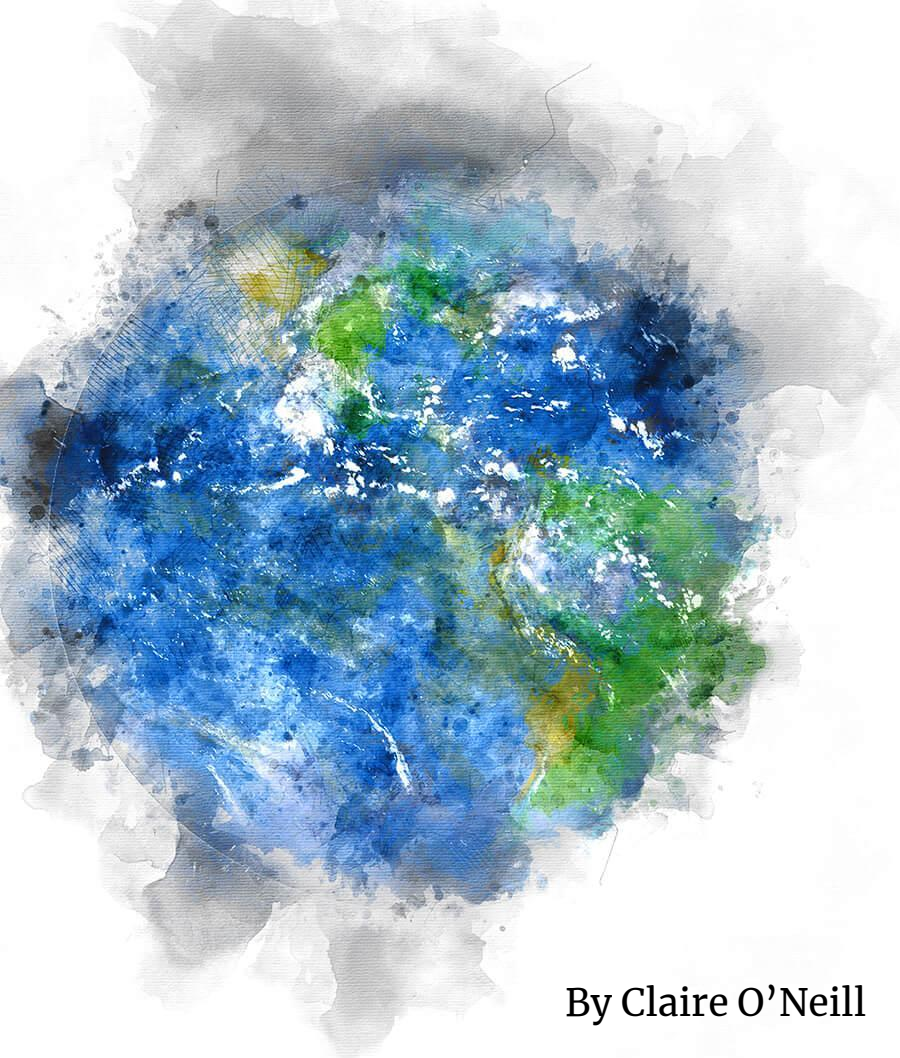


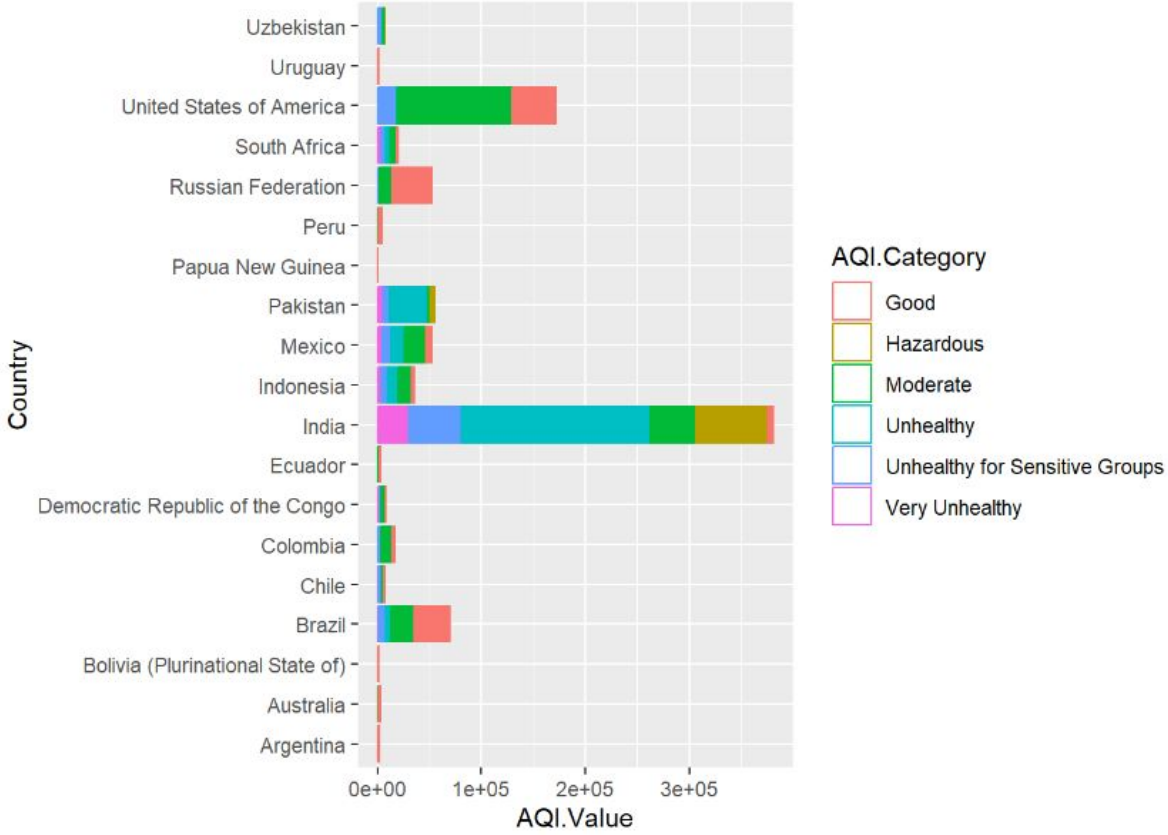
Should the US worry about Air quality?

By Claire O'Neill

Introduction

Our analysis provided a lot of insight into which factors contribute most to air quality level. After cleaning the data and looking at the visualizations provided, we were able to see trends that will allow us to properly give advice about reducing Air quality level in the US and ensure we remain within a healthy range.

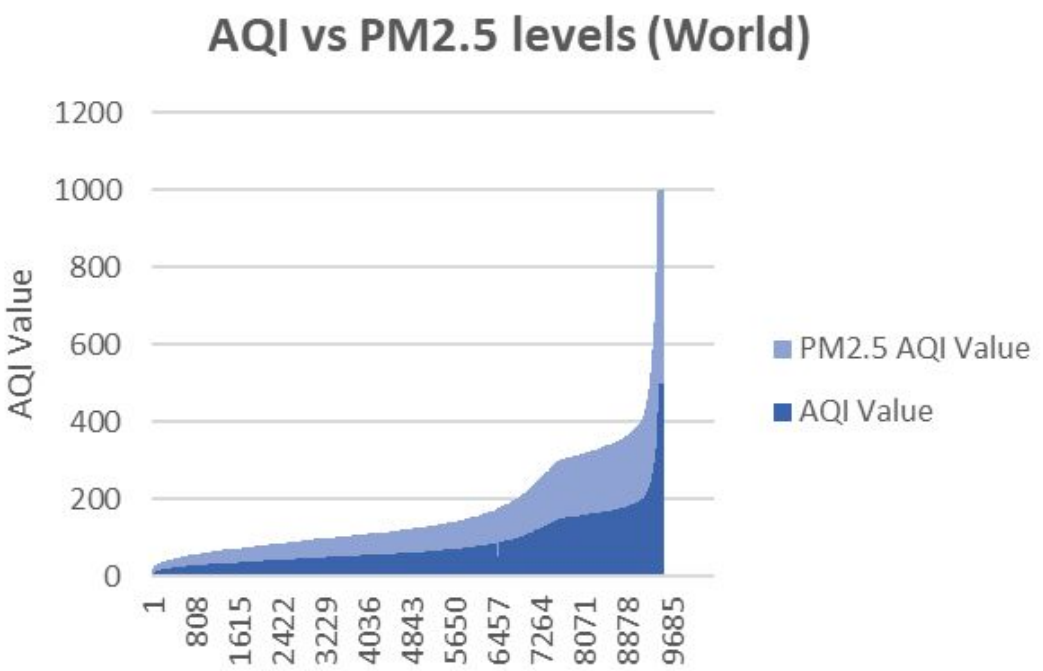
Plot of AQI Values in the world



This bar plot shows the frequency of Air Quality values, with the categories denoted by colors in the key. This model uses a subset of our data presenting the the best and worst air quality separated by countries. From this bar plot, we can see that countries with lower AQI have mostly red bars, in comparison to those with a higher AQI whose bars are a combination of other colors. India has the most extreme observations. While the US has many observations of moderate Air quality.

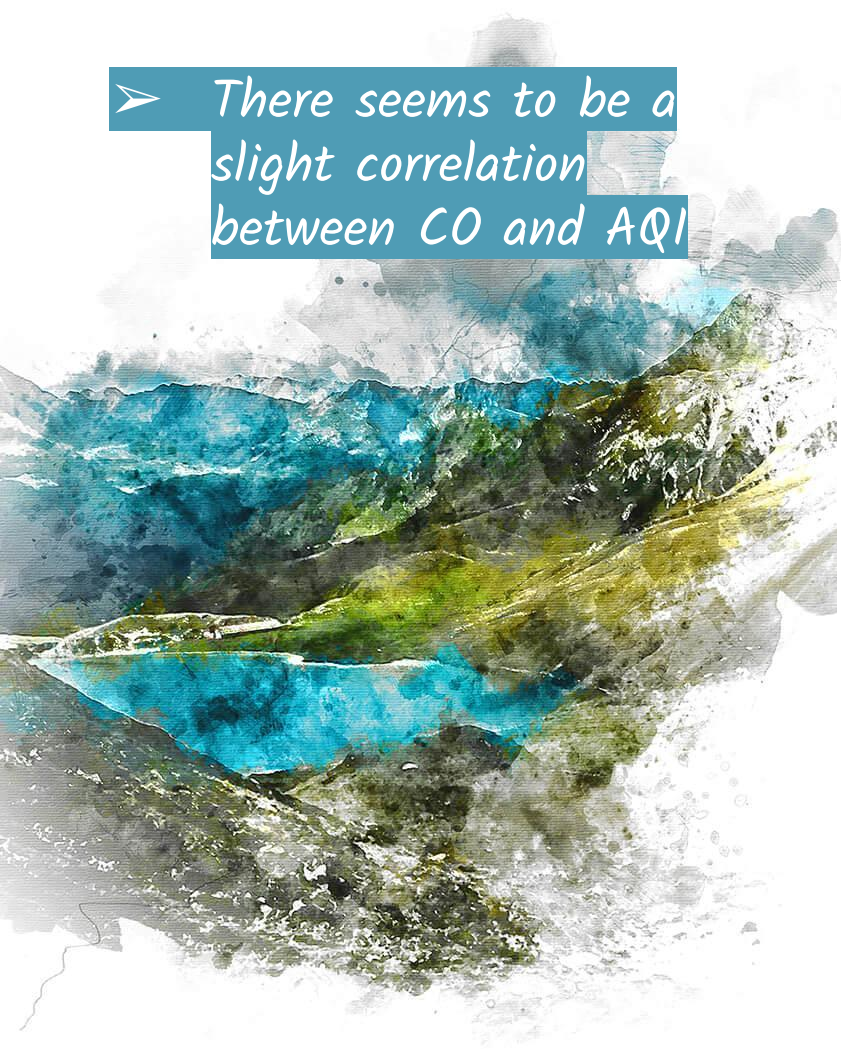
- High Correlation between PM2.5 Values and overall AQI value

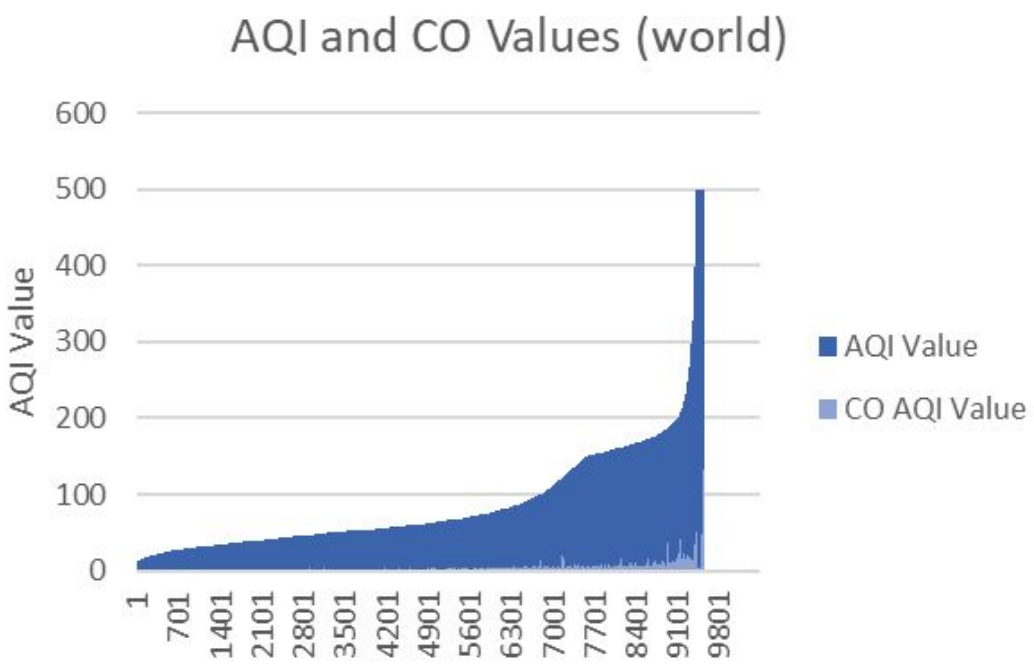
Impact of PM2.5 on AQI



Impact of CO Value on AQI

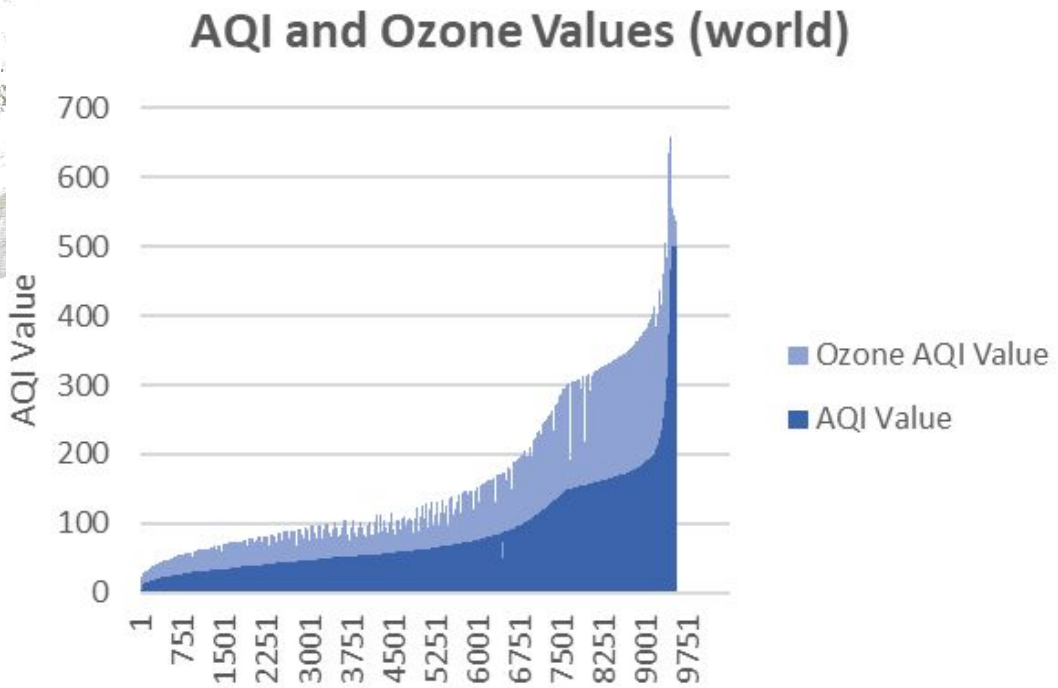
- There seems to be a slight correlation between CO and AQI





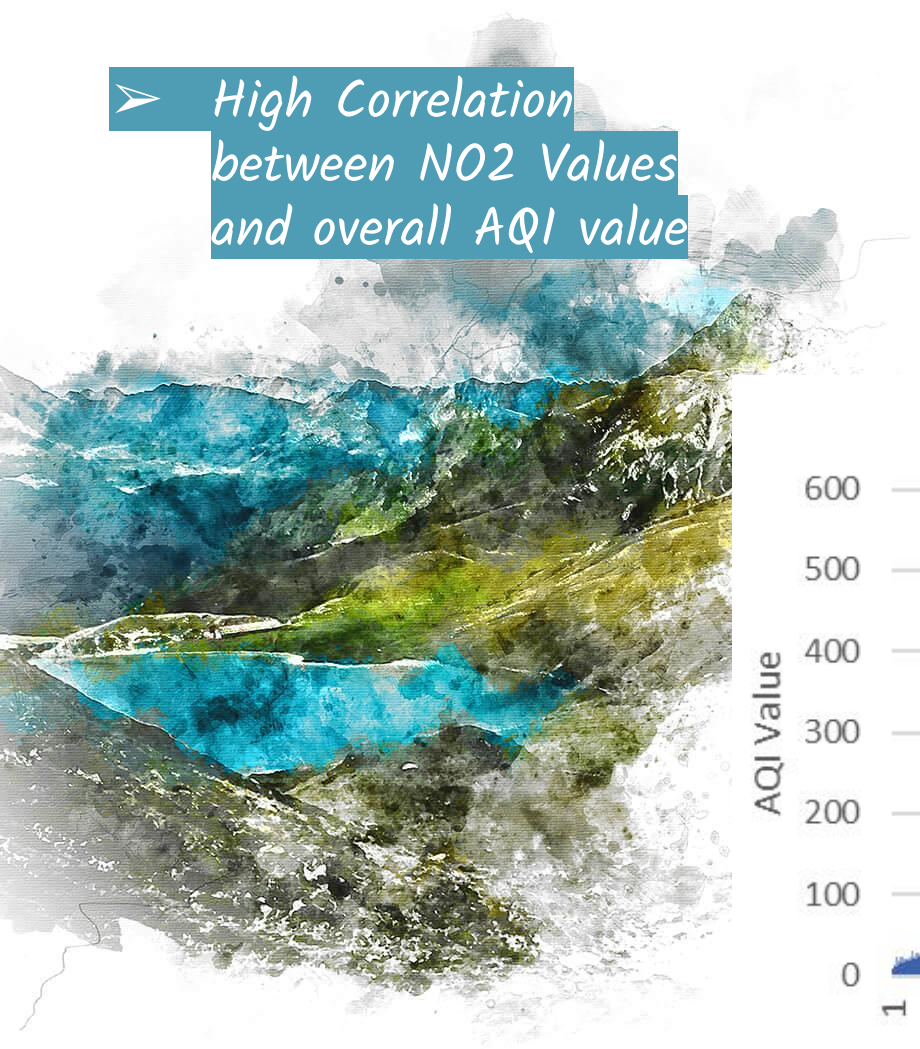
Impact of Ozone on AQI

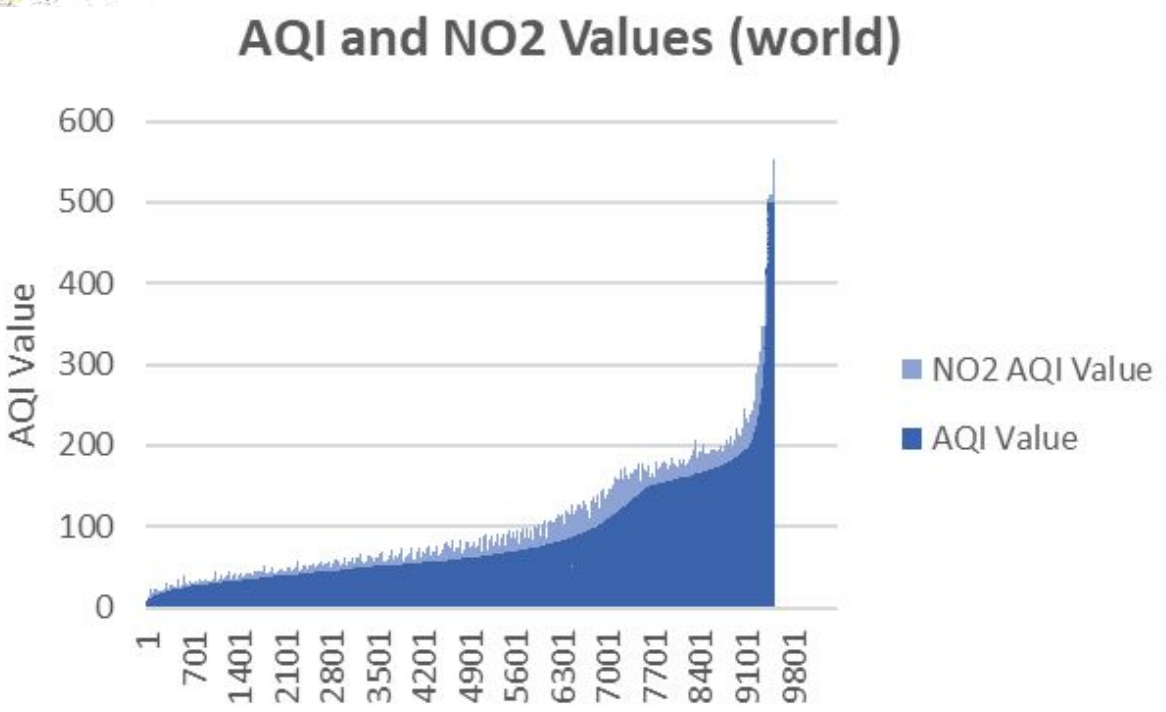
- There appears to be a relatively strong correlation between Ozone and AQI Value



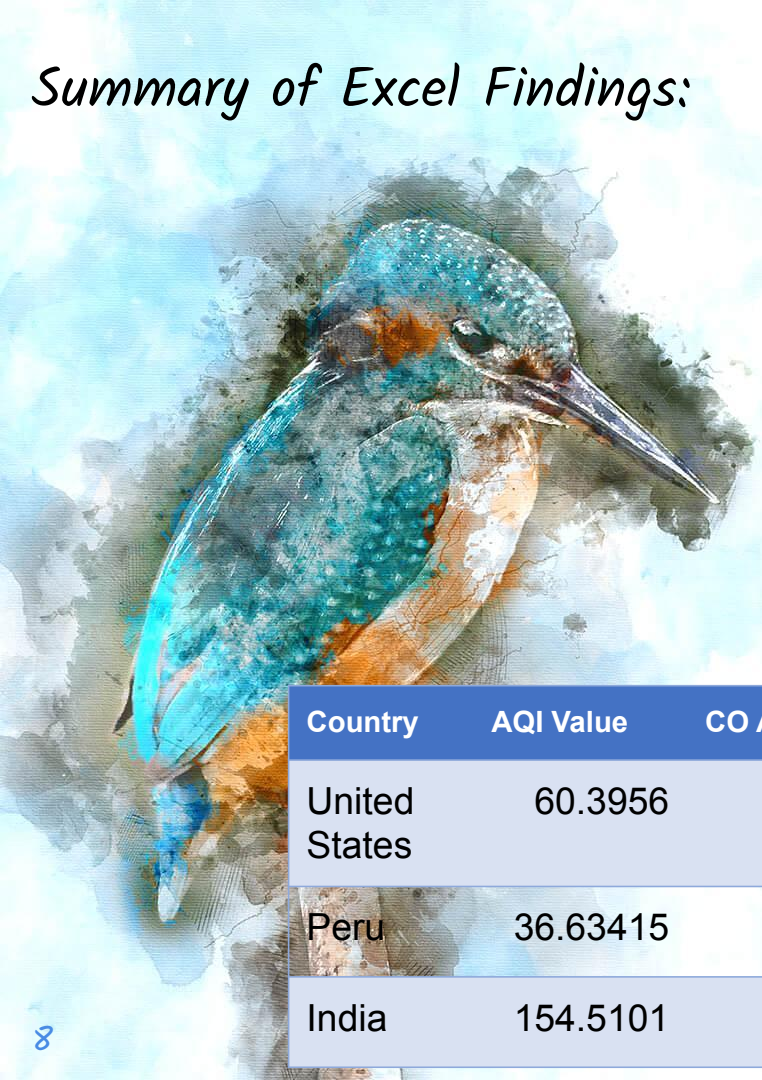
Impact of NO2 values on AQI

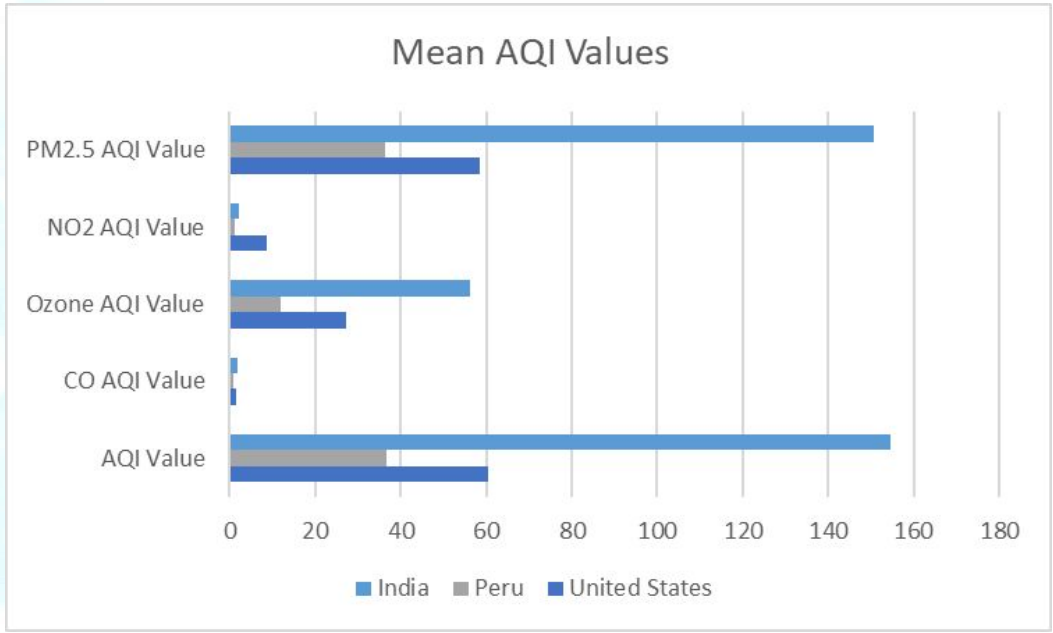
- High Correlation between NO2 Values and overall AQI value





Summary of Excel Findings:

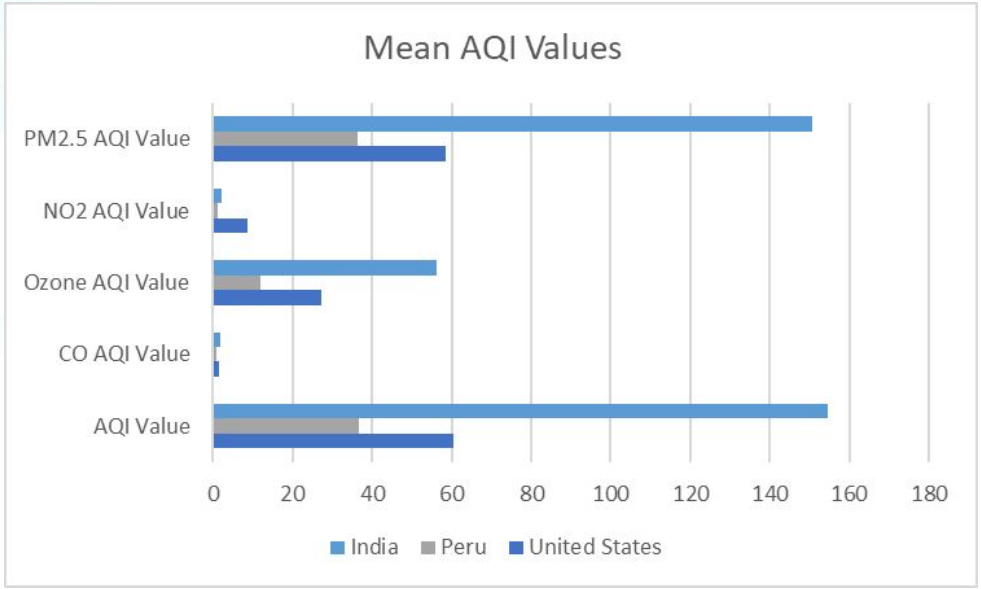




Country	AQI Value	CO AQI Value	Ozone AQI Value	NO2 AQI Value	PM2.5 AQI Value
United States	60.3956	1.487405	27.20076	8.799618	58.63359
Peru	36.63415	0.715447	12.03252	1.211382	36.31707
India	154.5101	1.807544	56.34205	2.059151	150.7994

Summary of Excel Findings:

- We can see a clear relation between high levels of PM2.5 and AQI Value
- We can also see a clear relation between NO2 and final AQI Value
- Note: Mean PM2.5 and mean AQI value have nearly identical bar plots
- We will continue to explore this relation by running a regression between these values



Regression model coefficients:

Mod2: Response AQI value Predictor NO2

- Intercept: 64.437
- Slope No2: 2.473
- R-squared: 0.054

Mod3: Response AQI value predictor PM2.5

- Intercept: 3.0157
- Slope PM2.5: 1.007
- R-squared: 0.969

Mod3: Response AQI value predictor Ozone

- Intercept: 43.554
- Slope Ozone: 0.809
- R-squared: 0.164

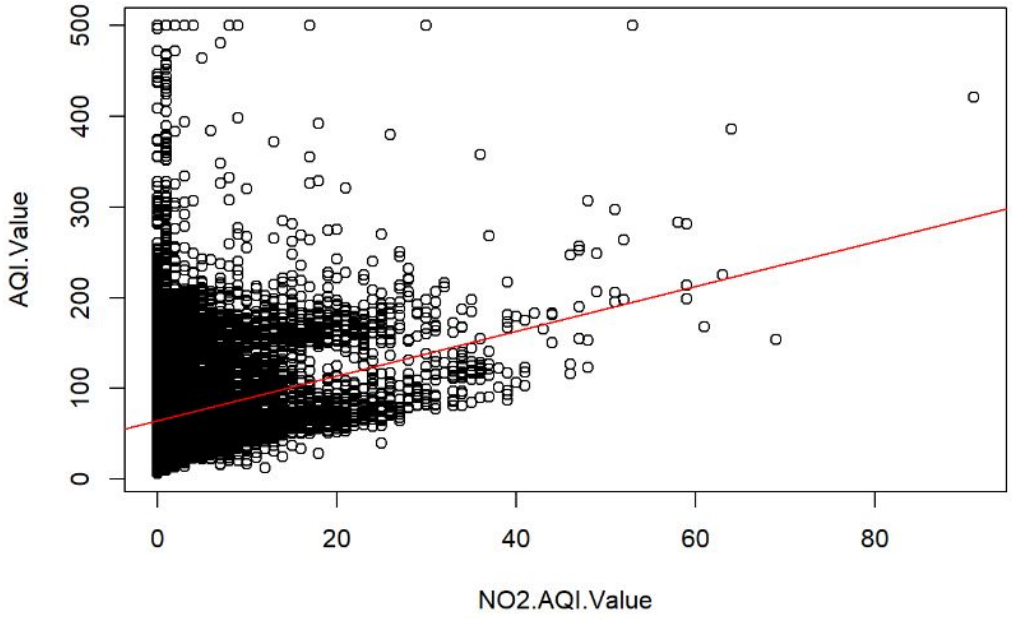
Mod1: Response AQI Value and all predictors and interaction term between NO2 and Ozone

- Intercept: -0.937
- Slope No2: 0.054
- Slope Ozone: 0.172
- Slope PM2.5: 0.980
- Interaction term: -0.005
- R-squared: 0.975

Regression Model (mod2):

Predictor: NO2 value

Response: AQI value

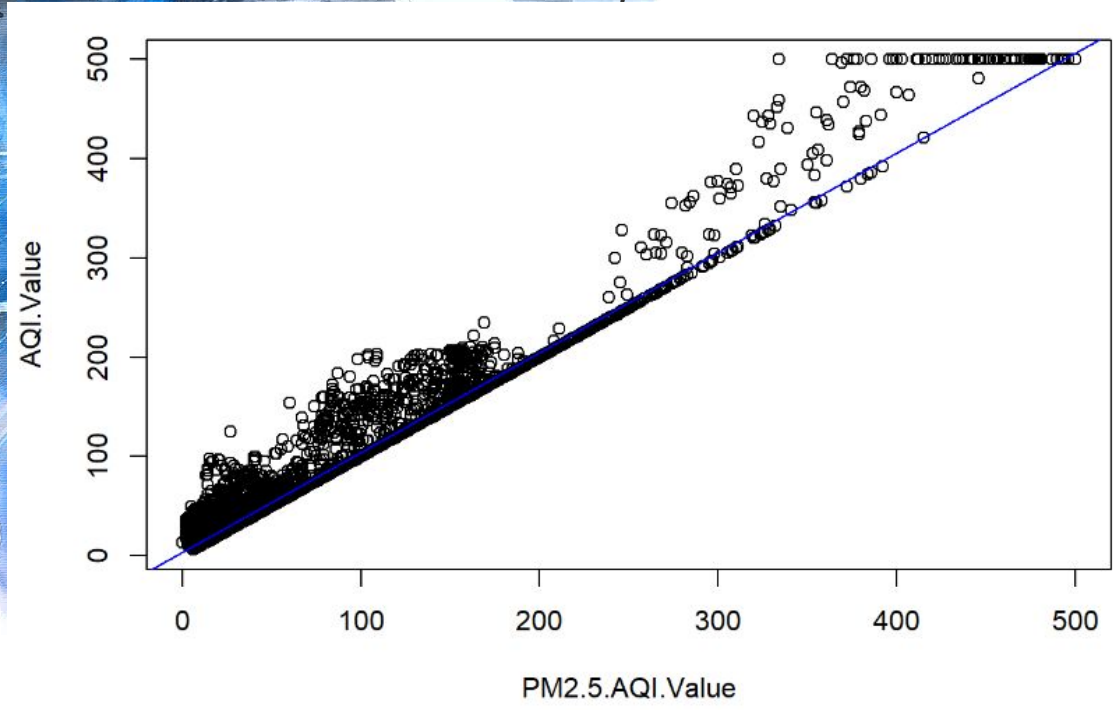


The regression run with AQI value as the response and NO2 value as the predictor has a mostly linear relation, however, it is not homoskedastic. While our model, denoted in red, does a good job at modelling the data given, it will be unable to make predictions since the variance is not constant. We can see that our model has difficulty modelling values at the end of NO2 value, since there are less observations, and the variance changes. Overall, our model reflects the moderate positive correlation between AQI and NO2.

Regression Model (mod3):

Predictor: PM2.5 value

Response: AQI value

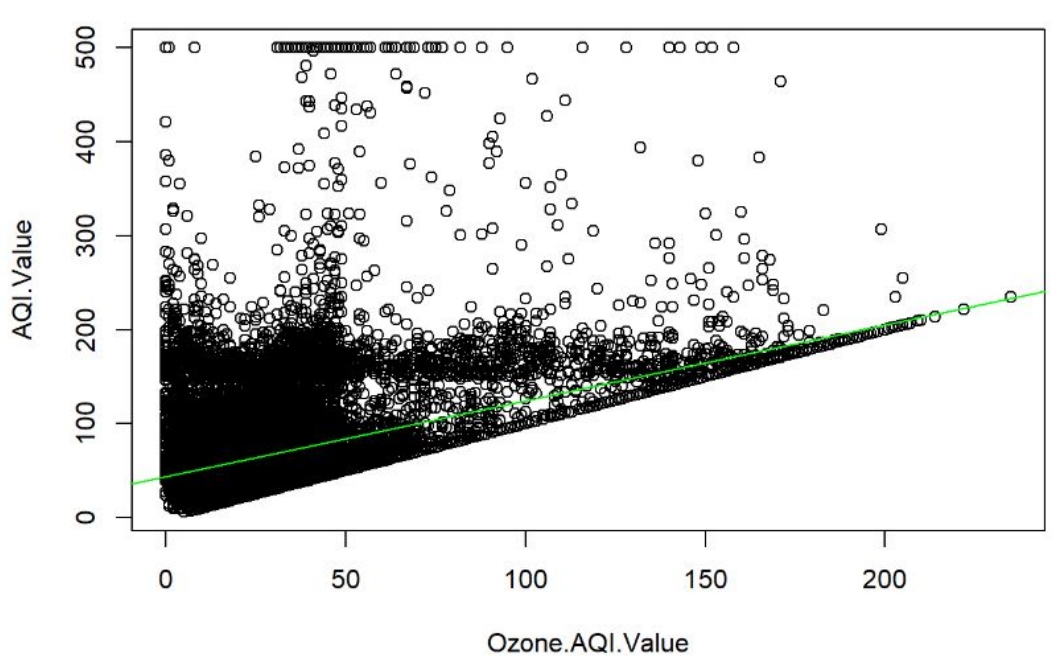


This particular model seems to model our data with PM2.5 AQI value as the predictor very well, as reflected in our plot including our regression model as well as the multiple r-squared value of 0.969. Since our values in our scatter plot appear to have relatively consistent residuals, we can expect our data from this model to be more accurate. Although this model does a better job in comparison to our prior model where NO2 was the only predictor, we need to take into account our models accuracy for larger values. Since our normal Q-Q model shows the observations deviating towards the end, we can see our model's accuracy is worse for data points towards the end of our data set. This inaccuracy is partially due to the spread of our observations, however, it should be noted for future reference.

Regression Model (mod4):

Predictor: Ozone Value

Response: AQI value



This final regression model with Ozone value as the predictor does not seem to have consistent variance, with values at the beginning having a much larger variance than those at the end. This also seems to have a low adjusted r-squared value of 0.1643, indicating it may not be the best model to make predictions with in the future. Overall, our regression model with Ozone as the predictor is able to model the linear relation, however the changes in variance impacted its overall accuracy. Plotting regressions with confidence intervals.

Regression Analysis

STRENGTHS

We found that PM2.5 and NO2 have a large impact on Air quality level

WEAKNESSES

Our data is not distributed with equal variance for all models, this makes certain predictions inaccurate

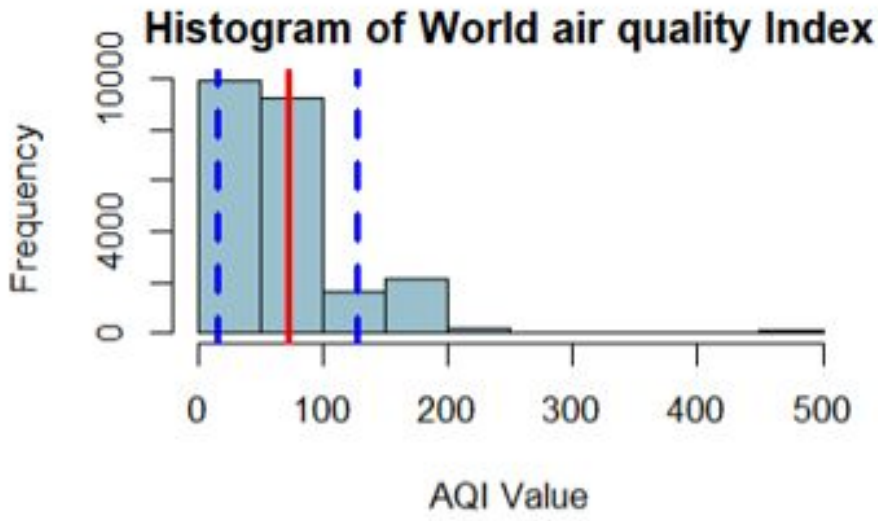
Improve our overall Air quality level by trying to reduce PM2.5 and NO2 levels in the future.

OPPORTUNITIES

Extreme value in PM2.5 have a strong impact on our air quality level according to our regression

THREATS

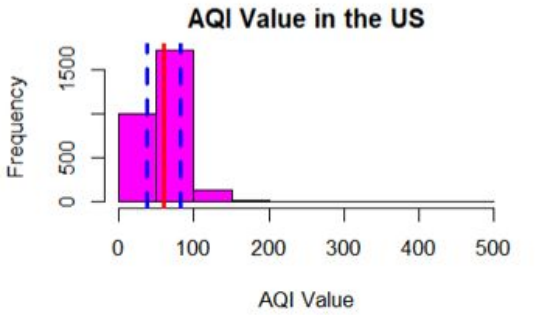
Barplot: World Air Quality Index

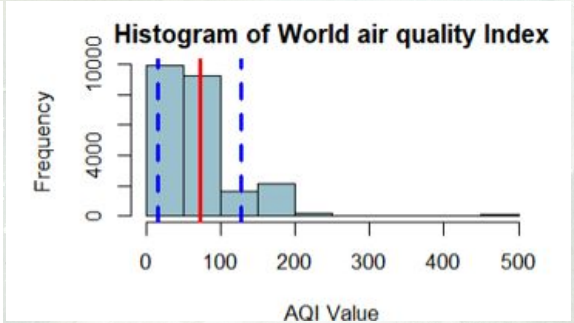


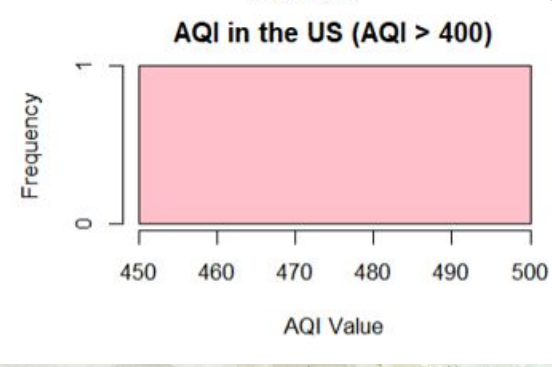
Analysis:

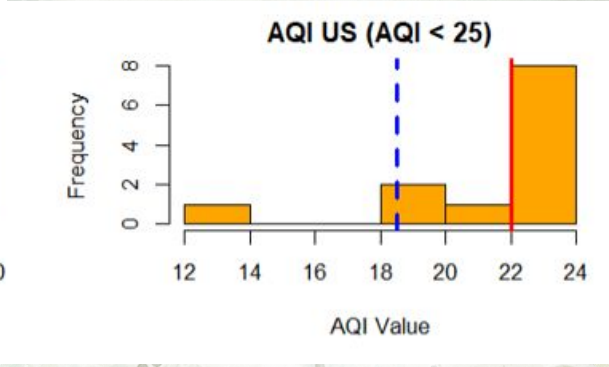
The blue dotted lines denote one sd above and below the mean denoted by red. This right skewed histogram tells us most countries fall within a AQI value range that is not considered hazardous.

Barplot: US Air Quality Index





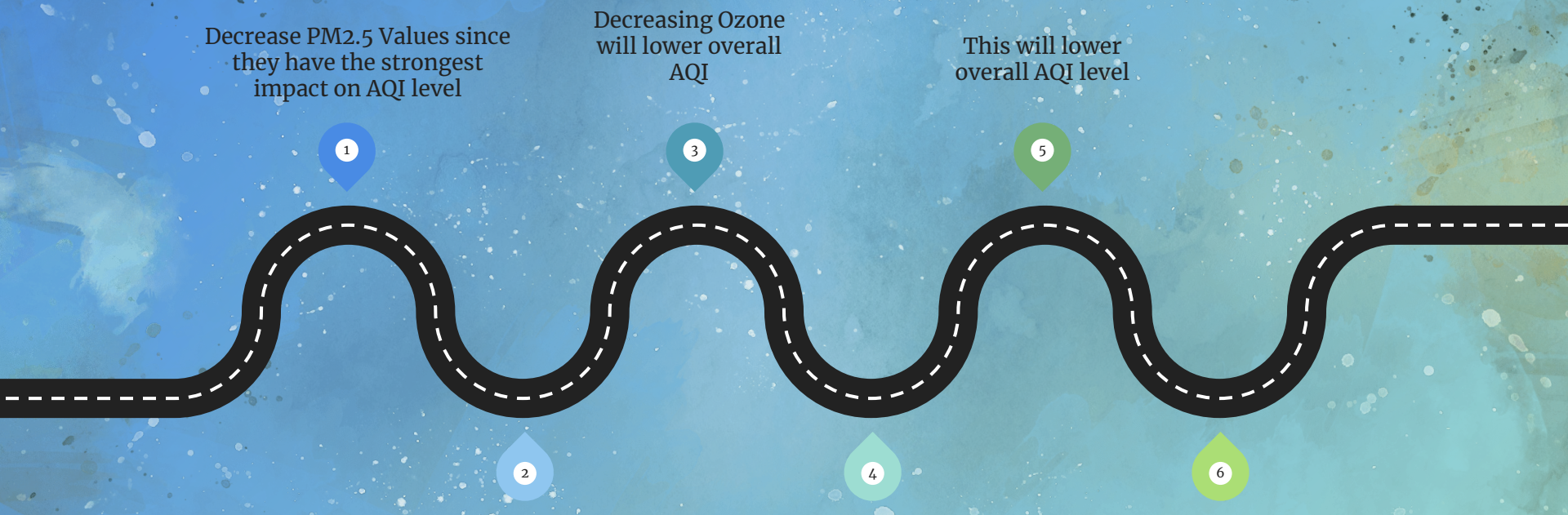




Analysis:

The US should not worry about their Air quality level, since the cities where air quality is hazardous is uniformly distributed. This means there are a minute number of cities in the US that should worry about their Air quality level. Overall, the US has a right skew distribution, with most values falling at or below 100. This is good news for the US, since most cities in the US are not categorized as having a hazardous air quality according to our data set.

Roadmap to improve air quality levels



Looking to the future

- Focus on lowering PM2.5
- Focus on lowering NO2

To improve the air quality value in the US we should work to lower PM2.5 and NO2 levels since they have the most impact on overall AQI value.

Conclusion

After analyzing the data set we have concluded that the US does not need to worry about their Air quality. While some cities in the US have an issue with air quality, we recommend addressing this by lowering the overall NO2 and PM2.5 levels as can be seen in our regression models.