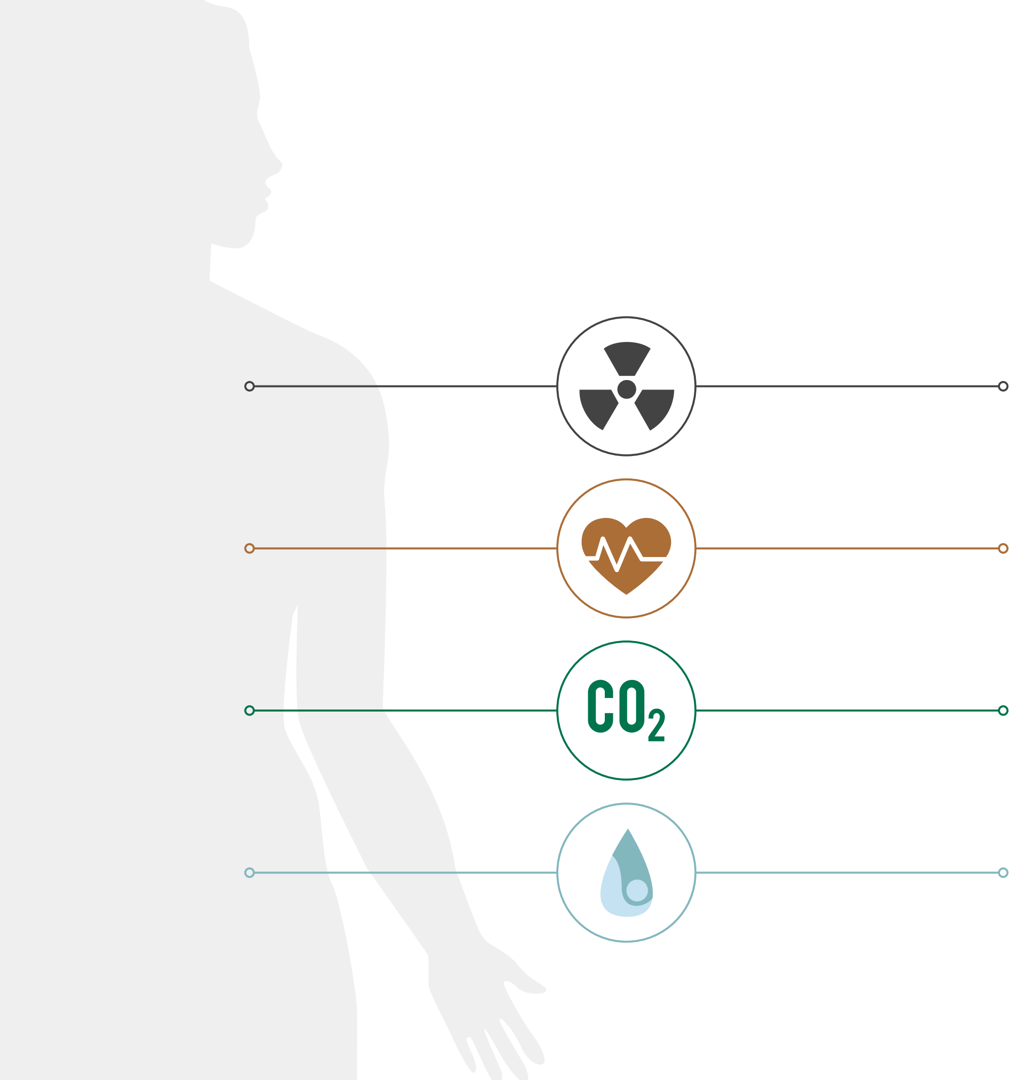


Asthma and Air Pollution

April 30, 2024

Ainsley Atherton
Nikki La
Jordan Ledbetter
Palaniappan Vijay Sithambaram

Introduction & Background

Current State

39% of Americans live in areas with poor air quality - an increase from last year

Importance

Continuous exposure may lead to respiratory diseases

Asthma Prevalence

7.6% of the U.S. population suffers from this condition

Impact

Air pollution contributes to 334 million asthma cases worldwide

Problem

Despite decades of effort and improvement, pollution levels are still unhealthy

Racial Disparity

56% of people of color in the U.S. live in poor air quality regions

Data Sources

Hospitalizations

Collected from the CDC, 2000-2021

ER Visits

Collected from the CDC, 2000-2021

Yearly Air Quality

Collected from the EPA, 2000-2021

Asthma Prevalence

Collected from the CDC, 2018-2021

Air Quality Index

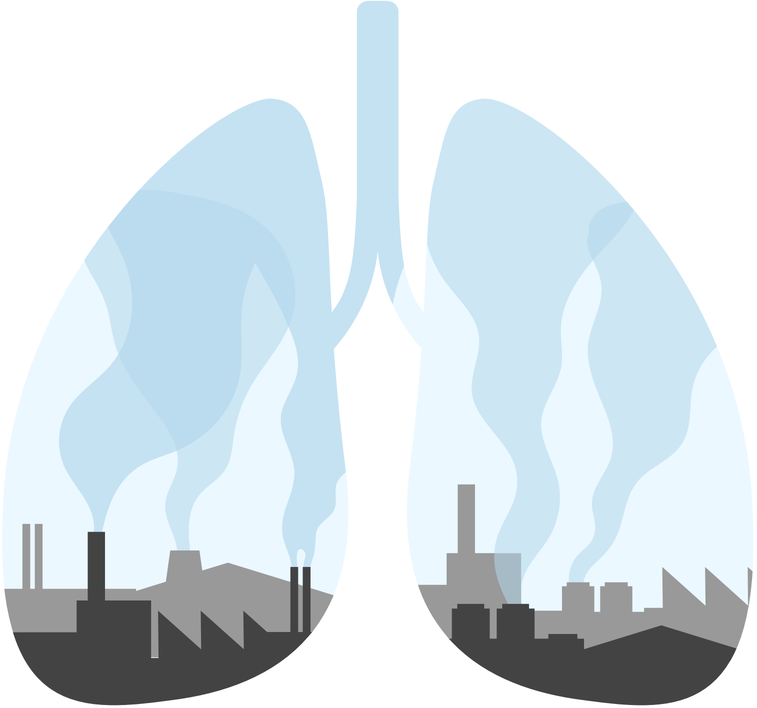
WHAT IS IT?

AQI is a standard indicator for daily monitoring of how clean or polluted the air is.

HOW IS IT USED?

Provided extensive data on the number of days each county records within each of these AQI categories.

Air Quality Index (AQI)	Levels of Health Concern
0-50	Good
51-100	Modest
101-150	Unhealthy For Sensitive Groups
151-200	Risky
201-300	Harmful
301-500	Dangerous



Key Air Quality Indicators

PM

Particulate Matter
Triggers inflammatory responses, including asthma attacks

NO2

Nitrogen Dioxide
Asthmatic conditions and reduced lung function

CO

Carbon Monoxide
Interferes with the blood's ability to carry oxygen

SO2

Sulfur Dioxide
Worsen respiratory conditions and asthmatic symptoms

O3

Ozone
Oxidative stress, airway inflammation, and lung failure

Guiding Research Questions

Main Question

What are the most significant environmental contaminants contribution to asthma incidence?

Sub-Question

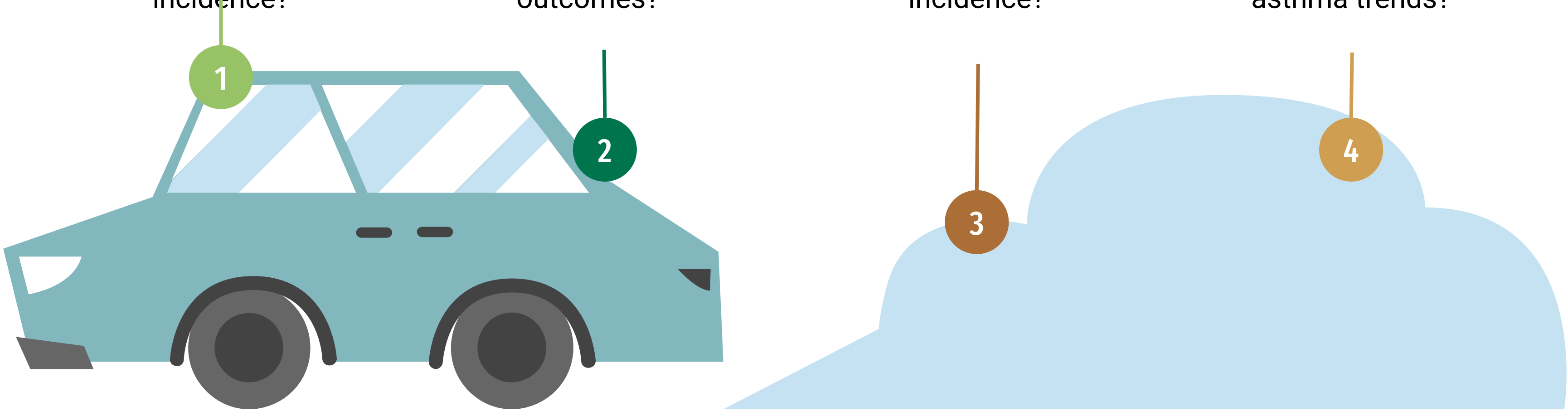
How do various environmental contaminants correlate differently with asthma incidence and health outcomes?

Sub-Question

In what ways can machine learning algorithms utilize air quality and health data to predict asthma incidence?

Sub-Question

Which machine learning algorithms demonstrate the highest accuracy and reliability in forecasting asthma trends?



Methods Used

PEARSON'S CORRELATION

1

Utilized to explore the relationship between air quality indicators and asthma-related health outcomes.

RANDOM FOREST

2

Utilized to predict health outcomes based on air quality indicators. Focusing on a single county, we prepared our dataset and split it into training and test sets for validation.

SARIMAX

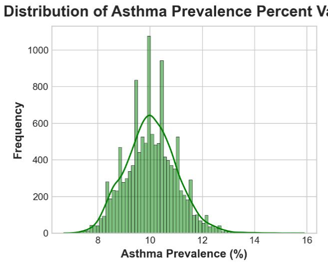
3

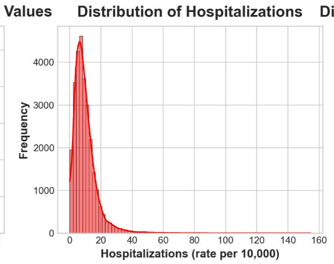
An extension of ARIMA, for time series forecasting, considering both seasonality and external values.

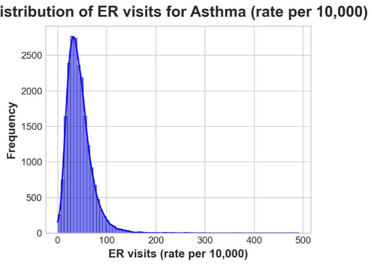
Exploratory Data Analysis

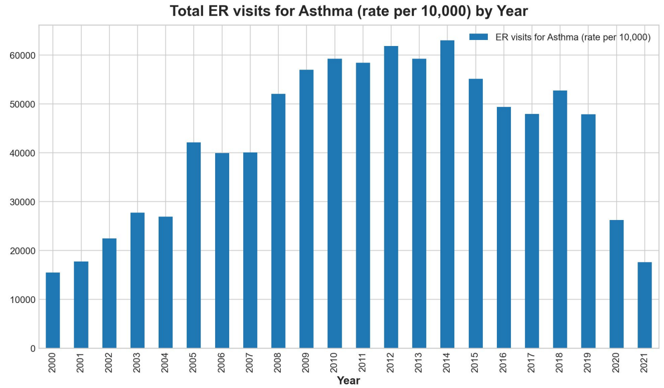
Datasets

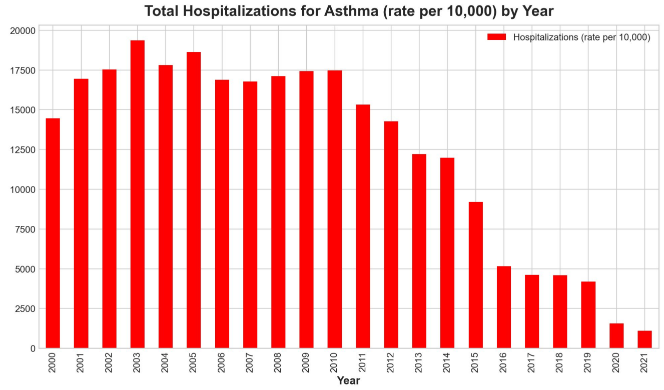
1. **ER Visits**
   - Dropped years outside of 2005-2019
2. **Hospitalizations**
   - Dropped years outside of 2005-2019
3. **Prevalence**
   - Dropped due to insufficient and incomplete data











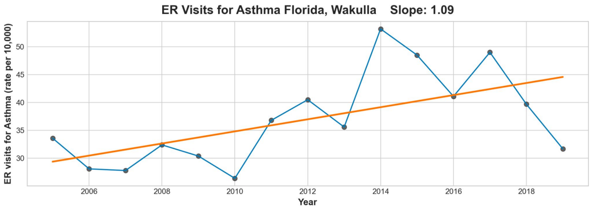
Data Processing

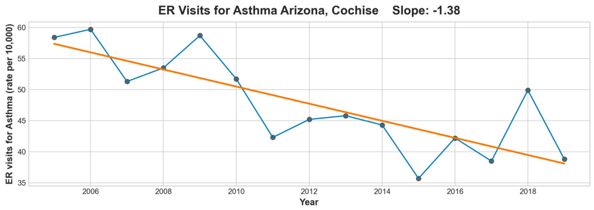
GROUPING

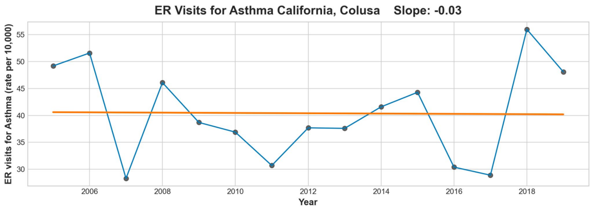
Group A - Increasing Slope >0.1

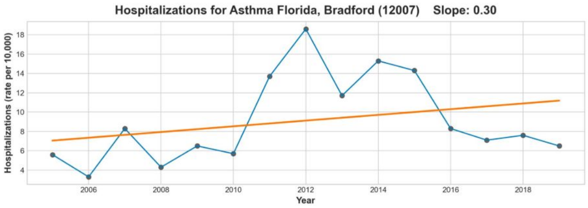
Group B - Decreasing Slope < -0.1

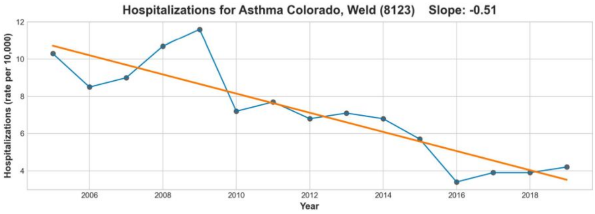
Group C - Neutral Slope -0.1<x<0.1

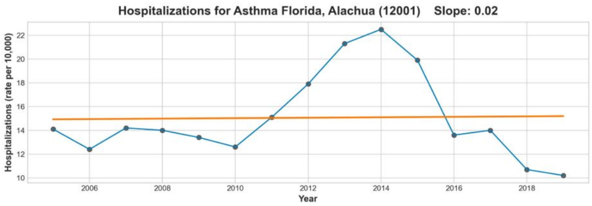










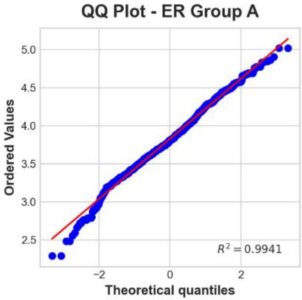


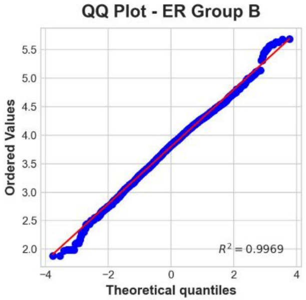
- Emergency Room visits group A consisted of 113 counties, group B contained 582 counties, and group C had 30.
- Hospitalizations group A contained data for 3 counties, group B had 671 counties, and group C contained 15 counties.
  - Due to the low number of counties in groups A and C for hospitalizations, these groups were dropped.

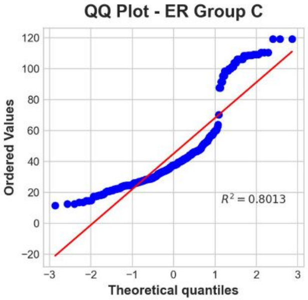
Data Processing

NORMALIZING

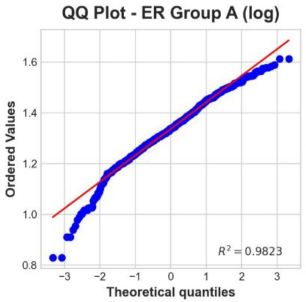
No Transformation

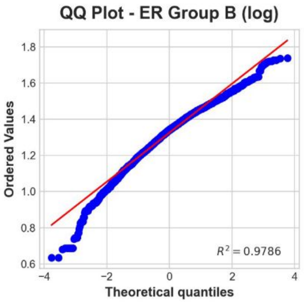


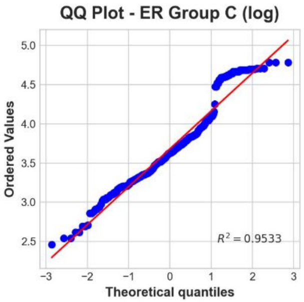


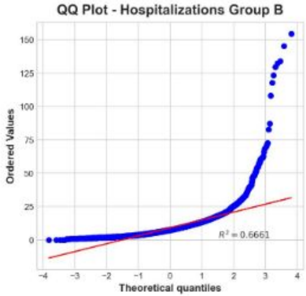


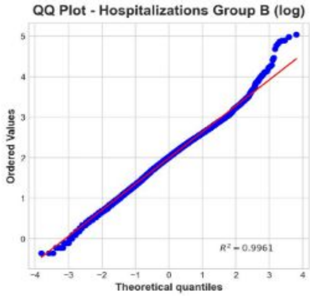
Log Transformation



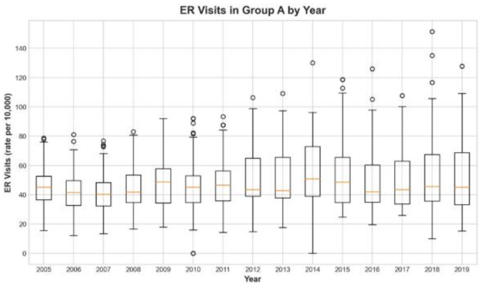


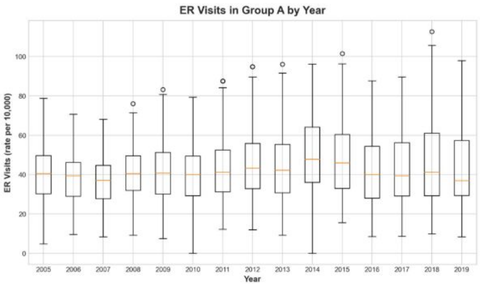


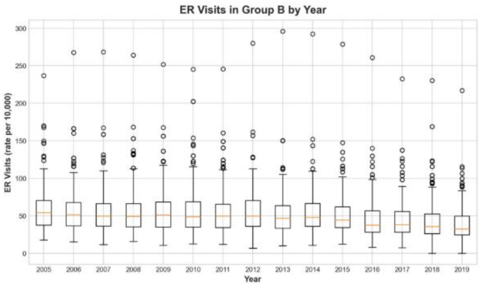


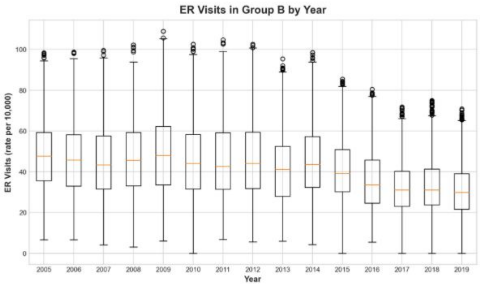


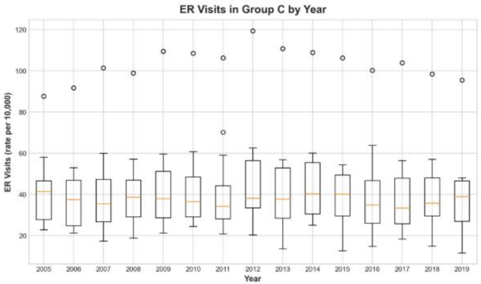
Outlier Removal

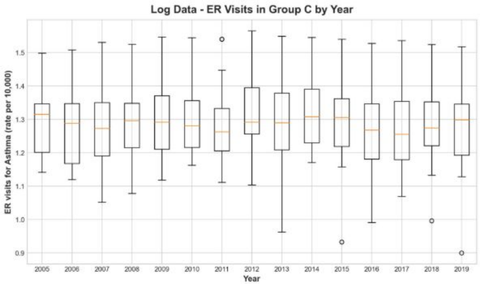


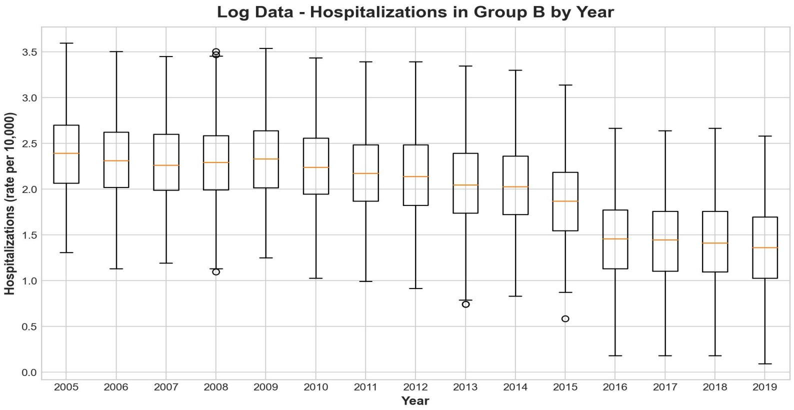


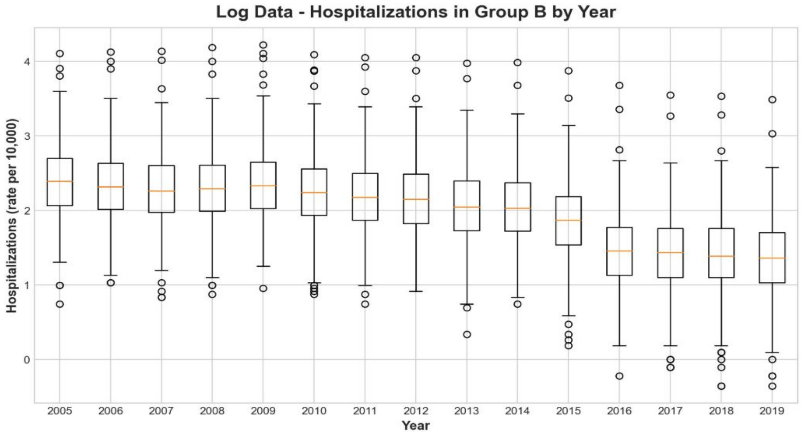












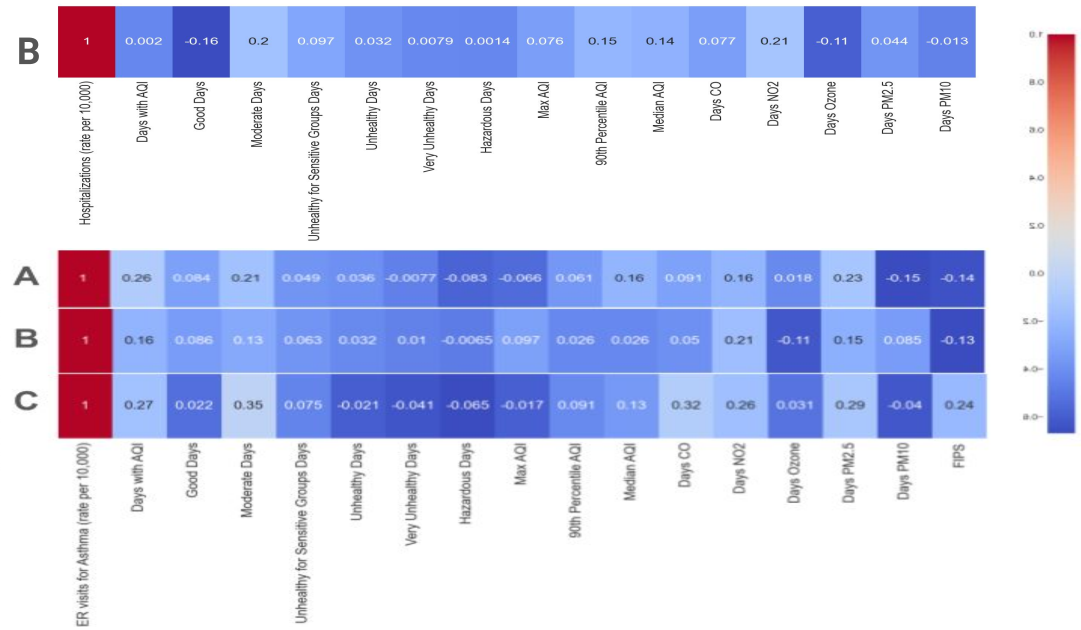
Feature Selection

Hospitalizations Group B:
Days NO2 and Moderate Days

ER Visits Group A:
Days AQI and PM2.5

ER Visits Group B:
Days AQI and NO2

ER Visits Group C:
Moderate Days and CO

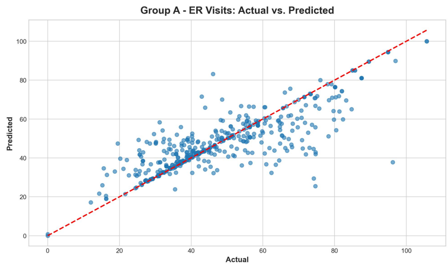


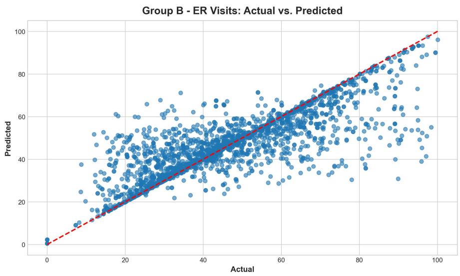
Random Forest

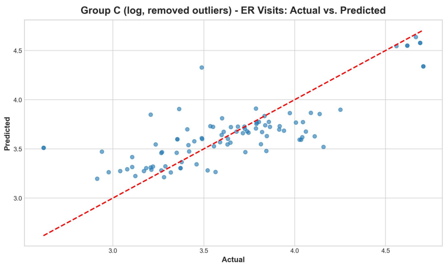
1. Test train split for each group:
   - Test size = 0.3
   - Random state = 42

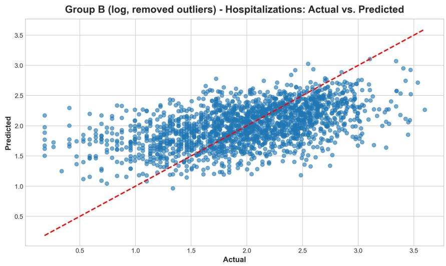
2. Random Forest Regressor
   - Features: all air quality metrics
   - Target: ER visits and hospitalizations

Random Forest Summary			
Group	R^2 Score	MSE	RMSE
ER Visits Group A	0.65	105.84	10.29
ER Visits Group B	0.68	119.77	10.94
ER Visits Group C	0.63	0.08	0.27
Hospitalizations Group B	0.25	0.26	0.51









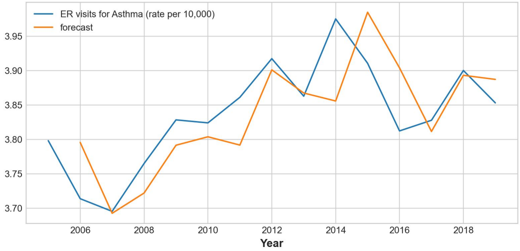
SARIMAX

Autoregressive Integrated Moving Average with Exogenous Regressors

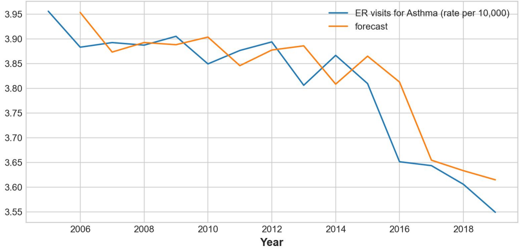
- Grouped by mean for each year
- Model order: (1,1,1) Seasonality Parameters: (1, 0, 0, 2).
- Exogenous variable: PM2.5

	ER Visits (A)	ER Visits (B)	ER Visits (C)	Hosp (B)
Ljung-Box (L1) (Q)	0.02	0.56	0.07	0.10
Prob(Q)	0.90	0.45	0.79	0.75
Heteroskedasticity (H)	0.22	4.01	3.06	10.64
Prob(H) (two-sided)	0.22	0.15	0.25	0.02
Jarque-Bera (JB)	0.60	1.01	0.78	3.24
Prob(JB)	0.74	0.60	0.68	0.20
Skew	0.44	-0.65	-0.02	-1.11
Kurtosis	2.51	3.11	1.84	3.81

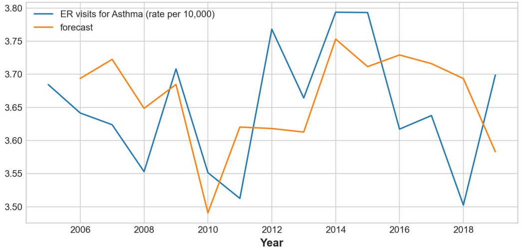
ER (A)



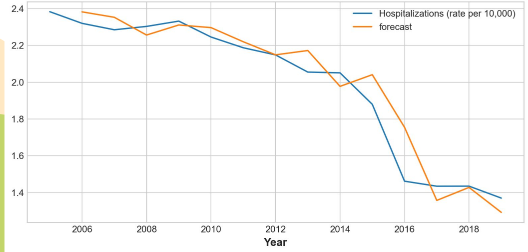
ER (B)



ER (C)



Hosp (B)



Implications & Challenges

Broader Air Quality Assessment

Our study highlights the importance of considering multiple air quality (AQI) for a more comprehensive understanding of asthma triggers

Targeted Interventions

By pinpointing key pollutants and AQI factors, healthcare professionals can develop better strategies to manage asthma and reduce healthcare burden

Incomplete Data

Missing values and inconsistent geographical information limited the generalizability of our findings and required data grouping strategies

Need For More Detailed Data

Integrating data specific to county demographics could provide deeper insights into the environmental factors affecting asthma outcomes

Discussion & Conclusions

The study confirms a link between air pollutants (PM2.5, NO2) and median AQI with asthma cases and hospitalizations.

While machine learning models offer promise for analyzing specific pollutants, their effectiveness in this study highlights the need for further refinement of theses models for broader public health

The study lays the groundwork for utilizing machine learning as a tool to inform public health strategies and develop community-specific solutions to address asthma

Mitigating air pollution is crucial, but interventions should also address determinants of healthcare accessibility to achieve a more substantial reduction in asthma related injuries

References

1. Andrew C. Leon. “3.12 - Descriptive and Inferential Statistics.” In: *Comprehensive Clinical Psychology*. Oxford: Pergamon, 1998, pp. 243-285. ISBN: 978-0-042707-2.
2. Ioannis Mansialidis et al. “Environmental and Health Impacts of Air Pollution: A Review”. In: 8.14 (Feb. 2020). DOI: 10.3389/fpubh.2020.00014.
3. Manuel Méndez, Mercedes G. Merayo, and Manuel Núñez. “Machine learning algorithms to forecast air quality: a survey”. en. In: *Artificial Intelligence Review* 56.9 (Sept. 2023), pp. 10031-10066. ISSN: 1573-7462. DOI: 10.1007/s10462-023-10424-4.
4. *State of the Air.* 2023. URL: https://www.lung.org/research/sota/key-findings#:~:text=The%20%E2%80%9CState%20of%20the%20Air,of%20ozone%20or%20particle%20pollution.