Exploratory Data Analysis (EDA)

Dataset: **Zomato Dataset**

Tools: Pandas, Matplotlib, Seaborn

1. Load Dataset

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load dataset
df = pd.read_csv("zomato.csv", encoding="latin-1")
print("Shape of dataset:", df.shape)
df.head()
```

2. Basic Info & Summary

```
print("\nDataset Info:")
df.info()

print("\nStatistical Summary:")
display(df.describe(include="all").T)

print("\nMissing Values:")
print(df.isnull().sum())
```

```
Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9551 entries, 0 to 9550
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Restaurant ID         9551 non-null   int64
 1   Restaurant Name       9551 non-null   object
 2   Country Code          9551 non-null   int64
 3   City                  9551 non-null   object
 4   Address               9551 non-null   object
 5   Locality              9551 non-null   object
 6   Locality Verbose      9551 non-null   object
 7   Longitude             9551 non-null   float64
 8   Latitude              9551 non-null   float64
 9   Cuisines              9542 non-null   object
 10  Average Cost for two  9551 non-null   int64
```

```
 11  Currency              9551 non-null   object
 12  Has Table booking     9551 non-null   object
 13  Has Online delivery   9551 non-null   object
 14  Is delivering now     9551 non-null   object
 15  Switch to order menu  9551 non-null   object
 16  Price range           9551 non-null   int64
 17  Aggregate rating      9551 non-null   float64
 18  Rating color          9551 non-null   object
 19  Rating text           9551 non-null   object
 20  Votes                 9551 non-null   int64
dtypes: float64(3), int64(5), object(13)
memory usage: 1.5+ MB

Statistical Summary:

                       count unique                          top  freq
\
Restaurant ID         9551.0    NaN                          NaN   NaN

Restaurant Name         9551   7446             Cafe Coffee Day    83

Country Code          9551.0    NaN                          NaN   NaN

City                    9551    141                    New Delhi  5473

Address                 9551   8918  Dilli Haat, INA, New Delhi    11

Locality                9551   1208             Connaught Place   122

Locality Verbose        9551   1265  Connaught Place, New Delhi   122

Longitude             9551.0    NaN                          NaN   NaN

Latitude              9551.0    NaN                          NaN   NaN

Cuisines                9542   1825                North Indian   936

Average Cost for two  9551.0    NaN                          NaN   NaN

Currency                9551     12          Indian Rupees(Rs.)  8652

Has Table booking       9551      2                          No  8393

Has Online delivery     9551      2                          No  7100

Is delivering now       9551      2                          No  9517

Switch to order menu    9551      1                          No  9551

Price range           9551.0    NaN                          NaN   NaN
```

Aggregate rating	9551.0	NaN		NaN	NaN
Rating color	9551	6		Orange	3737
Rating text	9551	6		Average	3737
Votes	9551.0	NaN		NaN	NaN

	mean	std	min	25% \
Restaurant ID	9051128.349178	8791521.282104	53.0	301962.5
Restaurant Name	NaN	NaN	NaN	NaN
Country Code	18.365616	56.750546	1.0	1.0
City	NaN	NaN	NaN	NaN
Address	NaN	NaN	NaN	NaN
Locality	NaN	NaN	NaN	NaN
Locality Verbose	NaN	NaN	NaN	NaN
Longitude	64.126574	41.467058	-157.948486	77.081343
Latitude	25.854381	11.007935	-41.330428	28.478713
Cuisines	NaN	NaN	NaN	NaN
Average Cost for two	1199.210763	16121.183073	0.0	250.0
Currency	NaN	NaN	NaN	NaN
Has Table booking	NaN	NaN	NaN	NaN
Has Online delivery	NaN	NaN	NaN	NaN
Is delivering now	NaN	NaN	NaN	NaN
Switch to order menu	NaN	NaN	NaN	NaN
Price range	1.804837	0.905609	1.0	1.0
Aggregate rating	2.66637	1.516378	0.0	2.5
Rating color	NaN	NaN	NaN	NaN
Rating text	NaN	NaN	NaN	

```
NaN
Votes                     156.909748      430.169145          0.0
5.0

                                50%          75%          max
Restaurant ID           6004089.0  18352291.5  18500652.0
Restaurant Name               NaN         NaN         NaN
Country Code                  1.0         1.0       216.0
City                          NaN         NaN         NaN
Address                       NaN         NaN         NaN
Locality                      NaN         NaN         NaN
Locality Verbose              NaN         NaN         NaN
Longitude               77.191964   77.282006  174.832089
Latitude                28.570469   28.642758    55.97698
Cuisines                      NaN         NaN         NaN
Average Cost for two        400.0       700.0    800000.0
Currency                      NaN         NaN         NaN
Has Table booking             NaN         NaN         NaN
Has Online delivery           NaN         NaN         NaN
Is delivering now             NaN         NaN         NaN
Switch to order menu          NaN         NaN         NaN
Price range                   2.0         2.0         4.0
Aggregate rating              3.2         3.7         4.9
Rating color                  NaN         NaN         NaN
Rating text                   NaN         NaN         NaN
Votes                        31.0       131.0     10934.0


Missing Values:
Restaurant ID             0
Restaurant Name           0
Country Code              0
City                      0
Address                   0
Locality                  0
Locality Verbose          0
Longitude                 0
Latitude                  0
Cuisines                  9
Average Cost for two      0
Currency                  0
Has Table booking         0
Has Online delivery       0
Is delivering now         0
Switch to order menu      0
Price range               0
Aggregate rating          0
Rating color              0
Rating text               0
```

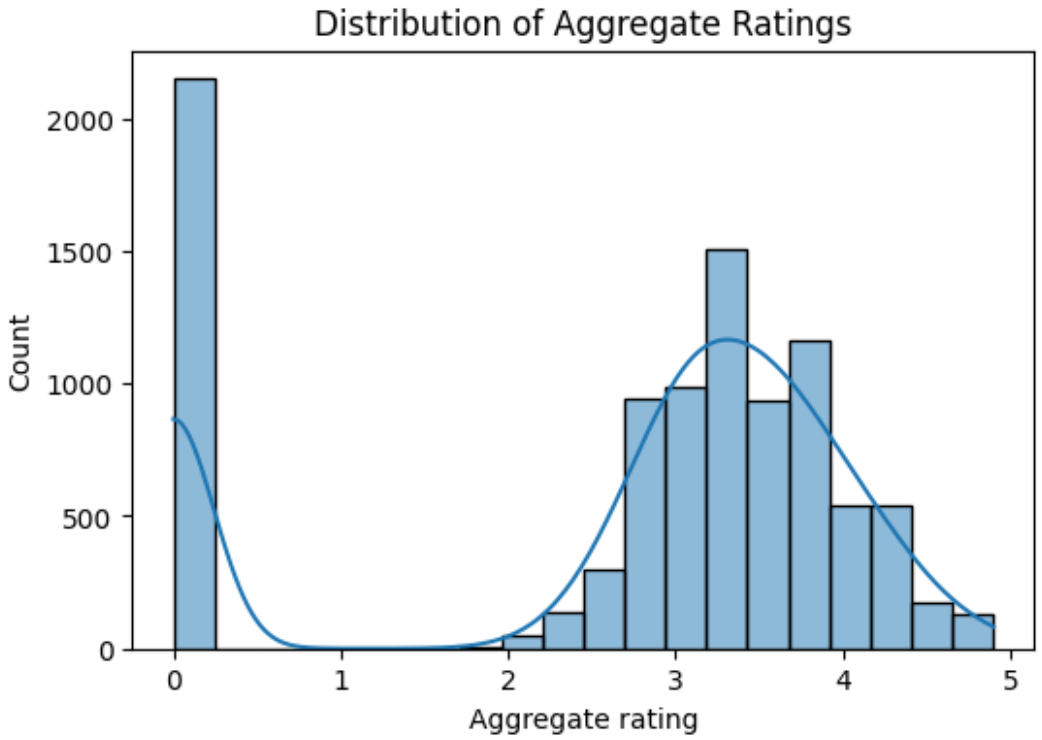
```
Votes                 0
dtype: int64
```

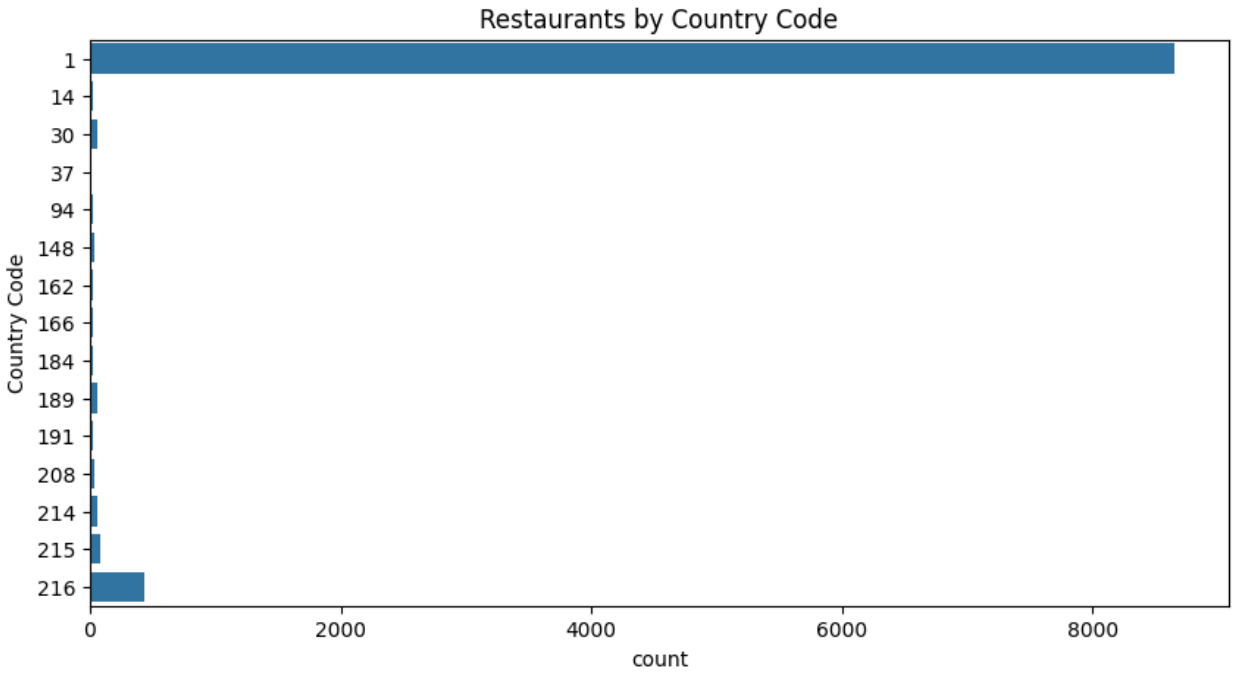
3. Univariate Analysis

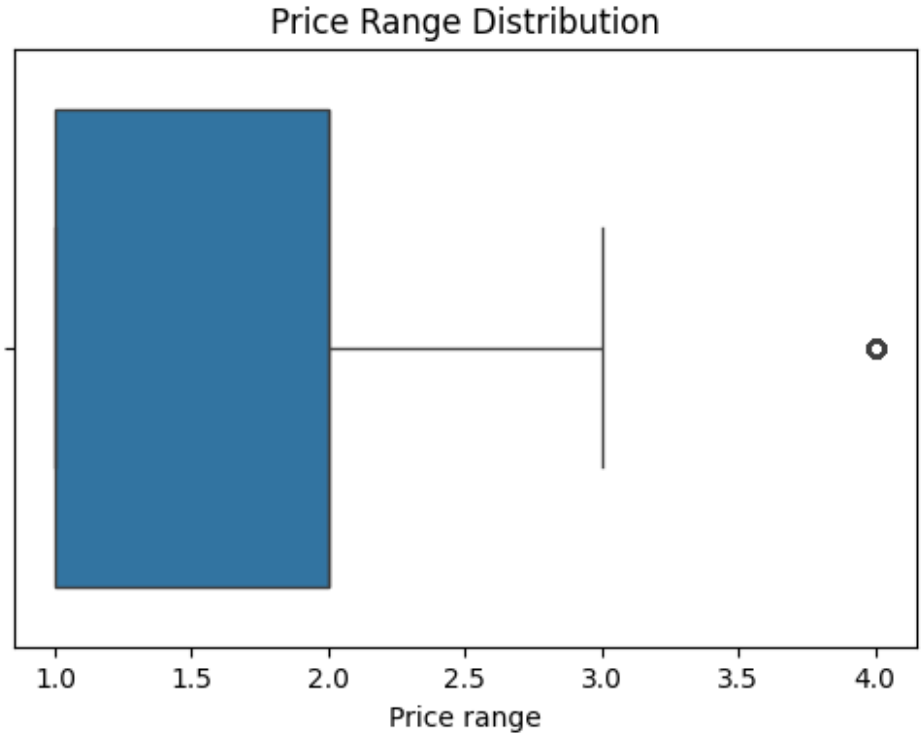
```
# Distribution of ratings
plt.figure(figsize=(6,4))
sns.histplot(df['Aggregate rating'], bins=20, kde=True)
plt.title("Distribution of Aggregate Ratings")
plt.show()

# Countplot for Country Code
plt.figure(figsize=(10,5))
sns.countplot(y=df['Country Code'])
plt.title("Restaurants by Country Code")
plt.show()

# Boxplot for Price Range
plt.figure(figsize=(6,4))
sns.boxplot(x=df['Price range'])
plt.title("Price Range Distribution")
plt.show()
```





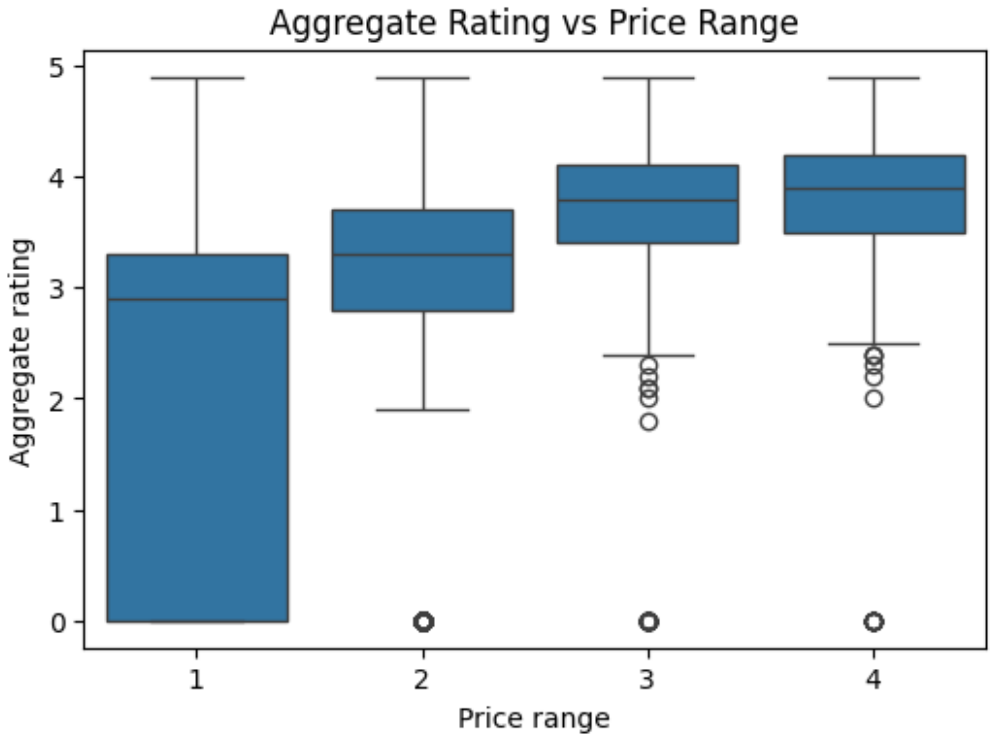


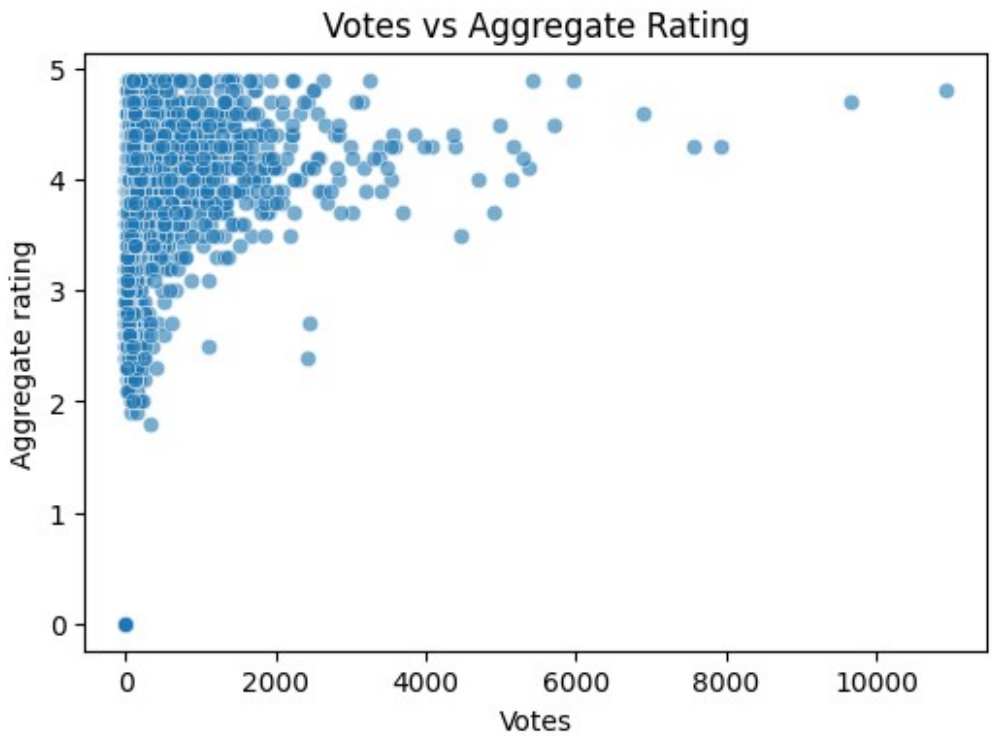
4. Bivariate Analysis

```
# Rating vs Price Range
plt.figure(figsize=(6,4))
sns.boxplot(x="Price range", y="Aggregate rating", data=df)
```

```
plt.title("Aggregate Rating vs Price Range")
plt.show()

# Votes vs Rating
plt.figure(figsize=(6,4))
sns.scatterplot(x="Votes", y="Aggregate rating", data=df, alpha=0.6)
plt.title("Votes vs Aggregate Rating")
plt.show()
```

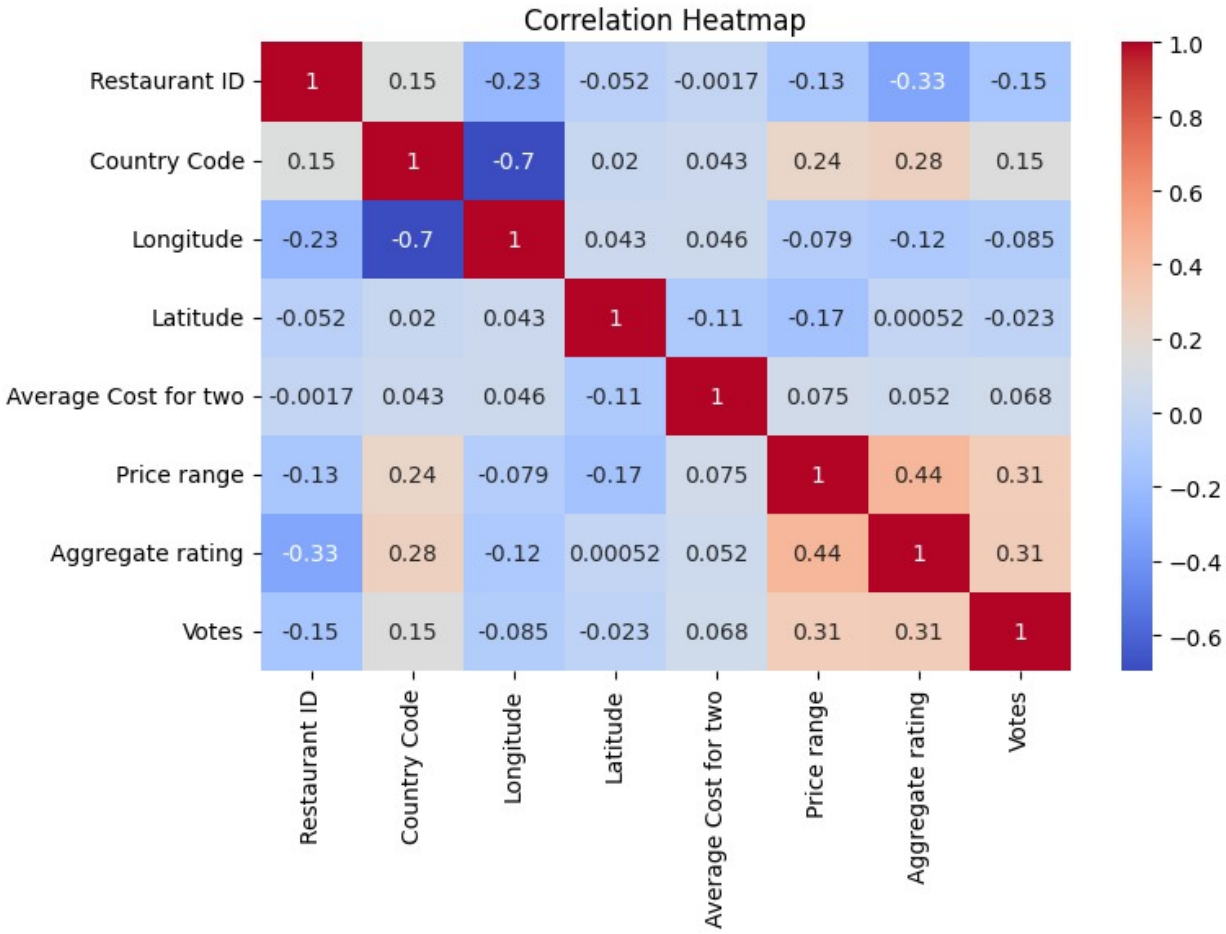


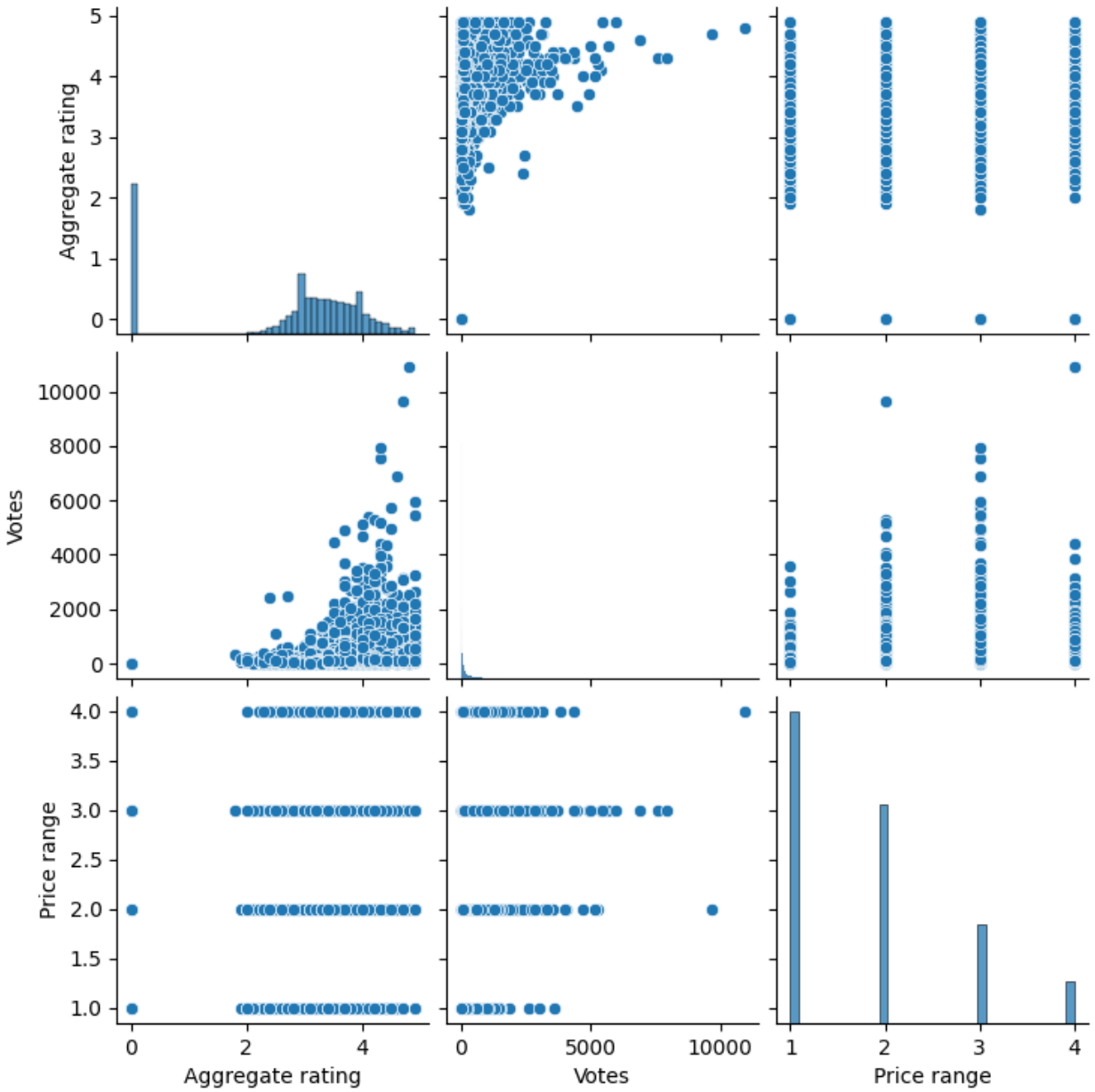


5. Correlation & Heatmap

```
plt.figure(figsize=(8,5))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()

# Pairplot for key numeric columns
sns.pairplot(df[["Aggregate rating", "Votes", "Price range"]])
plt.show()
```


Correlation Heatmap
Restaurant ID
Country Code
Longitude
Latitude
Average Cost for two
Price range
Aggregate rating
Votes
1 0.15 -0.23 -0.052 -0.0017 -0.13 -0.33 -0.15
0.15 1 -0.7 0.02 0.043 0.24 0.28 0.15
-0.23 -0.7 1 0.043 0.046 -0.079 -0.12 -0.085
-0.052 0.02 0.043 1 -0.11 -0.17 0.00052 -0.023
-0.0017 0.043 0.046 -0.11 1 0.075 0.052 0.068
-0.13 0.24 -0.079 -0.17 0.075 1 0.44 0.31
-0.33 0.28 -0.12 0.00052 0.052 0.44 1 0.31
-0.15 0.15 -0.085 -0.023 0.068 0.31 0.31 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

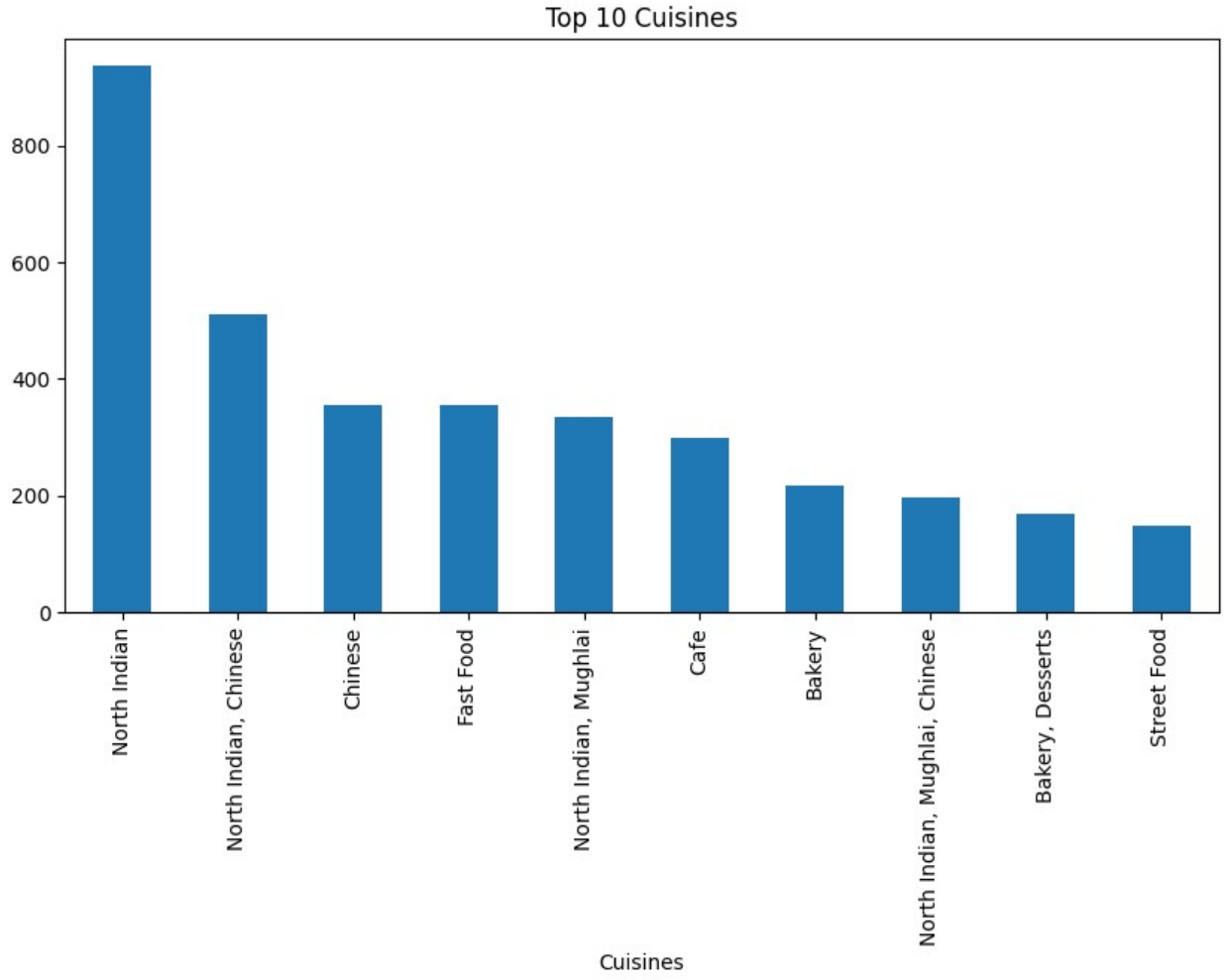


6. Categorical Insights

```
# Top 10 Cuisines
plt.figure(figsize=(10,5))
df['Cuisines'].value_counts().head(10).plot(kind='bar')
plt.title("Top 10 Cuisines")
plt.show()

# Online delivery vs Rating
plt.figure(figsize=(6,4))
sns.boxplot(x="Has Online delivery", y="Aggregate rating", data=df)
```

```
plt.title("Online Delivery vs Ratings")
plt.show()
```



7. Observations

- Most restaurants fall in mid-range ratings (3.0–4.0)
- Votes show moderate positive correlation with ratings
- Indian cuisine is among the most common
- Online delivery impacts ratings distribution
- Price range has weaker influence on ratings