

# Air Quality in the USA

A Group One Production  
Joe Serigano | Adriana Machado | Prathik Chukkapalli



# Table of Contents

01

## Problem Statement

Air quality and health

ProPublica  
EPA bill

Monitor placement

02

## Data

Satellite feature selection  
& PM 2.5 Target

Gathering,  
extraction,  
combining

03

## Modeling & Predictions

Model selection

Test results

EPA and prediction  
comparisons

04

## Conclusion & Recommendations

Monitor Placement

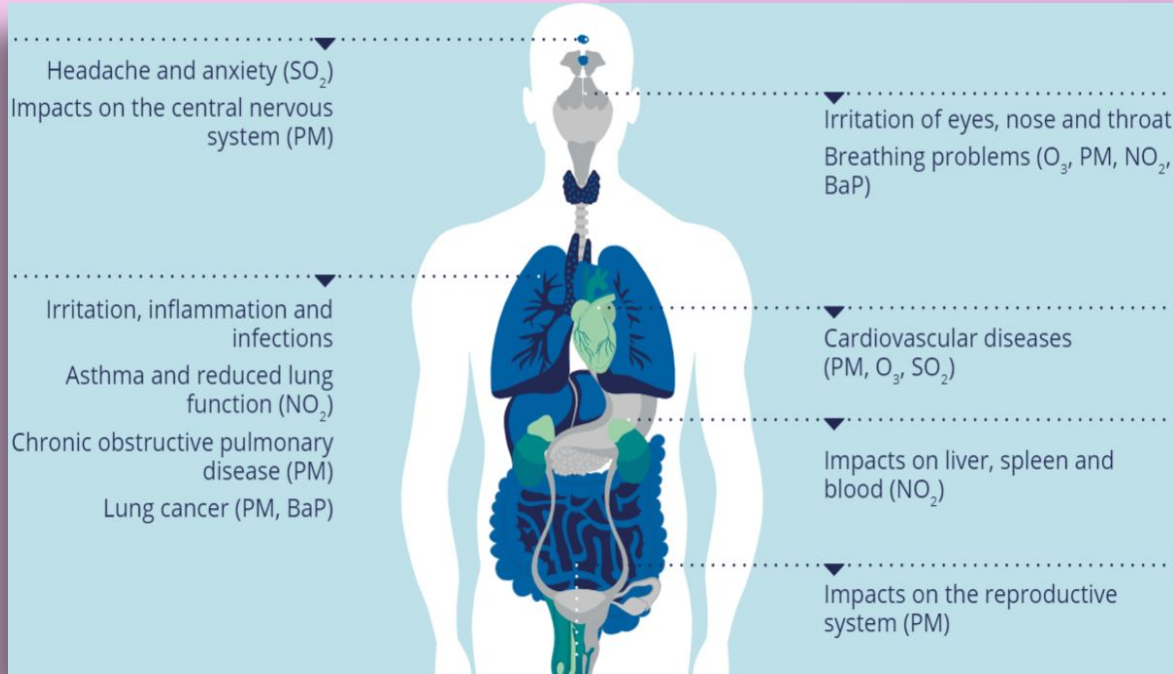
Modeling

Clean Data

# Background & Problem Statement



# Air Quality and Health

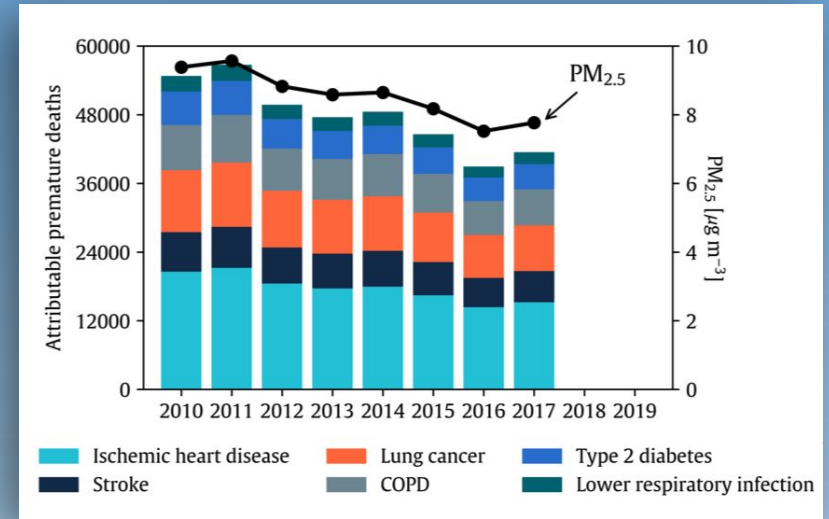
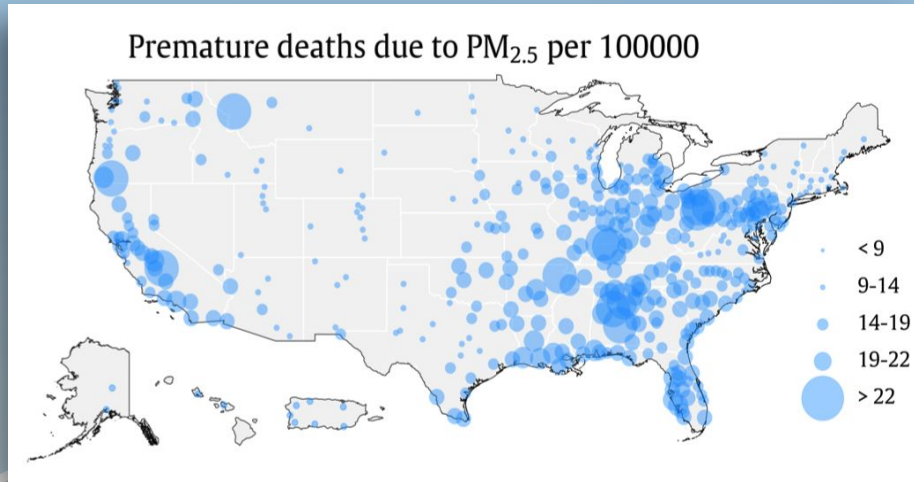


Fine particulate matter (PM<sub>2.5</sub>) is closely related to premature deaths caused by:

- ☁ Ischemic Heart Disease
- ☁ Stroke
- ☁ Type-2 Diabetes
- ☁ COPD
- ☁ Lower Respiratory Infection
- ☁ Lung Cancer

Source: European Environmental agency

- PM2.5 was associated with ~40,000 premature deaths in US in 2019 (~\$500 billion USD)
- ★ ● PM2.5 is responsible for ~8 million premature deaths annually



- ★ ● PM2.5 can be emitted directly from a source (construction, fires) or as a result of chemical reactions in the air due to every day pollutants (traffic, power plants, etc.)



1000+

hot spots of toxic industrial air  
pollution

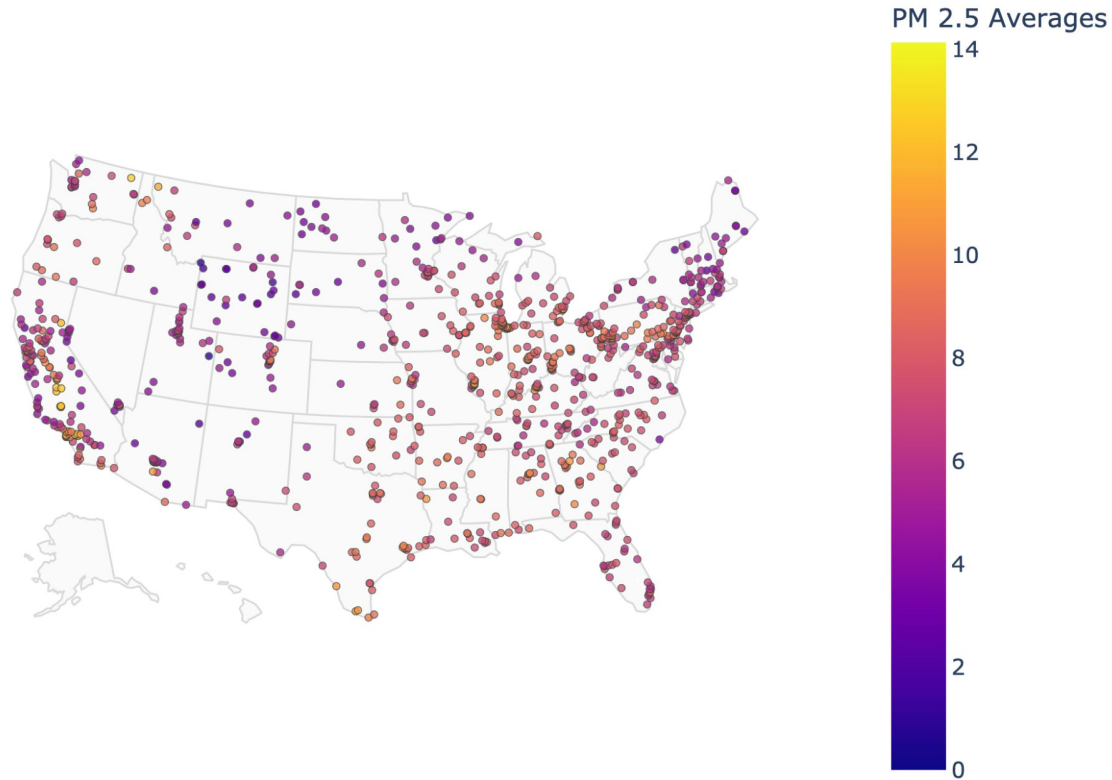
256,000

people live in high cancer risk  
areas

**Representatives Introduce  
\$500 Million Air Quality Bill,  
Citing ProPublica's  
Investigations**

Lawmakers introduced a House bill to fund air monitoring after ProPublica highlighted pollution in its "Black Snow" and "Sacrifice Zones" investigations. The bill is nearly identical to one introduced in the Senate last summer.

## 2019 Annual Contiguous USA EPA PM 2.5 Values



# Measurement Problems



CONTINUOUS MONITORING



LOCAL MONITORING



CLEAN DATA



TYPES OF PARTICULATES

The background of the slide features a solid light purple color. In the top right corner, there are two stylized clouds: a smaller yellow one and a larger blue one. In the bottom left corner, there are two more stylized clouds: a purple one and an orange one. The title 'Problem Statement' is centered in the upper half of the slide.

# Problem Statement

Air quality is a major contributor to overall health. Many underserved communities suffer the most from poor air quality in the USA. The EPA's current system for monitoring PM 2.5 levels is inconsistent and needs standardization, specifically in hyper-local monitor placement and consistency in monitoring.



# Satellite Data

(Features in our model)



**GATHER**



**EXTRACT**



**COMBINE**



**EDA**

# Feature Selection For PM 2.5 Predictions

## NO2

TROPOMI - Sentinel  
0.1° resolution

## AOD

(Aerosol Optical Depth)

MODIS-Aqua  
0.1° resolution

## Feature

Satellite  
Lat/Long resolution

## Precipitation

Global Precipitation Monitor  
0.1° resolution

## Temperature

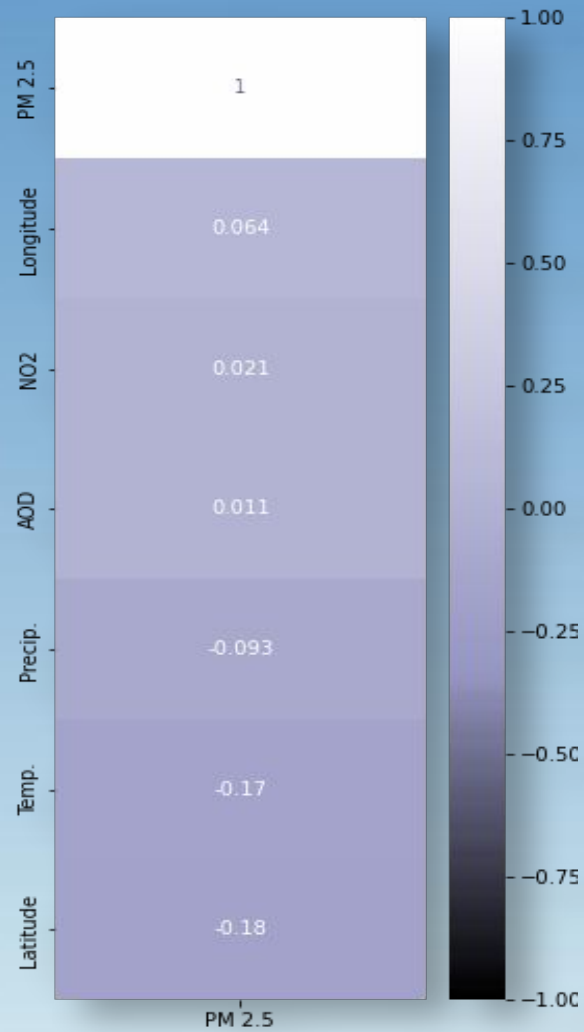
MODIS-Aqua  
0.05° resolution

MODIS = Moderate Resolution  
Imaging Spectroradiometer

TROPOMI = TROPOspheric  
Monitoring Instrument

# Extracting Variables From Multiple Datasets

- Data are from 2019, the last full year before COVID influenced global air quality.
- Temperature and precipitation data pulled from NASA GIOVANNI
- AOD data from *High-Resolution Gridded Level 3 Aerosol Optical Depth Data from MODIS* (Gupta et al. 2020) due to higher resolution.
- NO2 data from *COVID-19 pandemic reveals persistent disparities in nitrogen dioxide pollution* (Kerr et al., 2021) due to higher resolution.
- Individual data sets were combined by finding the nearest latitude/longitude matches to EPA monitoring sites using a Haversine function.



# Modeling & Predictions



# THREE BEST SIMPLE MODELS

**RandomForestRegressor()**

Training Score: 0.935

Test Score: 0.705

**BaggingRegressor()**

Training Score: 0.918

Test Score: 0.657

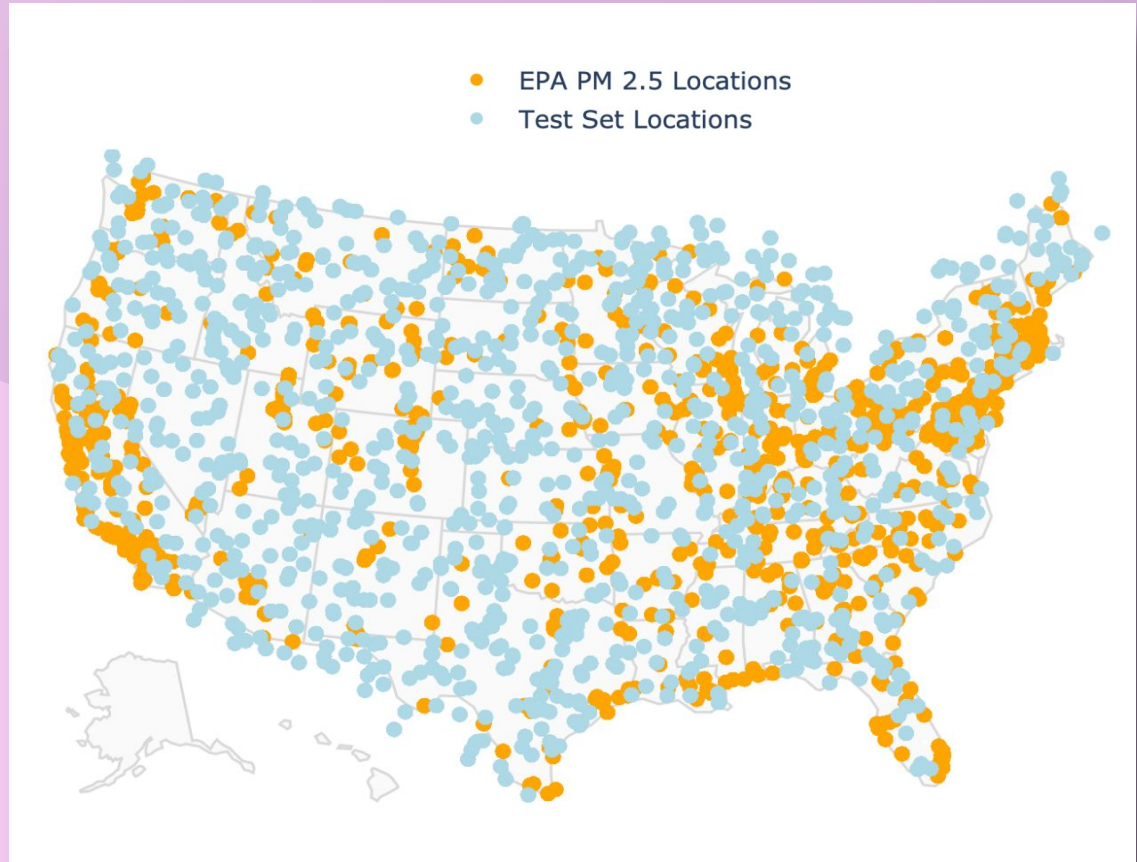
**AdaBoostRegressor()**

Training Score: 0.927

Test Score: 0.679

EPA PM 2.5 Locations  
are not evenly  
distributed.

Test Set random  
distribution across  
the US.



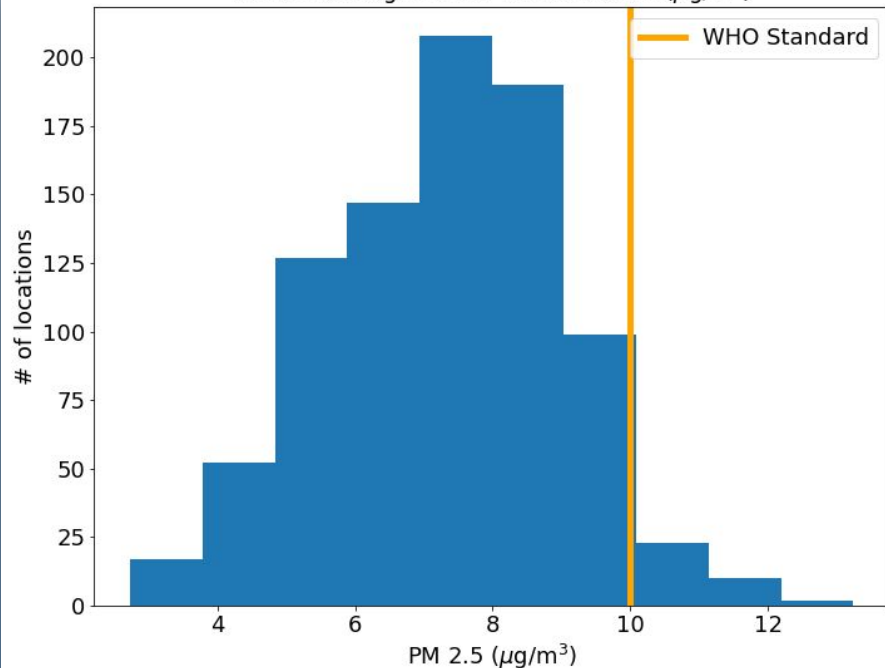
The background is a solid blue gradient. It features several stylized, light blue clouds of various shapes and sizes. Scattered throughout are small white dots representing stars, some with four-pointed starburst outlines. In the bottom right corner, there are two white curved lines suggesting a comet or a shooting star.

# Conclusion & Recommendations

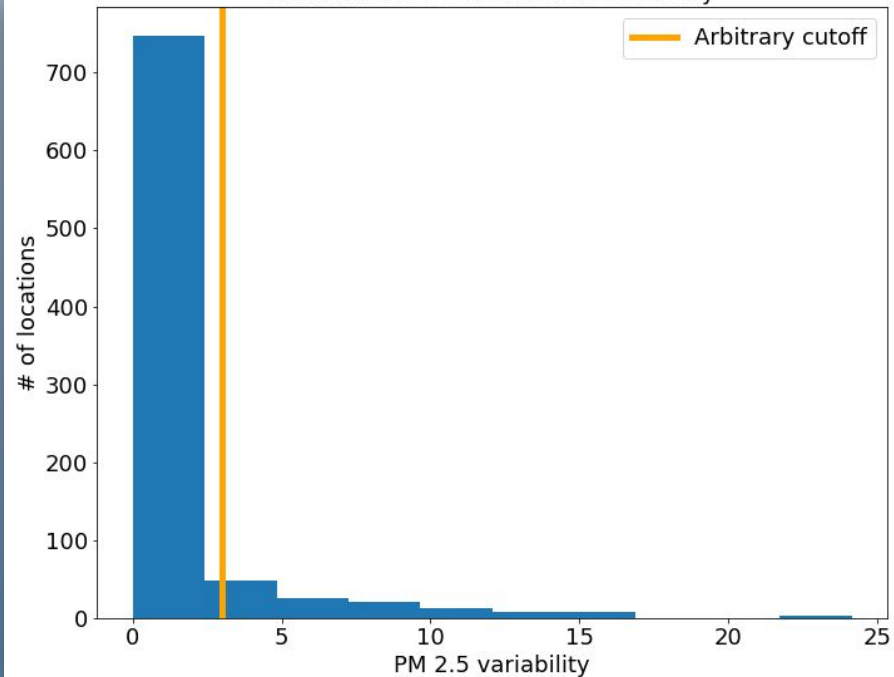


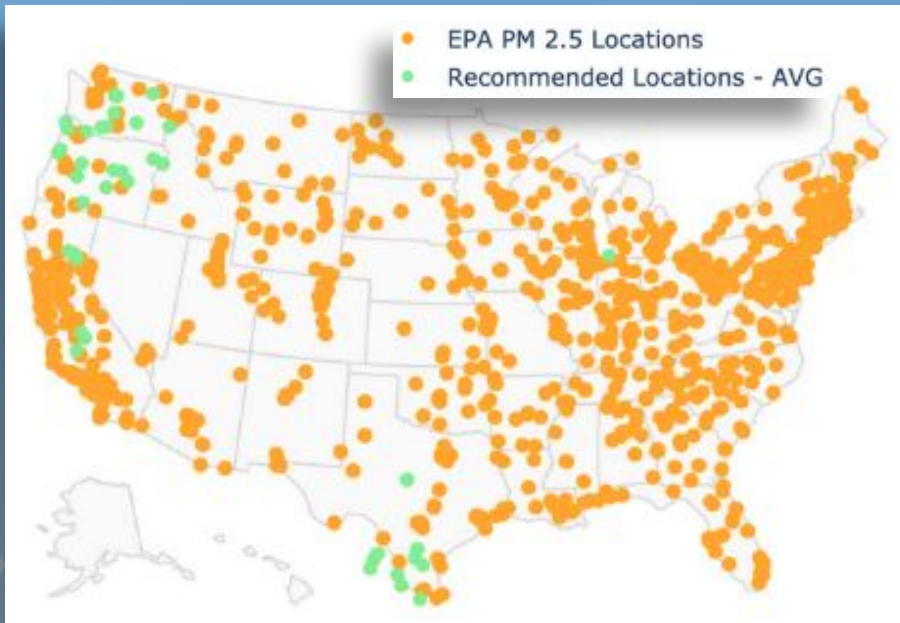
# Results & Standards

Annual average PM 2.5 concentration ( $\mu\text{g}/\text{m}^3$ )



Annual PM 2.5 concentration variability





AVERAGE

VARIABILITY





# Cleaner Data

Consistent location resolution  
between variables

Localized readings

Data across various years

# More Variables

Proximity to industrial areas

Population data

Feature relationships

THANK YOU

questions?

