Project Name

Zomato Data Analysis using python

Submitted by:
Devanshi Jindal
Himanshu Singla

MCA 2nd Year

Submitted to:

Dr. Anjula Mehto

Assistant Professor

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology, Patiala

November 2025

TABLE OF CONTENTS

Introduction or Project Overview

The food delivery and restaurant discovery ecosystem in India has grown rapidly over the last decade, with platforms like **Zomato** playing a central role in connecting customers with restaurants. As a result, understanding restaurant data has become extremely valuable for identifying customer preferences, predicting demand, analyzing market competition, and helping restaurant businesses make better decisions.

This project, **Zomato Restaurant Data Analysis**, focuses on performing an in-depth exploratory analysis (EDA) of restaurant data collected from Bangalore – one of the biggest food hubs in India. The aim is to uncover hidden patterns, extract meaningful insights, and understand the factors that influence restaurant performance on Zomato.

The project involves:

- Cleaning and preprocessing raw, messy restaurant data
- Performing statistical analysis and visualization
- Studying online ordering trends
- Analyzing rating patterns and customer behavior
- Understanding how cuisine variety, locality, pricing, and services influence success

The insights generated in this project can help:

- **Restaurant owners** understand what customers prefer
- **Food delivery platforms** improve recommendations
- **Consumers** make better decisions based on ratings, cost, and services
- **Researchers** build ML models such as price prediction, rating forecasting, or restaurant recommendation systems

Problem Statement

The restaurant industry produces massive amounts of data, but stakeholders often struggle to interpret it effectively due to inconsistent formats, missing values, and lack of analytical tools. Without proper analysis, restaurants may fail to understand customer expectations, leading to poor performance and low ratings.
The key problems addressed in this project include:

1. Lack of clarity about customer preferences:
Customers choose restaurants based on price, ratings, cuisines, and services. However, no structured analysis exists to understand which factors they value most.

2. Difficulty in identifying highly rated vs poorly rated restaurant patterns:
The dataset contains inconsistent rating formats, missing values (NEW, –), duplicate entries, and misleading values that require preprocessing.

3. Unclear relationship between cost and restaurant rating:
Many restaurants spend more on ambience and menu variety, but higher cost doesn't always mean higher rating.

4. Understanding the impact of services:

- Does *online ordering* increase customer ratings?
- Does *table booking* improve user experience?

5. Locality-based performance variation:
Different areas of Bangalore show different restaurant densities, cost trends, and popularity, which needs analysis.

Overview of the Dataset used

The dataset, **zomato.csv**, contains detailed information about restaurants listed on Zomato in Bangalore.

Dataset Size

- **Total Restaurants:** ~9,500+ entries
- **Columns:** 17
- **City:** Bangalore

Important Columns Explained

COLUMN	DESCRIPTION
URL	Web link to Zomato restaurant page
ADDRESS	Complete address of each restaurant
NAME	Restaurant name
ONLINE_ORDER	Whether restaurant offers online delivery (Yes/No)
BOOK_TABLE	Table booking availability
RATE	Rating in format "4.1/5", "NEW", "–"
VOTES	Number of ratings received
PHONE	Contact details
LOCATION	Area / locality within Bangalore
REST_TYPE	Type of restaurant (Casual Dining, Café, etc.)
DISH_LIKED	What customers liked most
CUISINES	Types of food served
APPROX_COST(FOR TWO)	Cost for two people (Messy format with commas)
REVIEWS_LIST	Actual customer reviews (text)
LISTED_IN(TYPE)	Buffet, cafés, desserts, etc.
LISTED_IN(CITY)	Locality group category

Project Workflow

Step 1: Importing Necessary Libraries
Libraries used:

- pandas – Data management
- numpy – Numerical operations
- matplotlib & seaborn – Data visualization
- wordcloud – text visuals
- collections – word count

Step 2: Data Loading and Basic Inspection

- Loaded zomato.csv using pandas
- Checked column names, dataset shape, datatypes
- Inspected missing values
- Verified duplicates

Step 3: Data Cleaning and Preprocessing
This is the most important step.
Important Cleaning Tasks

1. Removing duplicates
2. Fixing the 'rate' column
    - Removed "/5"
    - Converted "NEW" and "–" to NaN
3. Cleaning cost column
    - Removed commas
    - Converted strings to integers
4. Handling missing values
    - Filled or dropped where required
5. Cleaning textual columns (reviews, phone, cuisines)
6. Standardizing categories like online_order and book_table

Step 4: Exploratory Data Analysis
Performed detailed insights and visualizations:
a. Online Order vs Rating

- Restaurants offering online delivery have more customer engagement.

b. Table Booking Analysis

- Table booking is mostly available in higher-priced restaurants.

c. Cost Distribution

- Most restaurants fall between ₹200 – ₹700 cost range.

d. Cuisine Analysis

- North Indian, Chinese, and South Indian dominate.
- Multi-cuisine restaurants have higher votes.

e. Locality-Based Analysis

- Areas like Koramangala, Indiranagar, and Jayanagar have highest density.
- Locality affects cost and rating heavily.

f. Rating Distribution

- Most ratings lie between 3.0 to 4.5.
- Very few restaurants score above 4.5.

g. Votes Analysis

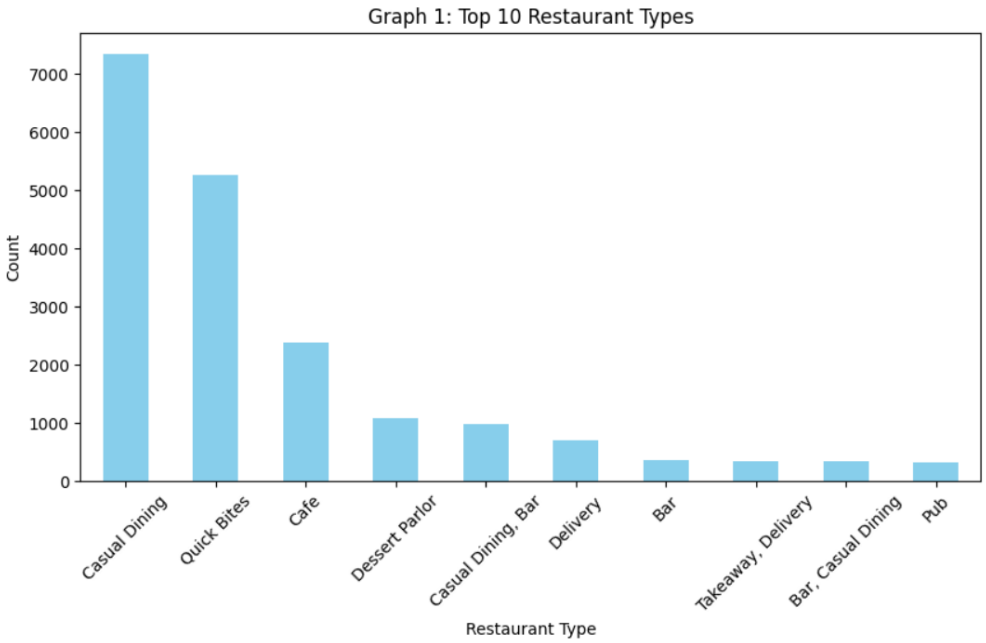
- Restaurants with online orders receive significantly more votes.
- Premium restaurants with high cost also tend to have higher votes.

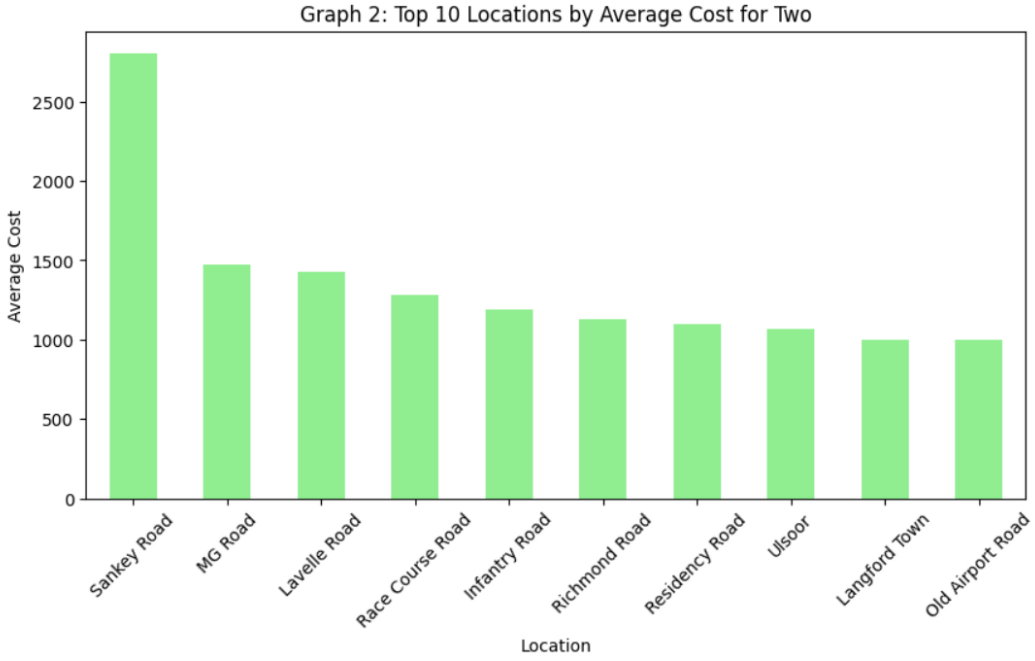
Step 5: Visualizations
Created several visual plots:

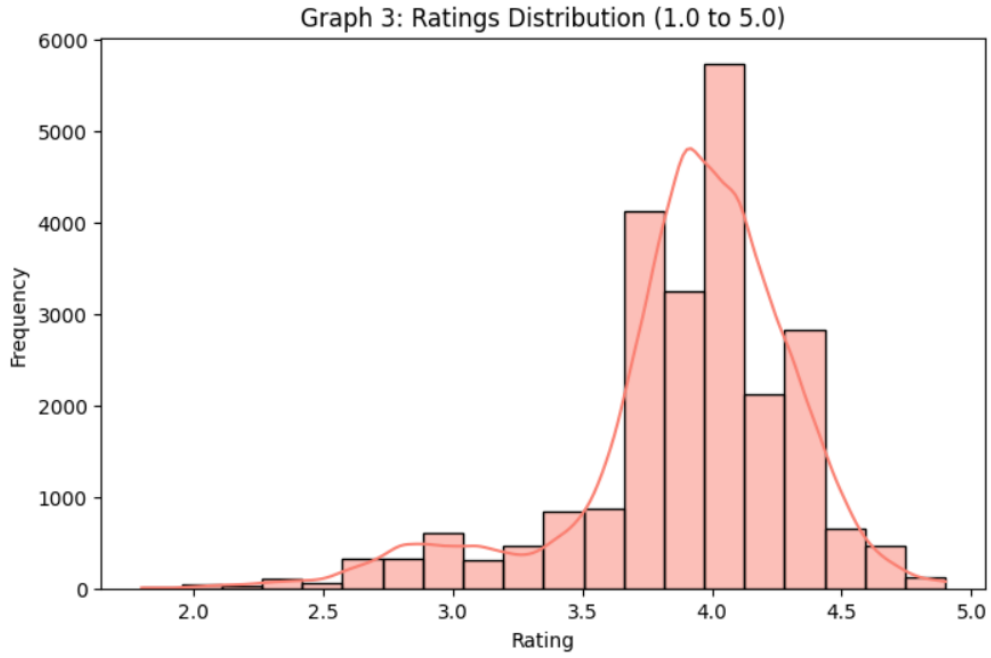
- Bar charts
- Countplots
- Pie charts
- Boxplots
- Heatmaps
- Scatter plots

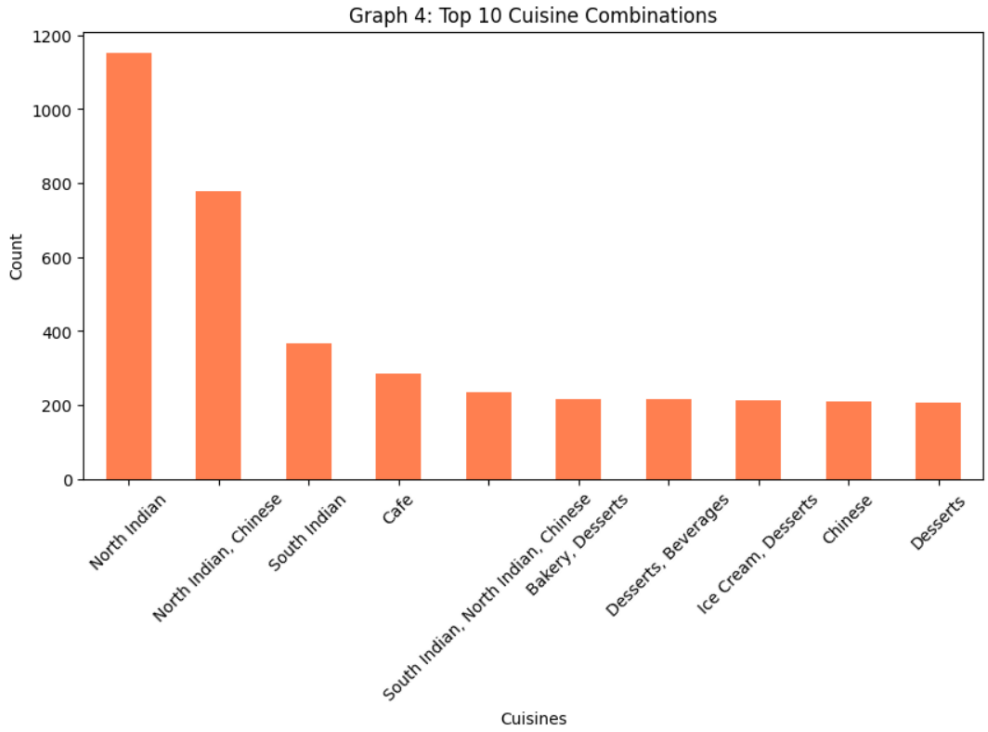
These helped identify correlations and patterns.

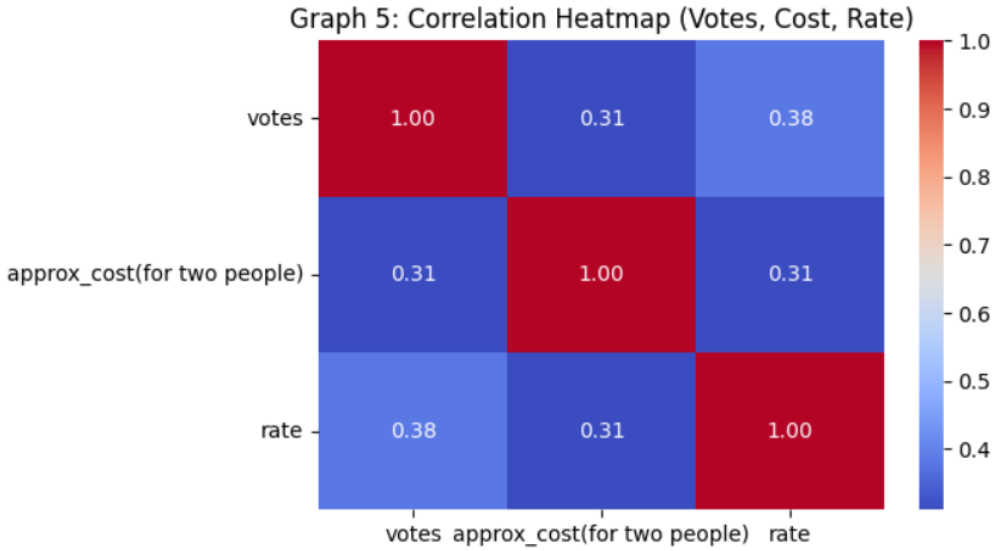
Results

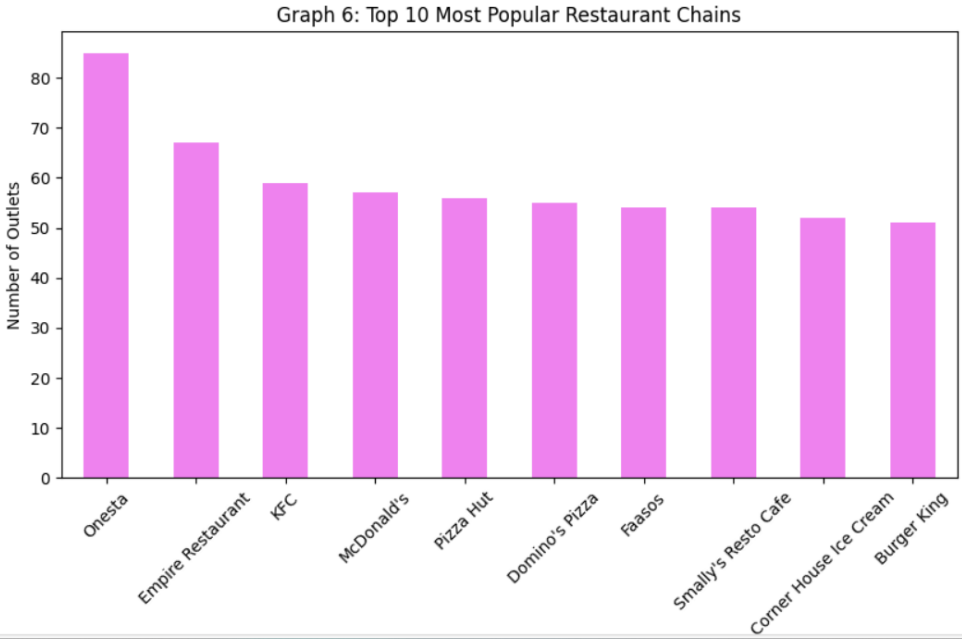

Graph 1: Top 10 Restaurant Types
Count
0
1000
2000
3000
4000
5000
6000
7000
Casual Dining
Quick Bites
Cafe
Dessert Parlor
Casual Dining, Bar
Delivery
Bar
Takeaway, Delivery
Bar, Casual Dining
Pub
Restaurant Type

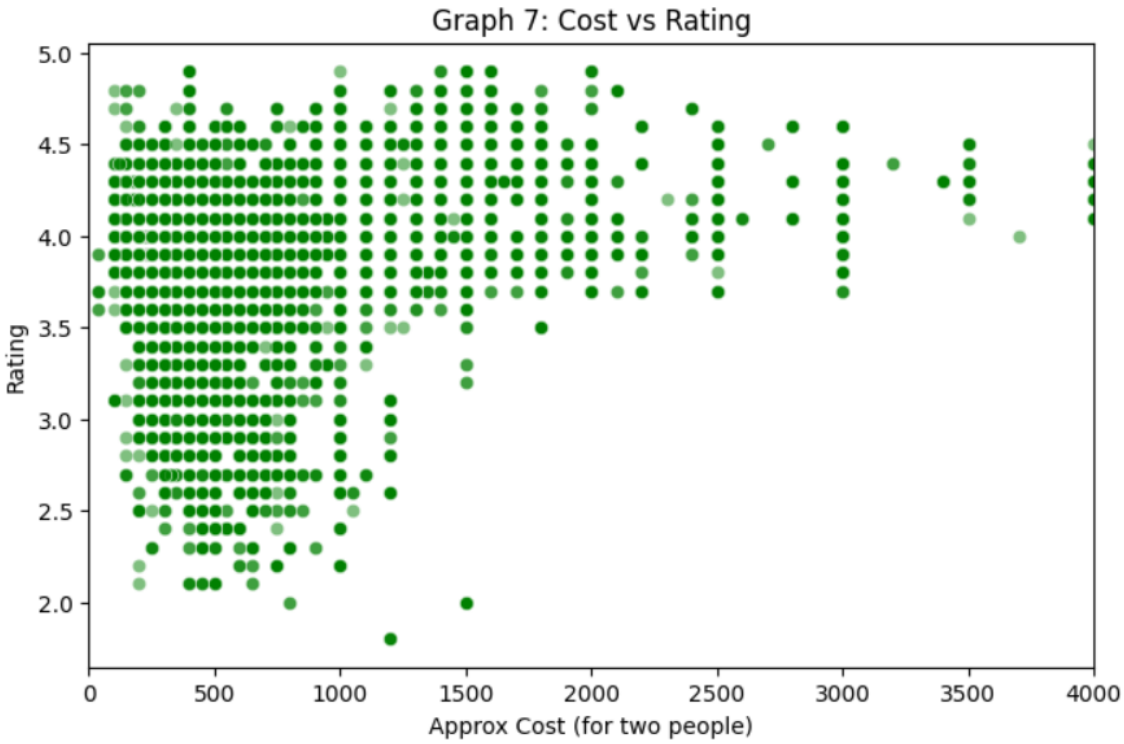

Graph 2: Top 10 Locations by Average Cost for Two
Average Cost
0
500
1000
1500
2000
2500
Sankey Road
MG Road
Lavelle Road
Race Course Road
Infantry Road
Richmond Road
Residency Road
Ulsoor
Langford Town
Old Airport Road
Location

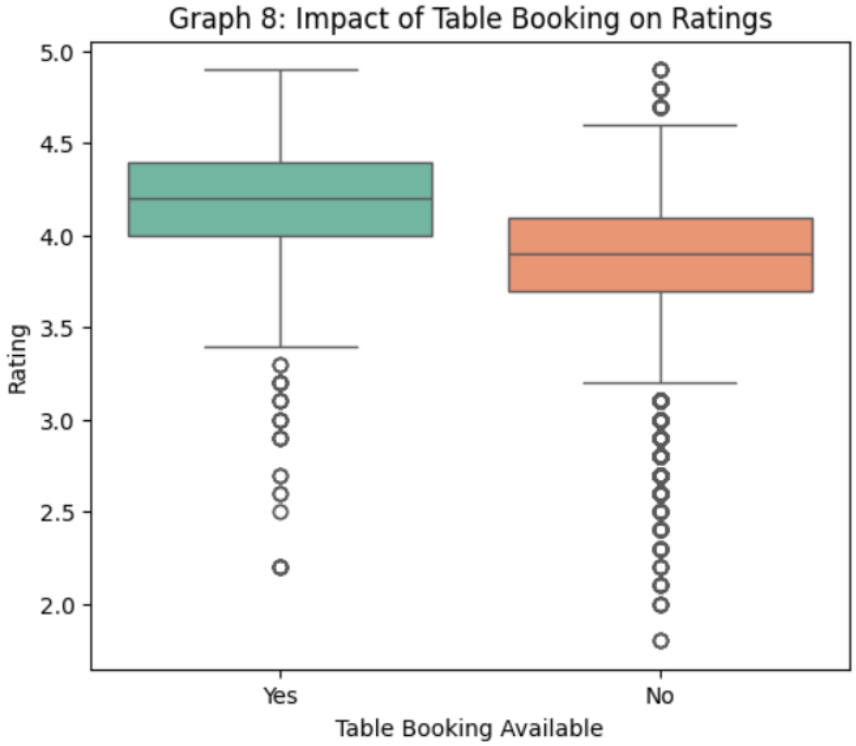

Graph 3: Ratings Distribution (1.0 to 5.0)
Frequency
Rating
0
1000
2000
3000
4000
5000
6000
2.0
2.5
3.0
3.5
4.0
4.5
5.0

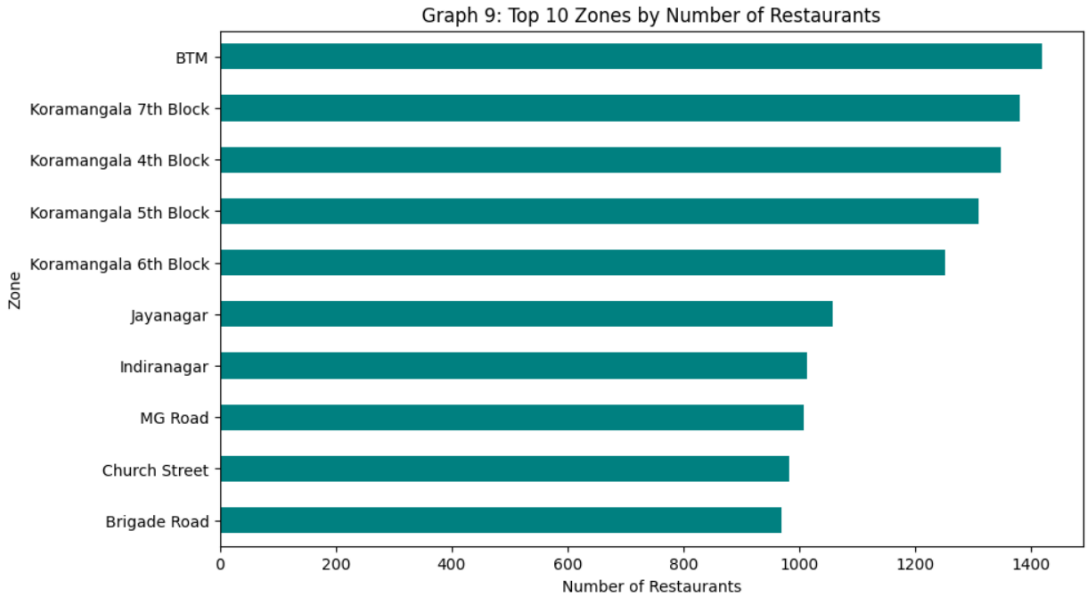

Graph 4: Top 10 Cuisine Combinations
Count
Cuisines
0
200
400
600
800
1000
1200
North Indian
North Indian, Chinese
South Indian
Cafe
South Indian, North Indian, Chinese
Bakery, Desserts
Desserts, Beverages
Ice Cream, Desserts
Chinese
Desserts


Graph 5: Correlation Heatmap (Votes, Cost, Rate)
votes
approx_cost(for two people)
rate
1.00
0.31
0.38
0.31
1.00
0.31
0.38
0.31
1.00
votes
approx_cost(for two people)
rate
1.0
0.9
0.8
0.7
0.6
0.5
0.4

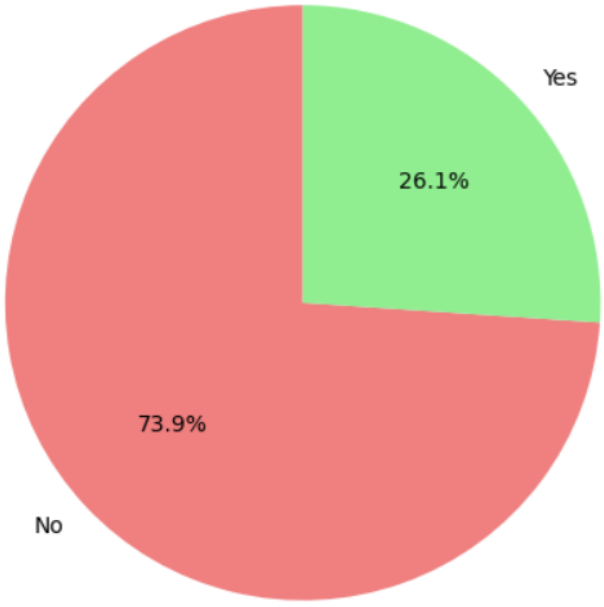

Graph 6: Top 10 Most Popular Restaurant Chains
Number of Outlets
0
10
20
30
40
50
60
70
80
Onesta
Empire Restaurant
KFC
McDonald's
Pizza Hut
Domino's Pizza
Faasos
Smally's Resto Cafe
Corner House Ice Cream
Burger King

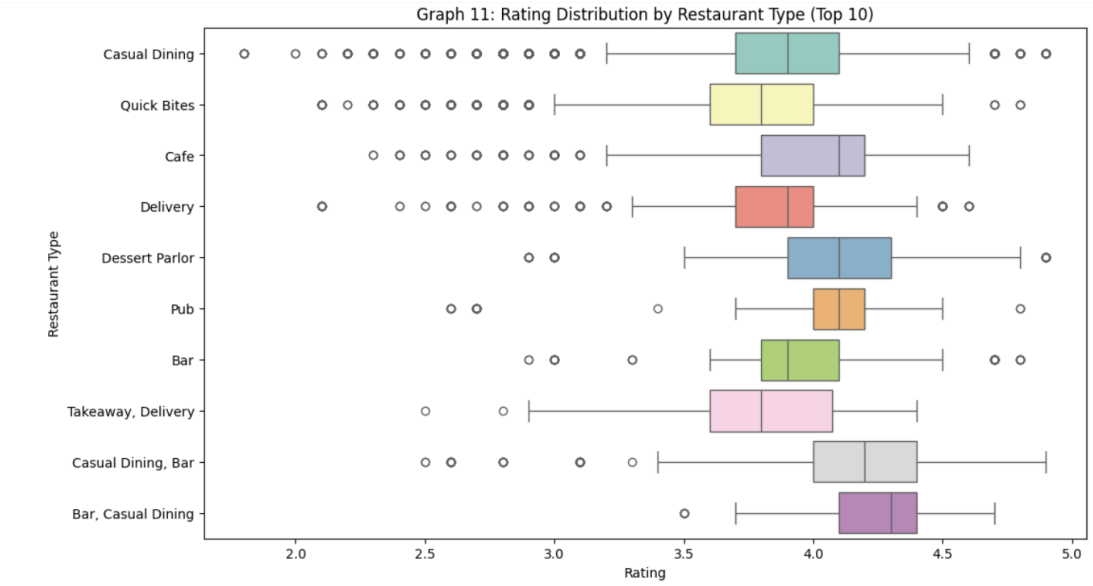

Graph 7: Cost vs Rating
Rating
Approx Cost (for two people)
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0
500
1000
1500
2000
2500
3000
3500
4000

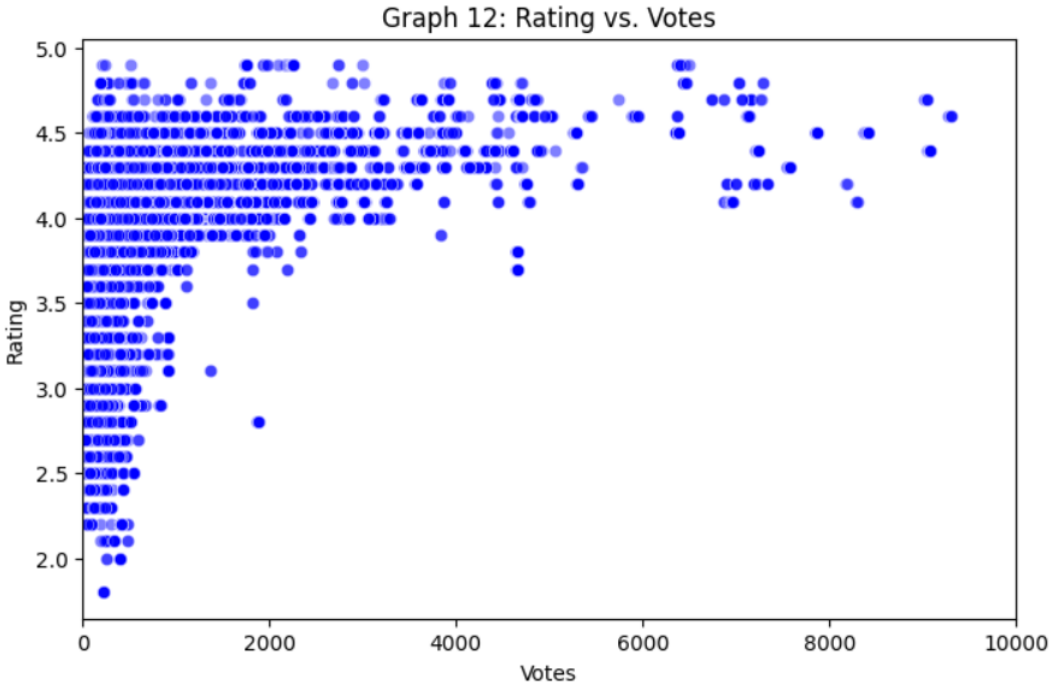

Graph 8: Impact of Table Booking on Ratings
Rating
Table Booking Available
Yes
No
5.0
4.5
4.0
3.5
3.0
2.5
2.0



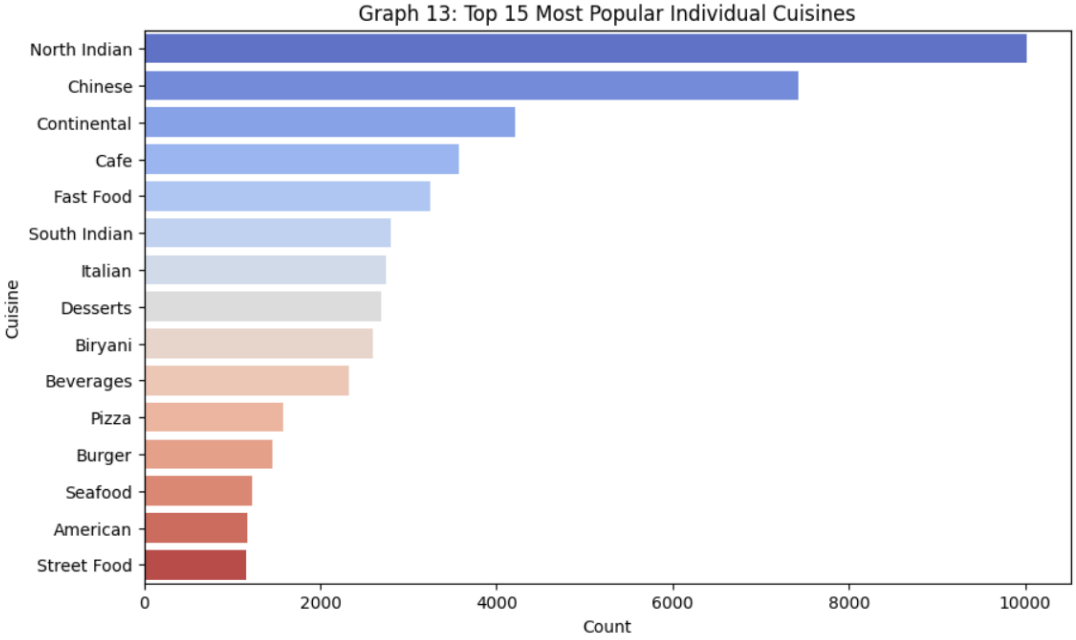
Graph 10: Percentage of Restaurants Offering Table Booking



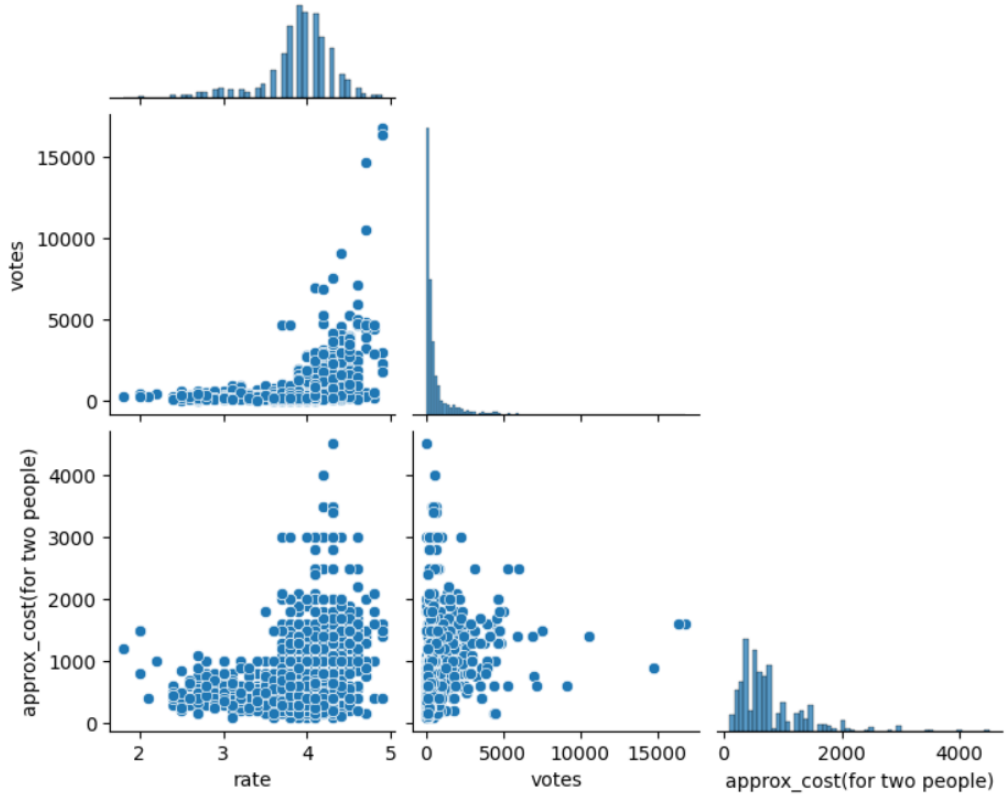

Graph 11: Rating Distribution by Restaurant Type (Top 10)
Casual Dining
Quick Bites
Cafe
Delivery
Dessert Parlor
Pub
Bar
Takeaway, Delivery
Casual Dining, Bar
Bar, Casual Dining
Restaurant Type
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Rating


Graph 12: Rating vs. Votes
Rating
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0
2000
4000
6000
8000
10000
Votes

Graph 13: Top 15 Most Popular Individual Cuisines



Graph 14: Word Cloud for Liked Dishes


votes
approx_cost(for two people)
rate
votes
approx_cost(for two people)
0
5000
10000
15000
0
1000
2000
3000
4000
2
3
4
5
0
2000
4000

Conclusion

This project successfully demonstrates how structured data analysis can transform raw restaurant data into meaningful business insights.

Key conclusions:

- Online ordering and table booking increase customer satisfaction and engagement.
- Bangalore has diverse food culture dominated by fast food, North Indian, and Chinese cuisines.
- Locality strongly affects restaurant performance, with premium areas having higher ratings and cost.
- Restaurants in highly commercial areas show tough competition and require better services to stand out.
- The data cleaning process plays a crucial role in improving data quality before analysis.
- The project provides a strong base for future extensions like:
    - Restaurant recommendation system
    - Price prediction models
    - Sentiment analysis from reviews
    - Rating prediction using ML

GitHub Link

https://github.com/Devanshi-Jindal11/zomato_data_analysis