Task: Data Exploration and Preprocessing

```
#importing the libraries
%matplotlib inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn

#importing the dataset
df = pd.read_csv(r"C:\Users\hp\Desktop\EDA\Dataset.csv")

#Explore the Dataset
num_rows, num_cols = df.shape
print(f"Number of rows: {num_rows}")
print(f"Number of columns: {num_cols}")
```

```
Number of rows: 9551
Number of columns: 21
```

```
df
```

```
      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
...             ...                       ...           ...
...
9546        5915730               Naml\ Gurme           208
��stanbul
9547        5908749              Ceviz A��ac\       208
��stanbul
9548        5915807                     Huqqa           208
��stanbul
9549        5916112               A���k Kahve           208
��stanbul
9550        5927402  Walter's Coffee Roastery           208
��stanbul
```

```
                                                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...
...                                                 ...
9546  Kemanke�� Karamustafa Pa��a Mahallesi, RıhtıM ...
9547  Ko��uyolu Mahallesi, Muhittin ��st�_nda�� Cadd...
9548  Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...
9549  Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...
9550  Cafea��a Mahallesi, Bademaltı Sokak, No 21/B, ...

                                        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
...                                          ...
9546                                    Karak�_y
9547                                   Ko��uyolu
9548                                 Kuru�_e��me
9549                                 Kuru�_e��me
9550                                        Moda

                                       Locality Verbose   Longitude  \
0     Century City Mall, Poblacion, Makati City, Mak...  121.027535
1     Little Tokyo, Legaspi Village, Makati City, Ma...  121.014101
2     Edsa Shangri-La, Ortigas, Mandaluyong City, Ma...  121.056831
3     SM Megamall, Ortigas, Mandaluyong City, Mandal...  121.056475
4     SM Megamall, Ortigas, Mandaluyong City, Mandal...  121.057508
...                                                 ...         ...
9546                                Karak�_y, ��stanbul   28.977392
9547                               Ko��uyolu, ��stanbul   29.041297
9548                             Kuru�_e��me, ��stanbul   29.034640
9549                             Kuru�_e��me, ��stanbul   29.036019
9550                                    Moda, ��stanbul   29.026016

       Latitude                          Cuisines  ...
Currency  \
0     14.565443        French, Japanese, Desserts  ...  Botswana
Pula(P)
1     14.553708                          Japanese  ...  Botswana
Pula(P)
2     14.581404  Seafood, Asian, Filipino, Indian  ...  Botswana
Pula(P)
3     14.585318                   Japanese, Sushi  ...  Botswana
Pula(P)
```

```
4     14.584450                  Japanese, Korean  ...  Botswana
Pula(P)
...         ...                               ...  ...               .
..
9546  41.022793                           Turkish  ...  Turkish
Lira(TL)
9547  41.009847   World Cuisine, Patisserie, Cafe  ...  Turkish
Lira(TL)
9548  41.055817            Italian, World Cuisine  ...  Turkish
Lira(TL)
9549  41.057979                   Restaurant Cafe  ...  Turkish
Lira(TL)
9550  40.984776                              Cafe  ...  Turkish
Lira(TL)

     Has Table booking Has Online delivery Is delivering now  \
0                  Yes                  No                No
1                  Yes                  No                No
2                  Yes                  No                No
3                   No                  No                No
4                  Yes                  No                No
...                ...                 ...               ...
9546                No                  No                No
9547                No                  No                No
9548                No                  No                No
9549                No                  No                No
9550                No                  No                No

     Switch to order menu Price range  Aggregate rating  Rating color
\
0                      No           3               4.8    Dark Green

1                      No           3               4.5    Dark Green

2                      No           4               4.4         Green

3                      No           4               4.9    Dark Green

4                      No           4               4.8    Dark Green

...                   ...         ...               ...           ...

9546                   No           3               4.1         Green

9547                   No           3               4.2         Green

9548                   No           4               3.7        Yellow

9549                   No           4               4.0         Green
```

```
9550                   No            2               4.0         Green


     Rating text Votes
0      Excellent   314
1      Excellent   591
2      Very Good   270
3      Excellent   365
4      Excellent   229
...          ...   ...
9546   Very Good   788
9547   Very Good  1034
9548        Good   661
9549   Very Good   901
9550   Very Good   591

[9551 rows x 21 columns]
```

```
X = df.iloc[:, :-1].values
y = df.iloc[:, -1].values

print(X)
```

```
[[6317637 'Le Petit Souffle' 162 ... 4.8 'Dark Green' 'Excellent']
 [6304287 'Izakaya Kikufuji' 162 ... 4.5 'Dark Green' 'Excellent']
 [6300002 'Heat - Edsa Shangri-La' 162 ... 4.4 'Green' 'Very Good']
 ...
 [5915807 'Huqqa' 208 ... 3.7 'Yellow' 'Good']
 [5916112 'A���k Kahve' 208 ... 4.0 'Green' 'Very Good']
 [5927402 "Walter's Coffee Roastery" 208 ... 4.0 'Green' 'Very Good']]
```

```
print(y)
```

```
[314 591 270 ... 661 901 591]
```

```
missing_values = df.isnull().sum()
print("Missing values per column:")
print(missing_values)
```

```
Missing values per 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
```

```
#Handling the missing values
df_clean = df.dropna()
```

```
#solution 1 : dropna
df1 = df.copy()
```

```
#summarize the shape of raw data
print("Before:",df1.shape)
#drop rows with missing values
df1.dropna(inplace=True)
#summarize the shape of the data with missing rows removed
print("After:",df1.shape)
```

```
Before: (9551, 21)
After: (9542, 21)
```

Solution 2 : Fillna

```
df2 = df.copy()
```

```
import warnings
warnings.filterwarnings('ignore')
```

```
#fill missing values with mean column values
df2.fillna(df2.mean(), inplace=True)
#count the number of NaN values in each column
print(df2.isnull().sum())

df2
```

```
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
```

```
      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   
...             ...                       ...           ...   
...   
9546        5915730               Namlı Gurme           208   
��stanbul   
9547        5908749              Ceviz A��acı       208   
��stanbul   
9548        5915807                     Huqqa           208   
��stanbul   
9549        5916112               A���k Kahve           208   
��stanbul   
9550        5927402  Walter's Coffee Roastery           208   
��stanbul   

                                                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...   
...                                                 ...   
9546  Kemanke�� Karamustafa Pa��a Mahallesi, Rıhtım ...
```

```
9547  Ko��uyolu Mahallesi, Muhittin ��st�_nda�� Cadd...
9548  Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...
9549  Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...
9550  Cafea��a Mahallesi, Bademalt\ Sokak, No 21/B, ...

                                        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
...                                          ...
9546                                    Karak�_y
9547                                   Ko��uyolu
9548                                 Kuru�_e��me
9549                                 Kuru�_e��me
9550                                        Moda

                                       Locality Verbose   Longitude  \
0     Century City Mall, Poblacion, Makati City, Mak...  121.027535
1     Little Tokyo, Legaspi Village, Makati City, Ma...  121.014101
2     Edsa Shangri-La, Ortigas, Mandaluyong City, Ma...  121.056831
3     SM Megamall, Ortigas, Mandaluyong City, Mandal...  121.056475
4     SM Megamall, Ortigas, Mandaluyong City, Mandal...  121.057508
...                                                 ...         ...
9546                                Karak�_y, ��stanbul   28.977392
9547                               Ko��uyolu, ��stanbul   29.041297
9548                             Kuru�_e��me, ��stanbul   29.034640
9549                             Kuru�_e��me, ��stanbul   29.036019
9550                                    Moda, ��stanbul   29.026016

       Latitude                          Cuisines  ...
Currency  \
0     14.565443        French, Japanese, Desserts  ...  Botswana
Pula(P)
1     14.553708                          Japanese  ...  Botswana
Pula(P)
2     14.581404  Seafood, Asian, Filipino, Indian  ...  Botswana
Pula(P)
3     14.585318                   Japanese, Sushi  ...  Botswana
Pula(P)
4     14.584450                  Japanese, Korean  ...  Botswana
Pula(P)
...         ...                               ...  ...                 .
..
9546  41.022793                           Turkish  ...  Turkish
Lira(TL)
9547  41.009847   World Cuisine, Patisserie, Cafe  ...  Turkish
Lira(TL)
```

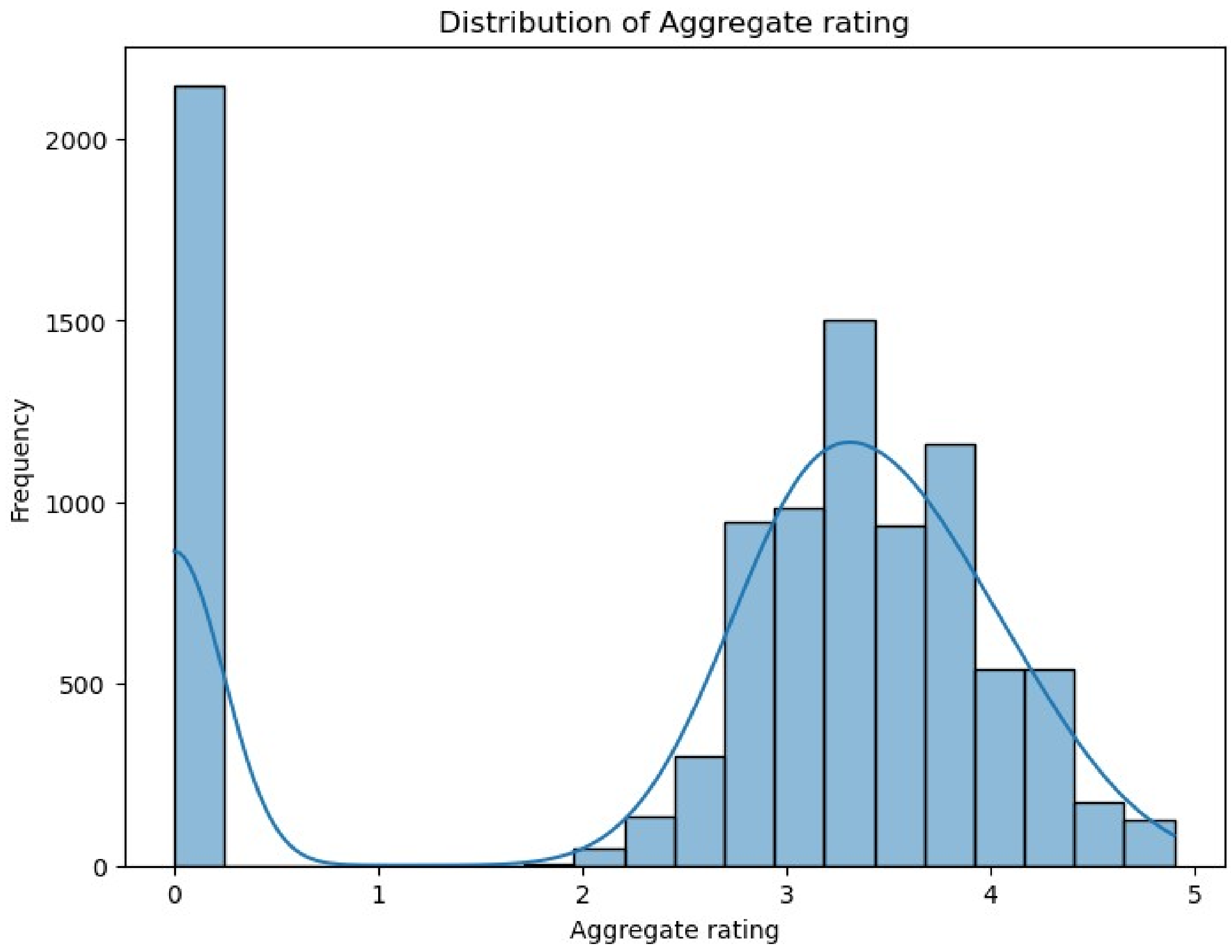
```
9548  41.055817           Italian, World Cuisine  ...  Turkish
Lira(TL)
9549  41.057979                  Restaurant Cafe  ...  Turkish
Lira(TL)
9550  40.984776                             Cafe  ...  Turkish
Lira(TL)

     Has Table booking Has Online delivery Is delivering now  \
0                  Yes                  No                No
1                  Yes                  No                No
2                  Yes                  No                No
3                   No                  No                No
4                  Yes                  No                No
...                ...                 ...               ...
9546                No                  No                No
9547                No                  No                No
9548                No                  No                No
9549                No                  No                No
9550                No                  No                No

     Switch to order menu Price range  Aggregate rating  Rating color
\
0                       No           3               4.8    Dark Green

1                       No           3               4.5    Dark Green

2                       No           4               4.4         Green

3                       No           4               4.9    Dark Green

4                       No           4               4.8    Dark Green

...                    ...         ...               ...           ...

9546                    No           3               4.1         Green

9547                    No           3               4.2         Green

9548                    No           4               3.7        Yellow

9549                    No           4               4.0         Green

9550                    No           2               4.0         Green


     Rating text Votes
0      Excellent   314
1      Excellent   591
2      Very Good   270
3      Excellent   365
4      Excellent   229
```

```
...             ...   ...
9546   Very Good   788
9547   Very Good  1034
9548        Good   661
9549   Very Good   901
9550   Very Good   591

[9551 rows x 21 columns]
```

```
#Analyze the Target Variable
target_summary = df['Aggregate rating'].describe()
print(target_summary)
```

```
count    9551.000000
mean        2.666370
std         1.516378
min         0.000000
25%         2.500000
50%         3.200000
75%         3.700000
max         4.900000
Name: Aggregate rating, dtype: float64
```

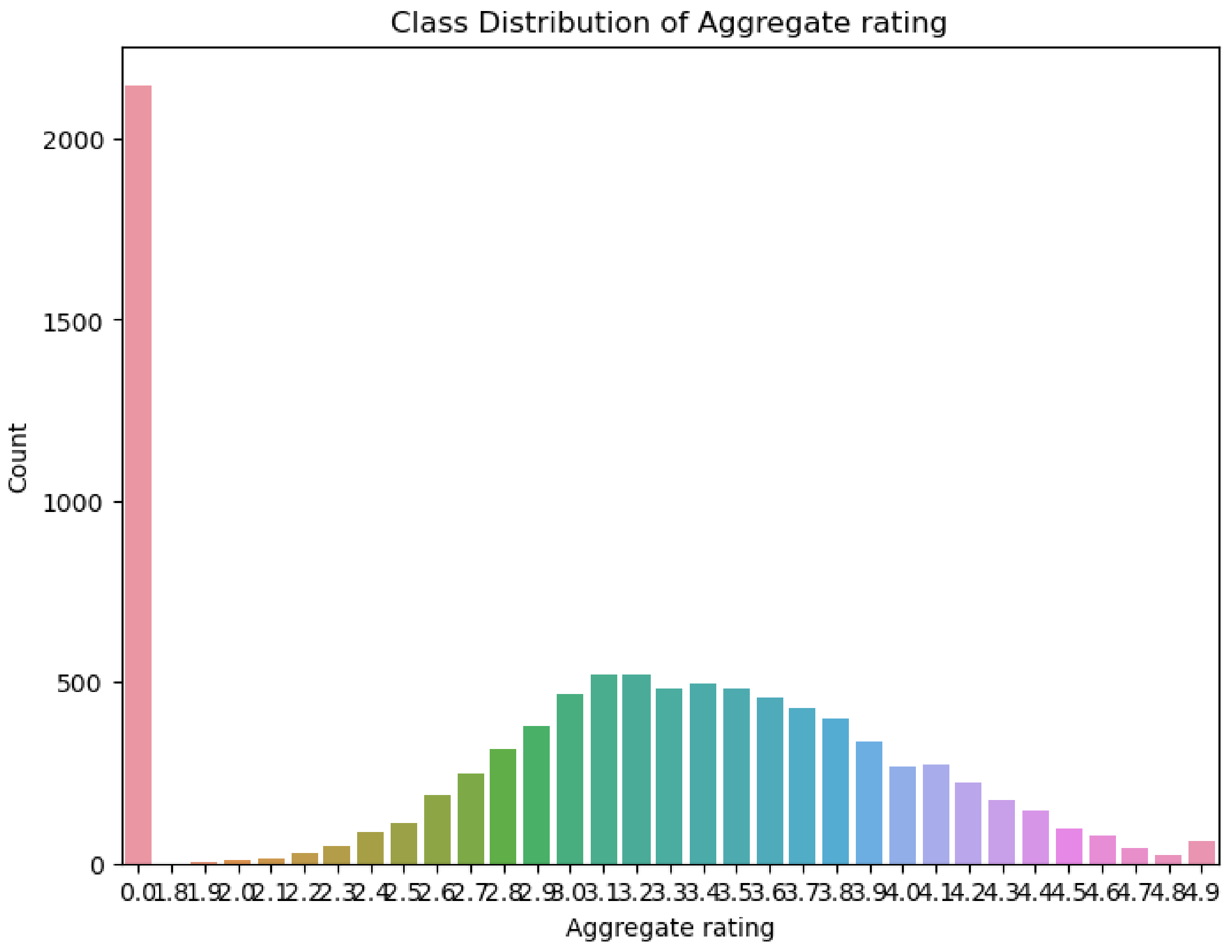
```
#plot the distribution
plt.figure(figsize=(8, 6))
sns.histplot(df['Aggregate rating'], bins=20, kde=True)
plt.xlabel('Aggregate rating')
plt.ylabel('Frequency')
plt.title('Distribution of Aggregate rating')
plt.show()
```



```
#Check for class imbalances
class_counts = df['Aggregate rating'].value_counts()
print("Class distribution")
print(class_counts)

Class distribution
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: Aggregate rating, dtype: int64
```

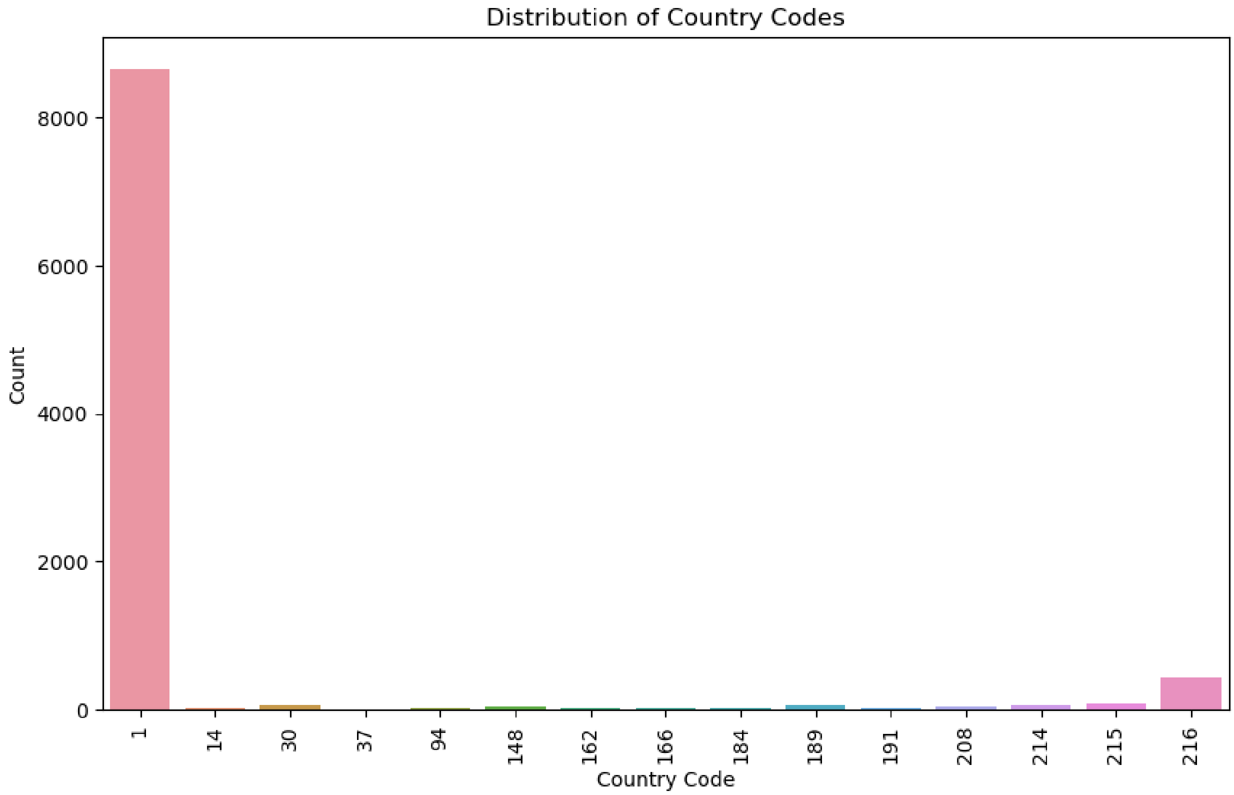
```
#Visualize class distribution
plt.figure(figsize=(8, 6))
sns.countplot(data=df, x='Aggregate rating')
plt.xlabel('Aggregate rating')
plt.ylabel('Count')
plt.title('Class Distribution of Aggregate rating')
plt.show()
```



Task : Descriptive Analysis

```
#Calculate mean, median, standard deviation, and more
numerical_stats = df.describe()
print(numerical_stats)

       Restaurant ID  Country Code    Longitude     Latitude  \
count   9.551000e+03   9551.000000  9551.000000  9551.000000
mean    9.051128e+06     18.365616    64.126574    25.854381
std     8.791521e+06     56.750546    41.467058    11.007935
min     5.300000e+01      1.000000  -157.948486   -41.330428
25%     3.019625e+05      1.000000    77.081343    28.478713
50%     6.004089e+06      1.000000    77.191964    28.570469
75%     1.835229e+07      1.000000    77.282006    28.642758
max     1.850065e+07    216.000000   174.832089    55.976980

       Average Cost for two  Price range  Aggregate rating
Votes
count           9551.000000  9551.000000       9551.000000
```

```
9551.000000
mean            1199.210763     1.804837          2.666370
156.909748
std            16121.183073     0.905609          1.516378
430.169145
min                0.000000     1.000000          0.000000
0.000000
25%              250.000000     1.000000          2.500000
5.000000
50%              400.000000     2.000000          3.200000
31.000000
75%              700.000000     2.000000          3.700000
131.000000
max           800000.000000     4.000000          4.900000
10934.000000

#Distribution of categorical variables
country_counts = df['Country Code'].value_counts()
print("Distribution of Country Codes:")
print(country_counts)

Distribution of Country Codes:
1      8652
216     434
215      80
30       60
214      60
189      60
148      40
208      34
14       24
162      22
94       21
184      20
166      20
191      20
37        4
Name: Country Code, dtype: int64

#Visualize the distribution
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='Country Code')
plt.xlabel('Country Code')
plt.ylabel('Count')
plt.title('Distribution of Country Codes')
plt.xticks(rotation=90)
plt.show()
```



```
#identifying the Top Cuisines and cities with the highest number of
restaurants
top_cuisines = df['Cuisines'].value_counts().head(10)
print("Top Cuisines:")
print(top_cuisines)

Top Cuisines:
North Indian                          936
North Indian, Chinese                 511
Chinese                               354
Fast Food                             354
North Indian, Mughlai                 334
Cafe                                  299
Bakery                                218
North Indian, Mughlai, Chinese        197
Bakery, Desserts                      170
Street Food                           149
Name: Cuisines, dtype: int64

top_cities = df['City'].value_counts().head(10)
print("Top Cities:")
print(top_cities)

Top Cities:
New Delhi         5473
Gurgaon           1118
```

```
Noida           1080
Faridabad        251
Ghaziabad         25
Bhubaneshwar      21
Amritsar          21
Ahmedabad         21
Lucknow           21
Guwahati          21
Name: City, dtype: int64
```

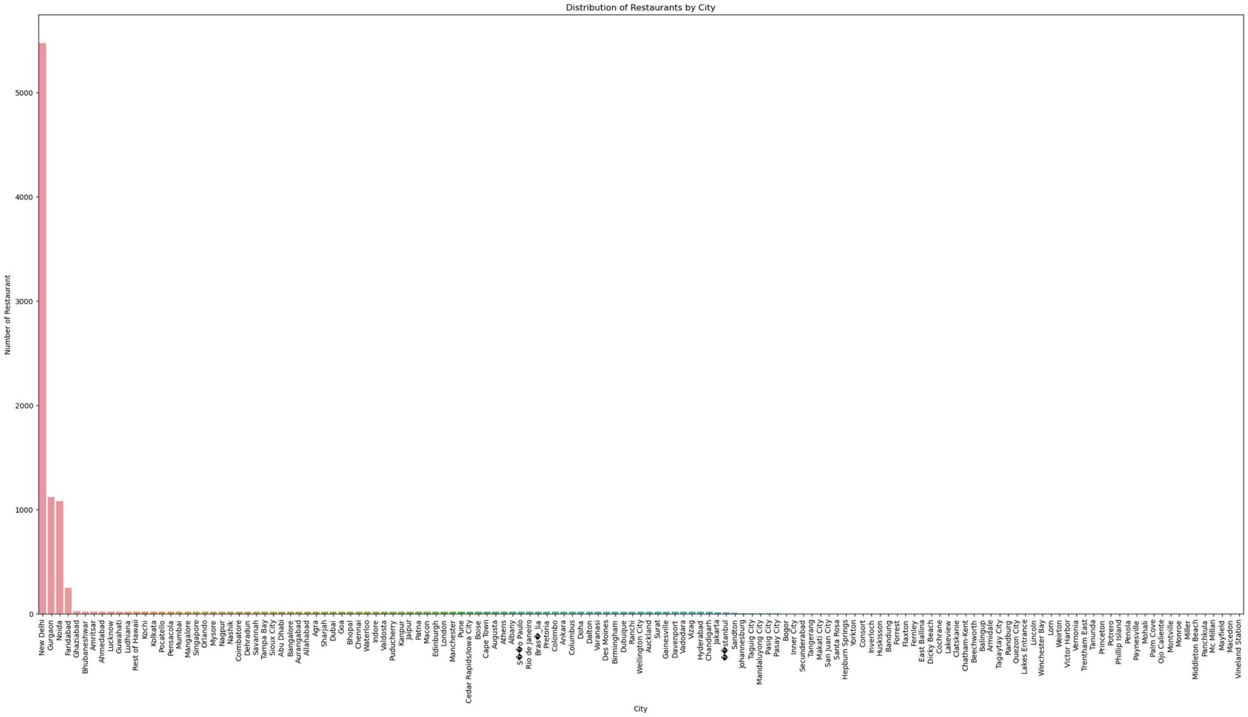
Task: Geospatial Analysis

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

average_latitude = df['Latitude'].mean()
average_longitude = df['Longitude'].mean()

#Create a map centered at a specific location
m = folium.Map(location=[average_latitude, average_longitude],
zoom_start=10)

#Add markers for each restaurants using latitude and longitude
for index, row in df.iterrows():
    folium.Marker([row['Latitude'], row['Longitude']],
popup=row['Restaurant Name']).add_to(m)

#Display the map
m.save('restaurant_locations.html')

#Analyze the Distribution of Restaurants
city_distribution = df['City'].value_counts()
print("Distribution of Restaurants by City:")
print(city_distribution)
```

```
Distribution of Restaurants by City:
New Delhi           5473
Gurgaon             1118
Noida               1080
Faridabad            251
Ghaziabad             25
                    ...
Panchkula              1
Mc Millan              1
Mayfield               1
Macedon                1
```

```
Vineland Station       1
Name: City, Length: 141, dtype: int64
```

```
#Visualize the distribution
plt.figure(figsize=(30, 15))
sns.barplot(x=city_distribution.index, y=city_distribution.values)
plt.xlabel('City')
plt.ylabel('Number of Restaurant')
plt.title('Distribution of Restaurants by City')
plt.xticks(rotation=90)
plt.show()
```



```
#Determine Correlation between location and rating
correlation = df[['Latitude', 'Longitude', 'Aggregate rating']].corr()
print("Correlation between Location and Rating:")
print(correlation)
```

```
Correlation between Location and Rating:
                  Latitude  Longitude  Aggregate rating
Latitude          1.000000   0.043207          0.000516
Longitude         0.043207   1.000000         -0.116818
Aggregate rating  0.000516  -0.116818          1.000000
```