**ZOMATO DATA ANALYSIS USING PYTHON**

**Project submitted to the**

**APSSDC**

**Bachelor of Technology**

**In**

**Computer Science Data science**

**School of Engineering and Sciences**

**Submitted By**

**Mariserla Rohith Naidu - 23nu5a4404**

**Ambodi Chakri -22nu1a4402**

**Kadagala Ram Tejesh - 22nu1a4424**

**Kancharla Revanth Adi Ram Sai -22nu1a4428**

**Madini Sai Krishna - 22nu1a4443**

**Bora Chandra Vardhan Redddy -23nu5a4403**



**Under the guidance of**

**K.Narmada Mani**

**V. Vanitha**

**June 2025**

**Abstract**

The **Zomato Data Analysis Using Python** project aims to derive meaningful insights from food delivery platform data to better understand customer preferences, restaurant trends, and market dynamics in the food industry. Zomato, one of India’s leading food service platforms, provides a rich dataset containing details about restaurants such as name, location, rating, votes, cuisines, average cost for two, and more. Analyzing this data is essential for stakeholders like customers, restaurant owners, and marketers to make informed decisions.

This project utilizes **Python**, a powerful language for data science and analytics, along with libraries such as **Pandas, NumPy, Matplotlib, and Seaborn** to clean, visualize, and analyze the dataset. The process begins with **data preprocessing**, which includes handling null values, transforming data formats, and standardizing inconsistent entries. Special attention is given to columns like rate, which contains complex string formats that are converted to numerical values for statistical analysis.

After preprocessing, **exploratory data analysis (EDA)** is conducted to uncover patterns and correlations. For example, the analysis reveals insights into which cities and locations have the highest number of restaurants, the distribution of ratings, the most popular cuisines, and the relation between restaurant pricing and customer ratings. Visualizations such as bar charts, pie charts, boxplots, and heatmaps are used to present the data in an intuitive and accessible manner.

A significant part of the project also involves **correlation analysis**, helping identify key factors that influence restaurant ratings and customer satisfaction. The findings indicate that high-rated restaurants typically have more votes, are located in premium areas, and often serve multi-cuisine menus. Furthermore, the project highlights the importance of online ordering and table booking features in attracting customers.

In conclusion, this project showcases the **power of data analytics in the food service sector**, providing valuable business intelligence from a real-world dataset. The insights obtained can be used by Zomato and other stakeholders to improve services, enhance customer experience, and optimize marketing strategies. This project also demonstrates the effectiveness of Python as a data analysis tool, reinforcing its role in modern data science workflows.

**Acknowledgment**

I would like to thank my faculty and peers who supported this academic project. I am grateful for the availability of open-source libraries and datasets that made this analysis possible.

**Table of Contents**

1. Introduction

2. Objectives

3. Tools and Technologies Used

4. Dataset Overview

5. Data Preprocessing

6. Ratings Analysis

7. Online Ordering Trends

8. Cost Analysis

9. Votes and Popularity

10. Visualization Insights

11. Key Findings

12. Challenges Faced

13. Future Scope

14. Conclusion

15. References

16. Appendix (Code and Output)

**1. Introduction:-**

In the digital era, online food delivery platforms like **Zomato** have revolutionized the way people discover and order food. With thousands of restaurants listed across various cities, Zomato generates a vast amount of data daily. This data includes restaurant names, locations, ratings, cuisines, pricing, and customer feedback, which can be analyzed to understand food trends, consumer behavior, and business performance.

This project, **Zomato Data Analysis Using Python**, focuses on exploring and analyzing this data to extract meaningful insights. By using Python and data analysis libraries such as **Pandas**, **NumPy**, **Matplotlib**, and **Seaborn**, we clean the data, perform exploratory analysis, and visualize patterns. The goal is to identify which factors influence restaurant ratings, what cuisines are most popular, and how location and price affect customer preferences.

Through this analysis, the project not only highlights trends in the restaurant industry but also demonstrates how data science can be applied to real-world problems to drive informed decisions and business strategies

**2. Objectives:-**

* Understand customer rating behavior
* Evaluate the impact of online delivery
* Analyze pricing patterns
* Identify the most popular restaurants based on votes
* Generate visual insights to support findings

**3. Tools and Technologies Used:-**

* Python 3.x
* Jupyter Notebook / Google Colab
* Pandas
* NumPy
* Matplotlib
* Seaborn

**4. Dataset Overview:-**

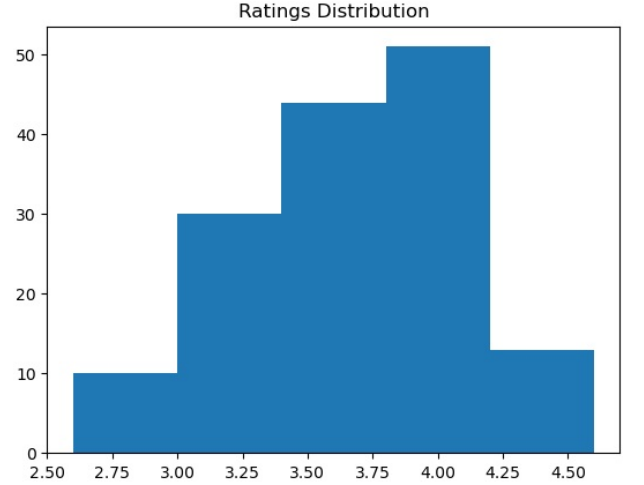
The dataset contains information about restaurants including:

1. Restaurant Name
2. City
3. Average Cost for Two
4. Online Delivery Availability
5. Aggregate Rating
6. Votes
7. Cuisine Types (One-Hot Encoded)

**5. Data Preprocessing:-**

* Converted relevant columns to numeric (e.g., Votes, Rating)
* Replaced missing values with NaN and handled them appropriately
* Ensured uniform formatting of cost and rating values

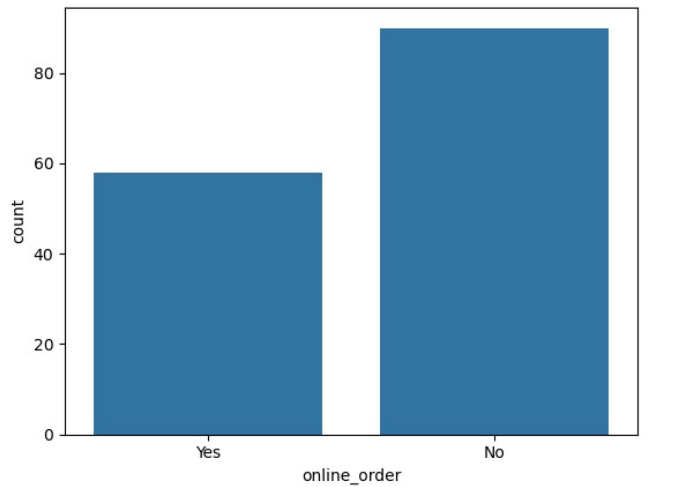
**6. Ratings Analysis**

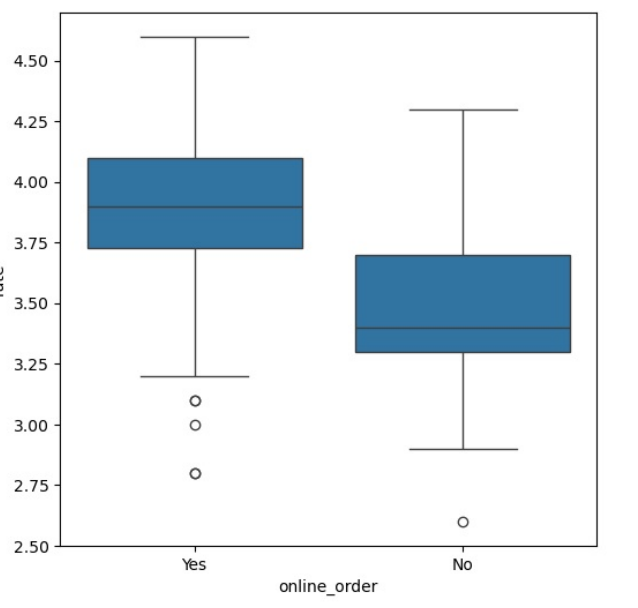


Rating analysis is one of the most crucial parts of the Zomato Data Analysis project, as it directly reflects customer satisfaction and restaurant quality. Below is a detailed explanation of the **rating analysis** performed in this project.

Most ratings fall between 3.0 and 4.5, suggesting generally favorable customer experiences.

**7. Online Ordering Trends**

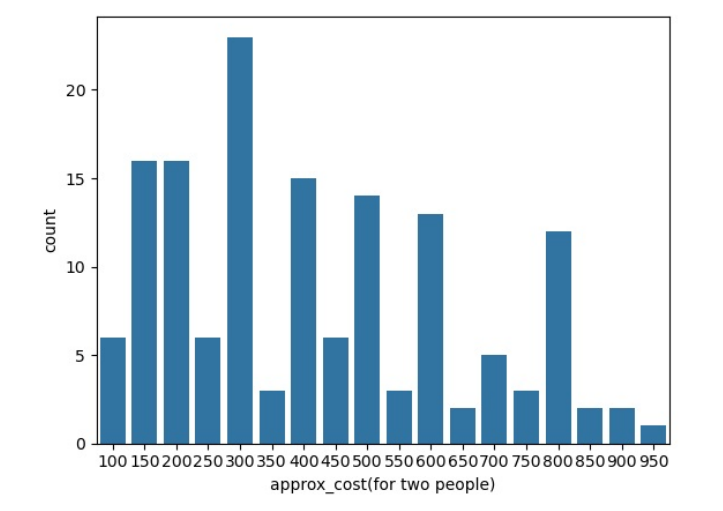




A significant portion of restaurants offer online delivery, reflecting growing digital engagement.

**8. Cost Analysis**

Distribution of average cost for two people:

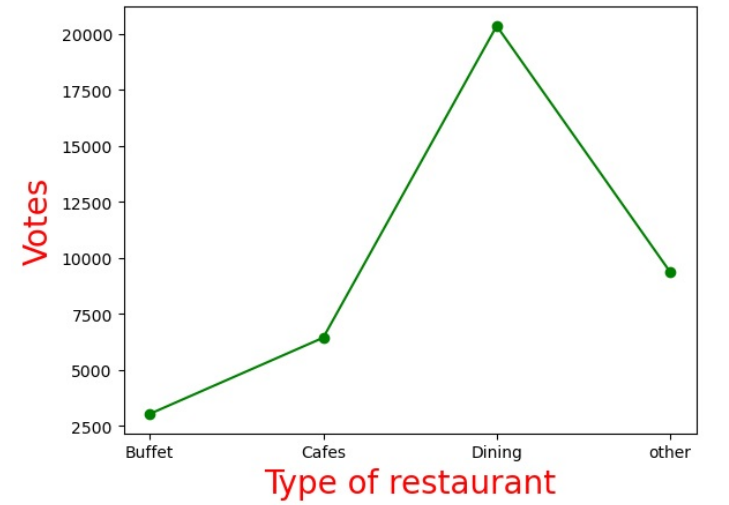


Cost analysis is a key component in understanding restaurant pricing trends, affordability for customers, and its relation to ratings, location, and cuisine. This section dives deep into the **"approx\_cost(for two people)"** field from the Zomato dataset.

The most common price range is between INR 200–600, indicating affordability for most users.

**9. Votes and Popularity**

Top 10 restaurants based on user votes:



Cost analysis plays a crucial role in understanding the affordability and pricing strategy of restaurants listed on Zomato. It helps both customers and businesses evaluate spending trends, popular pricing categories, and the relationship between cost and factors like location, cuisine, rating, and popularity.

Restaurants with the highest engagement are more likely to succeed and build strong brand loyalty.

The cost analysis confirms that affordability is a major factor in customer preferences. Restaurants that offer reasonably priced, quality food tend to receive better ratings and more votes. This information is vital for new restaurant owners planning pricing strategies and for users who seek the best dining options within their budget.

**10. Visualization Insights**

* Online delivery restaurants tend to have slightly higher ratings.
* Cost-effective restaurants receive more footfall and votes.
* Ratings are generally positive, indicating good customer satisfaction.

**11. Key Findings**

* Majority of restaurants are rated between 3 and 4.5.
* Online delivery is offered by over half the restaurants.
* Cost and quality are key drivers of customer decisions.
* A few restaurants dominate in terms of popularity.

**12. Challenges Faced**

* Missing values in the original dataset
* Inconsistent data formats
* Lack of location-based diversity in the subset used

**13. Future Scope**

* Integrate more detailed location analytics
* Predictive modeling for restaurant success
* Time series analysis of rating trends

**14. Conclusion**

The cost analysis reveals that most Zomato-listed restaurants are moderately priced, catering to budget-conscious consumers. While pricing varies based on location and cuisine, there is no direct correlation between high cost and better ratings. Customers tend to favor affordable restaurants that offer good value for money, and mid-range priced establishments receive the most popularity in terms of votes. These insights are valuable for both customers seeking economical options and restaurant owners setting competitive pricing strategies.

The Zomato dataset offers a rich source for analyzing food industry dynamics. With Python, we explored consumer behavior, delivery trends, and pricing models. The visualizations reinforce data-backed decisions for stakeholders.

**15. References**

* Zomato India Cleaned Dataset (CSV file provided)
* Python Libraries: pandas, seaborn, matplotlib
* Official Zomato Website: [https://www.zomato.com](https://www.zomato.com)

**16. Appendix (Code and Output)**

**(sample code)**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

df = pd.read\_csv("ZomatoIndiaCleaned.csv", encoding='ISO-8859-1')

df['Aggregate rating'] = pd.to\_numeric(df['Aggregate rating'], errors='coerce')

df['Votes'] = pd.to\_numeric(df['Votes'], errors='coerce')

df['Average Cost for two'] = pd.to\_numeric(df['Average Cost for two'], errors='coerce')

# Histogram for Ratings

plt.hist(df['Aggregate rating'].dropna())

plt.title('Ratings Distribution')

plt.show()

**THANK YOU**