AIR QUALITY INDEX FORECASTING

Air Quality Index (AQI) is a tool for effective communication of air quality status to people can easily understand and take action. The AQI is used by agencies to communicate to the public how polluted the air currently is or how polluted it is forecast to become. Public health risks increase as the AQI rises.

to import the important python libraries for analysis

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

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

Get the data from the renewable resources

```
air=pd.read_csv('/content/AQI and Lat Long of Countries.csv')
air.fillna(0)
```

	Country	City	AQI Value	AQI Category	CO AQI Value	CO AQI Category	Ozone AQI Value	Ozone AQI Category	NO2 AQI Value	NO2 AQI Category	PM2.5 AQI Value	Pl Cate
0	Russian Federation	Praskoveya	51	Moderate	1	Good	36	Good	0	Good	51	Mode
1	Brazil	Presidente Dutra	41	Good	1	Good	5	Good	1	Good	41	(
2	Brazil	Presidente Dutra	41	Good	1	Good	5	Good	1	Good	41	(
3	Italy	Priolo Gargallo	66	Moderate	1	Good	39	Good	2	Good	66	Mode
4	Poland	Przasnysz	34	Good	1	Good	34	Good	0	Good	20	(
...	...	...	...	...	...	...	...	...	...	...	...	
16690	United States of America	Highland Springs	54	Moderate	1	Good	34	Good	5	Good	54	Mode
16691	Slovakia	Martin	71	Moderate	1	Good	39	Good	1	Good	71	Mode
16692	Slovakia	Martin	71	Moderate	1	Good	39	Good	1	Good	71	Mode
16693	France	Sceaux	50	Good	1	Good	20	Good	5	Good	50	(
16694	United States of America	Westerville	71	Moderate	1	Good	44	Good	2	Good	71	Mode

16695 rows × 14 columns

Find the head and tail of the dataset

```
air.head(3)
```

	Country	City	AQI Value	AQI Category	CO AQI Value	CO AQI Category	Ozone AQI Value	Ozone AQI Category	NO2 AQI Value	NO2 AQI Category	PM2.5 AQI Value	PM2.5 AQI Category
0	Russian Federation	Praskoveya	51	Moderate	1	Good	36	Good	0	Good	51	Moderate
1	Brazil	Presidente Dutra	41	Good	1	Good	5	Good	1	Good	41	Good
2	Brazil	Presidente	41	Good	1	Good	5	Good	1	Good	41	Good

```
air.tail(3)
```

	Country	City	AQI Value	AQI Category	CO AQI Value	CO AQI Category	Ozone AQI Value	Ozone AQI Category	NO2 AQI Value	NO2 AQI Category	PM2.5 AQI Value	PM2. AQ Categor
16692	Slovakia	Martin	71	Moderate	1	Good	39	Good	1	Good	71	Modera
16693	France	Sceaux	50	Good	1	Good	20	Good	5	Good	50	Goc
16694	United States of America	Westerville	71	Moderate	1	Good	44	Good	2	Good	71	Modera

To explore the data from the dataset

```
air.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16695 entries, 0 to 16694
Data columns (total 14 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   Country             16393 non-null  object 
 1   City                16695 non-null  object 
 2   AQI Value           16695 non-null  int64  
 3   AQI Category        16695 non-null  object 
 4   CO AQI Value        16695 non-null  int64  
 5   CO AQI Category     16695 non-null  object 
 6   Ozone AQI Value     16695 non-null  int64  
 7   Ozone AQI Category  16695 non-null  object 
 8   NO2 AQI Value       16695 non-null  int64  
 9   NO2 AQI Category    16695 non-null  object 
 10  PM2.5 AQI Value     16695 non-null  int64  
 11  PM2.5 AQI Category  16695 non-null  object 
 12  lat                 16695 non-null  float64
 13  lng                 16695 non-null  float64
dtypes: float64(2), int64(5), object(7)
memory usage: 1.8+ MB
```

To describe the data from the dataset

```
air.describe()
```

	AQI Value	CO AQI Value	Ozone AQI Value	NO2 AQI Value	PM2.5 AQI Value	lat	lng
count	16695.000000	16695.000000	16695.000000	16695.000000	16695.000000	16695.000000	16695.000000
std	43.091971	2.371379	22.839343	5.880677	43.208298	22.947398	73.037148

To find the duplicates values present in the dataset

```
air[air.duplicated()]
```

Country	City	AQI Value	AQI Category	CO AQI Value	CO AQI Category	Ozone AQI Value	Ozone AQI Category	NO2 AQI Value	NO2 AQI Category	PM2.5 AQI Value	PM2.5 AQI Category	lat	ln

There is no any duplicated values present in the dataset

To find the shape of the dataset

```
air.shape
```

```
(16695, 14)
```

The dataset having the 16695 rows and 14 columns present in the dataset

Get the columns and datatypes present in the dataset

```
air.columns
```

```
Index(['Country', 'City', 'AQI Value', 'AQI Category', 'CO AQI Value',
       'CO AQI Category', 'Ozone AQI Value', 'Ozone AQI Category',
       'NO2 AQI Value', 'NO2 AQI Category', 'PM2.5 AQI Value',
       'PM2.5 AQI Category', 'lat', 'lng'],
      dtype='object')
```

```
air.dtypes
```

```
Country               object
City                  object
AQI Value              int64
AQI Category          object
CO AQI Value           int64
CO AQI Category       object
Ozone AQI Value        int64
Ozone AQI Category    object
NO2 AQI Value          int64
NO2 AQI Category      object
PM2.5 AQI Value        int64
PM2.5 AQI Category    object
lat                  float64
lng                  float64
dtype: object
```

Now the dataset will be ready to analysis

1.country

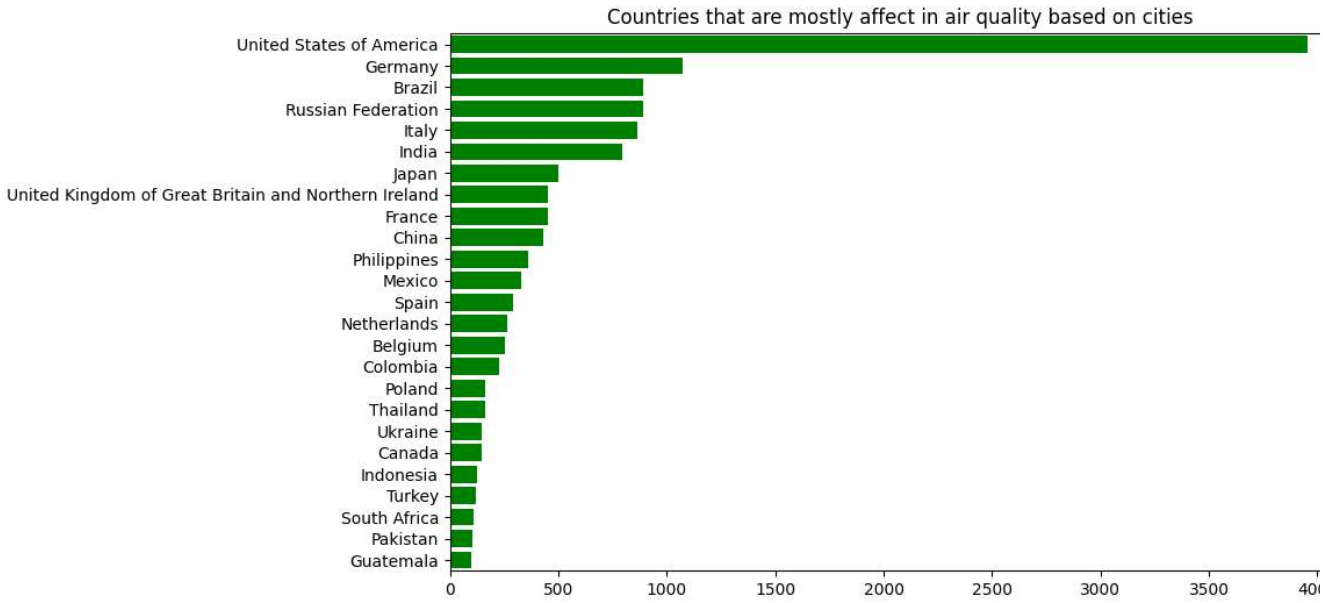
```
air['Country'].value_counts()
```

```
United States of America    3954
Germany                     1072
Brazil                       890
Russian Federation           889
```

```
Italy                         866
                             ...
Luxembourg                      1
Saint Lucia                     1
Republic of Korea               1
Solomon Islands                 1
Monaco                          1
Name: Country, Length: 174, dtype: int64
```

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

plt.title("Countries that are mostly affect in air quality based on cities")
sns.barplot(x=air['Country'].value_counts().head(25).values,
            y=air['Country'].value_counts().head(25).index,
            color='g',
            saturation=1)
plt.show()
```



The leading countries of affecting air quality are usa, germany, brazil, Russian federation are around the values of 3954, 1072,890,889 according based on the cities

The most of the air quality measurement are measured at america according to city wise compared to other countries

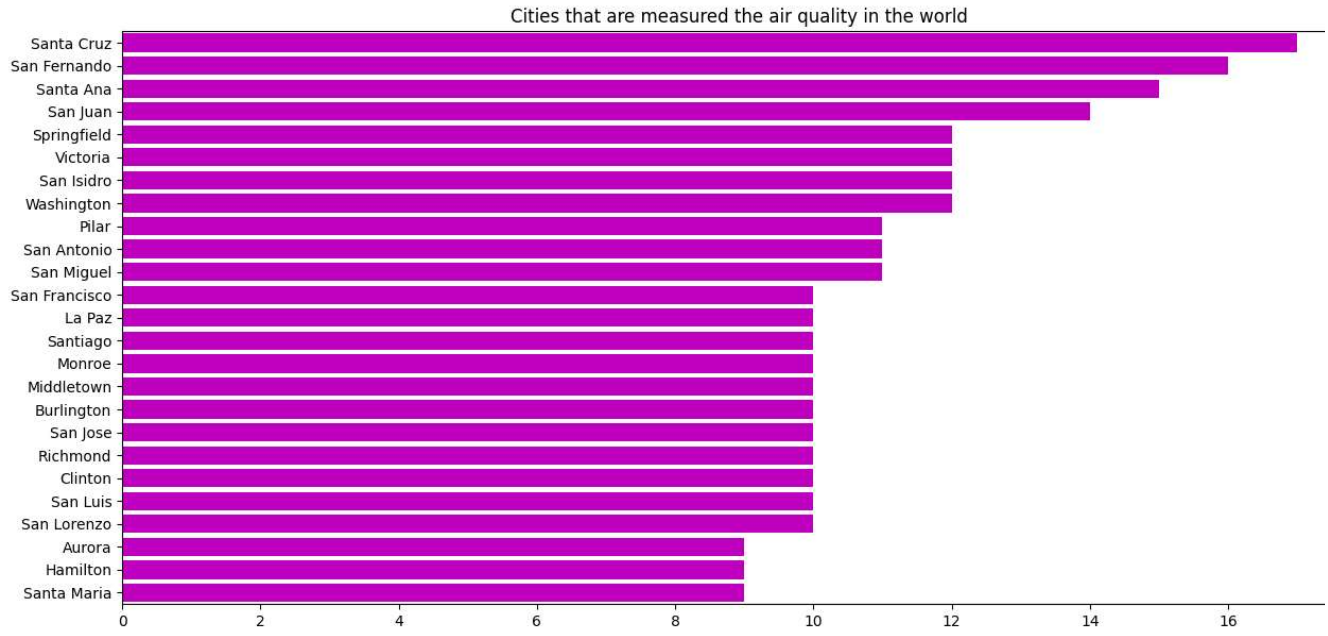
2.CITIES THAT ARE MEASURED AIR QUALITY

```
air['City'].value_counts()
```

```
Santa Cruz      17
San Fernando    16
Santa Ana       15
San Juan        14
Springfield     12
                ..
Hegang           1
Herxheim         1
Onalaska         1
Ostfildern       1
Westerville      1
Name: City, Length: 14229, dtype: int64
```

```
plt.figure(figsize=(15,7))
```

```
plt.title("Cities that are measured the air quality in the world")
sns.barplot(x=air['City'].value_counts().head(25).values,
            y=air['City'].value_counts().head(25).index,
            color='m',
            saturation=1)
plt.show()
```



santa cruz, san fernando, santa ana, san juan are cities that are mostly taken the air quality test by the organization at the times of 17,16,15,14 in the same region

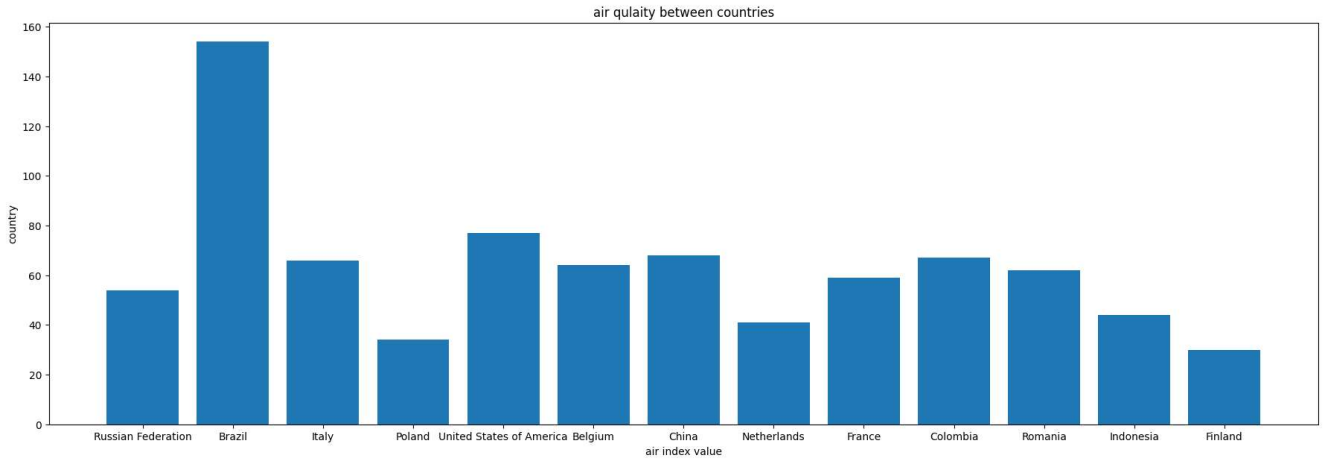
3.AIR INDEX QUALITY

```
air['AQI Value'].value_counts()
```

```
50     413
52     374
35     366
51     359
54     352
      ...
234      1
232      1
225      1
283      1
252      1
Name: AQI Value, Length: 282, dtype: int64
```

```
y=air['AQI Value'].head(25)
x=air['Country'].head(25)
```

```
plt.figure(figsize=(22,7))
plt.title("air qulaity between countries")
plt.xlabel("air index value")
plt.ylabel("country")
plt.bar(x, y)
plt.show()
```



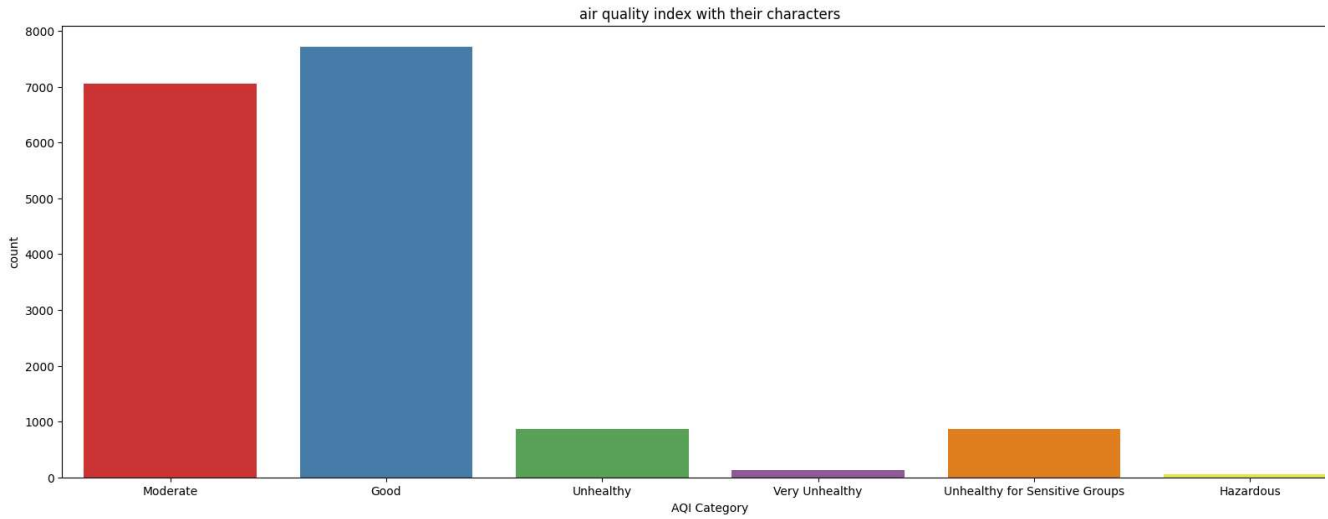
Brazil and united states of america are two countries has more air quality inde values with respect to other countries

4.AIR QUALITY CATEGORY

```
air['AQI Category'].value_counts()
```

```
Good                           7708
Moderate                       7054
Unhealthy                       871
Unhealthy for Sensitive Groups  869
Very Unhealthy                  131
Hazardous                        62
Name: AQI Category, dtype: int64
```

```
plt.figure(figsize=(20,7))
plt.title('air quality index with their characters')
sns.countplot(x='AQI Category',data=air,palette='Set1')
plt.show()
```



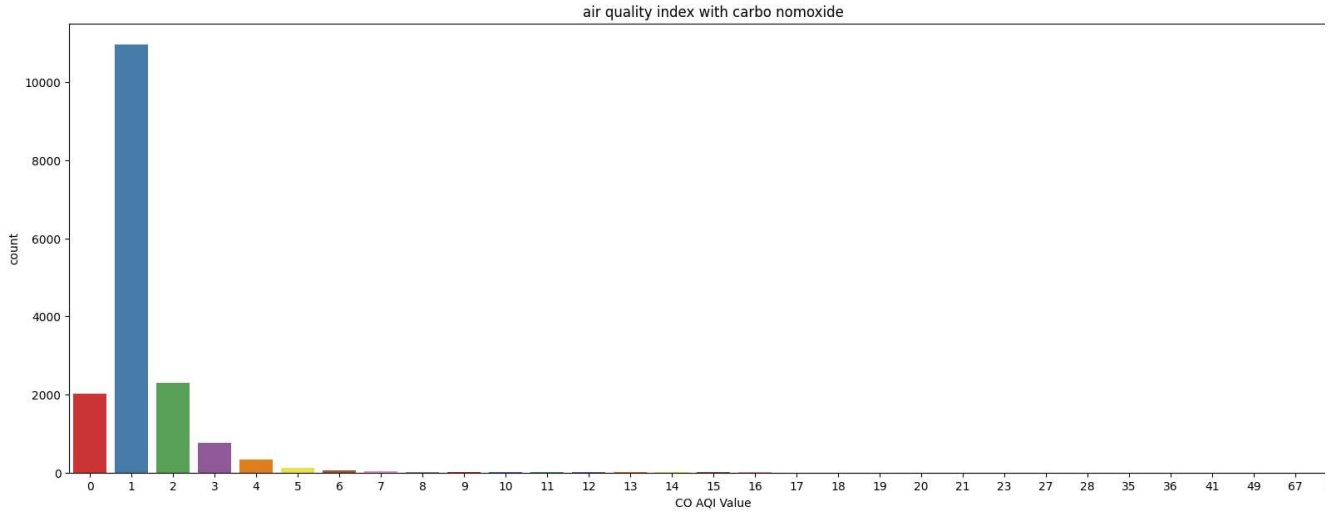
In air quality index are moderate and good from all the cities in the dataset and the charteristics represent in the cities values are 7054, 7708

5.AIR INDEX VALUES FROM CARBONMONO OXIDE

```
air['CO AQI Value'].value_counts()
```

```
1      10956
2       2292
0       2020
3        758
4        337
5        126
6         48
7         33
9         20
8         18
11        14
10        12
14         9
15         9
13         7
12         6
16         6
133        3
28         3
21         3
20         2
18         2
19         2
23         2
27         1
17         1
67         1
41         1
35         1
49         1
36         1
Name: CO AQI Value, dtype: int64
```

```
plt.figure(figsize=(20,7))
plt.title('air quality index with carbo nomoxide')
sns.countplot(x='CO AQI Value',data=air,palette='Set1')
plt.show()
```



The above graph shows the all the values present in the co2 with air quality index

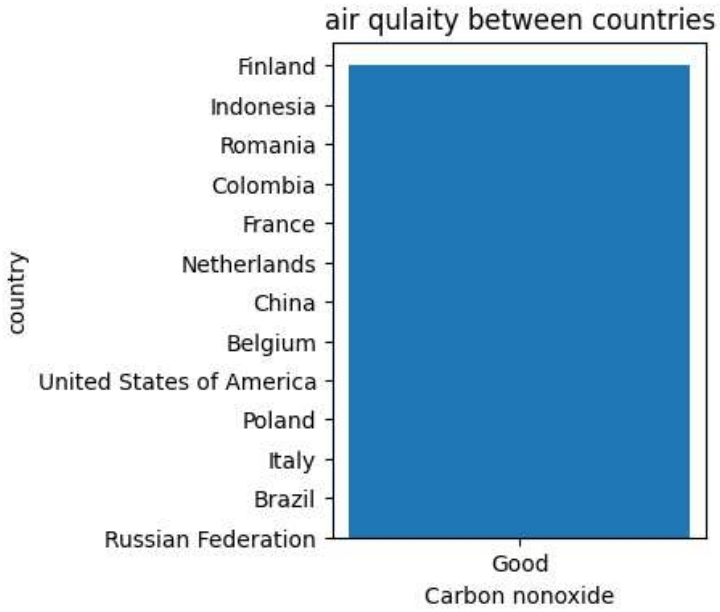
6.air quality index from carbonmonooxide category

```
air['CO AQI Category'].value_counts()
```

```
Good                              16691
Unhealthy for Sensitive Groups        3
Moderate                              1
Name: CO AQI Category, dtype: int64
```

```
x=air['CO AQI Category'].head(25)
y=air['Country'].head(25)
```

```
plt.figure(figsize=(3,4))
plt.title("air qulaity between countries")
plt.xlabel("Carbon nonoxide")
plt.ylabel("country")
plt.bar(x, y)
plt.show()
```




```
plt.figure(figsize=(7,5))
plt.title('air quality index with carbon ')
sns.countplot(x='CO AQI Category',data=air,color='red')
plt.show()
```

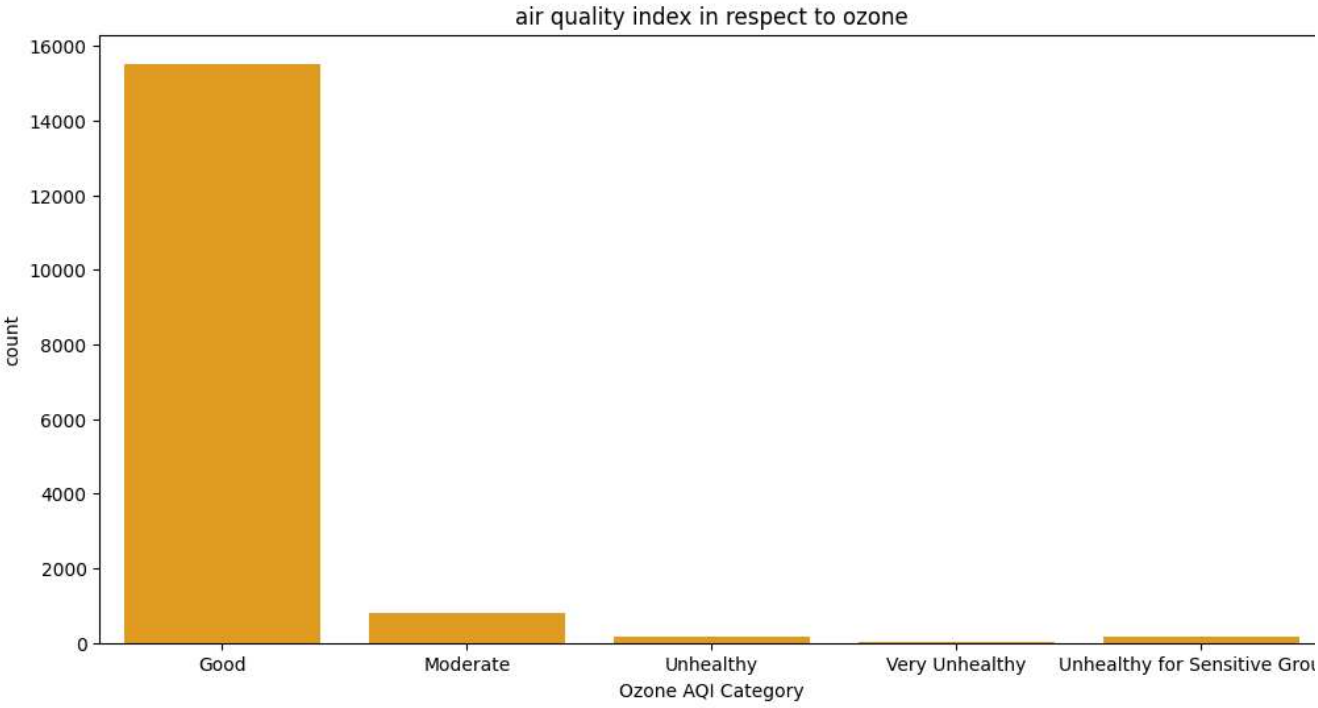
The air quality index of carbonmonoxide are good in all cities in the world are conformed in the above analysis

7.air quality index in ozone

```
air['Ozone AQI Category'].value_counts()
```

```
Good                            15529
Moderate                          806
Unhealthy for Sensitive Groups    176
Unhealthy                         159
Very Unhealthy                     25
Name: Ozone AQI Category, dtype: int64
```

```
plt.figure(figsize=(12,6))
plt.title('air quality index in respect to ozone')
sns.countplot(x='Ozone AQI Category',data=air,color='orange')
plt.show()
```



It shows only the aqi values present in O3 [ozone 3]

In most of the cities are good charcters of ozone content for air quality index are in the values of 15529

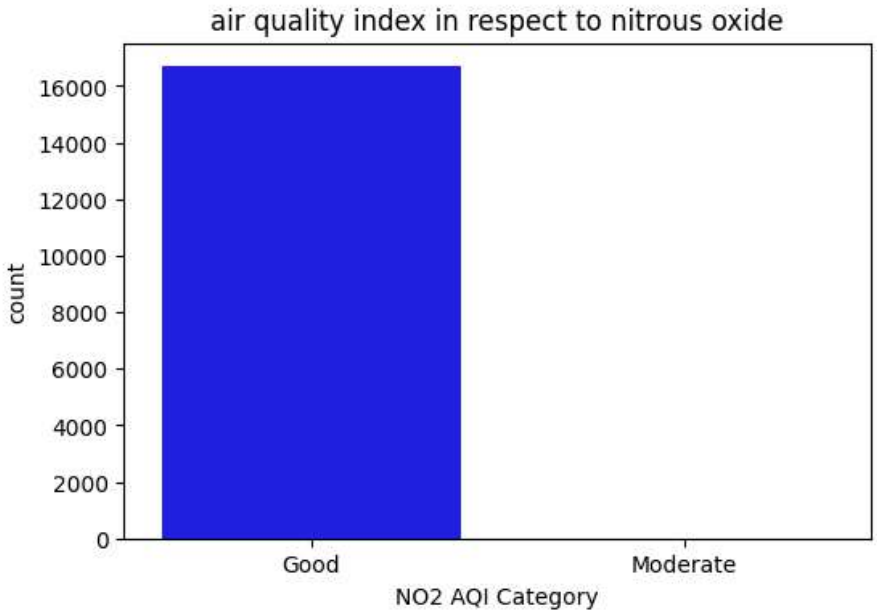
8.air quality index in nitrous oxide

Nitrous oxide is an odorless, colorless, non-flammable gas. While nitrous oxide is not flammable, it will support combustion to the same extent as oxygen. It leads to a state of euphoria

```
air['NO2 AQI Category'].value_counts()
```

```
Good        16684
Moderate       11
Name: NO2 AQI Category, dtype: int64
```

```
plt.figure(figsize=(6,4))
plt.title('air quality index in respect to nitrous oxide')
sns.countplot(x='NO2 AQI Category',data=air,color='blue')
plt.show()
```



Most of the countries has good content of nitrous oxide in the earth atmosphere are around the values of 16684 and moderate content present in the nitrous oxide are 11 are very very less are compared to good

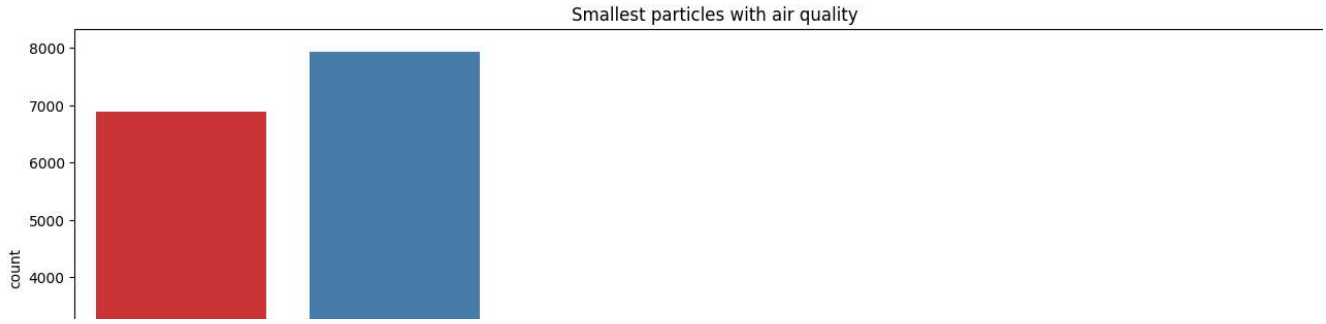
9.AIR QUALITY INDEX FOR SMALL PARTICLES

Fine particles in the air (measured as PM2.5) are so small that they can travel deeply into the respiratory tract, reaching the lungs, causing short-term health effects such as eye, nose, throat and lung irritation, coughing, sneezing, runny nose, and shortness of breath

```
air['PM2.5 AQI Category'].value_counts()
```

```
Good                             7936
Moderate                         6882
Unhealthy for Sensitive Groups    881
Unhealthy                         828
Very Unhealthy                    115
Hazardous                          53
Name: PM2.5 AQI Category, dtype: int64
```

```
plt.figure(figsize=(16,6))
plt.title('Smallest particles with air quality')
sns.countplot(x='PM2.5 AQI Category',data=air,palette='Set1')
plt.show()
```



There are good and moderate values of smallest particles That have also measured around the world and ranges from 7936, 6882

10.LATITUDE WITH RESPECT TO AQI

latitude is a coordinate that specifies the north–south position of a point on the surface of the Earth or another celestial body. Latitude is given as an angle that ranges from −90° at the south pole to 90° at the north pole, with 0° at the Equator. Lines of constant latitude, or parallels, run east–west as circles parallel to the equator.

```
air['lat'].value_counts()
```

```
51.2000    16
51.1000    13
51.3167    12
45.5500    11
50.9833    11
           ..
40.3412     1
36.2040     1
39.4254     1
44.5317     1
40.1241     1
Name: lat, Length: 14135, dtype: int64
```

```
plt.figure(figsize=(6,5))
plt.title('latitude with respect to air quality index')
sns.histplot(x='lat',data=air,kde=True,color='brown')
plt.show()
```

In air quality index the most of the values are ranges from the values of 40 to 60 from values of 1000 to 1750

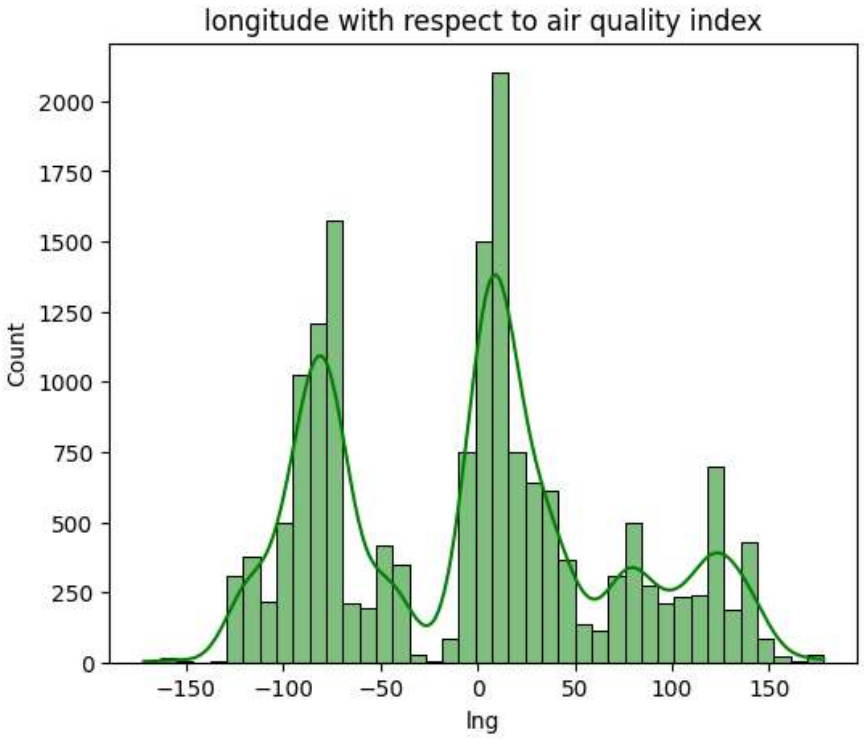
11.AIR QUALITY INDEX WITH RESPECT TO LONGITUDE

Longitude is a geographic coordinate that specifies the east–west position of a point on the surface of the Earth, or another celestial body. It is an angular measurement, usually expressed in degrees and denoted by the Greek letter lambda (λ). Meridians are imaginary semicircular lines running from pole to pole that connect points with the same longitude. The prime meridian defines 0° longitude

```
air['lng'].value_counts()
```

```
 9.2167      10
 8.7500       8
 4.5333       8
 10.8667      7
 8.6333       7
             ..
 119.1801     1
 77.5963      1
 34.8083      1
 120.5736     1
-82.9210      1
Name: lng, Length: 14896, dtype: int64
```

```
plt.figure(figsize=(6,5))
plt.title('longitude with respect to air quality index')
sns.histplot(x='lng',data=air,kde=True,color='green')
plt.show()
```



The above graph shows the longitude of air quality index and that are symmetyric with each other and highest longitude ranges from values(-100 to -50) and (50 to 150) and cities ranges from 1000 to 2000

Now the airquality forecasting can be analyzed sucessfully using python

✓ 0s completed at 3:00 PM