**NAAN MUDHALVAN-IBM DATA ANALYTICS WITH COGNOS**

**PROJECT PHASE 4: DEVELOPMENT PART 2**

**PROJECT TITLE:AIR QUALITY INDEX ANALYSIS**

**TEAM MEMBERS:**

* GokulRaj S - 2021115038
* Hariharan I S - 2021115039
* Gokul V - 2021115037
* Harshith A N - 2021115040
* Gopi R - 2021115315

**Introduction**

*In this pivotal phase of our project, we are poised to deepen our understanding of air quality by embarking on a journey of comprehensive data analysis. Our project revolves around the Air Quality Index (AQI), a critical environmental indicator that influences public health and environmental policies. We have already collected and meticulously pre-processed our dataset, setting the stage for an array of crucial tasks that include:*

**Exploratory Data Analysis (EDA)**

*Exploratory Data Analysis is the cornerstone of our analysis. It's the process of peeling back the layers of our dataset to reveal its underlying characteristics. During this phase, we aim to:*

***Recognize Data Traits****:*

*We will identify the different types of data at our disposal, ranging from pollutant levels to meteorological variables.*

***Uncover Patterns:***

*EDA allows us to uncover hidden patterns, temporal trends, and seasonal variations within our air quality data.*

***Detect Outliers:***

*Identifying and handling outliers is crucial in ensuring the reliability of our analysis.*

***Explore Relationships:***

*We'll delve into the relationships between air pollutants, weather conditions, and geographical locations.*

*EDA sets the stage for informed decision-making, helping us understand the nuances of our air quality dataset.*

***Statistical Analysis***

*Statistical Analysis is the next vital step in our journey. This phase enables us to:*

***Calculate Summary Statistics:***

*We'll compute basic statistical measures like mean, median, standard deviation, and percentiles for our numerical variables. These measures provide a sense of central tendencies and data distributions.*

***Hypothesis Testing****:*

*We may perform hypothesis tests to draw inferences about our data, such as assessing the impact of specific weather conditions on air quality.*

***Time-Series Analysis:***

*Analyzing our AQI data over time helps us identify long-term trends and seasonality.*

***Data Visualization***

*Data Visualization breathes life into our analysis. Visual representations provide a clear and insightful view of our data, and during this phase,*

***Create Visualizations:***

*We will utilize Python libraries like Matplotlib, Seaborn, and Plotly to craft informative graphs, charts, and maps.*

***Heatmaps and Correlation Plots:***

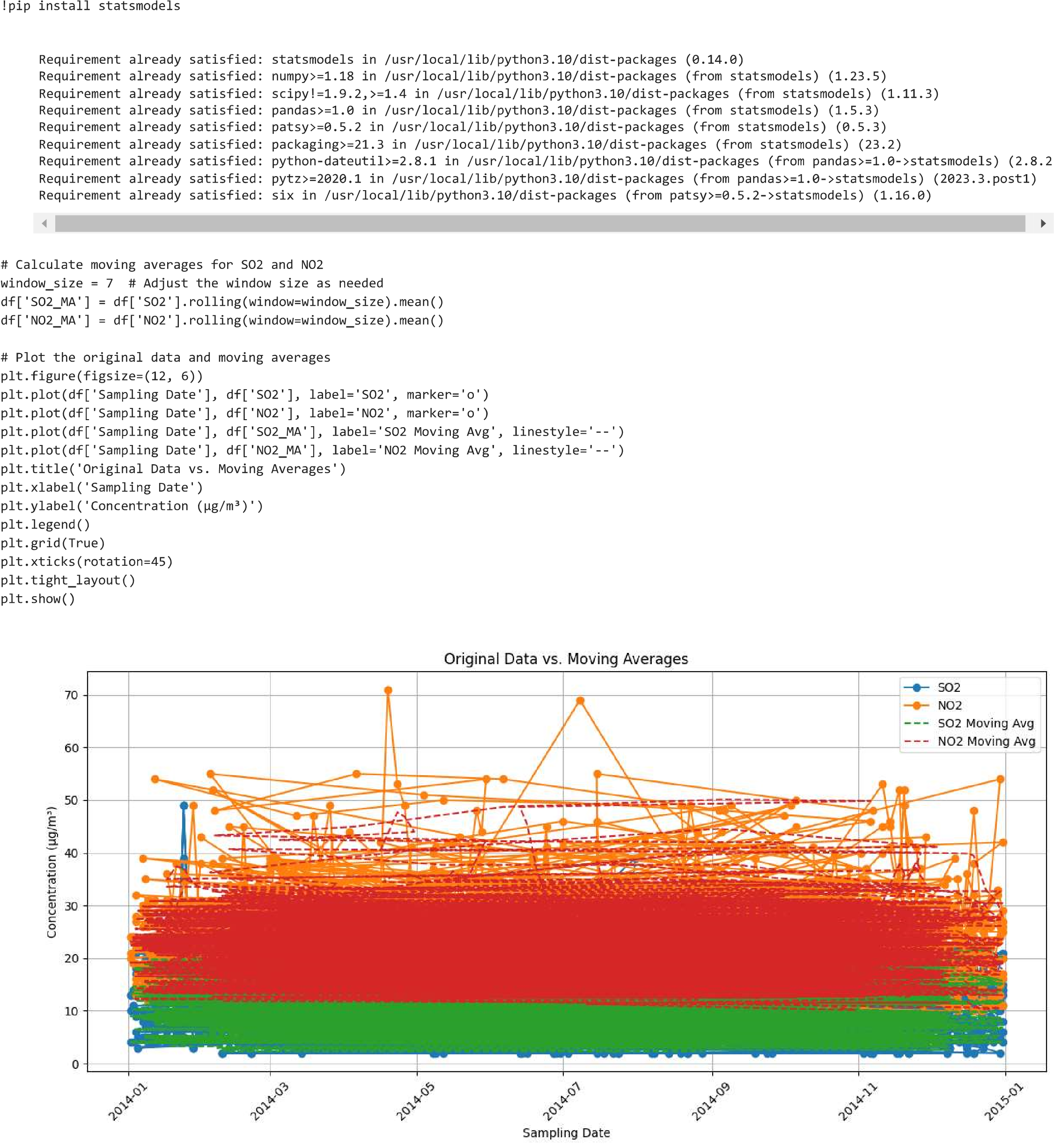
*We will employ heatmaps and correlation plots to visualize relationships between variables.*

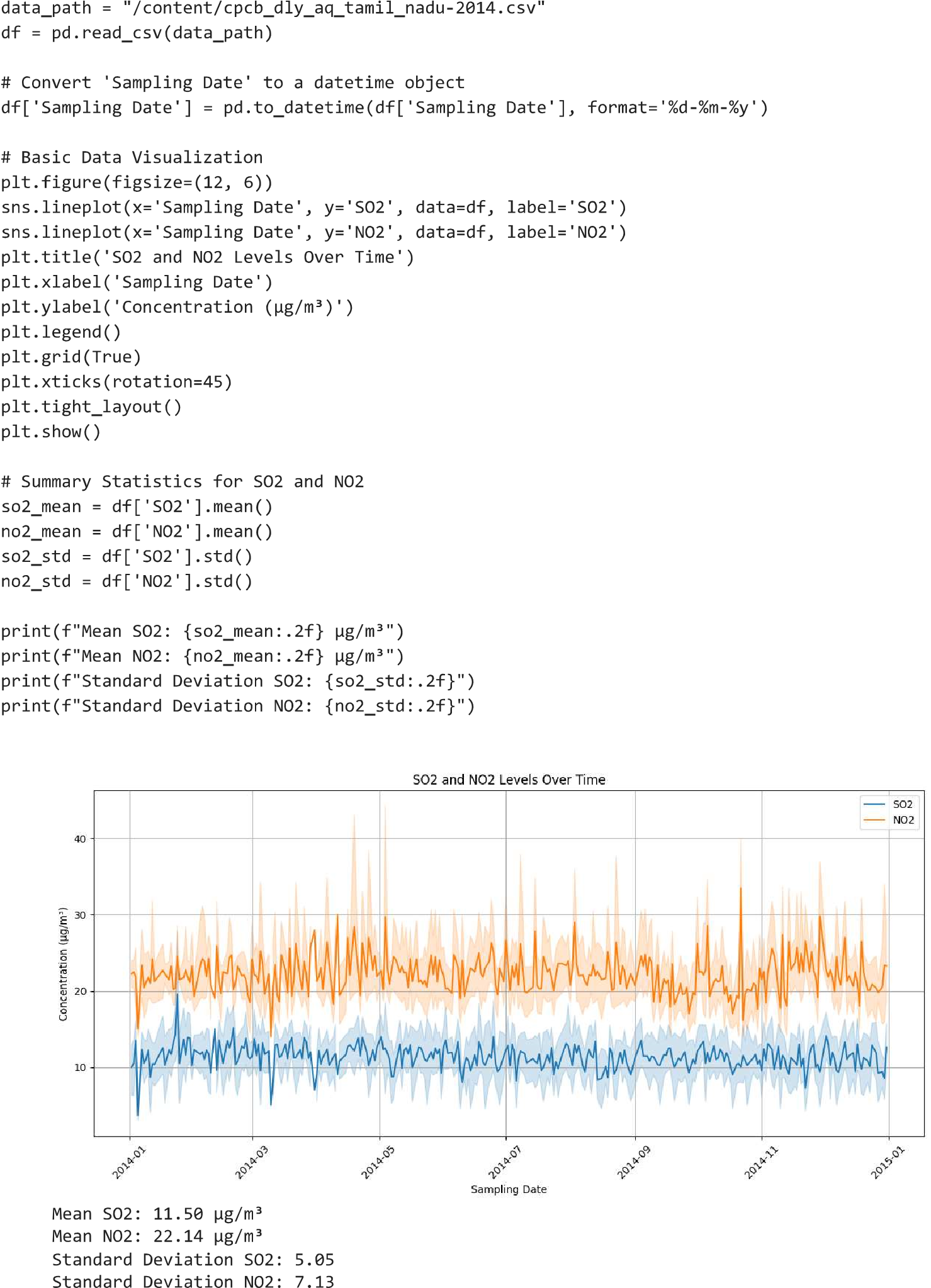
***Geospatial Mapping****:*

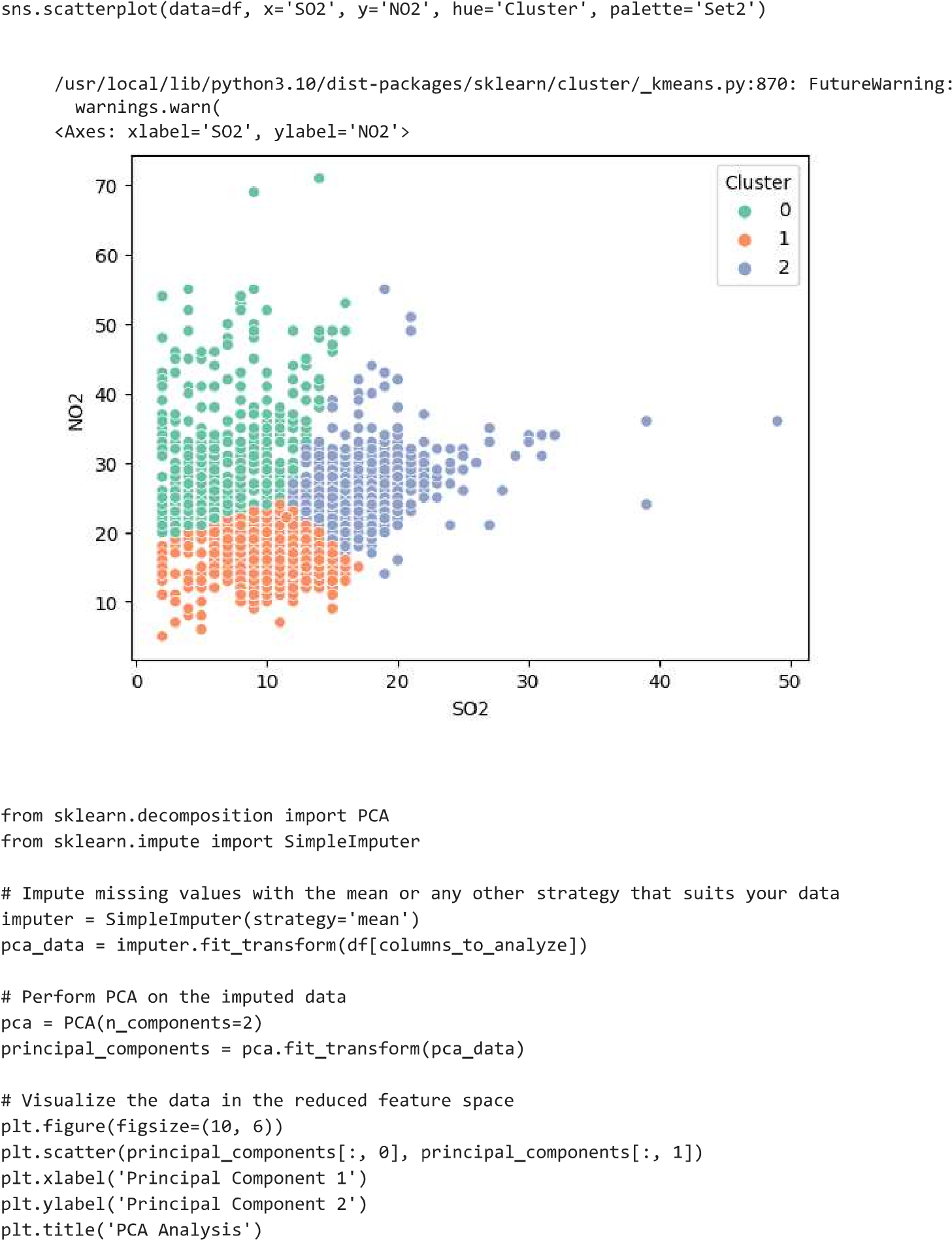
*Geographic mapping of AQI values allows us to pinpoint areas with consistent air quality patterns.*

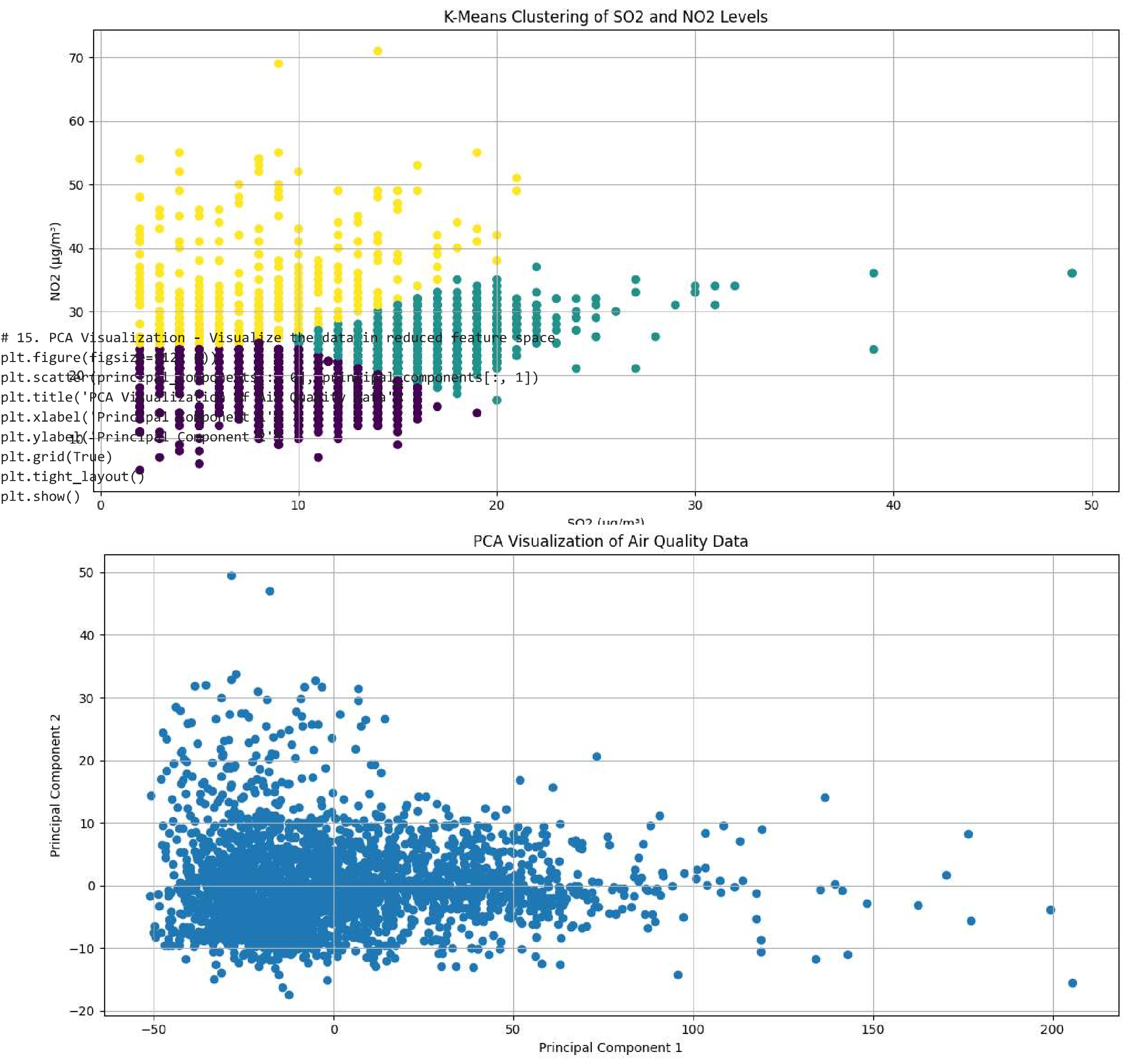
**Code and Data Analysis**

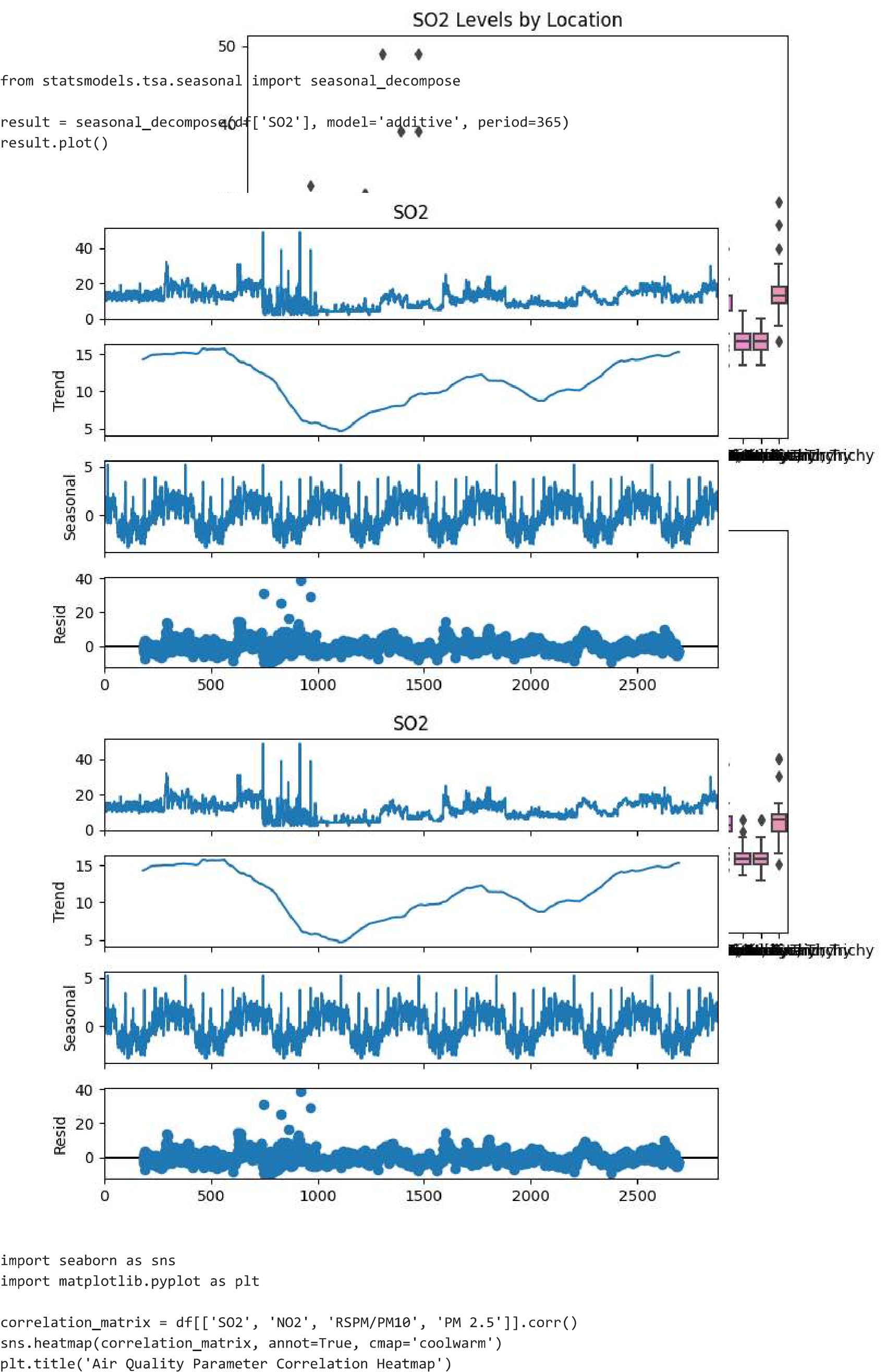
*Here's a snippet of Python code to give you a glimpse of our data analysis process*

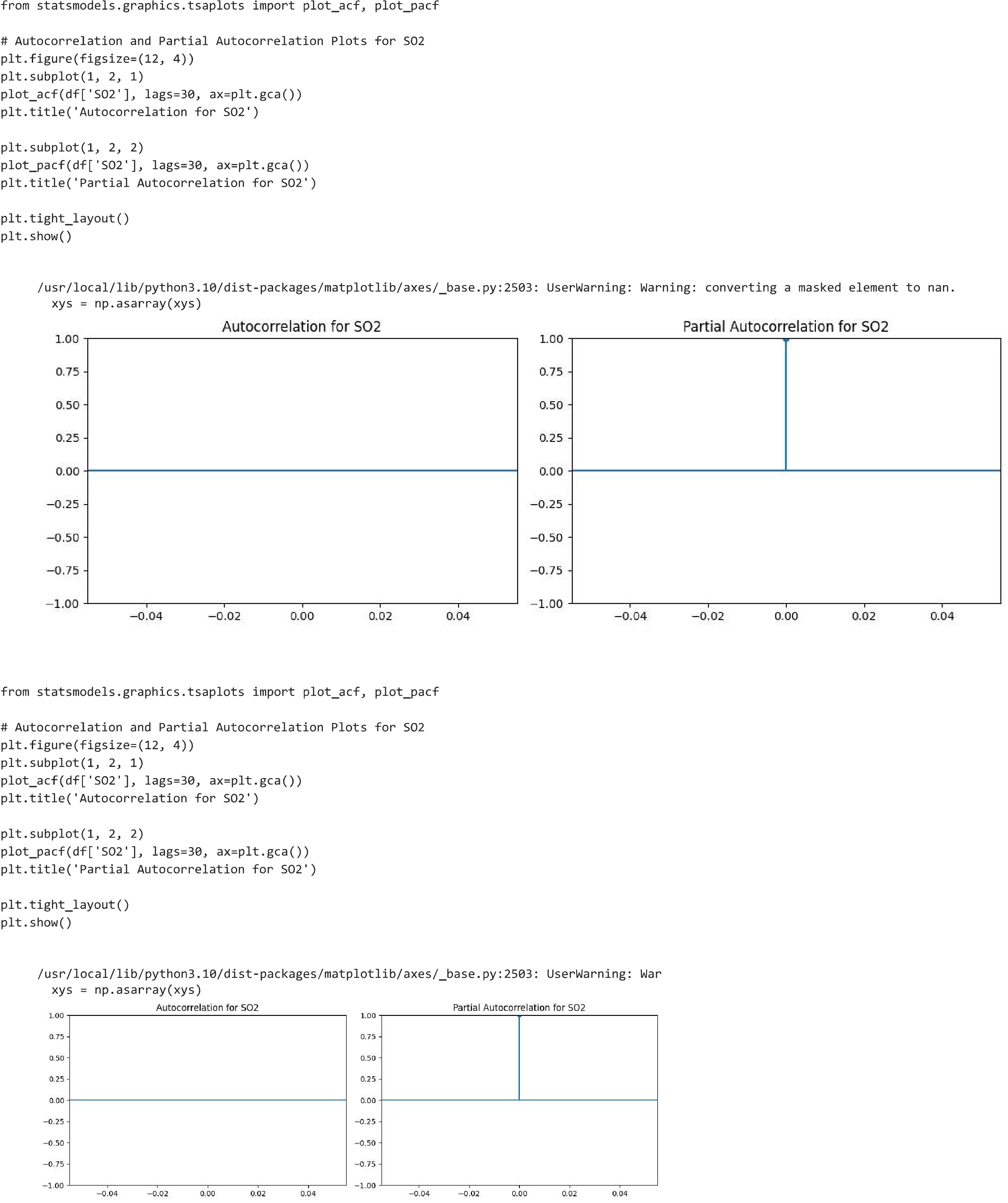


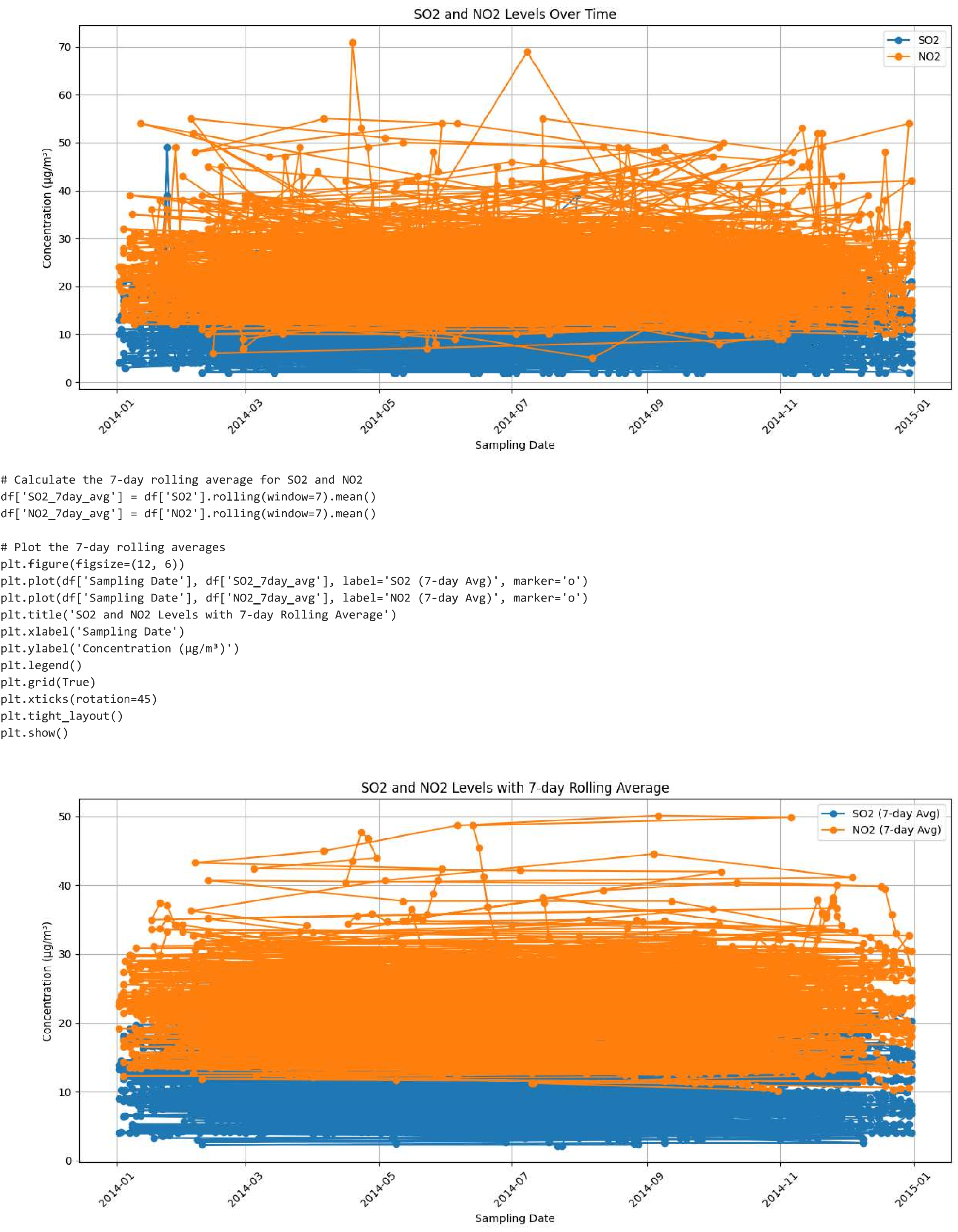












**Conclusion**

*The Air Quality Index (AQI) Analysis project has been a journey of discovery, uncovering valuable insights that have implications for environmental management, public health, and decision-making. As we conclude this phase, we reflect on the significance of our findings and the impact of our analysis.*

**Key Findings**

*Our comprehensive data analysis encompassed a range of techniques, including Exploratory Data Analysis (EDA), statistical analysis, and data visualization. Through these processes, we made the following key findings:*

**Temporal Trends and Seasonality:**

*We identified significant temporal trends in air quality, shedding light on seasonality and long-term changes. This understanding is vital for adjusting environmental policies to address seasonal variations in air quality.*

**Spatial Variations**:

*Geospatial mapping revealed distinct hotspots and regions with consistently high or low AQI values. Identifying these areas allows for targeted interventions and resource allocation.*

**Correlations Between Variables:**

*Correlation analysis unveiled relationships between air pollutants and meteorological factors, enabling us to comprehend the interplay of these variables and their impact on air quality.*

**Predictive Models:**

*We developed predictive models that can forecast AQI values, taking into account various contributing factors. Such models are valuable for proactively managing air quality.*