**IBM Capstone Project - The Battle of Neighbourhoods**

Analysis of Venues and Prices of Apartments in Bengaluru

Contents

[**1.** **Introduction** 3](#_Toc41417292)

[**1.1** **Background** 3](#_Toc41417293)

[**1.2** **Problem** 3](#_Toc41417294)

[**2.** **Data Description** 3](#_Toc41417295)

[**3.** **Methodology** 3](#_Toc41417296)

[**3.1** **Methodology Adopted** 3](#_Toc41417297)

[3.1.1 Problem Definition 4](#_Toc41417298)

[3.1.2 Data Collection 4](#_Toc41417299)

[3.1.3 Data Preparation 5](#_Toc41417300)

[3.1.4 Exploratory Data Analysis 5](#_Toc41417301)

[3.1.5 Clustering 11](#_Toc41417302)

[**4.** **Results** 12](#_Toc41417303)

[**5.** **Discussion** 13](#_Toc41417304)

[**6.** **Conclusion** 13](#_Toc41417305)

[**References** 14](#_Toc41417306)

# **Introduction**

## **Background**

Bengaluru, also known as Bangalore, is the capital city of the southern Indian state Karnataka. With a population of 12 million, it is the third-most populous city of India. This cosmopolitan ‘Garden City of India’ and the ‘Silicon Valley of India’ attracts scores of national as well as foreign professionals for the opportunities it offers. The population density of Bengaluru has seen a staggering growth of 47% over the past ten years. [1]

Having lived in this city for over two years, I know what a blessing it is to live close to work. Bengaluru has earned the top spot as the ‘Most Traffic Congested City’ amongst 416 cities cross 57 countries. [2]

As a professional moving to Bengaluru or someone wishing to move closer to work, one would find himself going through websites and contacting brokers to find a suitable accommodation. There are several things to keep in my mind while looking for a place to stay; whether you are buying or renting a house. The deciding factors are usually the price, neighbourhood, facilities around the neighbourhood and distance from work.

## **Problem**

We need a much bigger picture in order to buy or rent a house. While websites can give you an overview of the real estate prices in different neighbourhoods, you will have to research on the facilities around the neighbourhood by yourself. The objective of this project is to analyse neighbourhoods with respect to buying a house and access to different venues in the neighbourhoods.

# **Data Description**

Data for this project was collected from various sources which are listed below:

* Wikipedia: The list of neighbourhoods in Bengaluru was web scraped from Wikipedia. A few more neighbourhoods were also added as the list gathered missed some localities in Bengaluru.
* Google: Initially the library ‘Nominatim’ was used to fetch the geographical co-ordinates of neighbourhoods, but it was observed that some results were incorrect. Geographical co-ordinates of a few neighbourhoods were checked from Google and were found to be more accurate. Subsequently, a small process was developed in the RPA tool UiPath which picked neighbourhood names from a CSV file, fetched their respective co-ordinates from Google search and populated them back in the file.
* Foursquare: Details of venues in the neighbourhoods were gathered using Foursquare APIs.
* Makaan.com: The dataset of cost of buying an apartment in various neighbourhoods was web scraped from the website Makaan.com.
* 99acres.com: This website to fill in the values missing in the dataset prepared from Makaan.com.

# **Methodology**

## **Methodology Adopted**

The methodology used for the project is divided into 5 stages:

1. Problem Definition
2. Data Collection
3. Data Preparation
4. Exploratory Data Analysis
5. Clustering and Evaluation

### Problem Definition

At this stage the objectives and requirements of the project were formulated to define a problem.

### Data Collection

Data was collected from various sources as listed in section 2.

**Table 1: Neighbourhoods from Wikipedia**

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO. | ATTRIBUTE | TYPE | DESCRIPTION |
| 1. | Neighbourhood | String | The name of the neighbourhood. |
| 2. | Area | String | The area the neighbourhood belongs to – north, south, east, west, central, etc. |

**Table 2: Bengaluru neighbourhoods (Master table)**

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO. | ATTRIBUTE | TYPE | DESCRIPTION |
| 1. | Neighbourhood | String | The name of the neighbourhood. |
| 2. | Area | String | The area the neighbourhood belongs to – north, south, east, west, central, etc. |
| 3. | Latitude |  | Latitude of the neighbourhood. |
| 4. | Longitude |  | Longitude of the neighbourhood. |

**Table 3: Venue data from Foursquare**

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO. | ATTRIBUTE | TYPE | DESCRIPTION |
| 1. | Neighbourhood | String | The name of the neighbourhood. |
| 2. | Area | String | The area the neighbourhood belongs to – north, south, east, west, central, etc. |
| 3. | Neighbourhood Latitude | String | The latitude of the neighbourhood. |
| 4. | Neighbourhood Longitude | String | The longitude of the neighbourhood. |
| 5. | Venue | String | The name of the venue. |
| 6. | Venue Latitude | String | The latitude of the venue. |
| 7. | Venue Longitude | String | The longitude of the venue. |

**Table 4: Real estate data from Makaan.com**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.NO.** | **ATTRIBUTE** | **TYPE** | **DESCRIPTION** |
| 1. | Neighbourhood | String | The name of the neighbourhood. |
| 2. | Price Range per sqft | String | The price range of buying an apartment per sqft. |
| 3. | Average Price per sqft | String | The average price of buying an apartment per sqft. |

### Data Preparation

1. Data Cleaning:
   * Name correction: There were several names in the real estate dataset that were spelt differently and hence, were altered. For instance, ‘Indira Nagar’ was changed to ‘Indiranagar’, ‘Jeevan Bima Nagar’ to ‘Jeevan Bheema Nagar’, etc. This was done to keep the neighbourhood names consistent in all data sets.
   * Tuple duplication: There were certain neighbourhoods in the real estate data set that had multiple entries in the table. Such duplicate rows were removed.
   * Handling missing values: A few neighbourhoods had missing values in the columns ‘Price Range per sqft’ and ‘Average Price per sqft’. The values were filled in from 99acres.com.
2. Data Integration: The real estate dataset consisted of 5072 records. This contained many duplicate entries and very small localities which were a part of neighbourhoods that were already covered. The master neighbourhoods table and the real estate table were merged to create a dataset that contained only the real estate prices of the neighbourhoods in the master neighbourhoods table. The final dataset contained 68 records.
3. Data Transformation:
   * One-hot encoding was used to transform the venue dataset from Foursquare into a form suitable for calculating the distribution of types of venues in each neighbourhood

### Exploratory Data Analysis

1. Top 30 venue categories in Bengaluru:

The Fig.1 presents a bar chart of the top 30 types of venues in Bengaluru according the data gathered from Foursquare. The number of venue categories were found to be 198.

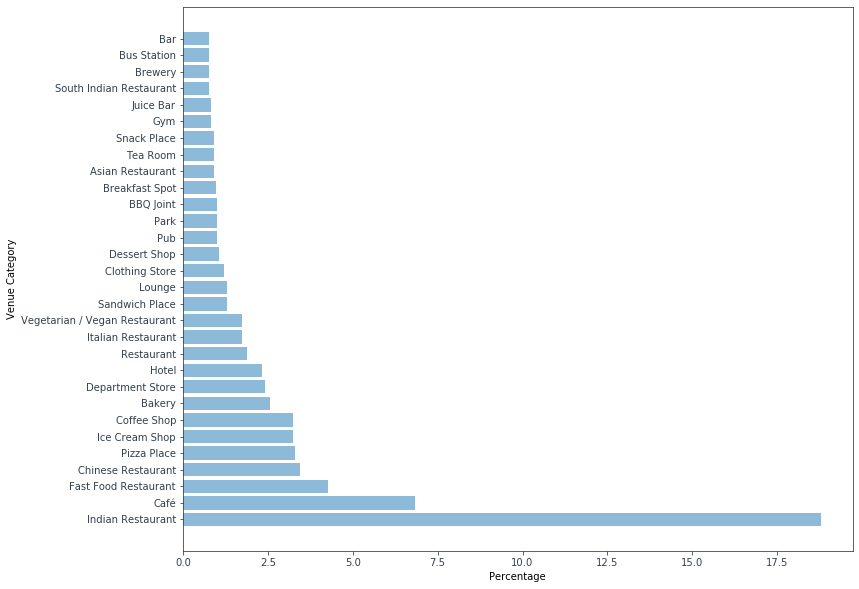


Figure 1: Top 30 types of venues in Bengaluru

1. Top 5 venue categories of neighbourhoods:

The top 5 venue categories for every neighbourhood in Bengaluru is shown in Fig. 2.

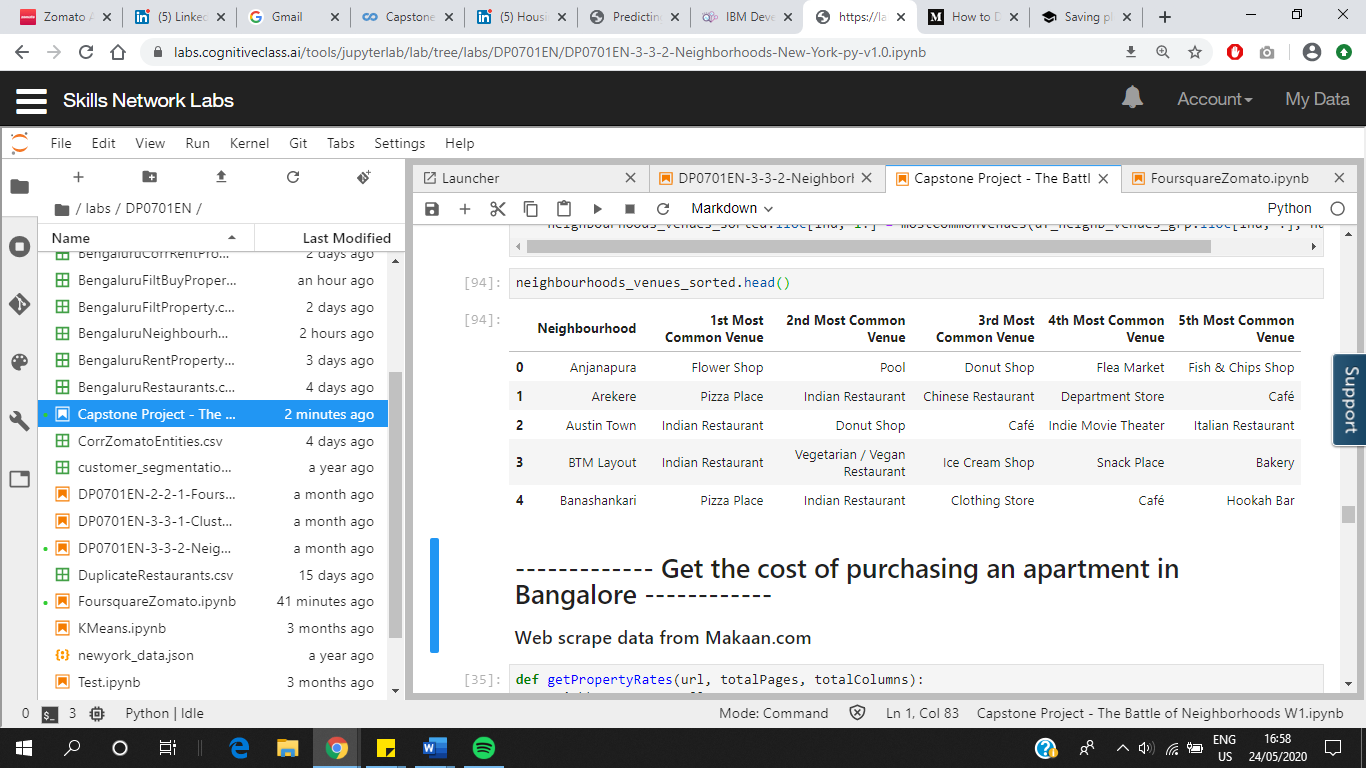


Figure 2: Top 5 types of venues in each neighbourhood

1. Overview of apartment prices:

The histogram in Fig.3 gives an overview of the apartment costs in 68 neighbourhoods of Bengaluru. For a majority of the neighbourhoods, minimum cost is less than Rs. 5000 per sqft, average is Rs. 5000 – Rs. 10,000 per sqft and maximum cost is Rs. 5000 – Rs. 10,000 per sqft.

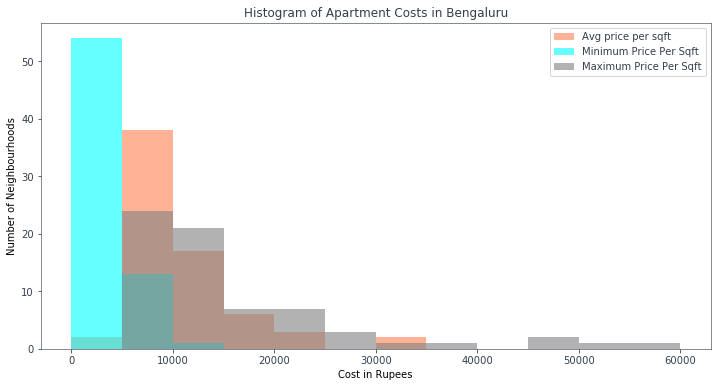


Figure 3: Histogram of apartment costs in Bengaluru

1. Average apartment rates:

The average per sqft cost of buying an apartment in a neighbourhood is shown in Fig. 4 and Fig. 5. Rajajinagar and Kammanahalli are considerably expensive than other neighbourhoods.

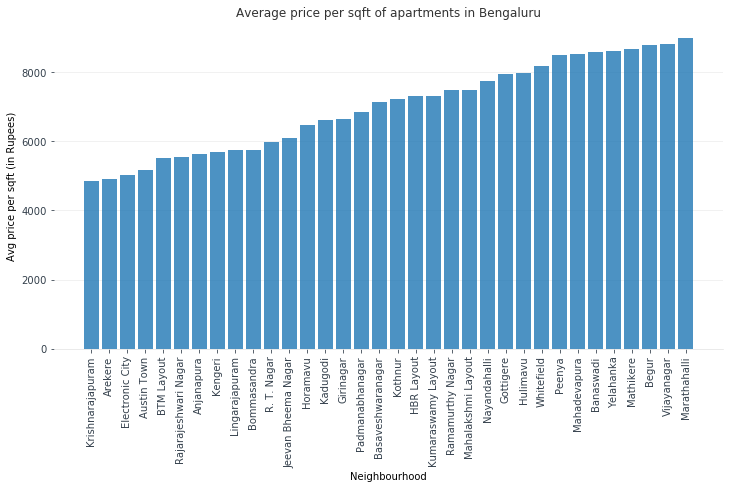


Figure 4: Average per sqft cost of buying an apartment

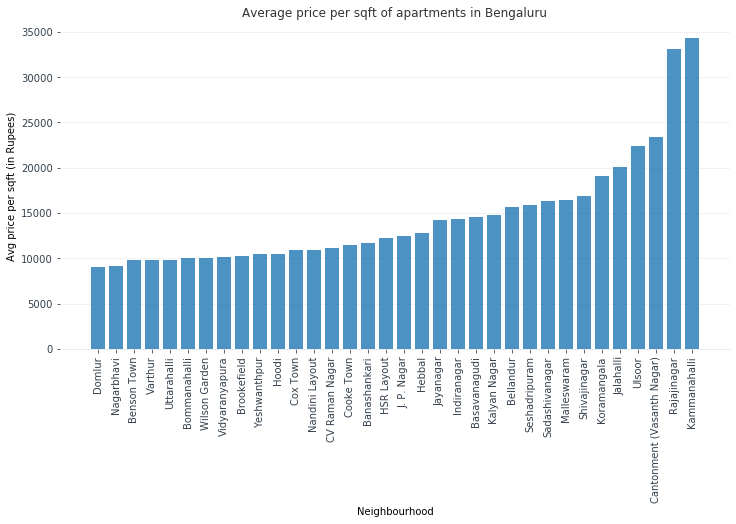


Figure 5: Average per sqft cost of buying an apartment

1. Minimum apartment rates:

The minimum per sqft cost of buying an apartment is shown in Fig. 6 and Fig.7. Sadashivanagar has a much higher lowest price per sqft.

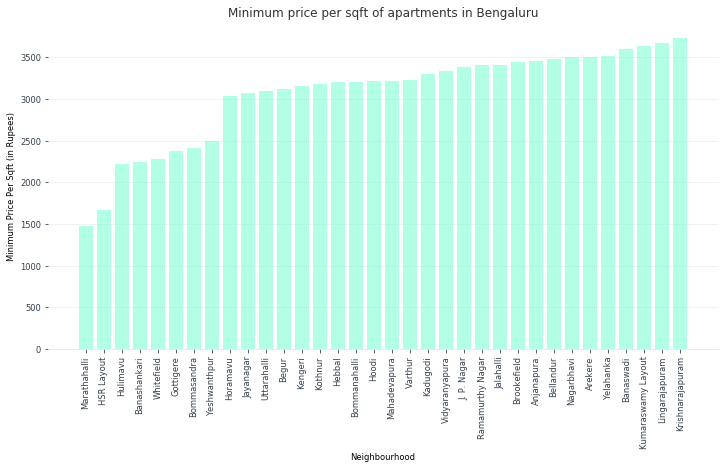


Figure 6: Bar chart of minimum cost of buying an apartment

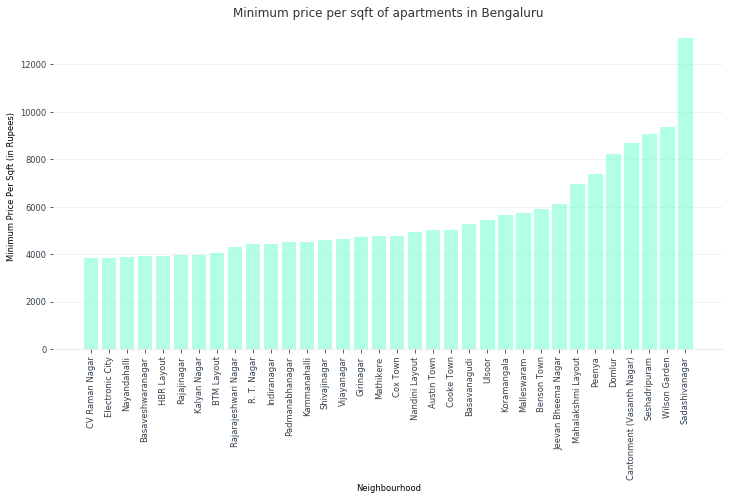


Figure 7: Bar chart of minimum cost of buying an apartment

1. Maximum apartment rates:   
   The maximum per sqft cost of buying an apartment is shown in Fig. 8 and Fig.9. Kalyan Nagar, Hulimavu, Kammanhalli, Rajajinagar and Jalahalli have the highest maximum cost per sqft.

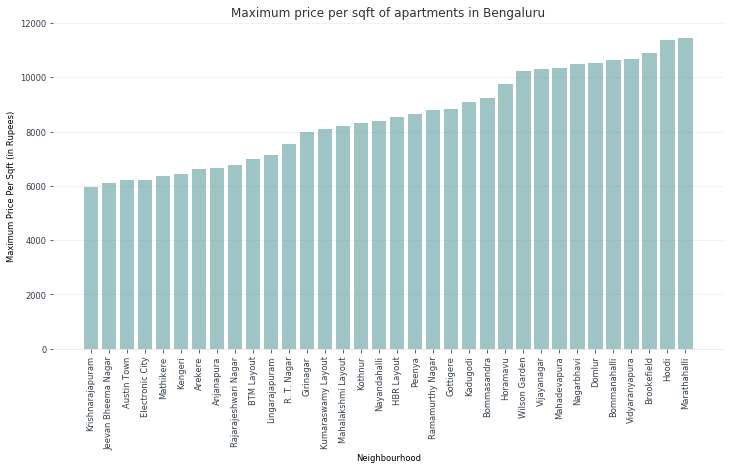


Figure 8: Bar chart depicting the maximum per sqft cost of buying an apartment

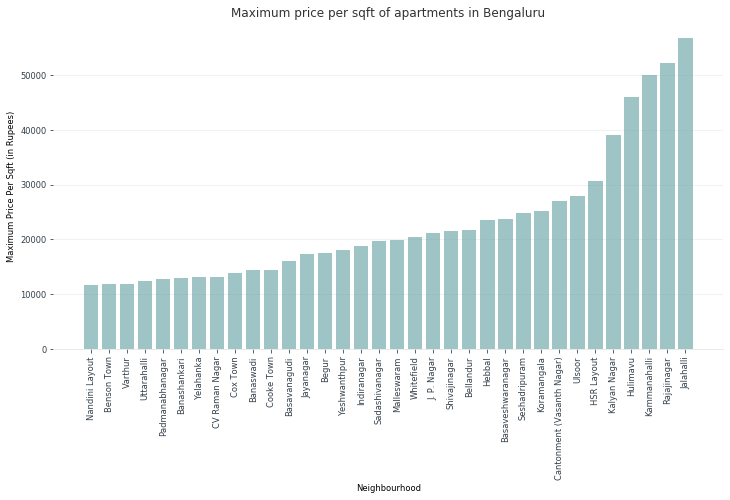


Figure 9: Bar chart depicting the maximum per sqft cost of buying an apartment

1. Overview of areas in Bengaluru:

The dataset was grouped by the attribute ‘Area’ to get analyse the price trends in different parts of Bengaluru as shown in Fig.10, Fig. 11 and Fig.12. It can be concluded that central Bengaluru is the most expensive and eastern Bengaluru the cheapest in terms of buying an apartment.

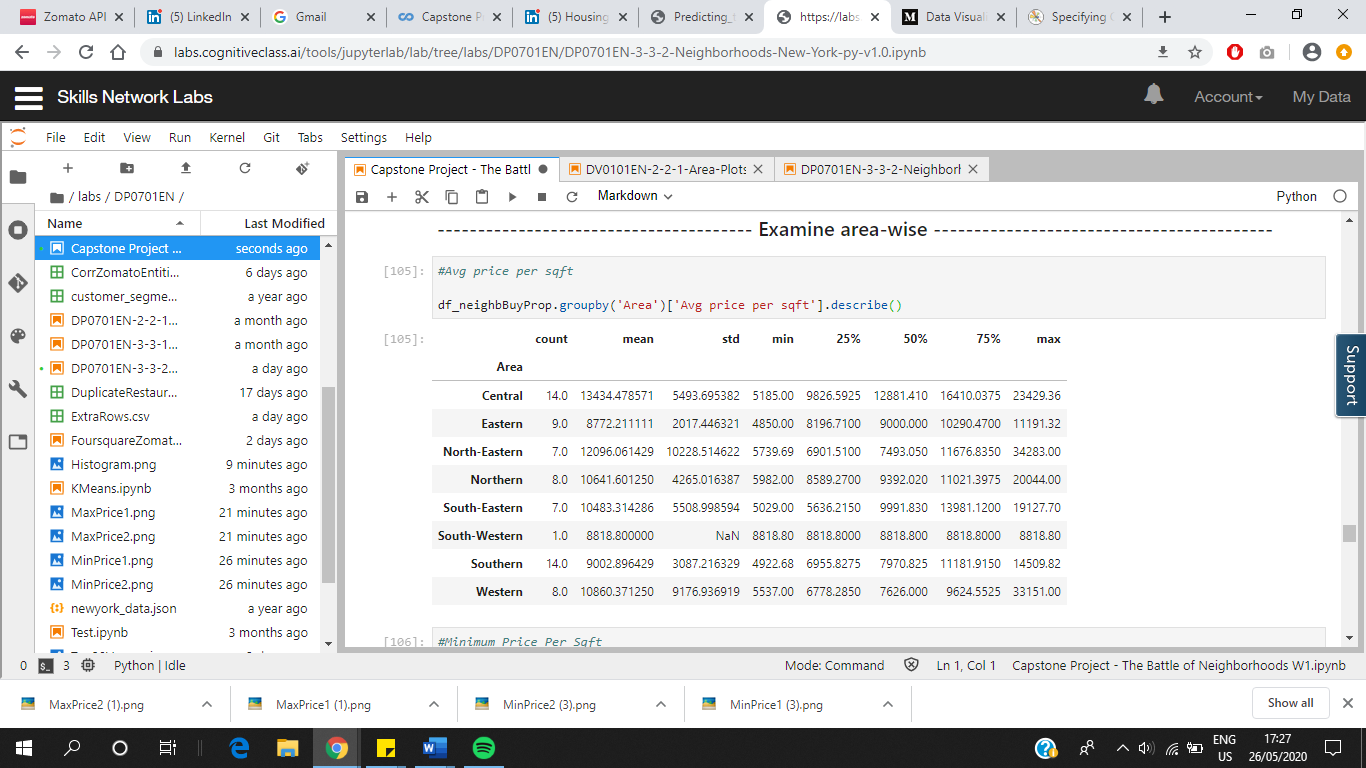


Figure 10: Average price per sqft analysis of areas

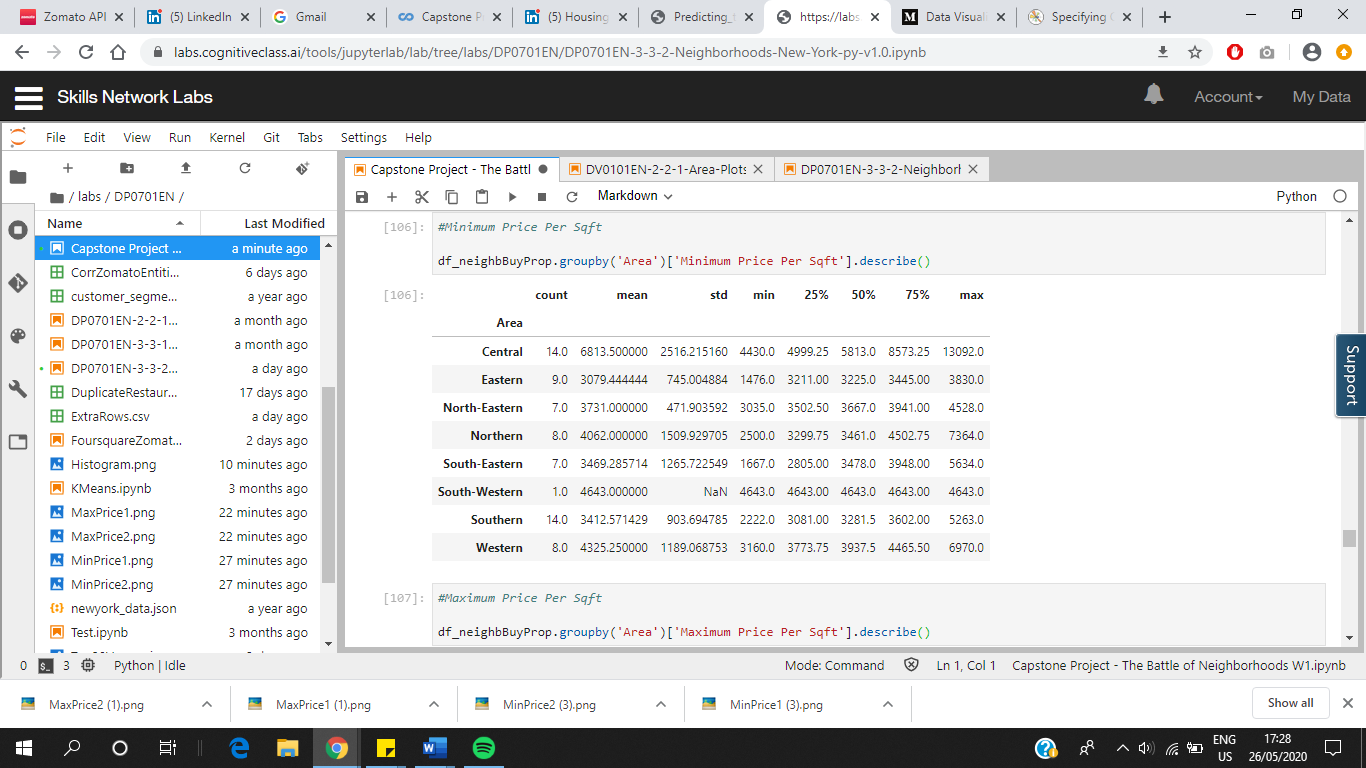


Figure 11: Minimum price per sqft analysis of areas

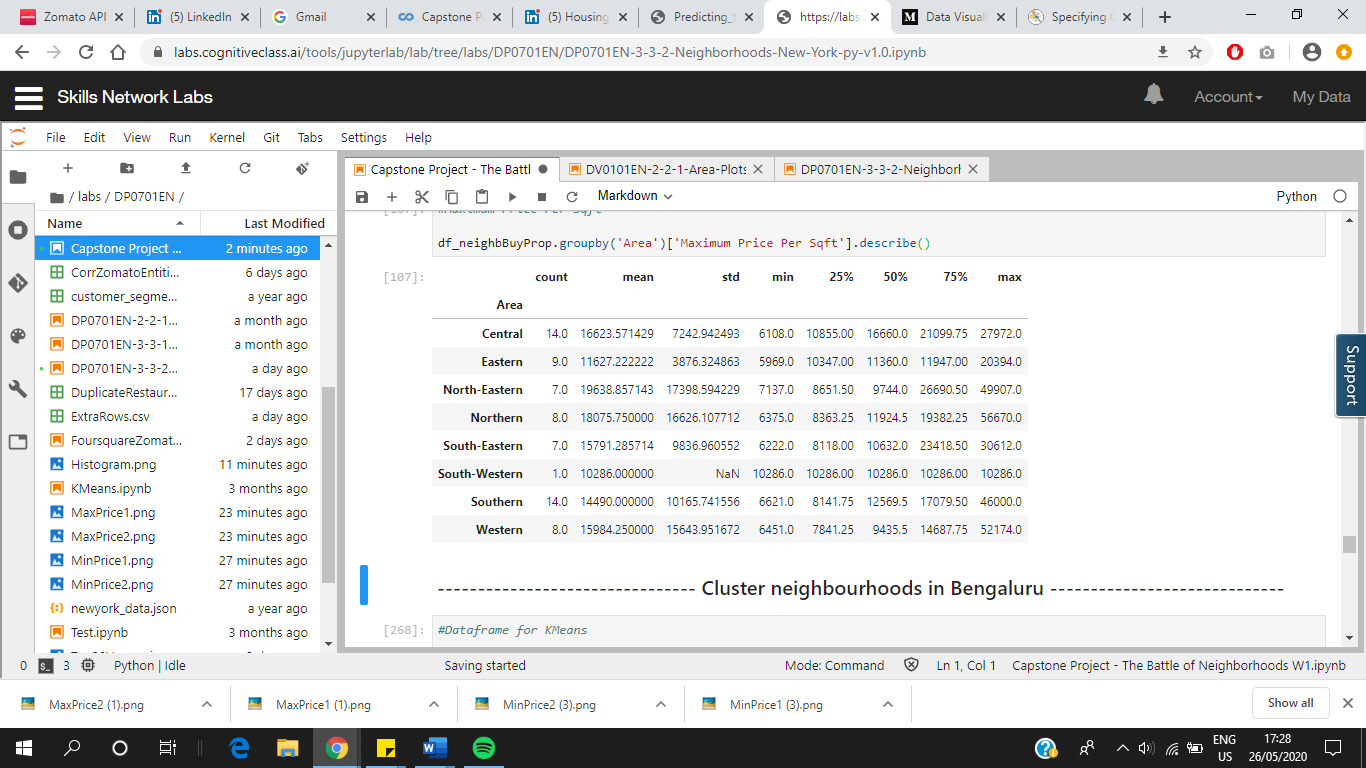


Figure 12: Maximum price per sqft analysis of areas

### Clustering

The neighbourhoods in Bengaluru had to be divided into groups on the basis of per sqft cost of purchasing an apartment. Therefore, the unsupervised machine learning technique K-Means clustering was used to cluster the neighbourhoods on the basis of average, maximum and minimum per sqft prices of apartments. The optimal value of k was found out to be 3 using the elbow method as shown in Fig. 10.

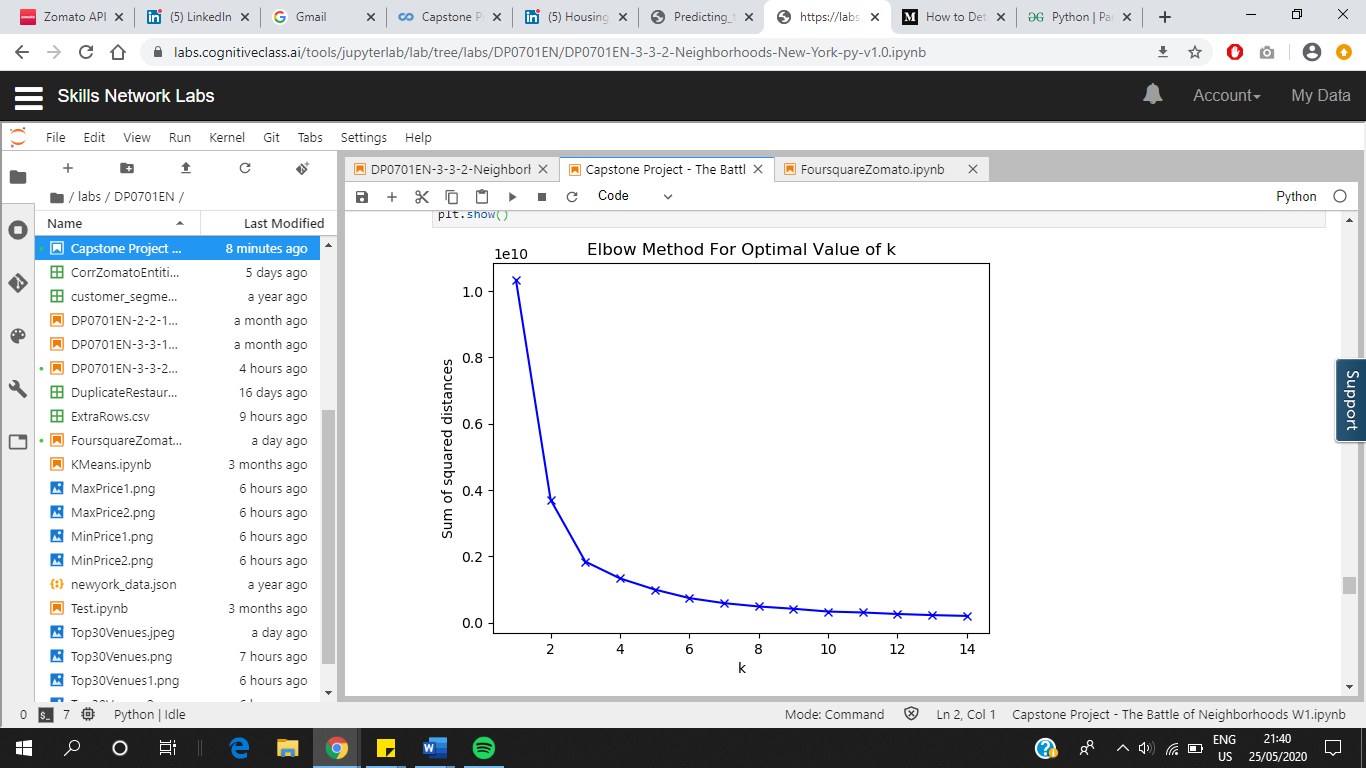


Figure 10: Sum of squared distance vs value of k from 1 to 14

# **Results**

The k-means clustering grouped the dataset into three clusters on the basis of the attributes – average cost per sqft, minimum cost per sqft and maximum cost per sqft. An overview of clusters is shown in Fig. 10, Fig. 11 and Fig.12. Cluster 0 contains localities with the cheapest apartments and accounts for 68% of the neighbourhoods. Cluster 1 has neighbourhoods which are expensive in Bengaluru. These neighbourhoods have a high average and maximum cost per sqft. Neighbourhoods in Cluster 2 are averagely priced.

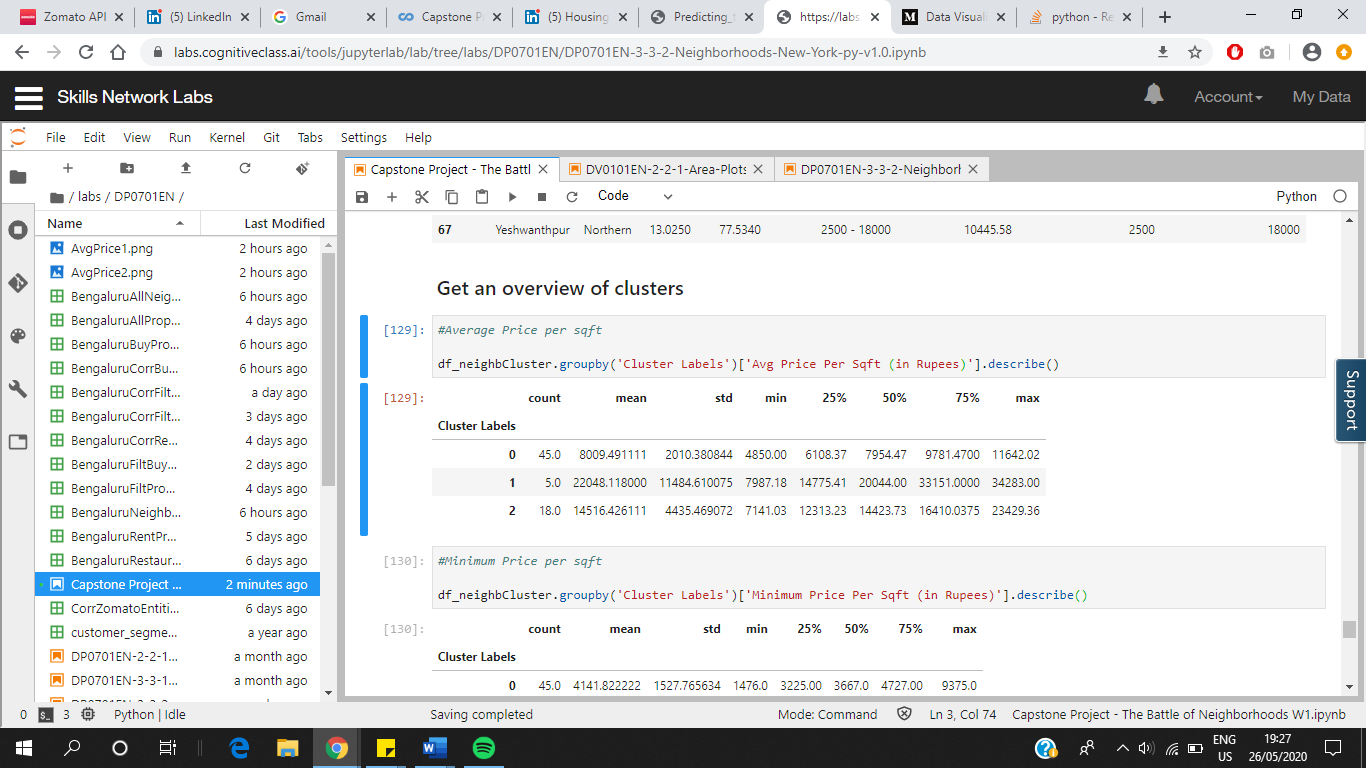


Figure 10: Average price per sqft analysis of clusters

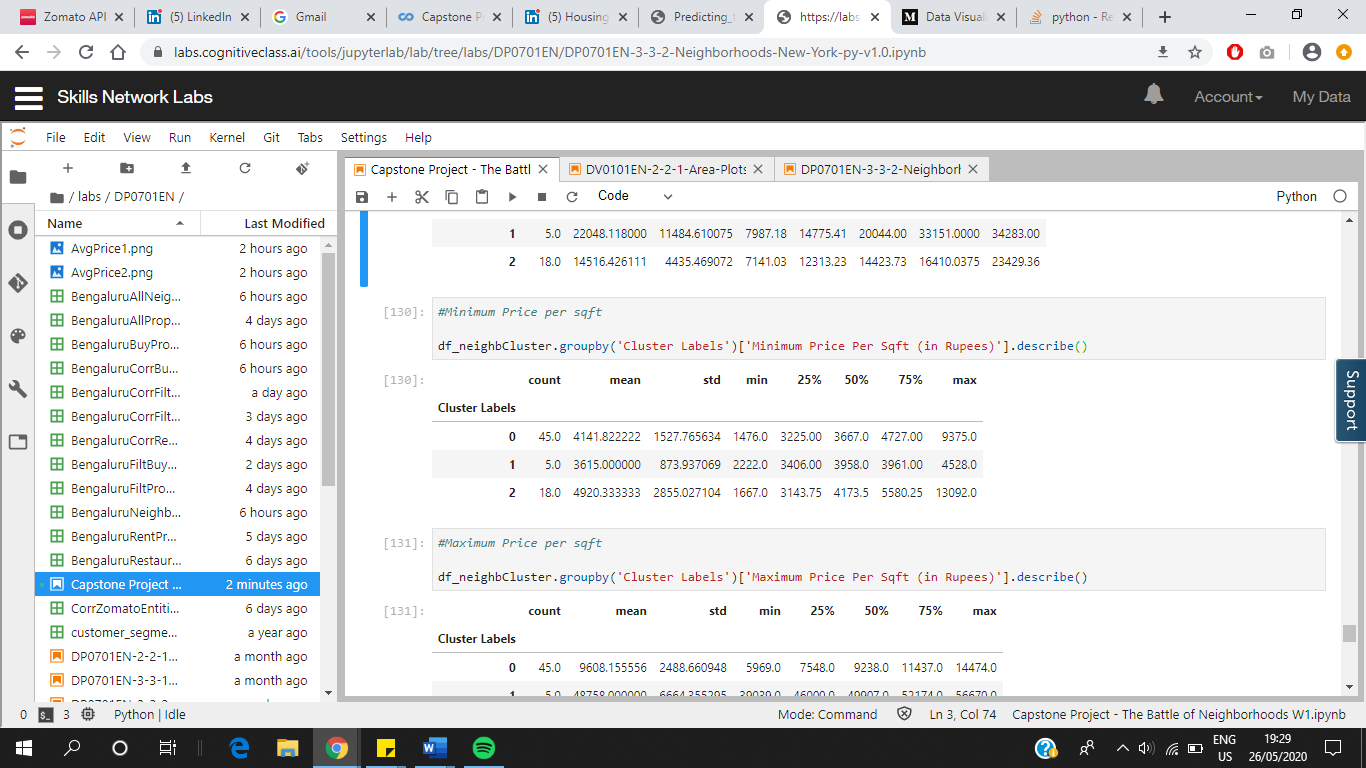


Figure 11: Minimum price per sqft analysis of clusters

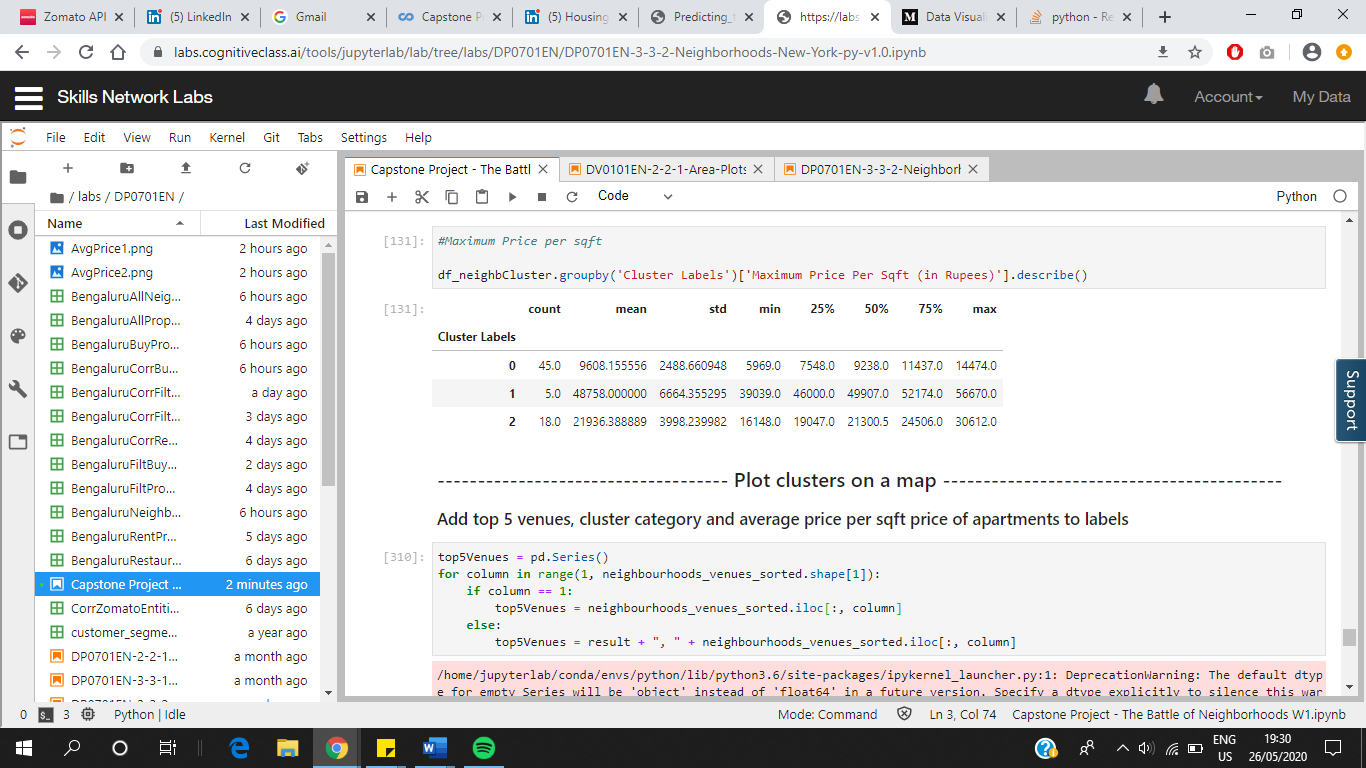


Figure 12: Maximum price per sqft analysis of clusters

The clusters were plotted on a map as shown in Fig. 13. Red, purple and green markers signify cluster 0, cluster 1 and cluster 2 respectively. The clusters were given labels as shown in Table 5.

**Table 5: Cluster labels**

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO. | CLUSTER | COLOUR | LABEL |
| 1. | Cluster 0 | Red | Affordable |
| 2. | Cluster 1 | Purple | Expensive |
| 3. | Cluster 2 | Green | Averagely priced |

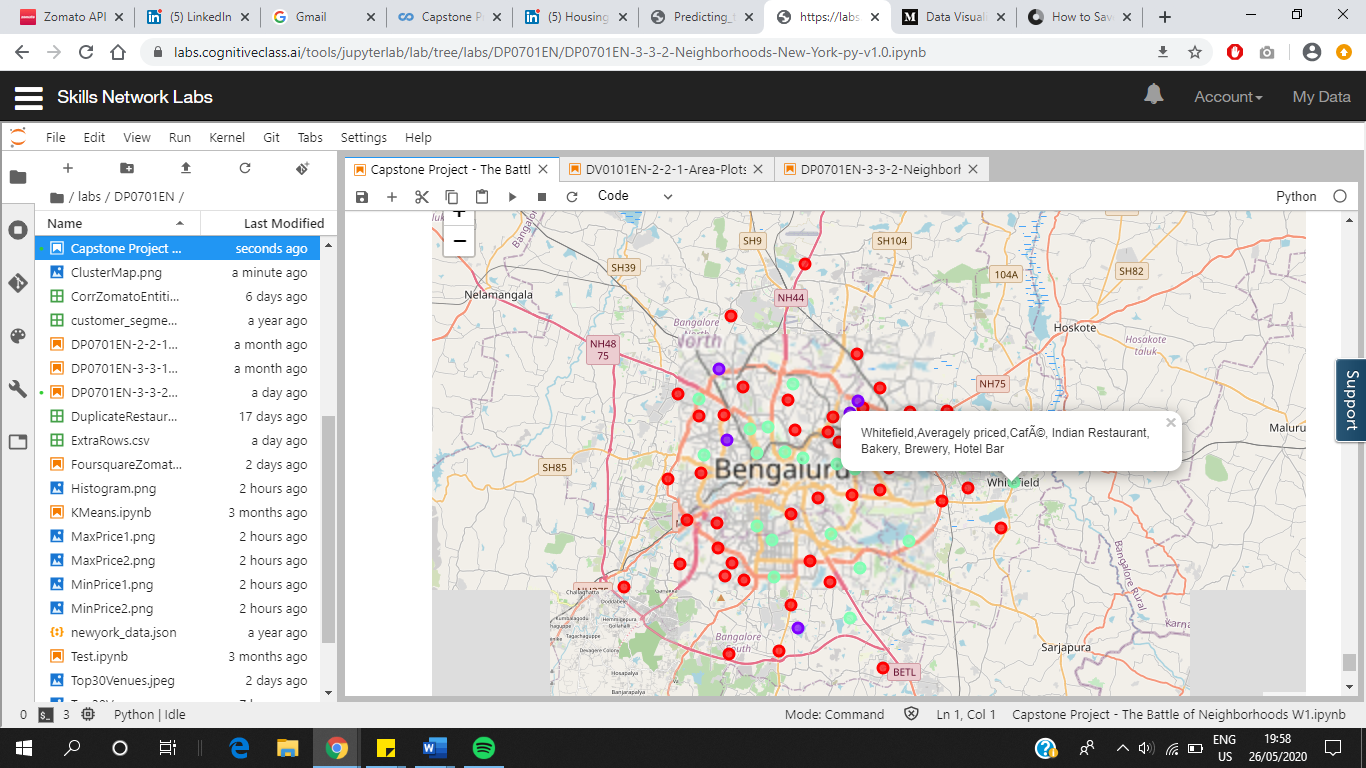


Figure 13: Clusters plotted on a map

# **Discussion**

Finding an accommodation is a challenge that many of us face. One has to consider several things like cost, neighbourhood, facilities nearby, etc while narrowing down houses. The task of finding an apartment in the metropolitan city Bengaluru was simplified by using K-means on a dataset gathered from Foursquare and real estate website Makaan.com. The elbow method implied that the optimum value of k was 3, i.e., number of clusters that can be formed from the dataset. A total of 68 neighbourhoods were clustered into three groups. It was observed that central Bengaluru is the most expensive and eastern Bengaluru is affordable in terms of buying an apartment.

This project can be extended in multiple ways. We can do the same analysis on datasets pertaining to costs of buying a plot, builder floor and independent house. Another useful study would be comparing neighbourhoods on the basis of rental costs of apartments.

# **Conclusion**

The vast quantity of data available in today’s world can be leveraged to get many hidden insights. In this study, data was obtained from Wikipedia, Foursquare, Makaan.com and 99acres.com for the purpose analysing neighbourhoods in Bengaluru with respect to venues and prices of apartments. K-means clustering technique was used to cluster neighbourhoods in Bengaluru on the basis of apartment prices. The top 5 types of venues of each neighbourhood were also found to assist in selecting neighbourhoods for buying an apartment.

# **References**

[1] https://worldpopulationreview.com/world-cities/bangalore-population/

[2] https://bangaloremirror.indiatimes.com/bangalore/others/bengaluru-is-the-most-traffic congested-city-in-the-world-report/articleshow/73698805.cms