Data Visualization Using Python

By: Anushka Sharma
(RA2111042010033)

Aim: To examine and visualize air quality over time in different cities and see if there is a correlation between industrialization and deteriorating air quality.

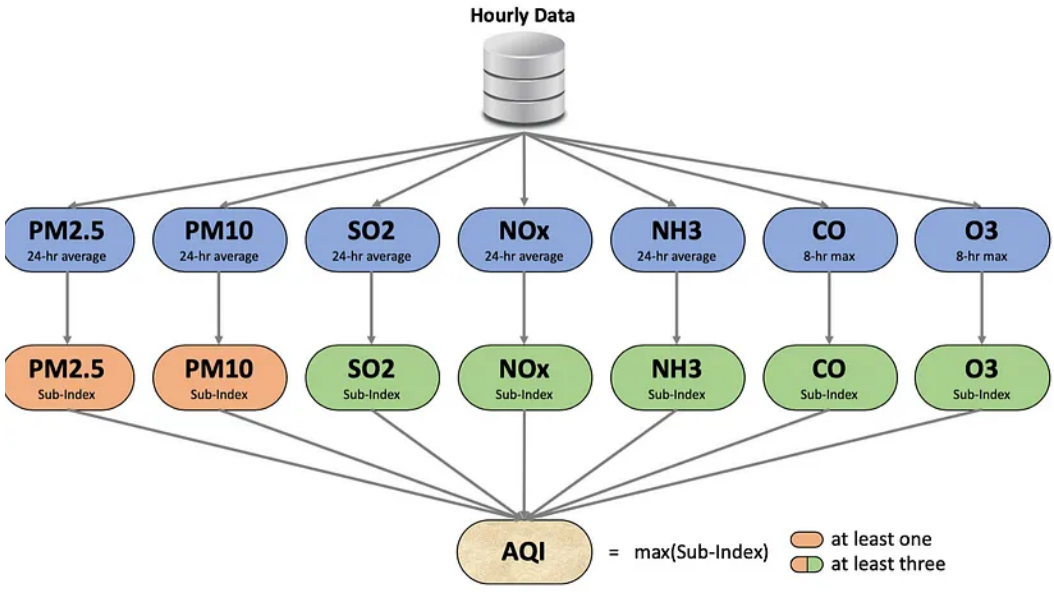
Introduction:

The daily air quality is reported using the Air Quality Index (AQI). It informs you of the cleanliness and pollution levels of your air and any potential health risks.

These Air Quality indices can be categorically broken down into 5 major air pollutants, namely:

- Ground-level ozone
- Particle Pollution (also known as particulate matter, including PM2.5 and PM10)
- Carbon Monoxide
- Sulfur Dioxide
- Nitrogen Dioxide

AQI Calculation:



- Seven measurements are used to calculate the AQI: PM2.5, PM10, SO2, NOx, NH3, CO, and O3.
- When there are at least 16 values, the average value from the previous 24 hours is utilized for PM2.5, PM10, SO2, NOx, and NH3.
- The highest value over the last 8 hours is utilized for CO and O3.
- Based on pre-established groupings, each measure is transformed into a Sub-Index.
- The final AQI is the maximum Sub-Index, provided that at least three of the seven components are present, including at least one each of PM2.5 and PM10.

Methodology:

- Getting a dataset.
- Checking the dataset.
- Doing visualization of the dataset.

Initialize Dataset:

Code:

```
import matplotlib.pyplot as plt
import pandas as pd

df = pd.read_csv('data.csv')


#Create a Pandas DataFrame from the data
df = pd.DataFrame(data)
```

Visualize Dataset:

- **Line Chart for each city differentiated by Industrial Area and Residential Area.**

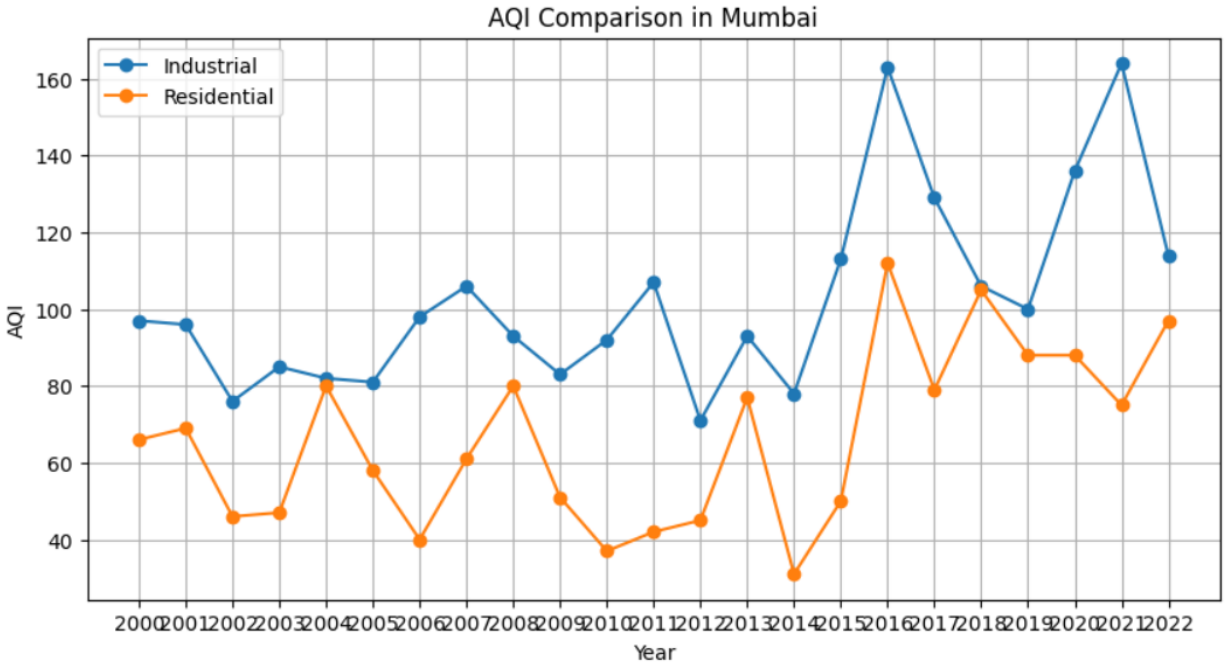
Code:

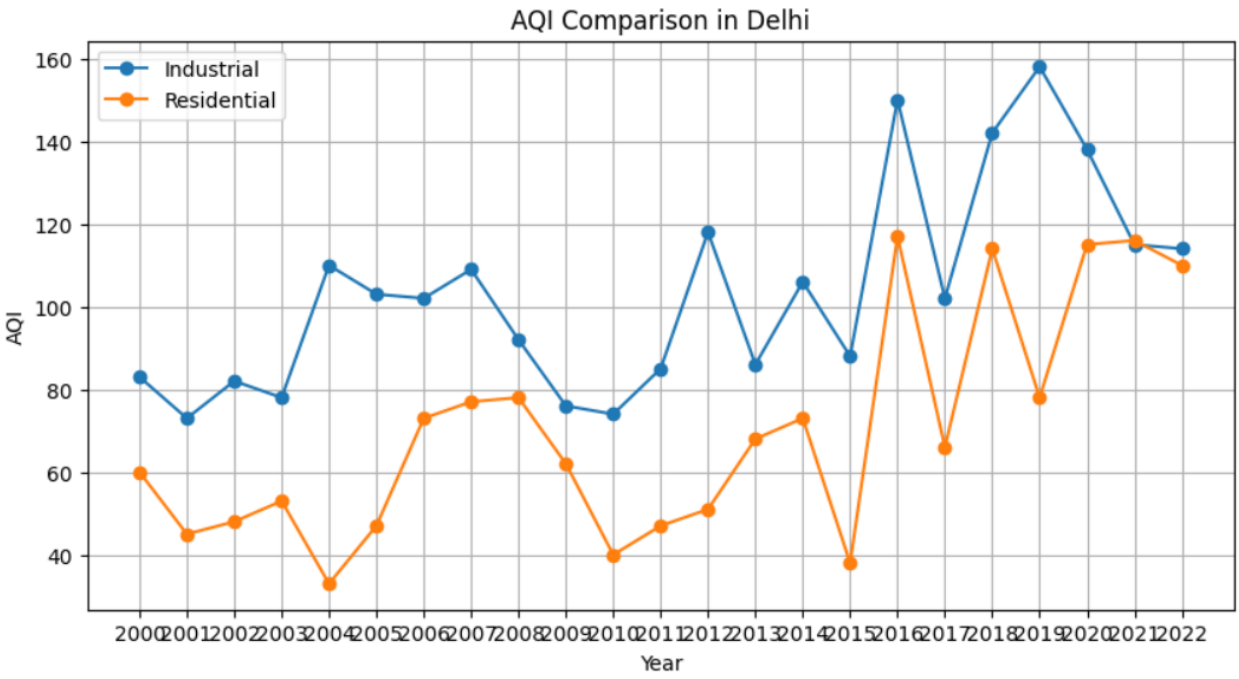
```
for city in cities:
    city_data = df[df['City'] == city]
    plt.figure(figsize=(10, 5))
    plt.title(f'AQI Comparison in {city}')
    plt.xlabel('Year')
    plt.ylabel('AQI')

    # Plot industrial AQI
    plt.plot(city_data['Year'], city_data['AQI_Industrial'],
label='Industrial', marker='o')

    # Plot residential AQI
    plt.plot(city_data['Year'], city_data['AQI_Residential'],
label='Residential', marker='o')

    plt.xticks(years)
    plt.legend()
    plt.grid(True)
    plt.show()
```



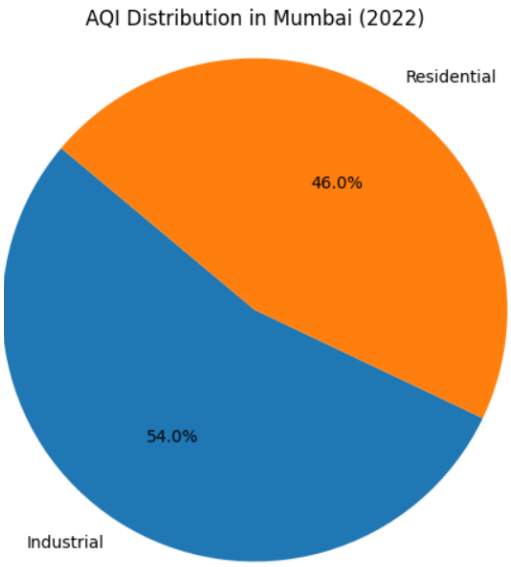


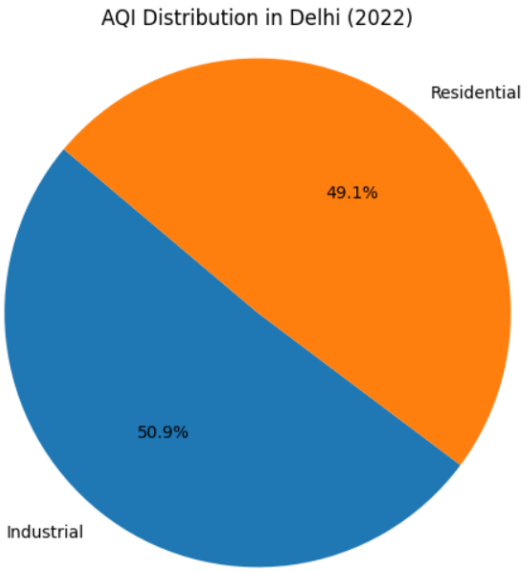
- **Pie Chart for each city differentiated by Industrial Area and Residential Area.**

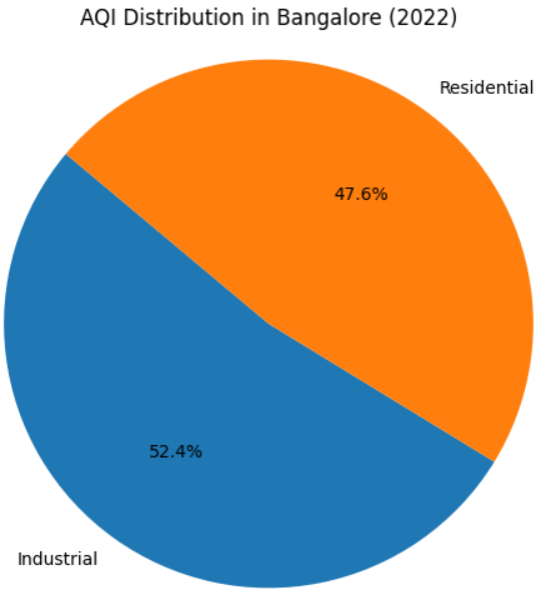
Code:

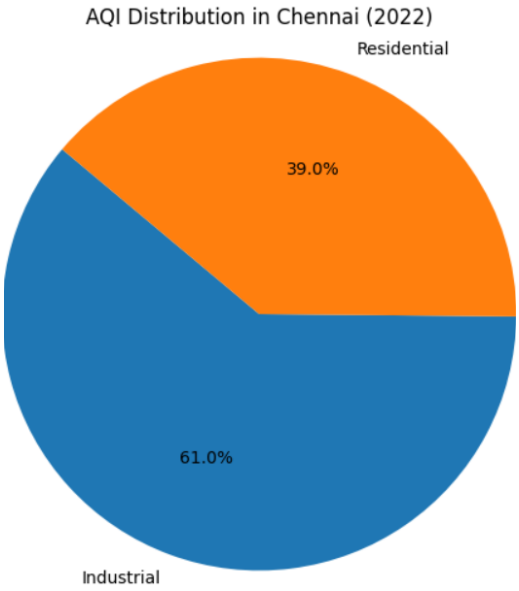
```
year_to_compare = 2022
pollution_data = df[df['Year'] ==
year_to_compare].groupby('City')[['AQI_Industrial',
'AQI_Residential']].sum()

for city in cities:
      plt.figure(figsize=(6, 6))
      plt.title(f'AQI Distribution in {city} ({year_to_compare})')
      labels = ['Industrial', 'Residential']
      sizes = pollution_data.loc[city].values
      plt.pie(sizes, labels=labels, autopct='%1.1f%%',
      startangle=140)
      plt.axis('equal')
      plt.show()
```







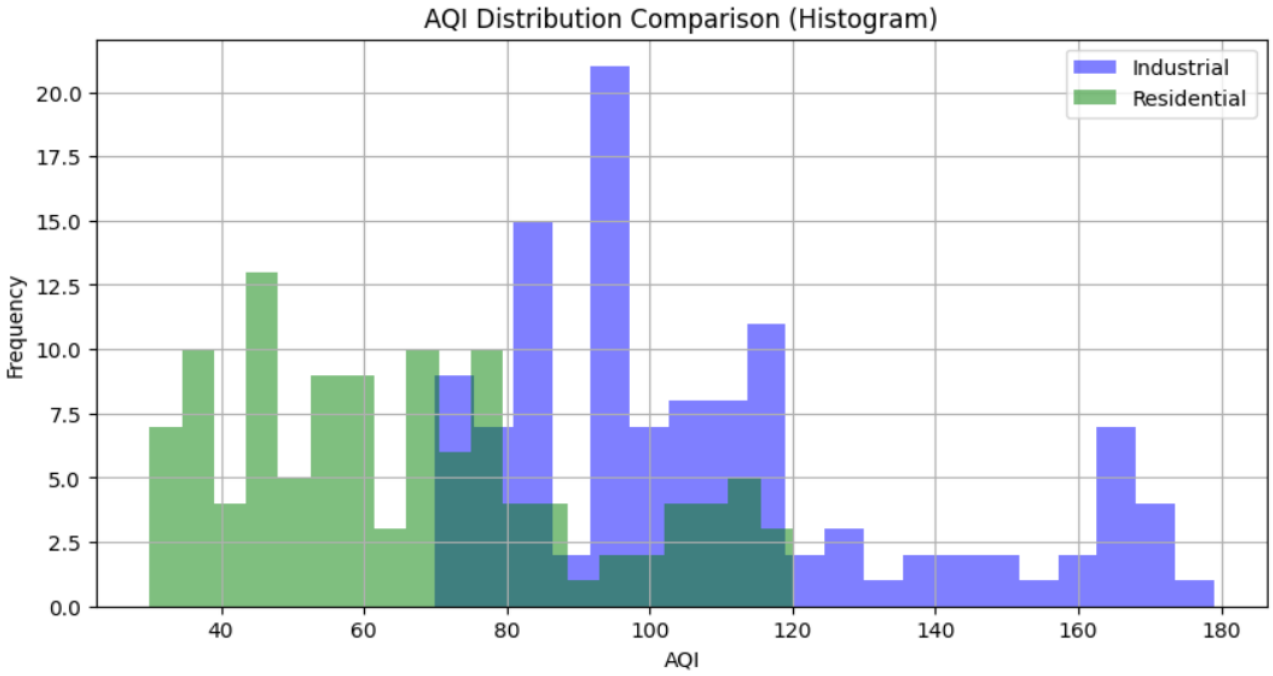


- **Histograms to visualize the distribution of AQI values in industrial and residential areas**

Code:

```
plt.figure(figsize=(10, 5))
plt.title('AQI Distribution Comparison (Histogram)')
plt.xlabel('AQI')
plt.ylabel('Frequency')
```

```
plt.hist(df['AQI_Industrial'], bins=20, alpha=0.5,
label='Industrial', color='blue')
plt.hist(df['AQI_Residential'], bins=20, alpha=0.5,
label='Residential', color='green')
plt.legend()
plt.grid(True)
plt.show()
```



- **Scatter plot to compare AQI levels between industrial and residential areas for a specific city.**

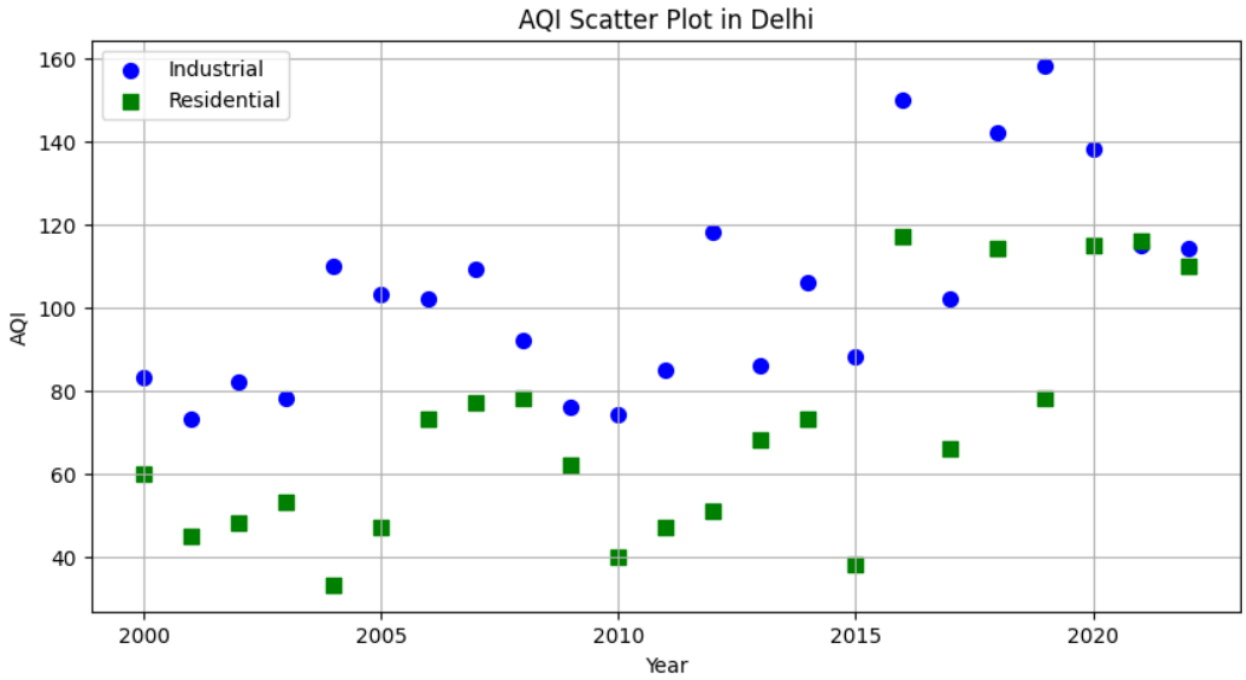
Code:

```
city_to_compare = 'Delhi'
city_data = df[df['City'] == city_to_compare]

plt.figure(figsize=(10, 5))
plt.title(f'AQI Scatter Plot in {city_to_compare}')
plt.xlabel('Year')
plt.ylabel('AQI')

plt.scatter(city_data['Year'], city_data['AQI_Industrial'],
label='Industrial', c='blue', marker='o', s=50)
```

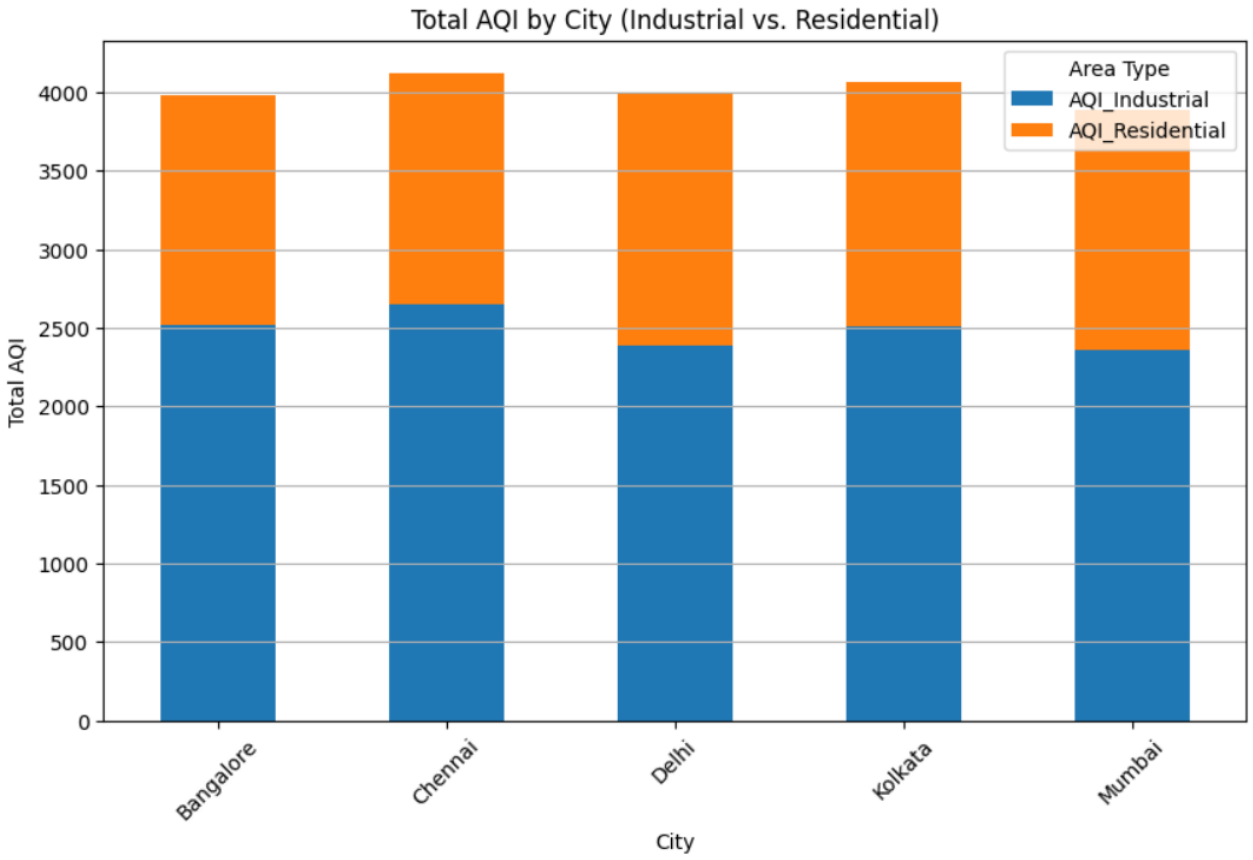
```
plt.scatter(city_data['Year'], city_data['AQI_Residential'],
label='Residential', c='green', marker='s', s=50)
plt.legend()
plt.grid(True)
plt.show()
```



- **Consolidated bar chart for total pollution levels in industrial and residential areas**

Code:

```
total_pollution = df.groupby('City')[['AQI_Industrial',
'AQI_Residential']].sum()
total_pollution.plot(kind='bar', stacked=True, figsize=(10, 6))
plt.title('Total AQI by City (Industrial vs. Residential)')
plt.xlabel('City')
plt.ylabel('Total AQI')
plt.xticks(rotation=45)
plt.legend(title='Area Type')
plt.grid(axis='y')
plt.show()
```

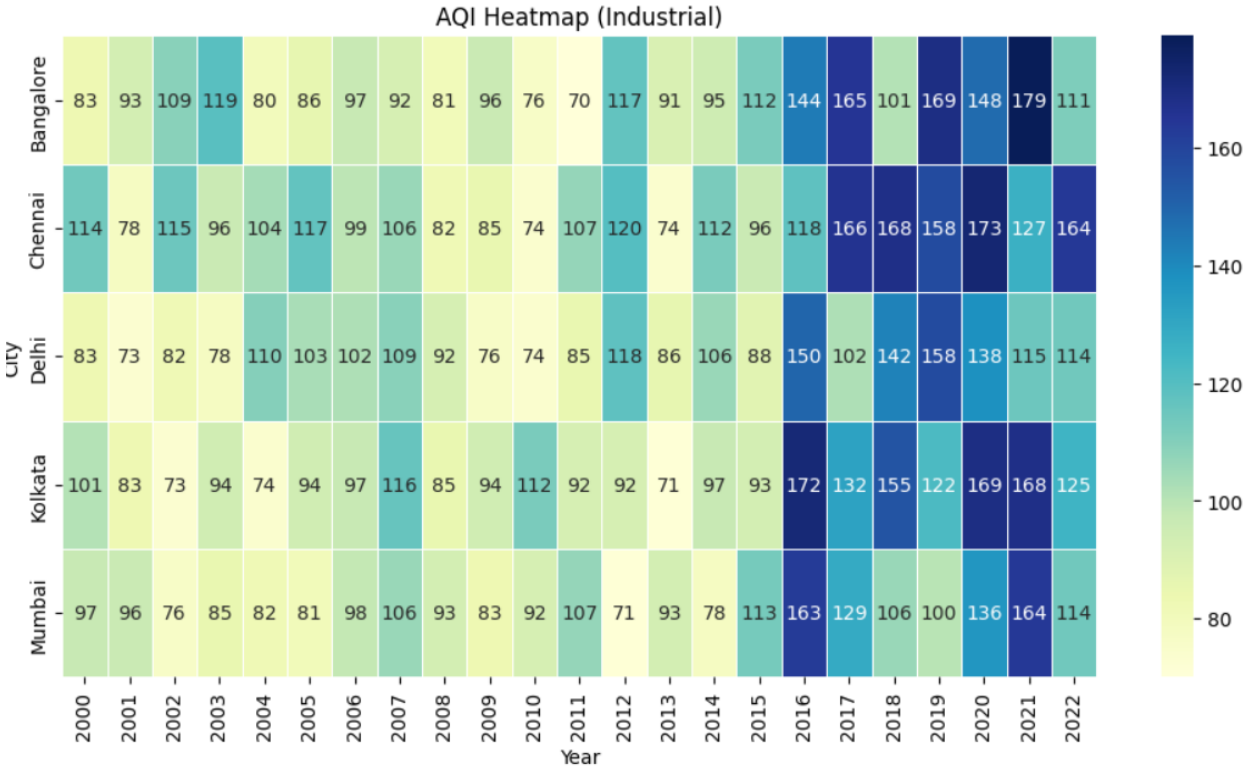


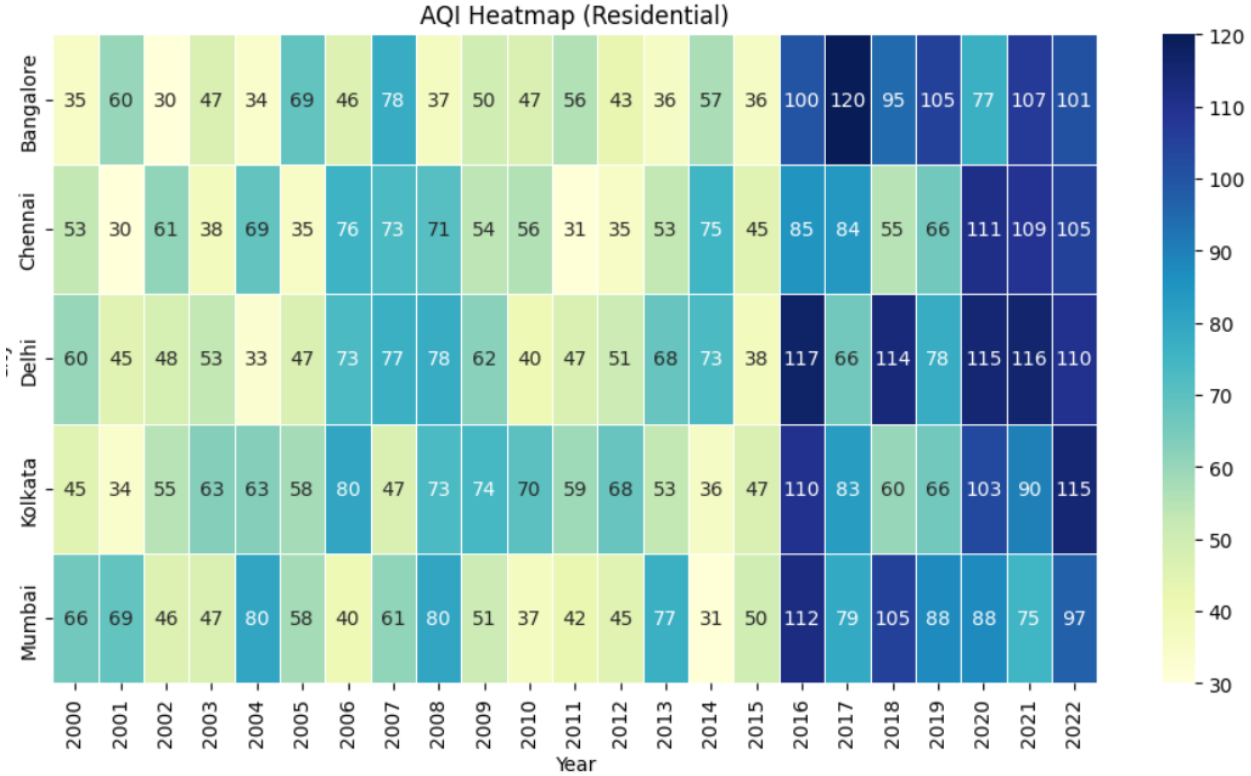
- **Consolidated heatmap for AQI comparison across all cities**

Code:

```
heatmap_data = df.pivot_table(index='City', columns='Year',
values='AQI_Industrial', aggfunc='mean')
plt.figure(figsize=(12, 6))
plt.title('AQI Heatmap (Industrial)')
sns.heatmap(heatmap_data, cmap='YlGnBu', annot=True, fmt='.0f',
cbar=True, linewidths=0.5)
plt.show()

heatmap_data = df.pivot_table(index='City', columns='Year',
values='AQI_Residential', aggfunc='mean')
plt.figure(figsize=(12, 6))
plt.title('AQI Heatmap (Residential)')
sns.heatmap(heatmap_data, cmap='YlGnBu', annot=True, fmt='.0f',
cbar=True, linewidths=0.5)
plt.show()
```


AQI Heatmap (Industrial)
City
Bangalore 83 93 109 119 80 86 97 92 81 96 76 70 117 91 95 112 144 165 101 169 148 179 111
Chennai 114 78 115 96 104 117 99 106 82 85 74 107 120 74 112 96 118 166 168 158 173 127 164
Delhi 83 73 82 78 110 103 102 109 92 76 74 85 118 86 106 88 150 102 142 158 138 115 114
Kolkata 101 83 73 94 74 94 97 116 85 94 112 92 92 71 97 93 172 132 155 122 169 168 125
Mumbai 97 96 76 85 82 81 98 106 93 83 92 107 71 93 78 113 163 129 106 100 136 164 114
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
Year
160
140
120
100
80


AQI Heatmap (Residential)
Bangalore 35 60 30 47 34 69 46 78 37 50 47 56 43 36 57 36 100 120 95 105 77 107 101
Chennai 53 30 61 38 69 35 76 73 71 54 56 31 35 53 75 45 85 84 55 66 111 109 105
Delhi 60 45 48 53 33 47 73 77 78 62 40 47 51 68 73 38 117 66 114 78 115 116 110
Kolkata 45 34 55 63 63 58 80 47 73 74 70 59 68 53 36 47 110 83 60 66 103 90 115
Mumbai 66 69 46 47 80 58 40 61 80 51 37 42 45 77 31 50 112 79 105 88 88 75 97
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
Year
120
110
100
90
80
70
60
50
40
30

The visualizations and data analysis provide insights into the correlation between industrial areas and the increase in pollution over time. We focused on five major Indian cities: Mumbai, Delhi, Bangalore, Chennai, and Kolkata.

The stacked bar chart reveals the cumulative AQI (Air Quality Index) for both industrial and residential areas in each city over a span of two decades.
We observe that industrial areas generally contribute significantly to the total AQI. In cities like Mumbai and Delhi, the impact of industrial areas on overall pollution levels is evident.

In the industrial AQI heatmap, we can observe that industrial areas tend to exhibit higher AQI values in cities where industrialization is prominent. For example, Delhi and Kolkata consistently show higher industrial AQI values.
In contrast, the residential AQI heatmap depicts relatively lower AQI values in residential areas compared to industrial areas, indicating that industrialization plays a significant role in pollution levels.

Result: Thus, AQI of different cities in India was examined, visualized and it can safely be inferred that there is a strong interconnection between the rise in pollution level and industrialisation.