Jene' Rigaud
ME 4953.003
Dr. K. Bhaganagar
11 May, 2021

Final Project Report

Link to data: https://search.earthdata.nasa.gov/search/granules?p=C1337992250-ORNL_DAAC&pg[0][gsk]=-start_date&g=G1422958404-ORNL_DAAC&q=vemap%20alaska&tl=1602683497!4!!&m=62.57812500000001!-149.625!1!1!0!0%2C2

Dataset Information

This dataset was found on NASA's Earth Data library, and originated from the National Center for Atmospheric Research in Boulder, Colorado. It contains the historical climate information of Alaska based on an annual 1% increase in greenhouse gases. Based on this historical data, future anticipated data is also included. This dataset begins on 01 January, 1997, and has a time step of 1 day between each data point. The climate information in this dataset includes the variables of irradiation, precipitation, minimum and maximum temperature, relative humidity, vapor pressure, irradiation, solar radiation, as well as the cartesian coordinates and bounds for each line of data. This dataset gives a look at the Alaskan weather conditions that can be expected for the next 80 years, and from it, many observations and assumptions can be made.

Hypothesis

It can be hypothesized that the location of each datapoint has an effect on each variable. It's expected that the average temperature will decrease, while the amount of precipitation, solar radiation, irradiation, vapor pressure, relative humidity, and ambient dewpoint will all increase with increasing latitude values. It's also hypothesized that throughout each of the year, the average temperatures will peak during the summer and spring seasons, and drastically decrease during the last half of the year. It's also assumed that with the 1% increase in greenhouse gasses each year, there will be an increased amount of solar radiation over time, resulting in higher average temperatures, less precipitation, greater humidity, and an increased vapor pressure over time. To solve this, a correlation will be conducted.

Methods

Modules Used

To analyze this data, the necessary modules had to be imported. For this project, numpy, pandas, matplotlib were used. In addition, the datetime module, seaborn, metpy, cartopy, xarray, and scipy were also used for the more complex data analysis requirements. To avoid overloading Jupyter Notebook, only a few specified functions were imported from the additional modules.

Data Reduction + Preparation

Firstly, xarray was used to open the NetCDF4 file and convert it to a dataframe. Once a necessary datetime conversion took place on the time coordinate, the dataframe was then converted to a working pandas dataframe. The original dataframe contained 7,7547,904 rows and 14 columns, so filtering the data was necessary. First, all rows containing Nan values were removed from the data. Next, scipy was used to remove all outliers from the data that were more than 3 standard deviations from the average value in each column. Once this was completed, the dataframe's length was reduced from 7547904 rows down to 3,642,126. This data reduction shows that the original data contained repeated instances, multiple Nan values, and many outliers. Although the amount of data was nearly cut in half, this ensures that the analysis will produce more accurate results.

Once the data was reduced to a workable size, new variables were created. Given the minimum and maximum temperatures or each month since January 1, 1997, a new "Avg. Temp" column was calculated. This was necessary for the ease of plotting and creating the remaining 2 new variables. Using metpy.calc, the heat index and ambient dewpoint were calculated and added to the dataframe.

Correlation Analysis

With the dataframe complete, the first correlation analysis was conducted using the DataA.corr() function. To visually compare the correlation of each variable against another, a seaborn heatmap was generated. Next, scatter plots were made for the average temperature, relative humidity, irradiation, solar radiation, precipitation, vapor pressure, ambient dewpoint, and the heat index against the varying latitude coordinates of the dataset. Additionally, using metpy and cartopy, a makeshift heatmap was created using the latitude and longitude coordinates, a color map, and the specified variables. This heatmap allowed for a visual understanding of the effect of the location when it comes to Alaskan Weather.

Time Series Analysis

Next, a time series analysis was conducted on the dataframe. Because the time coordinate was converted to a datetime format at the start of handling this data, it allowed for the simplistic grouping of variables. To do this, the ".reset_index()" command was called to remove each indices and allow for grouping. Once complete, the dataframe was grouped using the "Yearly_Mean = DataA.groupby([DataA['time'].dt.year]).mean('Avg_Temp')" and a separate dataframe was created. Additionally, another dataframe based on the Monthly mean was also created. The solar radiation, average temperature, amount of precipitation, relative humidity, and the vapor pressure was then plotted to see how their averages fluctuated throughout a single year as well as over the next 80 years.

Results

From the first correlation analysis for each variable against another, strong correlaions appear to be present at multiple instances. The Avg Temp appeared to correlate with many other values, but this was expected as the temperature plays a major roll in weather conditions. Additionally, a strong correlation occurred between solar radiation and irridiation, as well as the dewpoint and average temperature, vapor pressure, and minimum temperature. The original correlation analysis can be seen in Figure 1.

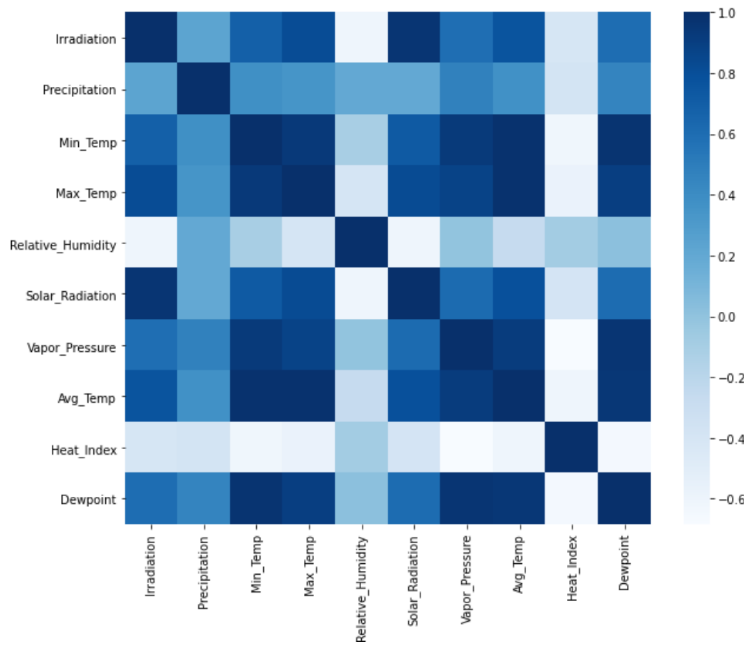

Figure 1:Variable Correlation Analysis

For the correlation between the latitude values and the selected variables, the average temperature, relative humidity, irradiation, solar radiation, precipitation, vapor pressure, and the ambient dewpoint had an overall decrease with increasing latitude coordinate values. However, the heat index increased with higher values (See figures 2-8).

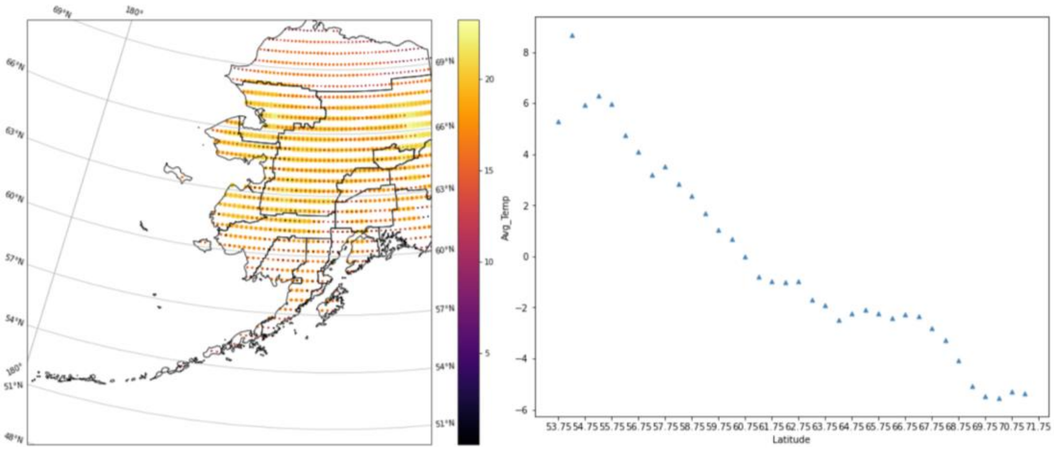

Figure 2:Average Temperature v. Latitude

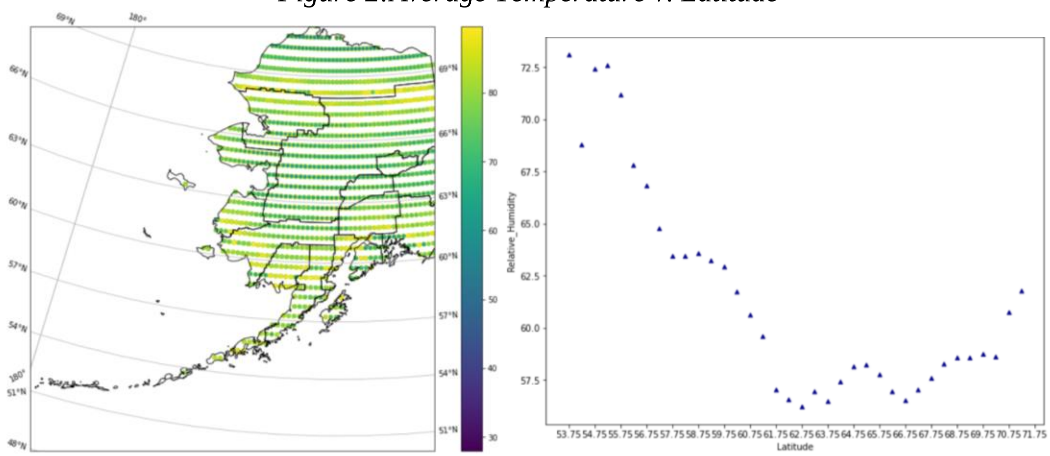

Figure 3: Relative Humidity v. Latitude

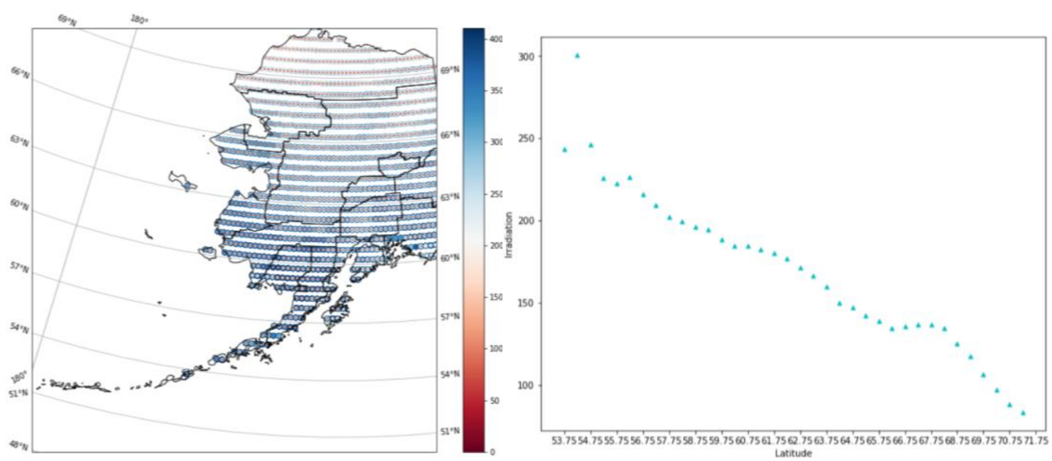

Figure 4: Irradiation v. Latitude

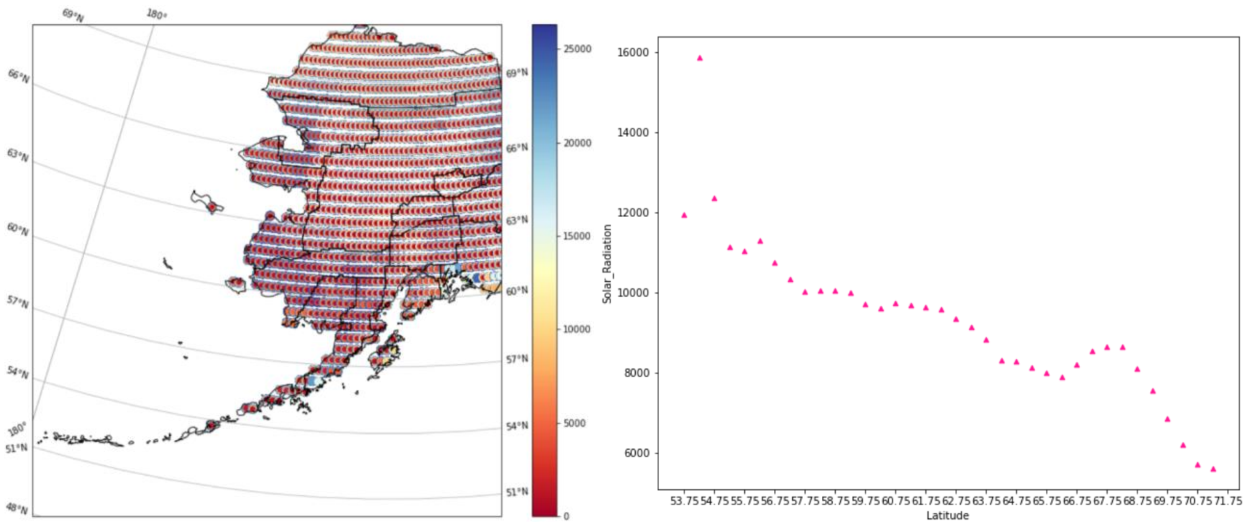

Figure 4: Solar Radiation v. Latitude

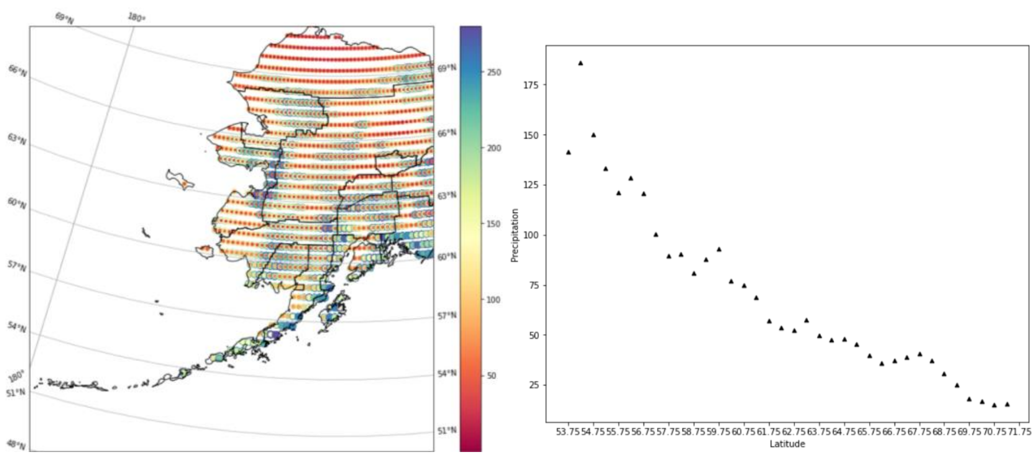

Figure 5: Precipitation v. Latitude

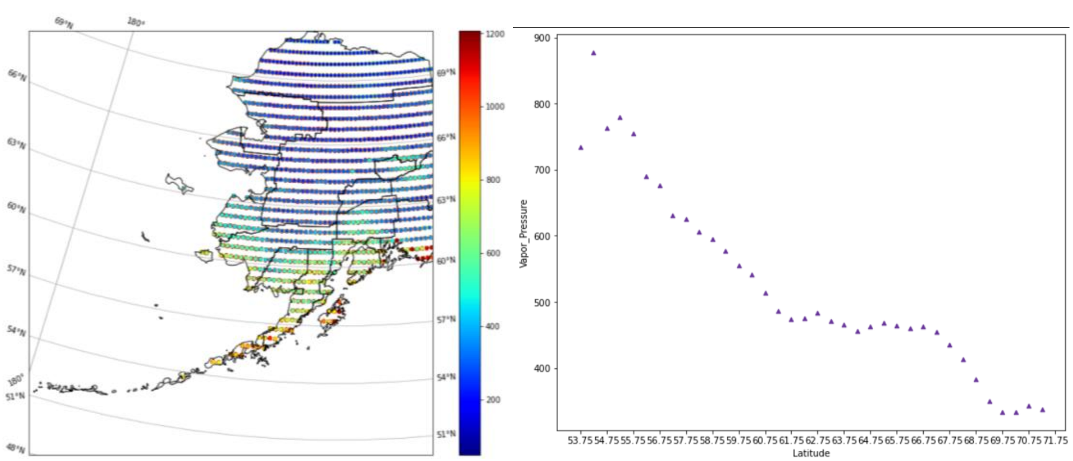

Figure 6: Vapor Pressure v. Latitude

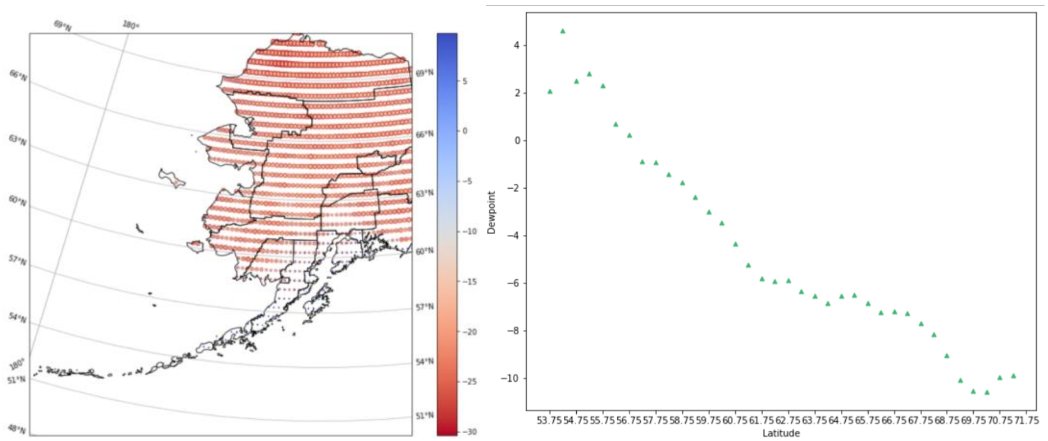

Figure 7: Dewpoint v. Latitude

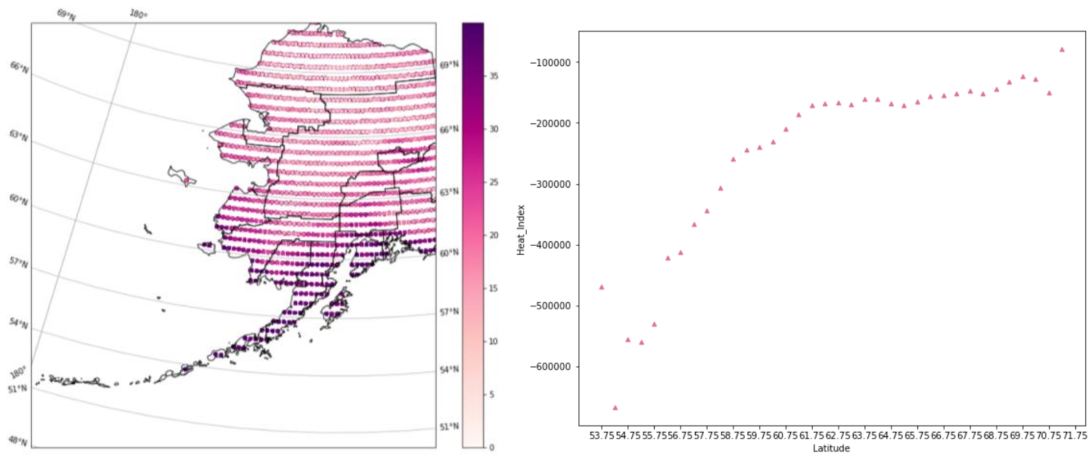

Figure 8:Heat Index v. Latitude

Upon analysis, it was found that the solar radiation had an overall decrease with increasing year, while the average temperature, relative humidity, and vapor pressure decreased. There was a slight decrease in precipitation with respect to the year, but overall there was no notable trend. This analysis can be observed in Figures 8-12.

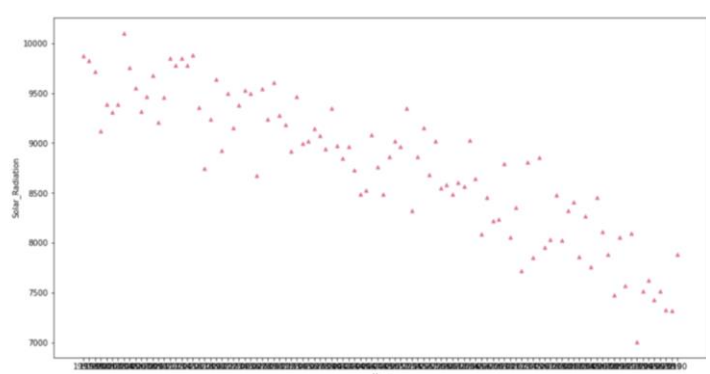

Figure 8:Solar Radiation v. Year

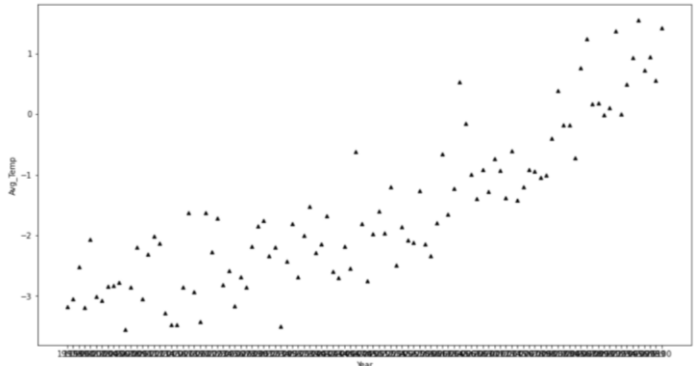

Figure 9: Average Temp v. Year

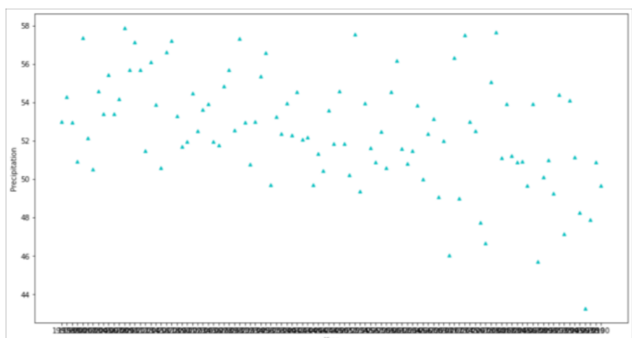

Figure 10: Precipitation v. Year

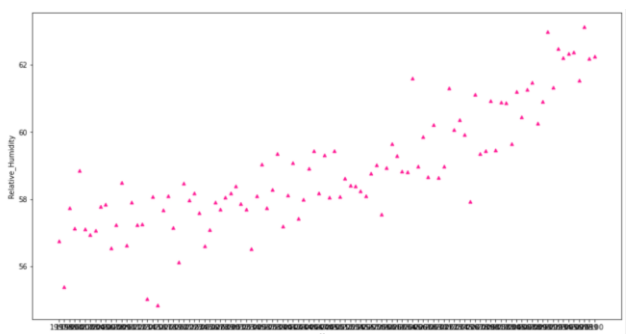

Figure 11: Relative Humidity v. Year

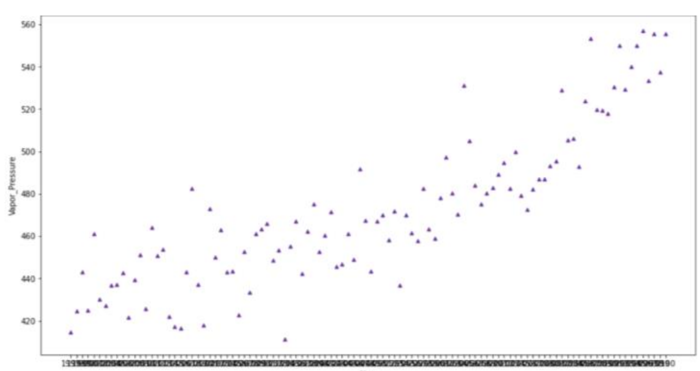

Figure 12: Vapor Pressure v. Year

For the monthly time analysis, it’s observed that due to the absense of sun during the colder seasons in Alaska, there is a decrease in solar radiation, average temperature, precipitation, and vapor pressure. However, the humidity spikes during the last half of the year.

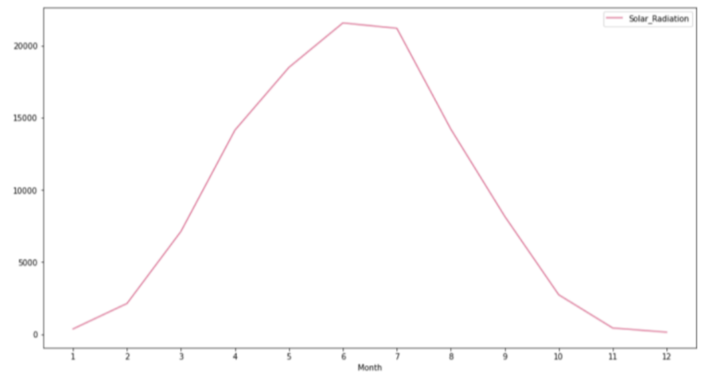

Figure 13:Solar Radiation v. Month

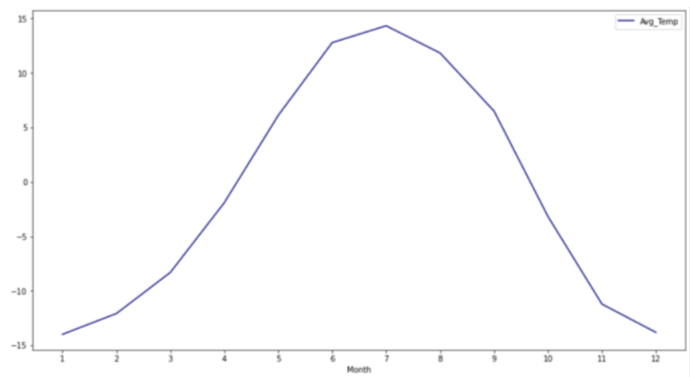

Figure 14: Average Temperature v. Month

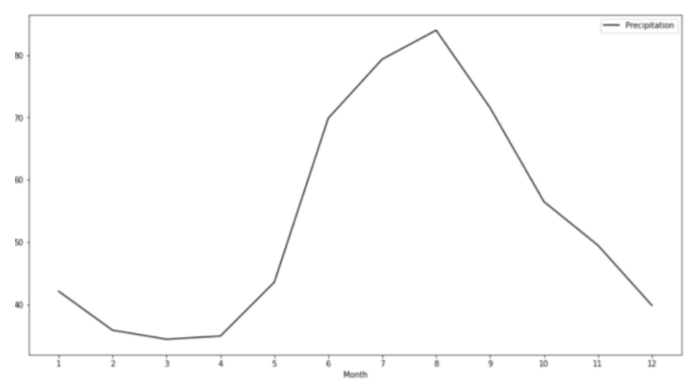

Figure 15: Precipitation v. Month

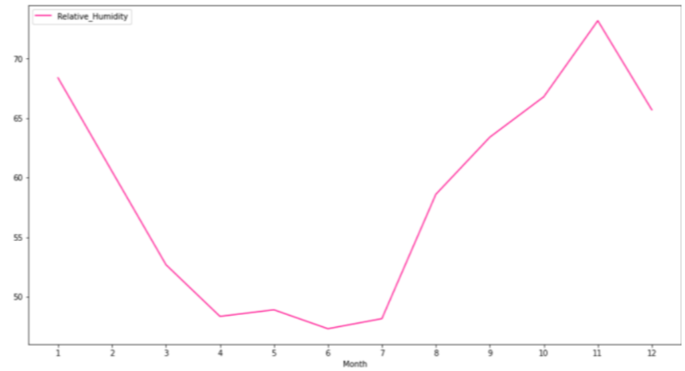

Figure 16: Relative Humidity v. Month

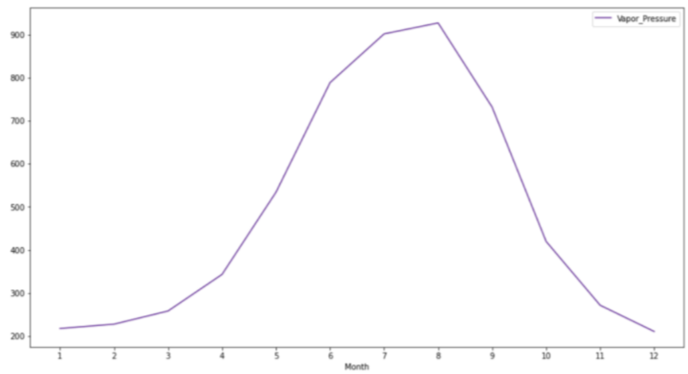

Figure 17: Vapor Pressure v. Month

Conclusion

From this analysis, it's been observed that the latitude coordinate does in fact have an effect on the many components of weather. The average temperature, relative humidity, irradiation, solar radiation, precipitation, vapor pressure, and the ambient dewpoint increase with increasing latitude values, while the heat index decreases. It was also found that throughout an average year, Alaska's absence of sun causes a drop in the vapor pressure, average temperature, solar radiation, and precipitation, but a spike in the relative humidity.