

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Zomato is an Indian restaurant aggregator and food delivery start-up founded by Deepinder Goyal and Pankaj Chaddah in 2008. Zomato provides information, menus and user-reviews of restaurants, and also has food delivery options from partner restaurants in select cities.

India is quite famous for its diverse multi cuisine available in a large number of restaurants and hotel resorts, which is reminiscent of unity in diversity. 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 Project focuses on Customers and Company, we have analyzed the sentiments of the reviews given by the customer in the data and made some useful conclusion in the form of Visualizations. Also, cluster the Zomato restaurants into different segments. The data is visualized as it becomes easy to analyze data at instant. The Analysis also solve some of the business cases that can directly help the customers finding the Best restaurant in their locality and for the company to grow up and work on the fields they are currently lagging in.

First I have found out the duplicates and null values present in both the data sets . Then I have removed the null values or have dropped the column if needed. Then performed EDA on the MetaData datasheet. Plotted graph of Top 10 costliest 'restraus' vs their 'cost' , then I also plotted 'top affordable' vs 'cost' graph. Then made word cloud of both the categories of the restrau (top costly and top affordable restraus). Then extracted the stopwords from "nltk" library and removed stopwords from 'cuisines'. Then I removed the punctuations and repeating words to make the data cleaner. After that plotted graph of 'most cuisines' served and made its wordcloud.

Then I began with 'Review' datasheet. First I cleaned the data by removing null values . Then I performed EDA . Plotted graph between 'ratings' vs 'count'. Plotted month wise distribution of reviews. Then I plotted the graph showing the top 15 Reviewers along with their number of reviews. Plotted few more graphs to simplify the data. Then further cleaned the data by removing punctuation, removing stopwords and Lemmatization.

After all this, I performed sentiment analysis of the 'Review' datasheet and plotted the scatterd graph of the same showing the polarity of the positive , negative and neutral reviews. Then I used various models such as Random Forest classifier , SVM , MultinomialNB , XGB classifier and compared their accuracy . Then performed clustering using K-means clustering

Conclusion –

The most popular cuisines are the cuisines which most of the restaurants are willing to provide.

The most popular cuisines in Hyderabad are North Indian, Chinese, Continental, and Hyderabadi.

The cheapest is the food joint called Mohammedia Shawarma and the costliest restaurant is Collage – Hyatt Hyderabad Gachibowli.

Sentiment Analysis was done on the reviews and a model was trained in order to identify negative and positive sentiments.

SVM and XGB both performed well and we can choose any one them.

SVM and XGB are having 0.9188 and 0.9369 of testing accuracy respectively.
We got best cluster as 5 in K-Means and Principal Component Analysis(PCA)

Made by – Madhavi Mali

GitHub Link - [https://github.com/madhavimali/
Zomoto Restaurant Clustering and Sentiment Analysis MM](https://github.com/madhavimali/Zomoto_Restaurant_Clustering_and_Sentiment_Analysis_MM)