Capstone Project - 4

Zomato Restaurant Clustering and Sentiment Analysis

Team Members
Akshada
Nikita

AI

Introduction

Zomato is an Indian multinational restaurant aggregator and food delivery company founded by Deepinder Goyal and Pankaj Chaddah in 2008. We'll be working on zomato restaurant datasets, where we'll go through the entire EDA process with data visualization, clustering, and sentiment analysis of reviews. This is an unsupervised learning problem. Clustering segregates groups with similar traits and assigns them into clusters. Sentiment analysis is one part of NLP, which is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback, and understand customer needs.

Objective

Zomato provides information, menus, and user-reviews of restaurants, and also has food delivery options from partner restaurants in select cities. The growing number of restaurants in every state of India has been a motivation to inspect the data to get some insights, interesting facts, and figures about the Indian food industry in each city. So, this project focuses on analysing the zomato restaurant data for each city in India. The given datasets are zomato restaurant names and metadata and zomato restaurant reviews. The zomato datasets consist of much information about restaurants and customers.

The main objective is to analyze the sentiments of the reviews given by the customers in the data and cluster the zomato restaurants into different segments and also draw some useful conclusions in the form of visualizations.

Why customer reviews are important?

Customer's online reviews have spawned a new domain in marketing and communication, bridging the gap between the traditional verbal and viral forms of feedback that can influence a consumer's opinion.

7 reasons why customer reviews are important:

1. To better understand the customers & improve customer service.

2. For credibility & social proof.

3. Reviews enable new businesses to stand with more established competitors and potentially gain a positive niche in people's estimations and expectations.

4. It allows consumers to have a voice and creates customer loyalty.

Why customer reviews are important? (Continued)

5. It helps businesses rank well, even if they have low quality link profiles.

6. Positive online business reviews are worth a great deal and can offer your business benefits that a simple marketing campaign can't. It creates continuous brand awareness that benefits the business in both the short and long term.

7. Reviews generate more reviews. The appearance of several reviews appears to be sufficient to provide new customers with the incentive and confidence to submit their own opinion on a specific product or service.

AI

Variable Breakdown

Zomato Restaurant names and Metadata:

We will use this dataset for the clustering.

- **Name : Name of Restaurants**

- **Links : URL Links of Restaurants**

- **Cost : Per person estimated Cost of dining**

- **Collection : Tagging of Restaurants w.r.t. Zomato categories**

- **Cuisines : Cuisines served by Restaurants**

- **Timings : Restaurant Timings**

Variable Breakdown(Continued)

Zomato Restaurant reviews:

We will merge this dataset with Names and Metadata and then use for sentiment analysis.

- **Restaurant : Name of the Restaurant**

- **Reviewer : Name of the Reviewer**

- **Review : Review Text**

- **Rating : Rating Provided by Reviewer**

- **MetaData : Reviewer Metadata - No. of Reviews and followers**

- **Time: Date and Time of Review**

- **Pictures : No. of pictures posted with review**

Steps involved

AI

01 **Data exploration and visualization**
- *pandas.DataFrame.info, pandas.DataFrame.describe*
- *Handled missing values, checked for duplicate values*
- *Used data visualizations (such as pie chart, bar chart, WordCloud, distribution plot, boxen plot etc.)*

02 **Clustering**
- *Multilabelbinarizer was used to encode multiple labels.*
- *StandardScaler() was used for scaling the data.*
- *The K-Means clustering algorithm was used for clustering the data and the number of clusters was decided using an elbow curve.*

03 **Text Preprocessing**
- *Data type correction*
- *Lower casing*
- *Removing punctuations, stopwords, URLs , html tags and emojis*
- *Lemmatization*

04 **Sentiment Analysis**
- *Text preprocessing*
- *TextBlob's function was used for analysing the polarity and subjectivity of the reviews.*
- *Pipelined model of TF-IDF vectorizer with logistic regression was implemented.*
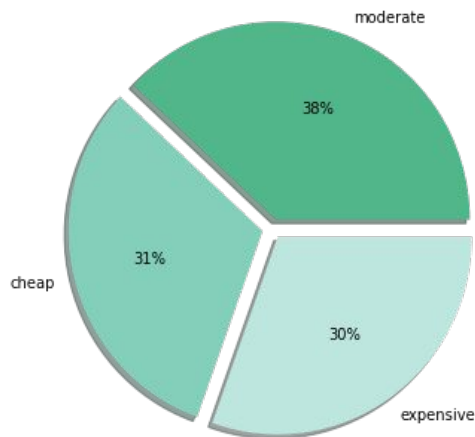
Descriptive statistics

- **The average per person estimated cost of dining for zomato restaurants is Rs. 861.428.**

- **The estimated minimum and maximum per person dining costs are Rs. 150 and Rs. 2800, respectively.**

- **The rating of restaurants ranges from 1 to 5.**

- **The average rating of zomato restaurants is 3.6.**

	Cost
count	105.000000
mean	861.428571
std	510.149730
min	150.000000
25%	500.000000
50%	700.000000
75%	1200.000000
max	2800.000000

	Rating	Pictures
count	9955.000000	9955.000000
mean	3.600402	0.751984
std	1.483565	2.575691
min	1.000000	0.000000
25%	3.000000	0.000000
50%	4.000000	0.000000
75%	5.000000	0.000000
max	5.000000	64.000000

Zomato restaurants distribution based on price range
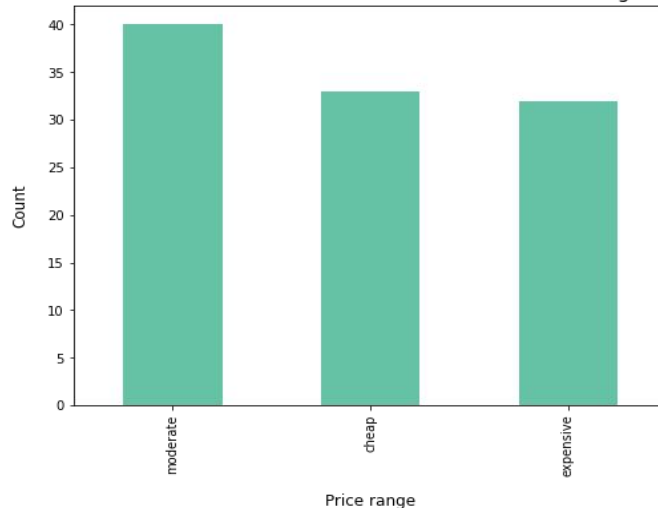
- **From the observations, we can conclude that our dataset has a balanced price range i.e. cheap is 33 (~ 31 %), moderate is 40 (~ 38 %), and expensive is 32 (~ 30 %).**
- **Zomato restaurants with a moderate price range are more.**
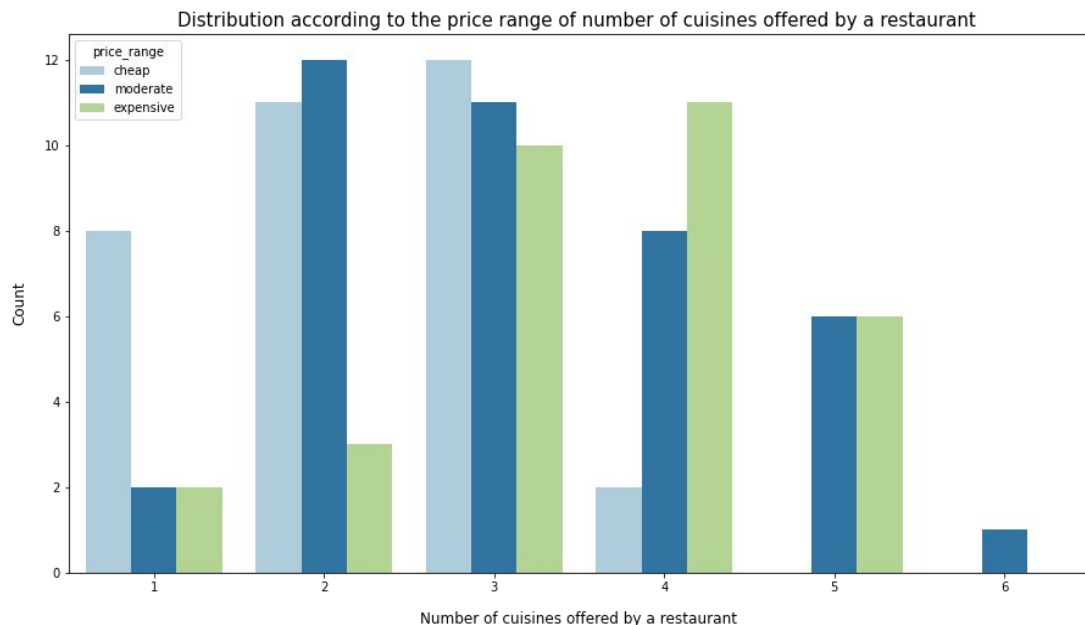
Top 20 cuisines in demand from zomato restaurants

- **North Indian is the most commonly offered cuisine, with 61 restaurants offering it.**

- **Chinese is the second most commonly offered cuisine, with 43 restaurants offering it.**

Top 20 Cuisines in demand from Zomato Restaurants
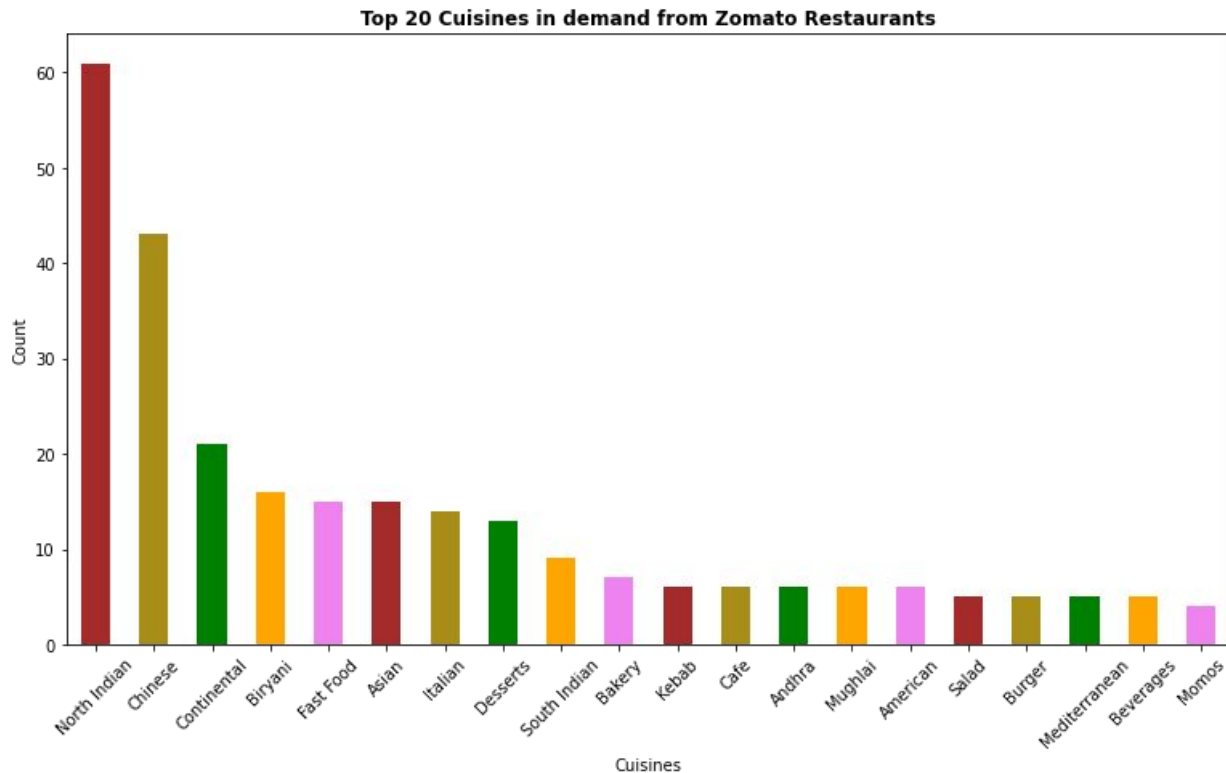
Distribution according to the price range of number of cuisines

- **Zomato restaurants offering 3 and 4 cuisines are more expensive, whereas the one offering six cuisines has a moderate price range.**

- **Zomato restaurants that are offering 2 and 3 cuisines have both low as well as moderate price ranges.**

Distribution according to the price range of number of cuisines offered by a restaurant

Zomato restaurants distribution based on ratings

- **From the pie chart, we can conclude that our dataset has an unbalanced rating range i.e. restaurants with exceptional is 3896 (~ 39 %), high is 2420 (~ 24 %), moderate is 1211 (~ 12 %), low is 639 (~ 7 %) and very low is 1735 (~ 17 %).**

- **Most customers are pleased with the food and services provided by Zomato restaurants, as evidenced by the majority of ratings of 5 stars.**

Zomato restaurants distribution based on ratings

Zomato restaurants distribution based on ratings

Top 10 zomato restaurants based on average ratings

AI
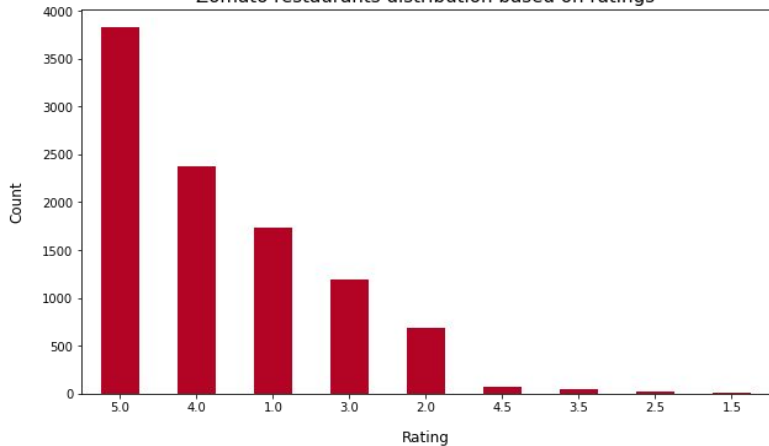
- **None of the top 10 zomato restaurants has got a full rating on average.**

- **AB's-Absolute Barbecues is the topper of zomato restaurants with close competition from B-Dubs.**

- **All the customers are not completely satisfied with any of the restaurants.**

Top 10 zomato restaurants based on average ratings
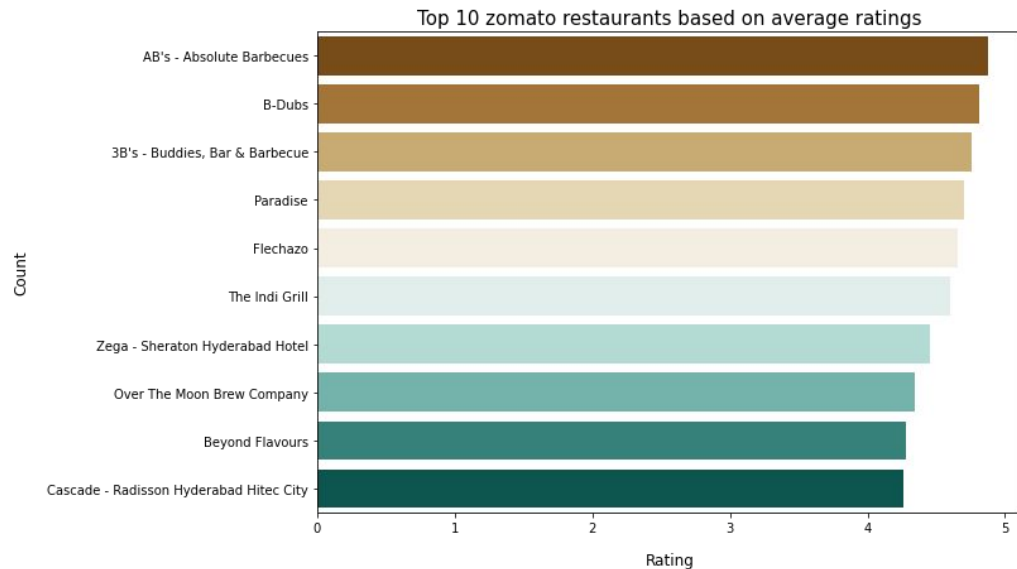
Top 10 expensive and affordable zomato restaurants

Top 10 affordable zomato restaurants based on average cost
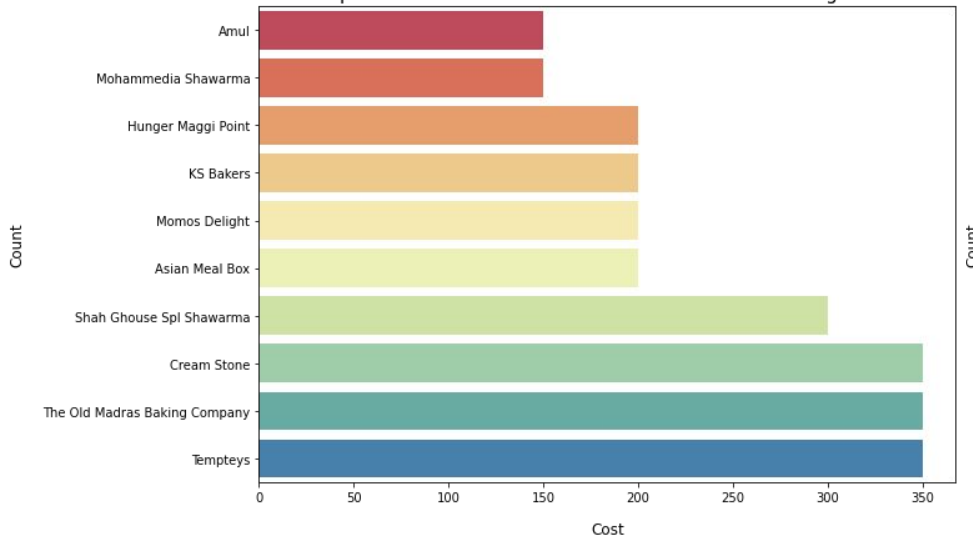
Top 10 expensive zomato restaurants based on average cost
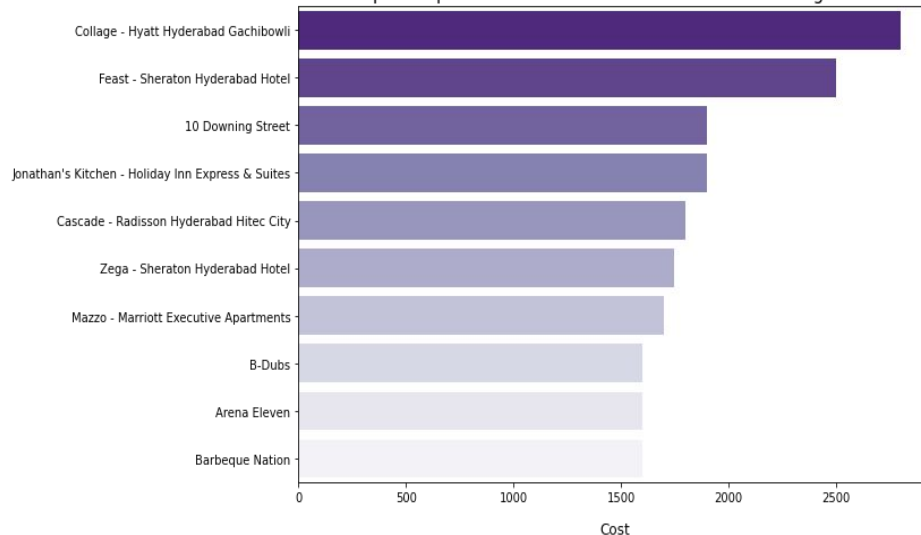
- **Amul and Mohammedia Shawarma are the cheapest zomato restaurants with an average cost of Rs. 150.**
- **Collage-Hyatt Hyderabad Gachibowli is the only restaurant to cross the average cost of Rs. 2500.**

Number of cuisines distribution w.r.t. cost

- **Restaurants with at least four cuisines have a starting price of Rs. 500 and above.**

- **Restaurants with three cuisines cover all the price ranges (i.e., cheap, moderate, and expensive).**

- **Restaurants with three cuisines may attract more customers because of their price range.**

Number of cuisines distribution w.r.t. cost
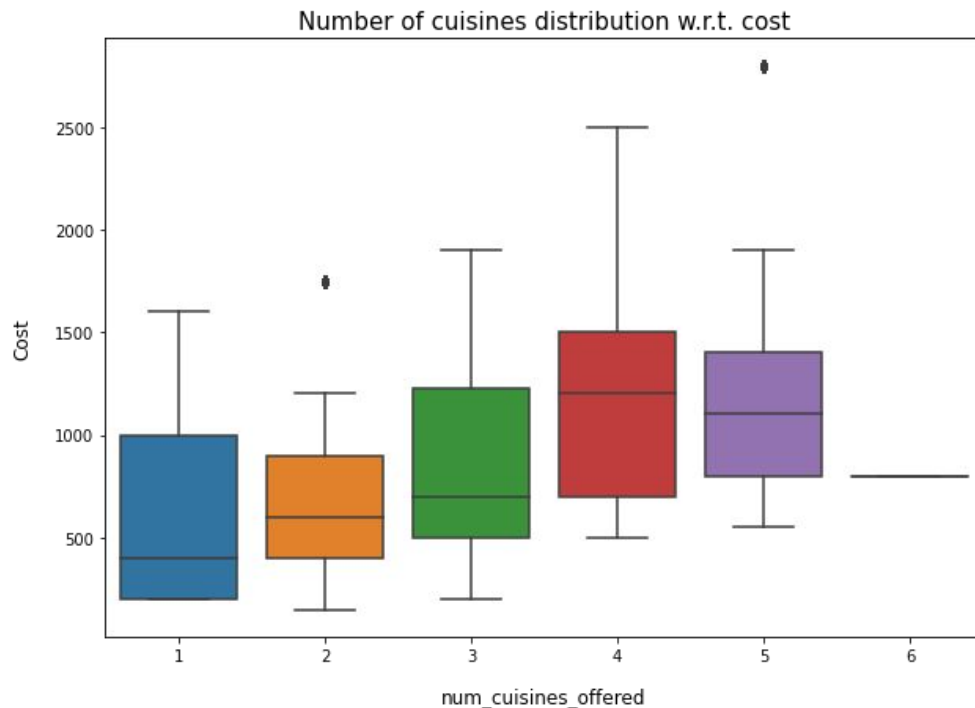
Top 10 reviewers on an average

- **On average, the total number of reviews reaches 700, except for the top reviewer with more than 1,000 reviews.**

- **Zomato can focus on top reviewers to know the reason behind so many reviews, whether it is due to satisfaction or dissatisfaction.**

- **If the review is due to dissatisfaction, zomato should try to resolve the issue.**

- **If the review is due to satisfaction, zomato can find the reason and continue with the same process.**

Top 10 reviewers on an average
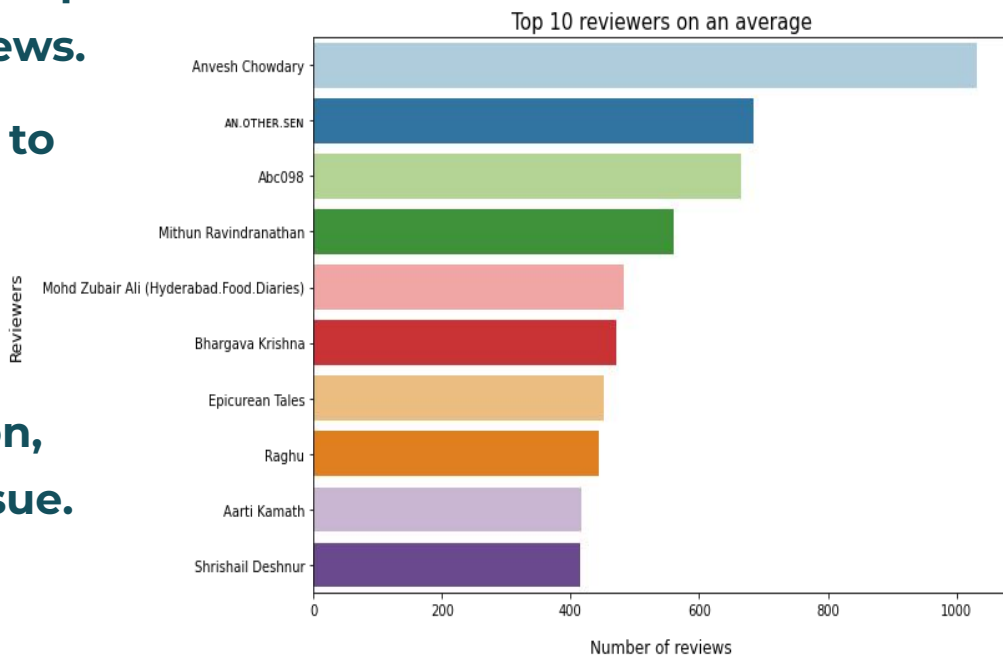
Distribution according to date and time of review

- **More reviews are there in the year 2018 and 2019.**

- **After 2017, there was a sudden rise in customer reviews.**

- **Reviewers are more active in later years.**

- **Customers prefer online ordering in later years.**

- **Zomato has got a sudden increase in popularity after the year 2017.**

Distribution according to date and time of review

Expanding and analysing 'Time' column

Reviews posted count per day of week

Reviews posted count per hour
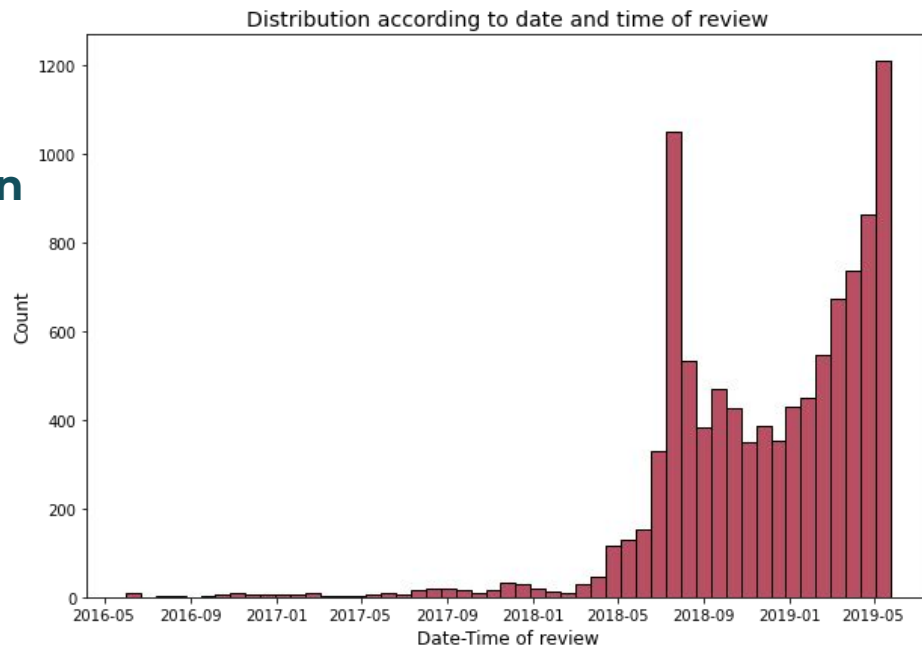
Average number of reviews per month
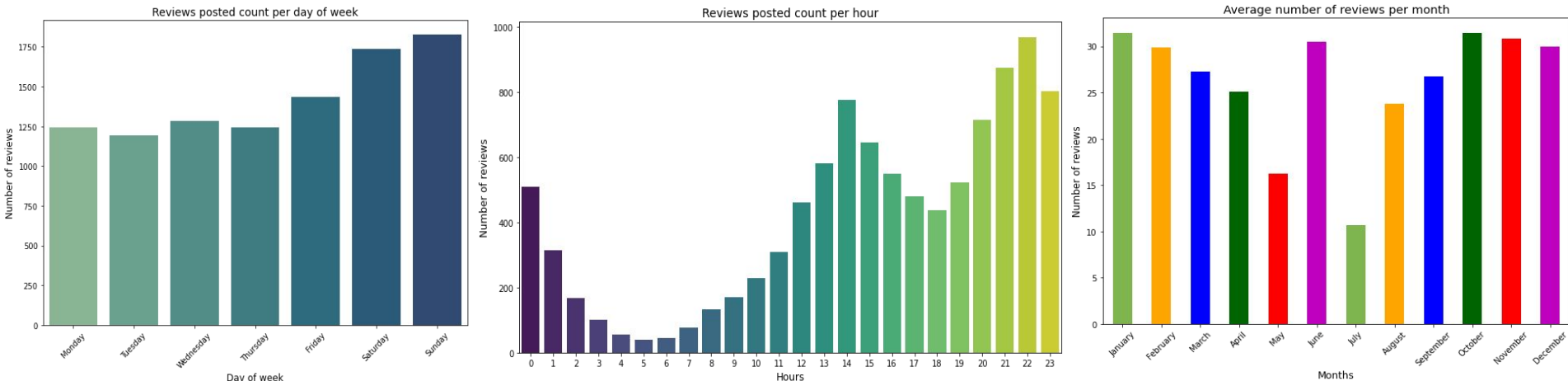
- **Maximum reviews are posted on weekends (i.e., Saturday and Sunday).**
- **The majority of reviews are posted at night (i.e., between 20 and 23 hrs).**
- **On average, the maximum number of reviews is in the month of January and October.**

Expanding and analysing 'Time' column(Continued)

- **From the heatmap, the highest and lowest average number of reviews per weekday/month are on the weekdays of June and July (i.e. Wednesday with approximately 47 reviews and Thursday with approximately 6 reviews).**

- **Customers are ordering less in May and July.**

Average number of reviews per weekday/month
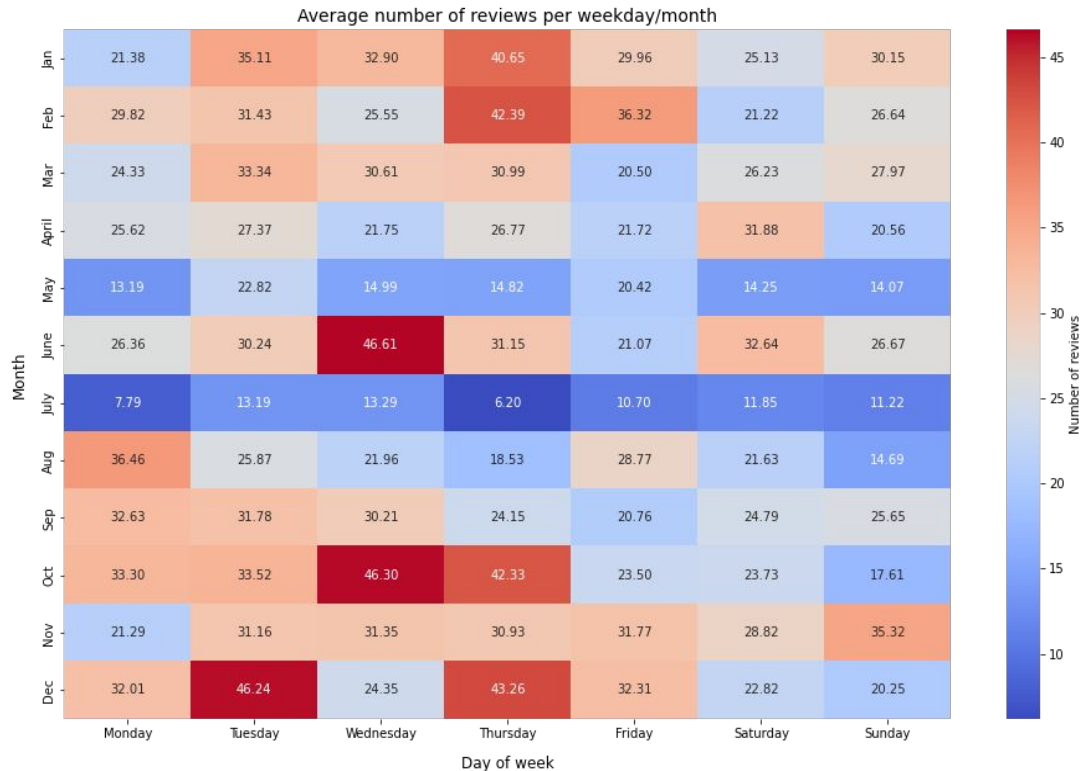
Recommended restaurants distribution based on cost, no. of reviews, and pictures posted

- **The most and least expensive restaurant is Collage-Hyatt Hyderabad Gachibowli and Amul, respectively, among the recommended restaurants.**

- **Cascade-Radisson Hyderabad Hitec City and Jonathan's Kitchen-Holiday Inn Express & Suites have got maximum and minimum number of reviews among the recommended restaurants.**


Recommended restaurants distribution based on cost, no. of reviews, and pictures posted

WordCloud

Most popular cuisines in zomato restaurants

Most famous zomato restaurants

Customer Reviews

- **The most popular cuisines in zomato restaurants are North Indian and Indian Chinese.**
- **Beyond Flavours and PourHouse7 are the most famous zomato restaurants.**
- **Popular words in zomato restaurant reviews are food and place.**
- **Zomato can suggest its restaurants to include North Indian and Indian Chinese in their cuisine to gain more orders.**
- **Customers are generally giving reviews based on food, place, taste, service, and ambience.**

Frequency distribution of all data using histogram
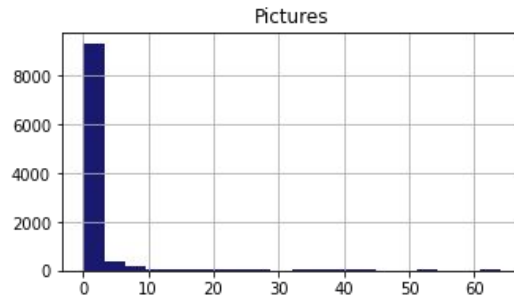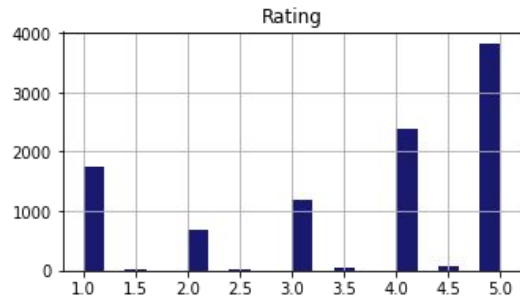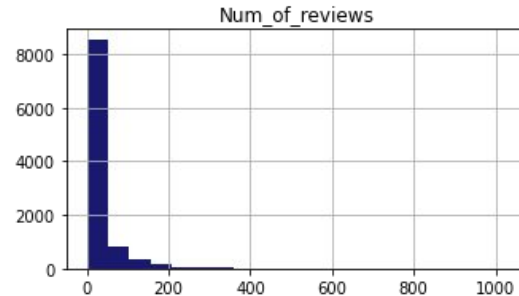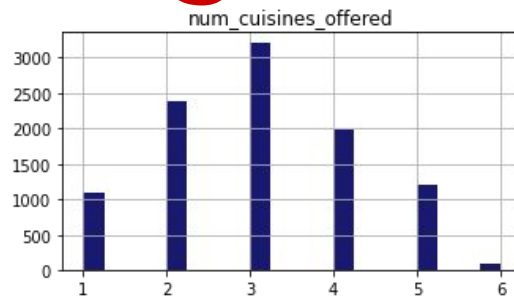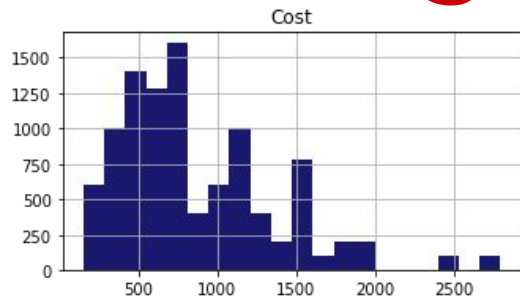
- **All variables show irregular distribution except "Pictures" and "Num_of_reviews".**
- **"Pictures" and "Num_of_reviews" are positively skewed.**

Heatmap
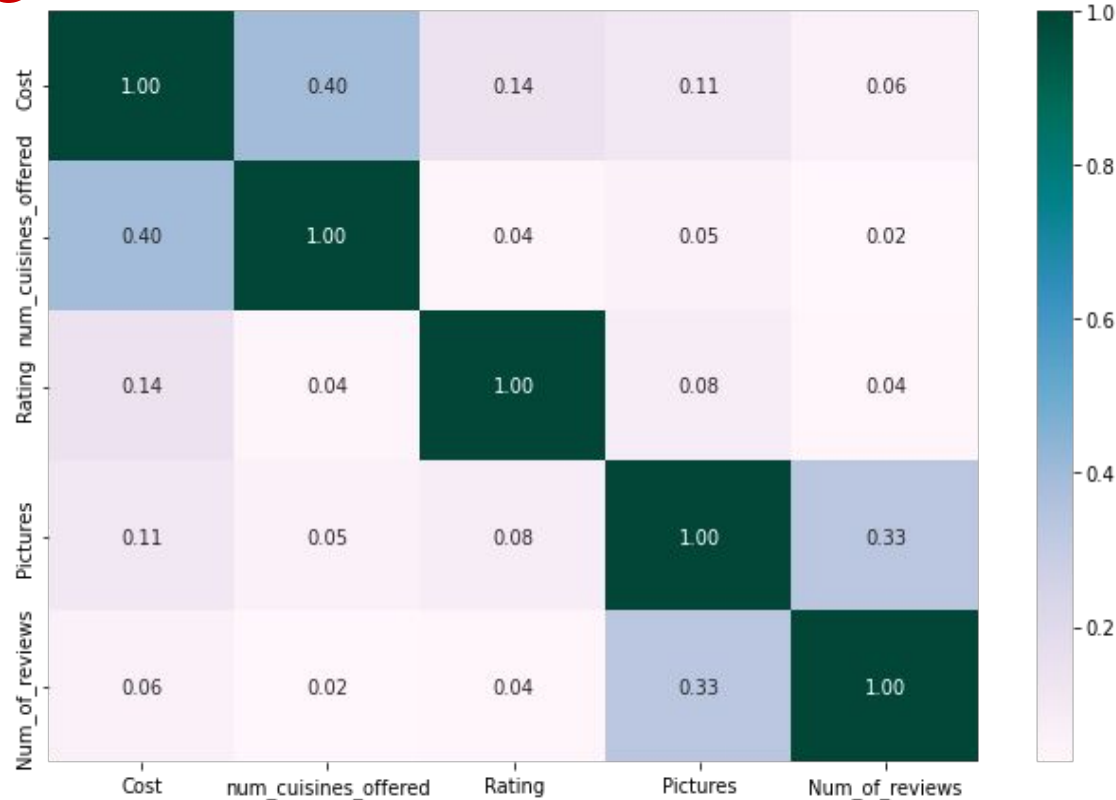
- **All variables are positively correlated with each other.**
- **None of the variables are highly correlated with each other.**

Clustering

Clustering is the process of segregating groups with similar traits and assigning them into clusters. It is essentially an unsupervised learning method. Clustering is important as it determines the inherent grouping of the unlabelled data present. There are no criteria for good clustering. It depends on the user and the criteria they may use to satisfy their needs. We used 'K-Means Clustering Algorithm' in this project.

K-means clustering algorithm is a type of partitioning method, it partitions the objects into k clusters and each partition forms one cluster. It divides 'n' observations into 'k' clusters, with the nearest mean serving as the cluster's prototype. 'K' here represents the number of groups/clusters we want to classify our items into.

Clustering (Continued)

- **The number of clusters was decided with the help of an Elbow Curve.**

- **WCSS (Within-Cluster Sum of Squares) is the sum of the squared distances between each point in a cluster and its centroid. When we plot the WCSS against the K value, the plot resembles an elbow.**

- **From Elbow Curve, we found that the best-suited k is 5.**
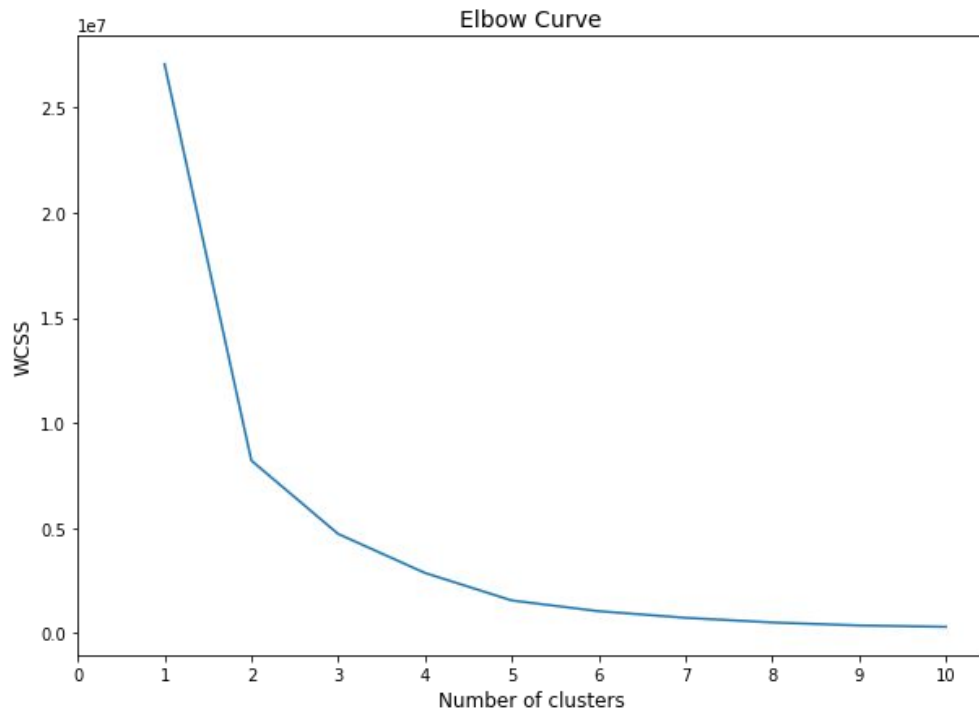
Clusters of restaurants based on price

- **From cluster distribution, restaurants with a low price have the largest cluster, and restaurants with an extremely high price have the smallest cluster.**
- **Zomato restaurants that have a low and affordable cost per person are in more demand.**

Clusters of restaurants based on price
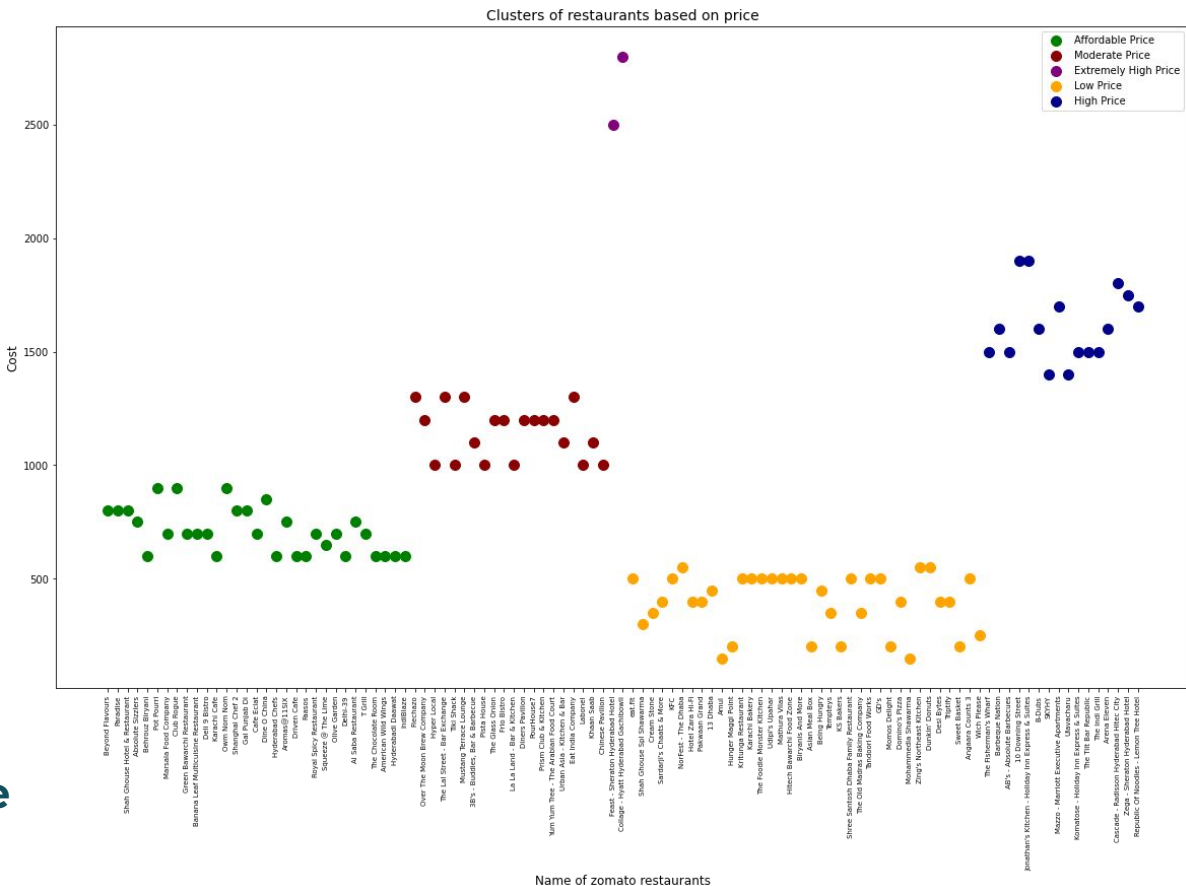
Clusters of number of cuisines and cost

- **From the cluster distribution, only restaurants with 4 and 5 cuisines offered have a different range of prices.**

- **Cuisines with a low and affordable cost are in more demand.**

Clusters of no. of cuisines and cost
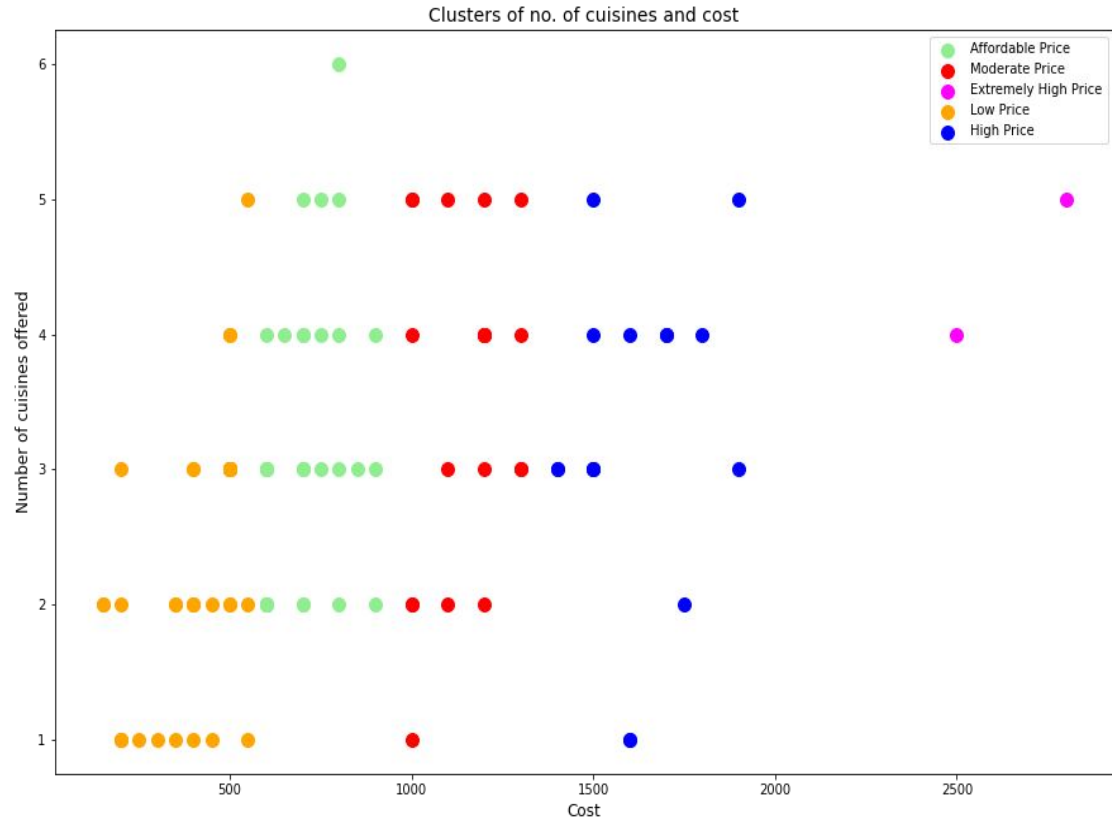
What is Sentiment Analysis?

Sentiment analysis is a subset of research areas such as natural language processing, data mining, and text mining, which are frequently used to analyse words based on people's writing patterns to discover positive or negative sentiments. The goal is to know how people feel about something from their text. Automatically analysing customer feedback, such as survey responses and social media conversations, enables brands to learn what makes customers happy or frustrated, allowing them to tailor products and services to their customers' needs.
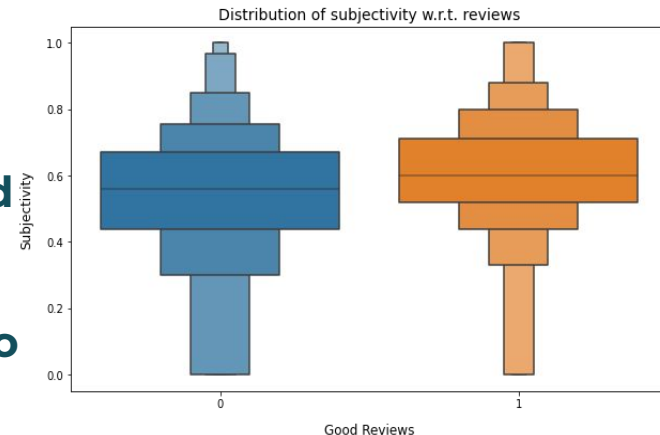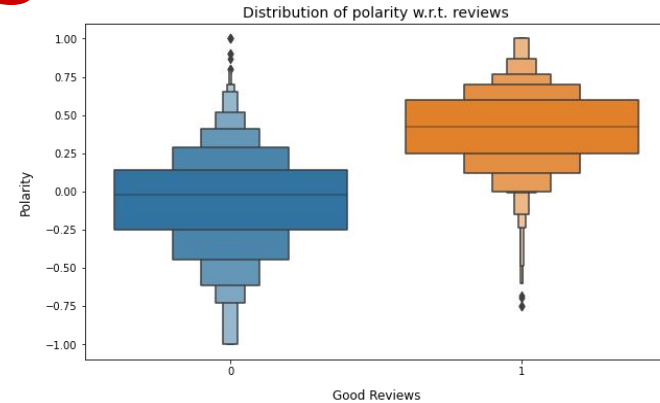
Distribution of polarity and subjectivity score

Distribution of polarity score

Distribution of subjectivity score
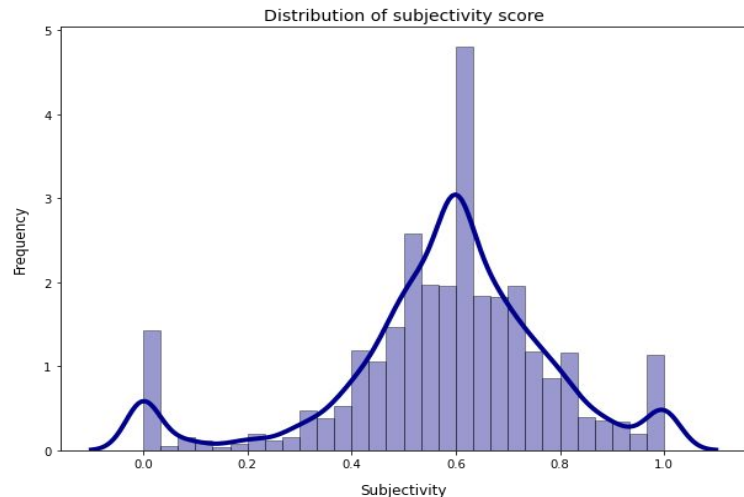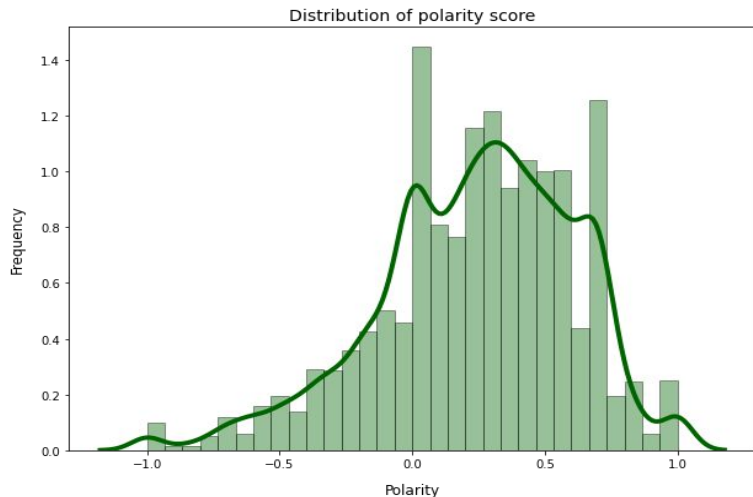
- **The density of positive feedback is more than negative feedback based on polarity distribution.**
- **The density of personal opinions is more than factual information based on subjectivity distribution.**

Distribution of polarity and subjectivity w.r.t reviews

- **We can observe that there are bad reviews with high polarity and good reviews with low polarity, also the median of good reviews is higher than the median of bad reviews for subjectivity.**

- **Customers giving positive feedback are not necessarily satisfied with the services offered by zomato restaurants.**

- **Similarly, customers giving negative feedback are not necessarily unsatisfied with the services offered by zomato restaurants.**

- **Zomato restaurants should improve their services to avoid negative feedback from customers.**

Distribution of polarity w.r.t. reviews

Distribution of subjectivity w.r.t. reviews

Positive and negative words in customer reviews

Positive words in Customer Reviews

Negative words in Customer Reviews

- **The most commonly used positive words in customer reviews are "good" and "nice" whereas the most commonly used negative words in customer reviews are "not good" and "less quantity".**

Relationship of subjectivity and polarity with reviews

- **Good reviews are more than bad reviews in polarity and subjectivity.**

- **Zomato restaurants should continue to provide the same service, but with some improvements to avoid bad reviews and gain more good reviews.**

Relationship of subjectivity and polarity with reviews
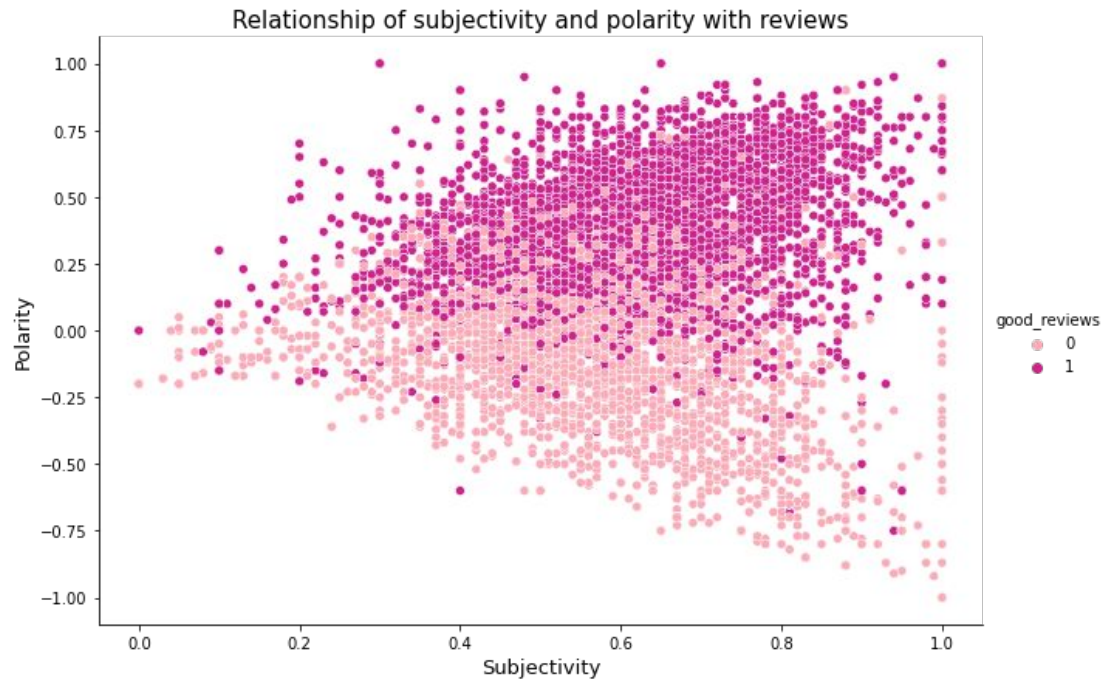
Conclusion

- From clusters, we observed that customers prefer zomato restaurants that have a low and affordable estimated cost per person, and there is more demand for cuisines with low and affordable prices.

- In sentiment analysis, we found that good reviews are more than bad reviews.

- The pipelined model of TF-IDF with logistic regression provides an accuracy of 0.89 and a ROC-AUC score of 0.95 for the validation set. The model predicted that 836/2489 were bad reviews and 1653/2489 were good reviews.

- The precision, recall, and f1-score for bad reviews in sentiment analysis of zomato restaurants are 0.89, 0.80, and 0.84, respectively, whereas for good reviews are 0.89, 0.94, and 0.91, respectively.

THANK YOU