# Zomato Dataset Analysis

**Submitted by:**
**Anubhav Singh Pathania-2024010015**
**Yash Arora-2024010120**

**MCA 2nd Year**

Submitted to:
Dr. Anjula Mehto
Assistant Professor

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

**Computer Science and Engineering Department**
**Thapar Institute of Engineering and Technology, Patiala**

**November 2025**

# TABLE OF CONTENTS

# Introduction or Project Overview

This project analyzes a large Zomato restaurant dataset for Bengaluru city to extract actionable insights about the local foodservice industry. We process and visualize over 50,000 restaurant listings to identify patterns in cuisine preferences, pricing, service offerings, and geographical distribution. By employing Python-based data science techniques (such as data cleaning, exploratory analysis, and visualization), the project highlights trends that can guide customers and businesses. The analysis aims to reveal how factors like location, cuisine type, and available services (e.g. online delivery or table booking) influence customer ratings and popularity in Bengaluru.

# Problem Statement

Bengaluru’s dining scene is diverse and rapidly growing, but understanding it quantitatively is challenging due to the volume and complexity of the data. Key questions — such as which neighborhoods have the most or best-rated restaurants, which cuisines dominate customer demand, and how pricing relates to popularity — are difficult to answer without systematic analysis. Customers lack clear guidance on finding affordable, high-quality dining options, and restaurant owners need data-driven insights to make strategic decisions about menus, pricing, and services. This project addresses these challenges by transforming raw restaurant listings into clear, data-driven observations. We systematically analyze the Zomato dataset to uncover trends that help stakeholders make informed choices.

# Overview of the Dataset used

1. **Scope:** The dataset contains over 50,000 restaurant entries in Bengaluru, India, spanning 17 columns of information. Each row represents a unique Zomato-listed restaurant in the city.

2. **Location Data:** Key fields include the restaurant's *name*, *address*, *locality*, and *city*. The dataset focuses specifically on Bengaluru's neighborhoods.

3. **Service Availability:** Attributes such as *online_order* and *book_table* indicate whether each restaurant offers online delivery or accepts reservations. These binary fields help distinguish service options available to customers.

4. **Cuisine and Cost:** The *cuisines* column lists the cuisine types served (e.g., North Indian, Chinese). There is also an *approximate cost for two* field (in Indian rupees), allowing analysis of price ranges across restaurants.

5. **Ratings and Popularity:** Each entry includes an aggregated *rating* (on a numerical scale) and the number of *votes* (user reviews). These metrics serve as proxies for customer satisfaction and restaurant popularity.

6. **Restaurant Type and Other Features:** Additional columns capture *restaurant type* (e.g., casual dining, cafe, quick bites) and other service details. Some entries also list popular *dishes* or provide links (e.g., menu URLs, phone numbers), though these were secondary for our analysis.

7. **Data Source:** This publicly available dataset (from Zomato/Kaggle) provides a comprehensive, real-world snapshot of Bengaluru's restaurant landscape suitable for exploratory analysis and modeling.

## Project Workflow

1. **Data Ingestion and Preprocessing:** We loaded the Zomato CSV dataset into Python (using Pandas) and performed an initial exploration of its structure. This involved verifying row and column counts and separating variables into categorical (e.g. *cuisines*, *online_order*) and numeric (e.g. *votes*). We cleaned the data by handling missing values (for example, imputing or removing entries without key information), removing duplicate records, and renaming columns for clarity. In particular, we standardized formats for columns like *approx_cost* (converted from strings to numeric values) and parsed ratings (cleaning entries such as “3.5/5” into a numeric scale). This step ensured the dataset was consistent and ready for analysis.

2. **Exploratory Data Analysis (EDA):** With a clean dataset, we conducted statistical summaries and visual explorations of major features. We calculated descriptive statistics (means, counts, distribution ranges) for numerical fields like cost and ratings. We then created visualizations to uncover patterns: bar charts and pie charts compared the count of restaurants offering online delivery versus those not, and those allowing table booking versus those that do not. We also examined how table-booking availability relates to ratings. Location-wise, we used bar plots to show the number of restaurants by locality and identified the areas with the highest concentration of establishments. We grouped data to analyze distributions across *restaurant types* (casual dining, cafes, fast food, etc.) and plotted rating distributions to understand customer satisfaction trends. Histograms of *approximate cost* illustrated the overall price range distribution. These EDA steps revealed where restaurants cluster, which services are common, and how basic features (like cost and rating) vary across the dataset.

3. **Cuisine and Cost Analysis:** We delved deeper into customer-centric factors by analyzing cuisines and pricing. Using string parsing on the *cuisines* column, we identified the most popular cuisine categories and common cuisine pairings (for example, how often North Indian cuisine is paired with Chinese, given both are frequently listed together). We found that North Indian and Chinese cuisines dominate the offerings in the dataset. We also examined the *approximate cost* metric in context: for example, grouping restaurants into low-, medium-, and high-priced brackets (e.g., ₹300–₹800 being mid-range). Within each price bracket, we calculated average ratings and vote counts. This allowed us to identify value-for-money restaurants (those with high ratings but relatively low cost) and to check whether mid-priced venues tend to attract more customer votes compared to luxury outlets. Such analysis helps highlight affordable yet highly-regarded options and shows how price correlates with popularity.

4. **Service Features and Chain Analysis:** We investigated how operational features affect restaurant performance. An important finding was that restaurants offering online delivery or table booking generally earn higher ratings and vote counts, suggesting these services positively impact customer satisfaction. We quantified this by comparing average ratings of restaurants with and without these features. Additionally, we analyzed the presence of major restaurant chains across Bengaluru. By filtering names, we identified that popular chains such as *Domino's Pizza*, *McDonald's*, and *Café Coffee Day* appear frequently in the dataset. We examined these chains' average ratings and vote counts; their widespread presence and high vote totals indicate strong market penetration. This step of the workflow a business perspective, showing that established chains consistently perform well and that mid-priced chains tend to accumulate more customer feedback than standalone luxury venues.

5. **Geographical Insights:** To capture spatial patterns, we employed geographical visualization. Using libraries like Plotly and Folium, we plotted restaurant locations on a map of Bengaluru. These interactive maps visualized restaurant density and overlaid performance metrics (such as average rating or vote count by area). From this mapping, we identified hotspots of premium dining — for example, neighborhoods like *Koramangala* and *Indiranagar* emerged as high-density areas for well-rated restaurants. We also observed that tier-2 or suburban localities tend to have lower average costs for two (suggesting more affordable dining) compared to central urban areas. Additionally, we visualized cuisine distribution by area: for instance, South Indian cuisine dominates in certain localities (often near traditional centers), whereas fast-food outlets are widespread around student-populated regions. These geographic analyses complement the statistical results by showing *where* certain trends occur, revealing the city zones that are underserved or saturated in various categories.

6. **Validation and Use Cases (Findings Interpretation):** Finally, we synthesized the findings and considered their implications. The patterns uncovered (e.g., popular cuisines, pricing tiers, service impact) were evaluated for logical consistency and relevance. We discussed potential use cases: for instance, entrepreneurs could use the insights to decide where to open a new restaurant and what type of cuisine to offer based on demand hotspots. We also noted limitations: the analysis is based on a static snapshot of Zomato listings, so it may not capture very recent changes or user reviews outside the dataset. Moreover, unlisted or new establishments are not represented. We flagged these as areas for future improvement (such as integrating live Zomato API data or extending the study to other cities). By reflecting on these considerations, we ensured the workflow not only performs the computations but also frames the results in practical and ethical context.

## Results

1. **Distribution by Location:** We found that a large proportion of restaurants are concentrated in central and high-density neighborhoods of Bengaluru. Major commercial and residential hubs host the majority of listings, while peripheral areas have fewer restaurants. Notably, Koramangala and Indiranagar stood out as prime dining districts with a high density of top-rated establishments.
2. **Service Offerings:** Over half of the restaurants in the dataset provide **online delivery**, whereas significantly fewer offer **table reservations**. Restaurants with delivery and booking services tend to have higher average ratings and vote counts, indicating these features enhance customer satisfaction and popularity.
3. **Cuisine Popularity:** Analysis of the cuisines column showed that **North Indian** and **Chinese** cuisines are the most commonly offered, reflecting broad local demand. Combinations of cuisines (e.g. "North Indian, Chinese") are also frequent. South Indian and continental cuisines appear regularly as well, but North/Chinese jointly dominate the market.
4. **Pricing Patterns:** The *approximate cost for two* is mostly in the moderate range (around ₹300–₹800), suggesting that Bengaluru's restaurant market caters primarily to mid-budget diners. Few restaurants fall into very low-cost or very high-cost extremes. Within each price segment, we saw that mid-priced restaurants generally accumulated more customer votes than upscale eateries, implying greater volume of patronage. We also identified *value-for-money* restaurants by cross-referencing low cost and high rating, highlighting affordable gems that rank highly.
5. **Ratings and Popularity:** A clear positive trend emerged between customer votes and ratings: higher-rated restaurants typically have more votes. This consistency suggests that popular consensus aligns with numeric ratings, validating the reliability of the rating system. Fast-food and casual dining venues (including major chains) often appear at the top in both votes and ratings, due to their large customer base.
6. **Restaurant Types and Chains: Casual dining** and quick-service restaurants are the most common types. Among them, well-known chains like **Domino's Pizza**, **McDonald's**, and **Café Coffee Day** feature prominently across neighborhoods. These chains collectively command a large share of customer reviews, indicating their strong presence. The analysis also showed that restaurants providing multiple services (e.g. home delivery and dine-in) tend to be more popular than single-service outlets.
7. **Key Insight Summary:** In summary, Bengaluru's dining trends include a concentration of restaurants in urban hubs, a preference for North Indian and Chinese food, a mid-range price point for most eateries, and a strong link between service offerings and higher ratings. These findings provide a quantitative overview of the city's restaurant landscape and can guide further strategic decisions.


Online Delivery Availability (%):
online_order
Yes 66.666667
No 33.333333

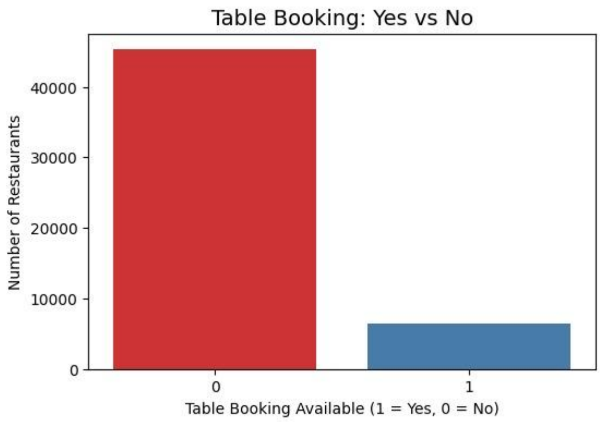
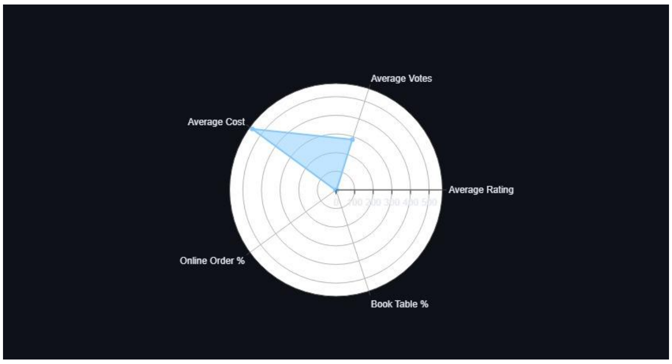
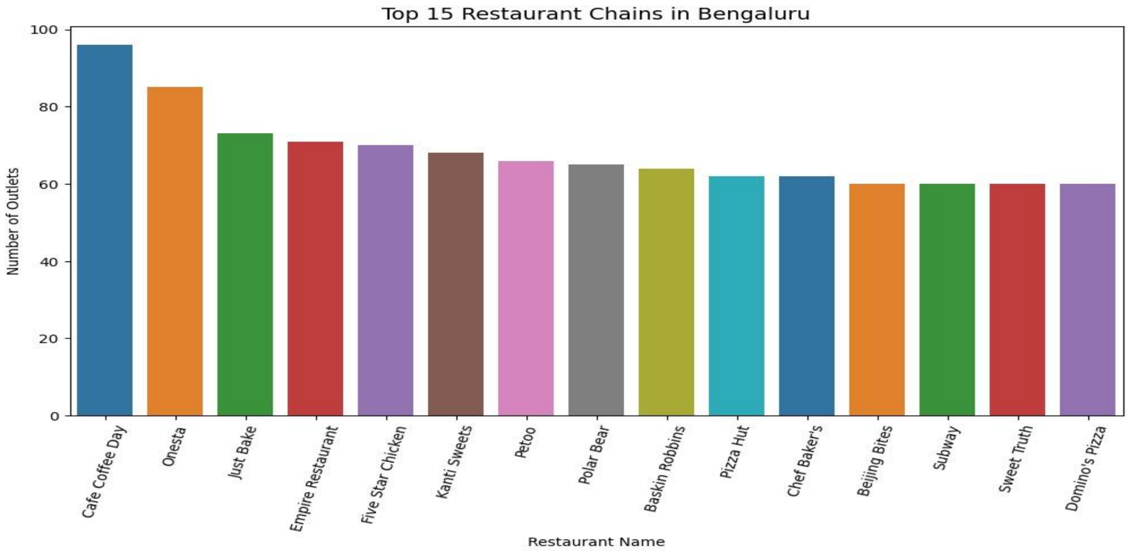
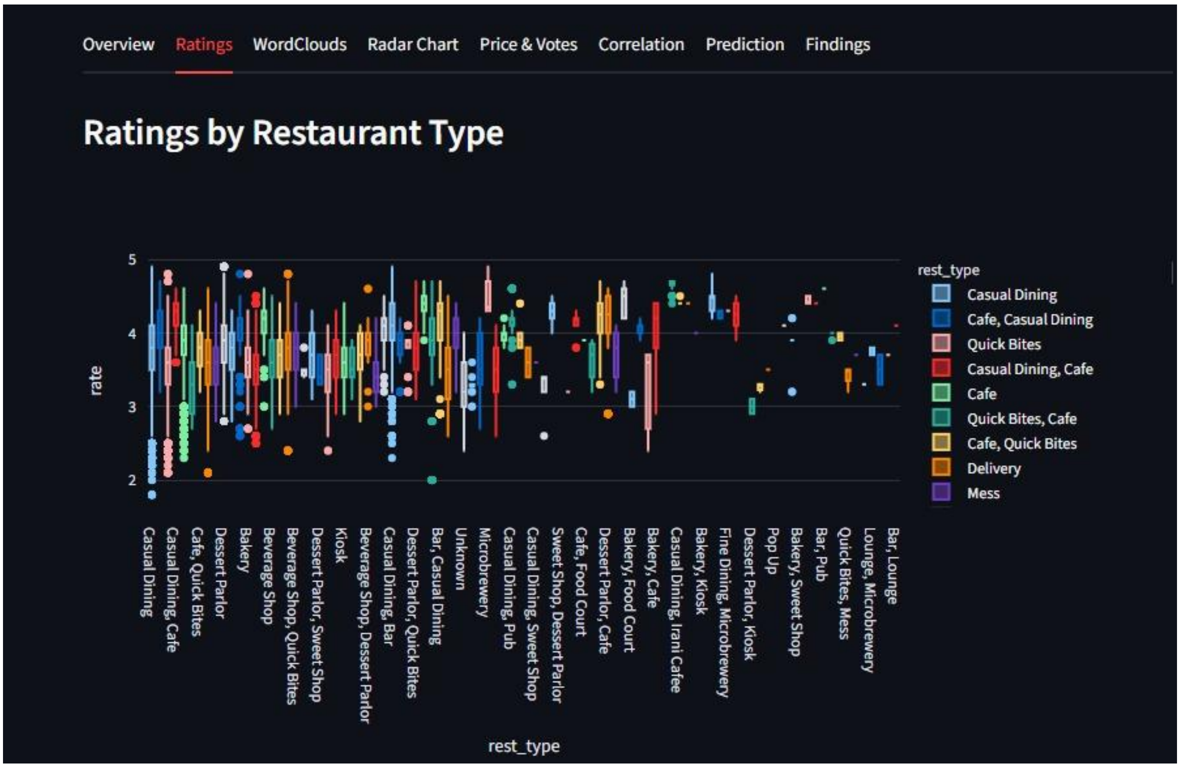
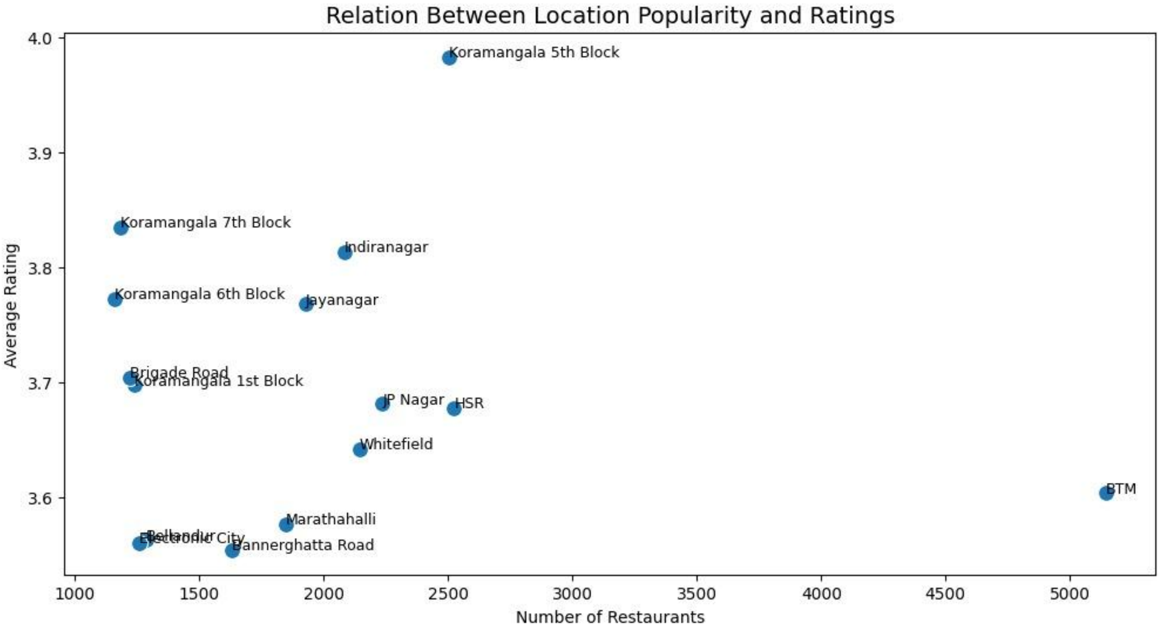



Table Booking Availability (%):
book_table
0 87.530213
1 12.469787




Top 15 Restaurant Chains in Bengaluru
Number of Outlets
100
80
60
40
20
0
Cafe Coffee Day
Onesta
Just Bake
Empire Restaurant
Five Star Chicken
Kanti Sweets
Petoo
Polar Bear
Baskin Robbins
Pizza Hut
Chef Baker's
Beijing Bites
Subway
Sweet Truth
Domino's Pizza
Restaurant Name


Overview
Ratings
WordClouds
Radar Chart
Price & Votes
Correlation
Prediction
Findings
Ratings by Restaurant Type
rate
5
4
3
2
rest_type
Casual Dining
Cafe, Casual Dining
Quick Bites
Casual Dining, Cafe
Cafe
Quick Bites, Cafe
Cafe, Quick Bites
Delivery
Mess
Casual Dining
Casual Dining, Cafe
Cafe, Quick Bites
Dessert Parlor
Bakery
Beverage Shop
Beverage Shop, Quick Bites
Dessert Parlor, Sweet Shop
Kiosk
Beverage Shop, Dessert Parlor
Casual Dining, Bar
Dessert Parlor, Quick Bites
Bar, Casual Dining
Unknown
Microbrewery
Casual Dining, Pub
Casual Dining, Sweet Shop
Sweet Shop, Dessert Parlor
Cafe, Food Court
Dessert Parlor, Cafe
Bakery, Food Court
Bakery, Cafe
Casual Dining, Irani Cafee
Bakery, Kiosk
Fine Dining, Microbrewery
Dessert Parlor, Kiosk
Pop Up
Bakery, Sweet Shop
Bar, Pub
Quick Bites, Mess
Lounge, Microbrewery
Bar, Lounge
rest_type


Relation Between Location Popularity and Ratings
Average Rating
Number of Restaurants
4.0
3.9
3.8
3.7
3.6
1000
1500
2000
2500
3000
3500
4000
4500
5000
Koramangala 5th Block
Koramangala 7th Block
Indiranagar
Koramangala 6th Block
Jayanagar
Brigade Road
Koramangala 1st Block
JP Nagar
HSR
Whitefield
BTM
Marathahalli
Bellandur
Electronic City
Bannerghatta Road


Sea Food
Biryani Chicken
Peri Peri
Salad
Masala Dosa
Chicken Tikka
Chicken Chicken
Panneer Butter
Chicken Grill
Chicken Wings
Paneer Tikka
Butter Chicken
Pav Bhaji
Peri Chicken
Vegetable Biryani
French Fries
Mutton Biryani
Hot Chocolate
Tandoori Chicken
Garlic Bread
Chicken Biryani
Hyderabadi Biryani
Filter Coffee
Chicken Curry
Chilli Chicken
Noodle
Dal Makhani
Ice Cream
Gulab Jamun
Long Island
Breakfast Buffet
Paratha
Craft Beer
Chicken Momo
Ginger Tea
Fish Curry
Murgh Ghee
Chocolate Fudge
Chicken Sandwich
Manchow Soup
Thick Shakes
Burgers Pasta
Veg Thali
Neer Dosa
Ghee Rice
Ghee Roast
Pasta Pizza
Lunch Buffet
Pizza Pasta
Butter Masala
Vada Pav
Dum Biryani
Roll
Biryani Mutton
Chicken Kebab
Chocolate Cake
Island Iced
Butter Naan
Potato Wedge
Spring Roll
Beer Cocktail
Cup Cake
Momo
Thali
Burger
Chicken
Iced Tea
Chaat
Roti
Fried Rice
Chop Suey
Chole Bhature
Nacho
Fish Fry
Pasta Burgers
Sweet Lassi
Rice Bowl
Onion Rings
Kerala Parotta
Aloo Paratha

# Conclusion

This project demonstrates how a real-world restaurant dataset can be cleaned, explored, and analyzed to extract actionable insights. By systematically processing the Zomato listings for Bengaluru using Python tools, we have bridged the gap between raw data and operational knowledge in the food industry. The analysis highlights that customers seeking value can focus on mid-priced, high-rated restaurants (often offering delivery services), and it reveals which cuisines and neighborhoods dominate the market. Customers can use these insights to discover affordable, top-rated dining options and popular cuisine pairings. Restaurant owners and investors can leverage the findings to refine menus, set competitive pricing, and choose strategic locations (for example, by identifying underserved localities). Identified market trends (such as emerging cuisine demands and price sensitivities) also suggest opportunities for new restaurant ventures in high-demand areas. Limitations of the study include its reliance on a static dataset for one city and the absence of real-time user feedback. Future work could integrate live data (e.g. via the Zomato API) and build predictive models (such as machine learning classifiers for restaurant success factors) to further enhance decision support. Overall, the project illustrates the value of data-driven analysis in the hospitality sector, providing a transparent approach that can adapt and extend to other cities or industries.

## GitHub Link

**Access the dashboard and source code:**

https://github.com/21anubhav/Zomato-EDA

https://github.com/Arora-Yash-coder/mca308/tree/main/zomato_eda