Business Location Analysis

Manas Ranjan Pradhan

1 Introduction

1.1 Background

In all kinds of business, finalizing location is one import factor for the success of business. This project is for someone who is looking for starting a small scale business/startup in Mumbai city of India. Mumbai is one of biggest city in India. The population in Mumbai is too huge and the cost of shop and office space are also high. On the other hand there is good scope of business in the city. Unless and until somebody is not planning for a completely new type of business and the type of business he has in mind is little common to the local region, choosing proper location and the nearby surrounding is necessary.

1.2 Problem

If someone has some business or shop idea in mind, the problem is to make dicision on which location to prefer and how to find market places which are better for him and his business, how to judge the surrounding etc. Also it is important to know whether the ideas in his mind are also relevant to the trend in locality.

1.3 Interest

Anybody who is an entrepreneur or wants to start a new small scale setup of shop, business in Mumbai city would be very interested in this analysis.

2 Data acquisition and cleaning

2.1 Data sources

I made use of Foursquare location data for solving this problem. I creates an account in Foursquare and used API of version '20180604' to downloaded location data for Mumbai city. The location points have properties like coordinates(combination of latitude-longitude),

category, name etc. 'Category' represents what type of location point is it and 'name' is the name of the location. While downloading, no category was specified. Because the objective is to analyze the trends of different business types in the area. Data was huge and downloading cannot be repeated, hence it was stored in a csv file during execution.

2.2 Data cleaning

As mentioned above, location data does not belong to any specific category for the purpose of general analysis. Still there are various categories of location points present in the downloaded set which seem to be outliers and better to be discarded before analysis. Examples of these categories are 'building', 'space' etc.

All the columns are not relevant from the location data, hence I filtered out other columns apart from latitude, longitude, name and category.

Only unique records were saved and the redundant ones were discarded.

Location records which seemed to be relevant to business spot, shop, office, restaurant are useful for analysis. Non alphabets and non-numbers were also removed in order to have meaningful words for ease of analysis.

2.3 Feature selection

The column 'name' and 'category' both have contents to indicate categories. Hence I made union of 'name' and 'category' column and considered both as category. e.g. If a location point has information as follows.

No	latitude	longitude	category	name
1	19.103378	72.848059	Indian Coffee	Coffee shop

Above was combined to form below 2 records.

No	latitude	longitude	category
1	19.103378	72.848059	Indian Coffee
2	19.103378	72.848059	Coffee shop

Analysis was done on the above type of data set. Also the above set was split through individual words to do extended analysis on the above words in the category column. So the new data for extended analysis becomes.

No	latitude	longitude	category
1	19.103378	72.848059	Indian
2	19.103378	72.848059	Coffee
3	19.103378	72.848059	Coffee
4	19.103378	72.848059	shop

Up course the above table contains some redundant rows now, those were deleted in the next step.

3 Solution

3.1 Business Hubs

From the above location data Business hubs in Mumbai city were discovered. Clusters were forms out of the coordinates in location data. Using Foursquare location data, points of interests for entire Mumbai are extracted.

We did not specify any category while downloading data for points of interest. After downloading this data I created clusters based on the coordinates of the points.

3.2 Clustering

There are various clustering techniques available. Density based algorithm comes to mind. This divided the points based on density. Instead, I used K-means algorithm and specified a certain number of clusters. It helped in dividing the points in more evenly.

I created 1000 numbers of clusters with the help of on K-means algorithm which are depicted in the following figure.

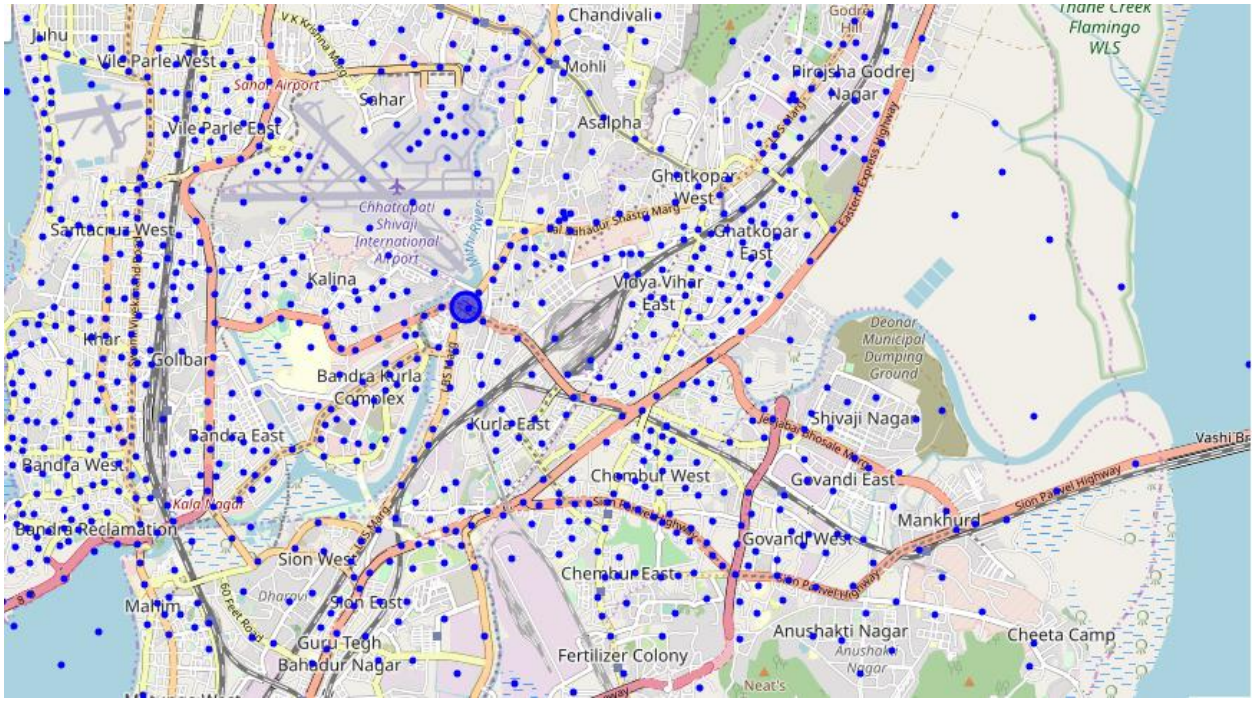

Figure 1 Business clusters

In the above figure the blue spots over the map denote the business clusters. The bigger circle is the center of Mumbai city. In the most simplest way a nearby or comfortable location indicated by the above depicted clusters can be short listed for business setup. On top of it, I shall verified whether the intended business type is present in the shortlisted cluster.

Dessert shop was one of the business ideas. So tried to find which clusters already have dessert shops.

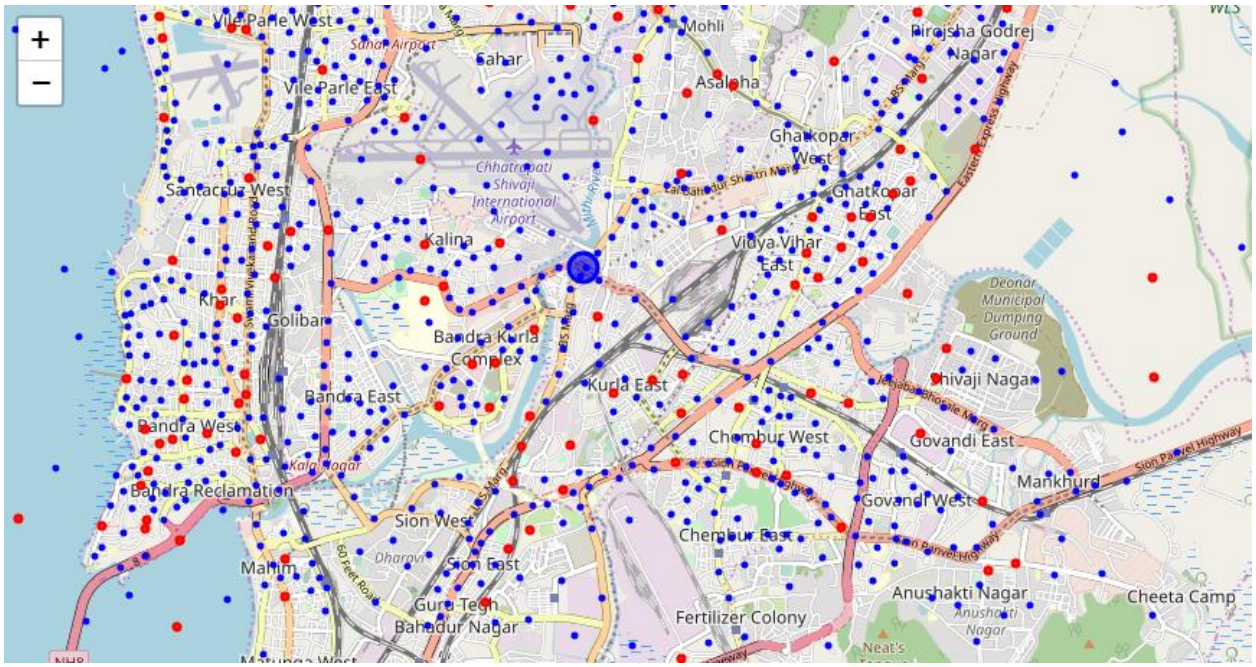

Figure 2 Clusters with or without dessert shops

In the above map the red circles denote the cluster where desserts shops are already there. So the locations which not nearer to the red spots can be shortlisted for the dessert shop business.

4 Conclusions

It can now be concluded with the list of locations which are suitable for opening a dessert shop.

e.g.

latitude	longitude
19.051154	72.89399
19.097033	72.850748
19.125291	72.921495
19.072219	72.866539
19.042699	72.862248
18.961169	72.931108
19.060156	72.853508

We examined with dessert shop here Similarly, other options can also be analyzed.

5 Future directions

The above decision can be strengthened more with market basket technique. I tried applying apriori algorithm to observe what are the market trends and which are the other business areas which remain together with the intended business. So we can find a place with similar fashion.