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 Wildfires Data**

Data Elements of Interest:

* Region
* Date (YYY-MM-DD)
* Estimated Fire Area (km^2)
* Mean Estimated Fire Brightness (K)
* Mean Estimated Fire Radiative Power (MW)
* Count (ie. Daily numbers of pixels for presumed vegetation fires)
  + 1. **Historical Weather Data**
    2. **Historical Weather Forecast Data**
    3. **Historical Vegetation Data**
    4. **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)**
   1. **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*

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.

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Min** | **Max** |
| **Precipitation** | **\*\*no trend; some spikes for some years, but overall minimum value sporadic, no seasonality or trend.** | **TREND:**  **Highest Jan – March; with an observable decrease beginning as early as March.**  **Leveling off April – October, and then beginning to spike as early as October** |
| **Relative Humidity** | **Highest Humidity seen to begin to rise from March through June/July and then decline at a steady level for remainder of the months.**  **Lowest Relative Humidity : Jan – Mar & Oct - December** | **Approximately the same trend as the minimum values.** |
| **Soil Water Content** | **\*\*no trend; each year it seems to be a bit different and no really trend between years.** | **Years are closer to each other in maximum values – seems to be :**  **Highest values in Jan/Mar**  **And Nov/Dec**  **Lowest(although not much lower) in April/May and Sept/Oct** |
| **Solar Radiation**  **&**  **Temperature**  **\*\* These two parameter trends, closely resemble each other.** | **Very consistent year to year, data close together.**  **Lowest – May – August**  **Hightest – Jan/Feb & Oct/Dec** | **High value trend is even closer together year by year and mimic the minimum value trends.** |
| **Wind Speed** | **Steady/similar in values until about October/November when values seem to be the highest.** | **No obvious trend; but seems highest values are possible to occur anywhere from June – October, but level out in maximum values Jan/Feb and Nov/December.** |

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 paramaters, 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'

|  |  |  |
| --- | --- | --- |
| **Region** | **Min** | **Max** |
| **NSW**  **& QL region (similar characteristic to NSW)**  **&TA region also seems to be similar to NSW and QL**  **WA regions also pretty similar to these.** | **Precipitation:**  **Spikes: Different depending on the years – there must be a cycle to the years in which the spikes occur. Need to investigate further.**  **QL – same aspect**  **TA – same aspect**  **WA – spike only one year in February**  **Relative Humidity:**  **Highest May-July**  **QL – same aspect**  **TA – same aspect**  **WA – same aspect**  **Soil Water Content:**  **\*\*No trend.**  **QL – same aspect**  **TA – same aspect**  **WA – seems to be an overall increase in May – October more than other months.**  **Solar Radiation:**  **Drops lowest in May / July**  **QL – the highest Sept – Nov – otherwise fairly low/steady over the rest of the months**  **TA – same as NSW**  **WA – highest Jan/Feb & Nov/Dec \*\* for min values of Solar radiation for WA – values aren’t so together, wonder if something unique about this region prevent the min values to be so close close.**  **Temperature:**  **Drops lowest in May – August**  **QL – same aspect.**  **TA – same aspect.**  **WA – same aspect.**  **Wind Speed:**  **Steady/similar in values until about October/November when values seem to be the highest.**  **QL – same aspect.**  **TA – no real trend obvious.**  **WA – no real trend obvious.** | **Precipitation:**  **Spikes: Some trend seems to be somewhat visible in Jan/Feb & June & Oct/Nov.**  **QL – Jan/Feb and Oct/Nov but not in June – steady May – September**  **TA – no obvious trend – similar to the min values.**  **WA – spikes seen Jan – Mar and Nov-Dec**  **Relative Humidity:**  **Highest May/July**  **QL – no obvious trend, somewhat pretty steady.**  **TA – same as NSW almost**  **WA – same as NSW almost, as well as highest in Jan - Mar**  **Soil Water Content:**  **\*\*Peaks June**  **\*\*Drops May/August/Oct**  **QL – Highest Jan – July and then the trend overall decreases except for a few years.**  **TA – Highest values stead from May - August**  **WA – Highest Jan – Mar and Nov-Dec**  **Solar Radiation:**  **Drops lowest in May / July**  **QL – Same aspect**  **TA – Same aspect**  **WA – Highest Jan/Feb & Nov-Dec**  **Temperature:**  **Drops lowest in May – August**  **QL – same aspect.**  **TA – same aspect.**  **WA – same aspect.**  **Wind Speed:**  **No obvious trend.**  **QL – same aspect.**  **TA – no real trend obvious.**  **WA – no real trend obvious.** |
| **NT** | **Precipitation:**  **ONLY peaks Jan/Mar & Oct/December and only for a few years between 2009 & 2012 it seems, otherwise low and steady.**  **Relative Humidity:**  **Highest Mar – July**  **Soil Water Content:**  **\*\*No trend.**  **Solar Radiation:**  **Not sure if there is a good trend but definitely peaks Sept – November**  **Temperature:**  **Highest Jan – March**  **And Oct – December**  **Windspeed:**  **Observable trend for highest windspeed May – July/Aug** | **Precipitation:**  **Obvious peaks in Jan/Mar and Nov/Dec for MOST years.**  **Relative Humidity:**  **Lowest April – August for most years.**  **Soil Water Content:**  **Highest Jan – April & Oct – December**  **Solar Radiation:**  **Year to year trend is close to one another with Highest at Jan – Feb & Oct – December**  **Temperature:**  **Highest Jan-Feb and Nov-Dec**  **Windspeed:**  **Observable trend for highest windspeed May – July/Aug AND Nov – Feb (for max)** |
| **SA** | **Precipitation:**  **One year it peaked in September**  **Relative Humidity:**  **Highest May – July**  **Soil Water Content:**  **\*\*No trend.**  **Solar Radiation:**  **Peaks Jan – Feb & Oct – Dec**  **\*\* and year to year is close to each other closely**  **Temperature:**  **Highest Jan – March & Oct – Dec**  **Windspeed:**  **In this region the Windspeed seems to be an obvious trend of highest Jan/Feb and Sept - Dec** | **Precipitation:**  **Only a couple of years it peaked in Jan-March**  **Relative Humidity:**  **Highest May – July**  **Soil Water Content:**  **Highest May – September**  **Solar Radiation:**  **Peaks Jan – Feb & Oct – Dec**  **\*\* and year to year is close to each other closely**  **\*\* this is a very consistent marker! Year to year!! It’s almost the same in the maximum values**  **Temperature:**  **Highest Jan – March & Oct – Dec**  **Windspeed:**  **Steady except for times around May – August/Sept when the max values are the highest.** |
| **VI** | **Precipitation:**  **No obvious trend, peaks in some years, but no pattern visible to the eye, perhaps needs more looking into this.**  **Maybe there’s a cycle of some sort.**  **Relative Humidity:**  **Highest May – July**  **Soil Water Content**  **No obvious trend.**  **Solar Radiation**  **Highest Jan/Feb & Oct – Dec**  **\*\*\* for all the regions the solar radiation values seem to be very predictable since they are so close to each other year to year.**  **Temperature:**  **Highest – Jan-Mar & Nov-Dec**  **Windspeed:**  **No obvious trend, but seem to be higher in April – September than the rest of the months.** | **Precipitation:**  **No obvious trend, peaks in some years, but no pattern visible to the eye, perhaps needs more looking into this.**  **Maybe there’s a cycle of some sort.**  **Relative Humidity:**  **Highest May – July**  **Soil Water Content:**  **No obvious trend.**  **Solar Radiation**  **Highest Jan/Feb & Oct – Dec**  **\*\*\* for all the regions the solar radiation values seem to be very predictable since they are so close to each other year to year.**  **Temperature:**  **Highest – Jan-Mar & Nov-Dec**  **Windspeed:**  **No obvious trend, but seem to be higher in April – September than the rest of the months.** |
|  |  |  |
|  |  |  |

* 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