zomato

Discover more places to eat around you

EXPLORATORY DATA ANALYSIS

CAPSTONE PROJECT

INTRODUCTION

In the age of digital dining, understanding the dynamics of restaurant choices and customer sentiments is pivotal. This project revolves around a thorough analysis of data from Zomato, a leading platform for restaurant discovery and food delivery.

Zomato's dataset provides a goldmine of information on restaurants, user reviews, and ratings. Our goal is to dig into this data, unveil meaningful insights, and offer practical recommendations for restaurants and Zomato.

Through our exploration, we aim to uncover patterns, correlations, and key trends in user behavior, ultimately contributing to a better understanding of what makes a restaurant successful on Zomato.

This report showcases our journey into the world of Zomato data analytics, emphasizing the actionable insights gained and their potential impact on the culinary industry.

OBJECTIVE

1. Business Improvement:-
Identify areas for improvement in restaurant offerings, - service, and overall customer experience to enhance business performance.

2. Market Insights:-
Gain insights into market trends, popular cuisines, and customer preferences to stay competitive in the food industry.

3. Strategic Decision-Making:-
Provide data-driven insights to guide strategic decisions for both individual restaurants and Zomato as a platform.

4. Customer Retention:- Understand customer sentiments and preferences to implement strategies that enhance customer satisfaction and loyalty.

5. Optimization of Resources:- Optimize menu offerings, pricing strategies, and operational efficiency to make better use of resources and maximize profitability.

6.Platform Enhancement:-

Contribute insights to improve the Zomato platform, making it more user-friendly and valuable for both customers and restaurant partners.

7. Marketing and Promotions:-

Develop targeted marketing and promotional campaigns based on seasonal trends and user behavior to attract and retain customers.

8. Competitive Analysis:-

Benchmark restaurant performance against competitors to identify strengths, weaknesses, and opportunities for differentiation.

9. Revenue Growth:

Identify strategies for revenue growth, such as upselling, cross-selling, and attracting new customers.10. Operational Efficiency:-Enhance operational efficiency by identifying areas for improvement and resource optimization.

10. Operational Efficiency:

Enhance operational efficiency by identifying areas for improvement and resource optimization.

1. NumPy ('import numpy as np`):

Numerical Operations: NumPy provides support for large, multi-dimensional arrays and matrices, along with a collection of high-level mathematical functions, enabling efficient numerical operations in Python.

2. Pandas ('import pandas as pd`):

Data Manipulation and Analysis: Pandas simplifies data manipulation and analysis with data structures like Series and DataFrame, making it a powerful tool for tasks such as cleaning, filtering, and exploring structured data.

3. Seaborn ('import seaborn as sns`):

Statistical Data Visualization: Seaborn simplifies the creation of complex statistical visualizations, providing a high-level interface for producing informative and aesthetically pleasing plots.

4. Matplotlib ('import matplotlib.pyplot as plt'):-

 2D Plotting Library: Matplotlib is a versatile 2D plotting library for Python, offering a wide range of customizable static visualizations, including charts, graphs, and figures.

5. Plotly Express ('import plotly.express as px`):-

Interactive Visualizations: Plotly Express is a high-level interface for generating interactive visualizations with minimal code, making it suitable for creating web-based interactive plots and charts.

6. Warnings (import warnings warnings.filterwarnigs('ignore')):-

Importing the warning Module: is a function call that sets up a filter for warning messages. The parameter 'ignore' specifies that any warning message issued during program exeecution should dbe ignored, meaning they won't be displayed to the user

Loading Dataset

```
In [2]:   1   # Loading Dataset
          2
          3   df=pd.read_csv('Indian-Resturants.csv')
```

I've load dataset named 'Indian_restaurants.csv' into a variable named 'df'. In order to import the dataset, I've used Pandas Libraries in which I've used 'pd.read_csv' command to import the respective dataset.

Viewing Column

```
In [5]:   1   df.columns
Out[5]: Index(['res_id', 'name', 'establishment', 'url', 'address', 'city', 'city_id',
               'locality', 'latitude', 'longitude', 'zipcode', 'country_id',
               'locality_verbose', 'cuisines', 'timings', 'average_cost_for_two',
               'price_range', 'currency', 'highlights', 'aggregate_rating',
               'rating_text', 'votes', 'photo_count', 'opentable_support', 'delivery',
               'takeaway'],
              dtype='object')
```

I've accessed and displayed the column names of the loaded dataset using df.columns. These columns encompass various aspects of 'Indian restaurants' information, providing a detailed overview of the dataset.

Basic Composition of Data

```
In [4]:   1  df.describe()
```

Out[4]:

	res_id	city_id	latitude	longitude	country_id	average_cost_for_two	price_range	aggregate_rating	votes	photo_c
count	2.119440e+05	211944.000000	211944.000000	211944.000000	211944.0	211944.000000	211944.000000	211944.000000	211944.000000	211944.00
mean	1.349411e+07	4746.785434	21.499758	77.615276	1.0	595.812229	1.882535	3.395937	378.001864	256.97
std	7.883722e+06	5568.766386	22.781331	7.500104	0.0	606.239363	0.892989	1.283642	925.333370	867.66
min	5.000000e+01	1.000000	0.000000	0.000000	1.0	0.000000	1.000000	0.000000	-18.000000	0.00
25%	3.301027e+06	11.000000	15.496071	74.877961	1.0	250.000000	1.000000	3.300000	16.000000	3.00
50%	1.869573e+07	34.000000	22.514494	77.425971	1.0	400.000000	2.000000	3.800000	100.000000	18.00
75%	1.881297e+07	11306.000000	26.841667	80.219323	1.0	700.000000	2.000000	4.100000	362.000000	128.00
max	1.915979e+07	11354.000000	10000.000000	91.832769	1.0	30000.000000	4.000000	4.900000	42539.000000	17702.00

I've generated descriptive statistics for the loaded dataset stored in the variable 'df' using the describe method. This Pandas function provides a summary of statistical measures, including count, mean, standard deviation, minimum, 25th percentile, median (50thpercentile), 75th percentile, and maximum values for each numeric column in the dataset. This summary aids in understanding the central tendency, dispersion, and distribution of the numerical features in the dataset.

Checking Dimentions

```
In [19]:   1  df.shape

Out[19]:  (60417, 26)
```

df.shape, revealing that it consists of 60,417 rows and 26columns.

Checking Information

```
In [6]:    1  df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 211944 entries, 0 to 211943
Data columns (total 26 columns):
 #   Column            Non-Null Count    Dtype
---  ------            --------------    -----
 0   res_id            211944 non-null   int64
 1   name              211944 non-null   object
 2   establishment     211944 non-null   object
 3   url               211944 non-null   object
 4   address           211810 non-null   object
 5   city              211944 non-null   object
 6   city_id           211944 non-null   int64
 7   locality          211944 non-null   object
 8   latitude          211944 non-null   float64
 9   longitude         211944 non-null   float64
 10  zipcode           48757 non-null    object
 11  country_id        211944 non-null   int64
 12  locality_verbose  211944 non-null   object
 13  cuisines          210553 non-null   object
```

I've obtained information about the dataset using df.info(). This method provides a concise summary, including the total number of entries, the data types of each column, and the count of non-null values.

Removing Duplicate

I've removed duplicate rows from the dataset using df.drop_duplicates(inplace=True). This operation modifies the DataFrame in-place, eliminating any duplicate entries based on all columns.

Brand Having Highest No. of Outlet
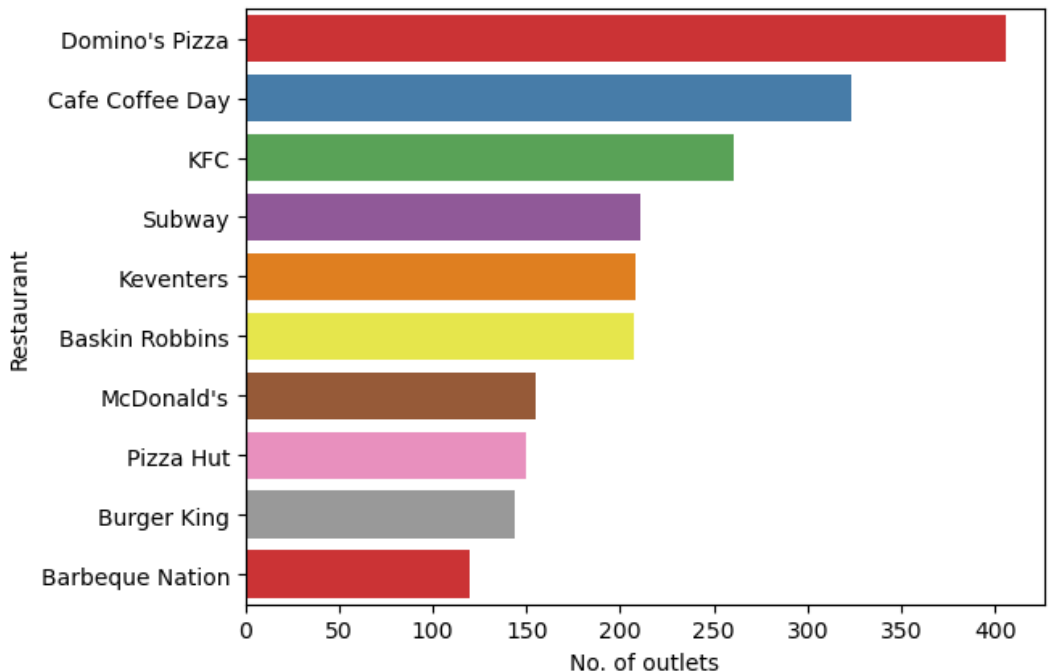
I've visualized the top 10 restaurants with the highest number of outlets using a seaborn bar plot (sns.barplot). The x-axis represents the number of outlets, and the y-axis displays the respective restaurant names. This visualization offers a clear comparison of the outlet counts for different restaurants, providing insights into the distribution and popularity of these establishments based on the available data.

Average Rating of Top 15 Outlet
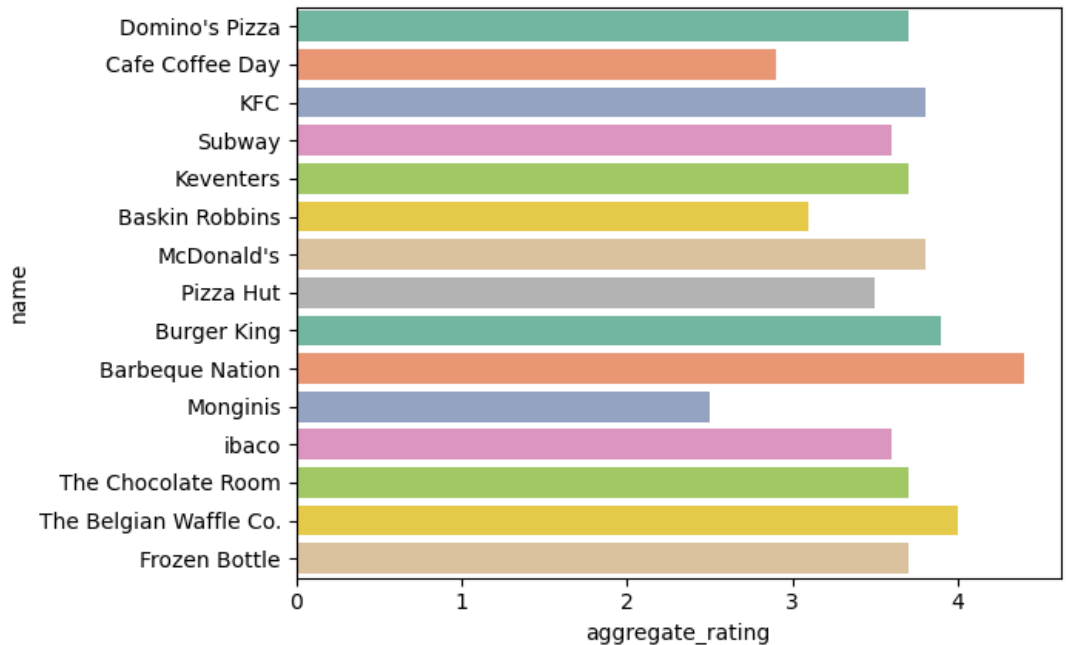
The presented data shows the top 15 restaurants along with their corresponding average ratings. Each row represents a restaurant, and the two columns display the restaurant name and its average rating. The average rating reflects the mean value of the 'aggregate_rating' column foreach restaurant in the dataset. This information allows users to quickly compare and identify the top-rated restaurants based on their average ratings. It can be useful for making informed decisions about where to dine or order food.

# Restaurants with high number of votes.
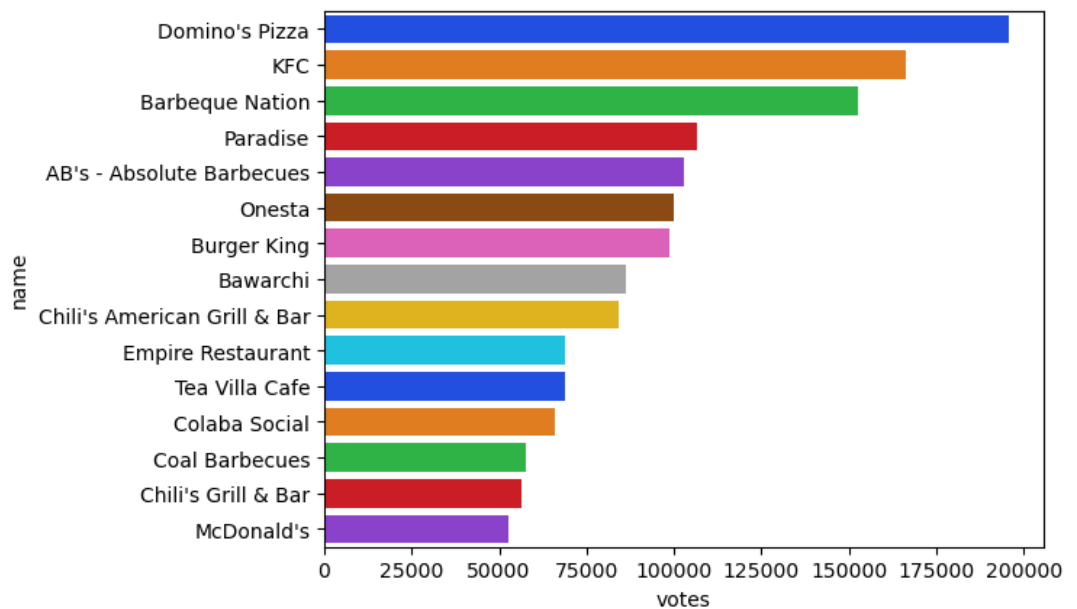
These numbers signify the level of customer involvement and satisfaction with these particular restaurants. From a business standpoint, this information indicates the popularity and positive reception of these establishments among customers. Restaurants can leverage high vote counts in their marketing strategies to attract a broader audience and strengthen their brand reputation. It's also essential for businesses to continue delivering quality services to maintain and enhance customer satisfaction.

DATA VISUALIZATION
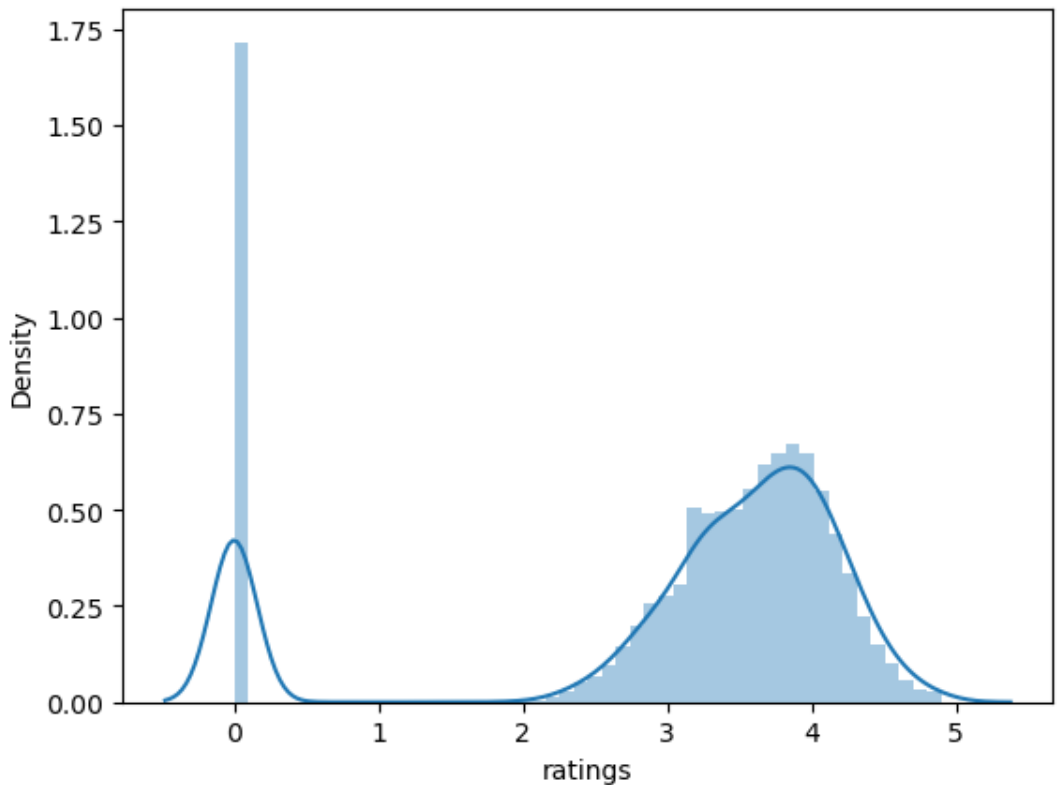
The distribution plot reveals a significant line at zero, indicating a substantial number of unrated restaurants. The bell curve starting at 2 and ending at 5 suggests most rated restaurants fall within this range. From a business perspective, addressing zero ratings is crucial, and leveraging positive ratings can enhance overall customer satisfaction and brand perception.

DATA VISUALIZATION

Cities With Highest Concentration of Restaurants
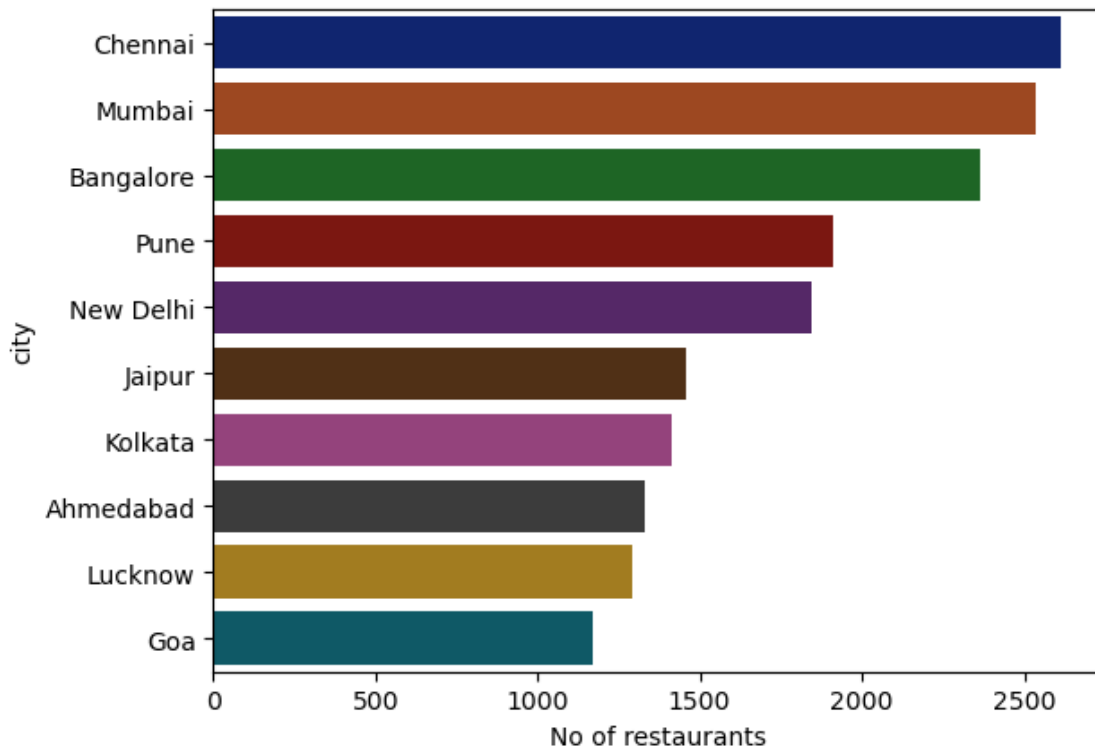
Chennai emerges as the city with the highest restaurant density, closely followed by Mumbai, Bangalore, and Pune in the dataset's top 10 cities. This insight provides a quick understanding of the regional distribution of dining establishments, offering valuable information for businesses and stakeholders in the restaurant industry.

DATA VISUALIZATION

Areas with high restaurant density
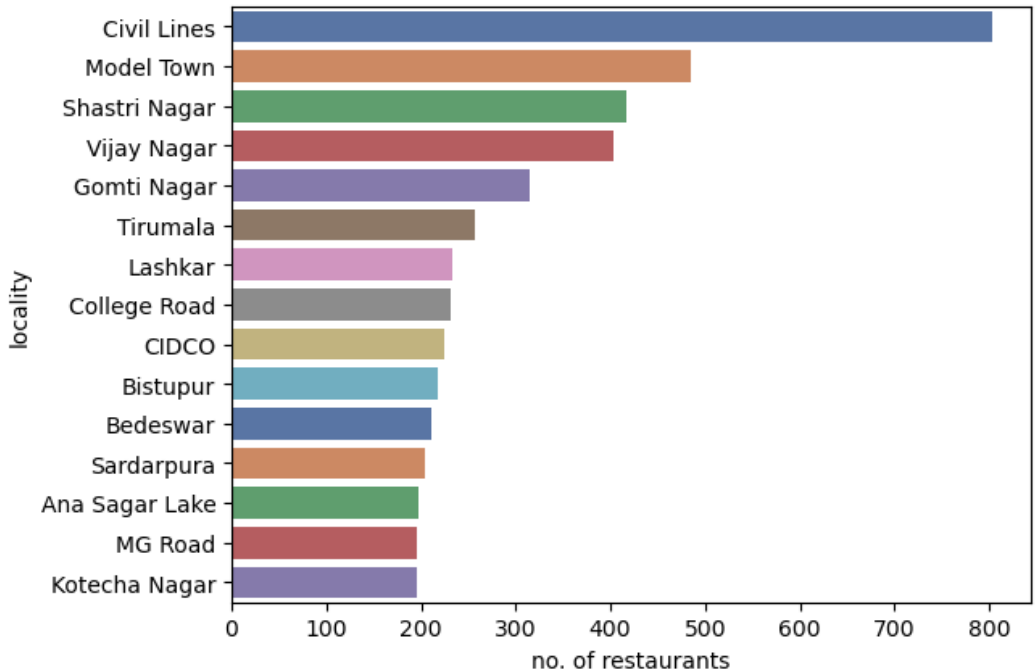
The bar plot reveals areas with a high concentration of restaurants. Notably, "Civil Lines" takes the lead, followed by "Model Town," "Shastri Nagar," and "Vijay Nagar" and so on. These insights highlight specific localities with a vibrant restaurant scene, providing valuable information for businesses and consumers seeking diverse dining options in these areas.

DATA VISUALIZATION

Distribution of Restaurant's Rating According to City


distribution of restaurant ratings in Jaipur
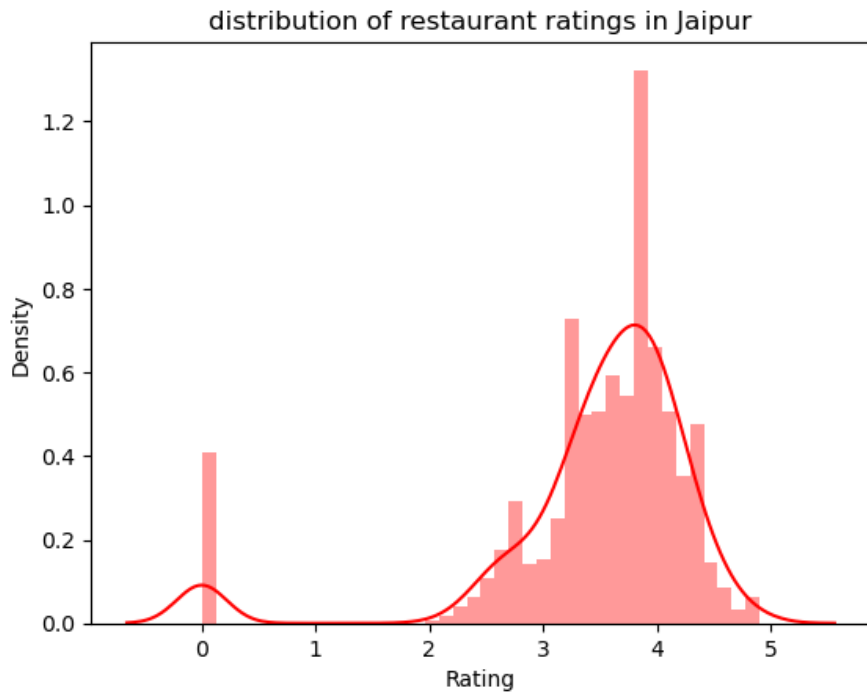
The image depicts the distribution of restaurant ratings in Jaipur. It combines a histogram with a line graph showing the density distribution of ratings. The x-axis represents ratings from 0 to 5, while the y-axis represents density (up to approximately 1.2). Notably, there are two peaks: one near a rating of 1 and a taller peak near a rating of 4.5.

Most Popular Cusins According to Restaurants

Most Popular cuisine in Barbeque Nation
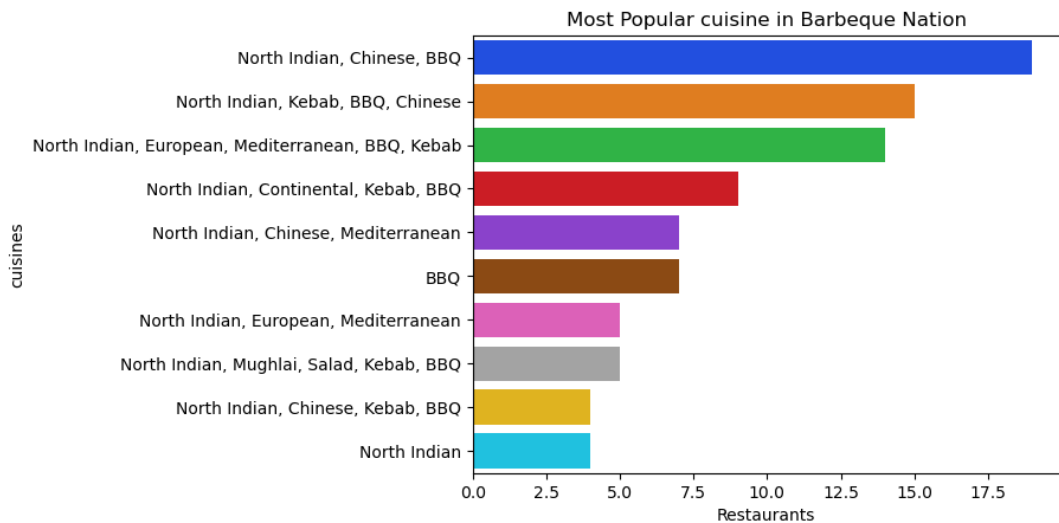
The most popular cuisines at Barbeque Nation. It features a horizontal bar graph with seven bars representing different cuisine combinations. The cuisines include North Indian, Chinese, BBQ, Kebab, Mediterranean, European, and Continental. Each bar's length corresponds to the number of restaurants offering that cuisine mix. Notably, the tallest bar represents the Chinese and BBQ combination.

Relationship between Price Range and Restaurant Ratings
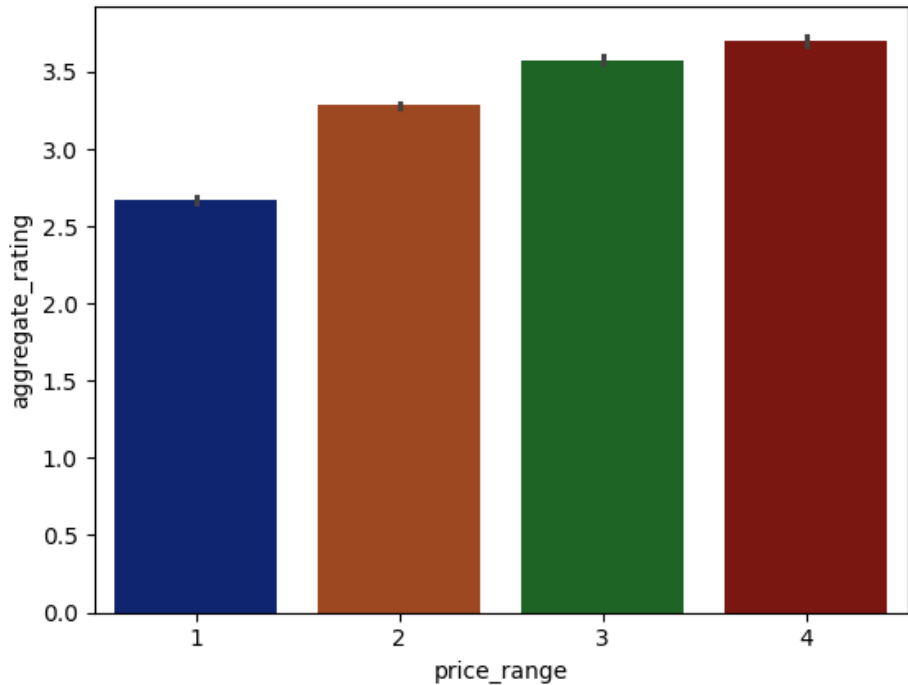
This bar chart with four vertical bars, each representing a different price range. The x-axis is labeled "price_range" (ranging from 1 to 4), and the y-axis represents the aggregate rating (ranging from 0 to 3.5). The bars correspond to specific price ranges: blue for price range 1, orange for range 2, green for range 3, and red for range 4. Each bar has a black line at the top indicating the precise rating value. The chart visually compares ratings across different price categories.

DATA VISUALIZATION

## Visualizing the Average cost for two people in different Price Categories
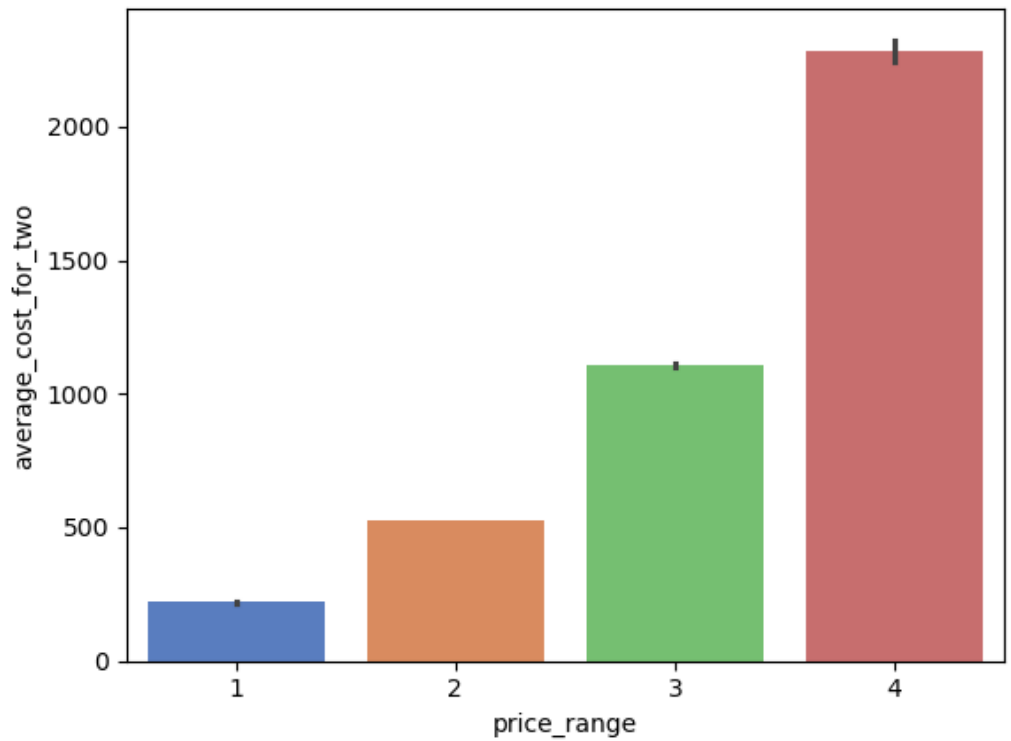
The bar plot offers a snapshot of how average costs for two diners relate to restaurant price ranges. It helps quickly identify trends, revealing whether higher-priced restaurants are associated with higher average costs. This insight is crucial for businesses in making informed decisions about pricing strategies and understanding customer perceptions of affordability.

Impact of Online Order on Restaurants Ratings

Restaurants offering delivery services (average rating: 3.5)tend to have higher ratings compared to those without delivery (average rating: 2.9). This indicates a positive correlation between providing delivery and customer satisfaction. Enhancing delivery services could be strategic for improving overall customer experience and ratings.

DATA VISUALIZATION

Restaurant outlets where delivery are available
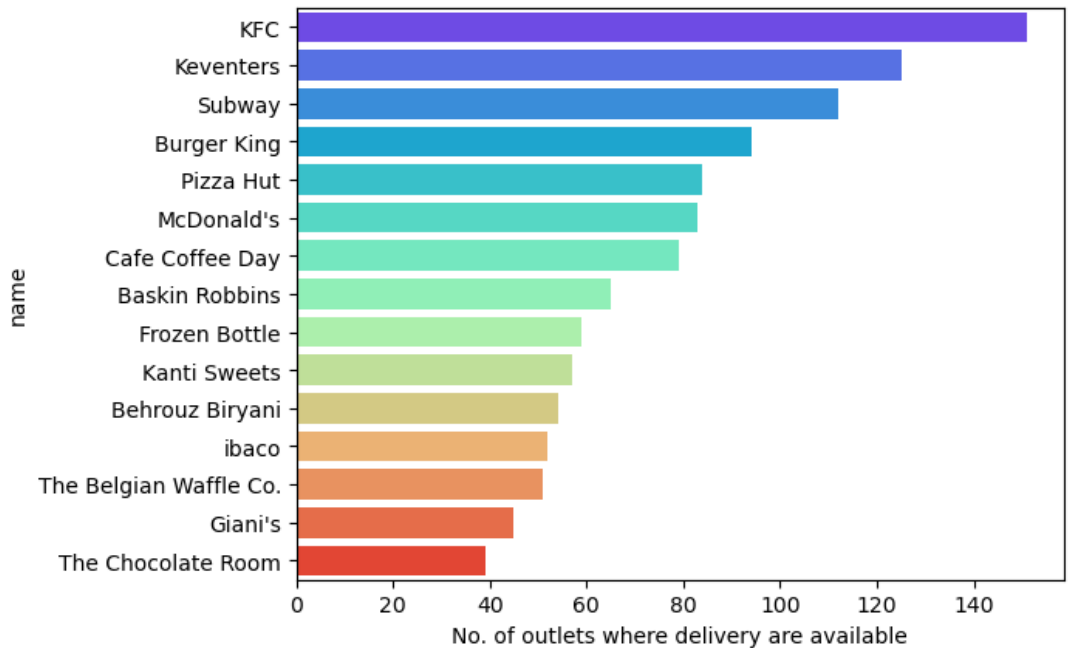
The bar plot shows the top 15 restaurants with the most outlets offering delivery services. It helps identify key outlets contributing to widespread delivery availability. This information is valuable for businesses to strategically optimize and promote delivery services, focusing on outlets with the highest impact.

The opening and closing timings to identify Peak hours
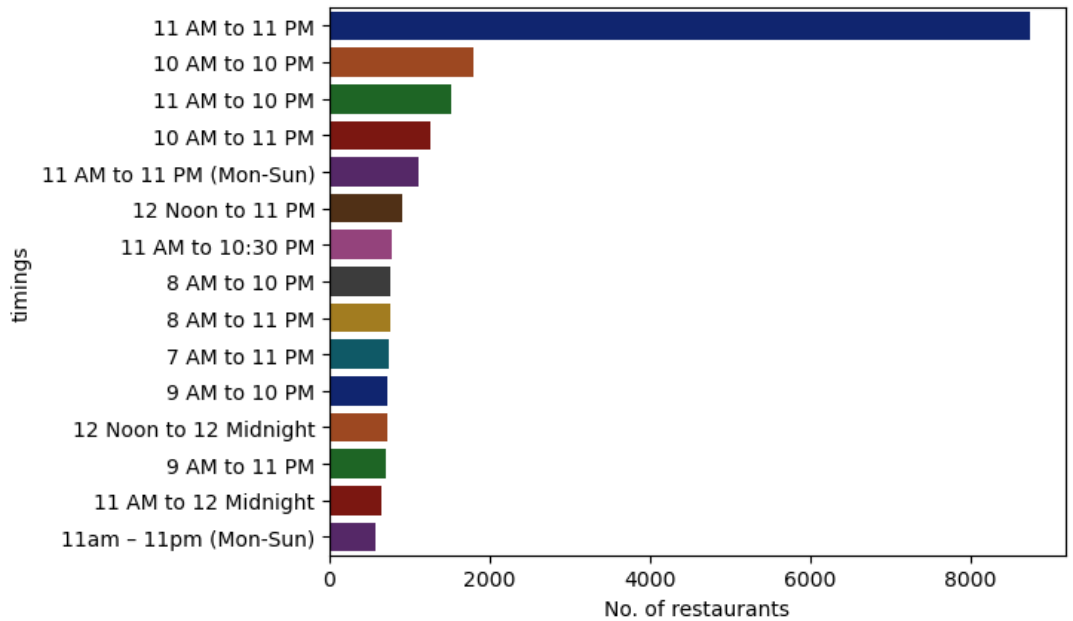
The bar plot illustrates the top 15 restaurant opening and closing timings, providing a quick overview of peak hours. This information is valuable for businesses to efficiently manage staff, allocate resources, and optimize services during high-demand periods. It aids in strategic decision-making for restaurant operations based on the popularity of specific timings.

DATA VISUALIZATION

Frequently mentioned words by customers
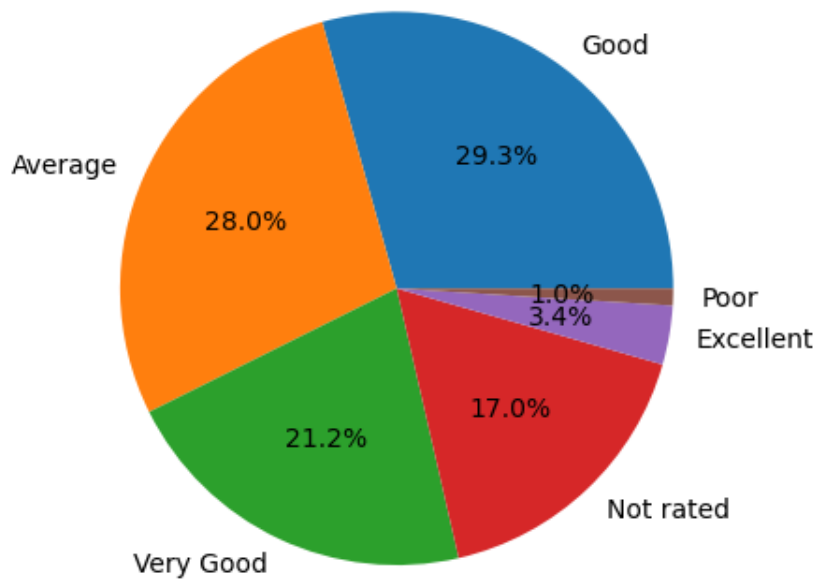
The pie chart illustrates the distribution of customer sentiments in reviews, categorizing them into sentiments like 'Good,' 'Average,' 'Very Good,' 'Not rated,' 'Excellent,' and 'Poor.' The chart's percentages provide a quick overview of the prevalence of each sentiment, helping businesses understand the overall sentiment landscape derived from customer feedback. This concise visualization is valuable for businesses to grasp the sentiments expressed by customers and tailor their strategies for service improvement accordingly. restaurants, such as outdoor seating, delivery options, and credit card acceptance. Businesses can leverage this information to understand customer preferences and tailor their services to meet popular demands.

1.Enhance Delivery Services:- Restaurants with delivery options tend to have higher ratings. Encourage and support restaurants to improve and expand their delivery services, ensuring a seamless and positive customer experience.

2.Focus on Peak Hours:- Identify and optimize resources during peak hours, especially for restaurants operating from 11 AM to 11 PM. Efficient staff management and service optimization during busy periods can enhance customer satisfaction.

3.Leverage Popular Features:- Highlight popular features such as outdoor seating, delivery, and credit card acceptance. Encourage restaurants to emphasize and promote these features to attract a broader customer base.

4.Improve Quality in Low-Rated Chains:- For chains with lower average ratings, consider implementing quality improvement initiatives. Analyze customer reviews to identify specific areas for enhancement and prioritize actions to boost overall satisfaction.

5.Engage with Customer Reviews:- Actively engage with customer reviews to understand specific feedback and address concerns promptly. Building a positive and responsive online presence can contribute to improved customer perceptions.

6.Strategic Marketing for Top Chains:- Leverage the popularity of top restaurant chains based on outlets for strategic marketing. Highlight the extensive presence of these chains to attract more customers and reinforce brand visibility.

7.Optimize Online Ordering:- Recognize the positive correlation between online ordering availability and higher ratings. Encourage and support restaurants in optimizing their online ordering systems for amore user-friendly experience.

8.Cuisine-Specific Strategies:- Tailor strategies based on the most popular cuisines in specific restaurants. Consider unique promotions, events, or menu additions that align with the preferences of the local customer base.

CONCLUSION

In conclusion, the analysis of the Zomato dataset has yielded valuable insights into various aspects of the restaurant industry, catering to both business and user perspectives. Key findings include:

1. Top Restaurant Chains and Ratings:

Identified the top restaurant chains based on outlet count and their corresponding average ratings. Barbeque Nation emerged as the highest-rated among the top chains.

2. Customer Sentiments:

Explored customer sentiments through reviews, creating a word cloud and pie chart to visually represent frequently mentioned words and sentiments. 'Good' and 'Average' were predominant sentiments, constituting a significant portion of reviews.

3.Delivery Services and Ratings:

Established a positive correlation between the availability of delivery services and higher ratings, suggesting the importance of efficient and reliable delivery options. Peak Hours and Operational

CONCLUSION

4. Peak Hours and Operational Timings:

Analyzed peak hours and operational timings, providing insights for businesses to strategically manage resources during busy periods, especially in the 11 AM to 11 PM timeframe.

5. Cuisine Preferences:

Explored the most popular cuisines and recommended tailoring strategies based on cuisine-specific insights to enhance customer satisfaction.

6. Photo Counts and Brand Visibility:

Investigated the impact of photo counts on brand visibility, highlighting the significance of creating visually appealing experiences to attract customers.

7. Platform and User Engagement:

Provided actionable recommendations for both restaurants and the platform to optimize services, engage with customer feedback, and enhance overall user experiences.

This analysis effectively addresses the project objectives by offering a comprehensive understanding of the Zomato dataset.

THANKS FOR READING

For Code
https://github.com/rishabhjain0310/Zomato-Data-Analysis