

Restaurant Business Analytics: Services, Customer Engagement, and Market Insights

Intern Name: Rishi Patel

Organization: Cognifyz Technologies

Project Type: Data Analytics Internship Project

Project Duration: 7-7-25 to 7-8-25

Overview of Project: -

This project, conducted as part of a Data Analytics Internship at **Cognifyz Technologies**.

Focuses on analyzing restaurant industry trends using a real-world dataset.

The analysis was performed using **Python** and key data science libraries including **Pandas**, **Matplotlib**, **Seaborn**, and **Plotly**.

The dataset includes restaurant listings from various cities, capturing details such as cuisine types, pricing, service availability (online delivery and table booking), customer ratings, votes, and geographic location.

Business Value: -

This analysis helps stakeholders understand customer trends, optimize service offerings, and strategize business expansion based on data-driven insights.

Tools Used: -

- Python 3.11.8
- Pandas
- Matplotlib
- Seaborn
- Jupyter Notebook
- GitHub (for version control & publishing)

Introduction

The goal of this project is to analyze restaurant industry data to uncover trends and generate actionable insights for strategic business decision-making.

The dataset contains detailed information about restaurants across various cities, including:

- Restaurant names and locations.
- Types of cuisines served.
- Pricing range categories.
- Availability of online delivery and table booking services.
- Customer ratings and votes.
- Geographic coordinates (longitude and latitude).

Project Objectives: -

- **Analyze Services Offered:**
Examine the availability of online delivery and table booking across different restaurants and pricing categories.
- **Study Customer Engagement:**
Evaluate customer votes and ratings to understand engagement levels and satisfaction trends.
- **Understand Pricing Impact:**
Explore how price range categories affect service offerings and customer behaviour.
- **Geographic Distribution Analysis:**
Map restaurant locations to identify city-wise and region-wise market concentrations.
- **Deliver Actionable Business Insights:**
Provide clear, data-driven recommendations to help restaurants optimize their service strategies, customer engagement, and market positioning.

Data Preparation & Cleaning

- **Initial Dataset Shape:** 9,551 rows × 21 columns.
- **Data Types Identified:**
 - float64: 3 columns
 - int64: 5 columns
 - object: 13 columns
(Verified using `.info()` method.)
- **Summary Statistics:**
Used `.describe()` method to understand numerical distributions, identify outliers, and analyze value ranges.

Steps in Data Cleaning: -

1. Handling Missing Values

- Detected using `.isnull().sum()`.
- Missing values primarily found in the **Cuisines** column.
- Addressed using:
 - Used `.dropna()` method and remove all the Missing values.

2. Removing Noisy & Duplicate Data

- Checked for:
 - **Duplicate rows** using `.duplicated()`.
 - **Whitespace-only entries** across object columns.
- Action Taken:
 - Duplicates and whitespace-only rows were removed to ensure data consistency.

3. Standardizing Text Columns

- Applied **lowercasing** and **whitespace stripping** to all object columns using:
- Ensured uniform representation for:
 - Restaurant names
 - Cities
 - Cuisine types
 - Currency and Rating Text columns

4. Outlier Detection & Treatment

- Used **IQR (Interquartile Range)** method to detect outliers in:
 - Average Cost for Two
- Treatment Strategy:
 - Extreme outliers were:
 - Replaced with **median values** in critical columns to prevent skewing analysis.

5. Final Dataset Verification

- Ensured:
 - No missing/null values.
 - No duplicate records.
 - No noisy/blank rows.
 - Consistent formatting of all columns.
 - Outliers handled appropriately.
- **Final Dataset Shape After Cleaning:**
9,542 rows × 21 columns

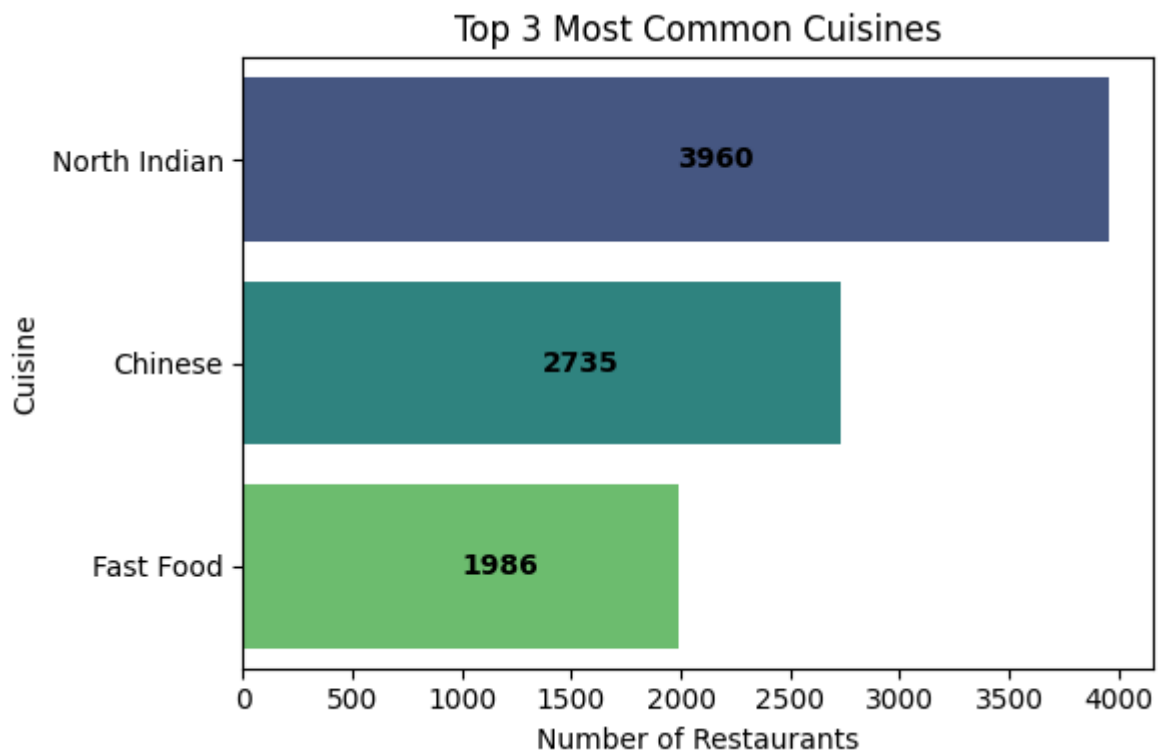
A fully cleaned, structured, and analysis-ready dataset was prepared, providing a strong foundation for trustworthy, noise-free analysis.

Data Analysis & Key Insights

Level 1: - Market Overview and Service Distribution

Task 1: - Top Cuisines

- Determine the top three most common cuisines in the dataset.



- Calculate the percentage of restaurants that serve each of the top cuisines.

Cuisines	
north indian	41.50
chinese	28.66
fast food	20.81

Summary: - The top 3 cuisines served across restaurants are **North Indian (3960)**, **Chinese (2735)**, and **Fast Food (1986)**, contributing approximately **41.50%**, **28.66%**, and **20.81%** respectively to the overall cuisine distribution.

Task 2: - City Analysis

- Identify the city with the highest number of restaurants in the dataset.

```
Top 5 City with Highest number of restaurants:-
City
new delhi      5473
gurgaon        1118
noida          1080
faridabad      251
ghaziabad      25
```

- Calculate the average rating for restaurants in each city.

```
Percentage contribution:-
City
new delhi      57.36
gurgaon        11.72
noida          11.32
faridabad      2.63
ghaziabad      0.26
```

- Determine the city with the highest average rating. (overall, 140 cities)

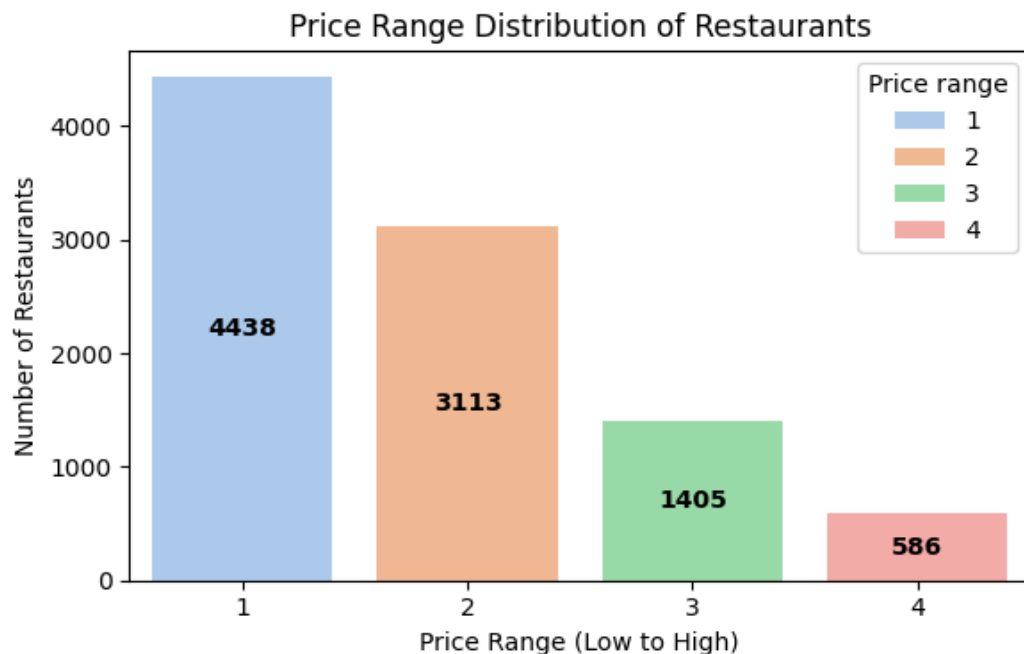
```
Top 10 highest avg. rating city:-
City Average Rating
0 inner city 4.900000
1 quezon city 4.800000
2 makati city 4.650000
3 pasig city 4.633333
4 mandaluyong city 4.625000
5 beechworth 4.600000
6 london 4.535000
7 taguig city 4.525000
8 lincoln 4.500000
9 tagaytay city 4.500000
```

Summary:

- Top 5 cities with the highest restaurants are New Delhi, Gurgaon, Noida, Faridabad, and Ghaziabad. New Delhi alone contributes around 57.35% of the total data.
- Top 10 highest avg. rating cities are Inner City (4.90), Quezon City (4.80), Makati City (4.65), and others with ratings above 4.50, showing strong customer satisfaction.

Task 3: - Price Range Distribution

- Create a histogram or bar chart to visualize the distribution of price ranges among the restaurants.



- Calculate the percentage of restaurants in each price range category.

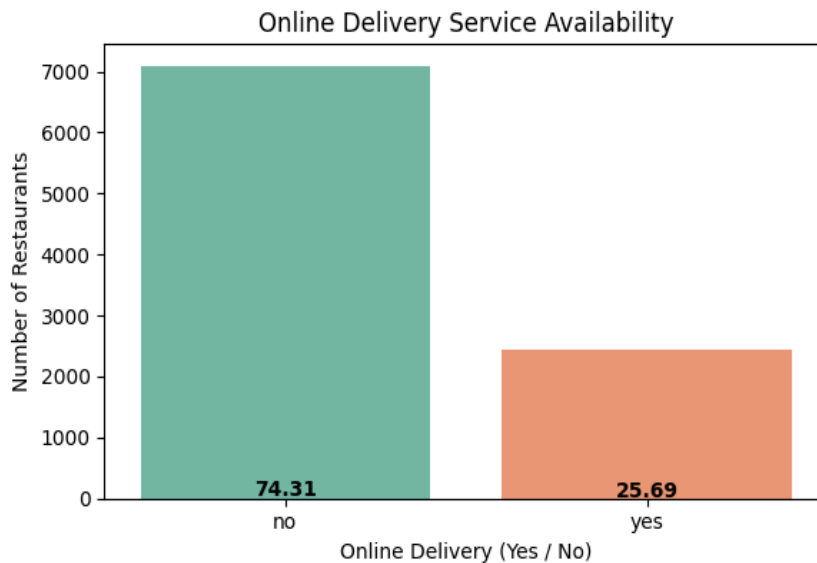
	Price range	Total restaurants	Percentage_restaurants
0	1	4438	46.51
1	2	3113	32.62
2	3	1405	14.72
3	4	586	6.14

Summary: -

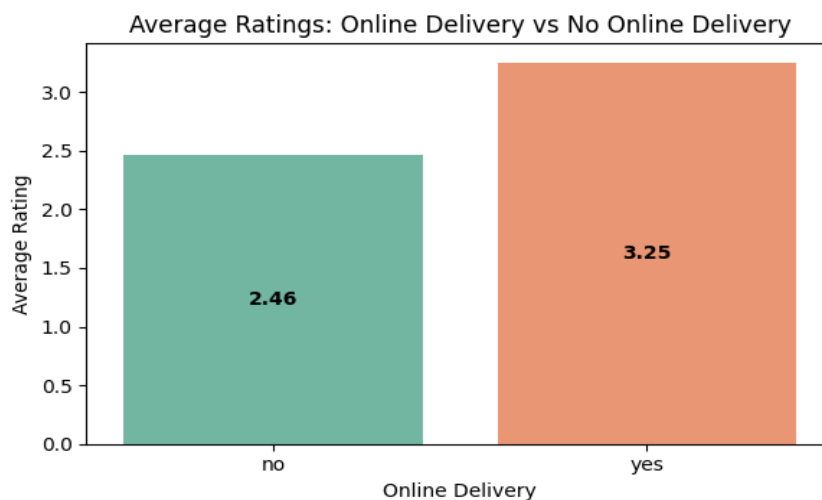
- Price range distributions low to high. (1 = low and 4 = high)
- There are Total **4438** restaurants in the Price range **1** which hold the total **46.51%** of market.
- in Price range 4 there are only **586** restaurants which are high Price range restaurants, which hold only **6.14%** market.
- it shows that around ****75%**** restaurants (price range 1 and 2) affordable.

Task 4: - Online Delivery

- Determine the percentage of restaurants that offer online delivery.



- Compare the average ratings of restaurants with and without online delivery.



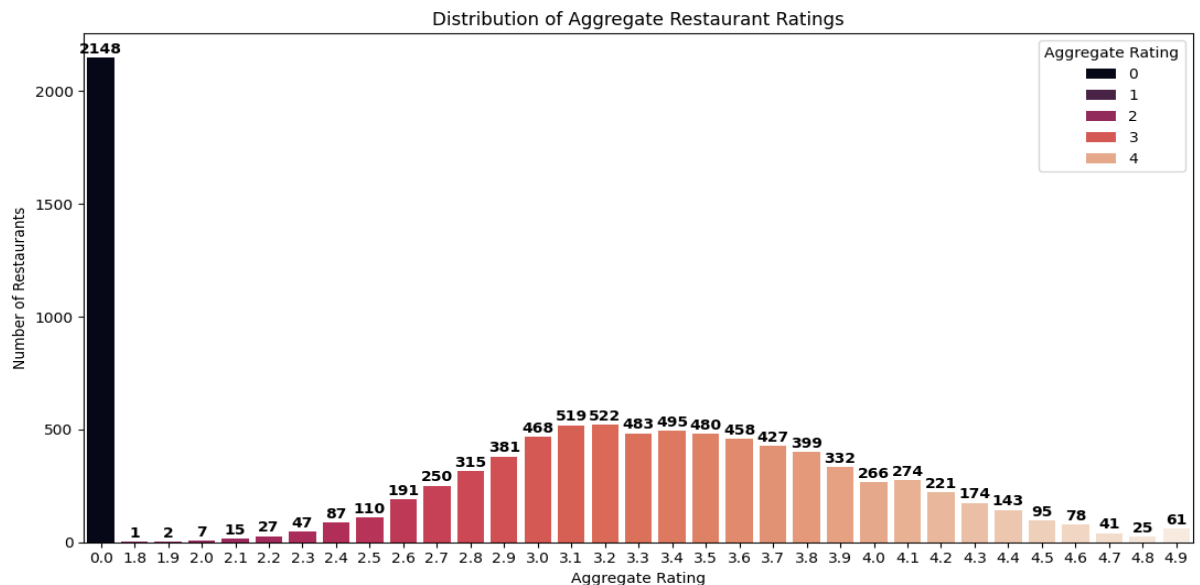
Summary: -

- The **74.31%** restaurant does not give online delivery.
- Only **25.69%** restaurant support online delivery.
- also, we found that the avg. rating of the **restaurant which support online delivery is 3.25**. Whereas which **restaurant does not support online delivery their avg. rating is 2.46**.

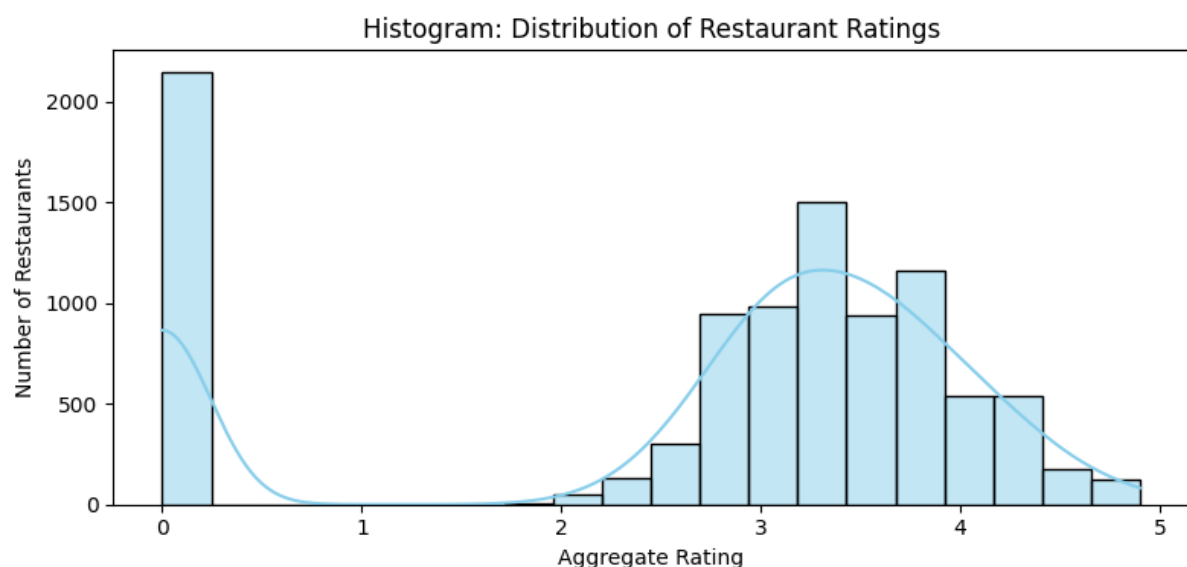
Level 2: - Strategic Insights on Ratings, Cuisines, Geography, and Chains

Task 1: - Restaurant Ratings

- a) Analyze the distribution of aggregate ratings and determine the most common rating range.



- Many restaurants (2,148) have a rating of 0.0, which likely means they were not rated at all, not poorly rated.
- The most common actual rating among rated restaurants is 3.2, followed closely by 3.1 and 3.3.
- Ratings between 2.5 and 4.0 cover most of the dataset — showing that most restaurants have moderate to good quality.



- There's a huge spike at **0.0**(not rated).
- Excluding 0.0, the data forms a **normal distribution** peaking around **3.0–3.5**, indicating that most customers give average to slightly above-average ratings.
- Very few restaurants achieve ratings above ****4.5****, making high ratings rare and prestigious.

b) Calculate the average number of votes received by restaurants.

➔ avg. votes of restaurant are around = **156.77**

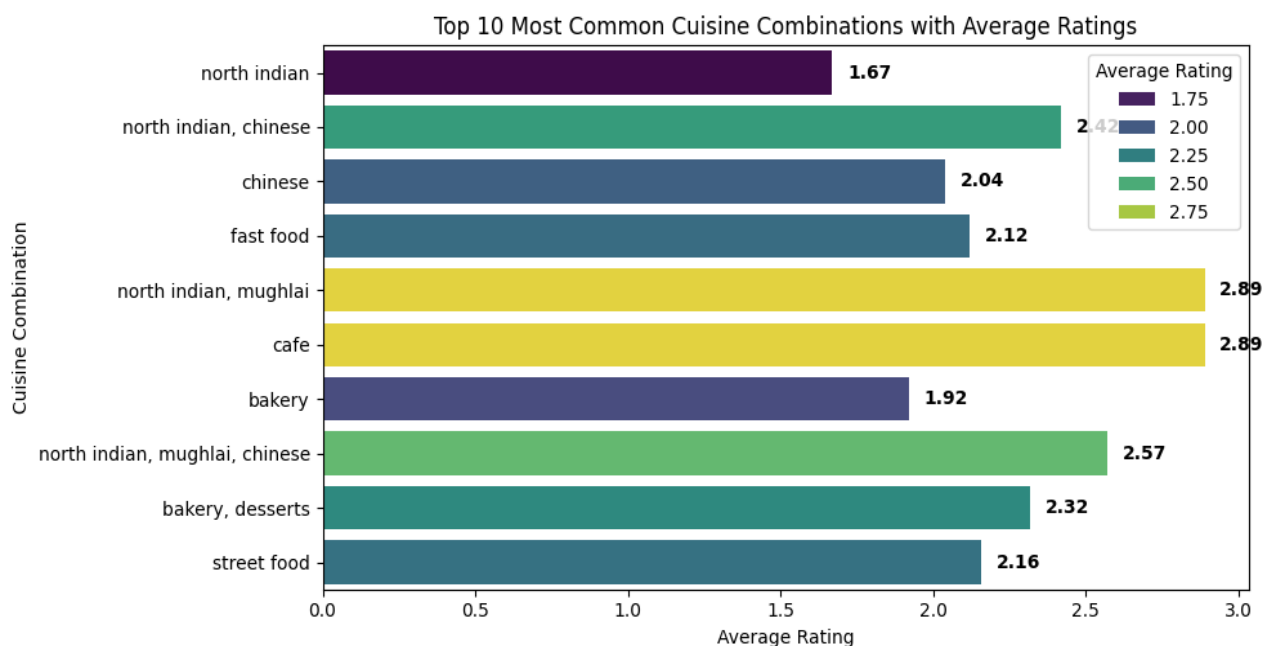
- This shows that, on average, restaurants have a moderate level of customer interaction through voting.
- Some highly rated or popular restaurants may have thousands of votes, while many unrated or new listings may have zero.

Task 2: - Cuisine Combination

- Identify the most common combinations of cuisines in the dataset.

Cuisines	
north indian	936
north indian, chinese	511
chinese	354
fast food	354
north indian, mughlai	334
cafe	299
bakery	218
north indian, mughlai, chinese	197
bakery, desserts	170
street food	149

- Determine if certain cuisine combinations tend to have higher ratings.

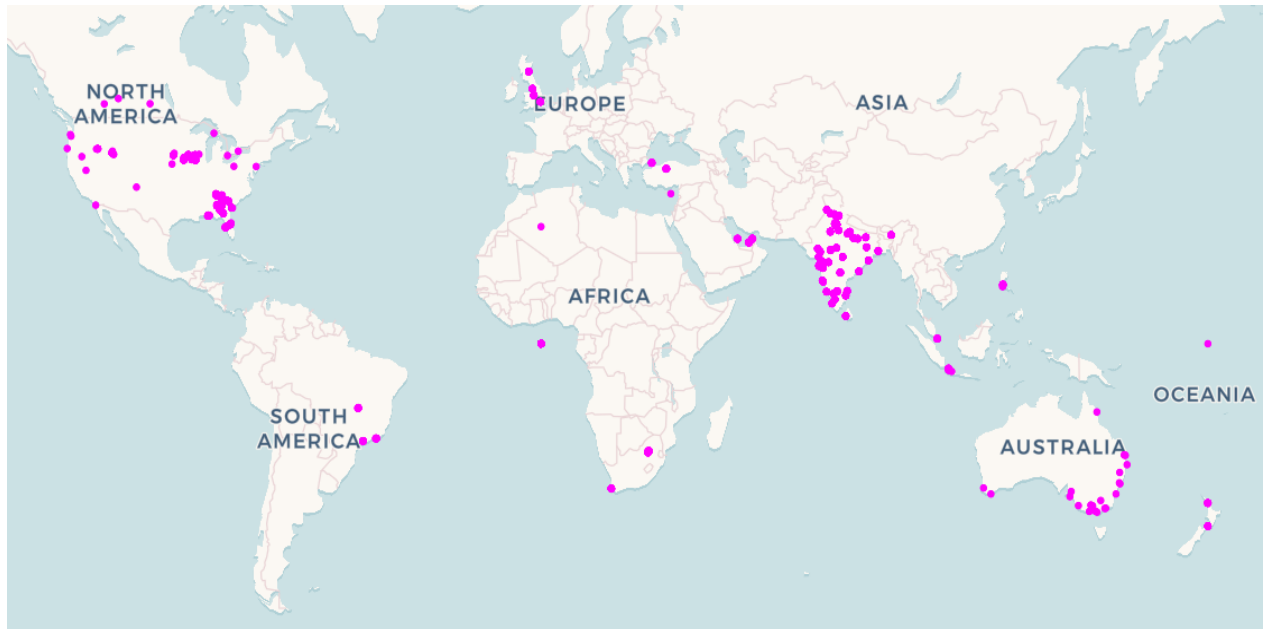


Summary: -

- The best cuisine combination based on customer ratings is North Indian, Mughlai and Cafe, both with an average rating of 2.89.
- This shows that customers prefer multi-cuisine combinations involving Mughlai dishes or cafe-style offerings.

Task 3: - Geographic Analysis

- Plot the locations of restaurants on a map using longitude and latitude coordinates.
- Identify any patterns or clusters of restaurants in specific areas.



LINK: - [Restaurant Map.html](#) (open in browser)

Summary: -

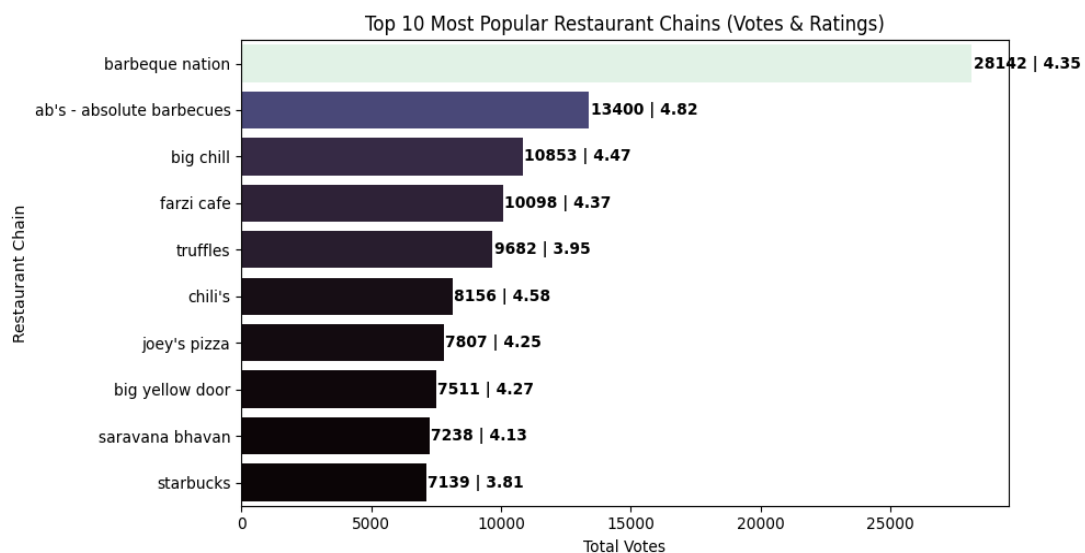
- India has the highest restaurant concentration, particularly in metropolitan cities like New Delhi, Mumbai, Bengaluru, and Chennai.
- Significant clusters are also visible in:
 - Australia (Sydney, Melbourne)
 - United States (New York, Los Angeles, Texas region)
 - Europe (London, some parts of Germany and France)
- Smaller restaurant clusters in Southeast Asia, Middle East, South Africa, and South America.

Task 4: - Restaurant Chains

- Identify if there are any restaurant chains present in the dataset.

Restaurant Name	
cafe coffee day	83
domino's pizza	79
subway	63
green chick chop	51
mcdonald's	48
keventers	34
pizza hut	30
giani	29
baskin robbins	28
barbeque nation	26

- Analyze the ratings and popularity of different restaurant chains.



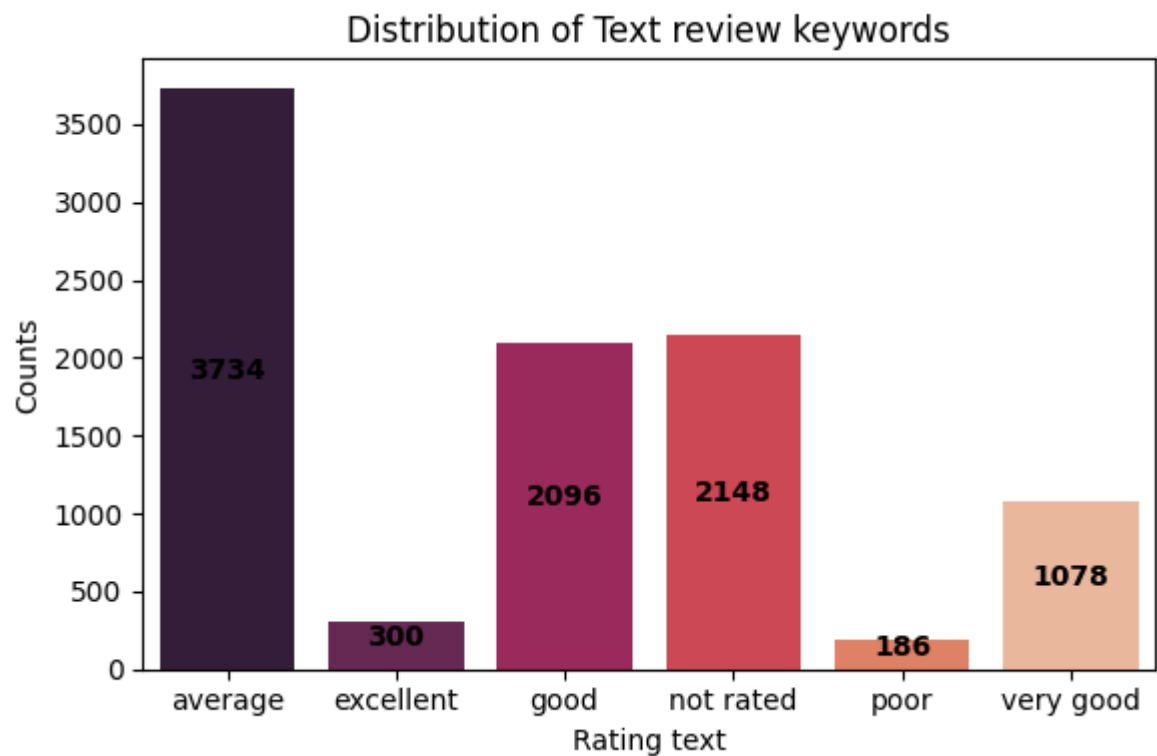
Summary: -

- **Barbeque Nation** dominates the restaurant chains with a total of **28,142 votes**, making it the most popular chain by a large margin.
- **AB's - Absolute Barbecues** holds the second position with **13,400 votes**, showing strong competition in the premium dining segment.
- Casual dining chains like **Big Chill (10,853)** and **Farzi Cafe (10,098)** are also among the top 5, reflecting their customer loyalty and consistent service quality.
- Chains like **Truffles, Chili's, Joey's Pizza**, and **Starbucks** show substantial popularity with votes ranging between **7,000 to 9,000**, indicating steady customer engagement.
- Most of the top chains fall under categories like **barbecue, cafes, casual dining, and pizza chains**, reflecting popular dining preferences.

Level 3: - Customer Engagement, Service Strategy, and Behavioural Insights

Task 1: - Restaurant Reviews

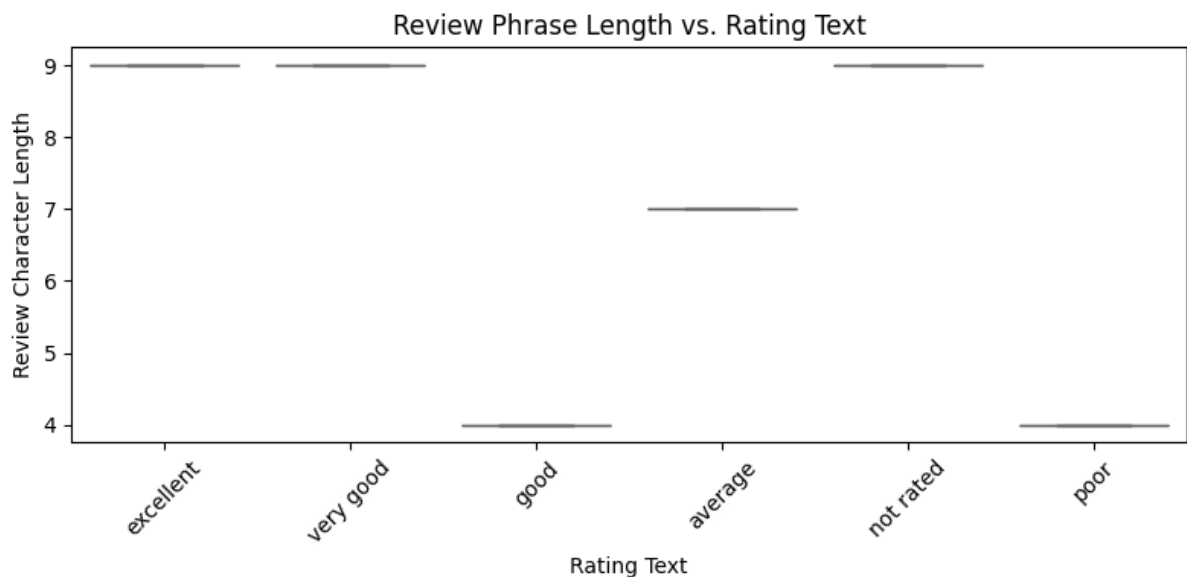
- i. Analyze the text reviews to identify the most common positive and negative keywords.



- The most common positive keywords are **Good** and **Very Good**, based on the Rating text column.
- This indicates that customers are generally satisfied with:
 - The **taste** of the food, The **service quality**, and The overall **restaurant environment**.
- The most common negative keywords are **poor** and **not rated**.
 - The **taste** of the food, **service quality** and the overall **restaurant environment**.
 - also, many restaurants are **not rated**, it means they open newly and not too many customers are visited the restaurants.
- The most common review category is the **average**, which indicating that the Customers experience is not too good also not too bad.

- ii. Calculate the average length of reviews and explore if there is a relationship between review length and rating.

➤ Average review phrase length: **7.02** characters



- The average length of review phrases (Rating Text) is approximately **7 characters**.
- Longer review phrases like **Very Good** and **Excellent** slightly increase the average length.
- However, there is no meaningful relationship between review character length and rating sentiment, since review categories are fixed and predefined.
- In this dataset, review text length doesn't indicate stronger or weaker opinions, as customers choose from fixed review labels rather than writing actual comments.

Task 2: - Votes Analysis

- Identify the restaurants with the highest and lowest number of votes.

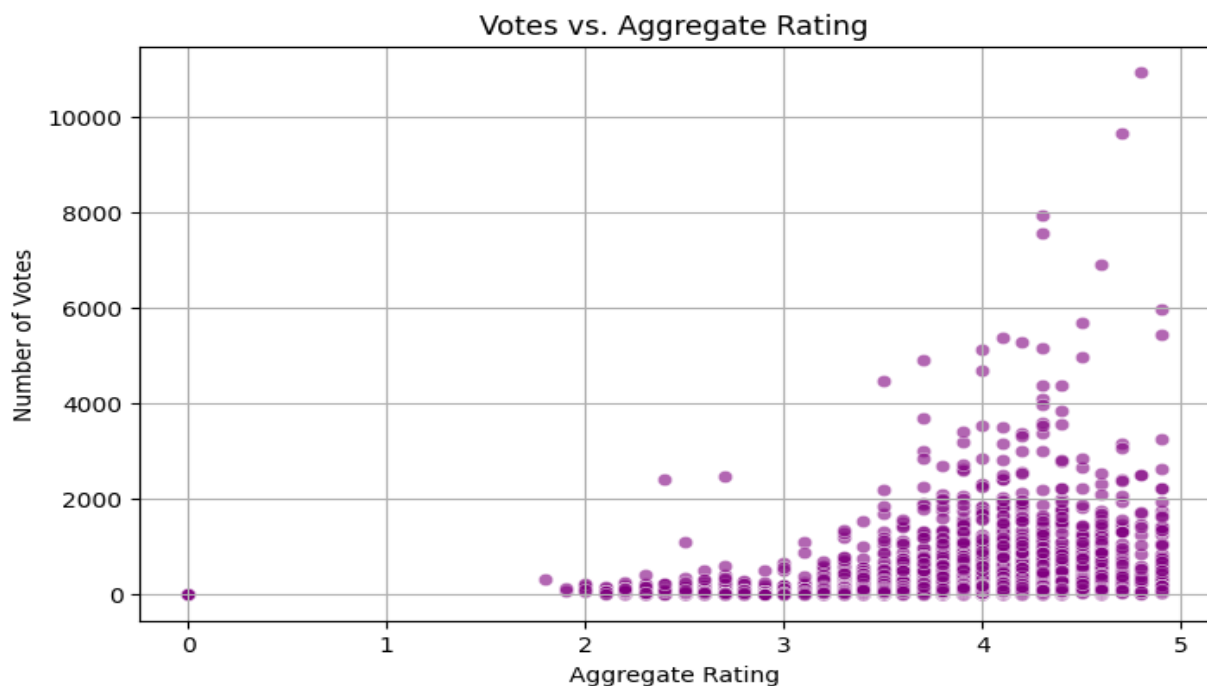
→ Highest number of voted restaurant: -

	Restaurant Name	Votes
719	toit	10934
726	truffles	9667
3985	hauz khas social	7931
2403	peter cat	7574
730	ab's - absolute barbecues	6907

→ Lowest number of voted restaurant: -

	Restaurant Name	Votes
1171	lavi foji dhaba	0
1169	gopi sweets & caters	0
1164	aha bites	0
1163	achoos food corner	0
9089	cafe' wow	0

- Analyze if there is a correlation between the number of votes and the rating of a restaurant.



- There is a **positive relationship** between a restaurant's rating and its number of votes:
- Higher-rated restaurants generally receive more customer votes.
- Poorly-rated or unrated restaurants tend to receive minimal engagement.
- However, the relationship is not perfectly linear — certain high-rated restaurants achieve exceptionally high votes, possibly due to brand recognition or marketing efforts.

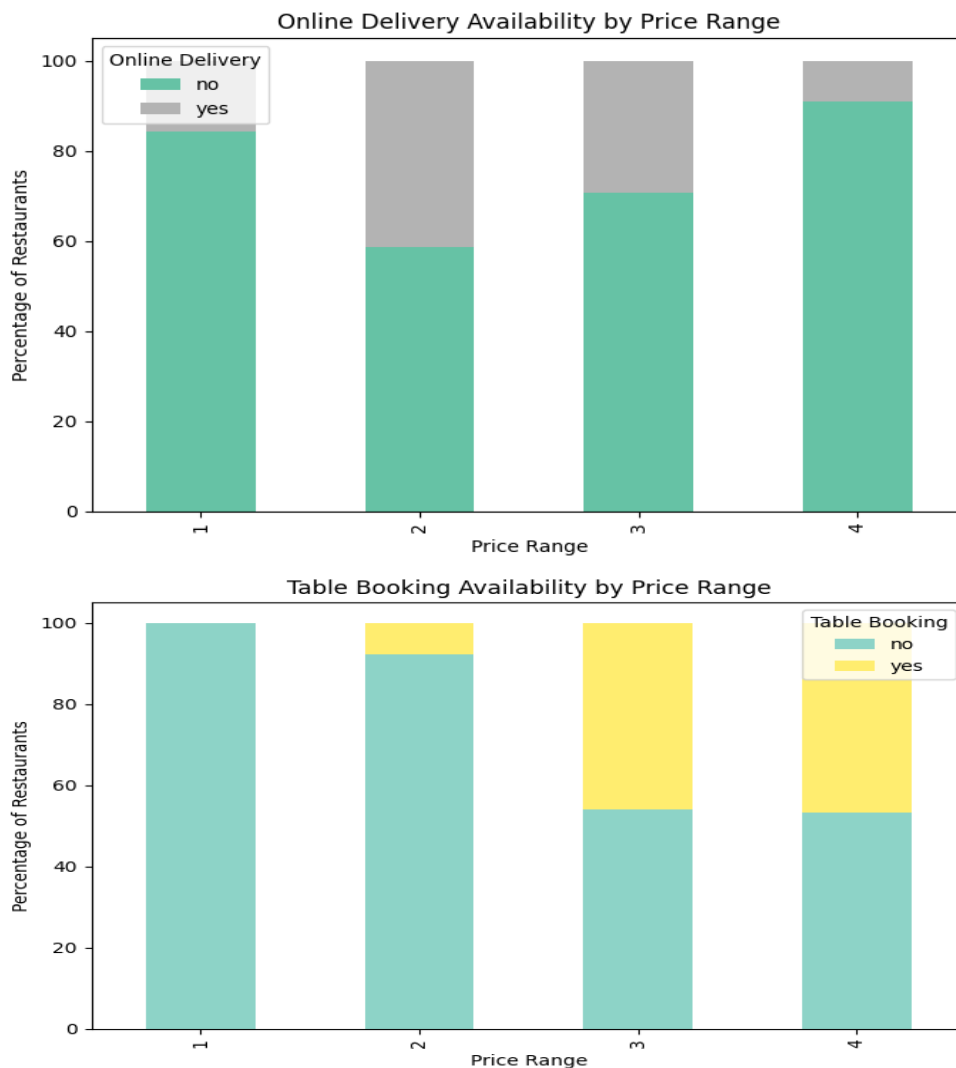
This analysis highlights that **customer satisfaction**(ratings) and **customer engagement**(votes) are related but influenced by other factors like visibility, brand strength, and customer base.

❖ **Correlation between votes and rating: 0.31**

- Restaurants with higher ratings generally attract more votes, but the relationship is weak to moderate.
- Correlation: 0.31 → Ratings help, but votes depend on brand, location, and visibility.

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

- 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.
- > Higher-priced restaurants prioritize table booking, focusing on exclusive dine-in experiences.
- > Mid-range restaurants lead in online delivery, balancing affordability with convenience.
- > Premium pricing signals dine-in service priority, while affordable restaurants focus on delivery growth.

Business Insights & Recommendations

Key Business Insights

- **Cuisine Preferences:**
Customers mainly prefer North Indian, Chinese, and Fast Food, covering over 90% of offerings. Mughlai and café-style combinations are also popular.
- **City Concentration:**
Restaurants are concentrated in metro cities, with New Delhi alone contributing over 57%.
- **Price Segmentation:**
Around 75% of restaurants fall into Price Range 1 & 2, showing a market dominated by affordable and mid-range restaurants.
- **Service Distribution:**
 - Online delivery is common in mid-range restaurants.
 - Table booking is mostly offered by higher-priced restaurants.
- **Ratings & Engagement:**
 - Most restaurants have average to good ratings (3.0–3.5).
 - Votes correlate weakly (0.31) with ratings.
 - Brand visibility and location drive customer engagement.
- **Restaurant Chains:**
Chains like Barbeque Nation, Domino's, and Starbucks lead in customer loyalty and consistent ratings.
- **Geographic Patterns:**
Restaurants cluster in urban hubs, visible via heatmaps.
- **Review Analysis:**
Positive keywords like "good" and "very good" reflect general customer satisfaction.

Strategic Recommendations

1. Promote top cuisines and offer multi-cuisine options to capture wider customer segments.
2. Focus online delivery on mid-range restaurants and promote table booking in premium restaurants.
3. Expand into underrepresented cities to avoid metro saturation.
4. Support smaller brands with branding, loyalty programs, and marketing.
5. Boost customer interaction via reviews, feedback programs, and promotions.
6. Use location-based marketing in metro clusters and plan expansion using geographic analysis.

Conclusion

This project analysed restaurant industry data using Python to uncover actionable insights across services, pricing, customer engagement, and geographic distribution. Key findings highlight that customer preferences focus heavily on popular cuisines like North Indian and Chinese, while service offerings differ based on restaurant pricing levels. Mid-range restaurants emphasize online delivery, whereas premium restaurants prioritize table booking for exclusive dining experiences.

Restaurant chains and metro cities dominate the market, but expansion potential exists in underrepresented regions. Customer engagement, while positively linked to ratings, also depends significantly on branding and visibility.

Data-driven insights from this analysis support strategic decisions in:

- Menu design and cuisine offerings,
- Service strategy based on pricing,
- Marketing efforts focused on customer engagement,
- Expansion planning using geographic trends.

Overall, this analysis enables restaurant businesses to optimize operations, enhance customer satisfaction, and develop competitive market strategies backed by data.