

```
In [1]: import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [9]: file_path = ('D:/cognifyz/Dataset .csv')  
df = pd.read_csv(file_path)
```

```
In [10]: print("Number of rows and columns:", df.shape)  
print("First few rows of the dataset:\n", df.head())
```

Number of rows and columns: (9551, 21)

First few rows of the dataset:

	Restaurant ID	Restaurant Name	Country Code	City \
0	6317637	Le Petit Souffle	162	Makati City
1	6304287	Izakaya Kikufuji	162	Makati City
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City
3	6318506	Ooma	162	Mandaluyong City
4	6314302	Sambo Kojin	162	Mandaluyong City

	Address \
0	Third Floor, Century City Mall, Kalayaan Avenu...
1	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...
2	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...
3	Third Floor, Mega Fashion Hall, SM Megamall, O...
4	Third Floor, Mega Atrium, SM Megamall, Ortigas...

	Locality \
0	Century City Mall, Poblacion, Makati City
1	Little Tokyo, Legaspi Village, Makati City
2	Edsa Shangri-La, Ortigas, Mandaluyong City
3	SM Megamall, Ortigas, Mandaluyong City
4	SM Megamall, Ortigas, Mandaluyong City

	Locality Verbose	Longitude	Latitude \
0	Century City Mall, Poblacion, Makati City, Mak...	121.027535	14.565443
1	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101	14.553708
2	Edsa Shangri-La, Ortigas, Mandaluyong City, Ma...	121.056831	14.581404
3	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.056475	14.585318
4	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.057508	14.584450

	Cuisines ...	Currency	Has Table booking \
0	French, Japanese, Desserts ...	Botswana Pula(P)	Yes
1	Japanese ...	Botswana Pula(P)	Yes
2	Seafood, Asian, Filipino, Indian ...	Botswana Pula(P)	Yes
3	Japanese, Sushi ...	Botswana Pula(P)	No
4	Japanese, Korean ...	Botswana Pula(P)	Yes

	Has Online delivery	Is delivering now	Switch to order menu	Price range \
0	No	No	No	3
1	No	No	No	3
2	No	No	No	4
3	No	No	No	4
4	No	No	No	4

	Aggregate rating	Rating color	Rating text	Votes
0	4.8	Dark Green	Excellent	314
1	4.5	Dark Green	Excellent	591
2	4.4	Green	Very Good	270
3	4.9	Dark Green	Excellent	365
4	4.8	Dark Green	Excellent	229

[5 rows x 21 columns]

```
In [13]: missing_values = df.isnull().sum()
print("Missing values in each column:\n", missing_values)
```

Missing values in each column:

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

```
In [15]: df['Cuisines'] = df['Cuisines'].fillna('Unknown')
missing_values_after = df.isnull().sum()
print("Missing values after handling:\n", missing_values_after)
```

Missing values after handling:

Restaurant ID	0
Restaurant Name	0
Country Code	0
City	0
Address	0
Locality	0
Locality Verbose	0
Longitude	0
Latitude	0
Cuisines	0
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

```
In [16]: print("Data types before conversion:\n", df.dtypes)
```

Data types before conversion:

Restaurant ID	int64
Restaurant Name	object
Country Code	int64
City	object
Address	object
Locality	object
Locality Verbose	object
Longitude	float64
Latitude	float64
Cuisines	object
Average Cost for two	int64
Currency	object
Has Table booking	object
Has Online delivery	object
Is delivering now	object
Switch to order menu	object
Price range	int64
Aggregate rating	float64
Rating color	object
Rating text	object
Votes	int64

dtype: object

```
In [18]: print("Data types after conversion:\n", df.dtypes)
```

Data types after conversion:

Restaurant ID	int64
Restaurant Name	object
Country Code	int64
City	object
Address	object
Locality	object
Locality Verbose	object
Longitude	float64
Latitude	float64
Cuisines	object
Average Cost for two	int64
Currency	object
Has Table booking	object
Has Online delivery	object
Is delivering now	object
Switch to order menu	object
Price range	int64
Aggregate rating	float64
Rating color	object
Rating text	object
Votes	int64

dtype: object

```
In [31]: target_variable = 'Aggregate rating'

# Set the aesthetic style of the plots
sns.set(style="whitegrid")

# Create the plot
```

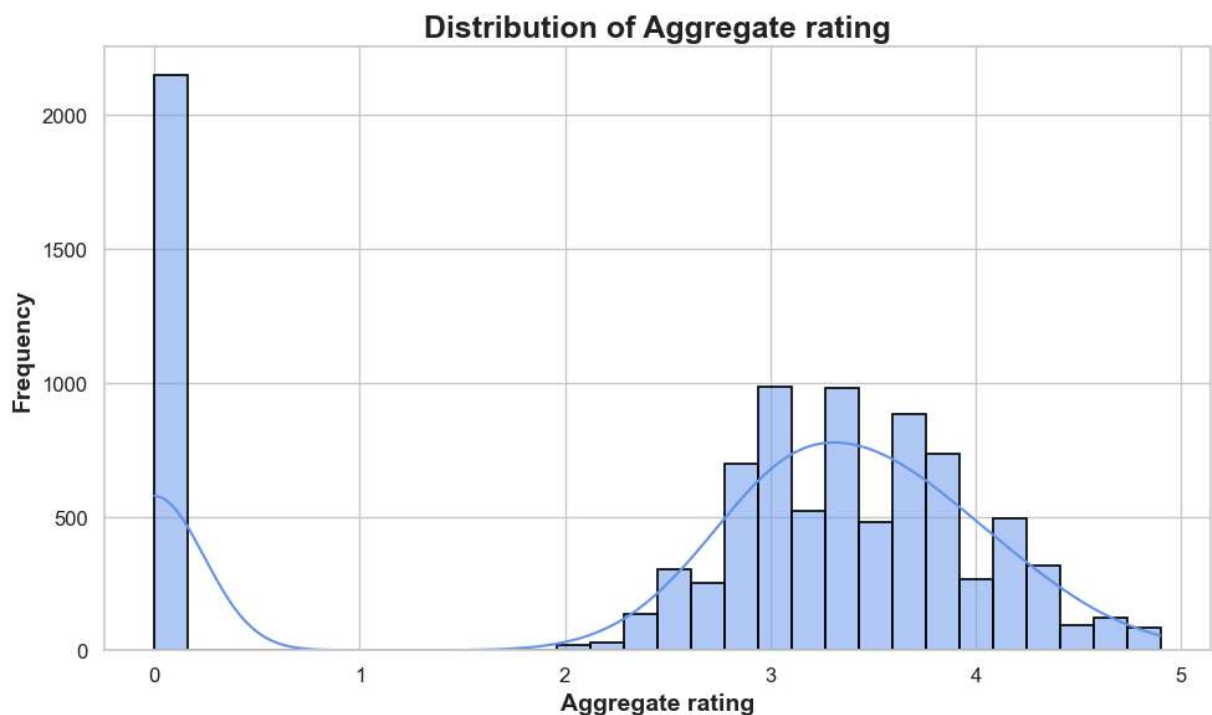
```
plt.figure(figsize=(10, 6))

# Plot histogram with KDE
sns.histplot(df[target_variable], kde=True, color='cornflowerblue', edgecolor='black')

# Add titles and labels
plt.title(f'Distribution of {target_variable}', fontsize=18, weight='bold')
plt.xlabel(target_variable, fontsize=14, weight='bold')
plt.ylabel('Frequency', fontsize=14, weight='bold')

# Customize ticks
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)

# Show the plot
plt.tight_layout()
plt.show()
```



```
In [20]: class_counts = df[target_variable].value_counts()
print("Class distribution:\n", class_counts)
```

Class distribution:

Aggregate rating

0.0 2148

3.2 522

3.1 519

3.4 498

3.3 483

3.5 480

3.0 468

3.6 458

3.7 427

3.8 400

2.9 381

3.9 335

2.8 315

4.1 274

4.0 266

2.7 250

4.2 221

2.6 191

4.3 174

4.4 144

2.5 110

4.5 95

2.4 87

4.6 78

4.9 61

2.3 47

4.7 42

2.2 27

4.8 25

2.1 15

2.0 7

1.9 2

1.8 1

Name: count, dtype: int64

In []: