**Analysis Report**

**Q1.**

**Section a.**

*1.1.1 Summary Statistics*

A table with numbers and letters

Description automatically generated

The levels of additives from 'b' to 'i' show greater variability compared to additive 'a'. This is because of the large standard deviation and range. Additive 'a' has a lower mean and narrower variability. Its standard deviation is 0.003037, and its range is 0.022780.

These summary statistics show that the petrol formulations have different levels of additives. Due to the significant differences in mean values, standard deviations, ranges, and IQRs, it is likely that the formulations have distinct characteristics in terms of burning pattern and additive compositions.

*1.1.2 Correlation Results*

A table with numbers and letters

Description automatically generated

* Additive 'a' tends to have a strong negative relationship with additives 'e' with coefficient of -0.54. This means that when the levels of additive 'a' are high, the levels of additive 'e' is usually low and vice versa.
* Additive 'b' has a weak negative relationship with additives 'c' and 'g' for coefficient of -0.27 and -0.28 respectively. This suggests that when the levels of additive 'b' change, the levels of additives 'c' and 'g' may change slightly in the opposite direction.
* Moderate negative relationship occurs between additives 'c' and 'd' of coefficient -0.48. This suggests that when the levels of additive 'c' increase, the levels of additive 'd' tend to decrease and vice versa.
* Additive 'd' shows a moderate positive relationship with additive 'h' of 0.48. This suggests that as the levels of additive 'd' change, the levels of additive 'h' may change too in the same direction.
* Additive 'h' has weak negative relationships with additive 'i' of coefficient -0.06 . This suggests that there is only slight changes in the opposite directions.
* A weak positive relationship occurs between additives 'i' and 'g' for coefficient 0.13. This suggests that when the levels of additive 'i' increase, the levels of additive 'g' may also increase slightly.

*1.1.3 ANOVA Results*

*statistic=168331.96885835085, pvalue=0.0*

Based on the analysis, a large F-statistic value and a low p-value indicate that there are differences among the petrol formulations. This suggests strong evidence against the null hypothesis. However, a large F-statistic value is not common and may suggest issues with the data or analysis. It's possible that outliers (Section b 1.2.2 Boxplot) in the data are increasing the F-statistic value.

In conclusion, the results show that at least one group of petrol formulation's mean is significantly different from the others. Therefore, it can be concluded that there are significant differences among the petrol formulations. However, it's important to consider conducting physical tests or experiments to complement the statistical analysis.

**Section b.**

*1.2.1 Histograms*

A graph of a number of blue and white bars

Description automatically generated with medium confidence

According to the histogram, the additives 'a', 'b', 'd', 'f', 'g', 'h', and 'i' have a positive skewness, indicating that these additives have more occurrences of lower values and fewer occurrences of higher values. This skewed distribution means that these additives tend to have higher values, which may suggest that some formulations are more effective at preventing engine knocking, gum formation or ensuring stability in storage at lower concentrations. In other words, positive skewness may indicate that certain additives work better at lower concentrations.

On the other hand, a negative skewness for additive 'c' and 'e' suggests that these additives have more occurrences of higher values and fewer occurrences of lower values, resulting in a distribution that is skewed towards lower values. A negative skewness may indicate formulations with higher additive concentrations are needed to achieve the desired effects.

*1.2.2 Boxplot*

A graph with a blue rectangle and black lines

Description automatically generatedA graph with a blue rectangle and black lines

Description automatically generatedA graph with a blue square

Description automatically generatedA graph with a blue rectangle and black lines

Description automatically generatedA graph with a blue rectangle and black lines

Description automatically generatedA graph with a bar graph and numbers

Description automatically generated with medium confidence A graph with a blue rectangle and black squares

Description automatically generatedA graph of a number of objects

Description automatically generated with medium confidenceA graph with a blue square and black lines

Description automatically generated

Based on the boxplot, we can further clarify the presence of outliers. We can see that there are more than one additives have outliers, specifically additives 'a', 'd', 'e', 'g' and ‘h’ have a relatively more outliers. The identification of outliers suggests that there are potential discrepancies in the data associated with these additives and variations in additive concentrations that may give impact on the consistency of petrol formulations.

*1.2.3 Heatmap*

A screenshot of a computer screen

Description automatically generated

Based on the heatmap, we can see that the lighter color indicates it has stronger positive correlation whereas the darker color indicates it has stronger negative correlation.

* Additive 'a' tends to have a stronger negative relationship with ‘e’ for correlation of -0.54 which can be represented with the black color cell.
* Additive 'b' has a dim purple color that shows a weak negative relationship of -0.27 and -0.28 with additives 'c' and 'g' respectively where there are slightly change in color in an opposite way.
* Both additive ‘c’ and ‘g’ exhibit a moderate negative correlation of -0.44 which indicated by the black colors.
* Additive 'd' shows orange color in heatmap which indicate there is moderate positive relationship with additive 'h' of coefficient 0.48 which both of them show changes in the color in the same direction.

**Section c**

*K-Means Clustering*

A graph with colored dots and red x

Description automatically generated

Upon observing the scatter plot and the distribution of clusters, it appears that the chosen number of clusters of 5 effectively captures the variations and patterns in the data to some extent. Each cluster exhibits unique characteristics as evidenced by the varying densities and patterns within clusters. However, there are a few aspects to consider:

* There are instances where clusters appear to overlap, especially light green and dark blue color. This suggest that the separation between clusters may not be optimal.
* The centroid is positioned relatively close to the purple and blue clusters at left side of the scatter plot which suggests that a closer relationship or similarity in their data points.
* The presence of outliers within the yellow cluster at the right side of the scatter plot. These anomalies might signify noise in the data where the K-Means algorithm's partitioning could be refined.

In summary, the scatter plot has provided valuable insights that reveal distinct groupings of data points into the formulations of petrol by looking at the pattern and distribution around centroids.

**Q2**.

*Summary Statistics (from section 2.5)*

Interpretation of each factor:

1. **Soil Moisture:**

|  |  |
| --- | --- |
| Avg\_SoilMoisture | 527.646923076923 |
| Min\_SoilMoisture | 380.7 |
| Max\_SoilMoisture | 647.3 |

Oil palm trees are a valuable crop that requires optimal soil moisture for healthy growth and high yields of Fresh Fruit Bunches (FFB). Proper moisture levels are essential to ensure good root development, nutrient uptake and overall plant health. Monitoring and managing soil moisture levels is crucial to prevent drought stress, root suffocation from waterlogging and reduced FFB production.

Maintaining ideal soil moisture levels is crucial for optimal nutrient availability and root health, which are necessary for palm growth and FFB yield. It is recommended to track both the minimum and maximum moisture levels to identify potential issues. Any deviations from the optimal range can negatively affect nutrient availability and plant health. For example, low moisture levels can lead to reduced growth and lower FFB yield whereas excessive moisture can suffocate roots, increase disease susceptibility and significantly impact FFB yield.

In addition, the average moisture level reflects the typical hydration state of the plantation. The summary statistics of the average soil moisture indicates that it is within the ideal range, then the overall health of the oil palm plantation is likely good, with roots functioning well and nutrient uptake happening efficiently.

However, it is important to note that even with an average moisture level in the good range, there could still be variations within the plantation. Some areas might be drier or wetter than others which could impact the overall health of the trees. Therefore, constant vigilance through monitoring and observation can help ensure optimal soil moisture conditions throughout the plantation. This way, any areas that are drier or wetter than others can be identified and addressed, helping to maintain the health of the trees and ensure optimal FFB yield.

1. **Temperature**

|  |  |
| --- | --- |
| Avg\_Average\_Temp | 26.849918480923062 |
| Min\_Average\_Temp | 25.15806452 |
| Max\_Average\_Temp | 28.58 |

Temperature is one of the critical factors in the physiology of oil palm trees and it directly affects their growth, development and fruit production. It has been observed that a moderate average temperature of 26.85°C is ideal for promoting healthy growth, flowering and yield development. This temperature range facilitates efficient metabolic processes, nutrient uptake and photosynthesis which are vital for successful yield development.

When the temperature is too low, it can slow down the metabolic processes and hinder the uptake of essential nutrients which can lead to stunted growth and delayed flowering and give impact on FFB yield. On the other hand, when the temperature is too high, it can cause increased water loss, leading to water stress and reduced photosynthetic efficiency which also affects the yield.

Effective temperature management is crucial for maintaining healthy oil palm plantations and maximizing yield. This involves monitoring the temperature regularly and providing suitable protection against extreme temperatures. For instance, in colder temperatures, oil palm trees can be protected by using heating systems while in hotter temperatures, shading or other cooling systems could be implemented to reduce heat stress.

In conclusion, it is advisable to pay close attention to temperature management to ensure that their plantations remain healthy and productive. By maintaining the ideal temperature range, high-quality yields can be produced which is essential for the sustainability of the palm oil industry.

1. **Precipitation**

|  |  |
| --- | --- |
| Avg\_Precipitation | 188.98076923076923 |
| Min\_Precipitation | 2.0 |
| Max\_Precipitation | 496.1 |

The health and fruit production of oil palm depend significantly on precipitation patterns as they influence the availability of water in the soil. The level of precipitation received by the oil palm plantation determines the amount of water available for plant growth and helps maintain soil moisture levels.

The minimum and maximum precipitation levels indicate the range of rainfall conditions experienced by the oil palm trees. Insufficient precipitation level can lead to drought stress which lowers the FFB yields whereas excessive precipitation level can cause waterlogging, leaching of nutrients and negatively impact FFB production. However, the average precipitation values which fall between the range of 188.98 mm is ideal for supporting soil moisture balance and overall plant hydration, resulting in a well-grown FFB yield.

To combat insufficient or excessive precipitation, some precautionary steps can be taken. For instance, the excessive precipitation level can be used as an extra supplement of water during dry periods. Additionally, it is important to monitor potential outbreaks due to an increase in humidity and design a good drainage system to prevent any waterlogging.

*Correlation (from section 2.6)*

|  |  |
| --- | --- |
| Corr\_SoilMoisture\_FFB\_Yield | -0.003182901 |
| Corr\_Average\_Temp\_FFB\_Yield | -0.005494353 |
| Corr\_Precipitation\_FFB\_Yield | 0.289603724 |
| Corr\_Workingdays\_FFB\_Yield | 0.116364072 |
| Corr\_Harvested\_FFB\_Yield | -0.350221838 |

1. **Correlation between soil moisture and FFB yield**

The analysis reveals a very weak negative correlation of -0.003 between soil moisture and FFB yield. This suggests that changes in soil moisture levels have minimal direct impact on the amount of fruit bunches harvested. Even if soil moisture levels fluctuate somewhat, it likely does not significantly affect FFB yield based on this data.

1. **Correlation between temperate and FFB yield**

Similar to soil moisture, the correlation coefficient of -0.005 indicates a very weak negative relationship between average temperature and FFB yield. This implies that average temperature doesn't strongly influence variations in FFB yield. Even if the average temperatures fluctuate a bit, it is unlikely to be a major factor driving changes in FFB yield within this dataset.

1. **Correlation between precipitation and FFB yield**

The correlation coefficient of 0.290 suggests a moderately positive relationship between precipitation and FFB yield. This means that higher levels of precipitation are associated with an increase in the amount of fruit bunches harvested. This makes sense because adequate rainfall likely improves soil moisture, keeps the plants well-hydrated and promotes ideal growing conditions which will then lead to a higher FFB yield.

1. **Correlation between working days and FFB yield**

The correlation coefficient of 0.116 indicates a weak positive relationship between the number of working days and FFB yield. This suggests that having more working days might be linked to slightly better plantation management practices such as harvesting and monitoring activities which could contribute to a small increase in FFB yield. However, the weakness of the correlation indicates that other factors likely have a stronger influence on FFB yield.

1. **Correlation between harvested and FFB yield**

The correlation coefficient of -0.35 suggests a moderately negative relationship between the area harvested and FFB yield. This means that larger harvested areas are associated with lower FFB yields. This relationship might be explained by factors like reduced plant density in larger areas. With fewer plants, there is more competition for resources like sunlight and nutrients. Additionally, larger harvested areas might be more susceptible to environmental stress which could lead to lower FFB productivity.

Overall, while precipitation and the area harvested show moderate correlations with FFB yield, soil moisture, average temperature and the number of working days exhibit weaker correlations.

*Trend Analysis (from section 2.7)*

Based on the correlation coefficients provided, the factors that exhibit a notable relationship with FFB yield are precipitation and the area harvested. Trend analysis is examined to observe the trend of precipitation and area harvested against FFB yield over year.

A graph with a line

Description automatically generated

*Graph 1: Average FFB yield over year*

A graph showing the average precipitation over time

Description automatically generated

*Graph 2: Average Precipitation over year*

A graph with blue line

Description automatically generated

*Graph 3: Average Area Harvested over year*

*(The line graph above obtained through Excel)*

The trend analysis revealed interesting insights into the relationship between external factors and FFB yield over the years. Based on the line graph shown above, the overall average FFB Yield and Area Harvested have perform an upward trend over the year while the overall average precipitation results in a downward trend over the year from 2008 to 2018.

Despite the moderate positive correlation between precipitation and FFB yield, the trend analysis showed a downward trend in precipitation levels over the years. This suggests that while precipitation may influence FFB yield positively in the short term, other factors such as soil fertility and agronomic practices may have mitigated the impact of decreasing precipitation on yield.

In examining the relationship between precipitation (Graph 2) and FBB yield (Graph 1), the moderate positive correlation between them is evident. This indicates that when the precipitation level increases, the FFB yield tends to increase as well and vice versa in the short term. This trend can be seen between the years 2010 to 2014 and 2016 to 2018.

However, the harvested area shows an upward trend over the years which may seem contradictory when considering its moderate negative correlation with FFB yield. This highlights the complexity of factors affecting yield including the time lags between changes in harvested area and their impact on yield.

The negative correlation between harvested area and FFB yield has validated through the trend analysis. This suggests that increases in harvested area tend to coincide with decreases in yield. The time lags between alterations in harvested area and their impact on yield further complicate this relationship.

It is important to note that expanding the cultivated area may initially lead to higher yields as newly planted palms mature and become productive. However, over time, the negative effects of increased competition for resources and potential declines in soil fertility may outweigh the benefits which will result in diminished yields.

*Conclusion*

To maintain healthy oil palm plantations and achieve high fresh fruit bunch (FFB) yields, there are several external factors that must be considered. These factors include soil moisture, temperature, precipitation, working days, and harvested area.

Soil moisture and temperature are crucial for plant health and indirectly impact FFB yield. Adequate rainfall is also important for good growing conditions and higher FFB yield. Plantation managers who spend more time on their plantations might have better management practices, contributing to a slight increase in FFB yield. However, larger harvested areas can lead to lower FFB yield due to reduced plant density and increased competition for resources.

Therefore, external factors play a significant role in influencing FFB yield alongside the internal factor of flowering. By understanding and managing these external factors effectively, plantation managers can create optimal conditions for both flowering and overall plant health.

**Q3.**

a. The probability of the word “data” occurring in each line is 0.1.

b. The distribution is consistent, with each line having 18 distinct words.

c. The phrase “data analytics” appears twice in the paragraph. In both instances, the word “analytics” follows the word “data.” Therefore, the probability of the word “analytics” occurring after the word “data” is 100%.