Built-in python tools were used to get a description of the data as below:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Features | count | mean | std | min | 25% | 50% | 75% | max |
| SoilMoisture | 130 | 527.6469 | 57.36784 | 380.7 | 488.625 | 538.3 | 571.025 | 647.3 |
| Average\_Temp | 130 | 26.84992 | 0.651413 | 25.15807 | 26.44229 | 26.93065 | 27.27073 | 28.58 |
| Min\_Temp | 130 | 21.37923 | 0.688971 | 18.9 | 21 | 21.5 | 21.8 | 22.6 |
| Max\_Temp | 130 | 33.85154 | 1.079638 | 31.1 | 33.1 | 33.9 | 34.6 | 36 |
| Precipitation | 130 | 188.9808 | 80.23721 | 2 | 140.3 | 182.15 | 226.1 | 496.1 |
| Working\_days | 130 | 24.75385 | 1.239289 | 21 | 24 | 25 | 26 | 27 |
| HA\_Harvested | 130 | 793404.5 | 34440.89 | 683431.9 | 768966.9 | 790036.2 | 821989.2 | 882254.2 |
| FFB\_Yield | 130 | 1.602231 | 0.281751 | 1.08 | 1.39 | 1.585 | 1.8075 | 2.27 |

From the above, we obtain the standardized deviations by dividing Standard Deviation by the Mean.

All have fairly reasonably-sized standardized deviations. FBB\_Yield has been grayed out, as it is the output variable.

|  |  |
| --- | --- |
| Feature | Standardized Standard Deviation |
| Precipitation | 0.424579 |
| FFB\_Yield | 0.175849 |
| SoilMoisture | 0.108724 |
| Working\_days | 0.050065 |
| HA\_Harvested | 0.043409 |
| Min\_Temp | 0.032226 |
| Max\_Temp | 0.031893 |
| Average\_Temp | 0.024261 |

When we calculate a Pearson correlation matrix, we find that the features have the following level of correlatedness:

Medium negative correlation between **FFB\_Yield and HA\_Harvested** of -0.35022183838334325

Low negative correlation between FFB\_Yield and SoilMoisture of -0.003182901354288915

Low negative correlation between FFB\_Yield and Average\_Temp of -0.0054943529235985545

Low positive correlation between **FFB\_Yield and Min\_Temp** of 0.10382969410528087

Low negative correlation between FFB\_Yield and Max\_Temp of -0.07120090454850743

Low positive correlation between **FFB\_Yield and Precipitation** of 0.28960372412286844

Low positive correlation between **FFB\_Yield and Working\_days** of 0.11636407232753451

We ignore the first and last Pearson Correlation Coefficient, because FBB yield cannot be correlated with itself.

From the correlation coefficient, it is clear that FBB yield is correlated positively with working days, precipitation, and minimum temperature. **The higher the working days, precipitation, and minimum temperature, the greater the FBB yield will be.** Additionally, there is a significant negative correlation between HA\_Harvested and FFB\_Yield. **The lower the HA\_Harvested, the higher the FFB\_Yield will be**.

Trying again with the Kendall correlation matrix for confirmation, we get the following levels of correlatedness:

Low negative correlation between FFB\_Yield and SoilMoisture of -0.026614232055472888

Low negative correlation between FFB\_Yield and Average\_Temp of -0.03200709960461089

Low positive correlation between FFB\_Yield and Min\_Temp of 0.05742958165068296

Low negative correlation between FFB\_Yield and Max\_Temp of -0.07554885173815311

Low positive correlation between **FFB\_Yield and Precipitation of 0.2182147845738709**

Low positive correlation between FFB\_Yield and Working\_days of 0.07569866343463635

Low negative correlation between **FFB\_Yield and HA\_Harvested of -0.2623789689787513**

Based on the results obtained for both the Spearman and Kendall correlation, the **most strongly-correlated features are Precipitation and HA\_Harvested, and the two more weakly-correlated features are Min\_Temp and Working\_Days.**