Contents

[**1.** **Introduction** 1](#_Toc59972720)

[**1.1.** **Objective** 1](#_Toc59972721)

[**1.2.** **Datasets** 1](#_Toc59972722)

[**1.2.1.** **Historical Wildfires Data** 1](#_Toc59972723)

[**1.2.2.** **Historical Weather Data** 1](#_Toc59972724)

[**1.2.3.** **Historical Weather Forecast Data** 1](#_Toc59972725)

[**1.2.4.** **Historical Vegetation Data** 1](#_Toc59972726)

[**1.2.5.** **Land Class Data** 1](#_Toc59972727)

[**1.3.** **Approach Introduction** 1](#_Toc59972728)

[**2.** **Cleaning and Preprocessing** 1](#_Toc59972729)

[**2.1.** **Normalizing Data** 2](#_Toc59972730)

[**2.2.** **Handling Missing (NaN) Values** 2](#_Toc59972731)

[**2.3.** **Handling Duplicate Data** 2](#_Toc59972732)

[**2.4.** **Transforming Data** 2](#_Toc59972733)

[**3.** **Exploratory Data Analysis (EDA)** 2](#_Toc59972734)

[**3.1.** **Trends & Seasonality** 2](#_Toc59972735)

[5. Modeling 3](#_Toc59972736)

[5.1. Regression Models 3](#_Toc59972737)

[5.1.1. Applying algorithm 3](#_Toc59972738)

[5.1.2. Solution to the problem 4](#_Toc59972739)

[5.1.3. Performance of Different Models 4](#_Toc59972740)

[5.2. Classification Models 4](#_Toc59972741)

[6. Conclusions 4](#_Toc59972742)

[7. Future Directions 4](#_Toc59972743)

1. **Introduction**
   1. **Objective**
   2. **Datasets**
      1. **Historical Weather Data**

Data Elements of Weather Data:

* All variables are aggregated to daily values from YYYY-mm-ddT01:00:00Z to YYYY-mm-(dd+1)T00:00:00Z
* **`Precipitation`**
  + is derived from total precipitation.
  + Hourly raw data is converted from m/hour to **mm/hour**
* **`Relative humidity`**
  + is derived from the temperature and dewpoint
* **`Soil Water Content`** 
  + is given for 0 - 7 cm below the surface
* **`Solar Radiation`** or Surface Solar Radiation Downwards.
  + Units are converted from J/h to **MJ/h**
* **`Temperature`**
* **`Wind Speed`** 
  + is calculated for every hour from the Easterly and Northerly 10 meter wind components
    1. **Historical Wildfires Data**

Data Elements of Wildfires Data:

* **`Region`**:
  + the 7 regions - 'NSW', 'NT', 'QL', 'SA', 'TA', 'VI', 'WA'
* **`Date`:**
  + in UTC and provide the data for 24 hours ahead
* **`Estimated\_fire\_area`:** **\*\* This data element can be used to train the model with however, other elements will also be considered.**
  + daily sum of estimated fire area for presumed vegetation fires
  + with a confidence > 75% for a each region
  + **in km^2**
* **`Mean\_estimated\_fire\_brightness`:**
  + daily mean (by flagged fire pixels(=count)) of estimated fire brightness for presumed vegetation fires
  + with a confidence level > 75%
  + in **Kelvin**
* **`Mean\_estimated\_fire\_radiative\_power`:**
  + daily mean of estimated radiative power for presumed vegetation fires
  + with a confidence level > 75% for a given region
  + in **megawatts**
* **`Mean\_confidence`:**
  + daily mean of confidence for presumed vegetation fires
  + with a confidence level > 75%
* **`Std\_confidence`:**
  + standard deviation of estimated fire radiative power
  + in **megawatts**
* **`Var\_confidence`:** 
  + Variance of estimated fire radiative power
  + in megawatts
* **`Count`:** 
  + daily numbers of pixels for presumed vegetation fires
  + with a confidence level of larger than 75% for a given region
* **`Replaced`:**
  + Indicates with an Y whether the data has been replaced with standard quality data when they are available (usually with a 2-3 month lag).
  + **Replaced data has a slightly higher quality** in terms of locations
    1. **Historical Weather Forecast Data**
    2. **Historical Vegetation Data**
    3. **Land Class Data**
  1. **Approach Introduction**

1. **Cleaning and Preprocessing**

***Related Juypter Notebooks*:**

*2.1 Historical\_Wildfires – Cleaning and Preprocessing*

*2.2 Historical\_Weather – Cleaning and Preprocessing*

*2.3 Historical\_Weather Forecasts – Cleaning and Preprocessing*

***Ouput Datasets “C&P\_Datasets”***

*C&P\_Wildfires.csv*

*C&P\_Weather.csv*

*C&P\_Forecasts.csv*

* 1. **Normalizing Data**

1 – Date data type in all three files (Historical Wildfires, Historical Weather, and Historical Wildfires Forecasts) set to YYYY-MM-DD

* 1. **Handling Missing (NaN) Values**

1 – Historical **Wildfires** Data:

2207 Missing (NaN) values for two columns, Std\_confidence and var\_confidence, replaced with the mean value of each respectively

* 1. **Handling Duplicate Data**
  2. **Transforming Data**
  3. **Asfd**

1. **Exploratory Data Analysis (EDA)**

**Types of Exploratory Data Analysis performed in this project:**

* Time Series:
  1. Time series was used to plot trends / seasonality of weather parameters, month by month, year over year comparisons for each parameter: **precipitation, relative humidity, soil water content, solar radiation, temperature and windspeed.** (Jupyter notebook 3.1 A)
  2. Time series was also used to plot trends/seasonality of weather parameters, month by month, year over comparison **for each region** by each parameter. (Jupyter notebook 3.1 B)
  3. Time series was further used in looking at the wildfires data, in a very similar manner to the weather data.
  4. **Trends & Seasonality**

***Related Juypter Notebooks*:**

*3.1 A) EDA – Historical Weather – Trends & Seasonality by Weather Parameter – Year by Year*

*3.1 B) EDA – Historical Weather – Trends & Seasonality by Region*

*3.1 C – 1) EDA – Historical Wildfires – Mean Estimated Fire Area – Trends & Seasonality Year by Year and Regions*

*3.1 C – 2) EDA – Historical Wildfires – Mean Estimated Fire Brightness – Trends & Seasonality Year by Year and Regions*

*3.1 C – 3) EDA – Historical Wildfires – Mean Estimated Fire Radiative Power – Trends & Seasonality Year by Year and Regions*

Wildfires occur under a set of specific circumstances and conditions, in a particular period of time during the year. Understanding the seasonality of weather year over year, as well as the trends for different parameters year over year, and then putting both weather seasonality and trends in parameter in combination with regions, can give us an insight into the features we should be associating to better understand how we could predict wilfires.

However, it must be remembered that **correlation does not mean causation**. Therefore, it will require further evaluation and feature engineering to get closer to causation and successful prediction.

* + 1. **Historical Weather Data -Trends & Seasonality**

1. Historical Weather – Trends & Seasonality **by Weather Parameters** – Year by Year:

Min & Max Values of All 6 Weather Parameters, month by month, year by year comparison.

**Background:**

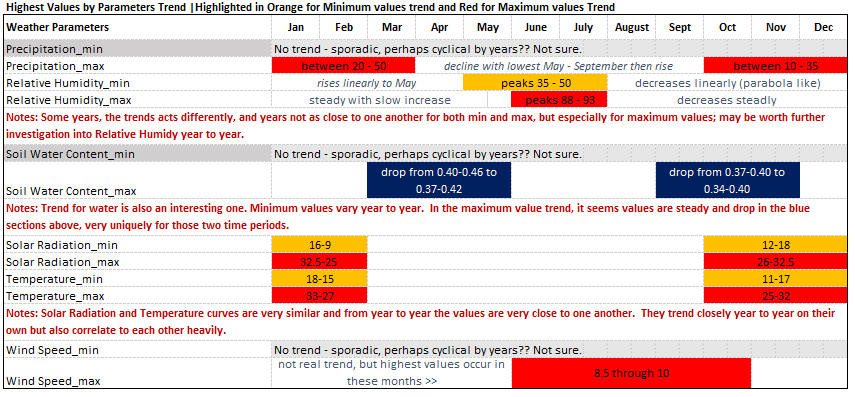
Australian Summer months are **December through January.**

*March – May is Autumn; June – August is Winter; and September – November is Spring*

**Objective of Analysis:**

To observe and gain insight on the overall trend, for all 6 parameters, Precipitation, Relative Humidity, Soil Water Content, Solar Radiation, Temperature and Wind Speed, over the period of a year month to month, with year-by-year comparison, by looking at both the minimum and maximum values.

**Findings:**



Note: This figure can also be found in the Tables excel spreadsheet [here](https://github.com/ElenaE873/WildfiresAustralia_CallforCode2020/blob/main/Trends%20%26%20Seasonality/Trends%20%26%20Seasonality%20Tables.xlsx).

1. Historical Weather – Trends & Seasonality **by Region with Parameters** – Year by Year:

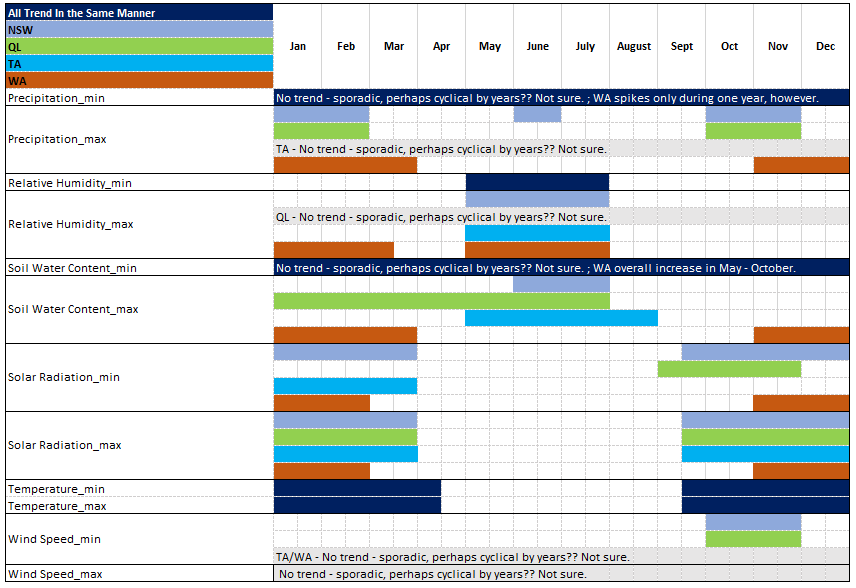
Min & Max Values of All 6 Weather Parameters, looking by Region, by month, year by year comparisons

**Objective of Analysis:**

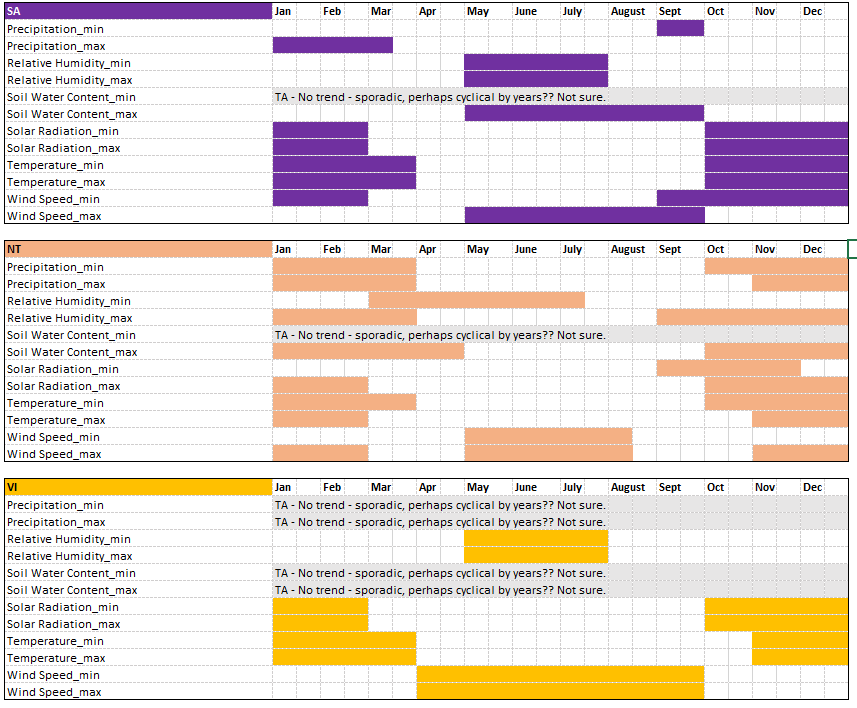
To examine how the trend or seasonality over the 6 parameters, by the 7 separate regions. This meaning that each of the 6 trend examinations are repeated area by area, to determine if weather differs and to what extent, area by area. \*\* With this analysis, the focus will only be on Max values.

Regions: 'NSW', 'NT', 'QL', 'SA', 'TA', 'VI', 'WA'

**Findings:**



Notes:



Notes:

1. Historical Wildfires - Trends & Seasonality Year by Year and Regions:

**Objective of Analysis:**

To observe and gain insight and an understanding of wildfires per region. Understanding the historical files year to year, as well as region to region, will aid in associating this data with weather and vegetation index.

**Findings:**

Findings are slightly inconclusive by looking at this type of time series, even when compared to weather parameter trends. It all just blends together. Need to examine at greater granularity with further questioning.

* 1. Asf

1. Asfd
   1. Calculation of Target Variables
   2. Relationship between Variables
   3. Asdfasdf
2. Modeling
   1. Regression Models
      1. Applying algorithm
      2. Solution to the problem
      3. Performance of Different Models
      4. AsdASD
   2. Classification Models
3. Conclusions
4. Future Directions