C O V E N T R Y
U N I V E R S I T Y

Faculty of Engineering, Environment and Computing
School of Computing, Electronics and Mathematics

7150CEM - Data Science Project - 2122SEPJAN

Analysis of Restaurant Business, rating prediction and sentimental analysis of reviews using Zomato data of Bengaluru
Author: Monish Nallagondalla Srinath

SID: 10969852

Supervisor: Kamal Bentahar

Submitted in partial fulfilment of the requirements for the Degree of Master of Science in Data Science

Academic Year: 2021/22

Declaration of Originality

I declare that this project is all my own work and has not been copied in part or in whole from any other source except where duly acknowledged. As such, all use of previously published work (from books, journals, magazines, internet etc.) has been acknowledged by citation within the main report to an item in the References or Bibliography lists. I also agree that an electronic copy of this project may be stored and used for the purposes of plagiarism prevention and detection.

Statement of copyright

I acknowledge that the copyright of this project report, and any product developed as part of the project, belong to Coventry University. Support, including funding, is available to commercialise products and services developed by staff and students. Any revenue that is generated is split with the inventor/s of the product or service. For further information please see www.coventry.ac.uk/ipr or contact ipr@coventry.ac.uk.

Statement of ethical engagement

I declare that a proposal for this project has been submitted to the Coventry University ethics monitoring website (https://ethics.coventry.ac.uk/) and that the application number is listed below (Note: Projects without an ethical application number will be rejected for marking)

Signed: Monish Nallagondalla Srinath   Date:10/12/2021

Please complete all fields.

First Name:	Monish
Last Name:	Nallagondalla Srinath
Student ID number	10969852
Ethics Application Number	P127360
1st Supervisor Name	Kamal Bentahar
2nd Supervisor Name	Mohamed Abdelshafy

This form must be completed, scanned and included with your project submission to Turnitin. Failure to append these declarations may result in your project being rejected for marking.

Abstract

Zomato is a meal delivery service as well as a restaurant aggregator with a global reach. As of 2019, Zomato's meal delivery service serves over 10,000 cities in 24 countries. We want to do exploratory data analysis, geographical analysis, sentiment analysis, and rating prediction using the Zomato dataset for the city of "Bangalore." Bangalore, Karnataka's capital city, is the state's biggest metropolis.

Natural Language Processing (NLP) is used by AI and Machine Learning to better understand human-computer interactions. Sentiment analysis is a prominent approach used in Natural Language Processing (NLP) to discover positive, negative, or neutral opinions based on how individuals write. Using sentiment analysis, you can learn how others feel about something. Sentiment analysis may be used to analyse the restaurant review in the rating.

A restaurant's rating may be predicted using five machine learning methods: lasso, support vector regression, linear regression, decision tree regressor and random forest regression.

<u>Acknowledgements</u>

To begin, I would want to thank my parents for assisting me in enrolling in the course and institution as an overseas student. I'd want to express my gratitude to all of the module teachers for assisting me in grasping the ideas and progressing to the project stage of the course.
I would like to express my gratitude to Kamal Bentahar, the project module supervisor, for guiding me throughout the project, assisting me in producing this report, resolving issues, and providing informative ideas during weekly meetings.

I'd also want to thank Himanshu Poddar for providing the dataset on Kaggle that enabled me to complete this project. And Zomato, from which Himanshu Poddar extracted data for educational reasons.

1. <u>Introduction</u>

Understanding a society starts with its food; the techniques used to collect, produce, clean, prepare, cook, and consume it convey a range of meanings on several levels (Levi Strauss1983:179, Counihan and Van Esterik 1997, Mintz 1996). According to Jack Goody (1982), food literature reveals structures of production and consumption, social structure and religious cosmological meaning, as well as class and hierarchy, and in this sense, food writings are potent expressions of ethnic identity, gender difference, power, desire, abundance, and want, as well as of societal values and expectations (Lupton 1996, Nestle 2002, Mintz 1996, Schlosser 2001).

Food has been and continues to be a marker of identity in India in terms of caste (Dumont 1980; Khare 1976); class (Breckenridge 1986; Marriott and Inden 1977; Yalman 1969); religious (Yalman 1969) and increasingly secular (coworker and friend) group identification (Yalman 1969). (Khare 1976,Appadurai 1981,Curtin and Heidke1992, Elias 1978). In other words, the adage "you are what you eat" holds true in Indian culture.

Bangalore is the capital city of the state of Karnataka. Bangalore has grown tremendously as a result of its heavy focus on the development of software and other high-level technological skills since India's economic liberalisation in the 1990s. Bangalore's software exports have grown from $50 million to $6 billion in less than a decade, accounting for 11.5 percent of India's total export revenuesiv. Bangalore's population grew from 3.4 million in 1985 to 5.5 million in 2000v and is predicted to reach 9 million in 2011 as a result of the inflow of professionals such as engineers (Heitzman 2004, quoting BDA 2000 statistics). Numerous Indian cities are becoming more integrated into the global economy, which makes Bangalore unique in certain ways but similar to any other city in the nation in others (Appadurai 1996, Wilk 1999).

Pankaj Chaddah and Deepinder Goyal founded Zomato in 2008 as an aggregator of Indian restaurants and meal delivery service. Zomato is a restaurant review site that also provides meal delivery from partner restaurants in a limited number of cities, in addition to menus and user reviews. As of 2019, the service is available in over 10,000 locations.

Nowadays, it costs a lot of money and effort to start a restaurant in a new location. Investors are necessary to fund a new restaurant business, and they assess a range of aspects before making an investment choice. Numerous aspects must be considered, including the restaurant's location, cuisine, and the general quality of its competition.

To begin, you'll need to choose an item from a menu. Following that, you'll want to choose a location where your cuisine will be in great demand. This gives investors confidence in the restaurant's performance. It's a good idea to visit nearby eateries to confirm that the new institution is not a failure. For those who

are concerned about location, they will need to adjust their menu to match the requirements of that area.

When deciding on the kind of food to offer at your restaurant, location is critical. Restaurants located near parks are great for breakfast and afternoon snacks, since the majority of park visitors arrive in the early morning or late evening. The option for customers to make online orders and book a table at your restaurant is also crucial for the future success of your company.

Similarly, the amount of votes, restaurant type, cuisines, projected cost for two people, and service style all contribute to the grade your company receives from clients. The objective of this study is to look at the factors that impact the establishment of a restaurant, such as the kind of cuisine served, the location, and so on.

This information is beneficial in assisting investors and owners in deciding if the restaurant will be a success or failure. To anticipate a restaurant's rating, the algorithm will collect data on numerical and categorical variables such as "online order," "book table," "voted," "approximate cost(for two people)," "location," "rest type," "cuisines," and "listed in(type)".

This may be accomplished via the use of machine learning techniques such as linear regression and random forest. All necessary data is prepared and analysed prior to a training session during which the model is fine-tuned and trained using real-world data. We explored with the Zomato Bangalore Restaurants database in Bangalore to see if we could make conclusions about how location, cuisine type, and other factors may affect a restaurant's rating in order to aid restaurant owners.

We are more than pleased to share our opinions on the internet. Additionally, leaving a review on a product or service enables others to learn more about it and gauge what others think. A frequent assessment of events, places, items, and other things by a person. When expressing our emotions about a particular item, place, or occurrence, we often resort to writing. the discrepancy between what consumers expect and what they get (Ilieska, 2013) Additionally, the act of leaving an internet review is beneficial since it enables other consumers to learn more about a product or service and to gauge what others think of it. Social networking sites such as Facebook and Twitter, as well as review services such as Zomato, Google My Business, and Yelp, are the most popular ways to communicate one's ideas. Customer feedback on online platforms such as Zomato may have an effect on a product's popularity.

Zomato is a website where users can provide restaurant evaluations, including their thoughts of the cuisine and the whole experience. You can determine a restaurant's customer satisfaction level by reading their Zomato review. Restaurants may see Zomato ratings but may be unaware of whether the opinions are favourable or unfavourable to their business.

Zomato reviews remain in text format and may be rated positively, negatively, or neutrally. Zomato has not conducted a research on how users interact with reviews and the phrases that indicate whether they like or hate something. We can learn more about how Zomato users interact with one another and how purchasers feel as a consequence of the evaluations by analysing the language used in the reviews.

The study focuses on Zomato restaurant reviews in Bangalore and employs a mechanism for analysing user sentiment.

1.1 Background to the Project

Bengaluru is where I was born and where I lived until I moved to the United Kingdom and began my life here nine months ago. Bengaluru is where I completed all of my schooling, college, and even my bachelor's degree, so it has been my home for more than 25 years. Since my youth, I've been a foodie; I used to frequent different places with family and friends. Due to Bengaluru's status as India's IT capital, residents from all parts of the country congregate there, resulting in each restaurant having its distinct menu. As India is a place of many cultures, there are several cultures in the city of Bengaluru, and as a result, I've sampled a variety of various foods each time I visited a new restaurant.

As soon as I saw this dataset on Kaggle, all I could think about was my days in Bengaluru and how much I missed the cuisine there, as well as all my family and friends. This prompted me to use this dataset for my project, which would allow me to delve into the food industry's insights in Bengaluru. I was also curious in how restaurants choose their cuisines and the price ranges of their establishments, among other things.

While working on the project, it brought up memories related with each establishment. This fuelled my desire to learn more about how the restaurant industry operates and which characteristics most accurately predict a restaurant's ranking. Restaurant ratings are the primary factor that a consumer considers when deciding whether or not to visit a restaurant. Everyone uses the internet to check restaurant ratings in today's digital age. In India, we have a fantastic application called Zomato that was founded by several foodies. This application was created to allow users to rate restaurants so that others could see the ratings and visit the restaurants. Generally, each restaurant has its own specialties; through this app, users could share the greatest cuisine they had at a certain restaurant, and those interested in that dish may order and enjoy it. This program's idea quickly gained popularity across the country, and as a result, I began using it on my mobile device. This application expanded so quickly that it now includes a meal delivery option similar to Uber Eats, making it one of the most successful applications in India today.

This software altered how people saw restaurants, and it did the same for me. I used to check the ratings and reviews before visiting a restaurant, in order to avoid wasting money and time at a substandard establishment. and sometimes, I

used to write reviews and score restaurants if I liked or disliked anything about them, which benefited both the restaurant and the consumers that visited.

Ratings and reviews have become critical to the restaurant industry in recent years. It assists restaurant owners and consumers in self-correcting by avoiding spending money and wasting money on poor eateries. My primary objective in this project is to utilise my data science abilities to get access to the ratings by forecasting them and doing sentiment analysis on the restaurant's reviews. Additionally, I gleaned several insights from the given data through numerous visualisations. This study may aid restaurant operators in their operations as well as assist consumers in making restaurant selections. Most importantly, assist me in selecting the first restaurant I will visit upon my return.

1.2 Project Objectives

Restaurants are springing up left and right. Currently, there are over 12,000 eateries in the Bengaluru. With so many eateries in the area. This market hasn't been fully exploited as of yet. Restaurants continue to open on a daily basis as well. Despite this, they are finding it more difficult to stay on par with more established rivals. They all serve the same thing. The city of Bengaluru is India's IT capital. The majority of residents in this town rely heavily on restaurant cuisine since they don't have the time to prepare. Since there is so much demand for restaurants, it is necessary to examine the demographics of a certain area. In a given area, what foods are the most widely consumed. Vegetarian cuisine seems to be popular in this area. Are there a lot of vegetarians in the area, such Brahmins, Jain, Marwaris, or Gujarat's? If so, what is the population of the area? It is possible to do these kinds of analyses utilising the data, such as by examining aspects such as:

- Whether or not to have a theme-based restaurant

- Which neighbourhood of the city has the most restaurants serving those cuisines?
- Foodies looking for the greatest food in the area's neighbourhood

- Is a neighbourhood known for its distinctive cuisine.

For the most part, the main challenges they face are related to growing food and real-estate prices, a lack of high-quality workers, a fragmented supply chain, and over-licensing. As part of this study, demographics and food culture of the area will be examined. One of the most significant benefits of this study is that it will aid new restaurants as they choose where they should open and what kind of food they should provide. It also seeks to identify commonalities amongst Bengaluru neighbourhoods based on their cuisine. Additionally, the project's findings will assist customers in making more informed decisions about where to eat. The initiative will mostly focus on restaurants and gastronomic interests in

order to address the issue. And what considerations must be made if a new restaurant is to be opened.

1. Does the demographics of an area have an impact on the food in an area?

2. In terms of the sort of restaurant that may be found in a neighbourhood, does it matter what type of people live there?

3. Is the restaurant's theme important?

4. Is a chain restaurant more likely to attract consumers than a local establishment?

5. Are there any neighbourhoods that share a same sort of cuisine?

6. Does a neighbourhood have a distinct culinary identity?

7. Assuming two neighbourhoods have a lot in common, does this suggest they are related?

8. Do these locations fall inside the boundaries of this neighbourhood?

9. Which meal is the most popular in a certain area?

10.   10. Is vegetarian cuisine popular in your neighbourhood? Are there a large number of individuals from a specific religious or ethnic group in that area? For example, Brahmins, Jain, Marwaris, and Guajarati's.

To answer these concerns, we'll look at what characteristics could influence a neighbourhood's ability to welcome a new restaurant. Each restaurant's reviews are included in the dataset, which may be used to calculate an overall rating. The reviews are also subjected to a sentimental analysis.

1.3 Overview of This Report

To begin, the introduction to the project is created, which includes an overview of all the themes covered in the project. Bengaluru is described in full, as is the data, which includes Zomato. The introduction also contains information on the algorithms, procedures, and processes. The project's purpose is then outlined, along with the research that will be conducted and the additional objectives. The literature review contains information on earlier studies that were discovered in relation to other articles, and what additional work was done in this study will be discussed. The methodology section details the framework used for the project and discusses the methodologies utilised to develop the machine learning algorithms.

The analysis portion of the project is concerned with exploratory data analysis of the data and its properties. This section discusses data visualisation and data inference. The whole design of the project is detailed in detail in the design

section. It is the overarching design framework within which a data science project may be carried out. The implementation stage of a project discusses how the intended project is carried out. The project is implemented in many stages, each of which is detailed in detail in relation to the dataset.

As models are implemented, we get outcomes; these findings are detailed in the testing and results section. The project management section discusses how the project is managed, including scheduling and all other management tasks necessary to complete the project. Critical assessment is the subsequent phase in which I incorporate what I've learned from the project.

The conclusion section discusses the project's accomplishments and conclusions. The future work section, which follows the conclusion section, discusses possible improvements to the project.

Student reflection is where students write on the positive and negative experiences they had while working on the project. The project concludes with a bibliography and list of references.

2. <u>Literature Review</u>

There are numerous restaurants in Bangalore, and Zomato Bangalore offers us with a wealth of information on each one. Additionally, it gives information on the various pricing, restaurant ratings based on cuisine, location of other characteristics, such as location. In order to reduce the quantity of null values in the data, we first clean it up using basic mathematical operations on a few columns. Reading the reviews and ratings is one of the most important aspects of visiting a restaurant.

Thus, we are able to predict our dataset's rating by identifying several components of our dataset that are dependent on different factors.

Analysis relies heavily on classification algorithms, one of the most extensively used and important data mining methods. There are a wide variety of data analysis techniques that involve classification algorithms. Many of them are quite good at classifying things. When analysing the data, we look for patterns that are most closely related to our objective variable and use a variety of visual aids to help us find them. Predicting the rate of individual restaurant based on specified set of factors is then accomplished using prediction models built on top of these co-relation features

Our Data Exploratory process, which involves handling Nan values and null values, eliminating duplicates, and performing other Transformations, would be the first step because to the nature of the real-time data. This column's "Rates" value is what we're after initially. We look at the relationship between Rates and the rest of the dataset's properties. Next, we'll show the connection between our objective variable and all of the other dependent attributes, in order to establish which characteristics are most closely associated to our goal variable. A variety of data modelling structures such as Random Forest, Linear Regression and Decision Tree would then be used to implement the data. We will next be able to identify which model gives us with the most accurate and exact outcomes by using these models.

"Online order, book table, rate, votes, rest kind, average cost, and city" are some of the most closely associated rate column factors used for rating prediction. More than eighty different kinds of food are available in Bengaluru, whereas the most popular restaurant serves up to eight different kinds of food. Data from cuisines was gathered by dividing the cuisines column into eighty columns. In order to increase the accuracy of rating prediction in certain algorithms, the column is read with the values 1 and 0. With this methodology, we were able to boost the number of predictions from seven to a whopping 104 when employing machine learning methods, as opposed to the seven that would have been possible without it.Clustering correct information on the restaurant's food might have a substantial impact on accuracy.

To put it another way, "sentiment analysis" is the study of people's thoughts and emotions regarding a wide range of topics and subjects. This includes anything from products and services to tourism to political concerns to personalities to events to themes (Liu, 2012). Surveys, opinion polls, and focus groups were

used by institutions and corporations when they needed to get feedback from the public or customers. Marketing, public relations, and political campaign businesses have long profited from the collection of public and customer opinion. People and companies are increasingly depending on the textual information included in social media for making decisions as the popularity of social media grows (Liu, 2012:8). Text mining, data mining, and sentiment analysis are all subfields of natural language processing, data mining, and sentiment analysis (Farhadloo & Rolland, 2016). Sentiment Analysis is typically used to detect positive, negative, or neutral opinions based on the patterns of persons writing. It is the goal of Sentiment Analysis to figure out how people feel about a certain topic based on what they say.

3. <u>**Methodology**</u>

First, a strategy was drawn up for the project's implementation and the business issues that needed to be addressed were included in this plan.
Following an agile methodology, a design was created for how to manage, clean and adjust data.
The design portion of this report goes into great detail about how the project was conceptualised.
Jupyter notebook is used to create all of the project's code, including the routines for cleaning data, EDA, feature scaling, and model development. A sentiment analysis model and a rating prediction model will be created.

To conduct model testing, the model is initially divided into two sections: test and train. Sentiment analysis and rating prediction models will be trained and evaluated in a 70:30 ratio. The data is tested using a variety of methods to ensure the greatest possible accuracy.

The most accurate model is chosen for deployment. Since there is only one model for sentimental analysis, the best model with a strong R-squared score will be used in the rating prediction model.

Review, Upon deployment, the models are assessed to see whether the project's goals have been met, and if not, the process is repeated until they are. A model's review stage might help us discover further ways to improve its efficiency by fine-tuning its qualities or discovering ways to streamline certain procedures.

The last stage is the launch procedure, which ensures that the results we have accomplished are preserved. This data science project may benefit from an Agile framework, as these procedures demonstrate..

3.1   Methods of MACHINE LEARNING ALGORITHMS

Linear regression
Linear regression is a method for predicting the relationship between two variables by fitting a linear equation to observed data. There are two types of variables: explanatory and dependent.
For example, a linear regression model may be used to link a person's weight and height. Modelers must determine whether the variables of interest are connected before they can fit a linear model to observed data. This does not necessarily mean that one causes the other, but rather that there is some indication of a strong correlation between the two variables.
It's doubtful that a linear regression model can be applied to data if there are no rising or falling patterns in the scatter plots of the indicated explanatory and dependent variables. Using the correlation coefficient as a numerical measure of the association between two variables is beneficial. The correlation coefficient is a value between -1 and 1 that reflects how strongly the observed data for the two variables are linked together.

Explanatory variables (X) and dependent variables (Y) are represented by the linear regression line equation, which reads, "Y = (a + (bX))." It is symbolised by b (the slope of the line) and a (the value of y when x equals 0).

Random Forest:
A classification method based on ensemble learning, random forests are also known as random choice forests. Under the supervised learning umbrella, Random Forest is a common machine learning approach. In machine learning, it may be used for both classification and regression.
Using several classifiers to tackle a complex problem and improve the model's performance is the premise of ensemble learning. For example, "Random Forest," as its name suggests, uses many decision trees to "comprise varied subsets of a given dataset" to "average" them to improve the dataset's prediction accuracy.
As opposed to using a single decision tree, the random forest gathers the predictions from each tree and anticipates the ultimate result based on the majority vote of predictions. Model accuracy and overfitting are both improved with more trees in the forest. The diagram below shows how the Random Forest algorithm works.

Decision Tree:
Compatibility with supervised learning techniques may be achieved by using decision trees. Classification and regression problems may be solved using Decision Trees, a supervised learning technique. However, categorization is the most common use for them. Each leaf node represents the conclusion in a tree-

structured classifier with internal nodes containing dataset properties, branches representing decision rules.

In a decision tree, there are just two nodes: the Decision Node and the Leaf Node (also known as a Branch). There are two types of nodes in a network: decision nodes, which have many branches, and leaf nodes, which have no branches. Based on the dataset's properties, a decision or test is made. Using this method, you may generate a visual representation of all possible solutions to a given problem/decision under particular constraints.

Decision trees are called such because they begin with a root node and branch out in a tree-like fashion. A tree is constructed using the CART method, which stands for Classification and Regression Tree. This kind of decision tree asks the user a question and then breaks the branch into subbranches based on their answer (yes/no).

Lasso regression
"LASSO" stands for Least Absolute Shrinkage and Selection Operator. Regularizing data models and picking features may be accomplished mathematically using this technique.
Lasso regression is a technique for ensuring uniformity. In order to get a more accurate prediction, it is preferable to use regression methods. This model incorporates shrinkage. It is the process of reducing the number of data points to a single average value that is known as "shrinkage." The lasso method encourages the adoption of simple, sparse models (i.e. models with fewer parameters). A high degree of multicollinearity or the necessity for automation in model selection, such as variable selection or parameter elimination, make this kind of regression a good choice for these models..

Support Vector Regression Algorithm:
Support vector machines include support vector regression as one of its components. That is to say, regression and classification data may be analysed with the use of a concept known as a support vector machine. It is called a support vector machine when used for classification, and a support vector regression when it is used to predict the outcome of a test.

4. <u>Analysis</u>

4.1   EDA   - Exploratory Data Analysis

As a result of the EDA, we have a better understanding of our data. Experimentation with various data analysis approaches (mainly visual) in order to gain more understanding into a dataset, find underlying structure, and draw conclusions is known as exploratory data analysis (EDA).

- identify the most significant variables.
- find outliers and abnormalities in the data
- do a thorough examination of the underlying assumptions
- create models that are economical.
- Determine the best factor parameters for your project.

1. <u>Kendall correlation heatmap:</u>

Kendall correlations are shown in Figure-1 as a heatmap. The Kendall rank correlation coefficient, also known as Kendall's tau coefficient, is used by statisticians to quantify the ordinal association between two measured variables (after the Greek letter). The rank correlation coefficient is a measure of the similarity of data orderings when ordered by each of the variables. Since 1938, it has been referred to as the Maurice Kendall Index, despite the fact that Gustav Fechner proposed a related statistic for use with time series data in 1897.

Similar (or identical, for a correlation of 1) observations in rank (i.e. the relative position label for observations within the variable: first, second, third, etc.) have a high Kendall correlation between the two variables; on the other hand, dissimilar (or completely different, for an absolute Kendall correlation of 1) observations have a low Kendall correlation.

The connection between dish likes and votes is 0.6, which is considered to be important information. While the URL and listed (city) are the most closely connected, this relationship has minimal practical significance.
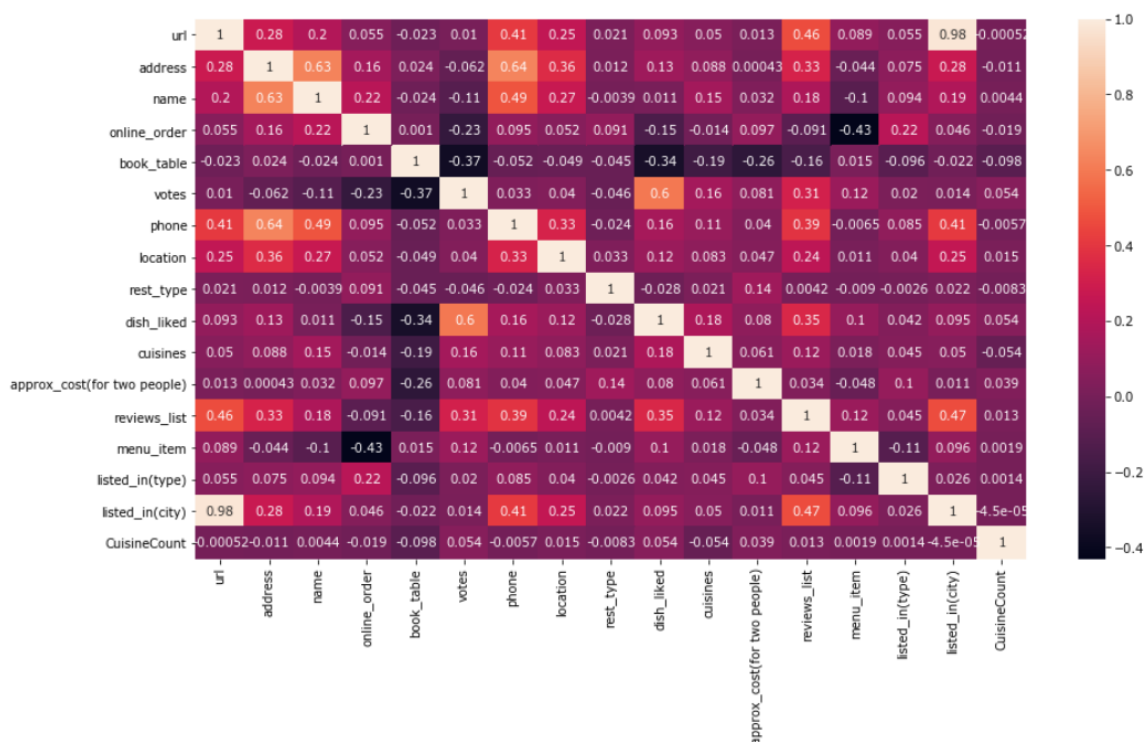
Figure – 1

2. **Famous resturants of Bengaluru :**

In a barplot, the data is represented by rectangular bars with heights or lengths that are proportionate to the values of the categories they cover. Figure 2 depicts the most popular restaurant chains in Bengaluru using the X-axis as the number of restaurants and the Y-axis as the name of the restaurant. Café-coffee day has the most outlets in Bengaluru, followed by Onesta and simply bake, based on a bar graph. Those that are owned by local businesses.

Bigger terms have more weight in a word cloud, which is an unique visual representation of text data. As seen in Figure 3, the restaurant with the most locations is shown as a word cloud with the largest size. Based on the amount of times a restaurant's name appears in the word cloud, the size of the text rises.

Figure – 2

Figure - 3

3. **Table booking** – FIGURES 4 and 5 exhibit the table reservation services provided by the restaurants; since the column table reservation has only two possible values (yes and no data), they are jointly tallied and depicted in two different kinds of charts below.

Figure 3 is a pie chart illustrating that, out of 51717 eateries, roughly 87 percent had a table reservation system.

Figure 5 illustrates that delivery restaurants do not provide table reservations since they simply deliver food to customers. We may assume that most delivery restaurants do not even offer a dining area where customers can sit down and eat. Even dessert restaurants do not need a table reservation system since their clients do not spend a significant amount of time there.

Table booking
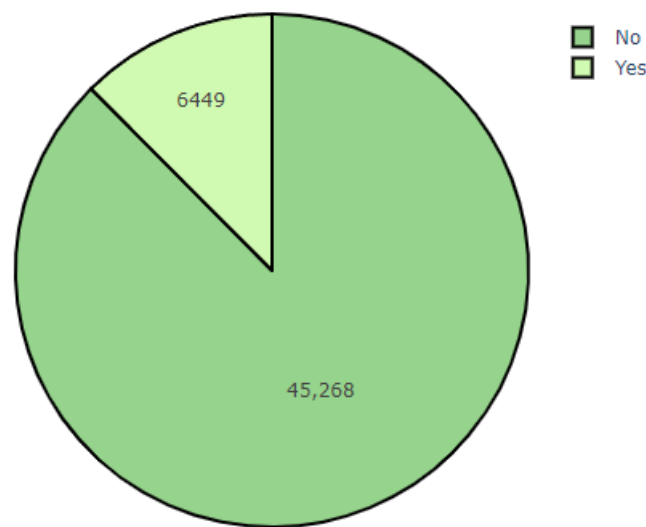
Figure - 4
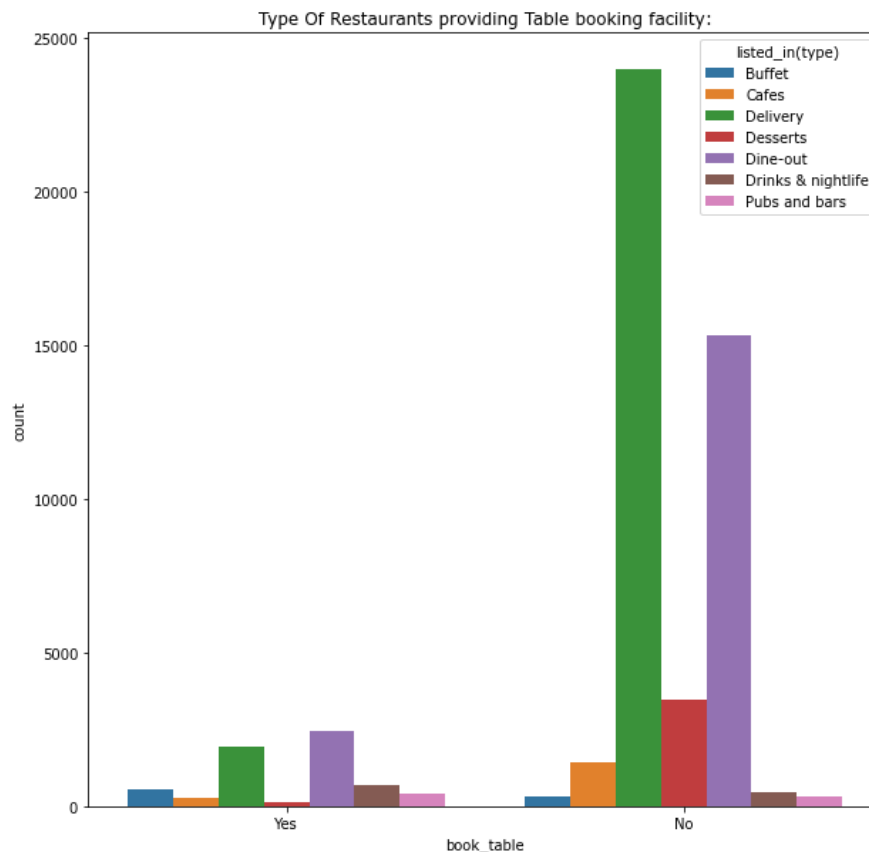
Figure - 5

4.Online order - Online ordering is now a standard part of every restaurant in the digital age, as seen by the large number of establishments that provide this service. This helps to enhance the revenue of the restaurant while also providing decent cuisine during these difficult times. Figure 6 shows that this function is available at more than 30000 restaurants, which is a positive thing. Figures 7 and 8 show that, in addition to delivery restaurants, dessert offering businesses are more likely to deliver via online orders than other types of restaurants.
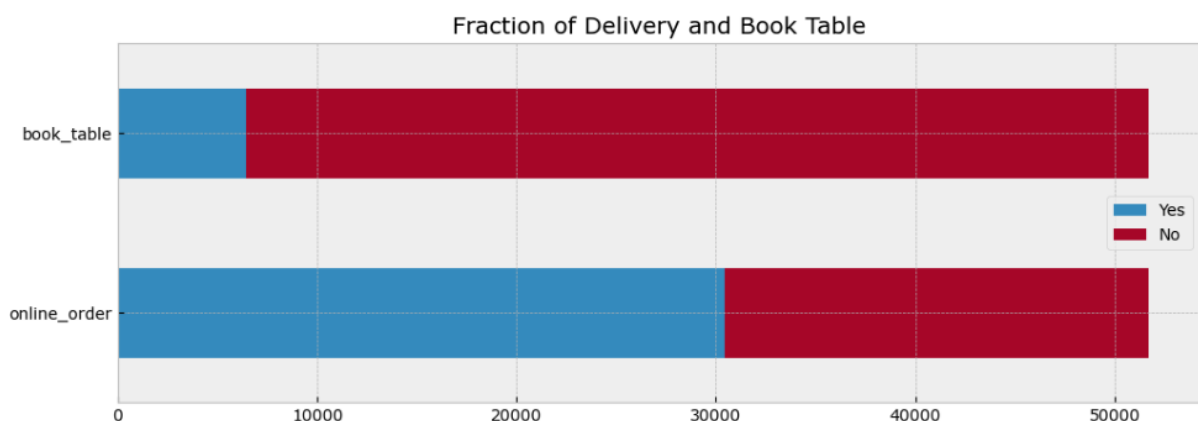
Figure – 6

Accepting vs not accepting online orders
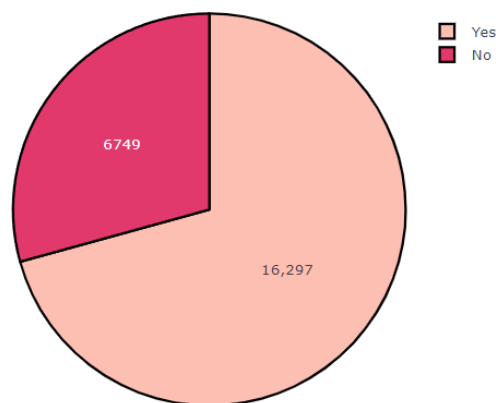
Figure – 7
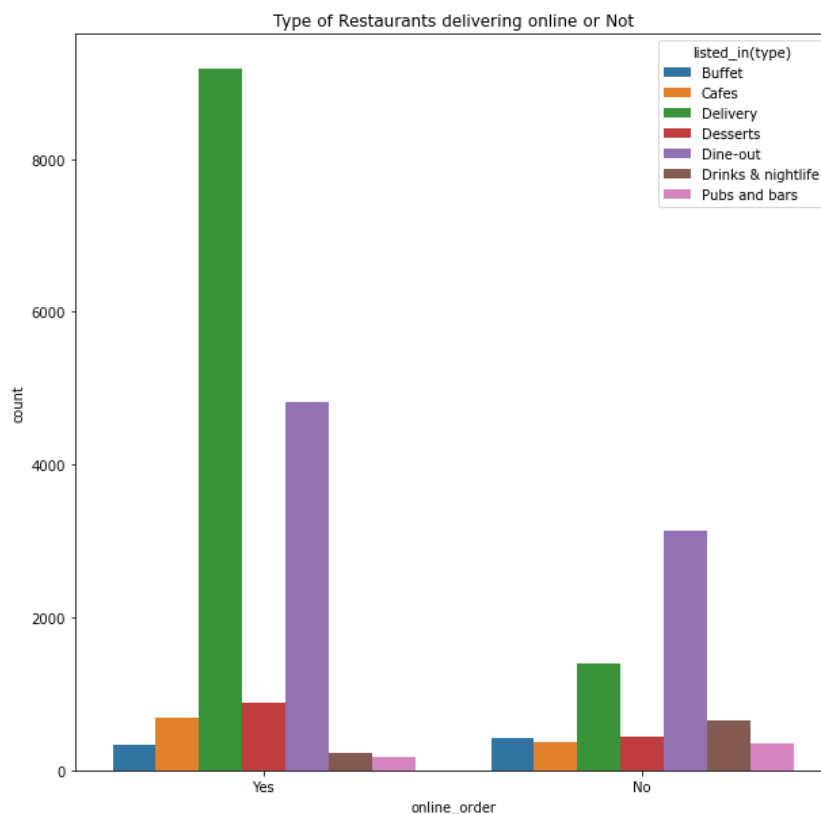
Figure – 8

5.Location – The location of a restaurant is critical to the success of the company since it serves as the foundation of the establishment. Many other characteristics of a restaurant, such as its prices and cuisines, are controlled by the location, which is the most important factor. Figure 9 shows that BTM has the biggest number of restaurants, followed by HSR, but it is not apparent which is which until we look at Figure 10, when it is evident. Figure 9 depicts the count, but figure 10 depicts a pie chart that aids in visualising the data.

BTM and HSR are two locations in Bengaluru that are considered to be among the city's most affluent neighbourhoods.
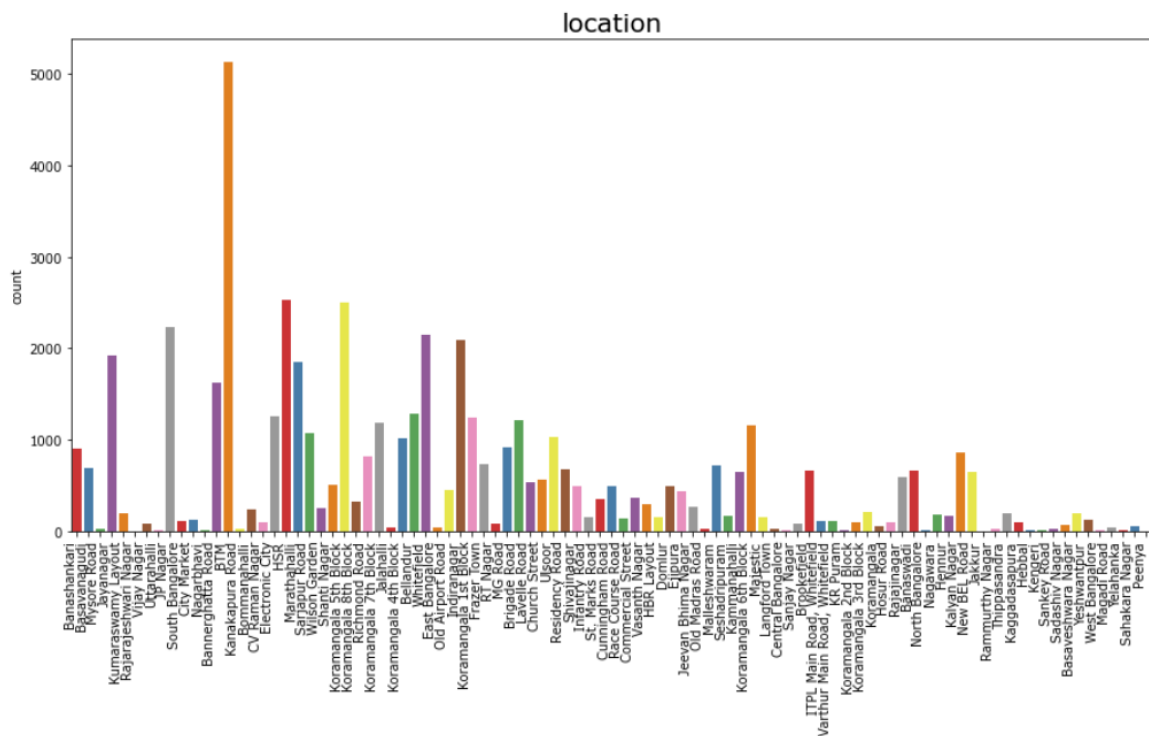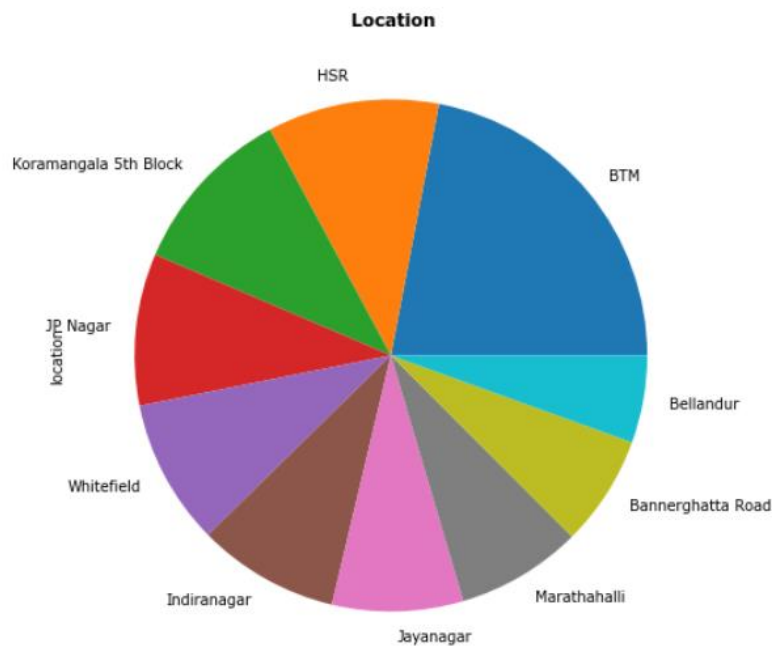

Figure – 9

Location

6.Restaurant type - Figuring out which restaurant category has the greatest number of establishments is shown in figures 11 and 12. As a result of the large number of values in figure 11, it can be a little confusing to look at, but the count is taken into consideration, which aids in visualising the data. By combining the two figures, we can conclude that quick bites restaurants have the highest number of customers, followed by casual dining and cafés. We can tell that consumers choose fast food over other types of meals.

Figure 13 demonstrates the relationship between the kind of restaurant and the foods that are served there. The larger the word in the word cloud, the greater the importance of the term. Because chicken was the most popular dish in all of the restaurants, we may deduce that there were more non-vegetarians in Bengaluru, as well as more eateries providing non-vegetarian fare.
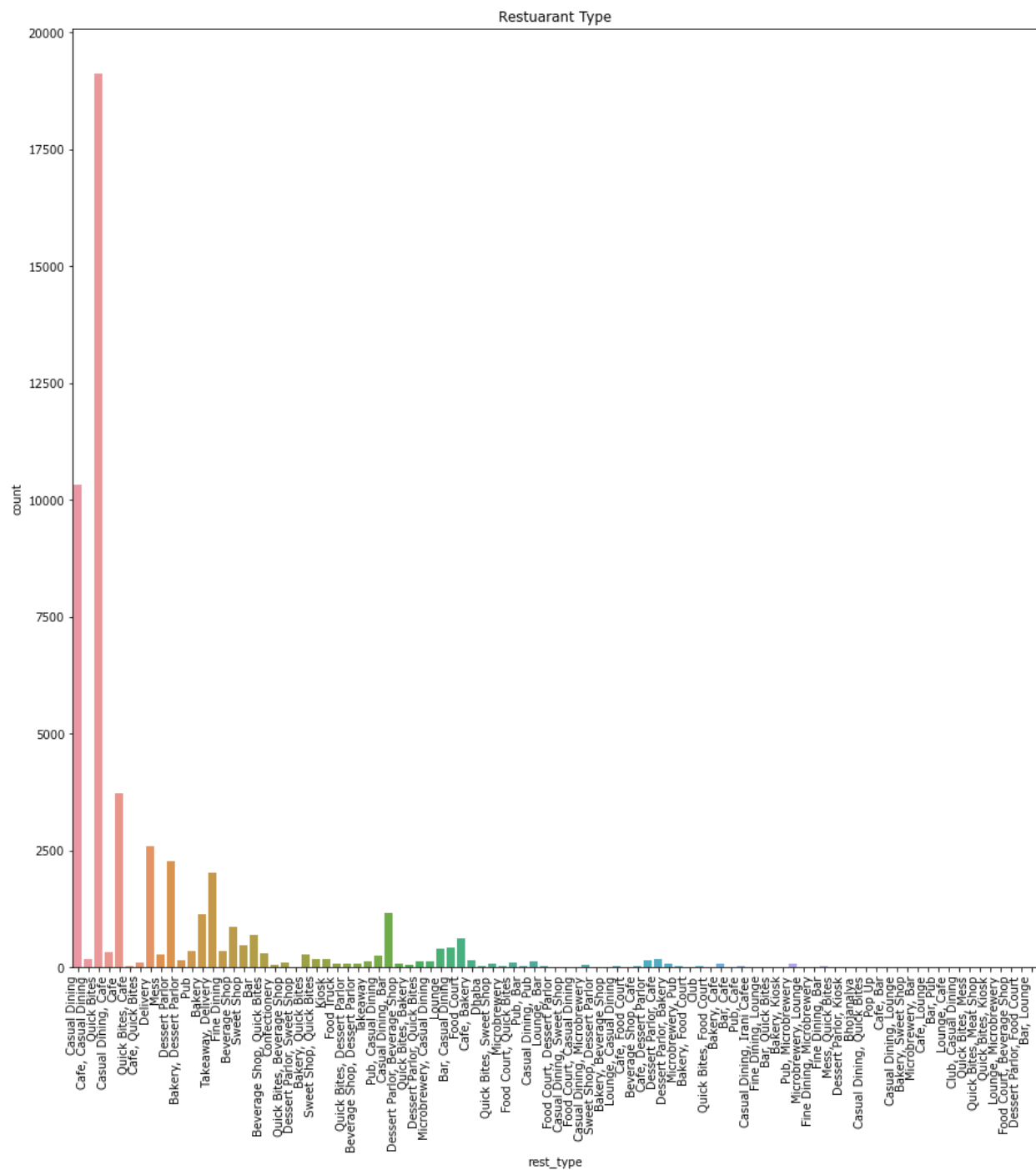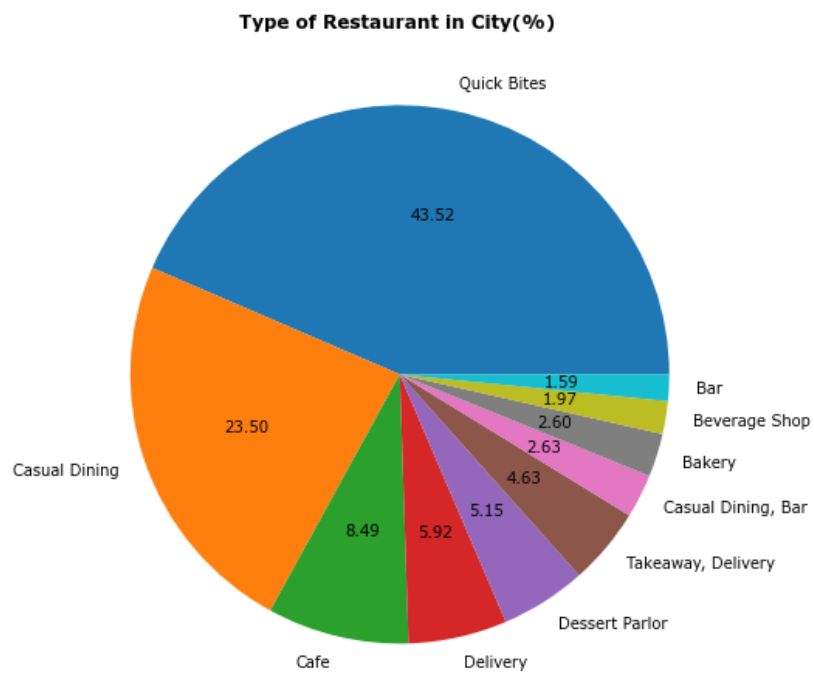
Figure -11

Type of Restaurant in City(%)

Figure – 12

Figure – 13

7.Average cost for two people : The average number of persons at a restaurant is two, making it simple to calculate the cost of the meal. Thus, in order to evaluate the price range of the restaurant, the average cost of two individuals is taken into consideration. In order to represent how many restaurants have the same price range, Figure 14 and Figure 15 are utilised. It turns out that about 20% of restaurants are in the price range of 300 Indian rupees, followed by 400 rupees at approximately 17 percent. We may deduce that there are more eateries that are really inexpensive from this information.
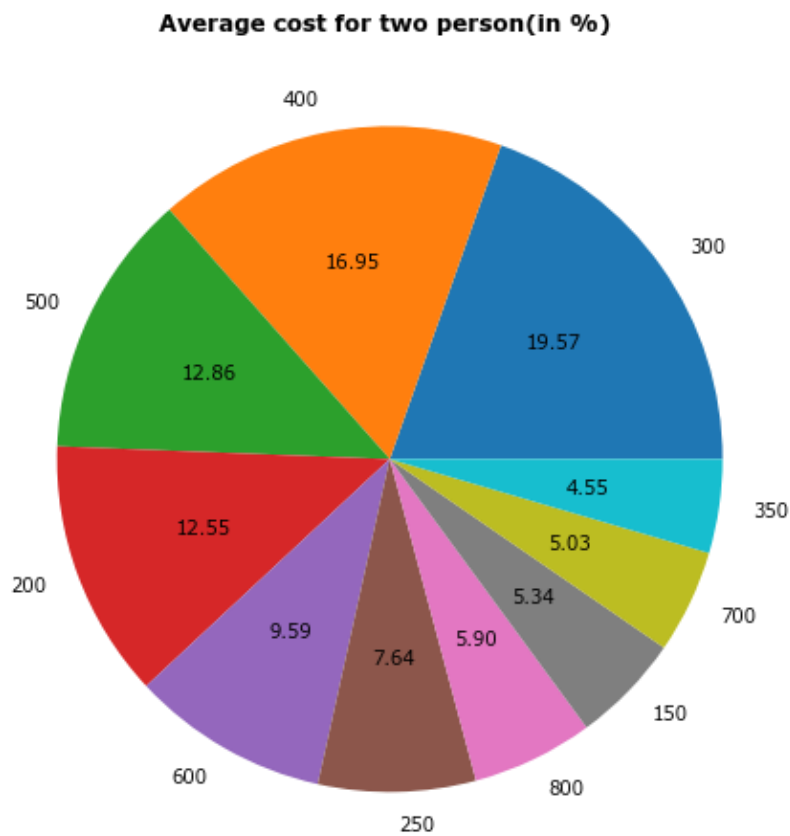
Average cost for two person(in %)
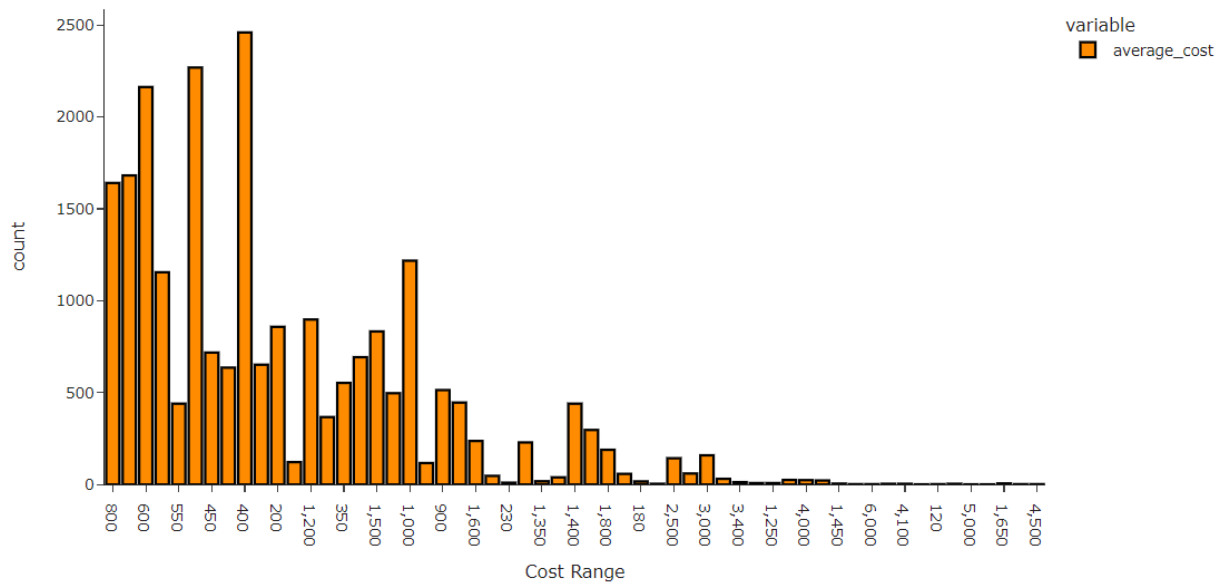

Figure −14

Cost of Restaurants

Figure −15

8.Cuisines: On any given day in Bengaluru, you may find approximately 80 distinct cuisines of being served . The maximum number of cuisines that may be served at a restaurant was eight. Figure 16 depicts the total number of restaurants serving the various cuisines, while figure 17 depicts

the proportion of restaurants serving the various cuisines. We may deduce that, despite the fact that Bengaluru is located in the southern region of India, North Indian Cuisine restaurants are the most numerous, followed by Chinese restaurants. Figure 18 indicates that there are a greater number of restaurants that offer more than one cuisine, with approximately 84 percent of restaurants serving more than one cuisine. This may be done because additional cuisines may attract a larger range of consumers and also provide more alternatives to customers.
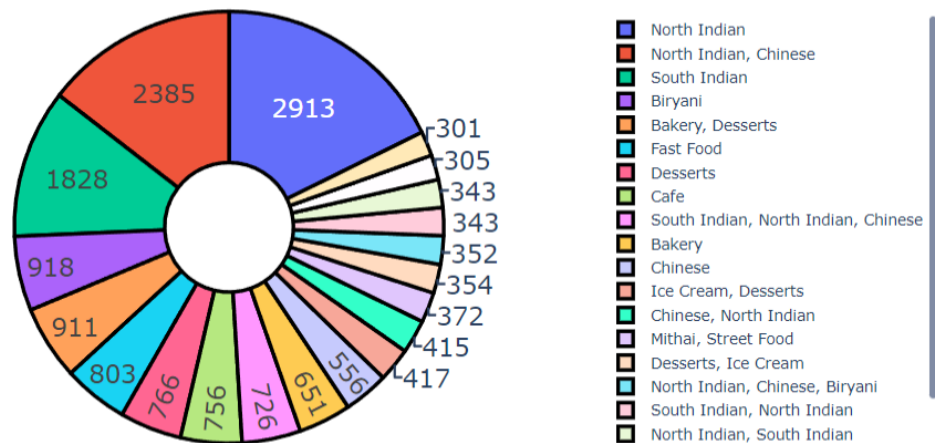

Most popular cuisines of Bangalore

Figure –16

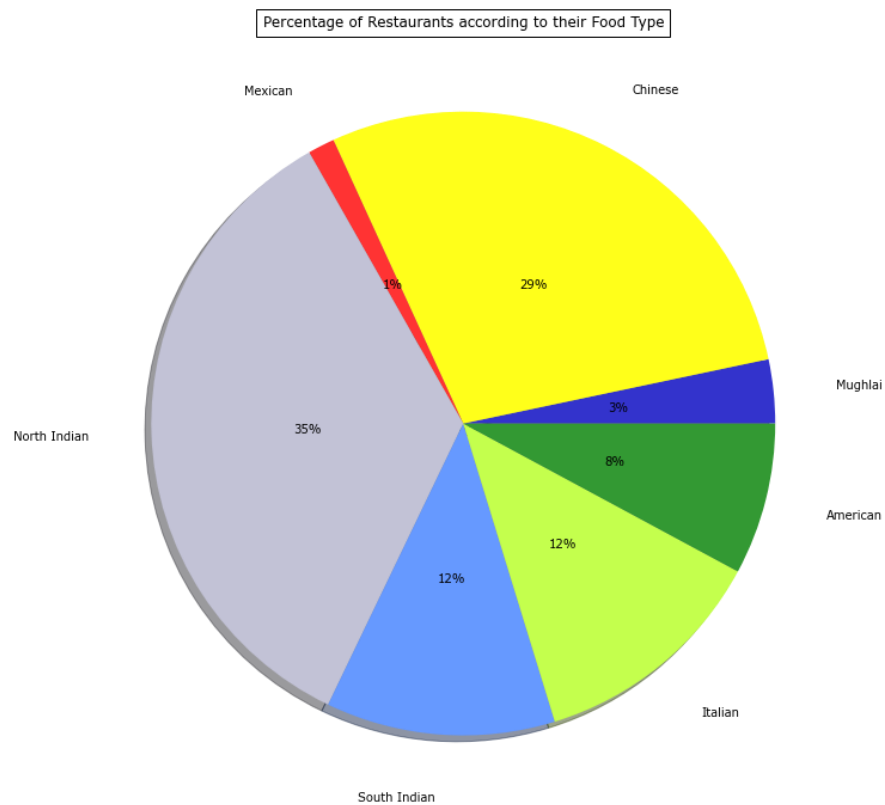Percentage of Restaurants according to their Food Type

Figure – 17

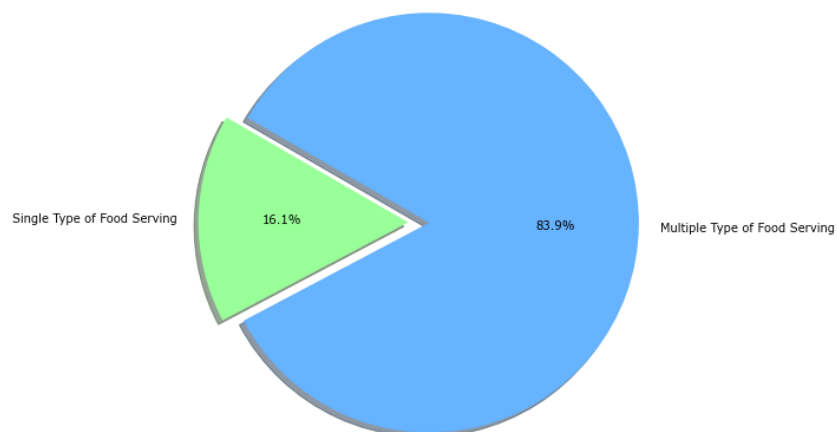Percentage of Restaurants serving single type vs Multiple type of Foods

Figure – 18

9.Rate : Figure 19 demonstrates that the largest number of restaurants have received 3.5 to 4 ratings, with 0 being the lowest possible rating and 5

representing the highest possible rating. This indicates that the overall experience was positive.

The opposite is reflected in figure 20, which shows that 46 percent of restaurants have a rating of 4 to 4.5 stars. Furthermore, in Figure 21, it is illustrated that the largest number of restaurants has a rating of 4. Based on these three data, we can deduce that the majority of the restaurants have a 4 rating, which indicates that they are really excellent.
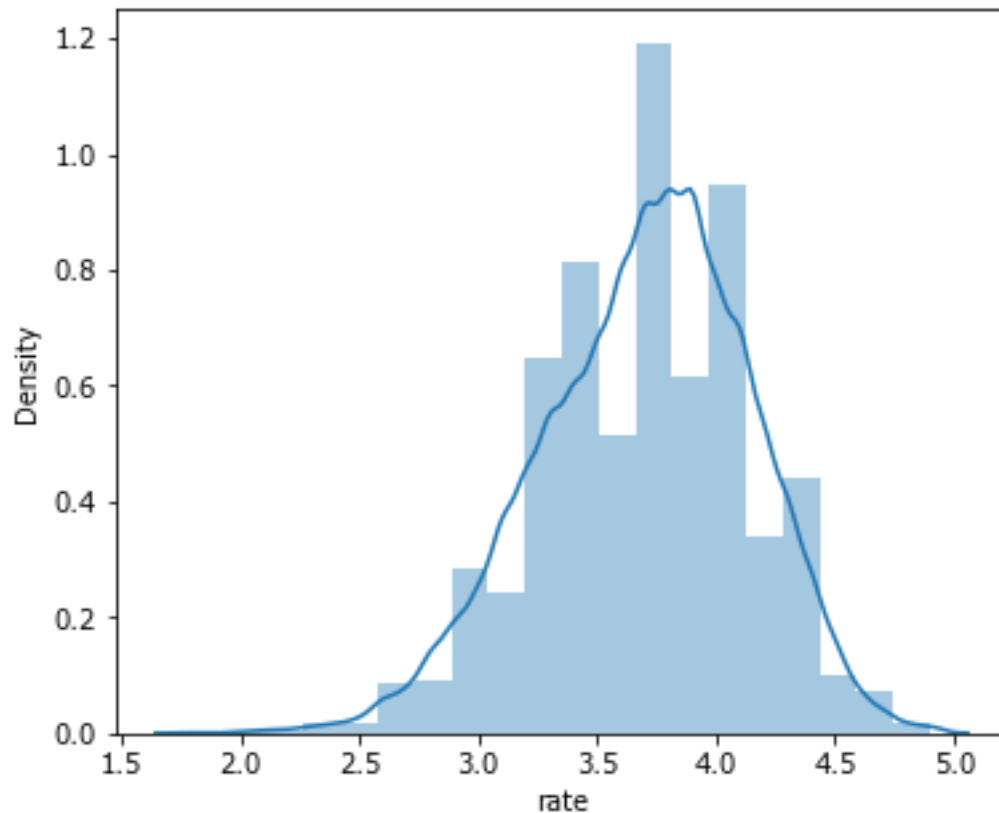

Figure -19

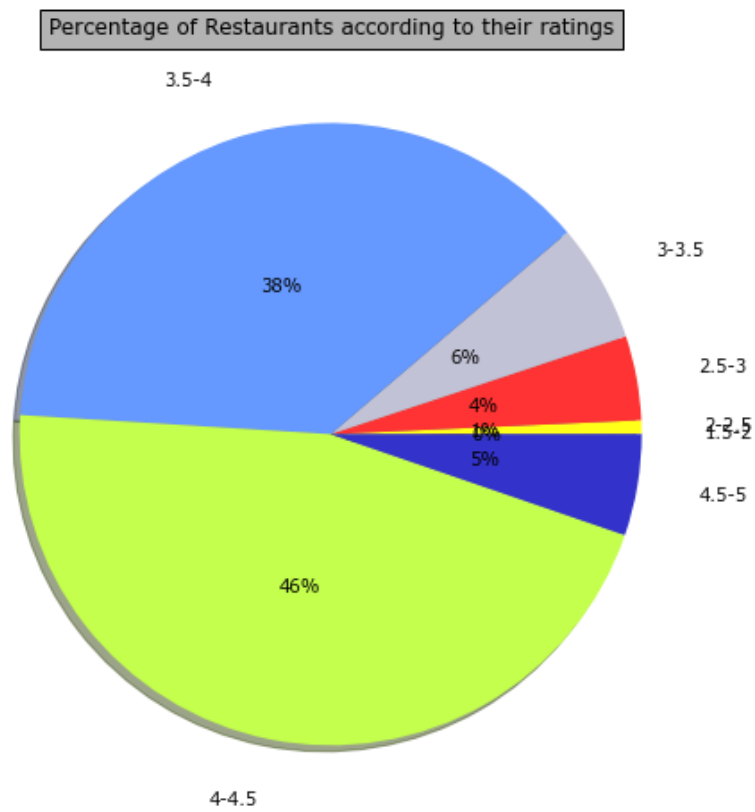Percentage of Restaurants according to their ratings
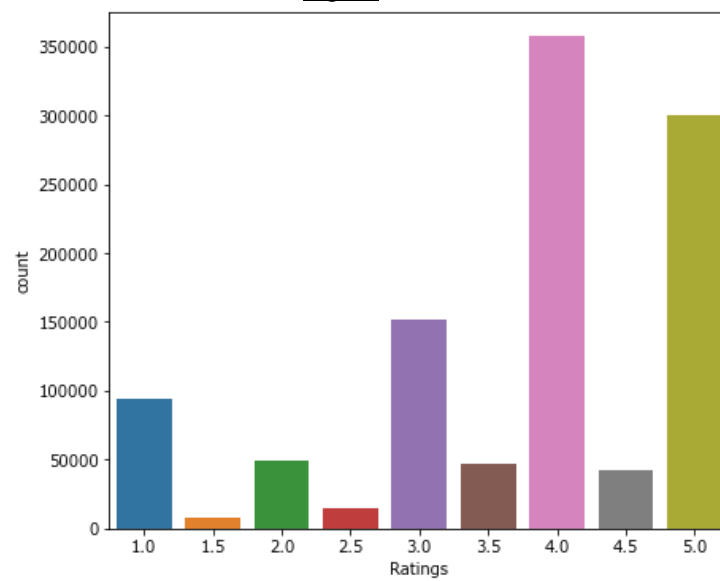
Figure – 20

Figure – 21

10.Reviews : Restaurants get both positive and negative feedback in the form of reviews. Figure 22 depicts a word cloud of reviews for restaurants that have a large number of locations around the city, which has resulted in their being more popular.

Figure – 22

11.Geospatial Analysis

 Geospatial Analysis may be used to locate a certain region's geographic area. The city of Bangalore's geographic region is seen in Figure 9. According to the

map, most of Bangalore's eateries are concentrated in the city's central region. We'll need the latitudes, longitudes, and total number of restaurants on the map if we're going to locate them. Python's built-in functions are used to pinpoint a geographic location on a map. With the aid of Python's "geopy" function, users may locate restaurant locations on maps using latitude and longitude coordinates and heatmaps. All of such eateries are concentrated in Bangalore's central. As we make our way out of Bangalore's core business district, the number of dining options drops precipitously. This is a useful resource for aspiring restaurateurs who are looking for suitable sites for their new enterprise. Using this map, you can see how many restaurants are in a certain area in Bangalore. Bengaluru's basic map is seen in Figure 23. Locations with their respective names may be seen.

Figure 24 shows a heat map with all of the restaurants marked on it based on the dataset's location information. We can confirm this by looking at the location maps shown above and seeing that they are correct. That region has a high concentration of eateries, as seen by Figure 25's heat signatures. Despite the fact that the marker cluster moves quickly.

Figure – 23 : Base map of Bengaluru

Figure – 24 : Heatmap of all restaurants
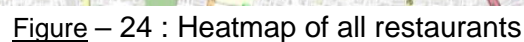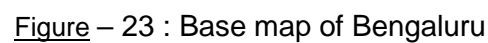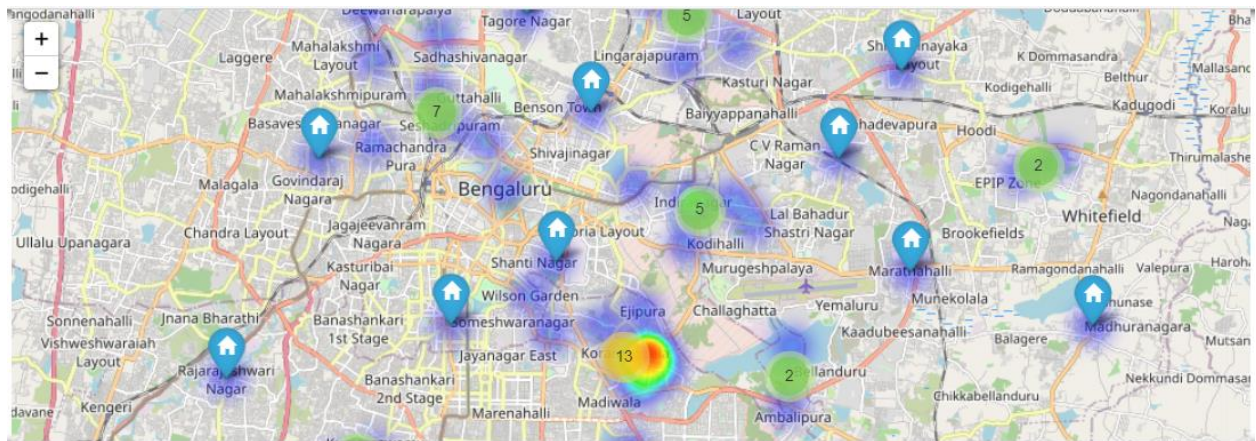
Figure – 25 : Fast marker cluster

North Indian restaurants are shown in Figure 26: the total number. Those who live in Bangalore prefer north Indian cuisine over southern Indian cuisine. Bangalore's core region is shown on the map with North Indian eateries. When compared to the overall number of restaurants in Bangalore, there are more than 15,000 eateries serving North Indian cuisine.

Restaurants serving south Indian cuisine are included in Figure 27. It's clear from the heat maps that there isn't a huge amount of difference between the two, but we can see that various regions like different sorts of food at restaurants.

Figure – 26: heat map of North Indian cuisines :

Figure – 27:Heat map of South Indian cuisines

Sentimental Analysis -

6. **Design**

Business Understanding
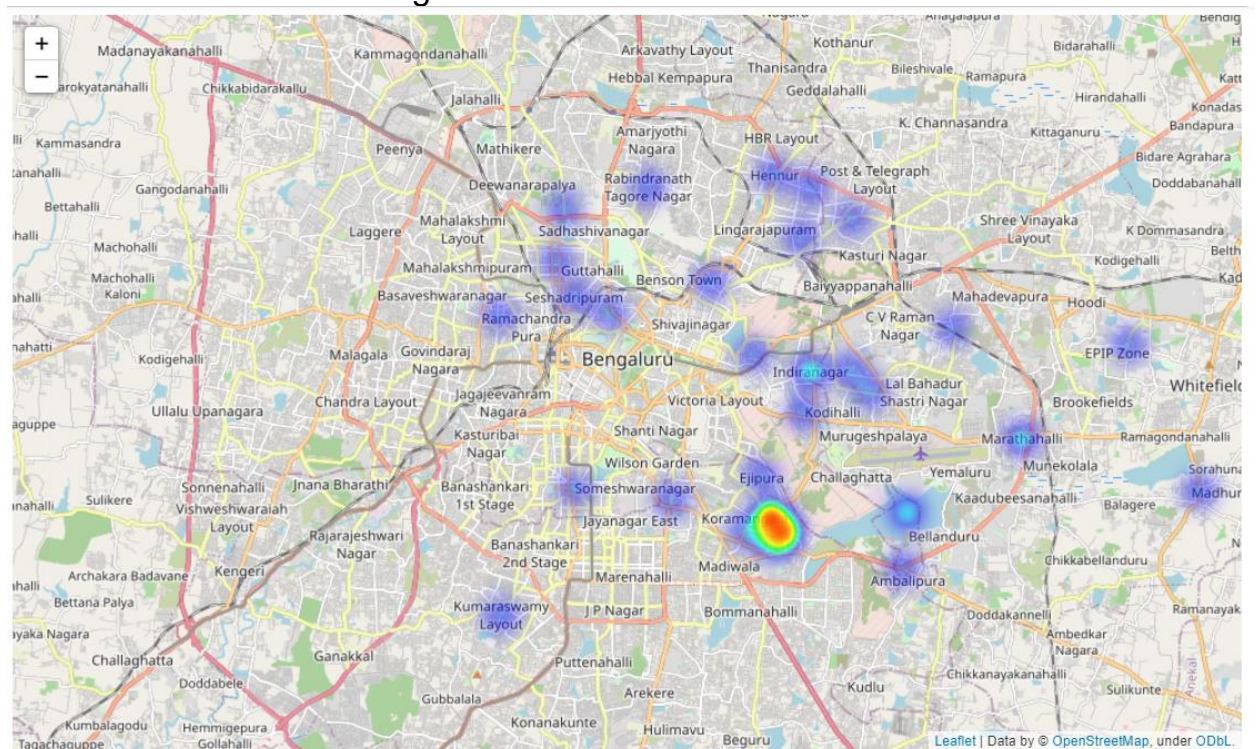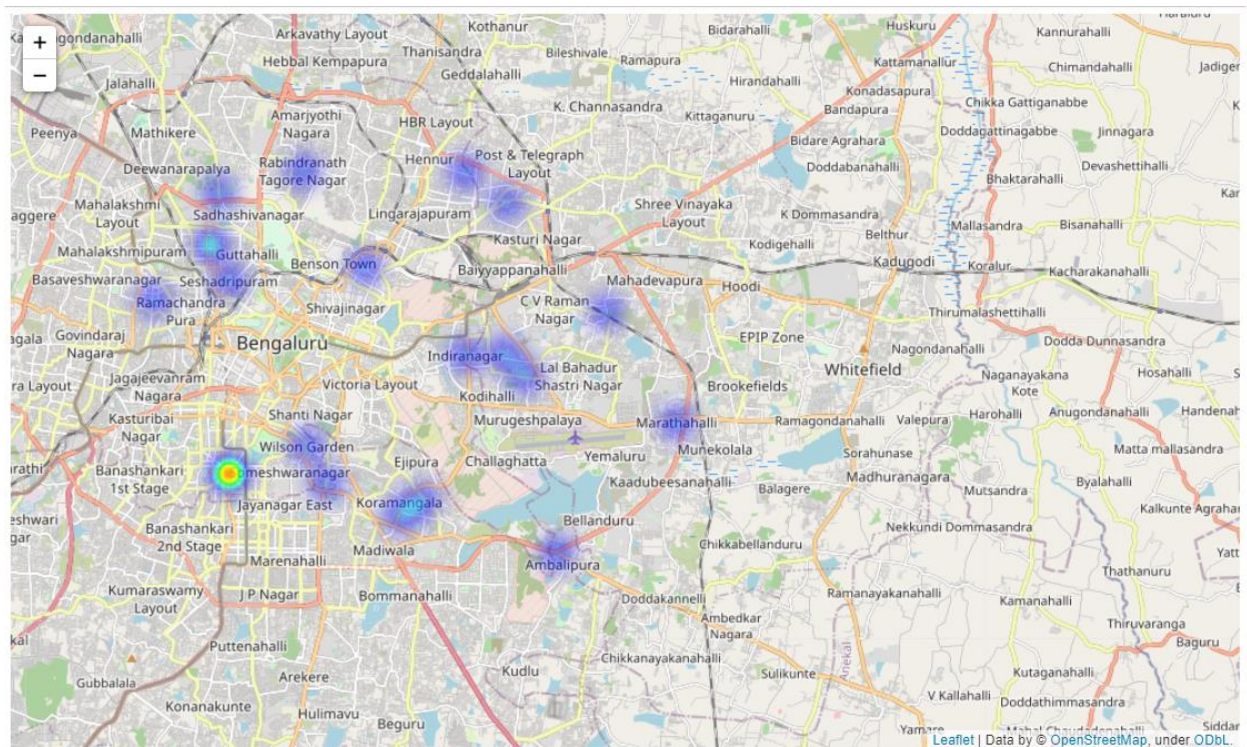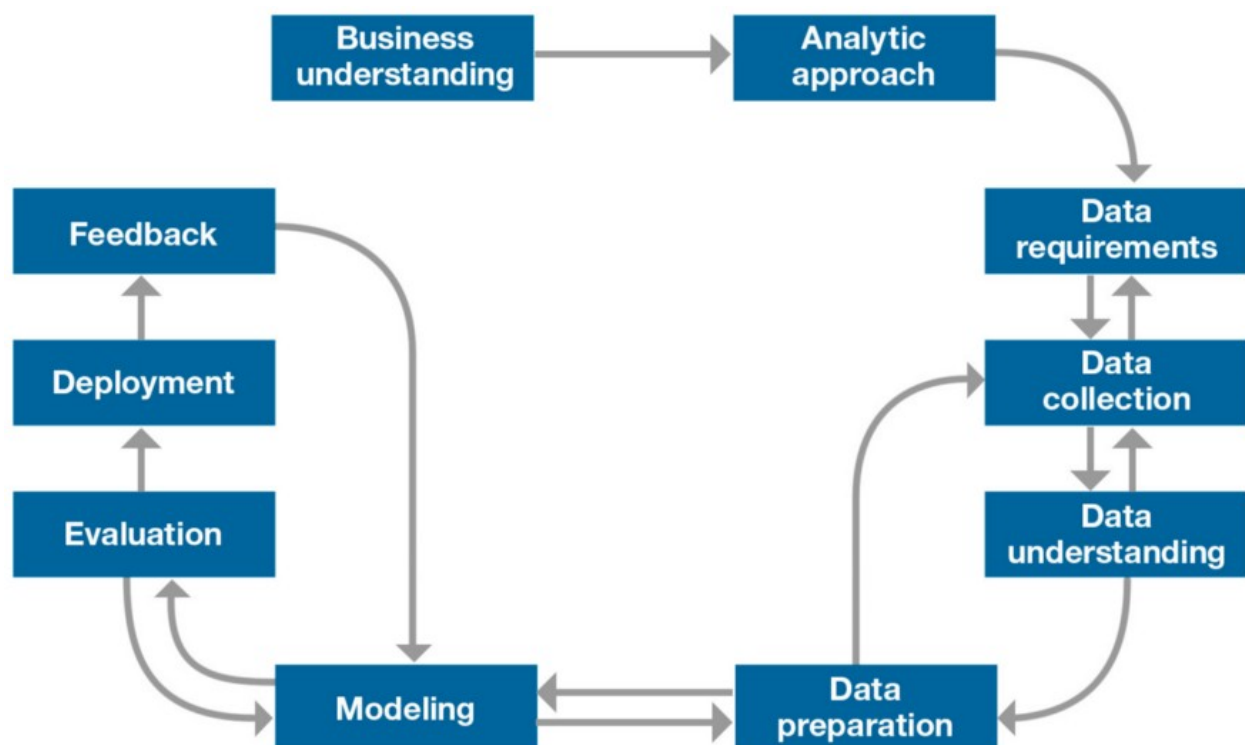Understanding the OBJECTIVE of the person who is asking the question is the first step in developing a well-defined inquiry strategy. Following the definition of the objective, the next stage is to establish the goals that will help the goal to be achieved. By breaking down the objectives, it is possible to have systematic discussions during which priorities may be established, culminating in the organisation and planning of how to address the problem. Depending on the nature of the issue, a large number of stakeholders will be necessary to engage in the discussion in order to aid in the definition of requirements and the clarification of difficulties.

Analytics Approach:
How may the information be used to answer the question? The question that is being asked determines the right analytic approach to be used. Following the identification of the problem to be solved, the most appropriate analytic approach is decided in light of the company's requirements. The second level of the data science method is represented by the symbol.

A predictive model may be used to determine the probability of an action occurring if the goal is to determine the likelihood of an action occurring. The use of a descriptive approach may be required if the query's goal is to illustrate linkages across data sets. A classification approach to predict response is useful in cases when counts are required. For example, if the inquiry requires a yes/no answer, a classification method to forecast response is appropriate.

Data Requirements
Prior to commencing the data collection and preparation parts of the technique, it is important to understand what is expected of you. This comprises determining the data content, format, and sources for the first data gathering.

Data Collection
This stage revises the data requirements and evaluates if the collection requires more or less data than was originally planned. Describing statistics and visualisation methods may be used to assess the content, quality, and early insights contained within the data collection. Gaps in the data will be identified, and plans will need to be put in place to either fill in the gaps or find other sources of information.

Data Understanding
The term "data comprehension" refers to all of the processes that took place throughout the data set's development. To put it another way, the data comprehension component of the data science method is concerned with the following question: Is the information you've gathered representative of the problem you're seeking to resolve?

Data Preparation
Data preparation, along with data collection and analysis, is the most time-consuming aspect of a data science project, accounting for 70 percent to 90 percent of the overall project time in most cases. When data is transformed during the data preparation phase, it means that it has been made more useable via the transformation process.

To work effectively with data, it must first be prepared in such a way that it can manage missing or incorrect information while also eliminating duplicates, all with the purpose of ensuring that everything is appropriately formatted throughout the process. The process of data preparation involves feature engineering as well. It is the process of leveraging domain knowledge about the data in order to produce the features needed by machine learning algorithms in order to improve their performance. A feature is a quality that may be beneficial in resolving a problem or addressing a problem more effectively. Predictive models are based on data characteristics, and these characteristics will have an impact on the results you want to achieve. In the case of data analysis including machine learning technology, feature engineering is essential.

In the phase of data preparation, you build the framework for the succeeding stages of the problem-solving process. While it may take some time to finish this phase, if it is done effectively, the results will be advantageous to the project as a whole. If you neglect this stage, the outcome will be substandard, and you may be forced to start again from the beginning.

Modelling
Data modelling is the process of creating descriptive or predictive models from data sets. A predictive model seeks to produce outcomes that are either yes or no, or stop or go. These models are constructed on the basis of the analytic approach that has been used, which may be statistical or machine learning-based. A thorough understanding of the problem and the use of the most appropriate analytical approach are required for efficient data collection, processing, and modelling. Each step must be polished, changed, and tuned on a continuing basis in order to produce a successful conclusion.

1. First and foremost, understand the nature of the problem at hand.

2. Second, decide on an analytical method or technique to use in order to resolve the problem.

3. Third, gather, understand, organise, and model data. This is the last step.

The model evaluation, deployment, and feedback loops included into the technique ensure that the response is near to and relevant at this moment in time. The greater the number of people who benefit from the repercussions of this approach, the more quickly the profession will grow and prosper.

Evaluation
Model evaluation takes place both during the model creation process and prior to deployment. When evaluating a model, not only is it possible to identify its overall quality, but it is also possible to determine whether it meets the requirements that were set down in the original demand.

Model evaluation may be separated into two phases that are unique from one another.

This is the initial phase in the process, and it is used to ensure that the model is running as intended. It is possible to utilise a decision tree to assess if the response that the model is capable of providing is compatible with the original design if the model is predictive in nature. It may be used to identify which areas of the body need repair. If the model is descriptive in nature, that is, it analyses relationships, it may be necessary to create the model using a testing set with known findings.

Statistical significance testing may be used as a second phase in the evaluation process, if necessary. This kind of evaluation may be used to validate that the model is handling and interpreting the data in an acceptable manner. This is done in order to avoid unwanted second thoughts once the answer has been revealed.

Deployment
As soon as the data scientist is certain that the model will work as expected, it is deployed and put through its final test. The model may be tested on a small group of users or in a test environment before being released to the general public to ensure that the output is applicable across a broad range of situations.

Feedback
Once the model has been put into operation, user feedback will be used to help refine it while also evaluating its effectiveness and impact. The value of the model will be measured by its capacity to absorb information and make adjustments in a timely manner for as long as the solution is required to be implemented. In the feedback process, the assumption is that the more you learn, the greater your desire to continue learning becomes.

7. <u>Implementation</u>

Step1 :
Python is the programming language that has been used to write the programmes for this project. It is necessary to write the programmes in a Jupyter Notebook since it is easier to see and contains distinct cells for each code to be executed. The following are the steps involved in putting the plan into action:
The first step is to create a plan.
Importing Library Resources
Some of the most important and distinctive libraries that are utilised are mentioned below.

In order to construct bar charts, histograms, and pie graphs, libraries such as Plotly, Matpolotlib, and Seaborn were utilised in conjunction with one another.

o Geopy.

Geocoders were utilised to get the longitudes and latitudes of the location data, which were then shown on a Bengaluru base map using the data.

o The Folium library was used to create maps and heat maps of Bengaluru's eateries, which were then superimposed on a base map of the city.

o The Sklearn library, which is an abbreviation for SCIKIT-Learn, is a free software machine learning library for the Python programming language that may be downloaded on the internet. It was utilised to apply all of the machine learning algorithms and to divide the data into two groups: training data and testing data (for testing purposes).

O A Python interface for artificial neural networks is provided by the Keras software package, which is free and open-source software. Keras serves as an interface between the user and the TensorFlow library. Tokenizer, pad sequences, and a few more things. The following algorithms are imported from the Keras library: Sequential Dense, Embedding, LSTM, and SpatialDropout1D. Tokenizer is used to separate individual words from sentences in the reviews, which allows for easier reading. It is necessary to verify that all of the sequences in a list have the same length, which is achieved by using the pad sequences function. Keras Embedding Layer is a kind of embedding layer. An Embedding layer is provided by Keras, which may be used to embed neural networks in text data. It is necessary for the input data to be integer encoded in order for each word to be represented by a different number. Hochreiter and Schmiduber were the first to propose Long Short Term Memory networks (LSTMs), which are now often referred to as "LSTMs." The applications of these include voice recognition, language modelling, sentiment analysis, and text prediction among others.

In order to do EDA, SENTIMENTAL ANALYSIS, and RATING PREDICTION, these libraries must be loaded into the project.

Step 2: Obtaining and analysing the dataset

The data set is downloaded from the source in the form of a CSV file, and the data is then read using Python code and shown on the screen.

Step 3: Pre-processing of the data:

Attributes were included in the dataset in total of 17.

• Records with null values were removed from the ratings columns and were replaced with a numerical value in the other columns, as described above.

• The values in the 'Rating' column have been modified. The string '/5' has been removed. For example, if a restaurant received a rating of 3.5/5, the rating was modified to 3.5.

• Encoding was performed on columns such as book table, online order, rest type, and listed in using the Label Encoding function from the sklearn package (city).

Step4: Using exploratory data analysis (EDA) as a fourth step:

As seen in the analysis section, exploratory data analysis is carried out via the use of algorithms that generate graphs, pie-charts, histograms, and other visual representations of data. The report's analysis section includes an explanation of how this data visualisation was accomplished.

Some data processing is also carried out in the EDA portion in order to get the best possible results from the data that is available.

Step 5: Sentimental Analysis of the Situation

Due to the fact that it is one of the two primary models used in this project, sentimental analysis of reviews begins with subject modelling. First we will split comments into two categories: negative and positive, based on the rating supplied. Comments with a rating of less than 2.5 are considered as negative, while comments with a rating more than 2.5 are regarded as positive.

Tokenizer is used to divide the sentences in the reviews into individual words once they have been identified as positive or negative. Then vectorizer is used to vectorize the words that have been tokenized. The words are then represented graphically in the form of plots, as seen in Figures 28 and 29. Figure 28 depicts

the favourable remarks, whereas Figure 29 depicts the negative comments, as well as the weightage of each word in each of the four subjects. The sentiment of the review is determined by using a machine learning algorithm that takes into account this weighting.

The dataset will be divided into 70 train data points and 30 test data points. Machine learning methods will be used to train the dataset, and the test data will be used to confirm the results of the training.

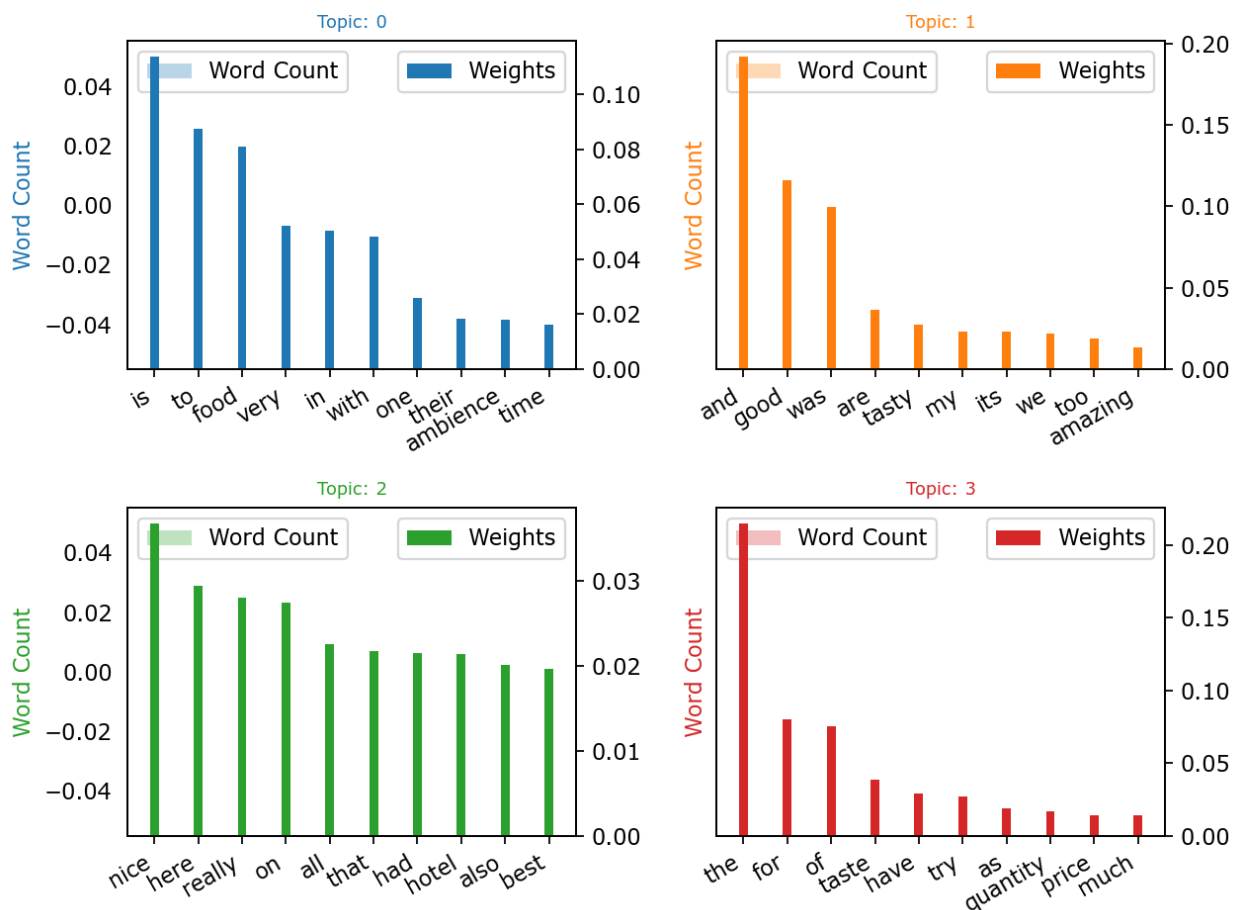
Figure – 28

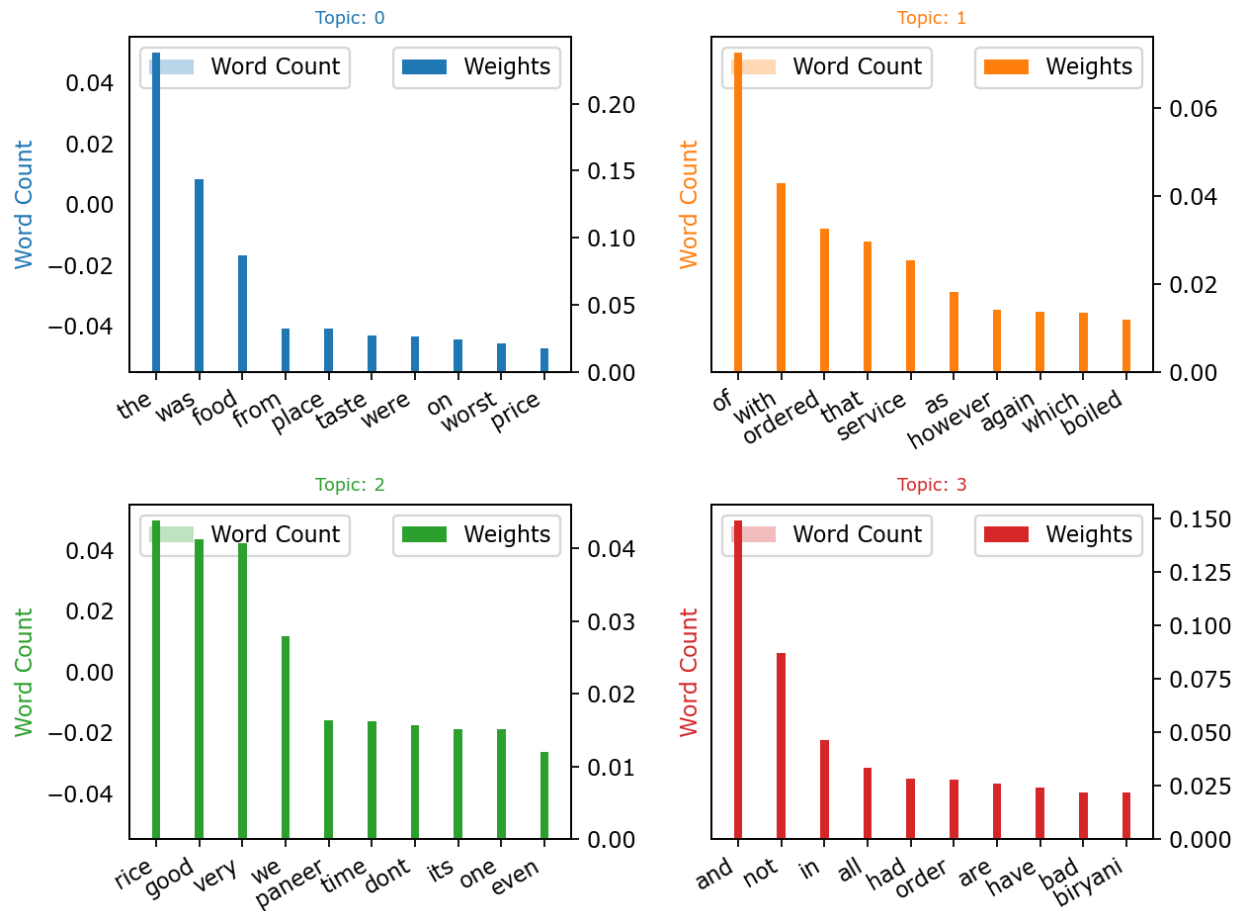Word Count and Importance of Topic Keywords

Figure – 29

T-SNE models are produced and plots are made , figures 30 and 31 show the plotted T-SNE models of positive and negative comments .
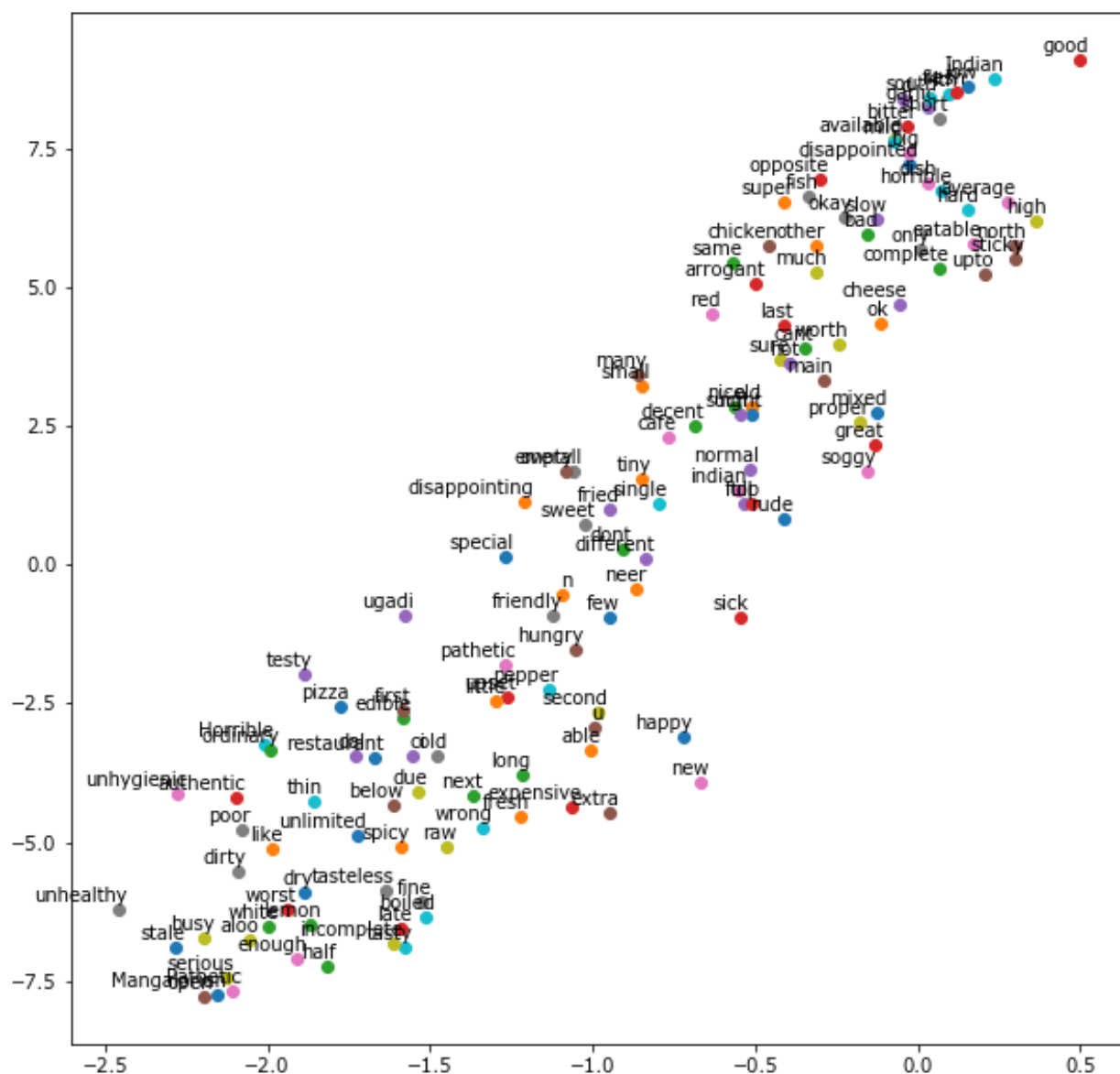
Figure – 29

Figure – 30

Step 6 – Rating Prediction :
This phase is divided into many substeps, each of which must be completed.

Step 6a – Data reading: The data is read a second time to ensure that all of the characteristics are present, after which the datatype information for each column is created to determine which columns need to be converted. With the exception of votes and Cuisine count, which have int64 data types, the most of the columns have object data types.

Columns are checked for the presence of null values. It is not necessary to include columns that are not needed for the prediction data, such as "link," "address," "phone," "reviews list," "menu item," "name," and "CuisineCount," since they are not required.

Step 6b: Removing duplicate values from the dataset and renaming certain columns to make them easier to find and utilise later on.

Similarly, the votes column, which has an int32 data type, is transformed to an object data type.

Step 6c: Data pre-processing, in which the data is tailored to be utilised for Machine learning algorithms, and label encoding is performed for chosen features, is performed . There are two feature lists created: one in which the cuisines are not divided into separate columns, and the other in which they are. Another one in which cuisines are arranged in eighty distinct rows is also available.

Step 6d: Feature scaling is accomplished by the use of standard scalar, which turns all of the values into a single scale. After that, the model is fitted, and the training and testing are divided into two parts: 80 percent train and 20 percent test.

Step 6e: The machine learning models are put into action. Five types of machine learning are employed: linear regression, random forest regressor, support vector regression, lasso, and decision tree regression. Linear regression is the most often utilised.

The results of the test and train are printed, and the R Squared values for both models are computed and shown.

Step 6f: The rating prediction is carried out with the help of the algorithm that attained the greatest accuracy. The predicted rating is compared to the actual rating in a table. Given that the majority of predicted ratings are in decimal values, they are rounded to three values. After the forecast has been made, it is saved and tabulated in CSV format for future reference.

8. <u>Testing and Results</u>

Sentiment analysis of reviews and rating prediction models are developed and evaluated in this project to ensure their correctness.

The table below summarises the accuracy of machine learning algorithms.

With UN-Sorted cuisines Column

ALGORITHM	ACCURACY
Linear Regression	21.33
Random Forest Regressor	78.74
Support Vector Regression	28.82
Decision tree Regression	66.99
lasso	-0.074

With Sorted cuisines Column

ALGORITHM	ACCURACY
Linear Regression	32.025
Random Forest Regressor	86.42
Support Vector Regression	46.42
Decision tree Regression	77.035
lasso	-0.0741

We may deduce from the two tables above that accuracy increases once the sorted cuisines column is introduced. Due to the fact that each restaurant could only provide a maximum of eight cuisines and a total of eighty cuisines, they were categorised into 80 cuisines. After the cuisines column was sorted, Algorithms received 103 columns as input, but only 7 columns were utilised when cuisines were not sorted. As a result, we can observe that the accuracy of all models has increased dramatically, except for Lasso.
Increased accuracy was reached, which aided in the prediction of the restaurant's rating with more precision. Because the Random Forest regressor has the maximum accuracy in this case, it is utilised to predict the restaurant's rating.
Answering the questions from objectives .

1. Does the demography of an area matters?
   Answer : Yes
2. Does location of a particular type of restaurant also depends on the people living in that area?
   Answer : Yes
3. Does the theme of the restaurant matters?
   Answer : not much , but yes
4. Is a food chain category restaurant likely to have more customers than its counter part?

Answer : not clear
5. Are any neighbourhoods similar based on the type of food?
   Answer : yes somewhat
6. Is a particular neighbourhood famous for its own kind of food?
   Answer : yes
7. If two neighbourhoods are similar does that mean these are related or particular group of people live in the neighbourhood or these are the places to it?
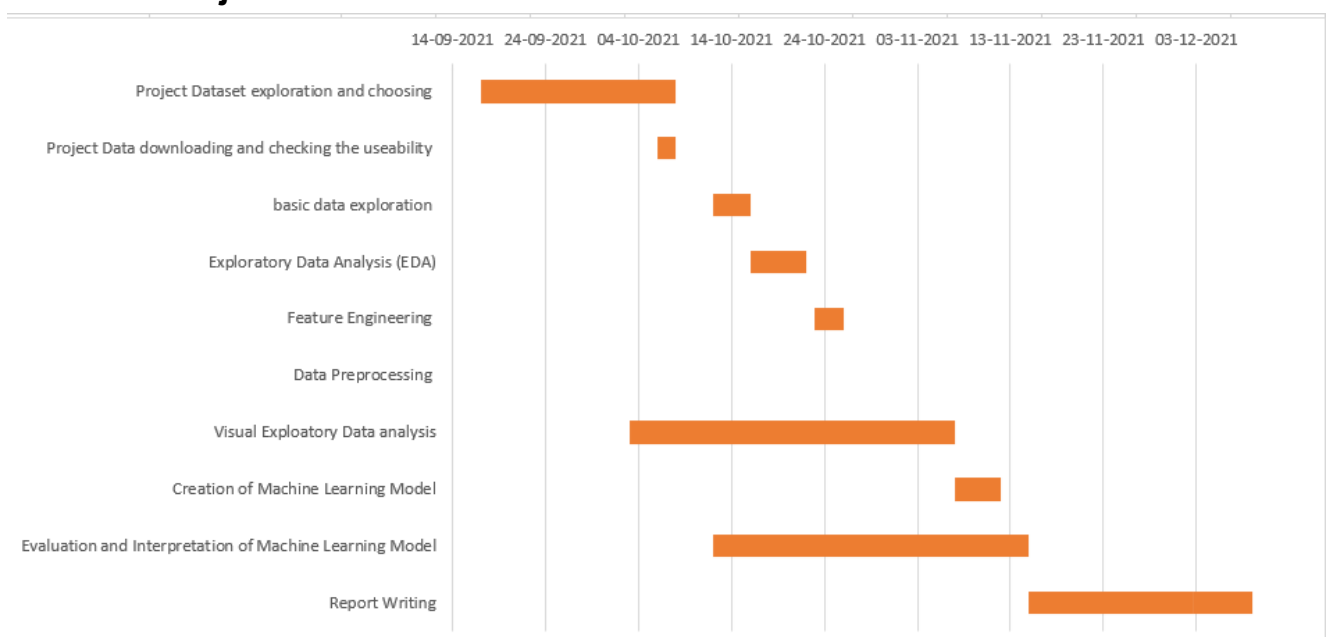   Answer : yes
8. Do the entire locality loves vegetarian food. If yes then is that locality populated by a particular sect of people for e.g. Brahmins, Jain, Marwaris, Guajarati's who are mostly vegetarian.
   Answer : not clear

8.Project Management

8.1 Project Schedule

The project schedule was planned using a Gantt chart; the majority of the work proceeded according to plan; however, a new proposed model could not be implemented due to a lack of funds for Google API usage for the project; however, the existing model was used and performance was increased in the case of Rating prediction.

Sentimental Analysis was implemented; however, since Epoch 5 was utilised, it took up to five hours to execute the code, causing it to sometimes run late. Coding took the majority of time because it is the foundation for getting the project working; once the dataset was selected, the steps in the coding took place as shown in the Gantt chart above; when the schedule was disrupted, I used to put up with late night sessions to get back on track; taking time off from part-time jobs was also necessary to get back on track. There were certain unavoidable circumstances that caused the timeline to be disrupted, yet it was still seen as a challenge to complete the project.

8.2 Risk Management

Sentimental analysis code would take five hours or more to run and would occasionally stop in the middle, necessitating the need to restart the entire code from the beginning. This was managed by checking the code again and again at intervals of half hour or fifteen minutes so that code runs and schedules would not be disrupted, posing a risk of not being able to submit the report on time.

Social, Legal, Ethical and Professional Considerations
Zomato Media Pt. Ltd.'s app, which was scraped for the data, included phone numbers and addresses of eateries. It was morally acceptable to utilise this information since it was publicly accessible information on venues that are open to the public.

The information that was scraped was only for the benefit of students. Please take note that I make no claim to ownership of the data. Zomato Media Pvt. Ltd. owns the copyrights to the data.

9.Critical Appraisal

o Several times over the course of developing the code, I had to rearrange the code in order to achieve the aim. The code always has a certain road map that must be followed, yet failing to do so will result in time being wasted in the process.
o Gaining an understanding of all of the algorithms that may be utilised for machine learning, as well as how they function when the features are sorted appropriately and applied.
To get more and more insights, co-relation plots and other graphs should be used in conjunction with an analytical and creative mindset. Eda's method provides opportunities to learn more about codes that may be used to create various data visualisation charts.

o Depending on the data, different codes were utilised, and it was learned how to use these codes with the dataset.

o Sentimental analysis was a new subject for me to learn about and experiment with. One thing that was very intriguing to learn about was how topic modelling worked, as well as its outputs and how the libraries that were imported, such as the tokenizer, functioned.

o Label encoding, how it works, and how it contributes to the production of higher accuracy in machine learning algorithms are discussed. The standard scalar procedure is the method through which the characteristics are converted to standard scale.

o the significance of all of the libraries that have been imported and their functionality. They are brought into the process in a variety of ways and at various stages.

Report writing and the method of writing, as well as the relevance of each area to be filled out, are discussed.

10.Conclusions

10.1 Achievements

- **The majority of the project's goals were addressed.**

- **A considerable boost in Machine Learning models' accuracy was achieved by sorting the Cuisines.**

10.2 Future Work

All restaurants' addresses are included in the data; eateries should be indicated on the map.

However, this could not be accomplished since it would need the usage of the GOOGLE MAPS API, which is a commercial service. By translating the addresses to longitudes and latitudes, we can plot individual restaurants on maps. Once plotted, we can forecast, using machine learning algorithms, what pricing range and cuisine should be employed if a customer needs to launch a new restaurant. The latitude and longitude of a newly opened restaurant will be input, and the machine learning model will output the pricing range and cuisine suggestion using clustering. A website is to be established that will allow for input and output.

11.Student Reflections

o Project management may have been more efficiently planned and performed.
o Additional relevant graphs might be created.
o Newly updated data may be scraped, and scraping techniques may be learned.

o Additional coding expertise is necessary to complete the project.

o The report could be worded more effectively.

12. Bibliography and References

- Sinha, N. (2018, March 3). *Understanding LSTM And Its Quick Implementation In Keras for Sentiment Analysis..* Medium. https://towardsdatascience.com/understanding-lstm-and-its-quick-implementation-in-keras-for-sentiment-analysis-af410fd85b47.

- Lasso (statistics) - Wikipedia. (n.d.). Lasso (statistics) - Wikipedia. https://en.wikipedia.org/wiki/Lasso_(statistics).

- What Is LASSO Regression Definition, Examples And Techniques. (2020, September 4). GreatLearning Blog: Free Resources what Matters to shape your Career!. https://www.mygreatlearning.com/blog/understanding-of-lasso-regression/.

- Rawat, A. S., & A. (2021, June 21). *What Is Support Vector Regression? | Analytics Steps.* What is Support Vector Regression? | Analytics Steps. https://www.analyticssteps.com/blogs/what-support-vector-regression.

- ISAR: Implicit Sentiment Analysis Of User Reviews. (n.d.). ISAR: Implicit sentiment analysis of user reviews. https://ieeexplore.ieee.org/abstract/document/7914994.

- Information for. (n.d.). https://www.tandfonline.com/doi/abs/10.2752/155280107780154141.

- Brownlee, J. (2017, October 3). *How To Use Word Embedding Layers for Deep Learning With Keras - Machine Learning Mastery.* Machine Learning Mastery. https://machinelearningmastery.com/use-word-embedding-layers-deep-learning-keras/#:~:text=2.-,Keras%20Embedding%20Layer,API%20also%20provided%20with%20Keras..

- Bar Chart - Wikipedia. (n.d.). Bar chart - Wikipedia. https://en.wikipedia.org/wiki/Bar_chart.

- Panuganty, H. (2021, May 4). *A Guide To Building Your First Data Science Project | By Harika Panuganty | Analytics Vidhya | Medium.* Medium. https://medium.com/analytics-vidhya/a-guide-to-building-your-first-data-science-project-e5f814b098a7.

- Panuganty, H. (2021, May 4). *A Guide To Building Your First Data Science Project | By Harika Panuganty | Analytics Vidhya | Medium.* Medium. https://medium.com/analytics-vidhya/a-guide-to-building-your-first-data-science-project-e5f814b098a7.

- Dataman, D. (2020, February 12). *Data Science Modeling Process & Six Consultative Roles | By Dr. Dataman | Dataman In AI | Medium.* Medium. https://medium.com/dataman-in-ai/data-science-modeling-process-fa6e8e45bf02.

- Gite, S., Udanshiv, A., Date, R., Jaisinghani, K., Singh, A., & Chetwani, P. (2021, May 25). *Sentiment Analysis Techniques On Food Reviews Using Machine Learning | SpringerLink.* Sentiment Analysis Techniques on Food Reviews Using Machine Learning | SpringerLink. https://link.springer.com/chapter/10.1007/978-981-16-0733-2_36.

- View Of Sentiment Analysis Of Customer Reviews In Zomato Bangalore Restaurants Using Random Forest Classifier. (n.d.). View of Sentiment Analysis of Customer Reviews in Zomato Bangalore Restaurants Using Random Forest Classifier. https://jurnal.unai.edu/index.php/isc/article/view/1003/1515.

- Choudhary, N., & Panwar, V. (2021, February 1). *Zomato Bangalore Restaurant Data Analysis.* zomato bangalore restaurant data analysis. https://www.jetir.org/view?paper=JETIR2102170.

- Dutta, K. B., Sahu, A., Sharma, B., Rautaray, S. S., & Pandey, M. (2020, July 31). *Machine Learning-Based Prototype for Restaurant Rating Prediction And Cuisine Selection | SpringerLink.* Machine Learning-Based Prototype for Restaurant Rating Prediction and Cuisine Selection | SpringerLink. https://link.springer.com/chapter/10.1007/978-981-15-5148-2_6.

Appendix A – Project Specification

Google drive 🔺 **- https://drive.google.com/drive/folders/1TTfl1LGgVY6oy8slHdac-jLl2Fd_UJPV?usp=sharing**

(works better)

Github ⬤ **- https://github.coventry.ac.uk/nallagondm/Zomato-Rating-Preddiction-and-sentimental-Analysis-/tree/main**

Appendix B – Interim Progress Report and Meeting Records

Meeting dates :

Sr. no	Meeting	Dates	Remarks
1.	First Meeting	15/09/21	Online, discussion of Ethics process and application.
2	Second Meeting	6/10/21	discussion of Dataset selection and Ethics form
3.	Third Meeting	13/10/21	discussion of Dataset exploration and Model discussion
4.	Fourth Meeting	20/10/21	Cancelled
5.	Fifth Meeting	03/11/21	discussion of Code methodology
6.	Sixth Meeting	17/11/21	discussion of EDA visualization
7.	Seventh Meeting	24/11/21	discussion of Model preparation and Report writing
8.	Eight Meeting	1/12/21	Help and doubt clearing sessions Report Writing
9.	Ninth Meeting	8/12/21	Report writing and verification.

Appendix C – Certificate of Ethics Approval

Analysis of Restaurant Industry   P127360

Coventry University

Certificate of Ethical Approval

Applicant: Monish Nallagondalla Srinath

Project Title: Analysis of Restaurant Industry

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval: 01 Nov 2021

Project Reference Number: P127360

Monish Nallagondalla Srinath (7150CEM)   Page 1   01 Nov 2021