Air Quality Data Analysis

This notebook analyzes global air quality index (AQI) data using Pandas, NumPy, and Matplotlib.

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
# Load the CSV file
df = pd.read_csv('data_date.csv')
df.head()
```

Out[2]:

	Date	Country	Status	AQI Value
0	2022-07-21	Albania	Good	14
1	2022-07-21	Algeria	Moderate	65
2	2022-07-21	Andorra	Moderate	55
3	2022-07-21	Angola	Unhealthy for Sensitive Groups	113
4	2022-07-21	Argentina	Moderate	63

In [3]:
```
# Convert 'Date' column to datetime
df['Date'] = pd.to_datetime(df['Date'])
```

In [4]:
```
# Dataset information
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19189 entries, 0 to 19188
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype         
---  ------     --------------  -----         
 0   Date       19189 non-null  datetime64[ns]
 1   Country    19189 non-null  object        
 2   Status     19189 non-null  object        
 3   AQI Value  19189 non-null  int64         
dtypes: datetime64[ns](1), int64(1), object(2)
memory usage: 599.8+ KB
```

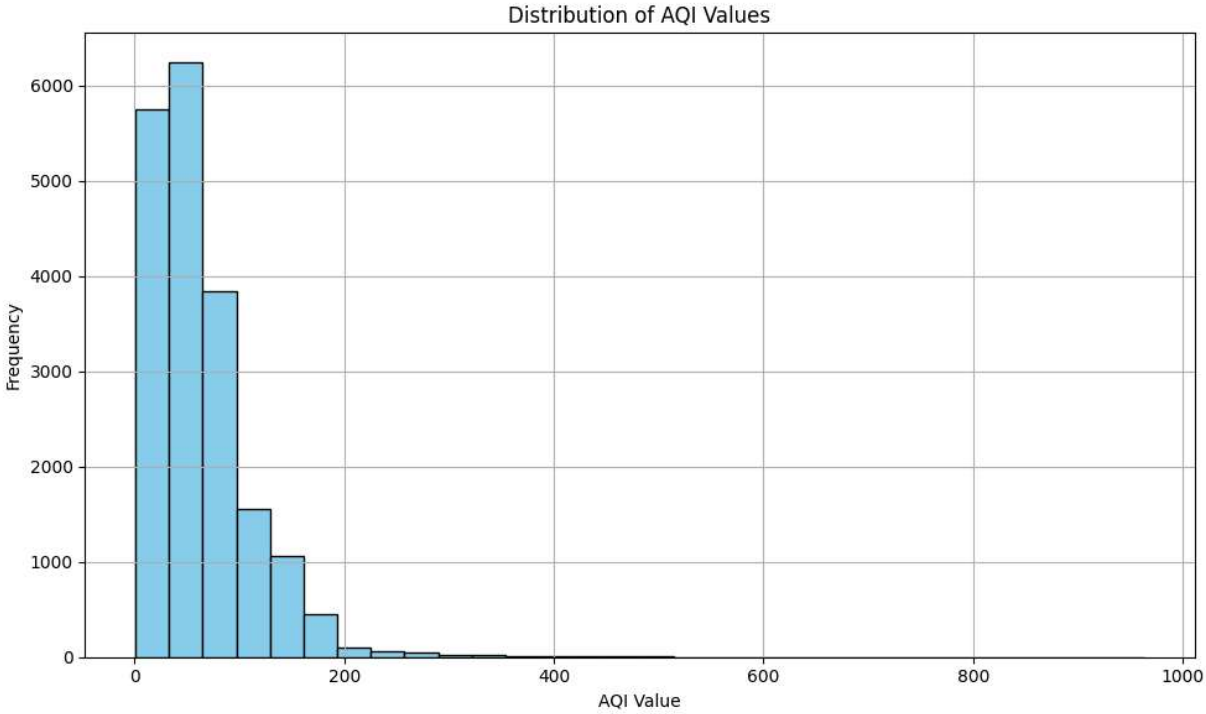
In [5]:
```
# AQI Summary statistics
df['AQI Value'].describe()
```

Out[5]:
```
count    19189.000000
mean        63.611653
std         50.525151
min          1.000000
25%         29.000000
50%         53.000000
75%         83.000000
max        963.000000
Name: AQI Value, dtype: float64
```

In [6]:
```
# Top 10 most polluted countries by average AQI
top_polluted = df.groupby('Country')['AQI Value'].mean().sort_values(ascending=Fals
top_polluted
```

Out[6]:
```
Country
India         220.493056
China         178.013889
Qatar         156.404255
Iraq          155.134921
Iran          151.447552
Bangladesh    150.770370
Uganda        132.052239
Ethiopia      131.843284
Bahrain       126.553719
Gabon         124.055556
Name: AQI Value, dtype: float64
```

In [7]:
```
# AQI Distribution Plot
plt.figure(figsize=(10, 6))
plt.hist(df['AQI Value'], bins=30, color='skyblue', edgecolor='black')
plt.title('Distribution of AQI Values')
plt.xlabel('AQI Value')
plt.ylabel('Frequency')
plt.grid(True)
plt.tight_layout()
plt.savefig('aqi_distribution.png')
plt.show()
```


Distribution of AQI Values
Frequency
AQI Value
0
1000
2000
3000
4000
5000
6000
0
200
400
600
800
1000