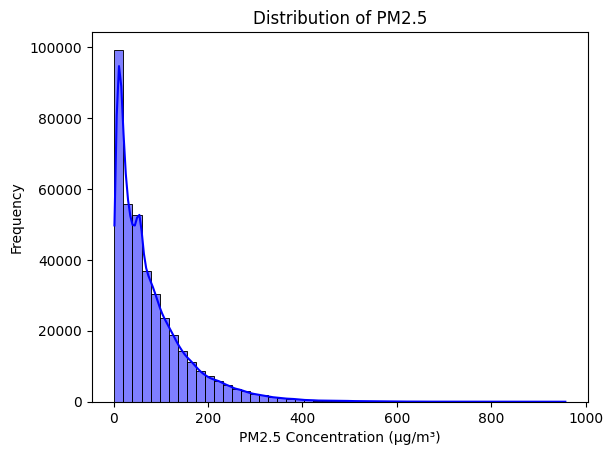
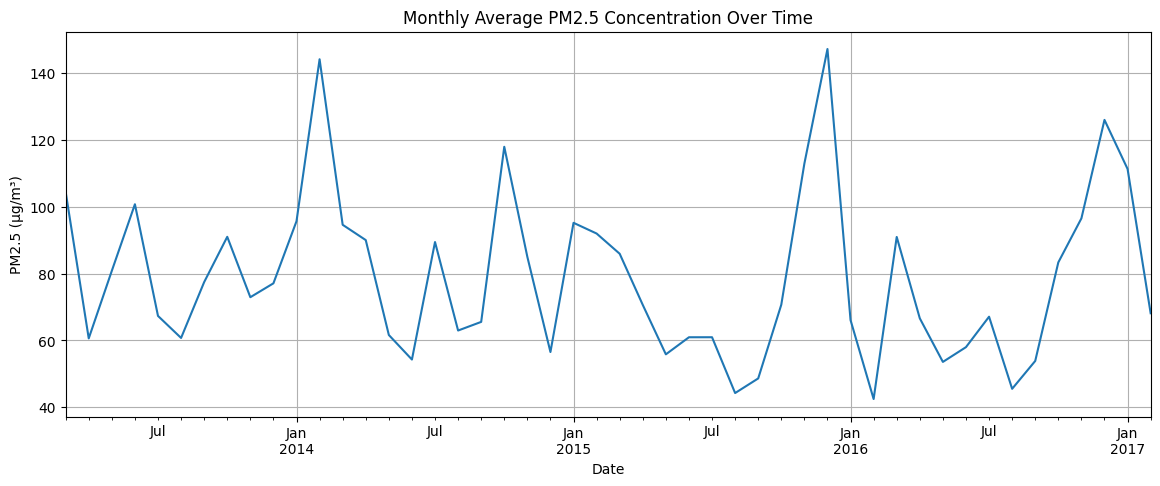
## Statistics/computation-based analysis and Visualisation



**Figure 9: Univariate Analysis: PM2.5 Distribution**

(Source: Google colab)

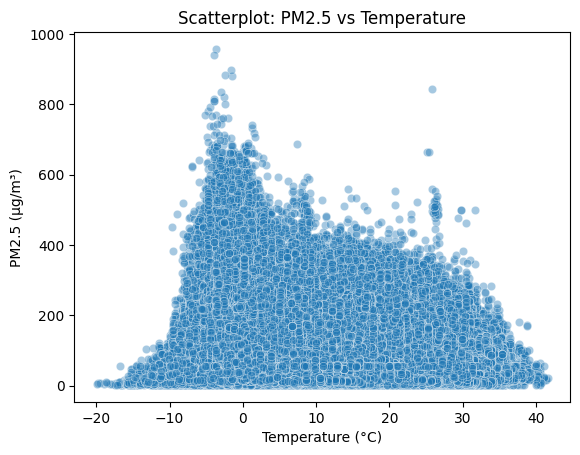
The histogram displaying the distribution of PM2.5 concentrations shows that the majority of the values fall between 0 and 200 µg/m³. This indicates relatively moderate pollution levels for most records, with fewer occurrences of extremely high PM2.5 concentrations in the dataset.



**Figure 10: Time Series Line Plot: PM2.5 Over Time**

(Source: Google colab)

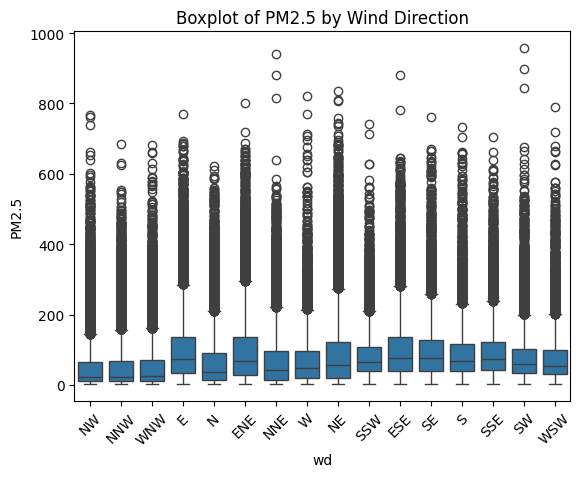
The time series line plot of monthly average PM2.5 concentrations reveals noticeable peaks during January of 2014, 2015, 2016, and 2017. These spikes suggest higher pollution levels in the winter months, likely due to increased heating activities and stagnant atmospheric conditions.



**Figure 11: Bivariate Analysis: PM2.5 vs Temperature**

(Source: Google colab)

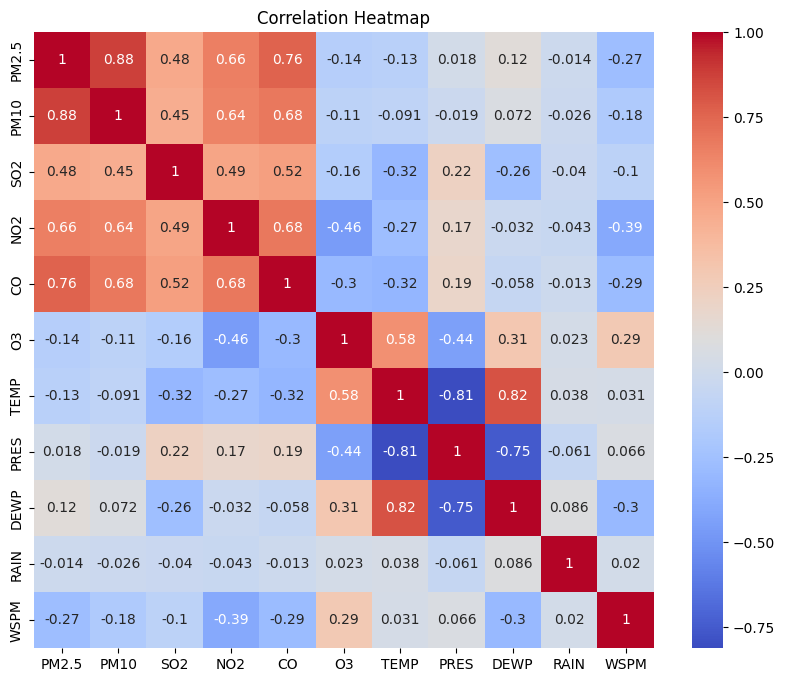
The scatterplot reveals a negative correlation between temperature and PM2.5 concentration. Higher PM2.5 levels are concentrated at lower temperatures, especially below 10°C, indicating winter-related pollution. As temperature rises above 20°C, PM2.5 levels decrease, suggesting better air quality in warmer conditions.



**Figure 12: Boxplot: PM2.5 by Wind Direction**

(Source: Google colab)

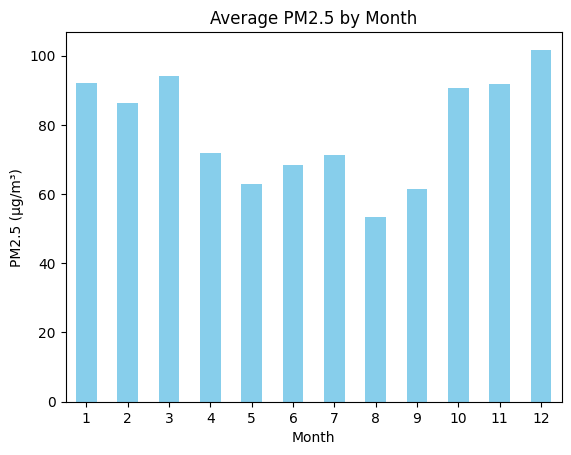
The boxplot shows that wind direction influences PM2.5 levels, with higher median concentrations observed for NE, NNE, E, and ENE. These directions also have greater variability and extreme outliers, suggesting pollution sources or dispersion patterns linked to specific wind directions.



**Figure 13: Heatmap: Correlation Matrix**

(Source: Google colab)

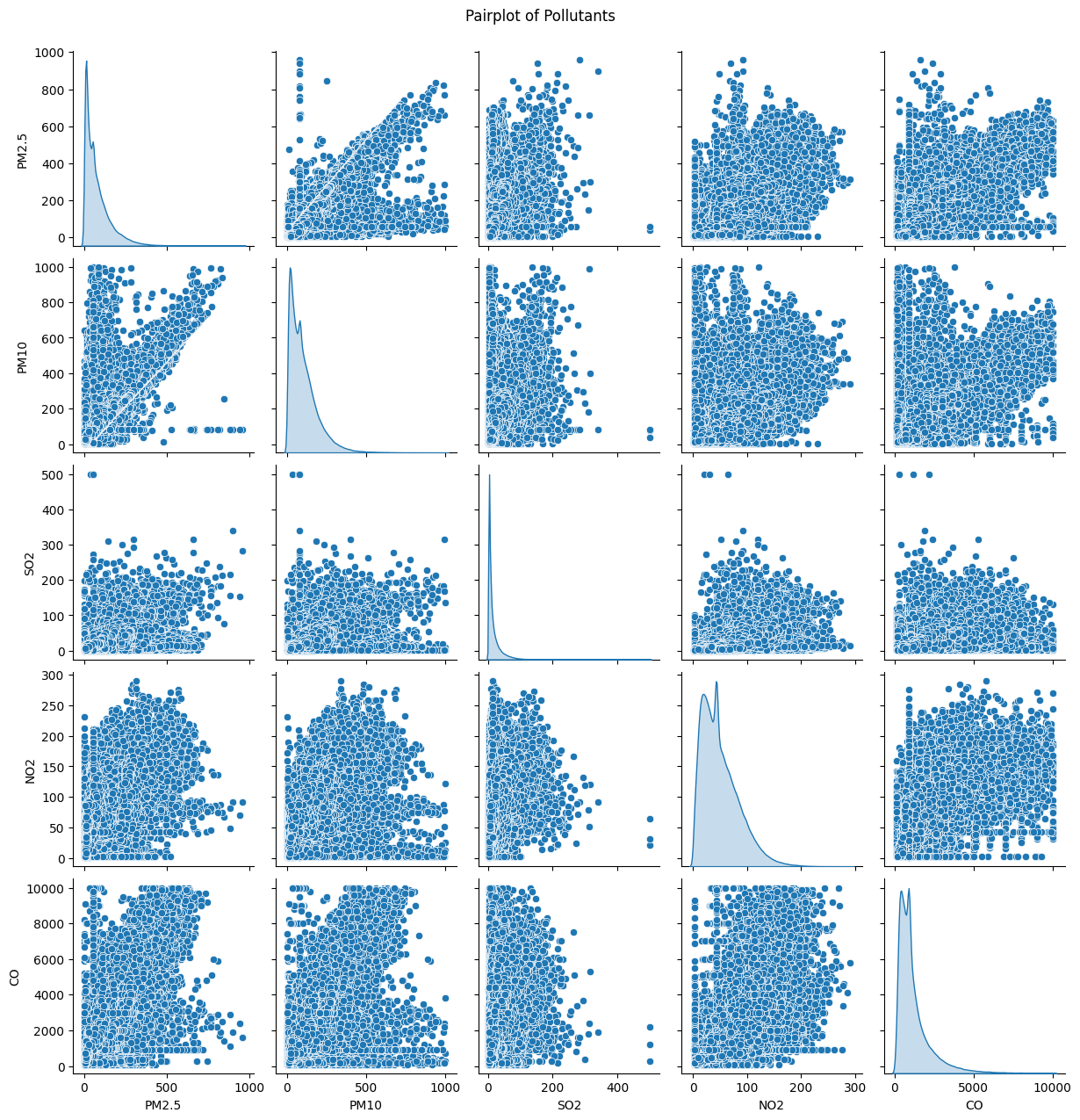
The heatmap shows strong positive correlations between pollutants like PM2.5, PM10, CO, and NO2, indicating common sources. Temperature inversely correlates with several pollutants, suggesting worse air quality in colder conditions. Wind speed mildly disperses pollutants, and ozone rises with temperature.



**Figure 14: Bar Chart: Average PM2.5 by Month**

(Source: Google colab)

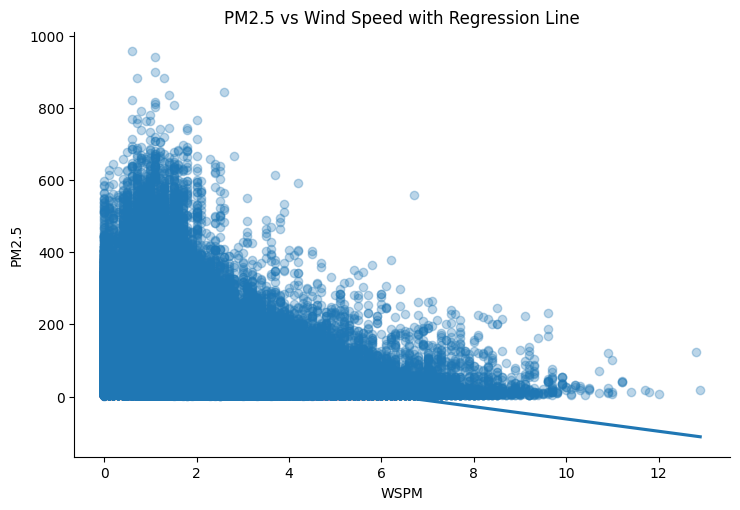
The bar chart illustrates the average PM2.5 concentration by month, with December (Month 12) showing the highest levels. This suggests that PM2.5 concentrations are typically elevated during the winter months, potentially due to seasonal pollution sources or weather patterns.



**Figure 15: Multivariate Analysis: Pairplot of Pollutants**

(Source: Google colab)

The pairplot provides insights into the relationships between PM2.5, PM10, SO2, NO2, and CO. PM2.5 and PM10 are strongly correlated, while other pollutants show more varied and less linear relationships. Right-skewed distributions and outliers suggest variability in pollution levels across observations.



**Figure 16: Wind Speed vs PM2.5: Scatter Plot with Regression Line**

(Source: Google colab)

The scatter plot with a regression line shows a clear negative correlation between wind speed and PM2.5 concentration. Higher wind speeds generally lead to lower PM2.5 levels, as wind helps disperse pollutants. However, some outliers and variability at low wind speeds suggest other influencing factors.



**Figure 17: Statistical summary**

(Source: Google colab)

The custom statistical summary provides a detailed overview of various pollutants and meteorological variables. Key highlights include high skewness and kurtosis in pollutants like PM2.5, CO, and NO2, indicating a right-skewed distribution with potential extreme values or outliers.