Exploratory Data Analysis (EDA)

Report on Zomato Restaurant data

Introduction

we aim to analyze Zomato restaurant data to identify key factors that contribute to the success of restaurants, as measured by their ratings. By exploring various features such as location, cuisine, pricing, and service offerings, we aim to provide insights that can help restaurant owners and Zomato users make informed decisions.

overview

1.Data overview

Data Characteristics:

Dimensions: Check the number of rows and columns in the dataset to understand its size.

Data Types: Identify the types of data (e.g., numerical, categorical) for each column.

Missing Values: Calculate the count and percentage of missing values in each column to address data quality issues.

2.Basic Statistics

Average Rating of Restaurants:

. Calculation of the average rating.

. Visualization of the average rating (e.g., bar chart)

Distribution of Restaurant Ratings:

. Histogram showing the distribution of ratings.

. Summary statistics (e.g., mean, median, mode).

3.Location Analysis

City with Highest Concentration of Restaurants:

Identification of the city with the most restaurants.

Visualization of the concentration of restaurants in different cities (e.g., pie chart).

Distribution of Restaurant Ratings Across Cities:

Box plots or bar charts showing restaurant ratings in different cities.

4.Cuisine Analysis

Most Popular Cuisines:

. Identification of the top cuisines.

. Visualization of the most popular cuisines (e.g., bar chart or word cloud).

Correlation Between Variety of Cuisines and Ratings:

Analysis of whether offering a variety of cuisines correlates with higher ratings.

5.Price Range and Rating

Relationship Between Price Range and Restaurant Ratings

Scatter plot or box plot showing the relationship between price range and ratings.

Average Cost for Two People in Different Price Categories:

Visualization of the average cost for two people in different price categories (e.g., bar chart).

6.Online Order and Table Booking

Impact of Online Order Availability on Ratings:

. Analysis of how the availability of online orders affects ratings.

. Visualization of ratings for restaurants with and without online order availability.

Distribution of Restaurants Offering Table Booking:

Analysis and visualization (e.g., pie chart) of the percentage of restaurants offering table booking.

7.Top Restaurant Chains

Identification of Top Restaurant Chains:

. Listing the top restaurant chains based on the number of outlets.

. Visualization of the top chains (e.g., bar chart).

Ratings of Top Restaurant Chains:

Analysis of the ratings of these top chains.

8.Restaurant Features

Distribution Based on Features:

. Analysis of restaurants based on features like Wi-Fi, Alcohol availability, etc.

. Visualization of feature distribution (e.g., bar charts).

Correlation Between Features and Ratings:

Investigation of whether the presence of certain features correlates with higher ratings.

9.Word Cloud for Reviews

Word Cloud Based on Customer Reviews:

Creation of a word cloud to identify common positive and negative sentiments.

Analysis of frequently mentioned words and sentiments.

10.Seasonal Trends

Seasonal Trends in Ratings or Reviews:

Exploration of any seasonal trends in restaurant ratings or user reviews.

Visualization of ratings distribution during different times of the year (e.g., line chart).

. This analysis aims to provide comprehensive insights into the factors contributing to restaurant success on Zomato. The visualizations and statistical analyses will help identify patterns and trends that can be valuable for restaurant owners and Zomato users alike.

objectives

The objective of this exploratory data analysis (EDA) on Zomato restaurant data is to identify key factors that contribute to the success of restaurants, as measured by their ratings. The analysis aims to provide insights that can help both restaurant owners and Zomato users make informed decisions.

Library using:

Pandas

Pandas is a powerful and widely-used library for data manipulation and analysis in Python. It provides data structures like Data Frame and Series to efficiently handle and analyze large datasets. With Pandas, you can:

Load data from various file formats (CSV, Excel, SQL, etc.).

Perform data cleaning, filtering, and transformation.

Conduct statistical analysis and generate summary statistics.

Handle missing data and perform data merging and joining operations

Matplotlib

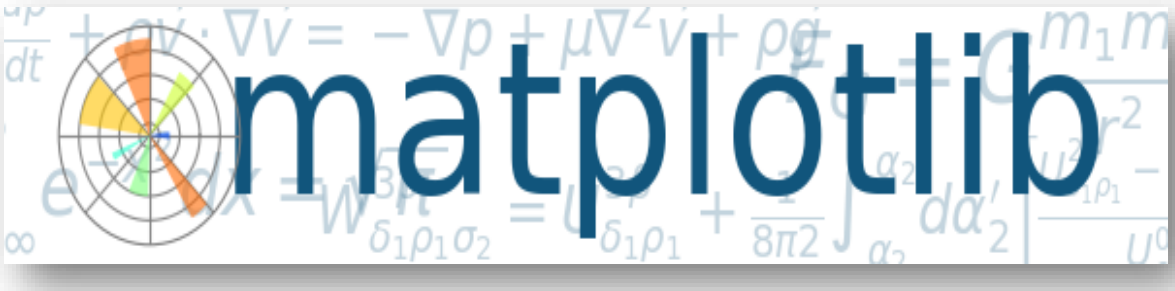
Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. It provides an object-oriented API for embedding plots into applications. Key features include:

. Creating a wide range of plots (line, bar, scatter, histogram, etc.).

. Customizing plots with titles, labels, legends, and colors.

. Exporting plots to various formats (PNG, PDF, SVG, etc.).

. Supporting interactive features like zooming and panning.



NumPy

NumPy (Numerical Python) is a fundamental package for scientific computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays. NumPy is known for:

. Efficient array manipulation and element-wise operations.

. Linear algebra, Fourier transform, and random number capabilities.

. Serving as the backbone for other libraries like Pandas and SciPy.

Seaborn

Seaborn is a statistical data visualization library built on top of Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Seaborn is particularly useful for:

. Creating complex visualizations with fewer lines of code.

. Visualizing relationships between variables (scatter plots, line plots, etc.).

. Displaying distributions of data (histograms, KDE plots, etc.).

. Enhancing visual appeal with built-in themes and color palettes.

Importance of eda (exploratory data analysis)

Exploratory Data Analysis (EDA) is a critical step in data analysis that helps in understanding the dataset before any modeling or deeper analysis is performed. Here are some key reasons why EDA is important:

1. Understanding Data Structure

EDA helps in understanding the basic structure of the data, including the types of variables, dimensions, and overall distribution. It provides insights into:

. The number of observations and variables

. Data types of each variable (numerical, categorical, etc.)

. Summary statistics (mean, median, mode, standard deviation)

2. Identifying Data Quality Issues

EDA helps in identifying and addressing data quality issues such as:

. Missing values

. Outliers

. Duplicates

. Inconsistent or incorrect data entries

3. Uncovering Patterns and Relationships

EDA allows us to uncover patterns, relationships, and correlations within the data:

. Correlations between variables

. Trends and patterns over time

. Grouping and clustering of similar data points

Preprocessing and exploration of data

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import seaborn as sns
import random
from wordcloud import WordCloud
```

After that we mount our drive

```
data = pd.read_csv("../input/zomato-restaurants-in-
india/zomato_restaurants_in_India.csv")
```

Mounting a drive means making the filesystem on the drive accessible to the operating system so that you can read from and write to it. It's like plugging in a USB drive and having it show up on your computer so you can access its files.

Here are some reasons why mounting a drive is necessary: -

Accessibility

File Access: You need to mount the drive to access the files stored on it. Without mounting, the system wouldn't know how to interact with the drive's filesystem.

Organization

Filesystem Structure: Mounting integrates the drive into the filesystem hierarchy. For example, in Linux, you might mount a drive to /mnt/external_drive, making it accessible as part of the larger file system.

-Head and tail method

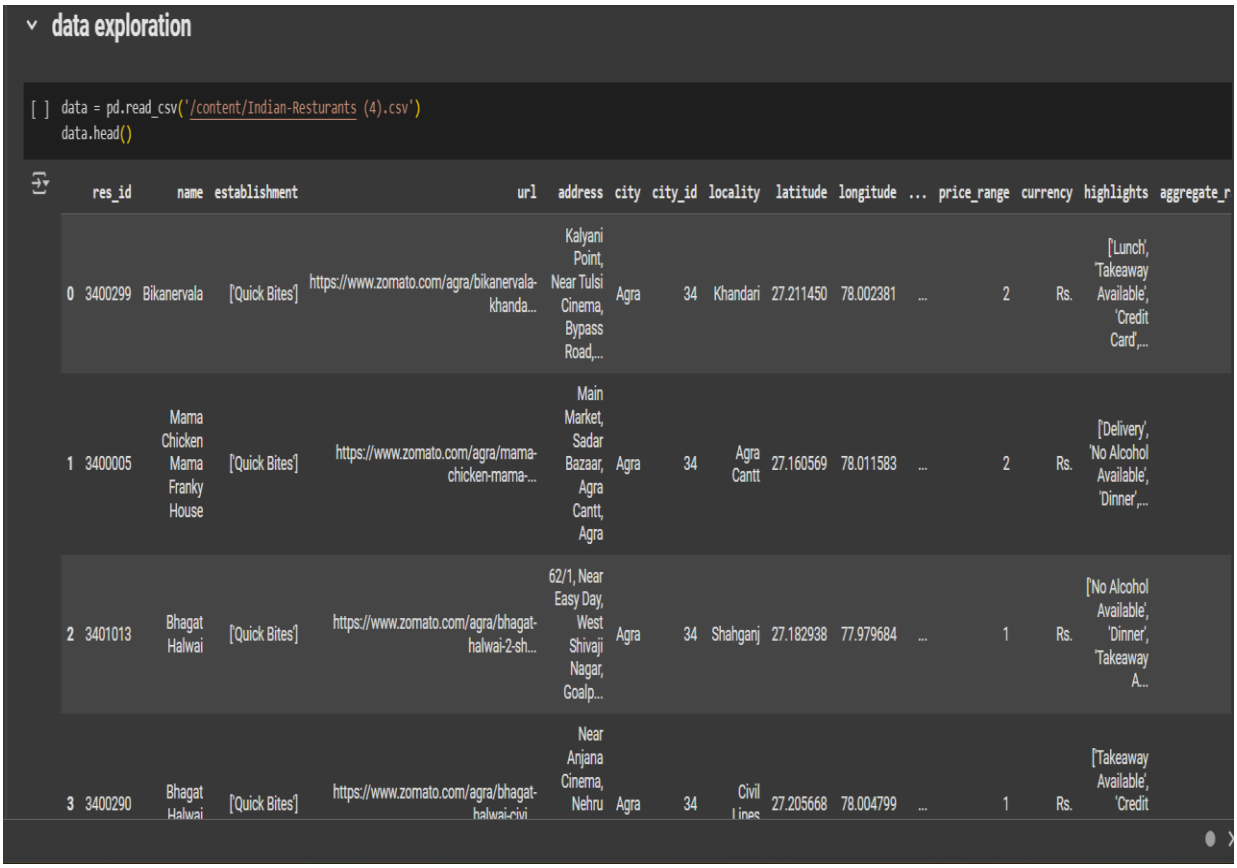
Head Method

The head method is used to display the first few rows of a dataset. It’s particularly useful for:

Getting a Quick Overview: Understand the structure and format of the data.

Inspecting Column Names: Ensure that the column names are correctly formatted.

Checking Initial Values: Verify that the initial data entries are as expected.



data exploration

```
[ ] data = pd.read_csv('/content/Indian-Resturants (4).csv')
    data.head()
```

	res_id	name	establishment	url	address	city	city_id	locality	latitude	longitude	...	price_range	currency	highlights	aggregate_r
0	3400299	Bikanervala	['Quick Bites']	https://www.zomato.com/agra/bikanervala-khanda...	Kalyani Point, Near Tulsi Cinema, Bypass Road,...	Agra	34	Khandari	27.211450	78.002381	...	2	Rs.	['Lunch', 'Takeaway Available', 'Credit Card',...	
1	3400005	Mama Chicken Mama Franky House	['Quick Bites']	https://www.zomato.com/agra/mama-chicken-mama-...	Main Market, Sadar Bazaar, Agra Cantt, Agra	Agra	34	Agra Cantt	27.160569	78.011583	...	2	Rs.	['Delivery', 'No Alcohol Available', 'Dinner',...	
2	3401013	Bhagat Halwai	['Quick Bites']	https://www.zomato.com/agra/bhagat-halwai-2-sh...	62/1, Near Easy Day, West Shivaji Nagar, Goalp...	Agra	34	Shahganj	27.182938	77.979684	...	1	Rs.	['No Alcohol Available', 'Dinner', 'Takeaway A...	
3	3400290	Bhagat Halwai	['Quick Bites']	https://www.zomato.com/agra/bhagat-halwai-civi...	Near Anjana Cinema, Nehru	Agra	34	Civil Lines	27.205668	78.004799	...	1	Rs.	['Takeaway Available', 'Credit	

tail

Tail Method

The tail method is used to display the last few rows of a dataset. It’s helpful for:

Inspecting the End of the Data: Check the format and values of the last entries.

Ensuring Data Completeness: Verify that the dataset doesn't have unexpected truncations or errors at the end.

```
data.tail()
```

	res_id	name	establishment	url	address	city	city_id	locality	latitude	longitude	...	price_range	currency	highlights
211939	3202251	Kali Mirch Cafe And Restaurant	['Casual Dining']	https://www.zomato.com/vadodara/kali-mirch-caf...	Manu Smriti Complex, Near Navrachna School, Gl...	Vadodara	32	Fatehgunj	22.336931	73.192356	...	2	Rs.	['Dinner', 'Cash', 'Lunch', 'Delivery', 'Indoo...
211940	3200996	Raju Omlet	['Quick Bites']	https://www.zomato.com/vadodara/raju-omlet-kar...	Mahalaxmi Apartment, Opposite B O B, Karoli Ba...	Vadodara	32	Karelibaug	22.322455	73.197203	...	1	Rs.	['Dinner', 'Cash', 'Takeaway Available', 'Debi...
211941	18984164	The Grand Thakar	['Casual Dining']	https://www.zomato.com/vadodara/the-grand-thak...	3rd Floor, Shreem Shalini Mall, Opposite Conqu...	Vadodara	32	Alkapuri	22.310563	73.171163	...	2	Rs.	['Dinner', 'Cash', 'Debit Card', 'Lunch', 'Tak...
211942	3201138	Subway	['Quick Bites']	https://www.zomato.com/vadodara/subway-1-akota...	G-2, Vedant Platina, Near Cosmos, Akota, Vadodara	Vadodara	32	Akota	22.270027	73.143068	...	2	Rs.	['Dinner', 'Delivery', 'Credit Card', 'Lunch',...
					Shop 7,									['Dinner',

-Data cleaning and preprocessing

Duplicate removal in data refers to the process of identifying and eliminating

repeated records or entries from a dataset. This step is crucial for ensuring data integrity and accuracy, especially before performing any analysis.

```
data.drop_duplicates(["res_id"], keep='first',inplace=True)
data.shape
```

```
(55568, 26)
```

Now we will handle the null values:

`.isnull().sum()` is a method in the Pandas library of Python that helps to identify and count missing values in a Data Frame.

```
data.isnull(). sum ()
```

Highlight columns with null values

. address: 18 nulls

.zip code: 44623 nulls

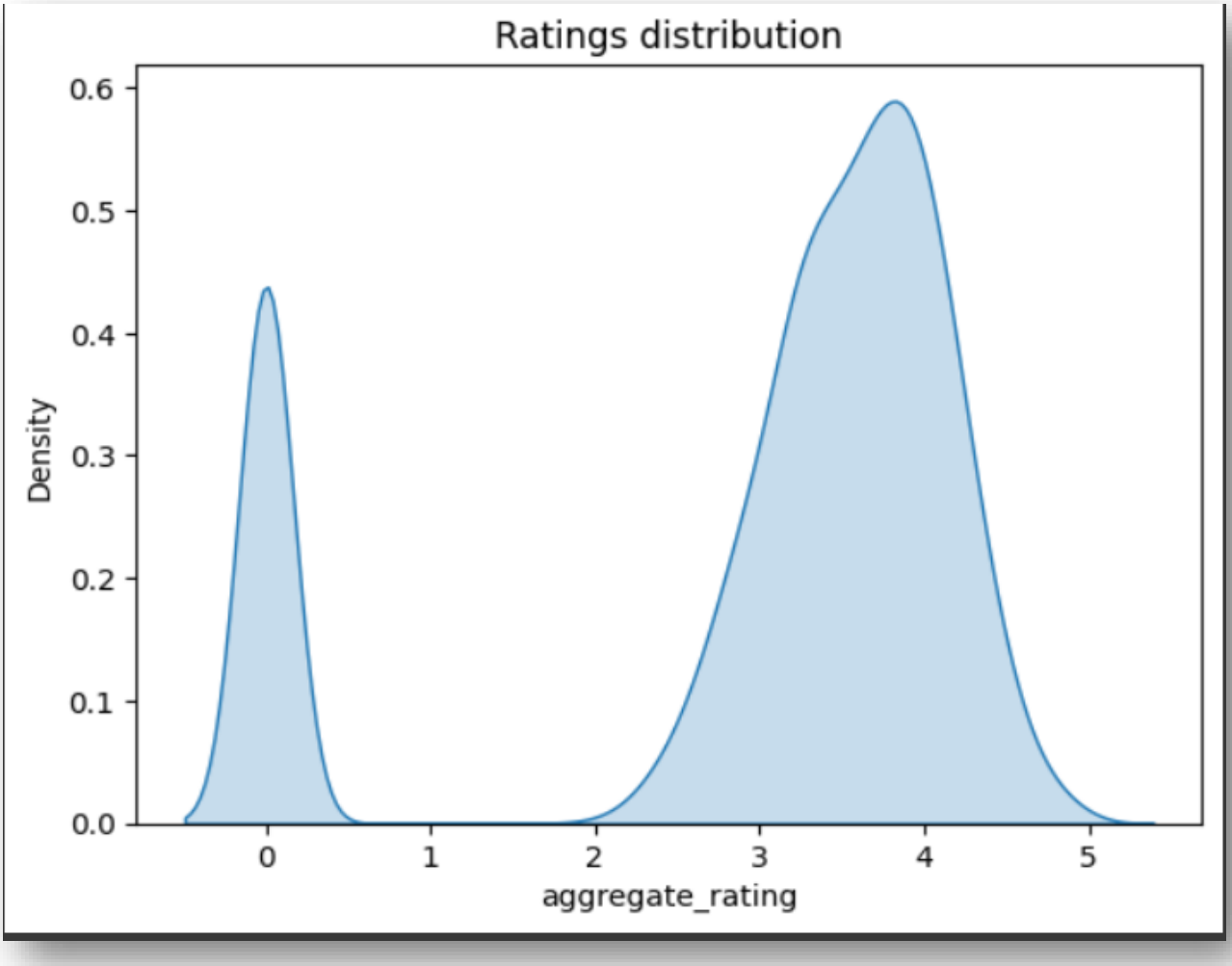
. cuisines: 470 nulls

. timings: 1003 nulls

.opentable_support: 12 nulls

	0
res_id	0
name	0
establishment	0
URL	0
address	18
city	0
city_id	0
locality	0
latitude	0
longitude	0
zip code	44623
country_id	0
locality_verbose	0
cuisines	470
timings	1003
average_cost_for_two	0
price_range	0
currency	0
highlights	0
aggregate_rating	0
rating_text	0
votes	0
photo_count	0
opentable_support	12
delivery	0
takeaway	0

-Calculate and visualize the ratings of restaurants



Analysis of Aggregate Rating Distribution

1.Understanding the distribution of aggregate ratings is essential for businesses, data scientists, and analysts to gauge user satisfaction and identify areas

for improvement. The density plot provided visualizes the distribution of aggregate ratings, offering insights into the overall rating trends.

2.The density plot showcases the distribution of aggregate ratings with the following characteristics:

.X-Axis: Represents the aggregate rating values ranging from 0 to 5.

. Y-Axis: Represents the density of these ratings.

. Peaks: There are two distinct peaks in the plot:

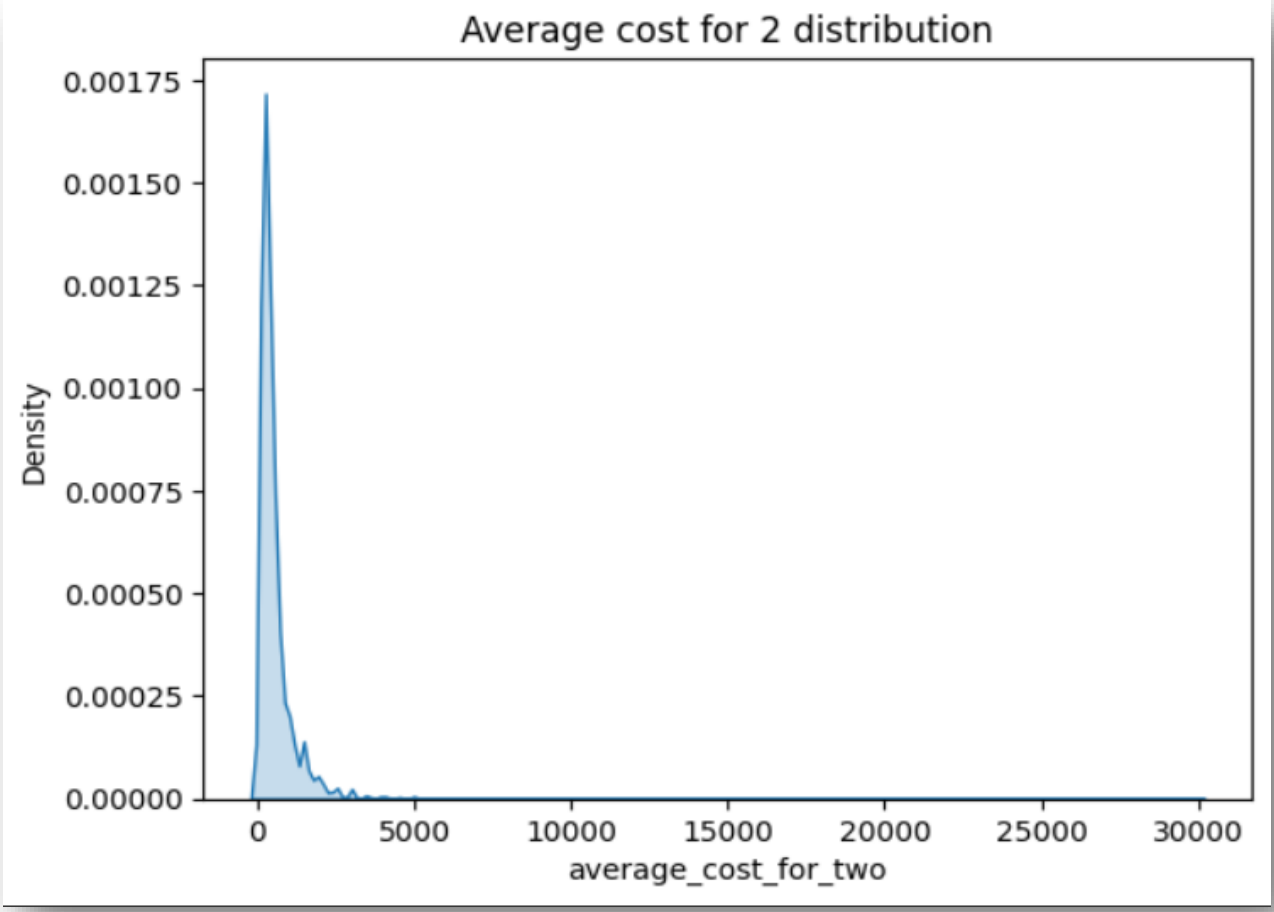
One peak around the rating value of 0.

Another peak around the rating value of 4.

The plot is shaded in blue to highlight the density areas.

4.

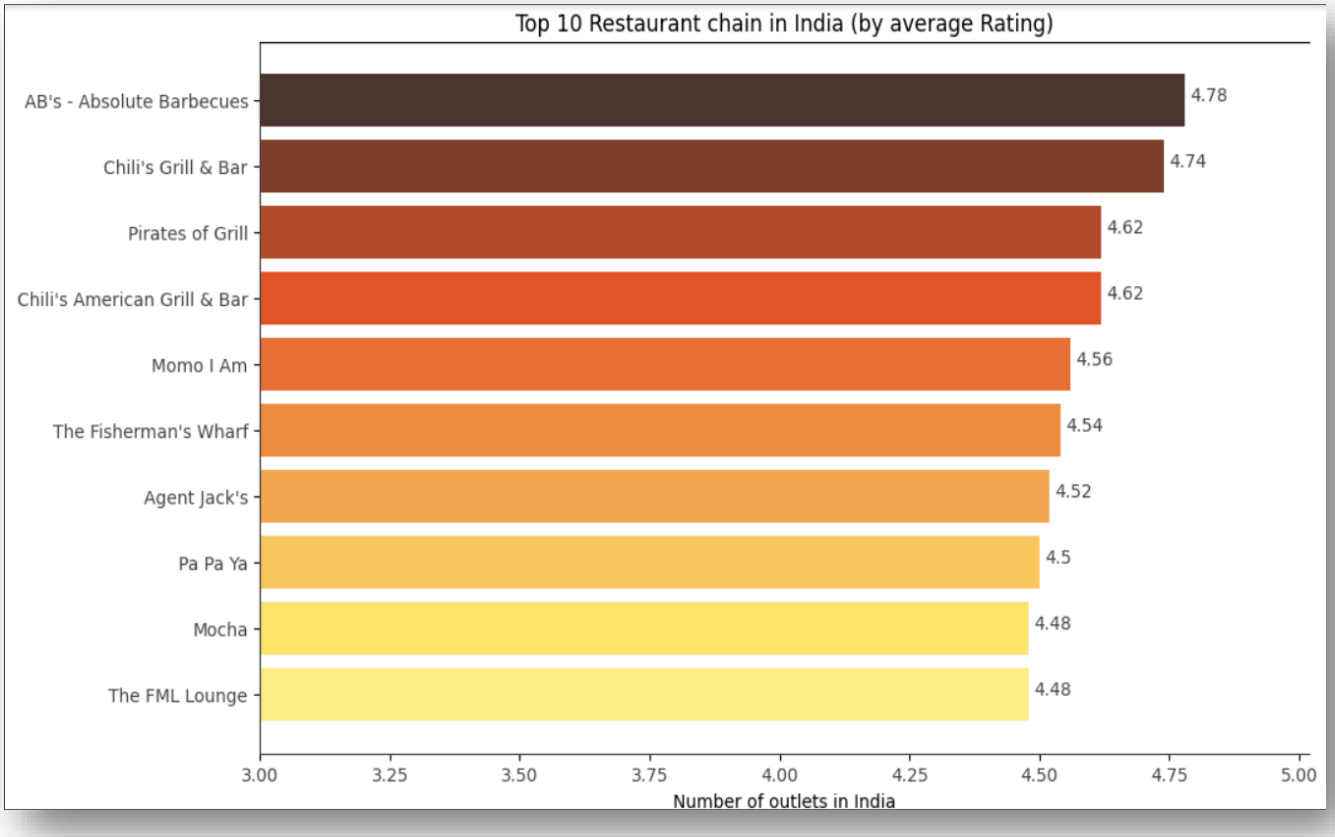
-Average cost of two distributions



The density plot showcases the distribution of the average cost for two people. The x-axis is labeled "average_cost_for_two," indicating the cost range

from 0 to 30,000. The y-axis is labeled "Density," representing the density of occurrences within the cost range.

- "An Analytical Overview of the Top 10 Restaurant Chains in India Based on Average Customer Ratings: A Comprehensive Study of Culinary Excellence and Consumer Preferences"



In the vibrant and diverse culinary landscape of India, restaurant chains have emerged as key players in shaping dining experiences. This study delves into the top 10 restaurant chains in India, meticulously ranked based on average customer ratings. By analyzing these ratings, we aim to uncover the factors that contribute to their success and popularity among consumers.

1.AB's - Absolute Barbecues (4.78): Renowned for its interactive dining experience, AB's offers a unique blend of live grills and a wide variety of dishes, making it a favorite among food enthusiasts.

2.Chili's Grill & Bar (4.74): With its American-inspired menu and vibrant ambiance, Chili's has carved a niche for itself, attracting a loyal customer base.

3.Pirates of Grill (4.62): Known for its extensive buffet and live grill concept, Pirates of Grill provides a delightful dining experience that keeps customers coming back for more.

4.Chili's American Grill & Bar (4.62): Another popular outlet of the Chili's brand, this restaurant continues to impress with its consistent quality and diverse menu.

5.Momo I Am (4.56): Specializing in momos and Asian cuisine, Momo I Am has gained a reputation for its flavorful dishes and cozy atmosphere.

6.The Fisherman's Wharf (4.54): Offering a taste of coastal cuisine, The Fisherman's Wharf stands out with its fresh seafood and charming decor.

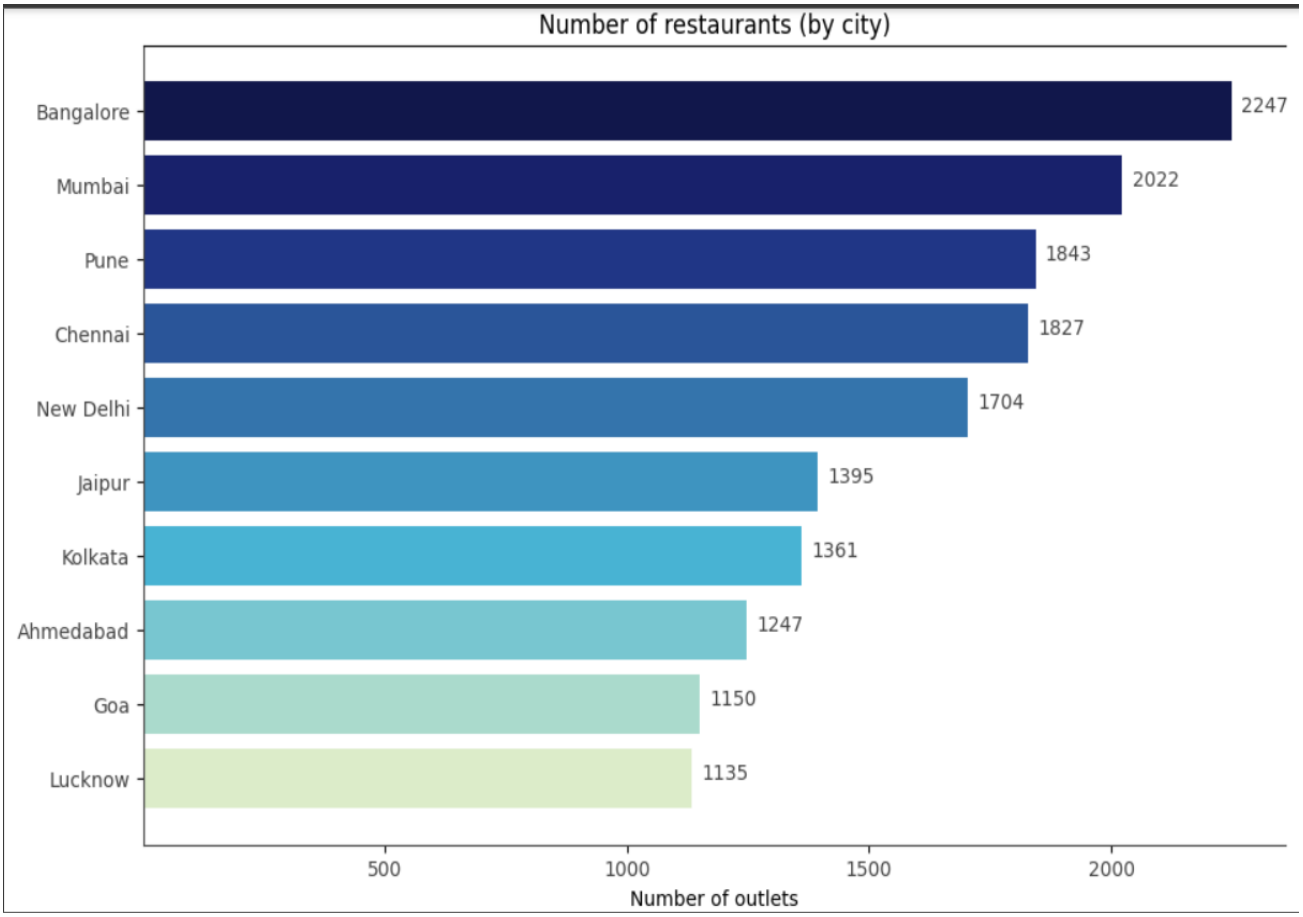
7.Agent Jack's (4.52): This unique bar and restaurant chain combines great food with an engaging atmosphere, making it a hit among patrons.

8.Pa Pa Ya (4.50): A modern Asian bistro, Pa Pa Ya is celebrated for its innovative dishes and stylish presentation.

9.Mocha (4.48): With its eclectic menu and relaxed vibe, Mocha has become a go-to spot for casual dining and social gatherings.

10.The FML Lounge (4.48): Known for its lively ambiance and diverse menu, The FML Lounge offers a memorable dining experience.

Location analysis-Identified cities with the highest concentration of restaurants



The bar chart titled "Number of restaurants (by city)" visualizes the number of restaurant outlets in ten prominent Indian cities. The cities and their corresponding number of outlets are:

Bangalore: 2247 outlets

Mumbai: 2022 outlets

Pune: 1843 outlets

Chennai: 1827 outlets

New Delhi: 1704 outlets

Jaipur: 1395 outlets

Kolkata: 1361 outlets

Ahmedabad: 1247 outlets

Goa: 1150 outlets

Lucknow: 1135 outlets

Leading City: Bangalore

Bangalore stands out as the city with the highest number of restaurant outlets, totaling 2247. This indicates a vibrant food culture and high consumer demand in the city.

High Competition in Major Metros:

Mumbai, Pune, and Chennai also exhibit a significant number of restaurant outlets, reflecting the intense competition and diverse culinary offerings in these metropolitan areas.

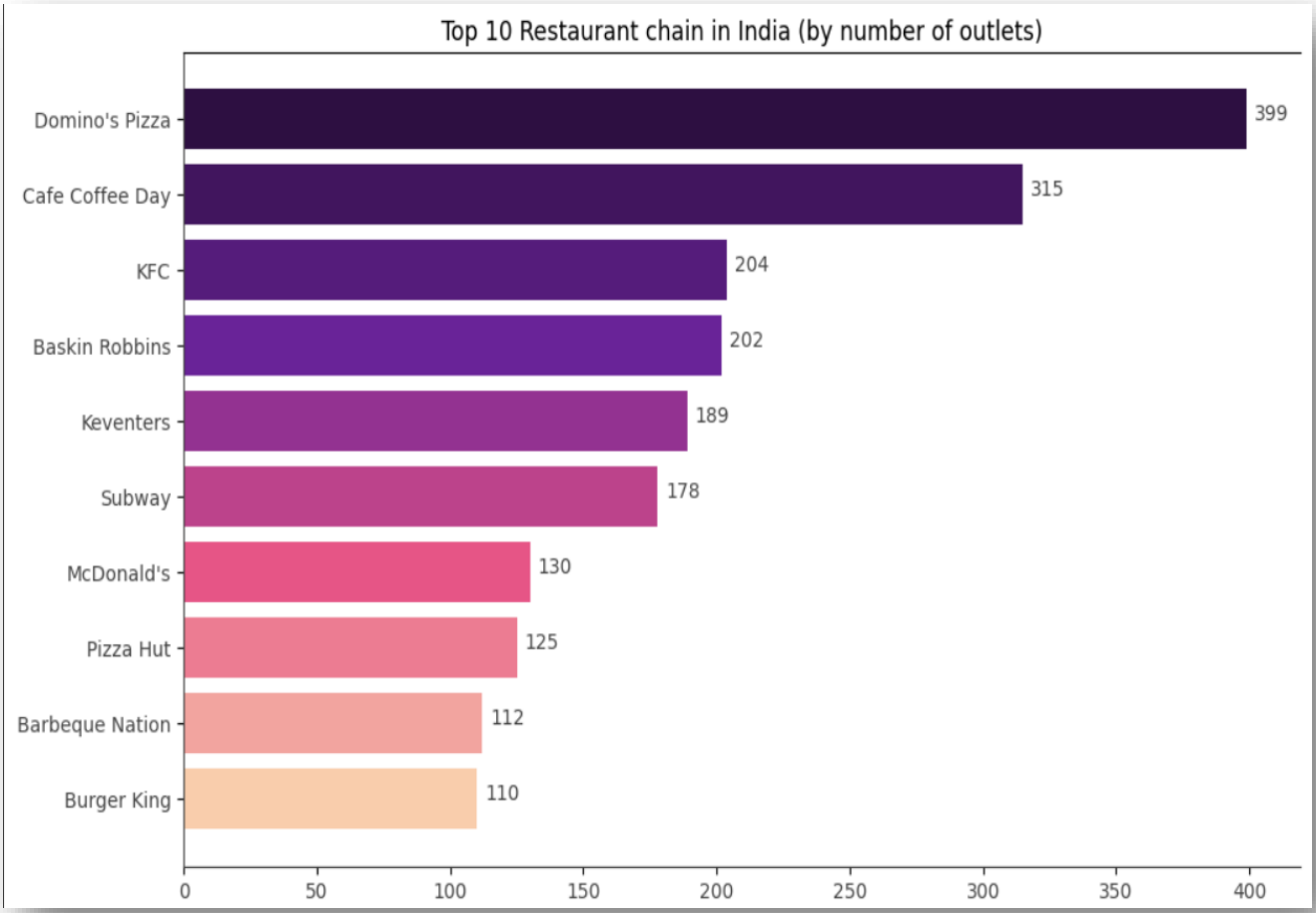
Emerging Food Markets:

Cities like Jaipur, Kolkata, and Ahmedabad, with over 1000 outlets each, showcase emerging food markets that are gradually expanding.

Tourist Hub: Goa

Goa, known for its tourism, has 1150 restaurant outlets. This highlights the importance of the hospitality industry in catering to the influx of tourists.

-top 10 restaurants chain in india (by number of outlet)



1.Highest Rated Restaurant Chains:

AB's - Absolute Barbecues stands out with the highest average rating of 4.78, indicating exceptional customer satisfaction.

Chili's Grill & Bar follows closely with an average rating of 4.74, suggesting strong customer approval.

2.Consistency in Ratings:

Both Chili's Grill & Bar and Chili's American Grill & Bar have high average ratings (4.74 and 4.62, respectively), showcasing consistency across their outlets.

Pirates of Grill also shares the same average rating as Chili's American Grill & Bar, highlighting its popularity.

3.Diverse Cuisine Preferences:

The list includes a variety of cuisines, such as barbecue, American, Asian, coastal, and more, reflecting diverse consumer preferences in India.

Momo I Am and The Fisherman's Wharf are notable for their specialized cuisine offerings, with average ratings of 4.56 and 4.54, respectively.

4.Close Competition:

The ratings of the restaurant chains are closely clustered, with only a 0.3-point difference between the highest and lowest ratings in the top 10.

Mocha and The FML Lounge have identical average ratings of 4.48, indicating a tie in customer satisfaction.

5.Customer Focus:

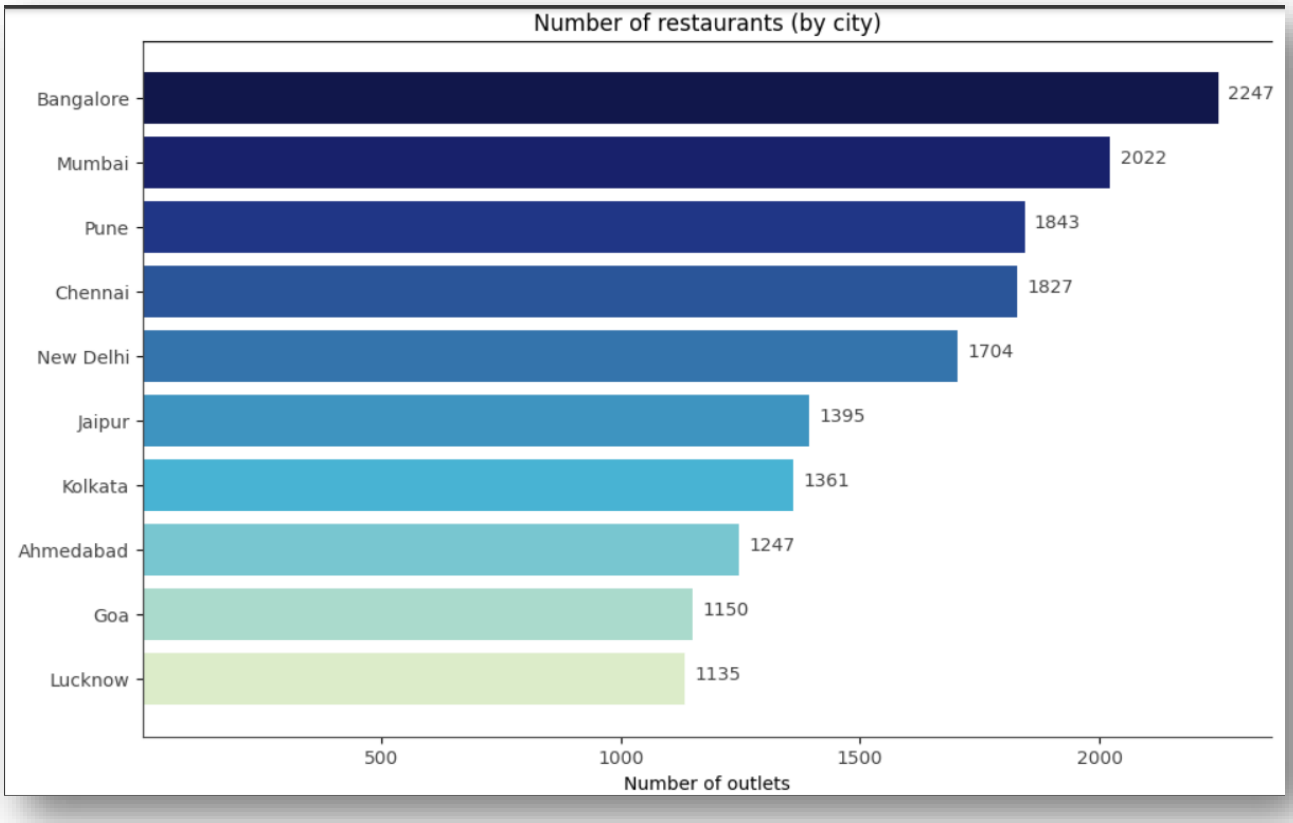
High average ratings suggest that these restaurant chains prioritize customer experience, quality of food, and service.

The consistent high ratings across different types of restaurants indicate that customers value these aspects regardless of cuisine type.

6.Potential Areas for Improvement:

While all the listed restaurant chains have impressive ratings, those with slightly lower ratings (around 4.48) could explore ways to further enhance their offerings and customer experience to reach the top.

-no of restaurants by cities



1.Bangalore: Has the highest number of restaurant outlets, totaling 2247. This indicates a bustling food scene and high demand for dining options in the city.

2.Mumbai: Follows with 2022 restaurant outlets, reflecting its status as a major metropolitan area with diverse culinary offerings.

3.Pune: Has 1843 restaurant outlets, showing significant growth in its dining sector.

4.Chennai: With 1827 restaurant outlets, Chennai showcases its rich culinary heritage and popularity among locals and tourists.

5.New Delhi: Features 1704 restaurant outlets, highlighting its vibrant and diverse food culture.

6.Jaipur: Has 1395 restaurant outlets, indicating a strong dining market in the city.

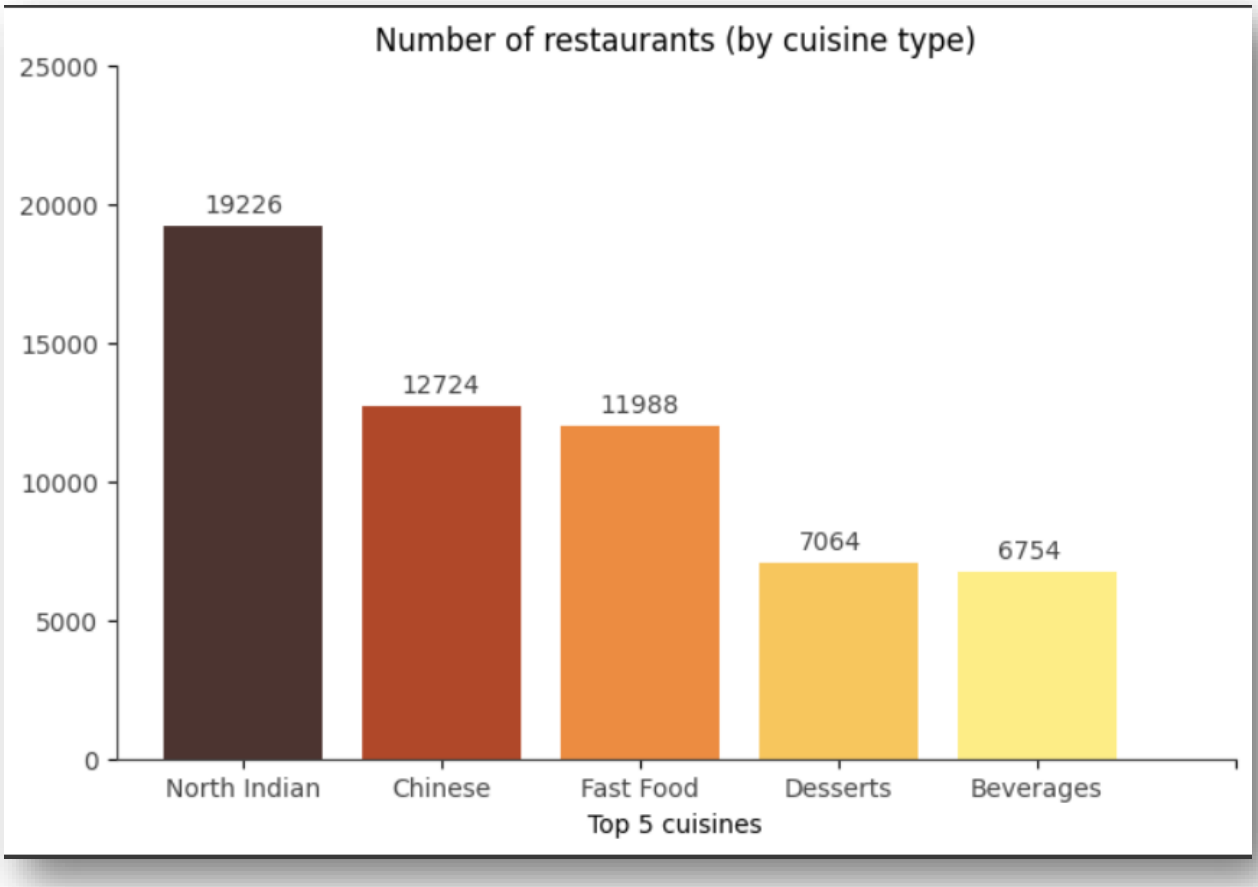
7.Kolkata: With 1361 restaurant outlets, Kolkata's love for food and traditional cuisine is evident.

8.Ahmedabad: Has 1247 restaurant outlets, reflecting the city's growing food industry.

9.Goa: Features 1150 restaurant outlets, catering to both local residents and tourists.

10.Lucknow: Has the fewest restaurant outlets among the listed cities, with 1135, but still indicates a considerable dining market.

Cuisine analysis-Determine the most popular cuisines among the listed restaurants. Investigate if there's a correlation between the variety of cuisines offered and restaurant ratings



1.Popularity of North Indian Cuisine:

North Indian cuisine has the highest number of restaurants (19226), indicating its immense popularity and widespread availability across India.

The significant lead in the number of North Indian restaurants suggests a strong preference for this cuisine among consumers.

2.Chinese Cuisine:

Chinese cuisine is the second most popular, with 12724 restaurants. This reflects its popularity and the demand for diverse and flavorful dishes.

The high number of Chinese restaurants highlights the adaptation and incorporation of Chinese flavors into the Indian culinary landscape.

3.Fast food:

Fast Food ranks third with 11988 restaurants, showcasing its appeal to a wide range of consumers seeking quick and convenient dining options. The prevalence of fast-food outlets indicates a growing trend towards fast and casual dining experiences.

4.Desserts and Beverages:

Desserts and Beverages have significantly fewer restaurants compared to the top three cuisines, with 7064 and 6754 restaurants, respectively.

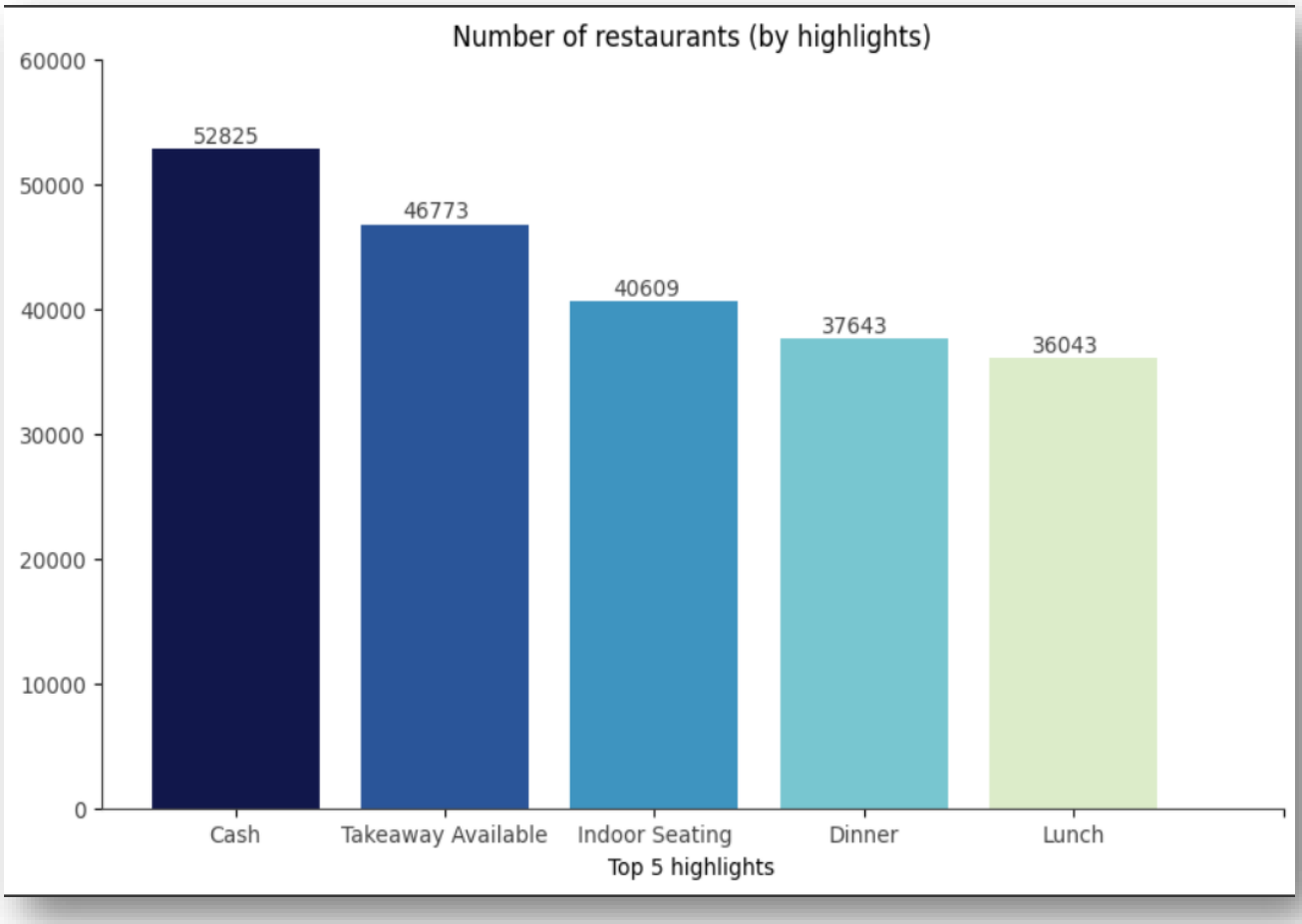
The lower number of dessert and beverage outlets suggests that while they are popular, they are often supplementary to the main dining experience rather than standalone options.

5.Culinary Diversity:

The graph highlights the culinary diversity in India, with a mix of traditional, international, and fast-food cuisines being widely available.

The variety of cuisine types reflects the eclectic tastes and preferences of Indian consumers.

Highlights and features of restaurant- Features of Premier Restaurant Chains in India: A Comprehensive Study on Culinary Excellence, Ambiance, and Service Quality"



1.Cash Payment is the Most Common Highlight:

A total of 52825 restaurants offers cash payment options. This indicates that despite the growing

popularity of digital payments, cash remains a widely accepted and convenient payment method in India.

2.Takeaway Services are Highly Popular:

46773 restaurants offer takeaway services. This reflects the increasing demand for convenience and the preference for dining at home or on the go.

3.Indoor Seating Availability:

40609 restaurants provide indoor seating arrangements, making it the third most common highlight. This suggests that many restaurants prioritize providing a comfortable and dine-in experience for their customers.

4.Dinner Service:

37643 restaurants offer dinner service, which places it fourth on the list of highlights. This indicates the importance of dinner service in the restaurant industry, catering to customers who prefer dining out in the evening.

5.Lunch Service:

36043 restaurants offer lunch service, making it the least common highlight among the top 5. However, the high number still indicates a substantial demand for lunch services, catering to the mid-day meal preferences of customers.

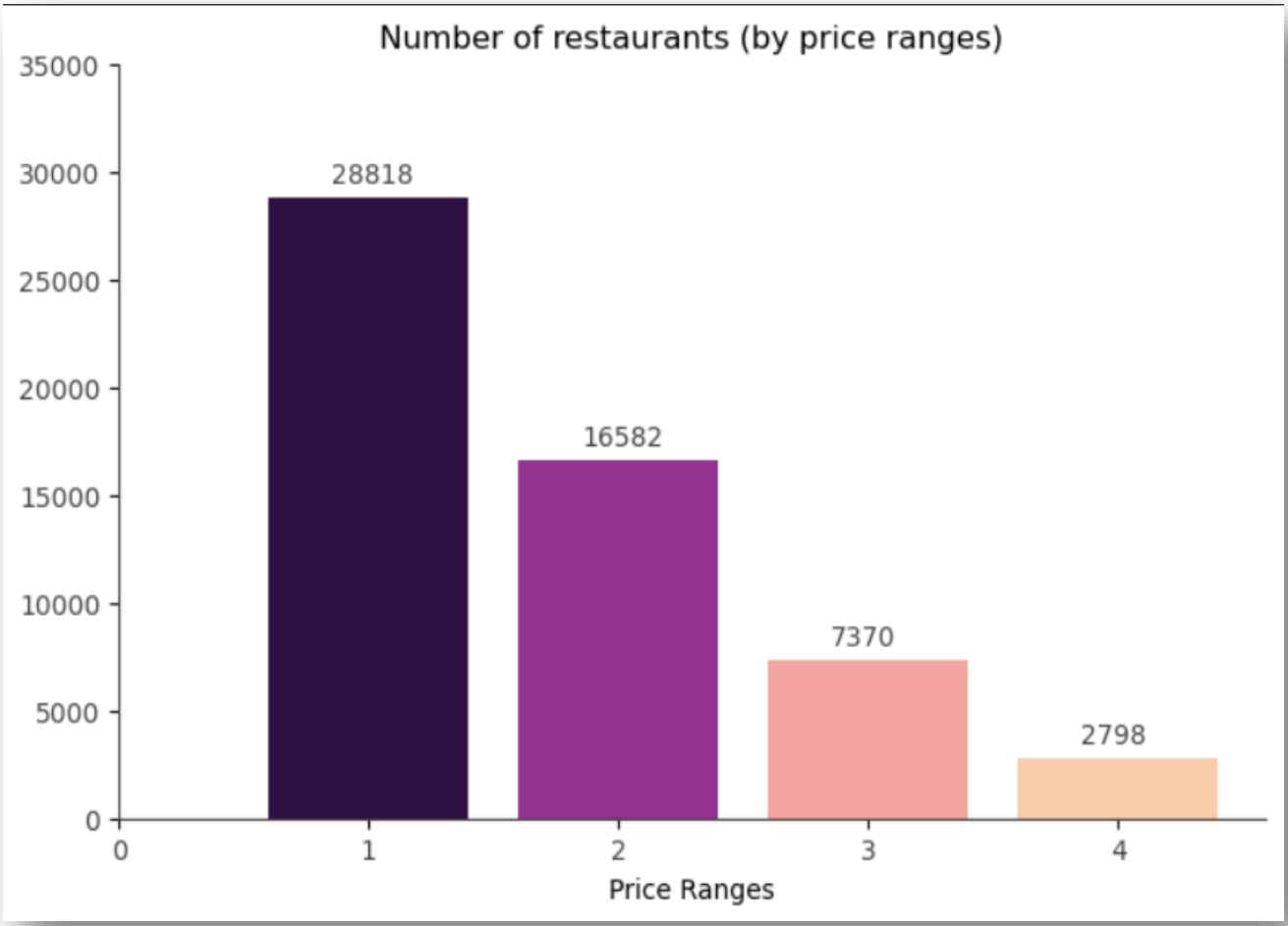
-Word cloud for reviews

```
# https://www.geeksforgeeks.org/generating-word-cloud-python/

hl_str = ""
for i in hl:
    hl_str += str(i) + " "
wordcloud = WordCloud(width = 800, height = 500,
                background_color ='white',
                min_font_size = 10, max_works=30). generate(hl_str)

plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)

plt.show()
```

a word cloud based on customer reviews to identify common positive and negative sentiments


Alcohol Available
Dinner Delivery
Seating Digital
Indoor Seating
Seating Air
Delivery Credit
Pure Veg
Dinner Takeaway
Dinner Cash
Available Debit
Credit Card
Card Lunch
Table booking
Digital Payments
Card Dinner
Card Indoor
Seating Available
Takeaway Available
Air Conditioned
Available Cash
Available Indoor
Lunch Cash
Payments Accepted
Available Dinner
Available Lunch
Debit Card
Cash Takeaway
Outdoor Seating
Cash Credit
Available Credit

-Price range count



The bars are color-coded to help differentiate between the price ranges:

Price Range 1: The tallest bar with a value of 28818, representing the most affordable restaurants.

Price Range 2: The second tallest bar with a value of 16582, indicating moderately priced restaurants.

Price Range 3: The third bar with a value of 7370, representing higher-priced restaurants.

Price Range 4: The shortest bar with a value of 2798, indicating the most expensive restaurants.

. Prevalence of Affordable Dining Options:

The majority of restaurants fall within the lowest price range (1), with 28818 restaurants. This suggests that affordable dining options are the most common and widely available in India.

. Moderately Priced Restaurants:

16582 restaurants fall within price range 2, indicating a significant number of moderately priced dining options. This reflects a substantial market for mid-range dining experiences.

. Higher-Priced Restaurants:

The number of restaurants in price range 3 is 7370, which is less than half of those in price range 2. This indicates that higher-priced dining options are less common but still present in the market.

. Least Common High-End Restaurants:

The fewest number of restaurants are in the highest price range (4), with only 2798 restaurants. This suggests that high-end dining options are the least common and cater to a niche market.

. Significant Drop in Numbers:

There is a significant drop in the number of restaurants as the price range increases. This distribution highlights the consumer preference for more affordable and moderately priced dining options.

Thank you