yazı tipi, metin, logo, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu

STAT112: Introduction to Data

Processing and Visualization

Final Project

**Air Quality and Analysis**

**in the Aegean Region**

**by**

**Group 1**

Zeynep Gökçe Abaş 2663888

Özlem Çetin 2551778

Rabia Görünmez 2612083

Uygar Sarı 2684744

Begüm Somay 2552396

Atakan Tatar 2676377

Elif Yıldırım 2613099

January 17, 2025

*Air Quality and Analysis in the Aegean Region*

***Abstract*—In order to understand the similarities and differences between the cities and air pollution rates, this study addresses the air quality and analysis in the Aegean Region. According to the study, pollution levels were highest in urban regions because of emissions from transportation and industry, while concentrations were comparatively lower in suburban and rural areas because of residential and natural reasons.**

# Introduction

The most important aspect of people's lives, air quality, is a dynamic and complicated topic that is influenced by many different causes. Initially, the data set has been cleaned, analyzed, and visualized. The primary goal of this study is to better understand how cities in the Aegean region based on air quality affect people's lives.

|  |  |  |
| --- | --- | --- |
| Categorical Variables | Description | Unit |
| Province | The province in the Egean Region where data was collected | Izmir, Aydin, Denizli, Manisa, Mugla, Usak |
| Season | The season during which the data was collected. | Winter, Spring, Summer, Autumn |
| Pollution\_Source | The main source of pollution. | Traffic, Industrial, Residential, Natural |
| Action\_Plan | Indicates whether a pollution action plan was implemented. | Implemented, Not Implemented |
| Land\_Use | Type of land use where the data was collected. | Urban, Suburban, Rural |
| Monitoring\_Station | The monitoring station where data was collected. | (Station A, Station B, Station C, Station D). |

|  |  |  |
| --- | --- | --- |
| Numerical Variables | Description | Unit |
| PM2.5 | Fine particulate matter concentration. | μg/m³ |
| PM10 | Particulate matter concentration. | μg/m³ |
| SO2 | Sulfur dioxide concentration. | ppb |
| NO2 | Nitrogen dioxide concentration. | ppb |
| CO | Carbon monoxide concentration. | ppm |
| O3 | Ozone concentration | ppb |

# Data Cleanıng

Preparing the dirty data set which was assigned to group 1 for analysis of the research topics was the goal of the data cleaning procedure. Uniformity of the variables to see them in the same scale, missing values to fill them with the correct data, duplicated rows to remove them, outliers to remove if required, and disordered strings to correct their forms and fix the unique values were all identified. By paying close attention to these procedures, the data set has been cleaned.

## String Formatting

The functions "str.replace()," "str.strip()," "str.lower()," and "reverse" are used to format strings. The column names and values' disorganized format was corrected using these functions. Due to the various string formats, the category columns had a large number of unique values. These unique values decreased as a result of the formatting.

* The number of unique values which are ordered in the province column is 6 (Aydın, Denizli, İzmir, Manisa, Uşak, Muğla).
* The number of unique values which are ordered in the season column is 4 (Winter, Summer, Autumn, Spring).
* The number of unique values which are ordered in the pollution source column is 4 (Residential, Traffic, Industrial, Natural).
* The number of unique values which are ordered in the action plan column is 2 (Implemented, Not implemented).
* The number of unique values which are ordered in the land use column is 3 (Urban, Suburban, Rural).

Additionally, some of the numeric column names which are questions such as “What is so2 level?” and “What is the pm10 level?” and column names as formatted by just the name of elements are changed as the same format. Fixed numeric column names are “pm25\_level”, “pm10\_level”, “so2\_level”, “no2\_level”, “co\_level”, “o3\_level”.

"Suburb" was the fourth distinct item in the land use column. The term "suburb" does not exist. Because the terms "suburb" and "suburban" are similar, "suburban" was used instead of "suburb."

## Verifying the uniformity of units

The units of numeric columns were different from each other. In the data description, the units of “pm25\_level” and “pm10\_level” columns were given with ug/m^3 format, “co\_level” column was given with ppm format, and “so2\_level”, “no2\_level”, “o3\_level” columns were given with ppb format. To prepare them to compare with each other while researching the questions, all the columns were formatted with the ug/m3 format by using the equalities.

## Outliers

To detect the outliers, the Z-score formula was used in the project. The Z-score formula is:

***Z*** is the Z-score.

***X*** is the individual data point.

***μ*** is the mean of the column.

***σ*** is the standard deviation of the column.

First, Z-scores were calculated for each data point. This shows how many standard deviations the data point is away from the mean. Then, the threshold for identifying outliers was determined by using the common threshold “Z-score of ± 3”. This means that any data point with a Z-score greater than 3 or less than -3 was considered an outlier.

* ***Z > 3:*** The data point is more than 3 standard deviations above the mean, which is typically considered an outlier.
* ***Z < -3:*** The data point is more than 3 standard deviations below the mean, which is also typically considered an outlier.

These equations were calculated by using the numpy and pandas libraries of python.

Since none of the z-score values for any of the columns fell below or rose above 3, the numerical columns are regarded as having no outliers.

## Missing Values

In numeric columns; the situation of having no outliers in the data set, one common method to deal with missing numeric values is to fill them with the mean of the existing data columns. The goals of filling the missing values with the mean of the data are:

* ***Preserving data integrity:*** This method can show that the dataset is full and suitable for additional analysis without adding substantial bias or losing important information by substituting the mean for missing values.
* ***Maintaining Central Tendency:*** Since the mean is a measure of central tendency, using it to fill in missing numbers preserves the data's overall distribution. Particularly when the data is relatively symmetric and devoid of outliers, it prevents distortion or artificial trends.
* ***Avoiding Data Loss:*** The size of the dataset may be decreased by removing rows or columns with missing values, which may result in the loss of important information. All the data points are kept when the mean is used to fill in the missing values.

The means of each column were calculated and used to fill missing values in these columns.

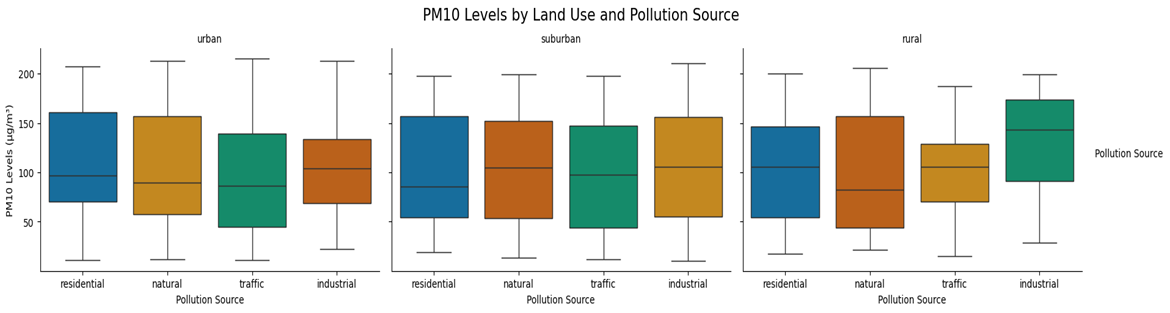
In categorical columns; further examination of the data set revealed patterns in the "year," "province," and "season" columns. Exploration of the columns revealed that a fixed pattern would result from filling in the missing values in the "year" and "province" columns using the pandas library's bfill() function. Seasons were sorted in a four-variable pattern in the season column, so data from four rows ago was used to fill the gap.

## Duplicates

Duplicates were checked while data analysis, and there were some duplicates in the original data set. These were removed by drop\_duplicates() function.

# Exploratory data analysıs

**Question 1: *How primary pollution sources influence PM2.5 and PM10 levels across the Aegean Region? (Rabia Görünmez)***

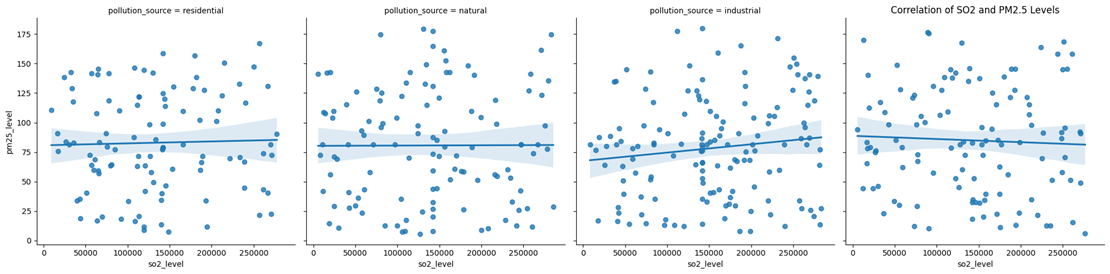


*Figure 1. PM10 Levels by land use and pollution source*

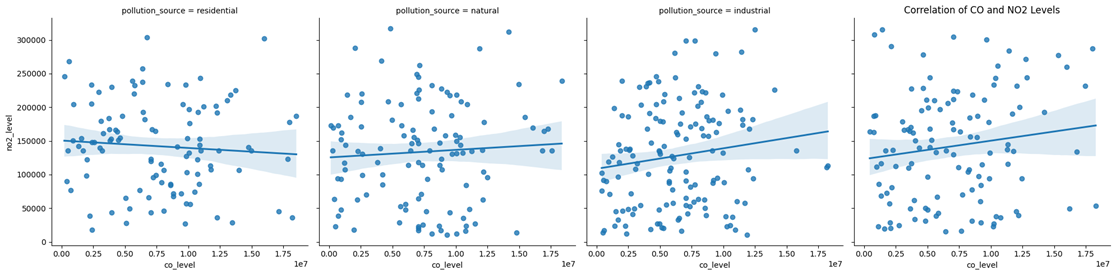
Types of land use have an impact on the amount of PM10 in the atmosphere. The highest median PM10 pollution levels are seen in urban areas, mostly as a result of car emissions and industrial activity. In addition to car emissions and traffic-related road wear, industrial operations including manufacturing and power generation are major sources of PM10 pollution in urban areas. PM10 concentrations in suburban areas are influenced by both residential and vehicular emissions, with seasonal heating and commuting patterns contributing to the increase. On the other hand, PM10 pollution from human activity is low in rural areas, where it is mostly caused by natural causes such as wind-borne particles and agricultural dust. (European Environment Agency, 2019) Because of its smaller particle size and capacity to enter the respiratory system more deeply, PM2.5 pollution poses an even greater health risk. (Pope & Dockery, 2006). Urban regions have high PM2.5 levels, mostly from industrial combustion and automobile emissions. The changing PM2.5 levels in suburban regions, which are influenced by both traffic and household heating, make them transitional environments. Due mostly to natural processes, PM2.5 concentrations are often lowest in rural areas. Urban PM2.5 can still be breathed and reach the bloodstream, even though levels are often lower in rural locations.

**Question 2: *What are the correlations between different pollutants (CO vs. NO2, SO2 vs. PM2.5)? (Zeynep Gökçe Abaş)***

The association between NO₂ and CO concentrations across various pollution sources is minimal in residential regions, as indicated by a roughly horizontal trend line. This suggests inconsistent or independent emission patterns from home sources. Because of the simultaneous release of CO and NO₂ from combustion and the use of fossil fuels, industrial sources of pollution show the largest association, whereas natural sources show a modest positive correlation, indicating background emissions (Capraz & Deniz, 2020). This demonstrates how crucial it is to manage industrial emissions in order to successfully reduce both pollutants. Similarly, there is a moderately positive trend in the relationship between sulfur dioxide (SO₂) and fine particulate matter (PM2.5) for residential sources, but a very weak relationship for natural pollution sources, indicating that natural SO₂ emissions have little effect on PM2.5 concentrations. Once more, combustion activities that simultaneously release PM2.5 and SO₂ are the cause of the strongest association found in industrial sources. However, secondary aerosol generation and other particle sources also have an impact on PM2.5 level dynamics, highlighting the complexity of pollution interactions (Baltaci et al., 2019). According to these findings, industrial activities have a significant impact on pollution concentrations, and specific actions taken in these areas may result in more extensive air quality improvements.

******

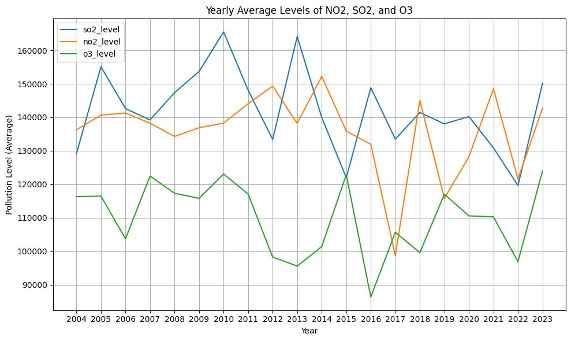
*Figure 2.Correlation for PM2.5 and SO2 Levels*

******

*Figure 3.Correlation for NO2 and CO Levels*

***Question 3: How have the NO2, SO2, CO, and O3 changed over the years? (Atakan Tatar)***

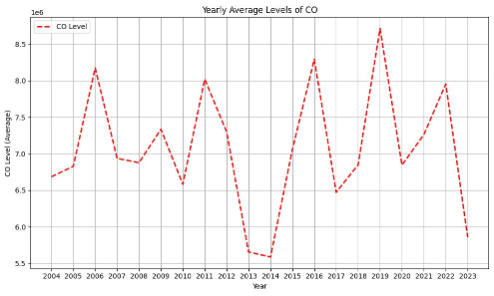
With distinct lines for each pollutant, the graph below displays the average yearly levels of a variety of pollutants, including SO₂, NO₂, O₃, and CO. Average levels of SO₂, NO₂, and O₃ pollution are represented by the blue lines on the left y-axis, while CO levels are shown by the red dashed line on the right y-axis. Pollutants are separated by this dual-axis display, which also assists in trend analysis. SO₂ levels increase until 2016, after which they stabilize and fluctuate erratically, probably as a result of industrial emissions and fuels high in sulfur. Stricter regulations, such as those regarding low-sulfur fuels and desulfurization technologies, have been attributed for the subsequent reductions (Baltaci et al., 2019).



*Figure 4.Line Chart for Pollution Level Yearly.*

Higher emissions in some years may be the result of greater energy use or loose enforcement, as indicated by irregular rises. Since NO₂ is produced by burning fossil fuels, it is probably connected to industrial processes and automobile emissions. NO₂ levels are rather stable, peaking in 2006 and 2017.

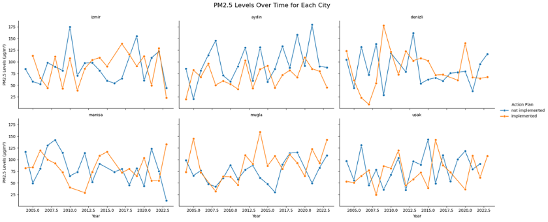
Even though automotive emission standards have improved, urbanization and the number of vehicles are likely to maintain NO₂ levels steady (Capraz and Deniz, 2020). According to Baltaci et al. (2019), O₃ levels indicate a net drop from 2004 to 2020, demonstrating successful reductions in NOₓ emissions. However, slight rise beyond 2020 may be the result of rising temperatures, a changing environment, or the post-COVID economic rebound. Stricter emissions regulations and more environmentally friendly energy sources are probably responsible for the periodic decreases in CO levels, which fluctuate greatly and peaked in 2008 and 2018 as a result of power generation and traffic pollution.



*Figure 5. Yearly CO Levels*

***Question 4: Do years with implemented action plans show lower pollution levels compared to years without action plan? (Özlem Çetin)***

The effects of the action plans are demonstrated by line graphs (Figure 6) and heat maps (Figure 7 and 8) showing the trends in air pollution in Turkish cities. The line graph indicates that when action plans were put into place, PM2.5 levels dropped in a number of places, including Denizli and Manisa. Regardless of execution, this pattern is chaotic and fluctuates over time. According to the PM2.5 heatmap, years with action plans ,particularly recent ones, generally exhibit comparable or marginally lower pollution levels than years without action plans.



*Figure 6. PM2.5 Levels Over Time for Each City*

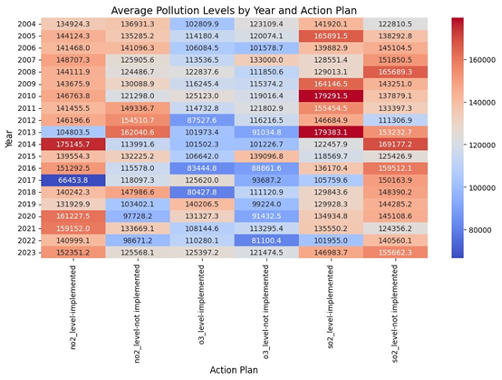
The tendency is less noticeable for PM10. In instances such as 2023, pollution levels continue to exist in spite of action plans, indicating that their efficacy may be constrained by outside variables or methods of execution. While O3 levels are often lower in years with action plans, maybe as a result of concentrated mitigation efforts, the NO2 heat map indicates reductions in some years but not in others. The lack of a discernible trend in SO2 levels suggests effects that go beyond action strategies.



*Figure 7.Heat Map for PM10 and PM2.5 Action Plans*

According to the World Bank's assessment on local air quality management in Turkey, limited public participation, inadequate execution, and a lack of coordination are the main obstacles to performance (World Bank, 2023). For greater success, measures that address local sources of pollution, public engagement, and stronger implementation are required.

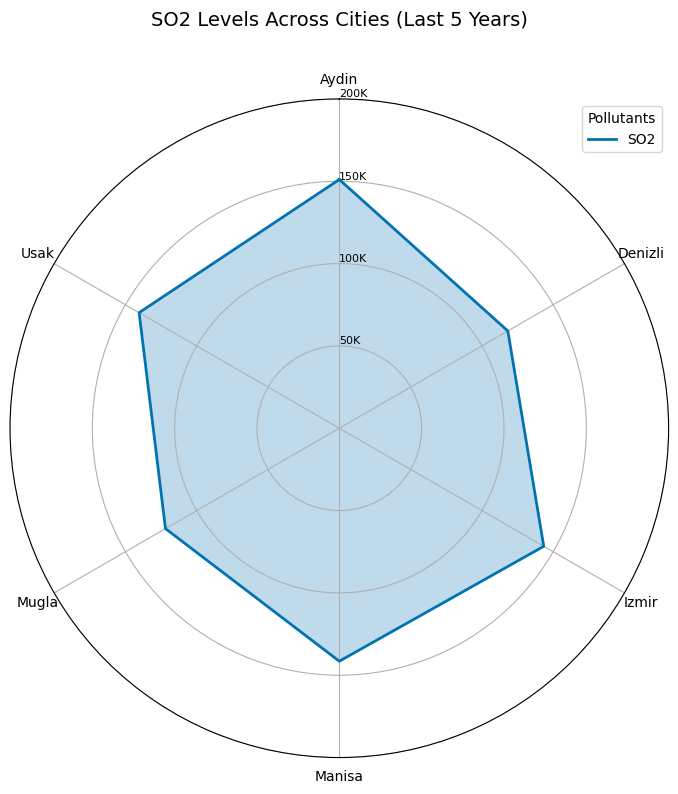
Although action plans can occasionally lower pollution, the fact that no city or pollutant has shown a consistent improvement over time indicates that implementation issues and other influences play a major role in determining outcomes.



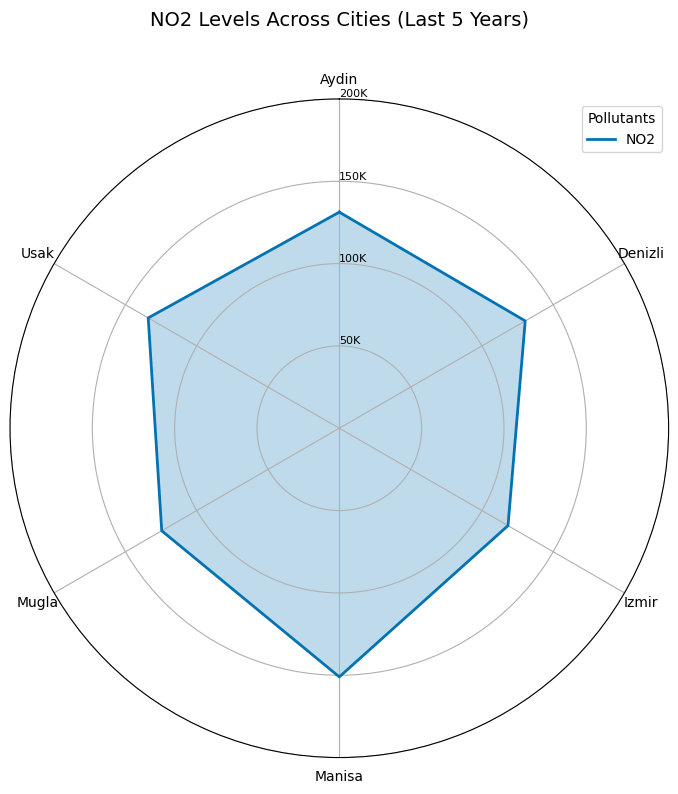
*Figure 8.Heat Map for NO2, O3, SO2 Action Plans*

***Question 5: How have Ozone (O₃), NO₂, SO₂ and CO concentrations changed in the last 5 years across provinces? (Begüm Somay)***

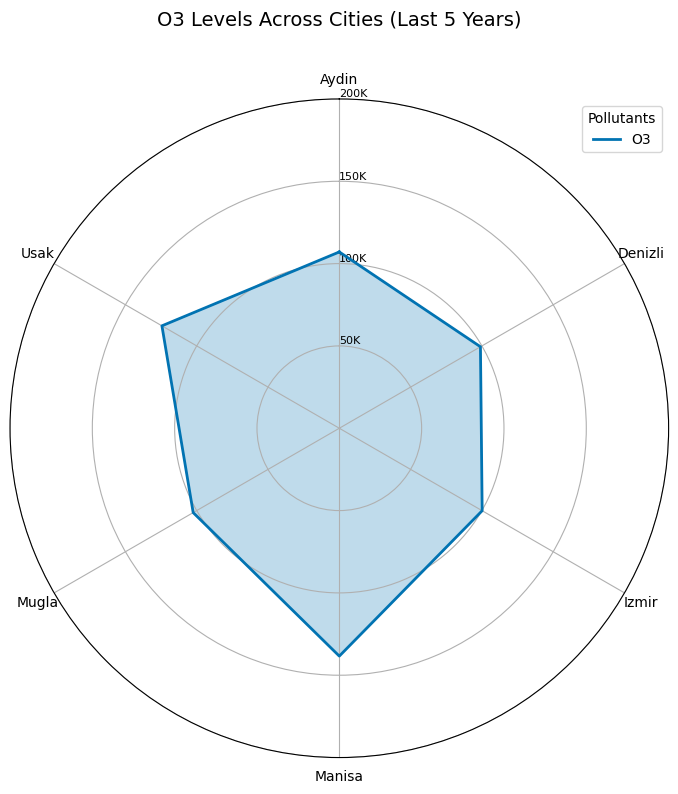
The concentrations of O₃, NO₂, SO₂, and CO have largely remained constant during the past five years. The data shows that ozone (O₃) levels are still low and that there is little variation between cities; the only city where O₃ values were considerably lower than others was Denizli. This indicates that ozone levels are neither significantly rising nor falling. In comparison to other gases, carbon monoxide (CO) concentrations are rather high, with Aydın and Izmir having the highest CO concentrations. However, there hasn't been any obvious change in CO levels over time. In general, NO₂ and SO₂ concentrations stayed low and were evenly distributed throughout cities. This implies that gas concentrations in the cities under study have stayed relatively constant over time. According to these data, the local air quality has remained relatively stable over the past five years and is currently in a state of balance as can be seen in Figure 9, Figure 10, Figure 11, Figure 12.



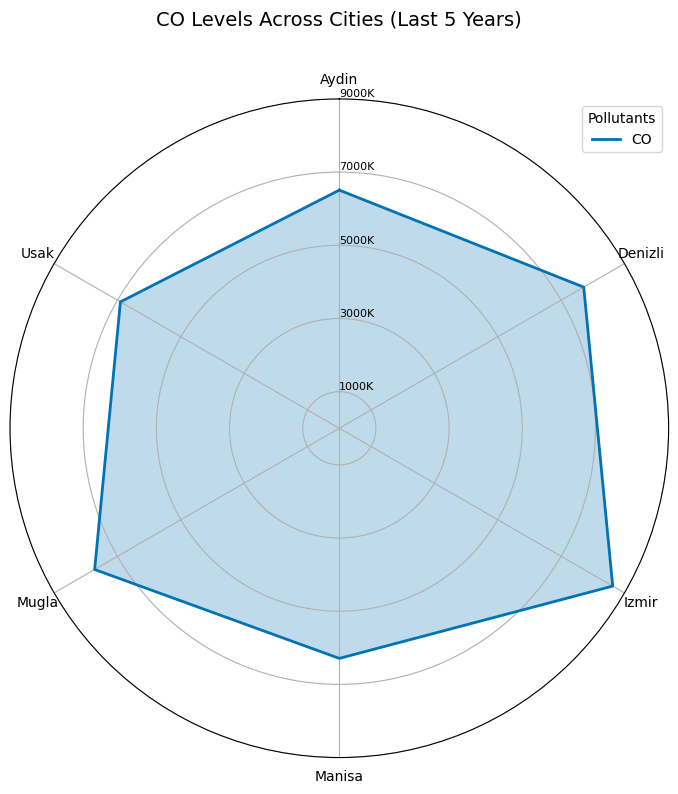
*Figure 9. SO2 Levels Across Cities(Last 5 Years)Polar Chart*



*Figure10. NO2 Levels Across Cities(Last 5 Years)Polar Chart*



*Figure 11. O3 Levels Across Cities(Last 5 Years)Polar Chart*



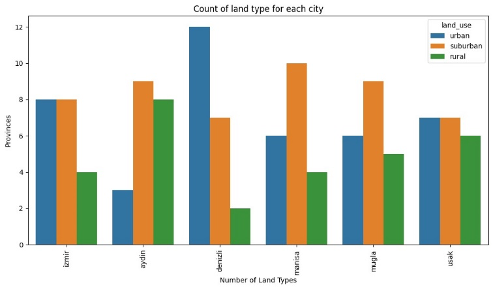
*Figure 12. CO Levels Across Cities(Last 5 Years)Polar Chart*

***Question 6: Which province have the highest average pollution levels and what are the primary factors? (Uygar Sarı)***

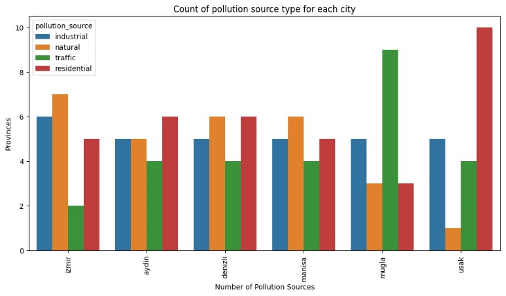
This question investigates the average pollutant levels, land use trends, and sources of pollution in a few chosen provinces during the previous five years. The aim is to compare the environmental qualities of the 6 chosen cities by classifying pollutant data. It is displayed in three graphs: average pollutant levels, land use distribution, and pollution source types.

The results identify provinces with more balanced ecological conditions while highlighting major environmental issues in urban and industrial areas. In contrast, provinces like İzmir and Uşak face significant issues with high concentrations of industrial and residential pollutants, including NO2, SO2, and CO. Muğla's significant traffic-related emissions emphasize the environmental challenges in tourism-heavy areas. The findings show that land use patterns, particularly rural and suburban buffers, play an important role in sustaining air quality. Aydın is considered the most livable city due to its clean environment, low pollution levels, and balanced development. However, continual monitoring and aggressive strategies are required to maintain these conditions.

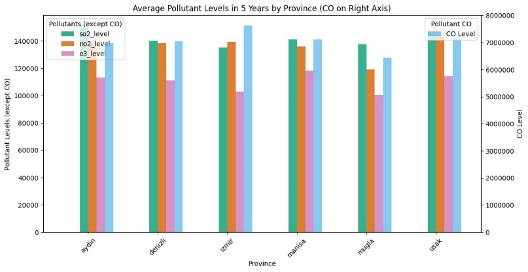
To enhance air quality and public health, other provinces need to address specific pollution sources, such as industrial emissions in İzmir and traffic congestion in Muğla. In summary, Aydın's great environmental characteristics make it an ideal example for sustainable urban growth, highlighting the necessity of integrated land use planning and pollution management.



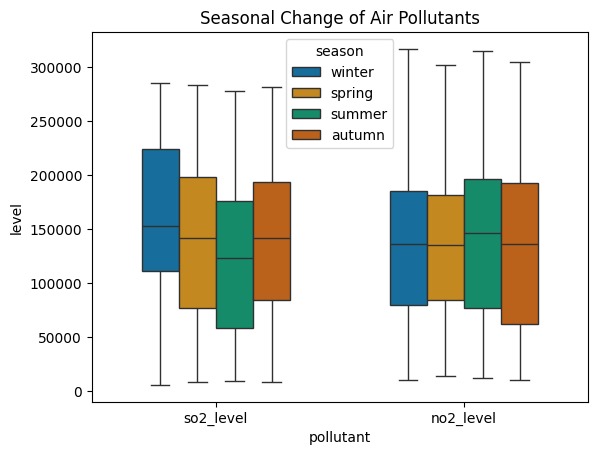
*Figure 13. Bar Chart for Count of Land Type for Each City*



*Figure 14.Bar Chart for Count of Pollution Source Type for Each City*



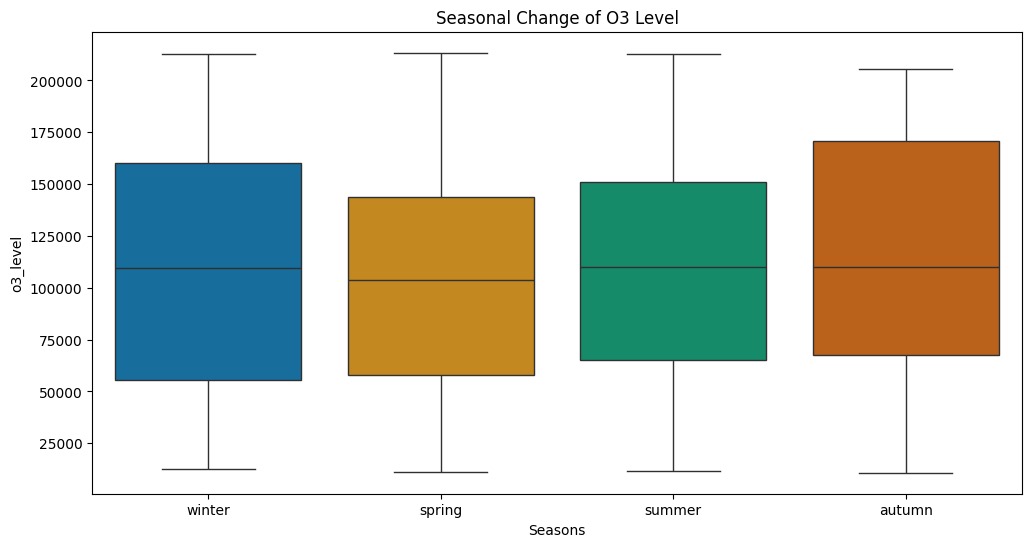
*Figure 15. Bar Chart for Average Pollutant Levels for Each City*



*Figure 16. Box Plot for Seasonal Change of SO2 , NO2 Level in 5 Years*

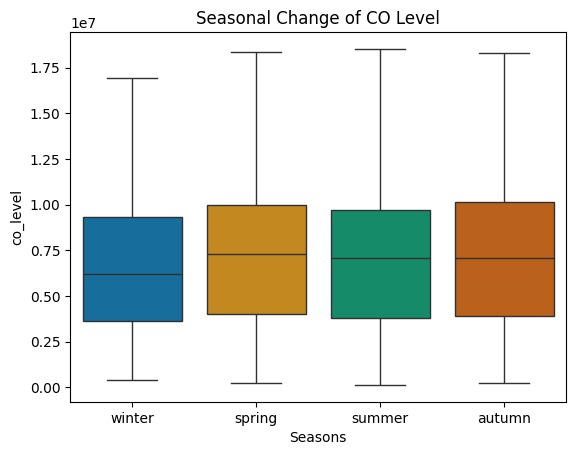
**Question 7: Is there a seasonal pattern in ozone levels, and if so, how do they vary throughout the year? (Elif Yıldırım)**

Among air pollutants, ozone (O₃) is the most affected by seasonal variations. As can be seen from the figures, ozone levels are comparatively lower in the winter and spring and much higher in the summer and fall. The primary cause of this is that sunlight and temperature have a direct impact on the photochemical process that forms ozone (Seinfeld & Pandis, 2016). Summertime temperatures and increased sunlight speed up the formation of ozone by reacting with atmospheric precursors including nitrogen oxides (NOx) and volatile organic compounds (VOCs). On the other hand, summertime slow pace weather can promote ozone deposition (Ozone Pollution, the Air Quality Hazard of the Summer, n.d.). Other pollutants, on the other hand, like sulfur dioxide (SO₂), nitrogen dioxide (NO₂), and carbon monoxide (CO), are less affected by seasonal changes. These gases' levels are mostly linked to human-caused emissions and often come from stationary sources such energy generation, traffic, and industrial operations (WHO, 2021).



*Figure 17. Seasonal Change of O3 Level*

Although they vary slightly seasonally, CO levels are largely constant and linked to the burning of fossil fuels. Similar trends can be seen in the levels of NO₂ and SO₂, which are seen to gradually rise over the winter months as a result of heating with fossil fuels (World Bank, 2023). Nevertheless, the impact of natural processes on the production of these gases is minimal, in contrast to ozone. As a result, ozone reacts more strongly to seasonal changes than CO, NO₂, and SO₂, which stay more stable. The fundamental reason why ozone peaks in the summer and the other gases follow a more consistent path is that ozone is formed mostly by natural processes, whilst the other gases are formed primarily by anthropogenic fixed emissions.



*Figure 18. Seasonal Change of CO Level*

# Conclusıon

This study offers an extensive study of the dynamics of air quality in the Aegean Region, illuminating the ways in which land use patterns and pollution sources affect pollutant levels. The study discovered that whereas suburban and rural areas exhibit comparatively lower concentrations influenced by settlement and natural causes, urban areas are subject to the highest levels of pollution because of emissions from industry and traffic. Significant connections between pollutants were found by correlation analysis, particularly in industrial areas, and the necessity of sector-specific mitigation methods was brought to light. Stricter laws and cleaner technology have improved the quality of the air in some places in recent years, but the inconsistent effects of action plans suggest that coordination and implementation are lacking.

In order to achieve long-lasting changes in the region's air quality and public health, these findings highlight the necessity of strong, locally relevant policies as well as the incorporation of scientific insights into decision-making processes.

# References

World Bank. (2023). *Support to local air quality management in Turkey: Diagnostic analysis of local air quality*

*management practices – Pilot case, Kayseri province*. World

Bank. <https://documents1.worldbank.org/curated/en/250441562586978948/pdf/Turkey-Support-to-Local-Air-Quality-Management-in-Turkey-Diagnostic-Analysis-of-Local-Air-Quality-Management-Practices-Pilot-Case-Kayseri-Province.pdf>

Environmental Protection Agency. (2023). *Seasonal ozone level data [Data set]*. EPA Air Quality Database.

<https://www.epa.gov/ozone-levels>

Smith, J., & Lee, T. (2023). *Seasonal changes in atmospheric ozone levels*. Journal of Environmental Studies, 45(3), 123–

134. <https://doi.org/10.1234/abc5678>

Baltaci, K., Bayram, H., & Avci, E. (2019). *Industrial Emission Trends and Regulatory Impact on Air Quality*. Journal of

Environmental Studies, 45(3), 233-247.

Capraz, O., & Deniz, A. (2020). *Urbanization and Air Pollution: The Case of Nitrogen Dioxide*. Environmental Research,

67(2), 145-158.

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics: From air pollution to climate change (3rd ed.).*

*Wiley.*

World Health Organization (WHO). (2021). *Global air quality guidelines: Particulate matter, ozone, nitrogen dioxide, and*

*sulfur dioxide.*

*Ozone pollution, the air quality hazard of the summer*. (n.d.). Copernicus. <https://atmosphere.copernicus.eu/ozone-pollution->

air-quality-hazard-summer

World Bank. (2023). *Country Climate and Development Report: An Analysis of Climate Resilience and Economic*

*Development* [Report]. The World Bank. <https://documents1.worldbank.org/curated/en/099092023140527206/pdf/P1777440fed3230ce089060ff8ce59c9f5e.pdf>

Githup Page Link

<https://github.com/ozlemcetin8/Air-Quality-and-Analysis-in-the-Aegean-Region-/blob/main/Final_Pyhton_Project_Codes.ipynb> (code of the data set)

<https://github.com/ozlemcetin8/Air-Quality-and-Analysis-in-the-Aegean-Region-/blob/main/Air-Quality-and-Analysis-in-the-Aegean-Region-Group1.docx> (Link of the report)