رواد مصر الرقمية

NEXT
TECHNOLOGY DEVELOPMENT

Urban Air Quality impact on Heart Health

Data Analysis Specialist Track

- Instructor Name:

  Abdullah Kamal
- Team Members:

  Julie Nader Fathy Attia

  Hania Essam AbdelHalim Mostafa

Content:

Objective:

The primary objective of this project is to analyze the relationship between urban air quality and its impact on human health, with a specific focus on heart diseases. By leveraging data analysis techniques, we aim to identify trends, correlations, and patterns between air conditions and the incidence of cardiovascular diseases. The ultimate goal is to provide actionable insights that can guide public health policies and interventions to mitigate the adverse effects of poor air quality on heart health.

Project Description:

This project investigates the growing concern of urban air pollution and its detrimental impact on human health, particularly the increasing rates of heart diseases. The analysis involves gathering and processing air quality data (such as levels of particulate matter, humidity, pressure, UV index and solar radiation) and health data (including heart disease rates and health risks). Using Python, Tableau, and other analytical tools, we examine the correlation between air quality metrics and cardiovascular health outcomes across different urban areas.

The findings from this project are expected to highlight key air conditions contributing to heart disease and provide recommendations for mitigating their effects through data-driven public health strategies. The project also emphasizes the importance of environmental conditions affecting health specifically heart diseases.

Members and Roles		
Process	Description	Name
1- Find Idea	Clearly identify the problem: understanding how urban air quality affects human health, specifically focusing on heart disease. • Search on the link between air quality and health. • Set the project's goal to find correlations between air conditions and heart disease. • Define measurable insights (e.g., the percentage increase in heart disease linked to poor air conditions).	Julie Nader Hania Essam
2- Data Collection	• "Urban Air Quality Data" • "Heart Data"	Julie Nader Hania Essam
3- Data Cleaning & Preprocessing	• Inspect the datasets for missing or incorrect values. • Remove outliers and irrelevant variables. • Merge datasets based on shared features. • Create new variables, like calculating average pollution exposure per region.	Hania Essam
4- Exploratory Data Analysis (EDA)	Gain a better understanding of the data through statistical summaries and visualizations. • Use descriptive statistics to summarize the data (mean, median, standard deviation, etc.). • Visualize the data .	Julie Nader

5- Hypothesis Testing, Correlation Analysis & Insights	• Test whether significant relationships exist between air quality and heart disease, and interpret the results to derive data-driven insights. • Use statistical analysis to test our hypotheses. • Evaluate the strength and significance of relationships. • Summarize key findings, such as conditions most strongly linked to heart diseases. • Use visualizations (e.g., heat maps, bar charts) to effectively communicate the results. • Highlight any trends.	Julie Nader
6- Documentation & Reporting	• Create a detailed report that covers all aspects of the project. • Describe the process, tools, data, and methods used. • Highlight key findings and include visualizations to support your conclusions.	Hania Essam

Working & Thinking Methodology

Throughout this project, my colleague and I adopted a ***parallel methodology*** in both our work and thinking. This approach allowed us to efficiently divide tasks and address multiple aspects of the project simultaneously. By working in parallel, we could explore different angles of the data analysis, share ideas quickly, solve problems together, and combine our findings more efficiently.

Before Analysis:

- Limited Understanding: There was uncertainty about the specific relationship between air quality conditions and heart diseases.
- Underutilized Data: Available air quality and health data were often isolated, preventing insights that could inform public health actions.
- Reactive Health Responses: Public health responses to air quality-related health issues were often reactive rather than proactive.
- Low Public Awareness: The general public lacked knowledge about the specific health risks associated with poor air conditions, limiting community involvement in initiatives.

After Analysis:

- Data-Driven Decisions: Integrated analysis enabled health organizations to develop evidence-based engagement and allocate resources effectively.
- Proactive Strategies: Insights led to proactive public health measures to mitigate health risks from air quality, improving overall community health.
- Increased Public Awareness: Clear findings communicated to the public raised awareness about the health impacts of air conditions, encouraging community engagement.

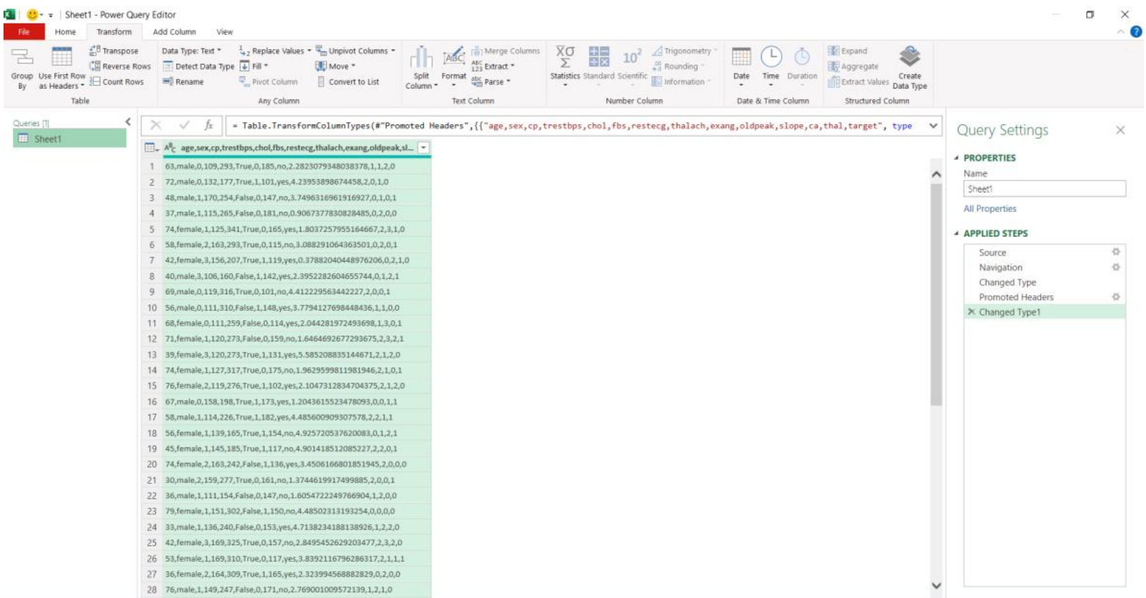
Overall, the analysis transformed data into actionable insights, leading to better health strategies, increased public awareness, and improved health outcomes related to air quality.

Project Stages:

- Data Source: [Kaggle: Your Machine Learning and Data Science Community]

- Cleaning:

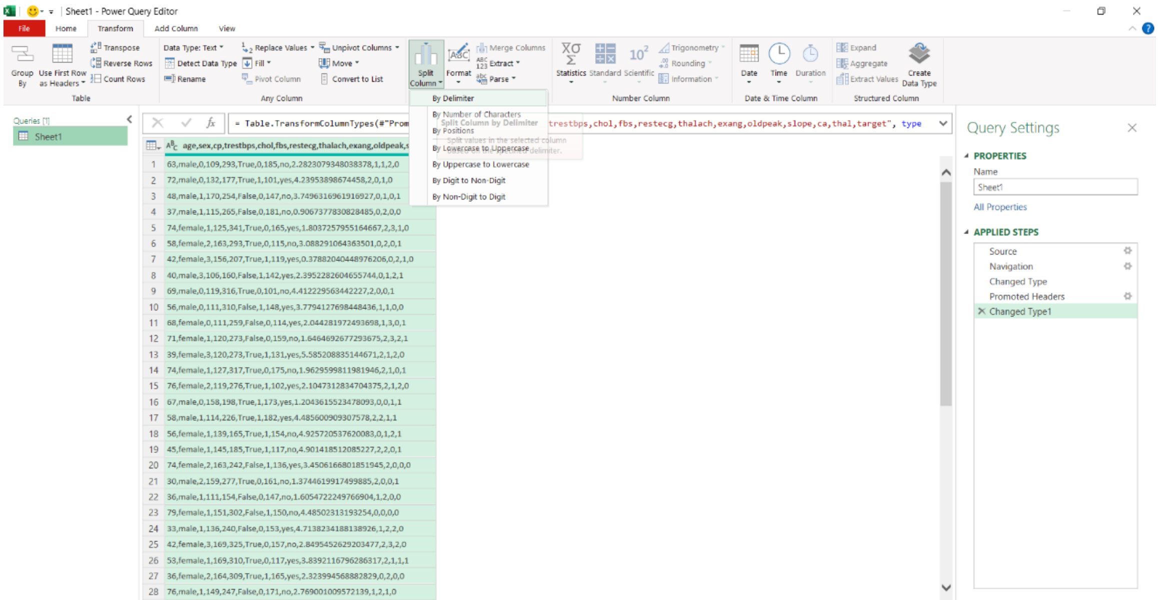
1. Excel Power Query



1. Open Power Query Editor

- In Excel, go to the **Data** tab and select **Get & Transform Data.**

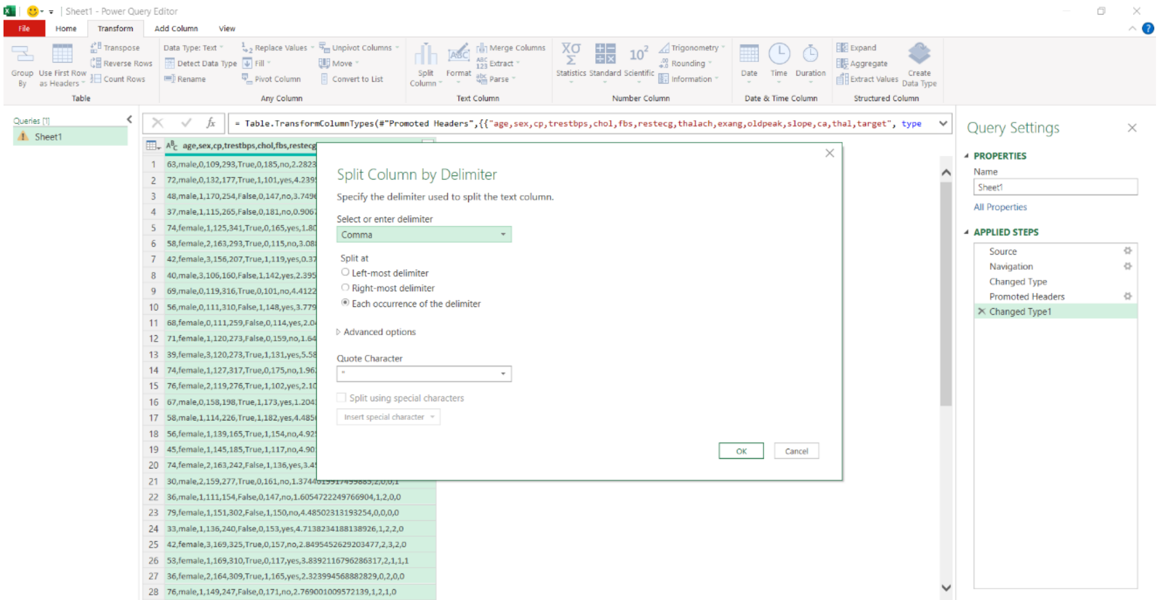
- Choose my dataset ‘ HeartData ’ and click **From Table/Range** to open it in Power Query Editor.



2. Select the Column to Split

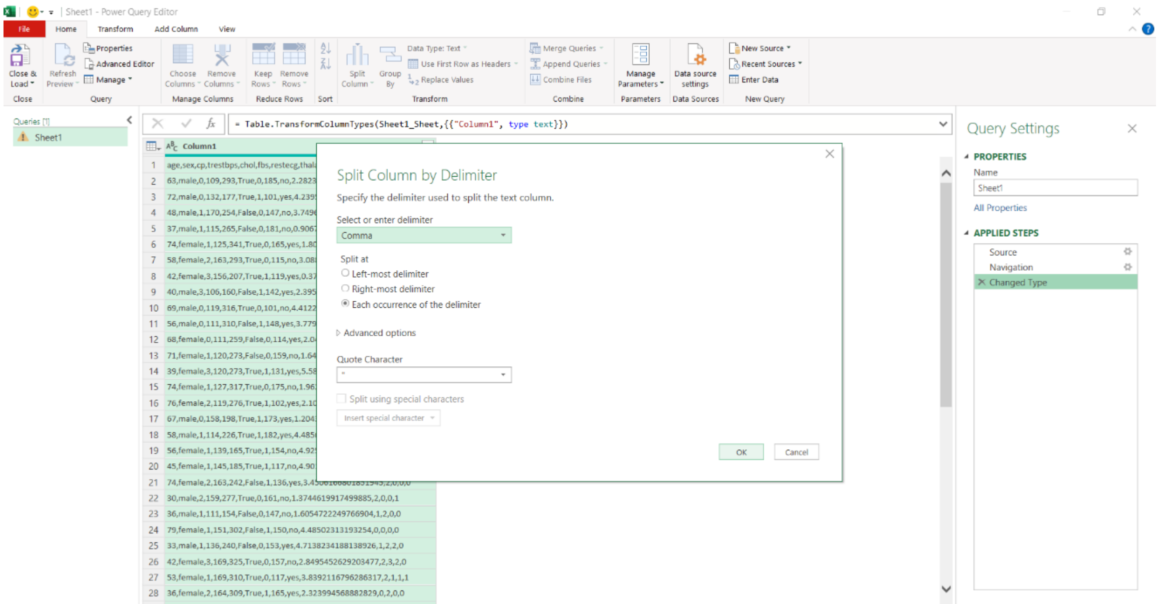
- In the Power Query Editor, locate the column that contains the combined data.

- Select the column by clicking on the column header



3. Apply Split by Delimiter

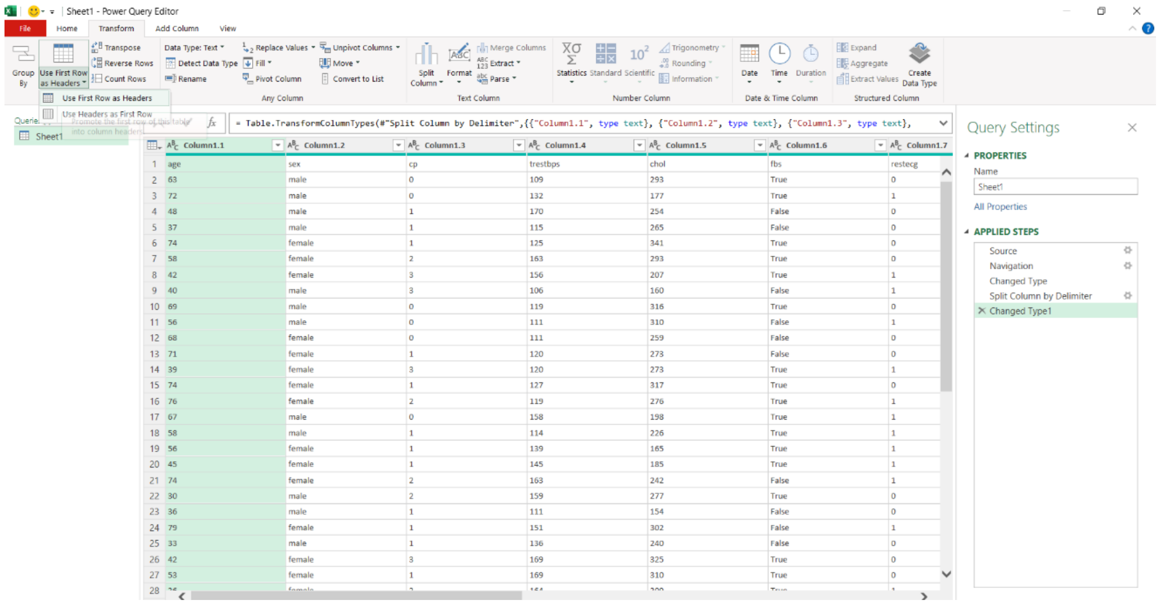
- Go to the **Transform** tab in the ribbon.
- Click on **Split Column** → By **Delimiter**.
- Choose the delimiter **comma** from the dropdown list.



4. Choose the Splitting Option

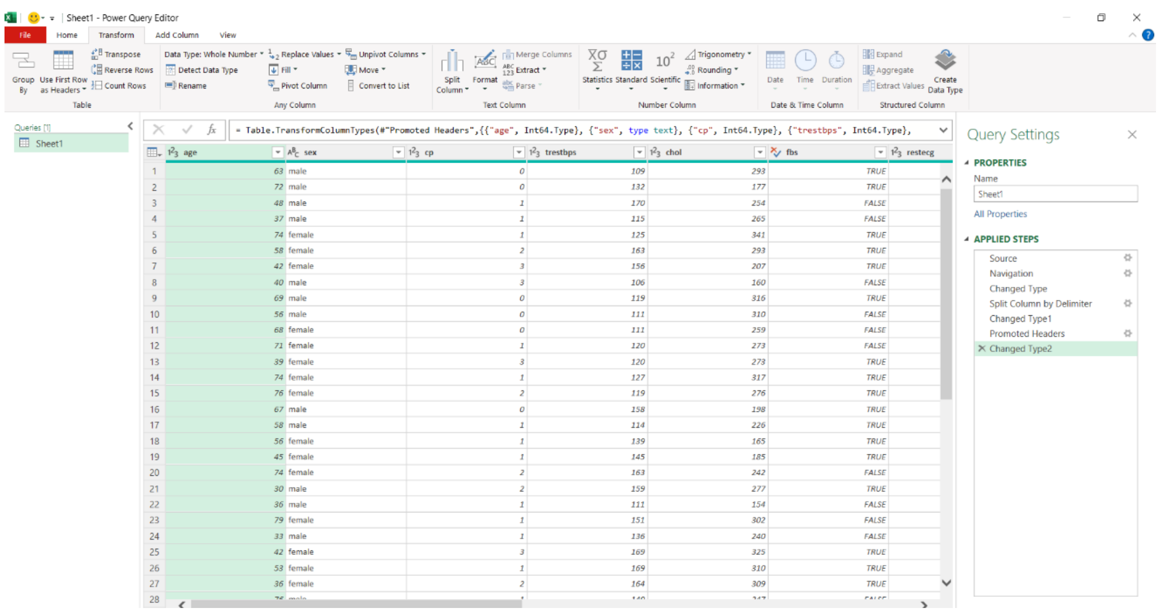
- We prompted to choose how to split the data:

At each occurrence of the delimiter: This option will create multiple columns based on every instance of the delimiter.

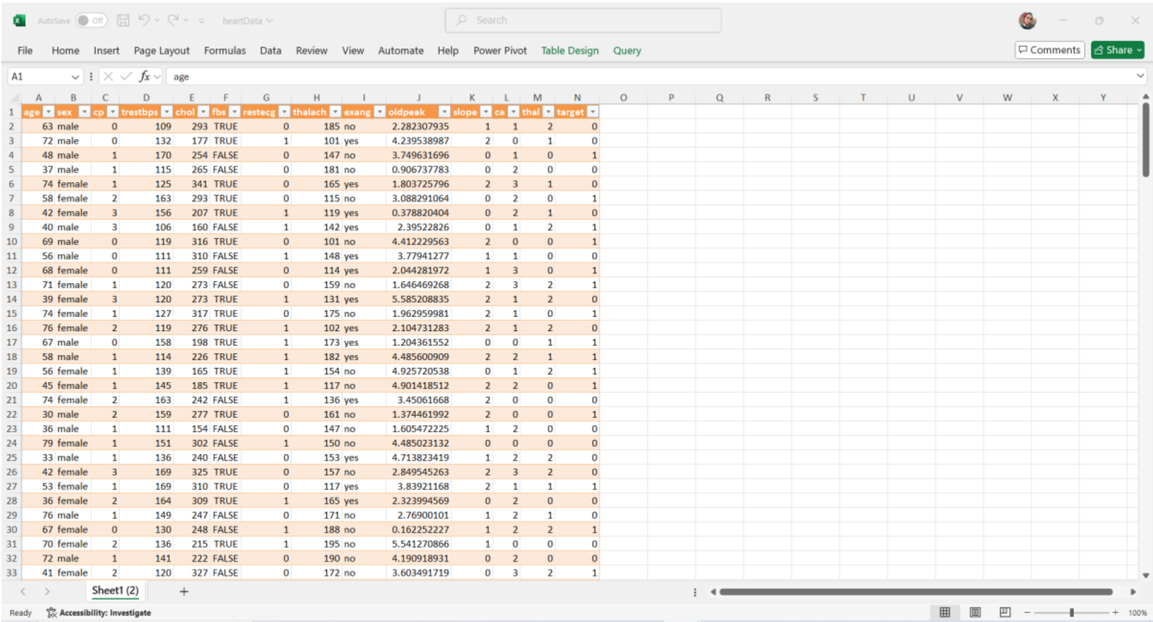


5. Adjust Column by using first row as Headers

- After the data is split, adjust columns and rename the new columns to more meaningful names.



Column names (age, sex, cp = chest pain, trestbps = resting blood pressure, chol = serum cholesterol, fbs = fasting blood sugar, restecg = resting electrocardiographic, thalach =maximum heart rate , exang = exercise induced angina, oldpeak = ST depression induced by exercise relative to rest, slope, chest pain, thalassemia, target)



age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	male	0	109	293	TRUE	0	185	no	2.282307935	1	1	2	0
72	male	0	132	177	TRUE	1	101	yes	4.239538987	2	0	1	0
48	male	1	170	254	FALSE	0	147	no	3.749631696	0	1	0	1
37	male	1	115	265	FALSE	0	181	no	0.906737783	0	2	0	0
74	female	1	125	341	TRUE	0	165	yes	1.803725796	2	3	1	0
58	female	2	163	293	TRUE	0	115	no	3.088291064	0	2	0	1
42	female	3	156	207	TRUE	1	119	yes	0.378820404	0	2	1	0
40	male	3	106	160	FALSE	1	142	yes	2.39522826	0	1	2	1
69	male	0	119	316	TRUE	0	101	no	4.412229563	2	0	0	1
56	male	0	111	310	FALSE	1	148	yes	3.77941277	1	1	0	0
68	female	0	111	259	FALSE	0	114	yes	2.044281972	1	3	0	1
71	female	1	120	273	FALSE	0	159	no	1.646469268	2	3	2	1
39	female	3	120	273	TRUE	1	131	yes	5.585208835	2	1	2	0
74	female	1	127	317	TRUE	0	175	no	1.962959981	2	1	0	1
76	female	2	119	276	TRUE	1	102	yes	2.104731283	2	1	2	0
67	male	0	158	198	TRUE	1	173	yes	1.204361552	0	0	1	1
58	male	1	114	226	TRUE	1	182	yes	4.485600909	2	2	1	1
56	female	1	139	165	TRUE	1	154	no	4.925720538	0	1	2	1
45	female	1	145	185	TRUE	1	117	no	4.901418512	2	2	0	1
74	female	2	163	242	FALSE	1	136	yes	3.45061668	2	0	0	0
30	male	2	159	277	TRUE	0	161	no	1.374461992	2	0	0	1
36	male	1	111	154	FALSE	0	147	no	1.605472225	1	2	0	0
79	female	1	151	302	FALSE	1	150	no	4.485023132	0	0	0	0
33	male	1	136	240	FALSE	0	153	yes	4.713823419	1	2	2	0
42	female	3	169	325	TRUE	0	157	no	2.849545263	2	3	2	0
53	female	1	169	310	TRUE	0	117	yes	3.83921168	2	1	1	1
36	female	2	164	309	TRUE	1	165	yes	2.323994569	0	2	0	0
76	male	1	149	247	FALSE	0	171	no	2.76900101	1	2	1	0
67	female	0	130	248	FALSE	1	188	no	0.162252227	1	2	2	1
70	female	2	136	215	TRUE	1	195	no	5.541270866	1	0	0	0
72	male	1	141	222	FALSE	0	190	no	4.190918931	0	2	0	0
41	female	2	120	327	FALSE	0	172	no	3.603491719	0	3	2	1

6. Load the Data Back to Excel

- The transformation is complete, click **Close & Load** to return the transformed data to Excel.

2. Python

1-Importing Data:

In the cleaning stage, we utilize Python for efficient data processing. Initially, we load two datasets "Urban Air Quality" and "Heart Data" using the pandas library, which allows for seamless handling and manipulation of CSV files.

```
import pandas as pd

df1 = pd.read_csv('heart-disease.csv')
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.300000	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.500000	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.400000	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.800000	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.600000	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
996	71	male	0	168	255	1	0	127	yes	1.470407	1	3	1	0
997	72	female	0	195	392	0	1	185	yes	1.678716	2	0	1	1
998	59	male	2	113	260	0	0	112	yes	1.335594	0	3	1	0
999	44	female	3	116	270	0	0	191	yes	5.448970	2	0	2	0
1000	47	female	0	184	361	1	0	129	no	2.398824	2	0	0	0

```
[ ] dfurban = pd.read_csv('Urban Air Quality and Health Impact Dataset.csv')
    dfurban
```

	datetime	datetimeEpoch	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity	...
0	2024-09-07	1.725692e+09	106.100000	91.000000	98.500000	104.000000	88.100000	95.900000	51.500000	21.000000	...
1	2024-09-08	1.725779e+09	103.900000	87.000000	95.400000	100.500000	84.700000	92.300000	48.700000	21.500000	...
2	2024-09-09	1.725865e+09	105.000000	83.900000	94.700000	99.900000	81.600000	90.600000	41.700000	16.900000	...
3	2024-09-10	1.725952e+09	106.100000	81.200000	93.900000	100.600000	79.500000	89.800000	39.100000	15.700000	...
4	2024-09-11	1.726038e+09	106.100000	82.100000	94.000000	101.000000	80.000000	90.000000	40.100000	15.900000	...
...	...	...	...	...	...	...	...	...	...	...	...
995	2024-09-18	1.726633e+09	76.060546	64.359387	69.002142	77.673823	63.510920	67.003338	59.936776	73.189130	...
996	2024-09-17	1.726550e+09	68.409198	65.939319	66.567410	68.956722	64.805635	65.992526	59.010257	74.137401	...
997	2024-09-12	1.726122e+09	69.756690	65.286919	65.919492	68.158536	63.662942	67.313322	62.024442	84.650482	...
998	2024-09-14	1.726284e+09	77.106797	61.481724	68.106569	76.426959	60.901526	68.094309	63.169608	86.860261	...
999	2024-09-18	1.726618e+09	90.923080	79.296868	81.636991	94.180423	78.071851	84.987113	73.393045	74.734715	...

2-Data Refinement:

During this phase, all unclear or inconsistent data is transformed into well-defined and uniform values, enhancing the quality and integrity of the dataset for accurate analysis.

```
df['Chest_Pain_type'].replace({0: 'typical angina', 1: 'atypical angina', 2: 'non-anginal pain', 3 : 'asymptomatic'}, inplace=True)
print(df['Chest_Pain_type'])
```

```
0          asymptomatic
1      non-anginal pain
2       atypical angina
3       atypical angina
4        typical angina
             ...
995    non-anginal pain
996      typical angina
997      typical angina
998    non-anginal pain
999        asymptomatic
Name: Chest_Pain_type, Length: 1000, dtype: object
```

```
df['resting_electrocardiographic'].replace({0: 'normal', 1: 'Having ST-T wave abnormality', 2: 'Left ventricular hypertrophy'}, inplace=True)
print(df['resting_electrocardiographic'])
```

```
0                            normal
1      Having ST-T wave abnormality
2                            normal
3      Having ST-T wave abnormality
4      Having ST-T wave abnormality
                   ...
995                          normal
996                          normal
997    Having ST-T wave abnormality
998                          normal
999                          normal
Name: resting_electrocardiographic, Length: 1000, dtype: object
```

```
df['slope'].replace({0: 'upsloping', 1: 'flat', 2: 'downsloping'}, inplace=True)
print(df['slope'])
```

```
0        upsloping
1        upsloping
2      downsloping
3      downsloping
4      downsloping
          ...
995           flat
996           flat
997    downsloping
998      upsloping
999    downsloping
Name: slope, Length: 1000, dtype: object
```

```
df['coronary_artery_disease'].replace({0: 'no coloured', 1: '1 coloured', 2: '2 coloured', 3 : '3 coloured'}, inplace=True)
print(df['coronary_artery_disease'])
```

```
0      no coloured
1      no coloured
2      no coloured
3      no coloured
4      no coloured
          ...
995     1 coloured
996     3 coloured
997    no coloured
998     3 coloured
999    no coloured
Name: coronary_artery_disease, Length: 1000, dtype: object
```

```
df['thalassemia'].replace({ 1: 'normal', 2: 'fixed defect', 3 : 'reversible defect'}, inplace=True)
print(df['thalassemia'])
```

```
0            normal
1      fixed defect
2      fixed defect
3      fixed defect
4      fixed defect
           ...
995          normal
996          normal
997          normal
998          normal
999    fixed defect
Name: thalassemia, Length: 1000, dtype: object
```

```
df1['exang'] = df1['exang'].astype(int)
df1['exang'].replace(0, 'no', inplace=True)
df1['exang'].replace(1, 'yes', inplace=True)
df1
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	male	3	145	233	1	0	150	no	2.3	0	0	1	1
1	37	male	2	130	250	0	1	187	no	3.5	0	0	2	1
2	41	female	1	130	204	0	0	172	no	1.4	2	0	2	1
3	56	male	1	120	236	0	1	178	no	0.8	2	0	2	1
4	57	female	0	120	354	0	1	163	yes	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	female	0	140	241	0	1	123	yes	0.2	1	0	3	0
299	45	male	3	110	264	0	1	132	no	1.2	1	0	3	0
300	68	male	0	144	193	1	1	141	no	3.4	1	2	3	0
301	57	male	0	130	131	0	1	115	yes	1.2	1	1	3	0
302	57	female	1	130	236	0	0	174	no	0.0	1	1	2	0

```
df1['sex'].replace(1, 'male', inplace=True)
df1['sex'].replace(0, 'female', inplace=True)
df1
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	male	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	male	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	female	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	male	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	female	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	female	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	male	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	male	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	male	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	female	1	130	236	0	0	174	0	0.0	1	1	2	0

3- Data Merging:

After cleaning, the two datasets are merged using an outer join to ensure that all records from both the "Urban Air Quality" and "Heart Data" datasets are included. The merged data is then saved as a single CSV file, ready for further analysis and download.

```
[ ] df_merged = pd.merge(dfurban, dfhearthealth, on='ID', how='outer')
    df_merged.to_csv('merged_file.csv', index=False)
```

```
[ ] df_merged
```

	datetime	datetimeEpoch	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity	...	chol	fbs	rest
0	2024-09-07	1.725692e+09	106.100000	91.000000	98.500000	104.000000	88.100000	95.900000	51.500000	21.000000	...	233	1	
1	2024-09-08	1.725779e+09	103.900000	87.000000	95.400000	100.500000	84.700000	92.300000	48.700000	21.500000	...	250	0	
2	2024-09-09	1.725865e+09	105.000000	83.900000	94.700000	99.900000	81.600000	90.600000	41.700000	16.900000	...	204	0	
3	2024-09-10	1.725952e+09	106.100000	81.200000	93.900000	100.600000	79.500000	89.800000	39.100000	15.700000	...	236	0	
4	2024-09-11	1.726038e+09	106.100000	82.100000	94.000000	101.000000	80.000000	90.000000	40.100000	15.900000	...	354	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
996	2024-09-17	1.726550e+09	68.409198	65.939319	66.567410	68.956722	64.805635	65.992526	59.010257	74.137401	...	255	1	
997	2024-09-12	1.726122e+09	69.756690	65.286919	65.919492	68.158536	63.662942	67.313322	62.024442	84.650482	...	392	0	
998	2024-09-14	1.726284e+09	77.106797	61.481724	68.106569	76.426959	60.901526	68.094309	63.169608	86.860261	...	260	0	
999	2024-09-18	1.726618e+09	90.923080	79.296868	81.636991	94.180423	78.071851	84.987113	73.393045	74.734715	...	270	0	

```
[ ] df_merged.to_csv('merged_file.csv', index=False)
```

4-Post-Merge Data Refinement:

Following the merging process, we identified null values and ambiguous data entries. A second cleaning phase was conducted to address these issues, filling the nulls and ensuring the data was further clarified for enhanced accuracy and completeness.

```
[ ] df_merged['fbs'] = df_merged['fbs'].astype(int)
    df_merged['fbs'].replace(0, 'false', inplace=True)
    df_merged['fbs'].replace(1, 'true', inplace=True)
    df_merged
```

	datetime	datetimeEpoch	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity	...	chol	fbs	res
0	2024-09-07	1.725692e+09	106.100000	91.000000	98.500000	104.000000	88.100000	95.900000	51.500000	21.000000	...	233	true	
1	2024-09-08	1.725779e+09	103.900000	87.000000	95.400000	100.500000	84.700000	92.300000	48.700000	21.500000	...	250	false	
2	2024-09-09	1.725865e+09	105.000000	83.900000	94.700000	99.900000	81.600000	90.600000	41.700000	16.900000	...	204	false	
3	2024-09-10	1.725952e+09	106.100000	81.200000	93.900000	100.600000	79.500000	89.800000	39.100000	15.700000	...	236	false	
4	2024-09-11	1.726038e+09	106.100000	82.100000	94.000000	101.000000	80.000000	90.000000	40.100000	15.900000	...	354	false	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
996	2024-09-17	1.726550e+09	68.409198	65.939319	66.567410	68.956722	64.805635	65.992526	59.010257	74.137401	...	255	true	
997	2024-09-12	1.726122e+09	69.756690	65.286919	65.919492	68.158536	63.662942	67.313322	62.024442	84.650482	...	392	false	
998	2024-09-14	1.726284e+09	77.106797	61.481724	68.106569	76.426959	60.901526	68.094309	63.169608	86.860261	...	260	false	
999	2024-09-18	1.726618e+09	90.923080	79.296868	81.636991	94.180423	78.071851	84.987113	73.393045	74.734715	...	270	false	

```
df['preciptype'].fillna('no rain', inplace=True)
df
```

	datetime	datetimeEpoch	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity	...	cholestrol	fasting_blood_sugar	resting_electrocardiogr
0	7/9/2024	1725692400	106.100000	91.000000	98.500000	104.000000	88.100000	95.900000	51.500000	21.000000	...	233	True	
1	8/9/2024	1725778800	103.900000	87.000000	95.400000	100.500000	84.700000	92.300000	48.700000	21.500000	...	250	False	
2	9/9/2024	1725865200	105.000000	83.900000	94.700000	99.900000	81.600000	90.600000	41.700000	16.900000	...	204	False	
3	10/9/2024	1725951600	106.100000	81.200000	93.900000	100.600000	79.500000	89.800000	39.100000	15.700000	...	236	False	
4	11/9/2024	1726038000	106.100000	82.100000	94.000000	101.000000	80.000000	90.000000	40.100000	15.900000	...	354	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
995	18/09/2024	1726632592	76.060546	64.359387	69.002142	77.673823	63.510920	67.003338	59.936776	73.189130	...	289	False	
996	17/09/2024	1726550178	68.409198	65.939319	66.567410	68.956722	64.805635	65.992526	59.010257	74.137401	...	255	True	
997	12/9/2024	1726121998	69.756690	65.286919	65.919492	68.158536	63.662942	67.313322	62.024442	84.650482	...	392	False	
998	14/09/2024	1726283774	77.106797	61.481724	68.106569	76.426959	60.901526	68.094309	63.169608	86.860261	...	260	False	
999	18/09/2024	1726618425	90.923080	79.296868	81.636991	94.180423	78.071851	84.987113	73.393045	74.734715	...	270	False	

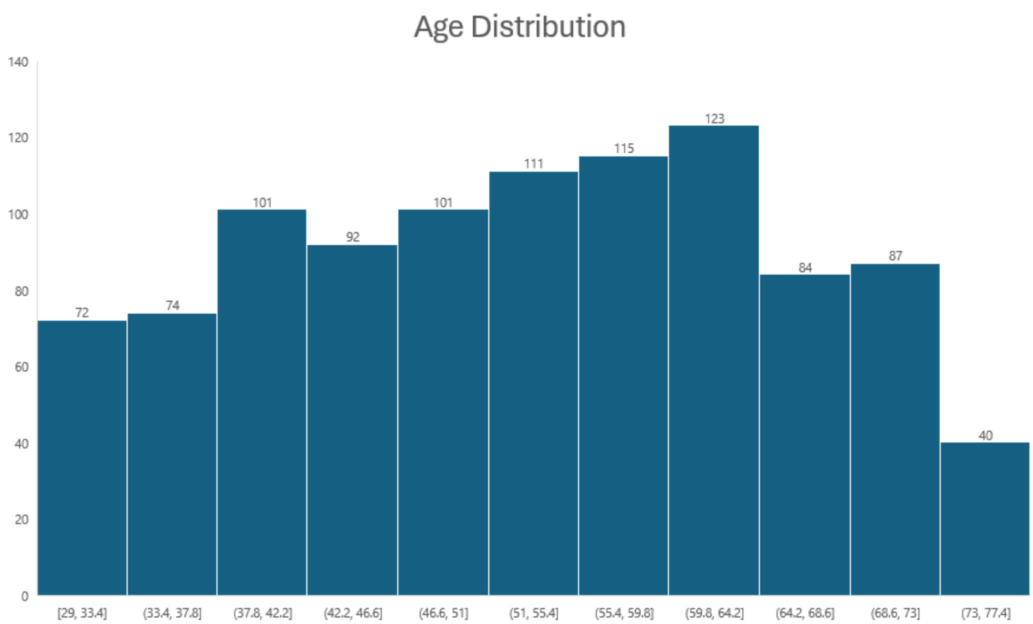
1000 rows × 58 columns

```
print(df['preciptype'])
```

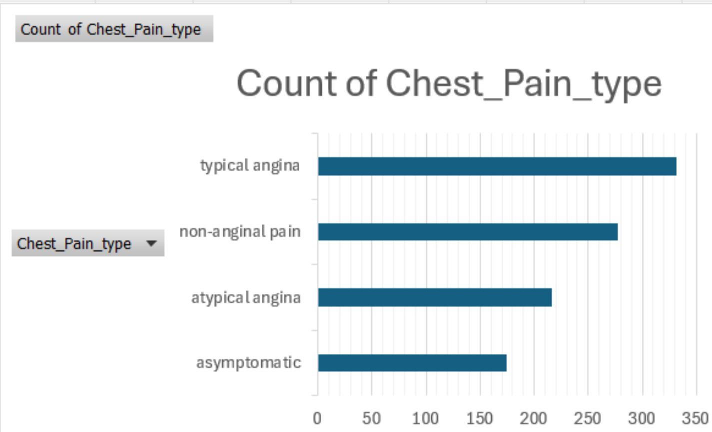
```
0       no rain
1       no rain
2       no rain
3      ['rain']
4      ['rain']
         ...
995    ['rain']
996    ['rain']
997     no rain
998     no rain
999    ['rain']
Name: preciptype, Length: 1000, dtype: object
```

- Exploratory Data Analysis (EDA)

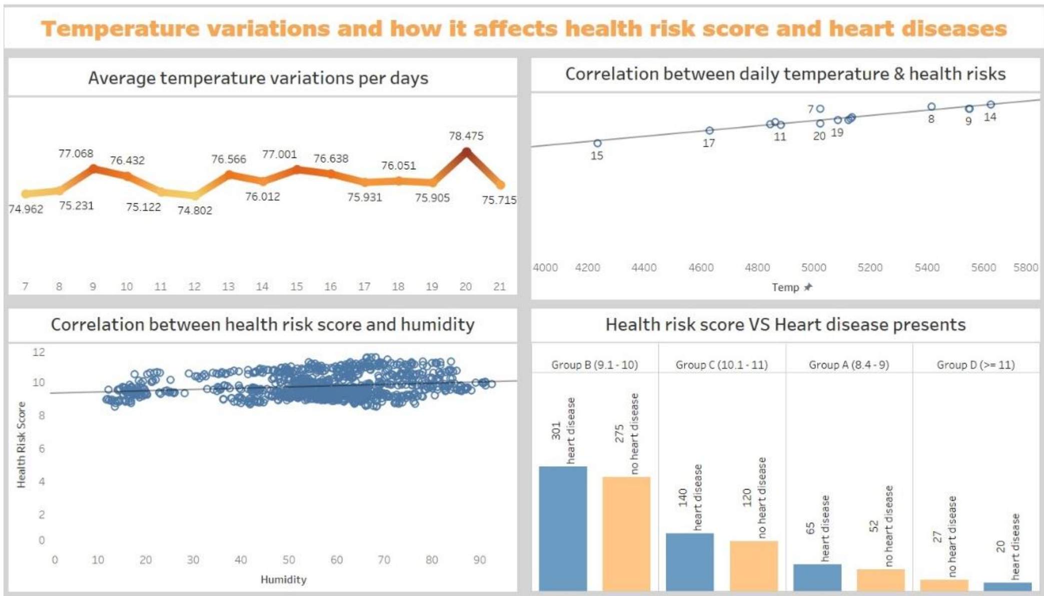
Column Name	Mean	Median	Standard Deviation
Temperature	76.11596739	75.2227509	8.717708116
Feelslike	85.19538425	84.2840775	9.492201592
Cloud Cover	24.16015046	17.3068815	22.42754986
Heat Index	80.19561315	78.5623453	6.050777393
Sever risk	12.92369467	10.0809192	8.834437439
Resting blood sugar	141.234	137	29.28301972
Cholesterol	256.628	251	71.40656564
Maximum heart rate	142.754	146.5	35.98807419



Chest_Pain_type	Count of Chest_Pain_type
asymptomatic	175
atypical angina	216
non-anginal pain	277
typical angina	332



Report (Visualization and Insights):



This dashboard explores the relationships between environmental factors like temperature and humidity and their impact on health risks.

1. Correlation between Health Risk Score and Humidity

- The scatter plot shows the relationship between health risk scores and humidity levels. There appears to be **a slight positive correlation**, indicated by the upward trend, where increasing humidity is associated with higher health risk scores.

- **Higher** humidity may **elevate** health risk scores, possibly due to the stress it places on the body's ability to regulate temperature. This can lead to **increased risks**, especially for people with heart conditions.

2. Health Risk Score vs Heart Disease Presence

- This bar chart categorizes individuals into four groups based on their health risk scores and examines how many within each group have heart disease.

  - Group A (8.4 - 9): 65 people with heart disease, 52 without.

  - Group B (9.1 - 10): 301 people with heart disease, 275 without.

  - Group C (10.1 - 11): 140 people with heart disease, 120 without.

  - Group D (>= 11): 20 people with heart disease, 27 without.

- As the health risk score increases, the number of people with heart disease tends to decrease after peaking in **Group B**. This suggests that individuals with moderate health risk scores are more likely to heart disease, while the very high scores may relate to other health issues beyond heart disease.

3. Average Temperature Variations per Days

-The line chart illustrates the fluctuation in daily average temperatures over a specific period. The values range from approximately **74.962°F to 78.475°F**, showing **slight** but consistent variations.

- The small daily temperature variations are likely impacting health risk scores. Periods of **slightly higher** temperatures (like on day 19 with 78.475°F) might be associated with **increased** health risks due to heat stress.

4. Correlation between Daily Temperature & Health Risks

- The scatter plot shows a **positive correlation** between daily temperatures and health risk scores. As the temperature **rises**, the health risk score also tends to **increase**, which is indicated by the upward trend line.

Conclusion

- The dashboard highlights clear correlations between environmental factors like temperature and humidity and health risks, especially concerning heart disease.

- Higher humidity appears to correlate with increased health risks, possibly due to the body's difficulty in cooling down in humid conditions.

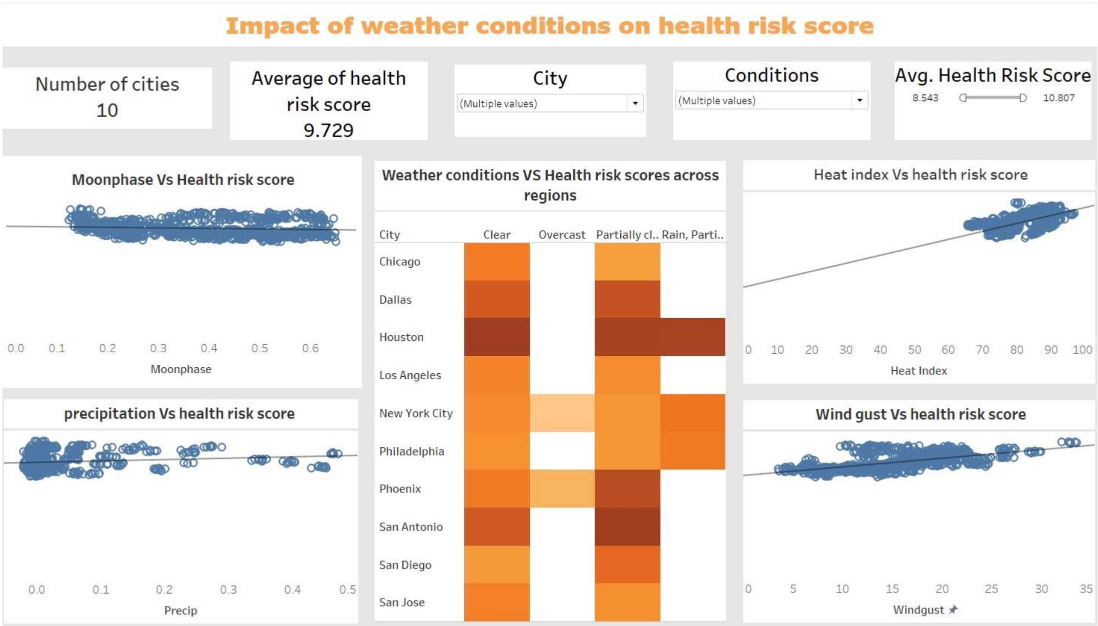
- Moderate health risk scores (Groups B and C) show a significant number of people with heart diseases, indicating that moderate levels of risk factors (temperature, humidity, etc.) might be the most dangerous.

- Temperature fluctuations have a noticeable impact, with slight increases in temperature leading to elevated health risk scores, further emphasizing the need for monitoring environmental conditions to prevent heat-related health issues.

Notes

Health Risk Score Groups (8.4 - 11.5):

1. **Low Risk (8.4 - 8.9)**:
   Individuals in this category present light health risk factors. Routine monitoring and preventive health measures are recommended to maintain well-being.
2. **Moderate Risk (9.0 - 9.4)**:
   This group presents moderate health risks. Periodic evaluations and proactive health management strategies are advised to prevent the progression of risk factors.
3. **High Risk (9.5 - 10.4)**:
   Individuals in this category are at elevated risk due to the presence of multiple or significant health risk factors. Comprehensive health assessments and targeted interventions are required.
4. **Very High Risk (10.5 - 11.5)**:
   The highest risk group, where immediate and intensive health management is essential to prevent severe health outcomes. Continuous monitoring and intervention are critical.



The relationship between moon phases, weather conditions (such as wind gusts, heat index, and precipitation), and health risks is a complex topic with various factors influencing human health. While there is no direct relation between these factors and (health risk score), research shows that environmental conditions do impact human health in different ways.

1. Moon Phases:

- Sleep Patterns: Some studies suggest that the full moon can influence sleep quality, potentially leading to sleep deprivation, which may affect mood, mental health, and immune function.

- Mental Health: There are **unscientific reports** and limited studies suggesting that certain moon phases, especially the full moon, can impact mood disorders, though scientific evidence is inconclusive.

2. Weather Conditions:

- Temperature and Heat Index: High temperatures and a high heat index (a measure that factors in humidity) can increase the risk of **heat-related illnesses** such as heat exhaustion or heat stroke.

- Heat can worsen respiratory and cardiovascular issues.

- Extreme heat can also affect mental health, causing irritability, fatigue, and anxiety.

- Cold Weather: Low temperatures can worsen cardiovascular diseases, particularly in older adults, by causing blood vessels to compress and raising blood pressure.

- Cold can also worsen respiratory conditions such as asthma.

3. Wind Gusts:

- Strong wind gusts can have both direct and indirect impacts on health.

- Allergies and Respiratory Issues: Winds can spread allergens like pollen and pollutants, leading to respiratory problems for people with asthma or allergies.

- Injury Risk: During strong winds, there’s an increased risk of accidents or injuries from flying waste or falls, especially during storms.

4. Precipitation (Rain):

- Increased Risk of Respiratory Infections: Wet and damp conditions can lead to an increase in mold, which can trigger respiratory problems. Also, sudden changes in temperature due to rain can increase the chances of catching a cold or flu.

- Accidents: Wet conditions can increase the risk of slipping and accidents, contributing to injury risks.

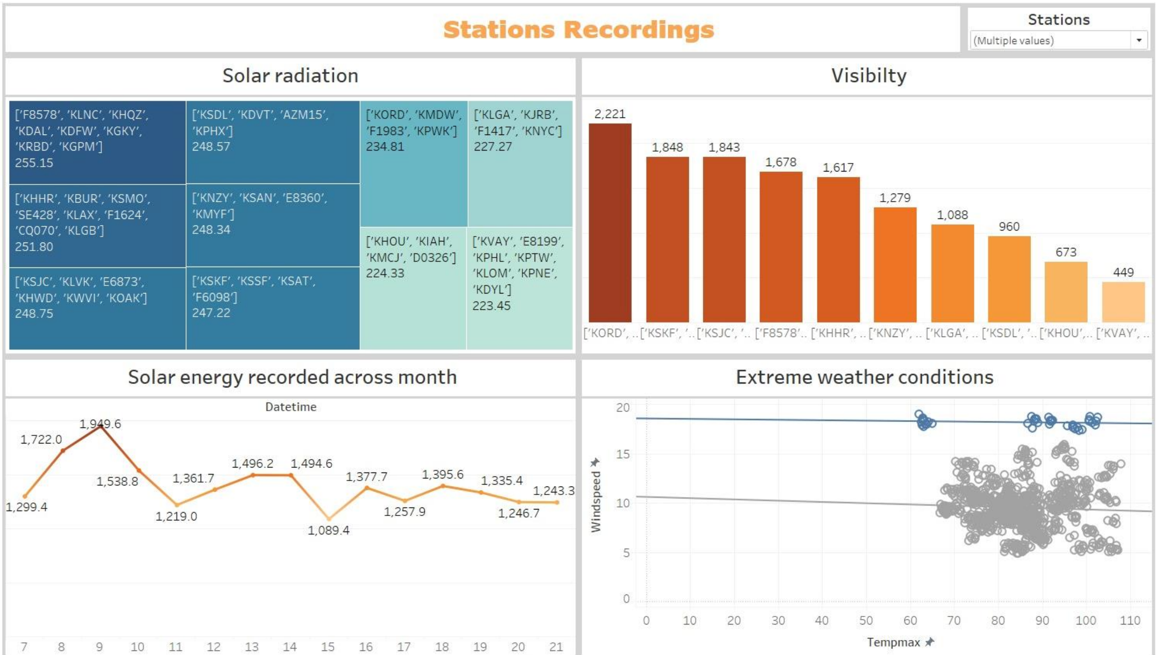
5. Health Risk Score:

- The concept of a "health risk score" typically incorporates various personal health metrics and environmental factors to assess an individual's risk for specific conditions.

- Environmental Factors: Incorporating weather conditions, moon phases, and climate factors can be part of a broader health risk model, particularly for people with heart conditions.

Conclusion

Environmental factors like heat index, wind gusts, and precipitation can significantly influence health, especially for those with pre-existing conditions. Moon phases, while less scientifically substantiated, are sometimes linked to mood and sleep disturbances.



This dashboard titled "**Stations Recordings**" focuses on solar radiation, visibility, solar energy over time, and extreme weather conditions. It provides an overview of various environmental recordings from different stations. Here's a breakdown and key insights from each section:

1. Solar Radiation by Station

-This treemap displays the average solar radiation values recorded by different stations.

- The **highest** solar radiation was recorded at stations like 'F8578', 'KLNC', 'KHQZ', 'KDAL', etc., with a value of **255.15**.

- The **lowest** recorded value is **223.45** by stations such as 'KVAY', 'E8199', 'KPHL', etc.

-Different regions present varied levels of solar radiation, with the **highest** levels found in stations from groups like '**F8578'** and the **lowest** in groups like '**KVAY'**. Higher solar radiation areas may *experience more extreme heat conditions, contributing to higher health risks* in those regions.

2. Visibility by Station

-This bar chart illustrates visibility recordings by different stations.

- **Station 'KORD'** recorded the **highest** visibility value of **2,221**, followed by stations like **'KSKF'** and **'KSJC'**, with values around **1,800+.**

- The **lowest** visibility was recorded by stations like 'KVAY', with a value of **449**.

- Stations with lower visibility may be experiencing different weather conditions, such as fog, heavy rain, or pollution, which could impact transportation and increase accident risks. Poor visibility can also worsen respiratory problems due to pollutants.

3. Solar Energy Recorded Across the Month

- This line chart shows how solar energy levels varied throughout the month.

- The solar energy **peaked** at **1,949.6** on day 9 and had the **lowest** value of **1,089.4** on day 14.

- The data shows an instable pattern, with higher values around day 8-10 and then **a gradual decrease**.

- This instability indicates varying sunlight exposure, which could be linked to cloud cover, weather patterns, or seasonal changes. **Higher** solar energy may correlate with **warmer** temperatures, *potentially increasing heat-related health risks*.

4. Extreme Weather Conditions

- This scatter plot shows the relationship between maximum temperatures (Tempmax) and wind speeds recorded by the stations.

- As temperatures **rise above 60°F**, wind speeds remain **consistent** between **15-20 mph**.

- Higher temperatures combined with consistent wind speeds may indicate that **extreme heat events** are not being eased by significant wind, which could pose additional stress on human health, especially in areas with little airflow, *sustained high temperatures without cooling winds can lead to heat stress and dehydration.*

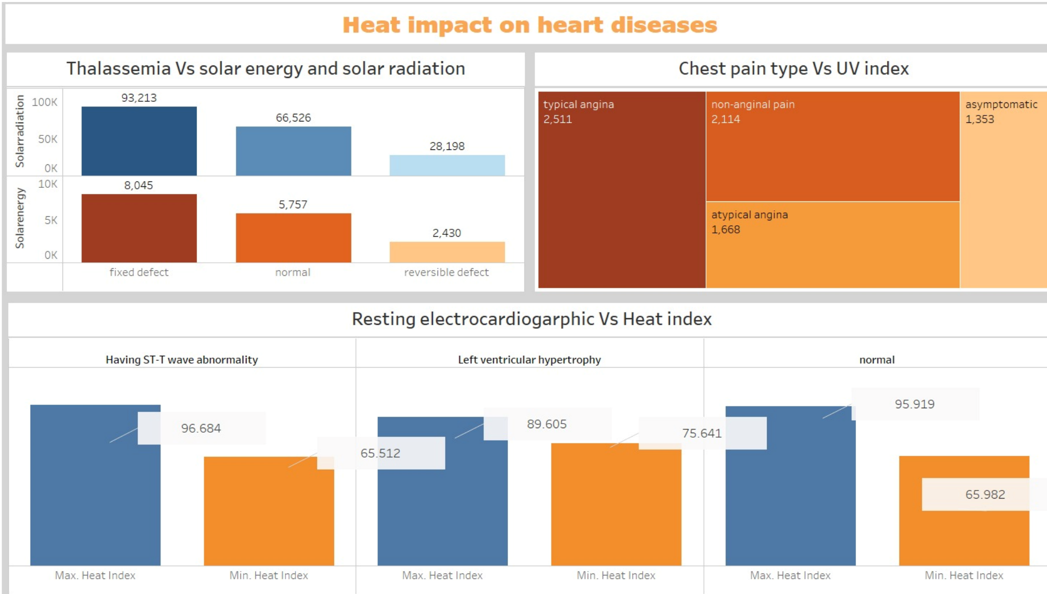
Conclusion:

- **Solar radiation** varies significantly across different stations, with some areas experiencing much higher levels than others, potentially leading to more extreme heat-related conditions.

- **Visibility** recordings reveal that certain stations have much lower visibility, suggesting different weather conditions that could impact both health and safety, particularly in terms of respiratory risks and traffic accidents.

- **Solar energy fluctuations** over the month suggest varying weather patterns, with spikes in solar energy potentially corresponding to warmer days, contributing to higher health risks.

- **Extreme weather conditions** show that, despite rising temperatures, wind speeds remain constant, which could amplify heat stress in populations during heat waves.



This dashboard provides a comprehensive analysis of the impact of heat, including solar energy, solar radiation, and heat index, on various heart conditions such as thalassemia, chest pain types, and electrocardiographic abnormalities.

1. Thalassemia vs Solar Energy and Solar Radiation:

- This bar chart compares the relationship between different types of thalassemia defects (fixed, normal, and reversible) and solar energy and solar radiation levels. For fixed defects, higher solar irradiation is observed (93,213) compared to normal and reversible defects, *where radiation levels drop significantly*. Solar energy is also higher for fixed defects (8,045) than for other categories.

2. Chest Pain Type vs UV Index:

- This treemap visualizes the relationship between various chest pain types and the UV index. **Atypical angina** and **asymptomatic** cases are shown with **lower** values, while **typical angina** and **non-anginal** pain **dominate** the chart. This suggests a variation in chest pain types depending on the UV index exposure.

3. Resting Electrocardiographic vs Heat Index:

- This bar chart describe how different heat index levels (both maximum and minimum) correlate with resting electrocardiographic conditions, such as ST-T wave abnormalities and left ventricular hypertrophy.

  - For **ST-T wave abnormalities**, the maximum heat index shows a higher impact (96.684), whereas the minimum heat index impact is lower (65.512).

  - Similarly, for **left ventricular hypertrophy**, higher impacts are seen with the maximum heat index (89.605), while lower impacts are seen with the minimum heat index (75.641).

- For individuals with **normal electrocardiograms**, the minimum heat index also tends to have a smaller impact (65.982) compared to the maximum (95.919).

Conclusion:

- Increased solar irradiation appears to be strongly associated with fixed thalassemia defects, whereas other conditions like reversible defects show lower correlations.

- The heat index seems to play a critical role in worsening electrocardiographic abnormalities, particularly with conditions like left ventricular hypertrophy.

This analysis provides valuable insights into how exposure to heat and solar **conditions might influence the risk** or **presence of various heart diseases**.

This dashboard titled "**Heart Disease Recordings**" focuses on various factors related to heart disease, including exercise-induced angina, chest pain types, and old peak values. Here's a breakdown of the main insights from the visualizations:

1. Number of Individuals Showing Signs of Exercise-Induced Angina

- The pie chart displays the number of individuals who show signs of exercise-induced angina.

- **430 individuals** showed signs of angina triggered by exercise.

- **570 individuals** did not exhibit this condition.

- A significant portion of individuals (about 43%) are experiencing exercise-induced angina, which is an indicator of potential coronary artery issues. This condition is often associated with heart disease, highlighting the need for closer monitoring of those showing symptoms.

2. Old Peak vs. Exercise-Induced Angina

- This bar chart compares the old peak (ST depression) for individuals with and without exercise-induced angina.

  - Individuals **without** exercise-induced angina have an average old peak of **2.2834**.

  - Individuals **with** exercise-induced angina have a slightly higher average old peak of **2.7255**.

- A higher old peak value among those with angina may suggest that these individuals have more clear heart issues, as old peak values are often indicative of heart muscle stress, which occurs when the heart is under stress during exercise.

3. Percentage of Chest Pain Types Among Individuals

- This pie chart shows the distribution of different chest pain types among the individuals.

  - **Typical angina** is the most common, affecting **33.2%** of individuals.

- **Non-anginal pain** affects **27.7%**, followed by **atypical angina** (21.6%) and **asymptomatic** individuals (17.5%).

- The most common chest pain type is typical angina, which is usually associated with heart disease. The significant percentage of individuals experiencing atypical or non-anginal pain indicates that heart disease can present with various symptoms, and not all cases will show the classical signs of angina.

4. Thalassemia vs. Gender

This stacked bar chart compares the occurrence of different types of thalassemia (fixed defect, normal, and reversible defect) across genders. The data highlights that males have a higher prevalence of reversible defects, while both genders are similarly affected by normal thalassemia.

5. Old Peak vs. Heart Disease

- This bar chart shows the comparison between old peak values in individuals with and without heart disease.

- Individuals **with** heart disease have an old peak value associated with **526 cases**.

- Individuals **without** heart disease represent **474 cases**.

- The old peak is more common among those with heart disease, as expected. Monitoring old peak values can help predict the likelihood of heart disease, making it crucial for diagnosing and managing cardiac issues.

Conclusions:

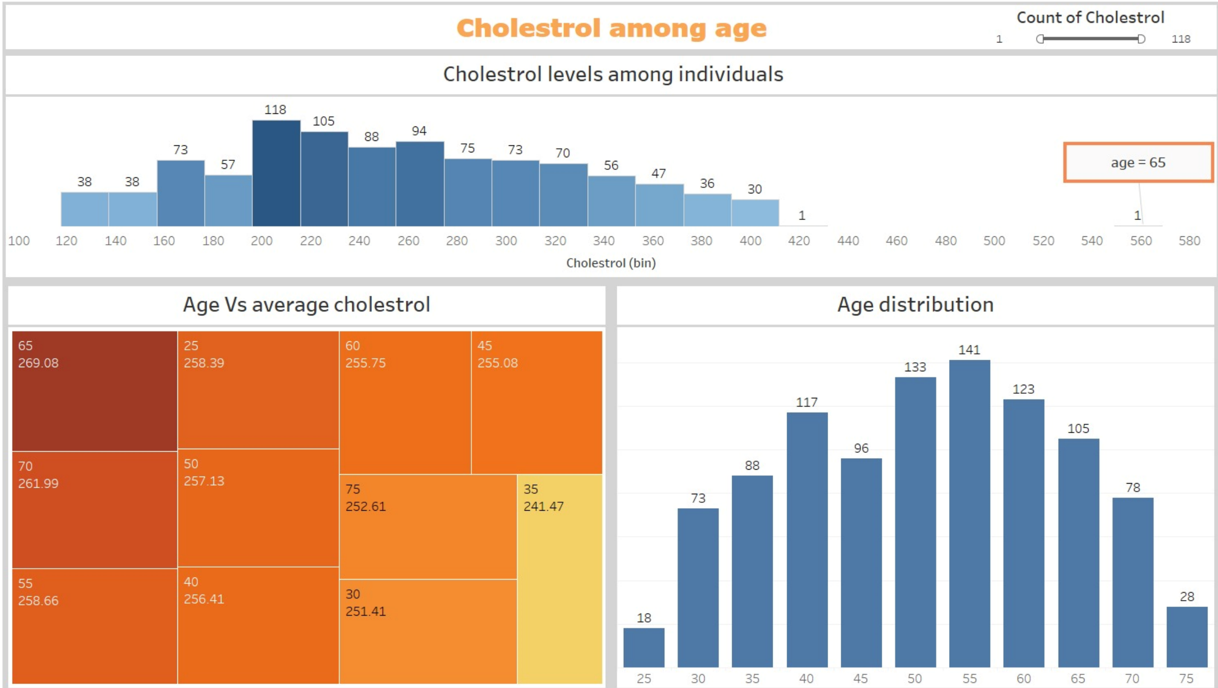
- A **significant proportion** of individuals show signs of **exercise-induced angina**, indicating a higher risk of heart problems for these individuals.

- The **old peak value** tends to be higher for individuals with both exercise-induced angina and heart disease, suggesting that those with higher old peak values are more likely to suffer from serious heart conditions.

- **Typical angina** is the most common type of chest pain, but a large number of individuals experience **non-anginal** and **atypical angina**, highlight the ways heart disease can present.

- The **average old peak value** indicates a moderate risk level across the population for heart-related issues, with a slight emphasis on those with existing heart conditions.

This dashboard effectively visualizes key indicators for heart disease, helping medical professionals assess risk factors like angina, chest pain types, and old peak values for better heart health management.



This dashboard provides an insightful analysis of cholesterol levels among individuals based on age distribution.

1. Cholesterol Levels Distribution:

- The top chart presents a histogram of cholesterol levels divided into bins. The **largest group** of individuals falls within the cholesterol range of **200-220**, **followed** by those in the **220-240** range. The histogram shows how cholesterol levels vary, *with a few individuals exhibiting extremely high cholesterol*.

2. Age vs. Average Cholesterol (Heatmap):

- This heatmap shows the correlation between age groups and their average cholesterol levels. The darker shades of orange indicate **higher cholesterol averages**, particularly around ages **65 and 70.** Younger age groups, like 30 and 35, tend to have lower average cholesterol levels.

3. Age Distribution:

- The bar chart displays the distribution of individuals across various age groups. The most common age group is between **60 and 70 years**, indicating a higher number of older individuals in the dataset. Ages 70 and above show a **gradual decrease** in population representation.

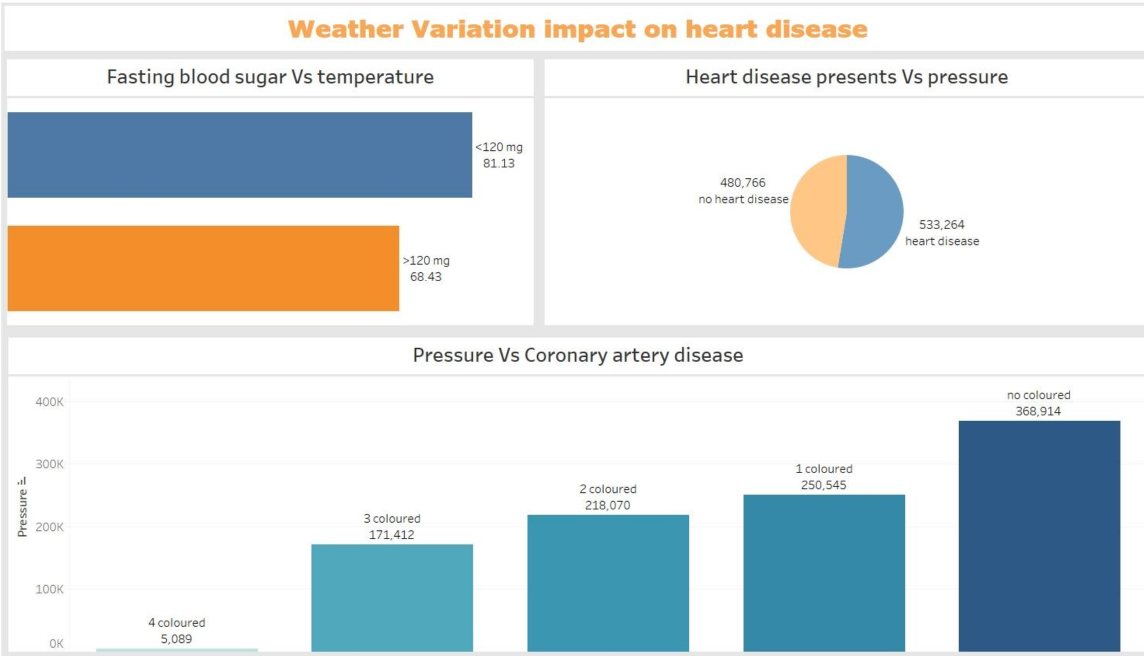
Conclusion:

- Higher cholesterol levels tend to be more common among older age groups.

- Individuals aged 65 exhibit the highest average cholesterol, suggesting a potential correlation between aging and cholesterol elevation.

- A considerable variation in cholesterol levels can be observed across different ages, emphasizing the importance of monitoring cholesterol as individuals age.

This dashboard serves as a comprehensive overview of how cholesterol levels relate to age, which is crucial for identifying at-risk age groups for cardiovascular conditions.



This dashboard provides a detailed overview of how weather variations, specifically pressure and temperature, impact heart disease and coronary artery disease.

1. Fasting Blood Sugar vs Temperature:

- This bar chart compares individuals with fasting blood sugar levels above and below 120 mg in relation to **weather temperature** variations. It shows that individuals with fasting blood sugar levels **under 120 mg** tend to be in areas with **slightly lower** temperatures compared to those with fasting blood sugar levels above 120 mg.

2. Heart Disease Presents vs Pressure:

- The pie chart shows the proportion of individuals diagnosed with heart disease compared to those without, highlighting that 53.8% (533,264) of individuals present with heart disease, while 480,766 individuals do not have heart disease.

3. Pressure vs coronary artery disease:

- The bar chart visualizes coronary artery disease cases based on pressure levels. It categorizes individuals into four pressure ranges, identified by the level of severity. The chart highlights that individuals with no pressure coloring (368,914 cases) are the most frequent, while those with 4 colored pressure bands (5,089 cases) are the least frequent, indicating a direct correlation between pressure and coronary artery disease.

- Weather, particularly pressure variations, shows a potential correlation with both heart disease and coronary artery disease.

- Individuals with higher fasting blood sugar levels tend to experience different weather conditions, possibly affecting heart health.

- A significant number of individuals with coronary artery disease are affected by higher pressure levels, suggesting the importance of monitoring both pressure and coronary artery disease symptoms together.

This dashboard effectively illustrates the relationship between weather variations, heart disease, and coronary artery disease, providing key insights for healthcare professionals and researchers.

Notes

Coronary Artery Disease Groups:

1. **No colored vessels**: This might indicate no significant blockage or a healthy state with no visible plaque in the coronary arteries.
2. **1 colored vessel**: One artery shows signs of plaque buildup or narrowing.
3. **2 colored vessels**: Two arteries are affected by plaque or narrowing.
4. **3 colored vessels**: Three arteries show signs of plaque or narrowing.
5. **4 colored vessels**: All four major coronary arteries are affected by plaque or narrowing , indicating severe CAD.

Conclusion:

- Link between Air Pollution and Heart Diseases:

Through our research and analysis, it is evident that prolonged exposure to urban air pollution poses a significant threat to cardiovascular health. The data clearly show an increase in heart attacks, strokes, and other cardiovascular conditions in areas with poor air quality.

- At-Risk Populations:

The elderly, children, and individuals with pre-existing heart conditions are disproportionately affected by poor air quality. In highly triggered cities, the frequency of heart diseases is considerably higher, making it clear that public health measures must prioritize these at-risk groups.

- Cost of Inaction:

The cost of ignoring urban air pollution is not only measured in terms of health but also economically. Treating pollution-related cardiovascular diseases places a heavy burden on healthcare systems globally. By improving air quality, we could prevent these diseases, reduce healthcare costs, and vastly improve the quality of life for millions of urban residents.

- Global and Local Impacts:

While air pollution is a global crisis, its impact is felt deeply in local contexts. Our analysis highlights the specific communities most affected, drawing a connection between global trends and the very real experiences of individuals living in polluted urban environments.

- Call to Action:

The time to act is now. By supporting green urban planning, stricter emission regulations, and clean technologies, we can mitigate the harmful effects of air pollution. Every step towards reducing pollution is a step toward saving lives and improving the future of our cities.

Further Work and Insights:

- Monitoring and Technological Innovation:

As we look ahead, the future lies in enhanced air quality monitoring. Integrating real-time sensors into city infrastructures would allow us to detect pollution spikes and respond quickly, potentially protecting thousands of people each day from dangerous levels of air pollution.

- Assessing Policy Impact:

More work is needed to measure the effectiveness of existing and new policies aimed at reducing pollution levels. This could include studies evaluating the impact of low-emission zones, electric public transportation systems, or stricter regulations on industrial emissions.

- Collaboration Across Disciplines:

The complexity of urban air quality demands collaboration across fields. Environmental scientists, data analysts, urban planners, and public health experts must continue to work together to develop comprehensive strategies that address quality from multiple angles.

- Public Health Campaigns:

Future efforts should focus on designing public health campaigns to raise awareness of the dangers of air pollution and offer practical advice for reducing personal exposure. These campaigns could empower individuals, especially those in high-traffic areas, to take control of their awareness of health by minimizing interactions during these conditions.

- Predictive Analytics for Air Quality Management:

Predictive analytics holds huge potential in urban air quality management. By analyzing weather patterns, traffic data, and industrial activity, we could develop models that forecast pollution levels, allowing cities to take preventive measures before conditions worsen.

- Global Cooperation and Data Sharing:

International cooperation in data sharing and policy implementation is essential to address this issue at a global scale. Cities can learn from one another's successes and challenges, creating more effective pollution control strategies.