



JUNE 20, 2020

GERMAN RESTAURANT

LOCATION PLANNING - BANGALORE

AVINASH KALARICKAL



TABLE OF CONTENTS

1) Introduction	2
a) <i>Background</i>	2
b) <i>Problem</i>	2
c) <i>Target group</i>	2
2) Business Requirements	3
a) <i>Where in Bangalore?</i>	3
b) <i>Where are most restaurants located?</i>	3
c) <i>How much do people spend for a meal on average in a restaurant?</i>	3
3) Data Analysis & Visualization	4
a) <i>PART I – Importing Data</i>	4
b) <i>PART II: Defining the location(s) of restaurant</i>	6
c) <i>PART III: Defining the "Average Cost For Two People"</i>	10
4) Conclusion	14
5) Future steps	14
a) <i>To define location</i>	14
b) <i>To define the average cost for two people</i>	14

1) INTRODUCTION

a) Background

I am a resident of Germany and would like to return to my home country, India. One of the plans on return is to start a restaurant serving German cuisines. The decision will be taken based on the business questions listed below. **So, the target audience is me who has decided to setup a restaurant in the city.**

The focus is on my hometown Bangalore, India as the city. The city has been growing migrant population due to the vast number of start-ups and IT companies. The population is also increasing depending on outside dining compared to self-cooking. The city also has Indian headquarters of some of the premium German companies like Siemens, Bosch, Continental, etc. This leads to high number of German Expats and families in the region.



"Bengaluru (also called Bangalore) is the capital of India's southern Karnataka state. The centre of India's high-tech industry, the city is also known for its parks and nightlife."

According to a newspaper article in 2016 –

"The city is home to around 200 German companies. An estimated 1,000 Germans live in the city. Karnataka has a tie-up with the German state of Bavaria for co-operation in sustainable agriculture, agricultural engineering, food processing, renewable energy and other allied sectors."

But the city spots only one German café, and no restaurant in the city.

b) Problem

The city of Bangalore is approximately 8000 sq.m, hence deciding the location for the restaurant is not easy. Further there are different localities in Bangalore known for Restaurants, there do not necessarily be downtown. So, the problem is to address the questions raised in the business requirement. This project aims to answer the question where to start a restaurant and what is the average price for dining that can be set as basis for menu planning of the restaurant.

c) Target group

I am being the Non-Residential Indian (NRI) planning to repatriate in the near future would like to gather the outcome of this study project and utilize for my restaurant planning. Hence the Target audience is me and probably other people who plan to start a restaurant in Bangalore.

2) BUSINESS REQUIREMENTS

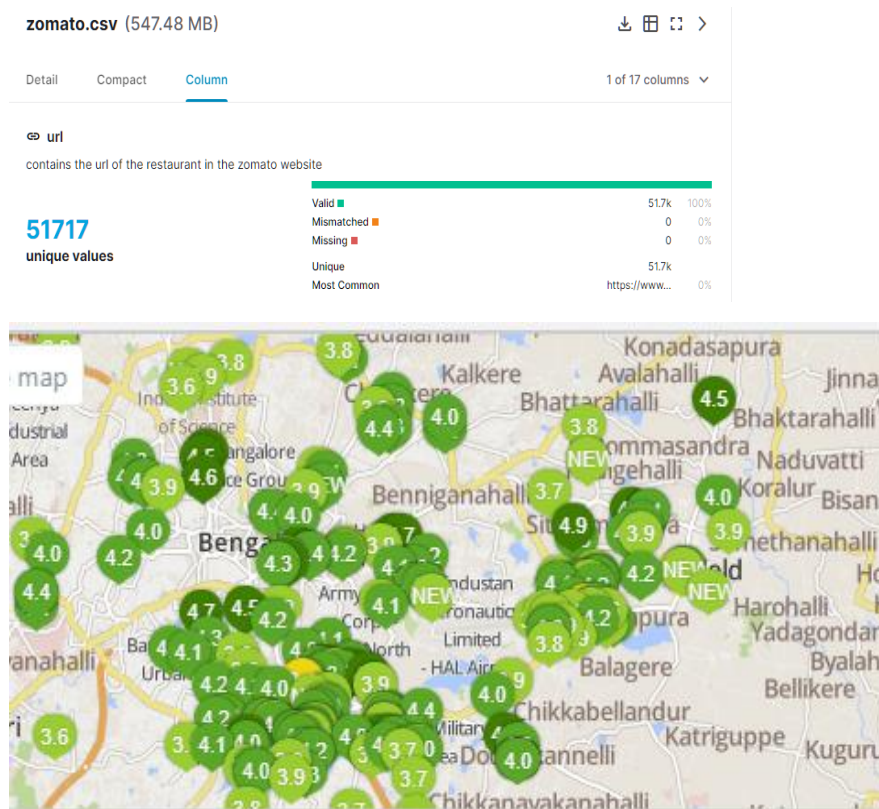
a) Where in Bangalore?

- Area of Bangalore, metropolis is 709 sq. km, with a population of nearly 10 Mio. which leads to the question which part of Bangalore is suitable for the Restaurant.

Area		Population (2011)	
• Metro 8,005 km2 (3,091 sq mi)		• Urban 10,456,000	
• Metropolis	709 km2 (274 sq mi)	• Metropolis	8,443,675
		• Density	12,000/km2 (31,000/sq mi)

b) Where are most restaurants located?

Bangalore city has approximately 52000 restaurants according to dataset from Zomato. This gives rise to the primary question where the density of restaurants is maximum. Which would reflect the market for restaurants.



c) How much do people spend for a meal on average in a restaurant?

- I. The cost of meal should be able to meet the average in the neighbourhood. This value will support in defining the meal cost of the restaurant. Considering the following to meet it:
 - Preparation cost (incl. ingredient, staff, rent, amenities ...)
 - Profit that can be generated
 - Control the portion of food served
- II. Does the cost per person depend on proximity to other venues in area?
 - a. Does the cost have correlation to having a metro train station nearby?
 - b. Does the cost correlate to the proximity to a shopping mall?
- III. Which localities have the greatest number of breweries?

This is essential, as the I do not plan to invest in brewing own beer, there focus on buying beer from a locally brewery.


3) DATA ANALYSIS & VISUALIZATION

a) PART I – Importing Data

1. The Wikipedia page “List of neighbourhoods in Bangalore” contains tables which needs to be scrapped to capture the neighbourhoods (here on called as **Location**). The four tables are categorized as regions. E.g of one of the regions is below:

url : https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore

Central [edit]

Name	Image	Summary
Cantonment area		The Cantonment area in Bangalore was used as a military cantonment during the British Raj in the 19th century. After the Indian independence, the area merged with rest of the city but still retains names for localities and streets used by the British as well as the colonial architecture in many buildings. Some localities which come under this area are Richmond Town , Austin Town , Murphy Town , Fraser Town , and Cox Town .
Domlur		Formerly part of the Cantonment area, Domlur has become a major hub for IT outsourcing firms and military establishments. The Embassy GolfLinks Business Park and HAL Airport are located in/near Domlur.

On scraping using BeautifulSoup html parse method we arrive at 65 locations in Bangalore city. This data contains some location names spelling incorrectly (old spelling vs new spelling issues).

The dataframe is cleaned to remove the columns “image” and “summary” as these do not add value to the analysis we will carry in the next step.

Region	Location
Central	11
Eastern	8
North-Eastern	7
Northern	8
South-Eastern	7
Southern	8
Southern suburbs	6
Western	10

2. The location Vs Pin code dataset. This dataset contains the following columns and **15570** lines of data as a TXT CSV. The Dataset was pre-processed to create a new CSV containing only fields necessary to be easier to handle and fasten the processing time.

url: <http://download.geonames.org/export/dump/IN.zip>

The dataframe has the following columns of which the **highlighted** are of interest for us in the analysis. It is observed that one location can have multiple pincodes which arises as there are more than one postal office in the location based on area of the location. Hence 65 locations from above dataframe do not mean 65 pincode.

Old Name	New name
country code	Country
postal code	Pincode
place name	Location
admin name1	
admin code1	State
admin name2	
admin code2	City
admin name3	
admin code3	
latitude	
longitude	
accuracy	

This Dataset is then pre-processed to filter only the highlighted columns and details for only data with value for City as "Bengaluru" (Alternative name for Bangalore). This leads to **266** locations (Pincode) in Bangalore city.

3. Restaurant dataset is available as a CSV file containing details about **51717** places in Bangalore. This list is generated by restaurant rating platform ZOMATO. The original dataset (csv) is from Kaggle courtesy Mr. Himanshu Poddar (see url below), it contained 17 columns and was 547MB. Due to the size restriction on cognitive class lab, the data was pre-processed on local instance of anaconda and trimmed to contain only the highlighted columns.

url: <https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>

Column	Description
url	Url of the restaurant on the Zomato website
address	Address of the restaurant
name	Name of the restaurant
online_order	Online order available or not (yes/no)
book_table	Restaurant allows booking a table (Yes/no)
rate	Rating of the restaurant on scale of 5
votes	Total number of votes the restaurant has
phone	Phgone number of the restaurant
location	location of the restaurant
rest_type	Type of restaurant
dish_liked	dishes people liked in the restaurant
cuisines	Cuisines served
approx_cost(for two people)	Approximate cost of meal for two people
reviews_list	2 reviews per restaurant
menu_item	list of menu items, if menu available
listed_in(type)	Type of meals
listed_in(city)	City of the restaurant

This list includes the following as type of restaurant (in column rest_type)

Casual Dining	Dessert Parlour
Cafe	Pub
Bakery	Bar
Fine Dining	Bhojanalya
Beverage Shop	Sweet Shop
Quick Bites	Mess
Food Truck	Kiosk
Microbrewery	Delivery
Food Court	Confectionery
Lounge	Takeaway
Meat Shop	Irani Café
Dhaba	PopUp
Club	

This analysis will focus on the 4 highlighted restaurant types "Casual Dining", "Fine Dining", "Microbrewery", and "Lounge".

The categorical columns Online_order and book_table are converted using transform to numeric values.

4. Alternative spelling dataset gives the alternative names of locations. This is relevant as the location names in above datasets differ at time and this dataset will help harmonize the names and thereby can be used to merge

dataframes in the analysis step. This dataset contains alternative spellings for 71 locations. The excel file named "correction.xlsx" contains these alternative spellings.

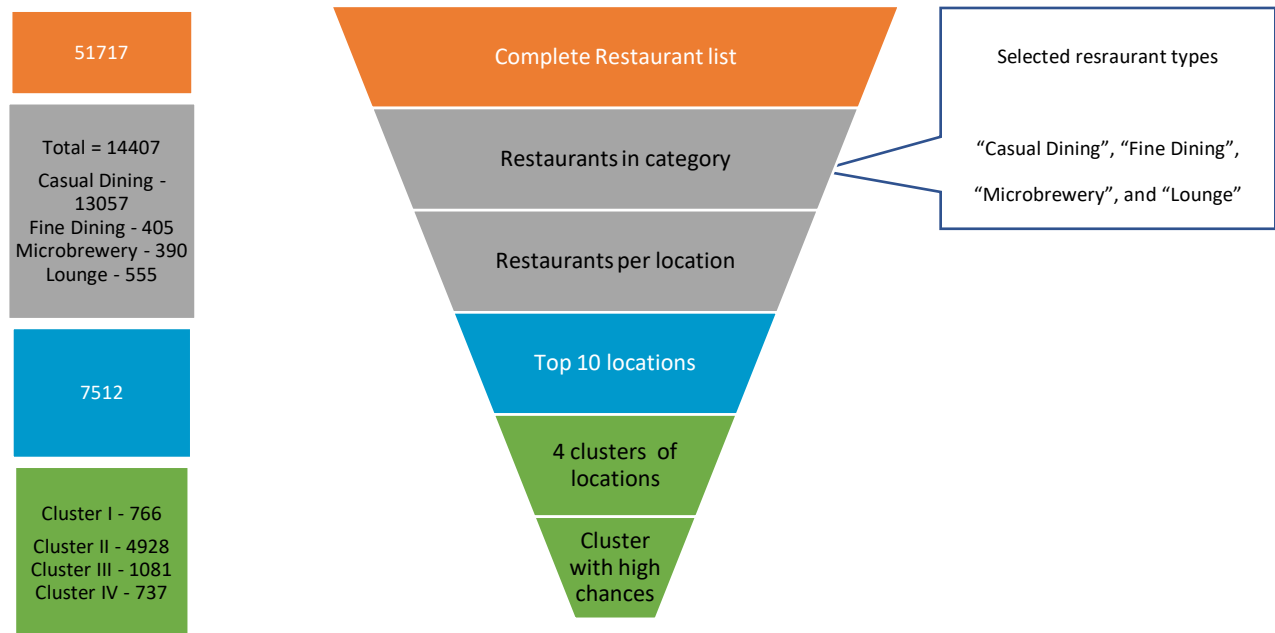
5. Merge datasets to create the following dataframe with following header:

address name online_order book_table rate votes location rest_type dish_liked cuisines approx_cost(for two people) listed_in(city) Country pincode State City

b) PART II: Defining the location(s) of restaurant

The data analysis will be following the following funnel approach.

The Data funnel is as described in figure below:



Counting the number of restaurants of each type per location

This will generate a dataframe with following information:

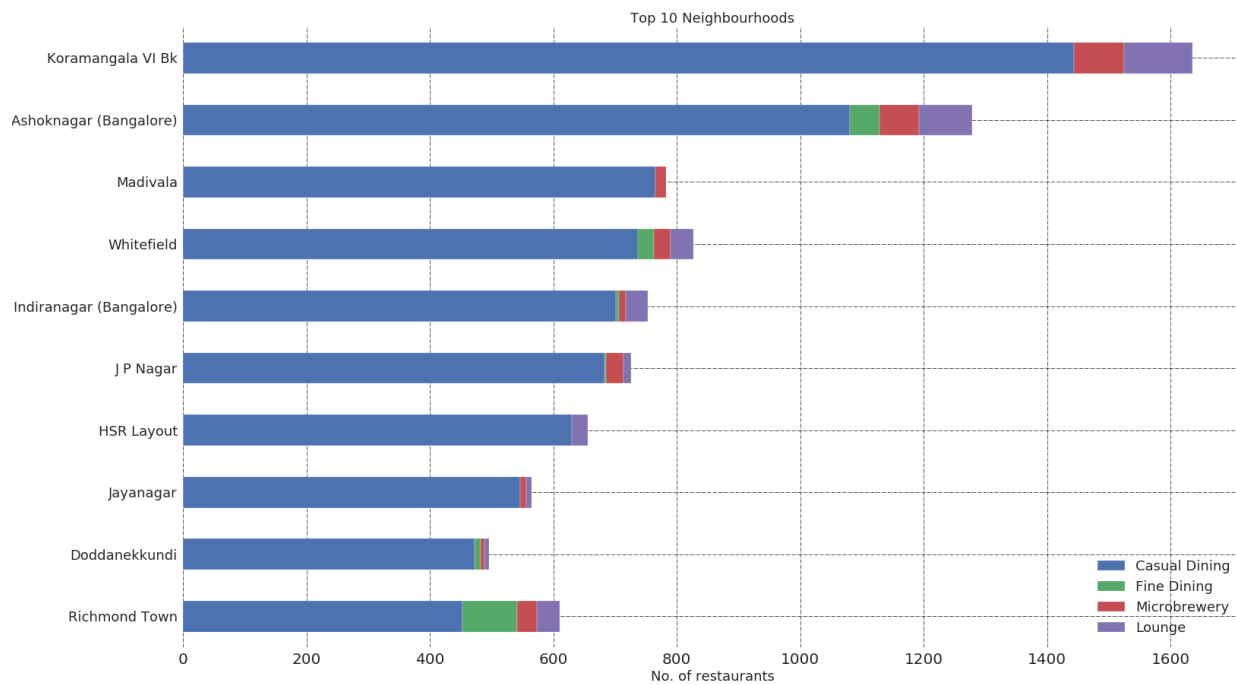
	location	Casual Dining	Fine Dining	Microbrewery	Lounge
0	Adugodi	37	0	0	0
1	Arabic College	76	0	0	0
2	Ashoknagar (Bangalore)	1081	47	65	86
3	Banashankari	170	0	0	4
4	Banawadi	285	0	0	0

Restaurant Type	Total Number of restaurants
Casual Dining	13057
Fine Dining	405
Microbrewery	390
Lounge	555

This list is then sorted descending based on the number of Casual dining restaurants per location. The Top 10 locations are selected. Which summarizes to the following:

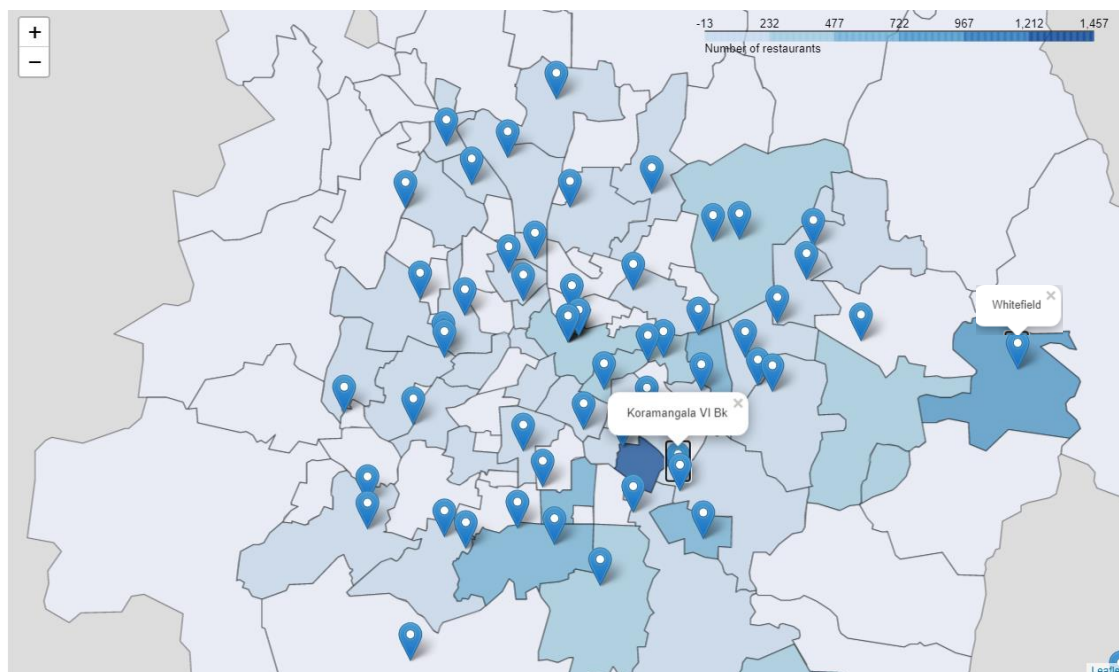
Index	location	Casual Dining	Fine Dining	Microbrewery	Lounge
30	Koramangala VI Bk	1443	0	82	110
2	Ashoknagar (Bangalore)	1081	47	65	86
33	Madivala	766	0	17	0
56	Whitefield	737	26	27	37
21	Indiranagar (Bangalore)	701	5	12	35
22	J P Nagar	683	3	28	12
19	HSR Layout	630	0	0	26
24	Jayanagar	546	0	10	9
15	Doddanekkundi	472	11	5	8
45	Richmond Town	453	88	33	36

Which can be represented by the bar plot below:



This gives a clear indication that the locality "Koramangala VI Block" is the hotspot for restaurants followed by Ashoknagar, the later also offering Fine dining options. The plot also informs us that the locality "Koramangala VI Block" also houses the greatest number of the Microbrewery according to total per location.

We now create a Choropleth of the restaurant density in the city of Bangalore. Which would suggest if we have a decentralized distribution or a more central distribution of restaurants in the city.



Map source : <https://github.com/openbangalore/bangalore/tree/master/bangalore/GIS>

The Choropleth shows us that the restaurants are not distributed only in centre of Bangalore but then there are pockets of locations with restaurant concentration. E.g. "Whitefield" also has above average concentration of restaurant while being an outskirt location in Bangalore city.

To further study the Top 10 locations it would be meaningful to study the top 100 venues of each location and then narrow down to top 10 venues per location.

We utilize the Foursquare API to collect the top 100 location of each location. The output gives us a dataframe with following venue categories (85 different):

Accessories Store	Andhra Restaurant	Asian Restaurant
BBQ Joint	Bakery	Bar
Beer Garden	Bengali Restaurant	Bistro
Bookstore	Boutique	Breakfast Spot
Brewery	Burger Joint	Café
Candy Store	Capitol Building	Chettinad Restaurant
Chinese Restaurant	Clothing Store	Cocktail Bar
Coffee Shop	Cupcake Shop	Department Store
Dessert Shop	Diner	Donut Shop
Eastern European Restaurant	Electronics Store	Farmers Market
Fast Food Restaurant	Flower Shop	Food Stand
Fried Chicken Joint	Garden	Gastropub
General Entertainment	Grocery Store	Gym
Gym / Fitness Center	Health Food Store	Hockey Arena
Hotel	Hotel Bar	Hotel Pool
Hyderabadi Restaurant	Ice Cream Shop	Indian Restaurant
Indie Movie Theater	Italian Restaurant	Juice Bar
Kerala Restaurant	Liquor Store	Lounge
Market	Mediterranean Restaurant	Mexican Restaurant
Middle Eastern Restaurant	Mobile Phone Shop	Motorcycle Shop
Movie Theater	Multiplex	North Indian Restaurant
Park	Performing Arts Venue	Pharmacy
Pizza Place	Pub	Punjabi Restaurant
Restaurant	Sandwich Place	Seafood Restaurant
Shopping Mall	Shopping Plaza	Snack Place
Soccer Field	South Indian Restaurant	Sports Bar
Steakhouse	Swiss Restaurant	Tea Room
Turkish Restaurant	Vegetarian / Vegan Restaurant	Women's Store
Yoga Studio		

Since the number of categories are so diverse, the top 10 venue categories are summarized per location.

The top 10 venues in the top 10 locations are as below:

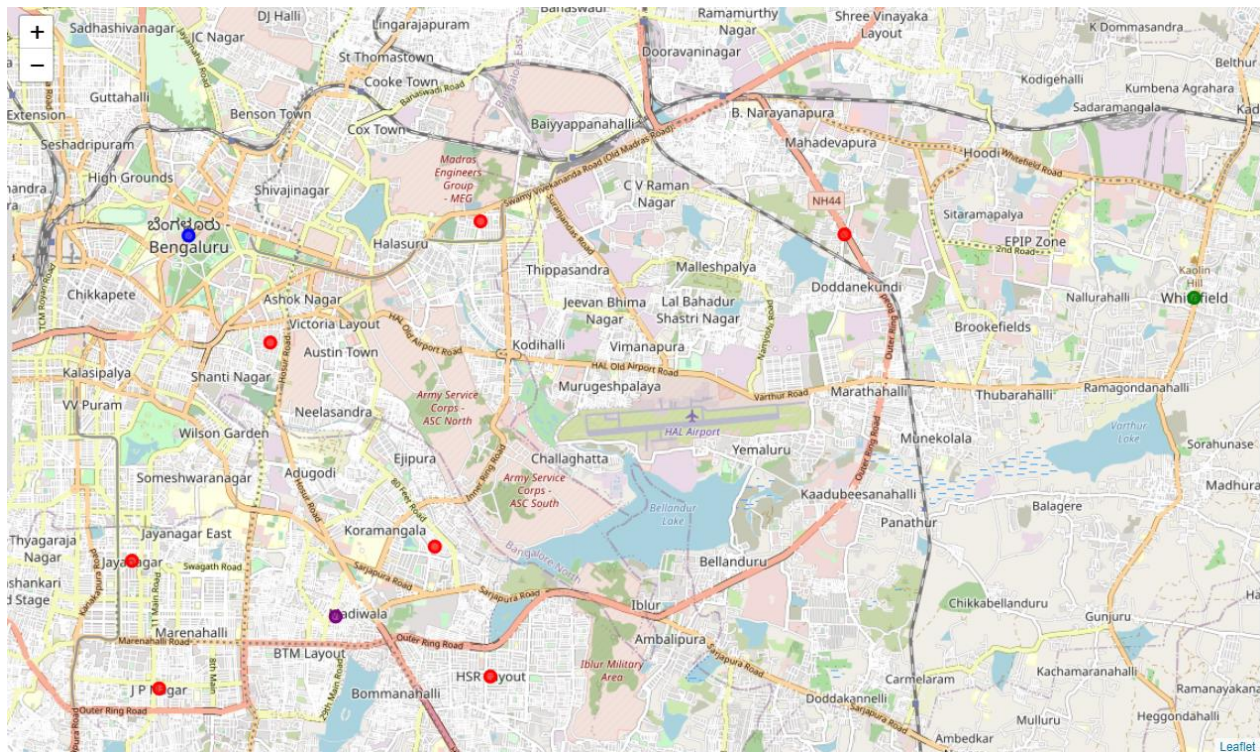
Sl. No.	location	Most common venue									
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
0	Ashoknagar (Bangalore)	Park	Capitol Building	Garden	Sports Bar	Res-tau-rant	Yoga Studio	Electronics Store	Cupcake Shop	Department Store	Dessert Shop
1	Doddanekkundi	Coffee Shop	Indian Res-tau-rant	Multiplex	Fast Food Res-tau-rant	Clothing Store	Res-tau-rant	Sandwich Place	Motorcycle Shop	Movie Theater	Sports Bar
2	HSR Layout	Indian Res-tau-rant	Café	Liquor Store	Coffee Shop	Farmers Market	Snack Place	Ice Cream Shop	Punjabi Res-tau-rant	Mediterranean Res-tau-rant	Chettinad Res-tau-rant
3	Indiranagar (Bangalore)	Indian Res-tau-rant	Café	Fast Food Res-tau-rant	Vegetarian / Vegan Res-tau-rant	Pub	Andhra Res-tau-rant	Chinese Res-tau-rant	Pizza Place	Department Store	Bar
4	J P Nagar	Indian Res-tau-rant	Fast Food Res-tau-rant	Snack Place	Chinese Res-tau-rant	Hotel	Diner	Performing Arts Venue	Electronics Store	Café	Brewery
5	Jayanagar	Indian Res-tau-rant	Café	Chinese	Juice Bar	Fast Food	Sandwich Place	Res-tau-rant	Gym / Fitness Center	Women's Store	Flower Shop

				Res- taur- ant		Res- taur- ant						
6	Koramangala VI Bk	Indian Res- taur- ant	Café	Italian Res- taur- ant	Bakery	Ice Cream Shop	Pizza Place	Juice Bar	Coffee Shop	Bar	Brew- ery	
7	Madivala	Indian Res- taur- ant	Pizza Place	Italian Res- taur- ant	Gym / Fitness Center	Tea Room	Park	Indie Movie Theater	Sand- wich Place	Res- taur- ant	Elec- tronics Store	
8	Richmond Town	Indian Res- taur- ant	Bakery	Chi- nese Res- taur- ant	Hotel	Coffee Shop	BBQ Joint	Middle East- ern Res- taur- ant	Mobile Phone Shop	Market	Gym	
9	Whitefield	Bakery	Hotel	Gym / Fitness Center	Hotel Bar	Swiss Restau- rant	East- ern Eu- ropean Res- taur- ant	Market	Café	Kerala Res- taur- ant	Farmers Market	

The above table shows us that the Restaurants and breweries are on top 10 list of all locations analysed. This assures that Restaurant business is promising in these locations. Nevertheless, the list does not give us a hint on other common venues in the locations. Hence it would be beneficial to look at the cluster of locations.

We will utilize the machine learning model for clustering known as K-Mean clustering method. This method partitions the locations based on their similarity of characteristics. This will support the location decision to know which locations are similar in their characteristics and thus offer wider range of options for final location.

The locations are clustered into 4 clusters as below, each colour represents one cluster.



Though geographically apart some locations combine into on cluster. The Red cluster is also around another cluster "purple" which is usually not easy to recognize.

The Cluster "Red" offers a wider option for location, this cluster would be worth evaluation to probable rental properties fitting the German restaurant setup.

The "Red" cluster – denoted as Cluster II contains the following locations (7 out of the top 10 locations) and top 10 venues.

Sl. no	location	Most Common Venue									
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
30	Koramangala VI Bk	Indian Restaurant	Café	Italian Restaurant	Bakery	Ice Cream Shop	Pizza Place	Juice Bar	Coffee Shop	Bar	Brewery
21	Indiranagar	Indian Restaurant	Café	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Pub	Andhra Restaurant	Chinese Restaurant	Pizza Place	Department Store	Bar
22	J P Nagar	Indian Restaurant	Fast Food Restaurant	Snack Place	Chinese Restaurant	Hotel	Diner	Performing Arts Venue	Electronics Store	Café	Brewery
19	HSR Layout	Indian Restaurant	Café	Liquor Store	Coffee Shop	Farmers Market	Snack Place	Ice Cream Shop	Punjabi Restaurant	Mediterranean Restaurant	Chettinad Restaurant
24	Jayanagar	Indian Restaurant	Café	Chinese Restaurant	Juice Bar	Fast Food Restaurant	Sandwich Place	Restaurant	Gym / Fitness Center	Women's Store	Flower Shop
15	Doddanekundi	Coffee Shop	Indian Restaurant	Multiplex	Fast Food Restaurant	Clothing Store	Restaurant	Sandwich Place	Motorcycle Shop	Movie Theater	Sports Bar
45	Richmond Town	Indian Restaurant	Bakery	Chinese Restaurant	Hotel	Coffee Shop	BBQ Joint	Middle Eastern Restaurant	Mobile Phone Shop	Market	Gym

CONCLUSION I: Cluster II location above is suitable as location for the restaurant. Other influencing factors need to be further studies like Rent of property, Parking availability, Staff availability etc. These are not part of this initial study and analysis. Further data collection for the before mentioned information is necessary.

c) PART III: Defining the "Average Cost For Two People"

The required columns for cost definition are filtered into a new dataframe as the header is below.

name	approx_cost(for two people)	Cluster Labels	online_order	book_table	rate	votes	location	rest_type
0	Empanada Restaurant	750	1	1	0	4.4/5	1001	Indiranagar, Casual Dining

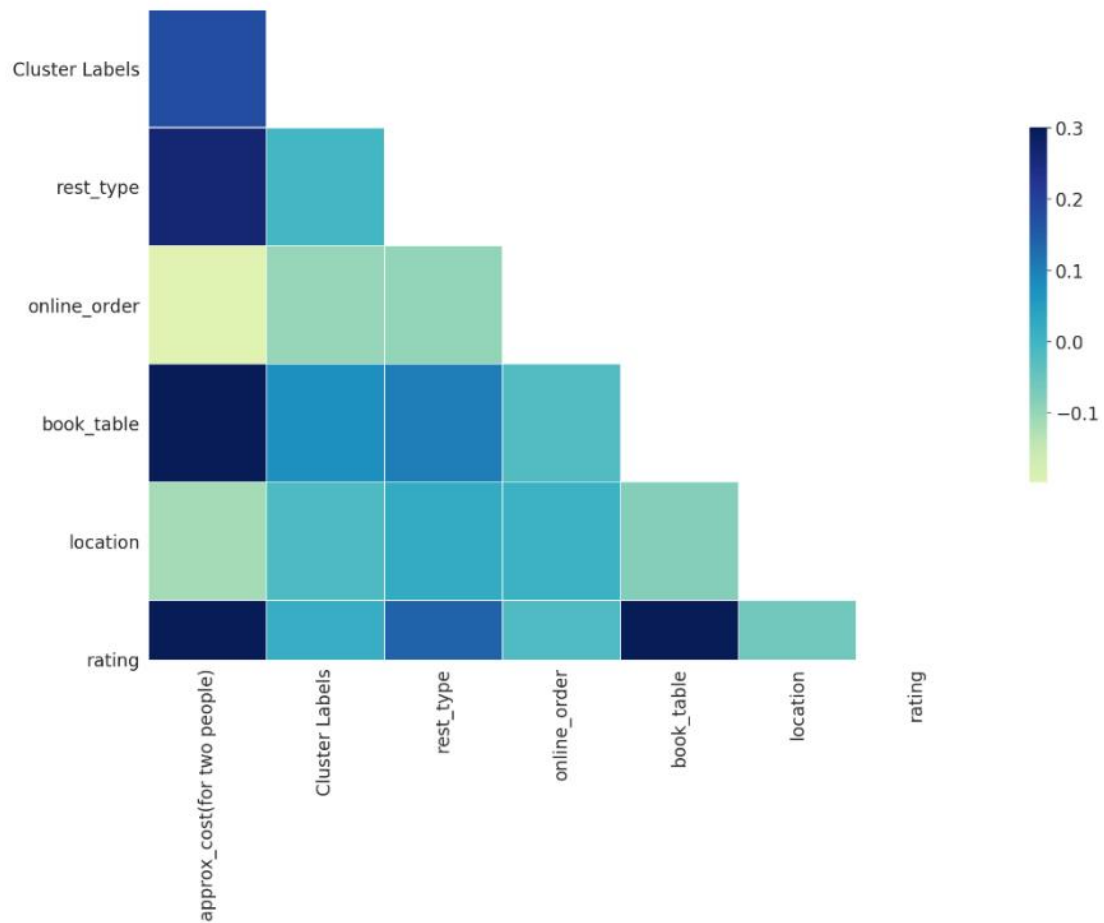
We will have to convert the Categorical values into numeric values to carry out statistical analysis.

Categorical values converted are Location and Restaurant type (denoted by column names location and rest_type)

The Rating is also converted from fraction to a whole number in range of 0-100. E.g 4.4/5 Rating is converted as 88.

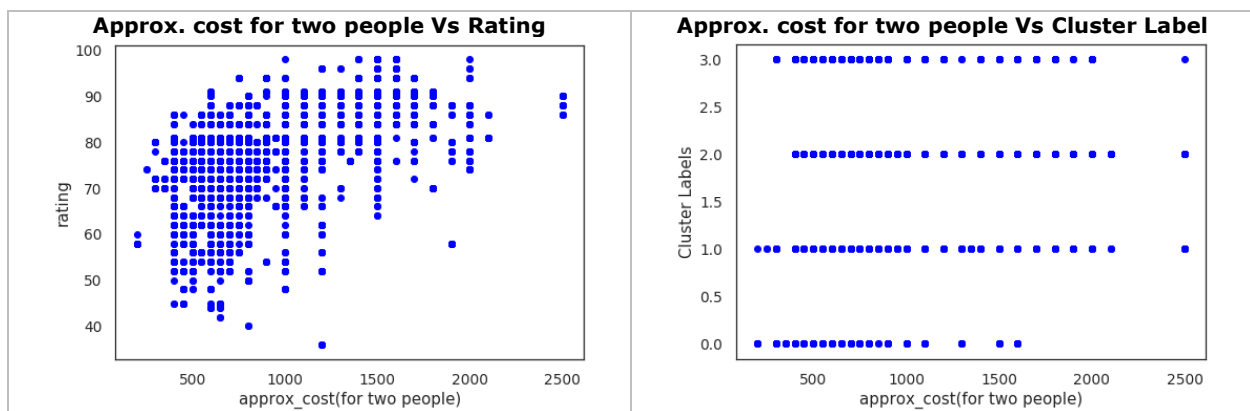
Now the correlation between the features (columns) are evaluated using a correlation matrix.

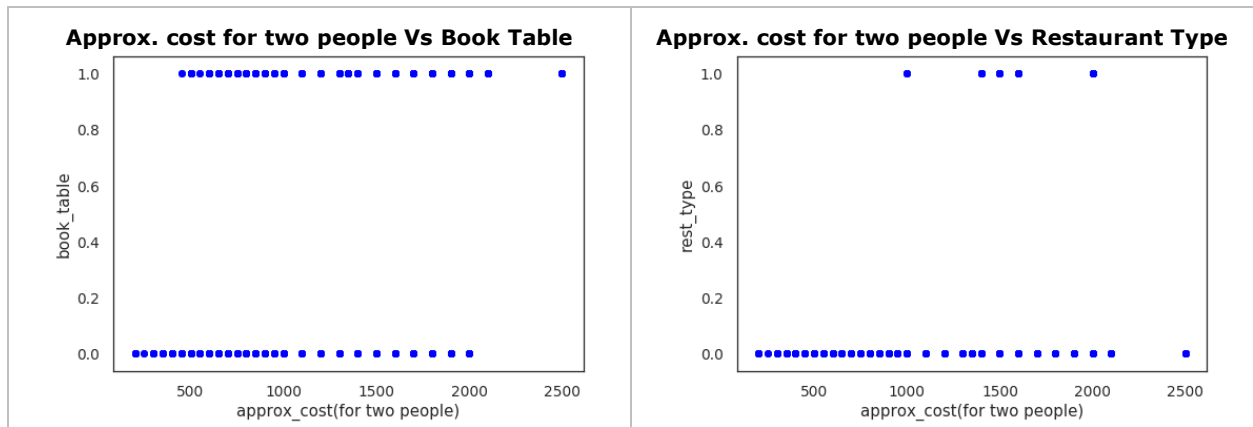
approx_cost(for two people)



This correlation matrix shows that the “**approx. cost for two people**” is related to **Restaurant type** (rest_type), **Table booking** (book_table) and **Rating**(rating) and to some extent Cluster of the restaurant(Cluster Labels). It should also be noted that the highest correlation is around 0.3 which does not show a high correlation in absolute terms.

The correlation of the dependent variable (**approx. cost for two people**) and independent variable is evaluated using scatter plots

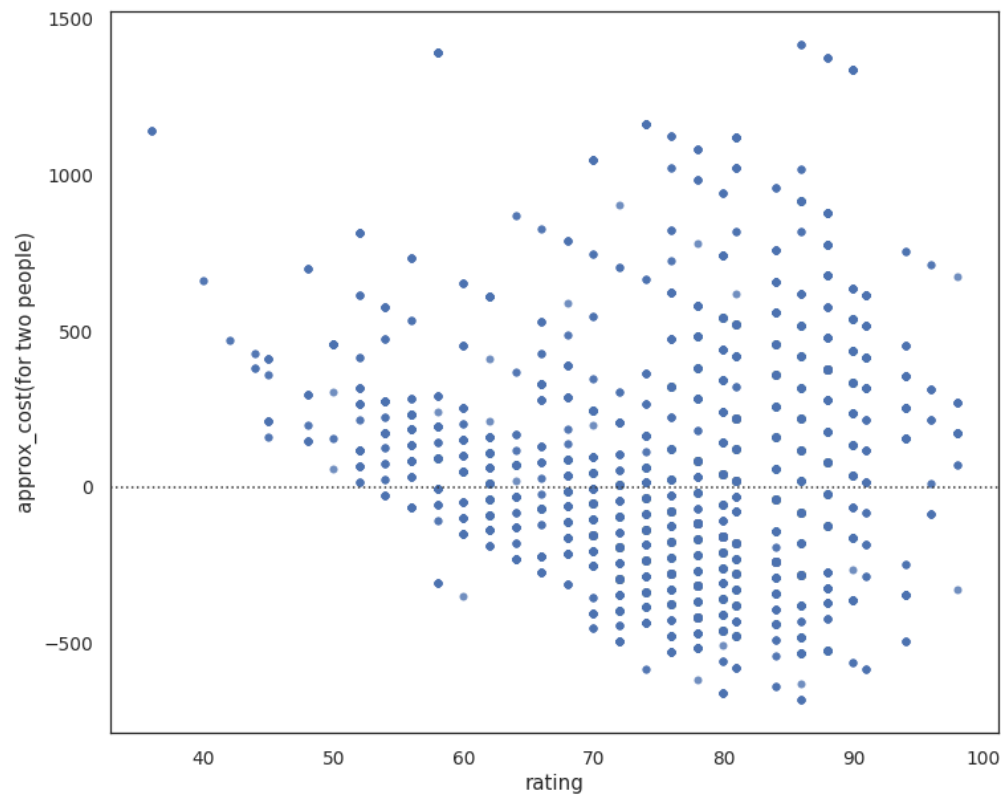




This shows that the Approx. cost for two people and Rating are the only combination which shows some trend of correlation.

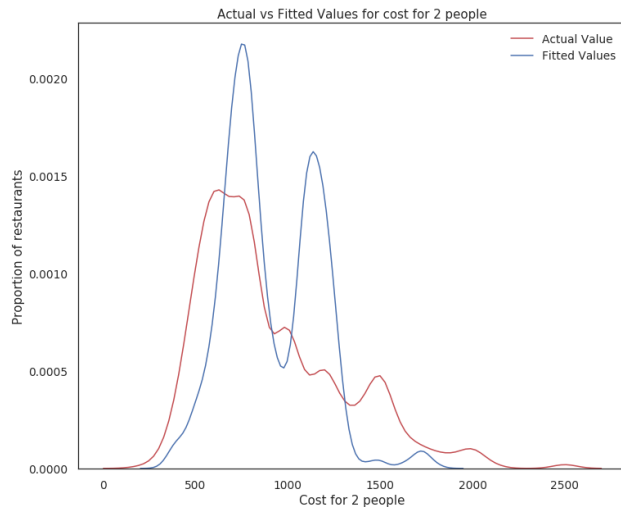
We will thus build different regression models and evaluate its R-square value to select a suitable model for predicting the "approx. cost for two people"

The residual plot of Rating Vs Approx. cost for two people looks as below:



Linear Regression:

The Actual Vs Fitted values using the linear regression is as follows:



The corresponding **R-Square value is 0.24177**, which means only ~25% of the values are explained with this model.

Multi-Linear Regression

Coefficients: [[300.27308264 532.46899228 11.3736346 74.92353334]]

Residual sum of squares: 89750.24

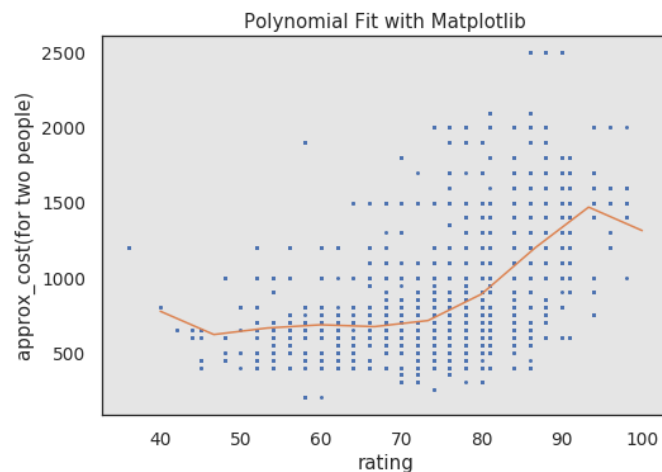
Variance score: 0.41

The **R-Square value is 0.4144**, which is better than the Linear regression but still only explains ~42% of the values.

Polynomial Regression

The polynomial equation of degree 6 is as follows:

$$7.903e+04 - 6592x + 221x^2 - 3.742x^3 + 0.03314x^4 - 0.0001404x^5 + 2.035e-07x^6$$



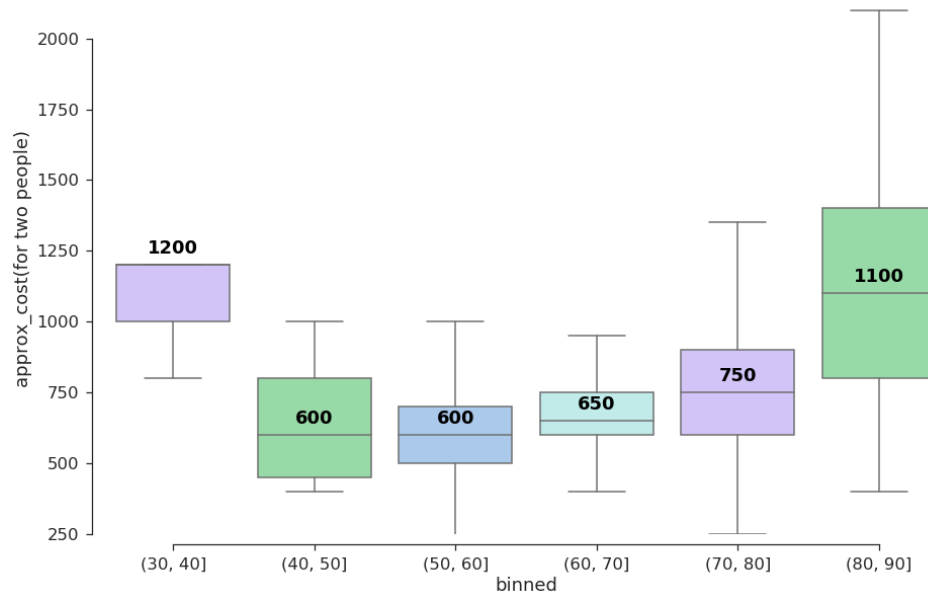
The **R-Square value is 0.3343**, which is better than the Linear regression but still only explains ~33% of the values.

Comparison of Models

Model	R-Square value
Linear Regression	0.24177
Multilinear Regression	0.4144
Polynomial Regression	0.3343

Since all three models are not delivering a above 60% confidence with their R-square values, we will have to conclude that the prediction of the “approx. cost for two people” is not feasible with the available data. Further data collection survey needs to be conducted to arrive at a model.

Hence as a recommendation we calculate the range of “approx. cost for two people” based on “ratings”. This is represented by the box plot below:



This box plot gives us the motivation to run the restaurant and earn a higher rating from the customers to be able to generate higher earnings as people are willing to pay higher for the higher rated restaurants.

4) CONCLUSION

1. Based on the top 10 location clustering and Choropleth the **CLUSTER II with 7 locations** is the most apt for starting the restaurant.
2. Based on R-square values (maximum of 0.41) of Linear, Multilinear and Polynomial regression it is evident that we do not have enough numerical data to predict the “Approx. cost for two people”. This also because the dataset is majorly categorical data.
3. Due to point 2., an average value for “Approx. cost for two people” is recommended based on the rating that the restaurant earns. The box plot also gives the hint on the min-max values for the same.
4. To facilitate a reliable value prediction further data collection (quantitative) should be organized.

5) FUTURE STEPS

a) To define location

Gather additional information as below to finalize the suitable location

- a. Availability of commercial buildings for rent meeting the space and surrounding requirements for the decided restaurant (Parking space, elevator, day light, noise levels, etc.)
- b. Check availability of working staff in the location
- c. Availability of accommodation for staff in the neighbourhood

b) To define the average cost for two people

Conduct a sample survey with diverse target groups to understand the willingness of the customer. The survey can include question to understand the deciding parameter while choosing a restaurant for dining. Survey can also be having open ended question to collect comments from customer towards the cuisine served.

Based on the outcome of the survey including the numeric parameters (number of customer willing to spend, number of willing to try German cuisines ,...) the models can be redone to predict the cost for two people.