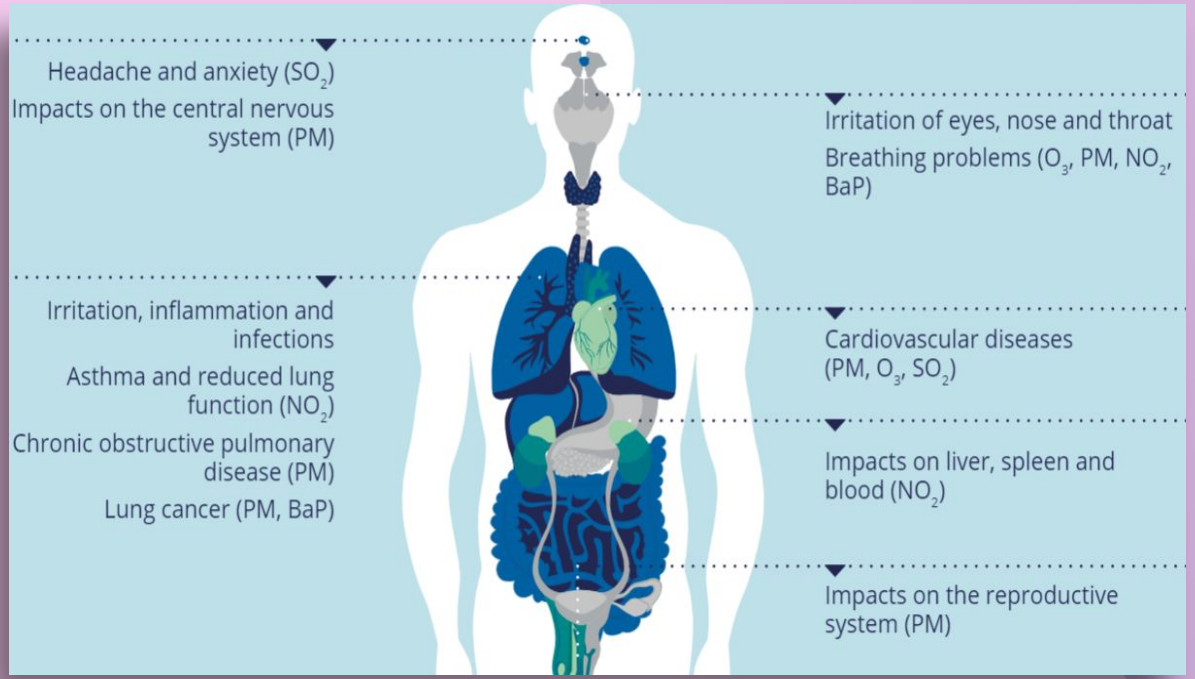
Air Quality in the USA

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

Table of Contents

Background & Problem Statement

Air Quality and Health

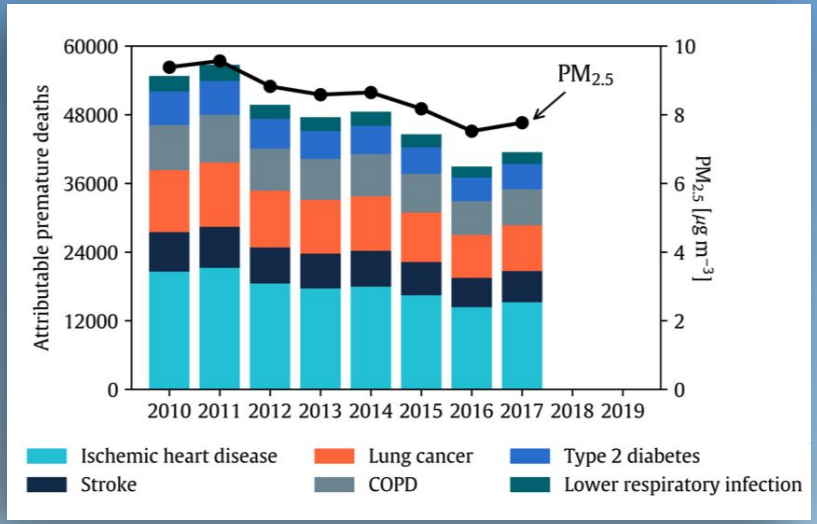


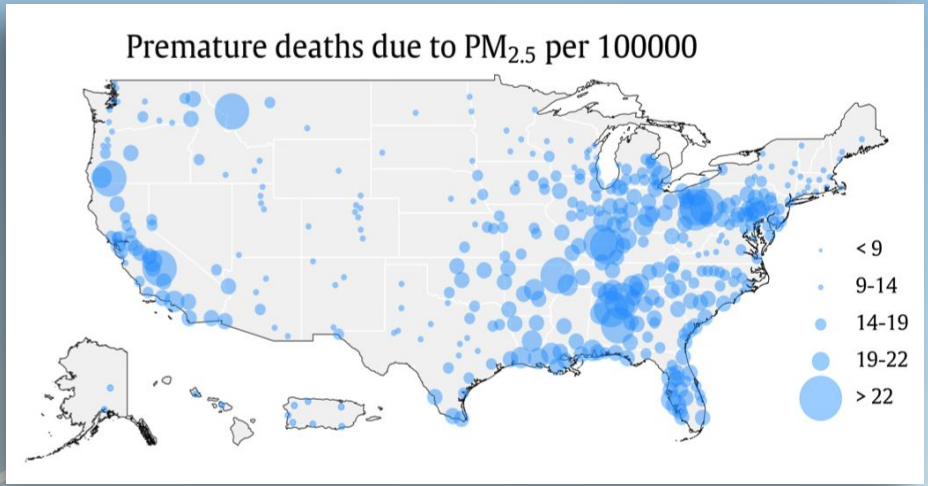
Source: European Environmental agency

Fine particulate matter (PM2.5) is closely related to premature deaths caused by:

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

- 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.)

Figures from Kerr et al., 2022

1000+

hot spots of toxic industrial air pollution

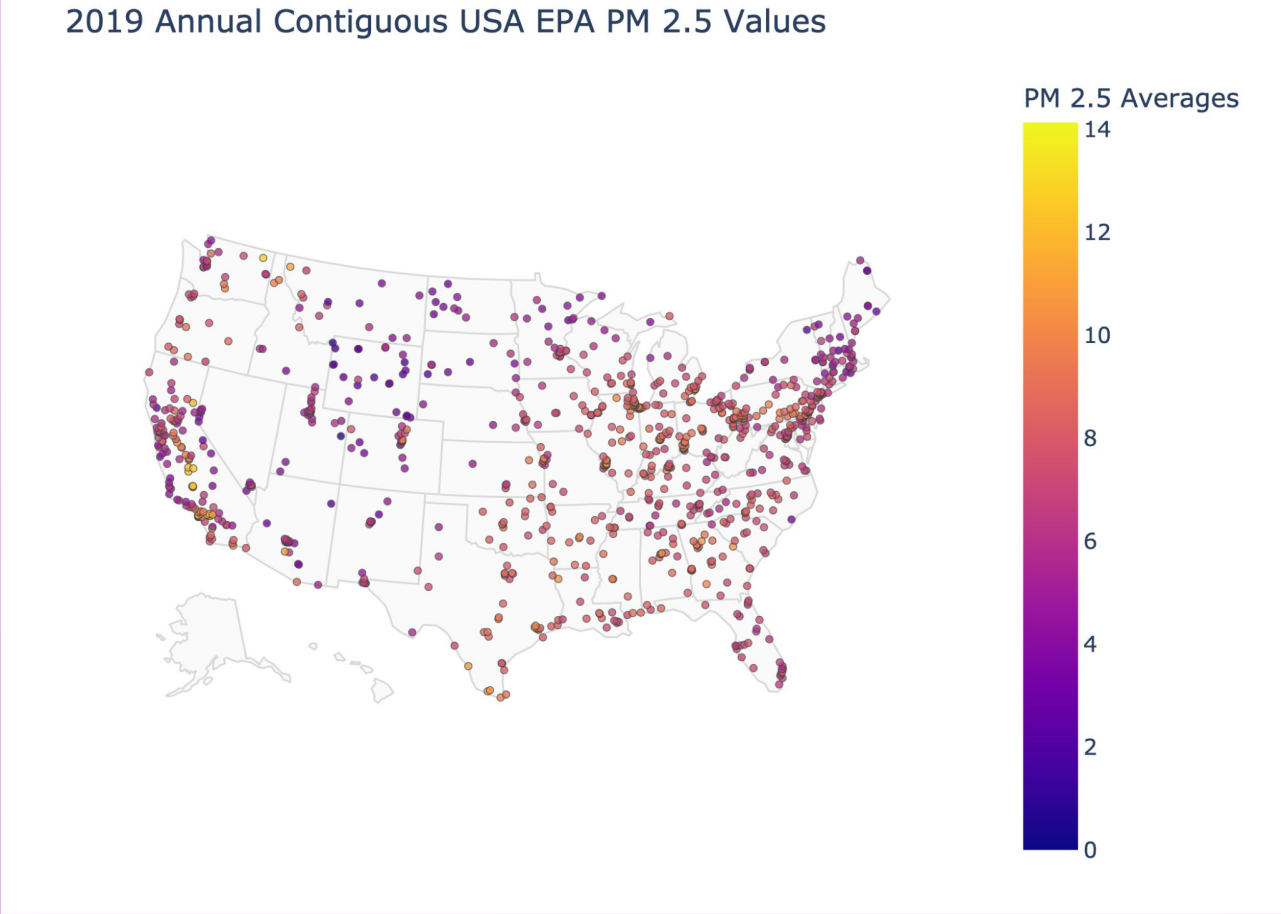
256,000

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.

Measurement Problems



- CONTINUOUS MONITORING
- LOCAL MONITORING
- CLEAN DATA
- TYPES OF PARTICULATES

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
molecules/cm^2
0.1° resolution

AOD

(Aerosol Optical Depth)

MODIS-Aqua
Dimensionless
0.1° resolution

Feature

Satellite
Units
Lat/Long resolution

Precipitation

Global Precipitation Monitor
inches/month
0.1° resolution

Temperature

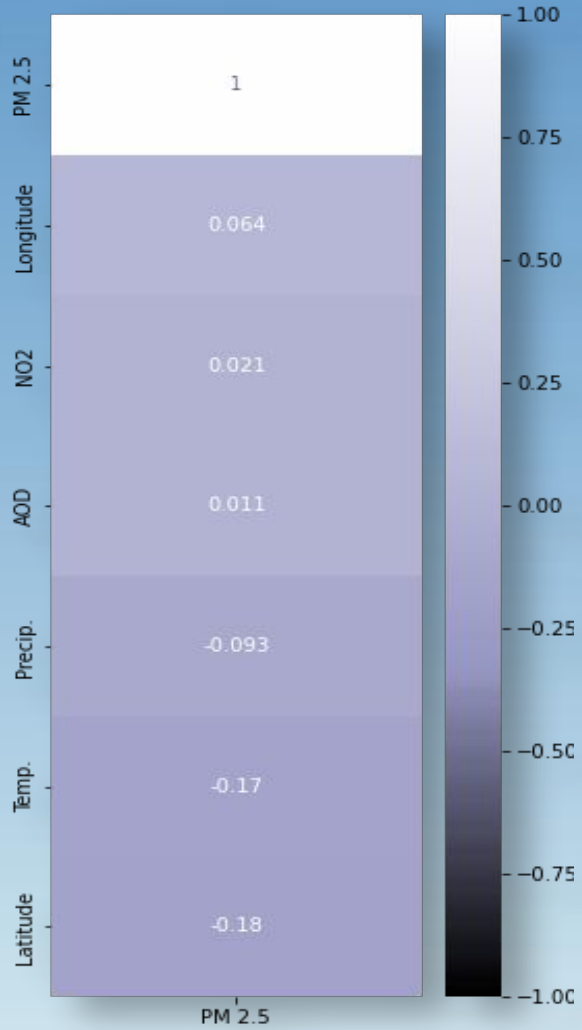
MODIS-Aqua
Kelvin
0.05° resolution

MODIS = Moderate Resolution Imaging Spectroradiometer

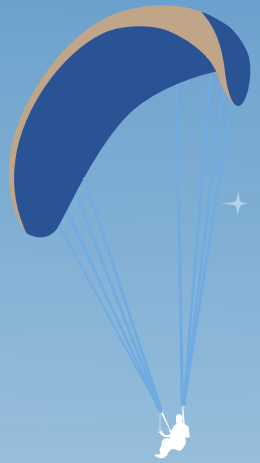
TROPOMI = TROPOspheric Monitoring Instrument

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.


PM 2.5
Longitude
NO2
AOD
Precip.
Temp.
Latitude
1
0.064
0.021
0.011
-0.093
-0.17
-0.18
PM 2.5
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Modeling & Predictions



THREE BEST SIMPLE MODELS

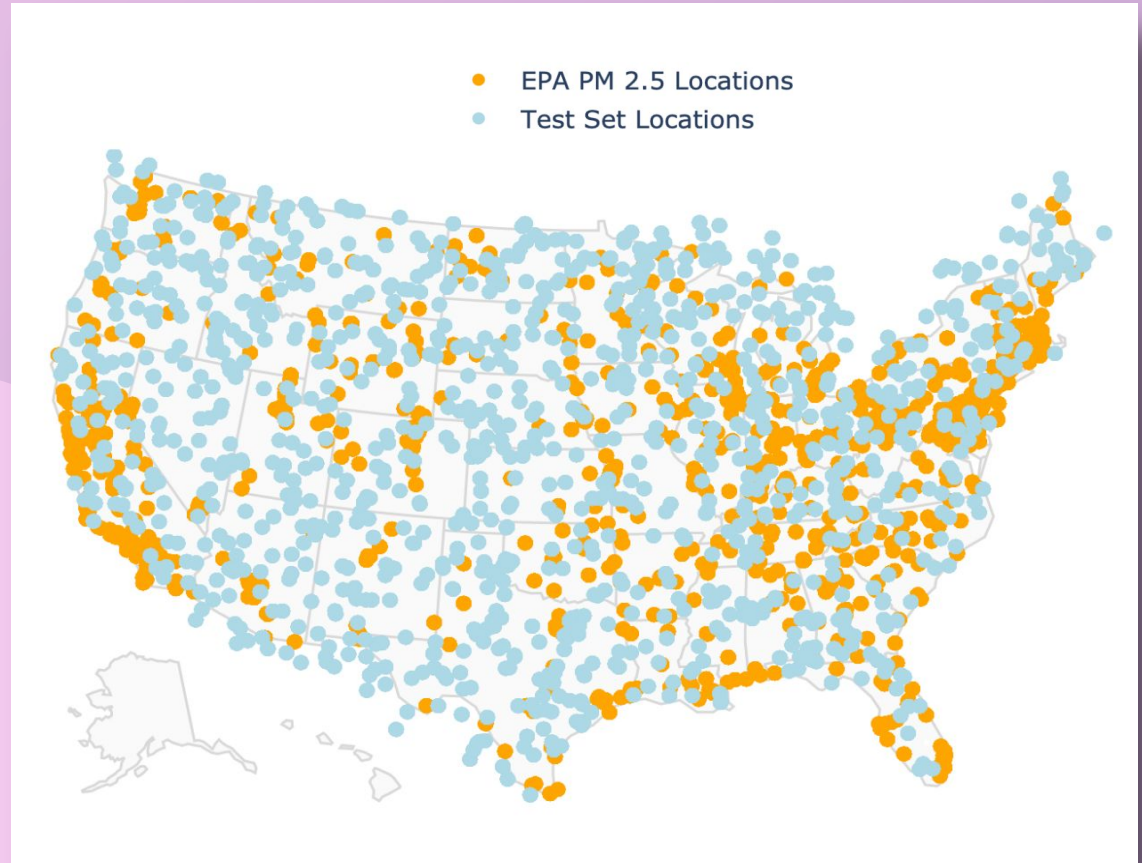
RandomForestRegressor()
Training Score: 0.935
Test Score: 0.705

BaggingRegressor()
Training Score: 0.918
Test Score: 0.657

AdaBoostRegressor()
Training Score: 0.927
Test Score: 0.679

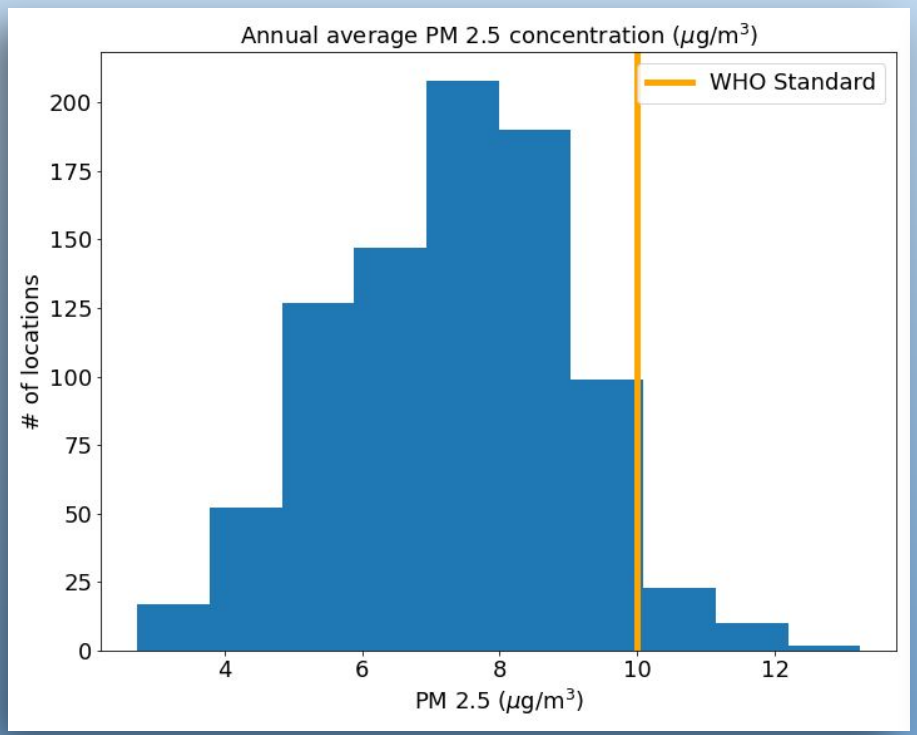
EPA PM 2.5 Locations are not evenly distributed.

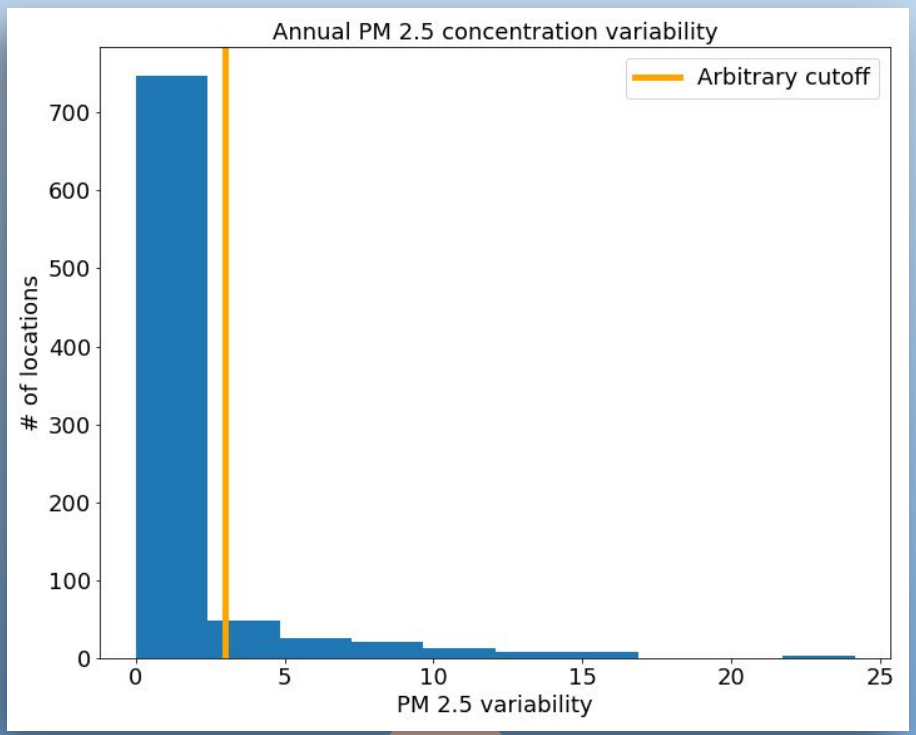
Test Set random distribution across the US.



Conclusion & Recommendations

Results & Standards





AVERAGE

VARIABILITY

Cleaner Data

Consistent location resolution between variables

Localized readings (ground data, not atmospheric)

Data across various years

More Variables

More environmental/industrial data

Population data

Feature relationships

THANK YOU

questions?