

The Battle Of Neighborhoods

Pandit Gangadhar

Capstone Projects Oct 2021

The Battle of Neighborhoods: Open a Hospital in Mumbai city India

Clustering and segmenting the neighborhoods to find the best location for opening a hospital in Mumbai city.

Table of contents

1. Introduction

1.1 Background

1.2 Business problem

1.3 Audience

2. Data

2.1 Data Requirement and Collection

2.1.1 Population Growth of Mumbai

2.1.2 Pollution rate dataset

2.1.3 Neighborhoods dataset

3. Data Cleaning and Feature Extraction

4. Methodology

4.1 Analytic Approach

4.1.1 Population data

4.1.2 Pollution data

4.1.3 Neighborhoods data

4.2 Exploratory Data Analysis of Population

4.2.1 Relationship Between Population and Growth

4.2.2 See the growth rate of population

4.3 Exploratory Data Analysis of Pollution

4.3.1 Poor Location

4.3.2 Satisfactory Location

4.3.3 Good Location

4.4 Clustering and Segmenting of Neighborhoods in Mumbai city

4.4.1 Scrap the dataset

4.4.2 K-means Clustering

5. Results and discussion

5.1 Conclusion

1. Introduction

1.1 Background

Mumbai is the capital of the Indian state of Maharashtra. According to the United Nations, as of 2018, Mumbai is the second most populous city in the country after Delhi and the seventh most populous city in the world with a population of roughly 20 million. Now it is above. Mumbai has population density of 20,482 person per square kilometer which is very high. So, there are more chance to grow many types of Business.

1.2 Business Problem

Now, imagine that we had a hospital in Delhi city called 'Public Hospital' that had been doing business success in Delhi city, But Now we wanted to expand this Business in Mumbai.

To success this project, our team has many problem for discussion. They wanted to know every things about this project because this is a challenging project. They asked many questions:

- What are the most popular venues where hospital is situated?
- What are probability to grow this business in that place?
- How many hospitals exist there?
- What is the health status in that neighborhoods? and so on.

Thus, we had a goal to open a hospital and we have to choose best location that met the above criteria.

1.3 Audience

This project is for business man and stakeholder who are passionate about opening a hospital in metropolitan city like Mumbai, Kolkata, etc. Also a business man who is searching this types of location to start a new business which is essential for people and never closed in future.

2. Data

2.1 Data Requirement and Collection

We needed the data about population of Mumbai, pollution data, and neighborhoods data. We collected the data from different website. This is common data. We could collect from Wikipedia and normally search in google. We would also use foursquare location data to segment the neighborhoods. Therefore we used following data.

2.1.1 Population Growth of Mumbai

Here, a questions arise in mind that why is population growth data is important and what is the connection between a hospital situated and population growth?

It's very important to know about population growth to start a new business in market. Population growth is big factor to impact on any types of business and opening a hospital is one of them. Suppose a city has low population and there are a number of hospital are opened there, what will happen, there is less chance to run a new business there also it will have to face most challenging environment. Now any one can think deeply that if population increases, more traffic occurs in that area, resultant pollution rate is also increased that effect on health of people. So it was also important to know about pollution rate of the neighborhoods of Mumbai city.

2.1.2 Pollution Rate Dataset

We collected the dataset from different sources and combine them together in a single dataset. We extracted the pollution data of each neighborhoods having the pollutants PM2.5, PM10, NO2, CO, NH3, and OZONE.

Again the same question why this data is important?

Pollution rate of the city help to know about health status of people in neighborhoods of the city. Lots of people are suffering from many types of diseases like heart diseases, cancer, breathing problem due to pollution and many more diseases and everyone wants to go to near hospital for treatment. So, we have to find such as location where no any hospital or less hospital in neighborhood. So our next step is to collect the neighborhoods dataset.

2.1.3 Neighborhoods data

We extracted the data from Wikipedia and scrap this data by pandas library.

2.1.4 Foursquare Location Data

Last data we will use foursquare location data. It will help to extract the venues within the range. In this case, we will use foursquare credential information like client id, client secret, and version. Also we take a limit of venues in a particular range.

3. Data Cleaning and Feature Extraction

1. Our first data of Population data is in csv file. It has 86 rows and 4 columns. Each row explains year, population, growth rate and growth of the population in that year. We will use all columns of the dataset and see the variation of each year.

2. Second data of pollution data is also in csv file. It has 114 rows and 5 columns. The first column represents location in which neighborhood exists.

Second column is pollutants that are PM 2.5, PM10, NO2, NH3, SO2, OZONE and CO. Each pollutant has average values in microgram per meter cube. Last two columns represent minimum and maximum value of pollutants.

3. Third dataset is in url form. And we will extract from Wikipedia with the help of pandas library. It has 93 rows and 4 columns.

First column represents Neighborhood of Mumbai city. Second column is Location detail where neighborhood exists and last two columns are latitude and longitude. These are coordinates where neighborhood is located.

First we will use geocoder python library and find the latitude and longitude of Mumbai city location.

4. Last dataset is used from foursquare location data where we will segment and cluster the neighborhood of Mumbai city. We will use our foursquare credential to create a url and extract the location data from it. We will use k means clustering to find the best k value so that we may find the best clusters in our map.

4. Methodology

4.1 Analytic Approach

4.1.1 Population Data

First we will explore the population data by using Plotly Library. Plotly is a python library that makes the graph more attractive. We can zoom in and zoom out in graph that is created by Plotly. We can see each value of the year on each points on the graph. We will see the growth rate of population that how increase the data over each year.

4.1.2 Pollution Data

Second step we will import pollution data. We will also use Plotly Library to plot the bar graph and decide to choose best location that is free from pollution or good or satisfactory in condition. In this dataset we will plot each location where neighborhood exists and see the AQI (stands for air quality index) decide which location is good for opening hospital.

4.1.3 Neighborhoods Data

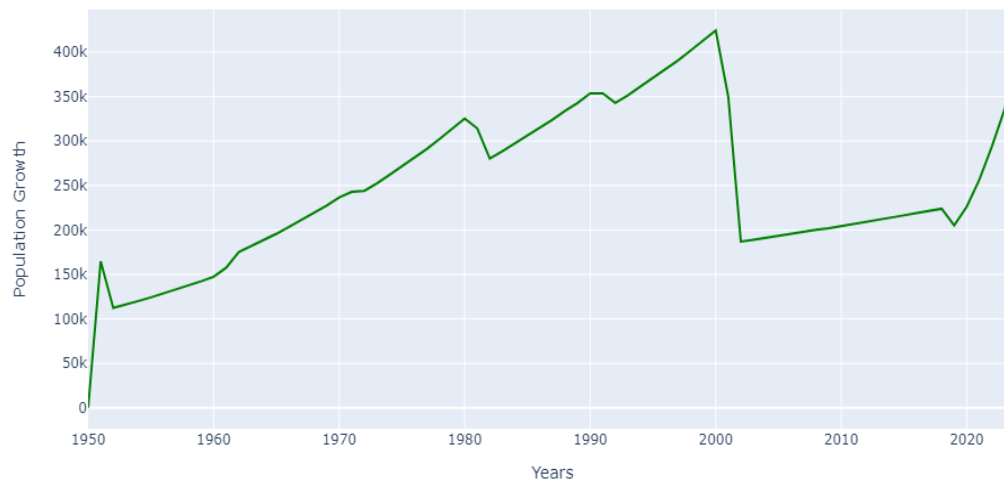
Last we will focus on neighborhood clustering. For analytical approach we will use clustering technique of k means machine learning algorithm and segment of neighborhoods. K means clustering technique is the best way for clustering the data points of similar characteristics. We plotted a graph of cluster to choose the best value of k and decided how to apply it to segment the neighborhood of Mumbai city. We decided the best cluster by the help of distance between the points within the cluster that is inertia. Inertia means that how far the points within the cluster.

4.2 Exploratory Data Analysis of Population

4.2.1 Relationship Between Population and Growth

We plotted a graph between population and growth. We took population in y axis and year in x axis. We used plotly library to plot this graph.

Years vs Population growth

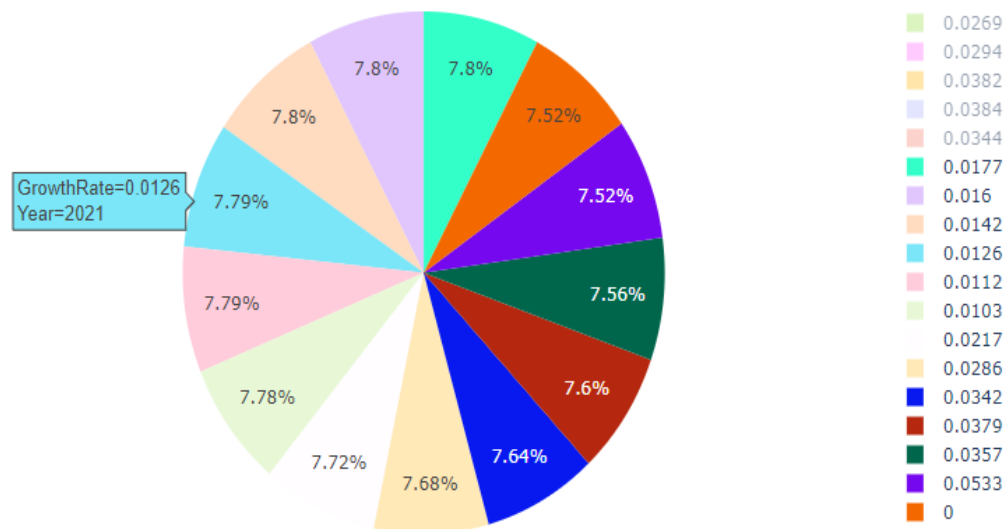


From above graph we can see that each point on line on the graph shows the year and average growth of population. On year 2000 we can see that the population reduced suddenly. This happened when landslide was occurred in many region of Mumbai on 12 July 2000 and other some reason, due to which no one went to Mumbai to live. But now all are fine. After the year 2020, we see that the line plot is going on upward direction. This shows that growth of population will increase in upcoming year but this graph does not show the growth rate of population. To see the growth rate we will use pie chart and where we will see how much growth on each year.

4.2.2 See the growth rate of population

We can see growth rate of population by pie chart. For that we have applied plotly library to plot the pie chart. Pie chart of the population shows the growth rate on each year. In 2019, population growth is 0.0103. It became 0.0112 in year 2020. Now it increased 0.0126 in 2021. It seems that it will increase in upcoming year.

Growth Rate By Year



4.3 Exploratory Data Analysis of Pollution

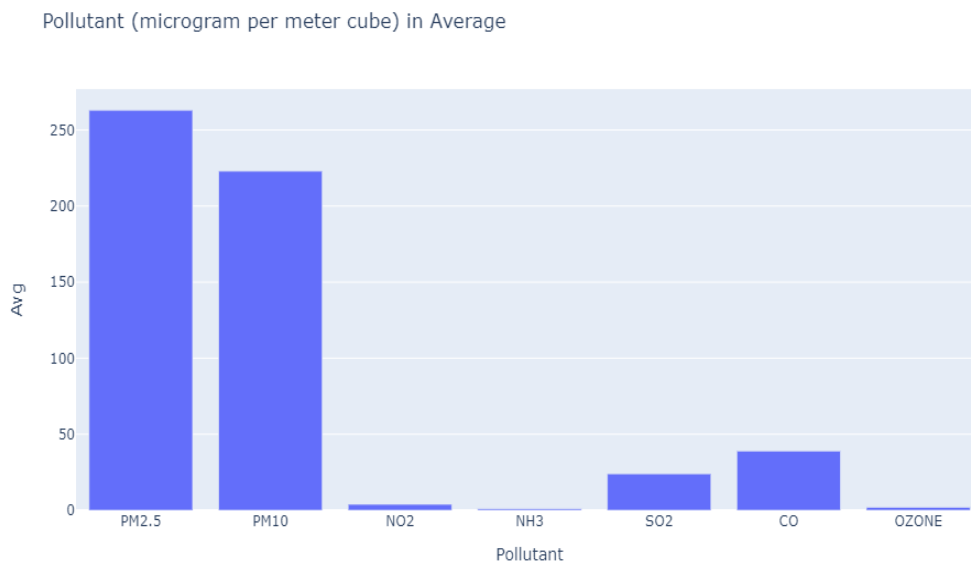
Now, we loaded the dataset of Mumbai pollution growth and explained in detail of each location where there is neighborhood of Mumbai city and try to choose good location where our goal is completed. We will open a new hospital in a safe place where pollution is low or location has good air quality. Also we will measure the AQI (stands for Air Quality Index). **what is AQI?** AQI is a index that is used by government agencies to communicate the public how polluted the air currently is or how polluted it is forecast to become. If AQI rises, public health risk will increase. If **Average (Avg) of any Pollutant was higher, we assumed the AQI is higher of that location this means this area is high polluted.** Opening a hospital in such area is not good for patients and anyone else. Lots of people will ignore. Our purpose is to choose best location as well as best environment in that location.

Here some measurement in AQI by which we will decide which place is better.

Remark	AQI
Good	0-50
Satisfactory	51-100
Moderate	101-200
Poor	201-300
Very Poor	301-400
Severe	401-500

4.3.1 Poor Location

We extracted '**Bandra kurla complex Mumbai IITM**', and saw its pollutant and average value. We used plotly library to visualize the pollutants. This is great library to plot the bar data. It provides the facility to zoom in and zoom out and we can see the values of pollutants on each bar easily.



Above plot shows that PM2.5 is greater than all pollutants. It means AQI is high and the condition of the Bandra Kurla complex IITM location is **Poor**. So we ignore this location.

4.3.2 Satisfactory Location

4.3.2.1 Location 'Mazgaon, Mumbai, IITM' in south Mumbai, has satisfactory air quality index that means this area has less polluted. In this area PM2.5(Particulate matter) is average 59 and PM10 is 43. AQI lies between 51 and 100.

Average pollutants in Mazgaon, Mumbai, IITM

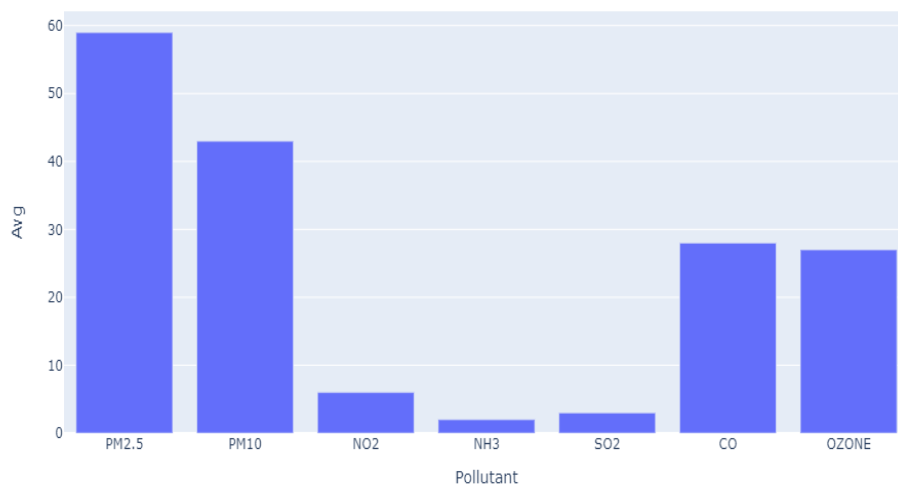


Fig. 4.3.2.1

Average pollutant in Khindipada-Bhandup West, Mumbai, IITM

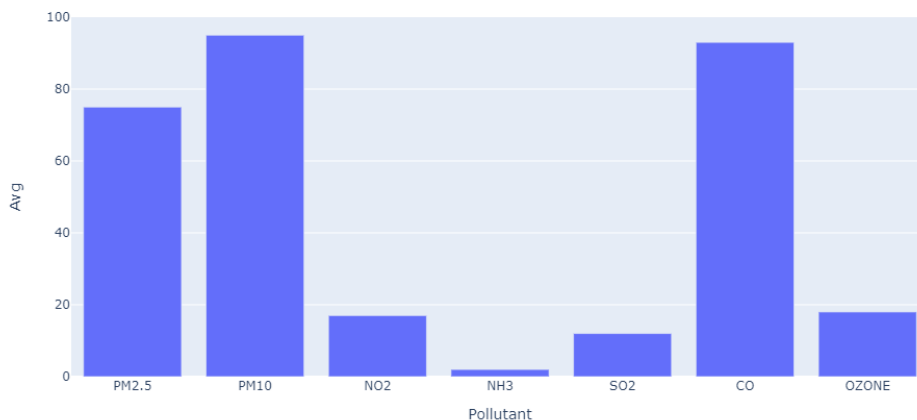
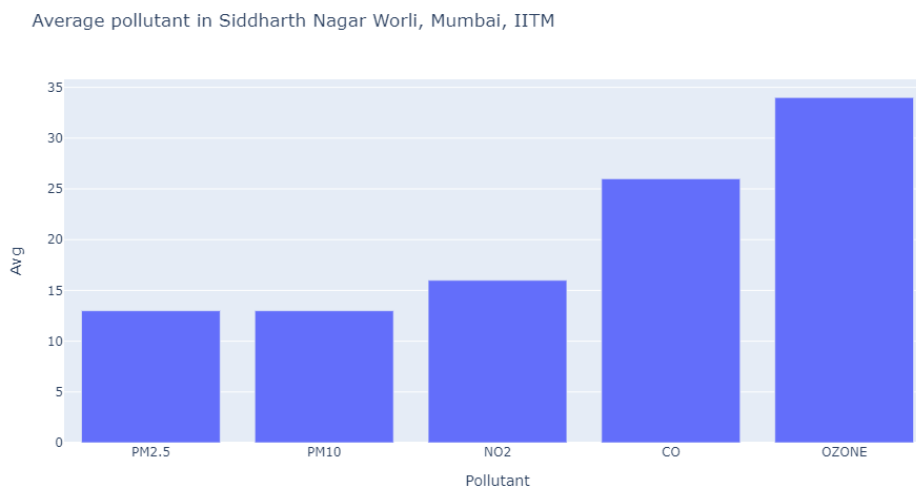


Fig. 4.3.2.2

4.3.2.2 Khindipada-Bhandup West, Mumbai, IITM also has satisfactory air quality shown in figure 7.2.2. AOI report in this location is 95 because particulates matter(PM) 10 is higher in this area.

4.3.3 Good Location

Siddharth Nagar Worli, Mumbai, IITM in south Mumbai has good air quality. It has less or no pollution. In this area all pollutants except ozone is less and Average of ozone is 34 which lies between AQI 0 to 50. It shows that the location has good environment.



PM 10 particles are less severe health effects mostly in the upper respiratory tract. These particles consist of smoke, dirt and dust from factories, farming, roads as well as mold. But If I talk about PM 2.5 then this is very small particles more than 100 times thinner than a human hair. The size of these particles is up to 2.5 micrometers. These consist of toxic organic compounds and heavy metals. They are made from automobile exhaust, burning garbage and some other smelting.

From Pollution data analysis we got that PM 10 is less harmful than PM 2.5 pollutants. So, our focus those area which is less PM 2.5 as well as PM 10. From above graph we get some location like west Mumbai, some location in south Mumbai are good and satisfactory in air quality. So, we kept these location in mind and go further to next analysis of neighborhoods to find best neighborhood.

4.4 Clustering and Segmenting of Neighborhoods in Mumbai city

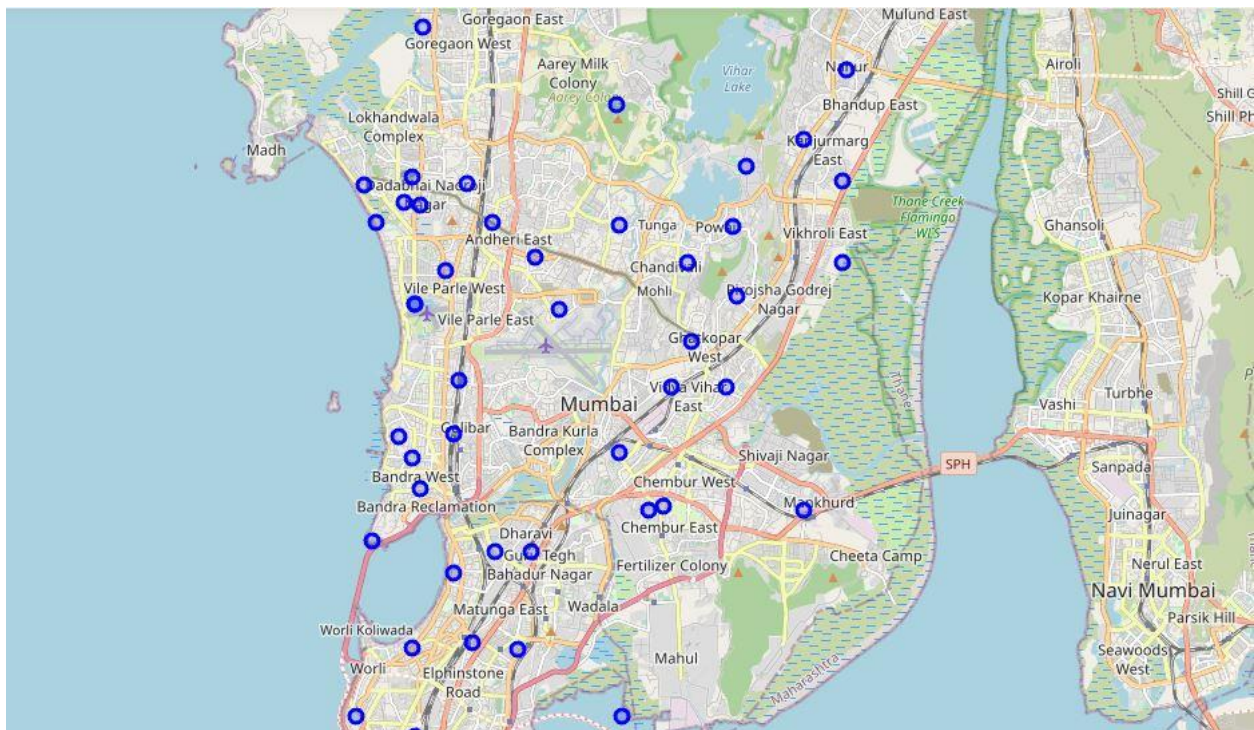
4.4.1 Scrap the dataset

First we scraped the data from Wikipedia and transformed it in Data Frame by Pandas python library. It gave us the data in table form having rows and columns.

After loading the data we checked the shape of the dataset. First column represent neighborhoods, and second column is location where neighborhoods exist and last two columns are latitude and longitude of that neighborhood.

Now we imported request library to get the data from url and also imported folium to create a map to make more attractive.

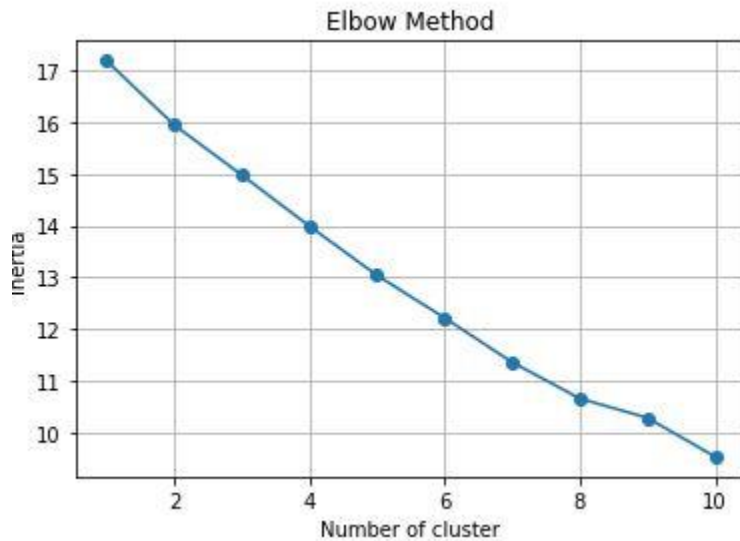
First we plotted a map of neighborhoods in Mumbai city. This look like.



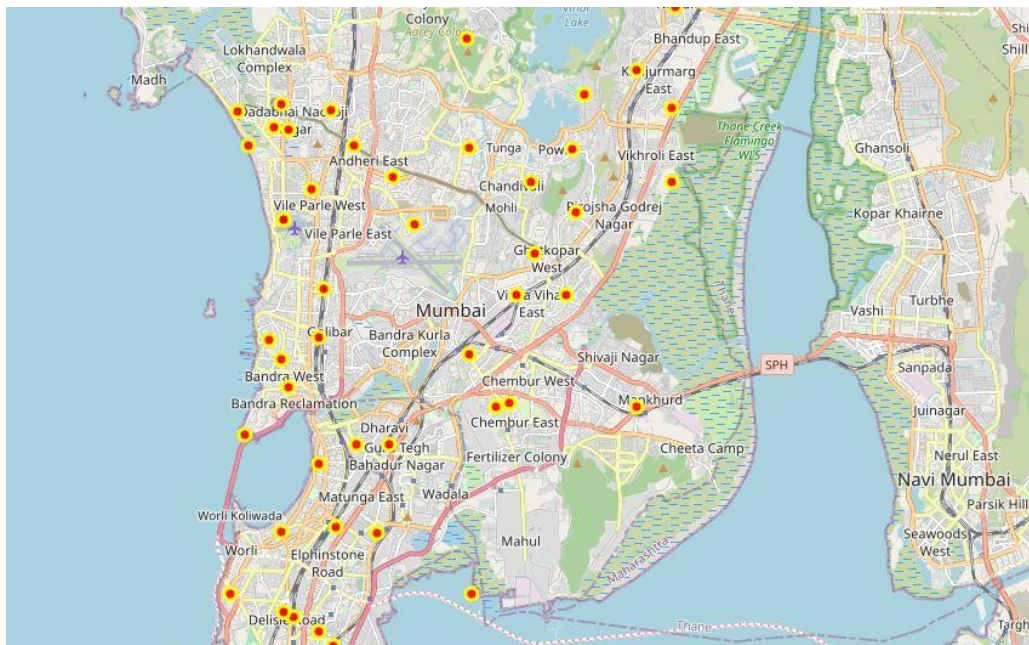
Now, we used foursquare location data to extract the data of venues. By the help of foursquare credential information we make a url and gave a limit of venues 100 in the range of 400 meters. We saw the results of venues. We got lots of venues in 400 range. We used one hot encoding to analyze the venues in each neighborhood and extracted 10 most common venues.

4.4.2 K-means Clustering

Now. We used analytics approach of k means for clustering these neighborhoods. First we plotted a graph and decide by inertia i.e. how far the points within the cluster and tried to take minimum value of k having less inertia. Here we took k = 4.



We fitted the cluster label in our original dataset and again created a map by folium to see how our clusters look like in map that is shown below.



At last, we found that there are four clusters and we analyzed the venues as well as neighborhoods then we decided which location is best to open a new business. We discussed in result.

5. Results and Discussion

Now we completed all analysis. As a result, we discussed about most common venue where our goal will become full fill. There are four clusters Having lots of venues.

Cluster 1 :- It has 43 rows and 12 columns. We can see that on 1st Most common Venue, there are lots of restaurant and coffee shop are available there. Only two pharmacy venue are available in 3rd most common venue.

Cluster 2:- It has 16 rows and 12 columns. There are lots of café, and coffee shop in 1st most common venue.

Cluster 3:- It has 33 neighborhoods with 12 columns. There are lots venues like coffee shop, Indian restaurant, Hotel, Electronics store etc. in 1st most common venue and other have same things.

Cluster 4:- Last cluster has only one venue of multiplex in 1st most common venue.

Our main goal is to find best location to open a hospital in neighborhood. First we consider the location that is free from pollution or less polluted and these are Siddharth Nagar Worli, Mumbai, IITM in South Mumbai, Khindipada-Bhandup West, Mumbai, IITM in West Mumbai, Mazgaon, Mumbai, IITM in South Mumbai and Kandivali East, Mumbai, MPCB in east mumbai. But we will consider three area that are **west Mumbai, south Mumbai towards worli nagar, and east Mumbai**. These locations have good or satisfactory or moderate air quality index that is fine for open a new business.

Now we will focus on neighborhood. After seeing the result of all clusters, we can say that there is good location where we can open a new hospital. In cluster 1 as we can see that in 3rd most common venue, there are only two pharmacy. First is in amboli neighborhood that is in Andheri, western suburbs and second pharmacy is in charkop neighborhood that is in Kandivali East, Mumbai. There is also less polluted location. So, we can consider to open a hospital near these pharmacy.

5.1 Conclusion

Purpose of this project to choose best location to open a hospital. For that we use foursquare location data and plot map to see neighborhood that where it exists. after that we extracted that 100 venues within 500 meters range then we decide

where we will open a new business. Last we saw that three locations are best for starting new business. First west Mumbai, second south Mumbai towards Worli nagar, and east Mumbai, these locations could be considered.

According to data that is present today gives us an idea to open a new business for any businessman or entrepreneurs. This data provides you only external information like location that is free from pollution, more populous, and best neighborhood. And we decide to open a new business. But It does not provide internal information of that area in which we're going to open a business and that is 1. what's on in market right now, 2. which types of people are living there, and 3. we will also take a approval from government if project is big. So that's all reason we will have to face difficulty and keep lots of patience while opening a new business. Starting a new business can be challenging and also frustrating. But we can gain meaningful insights quickly.

..... **Thanks**

