**A close up of a logo

AI-generated content may be incorrect.**

**ZOMATO RESTAURANTS ANALYSIS: MULTI-DIMENSIONAL INSIGHTS DASHBOARD**

**SINGH SHIVANG**

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN ANALYTICS

2025

**Statement of Originality**

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

A close-up of a blue signature

AI-generated content may be incorrect.

16-05-2025

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Date Singh Shivang

**Supervisor Declaration Statement**

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

[Input Date Here]

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Date Dr. Lim Kay Jin

**Table of Contents**

1. Abstract ............................................................. 5
2. Chapter 1: Introduction ................................................ 6
3. Chapter 2: Literature Review ........................................... 7
4. Chapter 3: Data Description........................................ 9
5. Chapter 4: Methodology ....................... 10
6. Chapter 5: Result and Discussion...... 19
7. Chapter 6: Conclusion ..................... 20
8. References....................... 21

**Abstract**

The booming restaurant industry in Bengaluru, driven by rapid urbanization and a diverse population, presents both opportunities and challenges for stakeholders seeking actionable insights. This practicum report delves into the Zomato Bengaluru dataset, comprising over 51,000 restaurant entries, to demonstrate a comprehensive analytics pipeline integrating data cleaning, exploratory data analysis (EDA), geospatial visualization, natural language processing (NLP), and sentiment analysis. We derive novel performance metrics, including a sentiment-weighted review score, and distil findings into an interactive Power BI dashboard that performs as a user-centric recommendation system. The outcomes highlight spatial economic patterns, uncover discrepancies between numeric ratings and textual sentiment, and illustrate how advanced analytics can guide urban planning, business strategy, and consumer decision-making.

**Chapter 1. Introduction**

Urban food ecosystems generate vast amounts of data from consumer reviews, operational information, and geographic distributions. Bengaluru, with its booming technology sector and rapidly evolving dining scene, exemplifies this complexity. Online food delivery and table reservation **platforms like Zomato** aggregate user ratings, textual reviews, pricing, cuisine types, and location data for thousands of restaurants. However, the raw data often contains inconsistencies, missing values, and unstructured text, which must be methodically processed before any meaningful analysis can occur.

This project addresses these challenges by constructing an end-to-end analytics framework tailored to the Bengaluru restaurant market. Our objectives span from meticulous data cleaning to advanced NLP-driven sentiment analysis, culminating in interactive dashboards that deliver insights to diverse stakeholders. In doing so, we demonstrate how to convert heterogeneous restaurant data into actionable intelligence, empowering urban planners, restaurateurs, and consumers alike.

**1.1 Background and Motivation**

Bengaluru, known as the Silicon Valley of India, is home to a varied demographic that relies heavily on dining and food delivery services. As the city’s population exceeds 15 million, its hospitality sector has expanded to more than 12,000 restaurants offering a spectrum of cuisines. Zomato serves as a central repository for customer feedback, but the platform’s rich data remains underutilized without systematic analysis.

Stakeholders in this ecosystem face distinct challenges:

* **Urban Planners** need to understand spatial patterns of restaurant density and quality to guide zoning, infrastructure development, and local economic policy.
* **Restaurant Owners and Investors** require market intelligence to identify high-demand areas, underserved cuisine segments, and factors driving customer satisfaction.
* **Consumers** benefit from recommendation systems that cut through noise and highlight establishments fitting their preferences for cuisine, budget, and location.

By tackling the entire workflow—from raw data to interactive dashboards—this project showcases how integrated analytics can bridge data complexity and real-world decision-making in a dynamic urban context.

**1.2 Objectives**

This study pursues the following objectives:

1. **Data Standardization & Enrichment**: Transform raw Zomato data into a clean, structured format ready for analysis.
2. **Exploratory & Geospatial Insights**: Uncover patterns in pricing, ratings, and neighbourhood distributions using statistical and mapping techniques.
3. **Review Sentiment Quantification**: Parse and analyse textual reviews to derive sentiment scores that complement numeric ratings.
4. **Composite Performance Metrics**: Create a weighted rating score that integrates rating, sentiment, and review volume for robust comparison.
5. **Interactive Dashboard**: Develop Power BI dashboard for personalized recommendations based on user inputs.

**Chapter 2. Literature Review**

A review of contemporary research provides a robust foundation for situating this study within the broader academic discourse. In recent years, the surge of publicly available restaurant and consumer review data has stimulated a diverse body of work in predictive analytics, spatial data analysis, natural language processing (NLP) for reviews, and dashboard development for decision support systems.

**2.1 Restaurant Rating Prediction**

**Wang, Shen, & Zhu (2018)** developed a gradient boosting machine approach to predict Yelp restaurant star ratings. By engineering features such as review counts, price level, and user interaction metrics, they achieved significant improvements over baseline models, demonstrating that structured metadata can effectively forecast user ratings in hospitality contexts **[1]** **.** **Jayaraman (2020)** compared multiple regression algorithms—including support vector regression, random forest, and linear regression—on the Yelp dataset. Their experiments showed that vote counts and average meal prices were among the most influential predictors of overall ratings, and that treating rating prediction as a continuous regression problem yielded better performance than classification-based formulations **[2]. Somashekar & Mallesh (2021)** focused specifically on Zomato data, applying linear and polynomial regression models to impute missing ratings. Their results confirmed that simple linear models using features like number of reviews and price estimates can accurately fill in sparse rating fields, validating our choice to impute missing Zomato ratings via linear regression on votes and approx\_cost(for two people) **[3]**

**2.2 Geospatial Analysis in Food Services**

**Shihab et al. (2018)** used DBSCAN clustering on restaurant geolocations in Seoul to identify high-density “food hubs” and to quantify access disparities across neighbourhoods. Their approach demonstrated how unsupervised clustering can reveal natural commercial corridors, an idea we adapted by using Folium maps and heat layers to expose high‑rating and high‑density zones in Bengaluru **[4].** **Bilen et al. (2018)** tackled the “site selection” problem by training classification models to predict the success of a new restaurant at a given location. They combined spatial coordinates with demographic and infrastructure variables, emphasizing the role of spatial patterns in service accessibility—a principle that underpins our neighbourhood‑aggregation of average ratings and cost to spotlight underserved areas **[5].**

**2.3 Sentiment Analysis of Customer Reviews**

**Liu (2012)** in *Sentiment Analysis and Opinion Mining* lays out foundational techniques for polarity scoring and opinion extraction from unstructured text. His lexicon‑ and machine‑learning‑based methods guide our use of TextBlob polarity scores as quantitative proxies for customer satisfaction **[6].** **Saha & Santra (2017)** performed sentiment classification on Zomato reviews, showing that rating polarity has only a moderate correlation (r ≈ 0.6) with numeric star ratings. This work highlighted the need to integrate both text‑derived sentiment and aggregate ratings, inspiring our dual‑aggregation strategy where sentiment and extracted per‑review ratings are combined into a composite score **[7].** **Pang & Lee (2008)** reviewed early opinion‑mining methods on movie and product reviews, illustrating best practices for text preprocessing, feature extraction, and evaluation metrics. Their recommendations for careful text cleaning and stopword removal inform our preprocessing pipeline for Zomato review texts, ensuring reliable sentiment computation [8].

**2.4 Recommendation Systems for Restaurants**

**Koetphrom et al. (2018)** compared collaborative‑filtering and content‑based restaurant recommenders, finding that metadata like cuisine, price, and location alone can yield strong recommendations when user histories aren’t available **[9]**. **Chen & Xia (2020)** developed a hybrid Yelp recommender combining collaborative signals with restaurant metadata, demonstrating that even purely content‑based approaches—when enriched with features like weighted scores and sentiment—can perform well without per‑user data **[10]**.

**3. Data Description**

**3.1** **Dataset Source**

The dataset was obtained from Zomato and contains information on Bengaluru restaurants as of March 2019. It includes approximately 51,696 entries and 17 columns.

**3.2 Key Variables**

**Original Columns from Zomato Dataset**

|  |  |
| --- | --- |
| Column name | Description |
| url | URL of the restaurant's Zomato page |
| address | Full address of the restaurant |
| name | Name of the restaurant |
| online\_order | Indicates if online ordering is available (Yes/No) |
| book\_table | Indicates if table booking is available (Yes/No) |
| rate | Original aggregate rating (text format, e.g., '4.1/5') |
| votes | Number of user ratings |
| phone | Phone number(s) of the restaurant |
| location | Locality or neighbourhood name |
| rest\_type | Type of restaurant (e.g., Casual Dining, Quick Bites) |
| dish\_liked | Popular dishes listed by customers |
| cuisines | Comma-separated list of cuisines served |
| approx\_cost(for two people) | Estimated cost for two people |
| reviews\_list | List of tuples: (Rating, Review Text) |
| menu\_item | Menu items available |
| listed\_in(type) | Type of service (e.g., Dine-out, Delivery) |
| listed\_in(city) | Neighbourhood or regional cluster |

### **3.3 Initial Inspection and Challenges**

An early review of data revealed multiple quality issues: textual ratings requiring parsing, missing values scattered across cost and review fields, inconsistent phone number formats, and multi-valued categorical fields (cuisines, listed\_in(type)). Additionally, raw textual reviews embedded as stringified lists necessitated careful parsing. These complexities motivated a tailored cleaning and enrichment process.

**4. Methodology**

**4.1 Data Cleaning and Imputation**

Our cleaning strategy combined automated scripts and statistical modelling. We first loaded the dataset into a Pandas Data Frame, inspecting column types and sample entries. Duplicate rows were removed to preserve unique observations. Handling missing values proceeded by column:

* **Ratings Imputation:** The rate column, originally in string format and containing blanks, was cleansed by extracting numeric values and converting them to floats. Missing ratings comprised approximately 10% of entries. We trained a linear regression model—using votes and approx\_cost(for two people) as predictors—to predict missing values, effectively imputing ratings grounded in observed relationships.
* **Phone Standardization:** Missing phones were set to empty strings. A regex filter stripped any unwanted characters, retaining digits, spaces, plus signs, and line breaks, ensuring consistency across records.
* **Categorical Imputation:** For rest\_type, dish\_liked, and cuisines, missing values were temporarily set to blank, then filled with the column’s statistical mode to maintain categorical integrity.
* **Cost Field:** Null values in approx\_cost(for two people) were substituted with the column mean, preventing bias in subsequent cost analyses.

Finally, Boolean indicators (online\_order, book\_table) and categorical columns (location, listed\_in(type), listed\_in(city)) were cast to category types to optimize performance. This rigorous cleaning yielded a complete, standardized dataset ready for feature engineering.

**4.2 Feature Engineering**

To unlock deeper insights, we engineered new variables:

1. **Geocoding:** Restaurant addresses were geocoded via the Google Maps API, producing latitude and longitude coordinates that enabled spatial analysis.
2. **Price Range Bucketing:** We segmented approx\_cost(for two people) into six discrete bins (e.g., 0–500, 501–1000 up to 3500+), stored in a new price\_range column.
3. **Service Score:** Binary encodings of online\_order and book\_table were summed into a composite service\_score, reflecting overall service capability. It’s clearly visible that restaurants with higher service score tend to have a higher rating on average.



Fig. 4.2.1. Service quality score vs rating boxplot

1. **Dimension Tables:** Multi-valued fields (cuisines, rest\_type) were exploded into separate dimension tables—CuisineDim and ServiceTypeDim—linked by restaurant name. This normalization facilitated flexible slicing in dashboards.
2. **Sentiment-weighted Score:** We allocated a new scaled\_review\_score metric—further detailed in Section 4.5.

**4.3 Exploratory Data Analysis (EDA)**

Our EDA phase sought to reveal underlying trends:

* **Univariate Distributions:** most restaurants cluster around 3.5–4.0 stars, with average two-person cost around ₹500. Count plots of neighbourhoods and cuisines revealed market concentration in central areas with popular cuisines like North Indian and Chinese dominating.

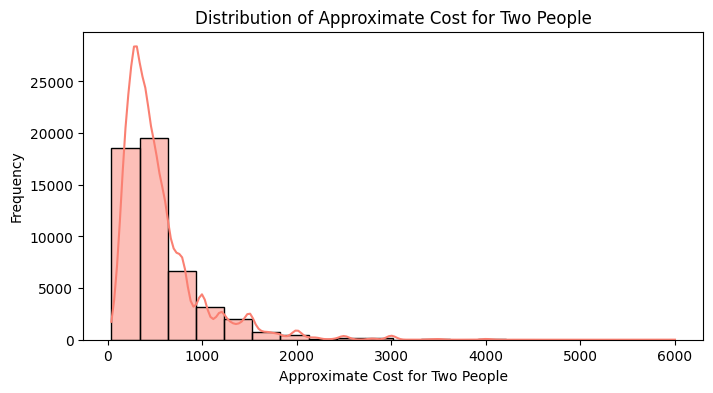
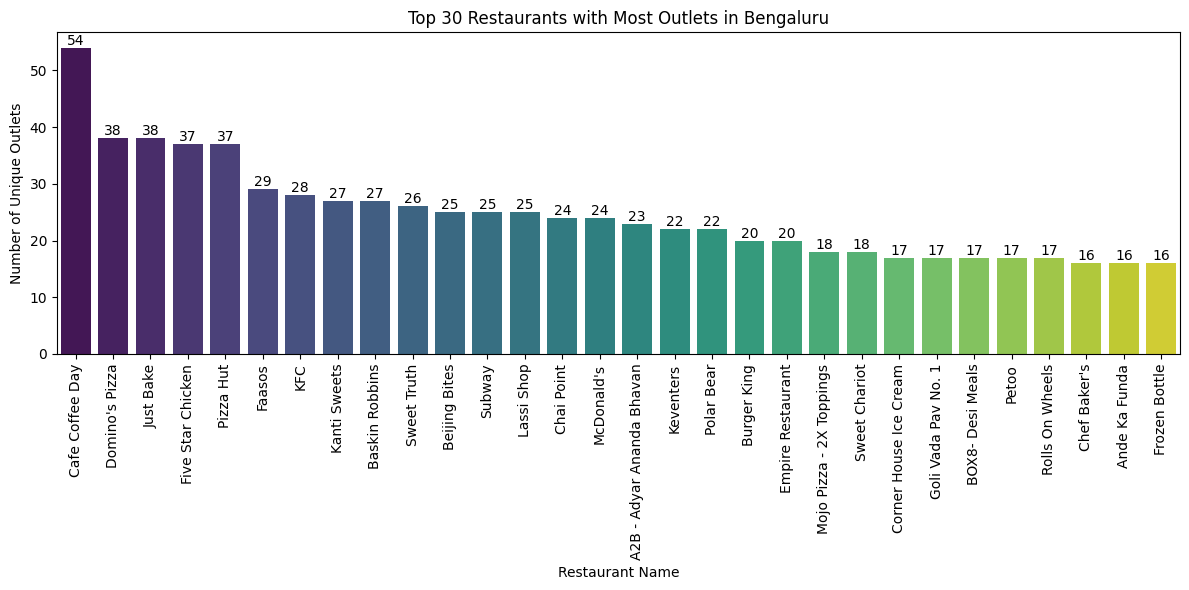


Fig. 4.3.1. Rating values Histogram Fig. 4.3.2. Approx cost for two Histogram

* **Categorical Counts:** Count plots of top neighbourhoods, restaurant types, and exploded cuisines highlighted market concentration and diversity. Fast food outlets and cafes turned out to be the most widespread category in terms of number of outlets with Café Coffee Day leading the chart. Fig 4.3.4 shows that BTM and HSR emerge as the city’s largest dining hubs (over 2,000 restaurants each), followed by key commercial and residential corridors like Koramangala and JP Nagar. In fig 4.3.5, we see that North Indian cuisine leads by a wide margin, with over 20,000 listings, trailed by Chinese and South Indian. Fast Food and Biryani also feature prominently.

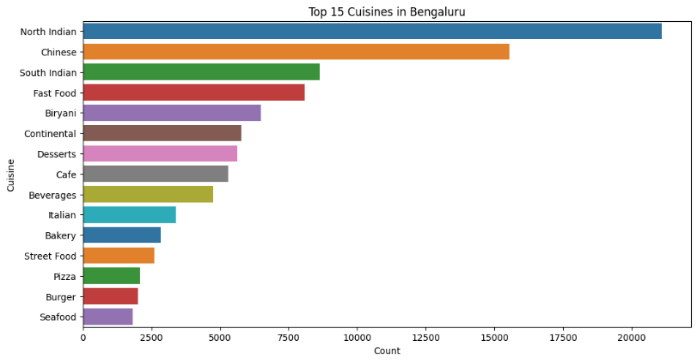
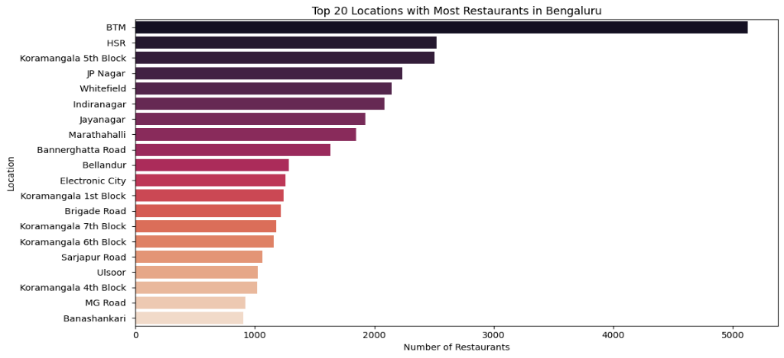
Fig. 4.3.3. Restaurant Outlets Bar chart

Fig. 4.3.4. Locations with most restaurants Fig. 4.3.5. Most listed cuisines in menus

* A graph with numbers and dots

  AI-generated content may be incorrect.**Bivariate Relationships:** The scatter plot of votes against average ratings fig 4.3.6 reveals a clear relationship between the volume of customer engagement and perceived quality. Restaurants with very few votes, typically under 500, exhibit a wide spread of ratings, ranging from as low as 2.0 to nearly 5.0 stars, suggesting that small sample sizes can yield unreliable or highly variable scores. In contrast, establishments with several thousand votes overwhelmingly cluster above

Fig. 4.3.6. Votes vs Ratings scatterplot

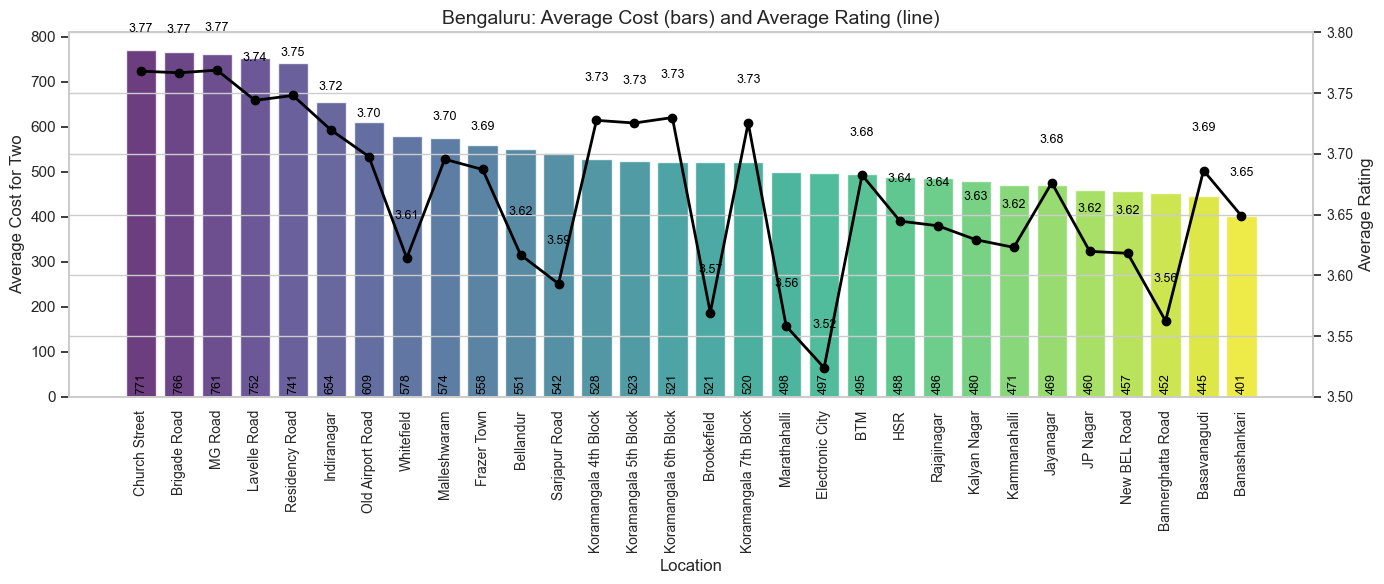
the 4.0‑star mark, indicating broad customer approval. Notably, restaurants offering table‑booking (indicated by orange markers) appear disproportionately in this high‑vote, high‑rating zone, implying that the ability to reserve a table may contribute both to greater popularity and to enhanced customer satisfaction. This pattern validates our choice to weight restaurant scores by vote count, ensuring that venues with substantial, consistent feedback are prioritized in our ranking and recommendations. Below, the combined bar–line chart compares average meal cost (bars) with average rating (line) across Bengaluru’s neighbourhoods. Central areas such as Church Street and MG Road

Fig. 4.3.7. Average cost & rating by location

command the highest average cost (approximately ₹750–800 for two people) and maintain strong average ratings around 3.75–3.80, illustrating a premium‑quality correlation. Mid‑range districts like Indiranagar and Frazer Town sit in the ₹550–650 cost bracket with solid ratings near 3.70–3.74, reflecting a balance between expense and customer satisfaction. On the other end, suburban localities—Electronic City and Banashankari, for example—average more affordable prices (₹450–500) but register slightly lower ratings (around 3.60–3.65), suggesting good value but fewer top‑tier experiences.

**4.4 Geospatial Visualization**

Leveraging Folium, we created layered, interactive maps:

1. **Heatmap:** A base heatmap rendered the overall density of restaurants. Subsequent layered heatmaps segmented by rating tiers (high, medium, low) illustrated spatial quality disparities with higher concentration in central Bengaluru.

A map with blue dots

AI-generated content may be incorrect.

Fig. 4.4.1. Heatmap of restaurant density

1. **Marker Clusters:** Markers coloured by no. of restaurants showcased budget and premium corridors. MarkerCluster plugins aggregated points at lower zoom and showed individual pins indicating restaurants when zoomed deeper. The clusters showed higher concentration of restaurants in the central and east parts of the city.

A map with many colored circles

AI-generated content may be incorrect.

Fig. 4.4.2. Marked clusters of restaurants by density

1. A map with many colored dots

   AI-generated content may be incorrect.**Neighbourhood Aggregation:** For each locality, we computed centroids and aggregated metrics—average rating, total outlets—and plotted CircleMarkers sized by outlet count. In our initial plot showing the number of outlets by location, BTM stands out with the largest bubble, indicating a high concentration of outlets paired with a moderate average rating, whereas Lavelle Road in central Bengaluru has fewer restaurants but boasts higher average ratings.

Fig. 4.4.3. Centroids bubble map aggregated by rating

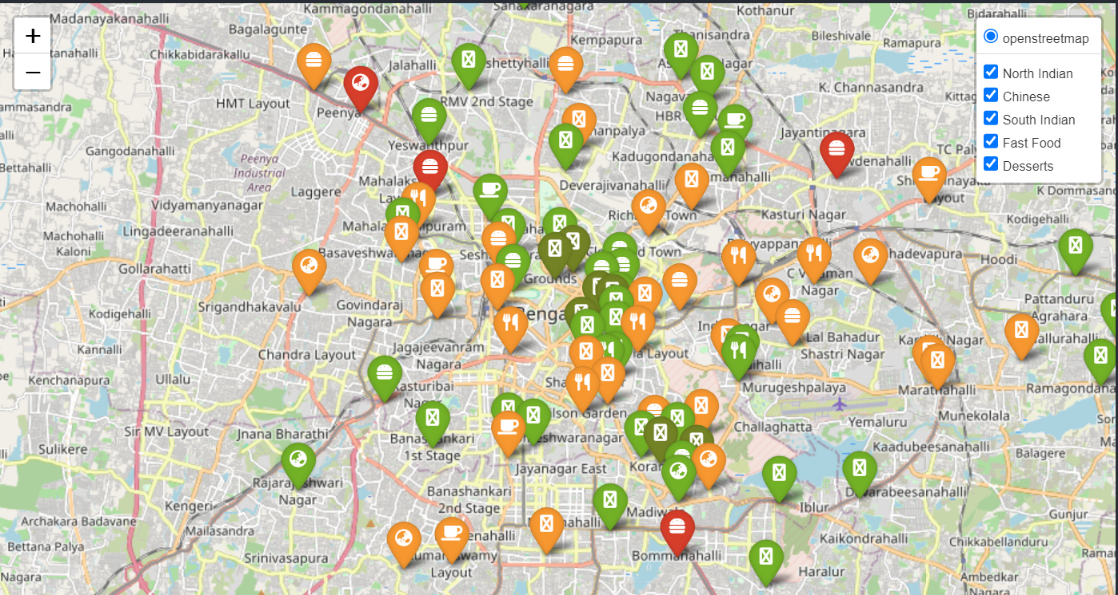
1. **Cuisine Insights:** Focusing on five major cuisines, we grouped exploded records by (location, cuisine), computed average ratings, and assigned each neighbourhood its top-rated cuisine. The plots use color‑coded markers (green for higher rating, orange for medium rating, red low rating) to display every outlet in the dataset, revealing where each cuisine predominates denoted by different icons. Central dining hubs such as Koramangala, Indiranagar,

Fig. 4.4.5. Highest rated cuisine by location from overall Top 5 cuisine

MG Road, and Jayanagar glow with a high density of North Indian and Chinese restaurants, while family‑oriented residential districts like BTM and HSR Layout are peppered with South Indian establishments. Fast food and dessert venues, in contrast, are more evenly dispersed—even into peripheral suburbs such as Electronic City and Banashankari—highlighting their broad appeal.

In fig. 4.4.6. the interactive pop‑ups reveal the location name, top rated cuisine type, and average rating for that cuisine (for example, “Sankey Road – Desserts – 4.20”), instantly guiding users to the premier destination for each cuisine in any locality.

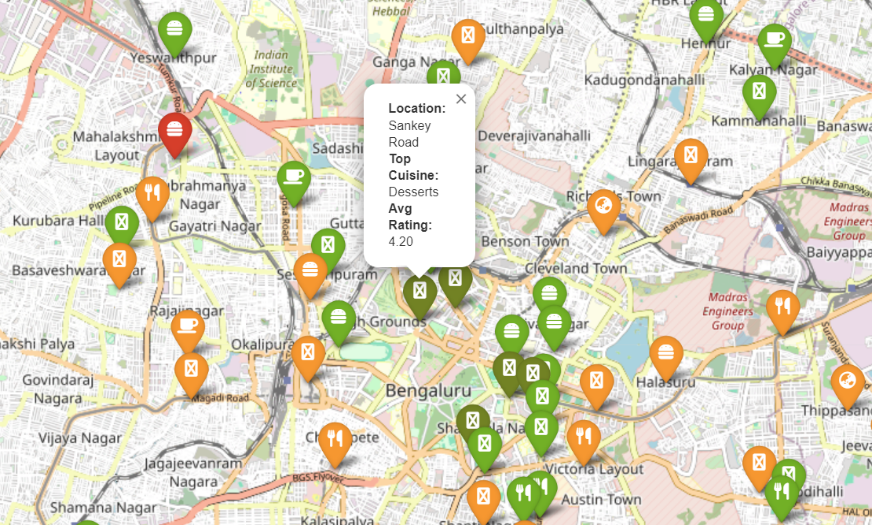


Fig 4.4.6. Sankey Road with high rating and dessert as top cuisine

This layered approach ensures that stakeholders can both explore overall culinary patterns and zero in on high‑quality venues tailored to their preferences.

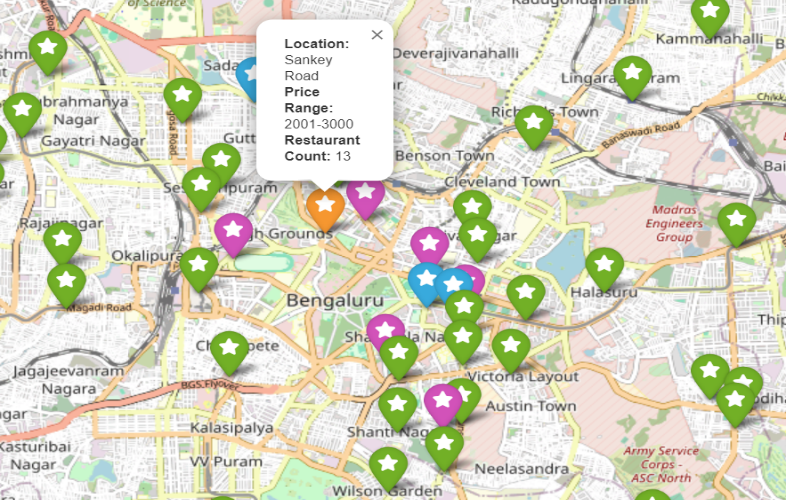
1. **Cost Map:** Neighbourhoods grouped by mode of price\_range received distinct marker colour, conveying average meal affordability briefly. This map gives a general overview of each location and the probable price range of any restaurant in that location. Once again posh areas like Sankey Road in central part of city have higher price bracket on average as compared to the outer regions.

Fig. 4.4.7. Location by price range map

**4.5 Sentiment Analysis and NLP**

We parsed reviews\_list—stringified Python lists of (rating, review) tuples—using ast.literal\_eval. Review texts underwent cleaning: encoding correction via ftfy, regex-based URL and punctuation removal, lowercase conversion, and whitespace normalization. We extracted numeric ratings from tuple labels and computed sentiment polarity (-1 to +1) using TextBlob. Aggregating by restaurant, we derived average extracted rating, average sentiment, and review counts. Merging these with original votes and rating, we calculated a raw review score (mean of extracted rating and sentiment) and applied **a weighted factor = 1 + log1p(votes)/5** to emphasize well-reviewed establishments. Min–max normalization scaled the final scaled\_review\_score to 0–5, aligning with original rating scales. Plots analysing and visualizing the scaled review scores with the original app ratings give us better insights into the importance of review as compared to blind or dummy ratings.

1. A screenshot of a graph

   AI-generated content may be incorrect.**Boxplots comparison:** Fig. 4.5.1. contrasts the original rating distribution (left) with the scaled review score (right) after min–max normalization to the 0–5 range. The median of the scaled review score sits lower (around 1.5) than the median star rating (~3.8), indicating that sentiment‑weighted scores tend to be more conservative—penalizing establishments with sparse or mixed reviews. The wider interquartile range and numerous outliers in the scaled scores reflect greater variability introduced by combining text sentiment and vote volume.

Fig. 4.5.1. Rating vs scaled review score

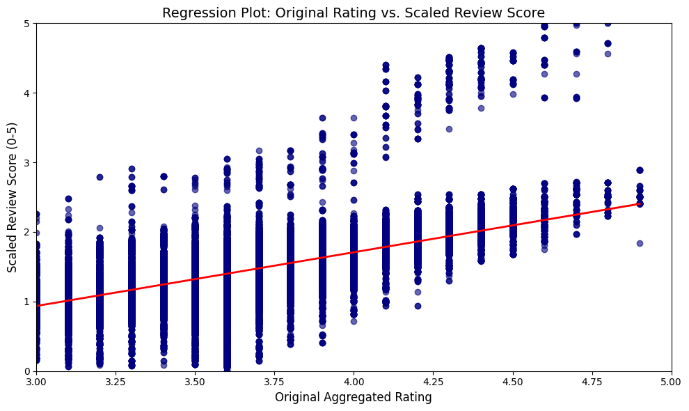
1. **Regression Scatter:** The scatter plot in fig. 4.5.2. places each restaurant’s normalized review score against its original star rating, with a fitted regression line in red. The positive slope confirms a moderate correlation: higher star ratings generally accompany higher review scores. However, the scatter’s vertical spread at each rating level highlights that two restaurants with identical star ratings can have quite different weighted scores—validating the added value of our sentiment and vote‑based adjustments.

Fig. 4.5.2. Regression scatter plot

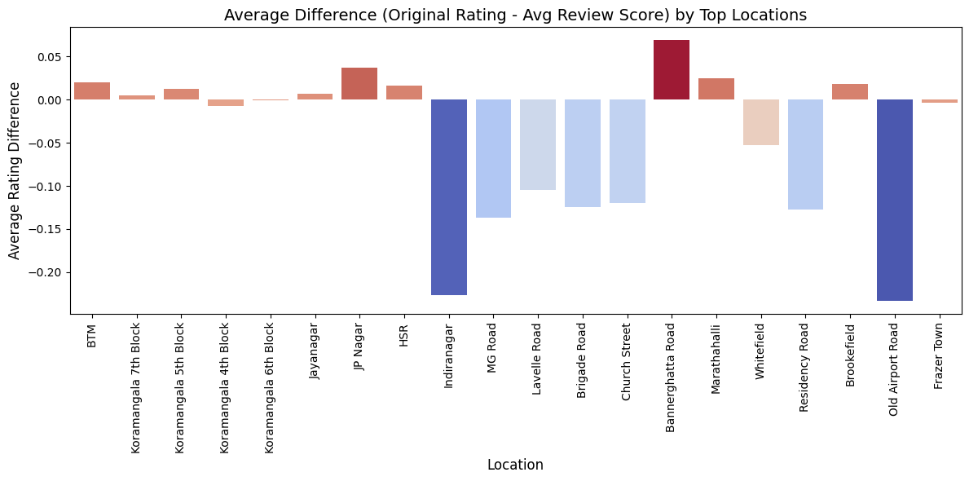
1. To assess systematic discrepancies between numeric star ratings and our more nuanced, sentiment‑weighted review scores, we calculated the average difference (star rating minus composite score) for the top neighbourhoods by outlet count (fig. 4.5.3).

Fig. 4.5.3. Star ratings VS Derived Review Scores

Positive differences—observed in areas such as BTM, JP Nagar, and Bannerghatta Road—indicate that the platform’s raw star ratings tend to exceed the sentiment‑driven scores, suggesting a mild inflation of stars relative to the textual feedback. Conversely, locations like Indiranagar, Old Airport Road, and MG Road exhibit notably negative gaps (approximately –0.20 to –0.25), revealing that customer reviews in these districts convey stronger positive sentiment than their star ratings alone would imply. This divergence highlights two practical insights: first, areas with positive gaps may harbour overrated venues whose actual customer experiences, as expressed in free‑text reviews, do not fully match their star appeal; second, neighbourhoods with negative gaps can be treated as “hidden‑gem” zones where textual feedback uncovers high satisfaction levels underrepresented by the numeric rating.

**4.6 Dashboard Development**

A screenshot of a computer dashboard

AI-generated content may be incorrect.Using Power BI Desktop (free), we built a dashboard:

Fig. 4.6.1. Power BI Recommendation Engine Dashboard

**Recommendation Engine** — Empowered users to select preferences via slicers for locality, ratings, approx. cost for two people, table reservation and view ranked lists of restaurants, accompanied by dynamic map makers and donut charts displaying the cuisine types for each respective slicer selections.

**Chapter 5. Results and Discussion**

**5.1 Findings and Summary**

Our in-depth investigation of Bengaluru's Zomato dataset leaves us with several actionable insights. After rigorous cleaning and imputation, we normalized 51,696 restaurant entries to enable healthy investigation of cost, ratings, and categorical features. EDA confirmed that central and southern corridors like BTM, HSR Layout, Koramangala, MG Road are where the densest outlet presence exists, and North Indian, Chinese, and South Indian cuisine take the lead. Table-booked restaurants concentrated among the highly rated, high-vote restaurants in a strong positive correlation between vote number and star rating. Geospatial mapping revealed a cost–quality gradient: prestige zones are more expensive and better rated, with suburbs offering less pricey options at slightly lower satisfaction. Sentiment analysis added further refinement to this snapshot, with a weighted review score based on the combination of text polarity and number of votes picking up on both under-rated gems and over-rated extremes. Power BI dashboard that produces tailored tips through cuisine, cost, location, and sentiment filters.

**5.2 Recommendations for Stakeholders**

* **Urban Policy Makers:** Use geospatial maps to prioritize infrastructure improvements (e.g., transit links, parking) in budding food corridors and to support micro‑enterprises in underrepresented neighbourhoods.
* **Restaurateurs & Investors:** Target mid‑density suburbs that show positive sentiment but lack sufficient high‑quality offerings; use the weighted review score to benchmark competitor performance beyond star ratings.
* **Platform Developers:** Enhance sentiment analysis with aspect‑level models (e.g., food, service, ambiance), integrate real‑time data feeds, and explore hybrid recommendation techniques to combine content‑ and collaborative‑filtering strengths.

**5.3 Limitations**

Although comprehensive, this study has several constraints. The dataset reflects a static snapshot from March 2019; it does not account for seasonality, recent market entries, or pandemic‑related shifts. Geocoding accuracy depends on address quality and the chosen mapping API, potentially misplacing outlier restaurants. Sentiment analysis via TextBlob captures general polarity but may overlook nuanced aspects—such as service vs. food quality—requiring aspect‑based NLP for deeper insights. Finally, the recommendation engine is content‑based and does not incorporate collaborative filtering, which limits its ability to capture community‑driven patterns and individual user tastes

**Chapter 6. Conclusion**

This study has illustrated a replicable, systematic pipeline for transforming raw restaurant platform data into strategic insight and customer‑facing tools. From meticulous data cleaning and imputation, through exploratory and spatial analysis, to advanced sentiment weighting and interactive dashboards, the project illustrates how heterogeneous datasets can inform urban policy, business strategy, and individual decision‑making in an increasingly dynamic food environment.

Future research directions include broadening the time horizon to examine pandemic‑caused and seasonal effects and applying aspect‑based sentiment analysis to distinguish food, service, and ambiance comments. Collaborative filtering methods can be incorporated once user‑level interaction data are available. Additionally, relating restaurant performance to other external data sources such as socioeconomic variables, pedestrian movement patterns, or property trends will enhance contextual analysis of restaurant-going behaviour. By publishing methodology and code online for free access, this work throws open the window to duplicate analysis of other metropolises and platforms as well as into the broader general area of urban hospitality analytics.

**References**

1. Wang, X., Shen, Y., & Zhu, Y. A data-driven approach to predicting rating scores for new restaurants. In Proceedings of the Asian Conference on Machine Learning (pp. 678–693) (2018).
2. Jayaraman, P. Predicting restaurant rating using machine learning and comparison of regression models. In 2020 International Conference on Emerging Trends in Engineering and Information Technology (ic-ETITE) (pp. 1–5). IEEE. (2020).
3. Somashekar, S., & Mallesh, S. Restaurant rating prediction using regression. In 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1139–1144). IEEE. (2021)
4. Shihab, I. F., Oishi, M. M., Islam, S., Banik, K., & Arif, H. A machine learning approach to suggest ideal geographical location for new restaurant establishment. In 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 1–5). (2018)
5. Bilen, T., Erel-Ozçevik, M., Yaslan, Y., & Oktug, S. F. A smart city application: Business location estimator using machine learning techniques. In 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 1314–1321). IEEE. (2018).
6. Liu, B. Sentiment analysis and opinion mining. Morgan & Claypool. (2012).
7. Saha, S., & Santra, A. K. Restaurant rating based on textual feedback. In 2017 International Conference on Microelectronic Devices, Circuits and Systems (ICMDCS) (pp. 1–5). IEEE. (2017).
8. Pang, B., & Lee, L. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135. (2008).
9. Koetphrom, N., Charusangvittaya, P., & Sutivong, D. Comparing filtering techniques in restaurant recommendation system. In 2018 2nd International Conference on Engineering Innovation (ICEI). IEEE. (2018).
10. Chen, Y., & Xia, F. Restaurants’ rating prediction using Yelp dataset. In 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA) (pp. 113–117). IEEE. (2020).