

# **CAPSTONE PROJECT**

Finding suitable neighborhood for opening a restaurant

# Predicting neighborhood for opening restaurant

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.

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

Food chain **management** has hired a Data Science firm to Identify the neighborhood in Navi Mumbai to set up the restaurant and wants the business to be **profitable** based on the neighborhood recommendation.

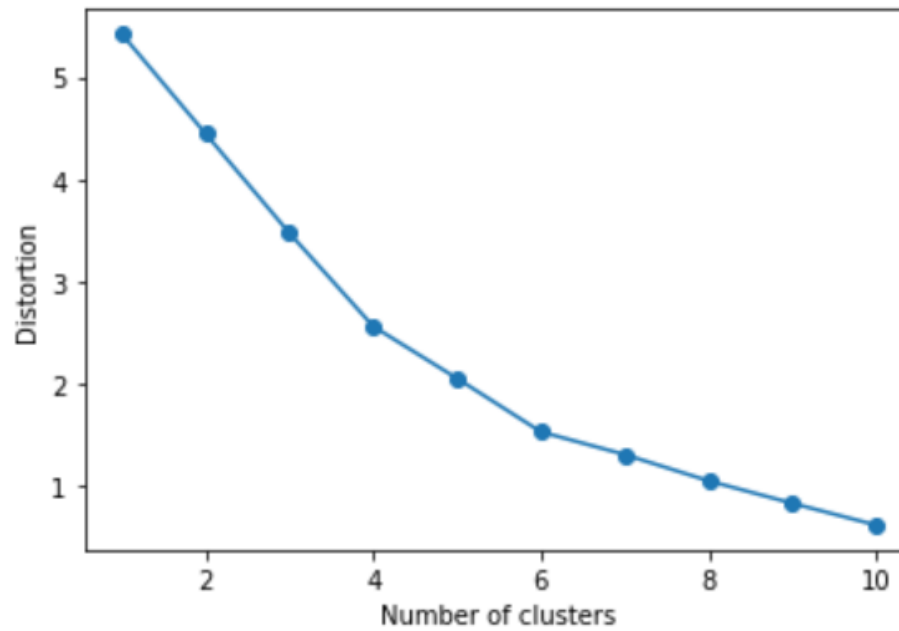
# Data acquisition and cleaning

- 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>
- The dataset does not have the latitude and longitude co-ordinates for the pincode. We have to 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** (or postal code) , **Borough** and **Neighborhood**.

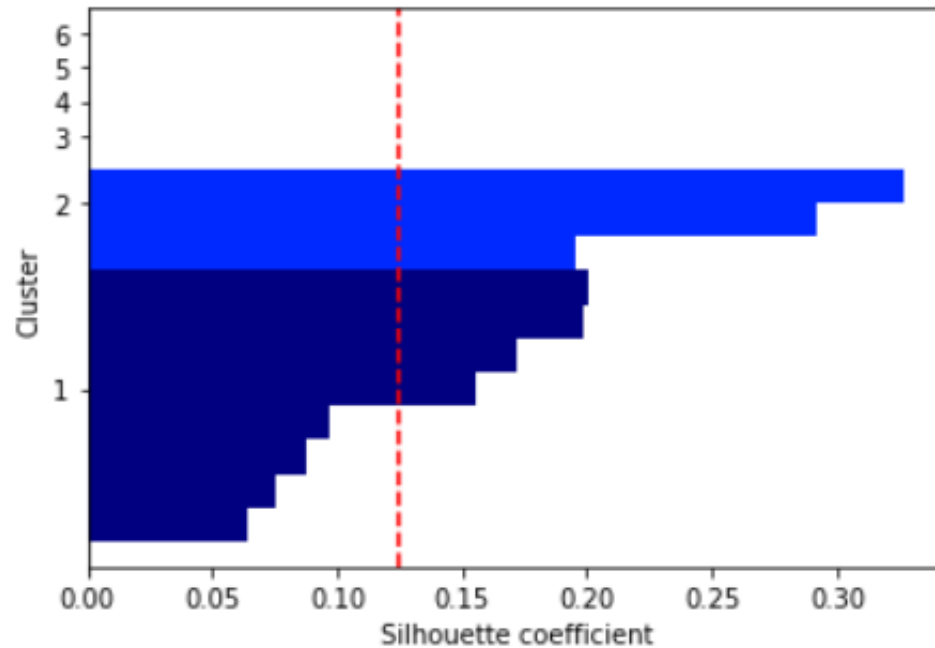
# Methodology

- 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.`
- Using K-Means Clustering algorithm, neighborhood were 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.

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

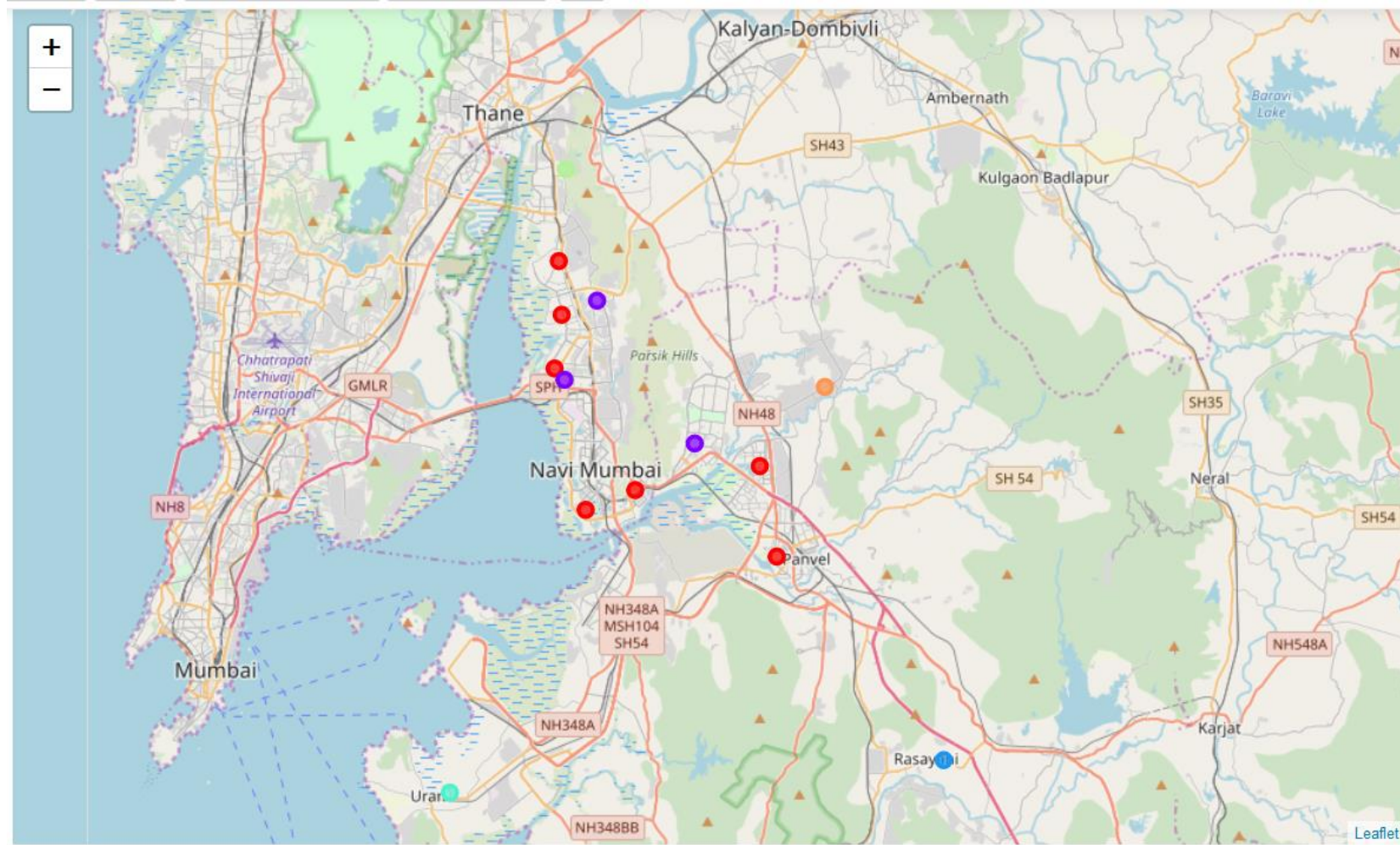


# Quantifying the quality of clustering via silhouette plots

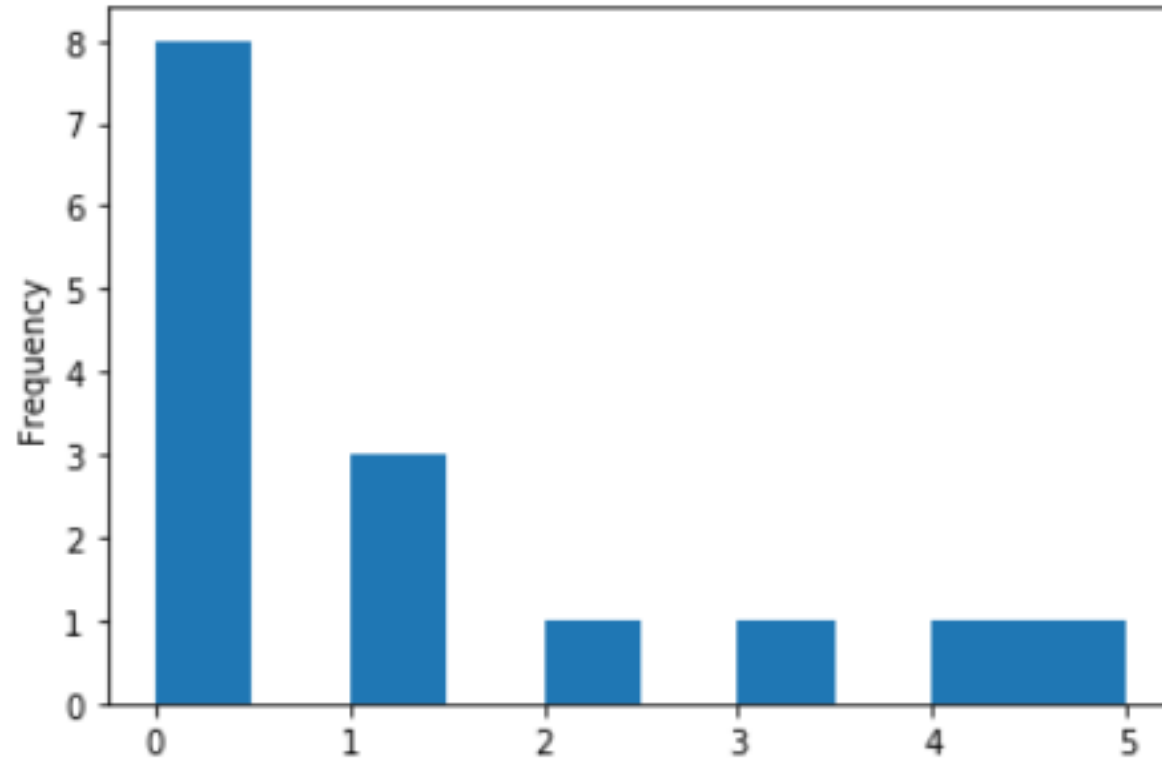


we can see that the silhouette coefficient is close to 0 meaning the cluster separation and cohesion are equal. Ideal value would be close to 1.

# Cluster mapping



# Histogram of clusters



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



# Cluster 1 analysis

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Navi Mumbai Division	1	<u>Indian Restaurant</u>	<u>Fast Food Restaurant</u>	Market	Farmers Market	Smoke Shop	Hotel	<u>Chinese Restaurant</u>	<u>Middle Eastern Restaurant</u>	<u>English Restaurant</u>	Electronics Store
10	Navi Mumbai Division	1	<u>Indian Restaurant</u>	Food Court	Hotel	<u>Restaurant</u>	Coffee Shop	<u>Fast Food Restaurant</u>	Farmers Market	<u>English Restaurant</u>	Electronics Store	Diner
20	Navi Mumbai Division	1	Gym	<u>Indian Restaurant</u>	<u>Vegetarian / Vegan Restaurant</u>	Coffee Shop	<u>Fast Food Restaurant</u>	Farmers Market	<u>English Restaurant</u>	Electronics Store	Diner	Dessert Shop

# Cluster 0 analysis

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Navi Mumbai Division	0	Fast Food Restaurant	Train Station	BBQ Joint	Bed & Breakfast	Bus Station	Indian Restaurant	Department Store	Fish Market	Farmers Market	English Restaurant
1	Navi Mumbai Division	0	Train Station	Grocery Store	Bus Station	Juice Bar	Vegetarian / Vegan Restaurant	Department Store	Fast Food Restaurant	Farmers Market	English Restaurant	Electronics Store
3	Navi Mumbai Division	0	Vegetarian / Vegan Restaurant	Café	Hotel	Dessert Shop	Indian Restaurant	Cocktail Bar	Middle Eastern Restaurant	Chinese Restaurant	Fast Food Restaurant	Theater
6	Navi Mumbai Division	0	Pizza Place	Café	Vegetarian / Vegan Restaurant	Train Station	English Restaurant	Sandwich Place	Indian Restaurant	Coffee Shop	Farmers Market	Electronics Store
9	Navi Mumbai Division	0	Department Store	Chinese Restaurant	Thai Restaurant	Diner	Lounge	Coffee Shop	Cocktail Bar	Fast Food Restaurant	Farmers Market	English Restaurant
15	Navi Mumbai Division	0	Food Truck	Bus Station	Electronics Store	Department Store	North Indian Restaurant	Movie Theater	Bistro	BBQ Joint	Fast Food Restaurant	Farmers Market
17	Navi Mumbai Division	0	Asian Restaurant	Snack Place	Ice Cream Shop	Flea Market	Vegetarian / Vegan Restaurant	Coffee Shop	Fast Food Restaurant	Farmers Market	English Restaurant	Electronics Store
22	Navi Mumbai Division	0	Toll Plaza	Fish Market	Breakfast Spot	Bus Station	Vegetarian / Vegan Restaurant	Coffee Shop	Fast Food Restaurant	Farmers Market	English Restaurant	Electronics Store

# Conclusion and future directions

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 ideal value of 1, indicating loose grouping.

We can include some other features for better neighborhood prediction.

THANK YOU