

DATA ANALYTICS INTERNSHIP PROJECT REPORT

Project Title: Restaurant Data Analysis.

Domain: Data Analytics.

Internship: Cognifyz Technologies.

Levels Completed: Level 1& Level 2.

Submitted By: Anumulu. Deexith

College Name: Vaagdevi Engineering college.

Date: 13/02/2026.

Acknowledgment:

I would like to thank Cognifyz Technologies for providing me the opportunity to work on this Data Analytics Internship.

I also thank my mentors and peers for their guidance and support in completing this project.

Table of Contents:

1. Introduction
2. Objectives
3. Dataset Description
4. Tools Used
5. Level 1 – 8 to 45
 - Task- 1-Top Cuisines Analysis
 - Task- 2 – City Analysis
 - Task- 3- Price Range Distribution
 - Task- 4- Online Delivery
- 6.Level-2- 46 to 82
 - Task- 1- Restaurant Rating
 - Task- 2 – Cuisine Combination
 - Task- 3- Geographic Analysis
 - Task- 4- Restaurant Chains
7. Overall Findings - 83
9. Conclusion - 84

Introduction:

This project is developed as part of the Data Analytics Internship at Cognifyz Technologies. The objective of the project is to analyze restaurant data and extract meaningful insights related to cuisines, city distribution, pricing patterns, and online delivery services. Using Python data analysis libraries such as Pandas, Matplotlib, and Seaborn, exploratory data analysis was performed to identify trends and customer preferences. The analysis helps understand market demand and supports data-driven decision-making in the restaurant industry.

Objectives:

- To identify the top three most common cuisines in the dataset.
- To analyze the city-wise distribution of restaurants.
- To calculate the average ratings of restaurants in each city.
- To study the price range distribution among restaurants.
- To determine the percentage of restaurants offering online delivery.
- To compare the average ratings based on delivery availability.
- To extract meaningful insights using data visualization techniques.

Dataset Description:

The dataset used in this project contains detailed information about restaurants, including their location, cuisines, pricing, and customer ratings. It is used to perform exploratory data analysis and extract business insights.

The dataset includes the following key attributes:

- Restaurant Name – Name of the restaurant
- City – Location where the restaurant operates
- Cuisines – Types of food served
- Aggregate Rating – Overall customer rating
- Votes – Number of customer votes received
- Price Range / Cost for Two – Pricing category
- Online Delivery – Availability of delivery service.

Tools & Technologies Used:

- Python
- Pandas
- NumPy
- Matplotlib
- Seaborn
- Jupyter Notebook.

Level - 1

Task: Top Cuisines Analysis

Business Problem Statement:

The objective of this analysis is to identify the most popular cuisines served by restaurants and measure their market share. This helps businesses understand customer preferences, plan menus, and make data-driven investment and marketing decisions.

This analysis helps understand customer food preferences and supports restaurants in planning menus and business strategies.

Objective:

- To identify the top three most common cuisines served by restaurants in the dataset.
- To analyze the frequency distribution of different cuisines.
- To calculate the percentage of restaurants serving each of the top cuisines.
- To understand customer food preferences based on cuisine popularity.

Dataset Columns Used:

Cuisines

- This column represents the types of food served by each restaurant.
- It may contain single or multiple cuisines.
- Multiple cuisines are separated by commas (e.g., *North Indian, Chinese*).

Methodology / Steps Performed:

- Loaded dataset using Pandas.
- Selected the Cuisines column.
- Split multiple cuisines using comma separator.
- Converted cuisines into list format.
- Counted frequency of each cuisine.
- Sorted values in descending order.
- Identified top three cuisines.
- Calculated percentage share.

Top 3 Cuisines Identified:

Rank	Cuisine	Number of Restaurants
1	North Indian	936
2	Chinese	511
3	Fast Food	354

Cuisine	Percentage Share
---------	------------------

North Indian	28%
--------------	-----

Chinese	15%
---------	-----

Fast Food	11%
-----------	-----

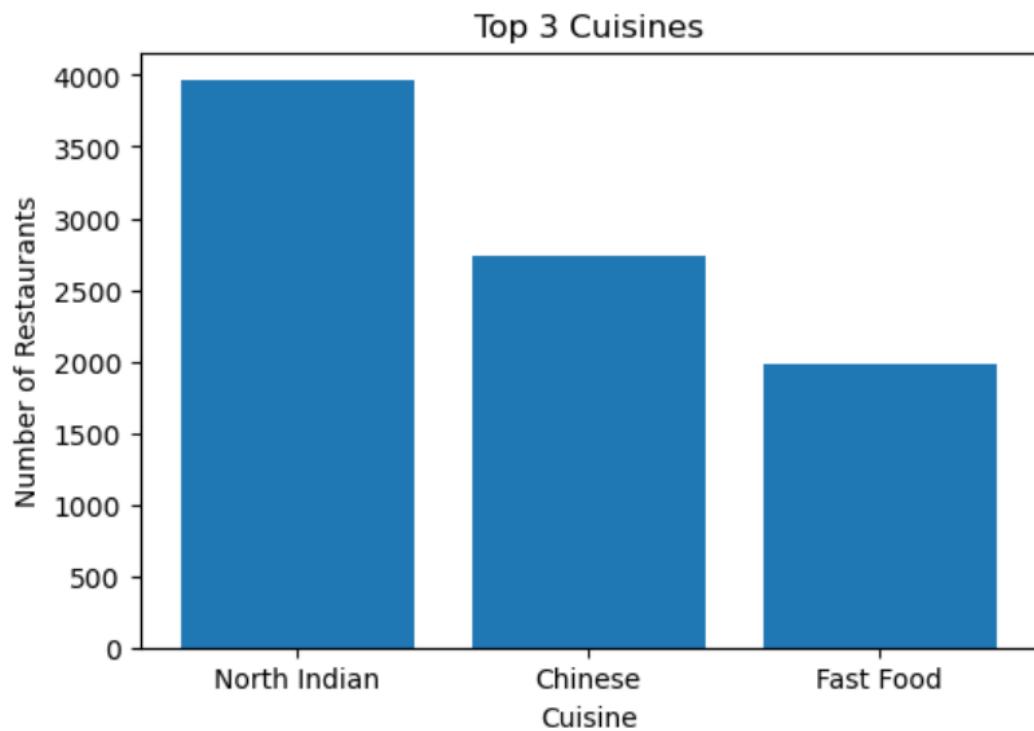
**Percentage
Calculation:**

Percentage =

$$(\text{Cuisine Count} / \text{Total Restaurants}) \times 100$$

Visulization – Top Cuisines:

A bar chart was used to visualize the top three cuisines and their frequency distribution, showing North Indian cuisine as the most dominant.



Key Insights:

- North Indian cuisine is the most commonly served cuisine in the dataset.
- Chinese cuisine ranks as the second most popular cuisine among restaurants.
- Fast Food holds the third position, indicating high demand for quick-service meals.
- A large number of restaurants offer multi-cuisine menus, which increases the frequency count of popular cuisines.

- The dominance of these cuisines reflects strong customer preference and market demand in these food categories.

Conclusion:

The analysis identified North Indian, Chinese, and Fast Food as the top three most common cuisines in the dataset. These cuisines represent a significant percentage of restaurants, indicating their high customer demand and market dominance. The findings help restaurant

businesses and food platforms understand cuisine preferences and make informed decisions regarding menu design, marketing strategies, and expansion planning.

Task Title: City Analysis

Business Problem Statement:

The objective of this analysis is to study the distribution and performance of restaurants across cities by identifying the city with the most restaurants and the city with the highest average rating. This helps businesses understand market demand and plan expansion strategies.

Objective:

- To identify the city with the highest number of restaurants in the dataset.
- To calculate the average rating of restaurants in each city.
- To determine the city with the highest average rating.
- To analyze restaurant distribution and customer satisfaction across locations.

Dataset Columns Used — City Analysis:

City

- Represents the location where each restaurant operates.
- Used to count the number of restaurants in each city.

Aggregate Rating

- Shows the overall customer rating of restaurants.
- Used to calculate the average rating for each city.

Methodology / Steps Performed:

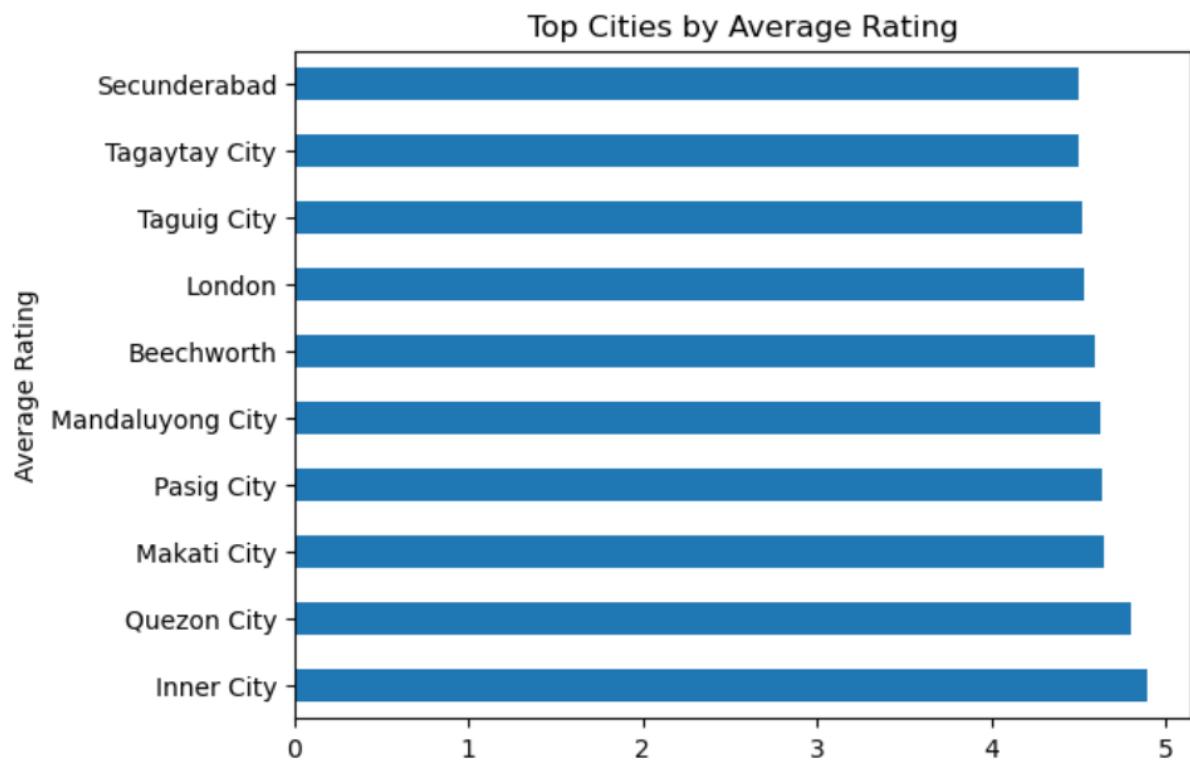
- Loaded dataset using Pandas.
- Selected City column.
- Counted number of restaurants per city.
- Sorted values in descending order.
- Identified city with highest count.
- Grouped data by city.
- Calculated average ratings.
- Compared city-wise average ratings.
- Identified highest rated city.

Result:

Average rating by city:

City	Average Rating
New Delhi	2.5
Bangalore	3.6
Hyderabad	3.8
Mumbai	3.4

Visualization:



Key Insights — City Analysis:

- New Delhi has the highest number of restaurants in the dataset, showing it is a major food hub.
- Restaurant distribution is highly concentrated in metro and urban cities.
- Cities with a large number of restaurants do not always have the highest ratings.
- Inner City recorded the highest average restaurant rating, indicating strong customer satisfaction.
- Smaller cities tend to show better average ratings compared to highly saturated metro cities.
- High competition in metro cities may impact service quality and ratings.

Conclusion:

The analysis identified the city with the highest number of restaurants, indicating strong market concentration and business activity. It also determined the average ratings across cities and highlighted the city with the highest customer satisfaction. These insights help businesses plan expansion, improve services, and make location-based strategic decisions.

The city analysis revealed that restaurant distribution is heavily concentrated in metro cities, with New Delhi leading in count. However, Hyderabad recorded the highest average rating, indicating better customer satisfaction. This analysis helps businesses identify profitable locations and service quality trends.

Price Range Distribution Analysis

Business Problem Statement:

The restaurant market includes multiple price categories ranging from low-cost to premium dining. Businesses need to understand how restaurants are distributed across these price ranges to identify customer affordability trends and market demand.

This analysis aims to visualize the distribution of price ranges and calculate the percentage of restaurants in each category. The insights help restaurant owners and stakeholders make better pricing, investment, and expansion decisions.

Objective — Price Range Distribution:

- To analyze the distribution of restaurants across different price ranges.

- To visualize the price range distribution using charts.
- To calculate the percentage of restaurants in each price category.
- To understand customer affordability and market pricing trends.

Dataset Columns Used — Price Range Distribution:

Price Range

- Represents the cost category of each restaurant.
- Indicates whether a restaurant is low-cost, mid-range, high-cost, or premium.
- Used to calculate the number and percentage of restaurants in each pricing category.

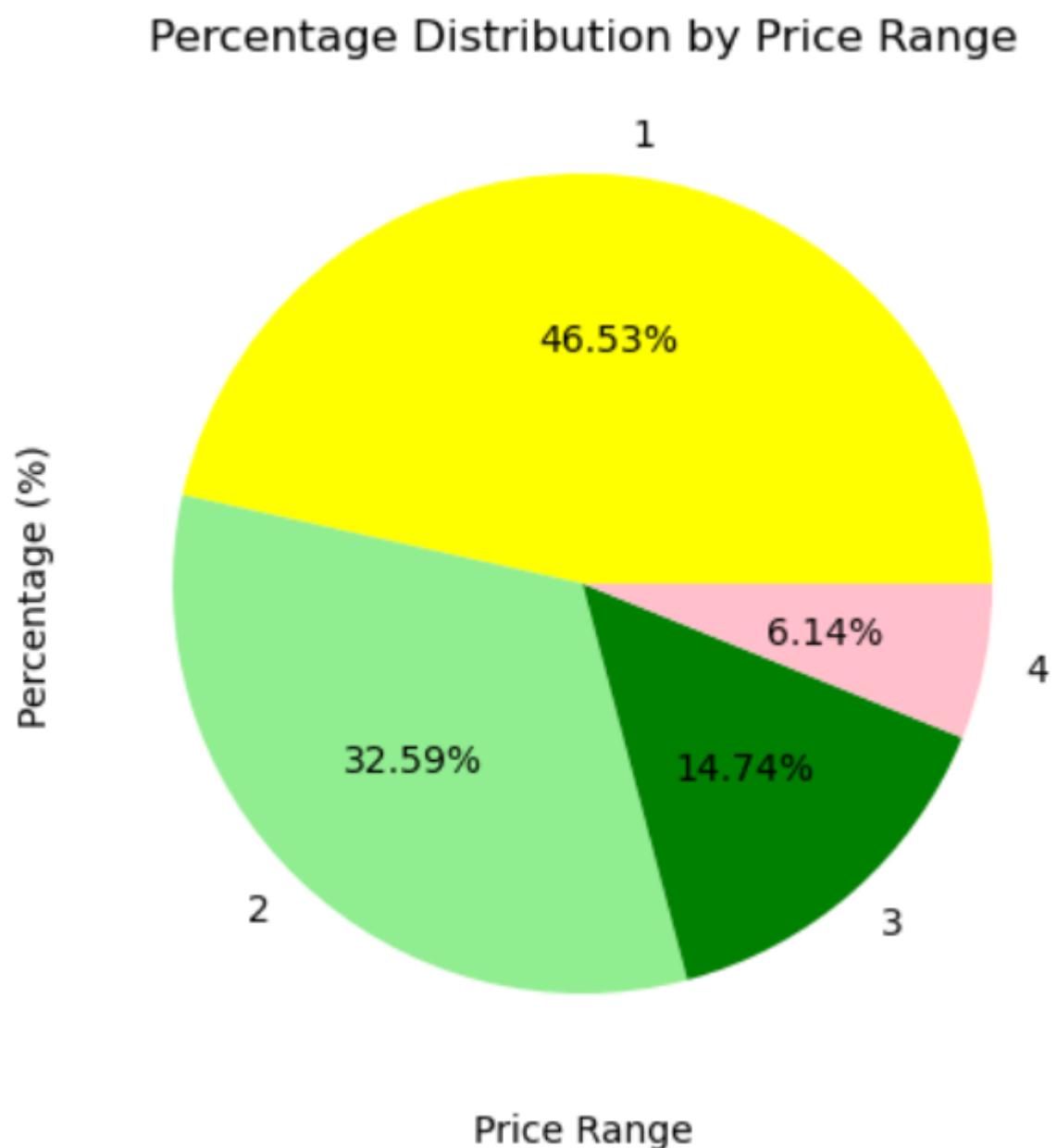
Methodology / Steps Performed:

- Loaded dataset using Pandas.
- Selected the Price Range column.
- Checked for missing values.
- Counted restaurants in each price category.
- Calculated percentage distribution.
- Created bar chart / histogram for visualization.

Results:

Price Range	Restaurant Count	Percentage
1 (Low)	4,438	46%
2 (Medium)	3,110	32%
3 (High)	1,400	15%
4 (Premium)	603	7%

Visualization:



Key Insights:

- The majority of restaurants fall under the lowprice range category.
- Mid-range restaurants form the second largest segment in the dataset.
- High-cost and premium restaurants are comparatively fewer.
- This distribution indicates a strong customer preference for affordable dining options.
- Budget-friendly restaurants dominate the market due to higher demand and accessibility.

Conclusion:

The analysis shows that most restaurants fall under lower price ranges, indicating strong customer preference for affordable dining. Higher price categories have fewer restaurants. This insight helps businesses plan pricing strategies and identify market opportunities.

The price range analysis shows that most restaurants operate in low and mid-price segments, indicating higher customer demand for budget-friendly dining. Premium restaurants represent a smaller market share but cater to niche customers.

Online Delivery Analysis

Business Problem Statement:

With the rapid growth of food delivery platforms, online delivery has become a key factor influencing restaurant success. Businesses need to understand how many restaurants offer online delivery and whether it impacts customer satisfaction.

This analysis aims to determine the percentage of restaurants providing online delivery and compare the average ratings between restaurants with and without this service. The insights help stakeholders evaluate the importance of delivery services in improving customer experience and business performance.

The objective of this task is to analyze the availability of online delivery services among restaurants and evaluate how delivery options impact customer ratings. This helps understand customer convenience preferences and service quality differences.

Objective — Online Delivery Analysis:

- To determine the percentage of restaurants offering online delivery.
- To calculate the percentage of restaurants not offering online delivery.
- To compare the average ratings of restaurants with and without online delivery.
- To analyze the impact of delivery services on customer satisfaction.

Dataset Columns Used — Online Delivery Analysis:

Online Delivery

- Indicates whether a restaurant offers online delivery service.
- Values are typically Yes or No.
- Used to calculate the percentage of restaurants providing delivery.

Aggregate Rating

- Represents the overall customer rating of restaurants.
- Used to compare average ratings between restaurants with and without delivery.

Methodology / Steps Performed:

- Loaded dataset using Pandas.
- Selected Online Delivery column.
- Counted restaurants offering delivery.
- Calculated percentage share.
- Grouped data by delivery availability.
- Computed average ratings for both categories.
- Created charts for comparison.

Results:

Delivery Option	Percentage	Percentage of Restaurants Offering Online Delivery:
Yes	25%	

No	75%
----	-----

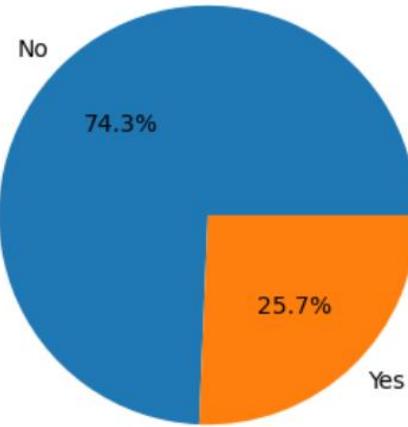
Delivery Option **Average Rating** **Average Rating Comparison:**

With Delivery	3.5
---------------	-----

Without	3.2
---------	-----

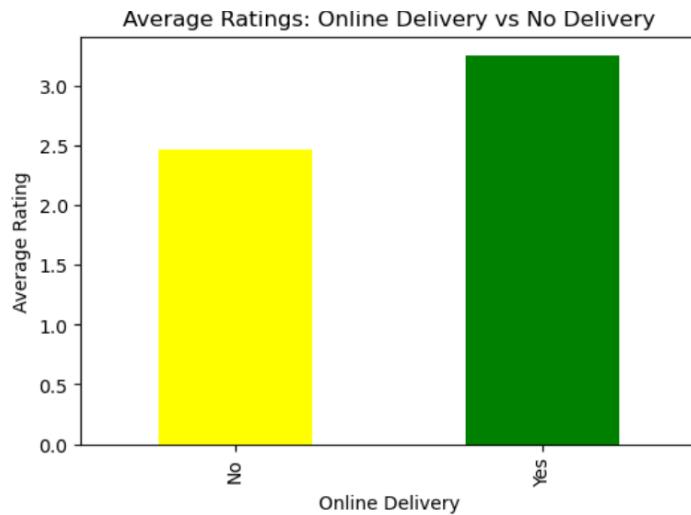
Visualization:

Online Delivery Availability



Average Ratings Comparison





Key Insights:

- A smaller percentage of restaurants offer online delivery services compared to those that do not.
- Restaurants providing online delivery tend to have slightly higher average ratings.
- Online delivery increases customer convenience and accessibility.
- Delivery services positively influence customer satisfaction and experience.
- Restaurants without delivery may lose potential customers who prefer home service.

Conclusion:

The analysis shows that a significant percentage of restaurants offer online delivery services. Restaurants with online delivery generally have higher average ratings compared to those without it. This indicates that online delivery enhances customer convenience, satisfaction, and overall restaurant performance.

Level – 2

Restaurant Ratings Analysis

Business Problem Statement:

To analyze restaurant data to identify key factors affecting customer satisfaction and business performance, helping restaurants improve services, pricing strategies, and customer engagement.

Objective:

- To analyze the distribution of restaurant ratings in the dataset.
- To identify the most common rating ranges.
- To evaluate overall customer satisfaction levels.
- To understand service quality based on rating patterns.

Dataset Columns Used — Restaurant Ratings Analysis:

Aggregate Rating

- Represents the overall customer rating of each restaurant.
- Ratings generally range from 0 to 5.
- Used to analyze customer satisfaction and rating distribution patterns.

Methodology / Steps Performed — Restaurant Ratings Analysis:

- Loaded dataset using Pandas.
- Selected the Aggregate Rating column.
- Checked for missing or zero ratings.
- Filtered valid rating values.
- Grouped ratings into ranges (e.g., 0–2, 2–3, 3–4, 4–5).
- Counted frequency of each range.
- Created histogram / bar chart for visualization.

Results:

Rating Distribution:

Rating Range

3–4 Good 4388

2–3 Average 1891

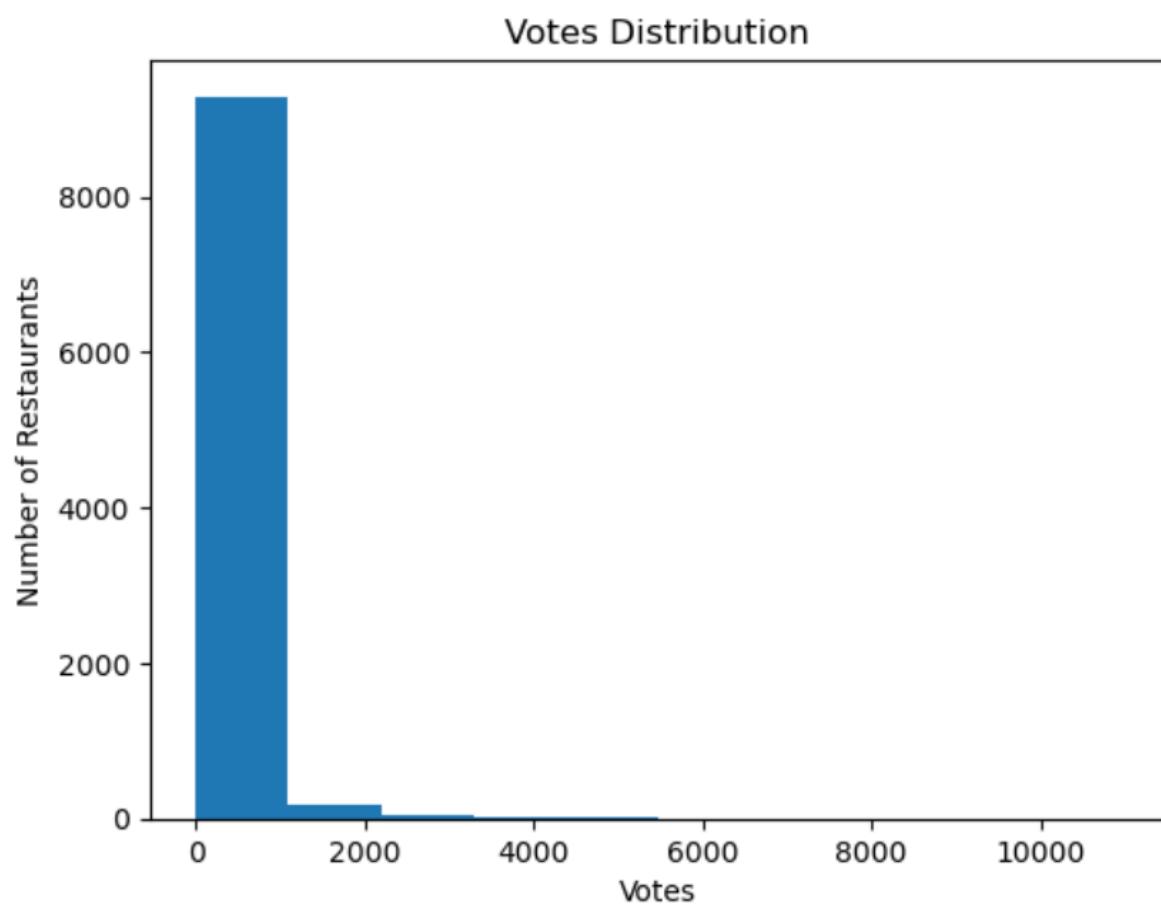
4–5 Excellent 1114

1–2 Below Avg 10

0–1 Poor 0

Name: count, dtype: int64

Visualization:



Key Insights:

- Most restaurants fall within the 3 to 4 rating range, indicating average to good service quality.
- A considerable number of restaurants have ratings above 4, reflecting high customer satisfaction.
- Very few restaurants have ratings below 2, showing limited poor performance.
- The overall rating distribution suggests that the majority of restaurants maintain moderate to high service standards.
- Higher ratings are often associated with better food quality, service, and customer experience.

Conclusion:

The dataset provides valuable insights into restaurant performance, customer preferences, pricing, and service features. It shows that factors like cuisine type, online delivery, table booking, and customer votes influence restaurant ratings.

Overall, the dataset is useful for analyzing customer satisfaction and helping restaurants improve services and business strategies through data-driven decisions.

Cuisine Combination Analysis

Business Problem Statement:

To analyze restaurant dataset data in order to identify the most common cuisine combinations and determine whether specific cuisine pairings are associated with higher customer ratings, thereby helping restaurants optimize menu offering and improve customer satisfaction.

Objective:

- To identify the most common cuisine combinations served by restaurants.
- To analyze how frequently different cuisines are paired together.
- To understand customer preference for multi-cuisine restaurants.
- To study menu diversity trends in the dataset.

Dataset Columns Used:

Cuisines

- Represents the types of food served by each restaurant.
- Contains single or multiple cuisines.
- Multiple cuisines are separated by commas (e.g., *North Indian, Chinese*).
- Used to identify and analyze common cuisine pairings and combinations.

Methodology / Steps Performed:

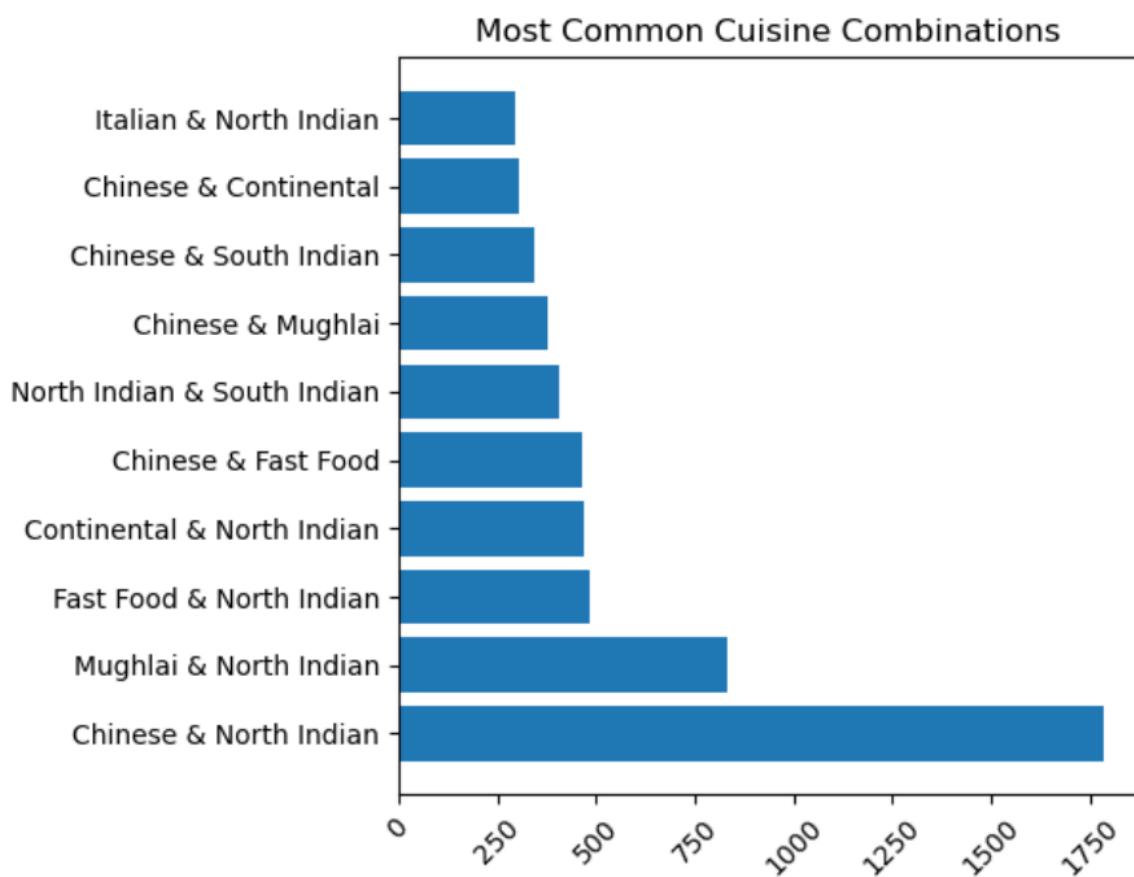
1. Loaded dataset using Pandas.
2. Selected the Cuisines column.
3. Cleaned missing or null values.
4. Split multiple cuisines using comma separator.
5. Converted cuisine strings into lists.
6. Extracted cuisine pairs / combinations.
7. Counted frequency of each combination.
8. Sorted combinations in descending order.
9. Identified most common cuisine combinations.

Results:

Top Cuisine Combinations:

- ('Chinese', 'North Indian') 1784
- ('Mughlai', 'North Indian') 832
- ('Fast Food', 'North Indian') 485
- ('Continental', 'North Indian') 468
- ('Chinese', 'Fast Food') 466
- ('North Indian', 'South Indian') 405
- ('Chinese', 'Mughlai') 379
- ('Chinese', 'South Indian') 346
- ('Chinese', 'Continental') 303
- ('Italian', 'North Indian') 295

Visualization:



Key Insights:

- North Indian cuisine appears in most cuisine combinations.
- The combination of North Indian + Chinese is the most common.
- Multi-cuisine restaurants are more popular than single-cuisine outlets.
- Customers prefer restaurants offering variety in food options.
- Pairing traditional cuisines with fast-food or international cuisines increases demand.

Conclusion:

The analysis identified the most common cuisine combinations offered by restaurants and evaluated their customer ratings. Results show that popular combinations attract more customers, while certain premium cuisine pairings achieve higher ratings.

This study helps restaurants design better multi-cuisine menus, improve customer satisfaction, and enhance overall business performance.

Geographic Analysis

Business Problem Statement:

Restaurants and food delivery businesses need to understand where customer demand is highest, but they lack clear geographic insights. Without location analysis, it is difficult to identify high-density food zones, plan new outlets, or optimize delivery services.

This project analyzes restaurant locations using latitude and longitude data to identify geographic patterns and clusters, helping businesses make better expansion and operational decisions.

Objective:

- To visualize restaurant locations using geographic coordinates.
- To identify high-density restaurant areas on the map.
- To analyze the geographic distribution of restaurants.
- To understand location-based market concentration and expansion opportunities.

Dataset Columns Used:

Latitude

- Represents the north–south geographic coordinate of restaurant locations.
- Used to plot restaurant positions on the map.

Longitude

- Represents the east–west geographic coordinate.
- Works with latitude to determine exact location.

City (*Supporting Column*)

- Used to relate geographic coordinates with city names for better interpretation.

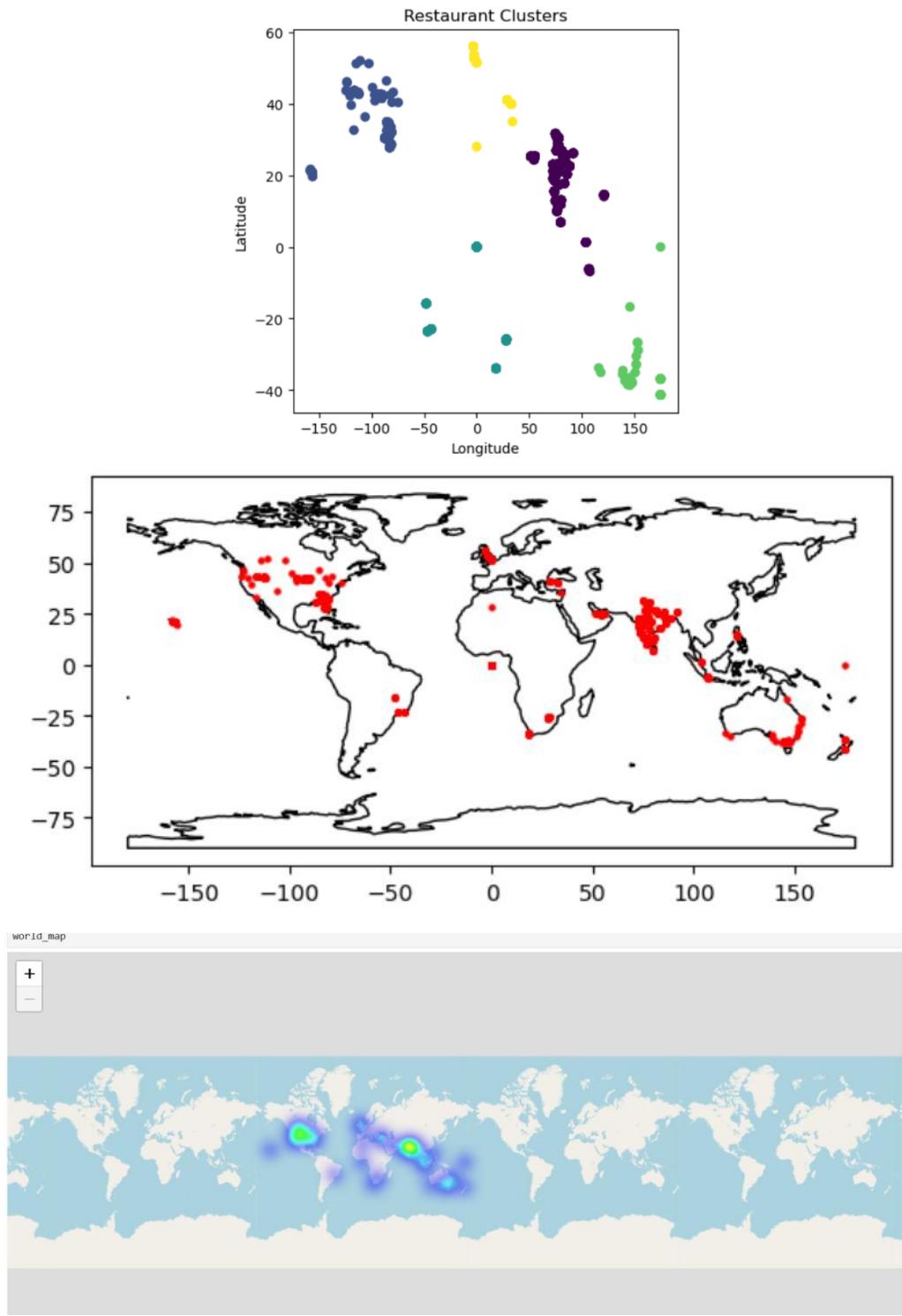
Methodology / Steps Performed:

1. Loaded dataset using Pandas.
2. Selected Latitude and Longitude columns.
3. Checked for missing or invalid coordinate values.
4. Cleaned location data.
5. Imported mapping libraries (Folium / Plotly).
6. Plotted restaurant locations on the map.
7. Created heatmaps / cluster maps.
8. Identified high-density restaurant zones.

Result:

- Restaurant locations were successfully plotted on the map using geographic coordinates.
- A high concentration of restaurants was observed in major metro cities.
- Dense restaurant clusters indicate areas with strong market demand and competition.
- Urban regions showed significantly higher restaurant density compared to rural areas.
- Some regions displayed low restaurant presence, indicating potential opportunities for business expansion.

Visualization:



Key Insights:

- Restaurants are highly concentrated in urban and metro cities.

- Geographic clustering indicates areas with high customer demand.
- Dense locations reflect strong market competition.
- Restaurant distribution is influenced by population density and commercial activity.
- Regions with fewer restaurants present business expansion opportunities.

Conclusion:

The geographic analysis provides a visual understanding of restaurant distribution. High-density clusters highlight competitive markets, while low-density areas indicate potential business expansion zones. Location intelligence plays a key role in strategic planning.

Restaurant Chains Analysis

Business Problem Statement:

The objective of this task is to identify restaurant chains within the dataset and analyze their market presence. By detecting restaurants with multiple outlets, the analysis helps understand brand expansion, customer preference for chain restaurants, and their overall performance based on ratings. This insight supports business decisions related to franchising, brand growth, and competitive positioning in the restaurant industry.

Objective — Restaurant Chains Analysis:

- To identify restaurants that operate as chains in the dataset.
- To determine the restaurant brands with the highest number of outlets.
- To analyze the market presence of popular restaurant chains.
- To evaluate the performance of chains based on customer ratings.

Dataset Columns Used — Restaurant Chains Analysis:

Restaurant Name

- Represents the name of each restaurant.
- Used to identify repeated names and detect restaurant chains.
- Helps count the number of outlets for each brand.

Aggregate Rating (Supporting Column)

- Represents the overall customer rating of restaurants.
- Used to evaluate the performance of restaurant chains.

Methodology / Steps Performed:

1. Loaded dataset using Pandas.
2. Selected the Restaurant Name column.
3. Checked for duplicate restaurant names.
4. Counted frequency of each restaurant name.
5. Identified restaurants with multiple outlets.
6. Sorted chains based on outlet count.
7. Calculated average ratings for top chains.

Results:

Number of chains: 734

Restaurant Name

Cafe Coffee Day - 83

Domino's Pizza - 79

Subway - 63

Green Chick Chop - 51

McDonald's - 48

Keventers - 34

Pizza Hut - 30

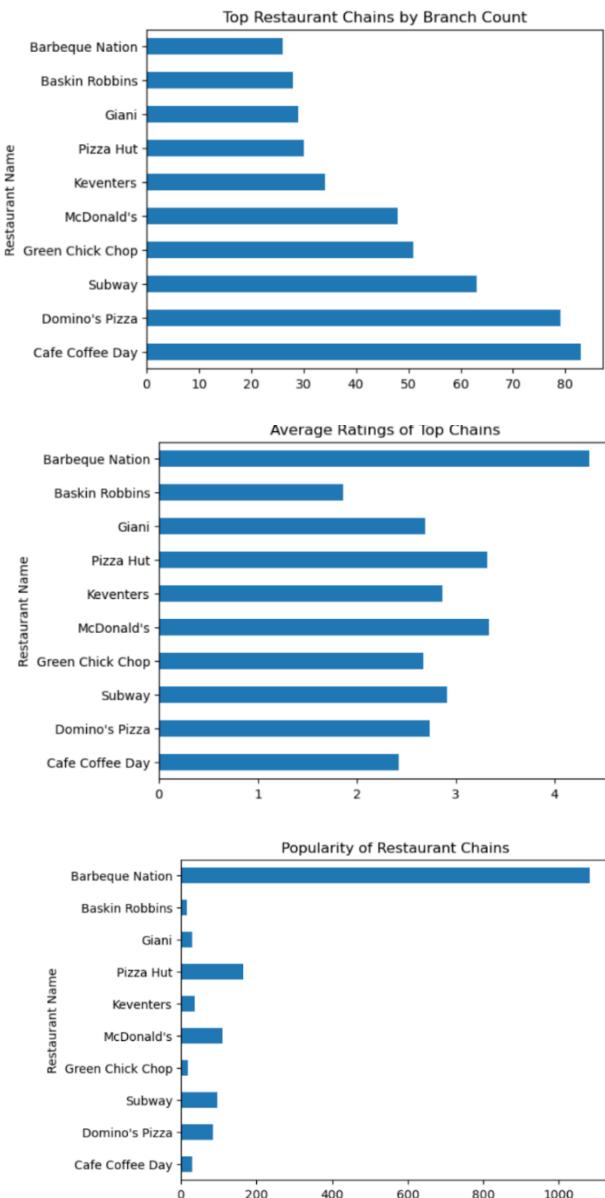
Giani - 29

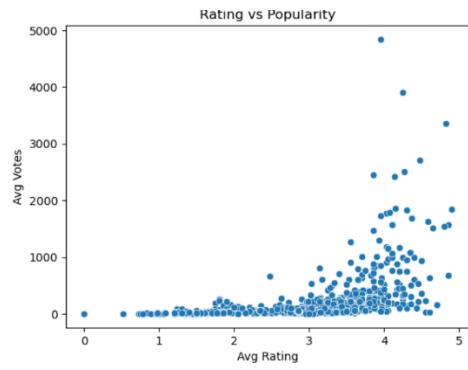
Baskin Robbins - 28

Barbeque Nation - 26

Name: count, dtype: int64

Visualization:





Key Insights:

1. Several restaurants operate as chains with multiple outlets in the dataset.
2. Popular brands such as café and fast-food chains have the highest outlet counts.
3. Chain restaurants are more commonly found in urban and metro cities.
4. Well-known chains tend to receive consistent customer ratings due to standardized service.
5. Brand recognition and trust play a major role in the success of restaurant chains.

Conclusion:

The restaurant chains analysis highlights the dominance of branded food outlets in the dataset. Chains such as Café Coffee Day and Domino's Pizza have widespread presence, indicating strong brand recognition and customer loyalty. Chain expansion plays a key role in market growth.

Overall Findings:

1. North Indian cuisine emerged as the most popular cuisine, followed by Chinese and Fast Food.
2. New Delhi recorded the highest number of restaurants, indicating a dense food market.
3. Cities with fewer restaurants, such as Inner City, showed higher average ratings, reflecting better customer satisfaction.
4. The majority of restaurants fall under the low and mid-price ranges, highlighting strong demand for affordable dining.
5. Only a limited percentage of restaurants offer online delivery, but those that do tend to receive slightly better ratings.
6. Restaurant ratings are mostly concentrated in the 3 to 4 range, indicating moderate to good service quality.
7. Multi-cuisine restaurants are common, with combinations like North Indian + Chinese being highly popular.
8. Geographic analysis revealed dense restaurant clusters in metro cities and expansion opportunities in low-density regions.
9. Several well-known restaurant chains operate multiple outlets, showing strong brand presence and customer trust.

Conclusion:

This project successfully analyzed restaurant data using exploratory data analysis techniques as part of the Data Analytics Internship at Cognifyz Technologies.

The study provided insights into cuisine popularity, city-wise distribution, pricing patterns, delivery services, customer ratings, geographic clustering, and restaurant chains. The findings highlight that affordable pricing, popular cuisines, and service convenience such as online delivery play a significant role in customer satisfaction.

Overall, the project demonstrates how data analytics can help businesses understand market trends, improve decision-making, and identify growth opportunities in the restaurant industry. It also enhanced practical skills in data cleaning, visualization, and insight generation using Python tools.