Capstone Project

Determining Ideal Location to Open a Coffee Shop in Mumbai

Purpose of the project

- To decide on a location to open a coffee shop in Mumbai
- To be able to make good business decisions using data science
- To give small and medium size business a competitive edge by using data to make smart decisions
- Increase the profitability of business by choosing good locations.

Methodology of the project

- Using beautifulsoup to scrape suburbs data
- Using geocoder and geopy to get the coordinates
- Using Foursquare API to get coffee shop data
- K-means clustering for getting insights from the venue data

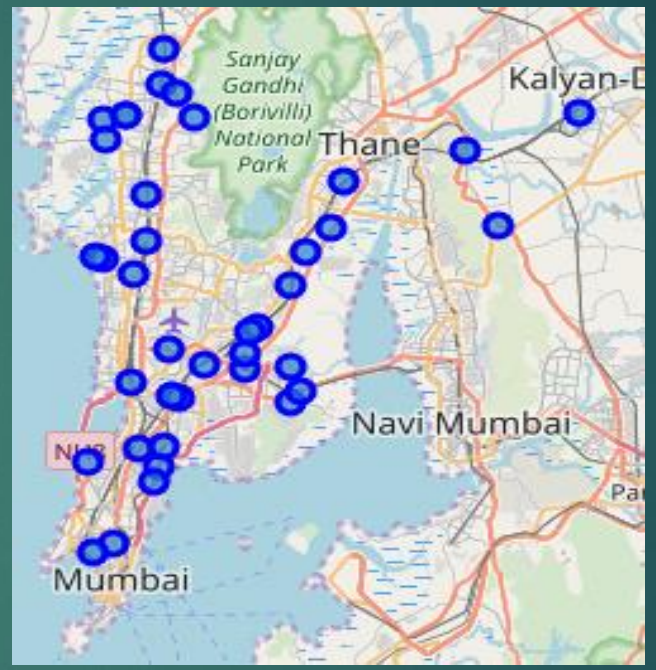
Data Source & Acquisition

- List of suburbs in Mumbai was acquired from a Wikipedia page using BeautifulSoup
- The latitude and longitude values was acquired using geopy and geocoder
- It was merged together using pandas

	Neighborhood	Latitude	Longitude
0	Andheri	19.118483	72.841774
1	Anushakti Nagar	19.042830	72.927340
2	Baiganwadi	19.062930	72.926660
3	Bandra	19.054220	72.840190
4	Bhandup	19.145560	72.948560

Visualization of locations data using folium

- The latitude and longitude areas of the suburbs was used to plot the locations on the map using folium



Venue data for Coffee shops

- Foursqaure API was used to get the coffee shop data for each suburb
- We groupby the data on neighbourhood to get the count of coffee shops in each suburb

Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
Andheri	2	2	2	2	2	2
Anushakti Nagar	1	1	1	1	1	1
Baiganwadi	1	1	1	1	1	1
Bandra	2	2	2	2	2	2
Borivali	2	2	2	2	2	2
Charkop	3	3	3	3	3	3

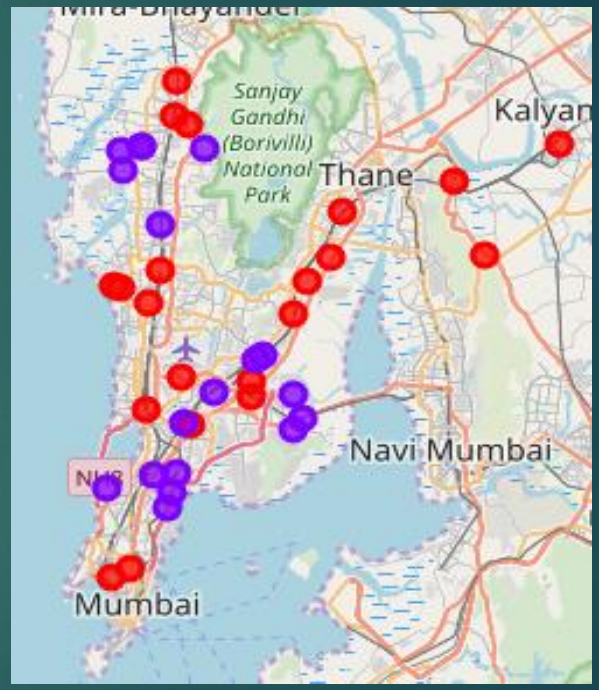
Venue data for Restaurants

- Foursqaure API was used to get the restaurant data for each suburb
- We groupby the data on neighbourhood to get the count of restaurant in each suburb
- We will use restaurant data for the analysis as well since having many restaurants near the vicinity of the coffee shop can effect the profitability.

Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
Bandra	2	2	2	2	2	2
Bhandup	2	2	2	2	2	2
Borivali	6	6	6	6	6	6
Charkop	1	1	1	1	1	1
Chembur	3	3	3	3	3	3
Dahisar	3	3	3	3	3	3

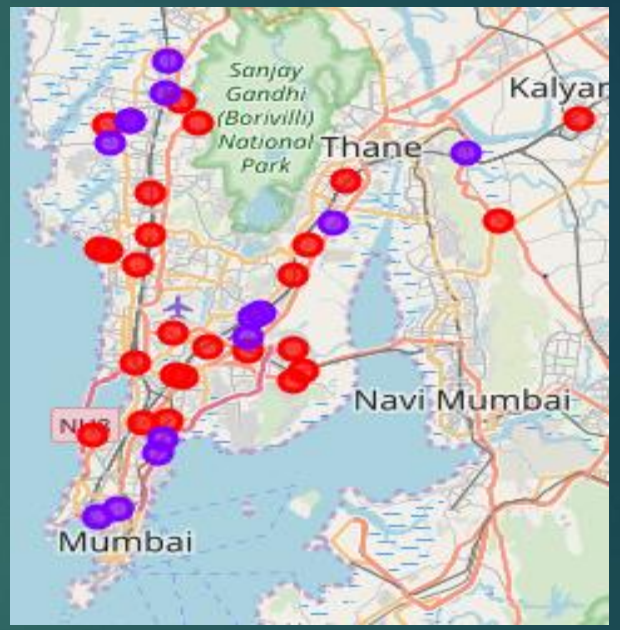
Visualization of Coffee Shop Clusters

- Cluster 0 (Red): Zero or lower number of coffee shops
- Cluster 1 (Purple): Higher number of coffee shops
- Locations with lower number of coffee shops will be more profitable due to lack of competition



Visualization of Restaurant Clusters

- Cluster 0 (Red): Zero or lower number of restaurants
- Cluster 1 (Purple): Higher number of restaurant
- Taken to further improve the location chosen for coffee shops



Ideal Locations

- The common locations in the cluster 0 from the coffee shop data and the restaurant data is being displayed
- This is done because the both coffee shops and restaurants have less frequency in those areas
- This will lead to lower competition and hence increases the profitability

	Suburb
0	Mulund
1	Kanjurmarg
2	Matharpacady, Mumbai
3	Andheri
4	Seven Bungalows
5	Kausa
6	Chembur
7	Mira Road
8	Jogeshwari
9	Devipada
10	Dombivli
11	Vikhroli
12	Shil Phata
13	Mogra Village
14	Bandra

Conclusion

- This analysis helps us in deciding ideal locations for coffee shops based on the competition in the area
- Similar method can be used for other venues as well
- This type of analysis is subject to the quality of data available
- More features can be used to improve the validity of the method