

Beatriz Veiga

# AIR QUALITY

A Data-Driven Look into Pollution and  
Weather in Portugal



# WHY AIR QUALITY MATTERS

- Air Quality has a direct impact on our health and daily life
- Pollution is a silent but a serious problem
- With urbanization and the change of weather patterns, understanding air quality has become more important than ever

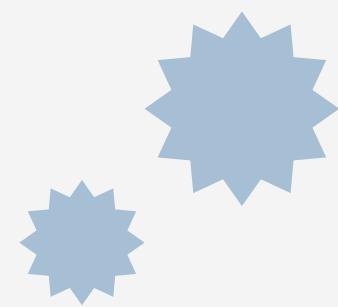






# PROJECT OBJECTIVES

- Analyze air quality trends across Portugal using hourly environmental data
- Uncover how temperature, humidity, and other factors affect pollution levels
- Provide actionable insights through a Power BI dashboard



# ABOUT THE DATA

## Dataset

Air Quality Data

## Time Frame

2023-2024

## Geographical Scope

Portugal, Lisbon  
(Districts, Municipalities, Parishes)

## Source

OpenMeteo  
National Statistics Institute  
Wikipedia



## Key Variables

- Temperature
- Rain
- Humidity
- Precipitation
- PM2.5\*, PM10\*, CO<sub>2</sub>
- AQI
- Time & location info
- Holidays and time of day (feature engineering)

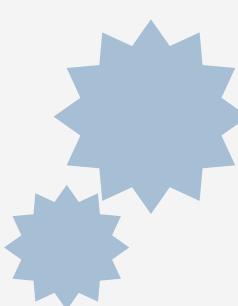
\* PM - Particulate Matters is the term for a mixture of solid particles and liquid droplets found in the air. These can be harmful to human health.

# DATA CLEANING

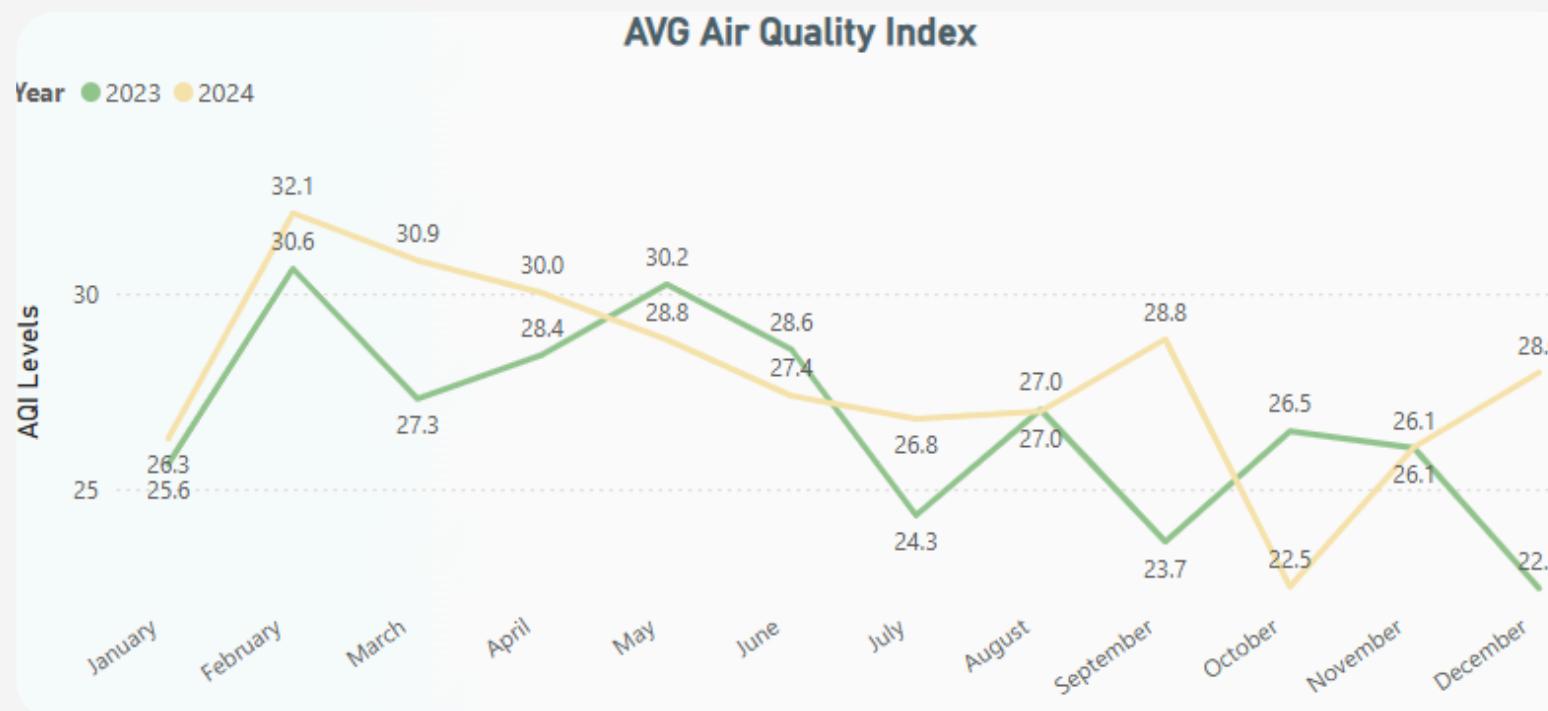
## Preparing the Data

- Split the datetime column
- Added holidays and time of day
- Did not remove outliers - important for pollution spikes
- Reordered and dropped unnecessary data





# HOW CLEAN IS PORTUGAL'S AIR?

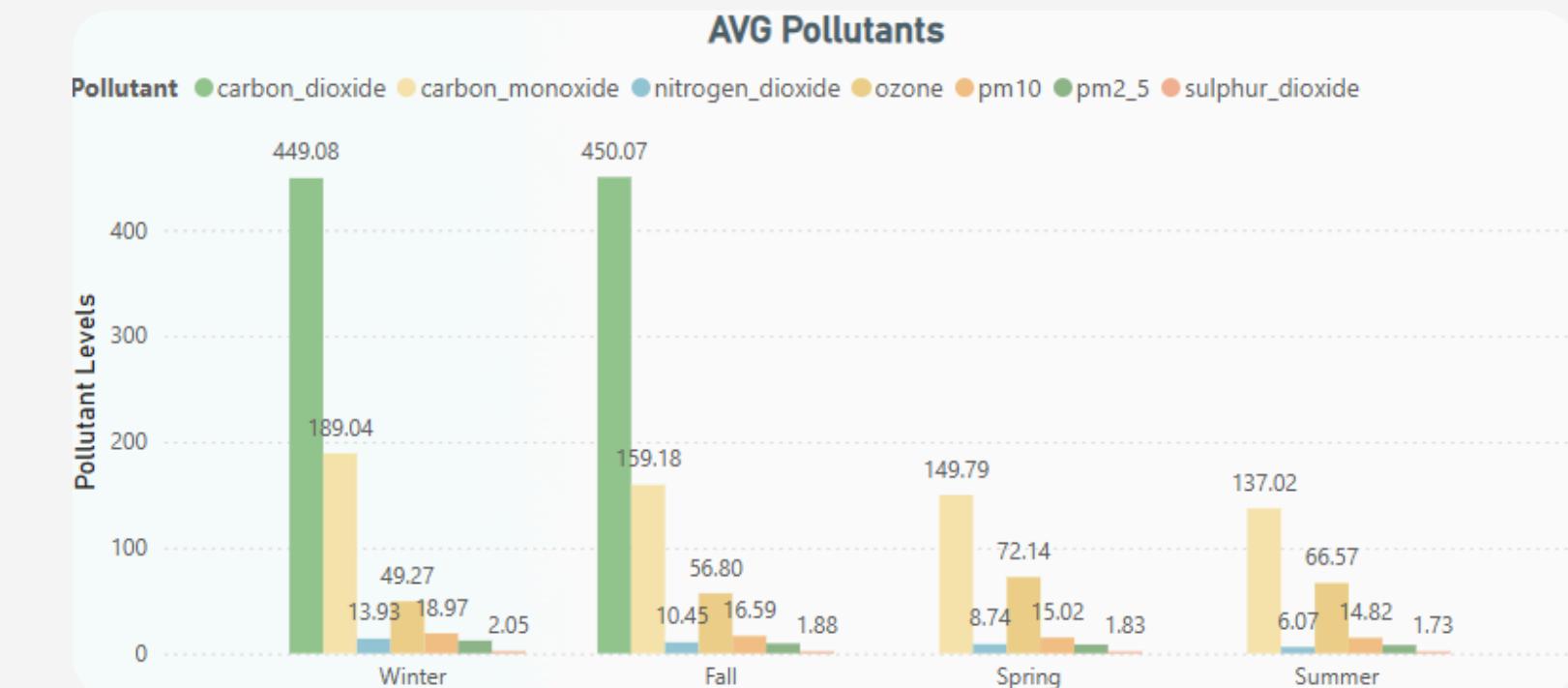


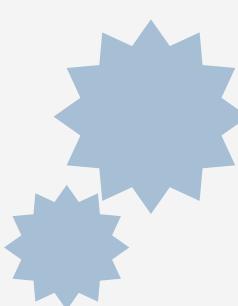
## AVG AQI Over time

- The **first half of 2024** generally experienced **higher average AQI levels** compared to the same period in 2023
- There are notable differences in AQI levels between the two years for certain months, such as **March, September, October, and December**

## AVG Pollutants Over time

- The average levels of **carbon dioxide** and **carbon monoxide** are notably higher in Winter and Fall compared to Spring and Summer.
- Other pollutants like **nitrogen dioxide**, **ozone**, **pm10**, and **pm2.5** remain at relatively low levels across all seasons shown





# HOW CLEAN IS PORTUGAL'S AIR?

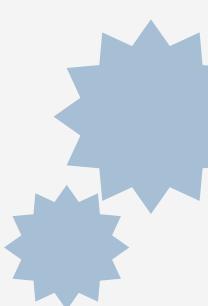
District	European AQI	Population	AQI Per Capita
Bragança	24.9	35.425	0.703743
Guarda	25.1	40.011	0.626336
Viseu	25.2	101.977	0.247359
Viana do Castelo	25.3	86.780	0.291456
Portalegre	25.4	21.914	1.157868
Vila Real	25.5	49.928	0.510022
Évora	25.6	53.937	0.474416
Castelo Branco	25.6	52.913	0.484461
Santarém	25.6	61.009	0.420589
Beja	25.7	33.838	0.759289
Coimbra	27.1	144.822	0.186984
Setúbal	27.1	123.548	0.219308
Faro	27.8	69.468	0.399848
Leiria	27.8	133.795	0.207855
Lisboa	27.9	567.131	0.049310
Braga	28.0	201.583	0.139061
Aveiro	28.7	86.037	0.333372
Porto	29.3	248.769	0.117906

## AQI vs Population (2023)

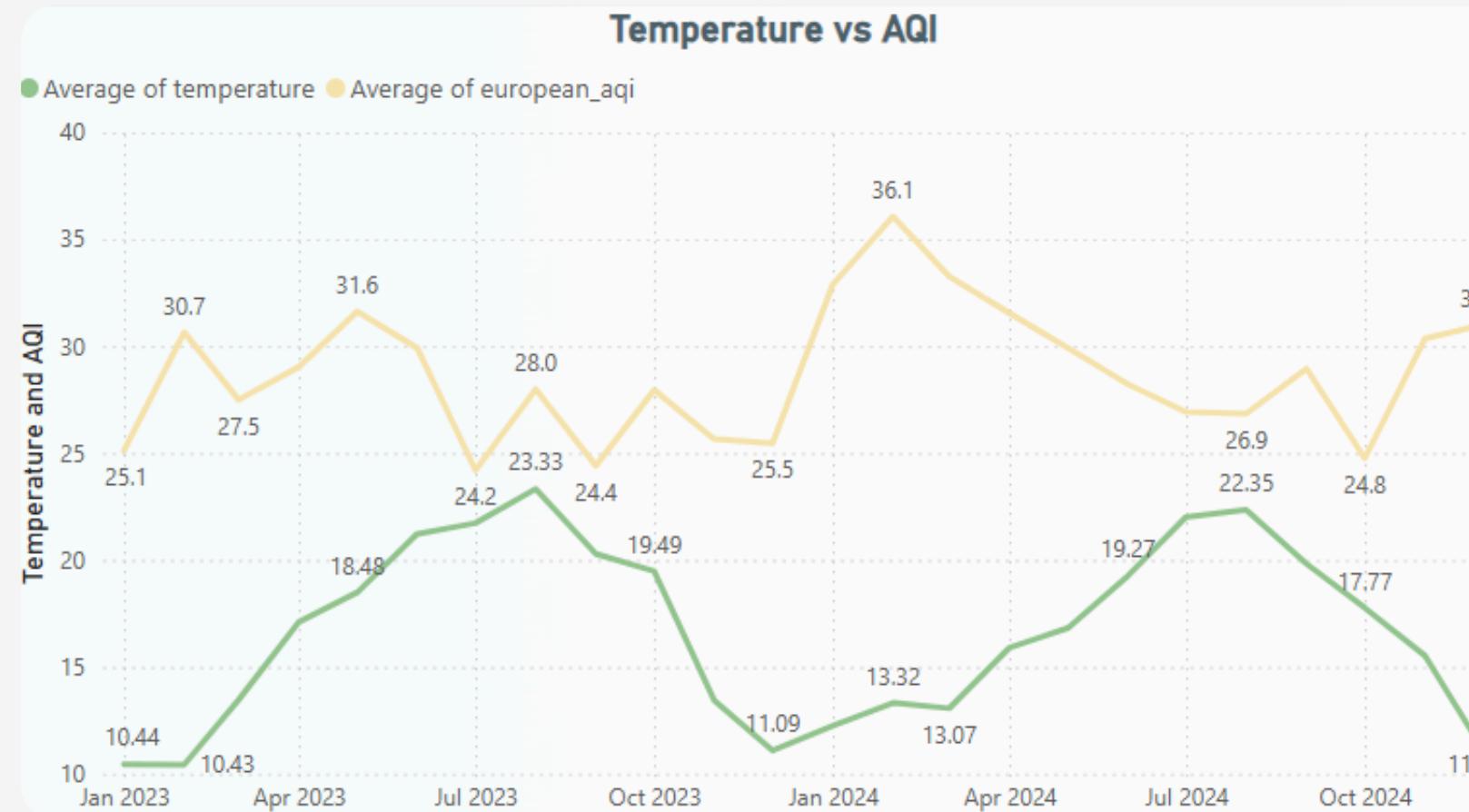
- Larger Cities Like **Lisboa and Porto**: Despite having the largest populations, their AQI per capita is relatively low
- Smaller Cities Like **Portalegre, Beja and Bragança**: These areas show higher AQI per capita despite having smaller populations
- Even with similar pollution levels **Castelo Branco and Évora's AQI Per Capita** differs

## What can we take away from this analysis?

- Bigger cities aren't necessarily more polluted per person
- Rural or less populated regions may have a heavier pollution load per person



# HOW CLEAN IS PORTUGAL'S AIR?

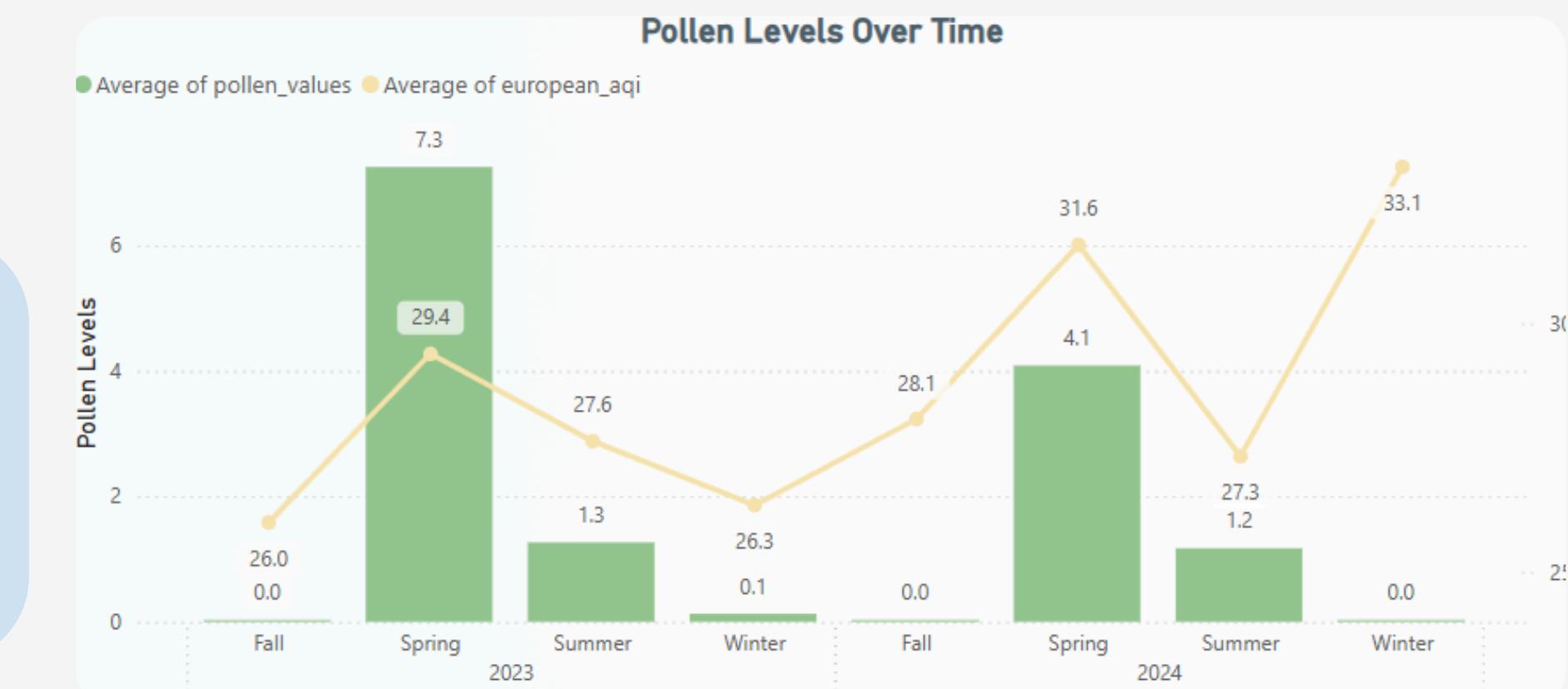


## Temperature vs AQI

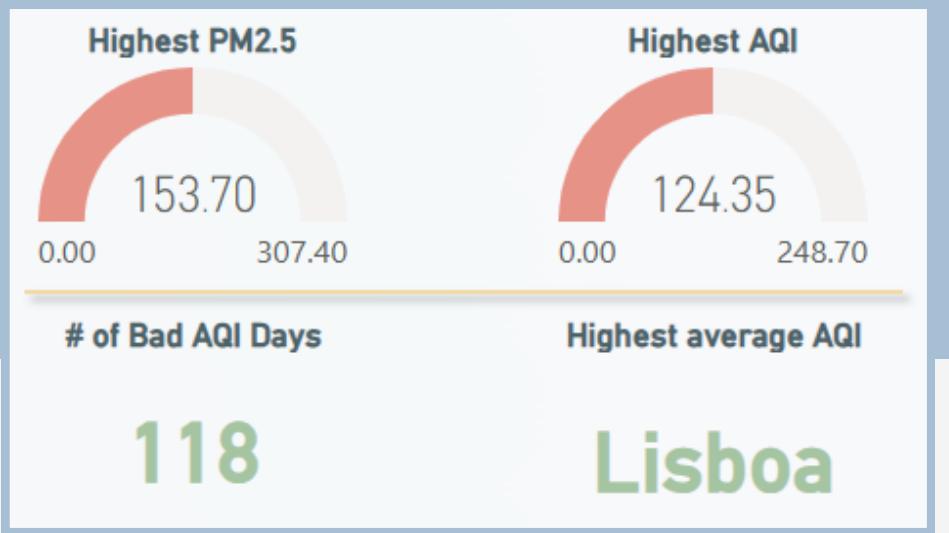
- The temperature shows a clear seasonal pattern
- The AQI tends to fluctuate more and doesn't show a consistent seasonal pattern as temperature
- Around July and October 2023 the AQI is generally lower than April
- Around January and April 2024 the AQI tends to be higher than July

## Pollen Levels and AQI Over Time

- Pollen levels exhibit a distinct peak in Spring, while AQI levels fluctuate across seasons without a clear direct relationship
- This suggests that other pollutants and environmental factors likely have a more significant impact on the average European AQI than the total average pollen values shown here



# KEY FINDINGS \*



## Highest PM2.5 Recorded

- This spike happened in Viseu during September, likely linked to the wildfires happening at the time.

## Highest AQI Recorded

- This was the worst air quality day in terms of AQI. It occurred in Vila Real in September.

## Number of bad AQI days

- Out of two years, 118 of them had a bad or worse AQI rating.

## Location with Worst AQI

- Lisboa had the highest average AQI rating across the dataset.



# HYPOTHESIS 1

## TIME OF DAY AND AIR POLLUTION LEVELS

**H0:** There is no significant difference in air pollution levels between different times of the day.

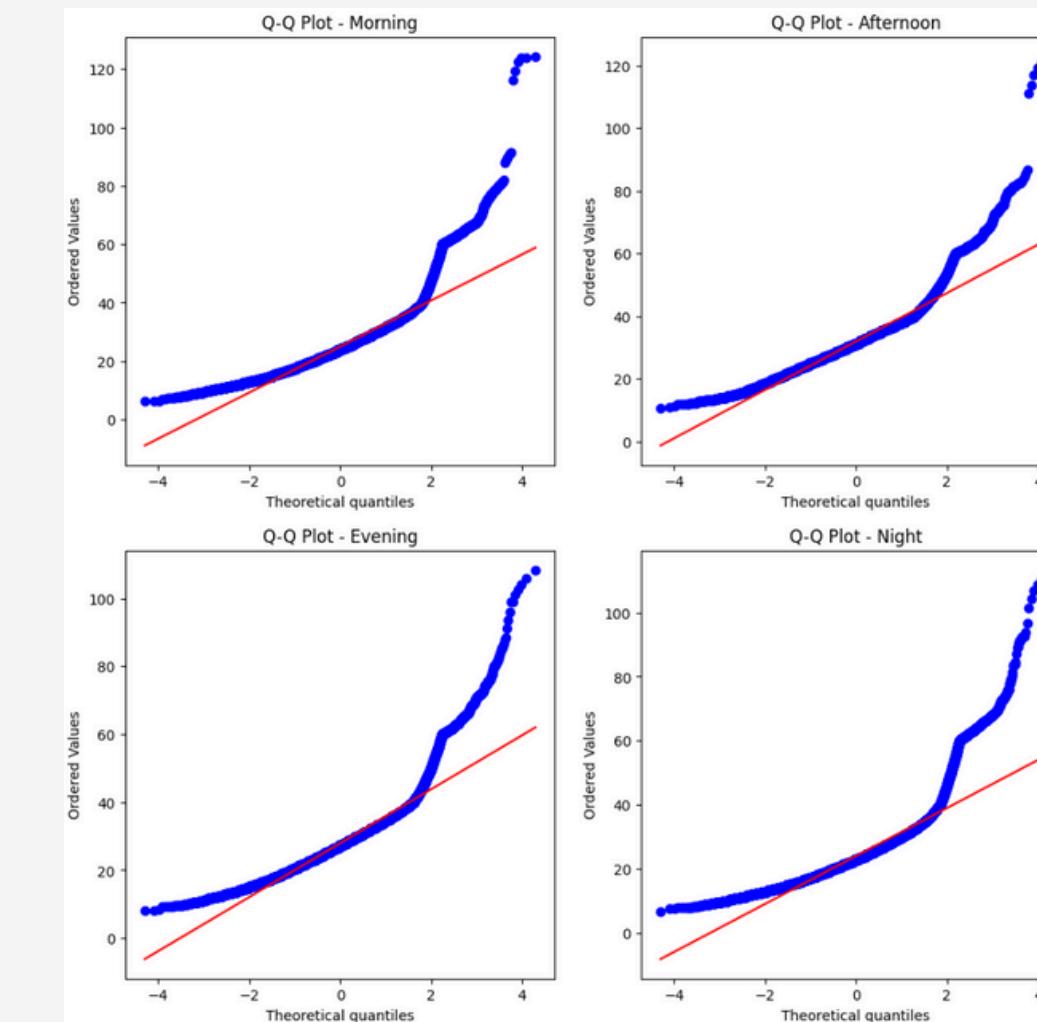
**H1:** Air pollution levels are significantly different when it comes to the time of the day.

### Testing

- Created QQ-Plots to check the distribution of my data
- Since the graphs indicate that there is not a normal distribution, I used the Kruskal-Wallis test

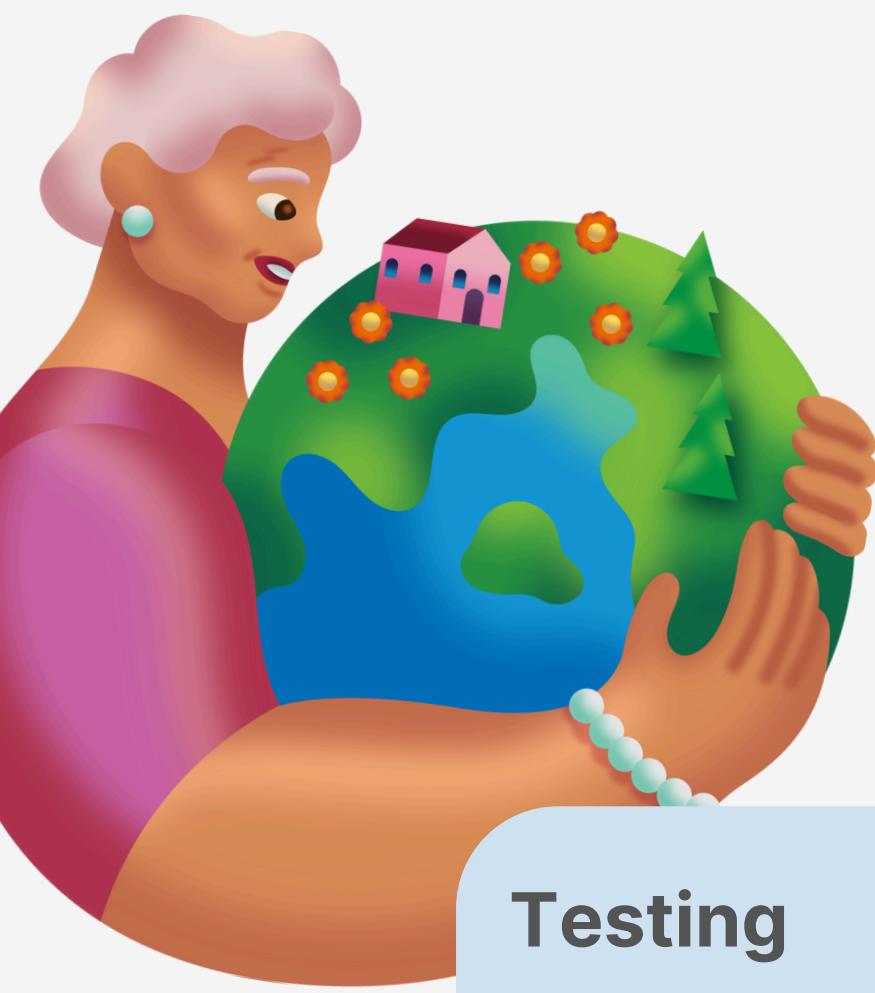
**P\_value < 0.05**

- Reject the null hypothesis** -There is a significant difference in terms of air pollution levels between different times of the day
- To further confirm this hypothesis test I performed a **Dunn's Statistical Test**. From this test I was able to verify the hypothesis result



### Dunn's Statistical Test

	1	2	3
1	1		
2	0	1	
3	0	0	1
4	0	0	0



# HYPOTHESIS 2

## HOLIDAYS AND AIR POLLUTION

**H0:** PM2.5 levels remain the same on holidays as on regular days.

**H1:** PM2.5 levels are lower on holidays.

### Testing

- Since the QQ Plots indicated that there is not a normal distribution, I used the Mann-Whitney U test as it's the best to compare two groups

**U-statistic**

**1606645695.5**

**P\_value**

**7.69e-28**

- Since the p-value is much smaller than 0.05, I can **reject the null hypothesis**. This means that there is a significant difference in PM2.5 levels between holidays and regular days.
- To further confirm this hypothesis test I computed the mean for both the holidays and regular days.
  - The mean PM2.5 levels are slightly lower on holidays (7.81) compared to regular days (8.05)



# CHALLENGES FACED

- Finding good reliable and up-to-date information was more difficult than expected
- Some APIs often had limited historical coverage and different update frequencies
- Some variables (like pollen and traffic) were inconsistent

# FUTURE WORK AND IMPROVEMENTS

- Add real time predictions
- Expand analysis to more localized data, such as neighborhood-level
- Add an interactive map based on the user's needs
- Develop real-time notifications for high-risk populations based on air quality thresholds



# CONCLUSION

Weather conditions and time significantly influence air quality across Portugal.

- Winter and Fall are the most polluted seasons
- Spring and Summer show improved air quality, largely due to higher temperatures
- Urban areas show lower AQI per capita despite high overall pollution
- The presence of holidays leads to a slight decrease in AQI
- Pollen levels do not have a significant impact on the AQI in Portugal, suggesting that other pollutants dominate air quality concerns in most regions



# CONCLUSION

## Actionable Insights

- During colder months switching to solar or wind sources could help reduce emissions
- Investing in and promoting cleaner, more efficient public transport options could help reduce car emissions, especially in urban areas
- Encouraging carpooling initiatives which could reduce the number of vehicles on the road
- Encouraging green roofs and vertical gardens on buildings could reduce air temperature and improve air quality



# DASHBOARD



# THANK YOU

