

AIR QUALITY

by Vasilisa, Tymur, and Sofia

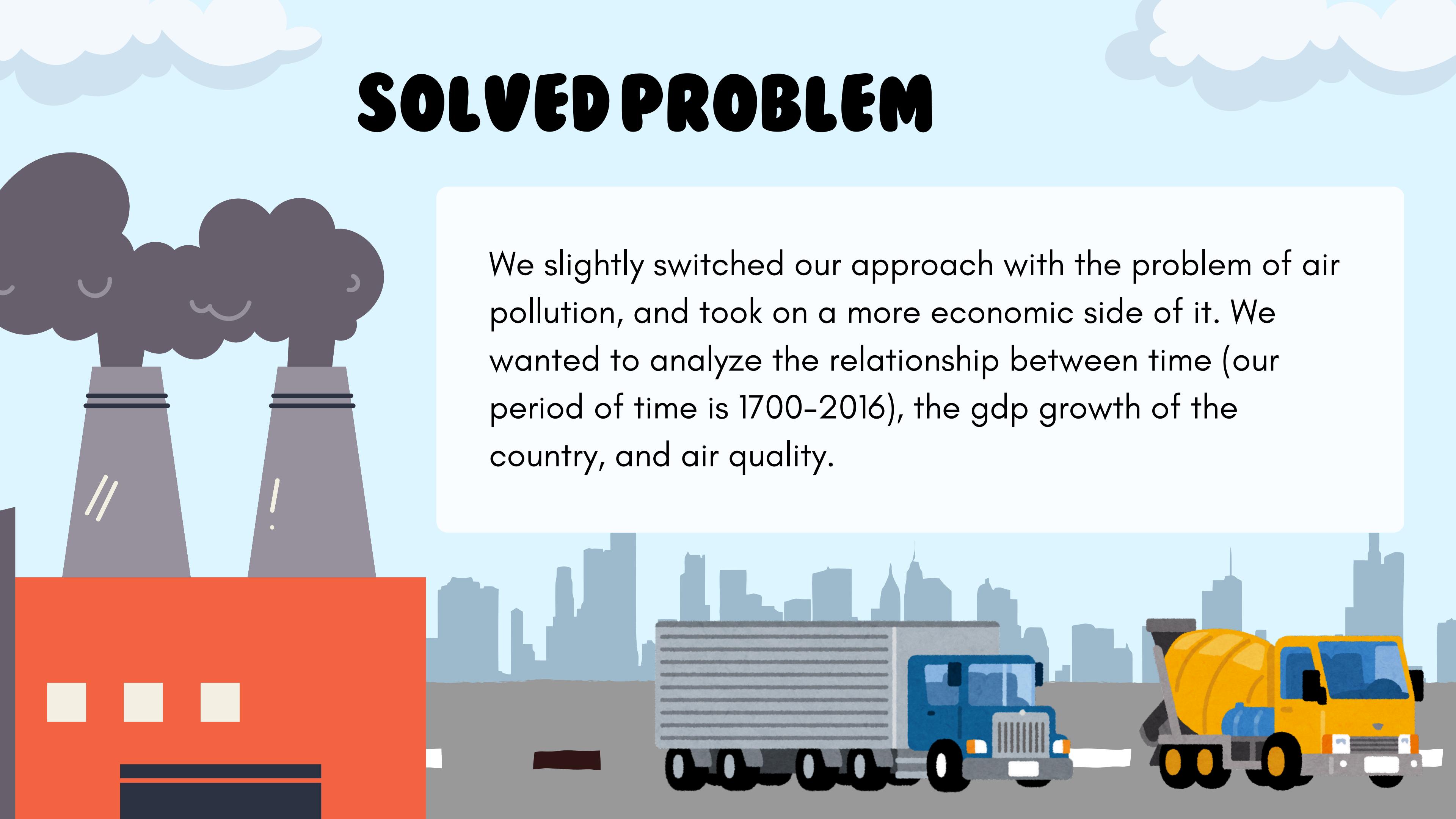


TABLE OF CONTENTS

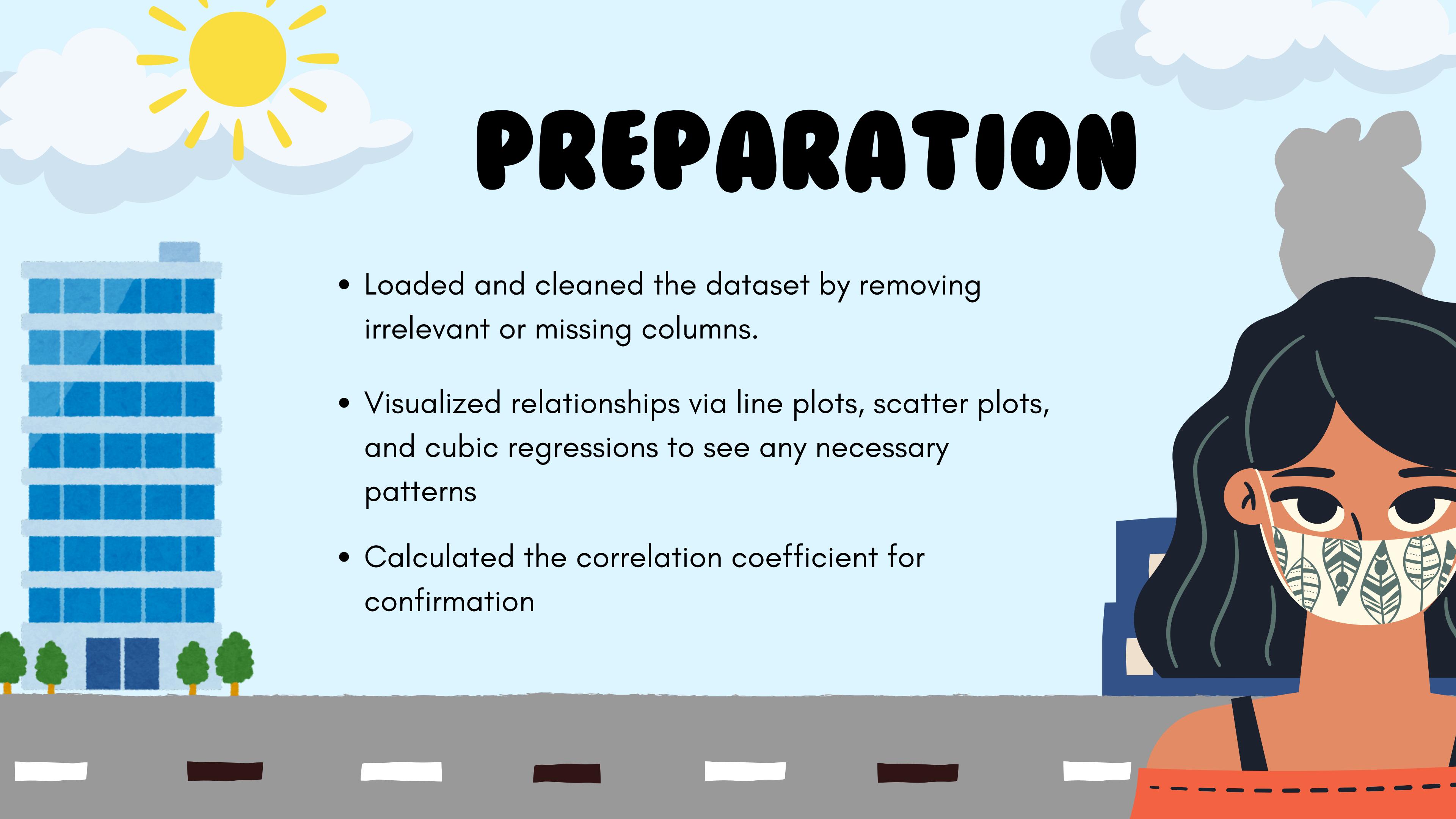
- Solved Problem
- ML models we used and their results
- Results of model exploration
- What we learned
- Key underlying factors and how to improve them
- Improvements



SOLVED PROBLEM



We slightly switched our approach with the problem of air pollution, and took on a more economic side of it. We wanted to analyze the relationship between time (our period of time is 1700-2016), the gdp growth of the country, and air quality.



PREPARATION

- Loaded and cleaned the dataset by removing irrelevant or missing columns.
- Visualized relationships via line plots, scatter plots, and cubic regressions to see any necessary patterns
- Calculated the correlation coefficient for confirmation

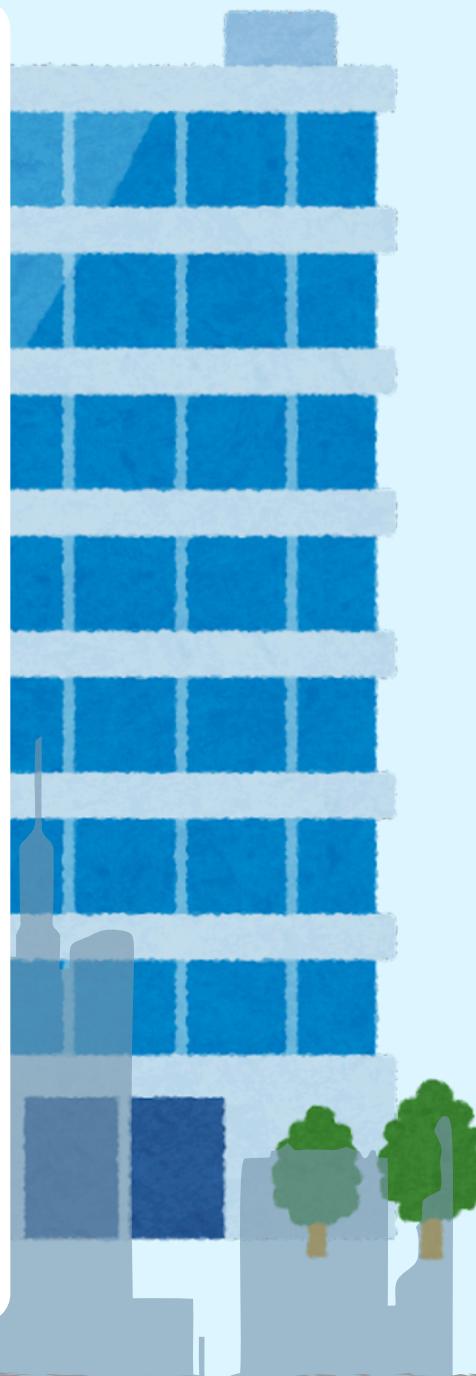
MODELS AND THEIR RESULTS

Model	Linear Regression	KNearestNeighbours	Random Forest	Decision Tree	Bagging
Performance	Very poor	poor	Best	Good	Excellent
Reasoning	Not suitable due to non-linearity of data.	Limited due to the data's time dependent nature.	Most reliable model, captures complex patterns effectively	Performs well but prone to overfitting.	better than decision tree but not random enough



RESULTS AND WHAT WE LEARNED

- The relationship between GDP and SPM is non-linear, following the EKC pattern—pollution peaks at middle-income levels then declines.
- Random Forest regression provided the most accurate predictions, with an R² close to 1, very low error metrics, and consistent cross-validation scores.
- Linear models are inadequate for such non-linear relationships. Tree-based ensemble models are highly effective for complex environmental and economic data.
- Visualizations confirmed the EKC pattern, giving more evidence for existing environmental-economic theories.

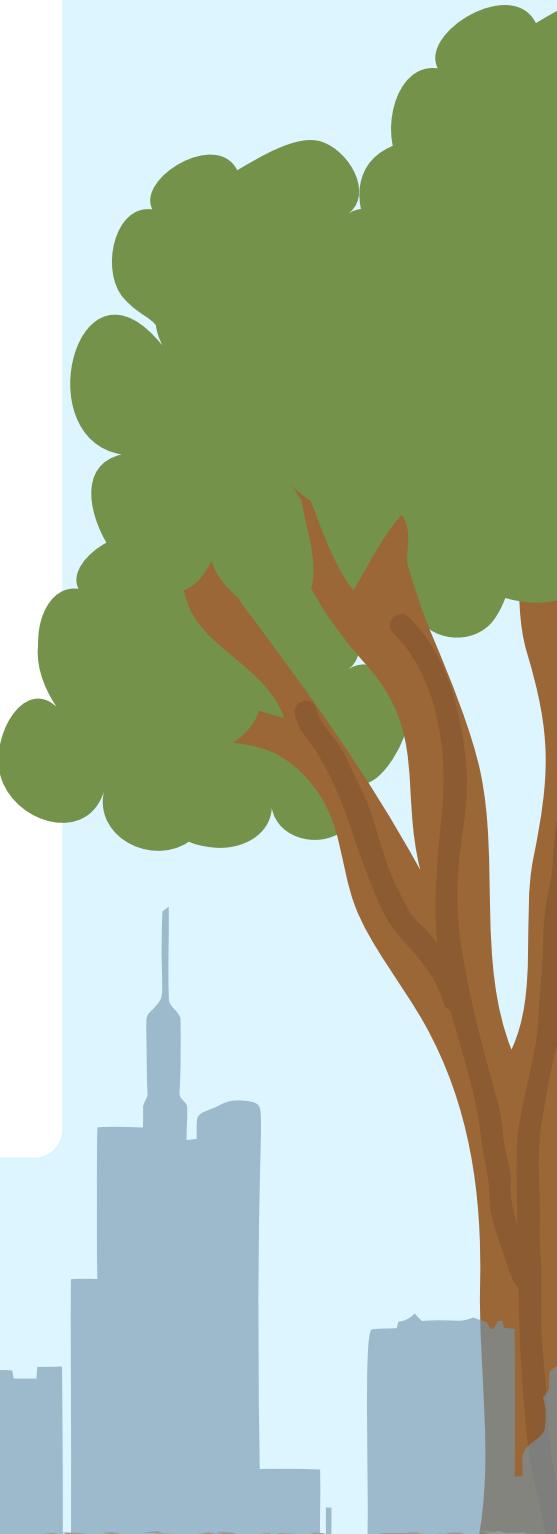


KEY FACTORS AFFECTING AIR POLLUTION BASED ON OUR RESULTS

AND SOLUTIONS FOR THEM



- In the early years the SPM was heightened due to coal mining and cold winters. Coal is still used as a source of heat and energy today, and a solution to that would be implementing more eco-friendly and green methods such as solar-panels or windmills. This would reduce the amount of released greenhouse gasses and help with the quality of the air we breathe
- With the increase in GDP due to industrialization, more coal usage was implemented, leading to even worse quality air (as we can see in the spike in SPM). This could be resolved using the same methods as stated above.

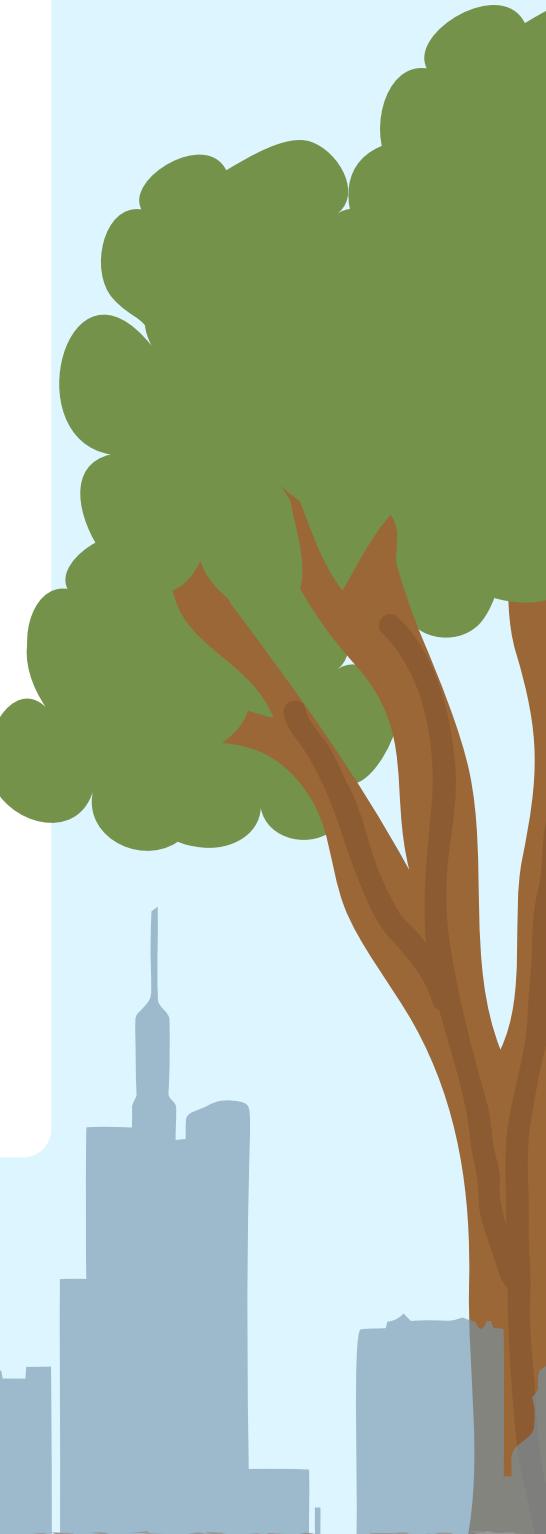


KEY FACTORS AFFECTING AIR POLLUTION BASED ON OUR RESULTS

AND SOLUTIONS FOR THEM



- Later on, with the progression of urban transport, now we have a problem with just how much CO₂ cars release on the daily into the air. This could be remedied by encouraging usage of public transport, which requires way less fuel per person than a personal car, or such transport as bikes, however that would also mean having to apply a better, walk-friendly structure to cities.
- In addition, previously, despite reinforcement of eco friendly ways, the monitoring of people following them was poor, so, using our modern technology, we can apply AI in order to not only monitor changes in factors and air quality, but enforce these methods.



POSSIBLE IMPROVEMENTS

- Data Enrichment: Incorporate more features (e.g., industrial activity, policy changes, population density) for richer modeling.
- Model Tuning: Optimize hyperparameters using Grid Search or Random Search.
- Temporal Models: implement models such as Autoregressive Integrated Moving Average for better work with cases when time is the dominant feature



THANK YOU!

