

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

```
In [2]: df=pd.read_csv("air_quality_index.csv")
df
```

Out[2]:

	DATE	COUNTRY	CITY	VALUE
0	2019-01-01	US	San Antonio	42.0
1	2019-01-01	US	Saint Paul	21.0
2	2019-01-01	US	Denver	42.0
3	2019-01-01	US	San Francisco	13.0
4	2019-01-01	US	Madison	14.0
...
68983	2021-05-31	US	Detroit	42.0
68984	2021-05-31	US	Jackson	30.0
68985	2021-05-31	US	The Bronx	12.0
68986	2021-05-31	US	Manhattan	13.0
68987	2021-05-31	IN	Patna	127.0

68988 rows × 4 columns

```
In [3]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 68988 entries, 0 to 68987
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   DATE        68988 non-null  object
1   COUNTRY     68988 non-null  object
2   CITY        68988 non-null  object
3   VALUE       68988 non-null  float64
dtypes: float64(1), object(3)
memory usage: 2.1+ MB
```

```
In [4]: df.describe()
```

Out[4]:

	VALUE
count	68988.000000
mean	54.766032
std	51.323430
min	1.000000
25%	23.000000
50%	35.000000
75%	63.000000
max	834.000000

```
In [5]: df.head()
```

```
Out[5]:
```

	DATE	COUNTRY	CITY	VALUE
0	2019-01-01	US	San Antonio	42.0
1	2019-01-01	US	Saint Paul	21.0
2	2019-01-01	US	Denver	42.0
3	2019-01-01	US	San Francisco	13.0
4	2019-01-01	US	Madison	14.0

```
In [6]: df.tail()
```

```
Out[6]:
```

	DATE	COUNTRY	CITY	VALUE
68983	2021-05-31	US	Detroit	42.0
68984	2021-05-31	US	Jackson	30.0
68985	2021-05-31	US	The Bronx	12.0
68986	2021-05-31	US	Manhattan	13.0
68987	2021-05-31	IN	Patna	127.0

```
In [3]: a=df[df.COUNTRY=="IN"]  
a
```

```
Out[3]:
```

	DATE	COUNTRY	CITY	VALUE
9	2019-01-01	IN	New Delhi	314.0
10	2019-01-01	IN	Chennai	135.0
13	2019-01-01	IN	Nashik	160.0
19	2019-01-01	IN	Chandigarh	163.0
21	2019-01-01	IN	Bhopal	175.0
...
68975	2021-05-31	IN	Mysore	50.0
68977	2021-05-31	IN	Bengaluru	36.0
68980	2021-05-31	IN	Hyderabad	65.0
68982	2021-05-31	IN	Lucknow	118.0
68987	2021-05-31	IN	Patna	127.0

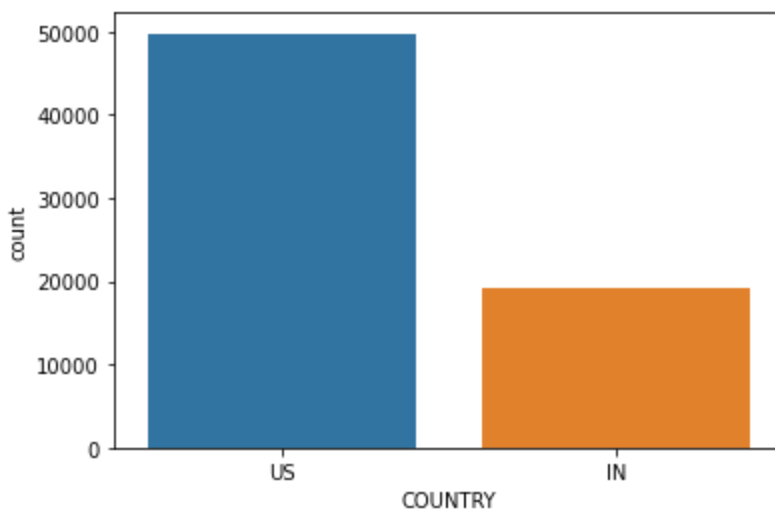
19179 rows × 4 columns

```
In [8]: sns.heatmap(df.isnull(),yticklabels=False,cbar=False,cmap='viridis');
```



There is no empty values in the dataset so it is a good dataset for further analysis

```
In [9]: sns.countplot(x='COUNTRY',data=df);
```



Count plot shows the number of count related to air pollution of different countries. In this dataset US country data is higher than India. So the pollution is more compared to India. It is only the assumption, it may be right or wrong based on analysis we can get the correct answer.

```
In [13]: a=pd.DataFrame(df[df.COUNTRY=="IN"])
a
```

```
Out[13]:
```

	DATE	COUNTRY	CITY	VALUE	YEAR	MONTH
9	2019-01-01	IN	New Delhi	314.0	2019	1
10	2019-01-01	IN	Chennai	135.0	2019	1
13	2019-01-01	IN	Nashik	160.0	2019	1
19	2019-01-01	IN	Chandigarh	163.0	2019	1
21	2019-01-01	IN	Bhopal	175.0	2019	1
...
68975	2021-05-31	IN	Mysore	50.0	2021	5
68977	2021-05-31	IN	Bengaluru	36.0	2021	5
68980	2021-05-31	IN	Hyderabad	65.0	2021	5
68982	2021-05-31	IN	Lucknow	118.0	2021	5

19179 rows × 6 columns

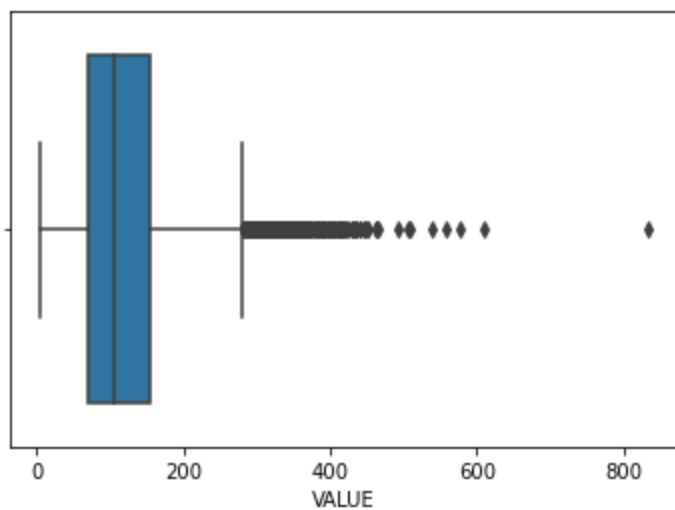
```
In [19]: b=pd.DataFrame(df[df.COUNTRY=="US"])
b
```

```
Out[19]:
```

	DATE	COUNTRY	CITY	VALUE	YEAR	MONTH
0	2019-01-01	US	San Antonio	42.0	2019	1
1	2019-01-01	US	Saint Paul	21.0	2019	1
2	2019-01-01	US	Denver	42.0	2019	1
3	2019-01-01	US	San Francisco	13.0	2019	1
4	2019-01-01	US	Madison	14.0	2019	1
...
68981	2021-05-31	US	Phoenix	22.0	2021	5
68983	2021-05-31	US	Detroit	42.0	2021	5
68984	2021-05-31	US	Jackson	30.0	2021	5
68985	2021-05-31	US	The Bronx	12.0	2021	5
68986	2021-05-31	US	Manhattan	13.0	2021	5

49809 rows × 6 columns

```
In [11]: sns.boxplot(x="VALUE", data=a);
```



```
In [9]: df['YEAR'] = pd.DatetimeIndex(df['DATE']).year
df['MONTH'] = pd.DatetimeIndex(df['DATE']).month
```

```
In [14]: df
```

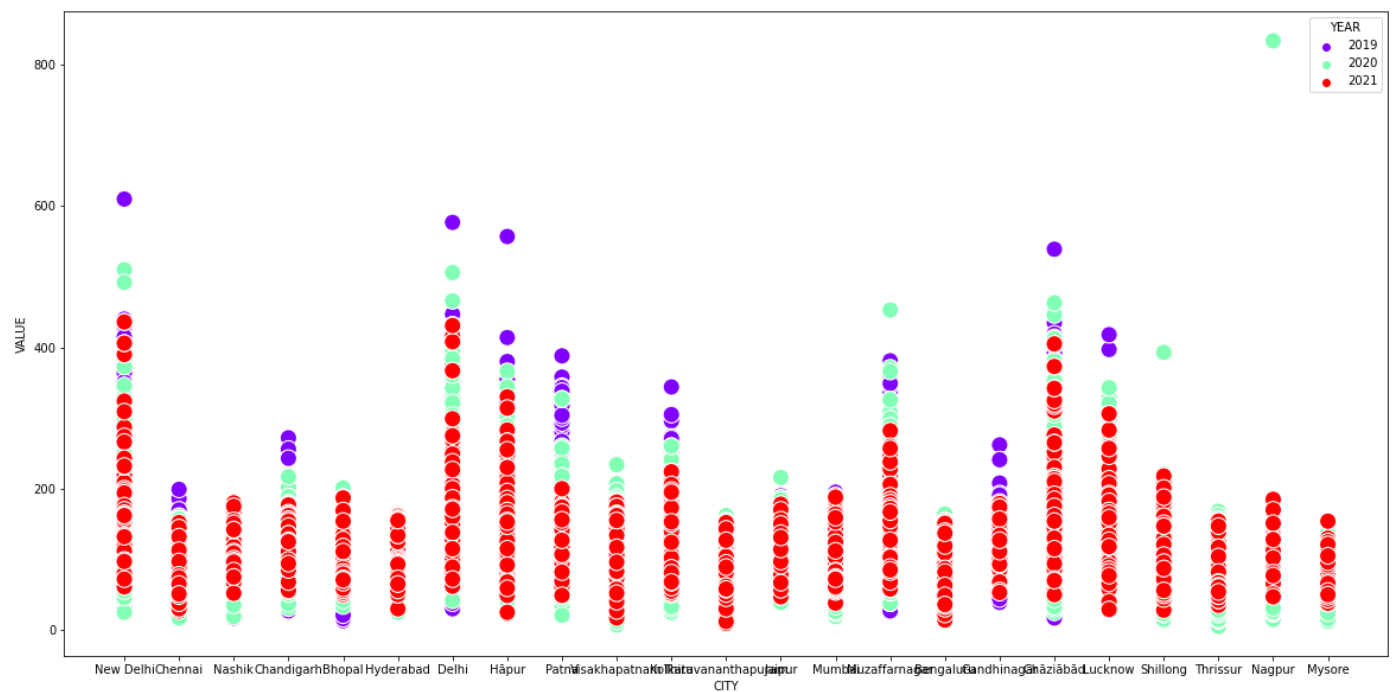
```
Out[14]:
```

	DATE	COUNTRY	CITY	VALUE	YEAR	MONTH
0	2019-01-01	US	San Antonio	42.0	2019	1
1	2019-01-01	US	Saint Paul	21.0	2019	1
2	2019-01-01	US	Denver	42.0	2019	1

3	2019-01-01	US	San Francisco	13.0	2019	1
4	2019-01-01	US	Madison	14.0	2019	1
...
68983	2021-05-31	US	Detroit	42.0	2021	5
68984	2021-05-31	US	Jackson	30.0	2021	5
68985	2021-05-31	US	The Bronx	12.0	2021	5
68986	2021-05-31	US	Manhattan	13.0	2021	5
68987	2021-05-31	IN	Patna	127.0	2021	5

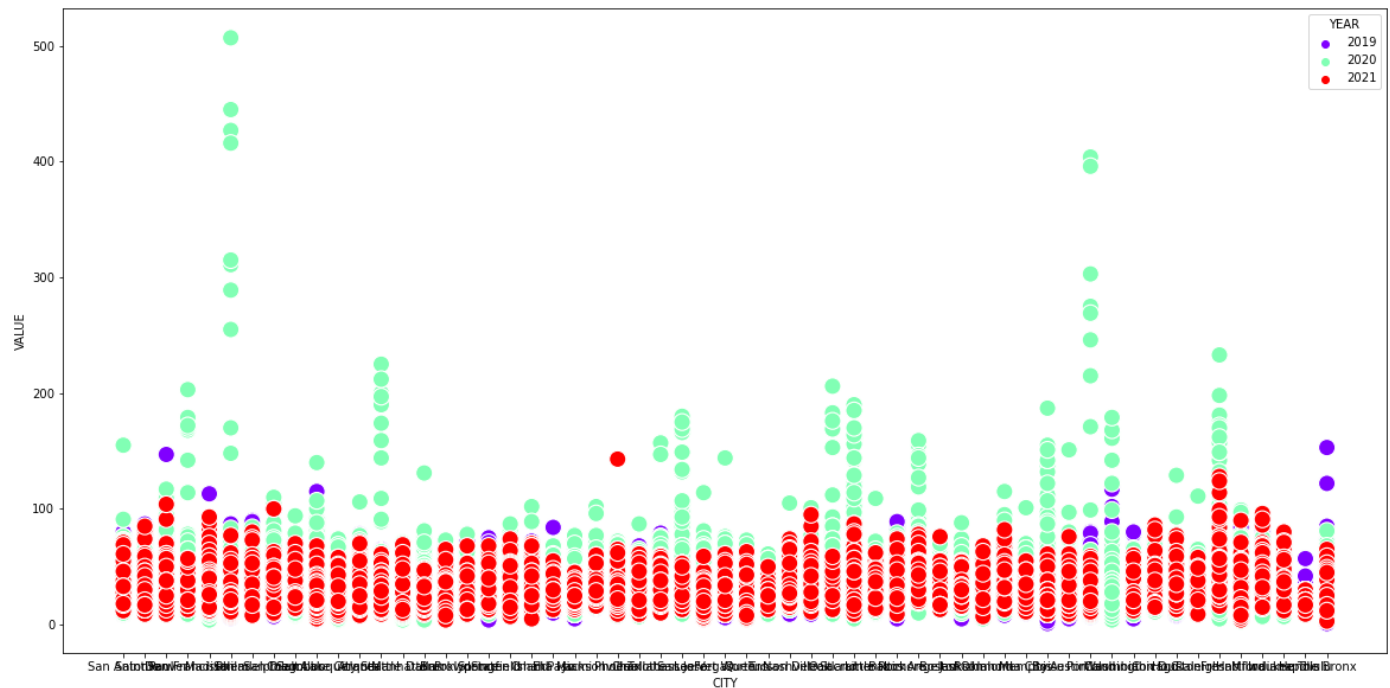
68988 rows × 6 columns

```
In [26]: plt.figure(figsize=(20, 10))
size=200
sns.scatterplot(x="CITY", y="VALUE", hue="YEAR", data=a, s=size, palette="rainbow");
```



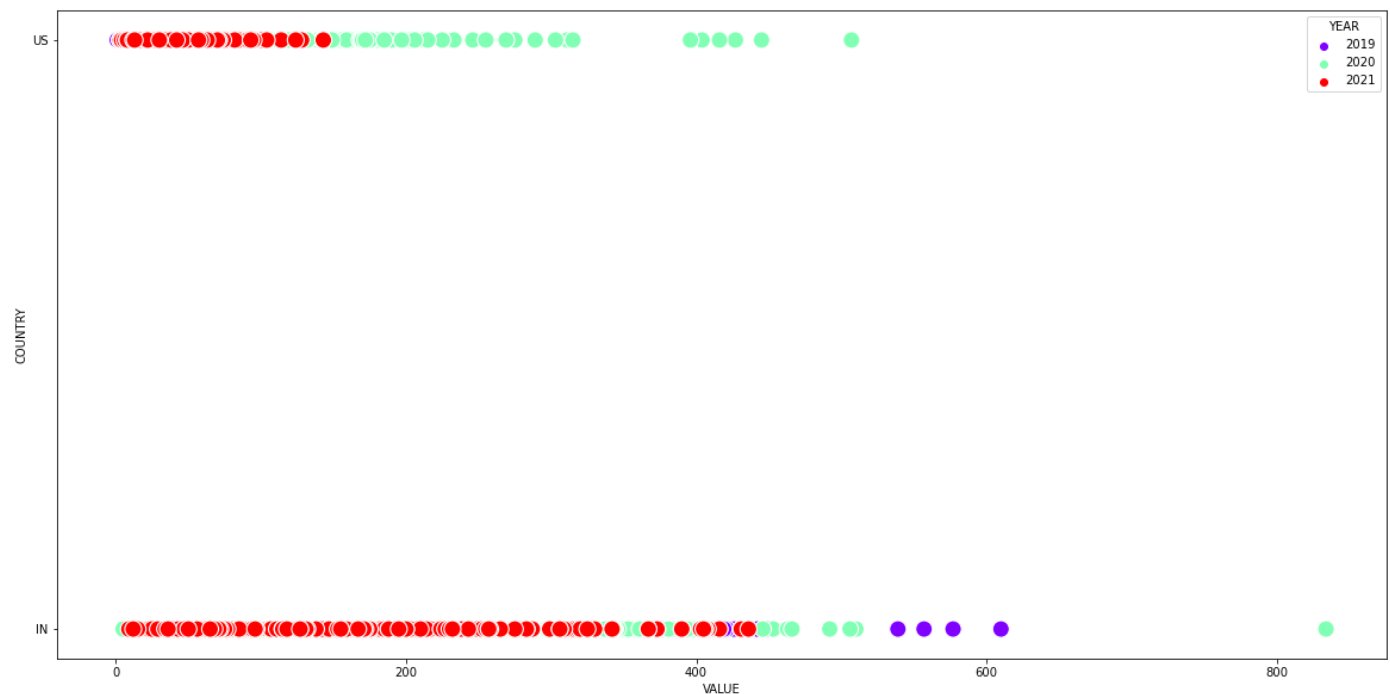
We can predict from the plot that air pollution in india is more in the year of 2019 and 2020 compared to 2021.

```
In [28]: plt.figure(figsize=(20, 10))
size=200
sns.scatterplot(x="CITY", y="VALUE", hue="YEAR", data=b, s=size, palette="rainbow");
```



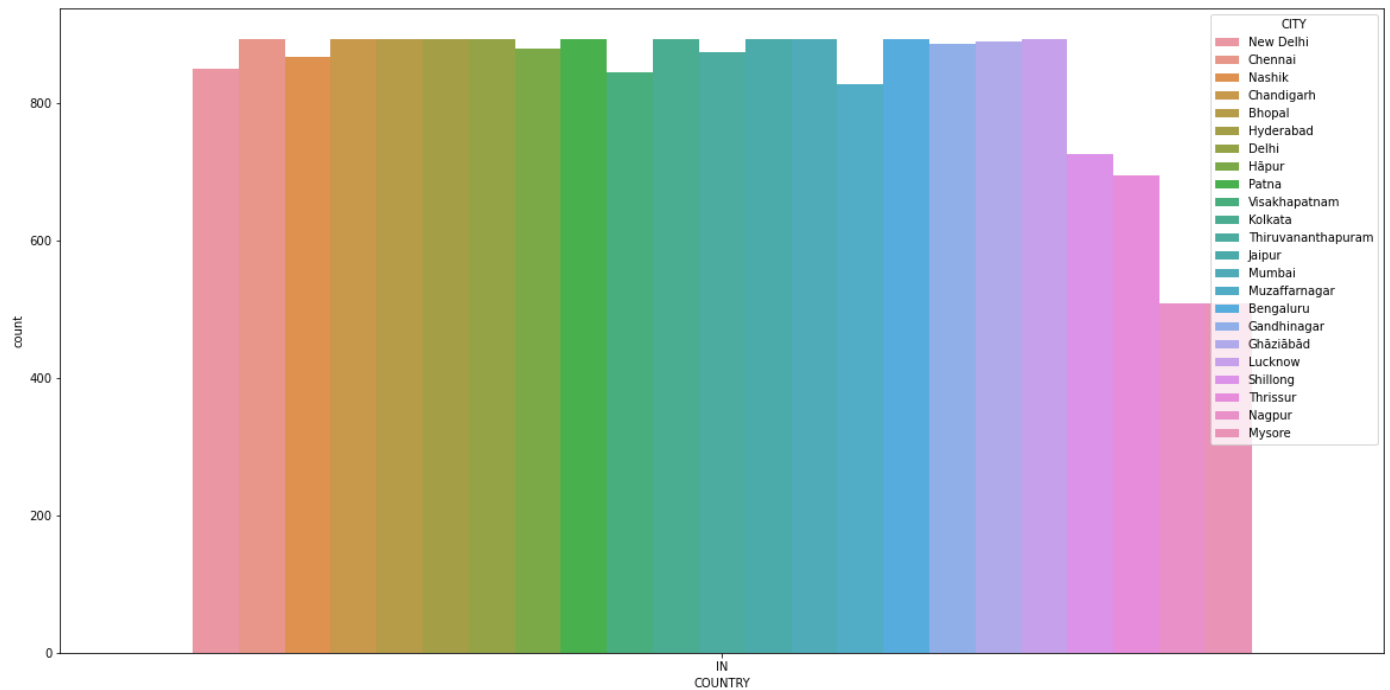
In US air pollution is more in the year of 2020

```
In [30]: plt.figure(figsize=(20, 10))
size=200
sns.scatterplot(x="VALUE", y="COUNTRY", hue="YEAR", data=df, s=size, palette="rainbow");
```

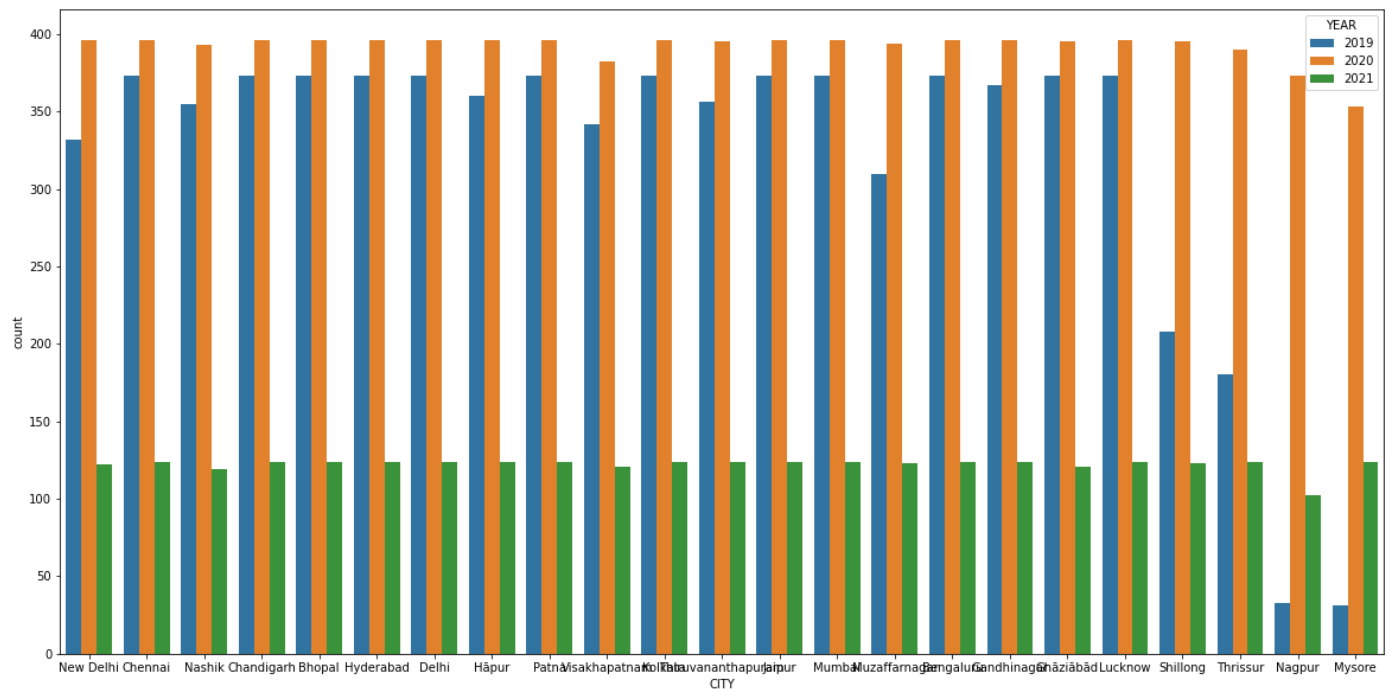


From the visualization we can see that the air pollution is more in the year 2020. So this may be the reason for some disaster happened in the year 2020

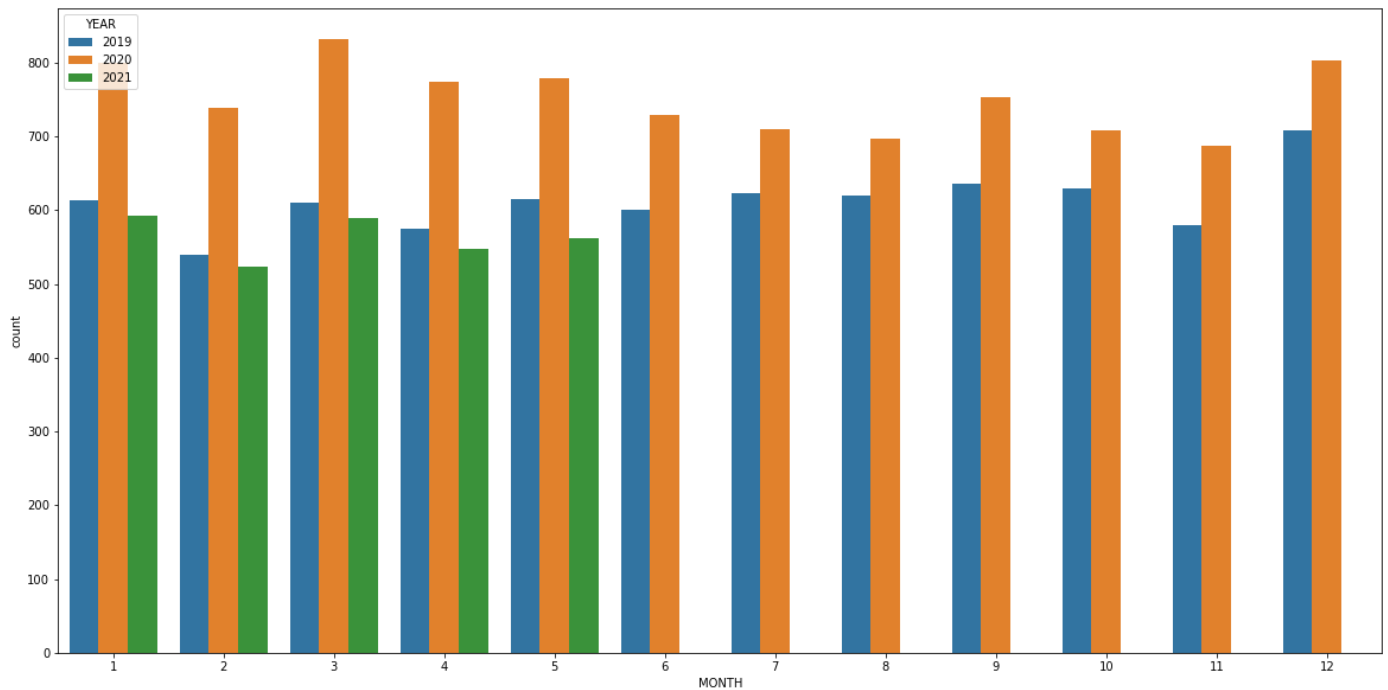
```
In [33]: plt.figure(figsize=(20, 10))
sns.countplot(x='COUNTRY', hue="CITY", data=a);
```



```
In [34]: plt.figure(figsize=(20, 10))
sns.countplot(x='CITY', hue="YEAR", data=a);
```



```
In [35]: plt.figure(figsize=(20, 10))
sns.countplot(x='MONTH', hue="YEAR", data=a);
```



```
In [41]: a[a.YEAR==2020]["VALUE"].max()
a[a.MONTH==3]["VALUE"].max()
```

```
Out[41]: 834.0
```

```
In [35]: df[df.COUNTRY=="IN"]["VALUE"].max()
```

```
Out[35]: 834.0
```

```
In [7]: df[df.CITY=="New Delhi"]["VALUE"].max()
```

```
Out[7]: 610.0
```

```
In [37]: a[a.MONTH==3]["VALUE"].max()
a[a.VALUE==610]
```

```
Out[37]:
```

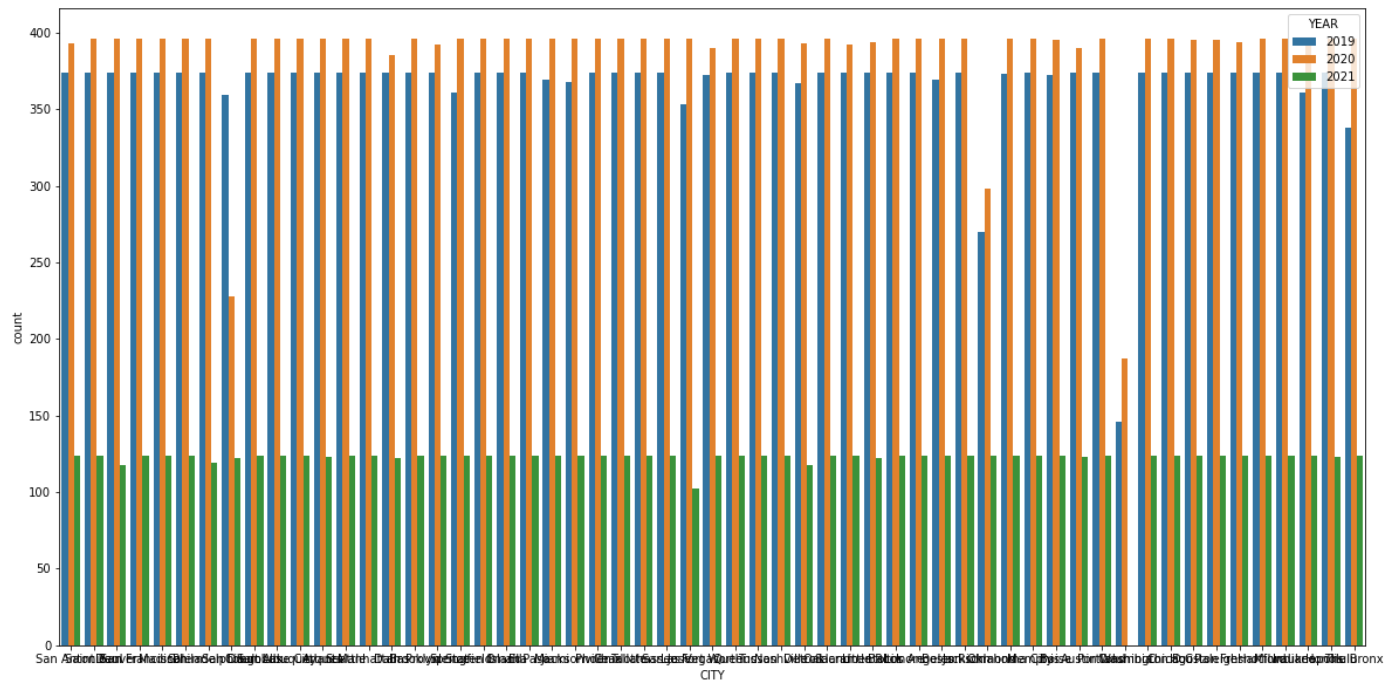
	DATE	COUNTRY	CITY	VALUE	YEAR	MONTH
5456	2019-03-11	IN	New Delhi	610.0	2019	3

```
In [23]: df[df.COUNTRY=="US"][["VALUE", "CITY", "MONTH"]].max()
```

```
Out[23]:
```

VALUE	507.0
CITY	Washington D.C.
MONTH	12
dtype:	object

```
In [26]: plt.figure(figsize=(20, 10))
sns.countplot(x='CITY', hue='YEAR', data=b);
```

```
In [28]: df[df.COUNTRY=="US"]["CITY"]
```

```
Out[28]:
0          San Antonio
1          Saint Paul
2           Denver
3        San Francisco
4           Madison
...
68981        Phoenix
68983        Detroit
68984         Jackson
68985        The Bronx
68986        Manhattan
Name: CITY, Length: 49809, dtype: object
```

```
In [34]: b["VALUE"].min()
c=b[b.VALUE==1]
c
```

	DATE	COUNTRY	CITY	VALUE	YEAR	MONTH
223	2019-01-03	US	The Bronx	1.0	2019	1
235	2019-01-04	US	The Bronx	1.0	2019	1
2586	2019-02-03	US	The Bronx	1.0	2019	2
2752	2019-02-05	US	The Bronx	1.0	2019	2
3791	2019-02-18	US	The Bronx	1.0	2019	2
4950	2019-03-06	US	The Bronx	1.0	2019	3
5752	2019-03-15	US	The Bronx	1.0	2019	3
5776	2019-03-16	US	The Bronx	1.0	2019	3
6071	2019-03-20	US	The Bronx	1.0	2019	3
6298	2019-03-23	US	The Bronx	1.0	2019	3
6799	2019-03-30	US	The Bronx	1.0	2019	3
6935	2019-03-31	US	The Bronx	1.0	2019	3
7255	2019-04-05	US	The Bronx	1.0	2019	4

7345	2019-04-06	US	The Bronx	1.0	2019	4
7960	2019-04-13	US	The Bronx	1.0	2019	4
8003	2019-04-14	US	The Bronx	1.0	2019	4
8388	2019-04-19	US	The Bronx	1.0	2019	4
8480	2019-04-20	US	The Bronx	1.0	2019	4
8571	2019-04-21	US	The Bronx	1.0	2019	4
8799	2019-04-24	US	The Bronx	1.0	2019	4
8873	2019-04-25	US	The Bronx	1.0	2019	4
8934	2019-04-26	US	The Bronx	1.0	2019	4
8976	2019-04-27	US	The Bronx	1.0	2019	4
9041	2019-04-28	US	The Bronx	1.0	2019	4
9605	2019-05-05	US	The Bronx	1.0	2019	5
10307	2019-05-14	US	The Bronx	1.0	2019	5
10414	2019-05-15	US	The Bronx	1.0	2019	5
10495	2019-05-16	US	The Bronx	1.0	2019	5
10553	2019-05-17	US	The Bronx	1.0	2019	5
10631	2019-05-18	US	The Bronx	1.0	2019	5
10768	2019-05-20	US	Boise	1.0	2019	5
10798	2019-05-20	US	The Bronx	1.0	2019	5
10877	2019-05-21	US	The Bronx	1.0	2019	5
11048	2019-05-23	US	The Bronx	1.0	2019	5
12018	2019-06-05	US	The Bronx	1.0	2019	6
14086	2019-07-02	US	The Bronx	1.0	2019	7
16506	2019-08-02	US	The Bronx	1.0	2019	8
16706	2019-08-05	US	The Bronx	1.0	2019	8
18843	2019-09-02	US	The Bronx	1.0	2019	9
18987	2019-09-04	US	The Bronx	1.0	2019	9
21419	2019-10-05	US	The Bronx	1.0	2019	10
23779	2019-11-05	US	The Bronx	1.0	2019	11
25918	2019-12-04	US	The Bronx	1.0	2019	12

```
In [18]: a["VALUE"].min()
a[a.VALUE==5]
```

```
Out[18]:
```

	DATE	COUNTRY	CITY	VALUE	YEAR	MONTH
49699	2020-09-08	IN	Thrissur	5.0	2020	9

```
In [15]: a["VALUE"].max()
a[a.VALUE==834]
```

Out[15]:

	DATE	COUNTRY	CITY	VALUE	YEAR	MONTH
56581	2020-12-01	IN	Nagpur	834.0	2020	12

In []: CONCLUSION:
The maximum polluted city in india is Nagpur and minimum polluted city is Thirssur i