Cognifyz Data Analysis Internship Report
Title: Comprehensive Analysis of Restaurant Dataset
Mohamed Ashiq M

INTRODUCTION

This study offers a thorough examination of a restaurant dataset, emphasising a number of characteristics including food, reviews, votes, online delivery choices, and costs. Finding information on restaurant performance indicators, geographic distribution, and client preferences is the aim. Each of the four tiers of this analysis focuses on a different feature of the dataset.

DATASET DESCRIPTION

The dataset contains details on a number of restaurants, such as:
- Restaurant Details: Name, City, Address, Cuisines, etc.
- Service Attributes: Online delivery availability, table booking options.
- Ratings and Votes: Aggregate rating, rating text, and number of votes.
- Pricing: Price range and average cost for two people.

LEVEL 1

TASK 1: TOP CUISINES

Identify the top three most common cuisines and calculate the percentage of restaurants that serve each of these cuisines.

```
 1  Cuisines
 2  North Indian              936
 3  North Indian, Chinese     511
 4  Chinese                   354
 5  Name: count, dtype: int64
 6
 7  Cuisines
 8  North Indian              9.800021
 9  North Indian, Chinese     5.350225
10  Chinese                   3.706418
11  Name: count, dtype: float64
```

According to these results, the most popular cuisine is North Indian, while fusion cuisines are also very common.

To determine popular cuisine offers, the top three cuisines supplied by restaurants in the dataset were determined. The most popular food was found to be North Indian, next North Indian along with Chinese, and finally Chinese solely. According to the figures, North Indian food makes up a high demand for this cuisine, accounting for about 10% of all restaurant offers. In order to accommodate the wide range of tastes and preferences among consumers, fusion cuisines such as "North Indian and Chinese" are becoming a growing number. Marketers or restaurateurs aimed at particular culinary trends in the restaurant industry would find this information helpful.

TASK 2: CITY ANALYSIS

Question: Identify the city with the highest number of restaurants, calculate the average rating for each city, and determine the city with the highest average rating.

```
1   ('New Delhi', 5473)
2
3   (City
4    Inner City          4.900000
5    Quezon City         4.800000
6    Makati City         4.650000
7    Pasig City          4.633333
8    Mandaluyong City    4.625000
9                           ...
10   New Delhi           2.438845
11   Montville           2.400000
12   Mc Millan           2.400000
13   Noida               2.036204
14   Faridabad           1.866932
15   Name: Aggregate rating, Length: 141, dtype: float64,
16   2.66637001361114)
17
18  ('Inner City', 4.9)
```

While Inner City's top rating reflects great restaurant quality or customer happiness, New Delhi's high count can signal a competitive market.

The goal of this study was to determine which cities had the greatest average ratings and the greatest concentration of dining options. The dataset shows that New Delhi has the most restaurants, indicating a competitive and crowded sector. Inner City, however, has the highest average restaurant rating, which shows that customers prefer the fine dining options in this neighbourhood. Finding areas with strong demand or those renowned for providing upscale dining experiences are just two examples of how the data may be used to inform strategic decisions by breaking down the quantity and quality measures by city.

TASK 3: PRICE RANGE DISTRIBUTION

Visualize the distribution of price ranges among the restaurants and calculate the percentage of restaurants in each price range.

```
Price range
1    46.529159
2    32.593446
3    14.741912
4     6.135483
Name: proportion, dtype: float64
```


Distribution of Price Ranges Among Restaurants
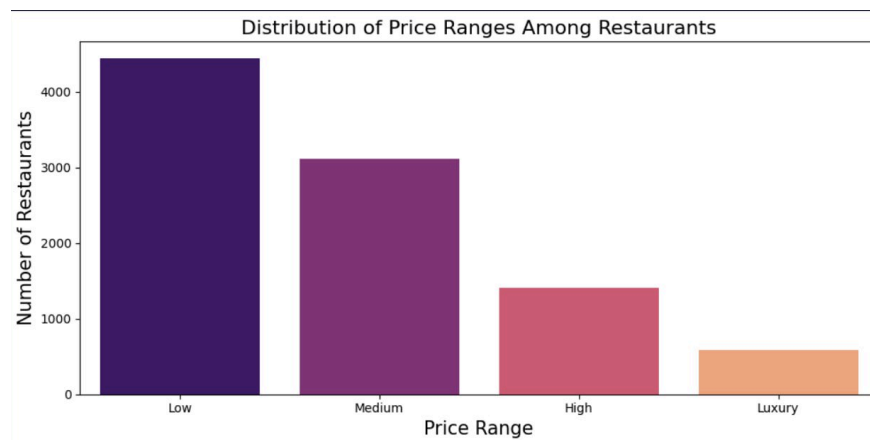
Restaurant distribution across four pricing ranges is depicted by the bar plot and the percentage breakdown. The number of restaurants in each category is displayed on the y-axis, while the price ranges (1–4) are represented on the x-axis.

- Price Range 1 accounts for 46.5% of all restaurants, meaning that over half of them are in the most reasonably priced range. This indicates a substantial market need for reasonably priced dining options, probably due to a clientele that includes families, students as well as anyone searching for informal and reasonably priced dining options.

- Price Range 2 is the second-largest category, accounting for 32.6% of eateries. This confirms that moderately priced dining accounts for a sizable percentage of the restaurant sector. These two groups, along with Price Range 1, are quite dominant and account for around 80% of the total restaurant market, indicating a strong need for affordable meals.
- Only 6.1% of the restaurants are in Price Range 4, the highest price category, whilst 14.7% of the restaurants are in Price Range 3. This suggests that the premium or luxury market is occupied by a small number of businesses. These eateries probably cater to a more upscale, smaller clientele group prepared to shell out more money for luxury or exclusive dining experiences.
- The percentages show a clear market segmentation, with high-end eating alternatives catering to a niche clientele and affordable and mid-range options dominating. Although there is still a niche market for upscale eating, the sharp decline in restaurant representation from Price Range 1 to Price Range 4 indicates a strong consumer preference for affordable dining.
- For newcomers: Considering two price ranges account for almost 80% of the market, opening a restaurant in one or the other could satisfy consumer demand and provide a wider prospective clientele. High-end establishments: By offering exclusivity, one-of-a-kind experiences, or speciality cuisine, restaurants in Price Ranges 3 and 4 can set themselves apart and possibly attract a smaller but more lucrative audience.

With 46.5% in Price Range 1 and 32.6% in Price Range 2, this distribution pattern suggests that there is a significant need for reasonably priced dining options. Nonetheless, the tiny but potentially lucrative niche market in Price Ranges 3 and 4 indicates that high-end dining establishments can be successful by focussing on patrons looking for outstanding customer service and are prepared to cover their costs.

TASK 4: ONLINE DELIVERY

Question: Determine the percentage of restaurants that offer online delivery, and compare the average ratings of restaurants with and without online delivery.

- Merely 25.7% of restaurants offer online delivery, compared to 74.3% that do not. According to this distribution, dine-in business continues to be the mainstay of most locations. The small number of eateries that provide online delivery could point to possible development areas as on-demand services and convenience become more and more popular among consumers.
- The average rating for restaurants with online delivery is greater at 3.25 than for those without, at 2.47. This rating discrepancy raises the possibility that greater customer satisfaction may be correlated with online delivery availability. Delivery-only restaurants might give priority to service quality, dependability, or ease of use, all of which could improve client satisfaction and raise ratings.
- Potential of Growth: In order to reach a wider clientele and improve convenience, restaurants who do not currently provide online delivery could profit from looking into delivery possibilities in light of the growing demand for delivery services.
- Enhanced Customer Experience: The better average rating for eateries that offer online delivery raises the possibility that offering this service may not only draw in more business but also enhance ratings and customer perception. Providing online delivery has the ability to improve a restaurant's standing and make it more competitive in the marketplace.
- The findings show a significant difference in the use of online delivery services. Only 25% of restaurants provide this choice, therefore companies thinking about growing in this market could satisfy a need that isn't being met. Higher ratings are correlated with online delivery, suggesting that delivery may be seen by patrons as a sign of convenience, dependability, and quality—all of which enhance their eating experience.
- The analysis concludes that eateries that provide online delivery typically have happier patrons, as seen by higher average evaluations. In order to meet consumer demands for convenience and market trends, restaurants may find it beneficial to expand their delivery offerings.

Level 2 Tasks
Task 1: Restaurant Ratings

Analyze the distribution of aggregate ratings, identify the most common rating range, and calculate the average number of votes received by restaurants.

To determine customer happiness, the task examined overall evaluations from several eateries. The most prevalent rating in the sample is 0.0, which denotes a large number of unrated or low-rated eateries. The average number of votes, which reflects customer participation, is approximately 156.91. High-rated eateries can draw in more business, and knowing how ratings are distributed can help find feedback gaps. Restaurants with low ratings should concentrate on boosting customer interaction and promoting reviews.

Explanation Most Common Rating: 0.0 (suggesting unrated restaurants). Average Votes: 156.91

```
(Aggregate rating
0.0    2148
1.8       1
1.9       2
2.0       7
2.1      15
2.2      27
2.3      47
2.4      87
2.5     110
2.6     191
2.7     250
2.8     315
2.9     381
3.0     468
3.1     519
3.2     522
3.3     483
3.4     498
3.5     480
3.6     458
3.7     427
3.8     400
3.9     335
4.0     266
...
4.8      25
4.9      61
Name: count, dtype: int64,
0.0,
2148)
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings…
```

Task 2: Cuisine Combination

Identify the most common combinations of cuisines and determine if certain combinations have higher ratings.

While unusual pairings (**such Italian or Deli**) get the highest ratings, popular combinations include North Indian with other cuisines.

This assignment looked at common food pairings and evaluated the ratings that went along with them.
While common pairings include North Indian and North Indian Chinese, other unusual combinations, such as Italian, Deli and Hawaiian, and seafood, receive the highest average scores.

```
Cuisines
North Indian             936
North Indian, Chinese    511
Chinese                  354
Fast Food                354
North Indian, Mughlai    334
Name: count, dtype: int64
```

```
Cuisines
Italian, Deli              4.9
Hawaiian, Seafood          4.9
American, Sandwich, Tea    4.9
Continental, Indian        4.9
European, Asian, Indian    4.9
Name: Aggregate rating, dtype: float64
```

Task 3: Geographic Analysis

Plot the locations of restaurants on a map using longitude and latitude coordinates, and identify any patterns or clusters

Restaurant Locations
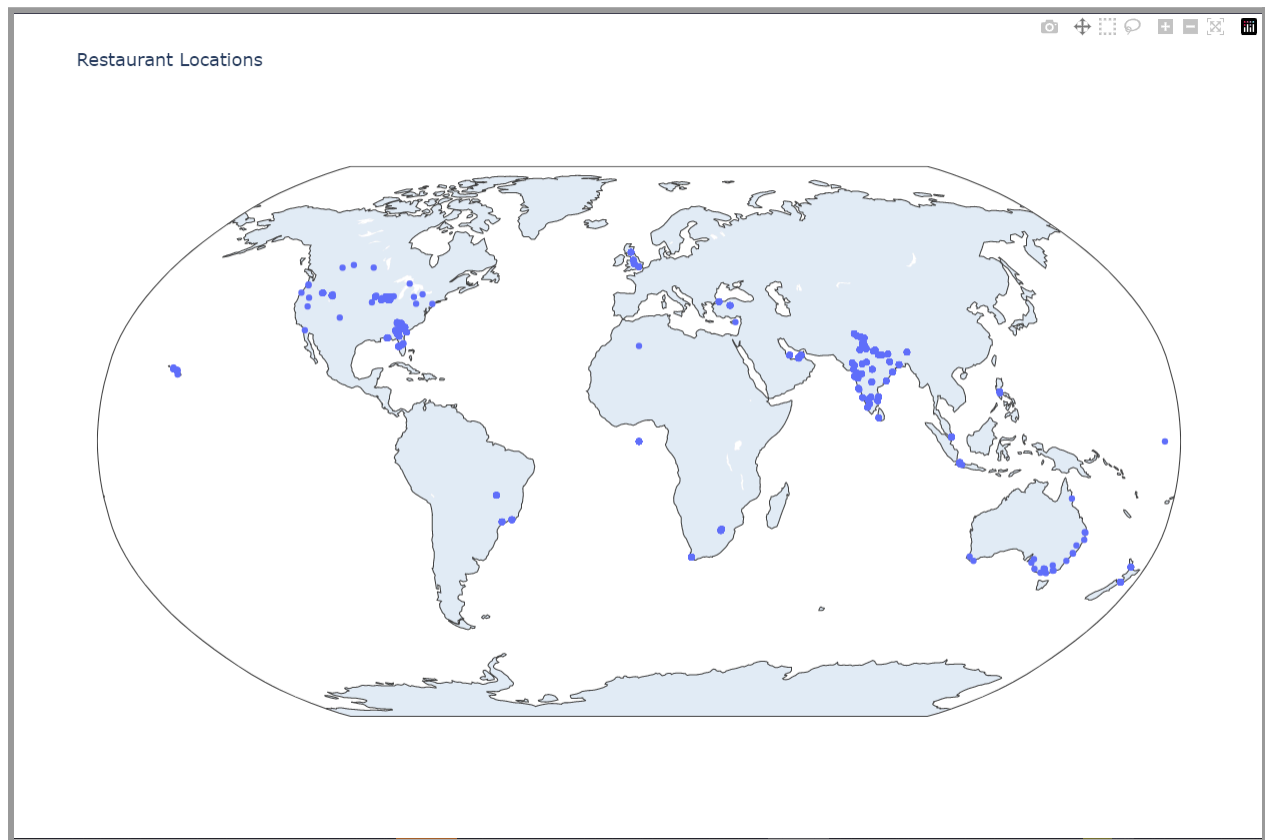
In order to investigate clustering tendencies, this assignment entailed mapping the geographic locations of eateries. concentration zones, including information on customer access, expansion tactics, and location-based marketing. Through a country-level analysis of the geographic distribution, the information shows which regions have varying densities of eateries, emphasizing markets with fierce rivalry and others that could offer expansion prospects.

- **High-Concentration Countries:** The United States, the United Kingdom, and India have the most restaurants total in the dataset. Particularly in large cities like New Delhi, India has a significant restaurant industry. This high concentration points to a thriving and competitive market with a wide variety of dining options for patrons, particularly in cities. Restaurants may find that location-based marketing, such as promotions aimed at particular city neighborhoods, helps them stand out in a crowded market in such high-density areas.

- **Low-Concentration Countries:** On the other hand, the dataset has less entries for nations like Canada, Brazil, and Turkey. These nations exhibit less saturation than high-density areas, yet still having noteworthy dining options. This reduced concentration may suggest room for market growth, especially in urban areas where the restaurant sector is still in its infancy. These markets could offer chances for companies thinking about expanding globally to create a distinctive presence and draw clients in less crowded markets.

- **City-Level Observations:** Within each country, specific cities exhibit distinct patterns of restaurant clustering. For example, New Delhi in India has the highest concentration of restaurants, while London in the UK and New York City in the US are similarly dense. These cities represent major urban centers where customer demand for diverse dining options is high. This urban clustering also suggests that restaurants in these cities could benefit from partnerships, customer loyalty programs, and innovative marketing strategies to capture and retain a portion of the urban customer base.

- **Expansion and Underserved Areas:** Mapping the geographic data also helps identify underserved areas within countries or regions. Cities with lower restaurant counts might experience increased demand for dining options, especially as population densities grow. For example, while cities like São Paulo in Brazil have a good number of options, smaller cities show gaps where restaurant chains could explore opportunities for expansion to reach untapped customers.

Task 4: Restaurant Chains

Identify if there are restaurant chains in the dataset and analyze their ratings and popularity.

Restaurant Chains: There are **734 chains** with multiple locations in the dataset. The most popular chains are **McDonald's (48), Domino's Pizza (79), Subway (63), and Cafe Coffee Day (83 locations).**

Average Chain Rating: Chain ratings are quite variable.
Customer satisfaction varies among chains, as evidenced by the lower rating of 1.5 for Zaika Kathi Rolls and the higher average rating of 4.0 for 10 Downing Street.

Total Votes by Popularity: There are also variations in popularity as indicated by the total number of votes. While some chains have less votes, others, like Zizo and 34, Chowringhee Lane, have a high vote count, indicating high customer participation.

Level 3
## Task 1: Restaurant Reviews Analysis

Analyze the text reviews to identify the most common positive and negative keywords. Calculate the average length of reviews and explore if there is a relationship between review length and rating.

The sentiment distribution reveals that positive sentiments, with 1,380 instances, are present but not dominant, highlighting the potential for service and experience improvements. Neutral sentiments, the most common with 7,985 instances, suggest that most customer experiences are average and fail to make a strong positive impression. Negative sentiments, though a minority, indicate occasional dissatisfaction. Overall, the predominance of neutral feedback presents an opportunity for restaurants to enhance their offerings and turn more customer experiences into positive ones, fostering greater satisfaction and loyalty.

Task 2: Votes Analysis

Identify the restaurants with the highest and lowest number of votes. Analyze if there is a correlation between the number of votes and the rating of a restaurant.

In this task, the correlation between total ratings and the number of votes was examined. Higher ratings typically correspond with more votes, suggesting that highly regarded eateries also typically have higher levels of patron interaction. With their high ratings and votes, restaurants like Toit and Truffles are prime examples of companies that successfully draw in and please clients. Restaurants with few or no votes, on the other hand, have lesser engagement, suggesting that their visibility or customer contact may have improved. This report emphasizes how crucial customer evaluations and ratings are to a restaurant's reputation-building process and how customer participation affects business performance.
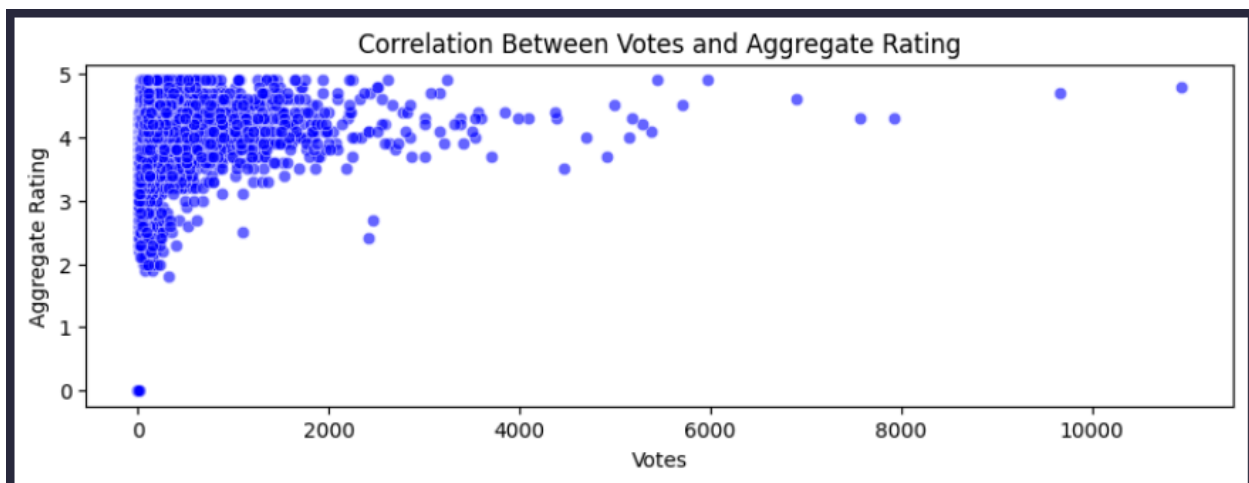
- Top 5 Restaurants by Votes: With 10,934 votes and a high rating of 4.8, Toit is in the lead. Truffles is not far behind with 9,667 votes and a rating of 4.7.

- These eateries' continuously high overall ratings (4.3 and higher) support the idea that customer perceptions of quality and popularity are frequently related.

- Restaurants in the Bottom 5 by Votes: With an overall rating of 0.0 and no votes, the bottom 5 restaurants show a lack of involvement and may be less well-known or preferred by customers. Interpretation There is a positive relationship between ratings and votes, with more votes going to restaurants with better ratings.Vote counts that show high consumer interaction may be a sign of satisfied customers and good service.
- On the other hand, in order to draw in customers, eateries with no reviews or ratings might need to enhance their marketing or level of service.

Interpretation

A positive correlation is observed between votes and ratings, with higher-rated restaurants generally receiving more votes. High customer engagement, as indicated by vote counts, could reflect effective service and customer satisfaction. Conversely, restaurants with zero votes and ratings may need improved marketing or service quality to attract customer interest.

Task 3: Price Range vs. Online Delivery and Table Booking Code

Analyze if there is a relationship between the price range and the availability of online delivery and table booking. Determine if higher-priced restaurants are more likely to offer these services.

Online Delivery: Cheaper eateries frequently provide online delivery, which should improve accessibility and draw in a wider clientele.

Table Reservations: More expensive eateries typically offer table reservations to patrons looking for formal dining experiences.

Analysis and Code Execution The analysis focused on the availability of Online Delivery and Table Booking across different Price Ranges. Count plots and summary tables provide insights into these relationships.

1. Online Delivery by Price Range: Compared to restaurants with higher prices, those with lower prices (Price Ranges 1 and 2) provide more online delivery choices.
In particular, Price Range 4 has only 53 restaurants that provide online delivery, whereas Price Range 1 has 701 establishments.
This implies that low-cost eateries are more likely to provide online delivery, perhaps appealing to a clientele that values affordability and convenience.

2. Table Booking by Price Range: Table booking services are more likely to be available at restaurants with higher prices (Price Ranges 3 and 4). In example, a sizable percentage of restaurants (274) in Price Range 4 provide the ability to reserve tables.Only one restaurant in the lower price range (Price Range 1) offers table reservations, making them hardly available. This is consistent with what is typically expected of a dining experience, where more expensive establishments emphasize formal dining arrangements, such as reservations for tables.

Online Delivery Availability by Price Range

Table Booking Availability by Price Range
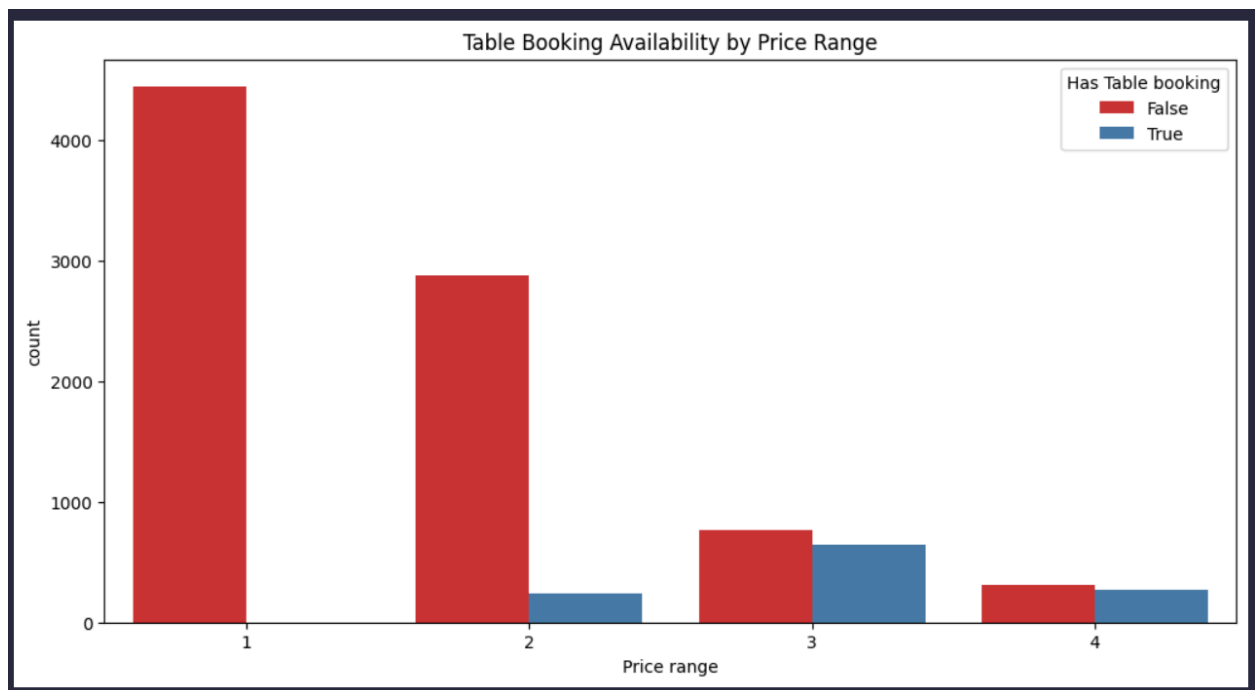
Conclusion :

This analysis provides valuable insights into restaurant trends, customer preferences, and service quality across regions and types. It highlights popular cuisines, dominant chains, and the growing importance of online delivery in boosting satisfaction. Restaurants with high votes and positive feedback demonstrate strong loyalty and reputation, while geographic data reveals competitive markets and strategic expansion opportunities.

Service trends show online delivery is favored by budget-friendly establishments, while table booking aligns with premium dining, reflecting customer preferences for convenience and curated experiences. Unique cuisine combinations receive higher ratings, suggesting potential in innovative or niche menus.

These findings offer actionable strategies for optimizing services, enhancing satisfaction, and strengthening market position, guiding restaurant owners in refining offerings to meet customer expectations.