Clustering & Segmentation of The Most Popular City in India Mumbai

Introduction

Mumbai (formerly called Bombay) is a densely populated city on India's west coast. A financial center, it's India's largest city. On the Mumbai Harbour waterfront stands the iconic Gateway of India stone arch, built by the British Raj in 1924. Offshore, nearby Elephanta Island holds ancient cave temples dedicated to the Hindu god Shiva. The city's also famous as the heart of the Bollywood film industry. The population of the city is about 18.41 million.

For anyone starting or running a business it is very important to analyze the statistics of that city based on the area and surrounding locations. People usually spend time by going to restaurants, spas, shopping malls, hotels, cinema, cafes, etc.

Business Problem

The objective of the project is to analyze and select the best location in Mumbai City to open a business like a café/restaurant. Various data science tools and methods were used to determine the data.

The data can be used by retailers, businessmen or construction developers and investors who are looking to start a business such as gyms, hotels, spas, etc.

Mumbai is one of the most commercially viable and the rise in the standard of living has led to increase in potential business opportunities in the city.

Methodology

The first steps were to web scrape the pin code of different areas within the city. This website https://www.mapsofindia.com/pincode/india/maharashtra/mumbai/ was used to retrieve all area codes and list of neighborhood locations in Mumbai.

Once the locations were retrieved the next steps was to get the geographical coordinates of Mumbai city and the neighborhood locations. Python's geocoder package was used to to get the latitude and longitude coordinates of the locations.

Foursquare API was used to get the most visited data for all the locations.

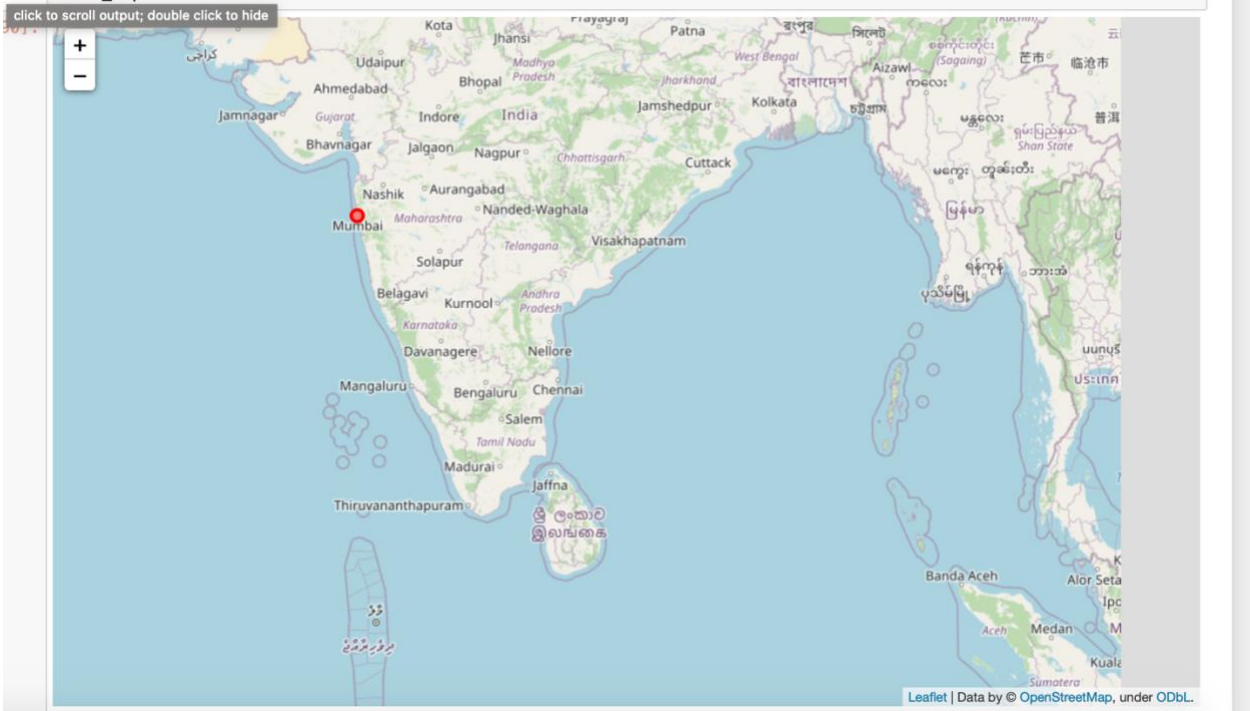
After performing the data cleaning and analysis, clustering method was used on the data.

The clustering method gives k=14 as the optimal number of clusters.

Results

Visualization of Mumbai City using geocoder library

```
world_map
```



The retrieved data from Foursquare API has 14,808 observations

```
(14808, 7)
```

:

	Location	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	A I staff colony	19.166115	72.859403	The Westin Mumbai Garden City	19.172654	72.860518	Hotel
1	A I staff colony	19.166115	72.859403	PVR Cinemas	19.174016	72.860485	Multiplex
2	A I staff colony	19.166115	72.859403	The Fern - An Eco Hotel	19.162362	72.856972	Hotel
3	A I staff colony	19.166115	72.859403	Starbucks	19.174177	72.860350	Coffee Shop
4	A I staff colony	19.166115	72.859403	MadOverDonuts	19.173902	72.860185	Donut Shop

There are 233 Unique Venue Category's in Mumbai

```
'Sports Club', 'Dim Sum Restaurant', 'Irani Cafe', 'Salad Place',
'Goan Restaurant', 'Brazilian Restaurant', 'Noodle House',
'South American Restaurant', "Men's Store", 'Steakhouse',
'Mexican Restaurant', 'Burrito Place', 'New American Restaurant',
'Shoe Store', 'Dumpling Restaurant', 'French Restaurant',
'Fish & Chips Shop', 'Road', 'Australian Restaurant',
'Spanish Restaurant', 'Harbor / Marina', 'Stadium',
'Other Great Outdoors', 'Opera House', 'Racetrack',
'Movie Theater', 'Track', 'Recreation Center', 'Roof Deck',
'Molecular Gastronomy Restaurant', 'Historic Site', 'Pet Store',
'Sporting Goods Shop', 'Lighthouse',
'Residential Building (Apartment / Condo)', 'Campground',
'Liquor Store', 'Big Box Store', 'Fried Chicken Joint',
'Cricket Ground', 'College Academic Building', 'Hockey Arena',
'Parsi Restaurant', 'Bridal Shop', 'Chaat Place', 'Comedy Club',
'Office', 'Cosmetics Shop', 'Toy / Game Store',
'Comfort Food Restaurant', 'Building', 'Monument / Landmark',
'Bike Rental / Bike Share', 'Auto Dealership', 'Soccer Field',
'Buffet', 'Frozen Yogurt Shop', 'Convention Center', 'Kids Store',
'Accessories Store', 'Pool Hall', 'Boat or Ferry', 'Theme Park',
'Airport Terminal', 'Airport Food Court', 'Indoor Play Area',
'Pier', 'College Gym', 'Camera Store', 'Neighborhood', 'Wine Shop',
'Platform', 'General College & University', 'Motorcycle Shop',
'Greek Restaurant', 'Trail', 'Recording Studio', 'Hotel Pool',
'Hostel', 'Duty-free Shop', 'Fish Market', 'Sculpture Garden',
'Lake', 'Toll Booth', 'Korean Restaurant', 'Beer Bar',
'Science Museum', 'Hyderabadi Restaurant', 'Tunnel', 'Train',
'Supermarket'], dtype=object)
```

scroll output; double click to hide

```
nt('There are {} unique categories.'.format(len(venues_df['VenueCategory'].unique())))
```

```
re are 233 unique categories.
```

Example of Top 5 Venue's in a Location

```
    print('\n')
```

```
------------ Central Building ------------
           Venue Cat  Freq
0  Indian Restaurant  0.13
1              Café  0.05
2            Bakery  0.05
3            Lounge  0.04
4        Coffee Shop  0.04


------------ Century Mill ------------
             Venue Cat  Freq
0     Indian Restaurant  0.26
1                 Café  0.06
2    Chinese Restaurant  0.04
3         Train Station  0.04
4  Fast Food Restaurant  0.03
```

```
1]: def return_most_common_venues(row, num_top_venues):
```

Top 10 Venue's in Each Location

Top 10 Venues in each Location

```
41]: venues_sorted.head()
```

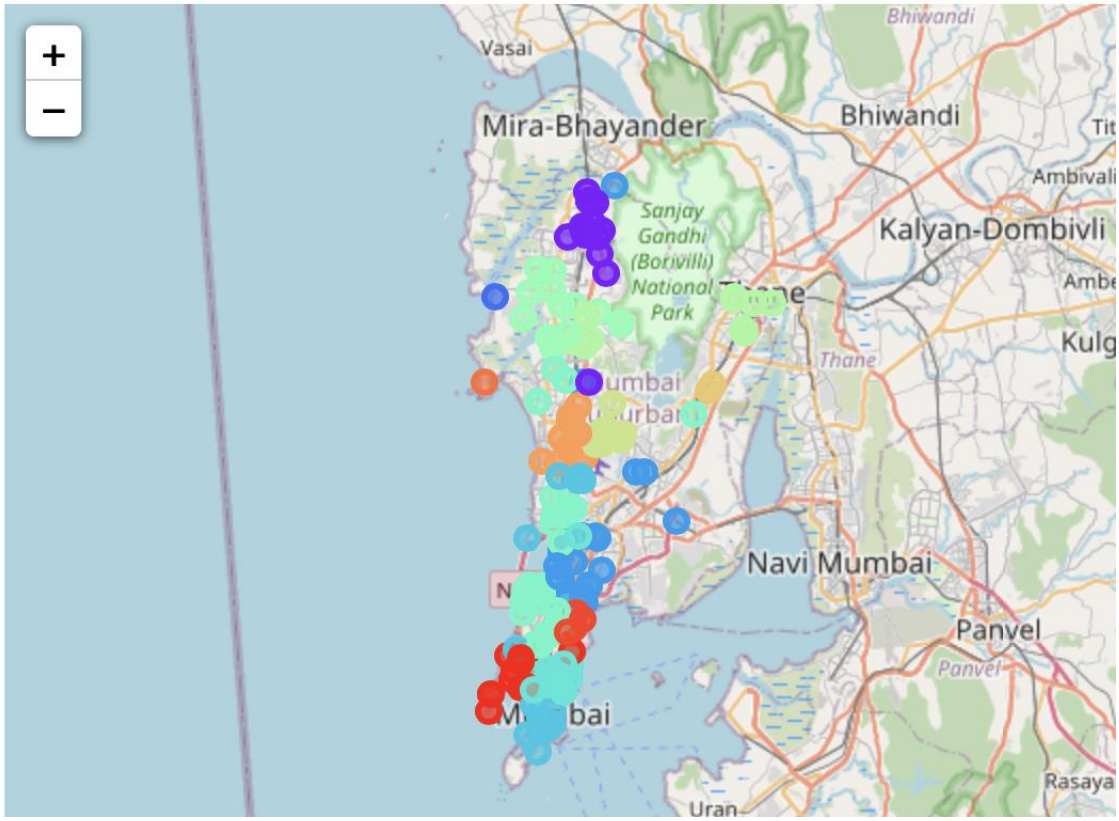
41]:

	Cluster Labels	Location	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	9	A I staff colony	Indian Restaurant	Fast Food Restaurant	Coffee Shop	Gym / Fitness Center	Vegetarian / Vegan Restaurant	Multiplex	Pizza Place	Sandwich Place	Bar	Bakery
1	0	Aareymilk Colony	Indian Restaurant	Restaurant	Bakery	Ice Cream Shop	Bar	Coffee Shop	Dessert Shop	Italian Restaurant	Art Gallery	Fast Food Restaurant
2	12	Agripada	Indian Restaurant	Café	Coffee Shop	Hotel	Sandwich Place	Chinese Restaurant	Fast Food Restaurant	Seafood Restaurant	Food Truck	Tea Room
3	7	Airport	Indian Restaurant	Café	Chinese Restaurant	Restaurant	Department Store	Lounge	Dessert Shop	Ice Cream Shop	Bar	Shopping Mall
4	12	Ambewadi	Indian Restaurant	Pub	Pizza Place	Ice Cream Shop	Bar	Chinese Restaurant	Snack Place	Fast Food Restaurant	Café	Seafood Restaurant

Cluster Analysis

```
map_clusters
```

[234]:



Observation

1. Cluster 11,12,13 and 14 have fewer restaurants
2. Indian Restaurants are the most popular in most locations
3. Hotels are most popular in cluster 10. This may be because the Airport falls in this sector
4. Fast Food restaurants are very common in cluster 9
5. Cluster 14 does not have a gym as a popular location. A franchise gym company can investigate whether this could be a good location to start a gym

Conclusion

The most popular visited places in Mumbai are analyzed through the data. I would like to expand this project in the future by including rent rates and other features in these areas which can help make the premise better. Machine learning methods can be used on these features to determine rent per square for business purposes.