Report

**I/ Identify regular customers**

First, we have 4 classes of customer income that are arranged as follows: customers with income less than $25,000, customers whose income ranges from $25,000 to $50,000, customers with income ranging from $50,000 to $ 5,000, and customers with incomes above $75,000 (above 80%). Next, on the Pivot sheet, we can see that regular bank customers are those whose income ranges between $25,000 and $75,000, with the range of $25,000 and $50,000 making up the largest proportion. Overall, between $ 25,000 and $ 50,000, 36.1 average age is the most frequent customer. On the other hand, in the range of $ 50,000 to $ 75,000, 35.3 average age is the most frequent customer. From that, it can be concluded that the potential customers of the bank are people aged 35 and 36 with incomes ranging from $ 25,000 to $ 75,000.

**II/ Recognize the correlation between independent variables in banking data**

The following is a result of the Correlation sheet that shows the correlation between the factors:

Correlation result between Income and Household Wealth is 0.947: this means that Income and Household Wealth have a very strong relationship in positive direction, which means when the Income increases, the Household Wealth surely increases.

Correlation result between Income and Average Bank Balance is 0.952: this means that Income and Average Bank Balance have a very strong relationship in positive direction, which means when the Income increases, the Household Wealth surely increases.

Correlation result between Income and Home Value is 0.795: this means that Income and Home Value have a pretty strong relationship in positive direction, which means when the Income increases, the Household Wealth also increases.

In conclusion, Income is a prerequisite and has an essential relationship with financial data variables in banking data in a positive direction. Thereby, it can be said that the higher the income customers have, the greater the value of assets and money in the bank.

**III/ Is it possible to use independent variables to predict the Household Wealth in each average age?**

First of all, I filtered the data about Household Wealth and the independent variables that I think are associated including Income, Home Value, and Average Bank Balance. Then perform the Regression in the Data Analysis tool to get the results returned in the Linear Regression sheet. In fact, the square root coefficient R is a function that does not decrease with the number of independent variables included in the model, the more we add the independent variables to the model, the more the R square increases. And because of this is a multivariate linear regression model as well as comparing Adjusted R Square value smaller than R Square value, I will use Adjusted R Square to evaluate the fit of the model that will be safer because the Adjusted R Square does not necessarily increase when we add independent variables into the model. Thus, the value we get is 0.928, which is nearly 93%. Because this value is over 50%, this linear regression model can be assessed very well and fits the data set at 93%. In addition, the sig value of the F-test is an extremely small number that is definitely less than 0.05. Thus, the linear regression model is consistent with the overall. Finally, we have an equation used to predict the Household Wealth value using the following formula:

y = 1.785x - 0.249y + 3.598z - 40993.818

x: Income

y: Home Value

z: Average Bank Balance