Title : Zomato Bengaluru dataset analysis

Name: Bhavya V

Reg No: P18BR23S126018

Department: MCA (BCU)

Semester:Third

Objective:

The objective of a case study based on Zomato's Bengaluru dataset analysis would typically involve extracting meaningful insights from the data to understand patterns, trends, and behaviors related to the food delivery and restaurant industry in Bengaluru. Here are some key objectives that could be part of such a case study:

1. **Restaurant Performance Analysis**: Analyze the performance of various restaurants in Bengaluru, including factors like average rating, price range, location, and cuisine type. This can help in identifying the most popular or successful restaurant categories.
2. **Customer Preferences**: Examine the preferences of customers, such as the types of cuisines they prefer, the time of day they order, and their spending behavior. This can help businesses cater better to their target audience.
3. **Location-based Insights**: Study the correlation between restaurant success and their geographic location within Bengaluru. Factors like proximity to business hubs, residential areas, or tourist spots could impact restaurant performance.
4. **Pricing Strategy**: Investigate how restaurant pricing impacts customer ratings and restaurant success. Does a higher price correlate with higher ratings, or is it better to have a more affordable pricing strategy?
5. **Sentiment Analysis of Reviews**: Use the available reviews and ratings to perform sentiment analysis and understand the customer satisfaction level. This can also highlight areas where restaurants can improve.
6. **Competition Analysis**: Identify key competitors in different locations and categories, and analyze their strengths and weaknesses in comparison to each other.
7. **Trends Over Time**: Investigate trends in the dataset, such as how customer ratings or the number of orders have changed over time, or if certain factors like special offers have an impact on demand.
8. **Business Recommendations**: Based on the findings from the analysis, propose strategic recommendations for restaurant owners, such as optimal pricing, marketing, and location strategies, to maximize growth and customer satisfaction.

The goal of the case study would be to derive actionable insights from data that can help in decision-making processes for businesses in the food service industry, particularly those operating within the Zomato ecosystem in Bengaluru.

Problem statement:

With the growing competition in the food delivery and restaurant industry in Bengaluru, restaurant owners and businesses need to make data-driven decisions to optimize their performance and improve customer satisfaction. However, there is a lack of comprehensive insights into the factors that influence restaurant success, customer preferences, and overall market trends in Bengaluru.

This case study aims to address the following key challenges:

1. **Performance Evaluation**: Understanding the performance of restaurants across different neighborhoods, cuisines, and price ranges to identify key drivers of success.
2. **Customer Behavior**: Analyzing customer preferences, including cuisine choices, average spending, and ordering patterns, to enable restaurants to cater better to customer needs.
3. **Location-Based Insights**: Identifying location-specific factors that contribute to restaurant popularity, helping restaurants target optimal locations for better business outcomes.
4. **Sentiment Analysis**: Assessing customer sentiment based on reviews and ratings to highlight areas where restaurants can improve their offerings and customer experiences.
5. **Competitor Benchmarking**: Understanding the competitive landscape within Bengaluru and identifying opportunities for restaurants to differentiate themselves based on performance and customer feedback.

The objective of this analysis is to leverage Zomato's Bengaluru dataset to extract actionable insights that will help restaurant owners and businesses develop effective pricing, marketing, and operational strategies, leading to improved customer satisfaction and business growth.

This problem statement identifies the core areas to be addressed and the purpose of analyzing the dataset. The insights gained from solving these problems can support better business decisions in the highly competitive food delivery market.

Solution:

The **solution** to the Zomato Bengaluru dataset analysis case study would involve a structured approach to addressing the key problems identified in the problem statement. Here's a potential solution framework:

1. Restaurant Performance Evaluation

Approach:

- **Data Analysis**: Analyze the restaurant data based on key metrics like **ratings**, **price range**, **cuisine types**, and **location**.
- **Grouping & Segmentation**: Categorize the restaurants into various segments based on their performance, such as "High Performers", "Moderate Performers", and "Low Performers."
- **Visualization**: Use **bar charts**, **scatter plots**, and **heatmaps** to show correlations between factors like price range and rating.
- **Insights**: Identify which factors are most closely correlated with higher ratings and customer satisfaction (e.g., higher ratings for certain cuisines or areas).

Outcome: This will help identify what makes a restaurant successful in Bengaluru, guiding businesses to improve their offerings accordingly.

2. Customer Behavior Analysis

Approach:

- **Trend Analysis**: Analyze **order patterns** over time to identify busy hours, peak days, and seasonal variations.
- **Cuisines Preference**: Investigate which cuisines are most popular, and whether this varies by time of day or location.
- **Customer Spend**: Analyze how average spending varies with customer ratings, price ranges, and cuisines.
- **Customer Demographics (if available)**: Investigate whether certain types of customers (e.g., families, young professionals) prefer certain types of food or services.

Outcome: This will help restaurants understand **customer preferences** better, allowing them to tailor their offerings (e.g., introducing popular cuisines or adjusting their menu pricing).

3. Location-Based Insights

Approach:

- **Geospatial Analysis**: Using the restaurant addresses or coordinates, map restaurants in Bengaluru to identify areas with higher concentrations of high-performing restaurants.
- **Proximity Analysis**: Analyze the relationship between restaurant success and factors like proximity to business districts, residential areas, and transport hubs.
- **Clustering**: Use **k-means clustering** to identify high-demand areas and pinpoint the best locations for new restaurant openings.

Outcome: Identify areas of opportunity for new restaurants or improvements in existing locations, enabling restaurants to expand strategically.

4. Sentiment Analysis of Reviews

Approach:

- **Text Mining**: Perform **sentiment analysis** on customer reviews to gauge the overall customer satisfaction and pinpoint areas for improvement.
- **Word Cloud & N-grams**: Use visual tools like **word clouds** or **n-grams** to find common phrases or keywords in positive and negative reviews.
- **Categorization**: Group common sentiments (e.g., "delivery time," "food quality," "service") to give restaurants actionable feedback on what customers appreciate or dislike.

Outcome: By identifying frequent pain points and positive feedback, restaurants can improve their operations (e.g., reducing delivery times, improving food quality).

5. Competitor Benchmarking

Approach:

- **Comparative Analysis**: Compare restaurant performance across various metrics (e.g., ratings, pricing, customer reviews) to understand the competitive landscape in Bengaluru.
- **SWOT Analysis**: For key competitors, conduct a SWOT analysis (Strengths, Weaknesses, Opportunities, Threats) based on their performance metrics.
- **Differentiation Strategy**: Identify areas where competitors excel and where they fall short, helping businesses find **differentiation opportunities** (e.g., offering unique cuisines, improving customer service).

Outcome: Restaurants can use these insights to identify competitive advantages and areas where they need to improve to stay ahead of the competition.

6. Trend Analysis Over Time

Approach:

- **Time Series Analysis**: Use time series analysis to study how ratings, orders, and customer sentiment have evolved over time. Identify any seasonal trends or patterns in customer behavior.
- **Impact of Promotions**: Analyze the impact of promotions (e.g., discounts, special offers) on customer ratings and order volume to determine the effectiveness of marketing strategies.

Outcome: This will allow restaurants to plan for **seasonal demand fluctuations** and optimize marketing strategies.

7. Business Recommendations

Based on the analysis, the following actionable recommendations can be provided:

- **Optimized Pricing Strategy**: Suggest price adjustments based on customer demand and competitor pricing, helping restaurants improve profitability.
- **Menu Customization**: Recommend adding or removing items based on popular cuisines and customer feedback.
- **Location Strategy**: Advise on the best areas for new restaurant openings, based on geospatial and competitor performance data.
- **Marketing and Promotions**: Recommend promotional strategies (e.g., discounts, limited-time offers) based on successful past trends.

Tools and Techniques

- **Data Cleaning & Preprocessing**: Handle missing data, incorrect entries, and outliers to ensure the analysis is accurate.
- **Statistical Techniques**: Use regression analysis, correlation matrices, and ANOVA tests to understand the relationships between variables.
- **Visualization Tools**: Leverage libraries like **Matplotlib**, **Seaborn**, and **Tableau** for creating clear, insightful visualizations.
- **Sentiment Analysis**: Implement Natural Language Processing (NLP) techniques using libraries like **NLTK** or **TextBlob** to analyze reviews.

Conclusion: By using data analytics techniques to explore the Zomato Bengaluru dataset, businesses can gain valuable insights into their operations, customer behaviors, and the competitive landscape. These insights can then be used to make data-driven decisions for enhancing restaurant performance, improving customer satisfaction, and driving business growth.

Implemented code:

```
#Importing Necessary Libraries
import numpy as np
import pandas as pd
import matplotlib
import seaborn as sns
from matplotlib import pyplot as plt
%matplotlib inline


#Reading dataset
df = pd.read_csv("E:/zomato.csv")
df.head()


#Explore the data
df.columns
df.shape
df.isnull().sum()
df.info()
df.duplicated().sum()

# Data Cleaning
df = df.drop(['url', 'phone', 'menu_item', 'reviews_list'], axis = 'columns')
df.shape
df = df.drop(['dish_liked'], axis = 'columns')
df.shape
df.rate.unique()
def rate_extractor(x):
    x = str(x)
    if (x == 'NEW' or x == '-'):
        return 0
    tokens = x.split('/')
    return str(tokens[0]).strip(' ')
df['rate'] = df['rate'].apply(rate_extractor)
df.head()
df.rate.unique()
df['rate'] = df['rate'].fillna(0)
df['rate'] = df['rate'].astype(float)
df.dropna(inplace = True)
df.isnull().sum()
df['approx_cost(for two people)'].unique()
def cost_extractor(x):
    x = str(x)
    tokens = x.split(',')
    cost = ''
    for i in tokens:
        cost += i
    return int(cost)
```

```
df['approx_cost(for two people)'] = df['approx_cost(for two people)'].apply(cost_extractor)
```

#Data visualization

1.Restaurant delivering online or not

```
df['online_order'].value_counts().plot(kind = 'pie', autopct = '%1.1f%%')
plt.ylabel('Percentage of Order Type')
Text(0, 0.5, 'Percentage of Order Type')
```

2.Restaurants allowing table booking or not

```
df['book_table'].value_counts().plot(kind = 'pie', autopct = '%1.1f%%')
plt.ylabel('Percentage for Table Booking Allowance')
Text(0, 0.5, 'Percentage for Table Booking Allowance')
```

3. Best Location:

```
df['listed_in(city)'].value_counts().plot(kind = 'bar')
plt.xlabel('City')
plt.ylabel('Count')
plt.show()
```

4.Relation between location and rating:

```
rate_location_group = df.groupby('location')['rate'].mean().reset_index()
rate_location_group = rate_location_group.sort_values(by = 'rate', ascending = False)
rate_location_group
plt.figure(figsize = (18, 16))

sns.barplot(data = rate_location_group, x = 'location', y = 'rate')
plt.xlabel('Location')
plt.ylabel('Average Rate')
plt.xticks(rotation = 90)
plt.tight_layout()
plt.show()
```

5. Restaurant type

```
df['rest_type'].value_counts()
df['rest_type'].value_counts()[:20].plot(kind = 'bar')
plt.xlabel('Restaurant Type')
plt.ylabel('Count')
plt.show()
```

6. Types of services:

```
df['listed_in(type)'].value_counts().plot(kind = 'bar')
plt.xlabel('Restaurant Service Type')
plt.ylabel('Count')
plt.show()
```

7.Relation between service types and ratings:

```
rate_service_type_group = df.groupby('listed_in(type)')['rate'].mean().reset_index()
rate_service_type_group = rate_service_type_group.sort_values(by = 'rate', ascending = False)
```

```
rate_service_type_group
sns.barplot(data = rate_service_type_group, x = 'listed_in(type)', y = 'rate')
plt.xticks(rotation = 45)
plt.xlabel('Restaurant Service Type')
plt.ylabel('Average Rate')
plt.tight_layout()
plt.show()
```

8. Cost of restaurant:

```
df['approx_cost(for two people)'].describe()
```

 # Cost Vs Online order

```
sns.barplot(data = df, x = 'online_order', y = 'approx_cost(for two people)')
plt.xlabel('Online Order Type')
plt.ylabel('Average Cost')
plt.show()
```

 # Cost Vs Book Table

```
sns.barplot(data = df, x = 'book_table', y = 'approx_cost(for two people)')
plt.xlabel('Book Table Type')
plt.ylabel('Average Cost')
plt.show()
```

 # Cost Vs Rate

```
sns.barplot(data = df, x = 'rate', y = 'approx_cost(for two people)')
plt.xlabel('Rate')
plt.ylabel('Average Cost')
plt.xticks(rotation = 90)
plt.tight_layout()
plt.show()
```

 # Cost Vs Location:

```
cost_location_group = df.groupby('location')['approx_cost(for two people)'].mean().reset_index()
cost_location_group = cost_location_group.sort_values(by = 'approx_cost(for two people)',
ascending = False)
cost_location_group
plt.figure(figsize = (18, 16))
sns.barplot(data = cost_location_group, x = 'location', y = 'approx_cost(for two people)')
plt.xlabel('Location')
plt.ylabel('Average Cost')
plt.xticks(rotation = 90)
plt.tight_layout()
plt.show()
```

#Cost Vs Restaurant type

```
cost_rest_type_group = df.groupby('rest_type')['approx_cost(for two people)'].mean().reset_index()
cost_rest_type_group = cost_rest_type_group.sort_values(by = 'approx_cost(for two people)',
ascending = False)
cost_rest_type_group = cost_rest_type_group.head(40)
cost_rest_type_group.shape
```

 # Cost Vs Service type

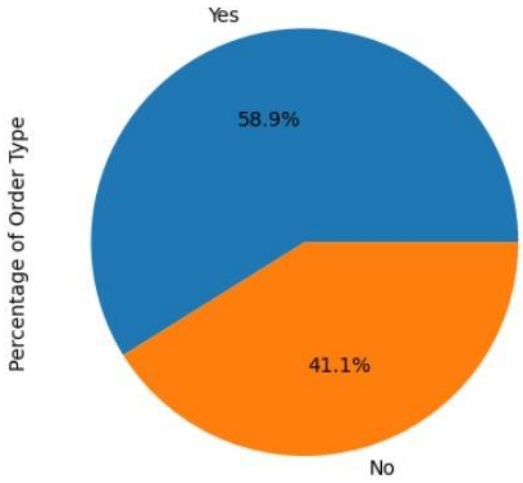
```
plt.figure(figsize = (18, 16))
```

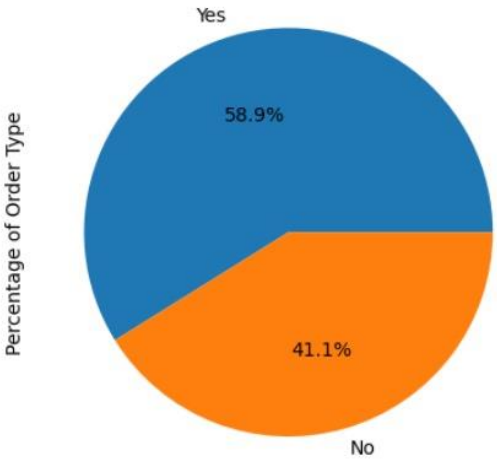
```
sns.barplot(data = cost_rest_type_group, x = 'rest_type', y = 'approx_cost(for two people)')
plt.xlabel('Restaurant Type')
plt.ylabel('Average Cost')
plt.xticks(rotation = 90)
plt.tight_layout()
plt.show()
```

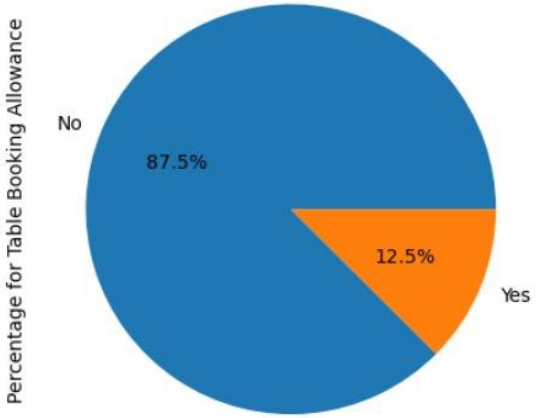
9. Number of restaurants in a location

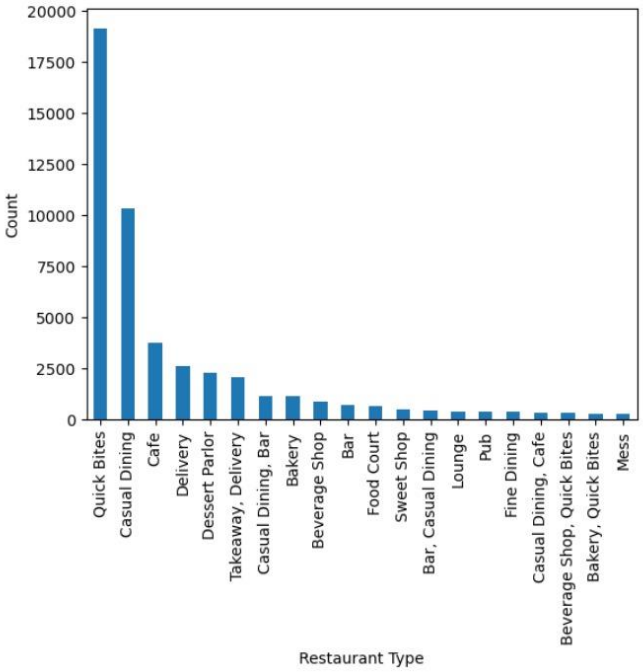
```
location_rest_group = df.groupby('listed_in(city)')['name'].nunique().reset_index()
location_rest_group
sns.barplot(data = location_rest_group, x = 'listed_in(city)', y = 'name')
plt.xlabel('Location')
plt.ylabel('No. of Restaurants')
plt.xticks(rotation = 90)
plt.tight_layout()
plt.show()
```

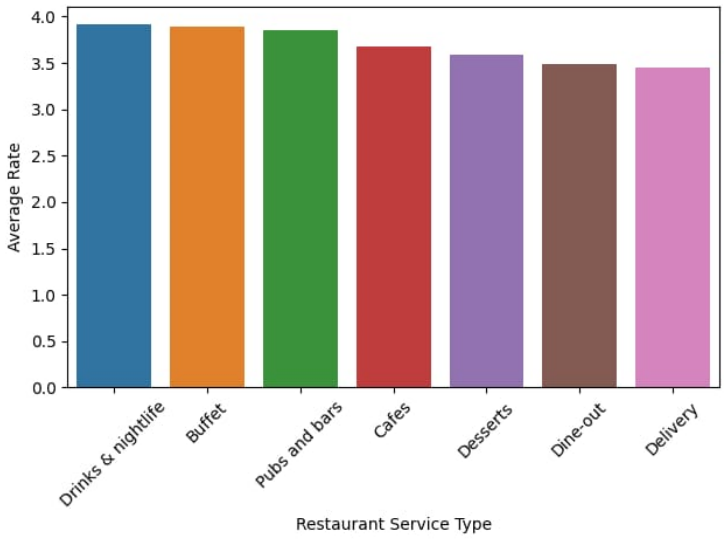
Output:

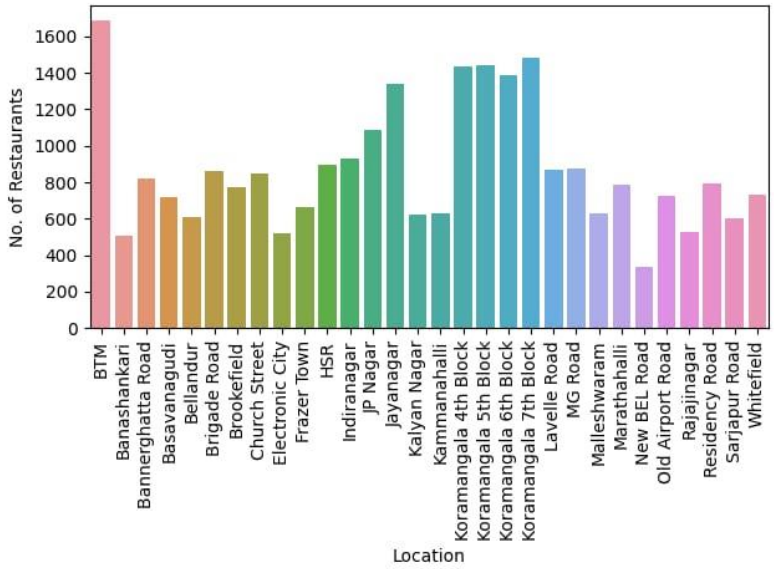





Percentage for Table Booking Allowance
No
87.5%
12.5%
Yes


Count
20000
17500
15000
12500
10000
7500
5000
2500
0
Quick Bites
Casual Dining
Cafe
Delivery
Dessert Parlor
Takeaway, Delivery
Casual Dining, Bar
Bakery
Beverage Shop
Bar
Food Court
Sweet Shop
Bar, Casual Dining
Lounge
Pub
Fine Dining
Casual Dining, Cafe
Beverage Shop, Quick Bites
Bakery, Quick Bites
Mess
Restaurant Type


Average Rate
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Drinks & nightlife
Buffet
Pubs and bars
Cafes
Desserts
Dine-out
Delivery
Restaurant Service Type


No. of Restaurants
0
200
400
600
800
1000
1200
1400
1600
BTM
Banashankari
Bannerghatta Road
Basavanagudi
Bellandur
Brigade Road
Brookefield
Church Street
Electronic City
Frazer Town
HSR
Indiranagar
JP Nagar
Jayanagar
Kalyan Nagar
Kammanahalli
Koramangala 4th Block
Koramangala 5th Block
Koramangala 6th Block
Koramangala 7th Block
Lavelle Road
MG Road
Malleshwaram
Marathahalli
New BEL Road
Old Airport Road
Rajajinagar
Residency Road
Sarjapur Road
Whitefield
Location