The Battle of the Neighborhoods

1.Introduction

Mumbai, dream city of India is well known for tourist attraction, hence a lot of places to eat opens in this city of amazing history, architecture, art, canals, red lights, and, well, coffeeshop culture. Finding the best coffee shop and restaurant for tourist has become a hectic task due to lot of options. Hence here will will try to find out the nearest and best coffee shop for tourist as well as local people.

Thus it is strongly advised by the locals and experienced tourists to plan your coffeeshop experience.But how does one avoid performing a double search for such a winning combination? Our research is to fill this niche.

Business Problem

Different groups of people would benefit from our project's results, namely:

1.Potential restaurant owners who provide good quality of food and coffee.

2.Tourists planning their first experience as described above.

3.Mumbai visitors or even locals looking for new ideas for their following from-coffeeshop-to-restaurant tour.

To achieve this, we will create a short and simple guide on where to eat in Mumbai based on Foursquare likes, restaurant category and geographical location data for restaurants and coffeeshops. We will also cluster all the restaurants of Mumbai by their proximity to coffeeshops so that our user could easily determine what is the best duo of places of their interest.

2.Required Data

Data Source = Zomato Mumbai Dataset (from Kaggle)

For this assignment, we will be utilizing the Foursquare API to pull the following location data on restaurants and coffeeshops in Mumbai:

1.Venue Name

2.Venue ID

3.Venue Location

4.Venue Category

3. Methodology Section

The Methodology section comprises of below stages:

1. Collect Data and load into pandas dataframe

2. Data Cleaning & Pre-processing

3. Exploratory Data Analysis

4. Use Foursquare API to get neighbourhood venues

5. Modelling using k-means clustering based on common neighbourhoods

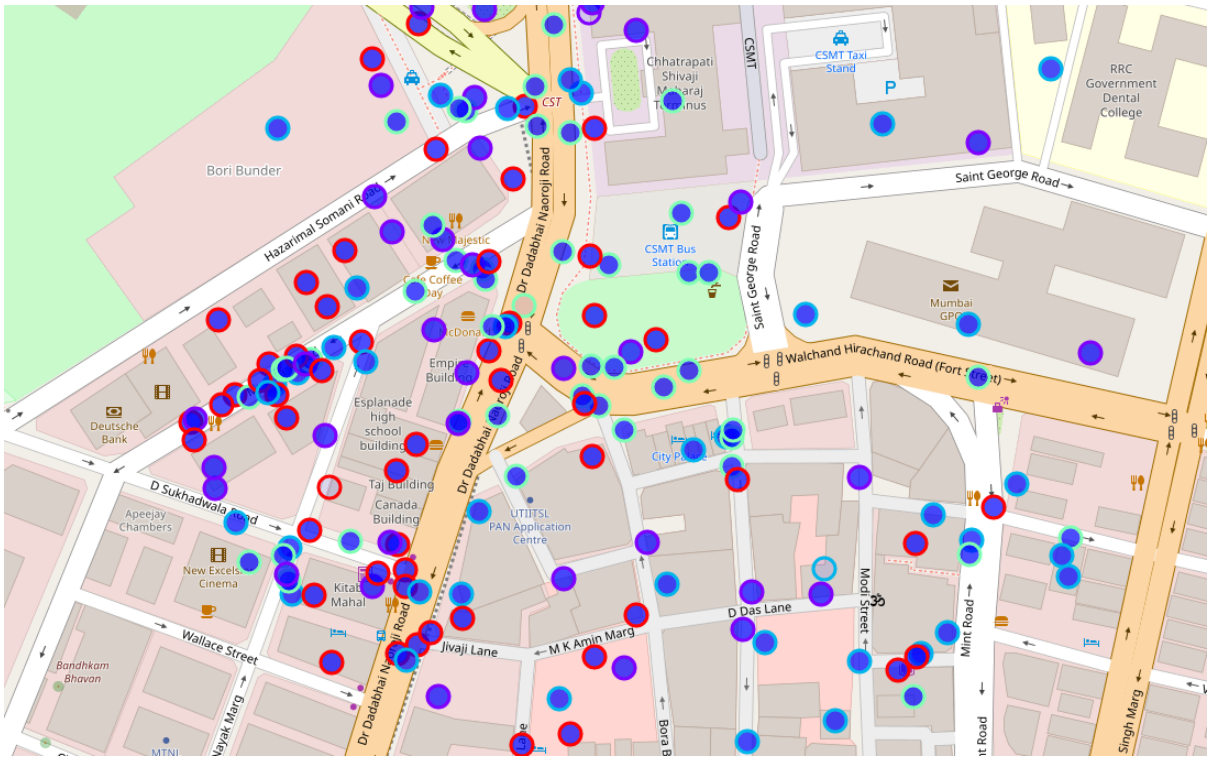
6. Derive conclusions from the model by analysing each cluster

4.Foursquare API (To get venues in each neighbourhood)

Venue	Venue Latitude	Venue Longitude
Shivala	18.938517	72.835514
CST Bus Stand	18.939009	72.835125
Chhatrapati Shivaji Maharaj Terminus	18.940088	72.835257
New Majestic Restaurant	18.938972	72.835517
City Palace Hotel	18.938970	72.835610
B.M.C. Headquarters	18.940565	72.834098
Marine Drive	18.941221	72.823261
McDonald's	18.938985	72.834504
Cafe Coffee Day	18.939029	72.834377
Pancham Puriwala	18.938214	72.835697
Central Railways Headquarters	18.940612	72.835073

.)	Cuisines	Features	Home_Delivery	Operational_hours	Restaurant_Location	Restaurant_Name	Restaurant_Type	V
0	Finger Food, Continental, European, Italian	Food Hygiene Rated Restaurants In Mumbai, Best...	False	12noon – 1am (Mon-Sun)	Kamala Mills Compound	Lord of the Drinks	Lounge,Casual Dining	
0	Pizza	Value For Money, Best of Mumbai	False	11am – 12:30AM (Mon-Sun)	Malad West	Joey's Pizza	Quick Bites	
0	Seafood	Super Seafood, Best of Mumbai	False	Closed (Mon),12noon – 3pm, 7pm – 12midnight...	Bandra West	Bastian	Casual Dining,Bar	
0	Finger Food, Continental	Where's The Party?, Best of Mumbai, Food Hygie...	False	12noon – 1am (Mon-Sun)	Lower Parel	Tamasha	Lounge,Bar	
0	North Indian, Street Food, Fast Food, Chinese	NaN	True	12noon – 4pm, 7pm – 11:45pm (Mon-Sun)	Vashi	Bhagat Tarachand	Casual Dining	

4.Data Visualization



4.Analysis of each section

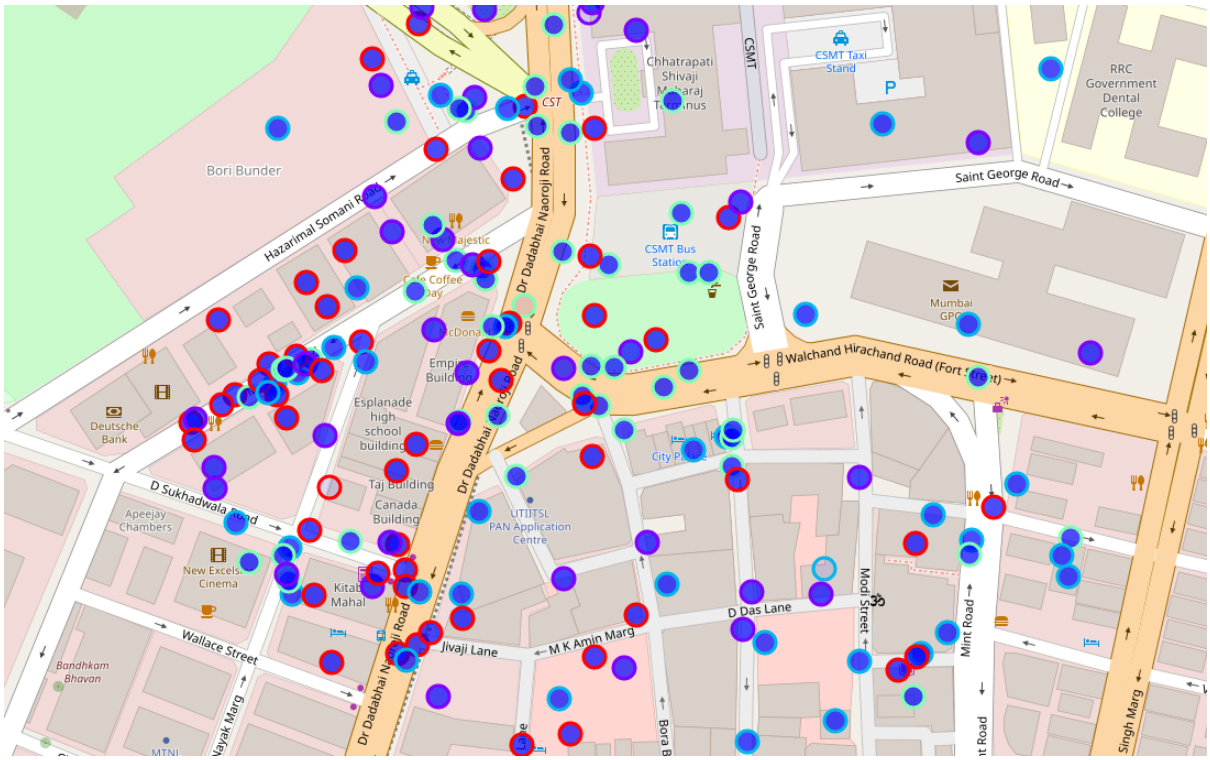
```
df_venue.describe()
```

3]:

	Venue Latitude	Venue Longitude
count	193.000000	193.000000
mean	18.938638	72.834749
std	0.001086	0.001349
min	18.936623	72.823261
25%	18.937729	72.834011
50%	18.938521	72.834619
75%	18.939539	72.835436
max	18.941221	72.837471

5. K means Clustering

Based on the venue, K-means clustering was conducted to group the neighbourhoods into 5 different clusters based on their similarity. The coloured dots below represent different clusters.



Results & Discussion

We analysed the locality data and neighbourhood data using Foursquare APIs. Using One-hot encoding to get most common venues around localities. Applied k-Means clustering algorithm with 5 clusters and we draw below results;

The closest Cluster we get from the tourist location is **Cluster 4**

By using this the tourist can choose the any of the restaurant from cluster 4