

CAPSTONE PROJECT - BATTLE OF NEIGHBORHOODS COIMBATORE

INTRODUCTION

BACKGROUND

Coimbatore is fast growing city and among the top ten districts in TamilNadu, India. It has a population of around 1.6 million people and the city proves to be a great place for living and also for starting new business. Many IT companies have opened their branch in Coimbatore and because of the that more people are moving with their families to Coimbatore which also led to many venues have started showing up in and around Coimbatore.

BUSINESS PROBLEM DESCRIPTION

Because of more people moving into Coimbatore, the rent rates become very high at some important areas in the city. People starting business considers only those places where population is high or they would start business near city centers. Day by day, as the traffic congestion becoming worse, people from different parts hesitate to travel to places with more venues. So only option for new people is to rent house of high cost in areas with more venues. Business people also does not have any data or report showing them the new areas to explore for starting new business.

Solution

This project intends to find out the rent rates in important areas in Coimbatore and the availability of number of venues in those areas.

Target Audience

This projects is intended for two audience.

One is the business poeple who owns the venues. The report will help them to explore new areas. E.G. If a person owns a venue in Gandhipuram which is the city center and people in and around Gandhipuram hesitates to go to Gandhipuram everytime due to heavy traffic, then it ultimately affects his business. What if he knows the place where he should consider opening another shop where new people move largely.

Two is the common people who is moving to Coimbatore with their families.

DATA DESCRIPTION

To get the desired result, I need the following Data -

Our goal of this project is to find two aspects. One is rent rates in important areas of Coimbatore for them to identify areas with lower rent rates and more venues. Second, for the business people to explore areas where venues are less and where more people will move based on rent rates, so that, they could think of business expansion.

We need data of rent rates in important areas of Coimbatore. I did a thorough search of such data and found out a structured data from Makkan website. The data contained four feature sets. - 'Neighborhood', '1 BHK', '2BHK' & '3BHK'. However, I thought it would be wise to find the mean of all the feature sets so as to let the people know the average rents for all kinds of house in an area.

Next, we need number of venues in those areas which I will pass a query to Four Square website to retrieve the data. Example, I want to move to area with low rent rates and availability of more venues, especially gym. In that case, retrieving venues would be appropriate because venues will be clustered and named as per its characteristics.

After retrieving the above data, I have planned to create a Choropleth map which will show the areas & their rent rates and number of venues and type of venues in those areas.

To create the Choropleth map to show the different rent rates, I need the Coimbatore geographical border data. But, I could not find data for the latitudes and longitudes. However, I found one file containing the geo data based on ward numbers and fortunately, I was able to find the ward numbers of the important areas of Coimbatore.

Next, as I have planned to cluster venues retrieved from Foursquare, I need to get the latitude and longitude data of each location to pass query and get venues data. For getting the latitude and longitude values of important areas, I used Python geocoder package.

METHODOLOGY & ANALYSIS

Neighborhood Data - Coimbatore:

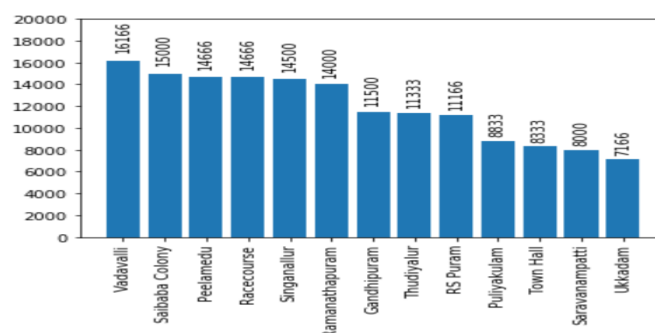
As I have described in my Data description content, I was able to collect rental figures of important neighborhoods in Coimbatore from Makaan website. Using pandas library, I downloaded the data.

Neighborhood	1 BHK	2 BHK	3 BHK	Ward No
Gandhipuram, Coimbatore	7500	12000	15000	51
Ramanathapuram, Coimbatore	8000	14000	20000	68
Singanallur, Coimbatore	7500	15000	21000	64
Peelamedu, Coimbatore	9000	15000	20000	39
Saravanampatti, Coimbatore	3500	7500	13000	31

The feature sets of the data frame consisting of four items which are 'Neighborhood', '1 BHK', '2BHK', '3BHK' & Ward No. Next, I have to find the average rental rates in important areas of Coimbatore because our objective is to include rent rates for all kinds of house collectively. So I added another column called “Average Rent” which shows the **mean** value of 1 BHK, 2 BHK & 3BHK.

Neighborhood	1 BHK	2 BHK	3 BHK	Ward No	Average Rent
Vadavalli, Coimbatore	8500	15000	25000	18	16166
Saibaba Colony, Coimbatore	8000	12000	25000	22	15000
Peelamedu, Coimbatore	9000	15000	20000	39	14666
Racecourse, Coimbatore	9000	15000	20000	71	14666
Singanallur, Coimbatore	7500	15000	21000	64	14500

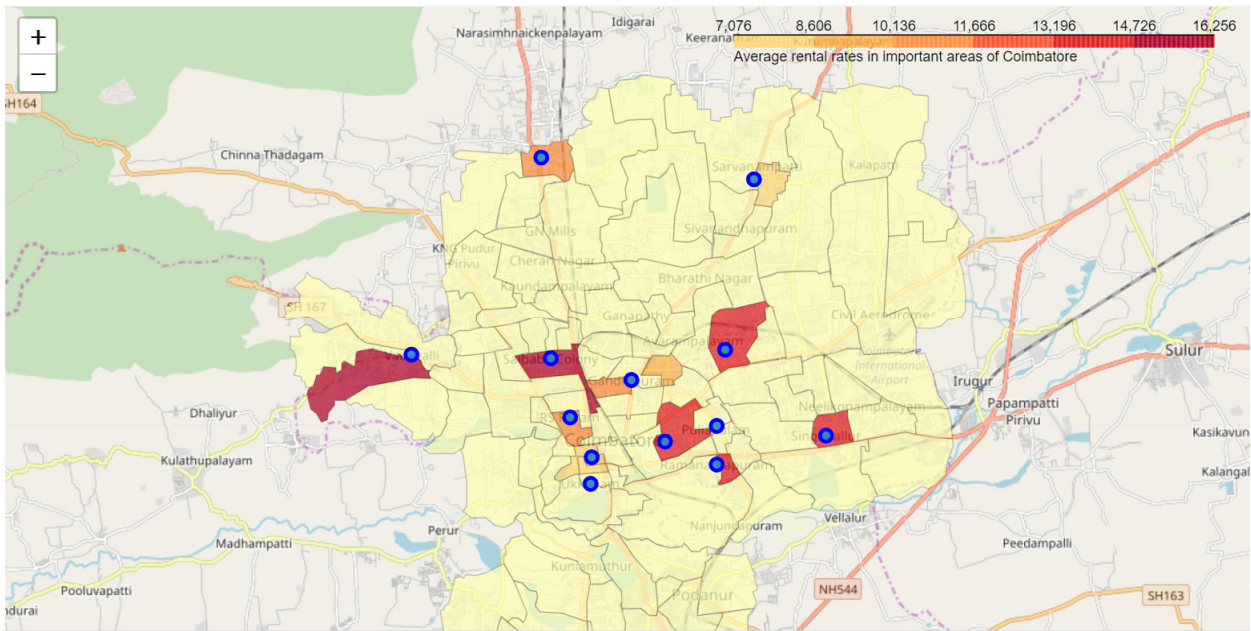
Next, I used **matplotlib** library to create a bar diagram to visually see the range of rent rates among different areas of Coimbatore.



From the above graph, “Vadavalli” has the highest rent rates and “Ukkadam” is the area with lowest rent rates.

Using folium library in python, I created a choropleth map to visualize the average distribution of rental rates in a different perspective. For creating a folium map, used the **geopy** library to get the latitude and longitude data of Coimbatore and its neighborhoods. Next,

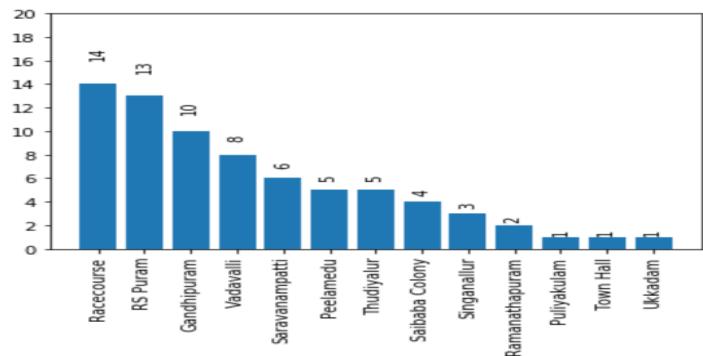
retrieved a geojson file containing the border geo data of Coimbatore based on Ward Number from **GITHUB**. Also, I superimposed the locations of important areas by marking them using blue ovals.



Next, my objective was to find out the number of venues in each area of Coimbatore. Hence, I passed query to **FOURSQUARE API** to fetch the venue data. Initially, I fetched the venue details of only “Gandhipuram” area. Later, I fetched venue details of all areas and it provided a total venue count of **73**. Then, created a dataframe which shows the count of venues in each area.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Racecourse, Coimbatore	14	14	14	14	14	14
RS Puram, Coimbatore	13	13	13	13	13	13
Gandhipuram, Coimbatore	10	10	10	10	10	10
Vadavalli, Coimbatore	8	8	8	8	8	8
Saravanampatti, Coimbatore	6	6	6	6	6	6
Peelamedu, Coimbatore	5	5	5	5	5	5
Thudiyalur, Coimbatore	5	5	5	5	5	5
Saibaba Colony, Coimbatore	4	4	4	4	4	4
Singanallur, Coimbatore	3	3	3	3	3	3
Ramanathapuram, Coimbatore	2	2	2	2	2	2
Puliyaikulam, Coimbatore	1	1	1	1	1	1
Town Hall, Coimbatore	1	1	1	1	1	1
Ukkadam, Coimbatore	1	1	1	1	1	1

The table shows the count of venues in descending order. However, I created a graph to visualize the data.



The above graph shows that Racecourse topping in the count of venues and Ukkadam has the lowest count. The foursquare data showed 41 unique categories of venues and using that I crated a table showing the top ten venues in each area.

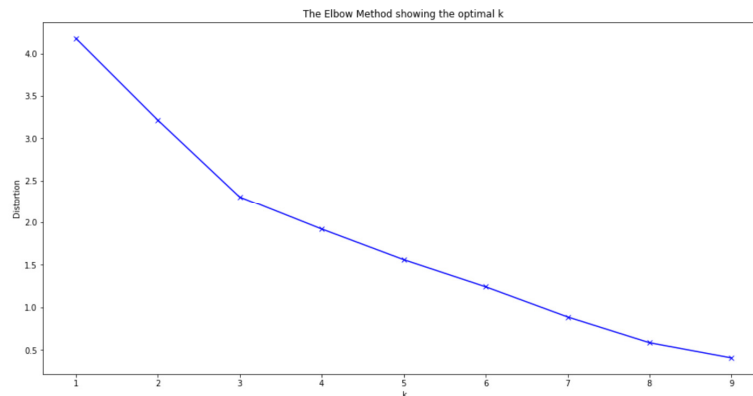
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
Gandhipuram, Coimbatore	Women's Store	Indian Restaurant	Multiplex	Asian Restaurant	Vegetarian / Vegan Restaurant	Miscellaneous Shop	Ice Cream Shop	Snack Place	Pool Hall
Peelamedu, Coimbatore	Airport Service	Convenience Store	Snack Place	Women's Store	Coffee Shop	Fast Food Restaurant	Farmers Market	Electronics Store	Diner
Puliyakulam, Coimbatore	Gym	Hotel	Fast Food Restaurant	Farmers Market	Electronics Store	Diner	Dessert Shop	Department Store	Convenience Store
RS Puram, Coimbatore	Fast Food Restaurant	Ice Cream Shop	Indian Restaurant	Asian Restaurant	Café	Coffee Shop	Dessert Shop	Diner	Electronics Store
Racecourse, Coimbatore	Ice Cream Shop	Bistro	Hotel	Vegetarian / Vegan Restaurant	Indian Restaurant	Department Store	Coffee Shop	Café	Park
Ramanathapuram, Coimbatore	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Farmers Market	Electronics Store	Diner	Dessert Shop	Department Store
Saibaba Colony, Coimbatore	Chinese Restaurant	Indian Restaurant	Bakery	Outdoors & Recreation	Coffee Shop	Farmers Market	Electronics Store	Diner	Dessert Shop
Saravanampatti, Coimbatore	Coffee Shop	Indian Restaurant	Kerala Restaurant	Pool Hall	Burger Joint	Dessert Shop	Women's Store	Farmers Market	Electronics Store
Singanallur, Coimbatore	Performing Arts Venue	Farmers Market	Multiplex	Chinese Restaurant	Electronics Store	Diner	Dessert Shop	Department Store	Convenience Store

Since I had the list of top ten venues in each area, I decided to cluster them using **k-means Clustering** technique to do that. **K-Means** is an unsupervised machine learning algorithm that groups data into **k** number of clusters. The number of clusters is user-defined and the algorithm will try to group the data even if this number is not optimal for the specific case.

For optimum clustering of venues, I need to find out the optimum **k**. I used **Elbow Method**.

The Elbow method is a very popular technique and the idea is to run k-means clustering for a range of clusters k (let's say from 1 to 10) and for each value, we are calculating the sum of squared distances from each point to its assigned center(distortions).

When the distortions are plotted and the plot looks like an arm then the “elbow”(the point of inflection on the curve) is the best value of k . Using the method, I found the optimum k which is **3**.



After clustering, provide cluster labels for each cluster and merged the data with our table containing top 10 venues in each area.

Neighborhood	1 BHK	2 BHK	3 BHK	Ward No	Average Rent	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Vadavalli, Coimbatore	8500	15000	25000	18	16166	11.025339	76.905125	0	Halal Restaurant	Street Food Gathering	Indian Restaurant	Market	Department Store
Saibaba Colony, Coimbatore	8000	12000	25000	22	15000	11.024334	76.944788	0	Outdoors & Recreation	Indian Restaurant	Bakery	Chinese Restaurant	Women's Store
Peelamedu, Coimbatore	9000	15000	20000	39	14666	11.026958	76.994581	0	Taco Place	Snack Place	Convenience Store	Women's Store	Coffee Shop
Racecourse, Coimbatore	9000	15000	20000	71	14666	11.001079	76.977557	0	Coffee Shop	Hotel	Café	Vegetarian / Vegan Restaurant	Ice Cream Shop
Singanallur, Coimbatore	7500	15000	21000	64	14500	11.002859	77.023495	0	Performing Arts Venue	Farmers Market	Street Art	Restaurant	Chinese Restaurant
Ramanathapuram, Coimbatore	8000	14000	20000	68	14000	10.994777	76.992130	0	ATM	Indian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Bakery
Gandhipuram, Coimbatore	7500	12000	15000	51	11500	11.018271	76.967774	0	Women's Store	Indian Restaurant	Multiplex	Asian Restaurant	Vegetarian / Vegan Restaurant

Based on the characteristics of each cluster, I named them as follows,

- Cluster '0' - “Top Class”
- Cluster '1' - “Students Class”
- Cluster '2' - “Low Budget Family Class”

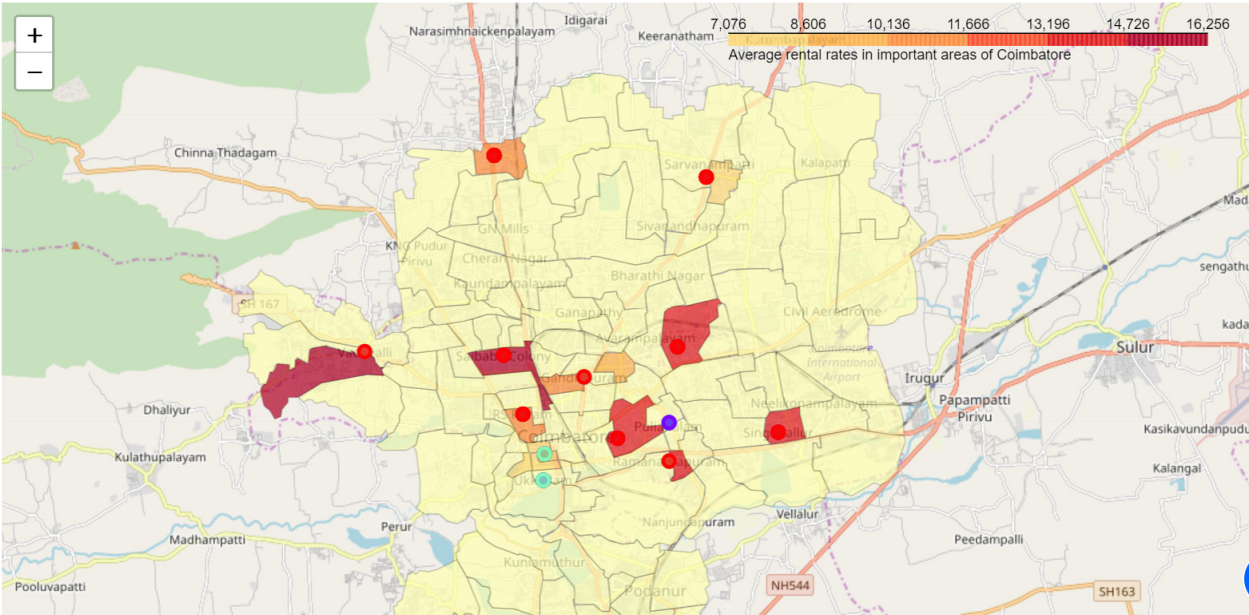
Then added the new cluster name to the data frame created based on the cluster numbers. Also, categorized the areas based on the rent rates,

- Rs 7000 - 9000 = “Low Range”
- Rs 9001 -12000 = “Middle Range”
- Rs 12001 – 17000 = “High Range”

Later, created a final table consisting of the above classification as well under the columns – Venue Category and Area Category.

Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Category	Area Category	Venue Category
0	Indian Restaurant	Street Food Gathering	Restaurant	Department Store	Pharmacy	Mobile Phone Shop	Women's Store	Chinese Restaurant	Electronics Store	Diner	Top Class	High Range	Top Class
0	Chinese Restaurant	Indian Restaurant	Bakery	Outdoors & Recreation	Coffee Shop	Farmers Market	Electronics Store	Diner	Dessert Shop	Department Store	Top Class	High Range	Top Class
0	Airport Service	Convenience Store	Snack Place	Women's Store	Coffee Shop	Fast Food Restaurant	Farmers Market	Electronics Store	Diner	Dessert Shop	Top Class	High Range	Top Class
0	Ice Cream Shop	Bistro	Hotel	Vegetarian / Vegan Restaurant	Indian Restaurant	Department Store	Coffee Shop	Café	Park	Bar	Top Class	High Range	Top Class
0	Performing Arts Venue	Farmers Market	Multiplex	Chinese Restaurant	Electronics Store	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Top Class	High Range	Top Class
0	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Farmers Market	Electronics Store	Diner	Dessert Shop	Department Store	Convenience Store	Top Class	High Range	Top Class
0	Women's Store	Indian Restaurant	Multiplex	Asian Restaurant	Vegetarian / Vegan Restaurant	Miscellaneous Shop	Ice Cream Shop	Snack Place	Pool Hall	Café	Top Class	Middle Range	Top Class

Finally created the Choropleth map with superimposed clusters showing the venue category and rent range.



RESULTS AND DISCUSSION

The final table shows the areas categorized under two different titles. One is categorized under the title "Average Rent" and another under "Venues" in each area. The table shows areas where people could move with low rent rates and more venues, or move to areas where venues are student oriented and yet high rent rates etc. It also shows the areas where business people could consider about opening new venues.

Initially, after downloading the data about different rent rates based on number of rooms in a house, I decided to find out the mean or average rent rates to keep the focus of including all kinds of house for all types of people. After that I created a bar graph to visually see the rent rates sorted in descending order. From that, I could see "Vadavalli" areas was having the highest rent rates and "Ukkadam" was having the lowest rent rate of all. The reason to collect this data was to find any correlation between rent rates and availability of number of venues in an area, to see, if they are directly proportionate to each other. Example - high rent rates in might be due to more number of venues available.

Then I decided to create choropleth map to see the shading of Coimbatore location with average rent rates. When I created the map, it was bit of shock to see "Gandhipuram" area, which is the city center, was showing moderate rent rate compared to areas far away from city center with high rent rates. As a resident of Coimbatore, I could assume the reason for "Vadavalli" being the place with high rent rates because it is a residential area. Along with map, to identify different areas, I used geopy package to find out the latitude and longitude of important areas of Coimbatore.

Then, my next step was obviously finding out the number of venues in each area. By passing query to "Foursquare API", I found out the count of venues in each area. Then, I created a bar graph to see if that graph matches with our another graph "Rent Rates". I compared the data and found that rent rates are not directly proportionate to number of venues completely. Because, I expected "Vadavalli" to be having more number of venues because of high rent rates. But Racecourse, RS Puram, Gandhipuram occupied the first three places moving Vadavalli to fourth place in count of venues. As I assumed earlier, Vadavalli's rent rate is high which might be due to its residential nature or some other factors. Now, we have the count, from the table data, I could say that, Racecourse, RS Puram, Gandhipuram would be optimum place for people who are moving to Coimbatore with their families because the rent rates are moderate and availability of venues also high. Likewise, Ukkadam, Townhall, Puliyakulam are the areas where business people should consider opening new shops. The reason being, the rent rates are moderate and if more venues are opened, there are high chances that new people who migrate will find these places appropriate.

As a additional information, I tried to cluster the venues and provided them name based on their characteristics, hence, people could select areas with venues which they like. Also, Business people could see whether they could open shops which are not available in these areas. Example, Pharmacy is not available in "Low Budget Family Class". Hence, opening a shop in those areas would be beneficial.

I used K Means Clustering technique to cluster the venues. For identifying the optimum K, I used "Elbow Method" which showed 4 as the optimum value. Hence, created 4 clusters and named them according to their characteristics.

Finally, I created a Choropleth map using "Average Rent" rates and superimposed clusters labels. To analyze the results, I created a data frame which shows all the areas with two columns - "Category" showing the venue category and "Area Category" showing rental category. From that, one could identify the areas where one could move with their family and business people could identify areas to explore. It can be further extended if we could get more data regarding rental rates in all areas of Coimbatore and by using various other data science methods or techniques, we could drill much more information.

CONCLUSION

The final result seems to be promising in fulfilling the objectives of our project. When we see the final table after naming the different clusters, we could find that there are few areas which are city centers with more number of venues and yet the rental rates are normal when compared to places with less venues and higher rent. However, there are many factors which affects the rent rates in different areas and "Vadavalli" is an example of that. If we have more information, we could drill down further as to see why the rental rates are high and what are the other factors affecting the rent rates. That might help people to look for areas with different characteristics which might suit them more than just venues. Also, by describing areas in detail could help business people to decide what kind of venue would suit those areas or to find out what are the other business prospects available in those areas.