**BIRD’S EYE VIEW OF RESTAURANT TYPES AND HOUSING PRICES IN BENGALURU**

**Prepared by Prudhvi**

1. **Introduction**

**1.1 Background**

Bengaluru is the metropolitan city and capital of the Indian state of Karnataka. The city consists of huge number of IT companies, government bodies, colleges and most importantly restaurants. In an effect, so many people are migrating to bengaluru for different reasons and planning to buy houses according to their wants and needs. Out of so many factors which lead to the decision whether to buy house or not, this project going to take two factors into consideration. First one is cost of the house area in a particular area. Second one is restaurant type in a particular area where house is situated.

**1.2 Business Problem**

Even though people can search for houses by travelling around the city with respective to the venues around the house, they are losing so much of time and energy and gaining confusion as they search in a metropolitan city like bengaluru. To circumvent the problem of losing time and energy and gaining confusion, this project going to create a folium map representing house areas as different colored dots which are grouped into different clusters based on type of restaurant surrounded the area and price/sqft of that house area. For each dot, a popup label is created which shows neighborhood’s name, top 3 common types of venues and price/sqft of that area. This is going to create a bird’s eye view for the investors to explore the properties of bengaluru more efficiently and swiftly

**2. Data Description**

Two data sets which are available in kaggle website were used to achieve the solution. First one is Zomato’s data (available in csv format) about different restaurants in bengaluru. Second one is housing data (available in csv format) having price/sqft information for each neighborhood. Zomato data consists of 50,000 restaurants, 88 neighborhoods and 17 columns (for each restaurant) such as restaurant name, restaurant theme and neighborhood of the restaurant etc. Housing data consists of 81 neighborhoods and 9 columns such as neighborhood name, price etc.

**3. Methodology**

**3.1 Data Cleaning and Preparation**

Even though there are so many columns in each dataset, only neighborhood name, restaurant type and price of the housing area was considered. Wherever missing information exists in any cell, whole row was deleted from the dataset.



Figure1: Pandas dataframe with essential columns which are extracted from the zomato.csv file.

Above figure represents the dataframe of the zomato data where only 4 columns were considered out of 17 columns. Similarly, housing price dataframe was prepared. To plot the neighborhoods on the folium map, one needs to have coordinates of neighborhoods. In this project googlemaps was used to extract the coordinates of the neighborhoods using neighborhoods names.

**3.2 Clustering of Neighborhoods**

There are around 50,000 restaurants in 88 neighborhoods of bengaluru.

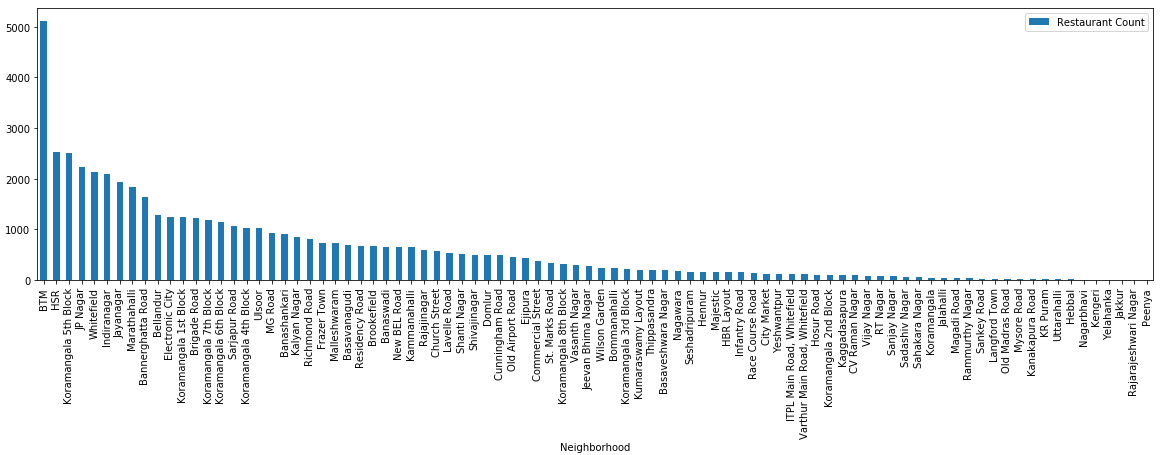


Figure2: Bar graph of Number of restaurants Vs Neighborhood’s names

As our aim is to portray the neighborhoods into different groups in different colors, neighborhoods needs to be clustered based upon the restaurants located in the neighborhoods. To execute clustering, firstly, categorical data is converted into numerical data. Secondly, after conversion, optimal K value is found in K Means algorithm using elbow method. In this case, K value is between 12 and 16. KneeLocator is used to find the optimal K value (elbow point).

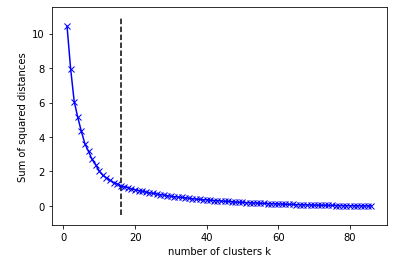


Figure 3: Plot of sum of squared distances between inside clusters points Vs number of clusters

In the above figure, as one can see the elbow point, optimal K value is between 14 and 16. After clustering with optimal K= 14 value, cluster labels were attached to each neighborhood. Finally, in each neighborhood, after finding top 10 most common restaurant types, our final data frame is completed with price/sqft is attached for every neighborhood.

To find the difference among the clusters based on the type of restaurants located inside neighborhood, top three most restaurant types are considered and a bar graph is plotted.

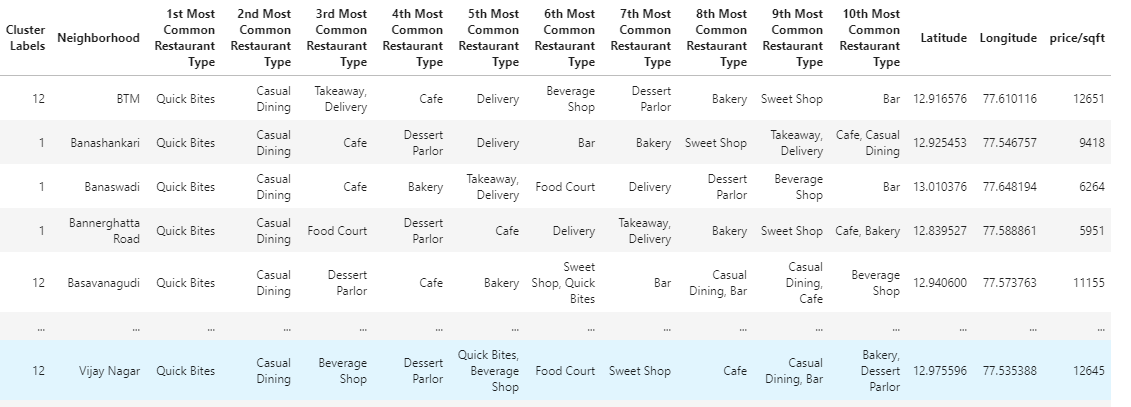


Fig 4: Final pandas data frame to develop folium map and for the analysis of clusters.

To find the difference among the clusters based on the type of restaurants located inside neighborhood, top three most restaurant types are considered and a bar graph is plotted.

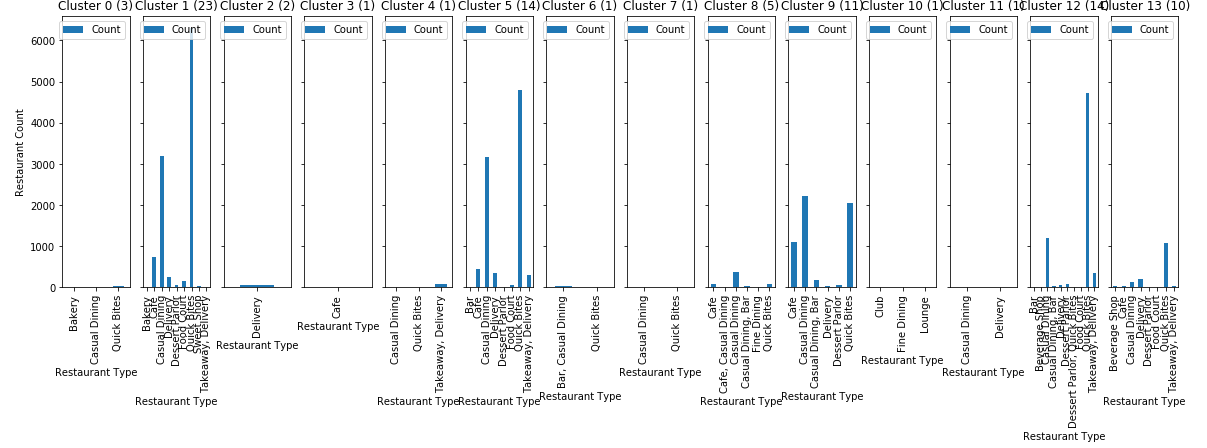


Fig 5: Bar graph of Number of restaurants Vs Restaurant type for all 14 clusters

As we can see, casual dining and quick bites are the most and highest common type of restaurants in most of the clusters. However, quick bites and casual dining too high in number which makes anyone difficult to distinguish the clusters. So, another bar graph is plotted by removing the quick bites and casual dining in every cluster.

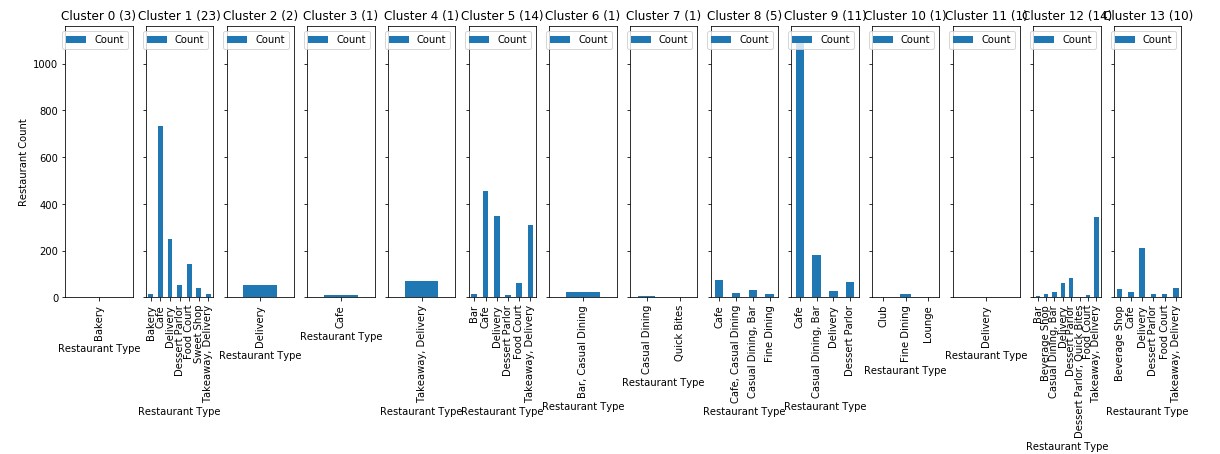


Fig 6: Bar graph of Number of restaurants Vs Restaurant type for all 14 clusters after removing quick bites and casual dining.

Now, one can label each cluster with different names. One can name cluster 9 as “Café Areas” as most of the café were located in this cluster after quick bites and casual dining restaurant types. Similarly, for other clusters too, one can name the cluster by analyzing the graph.

**4. Results**

Finally, bird eye view of bengaluru, the folium map is created by plotting each neighborhood on the begaluru map and a pop label is attached for each neighborhood to show the details such as neighborhood name, top 3 restaurant types and cost of the housing per square meter.

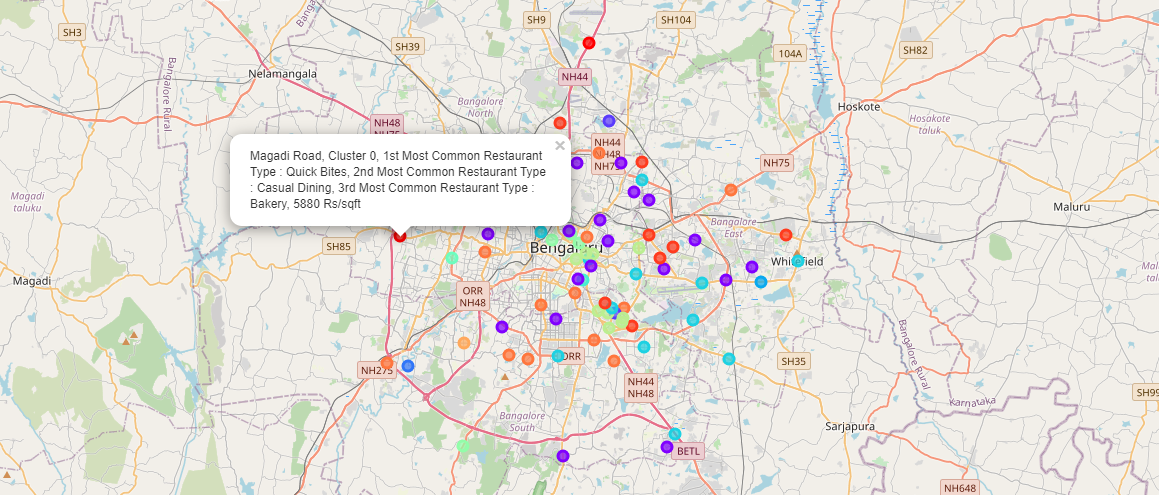


Fig 7: Folium map of bengaluru representing neighborhoods in different colors based on the cluster.

**6. Discussion**

Even though the zomato data is huge, there are some restaurants which are missing in the data frame when compared to present situation. Hence, Google Maps API can be used to extract the venues for about 1.5 km radius for each neighborhood. However, as the number of restaurant types keep increasing the unsupervised classification become harder. Yet that would also increase the accuracy of the folium map. Also, folium map can be enhanced by including a choropleth map for the price data which could lead to a better user interface.

This project can also be taken into the level where this project will suggest a restaurant location as per investor’s requirements and limitations. But that would need even more details of restaurant’s food menu details, ratings, reviews, type of people living around the restaurant and most importantly the type of venues around the restaurant such as colleges, IT hubs etc.

**7. Conclusion**

This bird’s eye view is not only useful for investors such as one who wants to buy house with his/her personal taste for surroundings but also for the investors who want to start a small food business.

**8. Reference**

1. Zomato data: <https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>

2. House price data: <https://www.kaggle.com/amitabhajoy/bengaluru-house-price-data>