# Analyzing the Impact of Air Quality on Crop Production in Germany

## Introduction

Several studies and research on air pollution and agriculture shows that air quality affects crop production because the toxic chemicals in air results in slower photosynthesis, thus lesser crops. Over the last few years, interest is growing on uncover significant impact of environmental factors like air quality on agricultural production. This project report explore the potential relationship between air pollutants and crop yields examining data from recent years, this report seeks to uncover any patterns or insights that could inform future agricultural practices and policy decisions.

## Used Data

1. Air Quality Data: **Structure:** Columns: City, Pollutant, Value, Last Updated (year), Country Label. 2. Crop Production Data: **Structure:** Columns: Area, Item (crop type), Element (production quantity), Y2018, Y2019, Y2020, Y2021, Y2022.
**Data Processing:** Filtering, Transformation, Merging: Filtered to include only measurements from Germany, with the Last Updated column converted to the year. Crop production data is filtered for Germany and transformed to a long-form DataFrame. The datasets are merged on the year to analyze the impact of air pollutants on crop production.

## Analysis

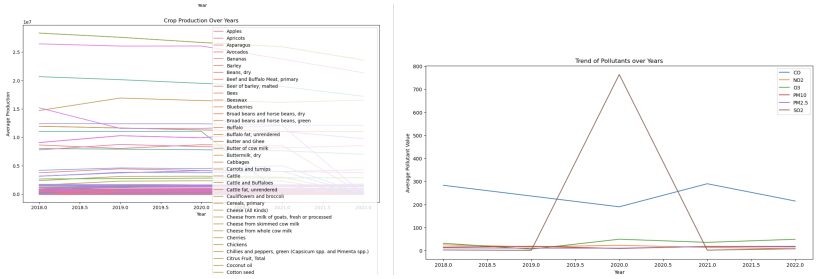
### Summary Statistics

The summary statistics signify high inconsistency in air quality data with critical outliers, by the time crop production data shows reliable averages across years.

### Data Distributions

**Air Quality Values:** The distribution is substantially skewed with most values concentrated at the lower end and a few extreme outliers just as recommending

that air quality is generally good with occasional periods of very poor air quality. **Crop Production Values:** Similarly skewed, indicating that many crops have low production levels while a few have very high production.




(a) Crop production over years (b) Trend on pollutants over years

Figure 1: Time-series Analysis

## Correlation Heatmap

The negligible correlation between air quality and crop production (0.0037) and the very weak negative correlation between year and crop production (-0.029) indicate no meaningful linear relationships between these variables. So, variations in air quality values do not linearly explain the variations in crop production within this dataset.

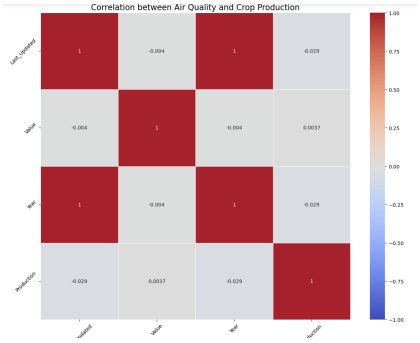
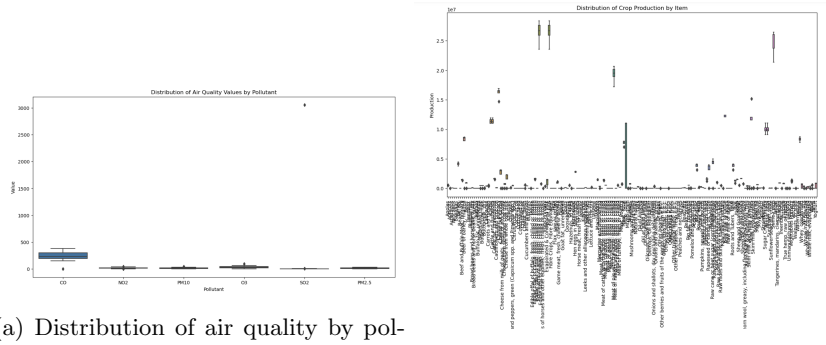



Figure 2: Correlation between Air Quality and Crop Production

## Trends Over Time

**Trend of Pollutants over Years:** Sharp increase in SO2 in 2020 suggests an anomalous event; overall decrease in CO levels might indicate improved air quality management.

**Crop Production Over Years:** Stability in crop production suggests resilience to varying air quality levels, indicating other factors are more influential on crop yields.




(a) Distribution of air quality by pollutants

(b) Distribution of crops by item

Figure 3: Comparing Distributions using box plots

## Comparing Distributions using Box Plots

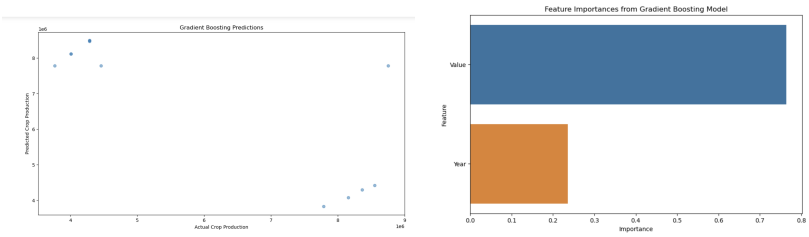
**Air Quality Values by Pollutant:** CO levels are more variable and prominent compared to other pollutants, which are tightly clustered near zero.
**Crop Production by Item:** Notable inconsistency with a few crops showing high production values and many crops having low production values, suggesting a need for focused analysis on high-producing crops.

## Applying Machine Learning Model and Finding Non-Linear Relationship

**Gradient Boosting Model for Crop Production Prediction:**
• Filtered Data: The data is filtered for specific crops (Wheat and Barley) and cleaned.
• Feature Engineering: The features used are Value (air quality value) and Year, while the target is Production.
• Model: A Gradient Boosting Regressor is trained to predict crop production.
**Output and Visualization:** The model underperformed with an R-squared of -2.513 and RMSE of 3882074.57 and faced complexity while predicting crop production. Hence, further tuning or additional features are suggesting for further actions.

## Feature Importance and Partial Dependence Plots

**Feature Importance:** Identify the key features influencing crop production predictions using the Gradient Boosting Model.




(a) Gradient-Boosting prediction (b) Feature importance from GBM

Figure 4

• Air quality (Value) is a significant predictor of crop production than the Year.
• Indicates air quality significantly impacts crop production predictions.
**Partial Dependence Plots:** Visualize and Understand the nature and direction of each feature's impact on crop production.
• Non-linear effect of value: Air quality shows a non-linear impact where initial increases in pollutants slightly boost production, but high levels reduce it.
• Year: Crop production generally improves over time due to better techniques and external factors.

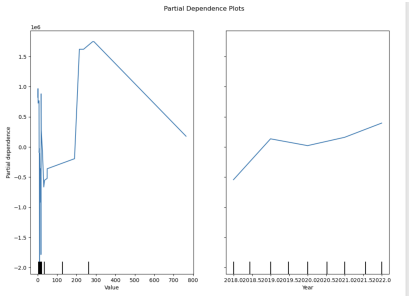



Figure 5: Partial dependence plots

# Conclusion

These visualizations and analyses collectively provide a comprehensive understanding of how air quality affects crop production. The resulted analysis reveals no significant linear relationship between air quality and crop production. However, a non-linear relationship is evident, with moderate pollutant levels not severely impacting production but higher levels reducing yields. The feature importance analysis indicates air quality as a crucial factor in predicting crop yields. Maintaining pollutant levels below harmful thresholds is essential for ensuring optimal crop production, necessitating stricter environmental regulations and improved pollution control measures.