**WIDS Datathon 2022 - Phase 2**

**Research Project Report**

# Research Topic

US Environmental Protection Agency (EPA): weather, air pollutant, and census data

# Research Project

Analysis of Pollutants between 2019 and 2020. Were Air Pollutant Concentrations different in 2019 and 2020 due to COVID-19 lockdowns

Team Name

EarthWise

**Members**:

ThripuraSundari Venugopalan

Sindhu S Chandran

Arvind Betrabet

# Statement of Purpose

This project is designed to understand the variation in Air Pollution levels in the United States of America between the years 2019 and 2020. 2020 is marked as the year of COVID-19, a pandemic that impacted globally.

# Research Questions

The study examines two major research questions:

1. 2019 Vs 2020 - PPM2.5 and Ozone:

Is there a variation in the Air Pollution levels especially Particulate Matter (PPM2.5) and OZONE between the years 2019 and 2020?

1. 2019 Vs 2020 –Pollution levels of PPM2.5 and Ozone on WeekDay Vs Weekend:

Analyse the Pollution levels of Particulate Matter (PPM2.5) and OZONE between the years 2019 and 2020 on Week days and Weekends

# Study Methodology

The methodology used in this project is to perform a Hypothesis based statistical analysis of the Air Pollution levels in the country of USA, between the years 2019 and 2020 to assess the impact of COVID-19 lockdowns on Air Pollution. The details are given below:

1. Null hypotheses: There is no difference in the Concentration levels of PPM2.5 and OZONE between the years 2019 and 2020
2. Alternative hypotheses: There is significant difference in the Concentration levels of PPM2.5 and OZONE between the years 2019 and 2020

## Data Source

The source of data used for the research were primarily the 2 data sets provided by WIDS team for this project sourced from the US Environment Protection Agency (EPA). The main contents of the data are

Air Quality System ID (AQSID)

Date of Measurement

Latitude & Longitude of various location by CBSA, County and State

Census data of Environmental Justice parameters such as

People of Color Fraction

Low Income Group

Fractions of Linguistically Isolated People

Fractions of people with less than Higher-Secondary Education

Climatic conditions such as

Temperature

Humidity

Wind Speed

Wind Direction

Air Pollutant Concentrations of various pollutants such as

PPM2.5 (in Micrograms)

Ozone (Parts per billion)

CO

NO2

SO2

Lead

Benzene

## Data Preparation

It is observed that the measurements for PPM2.5 and Ozone are available for all the counties of both 2019 and 2020. However, by a first view of the data, the max values are much higher than the mean. So, the outliers were removed to avoid one-off situations of extremes interfering with the more common values. Also, the number of samples of data for each county varies across the years. Hence the sample of pollutant data collected involves the same number of samples from each county for the years 2019 and 2020 have been considered.

## Analysis

**Scenario 1:**

2019 Vs 2020 - PPM2.5 and Ozone: Is there a variation in the Air Pollution levels especially Particulate Matter (PPM2.5) and OZONE between the years 2019 and 2020?

Since the objective is to identify the variation between 2 sets of samples for the year 2019 and 2020 of the same population, **“Paired- Sample T-test”** method of statistical analysis was performed.

The analysis was performed in Python and the details are as below:

1. **PPM 2.5 Levels**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (Index) | COUNTY | P | PPM25\_2019 | PPM25\_2020 | T |
| 0 | Kent | 0.030626 | [ 4.3 5.4 6. 8.9 4.4 5.3 3. 6.4 4.6 ... | [ 5.7 7.9 8.5 8.3 3.7 6.3 6.8 7.4 4.8 ... | 2.165759 |
| 1 | Jefferson | 0.008702 | [ 7.3 9.2 3.5 ... 8.2 10.2 15.9] | [5.4 5.9 5.5 ... 2. 3. 5.6] | 2.626428 |
| 2 | Lane | 0.284856 | [7. 6.5 5.9 3.3 3.4 3.6 1.8 3.9 3.3 7.7 5.1 2... | [3.2 2.5 5.4 2.5 2.5 1.5 1.8 4.7 6.2 4.5 4.6 3... | -1.082367 |
| 3 | Rowan | 0.069085 | [10. 9.3 2.9 3.2 8.6 7.2 3.1 4.3 8.5 ... | [11.4 9.7 7.4 2.8 2.6 5.3 6.7 2.7 7.5 ... | 1.854134 |
| 4 | Preble | 0.015386 | [ 4.6 8.4 7.7 10.2 11.3 12.6 6.3 5.8 4.6 ... | [ 3.2 5.4 6.8 10.9 4.3 4.5 6.6 7.2 4.5 ... | 2.436310 |
| ... | ... | ... | ... | ... | ... |
| 338 | Ottawa | 0.175875 | [ 7.2 7. 10.8 9.1 10.5 3.2 2.5 11.9 1.9 ... | [ 1.6 8.6 4.6 4.4 4. 3.1 8.3 13.9 7.6 ... | -1.366643 |
| 339 | Allegan | 0.596998 | [ 7.4 15.5 6.1 5.6 7.5 8.5 2.2 13.3 1.5 ... | [ 3.4 4.2 8.2 7.5 7. 6.5 3.5 4.3 6.5 ... | -0.530926 |
| 340 | Linn | 0.034033 | [ 7. 7. 5.4 10.5 16.2 13.5 9.7 12.6 12. ... | [ 8.7 4.9 6.5 3.6 3.3 4. 7.6 12.1 11. ... | -2.132579 |
| 341 | Sedgwick | 0.225876 | [ 6.1 5.9 4.6 4.9 4.9 7.4 1.8 4.5 2.7 ... | [ 3.9 5.8 5.1 6.1 8.4 7.8 11. 4.1 3.2 ... | -1.213875 |
| 342 | Madison | 0.080951 | [ 7.1 6.2 8.1 3.5 7.1 3.2 5.7 6.4 9.3 ... | [ 9. 7.2 5. 5.1 5.3 8.3 8.5 4.3 2.7 ... | 1.757122 |

The values of T and P are varying across the counties. Hence different counties experienced different variations in PPM2.5 levels.

**Median value of T - 0.18122514214768423**

**Median value of P - 0.04337803478005304**

2. OZONE\_PPM

The mean value of T comes to **2.498185289703127**

and the mean value of P is **0.005039655045701538**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (Index) | COUNTY | OZONE\_2019 | OZONE\_2020 | P | T |
| 0 | Kent | [0.029 0.027 0.02 0.027 0.021 0.034 0.032 0.0... | [0.035 0.032 0.029 0.026 0.031 0.031 0.034 0.0... | 4.047662e-06 | 4.641314 |
| 1 | Jefferson | [0.023 0.025 0.018 ... 0.037 0.039 0.044] | [0.034 0.029 0.033 ... 0.039 0.042 0.035] | 1.562729e-16 | 8.331395 |
| 2 | Lane | [0.043 0.043 0.05 0.027 0.014 0.026 0.029 0.0... | [0.039 0.041 0.047 0.035 0.039 0.028 0.033 0.0... | 8.250433e-02 | -1.775943 |
| 3 | Rowan | [0.038 0.04 0.038 0.039 0.025 0.024 0.025 0.0... | [0.039 0.026 0.014 0.032 0.034 0.037 0.034 0.0... | 9.656856e-01 | -0.043216 |
| 4 | Preble | [0.024 0.018 0.021 0.025 0.021 0.024 0.031 0.0... | [0.032 0.028 0.019 0.033 0.031 0.029 0.031 0.0... | 2.763619e-07 | 5.251722 |
| ... | ... | ... | ... | ... | ... |
| 338 | Ottawa | [0.036 0.045 0.047 0.054 0.037 0.043 0.045 0.0... | [0.04 0.046 0.041 0.039 0.034 0.041 0.035 0.0... | 2.464213e-01 | -1.168336 |
| 339 | Allegan | [0.032 0.038 0.044 0.045 0.051 0.039 0.042 0.0... | [0.036 0.036 0.036 0.044 0.039 0.045 0.045 0.0... | 1.095050e-02 | 2.607447 |
| 340 | Linn | [0.038 0.042 0.044 0.049 0.046 0.044 0.052 0.0... | [0.04 0.03 0.037 0.038 0.034 0.037 0.039 0.0... | 8.424305e-02 | 1.734184 |
| 341 | Sedgwick | [0.014 0.02 0.032 0.031 0.03 0.018 0.014 0.0... | [0.035 0.026 0.03 0.022 0.018 0.023 0.021 0.0... | 2.480975e-05 | 4.292406 |
| 342 | Madison | [0.023 0.038 0.048 0.046 0.052 0.047 0.052 0.0... | [0.056 0.032 0.042 0.033 0.033 0.018 0.034 0.0... | 1.673735e-03 | 3.201065 |

### Inference

This shows that there is a significant variation in the level of PPM2.5, thus proving that the Hypothesis that PPM2.5 levels in both the years is not similar across the country, but varies from county to county.

### Results

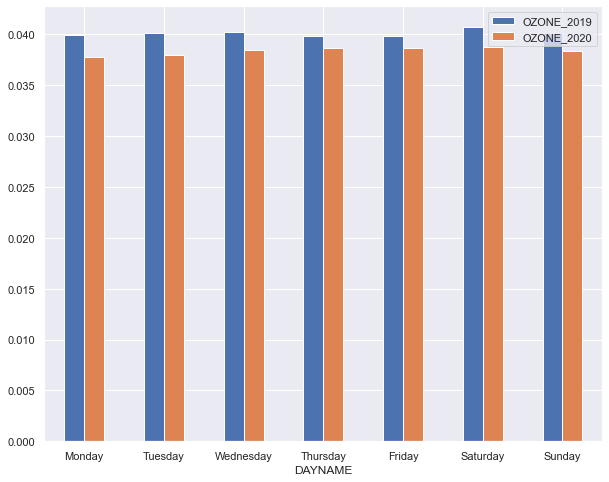
There is variation in the PPM2.5 and OZONE Concentrations of Air Pollution is not uniform across the counties in the 2 years being compared

**Scenario 2:**

2019 Vs 2020 –Pollution levels of PPM2.5 and Ozone on WeekDay Vs Weekend: Analyse the Pollution levels of Particulate Matter (PPM2.5) and OZONE between the years 2019 and 2020 on Weekdays and Weekends

|  |
| --- |
| DAYNAME DAYTYPE OZONE\_PPM\_2019 OZONE\_PPM\_2020 |
| Monday 0 0.039893 0.037802 |
| Tuesday 1 0.040183 0.037947 |
| Wednesday 2 0.040203 0.038510 |
| Thursday 3 0.039885 0.038618 |
| Friday 4 0.039880 0.038674 |
| Saturday 5 0.040719 0.038772 |
| Sunday 6 0.040148 0.038324 |

​

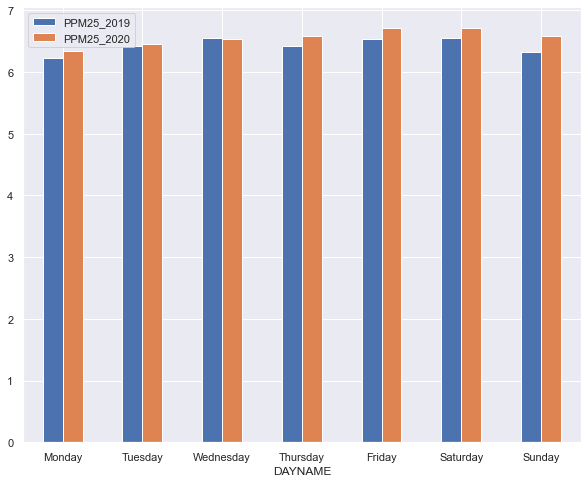


The bar graph shows that there is a considerable amount of reduction in OZONE concentrations during the weekdays.

Particulate Matter: PPM2.5

The week-day wise mean values of PPM2.5 in 2019 and 2020 are as follows. This has been calculated after eliminating the outliers.

|  |
| --- |
| DAYNAME DAYTYPE PPM25\_2019 PPM25\_2020 |
|  |
| Monday 0 6.230468 6.338489 |
| Tuesday 1 6.419479 6.452461 |
| Wednesday 2 6.544204 6.528337 |
| Thursday 3 6.424266 6.586072 |
| Friday 4 6.541763 6.711632 |
| Saturday 5 6.543928 6.709221 |
| Sunday 6 6.325170 6.578326 |



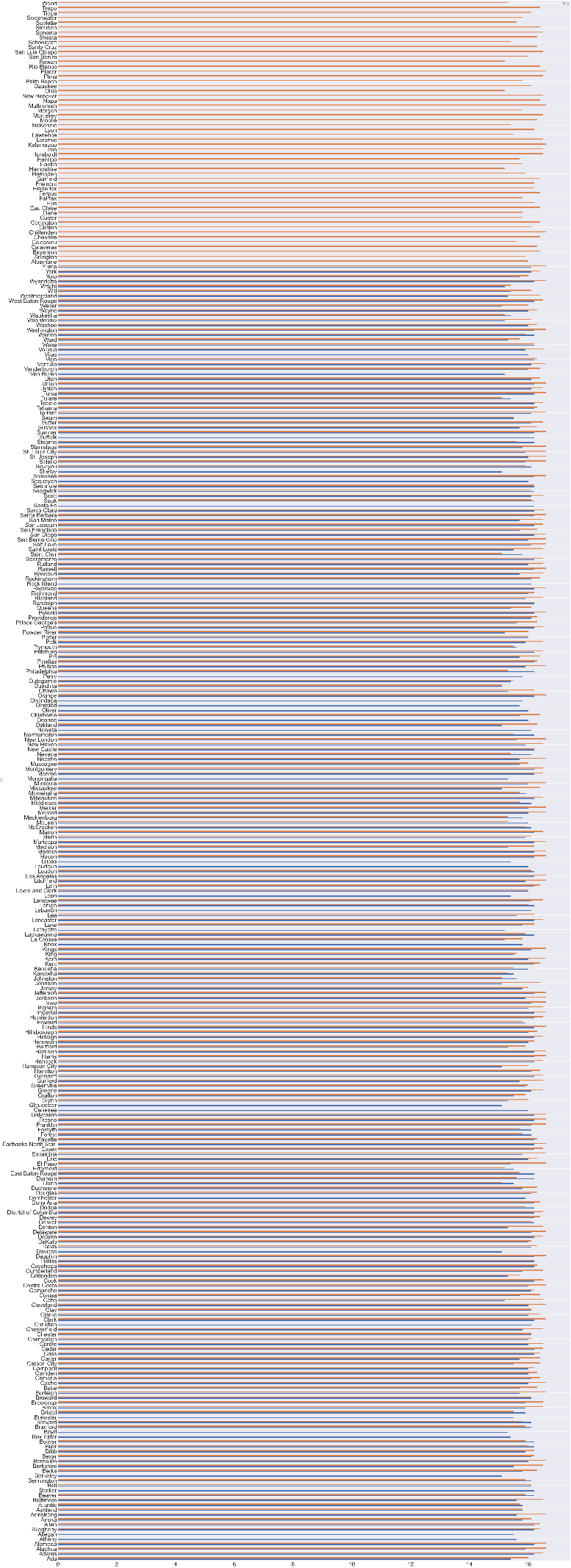
The observation shows that in 2020 there has been an increase in the PPM content compared to 2019 and this has happened more during the weekends.

Also, it is to be noted that the PPM concentration is relatively higher during the weekends than during the weekdays in 2020.

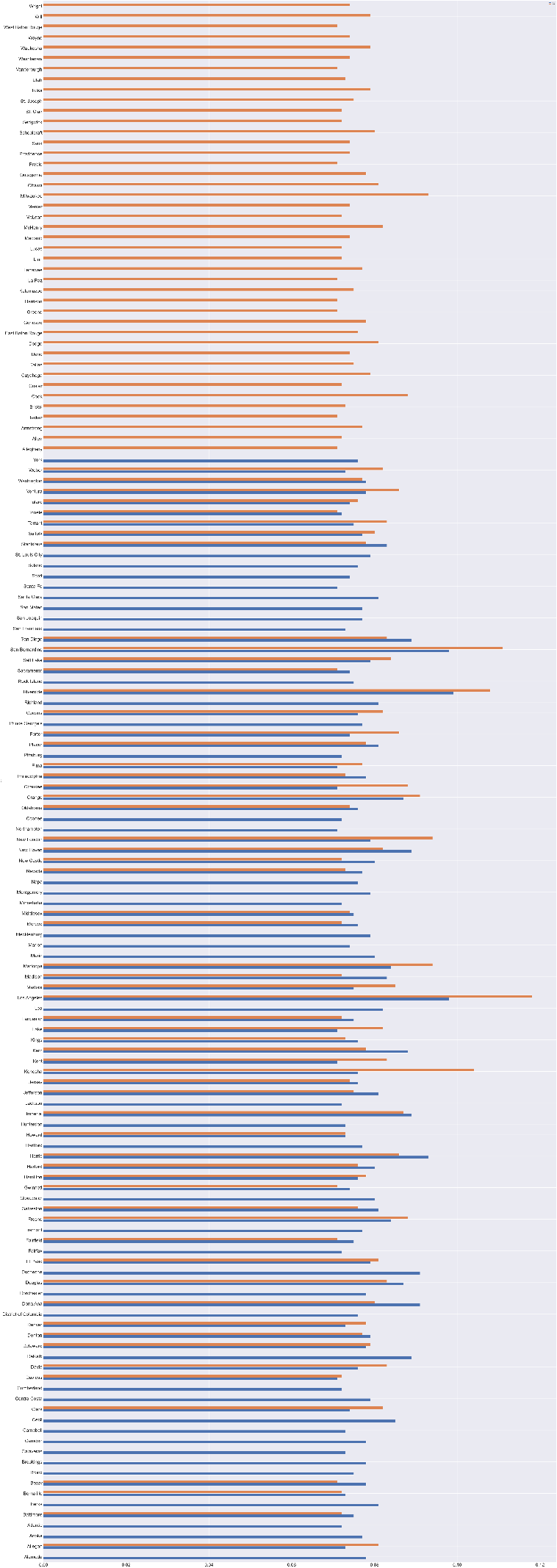
Discussion:

Further analysis shows that there were 41 counties which exceeded PPM2.5 levels of 15microgram per cubic meter in both 2019 & 2020, but there were only 10 counties which exceeded PPM2.5 levels of 15 micrograms per cubic meter in 2019 but not in 2020.

The following graph shows the PM2.5 levels in 2019 and 2020 only for those counties which exceeded the 35-microgram standard. It can be seen that there were a greater number of counties which exceed this standard level in 2020 than in 2019.



Ozone levels in 2019 Vs 2020 for those counties where the level was above the standard level of 0.07 Parts Per Billion.



Conclusion:

Air pollution varied at different counties in 2019 and 2020.