Steps:

1. **Data Loading and Preprocessing**:
    - Loaded the dataset from an Excel file.
    - Replaced negative values (indicating missing or erroneous data) with NaNs.
    - Dropped rows with NaN values to clean the dataset.
2. **Feature Selection**:
    - Selected the following features for the regression model: `PT08.S1(CO)`, `PT08.S2(NMHC)`, `PT08.S3(NOx)`, `PT08.S4(NO2)`, `PT08.S5(O3)`, `T`, `RH`, and `AH`.
    - Set `C6H6(GT)` (Benzene concentration) as the target variable.
3. **Data Splitting**:
    - Split the dataset into training and testing sets using an 80-20 split.
4. **Model Training**:
    - Trained a linear regression model using the training set.
5. **Model Evaluation**:
    - Made predictions on the test set.
    - Evaluated the model using Mean Squared Error (MSE) and R^2 Score.
    - Printed the coefficients of the model to understand the relationship between the features and the target variable.
6. **Visualization**:
    - Created a scatter plot to visualize the actual versus predicted Benzene (C6H6) levels.

Results:

- **Mean Squared Error (MSE)**: This metric indicates the average squared difference between the observed actual outcomes and the predicted outcomes. A lower value indicates a better fit.
- **R^2 Score**: This metric represents the proportion of the variance in the dependent variable that is predictable from the independent variables. A value closer to 1 indicates a better fit.
- **Model Coefficients**: The coefficients provide insight into the impact of each feature on the target variable. Positive coefficients indicate a direct relationship, while negative coefficients indicate an inverse relationship.

Example Output:

```
plaintext
Copy code
Mean Squared Error: [MSE value]
R^2 Score: [R^2 value]

Coefficients:
        Feature  Coefficient
0  PT08.S1(CO)    [value]
1  PT08.S2(NMHC)  [value]
2  PT08.S3(NOx)   [value]
3  PT08.S4(NO2)   [value]
4  PT08.S5(O3)    [value]
5  T             [value]
6  RH            [value]
7  AH            [value]
```

Visualization:

The scatter plot shows the actual Benzene (C6H6) levels on the x-axis and the predicted levels on the y-axis. A line representing a perfect prediction (where the actual and predicted values are equal) is also plotted for reference.

By following these steps, you can gain insights into the relationship between different air quality measurements and Benzene concentration, and assess the performance of a linear regression model in predicting Benzene levels.