A

DAV COURSE END PROJECT REPORT

on

DINING DYNAMICS

BE(IT)-IV Sem

By

J.Vinathi Reddy(160122737007)

M. Sai Varshitha (160122737015)

N.Manvika(160122737017)

Under guidance of

Dr Ramakrishna Kolikipogu Professor,

Assistant professor,

IT Department

DEPARTMENT OF INFORMATION TECHNOLOGY

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)

(Affiliated to Osmania University; Accredited by NBA(AICTE) and NAAC(UGC), ISO Certified 9001:2015)

KOKAPET(V), GANDIPET(M), RR District HYDERABAD - 75

2023-2024

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)

Kokapet(Village), Gandipet, Hyderabad, Telangana-500075. www.cbit.ac.in

ISO Certified 9001:2015

COMMITTED TO RESEARCH, INNOVATION AND EDUCATION 44 years

CERTIFICATE

This is to certify that the project work entitled "**Dining Dynamics**" submitted to CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY, in partial fulfillment of the requirements for the completion of Data Analysis and Visualization of IV Semester B.E. in Information Technology, during the Academic Year 2023-2024, is a record of original work done by J. Vinathi Reddy(160122737007), M. Sai Varshitha (160122737015) and Manvika(160122737017) during the period of study in the Department of IT, CBIT, HYDERABAD, under our guidance.

Project Guide	**Head of the Department**
Dr Ramakrishna Kolikipogu	**Dr. Rajnikanth Aluvalu**
Assistant Professor, Dept. of IT,	Professor, Dept. of IT,
CBIT, Hyderabad.	CBIT, Hyderabad.

ACKNOWLEDGEMENTS

We would like to express our heartfelt gratitude to Dr Ramakrishna Kolikipogu, our project guide, for his invaluable guidance and constant support, along with his capable instruction and persistent encouragement.

We are grateful to our Head of Department, Dr. Rajanikanth Aluvalu, for his steady support and for the provision of every resource required for the completion of this project.

We would like to take this opportunity to thank our Principal, Dr. C.V. Narasimhulu, as well as the Management of the Institute, for having designed an excellent learning atmosphere.

Our thanks are due to all members of the staff and our lab assistants for providing us with the help required to carry out the groundwork of this project.

TABLE OF CONTENTS

LIST OF FIGURES

DAV COURSE END PROJECT

TITLE OF THE PROJECT: Dining Dynamics

PROBLEM STATEMENT: Develop a decision analysis framework utilizing Zomato's data.

ABSTRACT

Zomato Data Analysis Using Python" embarks on a data-driven exploration of dining preferences and trends through the lens of Zomato's vast repository. Leveraging Python's powerful libraries— NumPy for efficient computations, Matplotlib for visualization, Pandas for data manipulation and Seaborn for enhanced graphics. This project aims to unravel intriguing insights. Through meticulous analysis, we tackle pivotal questions. Firstly, we investigate the prevalence of online delivery versus offline services among restaurants, shedding light on evolving consumer behavior. Next, we delve into the realm of culinary preferences, discerning the most favored restaurant types among the populace.

Firstly, it seeks to evaluate the prevalence and popularity of online delivery versus offline services among restaurants listed on Zomato, shedding light on evolving consumer behavior and preferences. Secondly, it endeavors to identify the most favored types of restaurants among users, leveraging user interactions, ratings, and reviews to discern culinary preferences and dining experiences preferred by the general public. Lastly, it aims to delve into the realm of dining preferences for couples, investigating preferred price ranges for romantic dinners to provide actionable insights for restaurant owners and marketers targeting this demographic segment.

Additionally, the project dives into the realm of dining dynamics for couples, scrutinizing preferred price ranges for romantic dinners. Utilizing Google Colab or Jupyter Notebook for seamless execution, our analysis promises to offer actionable insights for restaurateurs, diners, and Zomato alike.

INTRODUCTION

The "Zomato Data Analysis Using Python" project aims to leverage Zomato's extensive data to uncover key insights into dining preferences and trends. With the rise of digital platforms, understanding these trends is crucial for stakeholders in the food and hospitality industry. We explore fundamental questions such as whether more restaurants offer online delivery compared to offline dining, and which types of restaurants are most popular among users. Additionally, we analyze the preferred price ranges for couples dining out and identify the majority rating ranges received by restaurants. Using Python's powerful libraries—Numpy, Pandas, Matplotlib, and Seaborn—we perform data manipulation, visualization, and statistical analysis to uncover these insights. Our goal is to provide actionable information for restaurant owners, marketers, and food enthusiasts, enabling them to optimize offerings, enhance customer satisfaction, and drive business growth in a competitive market.

OBJECTIVES AND OUTCOMES

OBJECTIVES:

- Determine the proportion of restaurants offering online delivery compared to those providing offline services, aiming to identify the predominant mode of service in the restaurant industry.
- Investigate the popularity of different types of restaurants among the general public to discern which categories are most favored, contributing to a better understanding of consumer preferences in the dining landscape.
- Analyze the preferred price range for couples dining at restaurants, aiming to identify the typical spending habits of couples during dinner outings and provide insights for restaurant pricing strategies.

OUTCOMES:

- Majority Restaurant Type Identification:
  Clear identification of the predominant types of restaurants listed on Zomato, providing insights into the diversity of dining options available to users.
- Preference Analysis for Restaurants:
  Detailed analysis revealing which specific types of restaurants are preferred by a larger number of individuals, enabling stakeholders to understand popular culinary trends and preferences.
- Online vs. Offline Service Evaluation:
  Insightful comparison of the number of restaurants offering online delivery services versus those providing offline dining experiences, shedding light on evolving consumer behaviors and preferences in the digital age.

- Most Favored Restaurant Types:
  Identification of the types of restaurants that are most favored by the general public, based on factors such as user ratings, reviews, and interactions, providing valuable insights into popular dining experiences.
- Preferred Price Range for Couples:
  Analysis of price ranges preferred by couples for their dinner outings at restaurants, offering insights into dining dynamics and expenditure patterns within this demographic segment.
- Rating Distribution and Majority Ratings:
  Understanding the distribution of ratings received by restaurants listed on Zomato, including identification of the majority rating range, facilitating insights into overall customer satisfaction levels and restaurant performance.

Overall, the project aims to provide comprehensive insights into various aspects of dining preferences and behaviors, enabling stakeholders within the restaurant industry to make informed decisions, optimize their offerings, and enhance the overall dining experience for customers.

TECHNOLOGY STACK

i. Programming Languages:

- Python: Widely used for data analysis, machine learning, and scripting tasks.

ii. Data Preprocessing:

- Pandas: For data manipulation, cleaning, and transformation.
- NumPy: For numerical computations and array operations.

iii. Exploratory Data Analysis:

- Matplotlib, Seaborn: For data visualization and exploratory plots.

iv. Version Control and Collaboration:

- Git: For version control and collaboration.
- GitHub or GitLab: Platforms for hosting code repositories and managing projects.

v. Development Environment:

- Google Collaborate

vi. Documentation and Report:

- Microsoft Word

ALOGRITHM

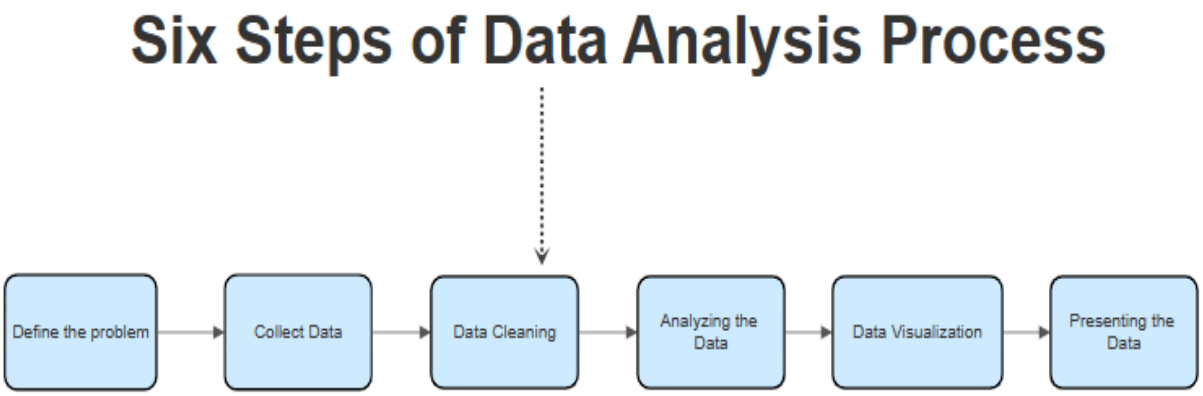

Figure 1: Algorithm

MATHEMATICAL ANALYSIS

The mathematical analysis for the "Zomato Data Analysis Using Python" project involves various statistical and computational techniques to extract meaningful insights from the data. Below is a breakdown of the key mathematical components involved:

i. **Descriptive Statistics:**

- Mean, Median, and Mode: Calculate these central tendency measures for numerical data such as ratings and price ranges to understand the typical values.
- Standard Deviation and Variance: Measure the dispersion of restaurant ratings and price ranges to understand the variability in the data.
- Frequency Distribution: Determine the frequency of different restaurant types, service types (online vs. offline), and rating ranges.

ii. **Comparative Analysis**:

- Proportion Comparison: Compare the proportions of restaurants offering online delivery versus offline services using a chi-square test to determine if the difference is statistically significant.
- Preference Analysis: Use bar charts and pie charts to visualize the popularity of different types of restaurants and price ranges preferred by couples.

iii. Correlation Analysis:

- Correlation Coefficient (Pearson's r): Measure the strength and direction of the linear relationship between different variables, such as price range and rating, or restaurant type and rating.

- Spearman's Rank Correlation: Use this non-parametric measure to assess the relationship between ordinal variables, such as ranking of restaurant types based on popularity.

iv. Hypothesis Testing:

- t-Tests: Conduct independent t-tests to compare the mean ratings of restaurants that offer online delivery versus those that do not, to see if there is a significant difference in customer satisfaction.
- ANOVA (Analysis of Variance): Use ANOVA to compare the means of ratings across multiple restaurant types to determine if certain types are rated significantly higher than others.

v. Regression Analysis:

- Linear Regression: Model the relationship between restaurant ratings (dependent variable) and factors like price range, type, and delivery option (independent variables).
- Logistic Regression: Use logistic regression to predict the likelihood of a restaurant offering online delivery based on its characteristics.

vi. **Visualization:**

- Histograms and Box Plots: Visualize the distribution of numerical variables like ratings and price ranges.
- Scatter Plots: Plot scatter diagrams to visualize the relationships between pairs of variables, such as price range and rating.

By applying these mathematical and statistical techniques, the project aims to provide a comprehensive analysis of the Zomato data, uncovering patterns and trends that can inform decision-making for stakeholders in the food and hospitality industry.

CODE, IMPLEMENTATION and RESULTS

CODE:

```
from google.colab import drive
drive.mount('/content/drive')
from google.colab import files
uploaded = files.upload()
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
dataframe = pd.read_csv("Zomato data .csv")
print(dataframe.head())
def handleRate(value):
  value=str(value).split('/')
  value=value[0];
```

```
  return float(value)

dataframe['rate']=dataframe['rate'].apply(handleRate)
print(dataframe.head())
dataframe.info()
sns.countplot(x=dataframe['listed_in(type)'])
plt.xlabel("Type of restaurant")
grouped_data = dataframe.groupby('listed_in(type)')['votes'].sum()
result = pd.DataFrame({'votes': grouped_data})
plt.plot(result, c="green", marker="o")
plt.xlabel("Type of restaurant", c="red", size=20)
plt.ylabel("Votes", c="red", size=20)
max_votes = dataframe['votes'].max()
restaurant_with_max_votes = dataframe.loc[dataframe['votes'] ==
max_votes, 'name']
print("Restaurant(s) with the maximum votes:")
print(restaurant_with_max_votes)
sns.countplot(x= dataframe['online_order'])
plt.hist(dataframe['rate'],bins=5)
plt.title("Ratings Distribution")
plt.show()
couple_data=dataframe['approx_cost(for two people)']
sns.countplot(x=couple_data)
plt.figure(figsize = (6,6))
sns.boxplot(x = 'online_order', y = 'rate', data = dataframe)
```

IMPLEMENTATION AND RESULTS:

Step 1: Uploading the Zomato data.csv file into google colab.

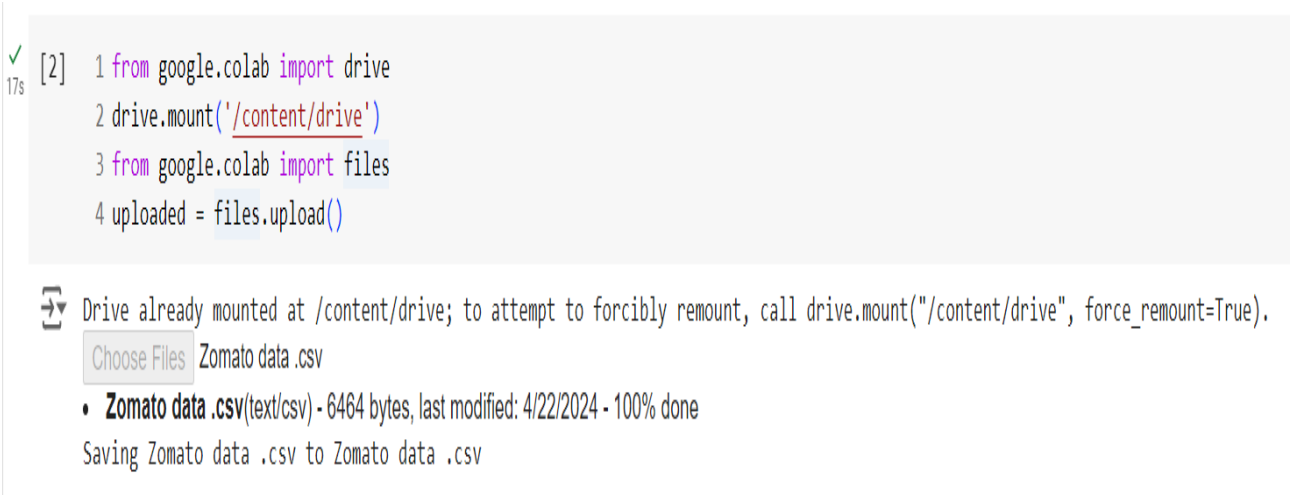

Figure 2: Uploading the files.

To run in google colab we should first upload the dataset i.e. Zomato data.csv from google drive.

Step 2: Importing the neccesary libraries

```
[21] import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
```

Figure 3: Importing the neccesary libraries

Step 3: Creating a Data frame

```
1 dataframe = pd.read_csv("Zomato data .csv")
2 print(dataframe.head())
```

```
                  name online_order book_table   rate  votes  \
0               Jalsa          Yes        Yes  4.1/5    775
1        Spice Elephant          Yes         No  4.1/5    787
2       San Churro Cafe          Yes         No  3.8/5    918
3  Addhuri Udupi Bhojana           No         No  3.7/5     88
4          Grand Village           No         No  3.8/5    166

   approx_cost(for two people) listed_in(type)
0                          800          Buffet
1                          800          Buffet
2                          800          Buffet
3                          300          Buffet
4                          600          Buffet
```

Figure 4: Creating the Data frame

Step 4: converting the data type of the “rate” column to float and remove the denominator.

```
1 def handleRate(value):
2   value=str(value).split('/')
3   value=value[0];
4   return float(value)
5
6 dataframe['rate']=dataframe['rate'].apply(handleRate)
7 print(dataframe.head())
```

```
                  name online_order book_table  rate  votes  \
0               Jalsa          Yes        Yes   4.1    775
1        Spice Elephant          Yes         No   4.1    787
2       San Churro Cafe          Yes         No   3.8    918
3  Addhuri Udupi Bhojana           No         No   3.7     88
4          Grand Village           No         No   3.8    166

   approx_cost(for two people) listed_in(type)
0                          800          Buffet
1                          800          Buffet
2                          800          Buffet
3                          300          Buffet
4                          600          Buffet
```

Figure 5: conversion of data type of the “rate” column to float and remove the denominator.

Step 5: summary of the data frame.

```
1 dataframe.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148 entries, 0 to 147
Data columns (total 7 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   name                         148 non-null    object
 1   online_order                 148 non-null    object
 2   book_table                   148 non-null    object
 3   rate                         148 non-null    float64
 4   votes                        148 non-null    int64
 5   approx_cost(for two people)  148 non-null    int64
 6   listed_in(type)              148 non-null    object
dtypes: float64(1), int64(2), object(4)
memory usage: 8.2+ KB
```

Figure 6: Summary of the Data frame

Step 6: Exploration of the listed_in(type) column.

```
1 sns.countplot(x=dataframe['listed_in(type)'])
2 plt.xlabel("Type of restaurant")
3
```

```
Text(0.5, 0, 'Type of restaurant')
```

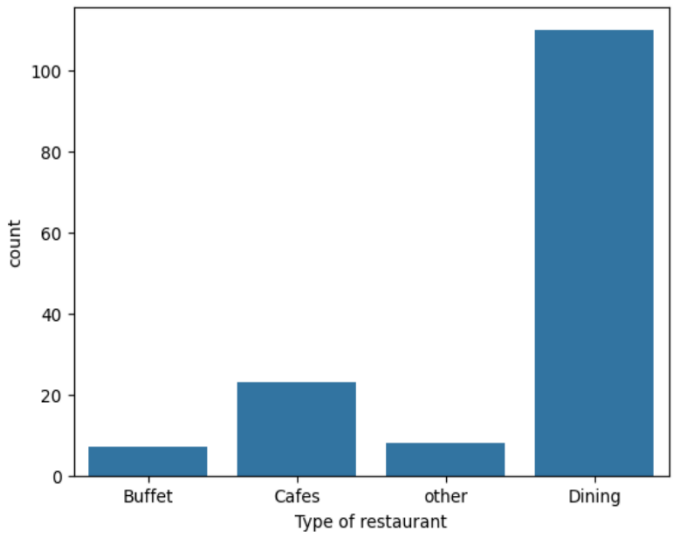

Figure 7: Exploring of the listed_in(type) column.

Step 7: Exploration of votes corresponding to each category of Restaurant

```
grouped_data = dataframe.groupby('listed_in(type)')['votes'].sum()
result = pd.DataFrame({'votes': grouped_data})
plt.plot(result, c="green", marker="o")
plt.xlabel("Type of restaurant", c="red", size=20)
plt.ylabel("Votes", c="red", size=20)

```

```
Text(0, 0.5, 'Votes')
```

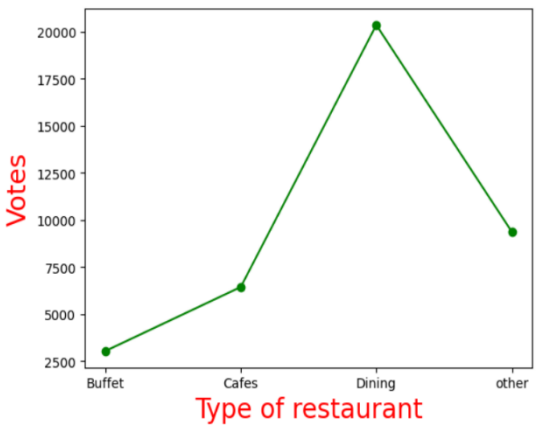

Figure 8: Exploring which category in restaurant got highest votes

Step 8: Determination the restaurant's name that received the maximum votes based on a given Data frame.

```
max_votes = dataframe['votes'].max()
restaurant_with_max_votes = dataframe.loc[dataframe['votes'] == max_votes, 'name']
print("Restaurant(s) with the maximum votes:")
print(restaurant_with_max_votes)
```

```
Restaurant(s) with the maximum votes:
38    Empire Restaurant
Name: name, dtype: object
```

Figure 9: Determining the restaurant's name that received the maximum votes based on a given Data frame.

Step 9: Exploring online orders of restaurant

```
sns.countplot(x= dataframe['online_order'])
```

```
<Axes: xlabel='online_order', ylabel='count'>
```

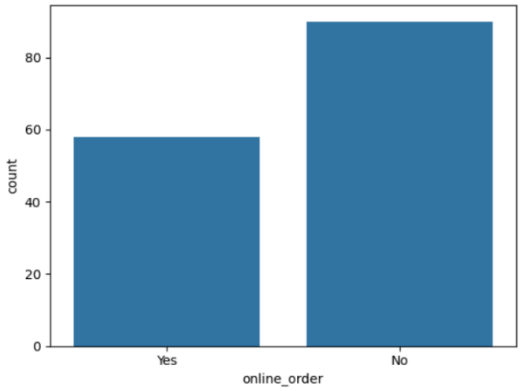

Figure 10: Exploring orders of restaurant

Step 10: Exploration of rate column

```
plt.hist(dataframe['rate'],bins=5)
plt.title("Ratings Distribution")
plt.show()
```

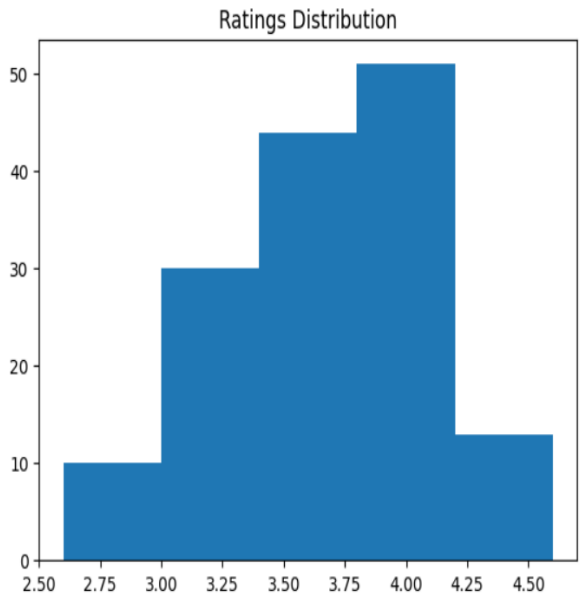

Figure 11: Exploring the rate column.

Step 11: Exploration of approx_cost(for two people) column.

```
couple_data=dataframe['approx_cost(for two people)']
sns.countplot(x=couple_data)
```

```
<Axes: xlabel='approx_cost(for two people)', ylabel='count'>
```

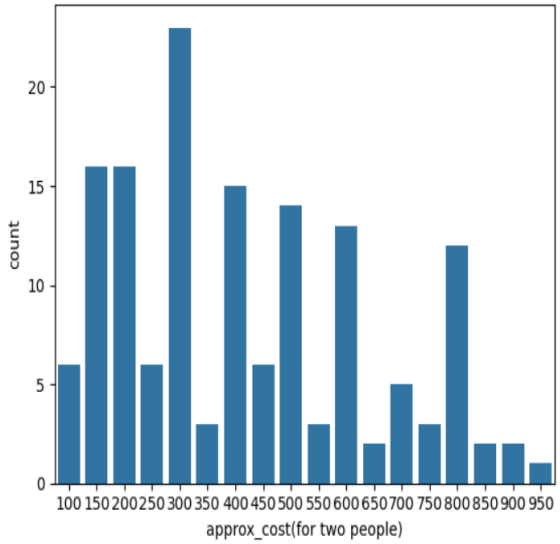

Figure 12: Exploring approx_cost(for two people) column.

Step 12: Determination of whether online orders receive higher ratings than offline orders.

```
plt.figure(figsize = (6,6))
sns.boxplot(x = 'online_order', y = 'rate', data = dataframe)
```

```
<Axes: xlabel='online_order', ylabel='rate'>
```

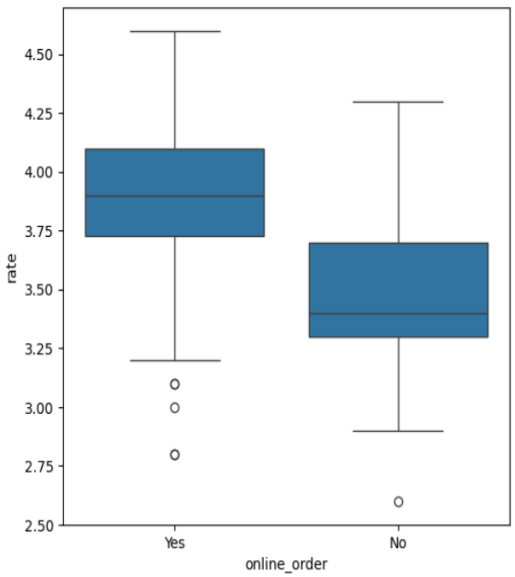

Figure 13: Determining whether online orders receive higher ratings than offline orders.

CONCLUSION

Conclusions at

Step 6: The majority of the restaurants fall into the dining category.

Step 7: Dining restaurants are preferred by a larger number of individuals.

Step 8: Empire restaurant got maximum votes.

Step 9: This suggests that a majority of the restaurants do not accept online orders.

Step 10: The majority of restaurants received ratings ranging from 3.5 to 4.

Step 11: The majority of couples prefer restaurants with an approximate cost of 300 rupees.

Step 12: Offline orders received lower ratings in comparison to online orders, which obtained excellent ratings.

Overall, Dining Dynamics- "Zomato Data Analysis Using Python" project successfully leverages Python's powerful data analysis libraries to uncover significant insights from Zomato's vast dataset. Our analysis reveals key trends and preferences in the restaurant industry, such as the predominance of online delivery services, the most popular types of restaurants, and the preferred price ranges for couples. Additionally, we identify the common rating ranges, providing a clear picture of customer satisfaction levels. These insights are invaluable for restaurant owners, marketers, and food enthusiasts, enabling them to make informed decisions and enhance their offerings. This data-driven approach not only improves customer satisfaction but also drives business growth in the competitive dining landscape.

REFERENCES

[1] Websites**:**

GeeksforGeeks, W3 schools.

[2] Python Libraries:

Requests Library Documentation: This documentation will help you understand how to make HTTP requests to the Zomato API.

Pandas Documentation: Pandas is a powerful library for data manipulation and analysis. The documentation provides detailed explanations and examples.

Matplotlib Documentation: Matplotlib is a popular library for creating visualizations in Python. The documentation covers various plotting functions and customization options.

Seaborn Documentation: Seaborn is built on top of Matplotlib and provides a high-level interface for creating attractive statistical graphics. The documentation includes tutorials and examples.

[3] GitHub Repositories:

Zomato Data Analysis: This GitHub repository contains code for analyzing Zomato restaurant data using Python and various libraries like Pandas, Matplotlib, and Seaborn.

Zomato Data Analysis Project: Another GitHub repository with code for analyzing Zomato data. It includes data cleaning, visualization, and insights.