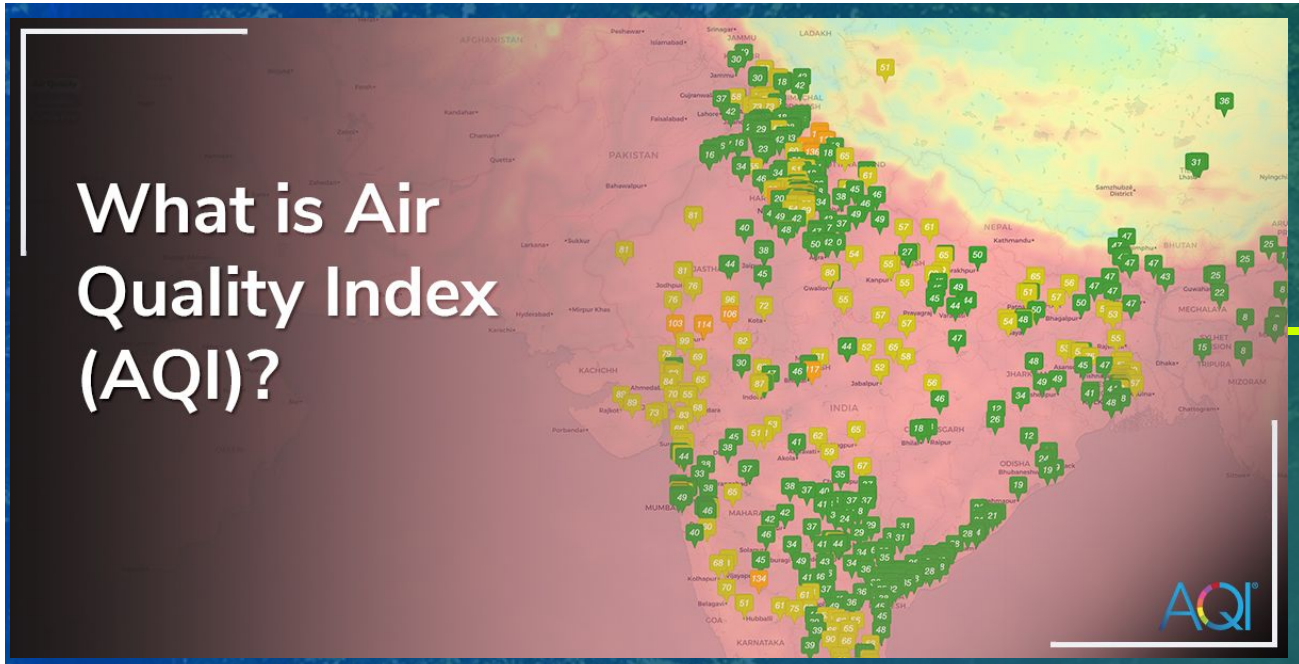
HYDERABAD AIR QUALITY INDEX

(How healthy is the air we breathe?)

What is Air Quality Index (AQI)?



An air quality index is an indicator developed by government agencies to communicate to the public how polluted the air currently is or how polluted it is forecast to become.

How do we know if Air Quality is poor?

AQI is an overall scheme that transforms individual air pollutant (e.g. SO_2, CO, PM_{10}) levels into a single number, which is a simple and lucid description of air quality for the citizens.

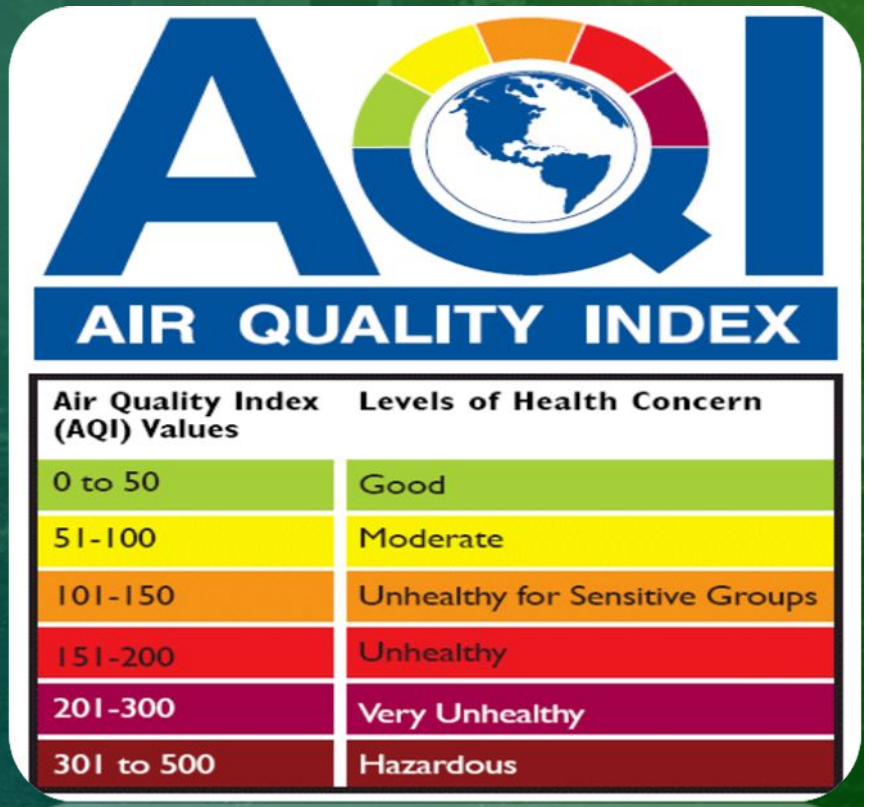
AQI relates to health impacts and citizens can avoid the unnecessary exposure to air pollutants;

AQI indicates compliance with National Air Quality Standards;

AQI prompts local authorities to take quick actions to improve air quality;

AQI guides policy makers to take broad decisions; and

AQI encourages citizens to participate in air quality management.



Are we affected by poor AQ?

1.The very young are at risk

-- Lungs are not fully developed

-- Faster breathing rate: more air volume/body weight

2.The very old are at risk

-- Undiagnosed lung or heart diseases

-- Pollution can exacerbate these conditions

3.Persons with chronic illnesses:

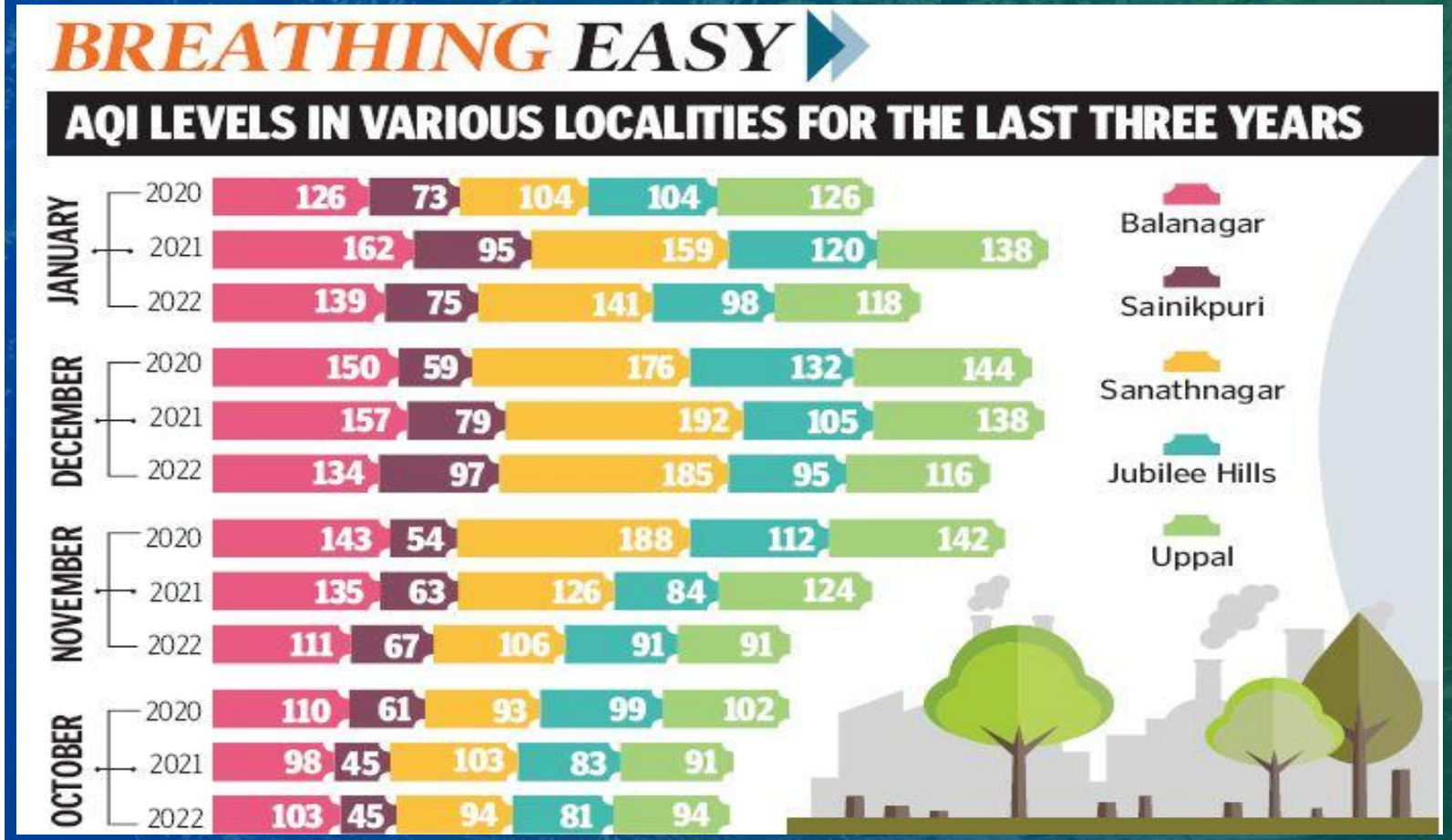
-- Respiratory circulatory, or cardiac diseases.

-- Even healthy persons can be affected when they exercise outdoors, or if the concentration of pollutants is very high

Pollutants Considered for AQI and Air Quality Standards

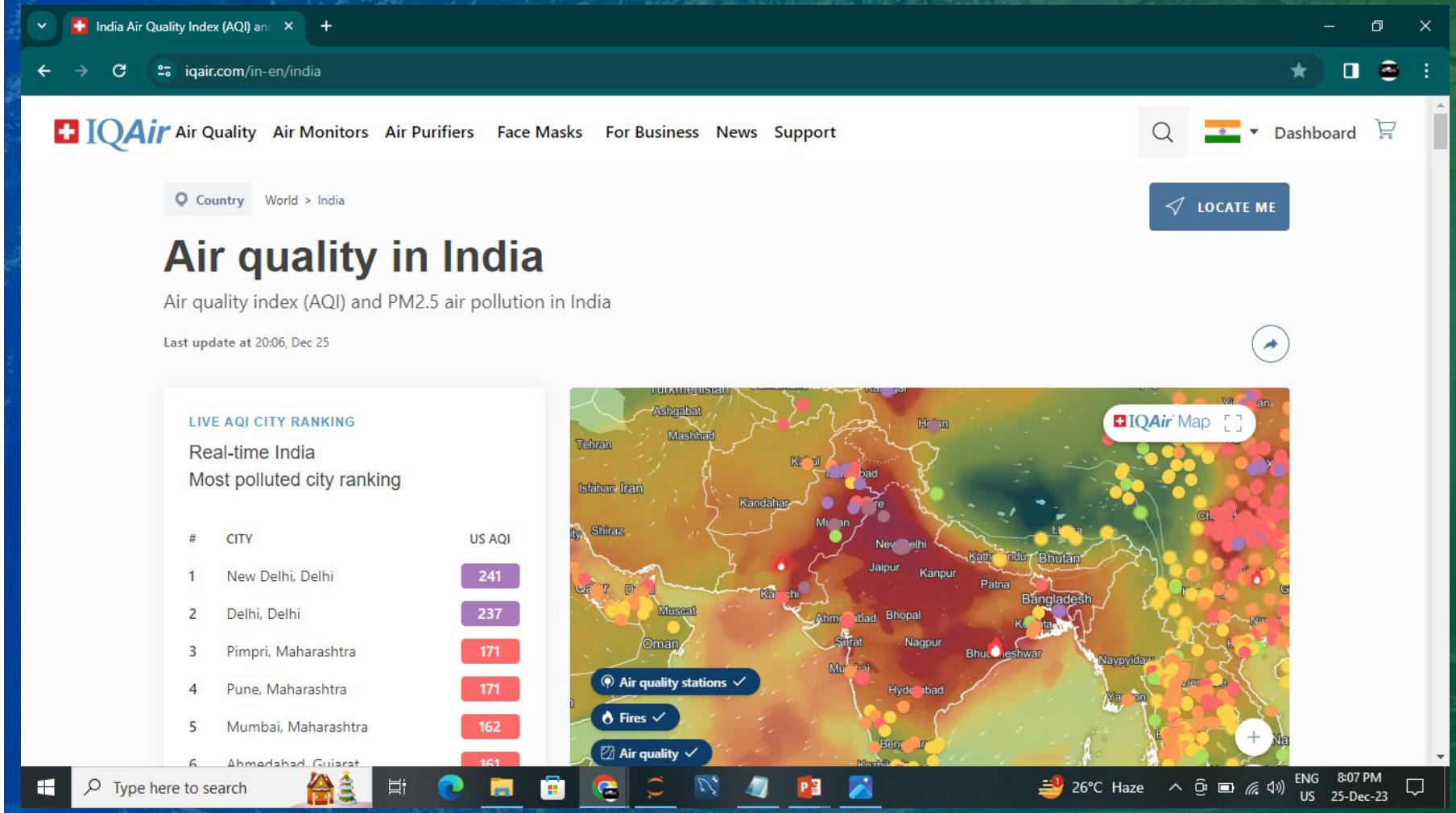
Pollutant	SO_2	NO_2	$PM_{2.5}$	PM_{10}	O_3		CO (mg/m^3)		Pb	NH_3
Averaging time (hr)	24	24	24	24	1	8	1	8	24	24
Indian Standard ($\mu g/m^3$)	80	80	60	100	180	100	4	2	1	400

Air quality in Hyderabad



https://telanganatoday.com/air-quality-improves-in-hyderabad-2

Web-based AQI dissemination



https://www.iqair.com/in-en/india

NowCast (Air quality index)

The PM **NowCast** is a weighted average of hourly air monitoring data used by the United States Environmental Protection Agency (USEPA) for real-time reporting of the Air Quality Index (AQI) for PM (PM_{10} or $PM_{2.5}$).

Reference: US

Reference: India

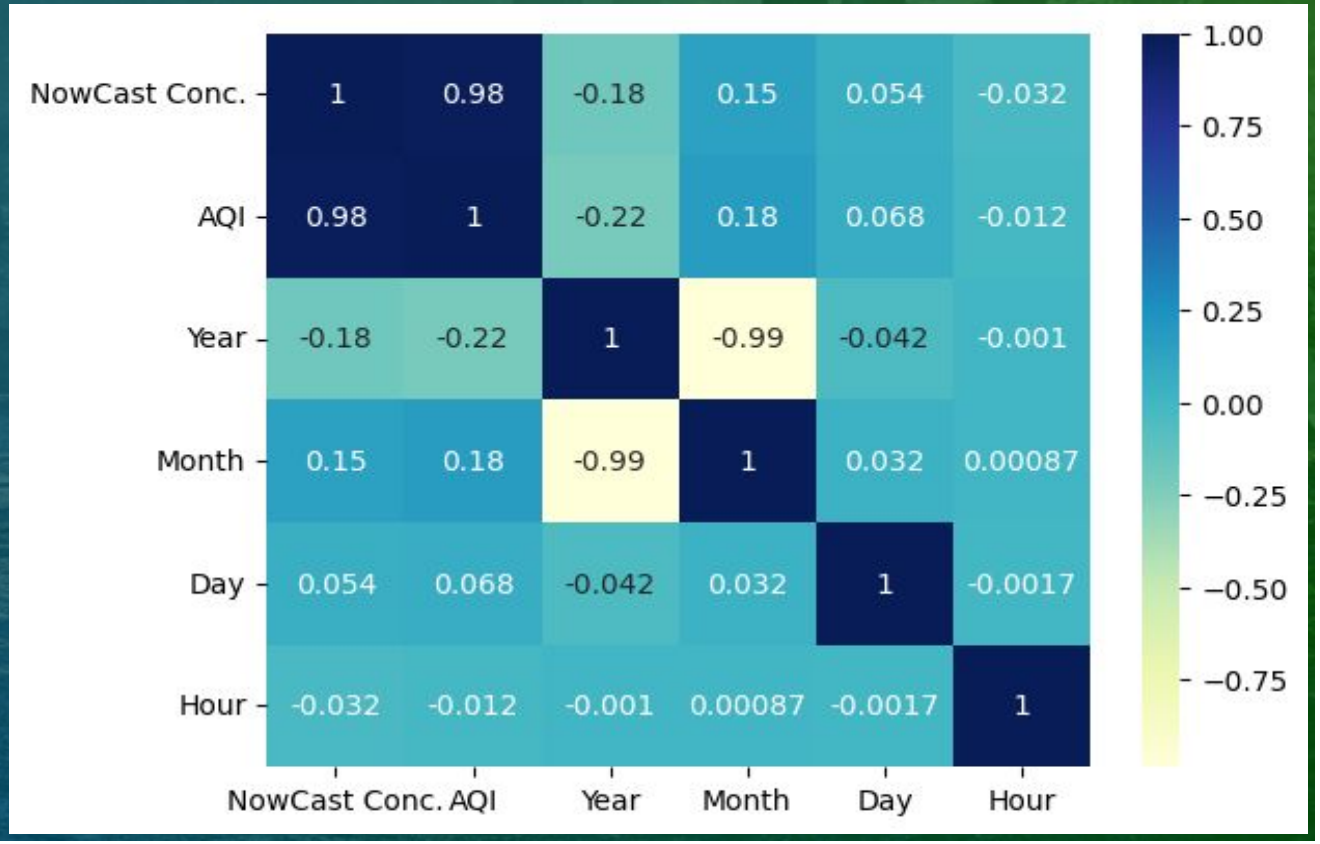
https://cpcb.nic.in/

Reference: Telangana

https://tspcb.cgg.gov.in/default.aspx

Correlation Matrix

The correlation matrix reveals strong positive correlation (0.98) between "NowCast Conc." and "AQI," indicating a significant relationship. Conversely, "Year" and "Month" exhibit negative correlations (-0.18 and -0.99, respectively), suggesting potential inverse associations. "Day" and "Hour" show weak correlations with other variables in the dataset.



Data Cleaning and Outlier Detection-NowCast Conc

Objective: Ensure data quality and handle outliers in NowCast Conc.

1.Quantile Calculation:

1. Q1 = 25th percentile
2. Q3 = 75th percentile
3. IQR (Interquartile Range) = Q3 - Q1

2.Outlier Thresholds:

1. Lower Bound = Q1 - 1.5 * IQR
2. Upper Bound = Q3 + 1.5 * IQR

3.Outlier Detection:

1. Identify outliers using threshold conditions

4.Outlier Handling:

1. Impute outliers in "NowCast Conc." with the mean

Result: Improved data quality for robust analysis.

Data Cleaning and Outlier Detection - AQI

Objective: Ensure data quality and handle outliers in "AQI"

1. **Quantile Calculation:**
    1. Q1 = 25th percentile
    2. Q3 = 75th percentile
    3. IQR (Interquartile Range) = Q3 - Q1

2. **Outlier Thresholds:**
    1. Lower Bound = Q1 - 1.5 * IQR
    2. Upper Bound = Q3 + 1.5 * IQR

3. **Outlier Detection:**
    1. Identify outliers using threshold conditions

4. **Outlier Handling:**
    1. Impute outliers in "AQI" with the mean

Result: Enhanced data quality for robust AQI analysis.

Boxplot Visualization - NowCast Conc.

Objective: Visualize the distribution of "NowCast Conc." to identify potential outliers

Boxplot Overview:

- Utilizing seaborn library for data visualization
- Focus on "NowCast Conc." variable

Boxplot Interpretation:

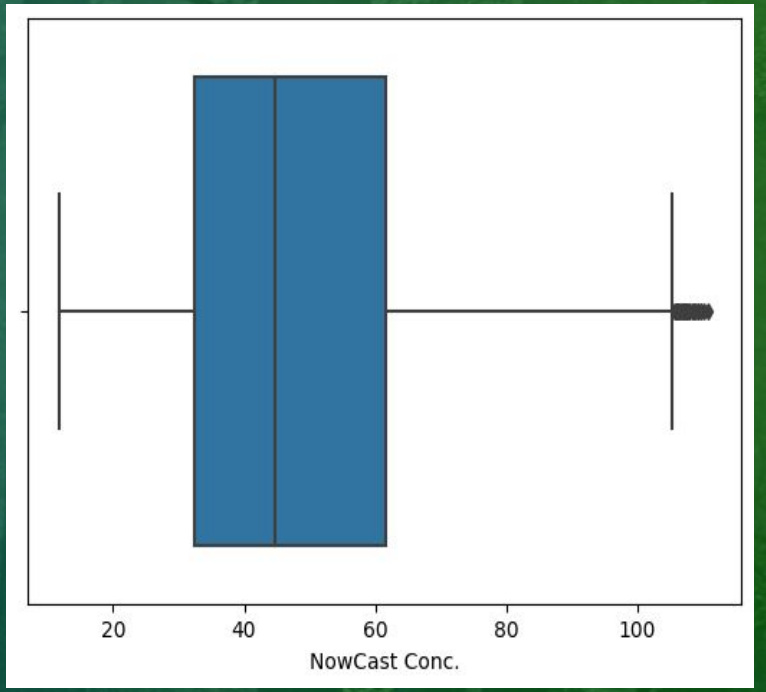
- Central box represents interquartile range (IQR)
- Whiskers show data distribution beyond the upper and lower quartiles
- Outliers may be identified outside the whiskers

Insights:

- Identify potential outliers in "NowCast Conc.“

Next Steps:

- Evaluate outlier impact on analysis
- Consider appropriate handling strategies



Result: Enhanced understanding of "NowCast Conc." distribution for data-driven decisions.

Scatterplot - NowCast Conc. vs. AQI

Objective: Explore the relationship between "NowCast Conc." and "AQI" using a scatterplot

Scatterplot Overview:

Utilizing seaborn library for data visualization
Axes represent "NowCast Conc." on the x-axis and "AQI" on the y-axis

Interpretation:

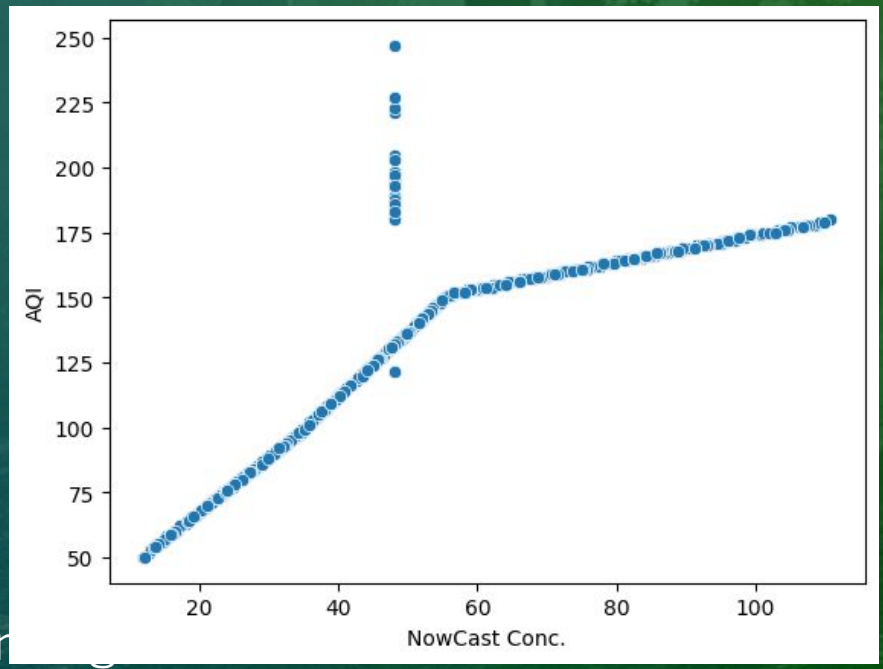
Each point represents a data pair (NowCast Conc., AQI)
Visualize the pattern and distribution of data points

Insights:

Identify any observable trends or correlations
Assess the strength and direction of the relationship

Next Steps:

Use insights to inform further analysis or modeling
Consider additional variables for a comprehensive understan



Result: Enhanced understanding of the relationship between "NowCast Conc." and "AQI" for informed decision-making.

Pairplot - Exploring Relationships Across Variables

Objective: Utilize seaborn pairplot to visualize relationships among variables in the dataset

Pairplot Overview:

Utilizing seaborn library for comprehensive data exploration
Grid of scatterplots for each pair of variables in "a“

Interpretation:

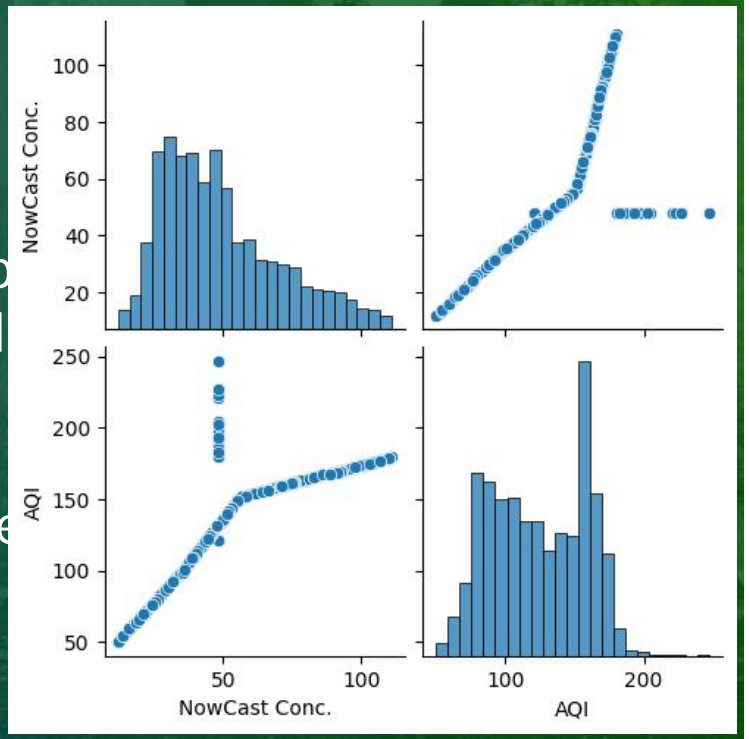
Diagonal: Histograms representing the distribution of each variab
Off-diagonal: Scatterplots showing relationships between variabl

Insights:

Identify patterns, trends, and correlations across multiple variable

Next Steps:

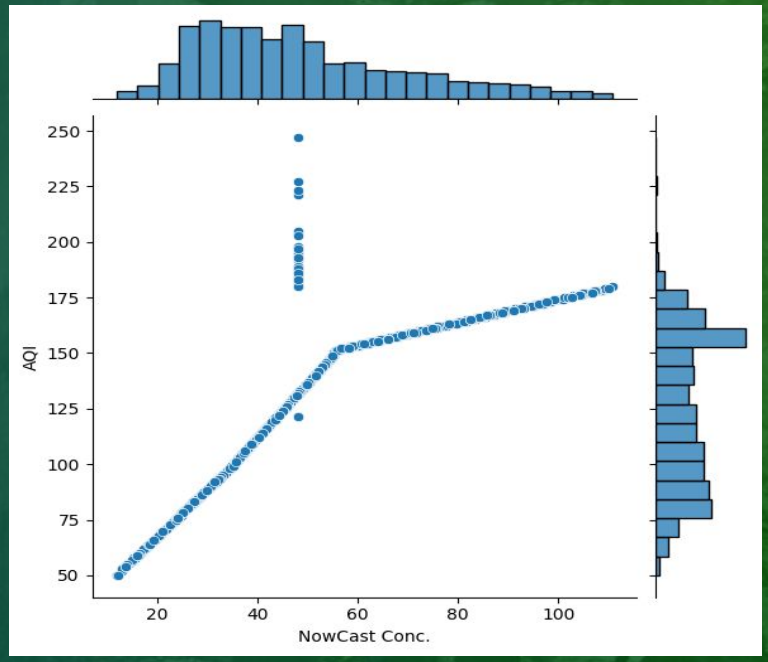
Use insights to guide further analysis and modeling
Assess variable interactions for a holistic understanding



Result: Comprehensive visualization aiding in the exploration of relationships among variables.

Jointplot - NowCast Conc. vs. AQI

Objective: Visualize the bivariate relationship between "NowCast Conc." and "AQI" using a seaborn jointplot



Jointplot Overview:

- Utilizing seaborn library for bivariate data exploration
- Axes represent "NowCast Conc." on the x-axis and "AQI" on the y-axis
- Central scatterplot highlights the joint distribution of the two variables

Interpretation:

- Assess the concentration of data points in the joint space
- Identify any patterns or trends in the relationship

Insights:

- Explore the strength and direction of the correlation between "NowCast Conc." and "AQI“

Next Steps:

- Use insights to inform further analysis or modeling
- Consider the visual representation for communication and reporting

Result: Enhanced understanding of the joint distribution of "NowCast Conc." and "AQI."

Linear Regression Modeling - NowCast Conc. vs. AQI

Objective: Utilize linear regression to model the relationship between "NowCast Conc." and "AQI“

Data Preparation:

Features (X): "NowCast Conc."
Target (Y): "AQI“

Data Splitting:

70% training set, 30% testing set
Random state set to ensure reproducibility (random_state=2)

Linear Regression Model:

Utilizing scikit-learn's LinearRegression module
Training the model on the training set

Prediction:

Making predictions on both the training and testing sets
Assessing model performance using R-squared (R^2) metric

Result: A trained linear regression model predicting "AQI" based on "NowCast Conc." with insights into model performance.

Model Evaluation - R-squared Scores

Objective: Assess the performance of the Linear Regression model on training and testing sets

Evaluation Metrics:

R-squared (R^2) Score

A measure of the proportion of variance explained by the model

Training Set Performance:

R-squared on the training set: 0.832

Testing Set Performance:

R-squared on the testing set: 0.830

Interpretation:

R^2 values close to 1 indicate a good fit

High consistency between predicted and actual values

Result: Strong model performance in explaining variability in "AQI" based on "NowCast Conc."

Model Evaluation - Regression Metrics

Objective: Calculate and assess additional evaluation metrics for the Linear Regression model

Evaluation Metrics:

Mean Absolute Error (MAE)
Mean Squared Error (MSE)
Root Mean Squared Error (RMSE)

Results:

Mean Absolute Error: 10.03
Mean Squared Error: 10.03
Root Mean Squared Error: 10.03

Interpretation:

Metrics quantify the average prediction errors
Useful for understanding model accuracy and performance
Low error metrics demonstrate model accuracy

Result: Insightful metrics providing a comprehensive view of the model's predictive accuracy.

Executive Summary

"Improve public health, ensure environmental compliance, and enhance community well-being with strategic air quality insights."

1. **Public Health Improvement**: Timely air quality insights lead to proactive measures, reducing health risks for the public.

2. **Environmental Compliance**: Monitoring ensures adherence to environmental regulations, safeguarding against penalties and legal issues.

3. **Community Well-being Enhancement**: Elevates community quality of life, fostering trust, and positive relationships through a commitment to environmental responsibility.


THANK YOU!
Done by :
Lokesh
AVS : Analytix Vidya and Solutions