Prioritizing Pollutant Reduction to Minimize Bee Colony Loss: Analyzing EPA and Survey Data

A project by Marta Fuentes-Filp and Prab Jaswal

Background, Problem and Methodology

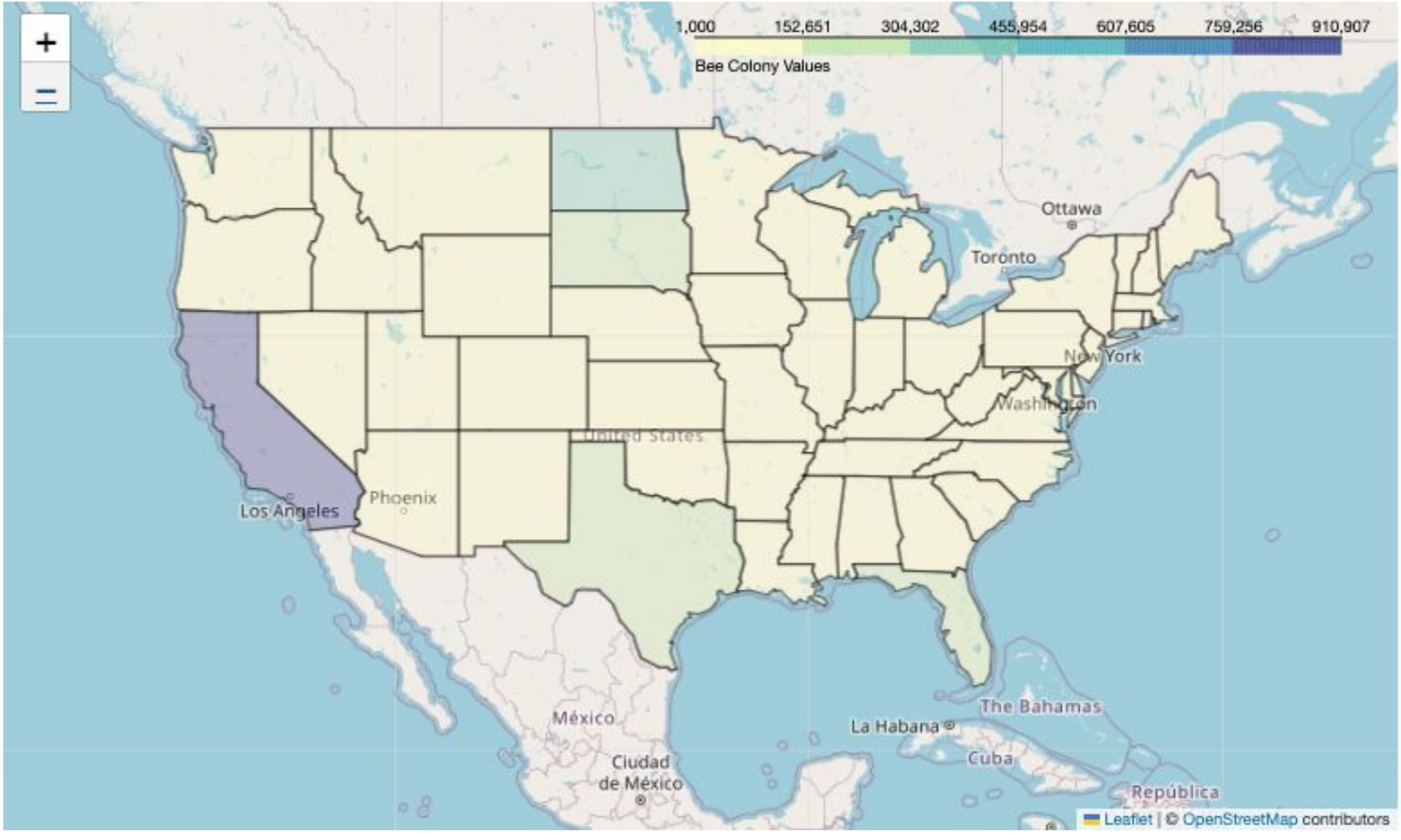
Pollutant Gases: Carbon Monoxide, Nitrogen Dioxide, Ozone, Particulate matter (PM2.5/PM10).

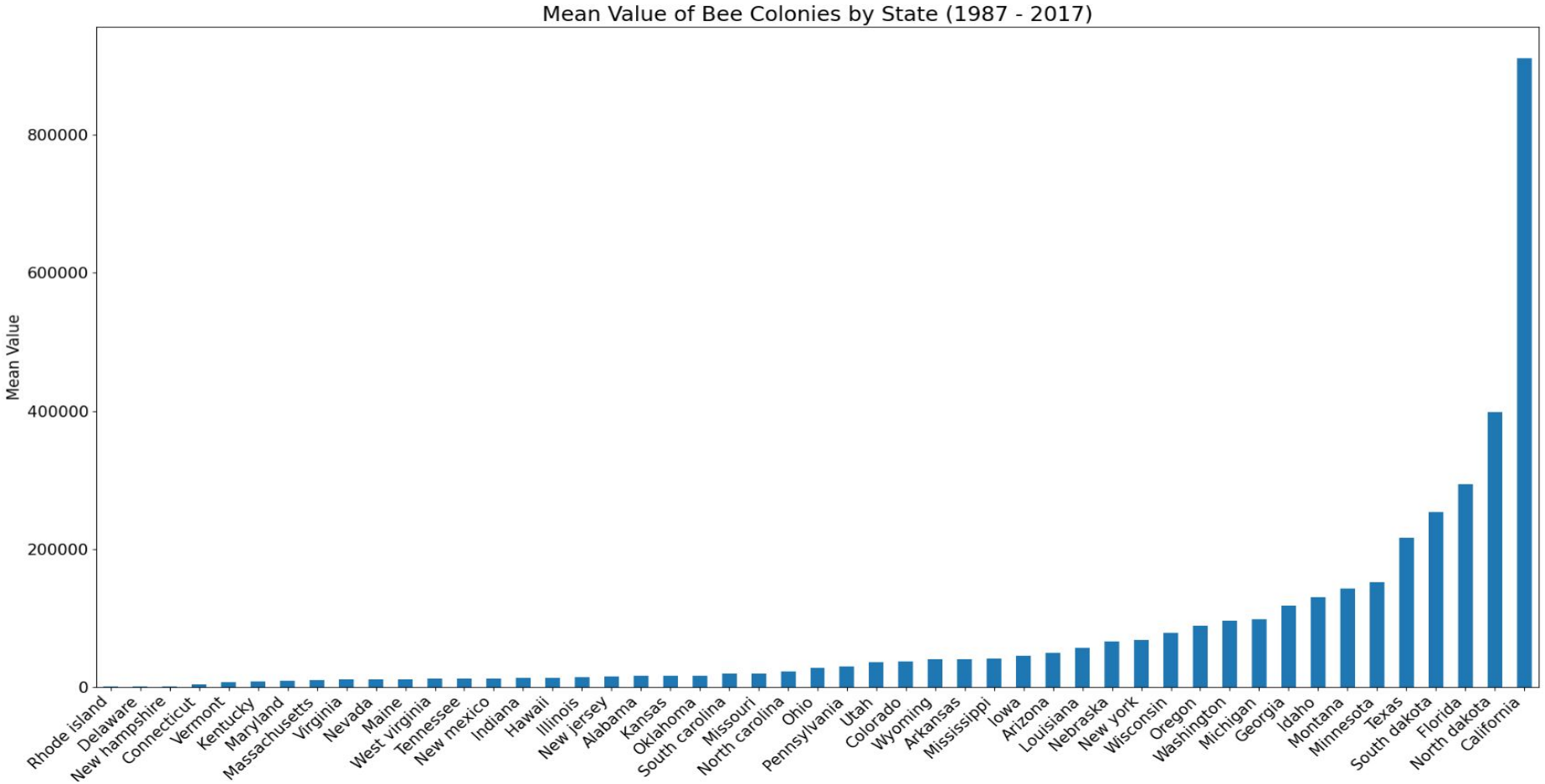
3 Aims

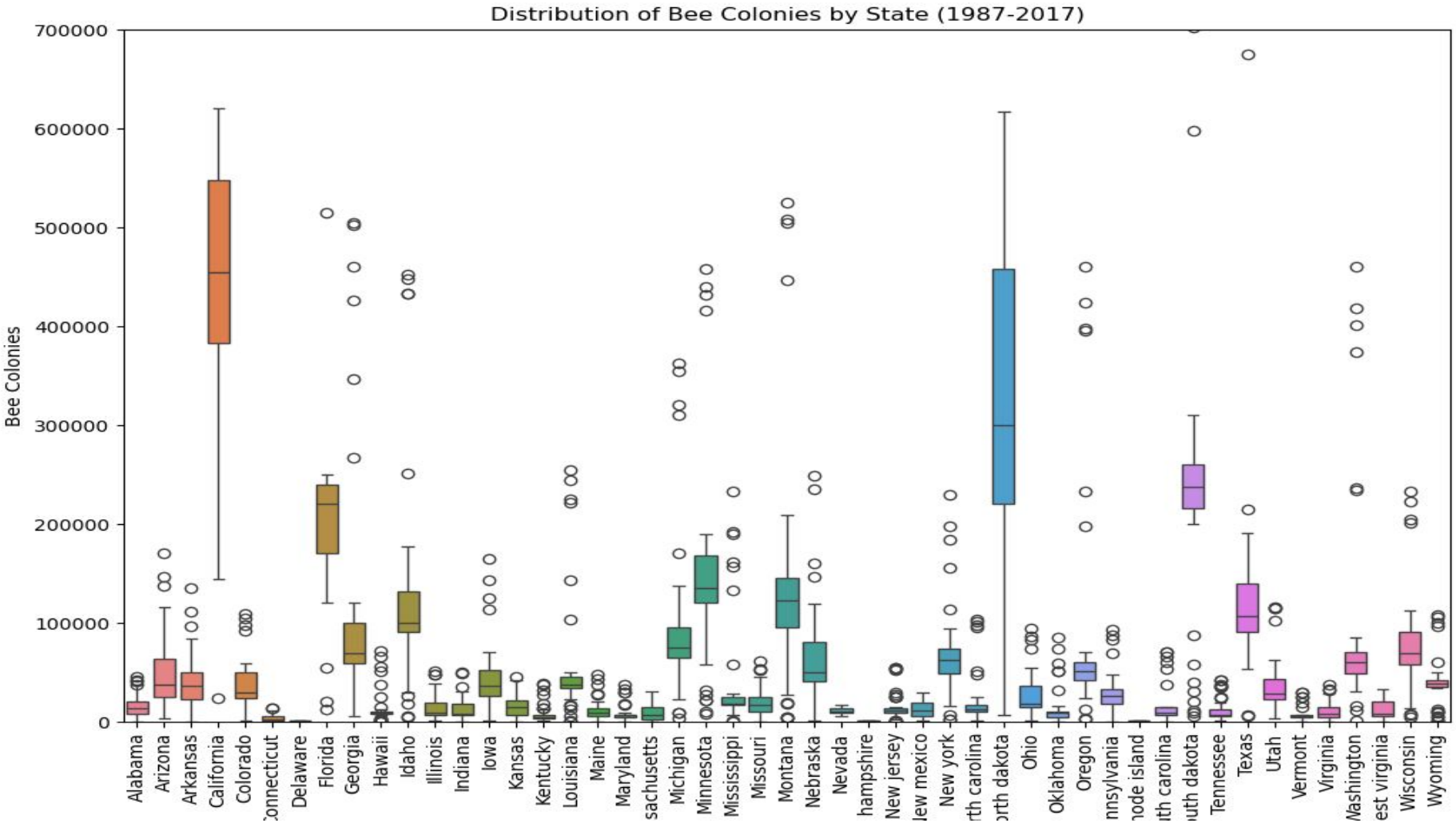
1. Investigate the trends and correlations between the presence of pollutant gases and bee populations over time for US states through exploratory data analysis and visualizations.
2. Use insights from statistical modelling to understand the extent to which each pollutant gas impacts bee colony numbers.
3. Use insights from statistical modeling to understand which pollutant gases should be prioritized for removal to maximize bee colony numbers and to provide recommendations to any interest bodies.

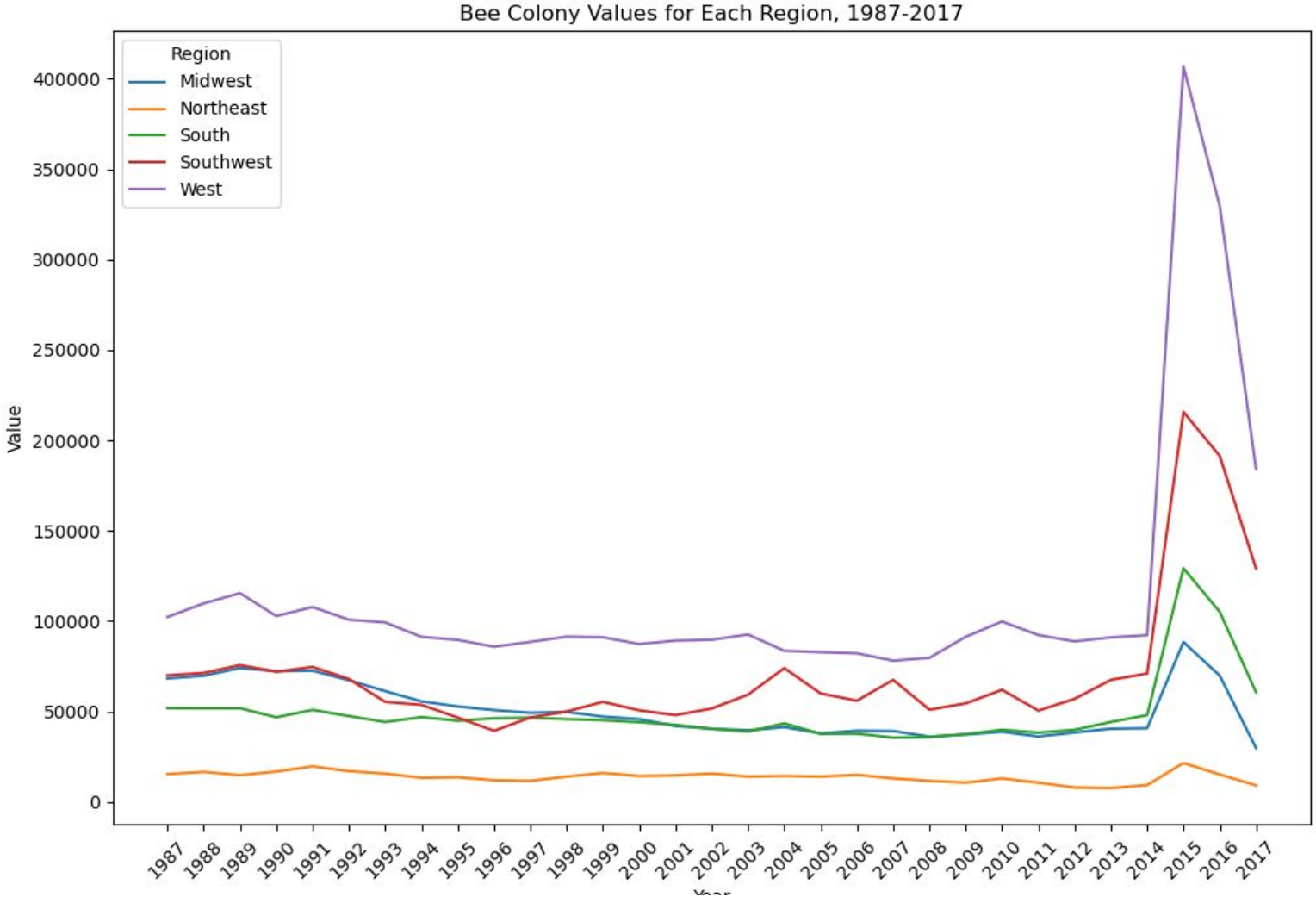
About the data

Exploring the data - Bee Colony Populations

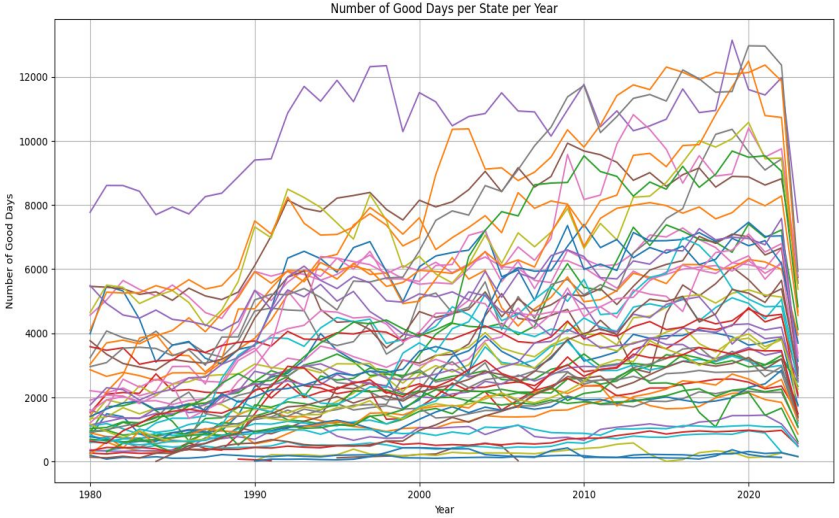

1,000
152,651
304,302
455,954
607,605
759,256
910,907
Bee Colony Values
Ottawa
Toronto
New York
Washington
United States
Phoenix
Los Angeles
México
Ciudad de México
La Habana
The Bahamas
Cuba
República
Leaflet | © OpenStreetMap contributors

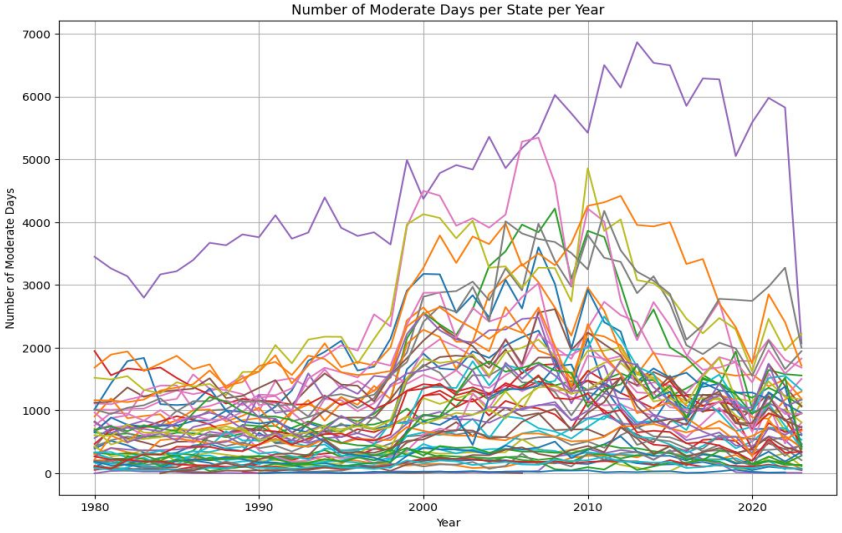

Mean Value of Bee Colonies by State (1987 - 2017)
Mean Value
0
200000
400000
600000
800000
Rhode island
Delaware
New hampshire
Connecticut
Vermont
Kentucky
Maryland
Massachusetts
Virginia
Nevada
Maine
West virginia
Tennessee
New mexico
Indiana
Hawaii
Illinois
New jersey
Alabama
Kansas
Oklahoma
South carolina
Missouri
North carolina
Ohio
Pennsylvania
Utah
Colorado
Wyoming
Arkansas
Mississippi
Iowa
Arizona
Louisiana
Nebraska
New york
Wisconsin
Oregon
Washington
Michigan
Georgia
Idaho
Montana
Minnesota
Texas
South dakota
Florida
North dakota
California

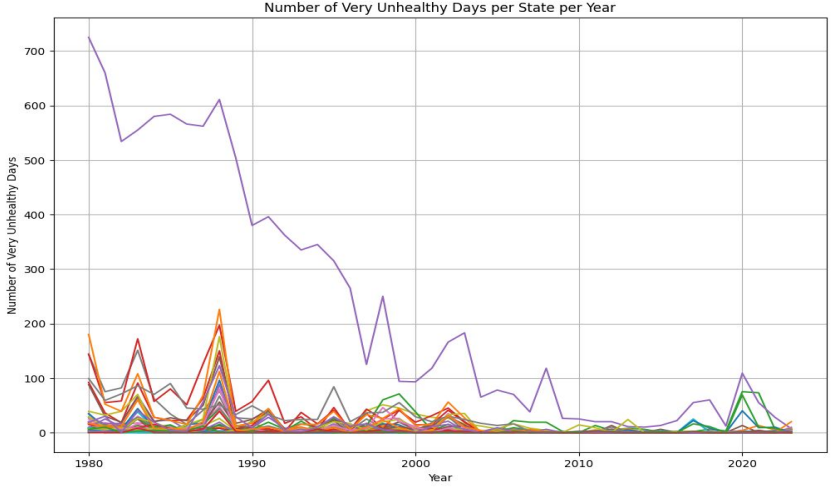

Distribution of Bee Colonies by State (1987-2017)
Bee Colonies
700000
600000
500000
400000
300000
200000
100000
0
Alabama
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
sachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
hampshire
New jersey
lew mexico
New york
rth carolina
rth dakota
Ohio
Oklahoma
Oregon
nnsylvania
node island
th carolina
uth dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
est virginia
Wisconsin
Wyoming

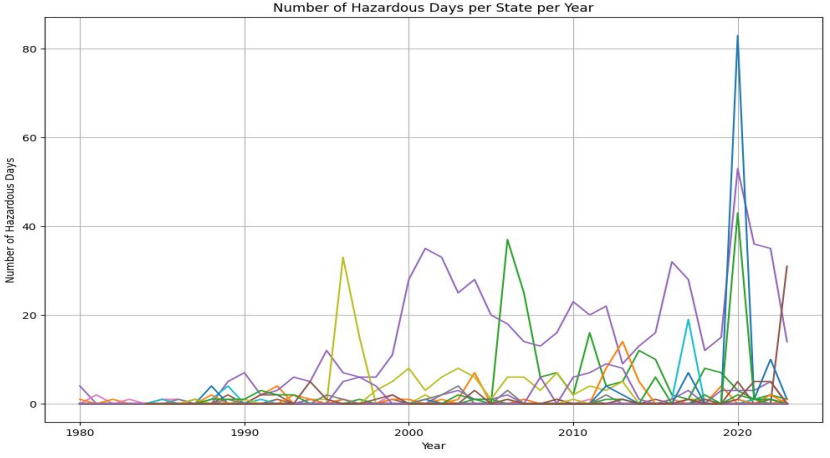

Bee Colony Values for Each Region, 1987-2017
Region
Midwest
Northeast
South
Southwest
West
Value
Year
0
50000
100000
150000
200000
250000
300000
350000
400000
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017

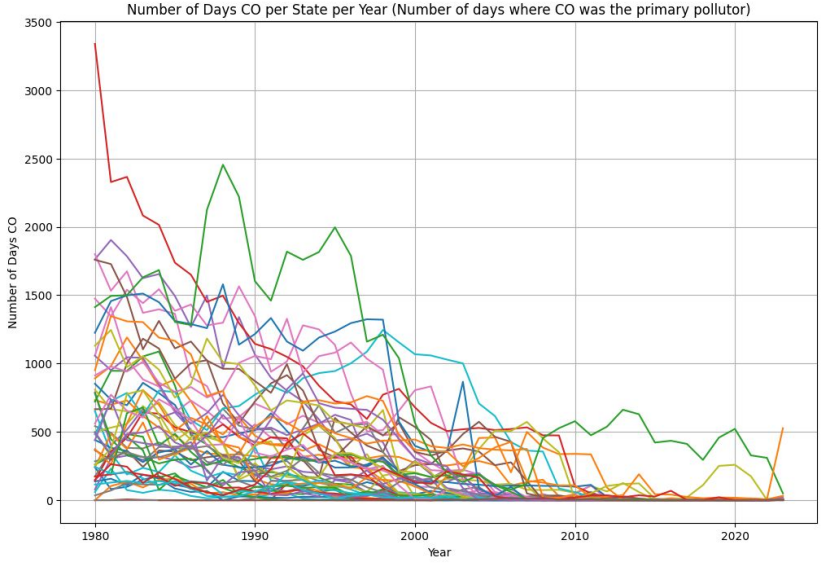
Exploring the data - Air Quality

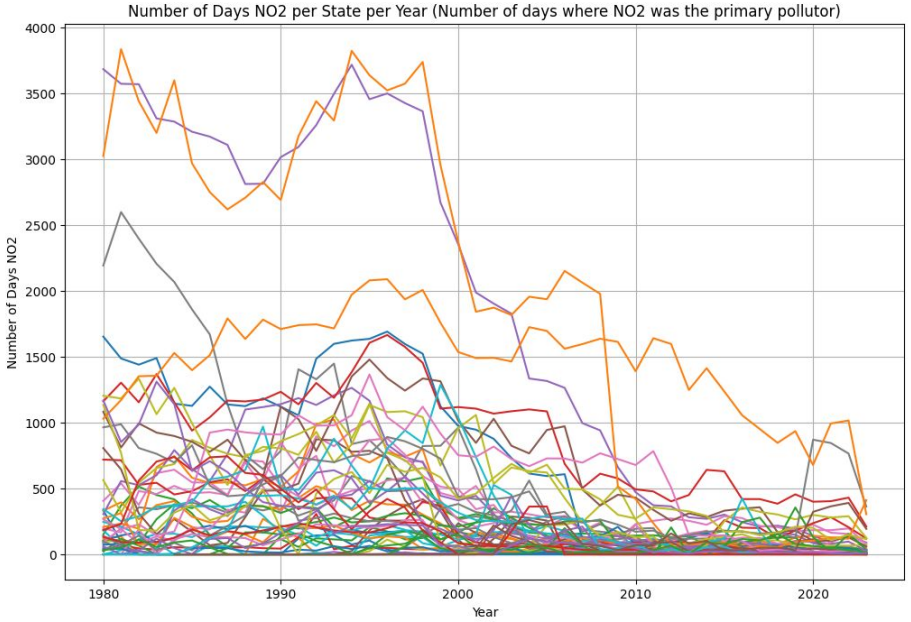

Number of Good Days per State per Year
Number of Good Days
Year
0
2000
4000
6000
8000
10000
12000
1980
1990
2000
2010
2020

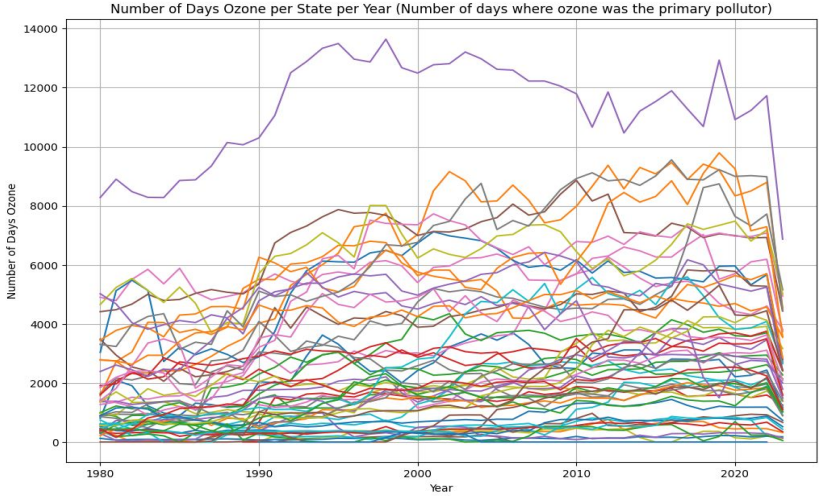

Number of Moderate Days per State per Year
Number of Moderate Days
Year
0
1000
2000
3000
4000
5000
6000
7000
1980
1990
2000
2010
2020

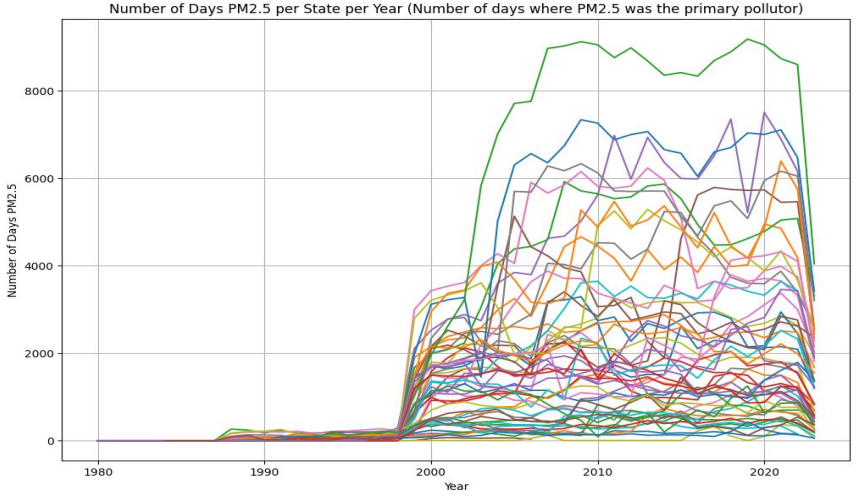

Number of Very Unhealthy Days per State per Year
Number of Very Unhealthy Days
Year
0
100
200
300
400
500
600
700
1980
1990
2000
2010
2020

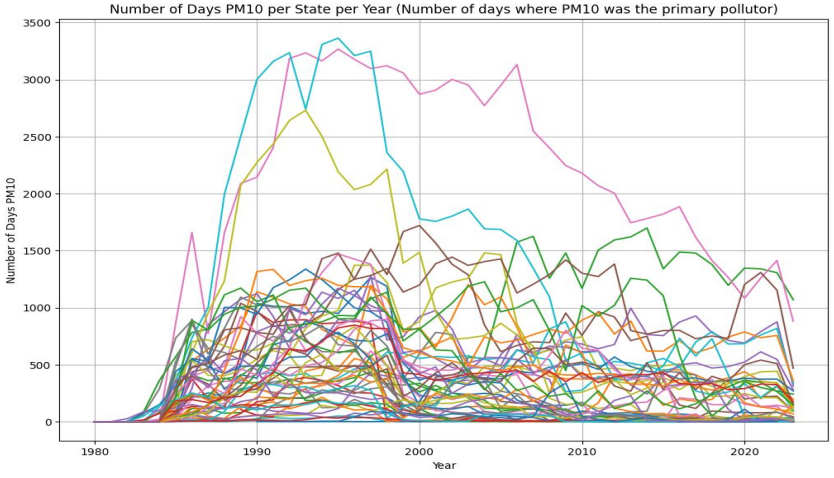

Number of Hazardous Days per State per Year
Number of Hazardous Days
Year
0
20
40
60
80
1980
1990
2000
2010
2020

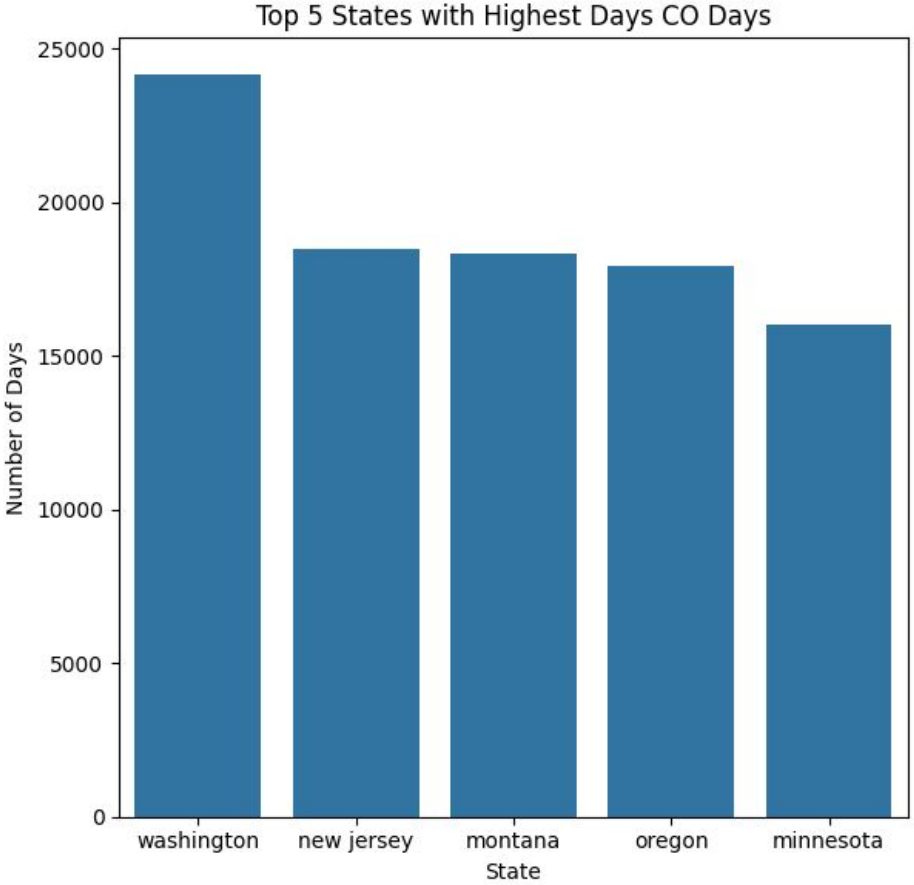

Number of Days CO per State per Year (Number of days where CO was the primary pollutor)
Number of Days CO
Year
0
500
1000
1500
2000
2500
3000
3500
1980
1990
2000
2010
2020

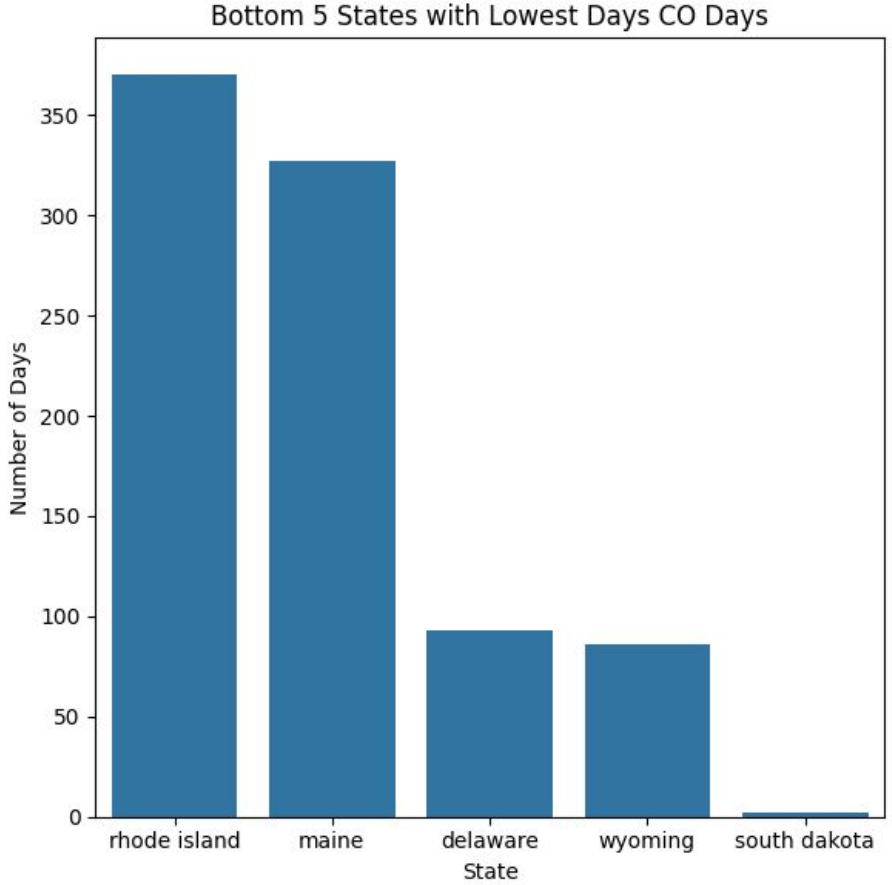

Number of Days NO2 per State per Year (Number of days where NO2 was the primary pollutor)
Number of Days NO2
Year
0
500
1000
1500
2000
2500
3000
3500
4000
1980
1990
2000
2010
2020

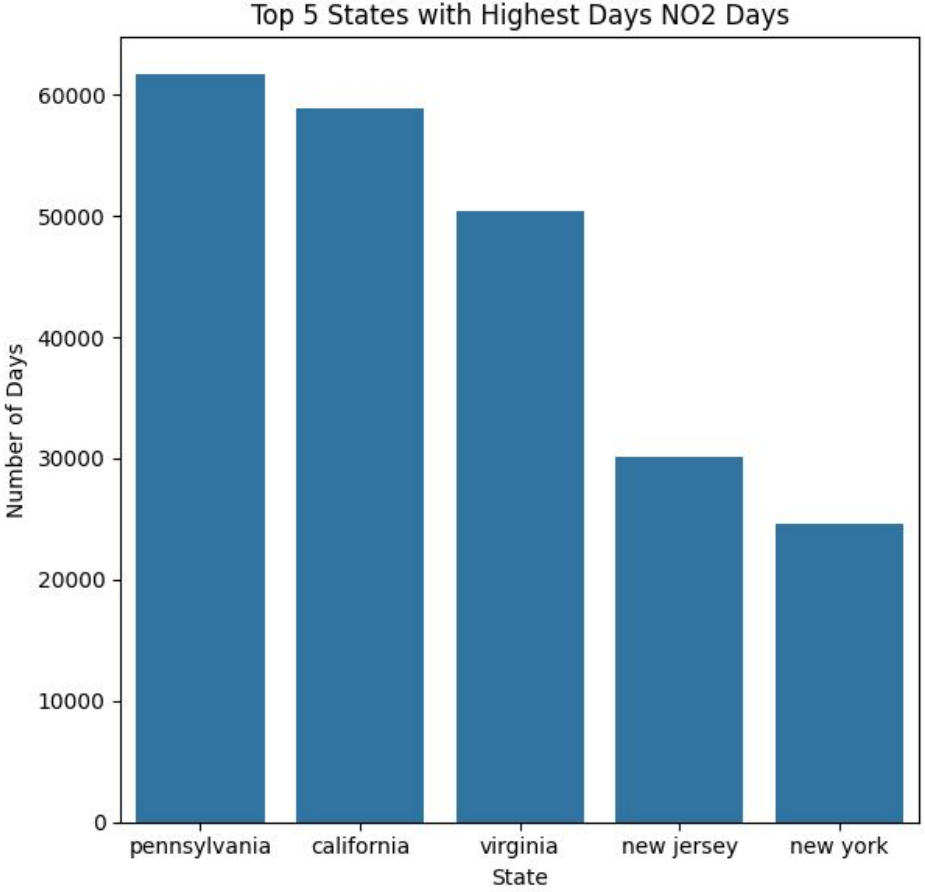

Number of Days Ozone per State per Year (Number of days where ozone was the primary pollutor)
Number of Days Ozone
Year
0
2000
4000
6000
8000
10000
12000
14000
1980
1990
2000
2010
2020

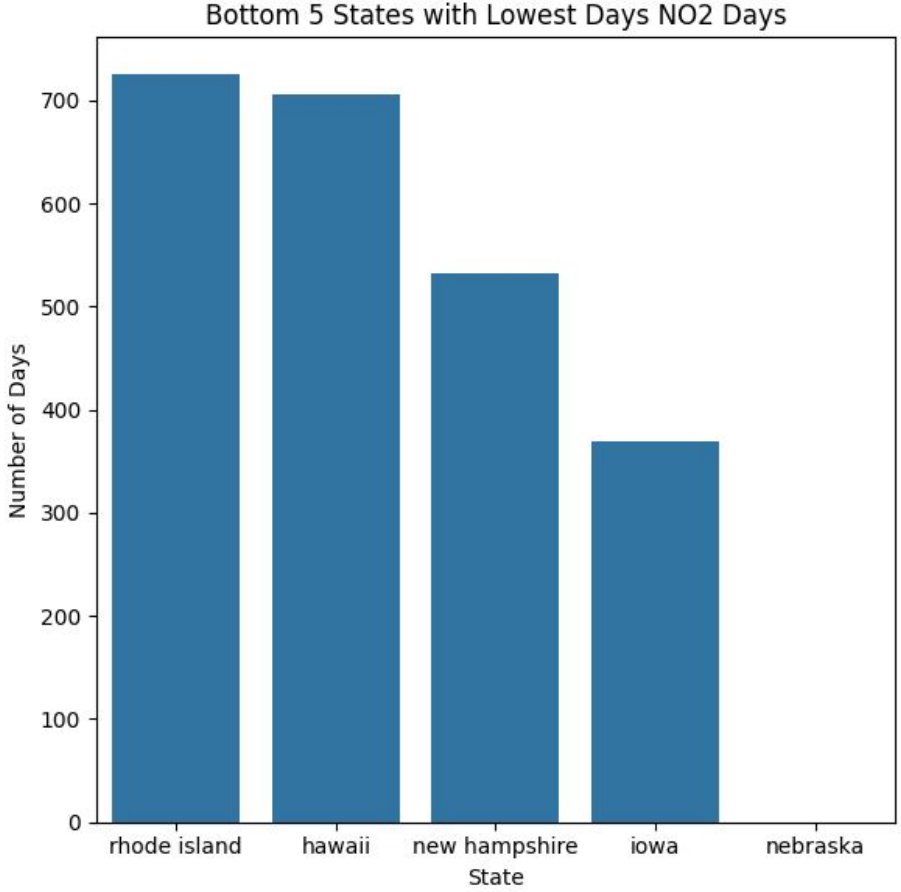

Number of Days PM2.5 per State per Year (Number of days where PM2.5 was the primary pollutor)
Number of Days PM2.5
Year
0
2000
4000
6000
8000
1980
1990
2000
2010
2020

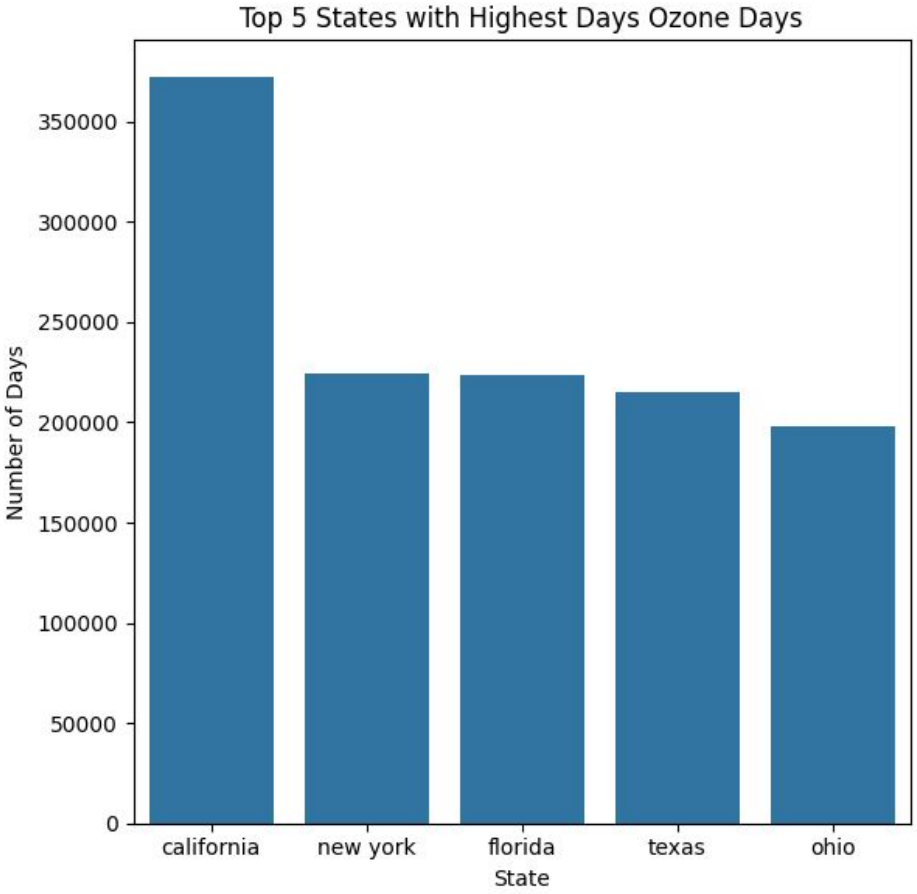

Number of Days PM10 per State per Year (Number of days where PM10 was the primary pollutor)
Number of Days PM10
Year
0
500
1000
1500
2000
2500
3000
3500
1980
1990
2000
2010
2020

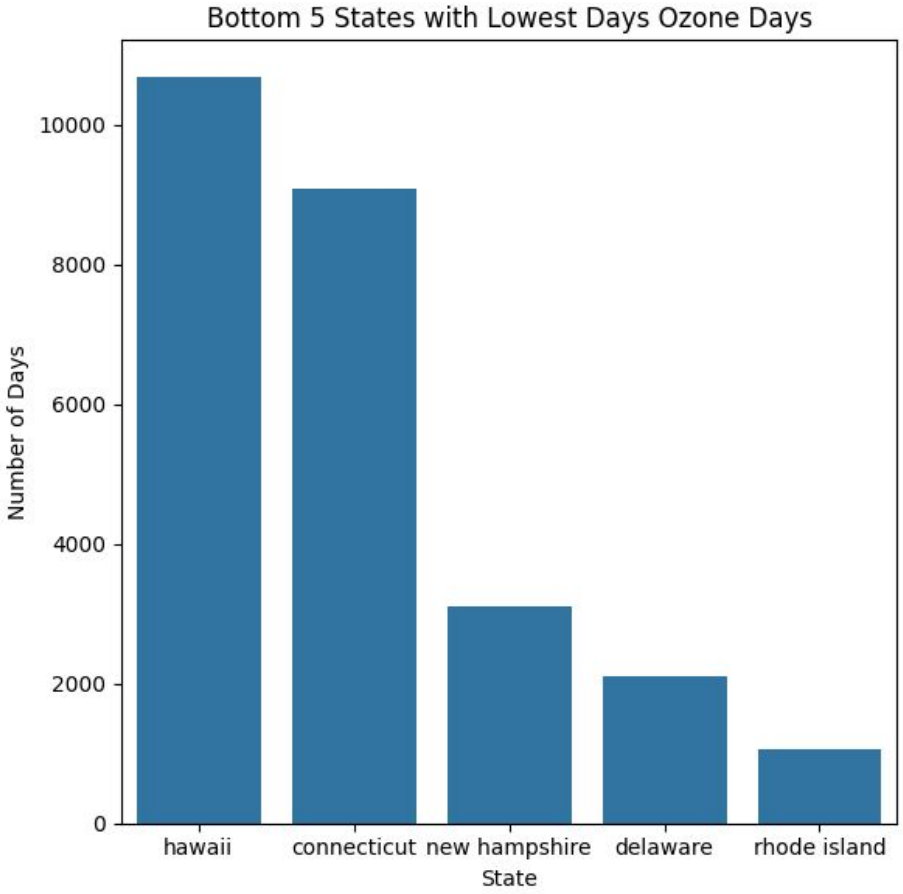

Top 5 States with Highest Days CO Days
Number of Days
25000
20000
15000
10000
5000
0
washington
new jersey
montana
oregon
minnesota
State

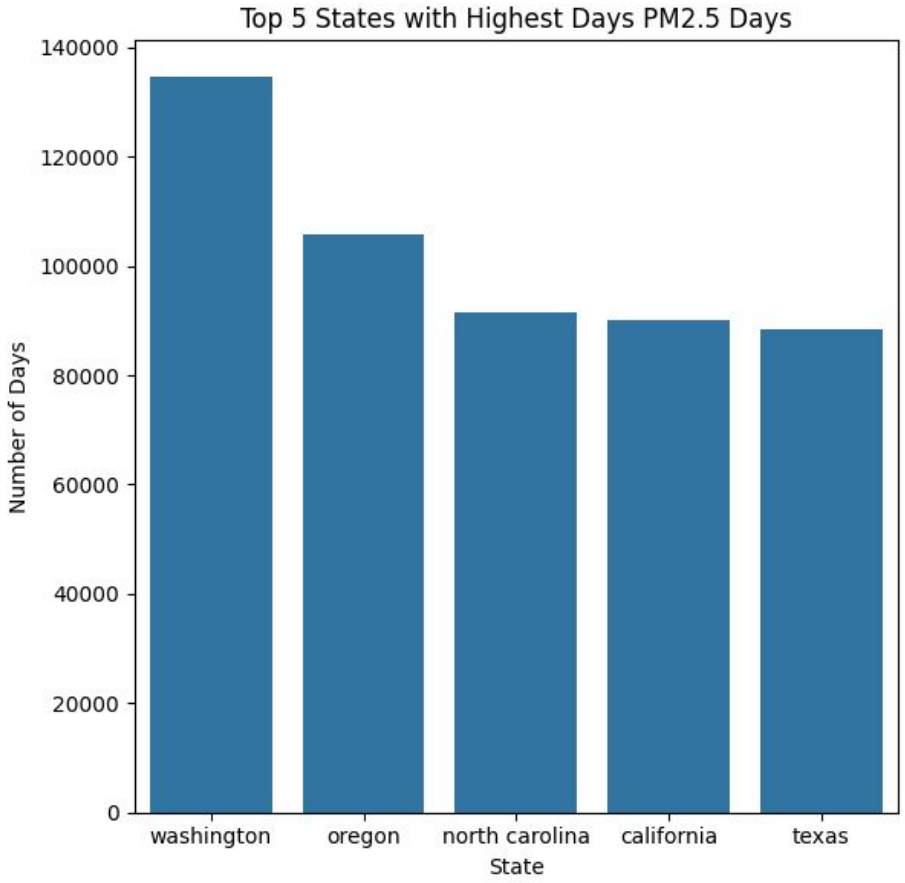

Bottom 5 States with Lowest Days CO Days
Number of Days
350
300
250
200
150
100
50
0
rhode island
maine
delaware
wyoming
south dakota
State

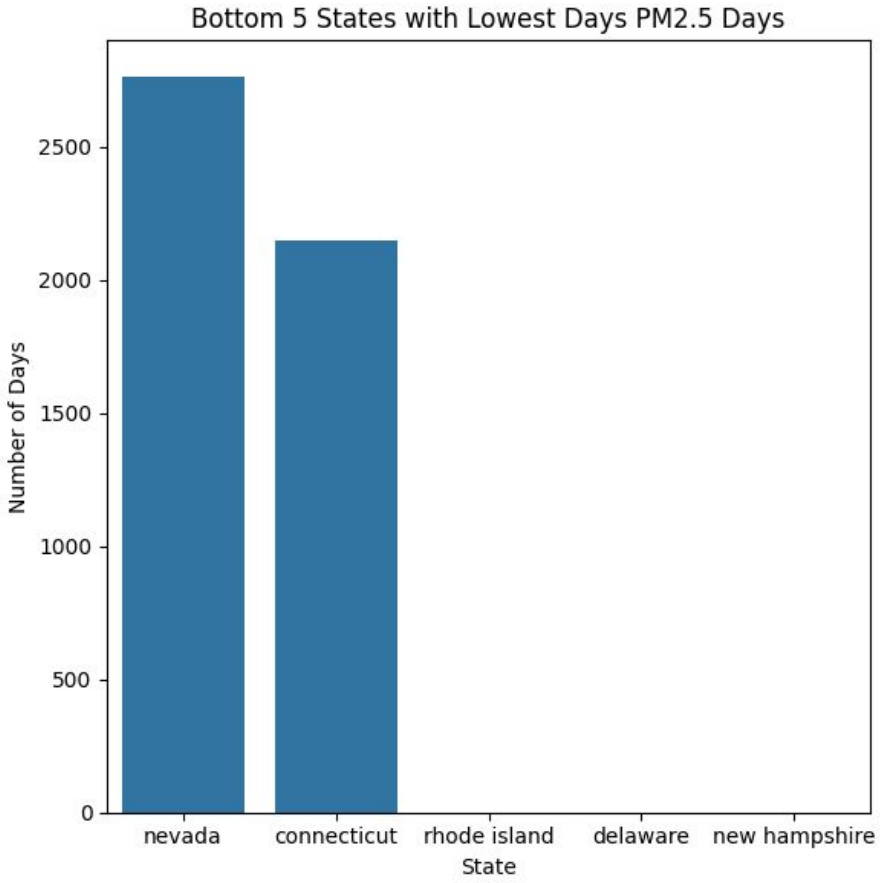

Top 5 States with Highest Days NO2 Days
Number of Days
0
10000
20000
30000
40000
50000
60000
pennsylvania
california
virginia
new jersey
new york
State

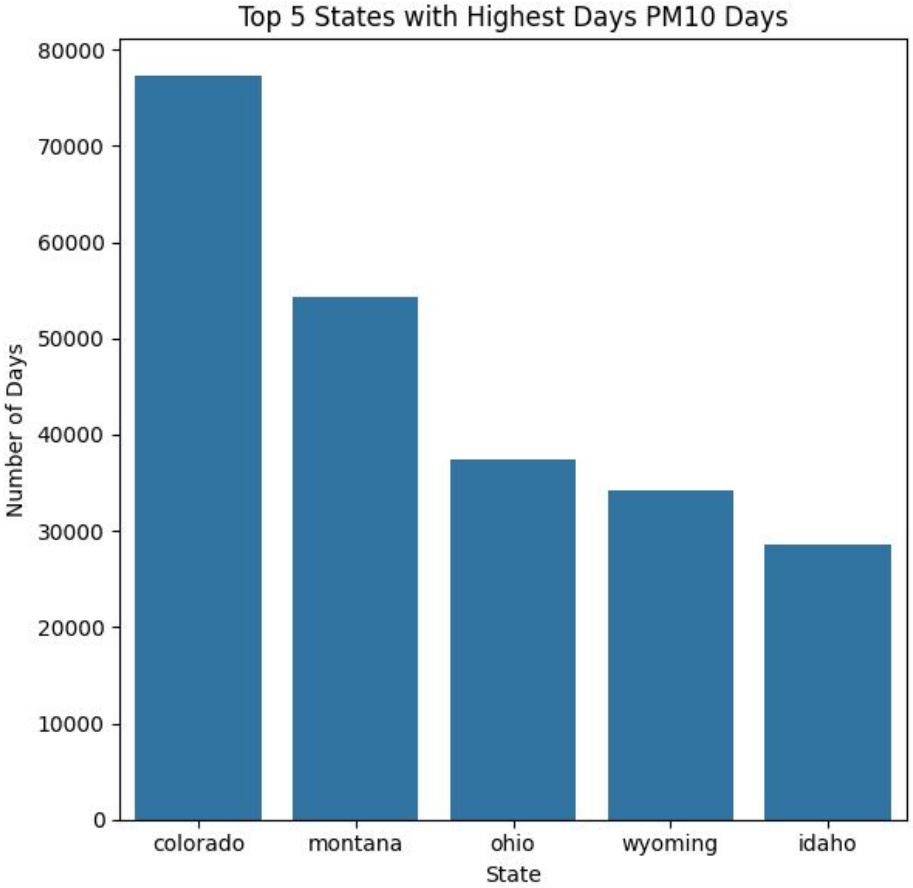

Bottom 5 States with Lowest Days NO2 Days
Number of Days
0
100
200
300
400
500
600
700
rhode island
hawaii
new hampshire
iowa
nebraska
State

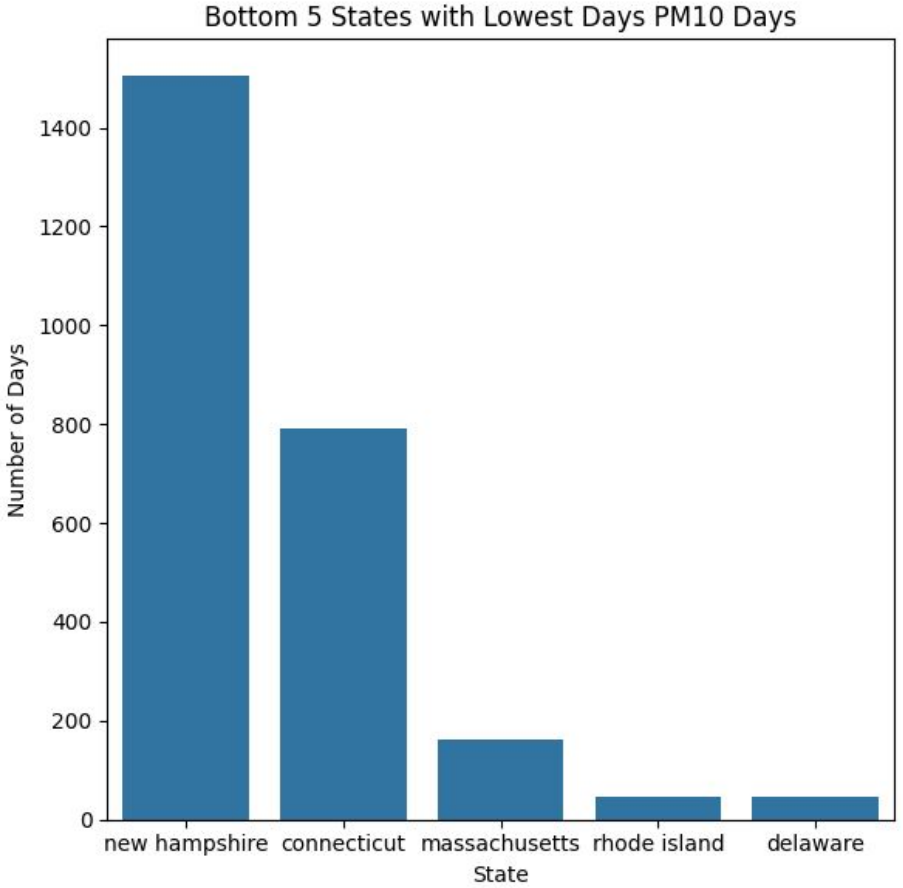

Top 5 States with Highest Days Ozone Days
Number of Days
0
50000
100000
150000
200000
250000
300000
350000
california
new york
florida
texas
ohio
State


Bottom 5 States with Lowest Days Ozone Days
Number of Days
0
2000
4000
6000
8000
10000
hawaii
connecticut
new hampshire
delaware
rhode island
State


Top 5 States with Highest Days PM2.5 Days
Number of Days
0
20000
40000
60000
80000
100000
120000
140000
washington
oregon
north carolina
california
texas
State


Bottom 5 States with Lowest Days PM2.5 Days
Number of Days
0
500
1000
1500
2000
2500
nevada
connecticut
rhode island
delaware
new hampshire
State


Top 5 States with Highest Days PM10 Days
Number of Days
80000
70000
60000
50000
40000
30000
20000
10000
0
colorado
montana
ohio
wyoming
idaho
State


Bottom 5 States with Lowest Days PM10 Days
Number of Days
1400
1200
1000
800
600
400
200
0
new hampshire
connecticut
massachusetts
rhode island
delaware
State

Model, Evaluation and Interpretation

Linear Regression

```
Root Mean Squared Error: 316651.77884869935
R-squared: 0.01718034841386784
Pollutants impact on Bee Colonies (sorted by impact):
    Pollutant  Coefficient
3  Days PM2.5   415.753639
2  Days Ozone   355.479014
4   Days PM10   249.542403
1    Days NO2    59.869507
0     Days CO    19.553487
```

Random Forest Regression Model
RMSE (testing data): 309866.742
R2 Score: 0.059

Conclusions & Recommendations

Questions?