Report : IBM Capstone

Coursera IBM Professional Data Science Project

JANUARY 3

Data Science

Authored by: Myron D'Souza

Introduction

The city of Mumbai consists of a large number of restaurants, but still there is always scope for new ones. Finding a suitable location for it to flourish is the most important factor for a restaurant. It has to be set up in a location where one can attract a good crowd as well as it must be located in an area where there is little or no competition.

We will be able to provide a solution for anyone looking to open an Chinese cuisine restaurant in the city of Mumbai.

The Goal of this problem is to find a location that suits the below criteria.

1) A location that has many restaurants in the vicinity (European, Intercontinental, Indian)

2) A location that has no or few Chinese cuisine restaurants, as this will ensure that there very little competition with other competitiors.

Data

The data that will be used in this projects is a csv file having data related to all neighborhoods in the city of Mumbai.

	Location	Pincode	State	District
0	A I staff colony	400029	Maharashtra	Mumbai
1	Aareymilk Colony	400065	Maharashtra	Mumbai
2	Agripada	400011	Maharashtra	Mumbai
3	Airport	400099	Maharashtra	Mumbai
4	Ambewadi	400004	Maharashtra	Mumbai
5	Andheri	400053	Maharashtra	Mumbai
6	Andheri East	400069	Maharashtra	Mumbai
7	Andheri Railway station	400058	Maharashtra	Mumbai
8	Antop Hill	400037	Maharashtra	Mumbai
9	Asvini	400005	Maharashtra	Mumbai
10	Azad Nagar	400053	Maharashtra	Mumbai
11	B P t colony	400003	Maharashtra	Mumbai
12	B.N. bhavan	400051	Maharashtra	Mumbai
13	B.P.lane	400003	Maharashtra	Mumbai
14	Bandra West	400050	Maharashtra	Mumbai
15	Bandra(east)	400051	Maharashtra	Mumbai
16	Bangur Nagar	400090	Maharashtra	Mumbai
17	Bazargate	400001	Maharashtra	Mumbai
18	Best Staff colony	400012	Maharashtra	Mumbai
19	Bharat Nagar	400007	Maharashtra	Mumbai
20	Bhawani Shankar	400028	Maharashtra	Mumbai
21	Bhawani Shankar rd	400028	Maharashtra	Mumbai
22	Borivali	400091	Maharashtra	Mumbai
23	Borivali East	400066	Maharashtra	Mumbai
24	Borvali West	400092	Maharashtra	Mumbai
25	C G s colony	400013	Maharashtra	Mumbai

Finding a suitable location for a restaurant to flourish is an important

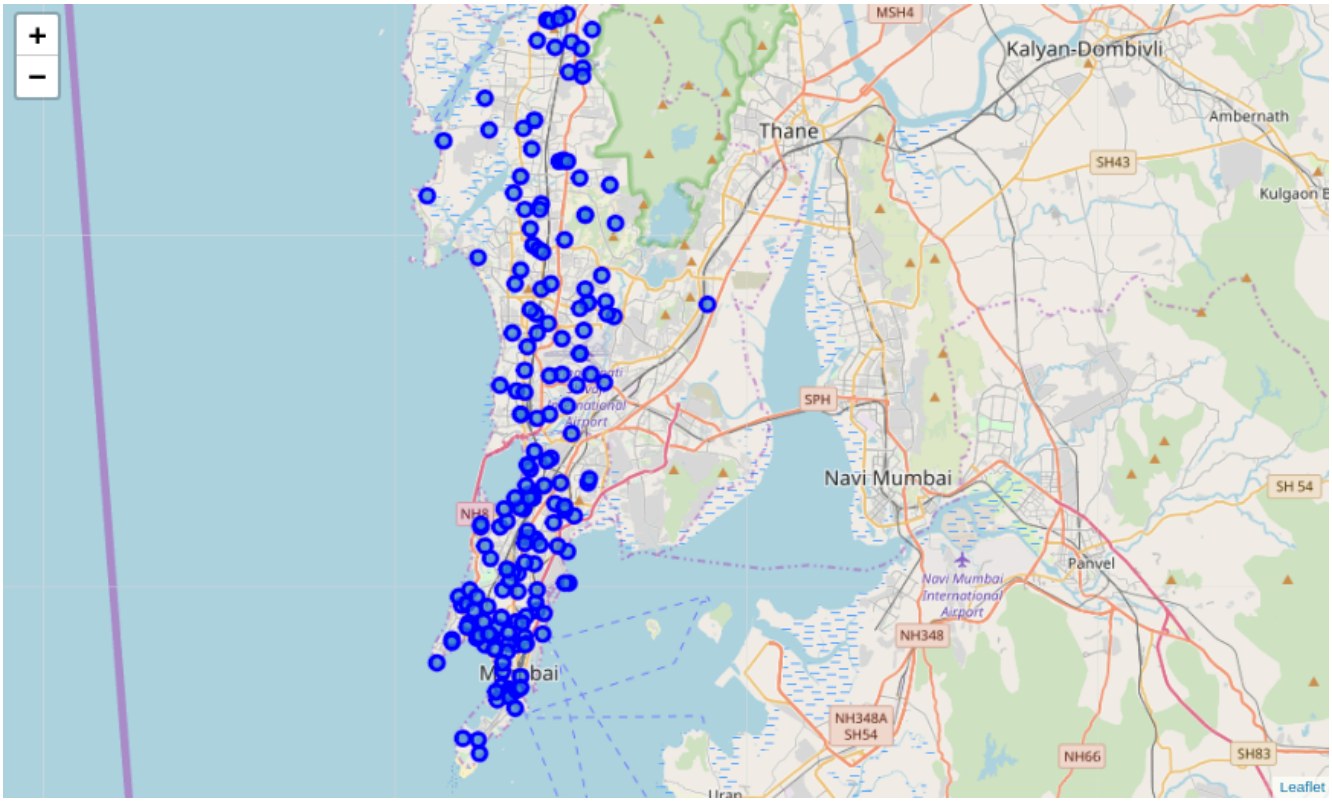
We explore the neighborhoods using Foursquare API to find the avenues within 500 meters of each neighborhood.

The Foursquare API that will be used to explore the neighborhoods is https://api.foursquare.com/v2/venues/explore.

This API returns json response which will be transformed into a Data Frame, taking only the required details into consideration.

Methodology

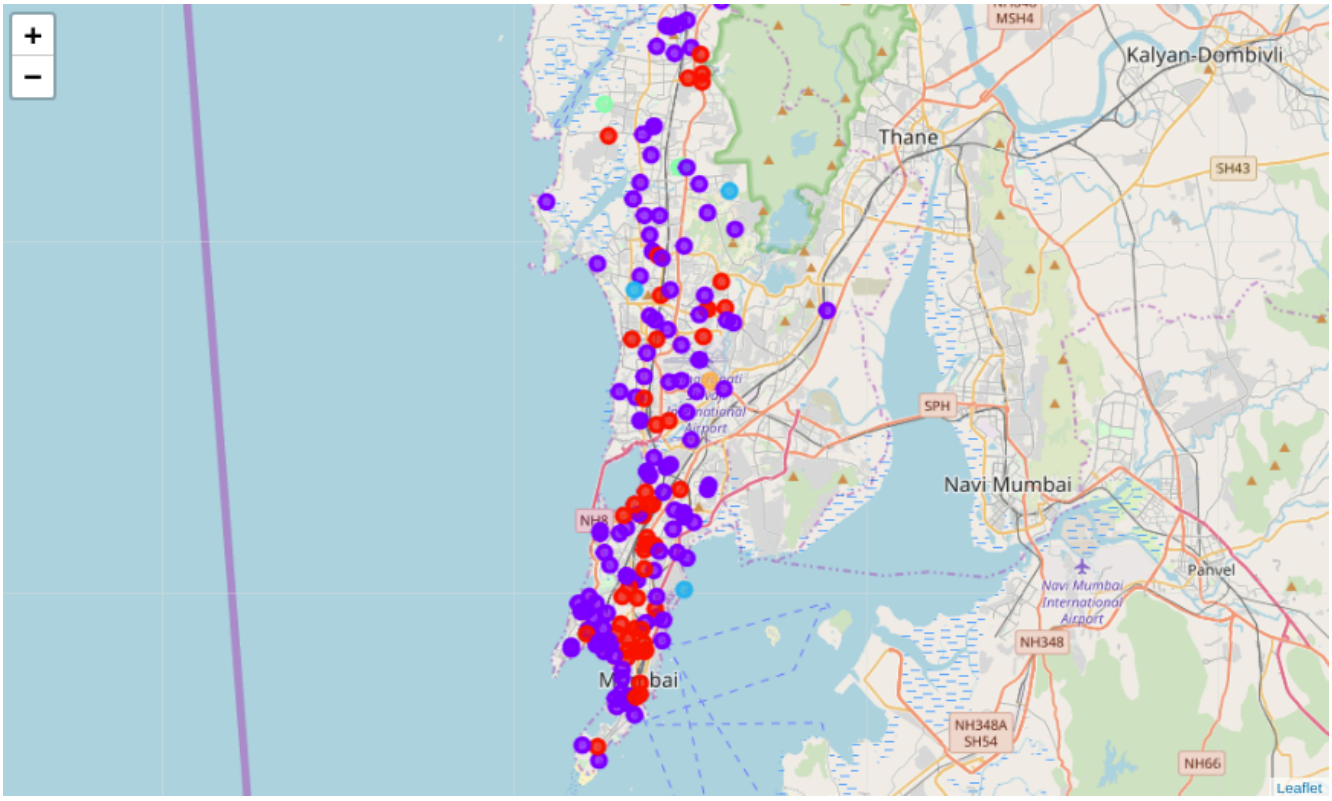
Using the Foursquare API venue information is obtained in nearby vicinity of postal locations in Mumbai.



A radius is set to cover large neighborhoods in a particular area in Mumbai.

This data is then merge with the Location dataset (Postal Codes) and a clustering algorithm is applied to the data.

K-Means Clustering : The data points are clustered into 4 clusters using K-Means algorithm.



The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K.

The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided.

Results

By exploring the requirements we found only two neighborhoods that match the requirements (Many restaurants in the vicinity & only a few Chinese restaurants)

	1st 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	Cluster Labels	Latitude	Longitude
57	Gym / Fitness Center	Field	Fast Food Restaurant	Falafel Restaurant	Event Space	Electronics Store	Eastern European Restaurant	2	19.122222	72.833484
115	Snack Place	Falafel Restaurant	Event Space	Electronics Store	Eastern European Restaurant	Duty-free Shop	Donut Shop	2	19.169408	72.880347
154	Snack Place	Falafel Restaurant	Event Space	Electronics Store	Eastern European Restaurant	Duty-free Shop	Donut Shop	2	18.980820	72.857592

```
Mumbai_merged.loc[Mumbai_merged['Cluster Labels'] == 3, Mumbai_merged.columns[[1] + list(range(5, Mumbai_merged.shap
```

	1st 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	Cluster Labels	Latitude	Longitude
30	Pizza Place	Falafel Restaurant	Event Space	Electronics Store	Eastern European Restaurant	Duty-free Shop	Donut Shop	3	19.210045	72.818349
95	Pizza Place	Falafel Restaurant	Event Space	Electronics Store	Eastern European Restaurant	Duty-free Shop	Donut Shop	3	19.180237	72.855415
96	Pizza Place	Falafel Restaurant	Event Space	Electronics Store	Eastern European Restaurant	Duty-free Shop	Donut Shop	3	19.180874	72.857456

Discussion

According to results we observe that most common venues (Top 10) come out to be restaurants and Pizza places and Snack places, which means any new chef/business man can start a restaurant provided that they need to compete with existing restaurants (Other Types) but only if he provides top class facilities to get to top.

Conclusion

Based on the Clusters formed it would be a good idea to open a restaurant in Clusters 2 & 3 since the other clusters already have Chinese Restaurants in their vicinities.
Also, cluster 2 & 3 have many restaurants in the vicinity (Pizza restaurants, European, Intercontinental, Indian) so one will be able to attract a good crowd.
Also, there is no other competitors in this neighborhood who have set up Chinese restaurants nearby.
These results have limitations - The venue data obtained is of top 10 venues in each neighborhood, where we might neglect Chinese restaurants with less frequency.