Albany New York Weather

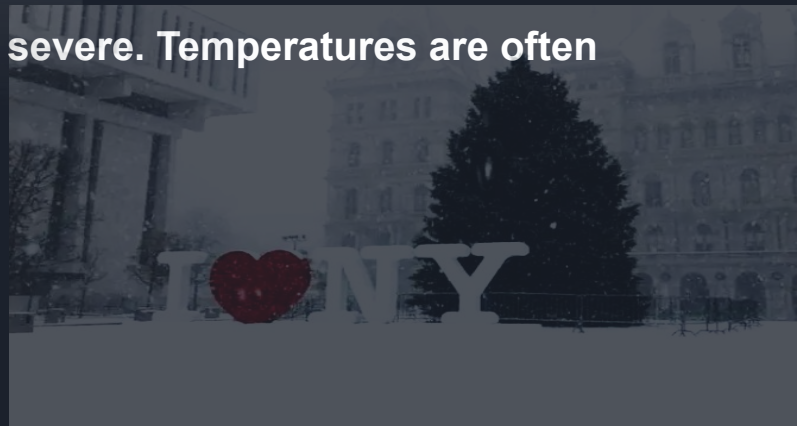
Contributors- Emily Turner, Sophie Seibure, Conner Dekok, and Oumar Diakite.

Objective of our dataset

- Analyze weather trends or patterns over a period of time.
- Find any discrepancies in the weather and possibly formulate some reasons to why.
- We examined various weather events such as precipitation, snowfall, and overall temperature changes over a seven year period (2015-2022).

Albany, New York has a primary continental climate with some influence from the Atlantic Ocean.

- The average annual temperature for Albany varies between 48.9 F and 50.6 F, as of 2021 and 2022.
- Warmer seasons include quick rising temperatures during the day that fall quickly after sunset.
    - Some periods of oppressive heat last for a week at a time.
- Winters are usually very cold and sometimes severe. Temperatures are often below freezing and below 10F° at night.
    - Snowfall throughout winter is variable.



Completing Analysis - Cleaning data

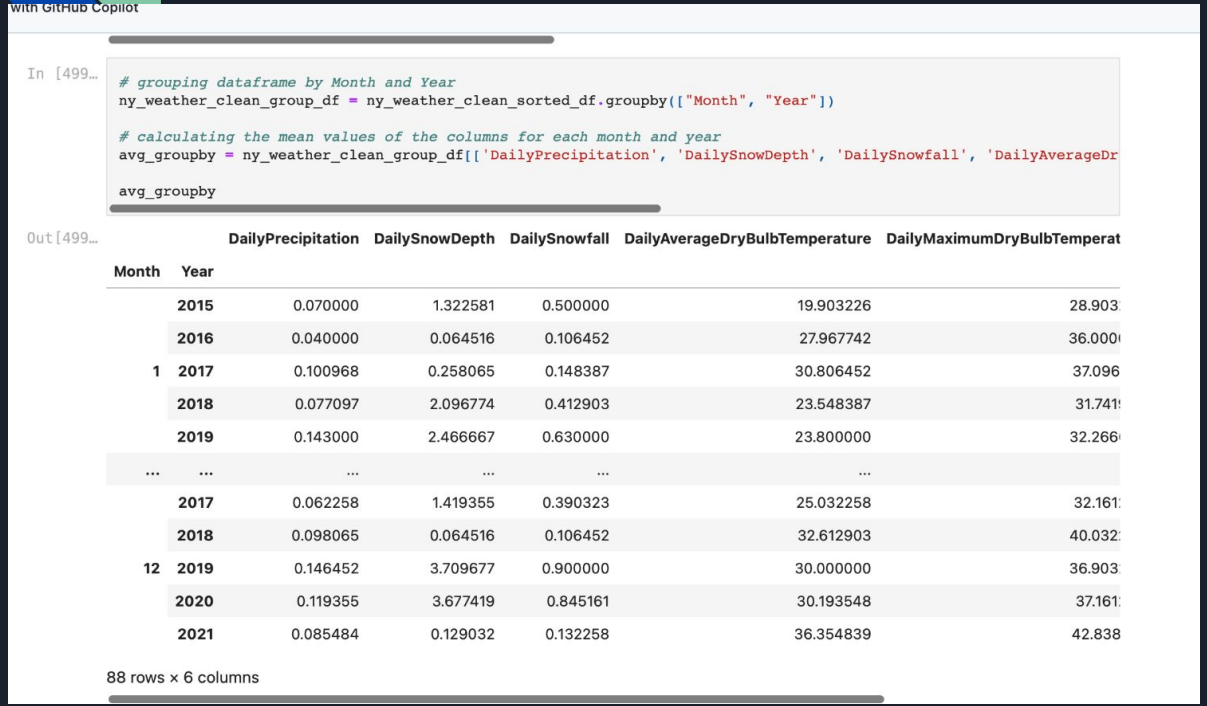
Out[498...

	DATE	DailyPeakWindDirection	DailyPeakWindSpeed	DailyPrecipitation	DailySnowDepth	DailySnowfall	DailySustainedV
0	2015-01-01 23:59:00	190.0	26.0	0.00	0.0	0.0	
1	2015-01-02 23:59:00	250.0	30.0	0.00	0.0	0.0	
2	2015-01-03 23:59:00	170.0	21.0	0.57	0.0	1.6	
3	2015-01-04 23:59:00	290.0	33.0	0.22	1.0	0.0	
4	2015-01-05 23:59:00	280.0	42.0	0.00	0.0	0.0	
...	...	...	...	...	...	...	
2663	2022-05-27 23:59:00	160.0	28.0	0.00	0.0	0.0	
2664	2022-05-28 23:59:00	310.0	26.0	0.04	0.0	0.0	
2665	2022-05-29 23:59:00	90.0	13.0	0.00	0.0	0.0	
2666	2022-05-30 23:59:00	200.0	15.0	0.00	0.0	0.0	
2667	2022-05-31 23:59:00	250.0	29.0	0.00	0.0	0.0	

2668 rows × 14 columns

- Selecting necessary columns
- Replacing “T” values or trace values (precipitation less than .005 in) with 0
- Changing data types to datetime & floats
- Created columns of year, month, & day to help with filtering

Completing Analysis - Calculating averages



With GitHub Copilot

In [499…

```
# grouping dataframe by Month and Year
ny_weather_clean_group_df = ny_weather_clean_sorted_df.groupby(["Month", "Year"])

# calculating the mean values of the columns for each month and year
avg_groupby = ny_weather_clean_group_df[['DailyPrecipitation', 'DailySnowDepth', 'DailySnowfall', 'DailyAverageDr

avg_groupby
```

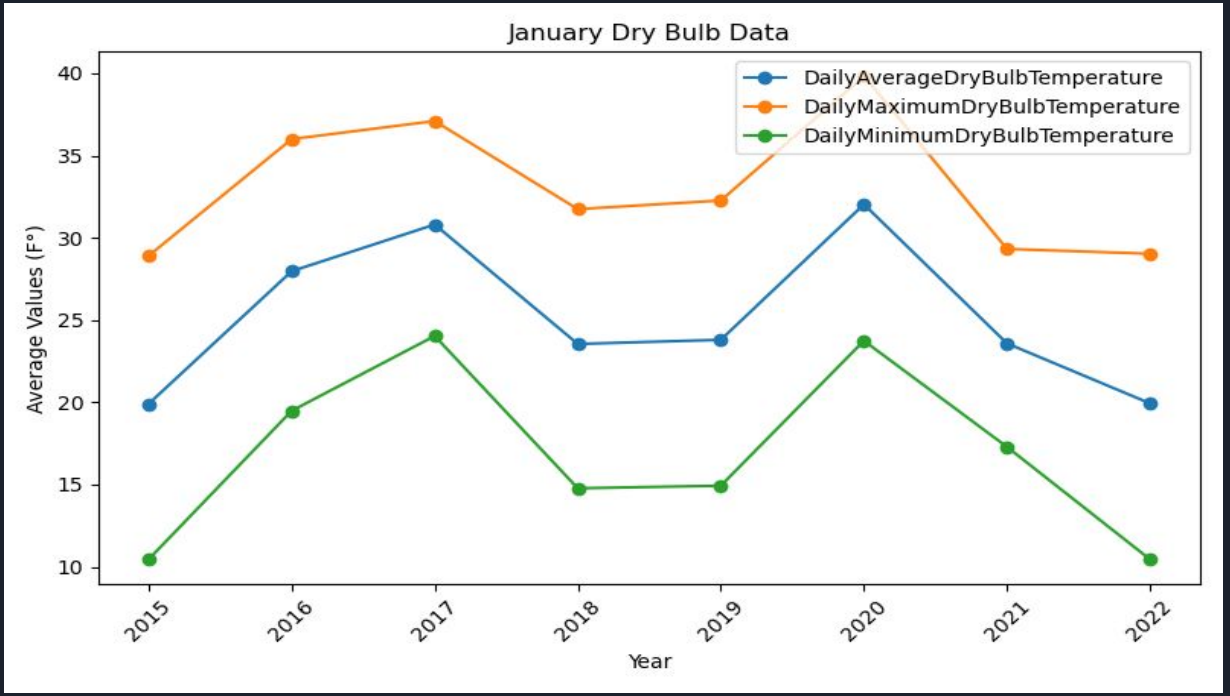
Out[499…

Month	Year	DailyPrecipitation	DailySnowDepth	DailySnowfall	DailyAverageDryBulbTemperature	DailyMaximumDryBulbTemperat
	2015	0.070000	1.322581	0.500000	19.903226	28.903
	2016	0.040000	0.064516	0.106452	27.967742	36.000
1	2017	0.100968	0.258065	0.148387	30.806452	37.096
	2018	0.077097	2.096774	0.412903	23.548387	31.741
	2019	0.143000	2.466667	0.630000	23.800000	32.266
…	…	…	…	…	…	
	2017	0.062258	1.419355	0.390323	25.032258	32.161
	2018	0.098065	0.064516	0.106452	32.612903	40.032
12	2019	0.146452	3.709677	0.900000	30.000000	36.903
	2020	0.119355	3.677419	0.845161	30.193548	37.161
	2021	0.085484	0.129032	0.132258	36.354839	42.838

88 rows × 6 columns

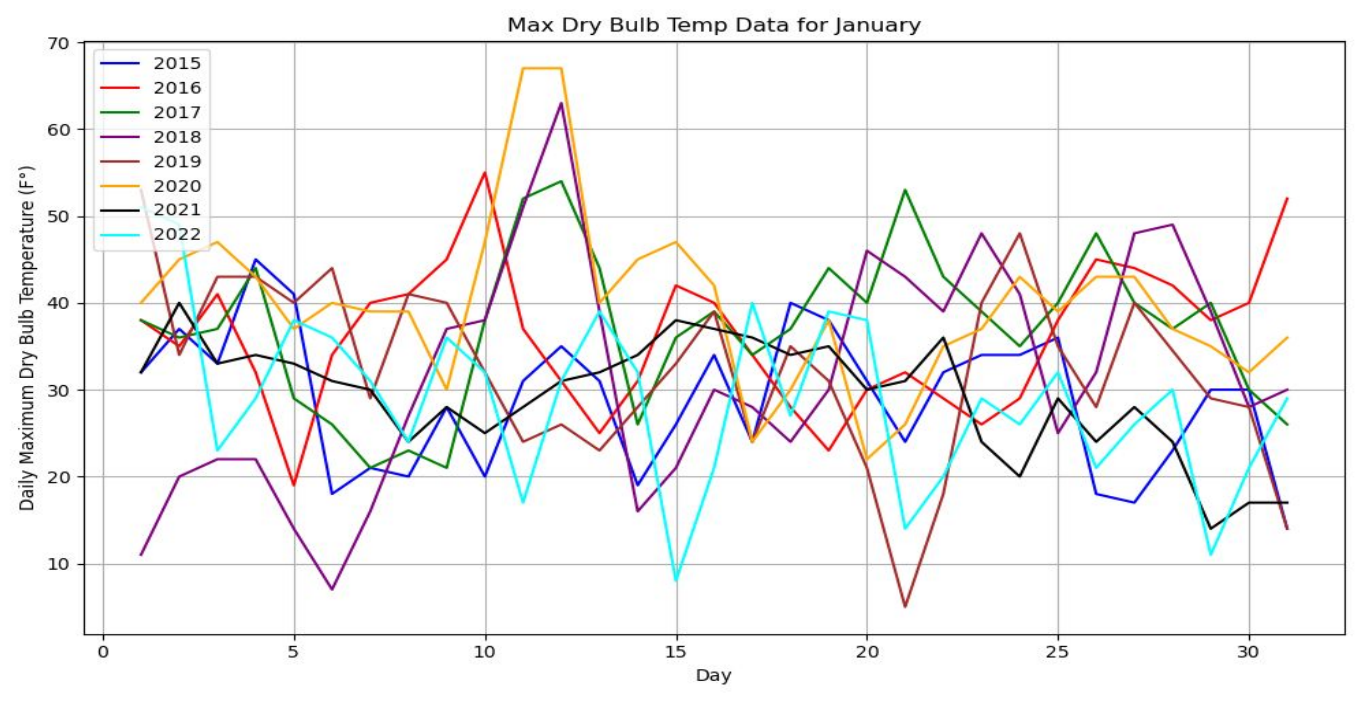
- SQLite used to store and extract clean data
- The averages for columns calculated using group by
- Using .iloc and .loc to filter data

January dry bulb data



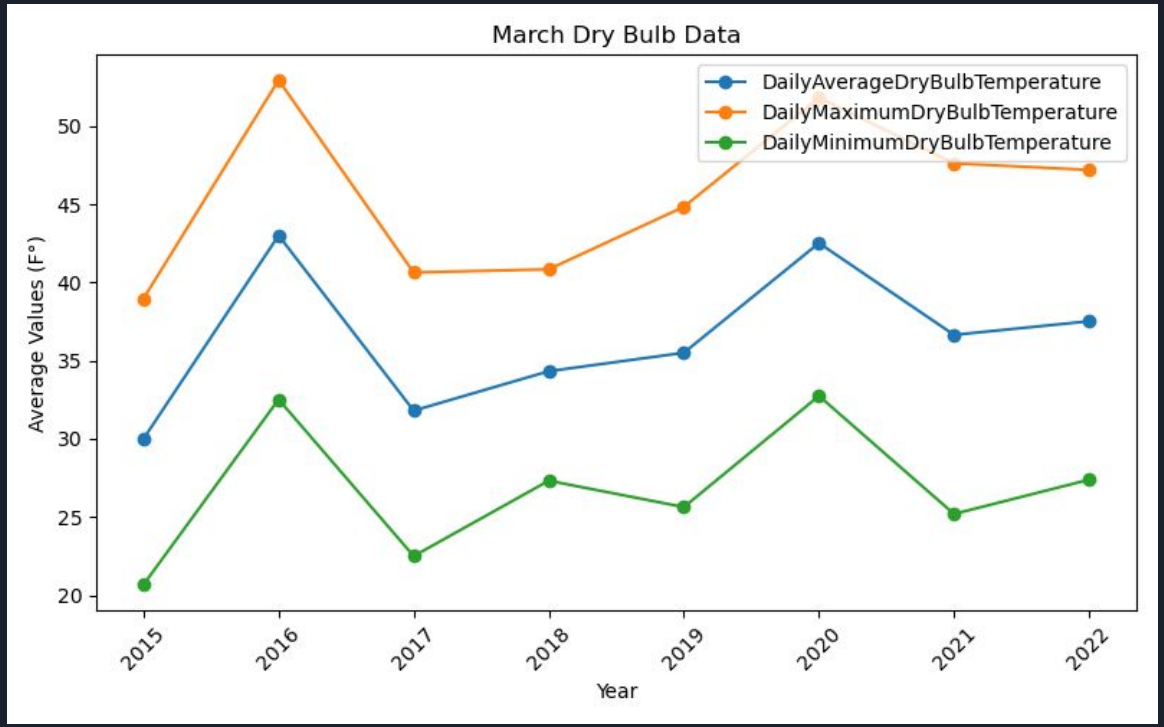
- The dry bulb temperatures for 2015 and 2022 was lower than the other years listed here. This could have resulted in a variety of other weather patterns throughout those years from the colder weather.

Max temperatures for January



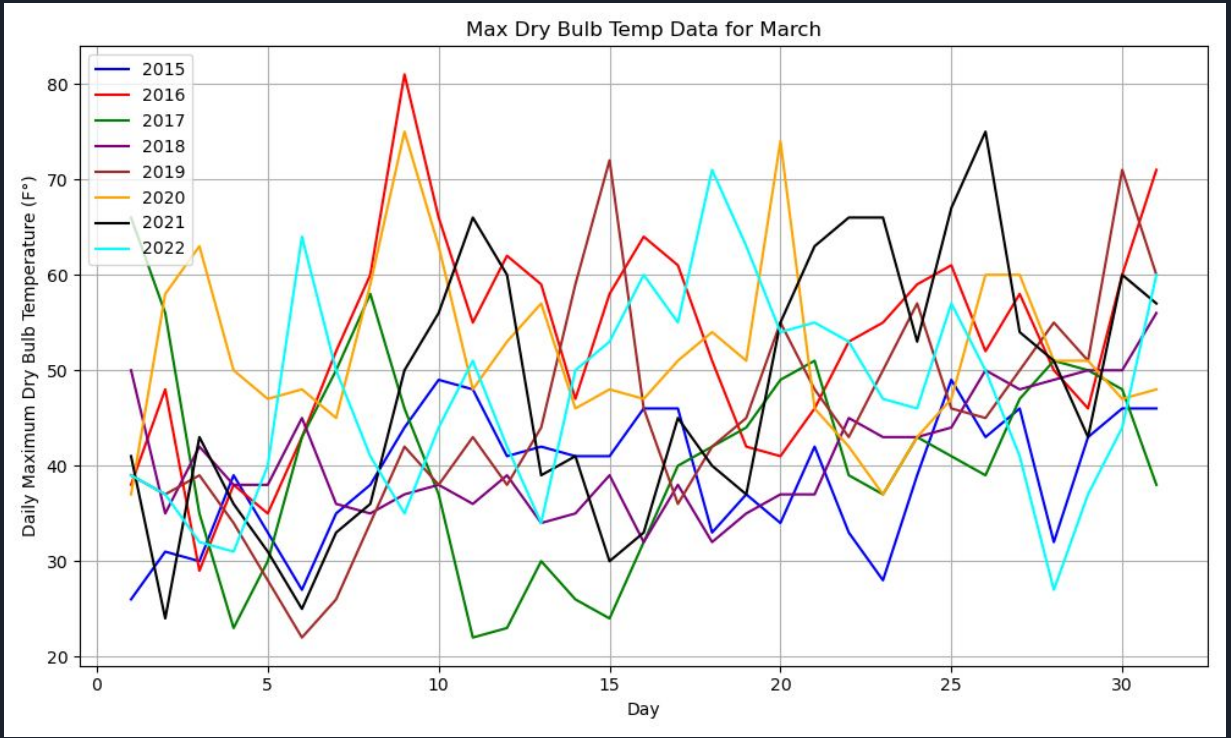
- Examining the max temperatures for January in Albany, New York this helps us see that 2020 and 2019 had a high spike around the same time compared to other years.

March dry bulb data



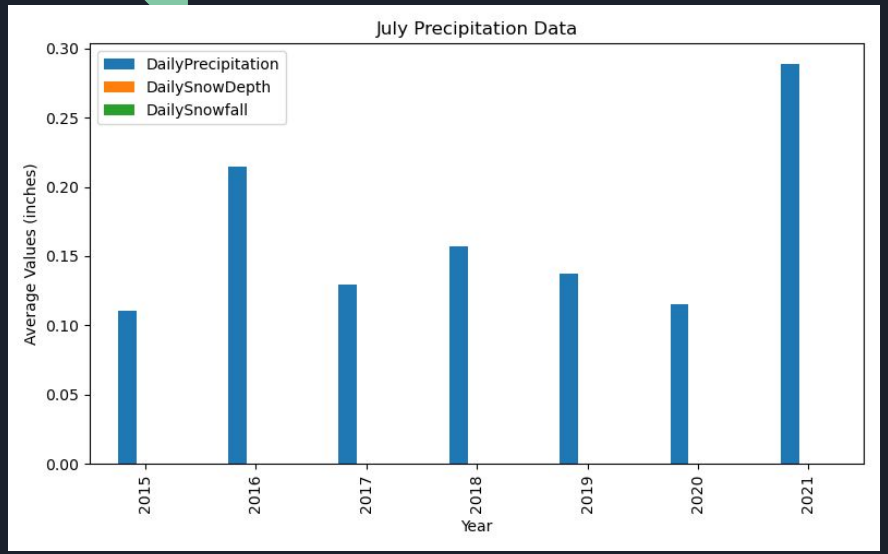
- This line chart helps show March dry bulb temperature for a total of 7 years in Albany, New York where it seems that in 2016 the all of the temperature values were higher than the other years.

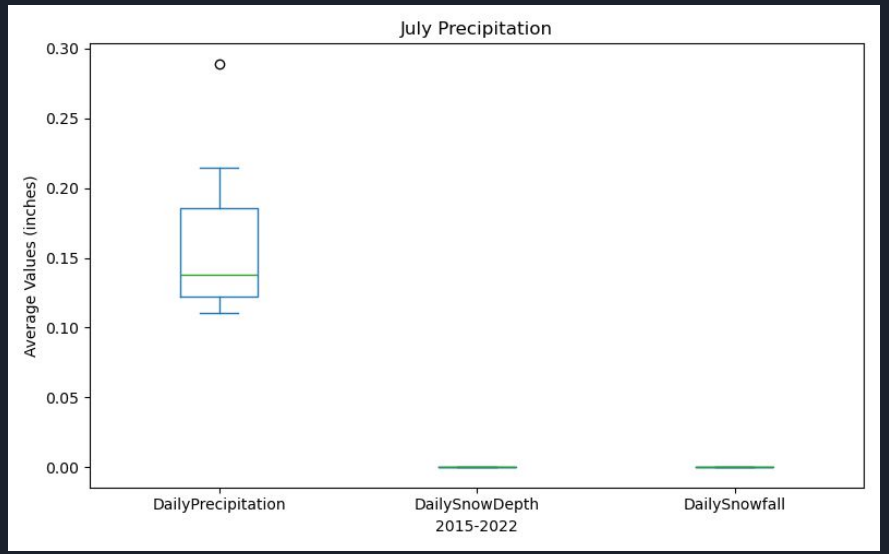
Max temperatures for March



- The March dry bulb temperatures had a few outliers when it came to temperature such as 2016 and 2020 having some spikes in temperature, and having a few low points with 2017 and 2019.

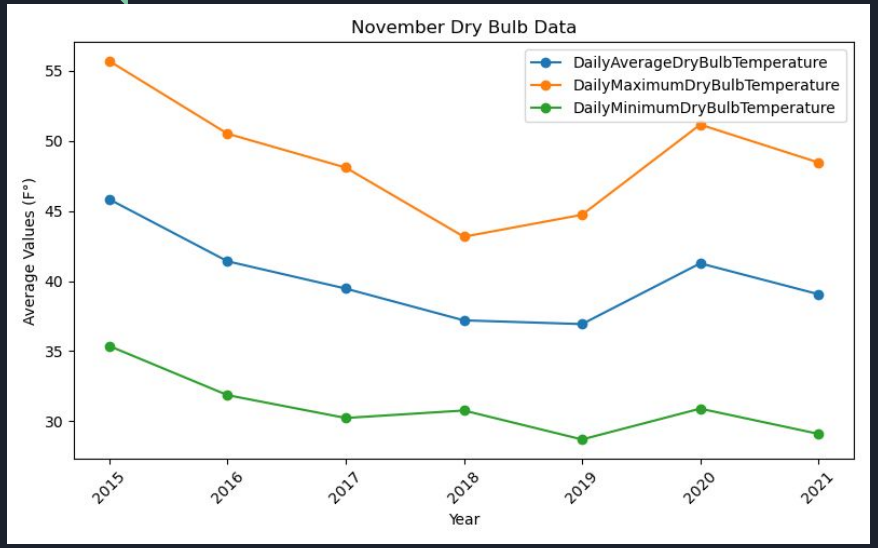
July precipitation and outliers

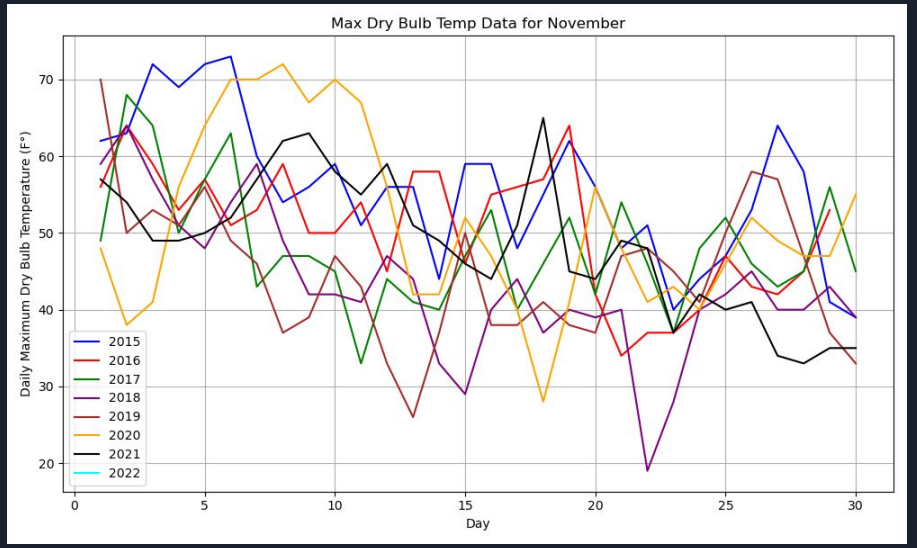




- The conjunction of these charts help show what year would have been the outlier in the July temperature data. In this case 2021 was the outlier of all of the years when it came to regular precipitation.

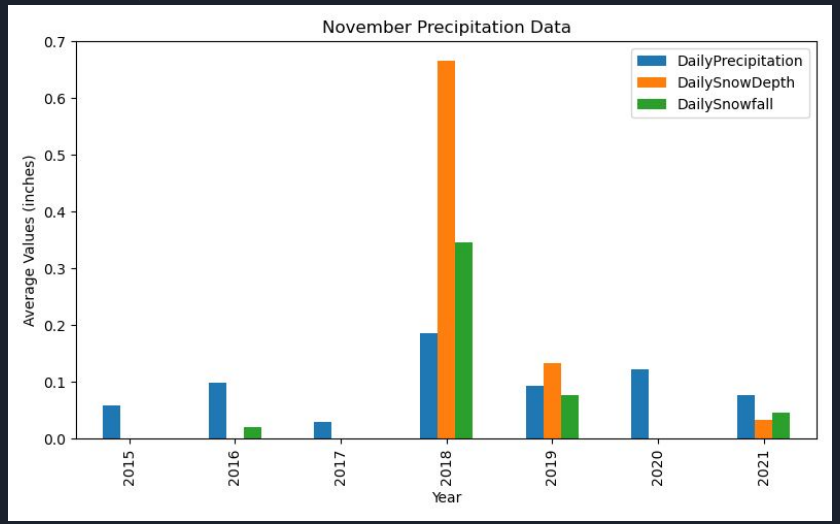
November averages and outliers

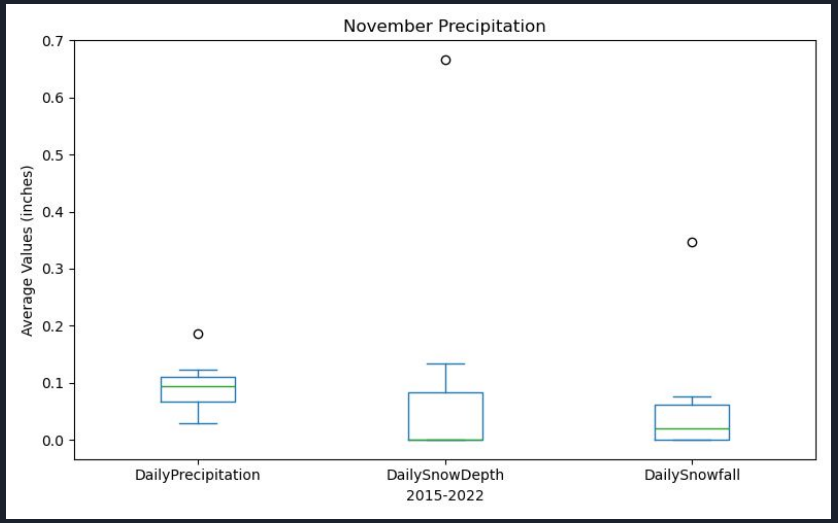




- These line charts help show how the dry bulb temperatures varied over time. For example, 2018 had a decrease in temperature compared to the other years.

November precipitation and outlier values





- These graphs help show where the data was varied when it came to differences in snow depth and total snowfall. These charts help show for 2018 that there was more snow than other years.

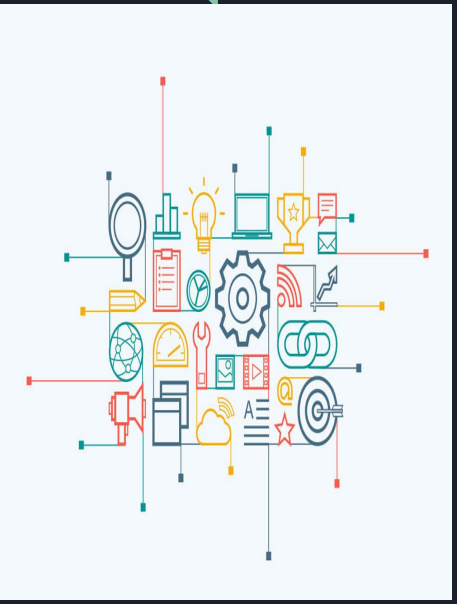
Data Ethics

- Legal Issues with Using Weather Dataset
  - Licensing restrictions
    - Ensure compliance with license agreements
    - Restrictions may apply for commercial use
  - Misrepresentation of Data
    - Inaccurate data representations
    - Misuse of weather data to influence markets or policies
  - Data Ownership
    - Verify proper authorization to use the dataset

Ethical issues with Using the Dataset

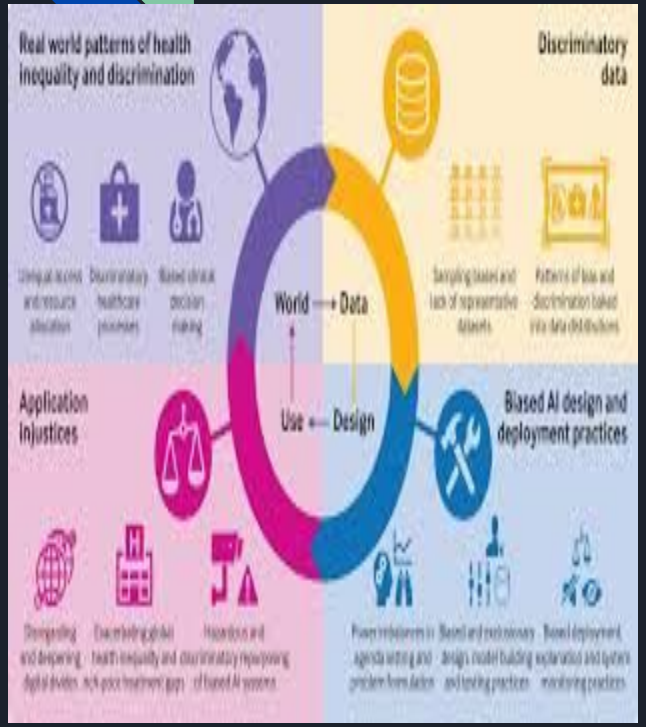
- Misleading Visualizations:
  - Avoid visual representations that exaggerate or minimize weather data trends.
  - Inaccurate charts or graphs can mislead decision-makers or the public.
- Public Policy and Impact:
  - Using weather data to influence public policy.
  - Misuse can harm communities or skew policy-making decisions.
- Privacy Concerns:
  - When combined with other datasets (e.g., geolocations, population data), weather data could unintentionally violate privacy.

Algorithmic Bias in Weather Data:



- Weather data is often used in decision-making systems (e.g., predicting energy demand, flood risks, disaster planning).
- Algorithmic bias can lead to inaccurate forecasts, which disproportionately harm certain communities (e.g., low-income areas).
- Biased models can exacerbate inequalities by failing to account for regional or socio-economic differences

Common Causes of Algorithmic Bias in Weather Data



- Historical Data Bias:
  - Algorithms trained on historical weather data may overlook changes in weather patterns due to climate change.
- Regional Inequities:
  - Weather stations are unevenly distributed, leading to better data in wealthier regions and biases in less-covered areas.

Best Practices

- Legal Considerations:
  - Always review licensing terms and ensure you have permission to use the dataset.
  - Ensure accuracy and avoid misrepresenting weather data in reports or visualizations.
- Ethical Considerations:
  - Be transparent in how the data is used, especially when influencing policy or public decisions.
  - When combining datasets, consider privacy risks and ensure ethical use of the data

Conclusion

- By analyzing the weather data of Albany, New York over a period of time we were able to find a variety of interesting trends.
- These fluctuations and trends could be a result of many issues such as random weather patterns or the overall changing of the climate due to climate change.
- Continuing to collect and analyze weather data will allow long term changes in weather patterns to be examined.