# Recommending neighborhood for opening restaurant

Anup Saurav

Jan 2,2020

# 1. Introduction

# 1.1 Background

# A global food chain wants to open a restaurant in Mumbai and is looking for suitable location/neighborhood. Location (nearby venues) is one of the most important factors to consider for opening a restaurant.

# **Location**: Wherever you decide to set up shop, be sure your restaurant is visible. It should be easily spotted from the street on which it resides.

# 1.2 Problem

# To Identify the neighborhood in Navi Mumbai to set up the restaurant and the business should to be **profitable** based on the neighborhood recommendation.

# 1.3 Interest

# Obviously, Food chain **management** will be interested in the business proposition as it would yield profit to them.

# 2. Data acquisition and cleaning

# 2.1 Data sources

# As part of the **Digital India initiative**, Govt of India has released dataset for All India Pincode Directory containing all the pin-code list across India with other relevant information like Division, Region, Circle, Taluk, Districts, States. The dataset has total 9 features.

# **Link for reference :** [**https://data.gov.in/resources/all-india-pincode-directory**](https://data.gov.in/resources/all-india-pincode-directory)

# The dataset does not have the latitude and longitude co-ordinates for the pincode. We must use a suitable API like geopy to get the location co-ordinates.

# The data was filtered based on the project requirement into 3 columns: **Pincode**, **Borough** and **Neighborhood**.

# 2.2 Data cleaning

Dataset has 155599 rows and 9 columns or features.

# 2.3 Feature selection

# 3. Exploratory Data Analysis

# 3.1 Calculation of the co-ordinates

# Using Opencage API, latitude and longitude co-ordinates were fetched for the corresponding Pincode.

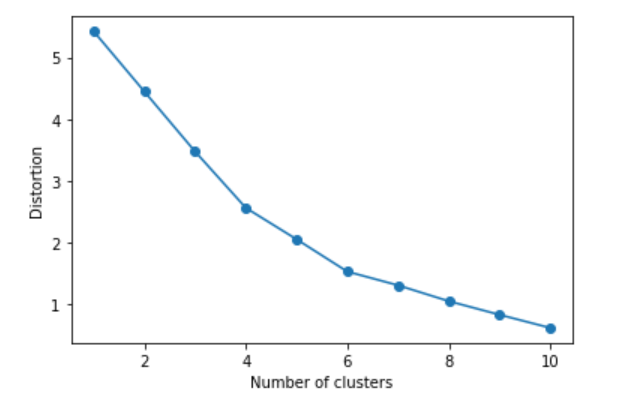
# Using the Foursquare Explore API, nearby venues were explored, and most common venues chosen for each neighborhood. This will make use of the latitude, longitude, CLIENT\_ID, CLIENT\_SECRET, VERSION

# 4. Clustering

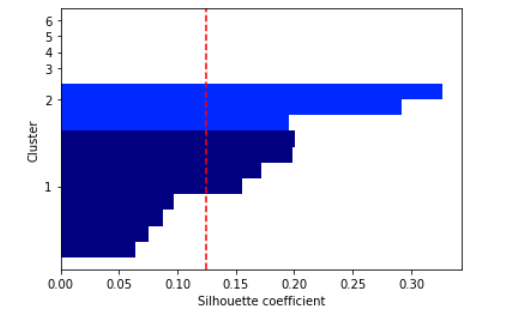
* Using K-Means Clustering algorithm, neighborhood was clustered based on the discriminating venue categories. This unsupervised machine learning technique was performed as the resulting data does not have any labels.
* Using the elbow method to find the optimal number of clusters.

# 5. Results

**Using the elbow method to find the optimal number of clusters**



**Quantifying the quality of clustering via silhouette plots**



we can see that the silhouette coefficient is close to 0 meaning the cluster separation and cohesion are equal. Ideal value would be close to 1.

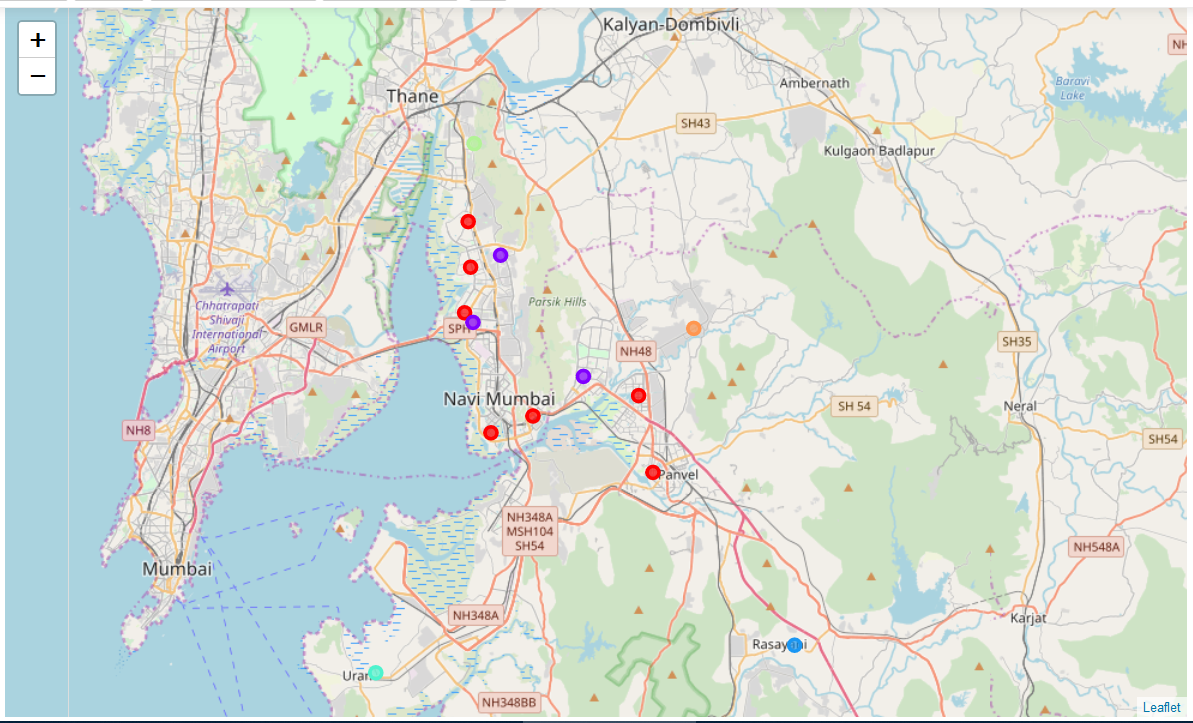
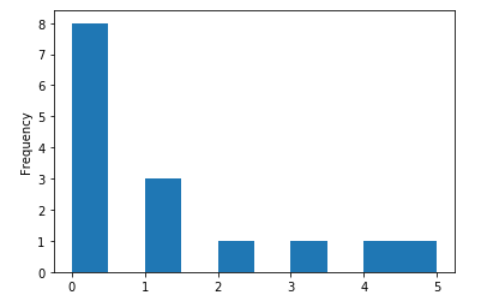


Figure showing cluster mapping



Cluster 2,3,4,5 each has only one observation/data point , so can be ignored.

# 6. Discussion directions

cluster 0 analysis



cluster 1 analysis:



# 7. Conclusion

Food chain can go either with cluster 0 or 1 for opening the restaurant as the nearby venues have similar category.

silhouette coefficient is bounded in the range -1 to 1 and is a measure of how tightly grouped the samples in the clusters are . Our silhouette analysis shows coefficient close to 0.12, which is far from idle value of 1, indicating loose grouping.

# 8. Future directions

We can include some other features for better neighborhood prediction.

# Links:

# <https://www.business.com/articles/7-things-to-consider-restaurant/>