# Analysis of Australian Rainfall

## On Which Australian Regions and Weather Components Do We Have to Focus to Predict and Improve The Rainfall Levels

Dr. Göktuğ Aşcı

**Professor**: Jorge Centeno
**Date**: 21.02.2021

# 1. Background/Scenario Description

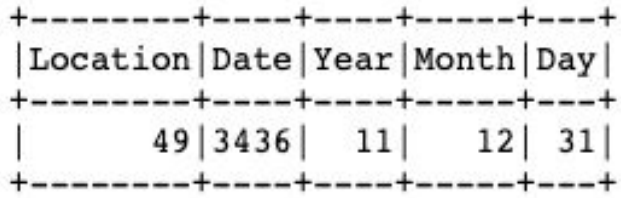
Rainfall in Australia has great importance since Australia has a huge ecosystem of animals and plants and it is home to many different species. Without rain, forests and animals suffer great losses. Results of these losses are beyond Australia and they may cause unwanted consequences in other parts of the world, too. Being able to understand contributing factors on rainfall in Australia helps authorities to take precautions before unwanted consequences emerge.

# The Dataset (Source)

This dataset contains **approximately 10 years of daily** weather observations from many locations across **Australia**.

"**RainTomorrow**" is the column that shows if it rained the next day. **Yes** or **No**? This column is Yes if the rain for that day was **1 mm or more.**

We are going to analyze this data to see the contributing **factors to the next day's rainfall in Australia**.



| Location | Date | Year | Month | Day |
|---|---|---|---|---|
| 49 | 3436 | 11 | 12 | 31 |

We have **3436** days' data from **49 different locations** in Australia.

# 2. Goal of the Analysis

We want to analyze the dataset to discover the regions with the highest and lowest levels of rainfall to compare both groups so that we can focus on improving found variables in the regions with the lowest levels of rainfall.

# 3. Analysis Deep Dive

# Descriptions of Columns

**Date**: The date of observation
**Location**: The common name of the location of the weather station
**MinTemp**: The minimum temperature in degrees celsius
**MaxTemp**: The maximum temperature in degrees celsius
**Rainfall**: The amount of rainfall recorded for the day in mm
**Evaporation**: The so-called Class A pan evaporation (mm) in the 24 hours to 9am
**Sunshine**: The number of hours of bright sunshine in the day.
**WindGustDir**: The direction of the strongest wind gust in the 24 hours to midnight
**WindGustSpeed**: The speed (km/h) of the strongest wind gust in the 24 hours to midnight
**WindDir9am**: Direction of the wind at 9am
**WindDir3pm**: Direction of the wind at 3pm
**WindSpeed9am**: Wind speed (km/hr) averaged over 10 minutes prior to 9am
**WindSpeed3pm**: Wind speed (km/hr) averaged over 10 minutes prior to 3pm
**Humidity9am**: Humidity (percent) at 9am
**Humidity3pm**: Humidity (percent) at 3pm
**Pressure9am**: Atmospheric pressure (hpa) reduced to mean sea level at 9am
**Pressure3pm**: Atmospheric pressure (hpa) reduced to mean sea level at 3pm
**Cloud9am**: Fraction of sky obscured by cloud at 9am. This is measured in "oktas", which are a unit of eigths. It records how many
**Cloud3pm**: Fraction of sky obscured by cloud (in "oktas": eighths) at 3pm. See Cload9am for a description of the values
**Temp9am**: Temperature (degrees C) at 9am
**Temp3pm**: Temperature (degrees C) at 3pm
**RainToday**: True if precipitation (mm) in the 24 hours to 9am exceeds 1mm, otherwise 0
**RainTomorrow**: The amount of next day rain in mm. Used to create response variable RainTomorrow. A kind of measure of the "risk".

# Dataset Rainfall Definition (1 mm or more) & "water.usgs.gov" Categorization

- Drizzle, very small droplets.
- Slight (fine) drizzle: Detectable as droplets only on the face, car windscreens and windows.
- Moderate drizzle: Windows and other surfaces stream with water.
- Heavy (thick) drizzle: Impairs visibility and is measurable in a raingauge, rates up to 1 mm per hour.
- Rain, drops of appreciable size and may be described as small to large drops. It is possible to have rain drops within drizzle!
- **Slight rain**: Less than 0.5 mm per hour.
- **Moderate rain**: Greater than 0.5 mm per hour, but less than 4.0 mm per hour.
- **Heavy rain**: Greater than 4 mm per hour, but less than 8 mm per hour.
- **Very heavy rain**: Greater than 8 mm per hour.
- Slight shower: Less than 2 mm per hour.
- Moderate shower: Greater than 2 mm, but less than 10 mm per hour.
- Heavy shower: Greater than 10 mm per hour, but less than 50 mm per hour.
- Violent shower: Greater than 50 mm per hour.

## Application to Our Dataset

```
"very low"    - <= 0.5 mm of Rainfall
"low"         - 0.5 to 1.0 mm of Rainfall
"medium"      - 1.0 to 4.0 mm of Rainfall
"high"        - 4.0 to 8.0 mm of Rainfall
"very high"   - > 8.0 mm of Rainfall
```

# Rainfall Column Simple Statistics

| summary | Rainfall |
|---|---|
| count | 142199 |
| mean | 2.3609181508908756 |
| stddev | 8.47805974281768 |
| min | 0.0 |
| 25% | 0.0 |
| 50% | 0.0 |
| 75% | 0.8 |
| max | 371.0 |

The mean rainfall is **2.36 mm** which is above the threshold of 1 mm. This must be a result of the rainfall days with very high level.

| RainfallLevel | NumDays | RoundedRatio |
|---|---|---|
| 1. very low | 91080 | 62.62 |
| 2. low | 6435 | 4.42 |
| 3. medium | 13459 | 9.25 |
| 4. high | 7004 | 4.82 |
| 5. very high | 24221 | 16.65 |

Lower than medium (**< 1.0 mm**) means no rainfall that day.

**67% of the days** are without rainfall whereas **33% of the days** are with rainfall.

# Top Correlations with Min/Max Temp and Humidity at 3 pm and 9 am and Rainfalls

```
df.stat.corr("Rainfall","MinTemp")
```

```
0.10308227743558883
```

```
df.stat.corr("Rainfall","MaxTemp")
```

```
-0.06707040954376242
```

```
df.stat.corr("Rainfall","Humidity9am")
```

```
0.2002553619853785
```

```
df.stat.corr("Rainfall","Humidity3pm")
```

```
0.224988156859171117
```

| MaxTemp | count |
|---|---|
| null | 1261 |
| 20.0 | 885 |
| 19.0 | 843 |
| 19.8 | 840 |
| 20.4 | 834 |
| 19.9 | 823 |
| 20.8 | 817 |
| 19.5 | 812 |
| 18.5 | 811 |
| 21.0 | 810 |

only showing top 10 rows

| MinTemp | count |
|---|---|
| null | 1485 |
| 11.0 | 899 |
| 10.2 | 898 |
| 9.6 | 896 |
| 10.5 | 884 |
| 10.8 | 872 |
| 9.0 | 872 |
| 10.0 | 871 |
| 12.0 | 866 |
| 8.9 | 861 |

only showing top 10 rows

| Humidity3pm | count |
|---|---|
| null | 4507 |
| 52 | 2751 |
| 55 | 2738 |
| 57 | 2728 |
| 53 | 2697 |
| 59 | 2690 |
| 58 | 2643 |
| 54 | 2642 |
| 50 | 2624 |
| 51 | 2621 |

only showing top 10 rows

| Humidity9am | count |
|---|---|
| 99 | 3391 |
| 70 | 3026 |
| 69 | 3023 |
| 65 | 3014 |
| 68 | 3011 |
| 71 | 2976 |
| 66 | 2973 |
| 67 | 2950 |
| 74 | 2917 |
| 72 | 2914 |

only showing top 10 rows

# Max Temperature Comparison

**'MaxTemp'** (in degrees celsius):

| RainfallLevel | AverageMaxTemp | LowestMaxTemp | HighestMaxTemp | StdDevMaxTemp |
|---|---|---|---|---|
| 1. very low | 24.86252850236239 | -2.1 | 48.1 | 6.9913464206633655 |
| 2. low | 20.346253308393138 | -1.7 | 44.9 | 6.357019826132103 |
| 3. medium | 19.909953063761726 | -4.1 | 46.3 | 6.327637734360568 |
| 4. high | 19.82801031612112 | -4.8 | 41.8 | 6.425530237851865 |
| 5. very high | 20.684054189296116 | -3.8 | 44.6 | 6.364049670083962 |

**There is a substantial difference** between the **average max temperatures** of the days with very low level rainfall and the days with medium or more level of rain.

# Temperature at 3 pm Comparison

**'Temp3pm'** (in degrees celsius):

| RainfallLevel | AverageTemp3pm | LowestTemp3pm | HighestTemp3pm | StdDevTemp3pm |
|---|---|---|---|---|
| 1. very low | 23.304804204747374 | -4.0 | 46.7 | 6.806170871279209 |
| 2. low | 18.829858801394877 | -3.0 | 42.6 | 6.1572952623697805 |
| 3. medium | 18.39319402943301 | -4.4 | 44.7 | 6.13526373501759 |
| 4. high | 18.30713137457193 | -5.4 | 40.2 | 6.271218431201189 |
| 5. very high | 19.17177249441582 | -5.1 | 43.3 | 6.199604232821436 |

**There is a substantial difference** between the **average temperatures at 3 pm** of the days with very low level rainfall and the days with low or more level of rain.

# Wind Gust Speed Comparison

**'WindGustSpeed'** (km/h):

| RainfallLevel | AverageWindGustSpeed | LowestWindGustSpeed | HighestWindGustSpeed | StdDevWindGustSpeed |
|---|---|---|---|---|
| 1. very low | 38.90988876894917 | 7 | 130 | 12.761121035053586 |
| 2. low | 40.44055237453688 | 6 | 113 | 13.92959458685912 |
| 3. medium | 41.97314629258517 | 9 | 135 | 14.207463267827045 |
| 4. high | 43.90278425882719 | 7 | 109 | 14.526739930842059 |
| 5. very high | 41.71483262108262 | 7 | 135 | 15.287392451713098 |

**There is a substantial difference** between the **average wind gust speed** of the days with very low level rainfall and the days with medium or more level of rain.

# Humidity at 3 pm Comparison

**'Humidity3pm'** (percent):

| RainfallLevel | AverageHumidity3pm | LowestHumidity3pm | HighestHumidity3pm | StdDevHumidity3pm |
|---|---|---|---|---|
| 1. very low | 44.857337354006795 | 0 | 100 | 19.410347322021217 |
| 2. low | 60.77081668531245 | 5 | 100 | 17.454898728730374 |
| 3. medium | 63.119221967963384 | 1 | 100 | 17.18607284132653 |
| 4. high | 65.93331370528485 | 6 | 100 | 17.27868124441239 |
| 5. very high | 63.36847248778415 | 1 | 100 | 18.133107232689323 |

**There is a substantial difference** between the **average humidities at 3 pm** of the days with very low level rainfall and the days with low or more level of rain.

# Top 20 locations with the highest rainfall level by ratio:

| Location | RainfallLevel | NumRainfallDays | TotalRainfallDays | RainfallRatio |
|---|---|---|---|---|
| Portland | 5. very high | 787 | 3009 | 26.15 |
| Witchcliffe | 5. very high | 723 | 3009 | 24.03 |
| MountGambier | 5. very high | 692 | 3040 | 22.76 |
| MountGinini | 5. very high | 680 | 3040 | 22.37 |
| NorfolkIsland | 5. very high | 667 | 3009 | 22.17 |
| Cairns | 5. very high | 673 | 3040 | 22.14 |
| Penrith | 5. very high | 670 | 3039 | 22.05 |
| Walpole | 5. very high | 660 | 3006 | 21.96 |
| Dartmoor | 5. very high | 657 | 3009 | 21.83 |
| Sale | 5. very high | 645 | 3009 | 21.44 |
| NorahHead | 5. very high | 633 | 3004 | 21.07 |
| CoffsHarbour | 5. very high | 626 | 3009 | 20.8 |
| Watsonia | 5. very high | 614 | 3009 | 20.41 |
| Sydney | 5. very high | 682 | 3344 | 20.39 |
| Bendigo | 5. very high | 598 | 3040 | 19.67 |
| Richmond | 5. very high | 578 | 3009 | 19.21 |
| Darwin | 5. very high | 611 | 3193 | 19.14 |
| Hobart | 5. very high | 605 | 3193 | 18.95 |
| Newcastle | 5. very high | 576 | 3039 | 18.95 |
| Brisbane | 5. very high | 592 | 3193 | 18.54 |

# Top 20 locations with the lowest rainfall level by ratio:

| Location | RainfallLevel | NumRainfallDays | TotalRainfallDays | RainfallRatio |
|---|---|---|---|---|
| AliceSprings | 1. very low | 2665 | 3040 | 87.66 |
| Woomera | 1. very low | 2603 | 3009 | 86.51 |
| Uluru | 1. very low | 1324 | 1578 | 83.9 |
| Mildura | 1. very low | 2424 | 3009 | 80.56 |
| Cobar | 1. very low | 2382 | 3009 | 79.16 |
| Katherine | 1. very low | 1229 | 1578 | 77.88 |
| Moree | 1. very low | 2267 | 3009 | 75.34 |
| PerthAirport | 1. very low | 2235 | 3009 | 74.28 |
| Townsville | 1. very low | 2232 | 3040 | 73.42 |
| Perth | 1. very low | 2293 | 3193 | 71.81 |
| WaggaWagga | 1. very low | 2153 | 3009 | 71.55 |
| Tuggeranong | 1. very low | 2149 | 3039 | 70.71 |
| Canberra | 1. very low | 2421 | 3436 | 70.46 |
| PearceRAAF | 1. very low | 2068 | 3009 | 68.73 |
| SalmonGums | 1. very low | 2033 | 3001 | 67.74 |
| Darwin | 1. very low | 2147 | 3193 | 67.24 |
| Nhil | 1. very low | 1052 | 1578 | 66.67 |
| Albury | 1. very low | 2015 | 3040 | 66.28 |
| Nuriootpa | 1. very low | 1943 | 3009 | 64.57 |
| BadgerysCreek | 1. very low | 1939 | 3009 | 64.44 |

# 4. Conclusions - I

1. The mean rainfall in all regions is **2.36 mm** which is above the threshold of 1 mm. This must be a result of the rainfall days with very high level.
2. The rainfall level has a higher correlation with **min/max temperature** during the day and and **humidity measured at 3 pm and 9 am**.
3. **Minimum temperature, maximum temperature, wind gust speed, and humidity at 3 pm and 9 am** components are remarkably different among the days with very low and very high levels of rainfall.
4. **Portland, Witchcliffe, MountGambier, MountGinini, NorfolkIsland, Cairns, Penrith, Walpole, Dartmoor, Sale** are the top 10 regions with very high levels of rainfall.
5. **AliceSprings, Woomera, Uluru, Mildura, Cobar, Katherine, Moree, PerthAirport, Townsville, Perth** are the top 10 regions with very low levels of rainfall.

# 4. Conclusions - II

1. Authorities can focus on strategies to improve the amount of average daily rainfall levels in **AliceSprings, Woomera, Uluru, Mildura, Cobar, Katherine, Moree, PerthAirport, Townsville, Perth**.
2. While implementing new strategies, improvement can be measured by components such as **average minimum temperature, average maximum temperature, average wind gust speed, and average humidity at 3 pm and 9 am**.

# Sources

https://www.kaggle.com/jsphyg/weather-dataset-rattle-package

https://water.usgs.gov/edu/activity-howmuchrain-metric.html

http://www.bom.gov.au/climate/data

http://www.bom.gov.au/climate/dwo/