



BATTLE OF NEIGHBORHOODS

APPLIED DATA SCIENCE CAPSTONE
PROJECT

Rahul Kumar Singh

BATTLE OF NEIGHBORHOODS

1. INTRODUCTION

Mumbai (formerly Known as Bombay) apart from being the busiest cosmopolitan city in India, is also the financial hub of the country, due to which people from different parts of the world live in the city. Due to the diversity in the cultural backgrounds of people, there are a thousand of restaurants in the city, that cater to a variety of cuisines.

BUSINESS PROBLEM

As a part of my 'Applied Data Science Capstone' final project, I shall survey the type of restaurants in different neighborhoods of Mumbai to determine the most suitable place to open an Italian restaurant. The use of Data science techniques and machine learning algorithms such as Clustering, shall help us answer the above described business problem.

TARGET AUDIENCE

The target audience of the above business problem is a restaurateur who is looking for the right neighborhood to open an Italian restaurant.

2. DATA

To solve the above business problem, we shall need the following data: -

- A list of neighborhoods in Mumbai

- Latitude and Longitude data of neighborhoods- Latitude and longitude data could be extracted using the Geocoder package.

- Venue-related data- Venue related data could be extracted using the Foursquare API.

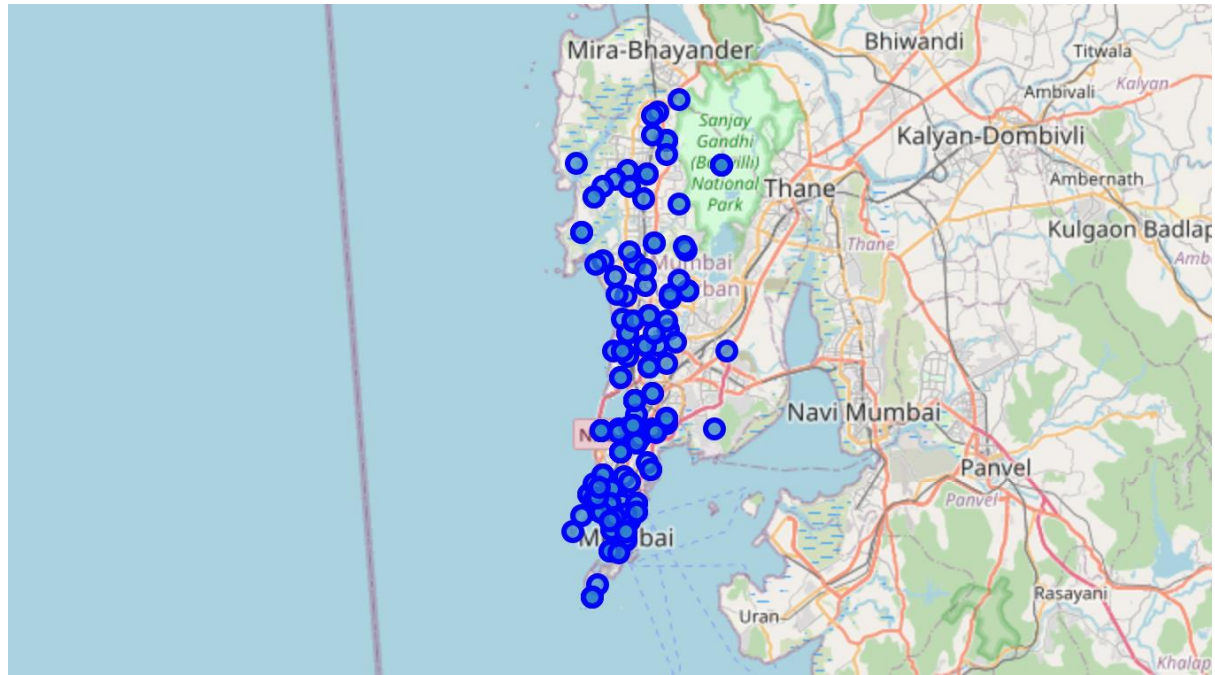
3. METHODOLOGY

-We begin by making a data-frame that contains a list of all the neighborhoods in Mumbai, along with their latitude and longitudes. The data frame looks something like this –

Mumbal_Neighborhood_data

Location	Pincode	State	District	Latitude	Longitude
A I staff colony	400029	Maharashtra	Mumbai	19.0787	72.86005
Aareymilk Colony	400065	Maharashtra	Mumbai	19.148509	72.88174
Agripada	400011	Maharashtra	Mumbai	18.98481	72.81686
Airport	400099	Maharashtra	Mumbai	19.0900516	72.8686722
Ambewadi	400004	Maharashtra	Mumbai	18.96301	72.82304
Andheri	400053	Maharashtra	Mumbai	19.1155251	72.8352615
Andheri East	400069	Maharashtra	Mumbai	19.113657	72.8697219
Andheri Railway station	400058	Maharashtra	Mumbai	19.113668	72.8697109
Antop Hill	400037	Maharashtra	Mumbai	19.02223	72.866554
Asvini	400005	Maharashtra	Mumbai	18.8952	72.80879
Azad Nagar	400053	Maharashtra	Mumbai	19.129	72.8268
B P t colony	400003	Maharashtra	Mumbai	19.02595	72.8664
B.N. bhavan	400051	Maharashtra	Mumbai	19.06295	72.85335
B.P.lane	400003	Maharashtra	Mumbai	18.9582222	72.8131886

-Then we find the latitude and longitude of Mumbai from the Nominatim package, so as to plot the map of Mumbai with all the neighborhoods using the Folium Package.



-Then we found out the nearby venues of each neighborhood. After doing so, we found the mean frequency of each Venue Category in every neighborhood.

[14]:

	Location	ATM	Accessories Store	Airport	Airport Lounge	Airport Terminal	Amphitheater	Antique Shop	Aquarium	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Automotive Shop
0	A I staff colony	0.0	0.0	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
1	Agripada	0.0	0.0	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.166667	0.0	0.0	0.0	0.0
2	Airport	0.0	0.0	0.25	0.25	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
3	Ambewadi	0.0	0.0	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
4	Andheri	0.0	0.0	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0

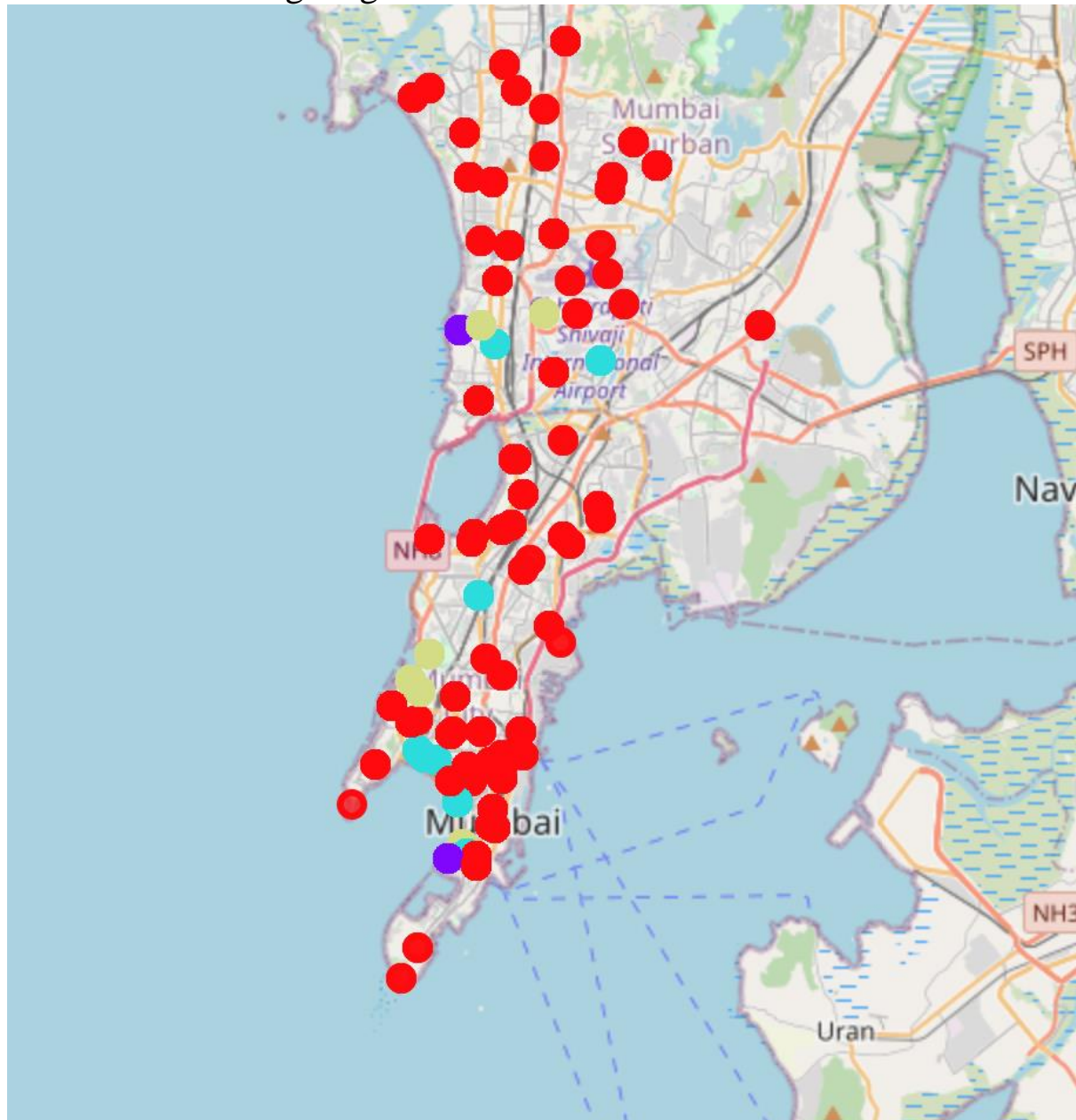
-Then we filtered out the venues that just had 'Italian Restaurant' as the Venue Category, as we just want to take Italian Restaurants into consideration.

	Location	Italian Restaurant
0	A I staff colony	0.000000
1	Agripada	0.083333
2	Airport	0.000000
3	Ambewadi	0.000000
4	Andheri	0.000000

-After this we used “KMeans” clustering to create clusters, and group the neighborhoods in Mumbai into different clusters, based on Italian Restaurants for further analysis of each area.

4. RESULTS

The final clustering diagram looks like this.



Cluster 0, Cluster 1, Cluster 2, Cluster 3.

The results of the clustering are as follows-

Cluster 0, has the least number of Italian Restaurants, so it is the most ideal place to open an Italian Restaurant.

```
j> mum_merged.loc[(mum_merged['Cluster_Labels'] == 0) & (mum_merged['Venue_Category'] == 'Italian Restaurant')]
```

	Location	Italian Restaurant	Cluster_Labels	Pincode	Latitude	Longitude	Venue_Name	Venue_Latitude	Venue_Longitude	Venue_Category
98	Mantralaya	0.013158	0	400032	18.929015	72.830121	Relish	18.93032	72.826361	Italian Restaurant

Clusters 1, 2 and 3 have a lot of Italian Restaurants, with Cluster 2 having the most.

5. DISCUSSION

The accuracy of determining the best places to open an Italian Restaurant can be improved by not just taking the number of Italian Restaurants into consideration, but also taking into consideration the demographic of a certain area, and determining the food preferences of the people in a certain neighborhood. Since, that data is not readily available, and can only be gathered through a widespread survey, for the time being only the number of restaurants have been used to find a solution to the posed Business problem. In the future, when such data is available we can update our findings by further incorporating the extra data.

6. CONCLUSION

Thus to conclude the above study, one can say that there is always, scope for more improvement, by finding more relevant data and increase its accuracy to narrow down the study, as discussed in the discussion section.