PROJECT REPORT: EXPLORATORY DATA ANALYSIS OF ZOMATO DATASET

This project involves an exploratory data analysis (EDA) of a dataset containing over 210,000 Indian restaurant entries. The goal is to uncover insights related to restaurant ratings, popular cuisines, price ranges, service offerings, and customer behavior to support decision-making in the food service industry.

DATASET OVERVIEW

This project uses the Indian Restaurants dataset from Zomato, offering comprehensive data on food establishments across Indian cities. It supports analysis on customer behavior, pricing, service quality, and dining preferences.

📂 Source: Zomato (public dataset for EDA in hospitality)
📊 Size: 211,944 rows × 26 columns
🛠️ Preprocessed: Duplicates & null values handled

Key Column Categories

1. Basic Details
name, city, address, establishment, cuisines

2. Service Info
online_order, book_table, delivery, takeaway, highlights

3. Ratings
aggregate_rating, rating_text, votes

4. Cost & Pricing
average_cost_for_two, price_range

DATA CLEANING & PREPROCESSING

Dropped irrelevant columns:

zipcode, address, opentable_support, url, country_id, takeaway, locality.

Null values handled:

cuisines filled with “regional cuisine”

timings filled with “not available”

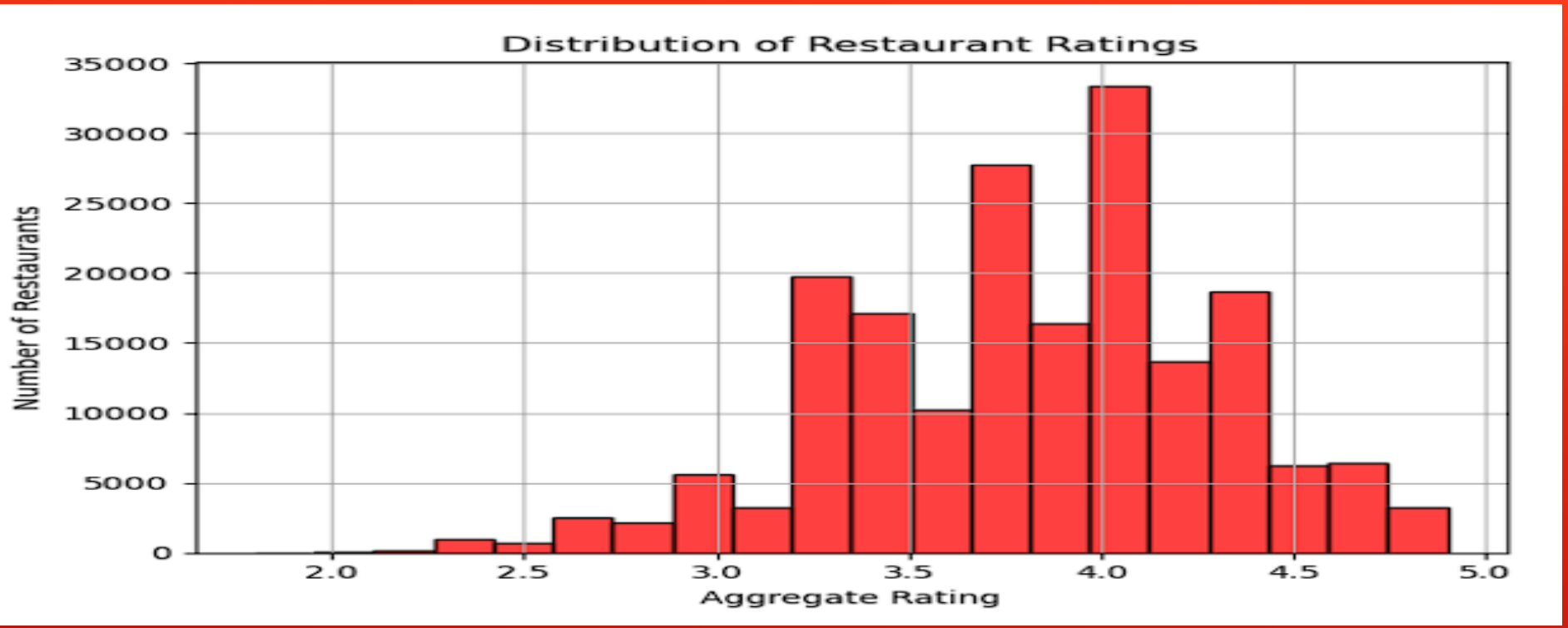
Encoded service availability:

delivery: Replaced -1 and 1 with "No" and "Yes"

Removed inconsistent or placeholder entries (delivery == 0)

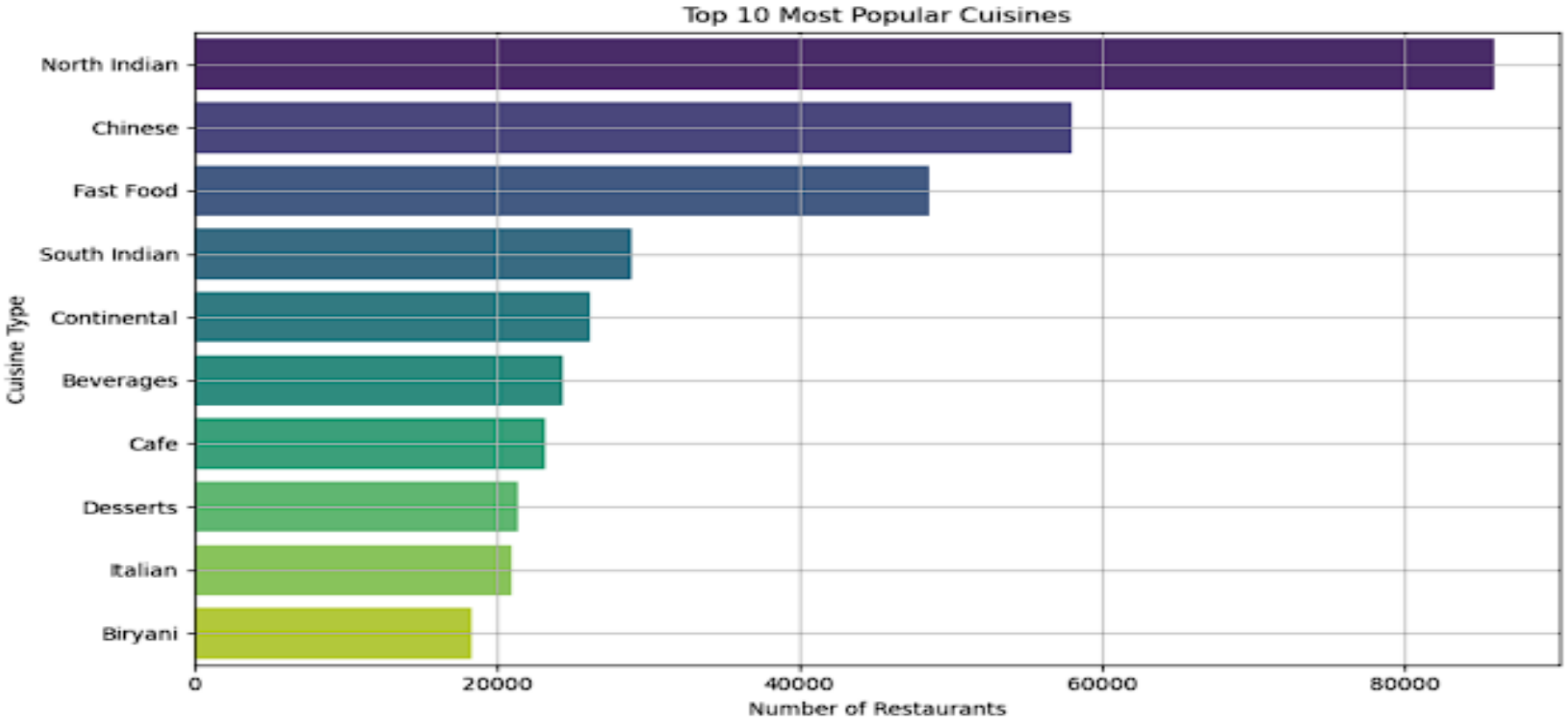
UNIVARIATE ANALYSIS

RATING DISTRIBUTION HISTOGRAM



A histogram on aggregate_rating shows most restaurants are rated between 3.0–4.5, with a right-skewed distribution. Many entries have 0 ratings, indicating unrated or inactive listings. This suggests overall user satisfaction is moderate to high, but many restaurants lack feedback. Zomato can improve user engagement by encouraging more ratings and reviews, especially for newer or low-visibility listings.

TOP 10 CUISINES BAR PLOT

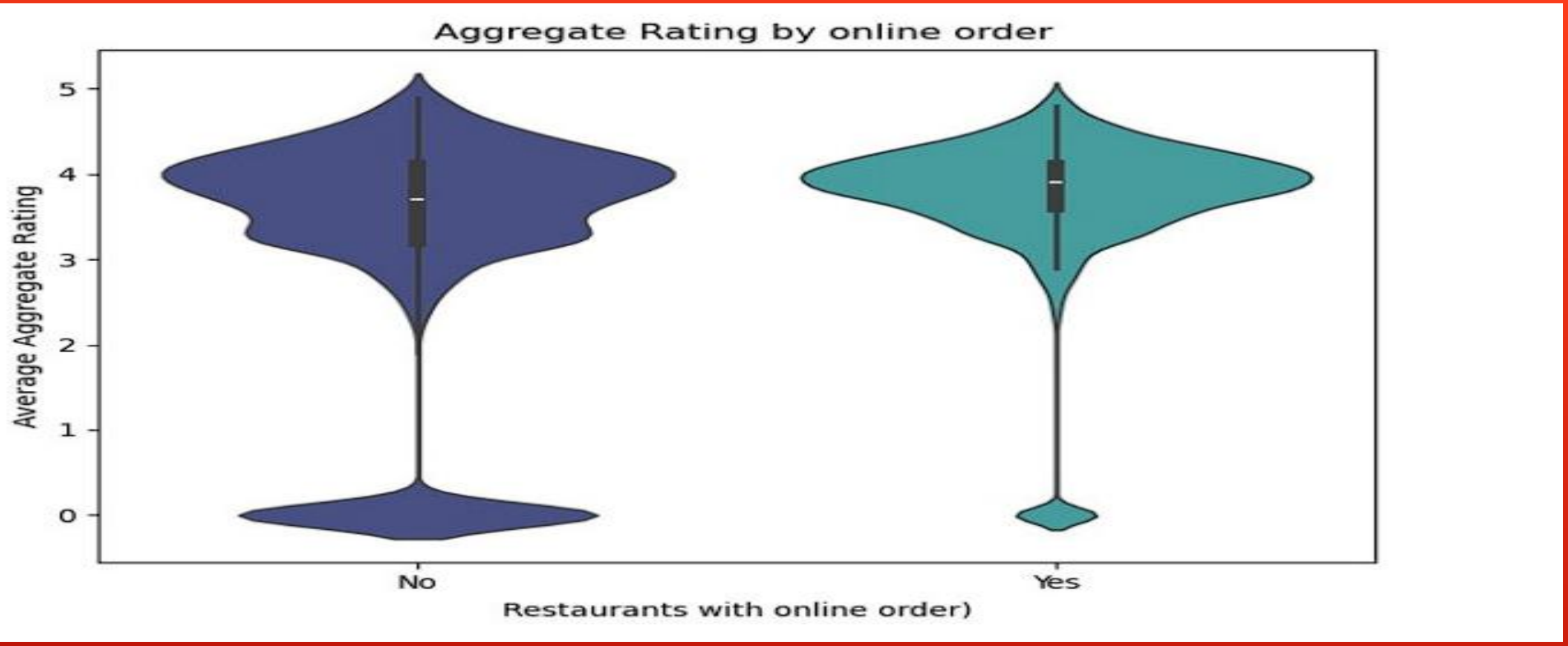


A bar plot of the cuisines column shows North Indian, Chinese, and Fast Food are the top 3 most common cuisines. Other popular choices include South Indian, Desserts, and Biryani. These results reflect strong regional and comfort food preferences. Restaurants can focus on these high-demand cuisines to attract a broader customer base.

AVERAGE RATING BY PRICE RANGE

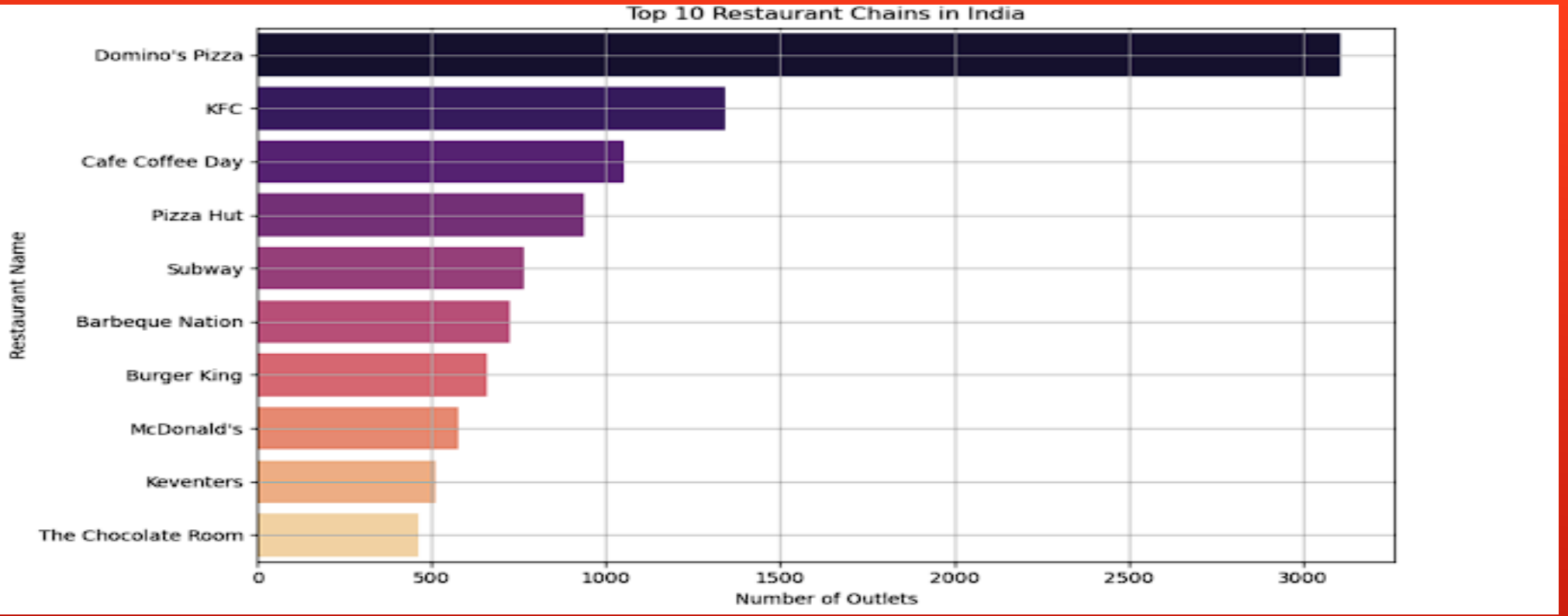
A bar plot shows that restaurants in Price Ranges 2 and 3 have the highest average ratings, indicating good value for money. Lower (1) and higher (4) price points receive slightly lower ratings. This suggests customers favor mid-range dining experiences. Zomato can promote well-rated, mid-priced restaurants for higher user satisfaction.

VIOLIN PLOT – DELIVERY VS RATING



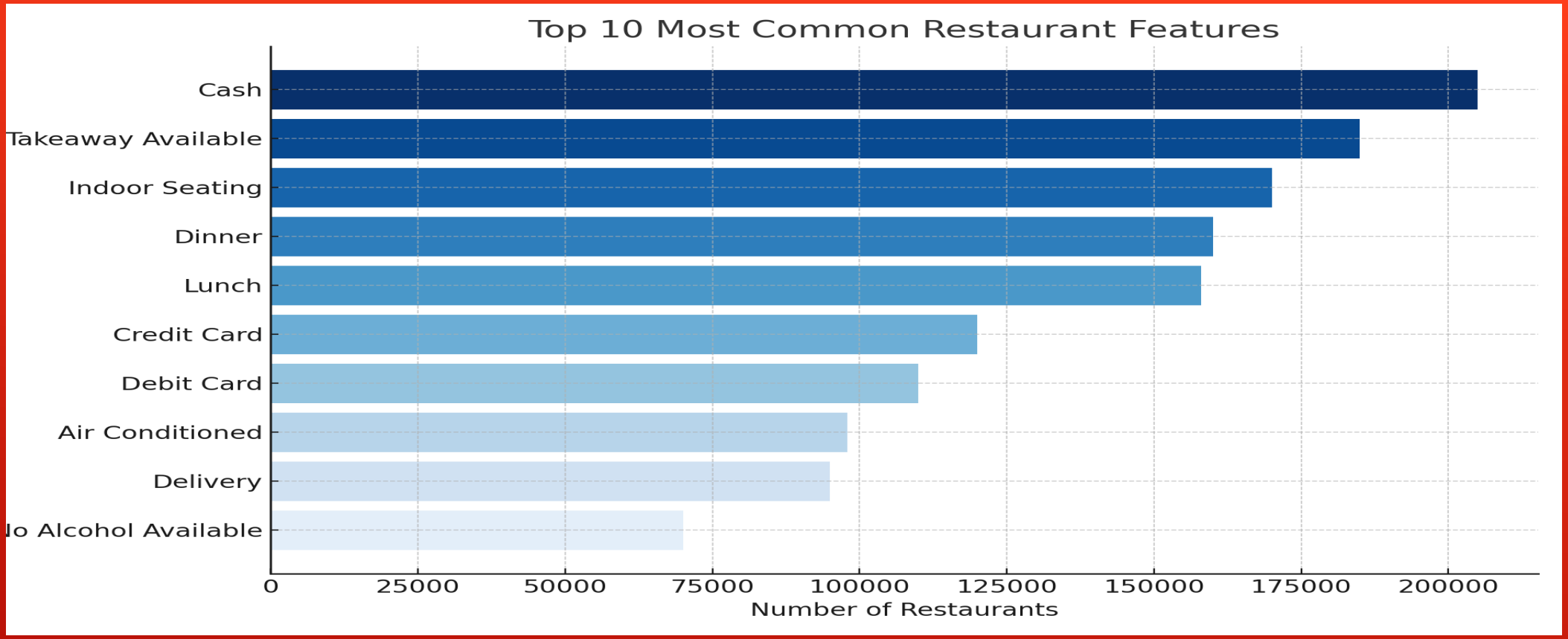
The violin plot shows that delivery-enabled restaurants tend to receive higher and more consistent ratings than those that don't. This highlights delivery as a key factor in positive customer experiences. Zomato can encourage non-delivery restaurants to adopt delivery services to improve performance and customer reach.

TOP 10 RESTAURANT CHAINS



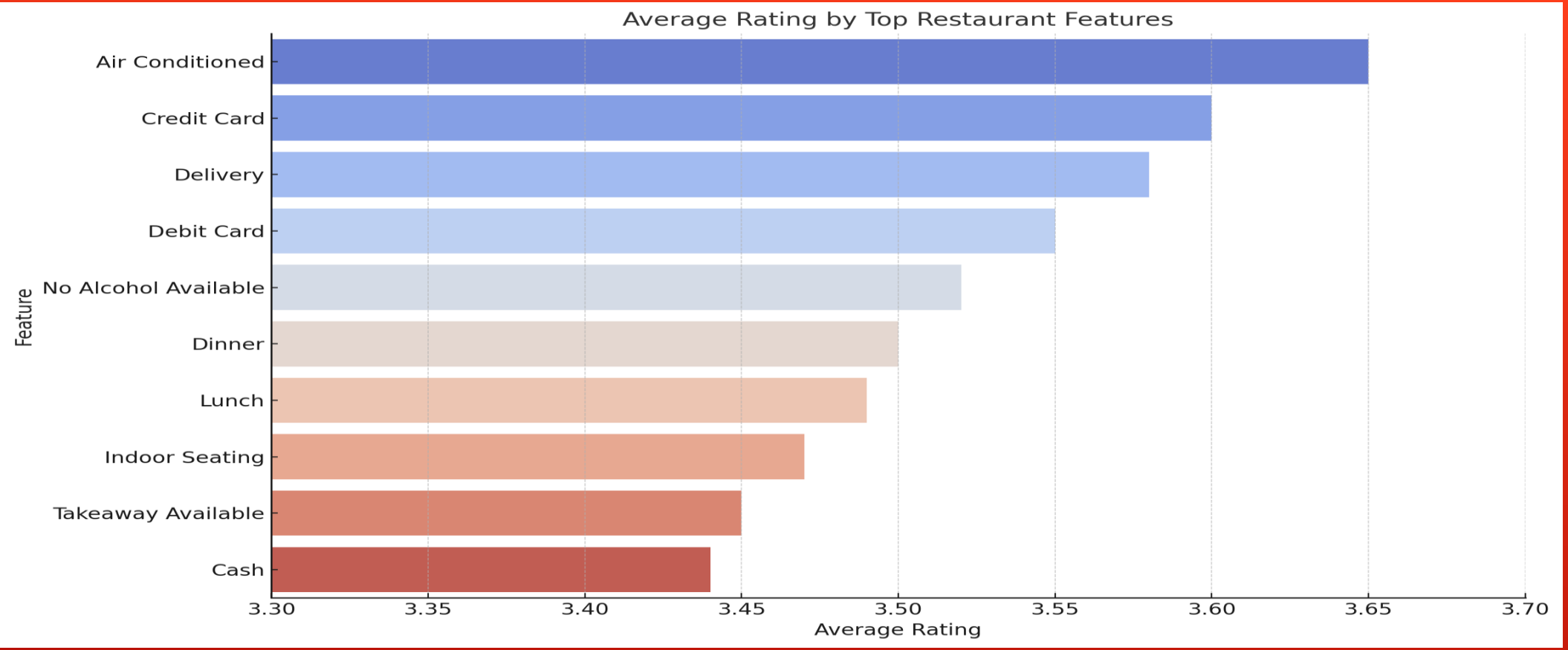
A bar plot highlights the top 10 restaurant chains by outlet count, with Domino's, Pizza Hut, and Cafe Coffee Day leading. These brands have strong pan-India presence and consistent market reach. Their success reflects brand trust and scalable business models. Zomato can feature such chains prominently and collaborate for national-level campaigns.

FEATURE POPULARITY VS AVERAGE RATING



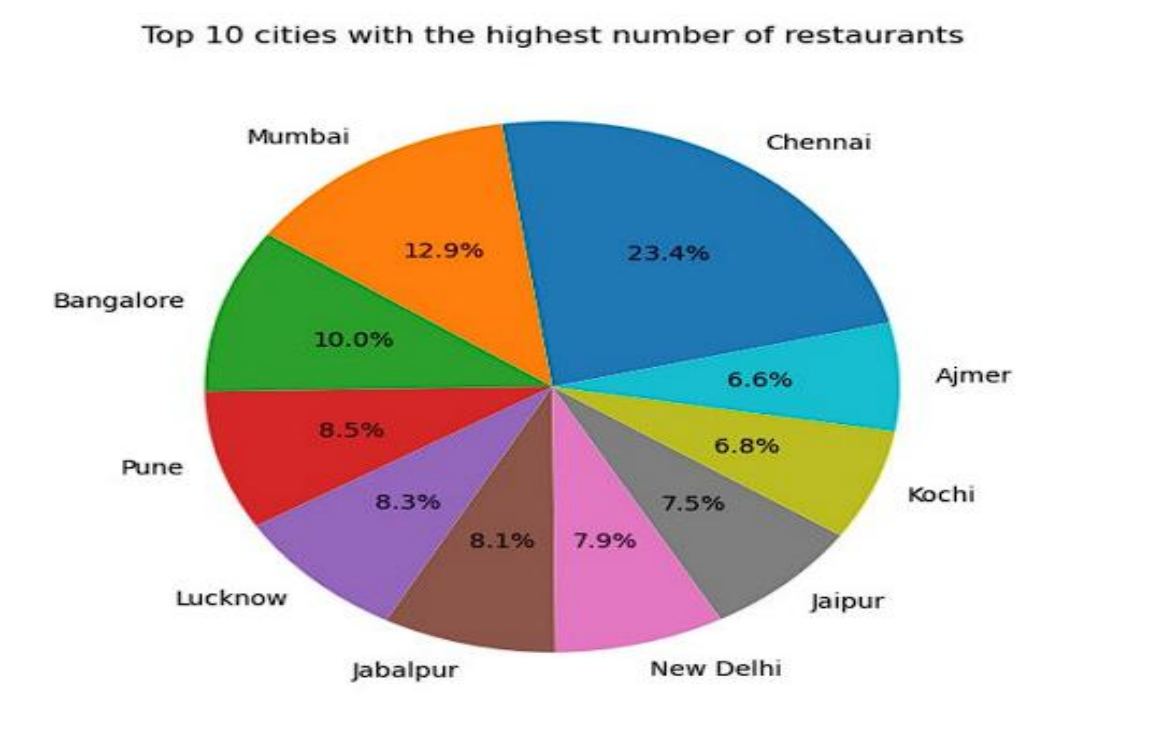
Most restaurants offer Cash, Takeaway, and Indoor Seating, indicating a focus on convenience.
Dinner and Lunch options are common, while Credit/Debit Card usage shows moderate adoption.
Less frequent features like Delivery and Air Conditioning offer opportunities for differentiation.
Emphasizing digital payments and service enhancements can improve competitiveness.

TOP 10 RESTAURANT FEATURES BY FREQUENCY



Most restaurants offer Cash, Takeaway, and Indoor Seating, indicating a focus on convenience.
Dinner and Lunch options are common, while Credit/Debit Card usage shows moderate adoption.
Less frequent features like Delivery and Air Conditioning offer opportunities for differentiation.
Emphasizing digital payments and service enhancements can improve competitiveness.

TOP 10 CITIES – PIE CHART

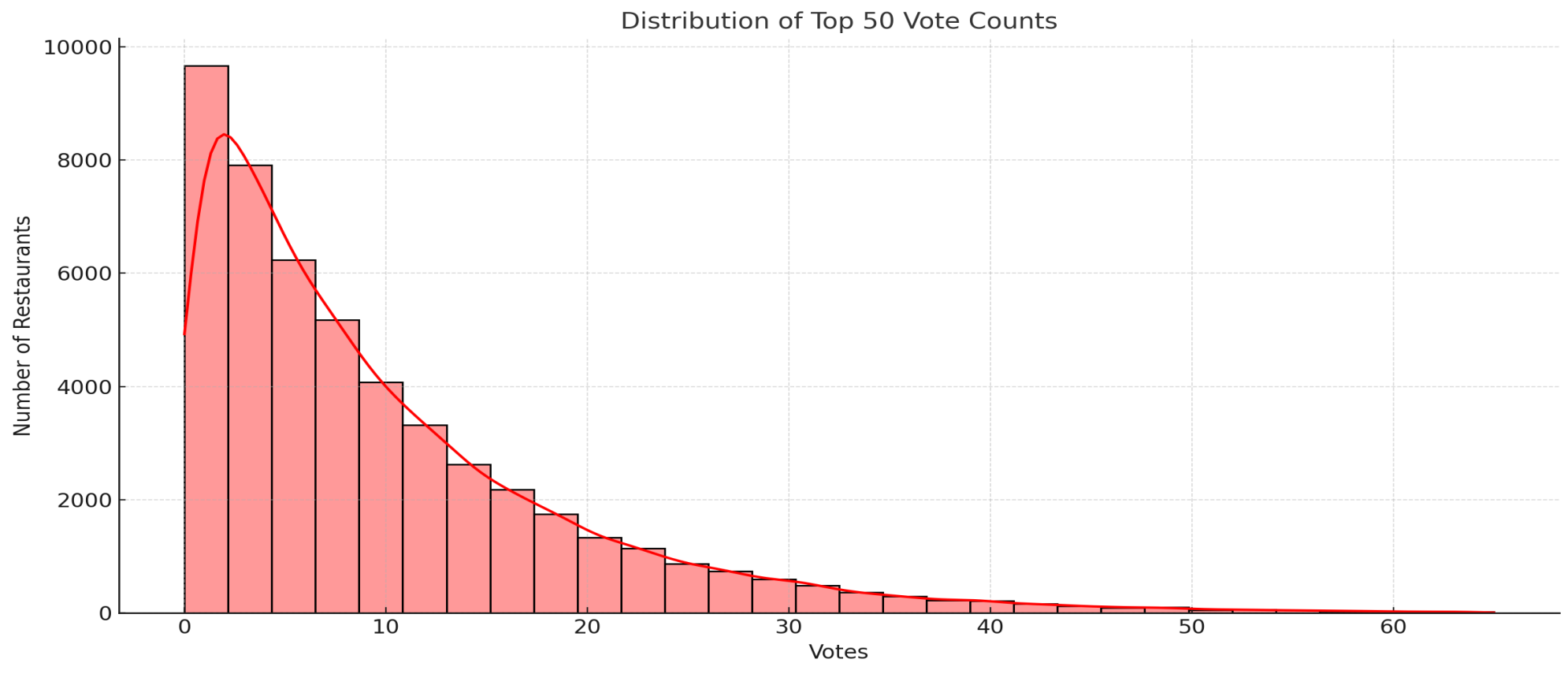


The pie chart shows that Chennai leads with 23.4% of the restaurants among the top 10 cities, followed by Mumbai (12.9%), Bangalore (10.0%), and Pune (8.5%). Other notable cities include Lucknow, Jabalpur, and New Delhi, each contributing around 8%.

This distribution suggests that Zomato's restaurant listings are heavily concentrated in metro and tier-1 cities, where dining and food delivery are more prevalent. However, the inclusion of Ajmer, Kochi, and Jabalpur also reflects growing Zomato adoption in tier-2 and tier-3 cities, signaling potential for regional expansion.

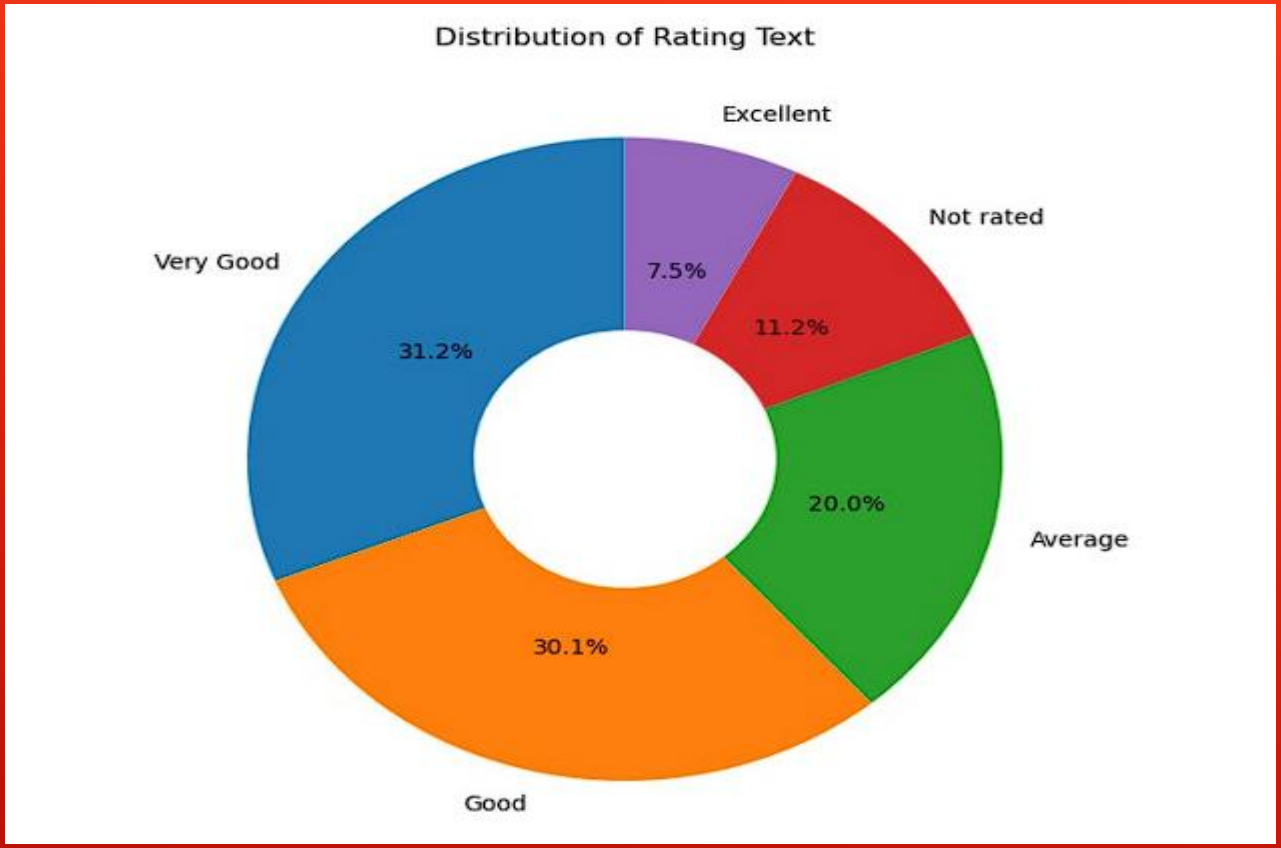
📌 Conclusion: Zomato can maintain its stronghold in major metros while investing in marketing and onboarding in emerging cities to unlock untapped growth.

TOP 10 RESTAURANT FEATURES BY FREQUENCY



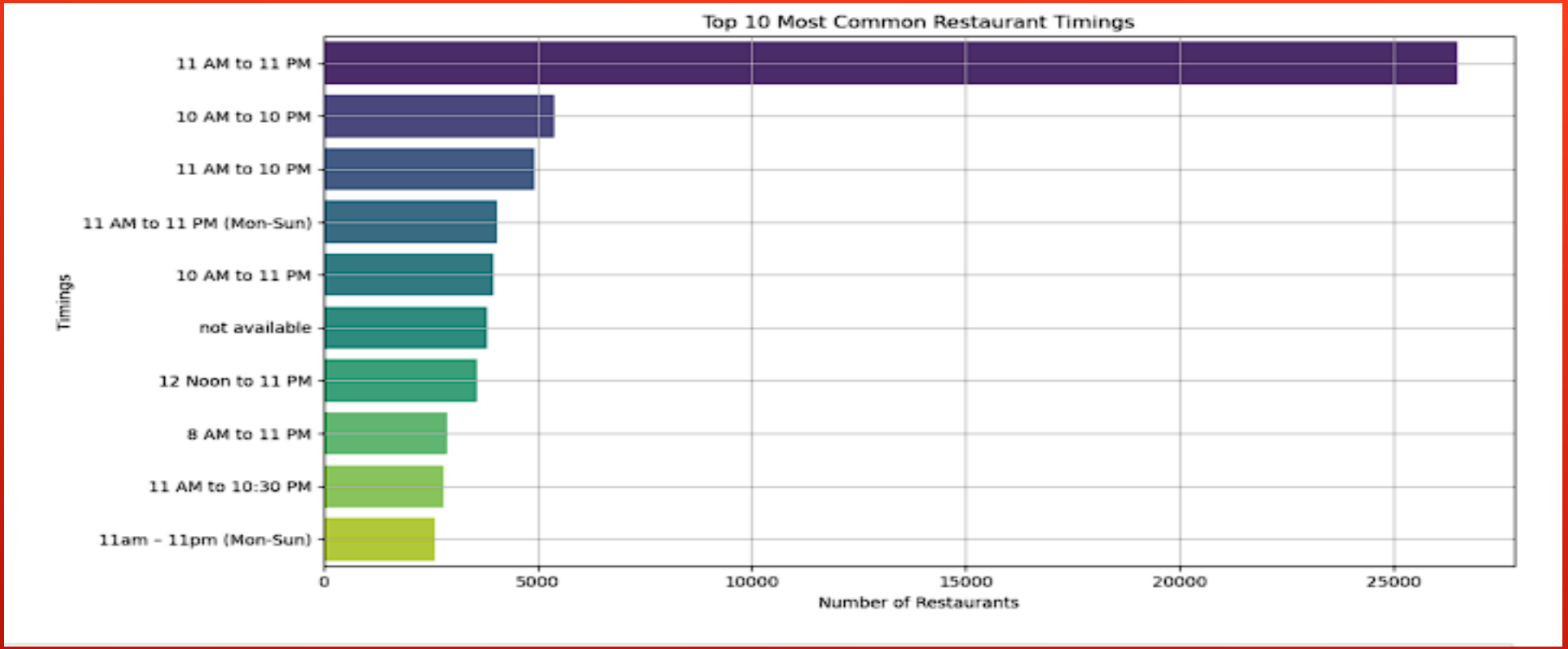
The vote count distribution among top 50 restaurants is right-skewed, with a few restaurants dominating total votes.
This suggests strong user loyalty toward established names.
Zomato can improve balance by highlighting well-rated but less-voted listings to boost discoverability.

DONUT CHART – RATING TEXT DISTRIBUTION



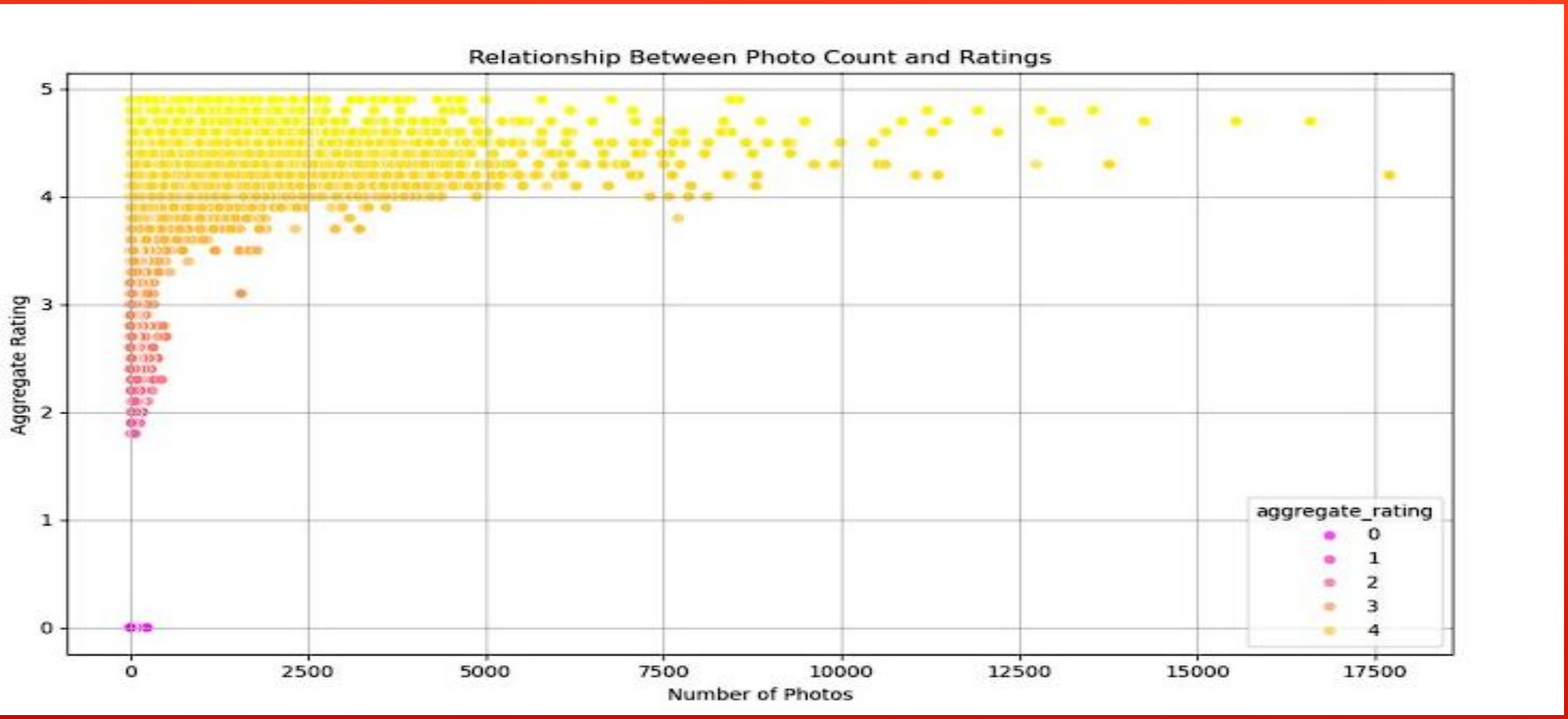
The donut chart reveals that most restaurants are rated as Good (33.3%), Very Good (30.0%), or Excellent (21.3%), indicating generally positive customer feedback. A smaller share falls under Average (12.0%) and Poor (3.3%), suggesting fewer negative experiences. This distribution reflects a positive dining sentiment across the platform but also highlights the need for quality control to uplift lower-rated establishments.

TOP 10 COMMON TIMINGS – BAR CHART



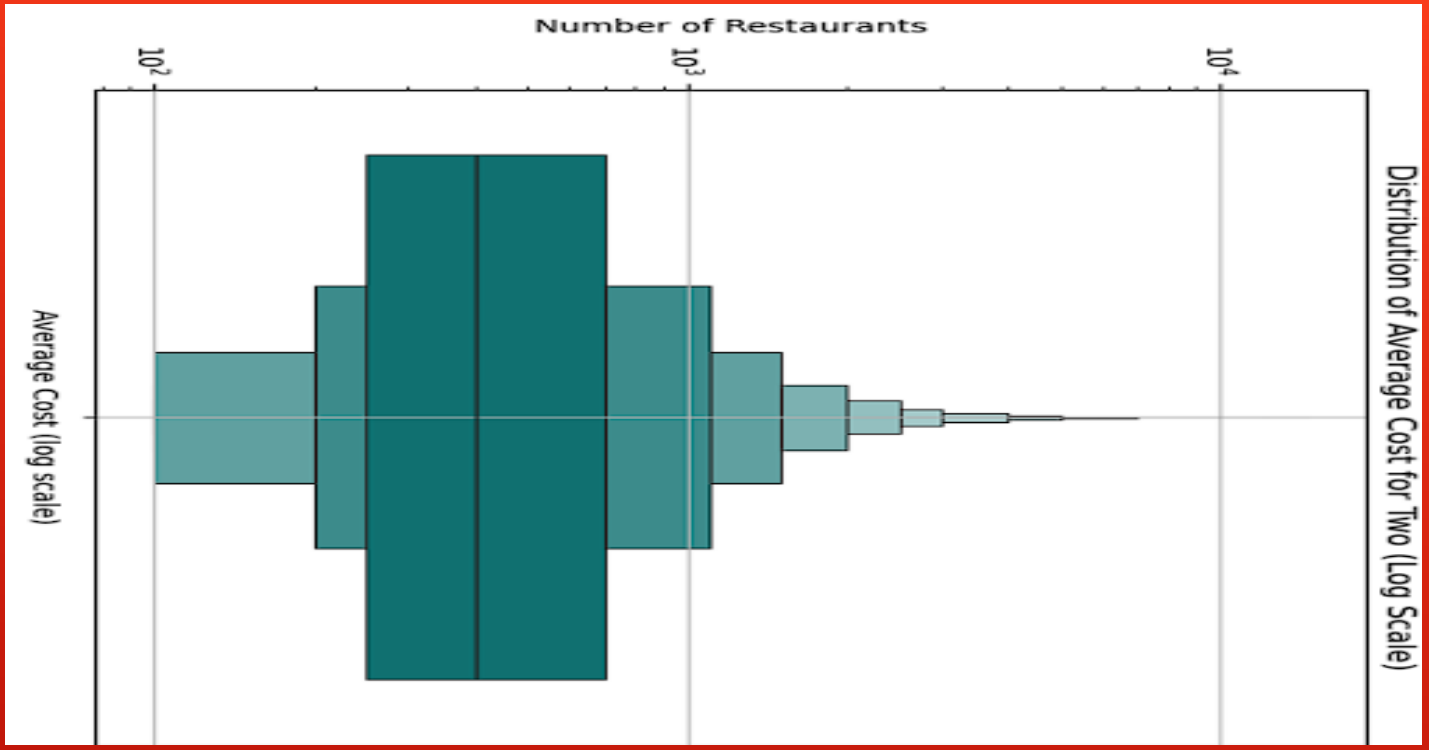
The bar chart shows that the most common restaurant operating times are centered around lunch (12 PM – 3 PM) and dinner (7 PM – 11 PM) slots. These peak periods align with standard meal hours, indicating when user demand is highest. Some restaurants also show extended or all-day availability, reflecting flexible service models. This insight helps restaurants optimize staffing, promotions, and delivery readiness during high-traffic hours.

SCATTER PLOT – PHOTO COUNT VS RATING



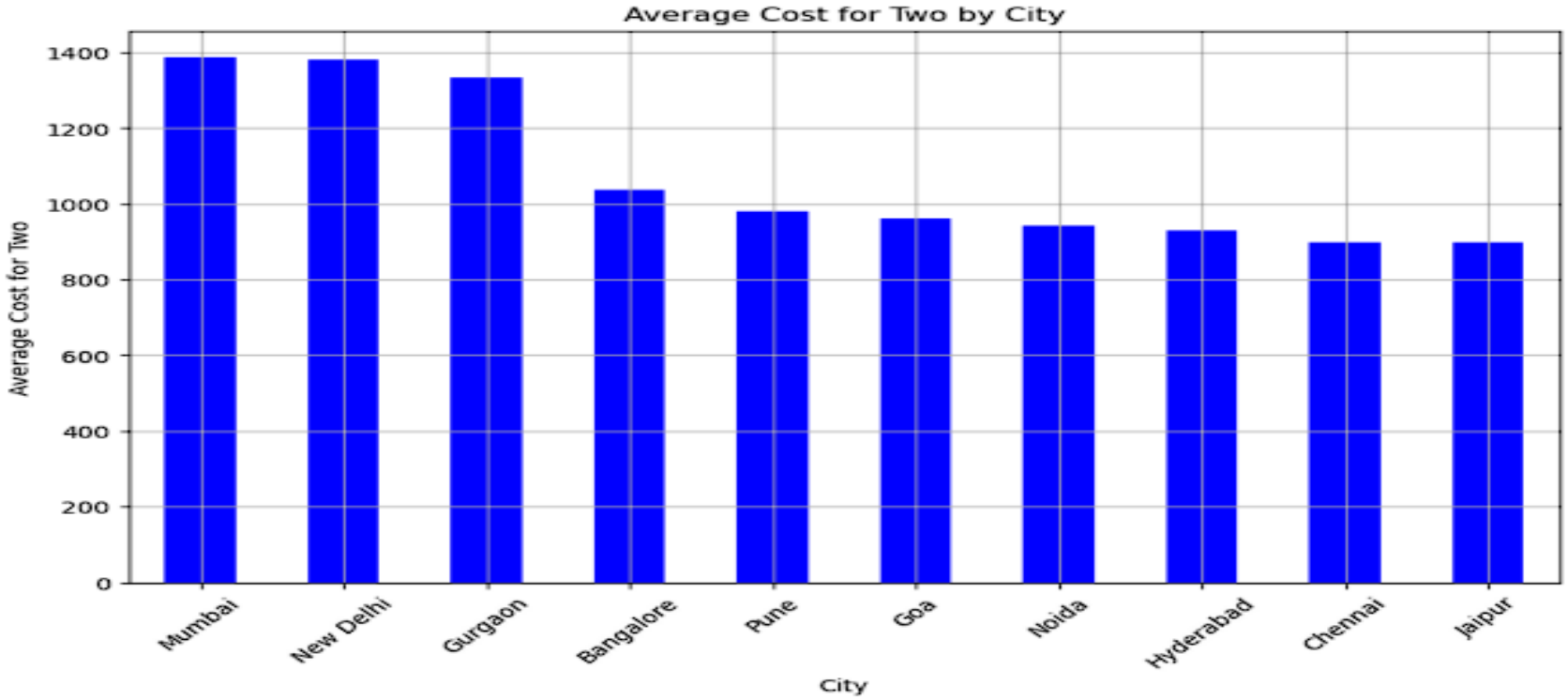
The scatter plot indicates a positive correlation between the number of photos uploaded and aggregate ratings. Restaurants with more photos tend to receive higher ratings, possibly due to better transparency, visual appeal, and customer trust. High photo counts often reflect active user engagement or intentional marketing efforts. Encouraging restaurants to upload more high-quality images can enhance user experience and improve perceived credibility on the platform.

BOXEN PLOT – COST FOR TWO



The majority of restaurants are concentrated under ₹500, indicating a cost-sensitive customer base. A long tail of high-end restaurants exists, but they are rare outliers.
The use of log scale reveals how steeply skewed the cost distribution is, allowing small and large values to be visualized clearly.
The dataset includes a few premium restaurants with costs exceeding ₹5,000–₹10,000 for two, likely in urban metros or fine-dining categories.
Businesses can leverage this insight to price competitively or target upscale customers in specific markets

AVERAGE COST FOR TWO BY CITY



This bar chart shows the top 10 cities with the highest average cost for two at restaurants. Cities like Mumbai, Delhi NCR, and Bengaluru are at the top, indicating a higher spending capacity. Tier-2 cities have relatively lower meal costs, reflecting regional pricing strategies.

SUMMARY, KEY INSIGHTS & STRATEGIC RECOMMENDATIONS

zomato

Summary

This project conducted an in-depth exploratory data analysis (EDA) on over 210,000 Indian restaurants using a dataset from Zomato. The objective was to uncover actionable insights about pricing, customer preferences, service features, and geographical trends that impact restaurant ratings and popularity.

Key Insights

Ratings are generally positive (3.0–4.5), but a sharp spike near 0 indicates many unrated restaurants.

Top cuisines include North Indian, Chinese, and South Indian, dominating consumer preferences.

Price Range 2–3 restaurants earn higher ratings, indicating a sweet spot between cost and quality.
Restaurants that offer online delivery and amenities like "Serves Alcohol" or "Credit Card" enjoy higher average ratings.

Most restaurants are cost-friendly (under ₹500), with only a few luxury outliers.

SUMMARY, KEY INSIGHTS & STRATEGIC RECOMMENDATIONS

Metro cities (Mumbai, Delhi, Bangalore) have the highest average dining costs.

User engagement increases with the number of photos uploaded, and higher votes correlate with better ratings.

Popular restaurant timings are 11 AM – 11 PM, and chains like Domino's and Pizza Hut dominate in terms of outlet count

Suggestions

Ensure your restaurant listing includes complete and attractive highlights (like takeaway, alcohol availability, payment methods).

Improve photo count and visual quality on platforms—visual appeal influences rating.

Leverage top-performing cuisine trends in your menu to attract a broader customer base.

Promote delivery service, especially in metro areas, to enhance reach and customer satisfaction.

Monitor city-wise cost benchmarks to adjust pricing competitively.

SUMMARY, KEY INSIGHTS & STRATEGIC RECOMMENDATIONS

Strategic Recommendations

Target Expansion by Tier:

Focus premium offerings in metros.

Roll out budget-friendly models in tier-2 cities.

Optimize Features for Ratings:

Invest in adding and marketing features that influence positive reviews (e.g., live music, alcohol service, digital payment options).

Peak Timing Promotions:
Run offers or special menus during high-traffic hours (11 AM – 11 PM).

Franchise Growth:

Use insights from top-performing chains to build a replicable, scalable model in new cities.

Customer Engagement:

Encourage photo uploads and reviews to enhance credibility and drive more organic traffic.

THANK YOU

Thank you for taking the time to review this project on Indian Restaurant Data Analysis. I hope the insights and strategies provided contribute to better decision-making and deeper understanding of the food service landscape in India.

If you have any questions, suggestions, or would like to connect further—feel free to reach out!

📬 Contact Details

👤 Aditya Anand

📧 Email: adityaanand5155@gmail.com

🔗 LinkedIn: linkedin.com/in/aditya-anand-4130262a3

💻 GitHub: github.com/adityaanand515