This question requires to analyse external factor influence fresh fruit bunch (FFB) yield. Before we start analysing first step would be to perform data exploration to understand the data. It can be seen that:

* Data does not contain null values.
* Data has 130 rows and 9 columns.

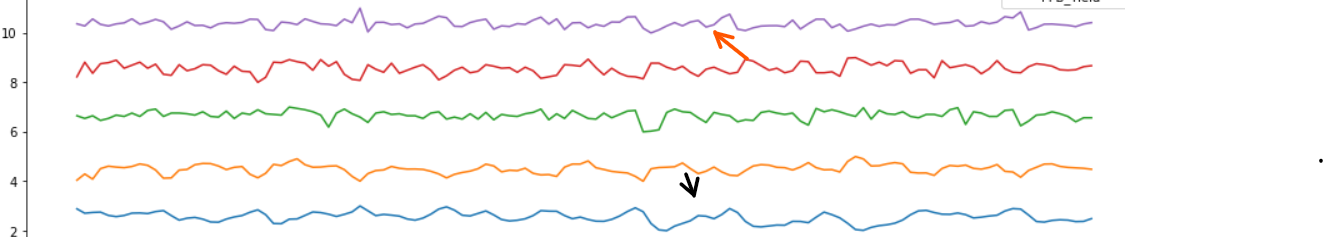
Secondly, to check the distribution of the target variable FFB density graph was plotted and it can be seen that FFB target variable is not skewed.

Chart, histogram

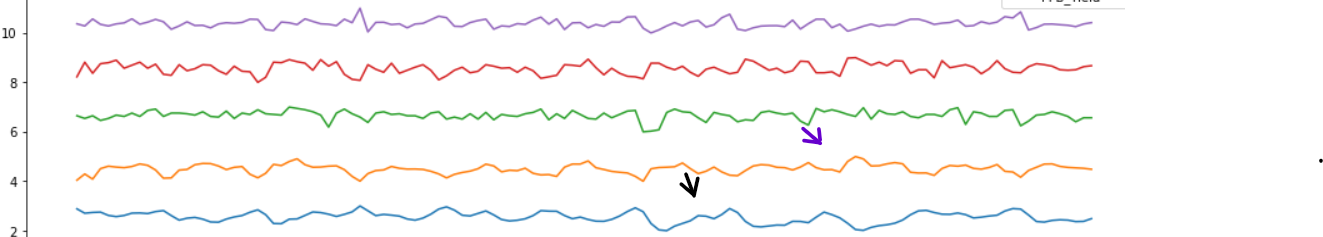
Description automatically generated

A picture containing background pattern

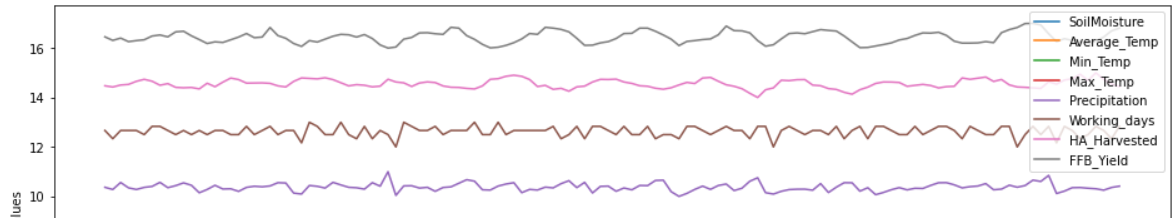
Description automatically generatedSince the date attribute is given we can plot a time-series visualization to check how variables varies from year 2008 to 2018. Before plotting the time-series the data is normalise from 0 to 1 to ensure values are in common scale without distorting difference in the ranges value.



From the time-series plot above we can see that there is a similar trend between the variable **“SoilMoisture”** and **“Precipitation”.** For instance, in year 2011 both of the variables have spike increase and in year 2014 both of the variable has a spike decrease.



On the other hand, variables **“SoilMoisture”** and  **“Average\_Temp”** shows an exactly opposite trend. Hence, in year 2014 while the value of FFB\_Yield was increasing the value of Average\_Temp was decreasing and in year 2010 while the value of Average\_Temp.



Lastly, the target variable “**FFB\_Yield”** also has similar trend with variable “**Precipitation”.** For instance, in year 2014 there is a decreasing trend for both of the variables.

Spearman correlation matrix was plotted to confirm the assumption made.

A picture containing table

Description automatically generated

As we can see from the correlation matrix above my initial assumption is valid. Among all the variables **FFB\_Yield** has highest positive correlation with **Precipitation**. On the other hand, **HA\_Harvested** (Harvested Area in hectare area) has highest negative correlation with **FFB\_Yield**.

Moreover, my initial assumption for relationship between **Average\_Temp** and **Precipitation** with **SoilMoisturiser** shown to be valid. As **SoilMoisturiser** has highest positive correlation with **Precipitation** and highest negative correlation with **Average\_Temp.**

Alternatively, I will also be using Linear Regression to investigate relationship between variables from which we can observe effect on independent variable on the dependent variables and determine the relative importance of variables in the model. Linear regression uses the formula **y = ax + b** where x and y are numerical variables and a and b are calculated to minimise the squared error between observe values and fitted values.

The coefficient values obtained for the variables importance are as shown as below

|  |  |
| --- | --- |
| Variables | Coefficient Values |
| **Precipitation** | 0.628 |
| **Working\_days** | 0.0587 |
| **Average\_Temp** | 0.00178 |
| **Max\_Temp** | 0.00653 |
| **Min\_Temp** | -0.0656 |
| **SoilMoisture** | -0.257 |

As shown above, the coefficient value of **Precipitation** is highest compared to all other variables. On the other hand, **SoilMoisture**  has the lowest coefficient value compared to other variables. To interpret the result:

* If the Precipitation increases by 1 unit, the FFB\_Yield increases by 0.628
* If the SoilMoisture increases by 1 unit, the FFB\_Yield decreases by 0.257

To summarise everything that has been stated, **Precipitation and SoilMoisture** are the 2 factors that has greatest influence on the fresh fruit bunch (FFB) yield. High precipitation and low soil moisture is favourable in order to produce quality FFB yield. However, study from 2015 has indicated that to produce quality FFB yield external factors such as harvesting frequency should be maintained at twice a month and all loose fruit should be collected and the FFB stalk trimmed less than 5 mm. The studies also shows that harvesting is a min factors affecting the production of quality FFB. Moreover, human factor such as education level and experience of a farmers plays an important role. Hence, the data provided is insufficient in order to give a concrete evidence on the FFB\_Yield production.

***Reference***: Zulkifli, A; Ayat, K; Abd Rahman, A K and Suboh, I (2015). Assessmentof oil palm seedlings assistance scheme on fresh fruit bunch (FB) yield and income of smallholders. Oil Palm Industry Economic J., 13(1): 35-44.