Department of Computer Science & Engineering

Report on Mini Project

Zomato Data Analysis

Course Code : CS3234-1
Course Name : Social and Web Analytics

Semester: VI SEM

Submitted To:
Dr. Anisha P Rodrigues
Associate Professor
Department of Computer Science and Engineering

Submitted By:
Abhijith P (NNM22CS005)
Adib Haris Neelambra (NNM22CS009)
B Gautam Mohan (NNM22CS037)

Date of submission:
10/04/2025

ABSTRACT

The rapid expansion of food delivery platforms like Zomato has generated large volumes of data that, when analyzed effectively, can reveal critical patterns in consumer behavior and restaurant performance. This mini project focuses on analyzing Zomato's Bangalore restaurant dataset to identify key factors that influence the success and popularity of restaurants across different locations. By leveraging Python for data preprocessing, transformation, and statistical exploration, the raw dataset was cleaned, structured, and enriched with derived features such as cost categories to enable deeper insights.

The analysis investigates how elements such as online ordering availability, restaurant type, price range, and customer engagement - measured through ratings and votes - vary across the city. In parallel, Microsoft Power BI was utilized to design an interactive dashboard that showcases visual summaries and dynamic filters for real-time data exploration. Visual tools like KPI cards, bar charts, box plots, and slicers were used to present findings in a user-friendly format, making it easier for stakeholders to interpret trends and draw conclusions.

Through this integrated analytical approach, the project successfully highlights location-based restaurant patterns, pricing trends preferred by couples, and the impact of service features on customer satisfaction. The insights derived not only provide a deeper understanding of user preferences but also serve as valuable input for restaurant owners, marketers, and platform analysts to make informed decisions based on actual consumer behavior.

TABLE OF CONTENTS

INTRODUCTION

Zomato is one of India's leading platforms for food discovery, restaurant reviews, and online food delivery. It serves as a bridge between customers and restaurants by offering a variety of useful information such as cuisines served, user ratings, approximate cost, service options, and customer feedback. As urban populations grow and food preferences become more dynamic, platforms like Zomato generate vast amounts of data that can be analysed to gain valuable insights into consumer behaviour and business performance.

With its widespread usage and ever-expanding database, Zomato provides a unique opportunity to study how different factors influence customer preferences and restaurant success. Understanding the relationship between service features—such as online ordering availability—and key performance indicators like ratings and customer votes can help stakeholders make data-driven decisions. Additionally, identifying which restaurant types are most common, how pricing affects popularity, and which locations attract the highest engagement can support both operational and strategic planning.

This mini project focuses on analysing Zomato's restaurant data for the Bangalore region. Using Python for data pre-processing and exploratory analysis, and Power BI for visual interpretation, the project aims to identify patterns and correlations among variables like location, cost, customer ratings, service features, and engagement levels. The goal is to provide actionable insights that can benefit restaurant owners, platform analysts, and marketers by offering a clearer understanding of what drives customer satisfaction and restaurant performance in a competitive food service environment.

PROBLEM STATEMENT

This project analyzes Zomato's dataset to uncover factors influencing restaurant establishment and popularity across locations by evaluating the impact of online ordering on ratings, identifying popular restaurant types, assessing cost preferences for couples dining out, and exploring the overall distribution of ratings and customer satisfaction.

OBJECTIVES

- To examine how online ordering availability affects restaurant ratings, helping to understand its influence on customer satisfaction.
- To identify the most popular restaurant types based on frequency and customer engagement.
- To assess preferred cost ranges for couples dining out, by grouping restaurants into cost categories and analysing their distribution.
- To explore the distribution of restaurant ratings across various locations, identifying patterns and outliers.
- To measure customer satisfaction and engagement by analysing rating trends and total vote counts.
- To visualize all findings effectively using Python and Power BI, ensuring clarity, interactivity, and accessibility of insights for decision-makers.

METHODOLOGY

This project followed a systematic approach combining Python programming and Power BI to extract, process, analyze, and visualize data from the Zomato Bangalore restaurant dataset. The methodology involved cleaning the raw dataset, transforming relevant fields, generating visual insights using Python, and creating an interactive dashboard in Power BI.

Data Source:

- The dataset was sourced from Kaggle and contains restaurant information specific to Bangalore, including details like name, city, type, rating, cost, votes, and online order availability.

Tools and Technologies Used:

- **Python**: Used for data preprocessing and generating static visualizations.
    - Libraries: pandas, NumPy, matplotlib, seaborn
- **Power BI**: Used for interactive dashboard development and data presentation.

Steps Followed:

1. **Data Preprocessing in Python:**
    - Loaded the dataset using Pandas.
    - Retained only necessary columns: Restaurant Name, City, Restaurant Type, Rate, Votes, Cost, and Online Order.
    - Renamed columns for better readability.
    - Removed duplicate records and dropped rows with missing values in important fields.
    - Cleaned the Cost column by removing commas and converting it to numeric format.
    - Processed the Rate column by extracting the numeric part from strings like “4.1/5”, and filtered out invalid entries.
    - Dropped any remaining null values after transformation.
2. **Feature Engineering:**
    - Created a new column Cost Category to categorize restaurants into pricing groups: < ₹300, ₹300–600, ₹600–900, and > ₹900.

3. **Visual Analysis using Python:**
    - Created a series of static visualizations to understand patterns and trends:
        - Top 10 restaurant locations (cities/areas) by count.
        - Average rating by city.
        - Box plot to show the effect of online ordering on customer ratings.
        - Most popular restaurant types by count.
        - Stacked bar chart of cost category distribution across cities.
        - Total votes per city to assess customer engagement.
        - Histogram showing the distribution of ratings.
4. **Dashboard Development in Power BI:**
    - Imported the cleaned dataset into Power BI.
    - Created KPI cards to display:
        - Total number of restaurants
        - Average rating
        - Average cost
        - Number of restaurant types
    - Built charts for:
        - Restaurant counts by city
        - Average rating per city
        - Popular restaurant types
        - Cost categories by city
        - Impact of online ordering on ratings
        - Total customer votes by city
    - Added slicers for filtering by **City**, **Restaurant Type**, and **Online Order**, enabling dynamic interaction and deeper exploration.

This integrated methodology ensured that the data was accurately prepared and effectively visualized. Python was used for in-depth statistical analysis and static insights, while Power BI transformed the findings into an interactive format suitable for both technical and non-technical users.

IMPLEMENTATION

The implementation of this mini project was carried out in two major phases:

1. Data preprocessing and visualization using **Python**, and
2. Interactive dashboard creation using **Power BI**.

This structured approach helped in cleaning the raw dataset, analyzing key metrics, and presenting the results visually for better understanding and interpretation.

1. Python-Based Analysis and Visualization

Python was used to clean the dataset, perform data transformations, and generate various graphs for exploratory data analysis. The following libraries were used throughout the process:

- pandas for data manipulation
- NumPy for numerical operations
- matplotlib and seaborn for plotting visualizations

1.1 Importing Required Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

1.2 Loading the Dataset

The dataset was loaded from a CSV file named zomato.csv.

```
# Load the dataset
df = pd.read_csv("zomato.csv", encoding='latin-1')

# Display initial shape and first few rows
print("Initial shape of dataset:", df.shape)
print(df.head())
```

1.3 Column Selection and Renaming

Only relevant columns were retained and renamed for better clarity.

```
# Keep only necessary columns
df = df[[
    'name',
    'listed_in(city)',
    'listed_in(type)',
    'rate',
    'votes',
    'approx_cost(for two people)',
    'online_order'
]]

# Rename columns for clarity
df.rename(columns={
    'name': 'Restaurant Name',
    'listed_in(city)': 'City',
    'listed_in(type)': 'Restaurant Type',
    'approx_cost(for two people)': 'Cost'
}, inplace=True)
```

1.4 Data Cleaning

Steps included removing duplicates, handling missing values, cleaning the cost column, and converting the rating to a usable numeric format.

```
# Drop duplicate rows
df.drop_duplicates(inplace=True)

# Drop rows with missing values in key columns
df.dropna(subset=['Restaurant Name', 'City', 'Restaurant Type',
'rate', 'votes', 'Cost', 'online_order'], inplace=True)

# Clean the 'Cost' column
df['Cost'] = df['Cost'].astype(str).str.replace(',', '')
df['Cost'] = pd.to_numeric(df['Cost'], errors='coerce')
```

```
# Clean the 'rate' column
df['rate'] = df['rate'].astype(str)
df = df[df['rate'].str.contains('/', na=False)]
df['rate'] = df['rate'].str.split('/').str[0]
df['rate'] = pd.to_numeric(df['rate'], errors='coerce')

# Drop remaining rows with null values after cleaning
df.dropna(inplace=True)
```

1.5 Feature Engineering

A new column called Cost Category was created to classify restaurants into pricing brackets.

```
# Define function to classify restaurants into cost
categories
def cost_category(x):
    if x < 300:
        return '< ₹300'
    elif x < 600:
        return '₹300-600'
    elif x < 900:
        return '₹600-900'
    else:
        return '> ₹900'

# Apply the function to create the 'Cost Category' column
df['Cost Category'] = df['Cost'].apply(cost_category)
```

1.6 Generating Visualizations

Multiple visualizations were generated to gain insights from the data:

- Top 10 restaurant locations (bar chart)

```
# Top 10 Restaurant Locations
plt.figure(figsize=(12,6))
df['City'].value_counts().head(10).plot(kind='barh', color='skyblue')
plt.title("Top 10 Cities by Number of Restaurants")
plt.xlabel("Number of Restaurants")
plt.gca().invert_yaxis()
plt.tight_layout()
plt.show()
```

- Average rating by city (bar chart)

```
# Average Rating by City
plt.figure(figsize=(12,6))
df.groupby('City')['rate'].mean().sort_values(ascending=False)
.head(10).plot(kind='bar', color='coral')
plt.title("Average Rating by City")
plt.ylabel("Average Rating")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

- Online ordering impact on ratings (box plot)

```
# Online Order Impact on Ratings
plt.figure(figsize=(8,6))
sns.boxplot(x='online_order', y='rate', data=df)
plt.title("Impact of Online Ordering on Ratings")
plt.xlabel("Online Ordering Available")
plt.ylabel("Rating")
plt.tight_layout()
plt.show()
```

- Most popular restaurant types (bar chart)

```
# Most Popular Restaurant Types
plt.figure(figsize=(12,6))
df['Restaurant Type'].value_counts().head(10).plot(kind='bar',
color='lightgreen')
plt.title("Most Popular Restaurant Types")
plt.ylabel("Number of Restaurants")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

- Preferred price range by city (stacked bar chart)

```
# Preferred Price Range by City
plt.figure(figsize=(14,6))
cost_city = pd.crosstab(df['City'], df['Cost Category'])
cost_city.plot(kind='bar', stacked=True, figsize=(14,6),
colormap='Set2')
plt.title("Preferred Cost Category by City")
plt.xlabel("City")
plt.ylabel("Number of Restaurants")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

- Customer engagement by city (votes bar chart)

```
# Customer Engagement by City (Votes)
top_votes =
df.groupby('City')['votes'].sum().sort_values(ascending=False).head(10)
plt.figure(figsize=(12,6))
top_votes.plot(kind='bar', color='orange')
plt.title("Customer Engagement by City (Total Votes)")
plt.ylabel("Total Votes")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

- Rating distribution (histogram)

```
# Distribution of Restaurant Ratings
plt.figure(figsize=(8,6))
sns.histplot(df['rate'], bins=20, kde=True,
color='purple')
plt.title("Distribution of Restaurant Ratings")
plt.xlabel("Rating")
plt.ylabel("Frequency")
plt.tight_layout()
plt.show()
```

2. Power BI Dashboard Development

After the Python-based data cleaning and transformation, the processed dataset was imported into Power BI for interactive visualization. The following elements were added to the dashboard:

- **KPI Cards** to show:
  - Total Restaurants
  - Average Rating
  - Average Cost
  - Number of Restaurant Types
- **Visual Charts**:
  - Pie chart for restaurant type distribution
  - Bar charts for restaurant count and average rating by city
  - Stacked bar chart showing rating distribution based on online ordering
  - Cost category distribution across different cities
  - Votes by location
- **Slicers/Filters**:
  - For City, Restaurant Type, and Online Ordering, allowing dynamic data filtering

RESULTS AND DISCUSSIONS

The results of the analysis provide insights into customer behavior, restaurant preferences, and service features using both Python visualizations and Power BI dashboards. This section presents key findings based on the cleaned and processed dataset.

1. Top Restaurant Locations

The bar chart showing the Top 10 restaurant locations revealed that areas like BTM, Koramangala 5th Block, and HSR have the highest concentration of restaurants. These are likely to be commercial hubs or popular social areas in Bangalore.

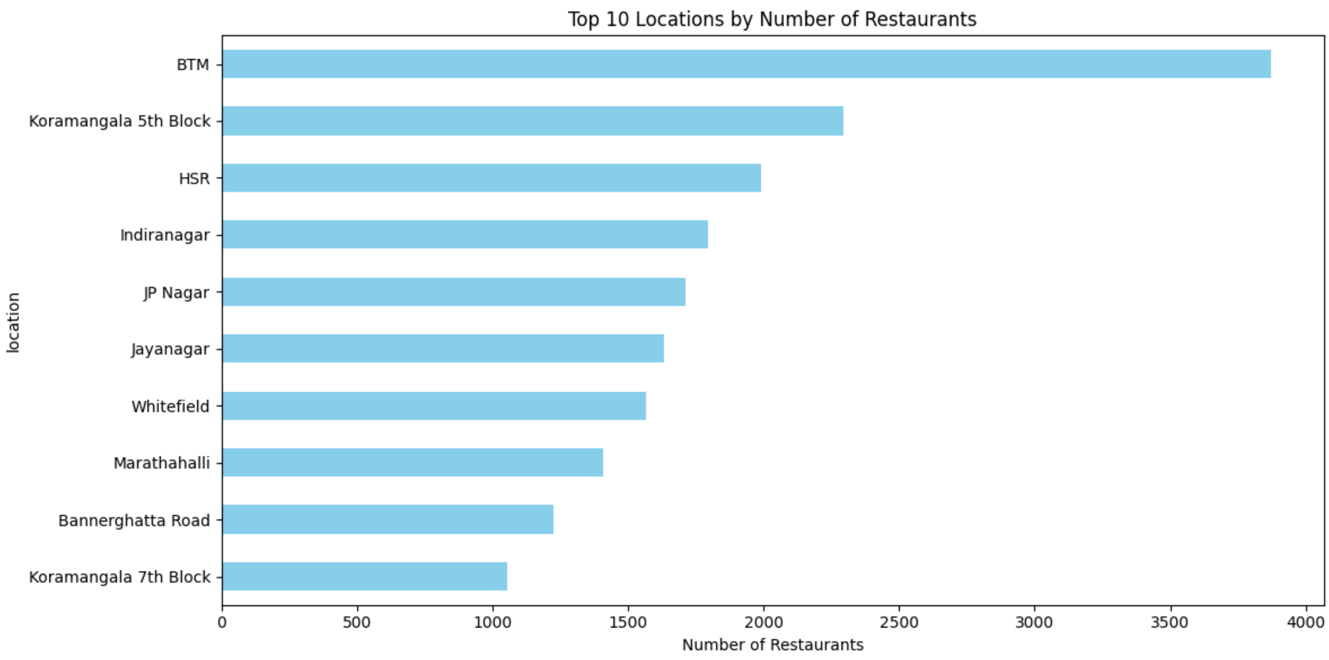

Figure 1: Top 10 Cities by Number of Restaurants

2. Average Rating by City

The chart depicting average ratings by city showed that although many areas have similar restaurant counts, some localities like Indiranagar and Whitefield tend to have higher average ratings, possibly due to better food quality or customer service.

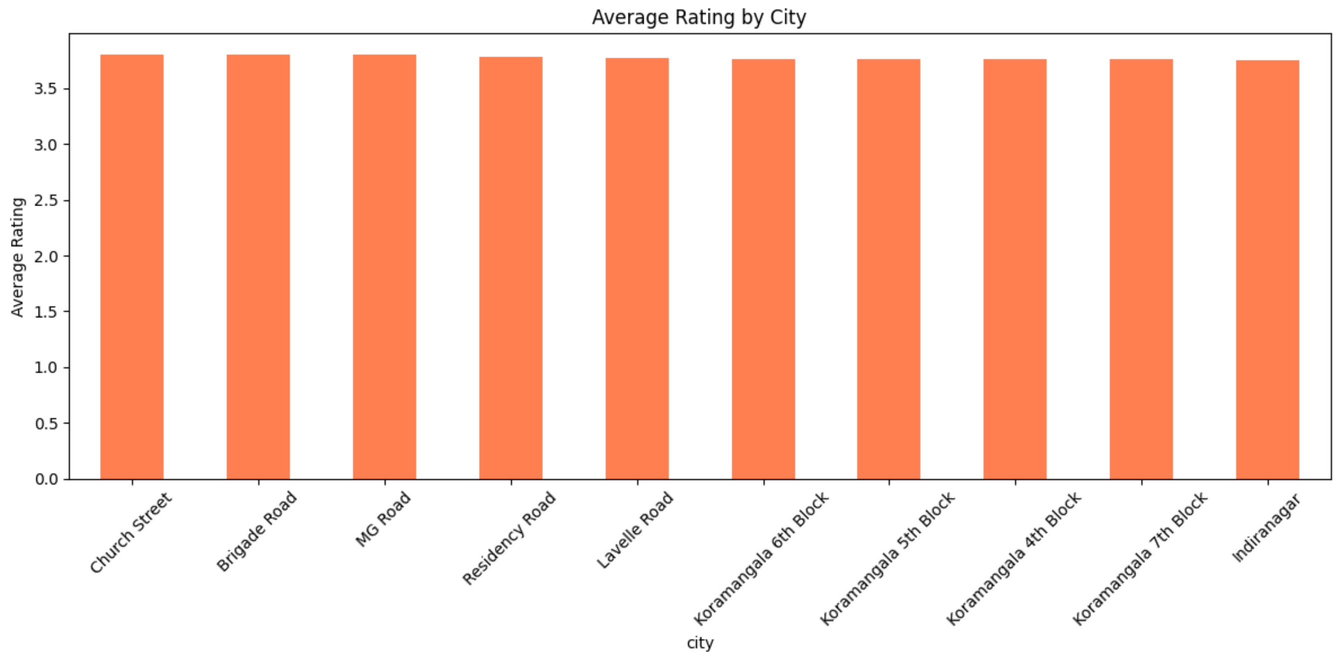

Figure 2: Average Rating by City

3. Impact of Online Ordering on Ratings

The boxplot comparing restaurants that offer online ordering versus those that don't showed that restaurants with online ordering tend to have slightly higher customer ratings, suggesting that customers value convenience and accessibility.

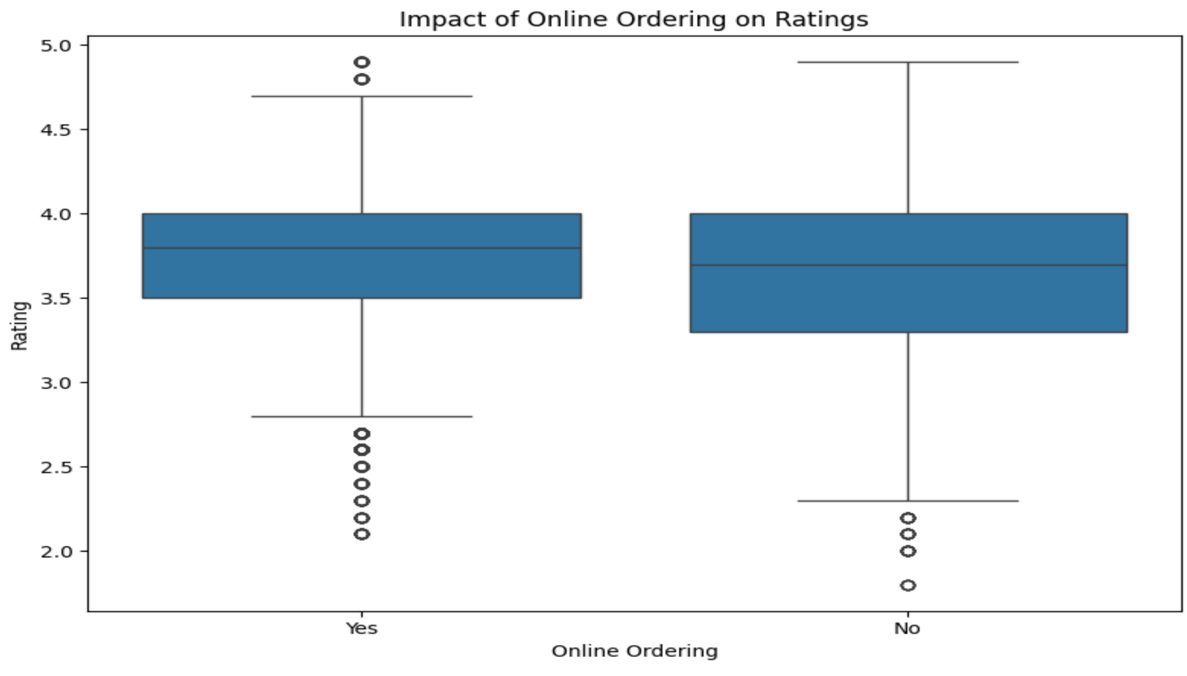

Figure 3: Impact of Online Ordering on Ratings

4. Popular Restaurant Types

Analysis of the most common restaurant formats indicated that Casual Dining, Cafes, and Quick Bites are the most popular types across Bangalore. These types are generally more flexible and cater to a broader customer base.

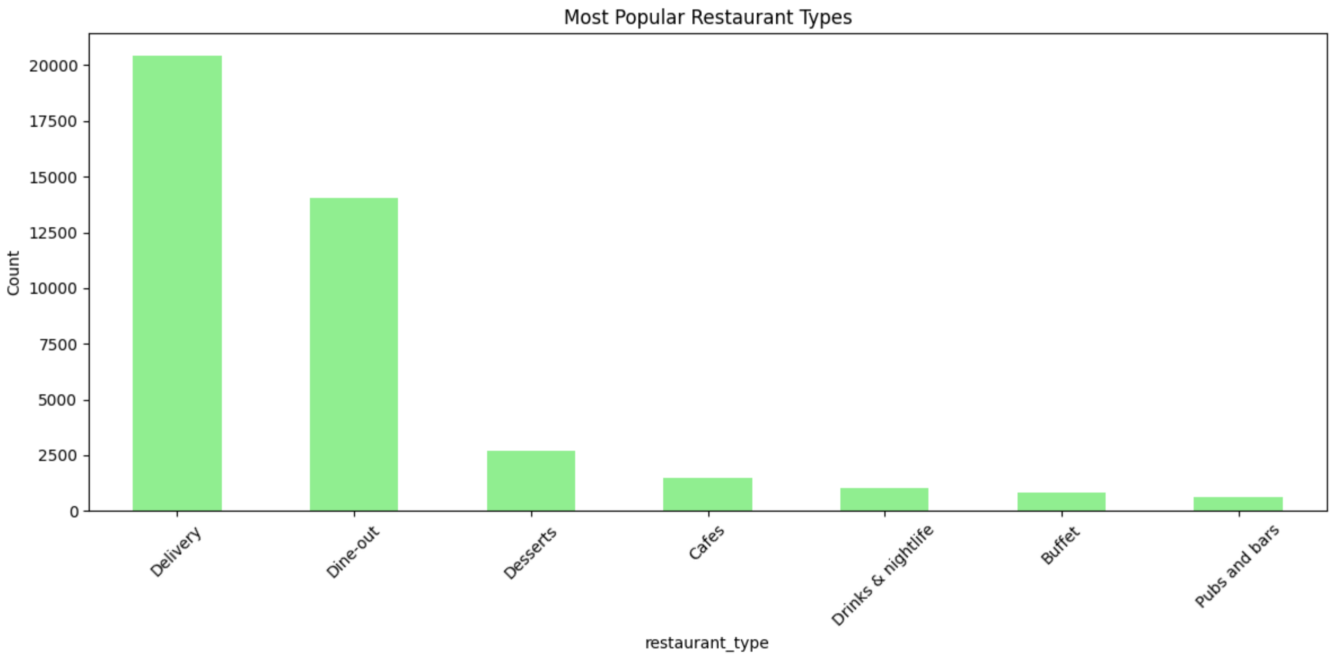

Figure 4: Most Popular Restaurant Types

5. Preferred Cost Range by City

The stacked bar chart of cost categories by city revealed that the majority of restaurants fall within the ₹300–600 price range, which appears to be the most preferred by customers, especially for couples dining out.

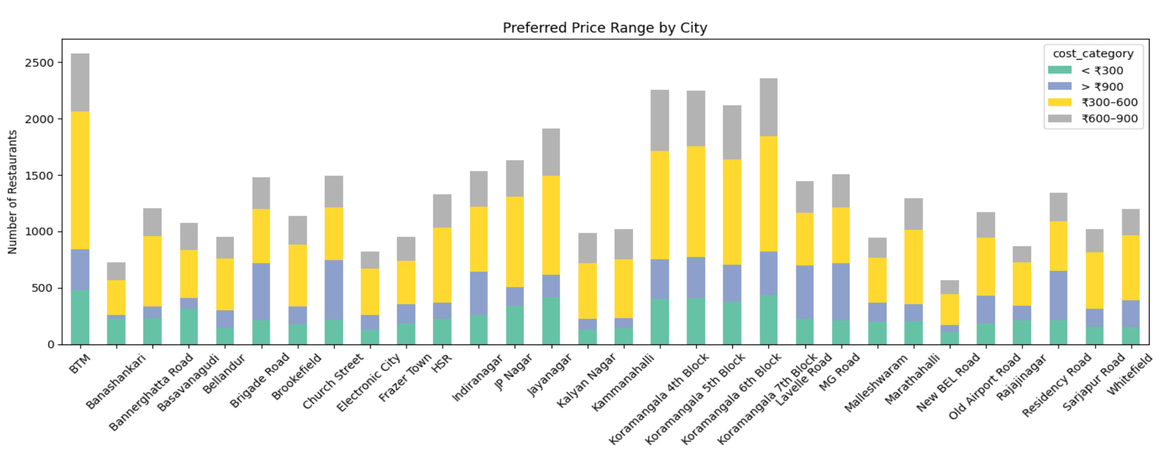

Figure 5: Preferred Cost Category by City

6. Customer Engagement (Votes) by City

Cities with the most restaurants also saw higher customer interaction in terms of total votes. This suggests that areas like BTM and Koramangala not only host more restaurants but also drive more customer engagement and feedback.

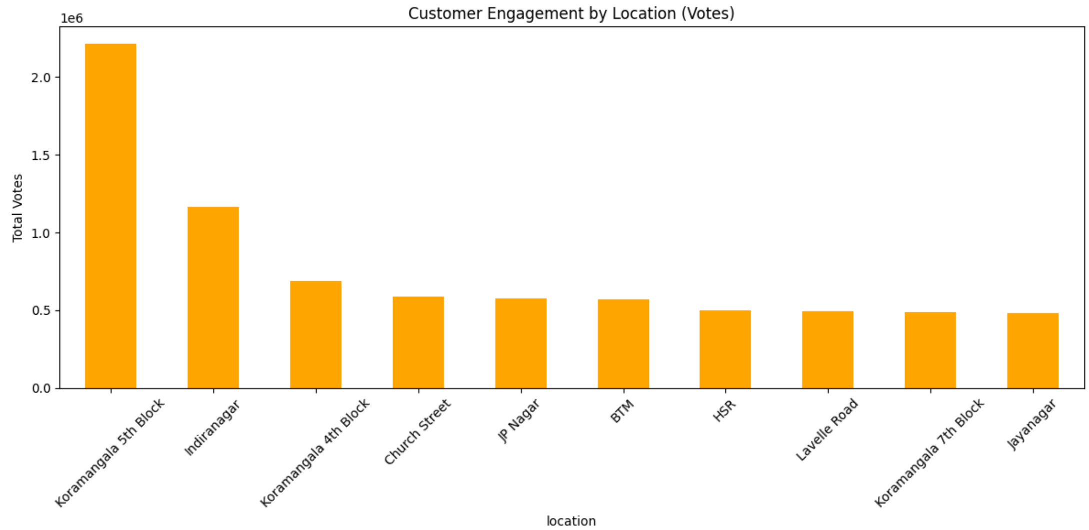

Figure 6: Customer Engagement by City (Total Votes)

7. Rating Distribution

The histogram of restaurant ratings showed that the majority of restaurants are rated between 3.0 and 4.5, with very few extremely low or high ratings. This indicates a relatively consistent level of service quality across most listed restaurants.

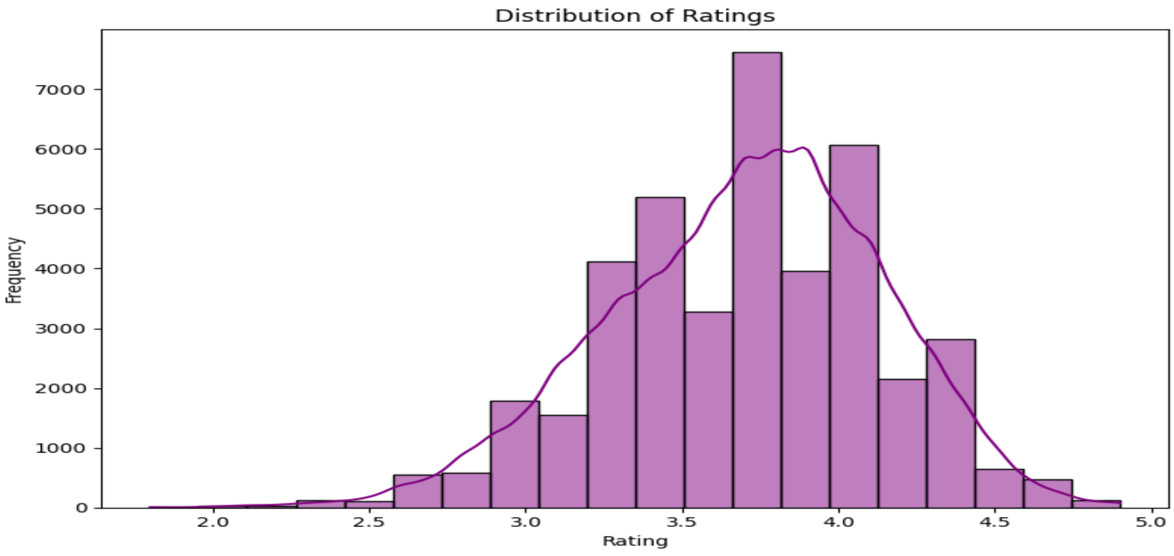

Figure 7: Distribution of Restaurant Ratings

8. Power BI Dashboard **Insights**

- The Power BI dashboard consolidated the key findings into an interactive format. It included:
- KPI cards showing total restaurants, average cost, average rating, and number of restaurant types.
- Charts visualizing restaurant count and average rating by city.
- A pie chart of restaurant types.
- A stacked bar comparing ratings with and without online ordering.
- Slicers to dynamically filter by City, Restaurant Type, and Online Ordering.

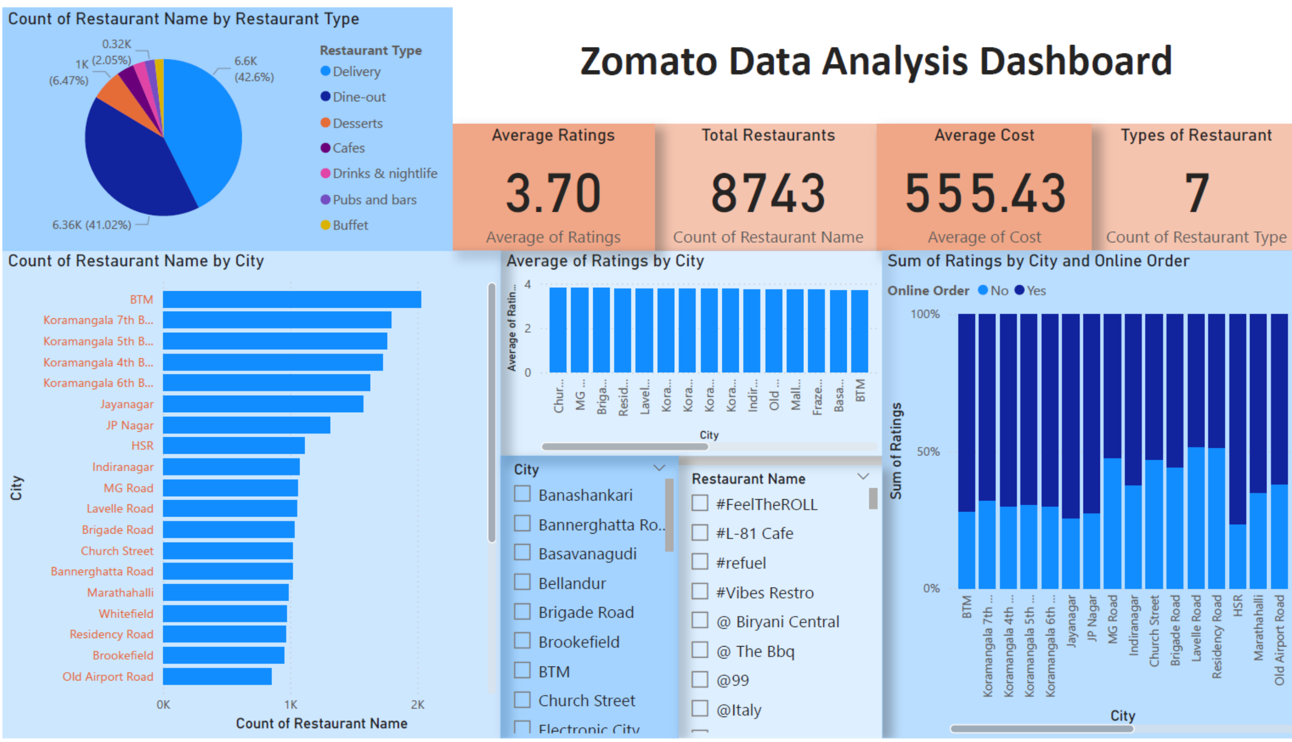

Figure 8: Power BI Dashboard Overview

The interactivity of Power BI allowed for real-time filtering and exploration, making it easier to spot localized trends and patterns in customer preferences.

CONCLUSION AND FUTURE SCOPE

This project successfully utilized Python and Power BI to analyze restaurant data from Zomato, specifically focusing on establishments located in Bangalore. Through data cleaning, transformation, and visualization, we uncovered key insights into restaurant popularity, customer preferences, and the influence of features such as online ordering and cost. The visualizations helped identify high-density restaurant areas, most preferred restaurant types, pricing trends, and customer engagement levels.

The analysis showed that online ordering availability is generally associated with higher ratings, while areas like Koramangala and BTM emerged as popular food hubs. Casual dining and cafes are the most prevalent types, and the ₹300–₹600 cost range appears to be the most favorable for couples dining out.

Future Scope

- **Geographic Expansion**: The same analysis can be extended to include data from other cities to identify nationwide patterns.
- **Time Series Analysis**: Incorporating date-based data can help study trends across different seasons or years.
- **Sentiment Analysis**: Analyzing customer review texts can provide deeper insights into satisfaction levels and service quality.
- **Real-time Dashboards**: Integrating real-time data with Power BI dashboards can allow restaurant managers to make quicker, data-driven decisions.

REFERENCES

[1] **Zomato Bangalore Restaurant Dataset – Kaggle:**

https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalore-restaurants

[2] **Python Documentation:**

https://pandas.pydata.org

https://numpy.org

https://matplotlib.org

https://seaborn.pydata.org

[3] **Microsoft Power BI Documentation:**

https://learn.microsoft.com/en-us/power-bi

[4] **Zomato Official Website:**

https://www.zomato.com