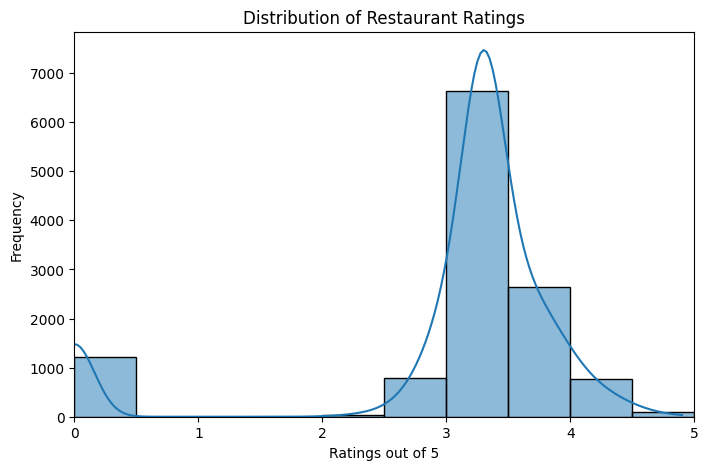
Zomato Pune Restaurant Analysis Report: Exploring Factors Influencing Ratings

Pune, a city celebrated as a cultural hub and educational center, boasts a vibrant and diverse culinary landscape. With a plethora of dining options ranging from traditional eateries to international chains, the restaurant scene is constantly evolving. This study delves into the dynamics of Pune's restaurant market by examining the distribution of Zomato ratings and their relationship with restaurant pricing. Utilizing data scraped from Zomato until December 1st, 2019, this analysis aims to uncover insights into customer preferences and market trends.

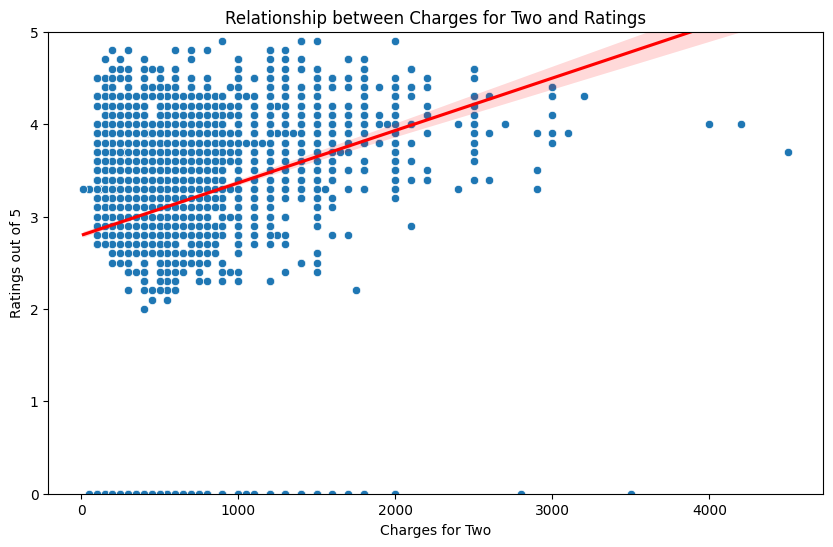
Table 1: Column details of the dataset

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| Restaurant\_Name | object | Name of the restaurant |
| Web\_Link | object | URL of the restaurant's website |
| Locality | object | Area or neighborhood where the restaurant is located |
| Sponsored | object | Indicates if the restaurant listing is sponsored |
| Ratings\_out\_of\_5 | object | Average rating given by customers (out of 5) |
| Number of votes | object | Total number of customer ratings |
| Phone\_number | object | Restaurant's phone number |
| Cuisines | object | Types of cuisine offered |
| Charges\_for\_two | object | Estimated cost for a meal for two people |
| payment\_modes | object | Accepted payment methods |
| Rest\_timming | object | Restaurant's operating hours |
| Detail\_address | object | Full address of the restaurant |
| 5\_star\_review\_percentage | object | Percentage of 5-star reviews |
| 4\_star\_review\_percentage | object | Percentage of 4-star reviews |
| 3\_star\_review\_percentage | object | Percentage of 3-star reviews |
| 2\_star\_review\_percentage | object | Percentage of 2-star reviews |
| 1\_star\_review\_percentage | object | Percentage of 1-star reviews |
| Wine and Beer | int64 | 1 if the restaurant serves wine and beer, 0 otherwise |
| Dance Floor | int64 | 1 if the restaurant has a dance floor, 0 otherwise |
| Lunch Menu | int64 | 1 if the restaurant offers a lunch menu, 0 otherwise |
| Outdoor Seating | int64 | 1 if the restaurant has outdoor seating, 0 otherwise |
| Seaside | int64 | 1 if the restaurant is located by the sea, 0 otherwise |
| 4/5 Star | int64 | 1 if the restaurant has a 4/5 star rating, 0 otherwise |
| Free Parking | int64 | 1 if the restaurant offers free parking, 0 otherwise |
| Pet Friendly | int64 | 1 if the restaurant is pet-friendly, 0 otherwise |
| Bulk Orders Accepted | int64 | 1 if the restaurant accepts bulk orders, 0 otherwise |
| Gaming Area | int64 | 1 if the restaurant has a gaming area, 0 otherwise |
| City View | int64 | 1 if the restaurant has a city view, 0 otherwise |
| Vegetarian Only | int64 | 1 if the restaurant is vegetarian only, 0 otherwise |
| Brunch | int64 | 1 if the restaurant serves brunch, 0 otherwise |
| Live Music | int64 | 1 if the restaurant has live music, 0 otherwise |
| Wheelchair Accessible | int64 | 1 if the restaurant is wheelchair accessible, 0 otherwise |
| Home Delivery | int64 | 1 if the restaurant offers home delivery, 0 otherwise |
| Celebrity Frequented | int64 | 1 if the restaurant is frequented by celebrities, 0 otherwise |
| Pre-Ordering Required | int64 | 1 if pre-ordering is required, 0 otherwise |
| Pool Table | int64 | 1 if the restaurant has a pool table, 0 otherwise |
| Keto Options | int64 | 1 if the restaurant offers keto options, 0 otherwise |
| Wifi | int64 | 1 if the restaurant offers wifi, 0 otherwise |
| Sports TV | int64 | 1 if the restaurant has sports TV, 0 otherwise |
| Wine Tasting | int64 | 1 if the restaurant offers wine tasting, 0 otherwise |
| Beer | int64 | 1 if the restaurant serves beer, 0 otherwise |
| Restricted Entry | int64 | 1 if the restaurant has restricted entry, 0 otherwise |
| Variable Menu | int64 | 1 if the restaurant has a variable menu, 0 otherwise |
| Serves Non Veg | int64 | 1 if the restaurant serves non-vegetarian food, 0 otherwise |
| Table booking for Groups | int64 | 1 if the restaurant accepts table bookings for groups, 0 otherwise |
| Poolside | int64 | 1 if the restaurant is poolside, 0 otherwise |
| Home Baker | int64 | 1 if the restaurant is a home baker, 0 otherwise |
| Buffet | int64 | 1 if the restaurant offers a buffet, 0 otherwise |
| Board Games | int64 | 1 if the restaurant has board games, 0 otherwise |
| Vegan Options | int64 | 1 if the restaurant offers vegan options, 0 otherwise |
| Table Reservation Not Required | int64 | 1 if table reservation is not required, 0 otherwise |
| Table reservation required | int64 | 1 if table reservation is required, 0 otherwise |
| Rooftop | int64 | 1 if the restaurant has a rooftop, 0 otherwise |
| BYOB | int64 | 1 if the restaurant allows bring your own beverage, 0 otherwise |
| Egg Preparations | int64 | 1 if the restaurant serves egg preparations, 0 otherwise |
| Live Sports Screening | int64 | 1 if the restaurant screens live sports, 0 otherwise |
| Luxury Dining | int64 | 1 if the restaurant offers luxury dining, 0 otherwise |
| Seating Not Available | int64 | 1 if seating is not available, 0 otherwise |
| Desserts and Bakes | int64 | 1 if the restaurant serves desserts and bakes, 0 otherwise |
| Card Upon Delivery | int64 | 1 if card payment is accepted upon delivery, 0 otherwise |
| Serves Jain Food | int64 | 1 if the restaurant serves Jain food, 0 otherwise |
| Mall Parking | int64 | 1 if the restaurant has mall parking, 0 otherwise |
| Available for Functions | int64 | 1 if the restaurant is available for functions, 0 otherwise |
| Gastro Pub | int64 | 1 if the restaurant is a gastro pub, 0 otherwise |
| Live Entertainment | int64 | 1 if the restaurant has live entertainment, 0 otherwise |
| Sugar Free Options | int64 | 1 if the restaurant offers sugar-free options, 0 otherwise |
| Couple Entry Only | int64 | 1 if the restaurant is for couples only, 0 otherwise |
| Above 18 Only | int64 | 1 if the restaurant is for those above 18 years old only, 0 otherwise |
| Drive in | int64 | 1 if the restaurant offers drive-in service, 0 otherwise |
| Drive Thru | int64 | 1 if the restaurant has a drive-thru, 0 otherwise |
| Valet Parking Available | int64 | 1 if valet parking is available, 0 otherwise |
| Indoor Seating | int64 | 1 if the restaurant has indoor seating, 0 otherwise |
| Free Wifi | int64 | 1 if the restaurant offers free wifi, 0 otherwise |
| Nightlife | int64 | 1 if the restaurant is known for its nightlife, 0 otherwise |
| Serves Cocktails | int64 | 1 if the restaurant serves cocktails, 0 otherwise |
| Kid Friendly | int64 | 1 if the restaurant is kid-friendly, 0 otherwise |
| Craft Beer | int64 | 1 if the restaurant serves craft beer, 0 otherwise |
| All Day Breakfast | int64 | 1 if the restaurant serves breakfast all day, 0 otherwise |
| Gluten Free Options | int64 | 1 if the restaurant offers gluten-free options, 0 otherwise |
| Group Meal | int64 | 1 if the restaurant offers group meals, 0 otherwise |
| Table booking not available | int64 | 1 if table booking is not available, 0 otherwise |
| Full Bar Available | int64 | 1 if the restaurant has a full bar, 0 otherwise |
| Members Only | int64 | 1 if the restaurant is members only, 0 otherwise |
| Standing Tables | int64 | 1 if the restaurant has standing tables, 0 otherwise |
| Karaoke | int64 | 1 if the restaurant has karaoke, 0 otherwise |
| Entry Fee | int64 | 1 if the restaurant charges an entry fee, 0 otherwise |
| Self Service | int64 | 1 if the restaurant is self-service, 0 otherwise |
| Delivery Only | int64 | 1 if the restaurant offers delivery only, 0 otherwise |
| LGBTQIA Friendly | int64 | 1 if the restaurant is LGBTQIA friendly, 0 otherwise |
| No Alcohol Available | int64 | 1 if the restaurant does not serve alcohol, 0 otherwise |
| Unlimited Pizz | int64 | 1 if the restaurant offers unlimited pizza, 0 otherwise |
| Private Dining Area Available | int64 | 1 if the restaurant has a private dining area, 0 otherwise |
| Smoking Are | int64 | 1 if the restaurant has a smoking area, 0 otherwise |
| Table booking recommended | int64 | 1 if table booking is recommended, 0 otherwise |
| Resto Bar | int64 | 1 if the restaurant is a resto bar, 0 otherwise |
| Serves Alcohol | int64 | 1 if the restaurant serves alcohol, 0 otherwise |
| Breakfast | int64 | 1 if the restaurant serves breakfast, 0 otherwise |
| Catering Available | int64 | 1 if the restaurant offers catering services, 0 otherwise |
| Disabled Friendly | int64 | 1 if the restaurant is disabled friendly, 0 otherwise |
| Serves Halal | int64 | 1 if the restaurant serves halal food, 0 otherwise |
| Takeaway Only | int64 | 1 if the restaurant offers takeaway only, 0 otherwise |
| BYOB Only | int64 | 1 if the restaurant allows only bring your own beverage, 0 otherwise |
| spam\_review | float64 | Score indicating the likelihood of a review being spam |

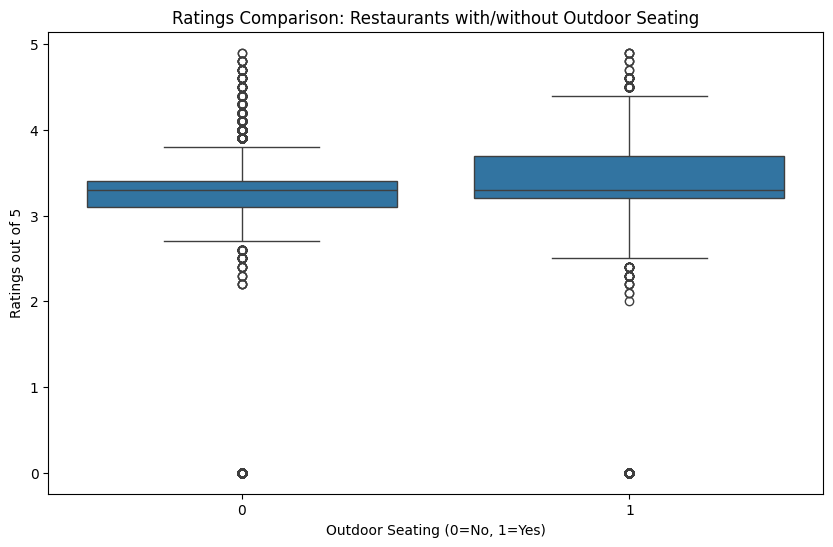
The dataset used in this study initially contained approximately 12,000 restaurants with a staggering 104 features. Each row represents an individual restaurant, encompassing a wealth of information, from basic details like name, location, and cuisine to more nuanced features like availability of outdoor seating, live music, and even LGBTQIA-friendly status. To prepare the data for analysis, several preprocessing steps were undertaken. Missing numeric values, a common occurrence in real-world datasets, were handled using median imputation. This method was chosen for its robustness against outliers, which could skew the mean. Further, outliers in the 'Charges\_for\_two' column were identified and removed using the four standard deviation rule. Any data point falling outside of four standard deviations from the mean was considered an outlier. This approach ensures that extreme values do not disproportionately influence the analysis, preserving the integrity of the findings. Several columns containing numerical information were initially stored as text. These were converted to numeric data types using regular expressions to extract numerical characters. This conversion is crucial for performing quantitative analysis.

Figure 1: Right skewed restaurant ratings

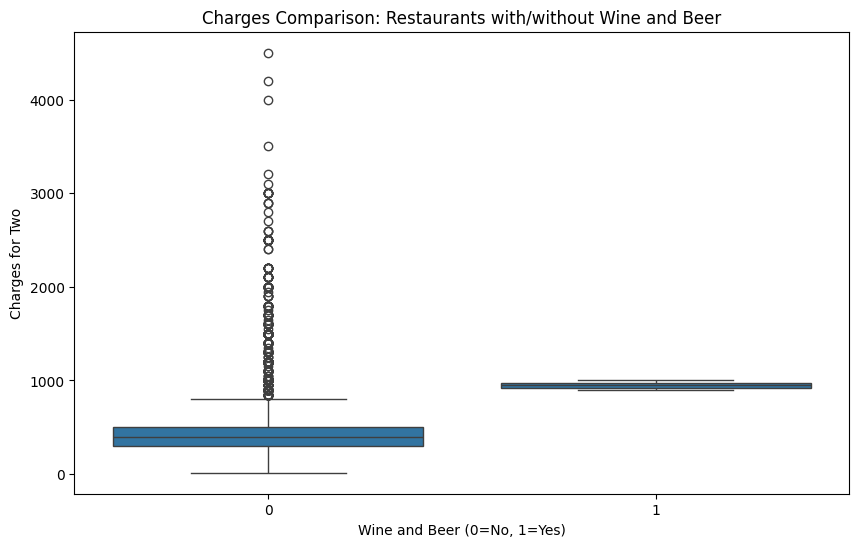
This study addresses the following research question: What is the distribution of restaurant ratings (out of 5) in Pune, and is it normally distributed? The null hypothesis posits that the distribution of restaurant ratings is normally distributed, while the alternative hypothesis proposes that the distribution is not normally distributed. These hypotheses provide a framework for our statistical investigation.

Figure 2: Relationship between charges and rating

The distribution of restaurant ratings is visualized using a histogram. The histogram reveals a left-skewed distribution, indicating a concentration of ratings towards the higher end of the scale (closer to 5). The KDE curve, overlaid on the histogram, smooths the distribution and confirms this left skew. The programmatic interpretation of the histogram, generated by the code, identifies the distribution as left-skewed. This observation is crucial in assessing the normality of the distribution and answering our research question. The distribution's left skew deviates from the perfect symmetry characteristic of a normal distribution.

Figure 3: Boxplot for sitting arrangement and Ratings

To investigate the relationship between restaurant ratings and prices, a scatter plot of 'Ratings\_out\_of\_5' versus 'Charges\_for\_two' was generated, along with a trendline. Visual inspection of the scatter plot suggests a potential positive correlation, although not particularly strong. The calculated Pearson correlation coefficient between ratings and charges is 0.178. This value indicates a weak positive correlation. Furthermore, the p-value associated with this correlation is extremely small (1.257e-87), far below the conventional significance level of 0.05. This signifies that the observed correlation is statistically significant, rejecting the null hypothesis of no correlation. In essence, higher-priced restaurants tend to have slightly higher ratings, but the relationship isn't dominant.

Figure 4: Comparison of retaurants prices serving wine and beer

Beyond the primary analysis, additional investigations were conducted using box plots. One box plot compares the ratings of restaurants with and without outdoor seating. While both groups show a similar median rating, the interquartile range (IQR) for restaurants with outdoor seating is slightly narrower, implying less variability in ratings. This suggests that outdoor seating might not dramatically impact average ratings but could contribute to more consistent customer experiences. Another box plot compares the average charges for two people at restaurants with and without the option of wine and beer. As expected, restaurants serving alcohol tend to have higher average charges. The difference in medians is substantial, indicating a clear price distinction between these two categories of restaurants.

In summary, this analysis provides insights into several aspects of Pune's restaurant market based on Zomato data. The distribution of ratings is left-skewed, suggesting a tendency towards higher ratings. A statistically significant, albeit weak, positive correlation exists between ratings and charges for two. Additional analyses reveal that the presence of outdoor seating might influence the consistency of ratings, and serving alcohol is clearly associated with higher prices. These findings must be interpreted considering the inherent limitations of the dataset. The data represents a snapshot in time (up to December 2019) and is sourced from a single platform (Zomato). It might not fully capture the complexity of the entire restaurant market or individual dining experiences. Future research could explore the impact of other factors on restaurant ratings, like cuisine type, location, specific amenities, or even analyze sentiment expressed in customer reviews. Investigating trends across different localities within Pune could also offer more granular insights into customer preferences.

# Appendix

GitHub repository link: <https://github.com/ma24aeu/7com1079-A309>

Group code: A309