

Zomato Data Analysis

Project Report

Table of Contents

1	PROJECT TITLE	3
2	PROBLEM STATEMENT	3
3	PROJECT DESCRIPTION	3
3.1	SCOPE OF THE WORK	4
3.2	PROJECT MODULES	7
3.3	CONTEXT DIAGRAM (HIGH LEVEL)	10
4	IMPLEMENTATION METHODOLOGY	12
5	TECHNOLOGIES TO BE USED	13
5.1	SOFTWARE PLATFORM	13
5.2	HARDWARE PLATFORM	13
5.3	TOOLS, IF ANY	13
6	ADVANTAGES OF THIS PROJECT	14
7	ASSUMPTIONS, IF ANY	14
8	FUTURE SCOPE AND FURTHER ENHANCEMENT OF THE PROJECT	14
9	DEFINITIONS, ACRONYMS, AND ABBREVIATIONS	15
10	CONCLUSION	15
11	REFERENCES	16

Appendix

A: Data Flow Diagram (DFD)

B: Screen Shots

1 Project Title

Zomato Data Analysis Using Python: Uncovering Restaurant Trends and Insights.

2 Problem Statement

We have given a zomato data set of 148 restaurant. This data set contain various attribute such as restaurant name, online_book(in which mode the customer has placed the order), book_table(Whether the customer has booked the table), rates, votes, approx_cost, listed_in(type). I have worked in a data- driven role for this data set. I have to perform EDA(Exploratory Data Analysis), visualization and draw following useful insights:

- 1) What type of restaurant do the majority of customer order from?
- 2) How many votes has each type of restaurant received from customers?
- 3) What are the ratings that the majority of restaurants have received?
- 4) Zomato has observed that most couples order most of their food online. What is their average spending on each order?
- 5) Which mode (online or offline) has received the maximum rating?
- 6) Which type of restaurant received more offline orders, so that Zomato can provide those customers with some good offers?

3 Project Description

Zomato has an average of 17.5 million monthly transacting customers for its food delivery business. The average monthly active food delivery restaurant partners on Zomato's platform have also increased by 8.7% year-on-year, from 208,000 to 226,000. We are working in a data- driven role at Zomato. We have a dataset of customers. As a data professional, we need to analyze the data, perform EDA (Exploratory Data Analysis) and visualization, and answer the following questions:

- 1) What type of restaurant do the majority of customer order from?
- 2) How many votes has each type of restaurant received from customers?
- 3) What are the ratings that the majority of restaurants have received?
- 4) Zomato has observed that most couples order most of their food online. What is their average spending on each order?
- 5) Which mode (online or offline) has received the maximum rating?

6) Which type of restaurant received more offline orders, so that Zomato can provide those customers with some good offers?

Data set of customer:

name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
Jalsa	Yes	Yes	4.1/5	775	800	Buffet
Spice Eleph	Yes	No	4.1/5	787	800	Buffet
San Churro	Yes	No	3.8/5	918	800	Buffet
Addhuri U	No	No	3.7/5	88	300	Buffet
Grand Villa	No	No	3.8/5	166	600	Buffet
Timepass I	Yes	No	3.8/5	286	600	Buffet
Rosewood	No	No	3.6/5	8	800	Buffet
Onesta	Yes	Yes	4.6/5	2556	600	Cafes
Penthouse	Yes	No	4.0/5	324	700	other

3.1 Scope of the Work

The scope of a project refers to the boundaries or limits of what will and will not be included in the project. It defines the specific objectives, deliverable, tasks, resources, timelines, and constraints that are part of the project.

1) Project Definition

Analyse Zomato restaurant data to extract insights regarding consumer preferences, restaurant performance, average spending, maximum rating, maximum votes and market trends.

2) Data Acquisition

Obtain the Zomato dataset (via API or downloadable CSV, Kaggle).Identify relevant features for analysis (e.g., restaurant ratings, cuisine types, location, votes, approx__cost , book__table , online__order,listed_in(type)).

3) Data Preparation

- Handle missing values (impute or remove).
- Remove duplicates and correct inconsistencies.
- Convert data types as needed (e.g., dates, numerical)

	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

	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	

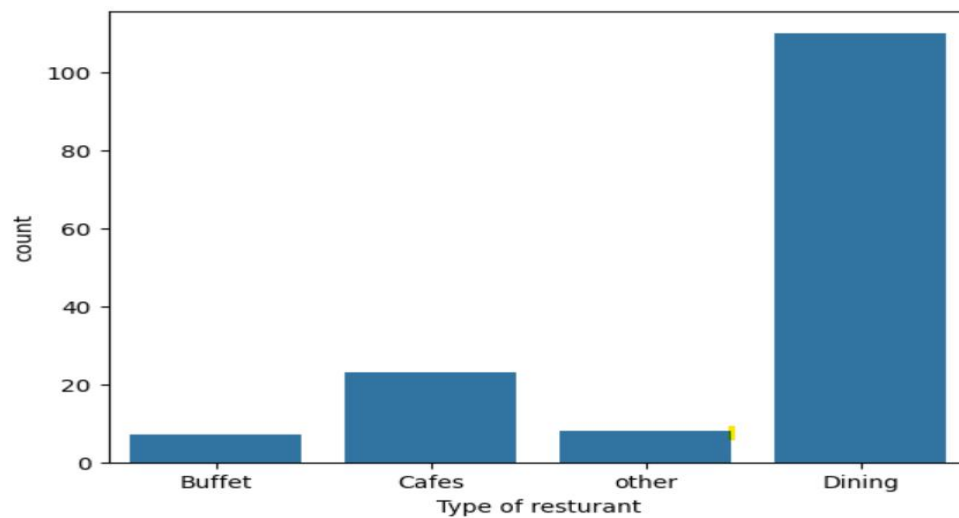
	approx_cost(for two people)	listed_in(type)
0	800	Buffet
1	800	Buffet
2	800	Buffet
3	300	Buffet
4	600	Buffet

4) Exploratory Data Analysis (EDA)

a) Summarize key statistics (mean, median, mode) for ratings and votes
i.e. `dataframe.describe()`

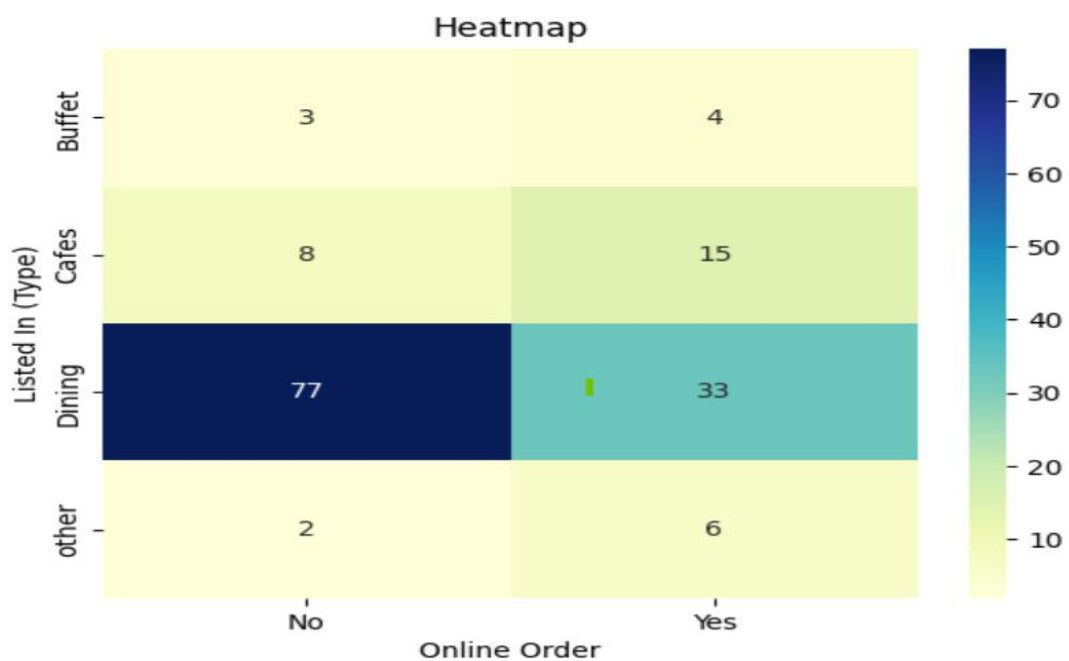
	rate	votes	approx_cost(for two people)
count	148.000000	148.000000	148.000000
mean	3.633108	264.810811	418.243243
std	0.402271	653.676951	223.085098
min	2.600000	0.000000	100.000000
25%	3.300000	6.750000	200.000000
50%	3.700000	43.500000	400.000000
75%	3.900000	221.750000	600.000000
max	4.600000	4884.000000	950.000000

b) Create visualizations (bar charts, pie charts, histograms, box plot, heat-map) to understand distributions and trends.i.e.



5) Correlation Analysis

a) Use heat maps to show correlation between features.



b) Analyse the relationship between price ranges and ratings.

c) Which mode (online or offline) has received the maximum rating.

d) Which type of restaurant has received more offline orders.

3.2 Project Modules

1) Data Collection Module

- Gather the Zomato data-set from available sources (e.g., API, CSV files, excel file).
- Identify the relevant features to include.
- Download and save the datasets for further analysis.

2) Data Preprocessing Module

- Clean and prepare the data for analysis.
- Handle missing values (imputation, removal).
- Remove duplicates and inconsistencies.
- Transform data types as necessary (e.g., converting strings to dates) e.g. 4.5/6->4.5.

	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

	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	

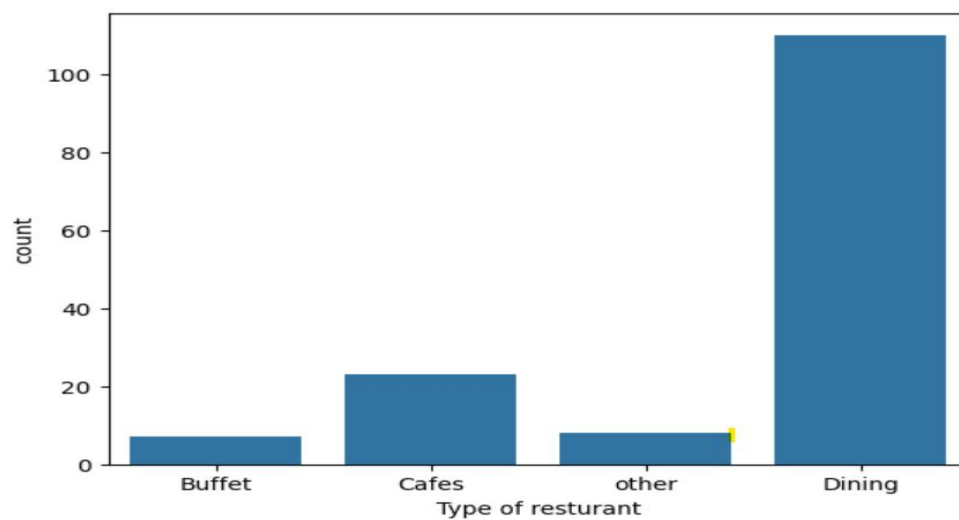
	approx_cost(for two people)	listed_in(type)
0	800	Buffet
1	800	Buffet
2	800	Buffet
3	300	Buffet
4	600	Buffet

3) Exploratory Data Analysis (EDA) Module

- a) Explore the dataset to identify patterns and insights.
- b) Generate descriptive statistics (mean, median, etc.).i.e.

	rate	votes	approx_cost(for two people)
count	148.000000	148.000000	148.000000
mean	3.633108	264.810811	418.243243
std	0.402271	653.676951	223.085098
min	2.600000	0.000000	100.000000
25%	3.300000	6.750000	200.000000
50%	3.700000	43.500000	400.000000
75%	3.900000	221.750000	600.000000
max	4.600000	4884.000000	950.000000

- c) Analyse trends in restaurant ratings, price, and type.
- d) Count the number of orders from each restaurant type.i.e.

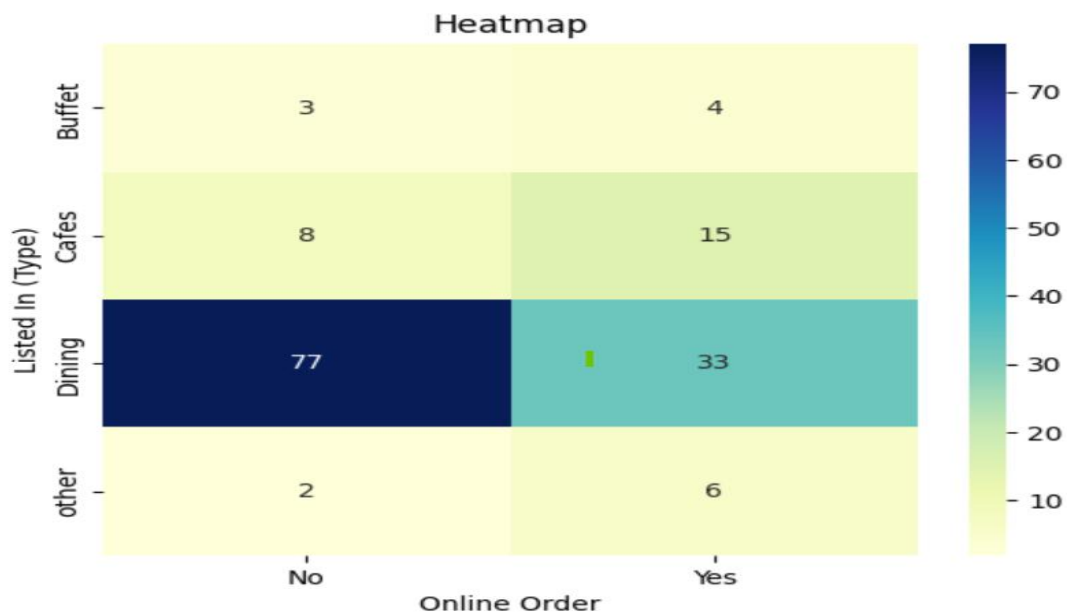


e) Count the votes for each type of restaurant.i.e.



4) Visualization Module

- Create visual representations of findings for better communication.
- Develop dashboards or reports using visualization tools (e.g., Matplotlib, Seaborn, Tableau).
- Create visualizations (histograms, bar charts, box plots) to understand data distributions.i.e.



5) Reporting Module

- Summarize and document the analysis and findings.
- Write a comprehensive report detailing methodologies, analyses, and insights.
- Highlight actionable recommendations based on the analysis.

6) Future Work Module

- a) Identify areas for further research or analysis.
- b) Suggest additional data sources for future projects.

3.3 Context Diagram (High Level)

Creating a Context Diagram for a Zomato data analysis project using Python would focus on representing the various interactions between external entities and the system that performs data analysis.

High-Level Context Diagram for Zomato Data Analysis:

1) External Entities

- a) Zomato Database/CSV/API/Kaggle:** The external source from which Zomato data is collected. This could be a database or kaggle or an API endpoint, or even CSV files containing restaurant details, votes, ratings, restaurant type , etc.
- b) Data Analyst/User:** The person who interacts with the data analysis system (you, as the user or analyst). The analyst provides inputs and receives insights or reports.
- c) Visualization Tools:** Tools like Matplotlib, Seaborn, Plotly, etc., that help generate visual insights (charts, graphs, etc.) based on the analyzed data.
- d) Report Generation System:** External systems or tools that generate reports, summaries, or dashboards based on the results of the analysis.

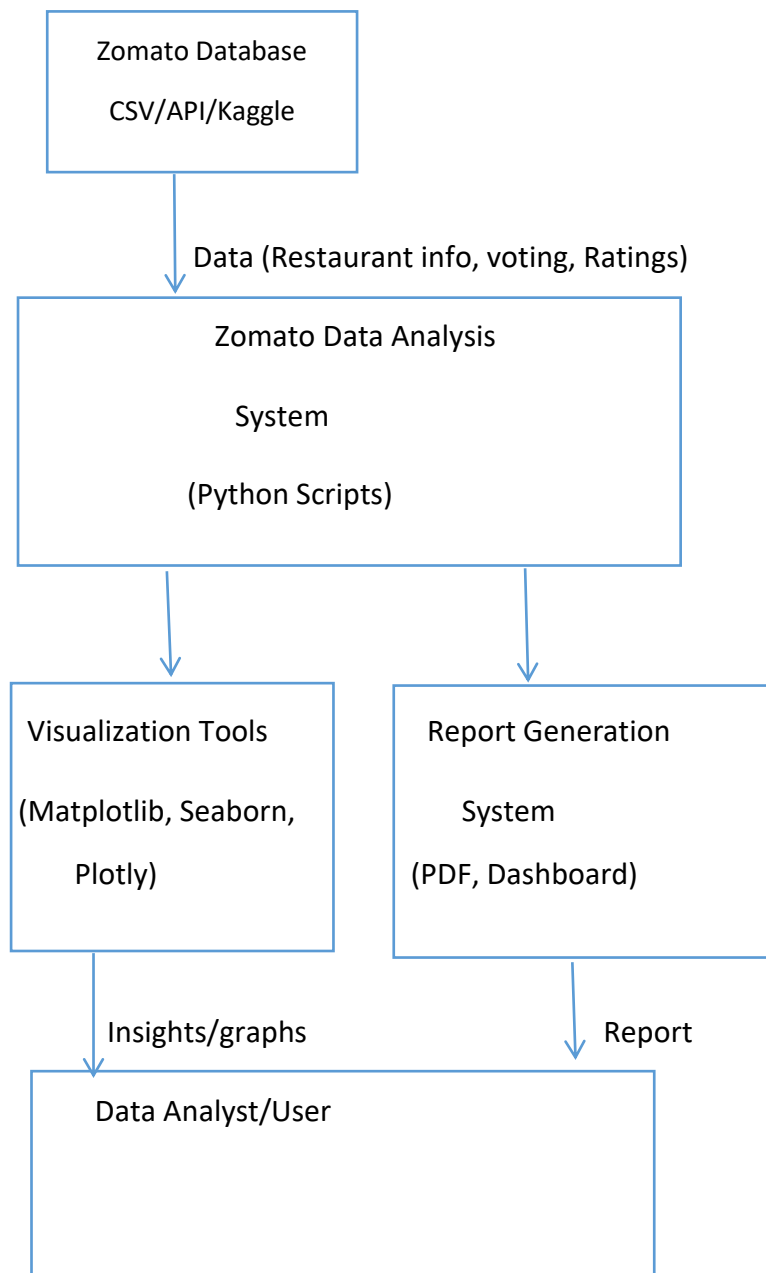
2) The System

- a) Zomato Data Analysis System (Python):** This is the central system where all the data analysis is performed, including data collection, cleaning, transformation, analysis, and reporting.

3) Data Flow

- a) Zomato Database/CSV/API → Zomato Data Analysis System:** The data is collected from kaggle or Zomato's API or a dataset (CSV/Excel). This data includes restaurant details, ratings, votes, price range, approximate cost, restaurant type etc.
- b) Zomato Data Analysis System → Data Analyst/User:** The system performs the analysis (e.g., finding top-rated restaurants, maximum vote) and provides insights or results to the user.
- c) Zomato Data Analysis System → Visualization Tools:** The data analysis system sends the processed data to visualization tools to create charts or graphs for better understanding.
- d) Visualization Tools → Data Analyst/User:** Visualization tools generate graphs (e.g., bar charts, pie charts) that are sent back to the user for easier interpretation.
- e) Zomato Data Analysis System → Report Generation System:** The system may send the results of the analysis to a report generation system that formats the findings into a formal report or dashboard.
- f) Report Generation System → Data Analyst/User:** The report generation system provides summarized insights or a detailed report.

Context Diagram:



4 Implementation Methodology

To implement a Zomato Data Analysis project using Python, several methodologies can be followed depending on the type of analysis you want to perform (e.g., exploratory data analysis (EDA), sentiment analysis, clustering, prediction). Below are the implementation methodologies that can be used to carry out such a project effectively. These methodologies will guide you through the steps required to analyze Zomato data using Python.

1) Data Collection

Zomato data can be obtained through multiple channels such as:

- a) **Zomato API:** If you're working with real-time or live data, you can use the Zomato API to retrieve restaurant details, reviews, ratings, and more.
- b) **CSV/Excel Datasets:** If you're working with a data-set available online (e.g., Kaggle), you can download CSV files containing restaurant information and reviews.
- c) **Web Scraping:** If the data isn't publicly available via an API or CSV file, you can scrape Zomato's website (or similar websites) using libraries like BeautifulSoup or Selenium.

2) Data Preprocessing

Data preprocessing is critical in any analysis, especially when working with real-world data. For Zomato data, preprocessing may include tasks such as cleaning missing values, handling duplicates, feature engineering, and converting categorical data. Use Pandas to load data from CSVs or APIs. Various tasks involve in data preprocessing are as follows:

- a) Handle missing values by imputing, removing rows/columns, or replacing with default values (e.g., mean, median).
- b) Remove duplicates.
- c) Convert categorical columns to a proper format.
- d) Create new features (e.g., restaurant ratings per type, approx_cost).

3) Exploratory Data Analysis (EDA)

EDA is the process of visually and statistically examining the data-set to summarize its main characteristics and discover patterns, anomalies, or trends.

- a) **Statistical Summaries:** Compute the basic statistics (e.g., mean, median, mode, variance) using Pandas.
- b) **Data Visualization:** Use Histograms for visualizing the distribution of numerical features like ratings, review counts, cost for two, etc.
- c) **Bar Charts:** For visualizing categorical features like restaurant type, cuisine, or city.
- d) **Box Plots:** To examine the spread and detect outliers in numerical variables.

- e) **Heat maps:** To visualize correlation matrices between different numerical features.
- f) **Correlation Analysis:** Check correlations between different numeric variables, such as price, rating, and review count.

5 Technologies to be used

5.1 Software Platform

1) Front-end

- a) **Web Frameworks:** Flask or Django (Python)
- b) **Data Visualization Libraries:** Pandas, matplotlib and seaborn
- c) **Dashboards:** Dash (by plotly) or Tableau

2) Back-end:

- a) **Programming Language:** Python
- b) **Data Manipulation:** Pandas, Numpy
- c) **Visualization:** Matplotlib, Seaborn
- d) **Machine Learning (if applicable):** Scikit-learn
- e) **Documentation:** Jupyter Notebook
- f) **Optional:** Tableau for enhanced visual storytelling.

5.2 Hardware Platform

- 1) **Processors:** AMD Ry-zen™ 3 processor
- 2) **Disk space:** 1 GB
- 3) **Min. RAM:** 128MB
- 4) **Operating systems:** Windows 11 or later, mac-OS, and Linux
- 5) **Python version:** 3.11.4

5.3 Tools, if any

- 1) Pandas
- 2) Matplotlib
- 3) Seaborn
- 4) Jupyter Notebook

6 Advantages of this Project

1) Insight into Consumer Preferences

Analyse restaurant ratings and reviews to identify trends in consumer preferences, helping businesses understand what customers value most.

2) Market Trends Identification

Discover emerging food trends and popular cuisines in specific regions, which can aid restaurant owners and marketers in strategic decision-making.

3) Data-Driven Decision Making

Provide actionable insights for restaurant owners to improve their services, menu offerings, and pricing strategies based on data rather than intuition.

4) Enhanced User Experience

Create tailored recommendations for users based on their preferences and past behaviors, improving user satisfaction and engagement on platforms like Zomato.

5) Visualizing Complex Data

Use data visualization techniques to simplify complex datasets, making it easier for stakeholders to understand and interpret results.

6) Predictive Insights

If predictive modelling is included, it can forecast future restaurant performance based on current data, helping businesses plan more effectively.

7) Research and Learning Opportunity

Offers a hands-on experience in data analysis, machine learning, and visualization, enhancing your technical skills and knowledge in these areas.

7 Assumptions, if any

No

8 Future Scope and further enhancement of the Project

1) Future scope

a) Incorporate data from social media, food delivery platforms, or customer demographics to enrich the analysis and provide more comprehensive insights.

- b) Develop a real-time dashboard that updates with live data, allowing users to see trends and changes as they happen.
- c) Create a mobile app that allows users to access the analysis and insights on the go, enhancing user engagement and accessibility.
- d) Expand predictive modeling to include various aspects, such as predicting customer footfall, revenue forecasting, or menu item success based on historical data.

2) Further Enhancement

- a) Implement features that allow users to customize their experience based on preferences (e.g., favorite cuisines, dietary restrictions).
- b) Use advanced visualization libraries (e.g., D3.js) to create more interactive and engaging visual representations of the data.
- c) Enable users to share their findings or favorite restaurants with friends via social media or within the app.

9 Definitions, Acronyms, and Abbreviations

Abbreviation	Description
EDA	Exploratory Data Analysis(Explore the data-set to identify patterns and insights.)
Approx_cost	Average spending by couple.
listed_in(type)	Type of restaurant.

10 Conclusion

The Zomato data analysis project has provided valuable insights into consumer preferences, restaurant performance, and market trends within the food industry. Through comprehensive data collection and analysis, we have achieved several key objectives such as Understanding Consumer Behavior, Identifying Trends, Data-Driven Recommendations and Visualizing Complex Data.

Overall, this project not only highlights the importance of data analysis in the food industry but also opens up numerous avenues for further research and development. By continuously refining our approach and leveraging emerging technologies, we can provide even more valuable insights to users and businesses, ultimately contributing to a more informed and responsive dining landscape.

11 References

<https://pandas.pydata.org/>

<https://matplotlib.org/>

<https://seaborn.pydata.org/>

<https://www.geeksforgeeks.org/seaborn-heatmap-a-comprehensive-guide/>

<https://www.geeksforgeeks.org/data-visualization-with-python/>

<https://www.geeksforgeeks.org/install-jupyter-notebook-in-windows/>

<https://www.w3schools.com/python/pandas/default.asp>

<https://www.python.org/downloads/>

<https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/>

<https://en.wikipedia.org/wiki/Zomato>

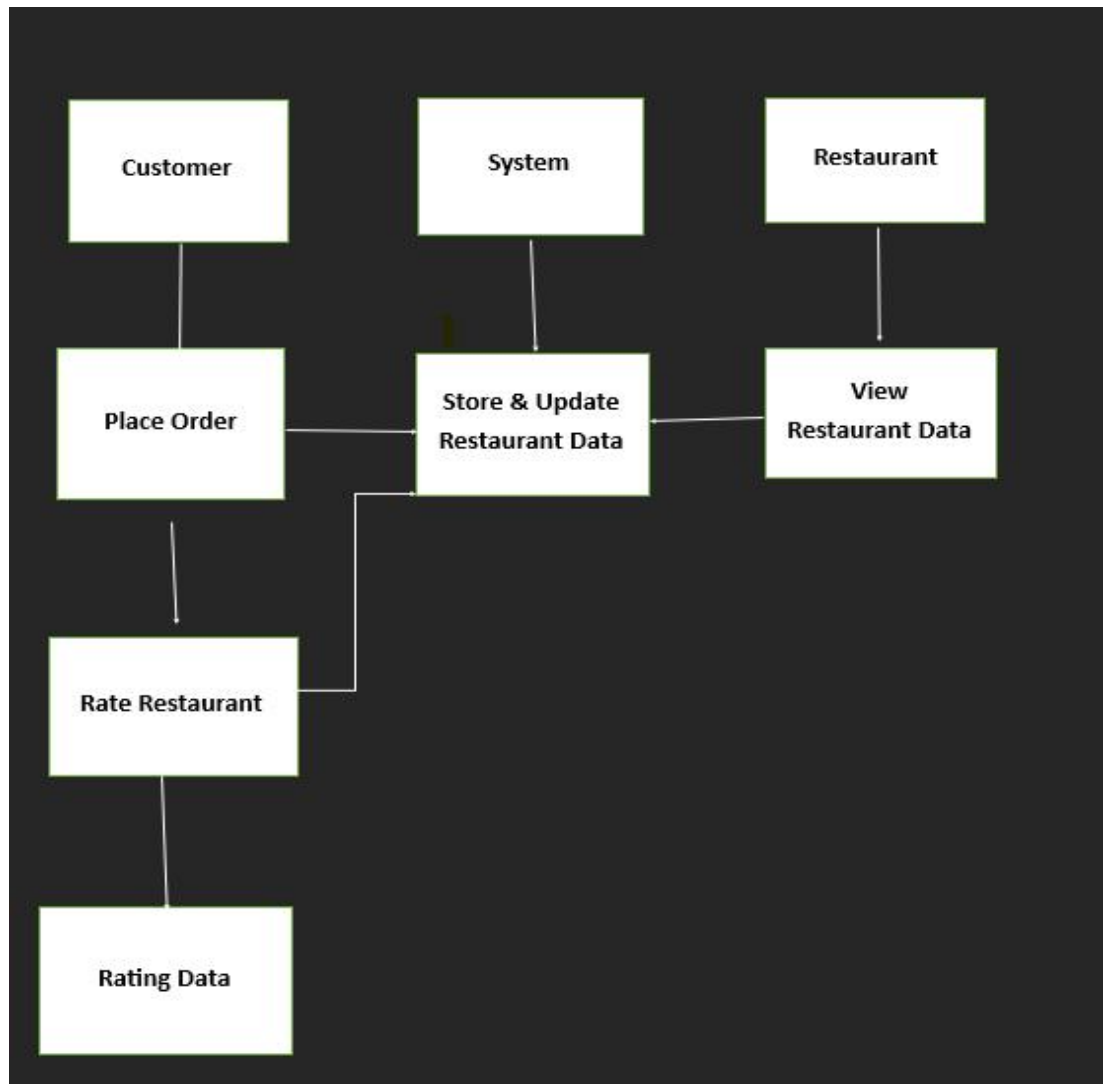
<https://www.geeksforgeeks.org/methods-of-data-collection/>

<https://www.geeksforgeeks.org/data-analysis-with-python/>

Annexure A

Data Flow Diagram (DFD)

Data Flow Diagram (DFD) is a visual representation of how data moves through a system. It shows the flow of data between different components of the system, how data is processed, and where it is stored.

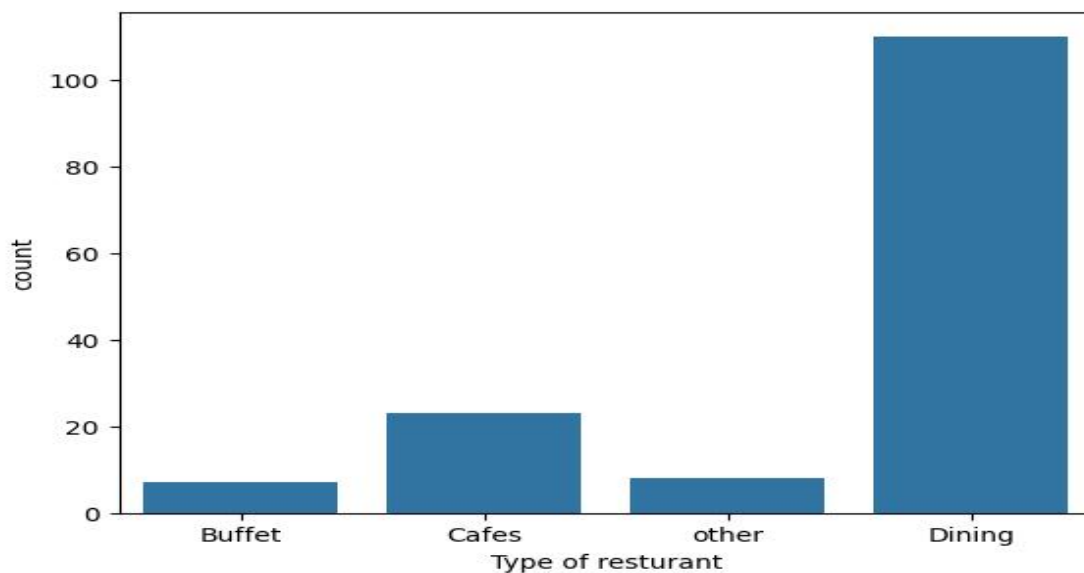


Annexure B

Screen Shots

Insights

1) What type of restaurant do the majority of customer order from?



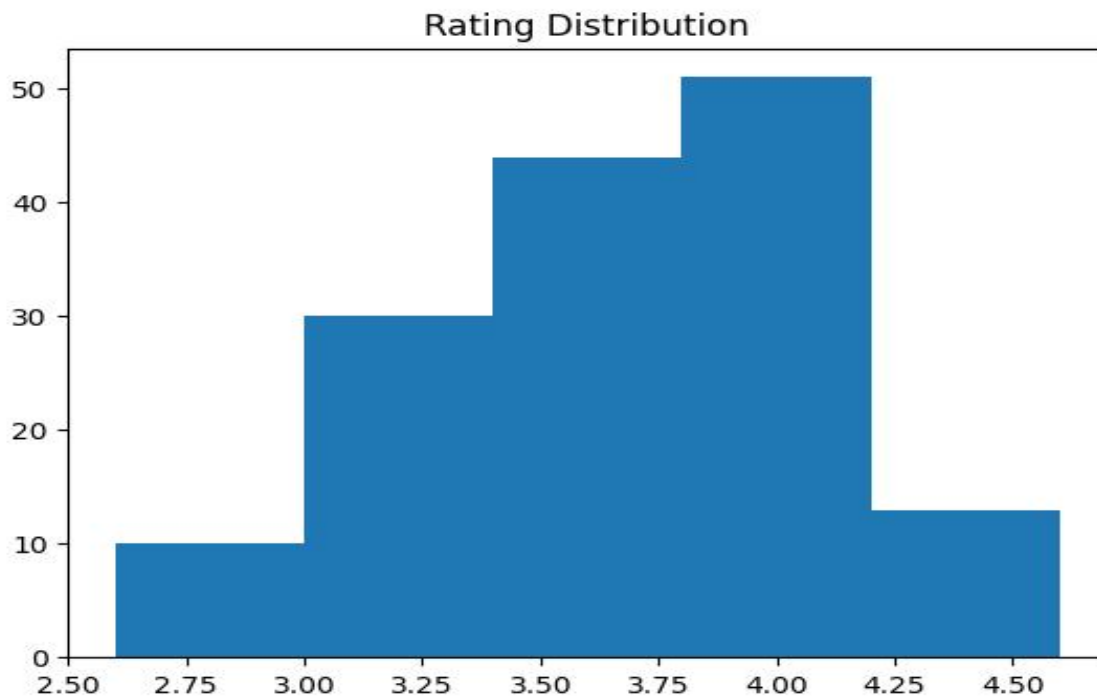
CONCLUSION: The majority of the restaurants fall into the dining category. Dining restaurants are preferred by a larger number of individuals. 📊

2) How many votes has each type of restaurant received from customers?



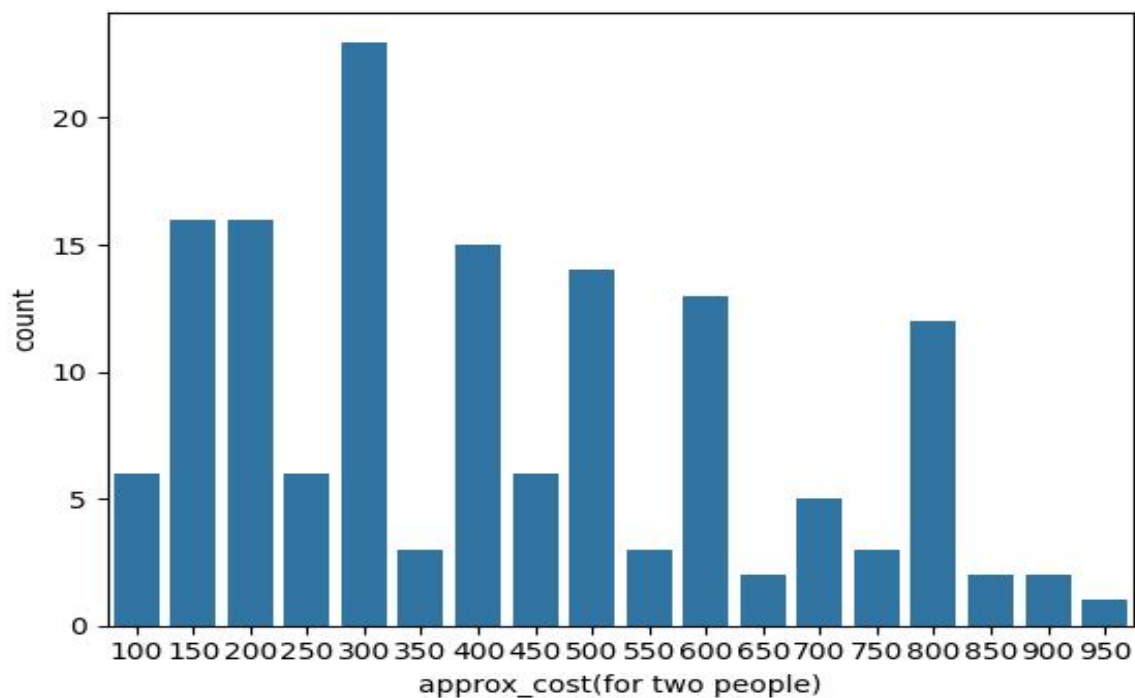
Conclusion: Dining restaurant has received maximum votes.

3) What are the ratings that the majority of restaurants have received?



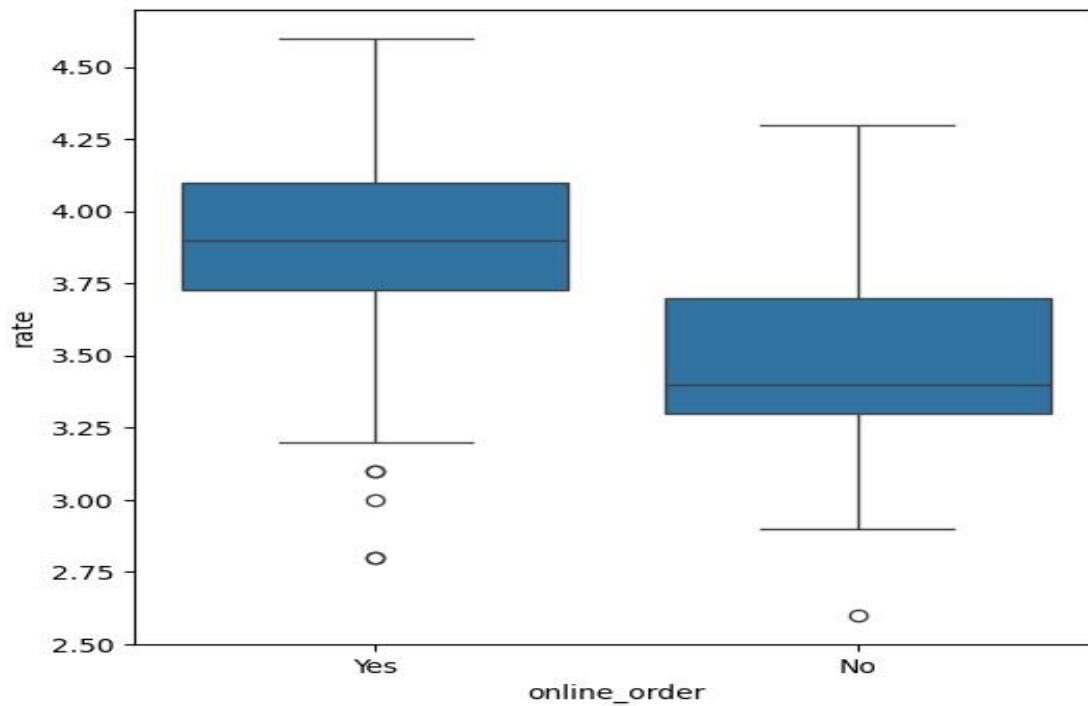
CONCLUSION: The majority of restaurants received ratings ranging from 3.5 to 4.

4) Zomato has observed that most couples order most of their food online. What is their average spending on each order?



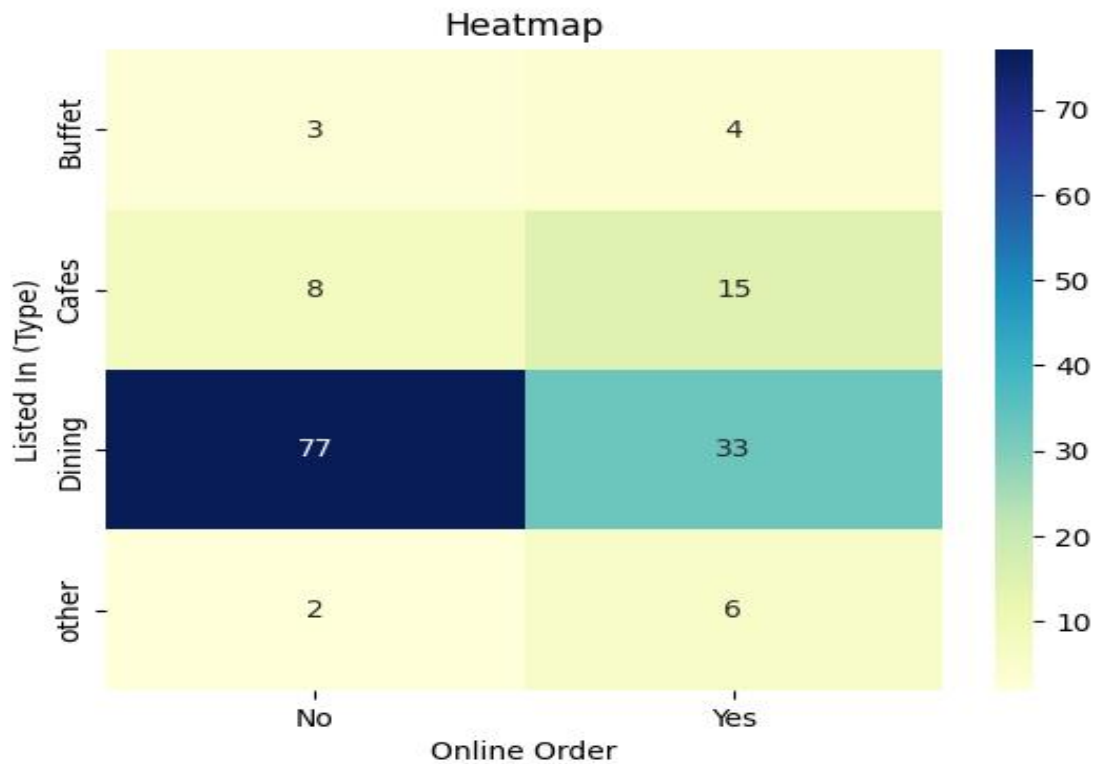
Conclusion: The majority of couples prefer restaurants with an approximate cost of 300 rupees.¶

5) Which mode (online or offline) has received the maximum rating?



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

6) Which type of restaurant received more offline orders, so that Zomato can provide those customers with some good offers?



CONCLUSION: Dining restaurants primarily accept offline orders, whereas cafes primarily receive online orders. This suggests that clients prefer to place orders in person at dining restaurant, but prefer online ordering at cafes.

In []:

