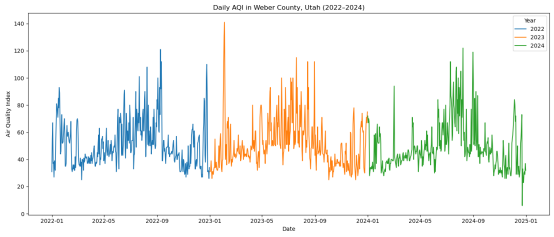
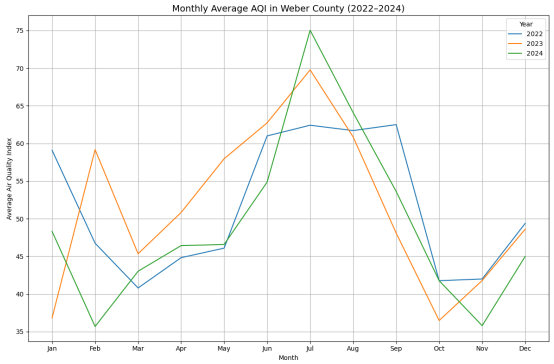
Math6450_Assignment3

October 13, 2025

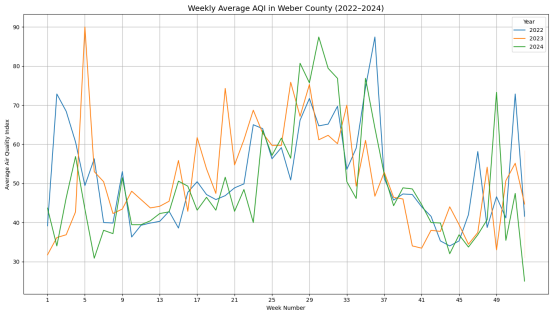
Exploratory Data Analysis (EDA): The data was loaded in and filtered to remove all columns except for State Name, and county Name. Then additional steps were taken to only grab rows with Utah, and Weber in their state and county name. 2022 had 365 rows, 2023 had 365 rows, and 2024 had 366 rows. This meant that there was a valid data entry for every single day, and this was proven to be true since there were no missing or NA values in these datasets.



The daily AQI shows that there is quite a bit of fluctuations day to day, and that each year follows a similar pattern.



Using the monthly average of the AQI shows a smoother line and a bigger picture view to how the data has changed over the years. However, it smoothed the line too much and it may not be able to capture short-term AQI fluctuations and changes.



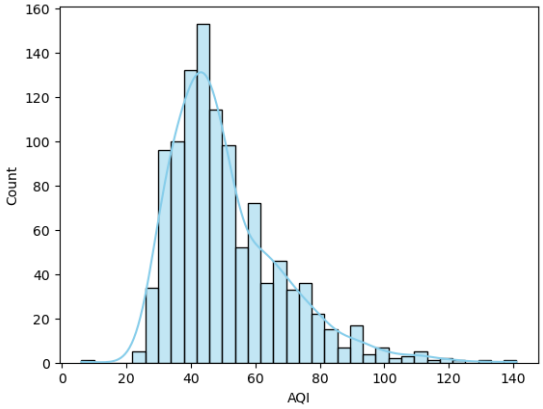
Using a weekly average for AQI seems to be the most promising due to the available data, and the graph isn't so smoothed over like it was when using monthly average AQI. This seems to be a nice middleground between daily, and monthly AQI averaging.

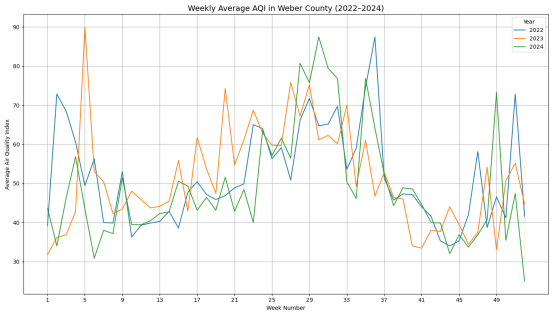
Reason for Aggregation (if applicable): Upon initial data exploration, while there is enough data for daily time series modeling, the distribution was skewed right, and the daily trends were quite jagged. So, when graphing both monthly, and weekly averages the monthly data looked a little to smooth, and weekly data looked to be the most promising since there are still obvious seasonly patterns and the data didn't look so smoothed over it wouldn't capture other trends.

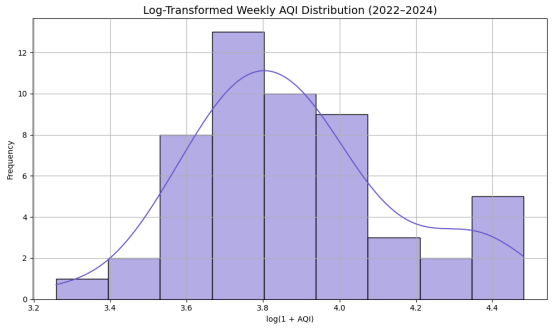
So, in an attempt to capture as much AQI patterns as possible this weekly model will capture less noise than a daily one, and more detail than a monthly one.

The aggregation performed gathers the weekly AQI for each dataset, and uses that as the new

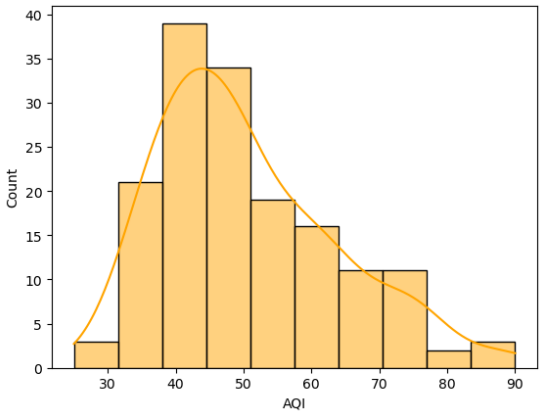
time series column.







The daily AQI variance is a little skewed to the right, and so aggregation will be a good idea for making the variance more normal.



The variance is not as prevalent, but still there, so we will move to log transformations to ensure the data is staionary.

Transformations:

```
ADF Test - Weekly AQI AVG w/ Log␣
↪Transformation
ADF Statistic: -1.1609283961297714
p-value: 0.6901735830954657
```

ADF Test - Weekly AQI AVG w/ Log Transformation ADF Statistic: -1.1609283961297714 p-value: 0.6901735830954657

These results show that the aggregation and transformation wasn't enough to make my series data stationary, so we will try first differencing before we look at seasonl patterns.

Seasonal Patterns:

Autocorrelation:

Correlation Coefficients:

Data Splitting:

Model Selection:

Model Parameters and Diagnostics:

Final Model Equation:

Forecasting:

Model Performance:

Conclusion: