

Finding The Hotspot For Mobile Phone Expo – 2020

Mudeet Jain

January 14, 2020

Applied Data Science Capstone – IBM Data Science Professional Certificate on Coursera

1. Introduction

1.1 Background

A Mobile Phone Expo in India for the year 2020 has been planned and the location for the expo is still undecided. This Expo will act as a marketing campaign for major mobile phone manufacturers and local mobile stores around the location. Customers can purchase a new mobile phone and also they can get to know the local stores, service points nearby from the expo.

1.2 Problem

The problem is to find the central hub-point, which should be located in middle of many mobile shops in a lesser distance. So, that expected amount of sponsors and customers around that area will be guaranteed to participate in the expo. Therefore, the cost of installing the expo can be reduced by more number of sponsors and profit of the expo will be increased.

2.Data Acquisition and Cleaning

2.1 Data Sources

I have used the Foursquare API for collecting the mobile shop location data's which includes latitude and longitude values for 5 major cities in India. The Cities are Bangalore, Chennai, Delhi, Hyderabad and Mumbai. The reason behind these 5 Cities are that these cities are well known for their daily market trade and populations.

2.2 Data Cleaning

The Data retrieved from the FourSquare API is in JSON format containing many information, which are not needed. The information only needed for this project are shop name, latitude, longitude and city name. By using python script, only the needed data columns are selected from the JSON results.

2.3 Feature Selection

Once the data is cleaned and ready for use. We can choose which rows and columns are needed. For our purpose, only name, city, latitude and longitude values are needed and it is available in the df_cities data frame. Below Fig-1 is the sample data screenshot.

	Name	Address	Lat	Lng	City
0	Ritche Street	Ritchie St	13.069477	80.271374	Chennai
1	Croma	2nd Ave,	13.084631	80.216891	Chennai
2	Reliance Digital Express	Anna Nagar	13.084756	80.213343	Chennai
3	Vodafone Store	Anna nagar	13.085272	80.203079	Chennai
4	Aircel Store	Aircel Towers, No.301,Poonamalle High Road,Kil...	13.075956	80.236309	Chennai

Fig. 1 – Screenshot of the Dataset.

3. Exploratory Data Analysis

3.1 Data Visualization

The Dataset which we collected needs to be analyzed. Since, it consists of latitude and longitude values. We can plot it in a graph for analysis. Therefore, the five cities are plotted individually with the data points on it by using folium maps.

Folium is a map visualization library available in python for visualization of geographical locations based on their latitude and longitude values. We have the latitude and longitude values of each city and it mobile shops. So, we can plot it for visualization. Below Fig- 2 to 6 are the screenshot of all the 5 cities mobile shop locations in the map.

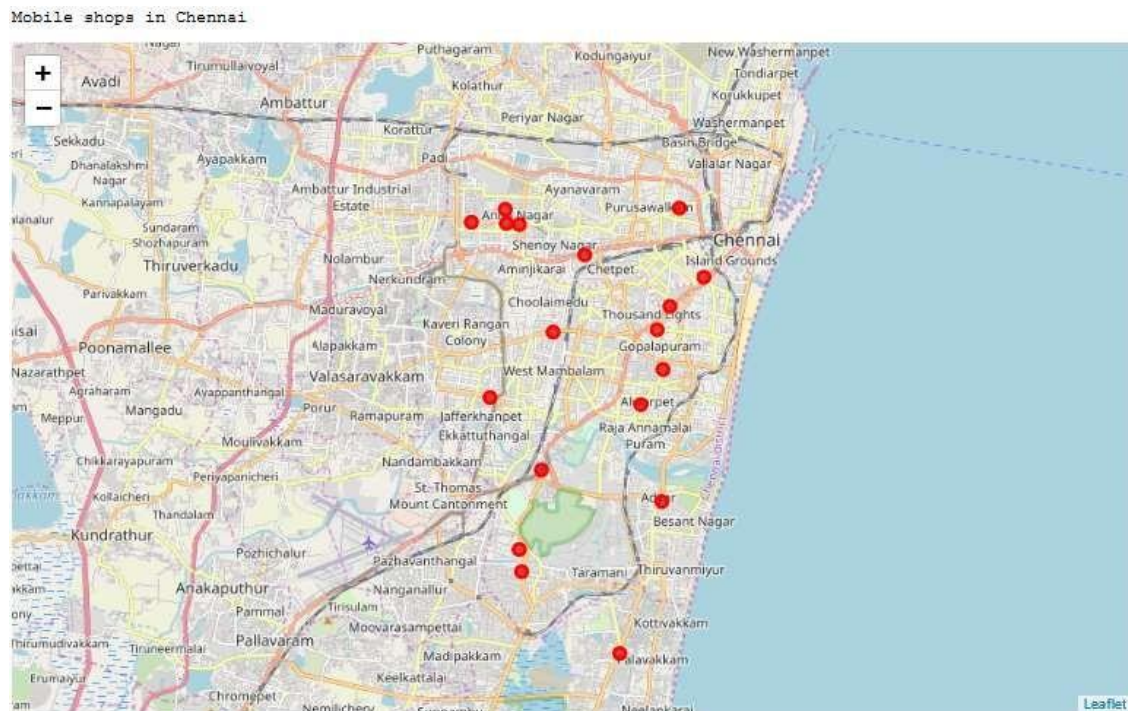


Fig 2 – Mobile shops in Chennai

Mobile shops in Hyderabad

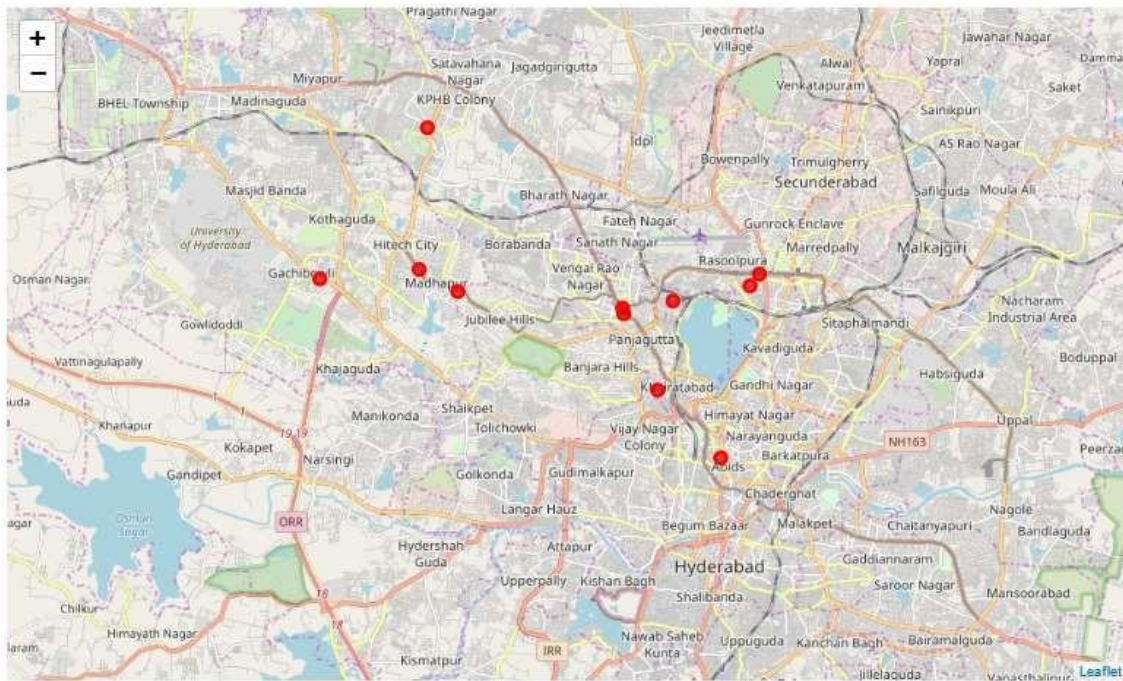


Fig 3 – Mobile shops in Hyderabad

Mobile shops in Bengaluru

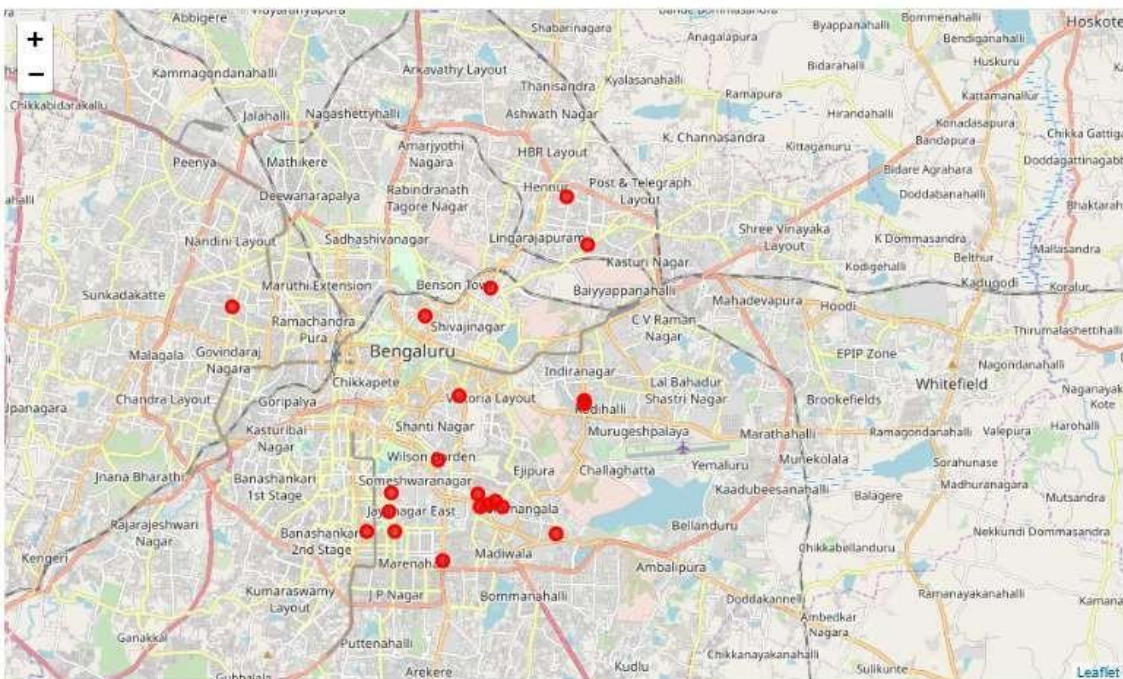


Fig 4 – Mobile shops in Bengaluru

Mobile shops in Mumbai



Fig 5 – Mobile shops in Mumbai

Mobile shops in Delhi

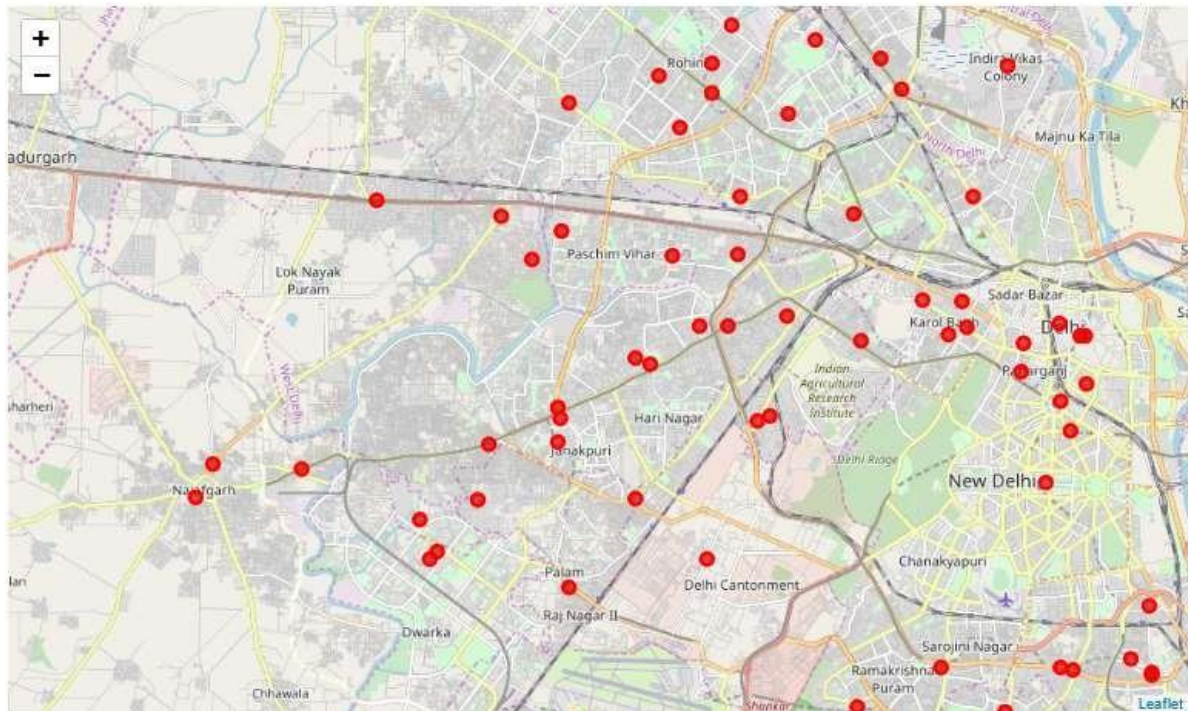


Fig 6– Mobile shops in Delhi

3.2 Calculation of Centre Point and Measure of Distance for Each Cities

Each city has n- number of mobile shops. The calculation of the distance from a central point of those mobile shops in each city would be helpful in identifying the best city with higher number of mobile shops and also helps to find the average distance to be travelled for the Expo-2020. Below Fig- 7 shows the screenshot of average distance to central point to all the mobile shops.

```
Chennai
Distance from the centre point : 0.05105306360595714
Hyderabad
Distance from the centre point : 0.043075877532875756
Bangalore
Distance from the centre point : 0.035131045589305084
Mumbai
Distance from the centre point : 0.08167181371420372
Delhi
Distance from the centre point : 0.09574720569597839
```

Fig. 7 – Distance from centre point in each cities.

These centre points are calculate by the mean average of the latitude and longitude values of each cities. The Central Hub point for Delhi seems bit far location for some mobile shops. So, We can create a cluster of 3 in the above map. We will find a central point for each cluster to determine the expo location for delhi.

4. Cluster Analysis

4.1 K-means Clustering Algorithm

K-Means algorithm is an unsupervised learning algorithm. It clusters, partitions or segments the available data points by means of similar characteristics based on labels. The available data points are divided into non-overlapping subsets called clusters without any internal cluster structure.

Algorithm :

1. Randomly placing k Centroids, one for each cluster
2. Calculate the distance of each point from each centroid
3. Assign each data point to its closest centroid, creating a cluster
4. Recalculate the position of the k centroids
5. Repeat the steps 2-4, until the centroids no longer move.

After applying the k-means algorithm to the data points for delhi city with a k cluster value. We got the cluster labels for each row in the delhi dataset. Below Fig-8 is the screenshot of delhi data which is merged with cluster labels.

	Cluster Labels	Lat	Lng	Name
0	0	28.683666	77.196670	My Idea Store
1	1	28.535816	77.196995	My Idea Store
2	0	28.713319	77.109163	My Idea Store
3	1	28.649738	77.304540	My Idea Store
4	1	28.641635	77.295433	Trendy cover for HTC

Fig 8. Delhi data with their cluster labels.

By using this cluster label, we can find the center point for each cluster to start expo in that place. The center point can be calculated by the mean average of latitude and longitude values for all the points available in each cluster.

5. Results

The Result is the center hub point for each segmented cluster in Delhi. There are 3 clusters of mobile shops located in this city and the center mid-point for each cluster can be used for starting the Expo-2020. Fig-9 is the screenshot of the Delhi city with its clusters and center points.

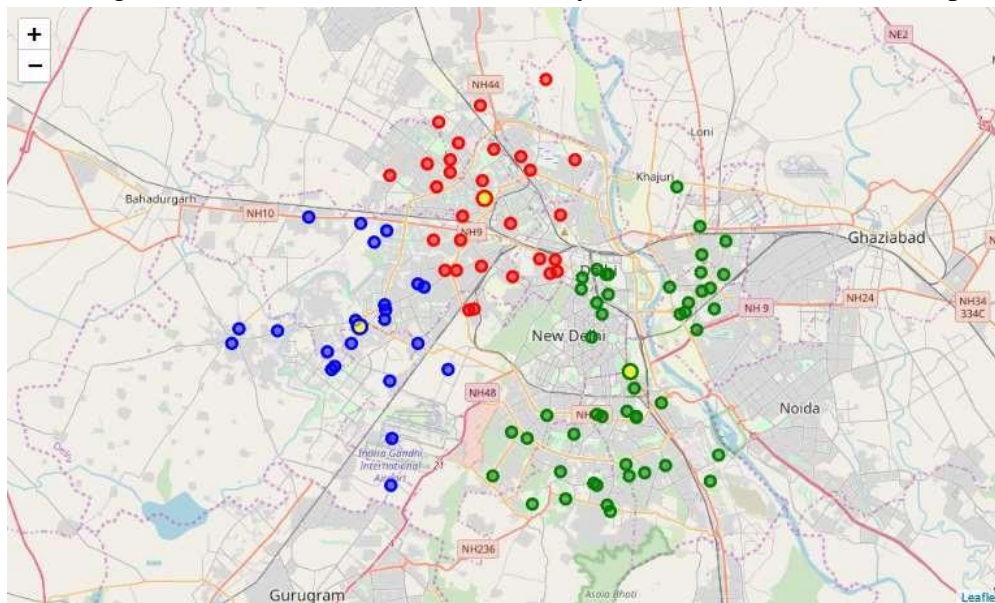


Fig 9. Cluster of mobile shops in Delhi city.

The three clusters are plotted in the graph having a differentiation of markings in colors as red, green, and blue. Each cluster has its own center points highlighted with yellow color.

6. Conclusion

The project work for finding the hotspot location to start a Mobile Expo-2020 in India has been found as Delhi. Therefore, the distance from the mobile shops to the expo will be lesser and will result in participation of higher number of mobile stores, sponsors, and customers. Which will increase the popularity and sales profit of all the mobile stores around Delhi.