**OBJECTIVES OF THE RESEARCH**

* **To describe the socio-economic characteristics of the respondent**
* **To determine the profitability of the farmers**
* **To determine the impact of weather condition on the profitability of the farmers**
* **To determine the impact of access to facilities on the profitability of the farmers**
* **To determine the effect of the socio-economic characteristics on the profitability of the farmers.**

**DATA CLEANING**

* Drop column ‘Q.I.D’ (done)
* Clean the column “Tribe” (done)
* Check for the number of null vales in column “Location” and fill the missing values (done)
* Fill the missing values for the column “Residence length” (done)
* Clean the column ‘Sex” (done)
* Fill the missing value for column “Age” with their mean (done)
* Fill the missing value for the column “Number active in Agric” with their mean (done)
* Fill the missing value for the column “Size(Ha)” with their mean (done)
* Fill the missing value for column “Size” with their mean(done)
* Fill the missing value for column “Size.1” with their mean (done)
* Fill the missing value for column “Size (Ha)” with their mean (done)
* Fill the missing value for column “Size(Ha).1” with their mean (done)
* Fill the missing value for the column “size (Trees)” with their mean (done)
* Fill the missing value for the column “P.O output (2018) (25L Keg)” with their mean(done)
* Fill the missing value for column “Price” with their mean (done)
* Fill the missing value for the column “Qty used in 2018” with their mean (done)
* Fill the missing value for the column “Cost per unit” with their mean (done)
* Fill the missing value for the column “Transport cost” with their mean
* Fill the missing value for the column “Qty used in 2018.1” with their mean
* Fill the missing value for the column “Cost per unit.1” with their mean
* Fill the missing value for the column “Transport cost.1” with their mean
* Fill the missing value for the column “Qty used in 2018.3” with their mean
* Fill the missing value for the column “Cost per unit.3” with their mean
* Fill the missing value for the column “Transport cost.3” with their mean
* Fill the missing value for the column “Qty used in 2018.4” with their mean
* Fill the missing value for the column “Cost per unit.4” with their mean
* Fill the missing value for the column “Transport cost.4” with their mean
* Fill the missing value for the column “Qty used in 2018.5” with their mean
* Fill the missing value for the column “Cost per unit.5” with their mean
* Fill the missing value for the column “Transport cost.5” with their mean
* Fill the missing value for the column “Qty used in 2018.6” with their mean
* Fill the missing value for the column “Cost per unit.6” with their mean
* Fill the missing value for the column “Qty used in 2018.7” with their mean
* Fill the missing value for the column “Cost per unit.7” with their mean
* Fill the missing value for the column “Transport cost.7” with their mean
* Fill the missing value for the column “Qty used in 2018.8” with their mean
* Fill the missing value for the column “Cost per unit.8” with their mean
* Fill the missing value for the column “Transport cost.8” with their mean
* Fill the missing value for the column “Qty used in 2018.9” with their mean
* Fill the missing value for the column “Cost per unit.9” with their mean
* Fill the missing value for the column “Qty used in 2018.10” with their mean
* Fill the missing value for the column “Cost per unit.10” with their mean
* Fill the missing value for the column “Transport cost.10” with their mean
* Fill the missing value for the column “Contribution to H.H income” with their mean (done)
* Fill the missing value for column “Labour” with their mean (done)
* Fill the missing value for column “cost” with their mean
* Fill the missing value for column “Qty of input” with their mean
* Fill the missing value for column “Cost/unit” with their mean

Complete this for other column (Cost/unit.1-4,7,9,11,15-17,21,36) (done)

**DATA EXPLORATION**

* Categorize the column “Residence length”
* Categorize the column “Age”
* Create new column “Totalchildren” = “Malechildrenborn” + “Femalechildrenborn” (done)
* Create new column “Totaldeprelative” = “DepMale RelativeNo” +”DepFemale RelativeNo” (done)
* Create a column “Total farmsize”= “Size”+ “Size (Ha)” +“Size(Ha).1” + “Size(Ha)” +“Size.1” (done)
* Create a column “Total cost” = “Transport cost” + (“Qty used in 2018” \* “Cost per unit”) + other column (“Transport cost”.1-10 except 2)
* Create a column “Total revenue” = “P.O output (2018) (25L Keg)” \* “Price”
* Create a column “Profit” = “Total revenue” - “Total cost”
* Categorize the column “Profit”

**To determine the effect of the socio-economic characteristics on the profitability of the farmers.**

**Linear Regression Model**

Y= Independent variables= Profit

X=Dependent variables=ResidenceLength, Age,Household size, Number active in agric, TotalChildren, TotaldepRelative, Education, TotalFarmsize, size (Trees), Contribution to H.H income, times a yearfarm under-brushed, Time it take to deliver harvested bunch, Time before comencement of processing, Sex, Estate land used to grow other crops, Do you apply fertilizer,

**Describing the Socio-economics Charateristics of the respondent**

**The variables needed for the description in Tableu includes:**

Year planted

Source ofVarietyPlanted, Variety, BirthLGA, BirthTown, BirthState, HomLGA, HomState, HomTown, LGA, Location, Tribe, Sex, Age, Household size, Number active in agric, TotalDepReleative, TotalChildren, Year of education, Total farm size, Total output, Total cost, Total revenue, Time it take to deliver harvested bunch, Time before comencement of processing, Price, Profit category, Home distance

**Interpretation of the Model and Recommendation**

Independent Variable Coefficients

Household size 83,738.0

Number\_active\_in\_agric 355,510.0

size (Trees) 83.0

Years\_of\_education 283,819.0

Age 14,430.0

Home\_distance -20,849.0

Total\_children -52,613.0

Total\_dep\_relative -32,286.0

Total\_farmsize 6,185.0

Estate\_land -118,947.0

Gender -570,858.0

Ferterlizer\_application 817,788.0

The table above shows the results of the linear regression model carried out where the profit of the farm

business is the dependent variable. A unit increase in the following independent variable will lead to a decrease in the profit of the farm: **Home Distance, Total Number Of Children, Total Dependent Relatives, Estate Land** and on the other hand, a unit increase in the following independent variables will lead to an increase in the profit of the farm business: **Household size, Number\_active\_in\_agric, Years\_of\_education , Tree size, Total\_farmsize , Ferterlizer\_application**

**Description Of The Socio-Economic Characteristics**

* Supare-akoko has the highest profit and output followed by erinje and ode-aye
* Yoruba and Apoi tribe are doing very good in terms of profit and total output on the other hand igala and idoma tribe are not doing good at all.
* So its recommended that Supare-akoko, Yoruba and Apoi tribe should engage more in the farming activities.

**Description of the breed of the seed used**

Dura breed is a very good breed because even though it is not the highest in terms of output and revenue it is still the highest in terms of profit, this is as a result of the low farmsize and total cost. So its is recommended that Dura seed should be used for planting and the farmsize should be increase to increase the output and the profit