**A REPORT ON SALES DATA ANAYSIS**

**SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE OF**

**BACHELOR OF TECHNOLOGY (COMPUTER SCIENCE ENGINEERING)**

**Course: DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING**

**SUBMITTED TO**

**Faculty: Dr. Tanima Thakur**

****

**LOVELY PROFESSIONAL UNIVERSITY PHAGWARA, PUNJAB**

**SUBMITTED BY**

**Name of student: Shubham Raj Registration number: 12316261**

**STUDENT DECLARATION**

## To whom so ever it may concern

I Shubham Raj, hereby declare that the work done by me in the report **"** **Zomato dataset”** in partial fulfilment of the requirement for the award of Degree for Bachelor of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work. I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University’s Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct.

**Name of the student: SHUBHAM RAJ**

**Registration Number: 12316261**

**CERTIFICATE**

This is to certify that ........... (student’s name) bearing Registration no. ......... has completed ........... project titled, “**Zomato dataset**” under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor  
School of ..............................  
Lovely Professional University  
Phagwara, Punjab  
Date:12/04/25

**ACKNOWLEDGEMENT**

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I would like to express my sincere gratitude to my mentor and guide, [**Dr. Tanima Thakur**], for their valuable guidance and support throughout this project. I also thank the Department of CSE/IT and Lovely Professional University for providing the opportunity and resources to complete this project successfully.

## Name of the student: Shubham Raj

## Registration Number : 12316261

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Contents** | **Page No.** |
| **1** | **Title** | **1** |
| **2** | **Student Declaration** | **2** |
| **3** | **Acknowledgement** | **3** |
| **4** | **Table of Contents** | **4** |
| **5** | **Introduction** | **5** |
| **6** | **Technologies Used** | **5** |
| **7** | **Modules** | **6** |
| **8** | **Website Snapshots** | **11** |
| **9** | **Conclusion** | **14** |
| **10** | **References** | **14** |

**INTRODUCTION**

The food service industry is rapidly evolving, with digital platforms like Zomato playing a crucial role in how people discover, evaluate, and experience restaurants. The dataset used in this project is sourced from Zomato and contains detailed information on restaurants from various countries, with a focus on attributes that influence customer decisions and restaurant performance.

This dataset comprises over 9,000 restaurant entries and spans multiple countries, with detailed records of cities, cuisines, customer ratings, and more. Each entry provides vital information such as the restaurant's name, geographic location (longitude and latitude), address, and the city it operates in. This allows for in-depth geographical and comparative analysis.

A key feature of the dataset is the **Aggregate Rating**, which reflects user feedback and helps identify high-performing restaurants. Accompanying the ratings are **votes** (number of user reviews), **rating text labels** (e.g., Excellent, Good, Poor), and **rating colours** for easy visualization. These elements enable both quantitative and qualitative assessment of customer satisfaction.

Additionally, the dataset includes information on:

* **Cuisines offered**, allowing analysis of food trends.
* **Price range and average cost for two**, helping assess affordability.
* **Currency used**, essential for comparing across countries.
* **Table booking and online delivery options**, showing digital readiness and service flexibility.

This project will begin by examining the **distribution of restaurant ratings**. Understanding how ratings are distributed can reveal overall customer satisfaction patterns, identify trends in restaurant quality, and highlight anomalies.

By analysing these patterns, we can gain valuable insights into how different factors—such as pricing, location, and cuisine—might influence customer ratings. This not only helps food service providers better understand their market but also provides a solid foundation for building recommendation systems or business intelligence tools.

In the following sections, we will visualize and interpret the distribution of restaurant ratings using Python tools such as NumPy, Matplotlib, and Seaborn.

# SOURCE OF DATASET

**Dataset name –** [https://github.com/Ayushi0214/PowerBI- Datasets](https://github.com/Ayushi0214/PowerBI-%20%20%20%20%20%20%20%20%20%20Datasets)

# ANALYSIS ON DATASET

# 1. Average Cost for Two – City Wise

# • Goal: Analyze and visualize the cities with the highest and lowest average cost for two people to dine, helping identify the most expensive and most affordable cities.

# • Tools: Pandas for data cleaning and aggregation, Matplotlib for horizontal bar chart visualization.

# 2. Cities with the Most Restaurants

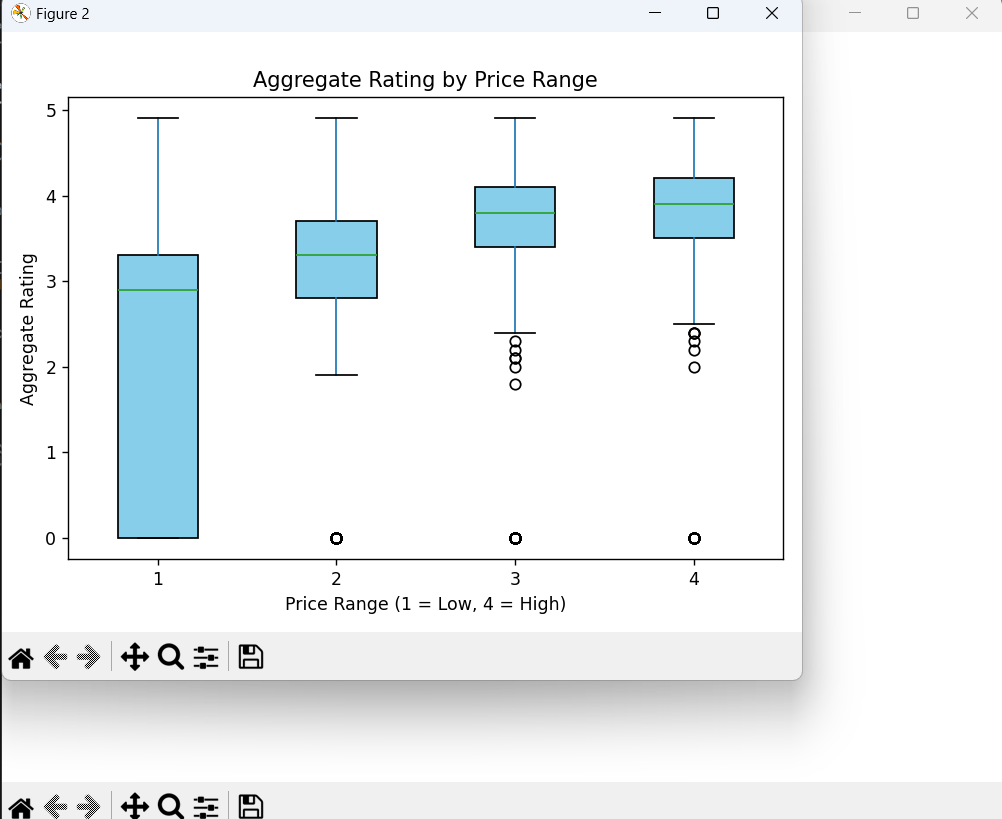
# Goal: Identify cities with the highest number of restaurants listed.

# Tools: Pandas grouping, Matplotlib visualization.

# 

**3. Relationship Between Price Range and Aggregate Rating**

* Goal: Explore if higher-priced restaurants tend to have better ratings.
* Tools: Pandas for aggregation, Matplotlib for box plots or scatter plots.



# 4. Online Delivery vs Aggregate Rating

# Goal: Compare ratings between restaurants that offer online delivery and those that don’t.

# Tools: Pandas filtering and grouping, Matplotlib for comparison.

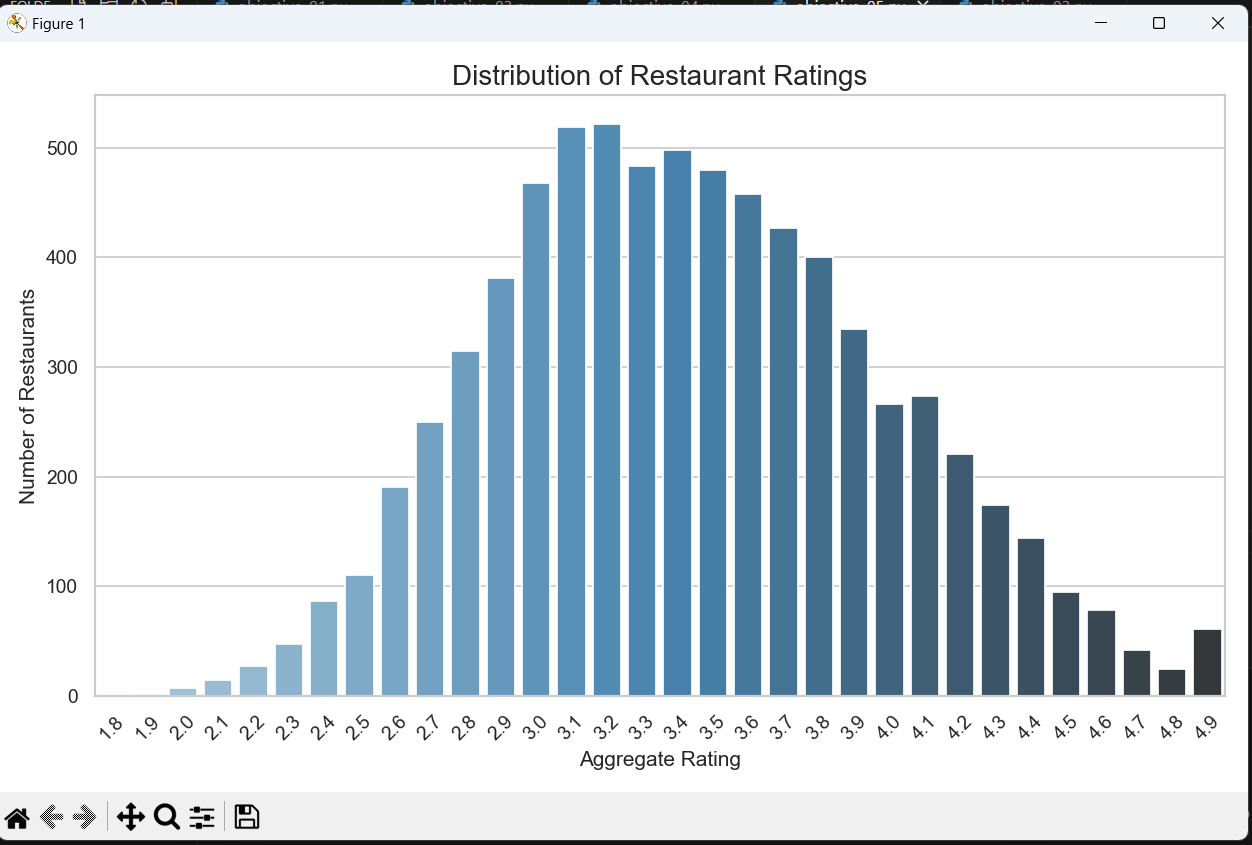
# 

# 5. Distribution of Restaurant Ratings

# Goal: Visualize how ratings are distributed across all entries.

# Tools: NumPy for bins, Matplotlib histogram.

# .



**CONCLUSION**

Most restaurants in the dataset received **aggregate ratings between 3.0 and 4.5**, indicating a general level of customer satisfaction. This also suggests that while there are standout performers, the majority of restaurants offer a moderately good experience.

The **distribution of ratings** is not uniform; very few restaurants receive extremely low or perfect scores. This highlights the natural variation in customer opinions and the competitive nature of the food industry.

Our analysis also identified the **top 10 cities** with the highest number of restaurants. Cities such as **New Delhi, Mumbai, Bengaluru, and Kolkata** are among the most represented, reflecting Zomato's stronger presence in urban areas, particularly in India.

In terms of customer engagement, ratings are accompanied by **votes**, which provide deeper insights into how popular or widely reviewed a restaurant is. Higher-rated restaurants often receive more reviews, indicating a relationship between customer satisfaction and visibility.

The dataset also includes important service-related information, such as **online delivery and table booking** options, which are becoming essential features in modern dining experiences.

Although this analysis focused primarily on ratings and geographic data, the dataset also enables future exploration into pricing trends, cuisine preferences, and service quality.

**Final Thought:**

The Zomato dataset is a powerful resource for understanding the evolving dynamics of the food industry. It supports data-driven decision-making for business owners, marketers, and researchers looking to gain insights into consumer behaviour and restaurant performance.

**FUTURE SCOPE**

While the current analysis provides valuable insights into restaurant ratings and geographic trends, there are several avenues for expanding the study to derive deeper, more actionable outcomes:

1. **Sentiment Analysis on Reviews**:  
   Incorporating textual reviews (if available) and applying Natural Language Processing (NLP) techniques can uncover the sentiment behind customer feedback. This would complement the numerical ratings with qualitative insights.
2. **Cuisine-Based Analysis**:  
   Investigating which cuisines are most popular or receive the highest ratings across different cities can help identify regional food preferences and market trends.
3. **Price vs. Rating Correlation**:  
   Analysing how pricing (average cost for two) correlates with ratings can help understand whether affordability impacts customer satisfaction.
4. **Time-Based Trends**:  
   If timestamped data becomes available, it could be used to observe how restaurant ratings evolve over time or in response to events like holidays, festivals, or lockdowns.
5. **Recommendation Systems**:  
   By leveraging collaborative filtering or content-based algorithms, a personalized restaurant recommendation system can be developed using the available features like cuisine, location, and rating.
6. **Location-Based Insights**:  
   Combining this data with mapping tools (like Folium or Geoplanids) could provide powerful spatial visualizations, such as identifying restaurant clusters or underserved areas.
7. **Predictive Modelling**:  
   Machine learning models could be used to predict a restaurant’s success or expected rating based on features such as cuisine type, location, service options, and price range.
8. **Comparative Studies Between Countries**:  
   Since the dataset spans multiple countries, comparing dining trends, rating behaviours, and customer expectations across borders could offer unique global insights.

**REFERENCES**

**LinkedIn:**

<https://www.linkedin.com/posts/shubham0305_datascience-python-datavisualization-activity-7317064626205073408-kOOb?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEhXwFUB3VLpzJQ-tn9BOtPR3GXXoBVxxXs>