Document Summaries

# Statistics for all documents

## Barometer Rainfall

|  |  |
| --- | --- |
|  | Barometer |
| Measure | Baro |
| sum | 358549.6 |
| min | 979.6 |
| max | 1035.6 |
| count | 355 |
| mean | 1009.999 |
| variance | 97.136 |
| standard deviation | 9.856 |

|  |  |
| --- | --- |
|  | Rainfall |
| Measure | mm |
| sum | 546.7 |
| min | 0.0 |
| max | 23.2 |
| count | 353 |
| mean | 1.549 |
| variance | 11.022 |
| standard deviation | 3.32 |

## Indoor Temperature

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Indoor Temp Readings | | | |
| Measure | Humidity | Temp | Temp (Low) | Temp (High) |
| sum | 17176 | 7727.071 | 7276.8 | 8330.9 |
| min | 37 | 18.04 | 14.9 | 19.7 |
| max | 59 | 29.21 | 28.2 | 31.1 |
| count | 354 | 354 | 354 | 354 |
| mean | 48.52 | 21.828 | 20.556 | 23.534 |
| variance | 26.848 | 4.225 | 5.768 | 2.887 |
| standard deviation | 5.182 | 2.055 | 2.402 | 1.699 |

## Outside Temperature

|  |  |  |  |
| --- | --- | --- | --- |
|  | Outdoor Temp Readings | | |
| Measure | Temp | Temp (Low) | Temp (High) |
| sum | 3954.301 | 2792.3 | 5511.1 |
| min | -1.81 | -4.1 | 1.5 |
| max | 26.38 | 18.7 | 38.5 |
| count | 355 | 355 | 355 |
| mean | 11.139 | 7.866 | 15.524 |
| variance | 28.596 | 23.737 | 49.344 |
| standard deviation | 5.347 | 4.872 | 7.025 |

# Statistical Differences

There are two files.

* A list of outdoor temperature readings
* An edited version of outdoor temperature readings

The edited version imagines that from 2017-06-09 to 2017-06-20 the thermometer was broken and only reported 0.0°C for all measures.

The statistics for the two files are this:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Outdoor Temp Readings | | | Edited Temp Readings | | |
| Measure | Temp | Temp (Low) | Temp (High) | Temp | Temp (Low) | Temp (High) |
| sum | 3954.301 | 2792.3 | 5511.1 | 3721.921 | 2626.4 | 5187.7 |
| min | -1.81 | -4.1 | 1.5 | -1.81 | -4.1 | **0.0** |
| max | 26.38 | 18.7 | 38.5 | 26.06 | 18.0 | 38.5 |
| count | 355 | 355 | 355 | 355 | 355 | 355 |
| mean | 11.139 | 7.866 | 15.524 | 10.484 | 7.398 | 14.613 |
| variance | 28.596 | 23.737 | 49.344 | 29.595 | 24.217 | 51.159 |
| standard deviation | 5.347 | 4.872 | 7.025 | 5.44 | 4.921 | 7.153 |

Comparing the stats between them only identifies one measure that looks a bit strange. Most of the stats are very similar. A minimum value of 0.0°C for the column of Temperature (High) doesn’t strike me as correct. If I saw this, I would try find that value in the source data. I would then obviously see the error.

However, in general, stats only give you a one-dimensional picture of your data. The best way to assess whether you have any data quality issues would be to visualise it.

![Chart

Description automatically generated]()

In the above image the solid blue line is the original data. The dashed grey line is the altered data set. They overlap completely for the most part, but the altered incorrect data is immediately obvious.