Level 3 Data Analysis Report

Cognifyz Technologies Internship Program

Date: October 2025

Intern Name: Shazia Afraz

Organization: Cognifyz Technologies Internship

Duration: 28 September 2025 – 28 October 2025

Program: Data Analysis Internship With Python

Introduction

This Level 3 Data Analysis Report presents a comprehensive examination of advanced customer behavior and operational strategies within the restaurant industry. It analyzes review text sentiment and length, explores the link between votes and ratings, and assesses the distribution of service offerings, such as online delivery and table booking, across budget and premium market segments. The findings provide actionable insights to help restaurants enhance customer experience, drive engagement, and optimize service strategies for sustainable success.

Task 1: Restaurant Reviews

Objective

- To analyze restaurant review texts, identify the most common positive and negative keywords, calculate the average length of reviews, and explore whether review length influences aggregate ratings.

Explain Code

- The review text is processed using text analysis (such as word tokenization and frequency counts) to determine which positive and negative keywords are most common.

- Positive and negative word lists are defined and used to filter relevant keywords.

- The length of each review is measured and averaged, then reviews are grouped by text length to study any correlation with rating.

• Visualizations are generated using Matplotlib: horizontal bar charts for top keywords and scatter plots for the relationship between review length and ratings.

Result:

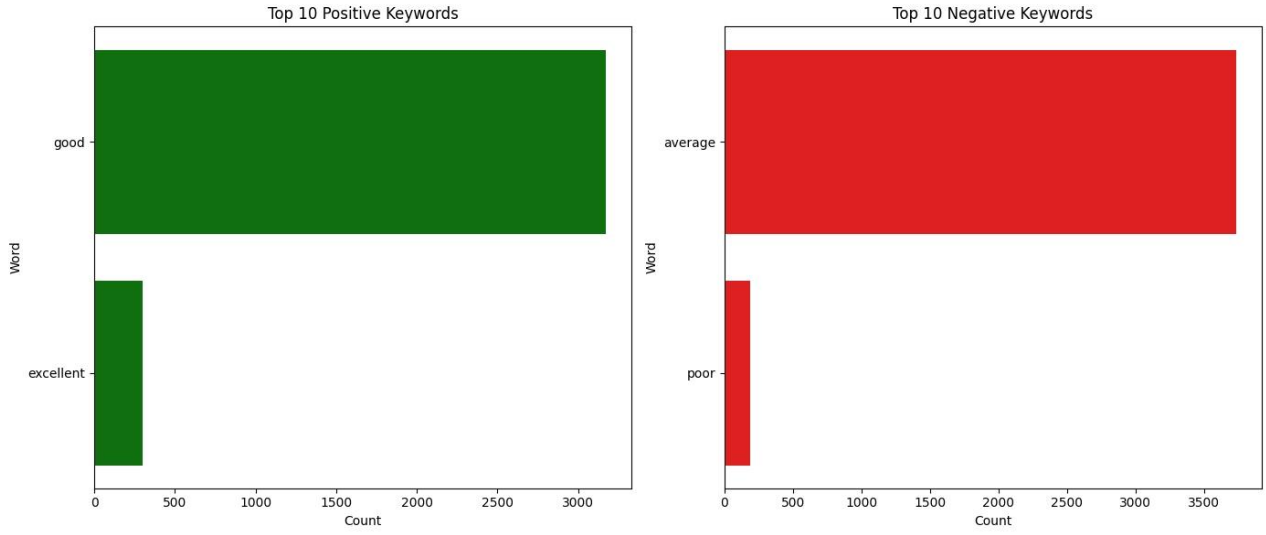
• "Good" and "excellent" are the most common positive keywords in customer reviews, while "average" and "poor" are the most frequent negative ones.

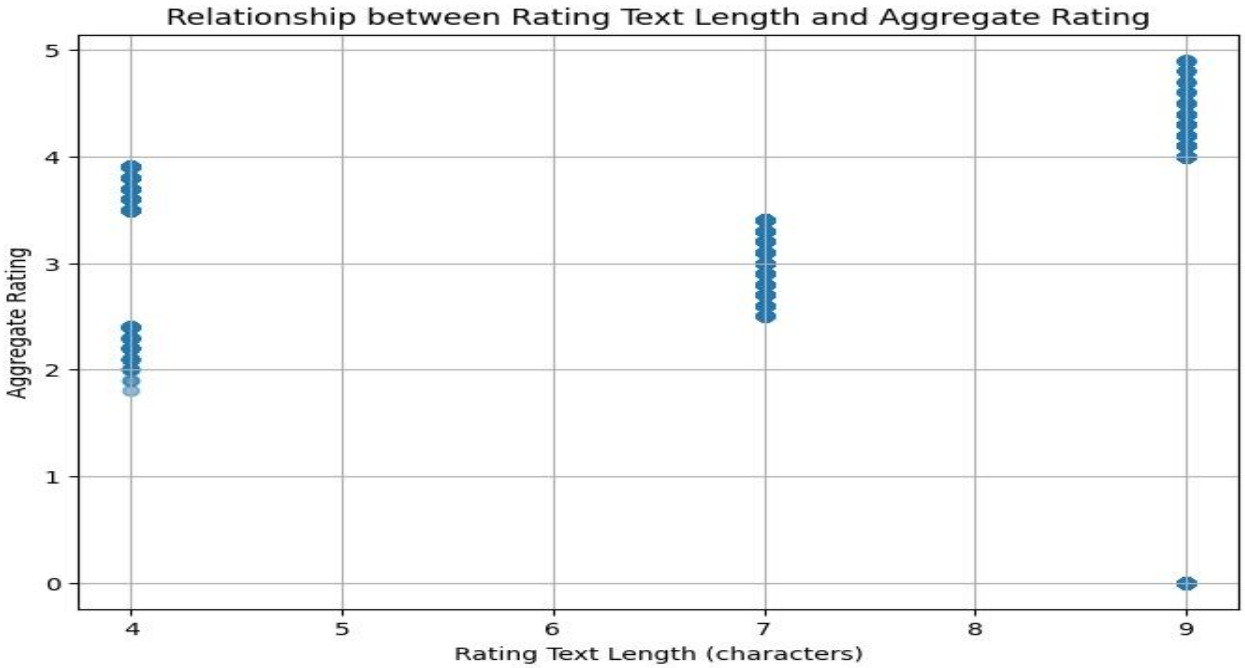
• Most reviews are concise, but those with longer text lengths may correlate with higher average ratings. Reviews with a text length of 9 characters received noticeably higher ratings.

```
Average Rating by Rating Text:
Rating text
Excellent    4.659667
Very Good    4.167904
Good         3.683063
Average      3.051339
Poor         2.297849
Not rated    0.000000
Name: Aggregate rating, dtype: float64

Most Common Positive Keywords:
good: 3174
excellent: 300

Most Common Negative Keywords:
average: 3734
poor: 186
```

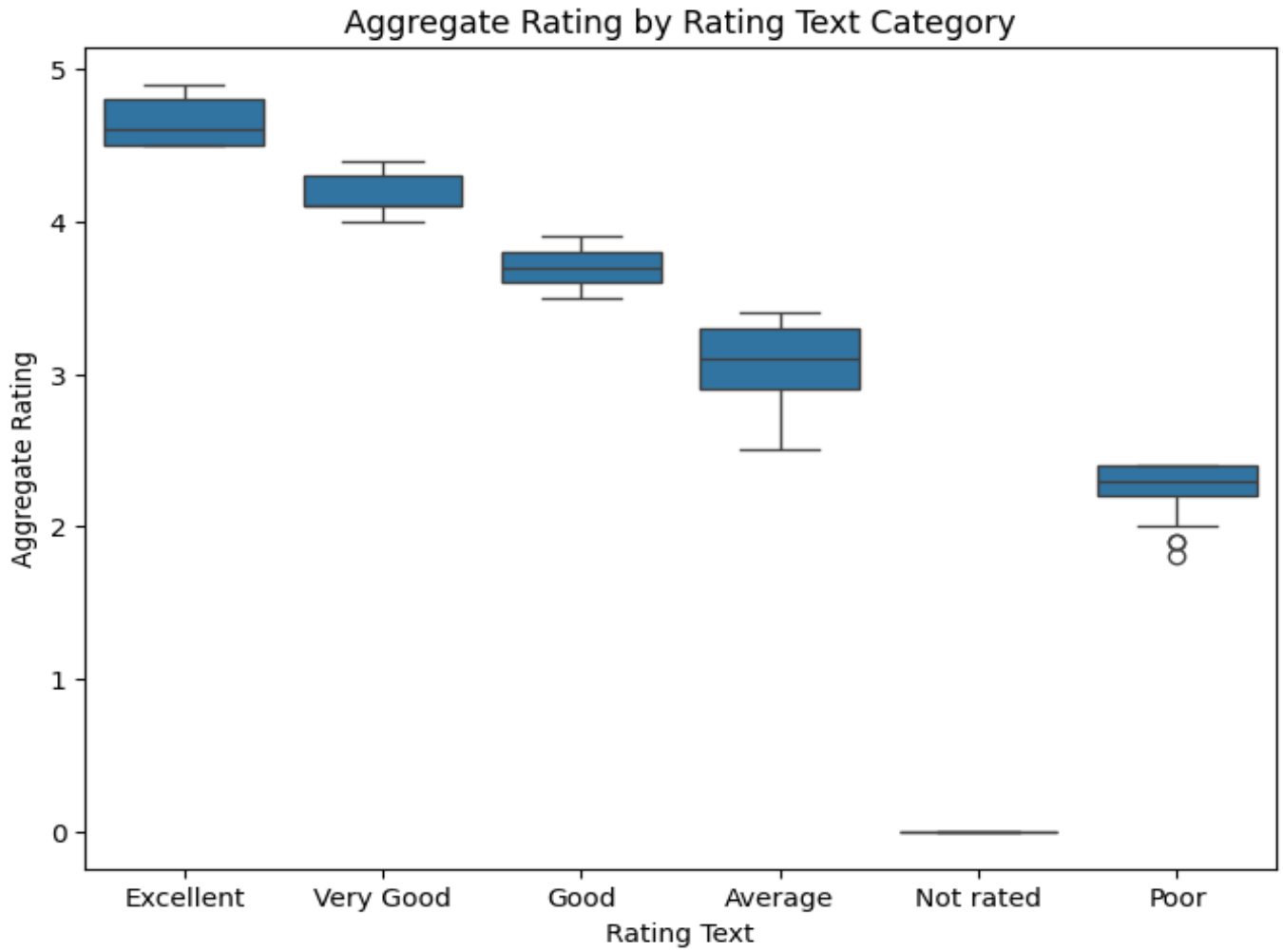





```
Average length of Rating text: 7.02 characters

Average length per Rating text category:
Rating text
Excellent    9.0
Not rated    9.0
Very Good    9.0
Average      7.0
Good         4.0
Poor         4.0
Name: Rating Length, dtype: float64

Correlation between Rating Length and Aggregate Rating: -0.48
```



Visualization:

- The first visual presents two horizontal bar charts: one shows the counts of the top 10 positive keywords (with "good" and "excellent" most frequent), and the other shows the top 10 negative keywords ("average" and "poor" stand out).

- The second plot is a scatter chart displaying review text length versus aggregate rating for each review. Groupings along the x-axis reveal if shorter or longer reviews tend to receive higher ratings.

- The Third plot is Higher aggregate ratings are consistently associated with more positive review text categories (such as "Excellent" and "Very Good"), while ratings drop sharply for neutral or negative categories like "Average", "Poor", or "Not rated".

Business Impact:

- Restaurants can use frequent keywords to understand what resonates with customers, adapting their marketing and service language.

- Insights into review length and rating patterns help tailor customer engagement strategies, encouraging more detailed feedback where it relates to higher satisfaction.

- Addressing negative feedback signaled by top keywords can directly improve reputation and customer experience.

Task 2 – Votes Analysis

Objective

- Identify which restaurants have received the highest and lowest number of votes.

- Analyze whether there is a correlation between the number of votes a restaurant receives and its aggregate rating.

Explain Code:

- The code calculates the sum of votes for each restaurant using grouping and aggregation.

- The top 20 restaurants by vote count are selected and plotted in a bar chart.

- To examine correlation, a scatter plot (bubble plot) is generated showing each restaurant's aggregate rating on the x-axis, votes on the y-axis, and bubble size indicating vote volume.

Result:

- Restaurants like Barbeque Nation and AB's Absolute Barbecues lead in total votes, indicating high customer engagement.

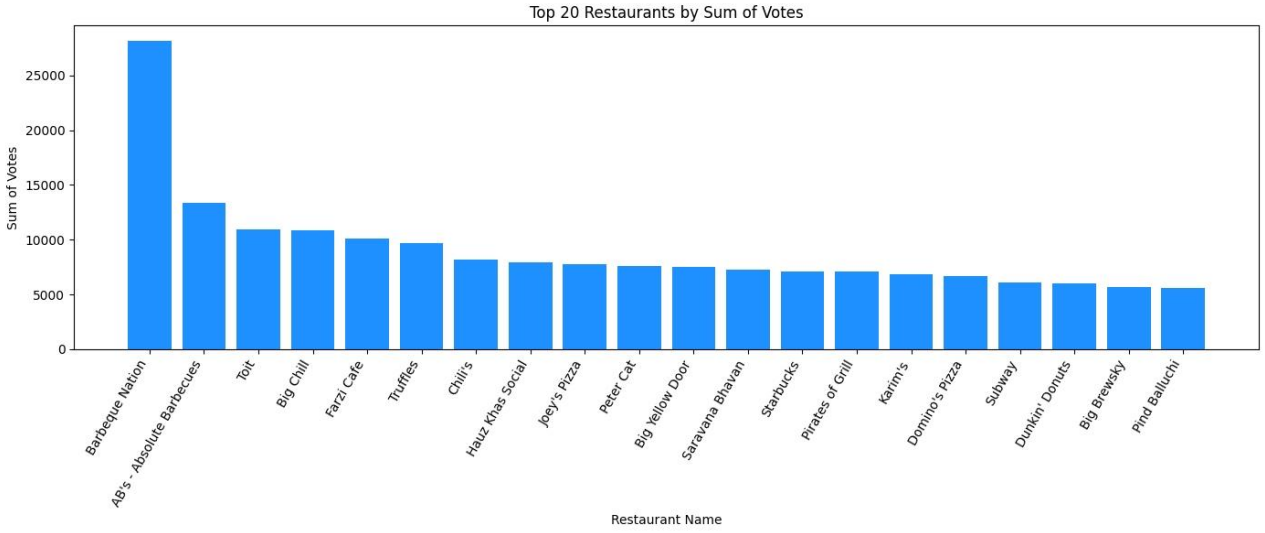
- The bubble plot shows a trend where restaurants with higher ratings typically receive more votes, suggesting that satisfied customers are more likely to provide feedback.

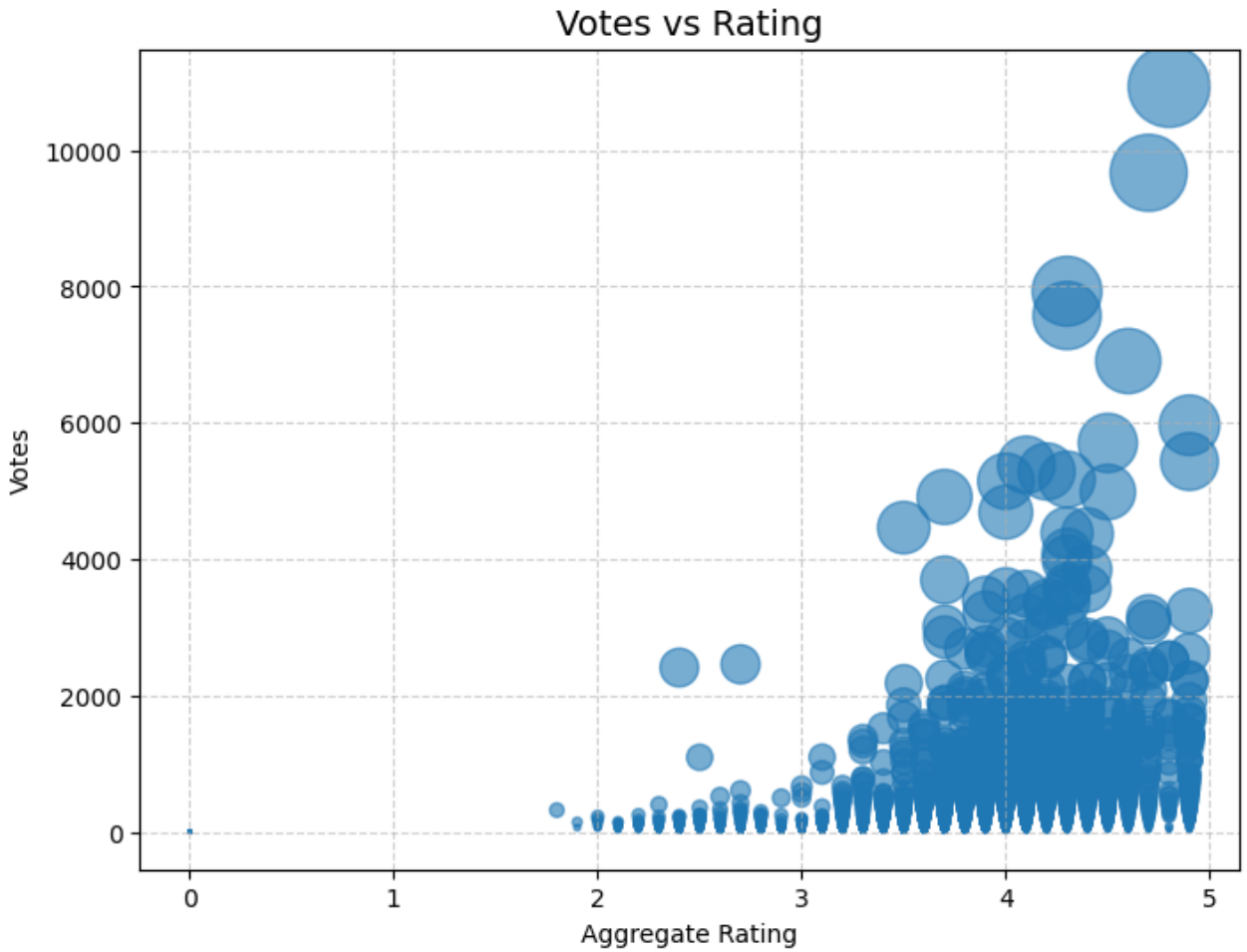
Top 10 Restaurants by Votes:

	Restaurant Name	Votes	Aggregate rating
719	Toit	10934	4.8
726	Truffles	9667	4.7
3985	Hauz Khas Social	7931	4.3
2403	Peter Cat	7574	4.3
730	AB's - Absolute Barbecues	6907	4.6
2405	Barbeque Nation	5966	4.9
734	Big Brewsky	5705	4.5
2298	AB's - Absolute Barbecues	5434	4.9
727	The Black Pearl	5385	4.1
2402	BarBQ	5288	4.2

Lowest 10 Restaurants by Votes:

	Restaurant Name	Votes	Aggregate rating
69	Cantinho da Gula	0	0.0
865	The Chaiwalas	0	0.0
870	Fusion Food Corner	0	0.0
871	Punjabi Rasoi	0	0.0
878	Baskin Robbin	0	0.0
891	Angaar	0	0.0
892	TcozY	0	0.0
893	The Retriever	0	0.0
896	Chill 'N Grill	0	0.0





Visualization:

• The first visual (bar chart) highlights the top 20 restaurants by their total votes, showing which establishments are most actively reviewed by customers.

• The second visual (bubble scatter plot) displays the relationship between aggregate ratings and votes, making patterns of popularity and satisfaction easy to observe.

Business Impact:

• High-vote restaurants can leverage their popularity for marketing, partnerships, or menu expansion.

• Understanding the vote-rating correlation helps businesses focus on delivering excellent service and food to boost both public reputation and engagement.

• Restaurants with low votes and ratings may need targeted improvements or promotional efforts to increase visibility and customer feedback.

Task 3 – Price Range vs. Online Delivery and Table Booking

Objective:

To analyze the availability of online delivery and table booking services across restaurant price ranges, highlighting how service offerings differ between budget and premium establishments.

Explain Code:

1. The code cleans relevant service columns, converts "Yes"/"No" answers to binary values, then groups the data by price range to calculate what percentage of restaurants at each price level offer online delivery and/or table booking.

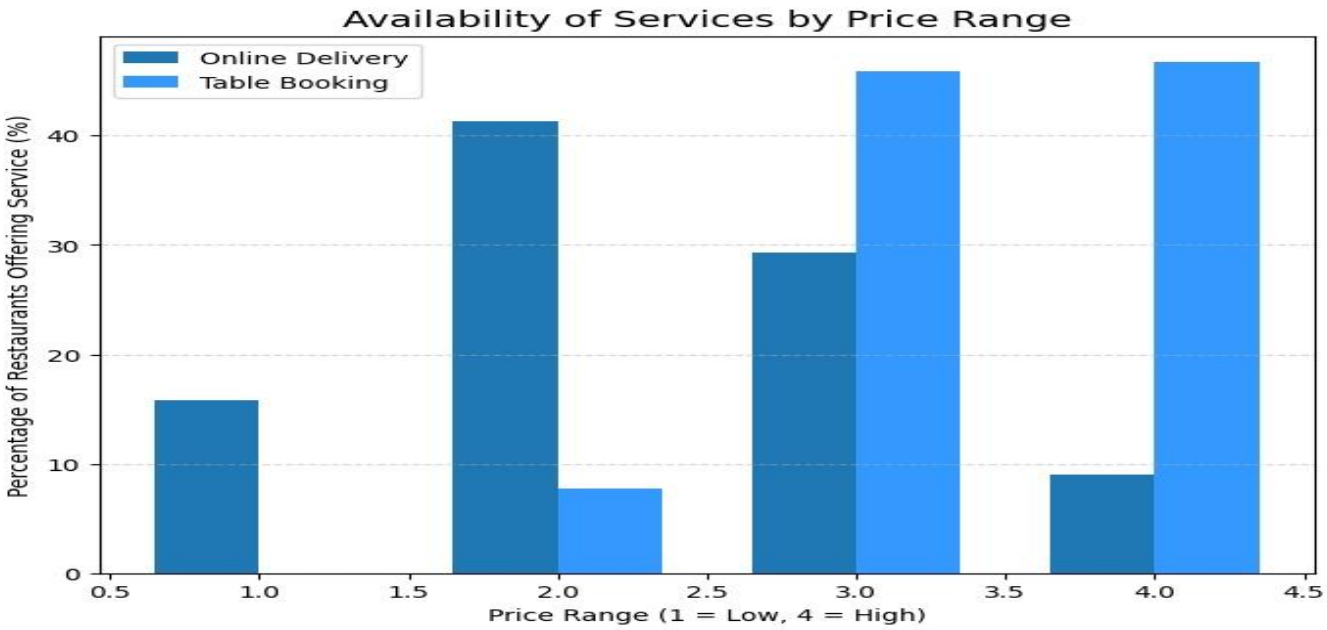
2. It then creates a grouped bar chart to visually compare the availability of these two services across different price ranges, helping spot trends in service offerings by price category.

Results:

- Online Delivery is most available in price range 2 (over 40%), with moderate availability in range 3 and lower presence in ranges 1 and 4.

- Table Booking is almost absent in lower price ranges (1 and 2), but jumps sharply in the premium segments (over 45% in price ranges 3 and 4).

- Thus, budget restaurants are more likely to offer online delivery, while high-end restaurants predominantly provide table booking services.



Visualization:

The grouped bar chart "Availability of Services by Price Range" displays:

- Price ranges on the x-axis and the percentage of restaurants offering each service on the y-axis.
- Two sets of bars per range: blue for online delivery, light blue for table booking.
- Clear contrast between service type prevalence in relation to price segments.

Business Implications:

- Targeted Service Offering: Budget and mid-range restaurants should continue prioritizing online delivery, tapping into the mass market's demand for convenience and digital access.
- Premium Differentiation: High-end establishments should emphasize exclusive experiences, with table booking as a critical component for personalization, event hosting, and hospitality.
- Expansion Strategies: Restaurant owners can tailor new outlets' service mix to match the expectations of their price segment—driving operational efficiency and meeting customer preferences.
- Customer Segmentation: Understanding these service trends enables sharper marketing strategies, ensures the right value proposition, and supports optimal technology investments for each market tier.

Conclusion:

The Level 3 analysis highlights key drivers of restaurant reputation and competitiveness. Positive keywords and longer reviews are associated with higher ratings, suggesting that satisfied customers tend to be more expressive. Popular restaurants with high ratings consistently attract more votes, confirming that quality leads to greater engagement. Service offerings vary by price segment—mid-range establishments focus on online delivery, while premium outlets specialize in table booking. Together, these insights enable businesses to tailor marketing, operations, and technology investment to meet the preferences of their target market, reinforcing loyalty and growth in a dynamic industry.

Prepared by: SHAZIA AFRAZ
Intern – Data Analytics
Cognifyz Technologies