PEOPLE VS. AQI

Anthonette Dominguez
Cecilia Aguayo
Clayton Nagle
Ken Sikes

MOTIVATION:

Our group was interested in the idea that if people adhered to the stay at home orders during the COVID-19 quarantine, there would maybe be some kind of measurable positive impact in the air quality, as measured by Air Quality Index (AQI).

HYPOTHESIS:

IF THE PEOPLE IN SAN DIEGO FOLLOWED THE STAY-AT-HOME ORDERS DURING THE PANDEMIC, THEN WE WOULD SEE A DECREASE IN AQI IN THE TIME PERIOD IMMEDIATELY FOLLOWING SHELTER IN PLACE ORDERS.

WHAT IS AQI?

Air Quality Index

AQI is the measure of pollutants in the air. A higher AQI level represents greater health concerns and vice versa.

WHAT AFFECTS AQI?

FIVE MAJOR POLLUTANTS:

O3

01

GROUND-LEVEL OZONE

PM10/ PM2.5

02

PARTICLE POLLUTION

CO

03

CARBON MONOXIDE

NO2

04

NITROGEN DIOXIDE

SO2

05

SULFUR DIOXIDE

QUESTIONS & MEASUREMENTS

QUESTIONS:

- DO HUMANS HAVE A HUGE IMPACT ON AIR QUALITY?
  - DID THE AQI SHOW ANY CHANGES WHEN GASSES INCREASED OR DECREASED?

MEASUREMENTS

- TIMEFRAME : 2018 - 2021
- MEASUREMENTS OF ATMOSPHERIC GASES THAT COMPRISE AQI
- AQI CHANGES OVER THE GIVEN TIME PERIOD

Data construction

In [1]:
```
import pandas as pd
import numpy as np
import requests
import json
from config import key_dict
```

In [2]:
```
data_2018 = pd.read_csv('Resources/aqidaily2018.csv')
data_2019 = pd.read_csv('Resources/aqidaily2019.csv')
data_2020 = pd.read_csv('Resources/aqidaily2020.csv')
data_2021 = pd.read_csv('Resources/aqidaily2021.csv')
```

In [3]:
```
aqi_data = data_2018.append(data_2019, ignore_index = True)
aqi_data = aqi_data.append(data_2020, ignore_index = True)
aqi_data = aqi_data.append(data_2021, ignore_index = True)
aqi_data
```

Out[3]:

	Date	Overall AQI Value	Main Pollutant	Site Name (of Overall AQI)	Site ID (of Overall AQI)	Source (of Overall AQI)	CO	Ozone	SO2	PM10	PM25	NO2
0	01/01/2018	139	PM2.5	Donovan	06-073-1014	AQS	11	54	0	.	139	33
1	01/02/2018	94	PM2.5	Donovan	06-073-1014	AQS	15	48	0	48	94	47
2	01/03/2018	94	PM2.5	Chula Vista	06-073-0001	AQS	14	40	0	.	94	42
3	01/04/2018	76	PM2.5	Donovan	06-073-1014	AQS	15	44	0	.	76	49
4	01/05/2018	78	PM2.5	Donovan	06-073-1014	AQS	16	35	0	26	78	43
...	...	...	...	...	...	...	...	...	...	...	...	...
1171	03/20/2021	42	Ozone	Chula Vista	06-073-0001	AirNow	NaN	42	NaN	NaN	23	NaN
1172	03/21/2021	48	Ozone	Alpine	06-073-1006	AirNow	NaN	48	NaN	NaN	44	NaN
1173	03/22/2021	50	Ozone	Alpine	06-073-1006	AirNow	NaN	50	NaN	NaN	48	NaN
1174	03/23/2021	44	Ozone	Alpine	06-073-1006	AirNow	NaN	44	NaN	NaN	42	NaN
1175	03/24/2021	52	PM2.5	Donovan	06-073-1014	AirNow	NaN	49	NaN	NaN	52	NaN

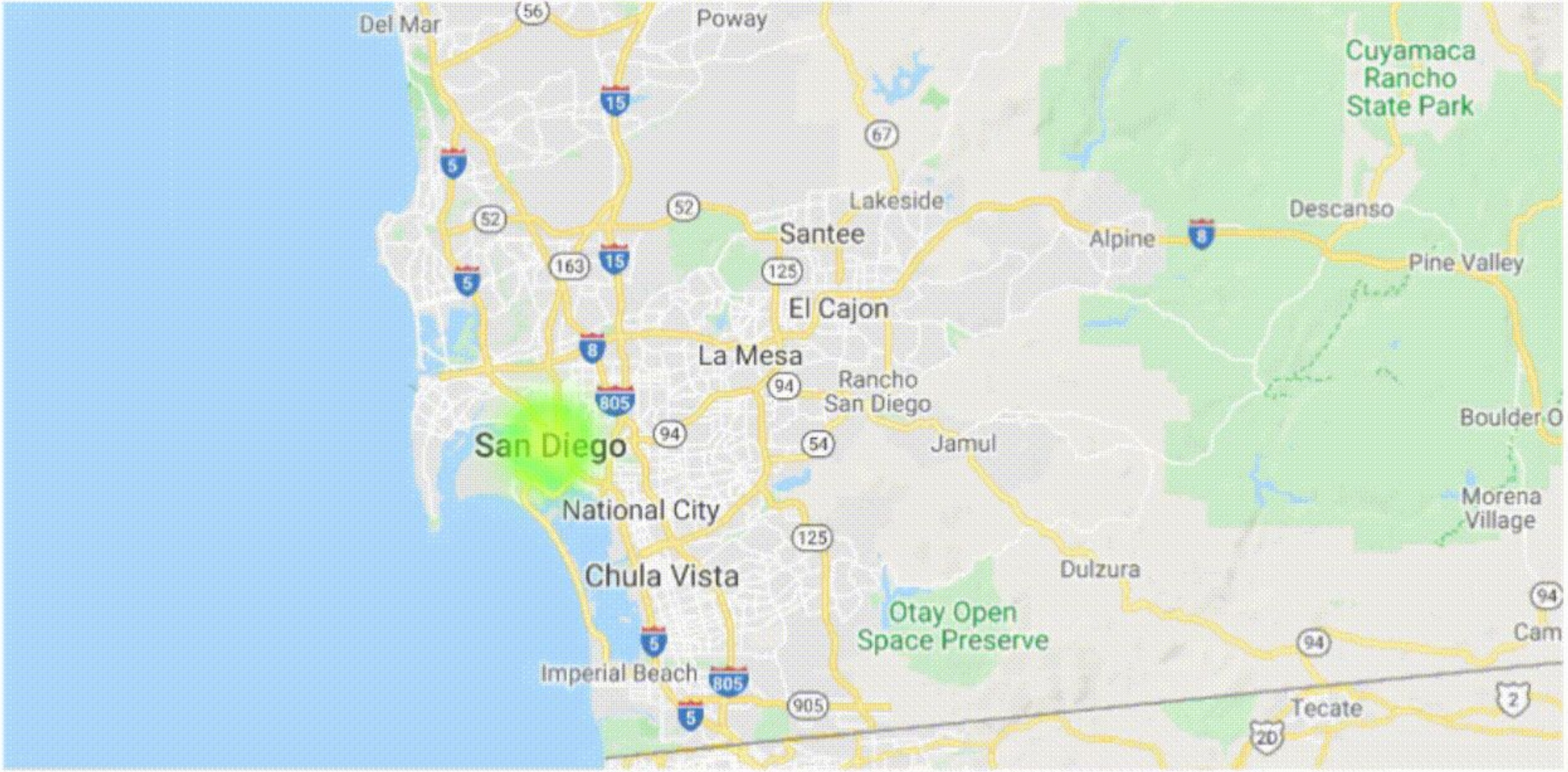
1176 rows × 12 columns

In [41]:
```
aqi_data.to_csv('Resources/aqi_data.csv',index=False)
```

Source limitations and data exploration

- Original source (owm) lacking historical data
- Second source had data from far enough back for meaningful analysis, but lacked consistency in reporting
- Final source (EPA) had very good records, but data required cleaning and manipulation specific to different vars

```
test_final_df = pd.read_csv('Resources/aqi_data.csv')
test_final_df.Date = pd.to_datetime(test_final_df.Date)
test_final_df = test_final_df.set_index('Date')
test_final_df['Lat'] = 32.7157
test_final_df['Long'] = -117.1611
test_final_df.rename(columns = {"Overall AQI Value":"AQI"}, inplace = True)
test_final_df['AQI'] = test_final_df['AQI'].astype(float)
test_final_df = test_final_df.replace(".",None)
test_final_df.iloc[0,8] = None
test_final_df[['CO','Ozone','SO2','PM10','PM25', 'NO2']] = \
    test_final_df[['CO','Ozone','SO2','PM10','PM25', 'NO2']].astype(float)
#final data set up for the whole data gathered
```

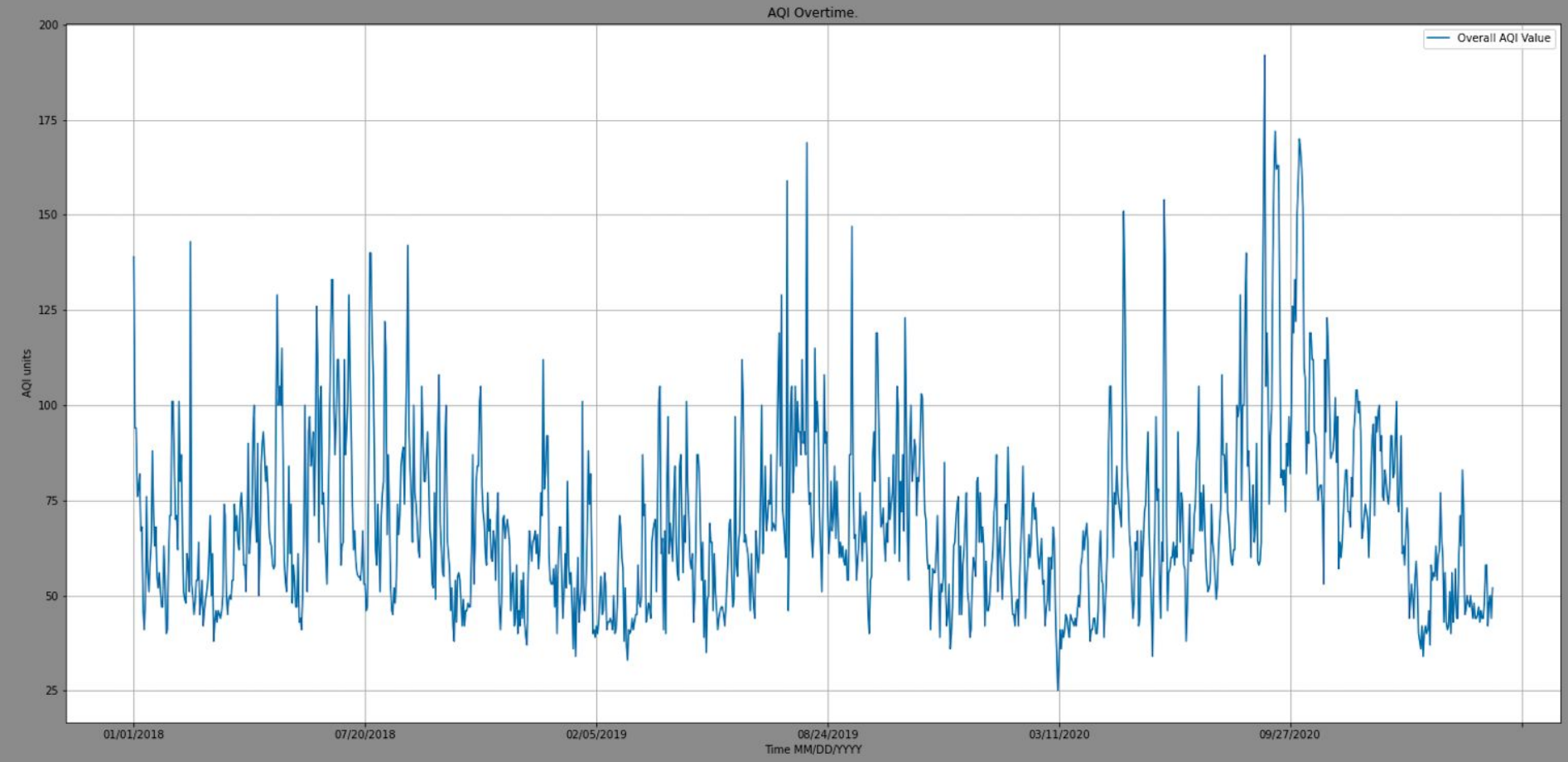

Del Mar
56
Poway
15
67
Cuyamaca
Rancho
State Park
5
52
52
Lakeside
Descanso
Santee
Alpine
8
163
15
125
Pine Valley
5
El Cajon
8
La Mesa
94
Rancho
San Diego
805
Boulder O
San Diego
94
54
Jamul
National City
Morena
Village
125
Chula Vista
Dulzura
94
Otay Open
Space Preserve
Cam
5
94
Imperial Beach
805
905
Tecate
2
5
20

2018-03-14

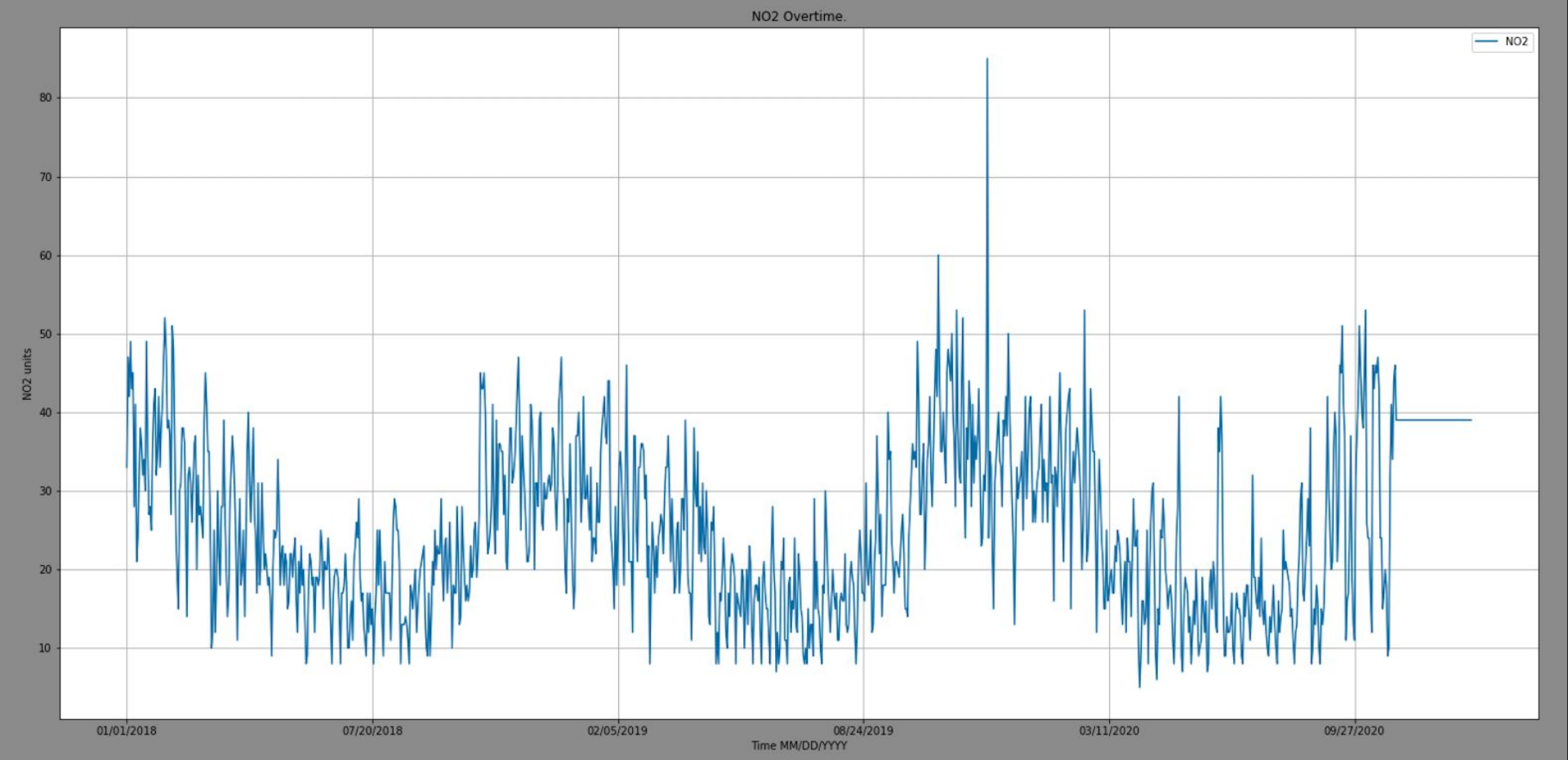
AQI LEVELS OF CONCERN

Daily AQI Color	Levels of Concern	Values of Index	Description of Air Quality
Green	Good	0 to 50	Air quality is satisfactory, and air pollution poses little or no risk.
Yellow	Moderate	51 to 100	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.
Orange	Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is less likely to be affected.
Red	Unhealthy	151 to 200	Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.
Purple	Very Unhealthy	201 to 300	Health alert: The risk of health effects is increased for everyone.
Maroon	Hazardous	301 and higher	Health warning of emergency conditions: everyone is more likely to be affected.

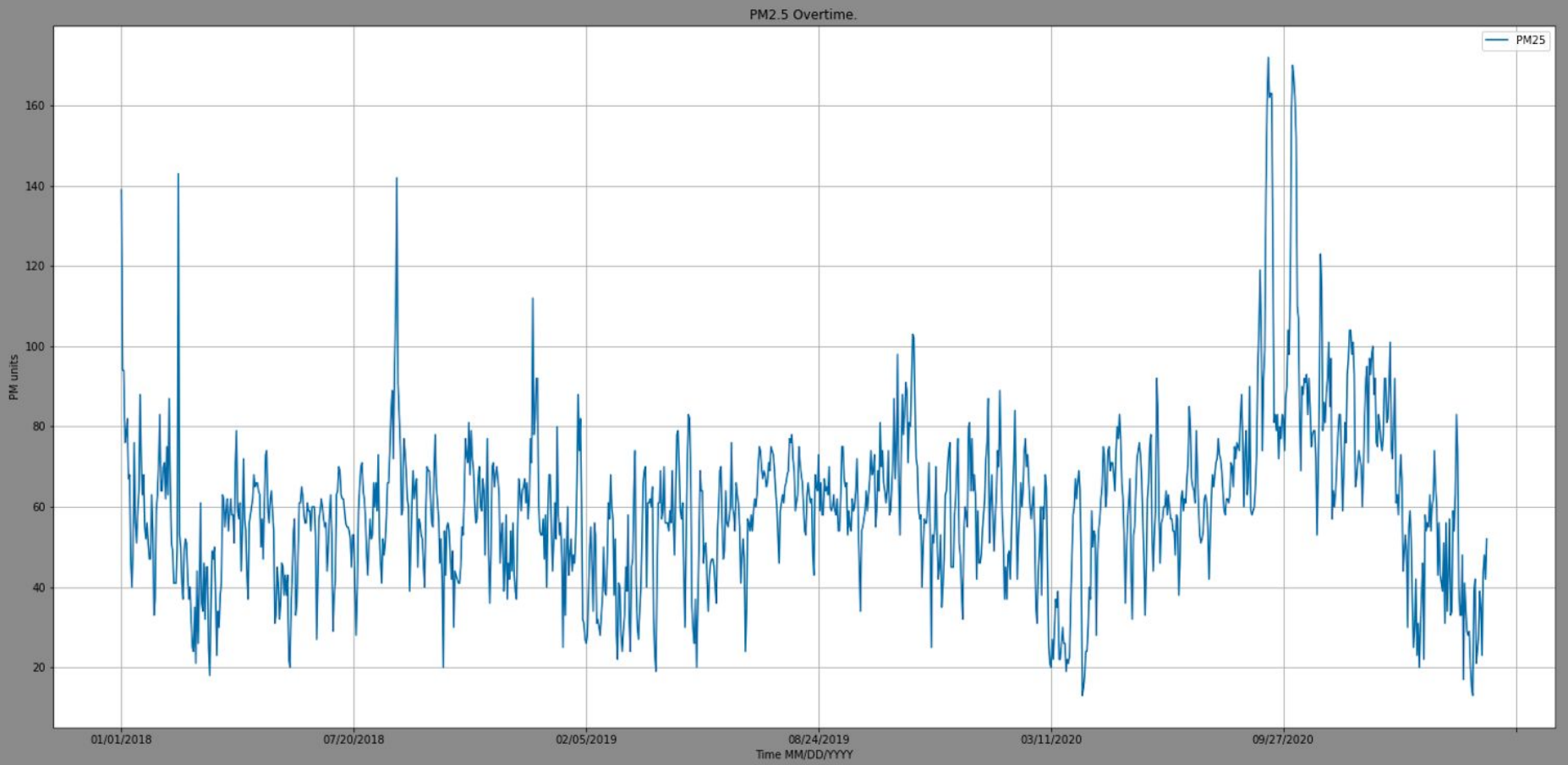
AQI Over Time.



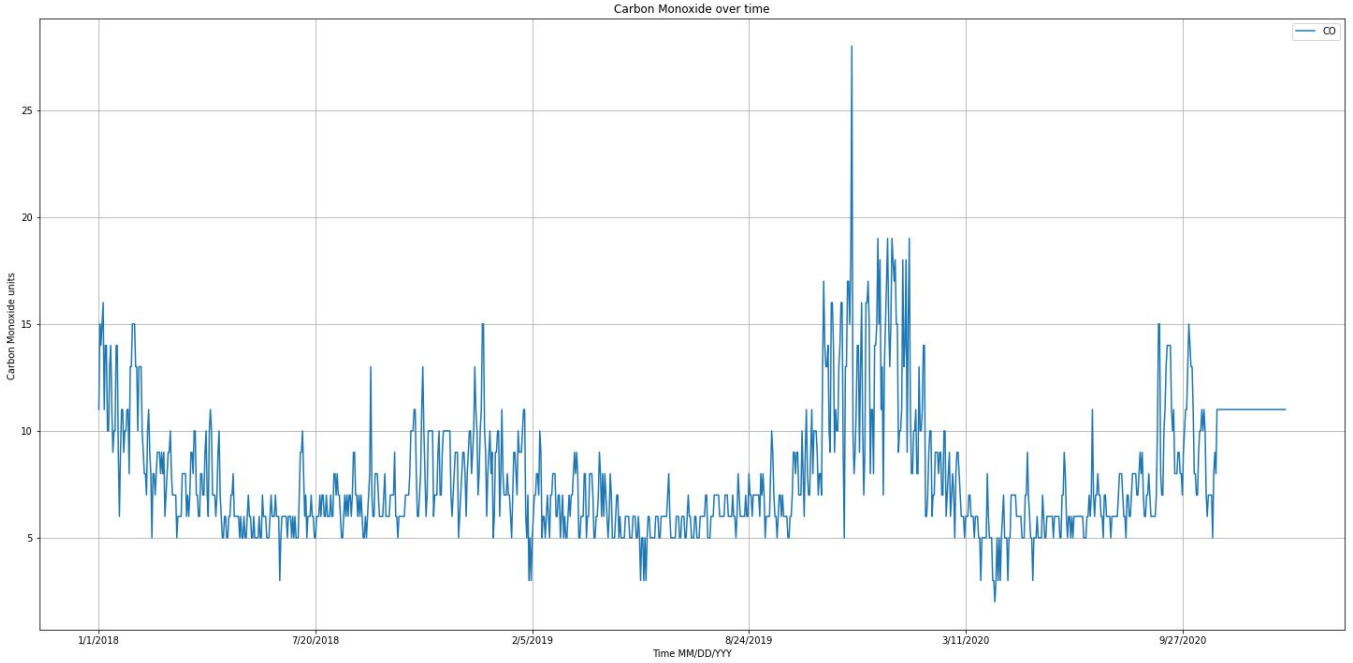
Nitrogen Dioxide Over Time.



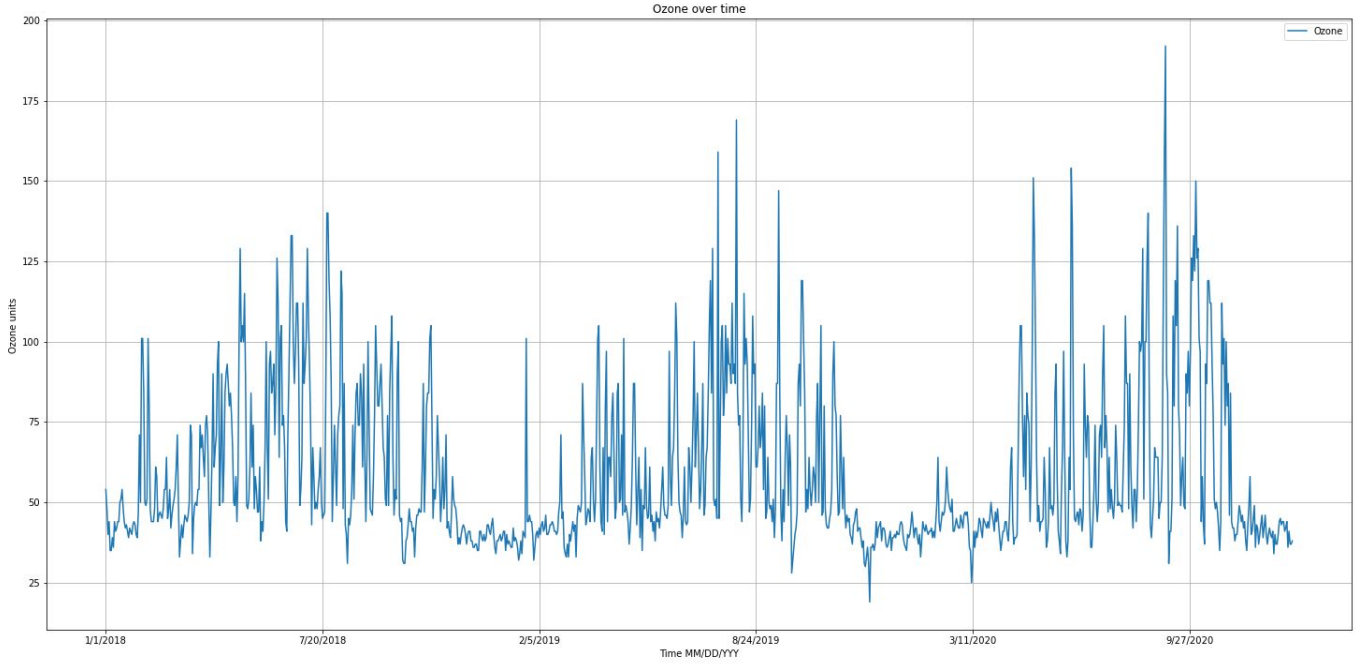
PM2.5 Over Time.



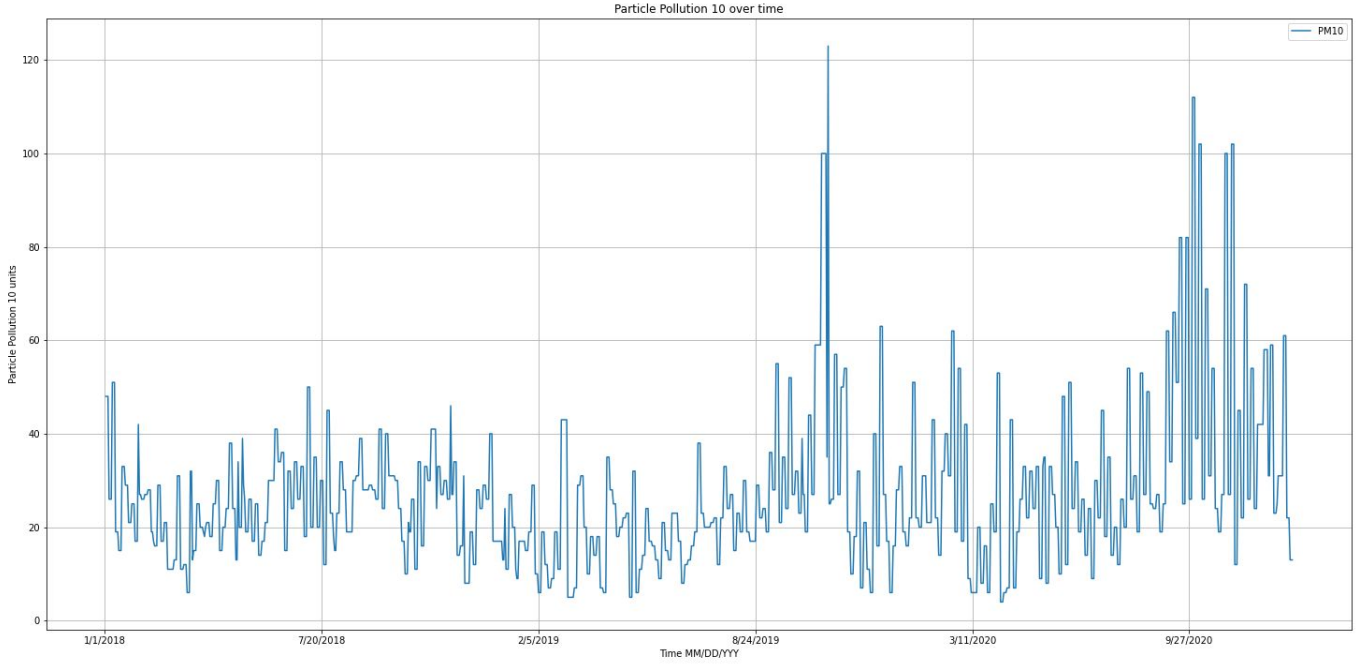
Carbon Monoxide Over Time.



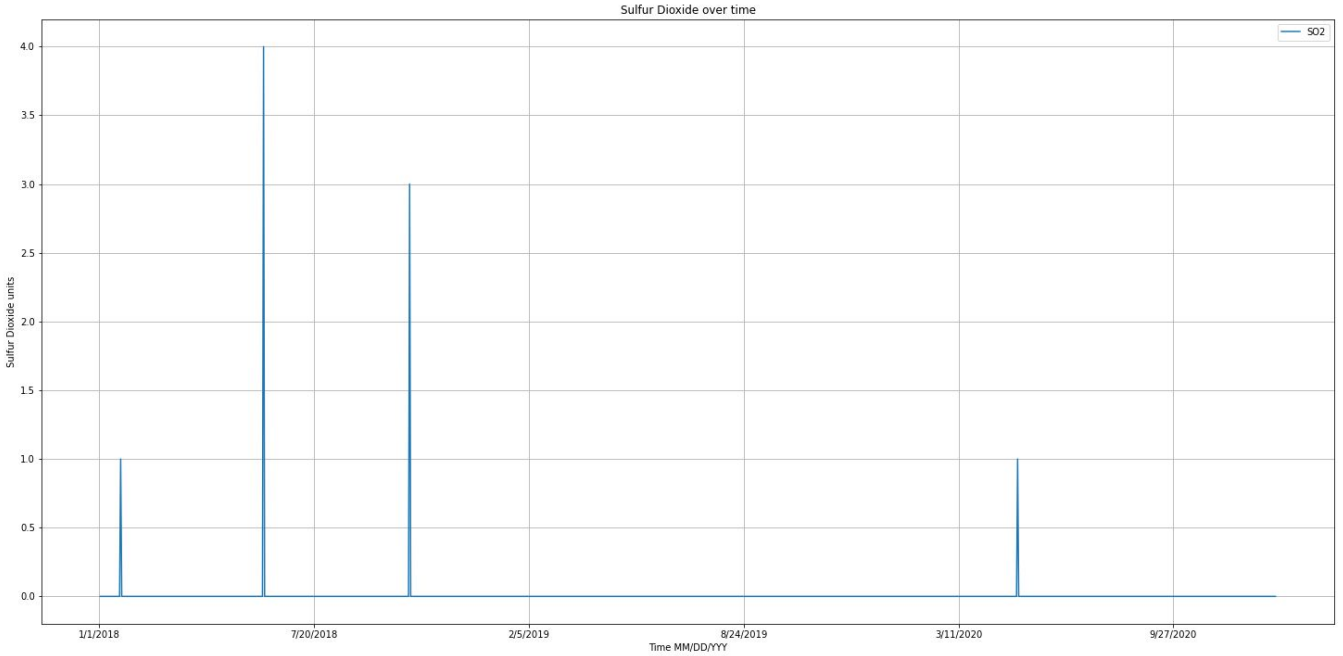
Ozone Over Time.



Particle Pollution (PM10) Over Time.



Sulfur Dioxide over time



DATA ANALYSIS

Date	Date	AQI	CO	Ozone	SO2	PM10	PM25	NO2	Lat	Long
2018	3	53.161	7.581	51.516	0.000	16.484	38.452	26.742	32.716	-117.161
	4	73.133	7.233	69.867	0.000	23.267	59.233	23.400	32.716	-117.161
	5	68.774	5.742	67.677	0.000	21.129	46.806	19.968	32.716	-117.161
	6	90.400	5.767	88.867	0.133	30.200	55.267	17.733	32.716	-117.161
2019	3	53.581	6.613	50.129	0.000	18.194	43.645	23.935	32.716	-117.161
	4	68.167	6.300	61.300	0.000	19.767	56.767	22.433	32.716	-117.161
	5	55.645	5.194	51.742	0.000	15.419	49.129	16.677	32.716	-117.161
	6	65.767	5.767	61.667	0.000	15.967	55.067	15.400	32.716	-117.161
2020	3	46.129	6.065	41.806	0.000	15.613	36.710	20.484	32.716	-117.161
	4	60.667	5.100	54.300	0.000	21.800	50.833	18.267	32.716	-117.161
	5	70.871	5.645	61.710	0.032	23.387	62.226	15.129	32.716	-117.161
	6	68.567	5.967	60.800	0.000	26.133	58.733	17.167	32.716	-117.161
2021	3	47.375	NaN	47.083	NaN	NaN	32.417	NaN	32.716	-117.161

DISCUSSION

- Although we lack a large enough sample size to make a strong statistical claim about the data, the preliminary exploration we conducted seems to indicate that the air quality did improve in the immediate weeks after shelter in place orders were initially sent out (in comparison to the same timeframe in past years)

POST MORTEM

- Beyond the source data problems, it likely would have been helpful to have a more robust understanding of the subject material which would have likely eased the data searching process.
- Each data source had limitations; more time could have allowed us to patch these holes.
    - Similarly, time was a factor in the depth of our analysis. More time could have allowed for more granular or broader exploration of data.
- Shaky grasp of statistical comparison tests means we did not implement any tests to mathematically check our hypothesis against the data; more experience with these tools could have allowed this to be done and generate a stronger conclusion
- As noted, the data was affected in some instances by natural events like wildfires; Although these are a part of the data in reality, a more complex analysis could have perhaps controlled for these outliers.

Questions?

DATA SOURCES

- Open Weather Map Historical Air Pollution API (limited)
- EPA.gov - Air Quality Index Daily Value Report
- Airnow - https://www.airnow.gov/aqi/aqi-basics/