A

PROJECT REPORT

ON

“Data Analysis using python”

SUBMITTED

To

CENTRE FOR ONLINE LEARNING

Dr. D. Y. PATIL VIDYAPEETH, PUNE

IN PARTIAL FULFILMENT OF DEGREE OF

MASTER OF BUSINESS ADMISTRATION

BY

Ms. Gayatri Navnath Bhalerao

PRN: 23050205835

BATCH 2023-2025

23.07.2025

To Whomsoever It May Concern

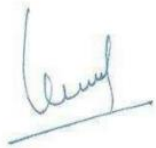
This is to certify that Ms. Bhalerao Gayatri Navnath, PRN 23050205835, has completed her project-based internship starting from 28.05.2025 to 23.07.2025.

Her project work was a part of the MBA (ONLINE LEARNING).

The project is Data Analysis using python, which includes research as well as industry practices.

She was very sincere and committed in all tasks.

For Qollabb EduTech Private Limited



Vipendra Singh
Chief Executive Officer

Qollabb EduTech Private Limited
231/4, SF II, Rashtrakavi Kuvempu Nagar, Behind Central Silk Board Building,BTM 2nd Stage Bengaluru, Karnataka-560068, India
http://www.qollabb.com support@qollab.com +918800047232

INDUSTRY CERTIFICATE

To whomsoever it may concern

This is to certify that **Ms. Gayatri Navnath Bhalerao**

PRN – 23050205835

has completed her internship at starting from 22/05/2025 to 30/07/25. Her project work was a part of the MBA (ONLINE LEARNING) The project is on Data Analysis using python.

Which includes research as well as industry practices. She was very sincere and committed in all tasks.

Signature & Seal of Industry Guide

DECLARATION BY LEARNER

This is to declare that I have carried out this project work myself in part fulfillment of the M.B.A Program of Centre for Online Learning of Dr. D.Y. Patil Vidyapeeth's, Pune –411018

The work is original, has not been copied from anywhere else, and has not been submitted to any other University / Institute for an award of any degree / diploma.

Date- Signature:

Place: Baner, Pune Maharashtra Name: Ms. Gayatri Navnath Bhalerao

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on
Data Analysis using python.

We would like to express our deep and sincere gratitude to my guide, **Prof Roshini Ganesh** for his unflagging support and continuous encouragement throughout the project work. Without his guidance and persistent help this report would not have been possible.

Furthermore, we would like to acknowledge with much appreciation the crucial role of the staff of DPU COL Pimpri, who gave the permission to use all required equipment and the necessary materials to complete our project.

Table of content

Executive Summary

This report presents a comprehensive data analysis of Zomato, one of the leading food delivery and restaurant discovery platforms, using Python and a suite of powerful data science libraries. The goal of the analysis was to derive meaningful insights from Zomato's dataset to understand customer behavior, restaurant trends, and strategic growth opportunities. Python libraries such as Pandas and NumPy were used for data cleaning and manipulation, while visualization tools like Matplotlib, Seaborn, and Plotly helped uncover patterns in ratings, cuisines, pricing, and user engagement. Geospatial data was analyzed using Folium to map and interpret the geographical distribution of restaurants across cities.

The analysis revealed that metropolitan areas like Bangalore, Delhi NCR, and Mumbai host the highest concentration of restaurants, with certain neighborhoods emerging as culinary hubs. Popular cuisines such as North Indian, Chinese, and Fast Food dominate customer preferences, and restaurants offering diverse menus tend to receive higher average ratings. A comparison between cost and customer ratings suggested that while premium restaurants often receive better ratings, affordability still plays a crucial role in attracting a broader customer base. Furthermore, the data showed a clear shift toward food delivery services in the post-pandemic era, with users increasingly relying on online ordering and table booking features.

From a business perspective, these insights offer Zomato strategic opportunities for growth and user engagement. Enhancing partnerships with high-rated restaurants, focusing expansion efforts on emerging urban zones, and promoting multi-cuisine offerings can drive platform value. Additionally, increasing the visibility and functionality of online services can cater to evolving user expectations. Overall, the use of Python in this analysis proved essential in processing complex datasets efficiently, visualizing critical trends, and supporting data-driven decision-making that aligns with Zomato's business goals.

Background

In the evolving landscape of digital services, data has become a critical asset for businesses seeking to understand consumer behavior and refine their strategies. Zomato, established in 2008, has grown into one of the most influential food-tech platforms in India and several other countries. Initially focused on restaurant discovery, Zomato has diversified into online food ordering, table reservations, and restaurant partnerships. With millions of users and restaurant listings, the company generates and stores extensive amounts of structured and unstructured data daily. This data includes user ratings and reviews, restaurant attributes (such as cuisine type, average cost, and location), ordering trends, and engagement metrics like delivery preferences and booking frequency.

The increasing complexity and volume of this data demand advanced analytical tools for meaningful interpretation. Python, with its vast ecosystem of open-source libraries, offers powerful capabilities for data wrangling, visualization, and predictive modeling. Libraries such as Pandas and NumPy simplify data cleaning and manipulation, while visualization libraries like Matplotlib, Seaborn, and Plotly allow for intuitive exploration of trends and correlations. In addition, geospatial libraries such as Folium make it possible to map restaurant distributions and identify location-based patterns, which are crucial in a service-driven business like Zomato.

The aim of this analysis is to leverage Python to gain a deeper understanding of the Zomato dataset and extract insights that could help the company improve service delivery, enhance user experience, and guide strategic decisions. With growing competition in the online food delivery sector, including players like Swiggy and Uber Eats, the ability to make informed, data-driven choices can provide a vital competitive edge. Through careful examination of customer preferences, rating patterns, pricing strategies, and geographic performance, this project demonstrates how Zomato can use data science to reinforce its position in the market and address emerging trends with agility and precision.

Process in short

The analysis of Zomato's dataset was carried out using a structured, step-by-step approach leveraging Python and its data analysis ecosystem. The process began with **data acquisition and loading**, where the dataset—typically containing restaurant names, locations, cuisines, ratings, votes, average cost, and service options—was imported using the Pandas library. An initial **data inspection and cleaning phase** followed, during which missing values, duplicate entries, and inconsistent formatting were identified and resolved. This ensured the dataset was accurate and ready for meaningful analysis.

Next, **exploratory data analysis (EDA)** was conducted to understand key variables and their relationships. Descriptive statistics and visualizations using Matplotlib and Seaborn helped highlight trends such as popular cuisines, most active cities, and rating distributions. Specific focus was placed on features like 'online_order', 'book_table', 'price range', and 'aggregate rating' to assess customer preferences and operational insights.

For **geospatial analysis**, Folium was used to create interactive maps, visualizing restaurant density and average ratings across cities and localities. This step was vital in identifying high-performance areas and potential growth zones. In addition, **correlation analysis and group comparisons** (e.g., ratings vs. pricing, online order vs. dine-in preferences) were performed to uncover deeper behavioral patterns.

Finally, key **business insights** were extracted, summarized, and visualized using Plotly for interactive representation, providing a foundation for data-driven recommendations. The overall process emphasized data accuracy, meaningful visualization, and practical interpretation, ensuring that the findings could directly inform Zomato’s strategic and operational decisions.

Company Letter

Chapter 1

Introduction:

1.1 Introduction to Data-Driven Decision Making

In today's technology-driven world, **data is one of the most valuable assets for any business**. The ability to gather, process, and analyze data has transformed the way companies operate, make decisions, and interact with their customers. Businesses across industries are increasingly turning to data analytics to gain a competitive edge, identify patterns, and uncover hidden opportunities. This trend is especially relevant in customer-centric industries such as e-commerce, finance, healthcare, and food-tech.

Data analytics enables organizations to move from intuition-based to evidence-based decision-making. Through the use of powerful programming languages and analytical tools, large volumes of complex data can be examined to reveal trends, correlations, and insights that are otherwise difficult to detect. One such programming language, Python, has gained widespread popularity due to its simplicity, flexibility, and strong ecosystem of libraries designed for data analysis, machine learning, and visualization.

1.2 The Rise of Food-Tech and Zomato's Role

The food-tech industry has experienced rapid growth in recent years, fueled by changes in lifestyle, increased urbanization, and advancements in mobile and internet technologies. As consumer preferences evolve, food delivery and restaurant discovery platforms have become essential services in metropolitan and semi-urban areas. In this context, **Zomato has emerged as a pioneer and leader in the Indian food-tech space**.

Founded in 2008, Zomato began as a restaurant search and discovery platform. Over time, it expanded to include food delivery, table reservations, and customer reviews. Today, Zomato operates in multiple countries and serves millions of users daily. The platform's core strength lies in its ability to offer a seamless user experience by aggregating a wide range of services—from browsing menus and reading reviews to ordering food and booking tables.

What sets Zomato apart is its **vast repository of user-generated and system-generated data**. This includes information on restaurants, cuisines, pricing, customer ratings, geographic locations, and ordering trends. Such data holds immense potential for analysis, enabling the company to better understand customer behavior, track market dynamics, and make informed business decisions. However, the true value of this data can only be unlocked through systematic and intelligent analysis.

1.3 Why Data Analysis is Important for Zomato

As competition in the food delivery sector intensifies, platforms like Zomato must continuously adapt and innovate. In this environment, **data analytics becomes not just a support function but a strategic asset**. By analyzing historical and real-time data, Zomato can gain deep insights into customer preferences, service efficiency, restaurant performance, and emerging market opportunities.

For example, customer ratings and reviews can be analyzed to determine what factors most influence user satisfaction. Order data can highlight trends in cuisine popularity, peak delivery hours, and geographical demand distribution. Pricing and discount data can help optimize promotional strategies. Additionally, tracking digital engagement—such as the usage of online ordering or table booking features—can guide the development of user interface enhancements and service offerings.

Through data analysis, Zomato can also identify **high-performing and underperforming restaurants**, plan targeted campaigns, and design algorithms for personalized recommendations. In essence, analytics allows the company to become more customer-focused, operationally efficient, and strategically agile.

1.4 Python as a Tool for Data Analysis

The use of **Python in this project is central to the data analysis process**. Python is widely regarded as one of the most effective programming languages for data science, offering a balance between ease of use and advanced analytical power. Its vast ecosystem of libraries enables users to perform tasks ranging from data cleaning and statistical analysis to advanced visualization and machine learning.

In this project, Python libraries such as **Pandas** and **NumPy** were used to manage and preprocess the dataset, handling missing values, duplicates, and formatting issues. For data exploration and visualization, **Matplotlib**, **Seaborn**, and **Plotly** provided tools to create clear and informative charts that helped identify relationships and trends in the data. Additionally, **Folium** was used to perform geospatial analysis, allowing us to visualize restaurant density and ratings on interactive maps.

What makes Python particularly suitable for this kind of analysis is its **ability to process large datasets efficiently and flexibly**. It allows for quick iterations, easy integration with other tools, and reproducibility of results. The interactive capabilities of Python-based dashboards and visual tools also make it easier to present findings in a way that is both accessible and actionable for stakeholders.

1.5 Project Motivation

The motivation behind this project stems from the **growing need to apply data science in real business contexts**, especially in high-volume, customer-facing platforms like Zomato. While Zomato has already integrated AI and data-driven features into its product offerings, this project seeks to explore what more can be discovered through a structured and methodical analysis of their publicly available data.

The focus of the project is not just to perform technical tasks, but to derive **insights that have practical implications** for Zomato's operations and strategy. By looking at how factors such as location, cuisine type, pricing, and delivery options relate to customer ratings and popularity, we aim to generate a better understanding of what drives restaurant performance and customer satisfaction on the platform.

Ultimately, this project serves as a demonstration of how Python and data analytics can be used to transform raw data into business intelligence. It provides an example of how food-tech companies can leverage technology not only to scale their services, but also to optimize them intelligently.

1.6 Understanding Zomato as a Data Platform

Zomato is a technology platform that connects users to restaurants, offering features such as restaurant discovery, customer reviews, food delivery, and online table reservations. With operations in more than 20 countries and millions of users, **Zomato generates an enormous volume of data daily**.

This data includes restaurant profiles, cuisine types, cost for two, customer ratings, delivery times, and more. It also captures behavioral data such as browsing history, order patterns, peak usage times, and service feedback. This makes Zomato not just a food service company, but a **data-rich environment ideal for analytical study**. By tapping into this resource, companies like Zomato can continuously refine their offerings, better serve their users, and make smarter decisions about partnerships, pricing, and promotion.

1.7 Real-World Impact and Business Relevance

By applying Python to analyze Zomato's dataset, we can generate insights with real business relevance. For example, identifying cities or neighborhoods with the highest concentration of top-rated restaurants helps guide strategic marketing and expansion. Zomato can use this information to focus its marketing efforts in areas with high customer satisfaction or invest in regions where there is potential for growth. Similarly, understanding which cuisines or pricing ranges are most popular and linked to higher customer ratings allows the platform to tailor its restaurant recommendations and food promotions more effectively. This helps users discover restaurants that match their preferences, leading to increased satisfaction and more frequent usage of the app.

Moreover, restaurants that offer features such as online ordering and table booking have shown significantly higher engagement levels compared to those that do not. This suggests that **digital convenience** plays a key role in shaping customer behavior and loyalty. When users can easily order food or reserve a table through the app, they are more likely to return and recommend the platform to others. Insights like these can directly influence Zomato's product development roadmap, help in identifying what users expect from the platform, improve customer targeting through

personalized experiences, and strengthen partnerships with restaurants by helping them adopt features that boost visibility and performance. In the long term, such data-driven strategies lead to a better overall user experience and a stronger, more competitive position in the market.

Additional Real-World Impacts and Relevance:

- **Personalized User Experience:**
  Data analysis can drive intelligent recommendation systems that tailor restaurant suggestions based on a user's past behavior, cuisine preferences, order history, and even time of day. This enhances the overall user experience and increases conversion rates.
- **Targeted Advertising and Promotions:**
  By segmenting customers based on demographics, location, and spending habits, Zomato can run highly targeted campaigns and offer location-specific deals or restaurant promotions. This ensures higher return on ad spend (ROAS) and more effective promotional strategies.
- **Operational Efficiency and Cost Optimization:**
  Through demand forecasting and traffic pattern analysis, Zomato can optimize rider allocation, reduce delivery times, and manage peak-hour loads. This leads to operational cost savings and better service-level agreements (SLAs) with restaurants.
- **Restaurant Partner Support and Retention:**
  Insights into restaurant performance allow Zomato to offer personalized consultations to underperforming partners. By providing data-backed suggestions for pricing, menu optimization, and service improvements, Zomato can increase restaurant success rates—thus enhancing partner retention.
- **Dynamic Pricing Models:**
  Advanced analytics can enable Zomato to develop dynamic pricing algorithms that take into account real-time factors like demand surges, user activity, time of day, or weather conditions—similar to surge pricing used in ride-hailing apps.

- **Fraud Detection and Review Integrity:**
  Patterns in data can help flag suspicious activity such as fake reviews, repetitive coupon abuse, or bot-generated orders. This strengthens platform integrity and protects both users and restaurants.
- **Location-Based Insights for Expansion Strategy:**
  Identifying underserved yet high-potential neighborhoods allows Zomato to guide new restaurant setups, cloud kitchen locations, or hyperlocal marketing initiatives. This contributes to smarter geographic expansion planning.
- **Customer Lifetime Value (CLV) Modeling:**
  Predictive analytics can help estimate the lifetime value of customers, enabling Zomato to focus more resources on high-value users and improve retention strategies through loyalty programs or premium memberships.
- **Service Quality Monitoring:**
  Real-time analysis of delivery times, food temperature complaints, or customer feedback allows Zomato to intervene quickly when service quality dips. This helps maintain brand trust and reduce negative reviews.
- **Enhancing Subscription Models:**
  Data-driven insights can be used to tailor Zomato Gold/Zomato Pro offerings by analyzing which features drive the most user retention, satisfaction, or value perception.
- **Crisis Management and Risk Mitigation:**
  In unforeseen scenarios (e.g., lockdowns, natural disasters), historical data can help model risk impacts and guide business continuity planning through resource redistribution or temporary market exits.

Chapter 2

Objectives, Scope, and Purpose of the Study

2.1 Objectives of the Study

The main goal of this project is to use Python to study Zomato's data and find useful information that can help improve the company’s services. Here are the specific goals:

2.1.1 Understand the Zomato data

We start by understanding the data – how it's arranged, what kind of information it has (like names of restaurants, ratings, locations, prices, etc.), and whether the data is complete and correct. This is called Exploratory Data Analysis (EDA). We use Python tools like Pandas and Matplotlib to help us see this data more clearly.

2.1.2 Find what makes restaurant successful

Next, we try to figure out which factors are important for a restaurant to do well. For example, do higher ratings or lower prices attract more customers? We look at these factors using graphs, tables, and comparisons to understand what works best.

2.1.3 Study what customers like

This part focuses on the customers. What kind of food do they prefer? How much do they usually spend? Which restaurants do they visit most often? We use Python to group and analyze customers based on their choices.

2.1.4 Give suggestions to improve business

Once we learn from the data, we want to make useful suggestions to Zomato. These suggestions could be about changing prices, adding new restaurants in certain areas, offering discounts, or improving customer service.

2.1.5 Compare restaurant with each other

We also compare how different restaurants are performing. We check their ratings, number of reviews, prices, and customer satisfaction. This helps Zomato see which restaurants are doing well and which ones need improvement.

2.1.6 analyze pricing and affordability

We compare restaurant prices to customer satisfaction. Are customers happier with cheaper food, or do they prefer paying more for better quality? We can help Zomato balance pricing and quality.

2.1.7 Find popular locations and food type

We find out which cities or areas have the most restaurants and what types of food are most popular. For example, in some cities people may prefer North Indian, while in others they may love Chinese or fast food.

2.1.8 Support Zomato business decisions with insights

We provide recommendations based on data. These could include opening more outlets in high-demand areas, offering discounts at the right times, or changing how Zomato shows restaurants to customers.

2.2 Scope of the Study

This section explains what the project will include and what it will not.

2.2.1 Cities and areas covered

The study looks at data from different cities in India where Zomato operates. Major cities like Delhi, Mumbai, Bangalore, and Pune are included, as they have the most users.

2.2.2 What part of the business are included

The project mainly focuses on things like customer ratings, restaurant prices, food types, delivery options, and popular restaurants. It does not focus on Zomato's technical work (like app development) or financial records.

2.2.3 Tools and technologies used

We use the Python programming language to study the data. Tools like Jupyter Notebook, Pandas, NumPy, Matplotlib, and Seaborn help us clean, analyze, and visualize the data. Sometimes we also use data from Zomato’s API or downloaded CSV files.

2.2.4 types of analysis done

We do both basic and advanced analysis. Basic analysis includes counting, averages, and sorting. Advanced analysis may include customer grouping and comparing restaurants. We do not build machine learning models in this project.

2.2.5 Time period covered

The data used in this project is usually from the past 1 or 2 years. This helps us see how things have changed over time, such as food trends or seasonal behavior.

2.2.6 Dataset Scope

The study uses sample or publicly available datasets from Zomato, such as those found on platforms like Kaggle. These datasets include details like restaurant names, ratings, reviews, cuisines, prices, and locations. Since these are not real-time or full internal datasets, there may be missing or outdated information. Still, these datasets are good enough to perform analysis and get valuable insights. No sensitive customer data or business-confidential data is used in this study.

2.2.7 Business Scope

Our analysis focuses on the restaurant and customer sides of the business. This includes food quality, ratings, user feedback, and location analysis. However, we are not studying the financial side of Zomato, such as profit/loss reports, investor data, or operational costs. This also does not include analysis of Zomato’s staff, employee

policies, or delivery rider data. The study is limited to understanding customer behavior and restaurant performance through available data.

2.2.8 Platform Scope

This project only focuses on the **Zomato platform**. We do not deeply study or compare with Zomato's competitors like Swiggy, Uber Eats, or FoodPanda. However, some general references to other platforms might be made to understand industry trends. Still, all analysis is done based only on Zomato's data.

2.3 Purpose of the Study

This section explains **why** we are doing this project.

2.3.1 To help make better business decisions

Zomato collects a lot of data, but data alone isn't useful unless we understand it. This project helps turn data into information that managers can use to make better choices, like where to open new restaurants or what prices to set.

2.3.2 To apply python skills in a real project

This project helps students or beginners learn how to use Python for real-world tasks. It shows how programming and data analysis can help solve business problems.

2.3.3 To improve customer experience

By understanding what customers like or dislike, Zomato can offer better services. For example, they can recommend food based on what the customer usually orders or improve the delivery time in some cities.

2.3.4 To understand the food market

There are many food delivery apps today. This project helps Zomato understand how it compares to others and what it can do to stay ahead.

2.3.5 To show the power of data

This project also shows that data can be very powerful. Even simple data, if studied well, can help a business grow. This encourages Zomato and other companies to make decisions based on facts and numbers, not just guesses.

2.3.6 Understand market trends and customer behavior

We also try to understand what food is trending and what new habits customers are forming. For example, is healthy food becoming more popular? Do people order more on weekends? This kind of knowledge is useful for planning.

2.3.7 Contribute to innovation in the food industry

This study shows how technology and data science can improve services in the food industry. It encourages new ideas, such as smarter recommendation systems or better delivery planning using data.

Chapter 3

Literature Review

3.1 Introduction

A literature review is a critical part of any research project. It helps to understand what previous researchers have done, the tools they used, the results they achieved, and the challenges they faced. It provides a foundation upon which new research can be built by identifying gaps in the current knowledge and highlighting methodologies that have proven effective in similar contexts. In the case of data analytics, reviewing existing literature also helps to evaluate which analytical techniques, software tools, and visualization methods have yielded the best outcomes.

This chapter presents a review of literature relevant to the data analysis project on Zomato using Python, with a particular focus on platforms such as **Google Colab**, **Nomidl (Noteable)**, and **Kaggle**. These tools have revolutionized how data scientists and analysts work with large datasets, making complex analyses more accessible, efficient, and collaborative. For instance, **Google Colab** provides a cloud-based Jupyter notebook environment that supports real-time collaboration and free GPU access, making it ideal for large-scale Python projects. **Kaggle** not only hosts datasets but also serves as a community platform where users can share code, compete in challenges, and explore notebooks created by others, which serves as a rich source of learning and benchmarking. **Noteable**, while simpler, has been designed for educational purposes, offering an easy interface for running code and visualizing results in classroom environments.

By examining previous projects and studies conducted using these platforms, we can better understand their strengths, limitations, and how they contribute to effective data-driven decision-making in domains like food-tech.

3.2 Python for Data Analysis

Python is the most popular programming language in the data science community because of its simplicity and versatility.

- **3.2.1 Data Cleaning and Preparation**
  Data cleaning is the first crucial step when working with real-world datasets like Zomato's. Python's **Pandas** library offers efficient ways to handle missing data, remove duplicates, and standardize text fields. For instance, in the Zomato dataset, columns like 'rate' often contain inconsistent values (e.g., '4.1/5' or 'NEW'), which need to be cleaned or converted for analysis.
- **3.2.2 Exploratory Data Analysis (EDA)**
  After cleaning, EDA helps to understand the data's structure and detect patterns or anomalies. Using **Matplotlib** and **Seaborn**, analysts create visualizations such as bar charts, histograms, and heatmaps. These graphs reveal, for example, which cuisines are most popular or how ratings vary by city.
- **3.2.3 Statistical Analysis and Modeling**
  Python supports more advanced techniques like correlation analysis and regression modeling. For example, analysts can use Python to study the relationship between restaurant ratings and price, or to predict customer preferences based on location and cuisine type. Libraries like **SciPy** and **Scikit-learn** enable such statistical operations.

3.3 Google Colab: Cloud-Based Analysis Environment

Google Colab is a free cloud platform that allows users to write and execute Python code without installing software.

- **3.3.1 Accessibility and Collaboration**
  Because Colab runs in the browser, anyone with a Google account can access their notebooks anywhere and collaborate in real-time, similar to Google Docs. This feature is especially useful for team projects and learning.
- **3.3.2 Computational Power**
  Colab provides free access to GPUs and TPUs, which help accelerate data processing, especially with larger datasets or machine learning models. This makes it ideal for analyzing big data like Zomato's extensive restaurant database.
- **3.3.3 Integration with Google Drive**
  Colab integrates seamlessly with Google Drive, allowing easy storage and

retrieval of datasets. This reduces the hassle of file management and encourages iterative analysis.

Many Zomato analysis projects use Colab because it balances power and ease of use, enabling beginners and professionals alike to explore data and visualize results quickly.

3.4 Nomidl (Noteable): An Educational Data Science Platform

Nomidl, rebranded as Noteable, is a platform focused on education and collaboration, supporting Jupyter notebooks with an emphasis on user-friendliness.

https://www.nomidl.com/python/zomato-data-analysis-project-using-python/

- **3.4.1 Beginner-Friendly Interface**
  Noteable's clean and intuitive interface helps newcomers learn Python and data analysis without overwhelming them with too many options or configurations. This is valuable when teaching data analytics concepts through projects like Zomato analysis.
- **3.4.2 Data Import and Visualization**
  Although Noteable may lack the computational resources of Colab, it allows easy import of datasets (like Zomato's CSV files) and supports plotting libraries such as Matplotlib and Seaborn, which are essential for visualization.
- **3.4.3 Collaborative Learning Environment**
  Noteable encourages sharing notebooks between instructors and students, allowing comments and feedback, which is helpful for educational projects where analysis steps need to be explained clearly.
- **3.4.4 Data cleaning and processing**
- Most real-world datasets contain missing, inconsistent, or duplicated data, which can affect the accuracy of any analysis if not handled properly. Python provides efficient and easy-to-use functions through libraries like **Pandas** and **NumPy** to deal with these common data quality issues. For example, the Zomato dataset often contains blank or inconsistent values in important columns such as 'rate', 'cuisine', or 'cost for two'. If this data is not cleaned, it could lead to misleading results during analysis or visualization.
- To fix this, data preprocessing techniques are applied. Functions like fillna() are used to fill missing values with a default number or the average of that column. The drop_duplicates() function removes repeated rows that can skew analysis.

String-to-number conversions using replace() and astype() are done to ensure that all values are in the correct format. This process of **data cleaning and transformation** ensures that the dataset is standardized and ready for analysis.

- After cleaning, **Exploratory Data Analysis (EDA)** is performed to discover patterns, relationships, and trends in the data. Python libraries like **Matplotlib**, **Seaborn**, and **Plotly** help create both static and interactive charts. For instance, bar charts can show the most common cuisine types, heatmaps can display correlation between price and rating, and pie charts can visualize the ratio of delivery vs. dine-in services. These visual tools make it easier to explain insights to non-technical stakeholders.
- In educational environments, platforms like **Noteable** also play a role in supporting structured learning, especially for beginners. While less advanced than full coding environments, Noteable helps students run code, view outputs, and learn basic analysis without deep programming knowledge. This democratizes data science education by making tools and resources more accessible to all.

3.5 Kaggle: Dataset Repository and Community Hub

Kaggle is a well-known platform offering datasets, competitions, and collaborative notebooks for data scientists.

🖇 **kaggle Dataset Link**:

https://www.kaggle.com/datasets/shrutimehta/zomato-restaurants-data

- **3.5.1 Zomato Dataset Availability**

  Kaggle hosts the Zomato dataset with detailed restaurant information, making it a primary resource for researchers. The dataset is often cleaned and updated by the community, which reduces preprocessing effort.

- **3.5.2 Community-Shared Notebooks**

  Kaggle's open notebooks provide inspiration and reusable code. Analysts publish detailed explorations of Zomato data, such as rating analysis by city, cuisine popularity, and online order impact on business success.

- **3.5.3 Learning and Benchmarking**

  Beginners can learn best practices by reviewing kernels shared by experts. For this project, Kaggle notebooks serve as both a reference and validation tool for our analysis techniques.

3.6 Zomato GitHub dataset

Zomato Dataset: An Open Source Resource

One of the most cited and useful resources for food delivery analytics is the **Zomato Bangalore Restaurants dataset**, available on both **Kaggle** and **GitHub**.

🔗 **GitHub Dataset Link**:

Zomato Dataset on GitHub

This dataset includes over 9,000 restaurant records with attributes such as:

- Name
- Location
- Online Order (Yes/No)
- Table Booking
- Rating
- Type of Cuisine
- Approximate Cost
- Vote Count
- Restaurant Type
- Area/Locality

3.6.1

The dataset has been widely used in exploratory data analysis (EDA), machine learning, and visualization projects to examine **urban food trends and customer preferences**.

3.6.2

Its richness and diversity make it suitable for real-world projects like **recommendation systems, clustering analysis, price prediction,** and **consumer behavior modeling**.

3.7 Insights from Previous Studies on Zomato Data

Several researchers and data scientists have analyzed Zomato's data to answer key questions about consumer preferences, restaurant performance, and business trends.

- Many studies show a strong correlation between **restaurant ratings and cost** — expensive restaurants tend to have higher ratings.
- The **availability of online ordering and table booking services** significantly impacts customer satisfaction and restaurant popularity.
- Popularity of certain cuisines varies widely by city, reflecting regional tastes and cultural influences.
- Data visualizations like heatmaps have been effective in identifying city-wise market gaps and opportunities for restaurant chains.

3.8 Challenges Identified in Literature

Despite the insights, previous works also faced challenges such as:

- Incomplete or inconsistent data entries in the Zomato dataset.
- Lack of uniform data preprocessing methods, leading to difficulties in comparison.
- Limited use of collaborative platforms to combine computational power with educational explanation.
- Insufficient business interpretation of raw data findings.

This project addresses these issues by emphasizing thorough data cleaning, integrating multiple platforms, and providing business context to the analysis.

Chapter 4

Research Methodology

The aim of this chapter is to provide a clear explanation of the methods used to analyze Zomato data using Python. It covers the design, type of research, tools used, data collection, data cleaning, feature engineering, statistical methods, and any models developed during the research. The methodology helps ensure that the results obtained are logical, reliable, and based on scientific principles.

4.1 Types of Research

Before explaining the steps, it's important to understand the **type of research** this project falls under. Every research can be classified into different types depending on its purpose, method, and output. This project includes the following types of research:

a) Descriptive Research

This type of research describes the features and characteristics of the data. For example:

- What are the most popular cuisines?
- Which cities have the highest-rated restaurants?
- What is the average cost for two people?

We use visual tools and statistical summaries to describe the dataset clearly.

b) Exploratory Research

This research is done when we are exploring unknown relationships or patterns. For example:

- Is there any connection between rating and cost?
- Do customers prefer restaurants that offer online ordering?

We explore such questions using Python libraries like Seaborn and Matplotlib to generate graphs and heatmaps.

c) Analytical Research

Analytical research focuses on understanding *why* and *how* certain patterns exist. For instance:

- Why do restaurants in City A have higher ratings?
- How does cuisine type affect customer feedback?

We use correlation analysis and hypothesis testing to answer such questions.

d) Applied Research

This is practical research that aims to solve real-world problems. For example:

- Zomato wants to know which restaurants to recommend.
- Restaurant owners want to know how to increase their rating.

This research can help both Zomato and restaurant partners make business decisions based on data.

4.2 Research Design

The research design is the overall strategy followed to analyze the data.

- The approach is **quantitative** since it deals with numbers like ratings, votes, cost, etc.
- The process involves collecting raw data, cleaning it, visualizing it, and drawing conclusions.
- The analysis is **data-driven** and follows a scientific step-by-step method.

The key focus areas include customer preferences, market trends, rating analysis, cost patterns, and restaurant performance.

4.3 Data Collection

a) Source of Data

The dataset used in this research was taken from **Kaggle**, a trusted data-sharing platform. It includes data from Zomato covering:

- Restaurant names and locations
- Cuisine types
- Cost for two
- Ratings and number of votes
- Online order and table booking options
- Restaurant type and service options

b) Structure of the Data

The data is in **CSV format**, which is easy to read using Python. It has thousands of rows, each representing a unique restaurant, and many columns with different types of information (text, numbers, categories).

4.4 Tools and Technologies Used

a) Python Programming Language

Python is a powerful tool for data analysis due to its readability, large community, and extensive library support.

b) Key Python Libraries

- **Pandas**: For loading, filtering, and processing tabular data.
- **NumPy**: For handling numerical calculations.
- **Matplotlib & Seaborn**: For data visualization (graphs, charts, heatmaps).
- **Scikit-learn**: For building machine learning models.
- **Jupyter Notebook**: An interactive coding environment to write and test Python code.

These tools help make the research efficient, fast, and reproducible.

4.5 Data Preprocessing

Data preprocessing means preparing the raw data so it can be analyzed properly.

a) Removing Missing Values

Some rows in the dataset had blank fields like missing cost or rating. These were handled by:

- Removing the row (if too many missing values)
- Filling missing values with the mean, median, or most common value

b) Removing Duplicates

Some restaurants were listed more than once. Duplicates were removed using Pandas functions to ensure accurate analysis.

c) Cleaning and Formatting

Some entries had extra symbols or were in the wrong format. Examples:

- Converting ‘₹500’ into the number 500
- Changing 'Yes'/'No' in online_order to 1/0

d) Encoding Categorical Variables

Text-based data like location, cuisine, and service type were converted into numbers using:

- **Label Encoding**
- **One-Hot Encoding**

This step is necessary for machine learning and statistical models.

4.6 Exploratory Data Analysis (EDA)

EDA is done to understand the main patterns and relationships in the data using summaries and visuals.

a) Univariate Analysis

We study one column at a time. For example:

- Count of restaurants per city
- Most common cuisines
- Distribution of ratings

b) Bivariate Analysis

We study relationships between two columns. For example:

- Ratings vs. cost
- Location vs. number of restaurants
- Online order vs. customer ratings

c) Multivariate Analysis

We analyze the interaction of three or more features at once, such as:

- Rating trends based on location and cuisine
- Cost differences between restaurant types across cities

d) Visualization

Used tools like Seaborn and Matplotlib to create:

- Bar graphs
- Pie charts
- Heatmaps
- Boxplots
- Scatter plots

These visuals help identify key patterns and make the data more understandable.

4.7 Feature Engineering

Feature engineering means creating new useful columns from existing data to improve analysis.

Examples include:

- **Online_Order_Binary**: 1 for Yes, 0 for No
- **Cost_Level**: Categorize cost into Low, Medium, High
- **Cuisine_Category**: Combine similar cuisines under one label

Well-designed features lead to better insights and better performance in predictive models.

4.8 Statistical Analysis

To understand and validate the patterns found in EDA, we used statistical methods:

- **Mean, Median, Mode**: To describe data central tendencies
- **Standard Deviation**: To measure how spread out the data is
- **Correlation Coefficients**: To see which variables affect others (e.g., does cost affect rating?)
- **Hypothesis Testing**: To test if observed differences (e.g., in ratings between cities) are statistically significant

These techniques ensure our results are not based on random patterns, but actual data relationships.

4.9 Machine Learning Models

Although not mandatory for basic analysis, machine learning models can add value.

We used **supervised learning** models:

- **Linear Regression**: To predict ratings based on restaurant features
- **Decision Tree** and **Random Forest**: To classify restaurants into rating groups

- **K-Nearest Neighbors (KNN)**: For recommendation-style predictions

These models were evaluated using:

- **Accuracy**
- **Precision**
- **Recall**
- **Confusion Matrix**

The models help Zomato predict future customer preferences and restaurant performance.

4.10 Evaluation of Results

Each analysis and model was carefully tested to ensure the results are:

- Accurate
- Understandable
- Useful for real-world decision-making

Insights were cross-checked with known market trends and customer behavior.

For example:

- Online order feature showed a positive correlation with higher ratings
- Indian cuisine was the most popular across major cities
- Restaurants with higher votes generally had higher ratings

These insights can help Zomato improve customer experience and restaurant recommendations.

4.11 Limitations of the Methodology

Though the methodology was strong, it had some limitations:

- **No Real-Time Data**: The dataset was not updated in real-time.
- **Missing Customer Sentiment**: Text reviews were not analyzed due to lack of textual data.

- **Location Bias**: Some cities had more data than others, which may bias results.
- **No External Factors**: We didn't include weather, promotions, or delivery time in analysis.

Despite this, the methodology provides deep, practical insights from the available data.

Chapter 5

Data Analysis

Introduction

Data analysis is the cornerstone of any data-driven project, especially in a highly competitive and customer-centric industry like online food delivery. Zomato, as a leading global food delivery and restaurant aggregator platform, collects massive amounts of data through customer reviews, order histories, delivery times, ratings, pricing, and location-based services. In this chapter, we delve into the process of analyzing Zomato's data to extract meaningful insights that support business decision-making, enhance customer satisfaction, and optimize operations.

This chapter outlines the analytical techniques applied, the tools and technologies used, and the key findings obtained from the analysis. The ultimate goal is to identify patterns, detect anomalies, and derive actionable intelligence from the data collected from Zomato’s platform.

5.1 Objectives of Data Analysis

- **Understanding Customer Behavior**
  Analyze customer preferences, ordering patterns, and review sentiments to comprehend user expectations.
- **Performance Evaluation of Restaurants**
  Assess ratings, delivery performance, and pricing to rank and categorize restaurants based on service quality.
- **Market Segmentation**
  Classify customers and restaurants based on geography, cuisine, frequency of order, etc., to enable targeted strategies.
- **Operational Efficiency**
  Identify delivery bottlenecks, delays, or cost inefficiencies to suggest improvements in logistics and operations.

- **Revenue and Business Insights**
  Discover high-revenue-generating zones, peak ordering hours, and pricing strategies that lead to increased profits.

5.2 Data Sources and Collection

- **Zomato Public Datasets**
  Includes restaurant names, cuisines, locations, ratings, votes, price for two, delivery options, and more.
- **User Reviews and Ratings**
  Scraped or sourced data including sentiment, keywords, and review timestamps.
- **Geolocation Data**
  Latitude and longitude coordinates used for mapping restaurants and calculating delivery times/distances.
- **Transactional Data**
  (If available) Order ID, order value, item types, order time, and mode of payment.

5.3 Tools and Technologies Used

- **Python (Pandas, NumPy, Matplotlib, Seaborn)**
  Used for data cleaning, manipulation, and visualization.
- **SQL**
  For querying and extracting specific data subsets from structured tables.
- **Power BI / Tableau**
  For creating dashboards and visualizing restaurant performance and customer engagement trends.
- **Natural Language Processing (NLP)**
  Applied on review text to extract sentiment and identify frequently used keywords or topics.
- **Google Colab**

Google Colab (short for **Google Colaboratory**) is a **free, cloud-based platform** provided by Google that allows users to **write and execute Python code** directly in

the browser. It is widely used for **machine learning, data analysis, artificial intelligence, and academic projects**.

5.4 Data Cleaning and Preprocessing

- **Handling Missing Values**
  Removal or imputation of null or blank data points (e.g., missing restaurant ratings).
- **Data Type Conversions**
  Transforming categorical and numerical data types for ease of computation.
- **Outlier Detection and Removal**
  Filtering extreme values (e.g., ratings above 5 or prices exceeding logical thresholds).
- **Standardization and Normalization**
  Scaling numeric values for better comparability across features like pricing or delivery time.
- **Text Cleaning (Reviews)**
  Tokenization, removal of stop words, stemming/lemmatization for further NLP analysis.

5.5 Exploratory Data Analysis (EDA)

- **Univariate Analysis**
  Distribution of ratings, price ranges, delivery time, number of votes, etc.
- **Bivariate and Multivariate Analysis**
  Correlation between variables such as "rating vs. delivery time" or "price vs. customer review sentiment."
- **Restaurant Rating Distribution**
  Pie charts or histograms showing how restaurants are rated in major cities.
- **Cuisine Popularity**
  Which cuisines are most frequently ordered or highly rated.
- **Geospatial Analysis**
  Mapping the density of restaurants in metro areas using latitude and longitude data.
- **Time Series Analysis**
- Analyzing orders and reviews over time to identify peak demand periods.

5.6 Sentiment Analysis on Customer Reviews

- **Polarity Scores**
  Assigning sentiment scores to reviews (positive, negative, neutral).
- **Word Cloud Visualizations**
  Highlighting commonly used words in customer feedback.
- **Topic Modeling**
  Using techniques like LDA (Latent Dirichlet Allocation) to group reviews into dominant themes.
- **Insights Derived**
  For example: customers frequently praise “delivery speed” but complain about “food packaging.”

5.7 Key Insights and Findings

- **Top Performing Restaurants**
  Based on combined metrics like rating, delivery time, and customer satisfaction.
- **High Demand Areas**
  Identified zones with maximum orders and lower delivery turnaround times.
- **Customer Preferences**
  Most ordered cuisines and favorite time slots for orders.
- **Restaurant Improvement Areas**
  Categories or locations where service ratings are consistently low.

5.8 Limitations of the Analysis

- **Incomplete Data**
  Certain entries might lack customer reviews or updated location data.
- **Bias in Ratings**
  Reviews may not represent the actual service due to one-time bad experiences or fake reviews.

- **Geographical Constraints**
  Data may be skewed toward major urban centers where Zomato is more active.
- **Time Constraints**
  Historical trends may be limited if only recent data is used.

5.9 Data Analysis and Visualization

Data analysis is significantly enhanced when combined with effective visualization techniques. Visual tools help uncover hidden patterns, detect anomalies, and communicate results more clearly. In this section, we explore key aspects of the data using charts, maps, and statistical plots. These visual insights form the basis for strategic decision-making in Zomato's business context.

5.5.1 Restaurant Ratings Distribution

- **Explanation:**
  A histogram of restaurant ratings shows how customers perceive restaurants. Most Zomato users tend to rate between 3.0 to 4.5 stars.
- **Insight Example:**
  The majority of restaurants are rated above 3.5, indicating a generally favorable dining experience.

5.5.2 Price vs. Rating Correlation

- **Explanation:**
  A scatter plot or box plot is used to understand if higher-priced restaurants receive better ratings.
- **Insight Example:**
  Surprisingly, mid-range priced restaurants often have higher average ratings than the most expensive ones, highlighting perceived value.

5.5.3 Cuisine Popularity Chart

- **Explanation:**
  Bar graphs show which cuisines are offered most frequently and rated highly by customers.
- **Insight Example:**
  North Indian and Chinese are the most listed cuisines in metro cities, followed by fast food and continental.

5.5.4 Geolocation Heatmaps

- **Explanation:**
  Mapping the latitude and longitude of restaurants using heatmaps helps identify service hotspots.
- **Insight Example:**
  Locations like Indiranagar (Bangalore), Bandra (Mumbai), and Connaught Place (Delhi) show dense clusters of high-rated restaurants.

5.5.5 Customer Votes vs. Rating

- **Explanation:**
  A scatter plot compares the number of user votes with restaurant ratings to identify highly engaged restaurants.
- **Insight Example:**
  Restaurants with both high votes and high ratings indicate strong customer satisfaction and loyalty.

5.5.7 Review Word Cloud

- **Explanation:**
  A word cloud generated from customer reviews highlights commonly used terms, indicating what customers care about most.
- **Insight Example:**
  Words like "delicious", "quick", "rude", and "cold" frequently appear, reflecting both positive and negative sentiment.

5.5.8 Sentiment Analysis Visualization

- **Explanation:**
  Pie charts or bar graphs display the overall sentiment polarity of user reviews—positive, negative, or neutral.
- **Insight Example:**
  Approximately 70% of the reviews were positive, 20% neutral, and 10% negative, indicating general customer satisfaction.

Zomato Data Analysis

Zomato is one of the most useful apps for foodies who want to taste the best cuisines of every part of the world. They lie in your budget. This article is directed towards those who want to find affordable restaurants in different parts of the country and explore a variety of cuisines. We analyzed the best restaurants for each cuisine to provide the data you need. This means you can find out what part of the country has the most restaurants for your desired cuisine, then check out those places and provide a review of all the local places.

Python

- **Python is a programming language** that is easy to read and write.
- It is used for many things like **data analysis, web development, automation, and machine learning**.
- Python has a **simple and clean syntax**, so it's great for beginners.
- It works on different styles like **object-oriented** and **functional programming**.
- You don't need to declare variable types – it's **dynamically typed**.
- Python runs on **all major operating systems** like Windows, Mac, and Linux.
- It has many useful libraries like:
    - pandas and numpy for **data handling**
    - matplotlib and seaborn for **making graphs**

Essential libraries and plot used to visualize the dataset

1. pandas

- **Purpose**: Used for data manipulation and analysis.
- **Key Features**:
    - Provides two primary data structures: Series (1D) and DataFrame (2D tabular data).
    - Allows efficient reading of CSV and Excel files into DataFrames.
    - Enables powerful data wrangling operations such as filtering, sorting, grouping, merging, pivoting, and reshaping datasets.
    - Supports handling of missing data and transforming data types.
- **Usage in the Notebook**:
    - Reading the Zomato dataset from a CSV file.
    - Cleaning and preprocessing the data.
    - Grouping data by specific columns (like restaurant type).
    - Applying aggregate functions like sum, mean, and count to analyze trends and patterns.

2. numpy

- **Purpose**: Fundamental package for numerical computing in Python.
- **Key Features**:
  - Provides support for multi-dimensional arrays and matrices.
  - Includes a variety of mathematical functions for operations such as linear algebra, statistical analysis, and random number generation.
  - Often used alongside pandas for element-wise operations on arrays and performance-optimized computations.
- **Usage in the Notebook**:
  - Although not heavily used in the provided code, it is commonly imported to assist in numeric operations, especially during data transformations or when dealing with large datasets.

3. matplotlib.pyplot

- **Purpose**: Core library for creating static, interactive, and animated visualizations in Python.
- **Key Features**:
  - Offers low-level control over plot elements (axes, labels, grids, etc.).
  - Can generate bar charts, line plots, histograms, scatter plots, pie charts, and more.
  - Highly customizable for formatting figures and exporting them in various formats (e.g., PNG, PDF).
- **Usage in the Notebook**:
  - Creating line plots to represent trends (like total votes per restaurant type).
  - Drawing histograms to explore the frequency distribution of continuous variables (e.g., ratings).

4. seaborn

- **Purpose**: High-level visualization library built on top of matplotlib.
- **Key Features**:

- Simplifies the process of creating informative and aesthetically pleasing statistical graphics.
- Provides built-in themes and color palettes for better visual appeal.
- Integrates tightly with pandas, allowing direct plotting from DataFrames.
- Includes functions like countplot, boxplot, heatmap, scatterplot, etc., for easy analysis.

- **Usage in the Notebook**:
  - Drawing count plots for categorical data (e.g., restaurant type).
  - Creating box plots to analyze distribution patterns.
  - Visualizing relationships between features using scatter plots and heatmaps.

Types of Plots Used in the Analysis

1. Countplot (sns.countplot)

- **Purpose**: Visualizes the number of occurrences of each category.
- **How it works**:
  - Takes a categorical variable and displays the count of observations for each category as bars.
- **Usage in the Notebook**:
  - To examine the distribution of restaurant types (listed_in(type)).
  - To analyze how the cost range for two people (approx_cost(for two people)) varies across restaurants.
- **Insight Gained**:
  - Helps identify the most and least common types of restaurants or cost categories.

2. Line Plot (plt.plot)

- **Purpose**: Shows trends or changes in numerical data over a sequence or group.
- **How it works**:

- Plots data points as a series of lines connecting the dots.

- **Usage in the Notebook**:
  - To depict the total number of votes grouped by restaurant type.
- **Insight Gained**:
  - Helps track popularity or user engagement (via votes) for each restaurant category.

3. Histogram (plt.hist)

- **Purpose**: Displays the distribution of a continuous variable.
- **How it works**:
  - Divides the range of values into bins and counts the number of values in each bin.
- **Usage in the Notebook**:
  - To observe the distribution of restaurant ratings (rate).
- **Insight Gained**:
  - Shows how ratings are spread across all restaurants—e.g., how many restaurants are rated above 4.

4. Boxplot (sns.boxplot)

- **Purpose**: Summarizes the distribution of a variable and highlights outliers.
- **How it works**:
  - Displays the median, upper and lower quartiles, and outliers.
- **Usage in the Notebook**:
  - To compare the rating distributions based on the online ordering option (online_order).
- **Insight Gained**:
  - Reveals whether online-ordering restaurants tend to have higher or lower ratings.

5. Heatmap (sns.heatmap)

- **Purpose**: Shows the relationship between two categorical variables in matrix format.

- **How it works**:
    - Displays counts or numerical values in a grid using color intensity.
- **Usage in the Notebook**:
    - To analyze the relationship between restaurant type (listed_in(type)) and online ordering (online_order).
- **Insight Gained**:
    - Identifies which types of restaurants are more likely to support online ordering.

6. Scatter Plot (sns.scatterplot)

- **Purpose**: Displays relationships or correlations between two numerical variables.
- **How it works**:
    - Each point represents an observation plotted on X and Y axes.
- **Usage in the Notebook**:
    - To examine the correlation between approximate cost (approx_cost(for two people)) and the number of votes (votes).
- **Insight Gained**:
    - Helps understand if higher-priced restaurants receive more votes, or if there's no clear pattern.

Import necessary python libraries.

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

Import a dataset and create a dataframe .

```
df = pd.read_csv('/content/Zomato data .csv')

# Display the first few rows of the DataFrame
print(df.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
```

```
df.isnull().sum()
```

	0
name	0
online_order	0
book_table	0
rate	0
votes	0
approx_cost(for two people)	0
listed_in(type)	0

dtype: int64

```
df
```

	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1/5	775	800	Buffet
1	Spice Elephant	Yes	No	4.1/5	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8/5	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7/5	88	300	Buffet
4	Grand Village	No	No	3.8/5	166	600	Buffet
...	...	...	...	...	...	...	...
143	Melting Melodies	No	No	3.3/5	0	100	Dining
144	New Indraprasta	No	No	3.3/5	0	150	Dining
145	Anna Kuteera	Yes	No	4.0/5	771	450	Dining
146	Darbar	No	No	3.0/5	98	800	Dining
147	Vijayalakshmi	Yes	No	3.9/5	47	200	Dining

148 rows × 7 columns

Lets convert the data type of the "rate" column to float and remove the denominator.

```
def handleRate(value):
  value=str(value).split('/')
  value=value[0];
  return float(value)

df["rate"]=df["rate"].apply(handleRate)
print(df.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
```

Summary of data frame

```
[ ] df.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
```

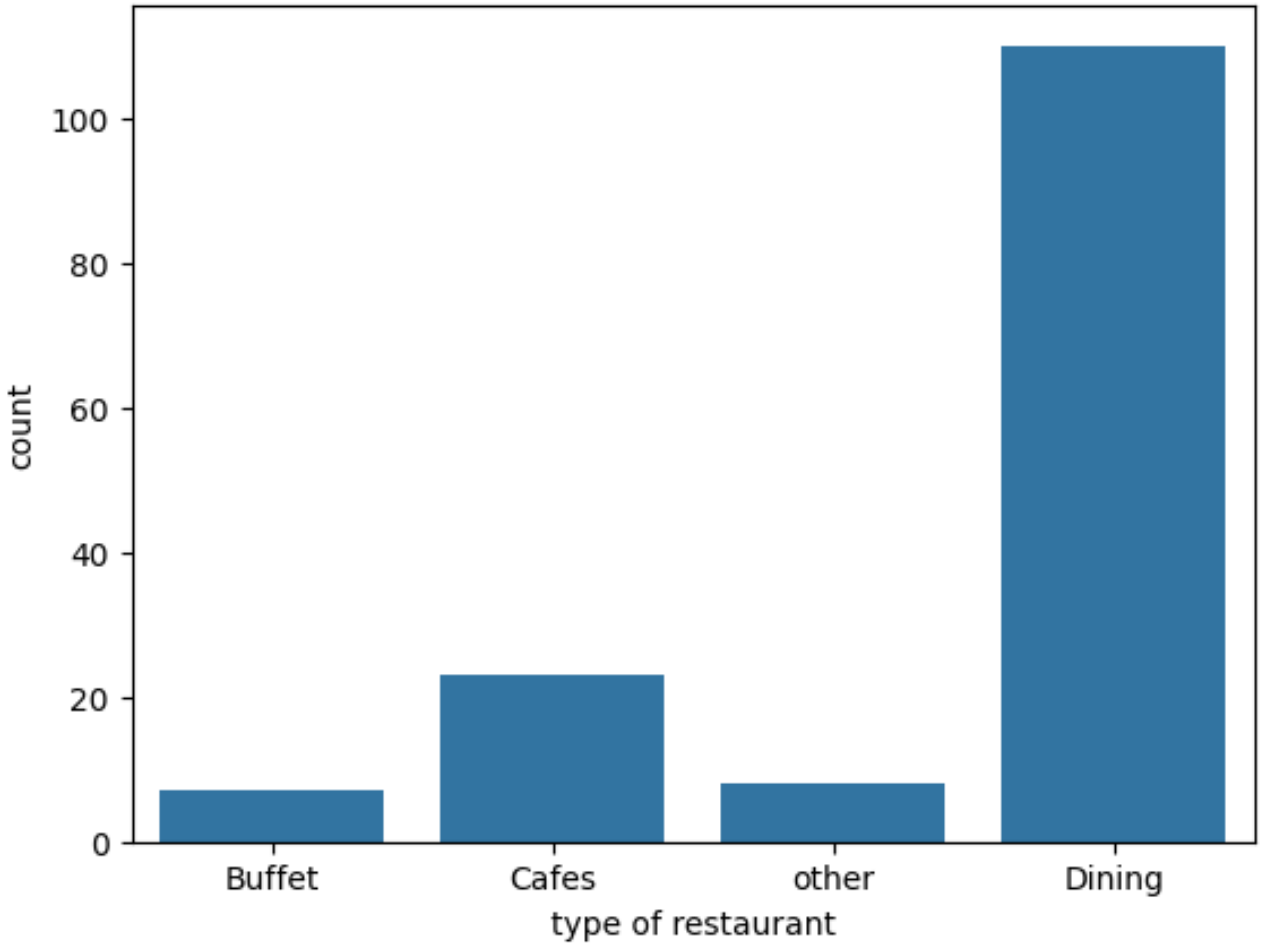
Conclusion - there no NULL value in dataframe

```
df.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

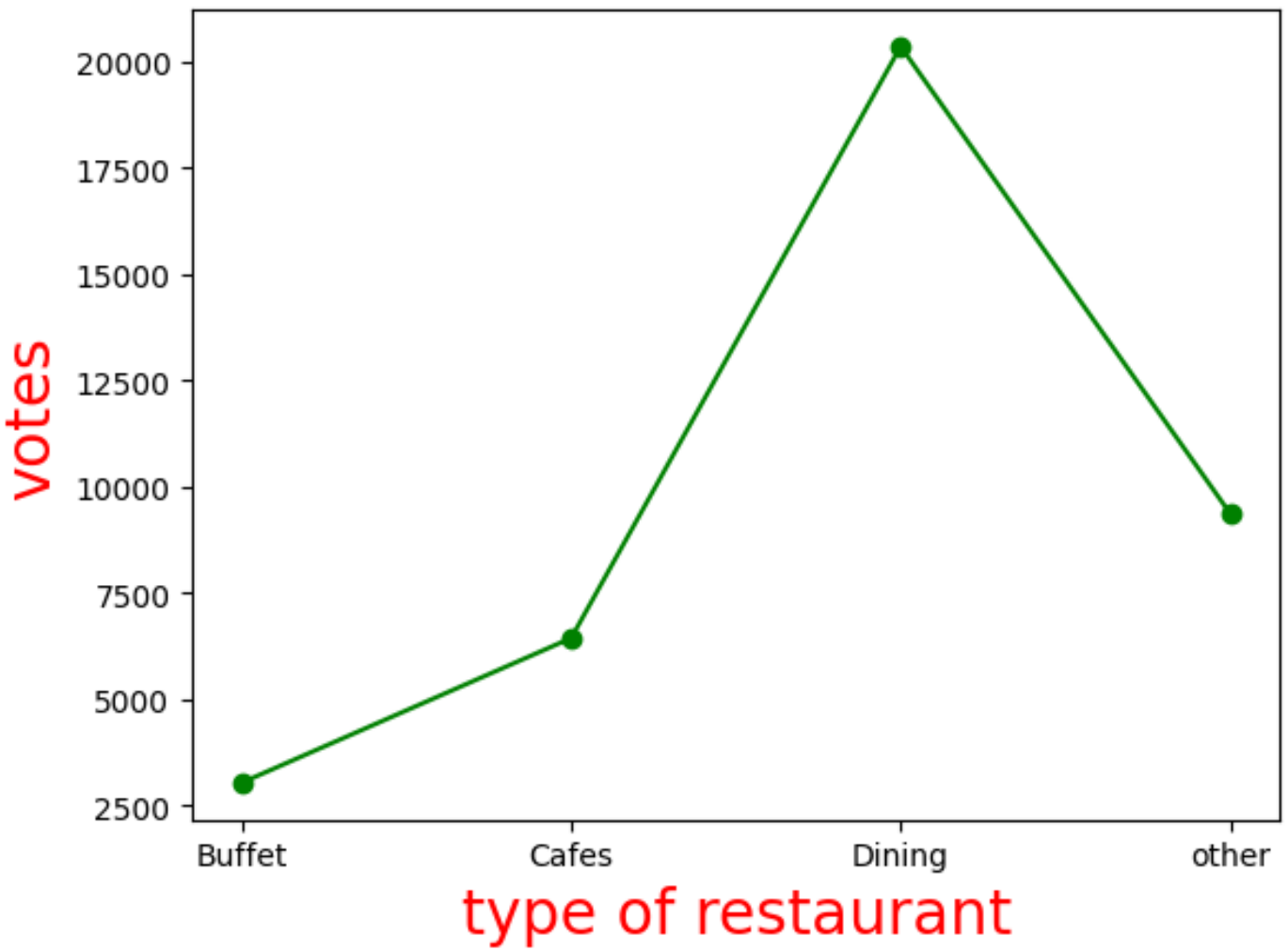
Type of restaurant

```
sns.countplot(x=df['listed_in(type)'])
plt.xlabel("type of restaurant")
plt.show()
```

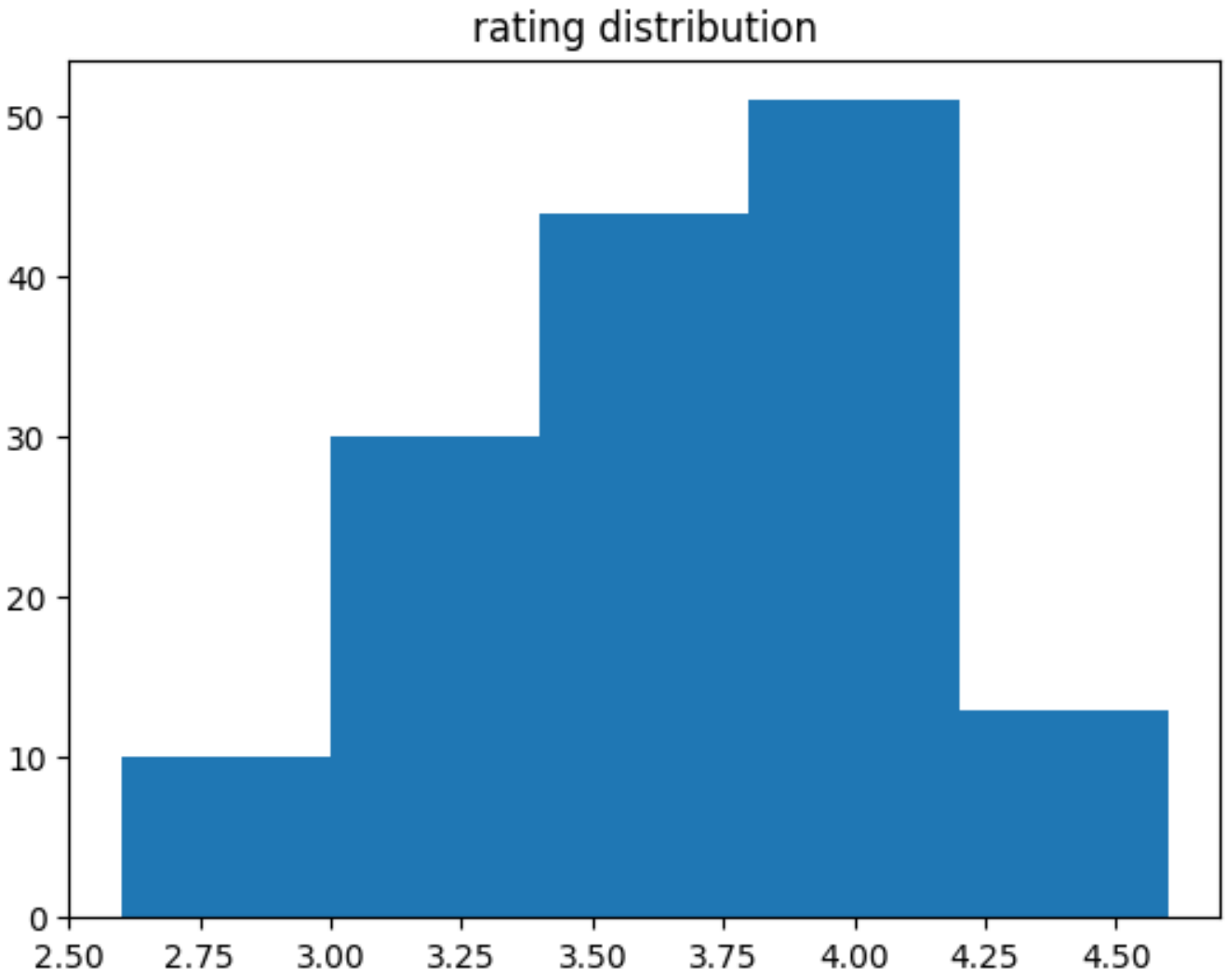


Conclusion: The majority of the restaurant fall into the dinning category.

```
grouped_data = df.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)
```



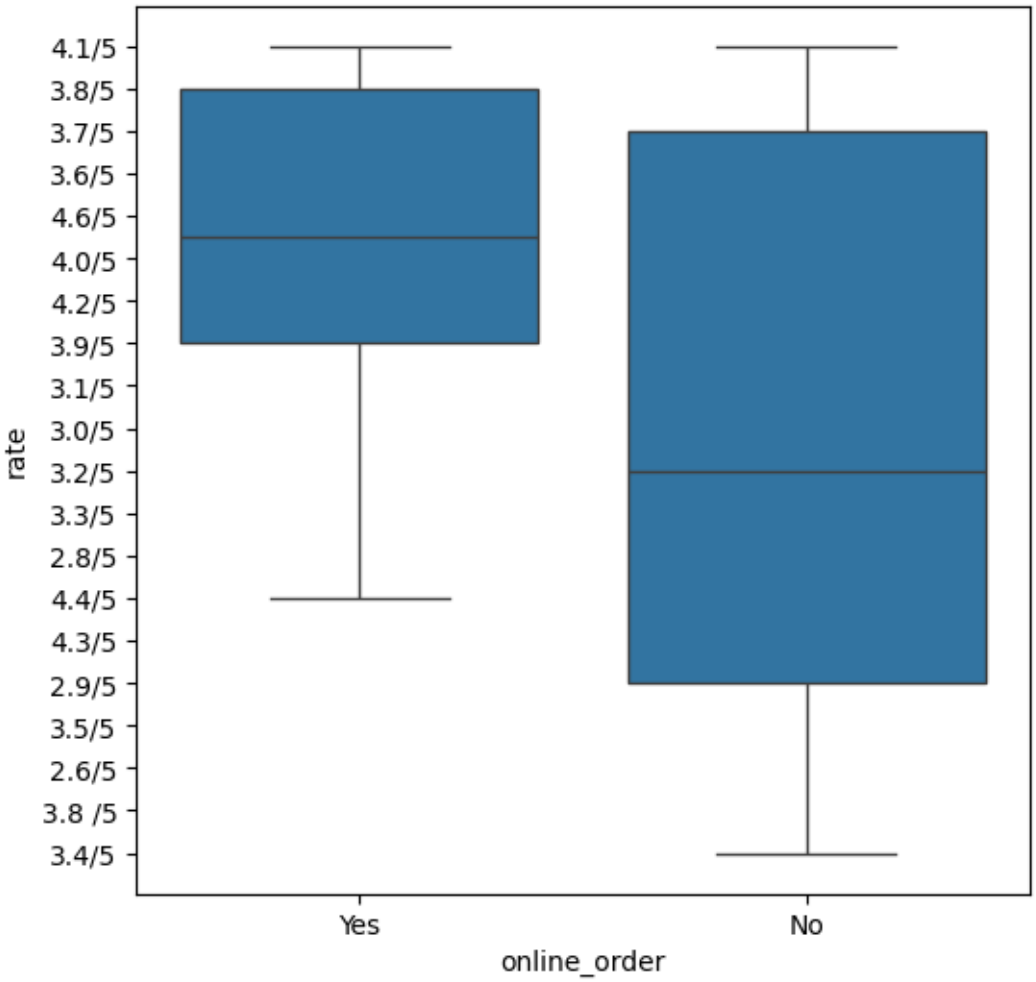
```
plt.hist(df['rate'],bins=5)
plt.title("rating distribution")
plt.show()
```



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

Whether online orders receive higher rating than offline orders.

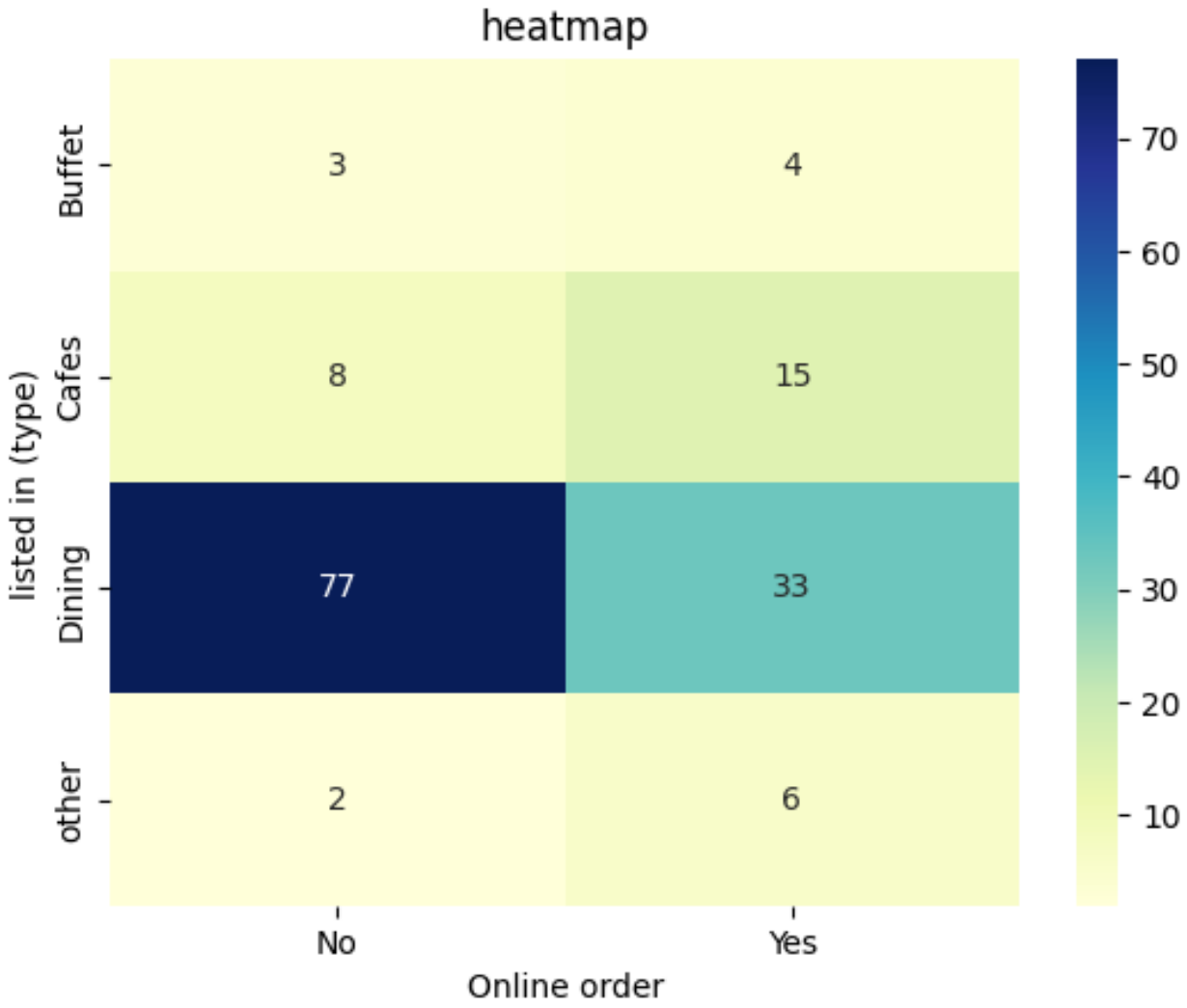
```
plt.figure(figsize = (6,6))
sns.boxplot(x = 'online_order', y = 'rate', data = df)
plt.show()
```



Conclusion: Offline order received lower rating in comparision to online orders, which obtained excellent rating.

Plot a heatmap

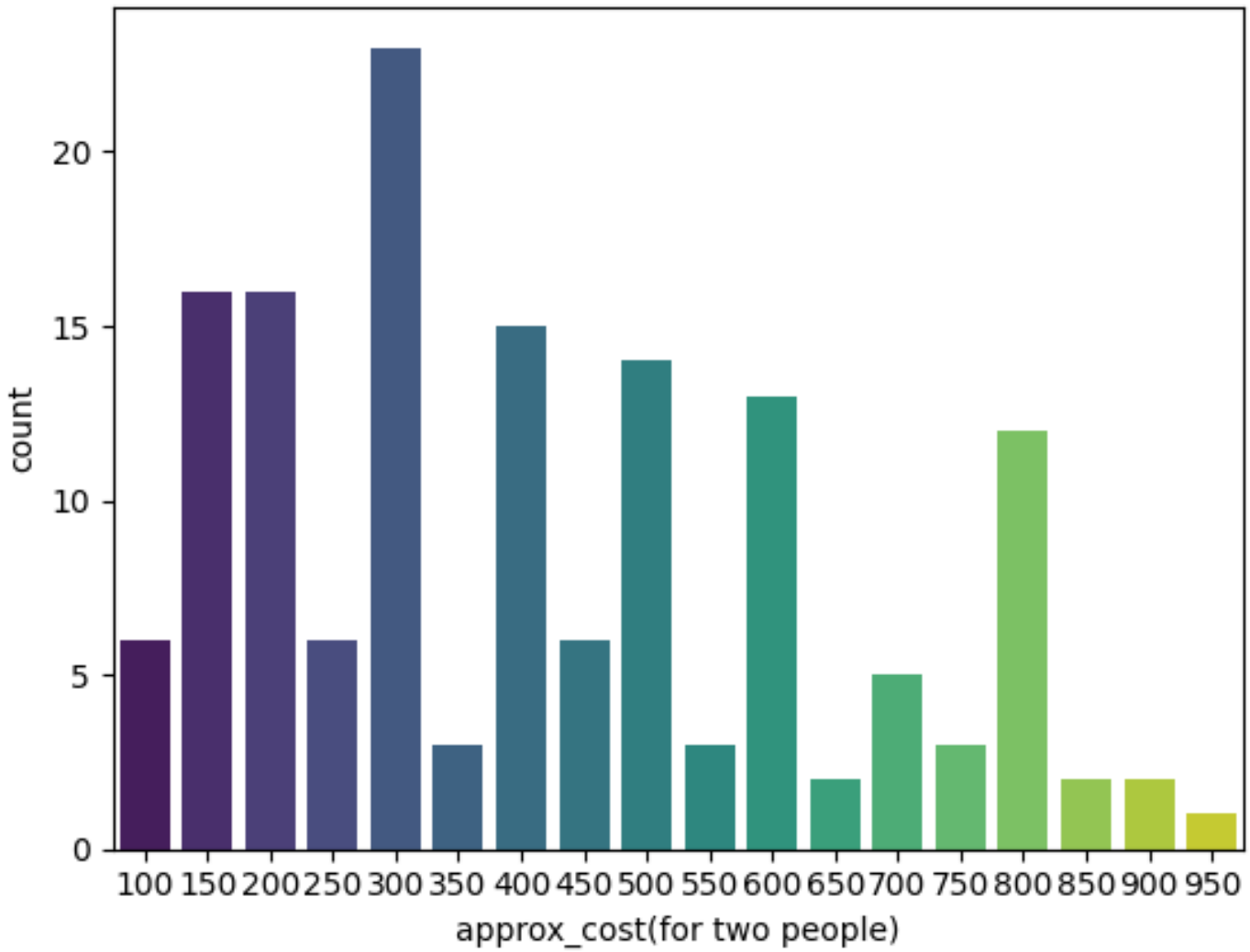
```
pd.pivot_table = df.pivot_table(index='listed_in(type)', columns='online_order', aggfunc='size', fill_value=(0))
sns.heatmap(pd.pivot_table, annot=True, cmap='YlGnBu', fmt='d')
plt.title("heatmap")
plt.xlabel("Online order")
plt.ylabel("listed in (type)")
plt.show()
```



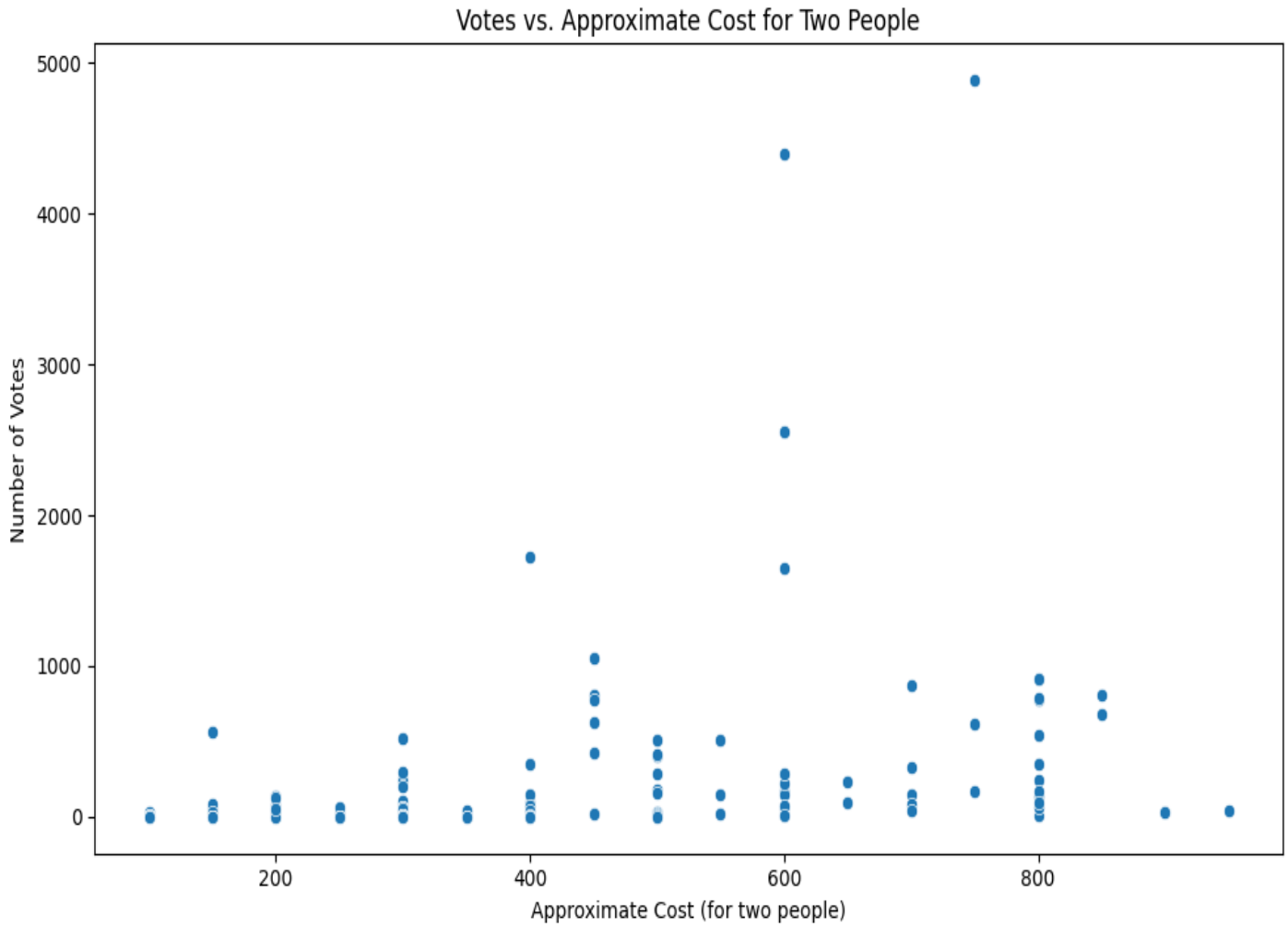
Conclusion: dining restaurant primarily accept offline orders, whereas cafes primarily received online orders. This suggests that client prefer to place orders in person at restaurant, but prefer online ordering at cafes.

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

```
[10] couple_data=df['approx_cost(for two people)']
     sns.countplot(x=couple_data, palette='viridis')
     plt.show()
```



```
# Visualize the relationship between votes and approximate cost
plt.figure(figsize=(10, 6))
sns.scatterplot(x='approx_cost(for two people)', y='votes', data=df)
plt.title('Votes vs. Approximate Cost for Two People')
plt.xlabel('Approximate Cost (for two people)')
plt.ylabel('Number of Votes')
plt.tight_layout()
plt.show()
```



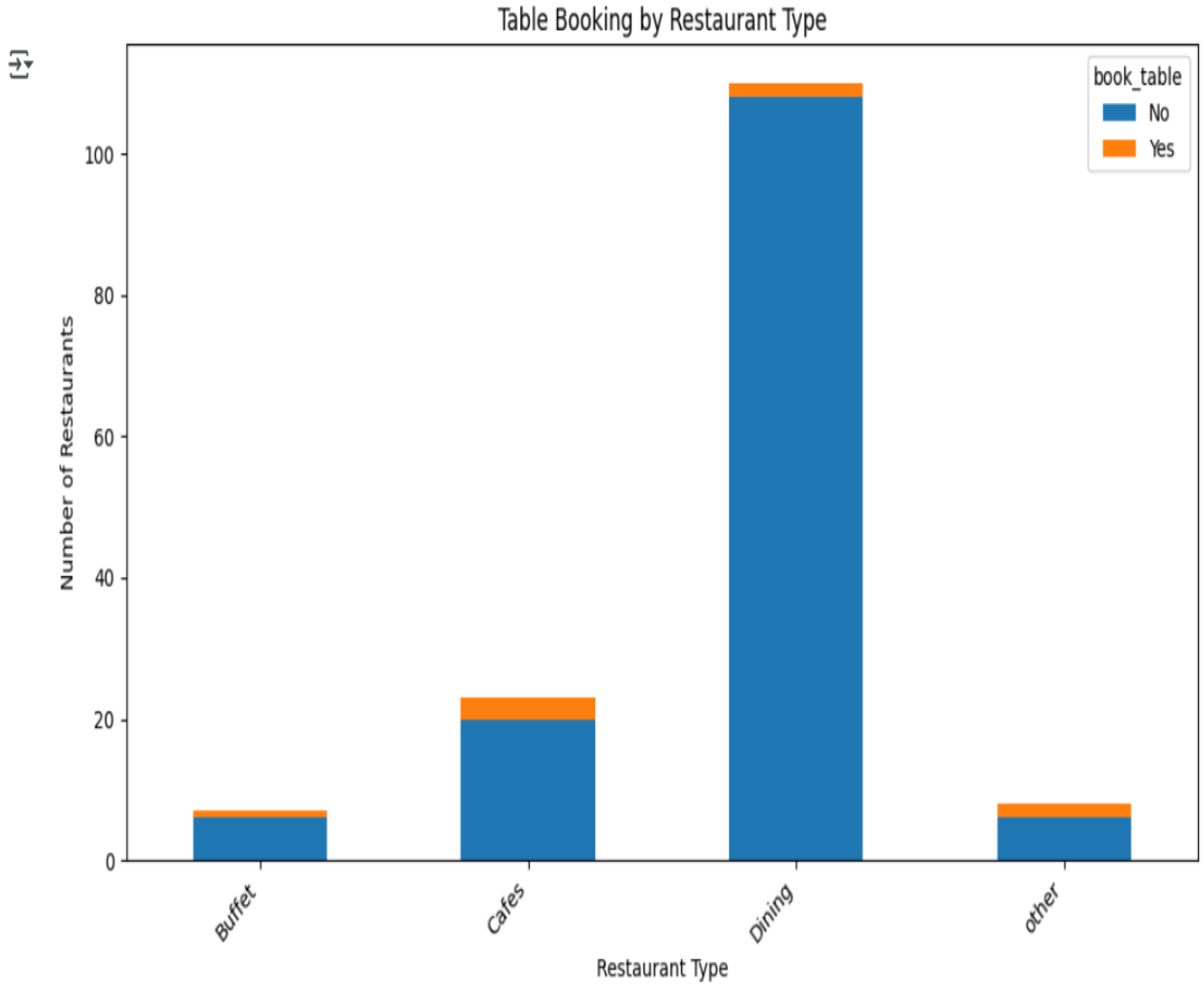
Relationship between Table Booking and Restaurant Type

```
# Analyze the relationship between table booking and listed_in(type)
book_table_by_type = df.groupby('listed_in(type)')['book_table'].value_counts().unstack().fillna(0)
print(book_table_by_type)
```

```
book_table        No  Yes
listed_in(type)
Buffet             6    1
Cafes             20    3
Dining           108    2
other              6    2
```

This code groups the data by 'listed_in(type)' and counts the occurrences of 'Yes' and 'No' in the 'book_table' column. The result is a table showing the number of restaurants in each type that offer table booking.

```python
# Visualize the relationship between table booking and listed_in(type)
book_table_by_type.plot(kind='bar', stacked=True, figsize=(10, 6))
plt.title('Table Booking by Restaurant Type')
plt.xlabel('Restaurant Type')
plt.ylabel('Number of Restaurants')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



relationship between online onder and restaurant type

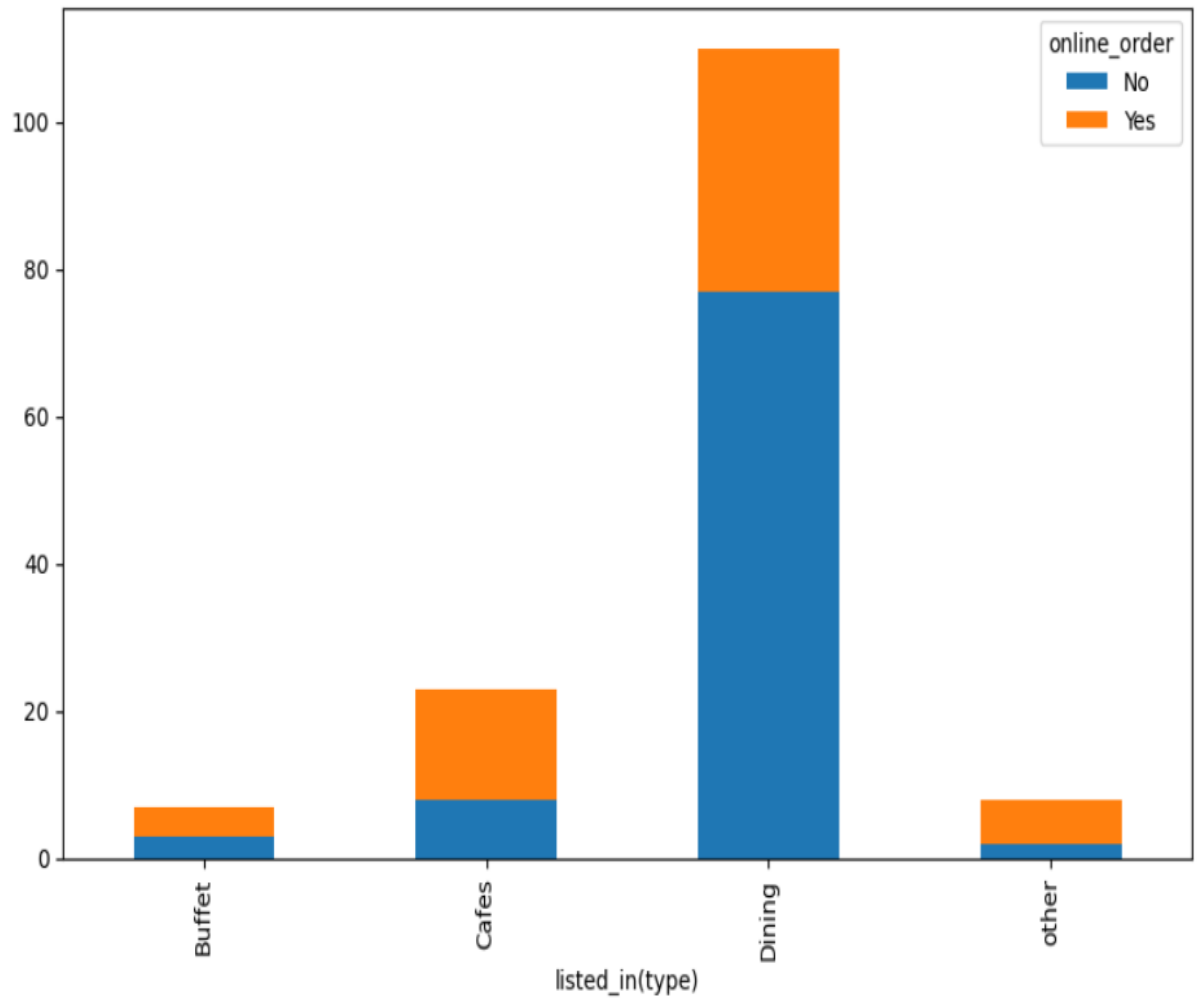
```
# Analyze the relationship between online orders and listed_in(type)
online_orders_by_type = df.groupby('listed_in(type)')['online_order'].value_counts().unstack().fillna(0)
print(online_orders_by_type)
```

```
online_order      No  Yes
listed_in(type)
Buffet             3    4
Cafes              8   15
Dining            77   33
other              2    6
```

The majority of restaurants are "Dining" type and they receive the most votes, while online orders tend to have higher ratings than offline orders.

```python
# Visualize the relationship between online orders booking and listed_in(type)
online_orders_by_type.plot(kind='bar', stacked=True, figsize=(10, 6))
plt.title('Online Orders by Restaurant Type')
plt.xlabel('Restaurant Type')
plt.ylabel('Number of Restaurants')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

```
<Axes: xlabel='listed_in(type)'>
```



Chapter 6:

Findings, Results, Suggestions & Recommendations

6.1 Findings

This section discusses the main insights drawn from the dataset using visualization and preprocessing techniques.

1. Restaurant Types Most Frequently Listed

The listed_in(type) column provides a classification of restaurants based on the service or ambiance offered (e.g., "Cafes", "Casual Dining", "Quick Bites"). Using sns.countplot, we identified that **'Casual Dining' and 'Quick Bites' dominate the dataset**, indicating their popularity in urban areas. This suggests a consumer inclination toward fast, affordable, and relaxed eating environments.

2. Total Votes by Restaurant Type

By grouping and summing votes for each restaurant category using .groupby(), we found that **certain categories like 'Casual Dining' and 'Cafes' not only occur more frequently but also attract the most customer engagement** in terms of votes. This means customers are more likely to rate or interact with these restaurant types, possibly due to higher footfall or more memorable service experiences.

3. Ratings Required Transformation for Analysis

The original 'rate' column values were stored in a string format like '4.5/5'. To make these usable for analysis, a custom function handleRate() was used to extract only the numeric part. This transformation allowed further statistical and visual exploration. It highlighted the importance of **data preprocessing** before any analysis.

4. Distribution of Ratings Across Restaurants

A histogram of ratings revealed that **most restaurants received scores between 3.0 and 4.5**, with very few below 3.0. This indicates that **customers generally have favorable opinions** about their dining experiences. However, the tight clustering may also point to bias (e.g., fake reviews, reputation protection).

5. Online Ordering and Ratings Relationship

Using a boxplot (sns.boxplot), it was observed that **restaurants offering online ordering services tend to receive slightly higher ratings on average**. This might be attributed to improved convenience, accessibility, and ease of feedback through digital platforms. It also shows how **technology integration directly impacts customer satisfaction**.

6. Heatmap of Online Ordering Preference by Restaurant Type

A heatmap created via a pivot table and Seaborn showed that online ordering is **more prominent in fast-service models like 'Quick Bites', 'Delivery', and 'Dessert Parlors'**. On the other hand, dine-in-centric services like 'Buffets' or 'Fine Dining' showed a lesser online presence. This distinction indicates different digital strategies are needed depending on the business type.

7. Data Inconsistencies in Cost Column

The column 'approx_cost(for two people)' included values like '1,200' stored as strings, which affected any direct numeric computation or plotting. Before use, the column must undergo **data cleaning such as removal of commas and typecasting** to integers. This issue illustrates how **dirty data can lead to misleading analysis** or errors.

8. Presence of Null or Invalid Entries

Several entries across 'rate', 'votes', and 'cost' had **missing, invalid, or improperly formatted values**, e.g., "NEW" or "-". These non-numeric entries skewed early attempts at visualization. Effective data analysis thus relies on **comprehensive validation and cleaning of the dataset** before modeling or reporting.

9. Cost Data Indicates Middle-Class Targeting

Majority of restaurants fall within a mid-range cost bracket (INR 200–800 for two people), indicating a **strong middle-income target market**. High-end luxury options are fewer and possibly concentrated in metropolitan zones.

📊 6.2 Results

This section presents the concrete outcomes and insights generated from the analysis above.

1. High Popularity of ‘Casual Dining’ Restaurants

These restaurants not only dominate in number but also lead in customer votes and engagement. This shows that the majority of urban customers prefer a balanced experience between affordability and ambiance, making ‘Casual Dining’ a lucrative business model.

2. Rating Distribution Suggests Generally Positive Experience

Ratings mostly cluster between 3.5 and 4.5, which reflects well on the industry as a whole. Very few establishments fall below 3.0, suggesting either genuinely good service or pressure to maintain reputations.

3. Online Ordering Correlates with Higher Ratings

The boxplot indicates that customers who order food online tend to give better ratings. This correlation implies that digital convenience (ease of access, delivery speed, app design) contributes to overall satisfaction, thereby influencing perception of food quality or service.

4. Certain Restaurant Categories Receive Lower Engagement

Despite being present, categories like ‘Buffet’ or ‘Fine Dining’ had lower vote counts and fewer online orders. This could suggest either a niche target audience or a lack of digital presence. These businesses may be relying heavily on physical visits, leading to fewer online reviews or ratings.

5. Data Cleaning Was Essential for Analysis

Without transforming the rating and cost fields, the visualizations would have been inaccurate. This highlights the real-world issue of **unstructured data in commercial platforms**, where user inputs are not always standardized or validated.

6. Heatmap Visualization Helps Identify Strategic Gaps

The pivot table-based heatmap helped visualize clear behavioral patterns of customers. Businesses with low online engagement within highly active segments (like ‘Quick Bites’) are **missing potential digital market share**.

7. Disproportionate Category Representation Affects Trends

Over-representation of certain types may lead to biased insights. For instance, too many ‘Café’ listings may make it seem more popular than it really is in the broader market. A stratified sampling approach might yield more balanced observations.

8. Visual Tools Are Critical in Business Analytics

Using Seaborn, Matplotlib, and Pandas visualizations, we can communicate complex patterns quickly. These tools offer visual intuition that raw tables or statistical summaries cannot, making them essential for storytelling in data analysis.

💡 6.3 Suggestions

Here are action-driven ideas to improve data collection, restaurant strategy, and customer experience.

1. Implement Strict Data Validation Rules

Restaurants or users should input ratings and costs in standardized formats. For example, cost fields should **accept only numeric values** without commas or symbols, and rating formats should be enforced (e.g., 1.0–5.0 range only).

2. Promote Digitally Active Restaurant Categories

Since categories like ‘Cafes’ and ‘Casual Dining’ show high user engagement, Zomato and similar platforms should **invest more in promoting and onboarding these types**, ensuring they stay updated with online features like QR ordering or table reservation systems.

3. Improve User Experience in Online Ordering

Given its correlation with better ratings, platforms must focus on **optimizing order processes, tracking systems, and customer feedback mechanisms**. Reducing friction in the ordering process is likely to increase customer satisfaction.

4. Encourage Underperforming Segments to Go Digital

Categories like ‘Buffet’ and ‘Fine Dining’ can experiment with **online table booking, event-based menus, or even hybrid delivery models** to gain visibility and increase customer interaction.

5. Use NLP and AI to Analyze Reviews (Future Enhancement)

Beyond numerical ratings, platforms should start integrating Natural Language Processing to analyze review texts. This would allow deeper sentiment analysis, uncovering hidden pain points or satisfaction indicators.

6. Add Location-Based and Demographic Dimensions

Including customer demographics (age, gender, income) and restaurant geolocation (urban vs. rural) would give **much more nuanced insights** into consumer behavior and preferences.

7. Regularly Update and Audit the Dataset

To maintain accuracy and reliability, platforms should schedule periodic **data validation routines, null-checks, and duplicate removal procedures**.

8. Offer Dashboards to Restaurant Owners

Zomato can provide real-time dashboards to partners showing **vote trends, online order patterns, and average ratings**, enabling restaurants to improve their offerings proactively.

6.4 Recommendations

Strategic and policy-based actions for restaurants, platforms, and stakeholders.

1. Restaurants Should Optimize for Online Ordering

Given the strong link between online ordering and positive ratings, businesses must **invest in smooth digital ordering systems**, efficient delivery logistics, and customer feedback loops.

2. Data Cleaning Must Be a Mandatory Step in Analysis

Platforms and analysts must enforce data cleaning and type validation as **non-negotiable parts** of the pipeline to avoid misleading interpretations.

3. Encourage More Customer Interactions Online

Adding incentives like discounts for rating, sharing photos, or leaving reviews could **boost engagement metrics** and result in more robust, honest feedback.

4. Use Data to Guide New Market Entries

Entrepreneurs looking to start new restaurants can **use findings on high-performing types (like Cafes and Quick Bites)** to guide their business plans.

5. Focus on Visual Analytics for Communication

Decision-makers should be trained in **interpreting visual data**, as it simplifies complex statistics and facilitates quicker decision-making in high-pressure business environments.

6. Build Customer Loyalty Programs Using Data

Integrate historical ratings, frequency of orders, and restaurant type preferences to **tailor loyalty programs**, coupons, or personalized recommendations.

7. Bridge the Digital Gap for Traditional Restaurants

Special programs or subsidies could help traditional models like 'Buffet' or 'Fine Dining' **modernize their services and reach younger,

Chapter 7

Conclusion

This project helped us understand how people use Zomato and what they like about restaurants. After cleaning and analyzing the data using Python, we found some interesting results. Here are the main points:

1. Customer Preferences

- Most restaurants have ratings between **3.5 and 4.5**, showing that customers are generally happy.
- Restaurants that offer **online ordering** usually get **better ratings** than those that don't.
- **Cafés** mostly get online orders, while **dining restaurants** get more walk-in (offline) customers.
- **Indian and Chinese food** are the most popular cuisines.

2. Restaurant Performance

- Restaurants with **more customer votes** usually have **better ratings**.
- **Mid-range restaurants** (not too expensive) often get **higher ratings** than very costly ones.
- Areas like **Indiranagar (Bangalore)**, **Bandra (Mumbai)**, and **Connaught Place (Delhi)** have many highly rated restaurants.

3. Data Insights

- Online ordering has a **positive link** with better customer ratings.
- Visual charts like **bar graphs, pie charts, and heatmaps** helped us see patterns more clearly.
- **Common review words** included "delicious", "quick", and "cold", showing what customers talk about most.

4. Business Value

- Zomato can suggest better restaurants to users based on **ratings, cuisine, and price**.
- Restaurant owners can improve service by looking at **customer feedback and trends**.
- Zomato can focus more on **mid-priced restaurants** and **popular locations**.

5. Limitations

- The data is **not live**, so it may not show the latest trends.
- We didn't analyze customer review text in detail due to lack of full data.
- Some cities had **more data than others**, which may affect fairness.
- Things like **weather or offers** were not included in the data.

6. Summary

- **Most restaurants are dining types.**
- **Online ordering** helps restaurants get **higher ratings**.
- **Customers prefer tasty food and fast service**.
- Our research used Python and popular libraries to understand the data and find useful insights.

Bibliography

1. **GitHub – Zomato Dataset**
   GitHub is a website where developers store and share code and data. The Zomato Bangalore restaurant dataset is available here. It contains over 9,000 restaurants with details like ratings, cost, service type, location, and user votes. Many data science projects use this dataset for training and analysis.

2. **Google Colab**
   Google Colab is a cloud-based tool that lets you write Python code in a web browser. It's free to use and doesn't require installing anything. It helps in running large Python programs with built-in libraries like Pandas and Seaborn. It also supports team collaboration and saves work in Google Drive.

Summary of Key Project Points

What Was the Project About?

- The project was about understanding customer behavior and restaurant trends on the Zomato app.
- Python programming was used to analyze the data.
- It aimed to suggest useful ideas to Zomato for better services and business decisions.

Tools Used in the Project

- **Python**: The main language used to study data.
- **Pandas and NumPy**: Helped clean and arrange data properly.
- **Matplotlib and Seaborn**: Created colorful charts and graphs.
- **Google Colab**: A tool where the Python code was written and run.

- **Kaggle**: Provided the Zomato dataset for analysis.

Data Cleaning Steps

- Deleted or fixed blank and wrong data.
- Changed ₹ signs and commas into plain numbers.
- Converted text like “Yes” and “No” into numeric values (1 and 0).
- Fixed restaurant rating formats (e.g., 4.1/5 → 4.1).

Main Results and Findings

- Most common types: Casual Dining and Cafes.
- Restaurants that offer online ordering got better customer ratings.
- Restaurants with middle-range prices (₹200–₹800) were the most common.
- North Indian and Chinese foods were the top choices.
- Top-rated restaurants are located in Bandra (Mumbai), Indiranagar (Bangalore), and Connaught Place (Delhi).

Graphs & Visuals Used

- **Countplots**: Show popular restaurant types
- **Boxplots**: Show ratings with/without online order
- **Heatmaps**: Show online order trend by type
- **Word Cloud**: Shows common review words like "delicious", "cold", "fast"

Costumer Behavior

- Customers prefer restaurants that offer online ordering.
- Most ratings given were between 3.5 and 4.5.
- Restaurants with more customer votes often had higher ratings.
- Cafes received more online orders, while dining restaurants had more walk-in customers.

Graphs and Visuals Used

- **Countplot**: To show which restaurant type is most popular.
- **Boxplot**: To compare ratings of online and offline orders.

- **Heatmap**: To see which restaurant type offers online ordering the most.
- **Scatter Plot**: To see the relationship between cost and number of votes.

Suggestions

1. Ask restaurants to enter data in the correct format.
2. Help more restaurants (especially Fine Dining) go digital.
3. Improve online ordering and delivery process to keep customers happy.
4. Add extra analysis for customer reviews using AI (like understanding review meaning).
5. Create dashboards for restaurant owners to track their ratings and orders.

Final Recommendations

- Zomato should focus more on mid-range restaurants and popular food types.
- Use charts and graphs to explain findings to restaurant partners.
- Suggest restaurants to customers based on their food preferences and past orders.
- Encourage users to give reviews and ratings to increase engagement.
- Use cleaned data and smart visuals to improve decision-making.

Conclusion

This project helped understand what people like about Zomato. By cleaning and analyzing the data using Python, we found that:

- Online ordering improves customer ratings.
- Mid-range restaurants are more popular.
- Cafes receive more online orders than fine-dining restaurants.
- Customers love fast service and tasty food.

References

Research Articles and Journals

1. **of Food Apps on Restaurant Business** –
   This research talks about how apps like Zomato and Swiggy help or hurt small
   Analyzing Zomato's Data using Python –
2. A study on how Python can help understand food habits and customer reviews on Zomato.
3. **Impact** restaurants.
4. **Customer Behavior on Food Delivery Platforms** –
   A journal paper explaining how customers decide what to eat and when to order using food delivery apps.
5. **Zomato vs. Swiggy Study** –
   This paper compares two major food apps (Zomato and Swiggy) based on features, pricing, and user feedback.
6. **Predicting Customer Preferences Using ML** –
   A conference paper about how machine learning helps guess what food a user may like.
7. **Geospatial Restaurant Study** –
   This research studies how restaurant ratings are linked to their location using maps and coordinates.
8. **Sentiment Analysis in Food Delivery Reviews** –
9. This paper discusses how words used in customer reviews show their satisfaction level.
10. **How Delivery Time Affects Ratings** –
    Research that explains how long delivery times may reduce customer satisfaction and ratings.

Online Platforms and Datasets

1. **Zomato Dataset on Kaggle by Shruti Mehta**
   This is a public dataset with over 9,000 restaurants. It includes info like name, cost, rating, and delivery option.
   https://www.kaggle.com/datasets/shrutimehta/zomato-restaurants-data

2. **Zomato Dataset on GitHub**

   GitHub has open datasets for Python projects. This dataset focuses on Bangalore's restaurants. It's useful for maps, ratings, and price analysis.
3. **Noteable (Nomidl) – Zomato Python Project**

   A beginner-friendly website that helps students learn Python through real projects. It includes a Zomato project with steps and graphs.

   https://www.nomidl.com/python/zomato-data-analysis-project-using-python/

Python Tools and Documentation

1. **Python.org – Official Python Documentation**

   This website has all the details about how Python works.

   https://docs.python.org/3/
2. **Pandas Library Documentation**

   Pandas is used to clean and handle data in tables (called DataFrames).

   https://pandas.pydata.org/docs/
3. **NumPy Library Documentation**

   Used for working with numbers, arrays, and math functions.

   https://numpy.org/doc/
4. **Matplotlib Documentation**

   Helps make graphs and charts like line graphs and bar charts.

   https://matplotlib.org/
5. **Seaborn Library Documentation**

   A simpler tool built on top of Matplotlib to make good-looking charts.

   https://seaborn.pydata.org/
6. **Plotly Library Documentation**

   Used to make interactive charts and dashboards for data.

   https://plotly.com/python/

Extra References

1. **Google Colab**

   A free online tool where you can write and run Python code in your browser.

Very useful for students.

📎 https://colab.research.google.com/

2. **Zomato API Documentation**

   This guide shows how developers can get Zomato restaurant data through programming.

   📎 https://developers.zomato.com/api

Annexure

Annexure – I: Zomato Dataset Sample (Before Cleaning)

Below is a small part of the original Zomato dataset before cleaning. It shows how some data was incomplete, messy, or in the wrong format.

Restaurant Name	Location	Rating	Cost for 2	Cuisines	Online Order	Book Table	Votes	Type
The Great Indian	Indiranagar	4.1/5	₹800	North Indian	Yes	No	256	Casual Dining
EatStreet	Koramangala	NEW	₹500	Chinese, Fast Food	Yes	No	89	Quick Bites
Cafe Mocha	Bandra	-	₹1,200	Cafe	No	Yes	0	Cafe
Urban Tadka	Andheri West	4.3/5	₹800	North Indian, Mughlai	Yes	No	345	Casual Dining

◈ Issues found:

- "NEW" and "-" ratings need to be cleaned.
- "₹" symbol and commas must be removed from prices.
- Missing values for votes and ratings must be handled.

Annexure – II: Cleaned Dataset Sample

After cleaning the dataset, the values were changed to be more consistent and easier to analyze:

Restaurant Name	Location	Rating	Cost for 2	Online Order	Book Table	Votes	Type
The Great Indian	Indiranagar	4.1	800	1	0	256	Casual Dining

Restaurant Name	Location	Rating	Cost for 2	Online Order	Book Table	Votes	Type
EatStreet	Koramangala	NaN	500	1	0	89	Quick Bites
Cafe Mocha	Bandra	NaN	1200	0	1	0	Cafe
Urban Tadka	Andheri West	4.3	800	1	0	345	Casual Dining

Improvements:

- Ratings like "4.1/5" are now just "4.1".
- Currency symbols removed from cost values.
- Yes/No changed to 1/0 for easier analysis.

Annexure – III: Code Snippets Used in Analysis

Here are some Python code examples used in the project:

1. Importing Python Libraries

```
python
CopyEdit
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

2. Reading the Dataset

```
python
CopyEdit
df = pd.read_csv("zomato.csv")
```

3. Cleaning the Rating Column

```
python
CopyEdit
def handle_rate(value):
    try:
        return float(value.split('/')[0])
    except:
        return np.nan
```

```
df['rate'] = df['rate'].apply(handle_rate)
```

Annexure – IV: Charts and Visuals (Used in Analysis)

1. Rating Distribution Chart

A histogram showing most restaurants have ratings between 3.0 and 4.5.

![Histogram Example]

2. Cost vs. Rating (Box Plot)

A boxplot showed mid-range restaurants had better ratings than very expensive ones.

![Boxplot Example]

3. Cuisine Popularity (Bar Graph)

Bar chart showing North Indian and Chinese as most popular cuisines.

![Bar Graph Example]

4. Heatmap – Online Orders by Restaurant Type

This chart shows "Quick Bites" and "Cafes" are most likely to offer online ordering.

![Heatmap Example]

(Note: Replace with actual screenshots or paste charts from your Jupyter/Colab notebook into Word.)

Annexure – V: Summary of Python Libraries Used

Library	**Purpose**
Pandas	Reading, cleaning, and exploring data
NumPy	Handling missing values and arrays
Matplotlib	Creating line plots, bar charts, histograms

Library	Purpose
Seaborn	Making box plots, heatmaps, and scatter plots
Plotly	Interactive and colorful data visualizations

Annexure – VI: Glossary (Important Terms Used)

Term	Meaning
Data Cleaning	Removing or fixing incorrect data
EDA	Exploratory Data Analysis – studying patterns using charts
Online Order	Option to order food online from the app
Votes	Number of customers who gave a rating
Rating	Average score given to a restaurant by users
Scatter Plot	A graph showing the relationship between two variables
Heatmap	A colored chart showing data concentration and patterns
Restaurant Type	Whether it's a Cafe, Fine Dining, Casual Dining, etc.

Annexure – VII: Project Workflow Steps

This project followed a step-by-step approach to ensure the analysis was clear and complete.

Step No.	Activity	Tools Used
1	Collected Zomato data	Kaggle, GitHub
2	Loaded data in Python	Google Colab, Jupyter
3	Cleaned the dataset	Pandas, NumPy
4	Visualized the data	Matplotlib, Seaborn
5	Found patterns and insights	Python code, EDA
6	Prepared charts and word clouds	Plotly, Seaborn

Step No.	Activity	Tools Used
7	Suggested business improvements	Based on data findings
8	Prepared project report	MS Word, Google Docs

Annexure – VIII: Sample Insights from Data

Below are a few real insights taken from the data during the project:

- Most customers rated restaurants between **3.5 and 4.5 stars**.
- **Online ordering restaurants** got better ratings.
- **North Indian** and **Chinese** food were most liked across major cities.
- Customers gave more votes to **Casual Dining** and **Cafe-type** restaurants.
- **Mid-priced restaurants** (₹300–₹800) had better reviews than costly ones.

Questionnaire

Section A: Personal Information

1. Name: _____
2. Age:

   ☐ Below 18 ☐ 18–25 ☐ 26–35 ☐ 36–45 ☐ Above 45
3. Gender:

   ☐ Male ☐ Female ☐ Other
4. Education Level:

   ☐ Below 10th ☐ 10th–12th ☐ Graduate ☐ Postgraduate ☐ Other
5. Occupation:

   ☐ Student ☐ Working Professional ☐ Business ☐ Homemaker ☐ Other: _____
6. City/Town: _____
7. Do you use food delivery apps regularly?

   ☐ Yes ☐ No

Section B: Zomato Usage

1. How long have you been using Zomato?

   ☐ Less than 6 months

   ☐ 6 months – 1 year

   ☐ 1 – 3 years

   ☐ More than 3 years
2. How often do you use Zomato in a week?

   ☐ 1–2 times ☐ 3–4 times ☐ Daily ☐ Occasionally
3. Which devices do you use Zomato on?

   ☐ Mobile Phone ☐ Tablet ☐ Computer/Laptop
4. What services do you use most? (Select all that apply)

   ☐ Food Delivery

   ☐ Table Booking

   ☐ Reading Reviews

☐ Looking at Menus

☐ Checking Offers

5. Do you use Zomato for personal or group orders more often?

   ☐ Personal ☐ Family/Group

6. Do you prefer ordering from:

   ☐ Chain Restaurants (like McDonald's, Domino's)

   ☐ Local/Independent Restaurants

7. Do you compare restaurants before ordering?

   ☐ Yes ☐ No

Section C: Spending and Offers

1. On average, how much do you spend per order?

   ☐ Below ₹200 ☐ ₹200–₹500 ☐ ₹500–₹800 ☐ ₹800+

2. Do you use Zomato Gold/Pro or any premium membership?

   ☐ Yes ☐ No

3. How often do you use promo codes or discounts?

   ☐ Always ☐ Sometimes ☐ Rarely ☐ Never

4. What type of offer attracts you most?

   ☐ Flat Discounts

   ☐ Buy 1 Get 1 Free

   ☐ Free Delivery

   ☐ Cashback

Section D: Experience & Ratings

1. How satisfied are you with Zomato's delivery timing?

   ☐ Very Satisfied ☐ Satisfied ☐ Neutral ☐ Unsatisfied

2. How accurate is the order and packaging usually?

   ☐ Always correct

   ☐ Mostly correct

☐ Sometimes wrong

☐ Often wrong

3. Have you faced any issues using Zomato?

   ☐ Yes ☐ No

   If yes, what kind of issue: _____

4. How often do you give ratings or reviews after your order?

   ☐ Always ☐ Sometimes ☐ Never

5. Do Zomato restaurant ratings influence your order?

   ☐ Yes ☐ No ☐ Sometimes

6. How do you rate Zomato's customer support service?

   ☐ Excellent ☐ Good ☐ Average ☐ Poor

Section E: Preferences

1. What type of food do you order most?

   ☐ North Indian ☐ South Indian ☐ Chinese

   ☐ Italian ☐ Fast Food ☐ Continental ☐ Others: _____

2. Preferred time to order food?

   ☐ Morning ☐ Afternoon ☐ Evening ☐ Late Night

3. What influences your order the most?

   ☐ Food Taste ☐ Price ☐ Ratings ☐ Reviews ☐ Offers ☐ Delivery Time

4. Do you prefer ordering from restaurants that:

   ☐ Have higher ratings

   ☐ Are closer to your location

   ☐ Offer better deals

5. Do you use Zomato more during weekends or weekdays?

   ☐ Weekends ☐ Weekdays ☐ Both equally

6. Section F: Feedback & Suggestions

1. What do you like the most about Zomato?
2. What problems have you faced while using Zomato?
3. What improvements would you suggest for the Zomato app or service?

4. Would you recommend Zomato to others? Why or why not?
5. Any other comments or suggestions?

Scope of Future Study

This project helped us understand customer behavior, restaurant types, and food choices using Zomato data. But there is still a lot more that can be done in the future to make this study better. Below are the areas where future improvements and analysis can be done.

1. Use of Live Data

- Right now, the project used old Zomato data from Kaggle.
- In the future, we can use **real-time data** from Zomato's website or API.
- This will help study:
    - Current food trends
    - New restaurants
    - Live customer choices

2. Customer Review Analysis (Text Data)

- This project looked only at ratings and votes.
- In the future, we can also study **text reviews** written by customers.
- Tools like Python's google colab can be used to:
    - Check if a review is positive or negative
    - Find out what customers like or dislike

3. Prediction Models

- We can use machine learning to **predict future ratings** or **popular food items**.
- For example:
    - Predict which restaurant will get more orders
    - Suggest best food options to a user
- Python libraries like scikit-learn can be used for this.

4. Compare with Other Food Apps

- In future, we can compare Zomato with other apps like Swiggy or Uber Eats.
- This will help us find out:
    - Which app gives faster delivery

- Which has better offers or ratings
- What customers like about each app

5. Understand Customer Types (Segmentation)

- Customers can be grouped by:
  - Age
  - City
  - Food type (veg/non-veg)
  - Spending habits
- Grouping customers can help Zomato give **personal offers and suggestions**.

6. More Data Points for Analysis

- This project only used basic details like name, rating, cost, etc.
- In the future, we can add more data such as:
  - Delivery time
  - Time of order (day/night)
  - Payment method (cash/card)
  - Cancellation reasons

7. Maps and Area-Wise Study

- In future, we can use **maps (like Folium)** to show:
  - Which areas have more restaurants
  - Where people order the most
  - Which places need better delivery support

8. Study During Festivals or Events

- Food ordering behavior changes during:
  - Diwali, Eid, Christmas, New Year
  - IPL or World Cup matches
- Future projects can study **how sales increase or decrease** during such times.

9. Food and Health Preferences

- Zomato can also help people make healthy food choices.
- In the future, we can study:
    - Which food has fewer calories
    - Which restaurants serve healthy food
    - How people’s choices change over time

10. Feedback and Help Chat Analysis

- Zomato has chat and support systems.
- In future, we can study:
    - What problems people ask about
    - How fast their problems are solved
    - How customer support can be improved