Haleakalā Viewing Conditions

DAI Capstone II Project Proposal

Submitted by Richy Peterson

Objective

The Air Force Maui Space Surveillance Complex (MSSC) is located at the summit of a dormant volcano, Haleakalā, on Maui Island, Hawaiʻi. The location is one of the best in the world for both daytime and nighttime sky conditions due to its height of 10,023 feet and relatively good weather year-round. Despite this, bad weather routinely stops operations, disrupting collection campaigns that were often scheduled months in advance. There are three categories of weather for MSSC operations, red, yellow, and green. Weather that prevents telescope operation is called red weather. Weather that degrades performance, but does not stop operation is called yellow. Green weather refers to normal operating conditions. The purpose of this analysis is to determine if there is a specific month of the year that provides significantly more green weather hours per day. This could help inform researchers when to plan collection campaigns that are most sensitive to schedule changes.

The Data Set

The data that will be used for this analysis comes from the Institute for Astronomy (IfA) weather archive. The IfA has Haleakalā weather data archived in csv files for each year from 1994-2021. There are 13 columns in each file and the seven of interest to this analysis are shown below. These columns are of interest because they are weather measurements that tie to operating threshold ranges.

Date/Time	Temperature (°C)	Humidity (%)	Wind Speed (m/s)	Visibility (m)	Precipitation (inches)	Dewpoint (°C)
2019-01-01 00:00:01	11.27	17	10.5	49425	nan	7.13
2019-01-01 00:00:11	11.24	15.7	9.4	48363	nan	4.46
2019-01-01 00:00:22	11.26	15.5	10.2	48363	nan	3.08
2019-01-01 00:00:32	11.25	16.2	9.6	48000	nan	3.26
2019-01-01 00:00:42	11.27	17.8	9.8	47351	nan	5.28

All columns of interest are numerical other than the date-time. Up to Nov 2006, the recorded data represents 10 min averages. Since then the raw data was recorded every 10 seconds. As a result, the number of rows of data ranges from around 50 thousand to around 3 million in each file.
In my initial look at five years (1994-1995 and 2018-2020), I found several null values. For the early years, all values for humidity, visibility, and dewpoint are null and for the later years, all values for precipitation are

null. There are also several other null values in all columns except the date-time. For the 2020 data, there is another issue where the data has formatting problems. There are extra columns at times it appears some measurements are in the wrong columns. At present, I plan to leave this set out unless I have more time to clean it.

Minimum Viable Product

My plan for this analysis is to:

1. Check all the weather measurements against the respective operating thresholds.
   To be considered green weather all conditions must be met. For small groups of null values, I will assume the state has not changed since the last recorded value. For the measurements that are missing for the entire year, I will assume the status is green.

2. Build a data frame for each year with the number of hours of green/yellow/red weather each day.

3. Combine all the data for each year, grouped by month.

4. Run a hypothesis test to check if the average hours of green weather per day are higher in one month than in another.
   The null hypothesis will be there is no significant difference between the average hours of green weather per day each month.
   There are multiple alternate hypotheses as each month will be compared to all others. As a result, a Bonferroni correction will be applied.

While there is a large amount of data planned to be included in this analysis, once step two above is reached it will reduce each year of data down to 365-366 rows. That being said the MVP will aim to include four years of data. This will give a manageable size to develop the analysis pipeline. Once the pipeline is developed additional years will be added as long as they do not have formatting (such as the year 2020) that require more cleaning.

MVP+

1. Perform additional cleaning (if feasible on further inspection) on any applicable years left out due to formating. Include those years in the analysis.

2. Divide the analysis into hours of daylight, twilight, and nighttime.
   This will require writing a web scraper to get the sunrise, sunset, and astronomical twilight times for each corresponding day. Then the data frames would need to be split according to those times. The rest of the analysis would be the same as above but the process above would be repeated 3 separate times.