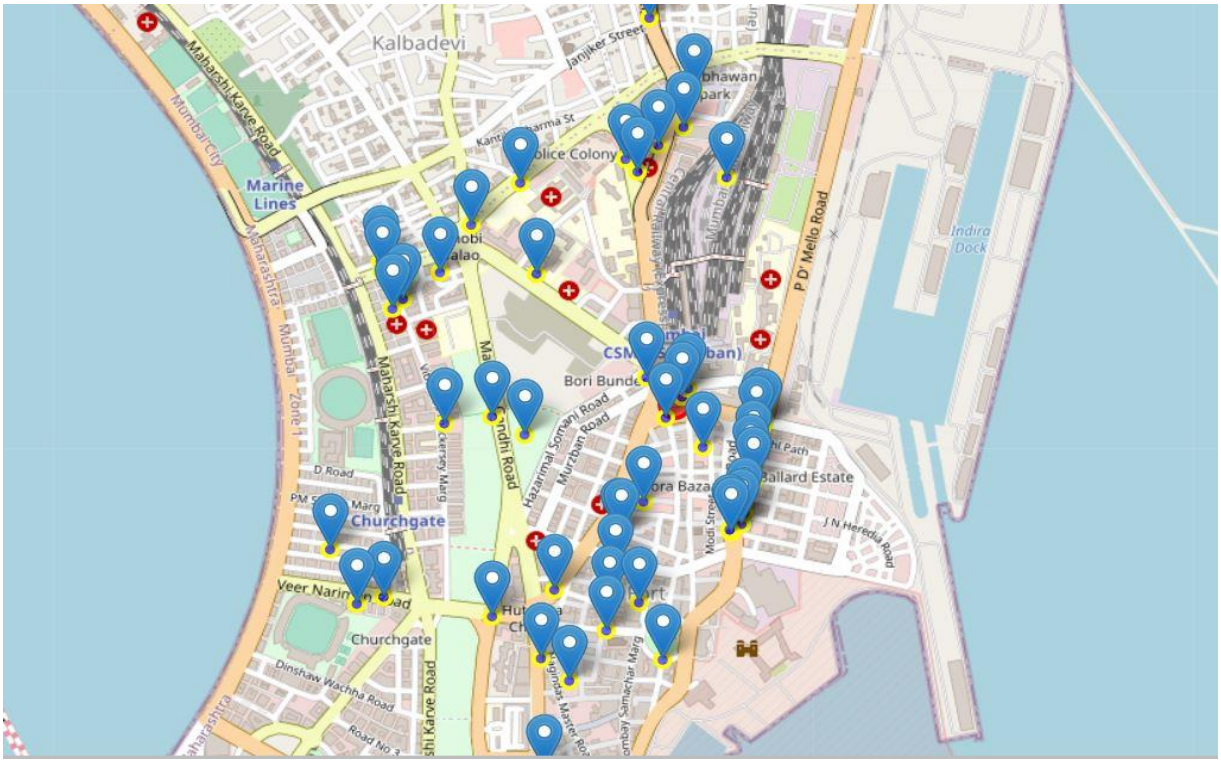
The venues returned are present in the venue list which is present in the items key of the response key.

The result is returned in the form of JSON object which can easily be transformed to data frame, and then the relevant information can be extracted from it.

Exploratory Data Analysis

Visualization of Toronto Neighborhood

Visualization of Mumbai Venues

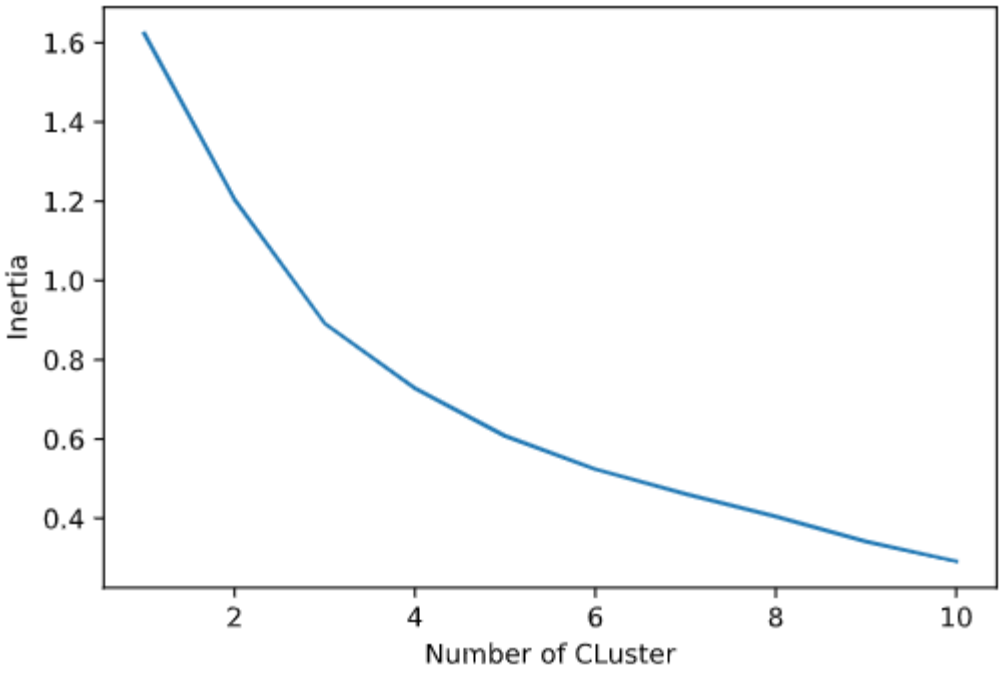


Modeling

The clustering of neighborhood is done using K - Means Clustering Algorithm.

Choosing the Best Number of Cluster

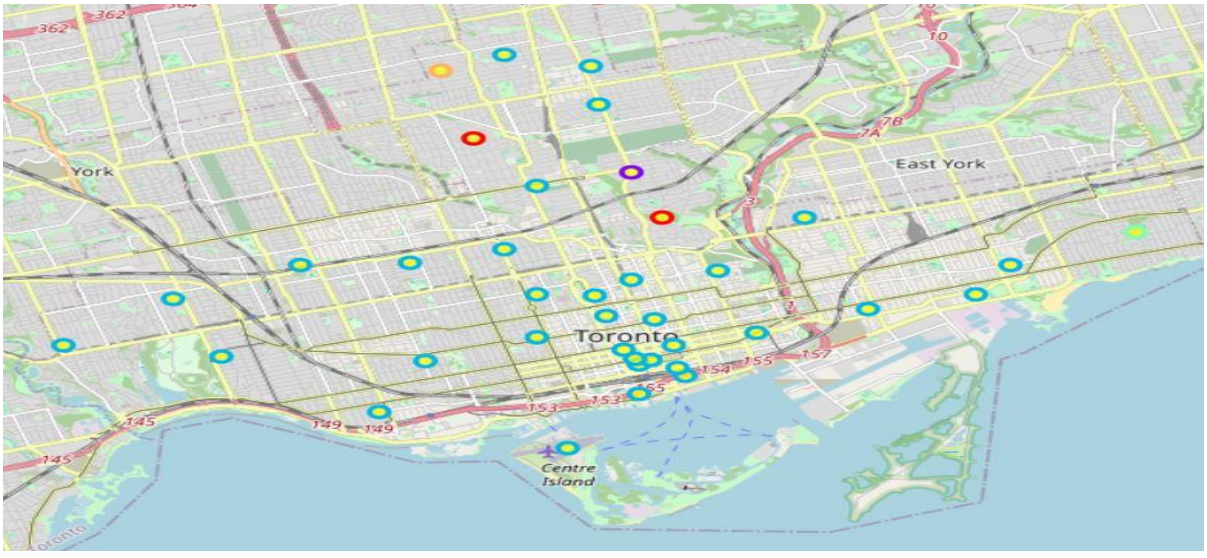
The right number of clusters is necessary to build a clustering model. The model uses the elbow method to determine the within cluster sum of square using the inertia parameter of the K - Means algorithm. The best number of clusters is choses from the plot.



Clearly from the elbow method number of clusters should be 5, as from the graph increasing the number of clusters further will not decrease the within cluster sum of squares much.

Model Evaluation

Clustering of the Toronto Neighborhood



Clustering of Mumbai Venues

For the Toronto Region:

- Total number of Neighborhood belonging to cluster 0 are 3
- Total number of Neighborhood belonging to cluster 1 is 1
- Total number of Neighborhood belonging to cluster 2 are 38
- Total number of Neighborhood belonging to cluster 3 are 1
- Total number of Neighborhood belonging to cluster 4 is 1
  Total venues: 44

For the Mumbai Region:

- Total number of Neighborhood belonging to cluster 0 are 12
- Total number of Neighborhood belonging to cluster 1 is 10
- Total number of Neighborhood belonging to cluster 2 are 17
- Total number of Neighborhood belonging to cluster 3 are 6
- Total number of Neighborhood belonging to cluster 4 is 5
  Total venues: 50