BrainStation®

Delhi Air Quality Prediction

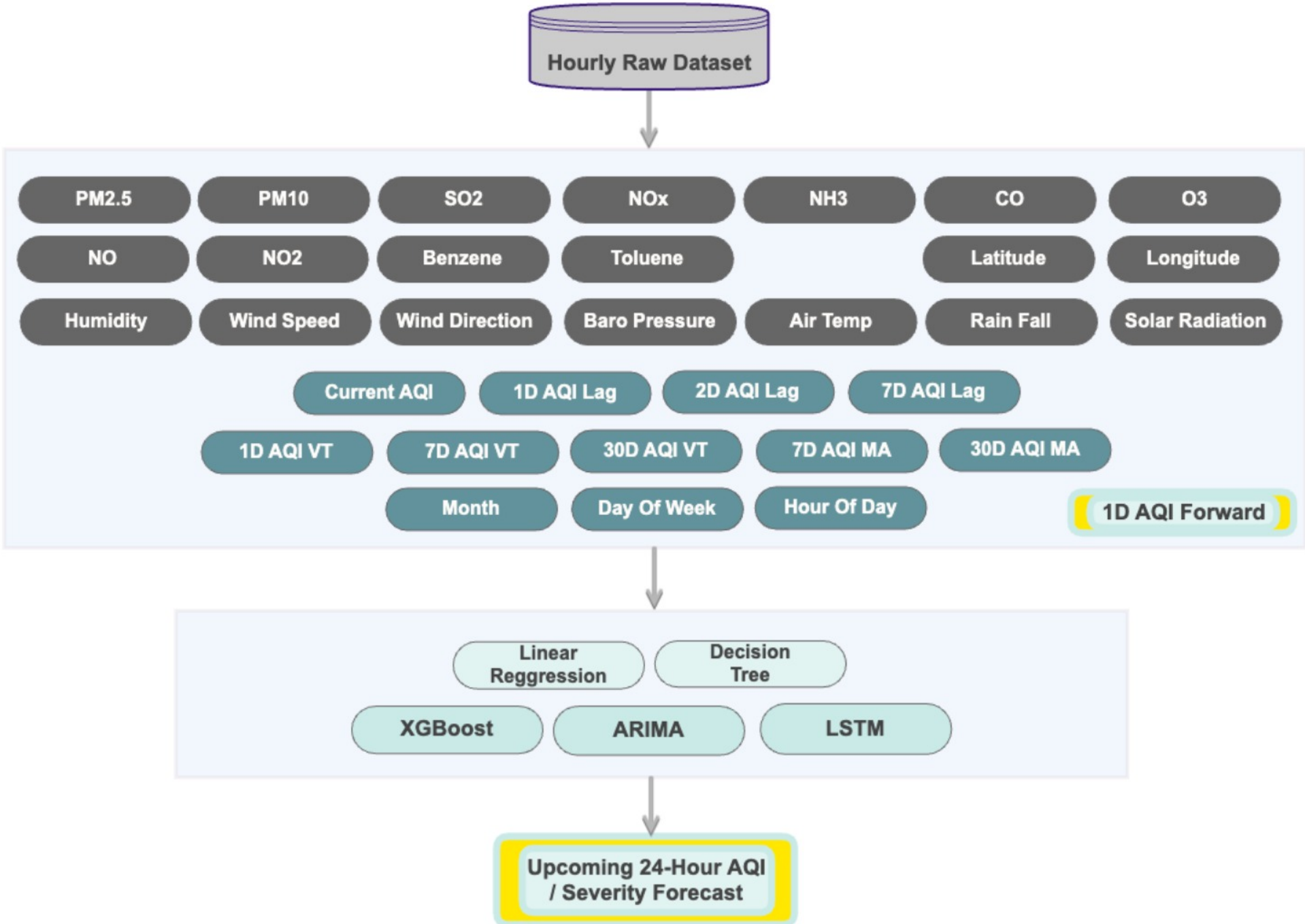
Data-Driven Insights

Prepared by Lipsita Tripathy

Agenda:

- Introduction
- Dataset Overview and Preprocessing
- Exploratory Data Analysis (EDA)
- Baseline Modeling and Evaluation
- Next Steps

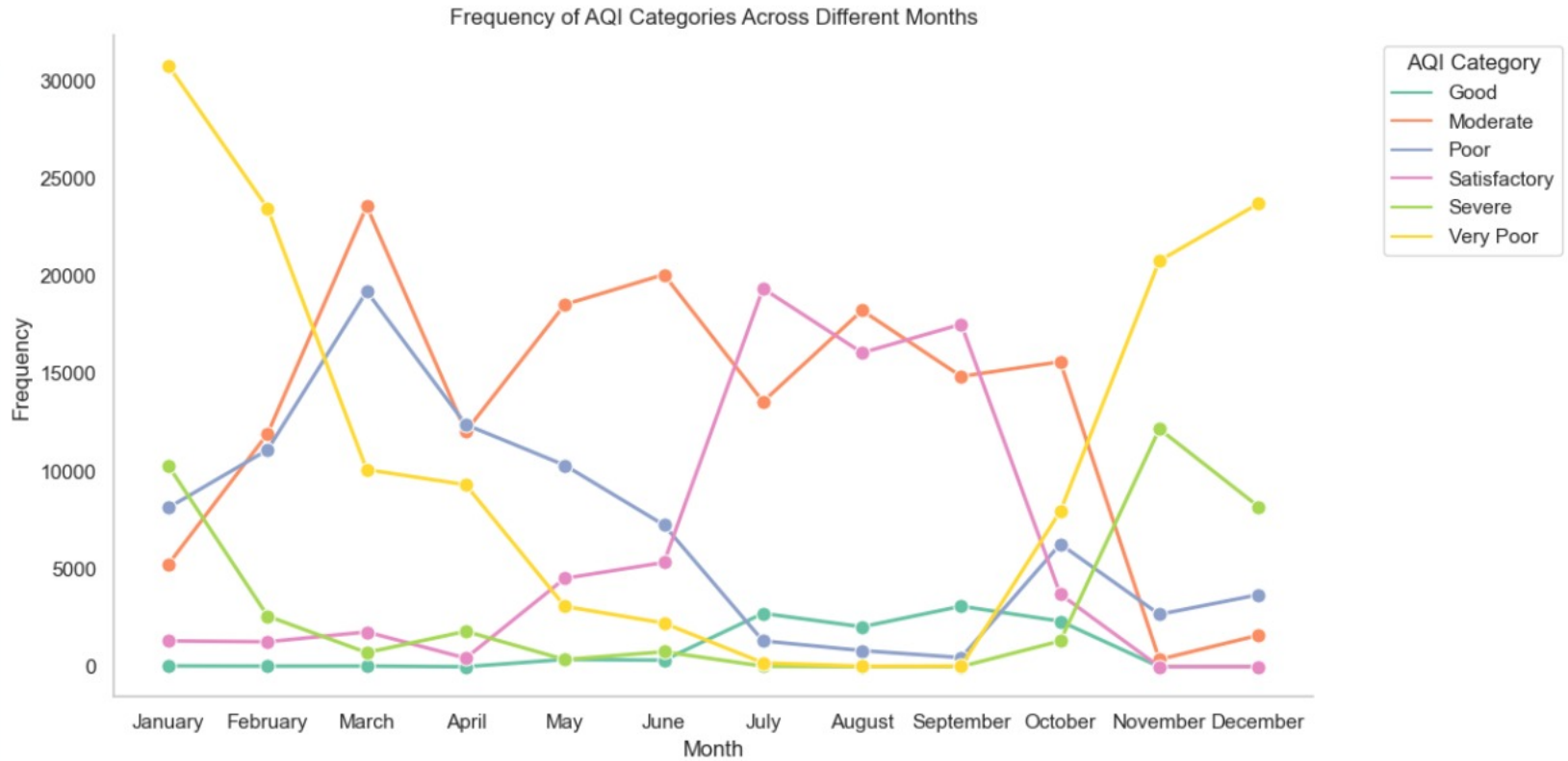
Data:

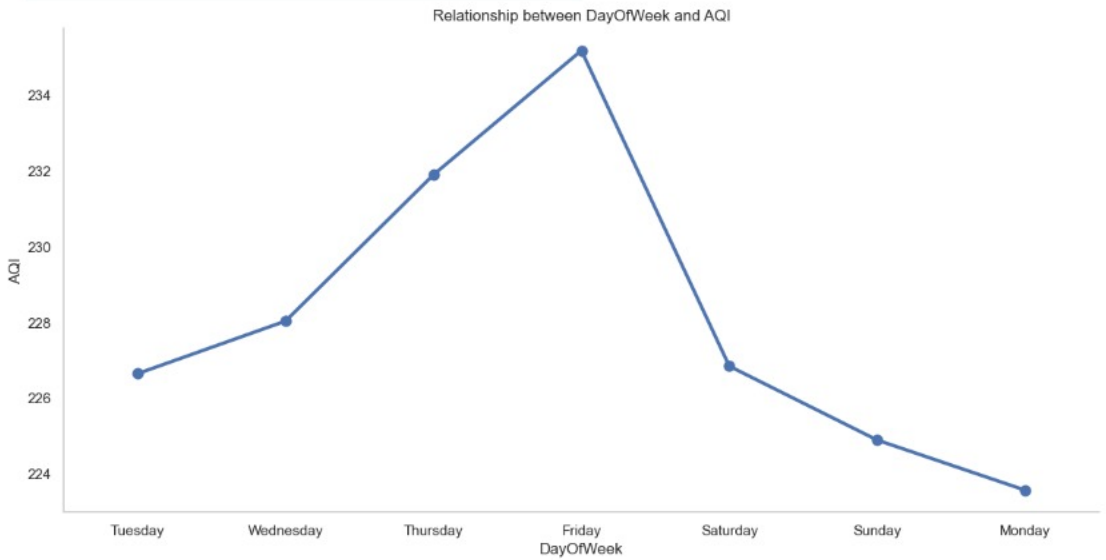

Hourly Raw Dataset
PM2.5
PM10
SO2
NOx
NH3
CO
O3
NO
NO2
Benzene
Toluene
Latitude
Longitude
Humidity
Wind Speed
Wind Direction
Baro Pressure
Air Temp
Rain Fall
Solar Radiation
Current AQI
1D AQI Lag
2D AQI Lag
7D AQI Lag
1D AQI VT
7D AQI VT
30D AQI VT
7D AQI MA
30D AQI MA
Month
Day Of Week
Hour Of Day
1D AQI Forward
Linear Reggression
Decision Tree
XGBoost
ARIMA
LSTM
Upcoming 24-Hour AQI / Severity Forecast

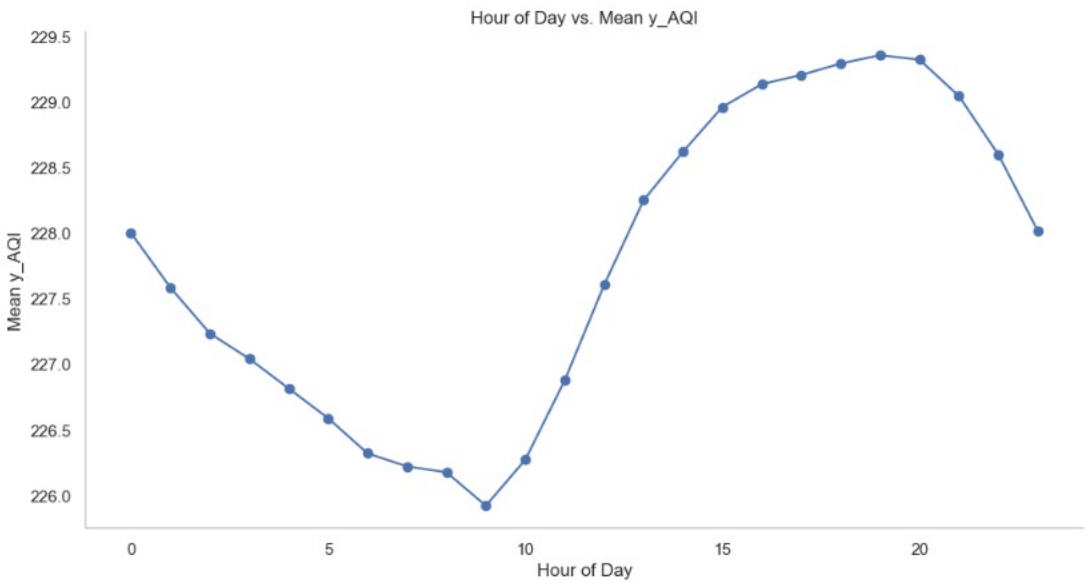
Trends:

AQI over

- Months
- Days of the week
- Hour of the day







Baseline Modeling & Evaluation:

Linear Regression:

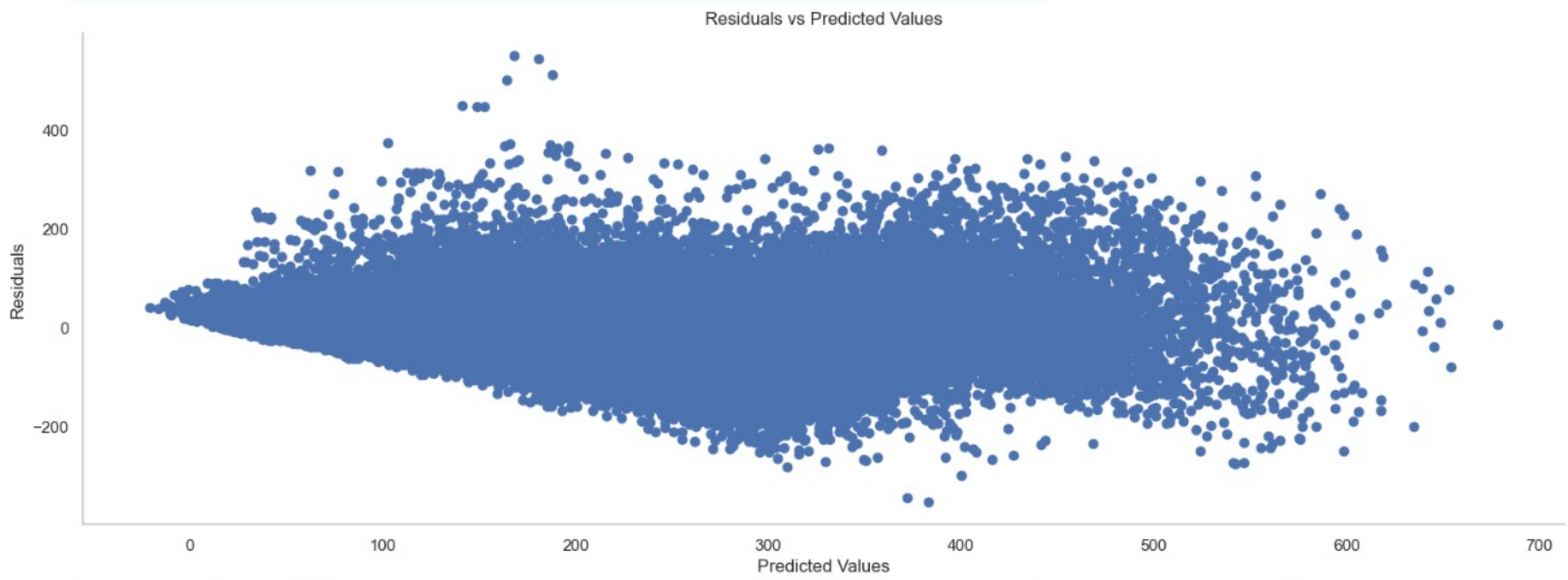
- Achieved R-squared: 0.79
- MSE: 844.70

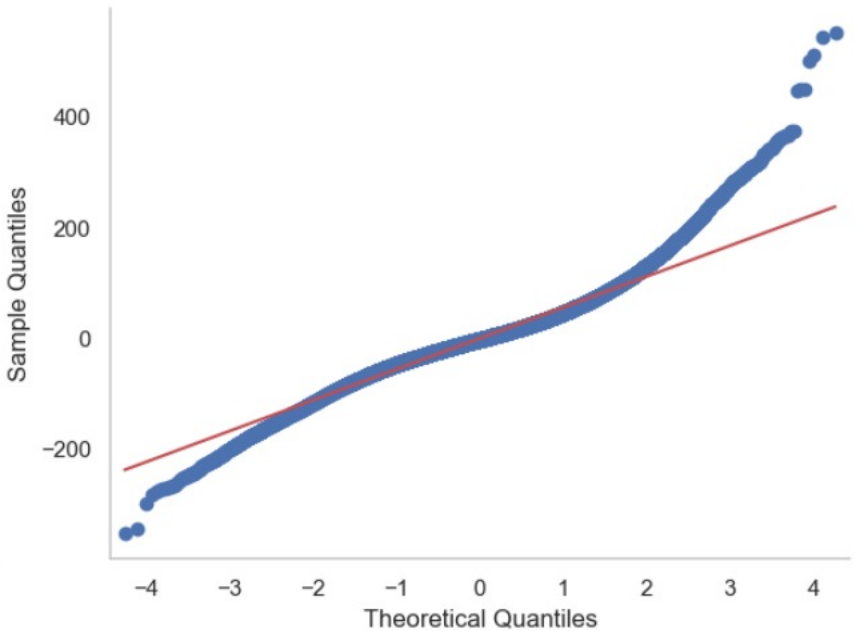
Decision Tree:

- Baseline model, max depth 3.
- R-squared: 0.71
- MSE: 4408.77

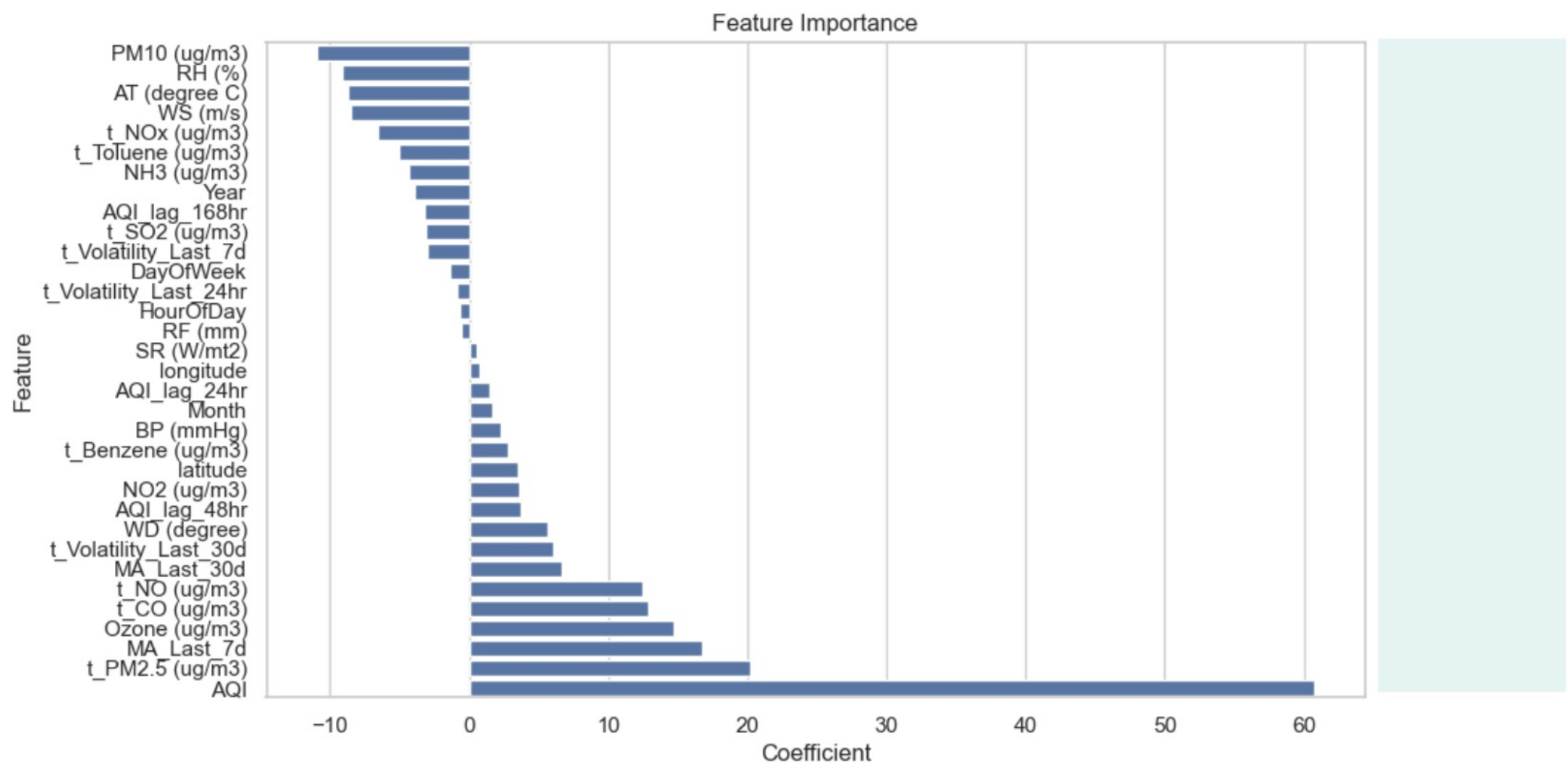
Evaluation Plots:

- Residual Plot
- Q-Q Plot





Feature Importance:



Moving Forward:

- Time Series Modeling(ARIMA)
- Advanced Modeling(XGBoost and LSTM)
- Model Evaluation
- PCA
- In-depth Feature Importance Analysis

Thank you!
Questions?