# MODULE 1 UNIT 3

## Activity submission

Learning outcome:

LO5: Use Tableau to summarise and visualise a data set.

### Name: Matlotlo Magasa

#### 1. Instructions and guidelines (Read carefully)

##### Instructions

1. Insert your name and surname in the space provided above, as well as in the **file name.** Save the file as: **First name Surname M1U3 Activity Submission** – **e.g. Lilly Smith M1U3 Activity Submission.** **NB:** *Please ensure that you use the name that appears in your student profile on the Online Campus.*

2. Write all your answers in this document. There is an instruction that says, “Start writing here” under each question. Please type your answer there.

3. Submit your assignment in **Microsoft Word only**. No other file types will be accepted

4. Do **not delete the plagiarism declaration** or the **assignment instructions and guidelines**. They must remain in your assignment when you submit.

**PLEASE NOTE:** **Plagiarism cases will be investigated in line with the Terms and Conditions for Students.**

**IMPORTANT NOTICE:** Please ensure that you have checked your course calendar for the due date for this assignment.

##### Guidelines

1. There are 8 pages and 4 questions in this assignment.

2. You will be required to download a data file, titled LOANS.sav, in order to interact with the data in Tableau. This is available in the Module 1 downloads folder. Ensure you have access to this file before attempting to answer the questions.

3. Make sure that you have carefully read and fully understood the questions before answering them. Answer the questions fully but concisely and as directly as possible. Follow all specific instructions for individual questions (e.g. “list”, “in point form”).

4. Answer all questions in your own words. Do not copy any text from the notes, readings or other sources. **The assignment must be your own work only.**

|  |
| --- |
| **Plagiarism declaration:** |
| **1. I know that plagiarism is wrong. Plagiarism is to use another’s work and pretend that it is one’s own.**  **2. This assignment is my own work.**  **3. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.**  **4. I acknowledge that copying someone else’s assignment (or part of it) is wrong and declare that my assignments are my own work.** |

#### 2. Mark allocation

Each question receives a mark allocation. However, you will only receive a final percentage mark and will not be given individual marks for each question. The mark allocation is there to show you the weighting and length of each question.

Question 1 5

Question 2 5

Question 3 5

Question 4 5

**TOTAL 20**

#### 3. Questions

Note:

The questions included in this activity submission are based on the data contained in the data set UCT BAN M1U3 Activity submission data, available in the Module 1 “Downloads” folder. Ensure that you have downloaded and connected to this data in Tableau before answering the questions.

You are working for a financial services company called Speedy Loans that provides loan services to its customers. The board of directors has asked you to analyse the company’s data and provide specific information about its customer base. You have been provided with a data set containing information on several customer attributes, as shown in Table 1. After your analysis, you will be tasked with presenting a report to upper management.

Table 1: Description of data set variables.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Variable Name** | **Variable Description** |
| 1 | Age | The age of the customer (in years). |
| 2 | Education | The education level of the customer (Primary, Secondary, Some Tertiary, and Tertiary). |
| 3 | Year\_emp | The number of years the customer has worked for their current employer. |
| 4 | Income | The customer’s household income (in months). |
| 5 | Debt\_income | The customer’s debt-to-income ratio. |
| 6 | Cred\_debt | The customer’s current credit card debt. |
| 7 | Other\_debt | The sum of the customer’s remaining debt, excluding credit card debt. |
| 8 | Loan | An indication whether the customer’s loan application was approved or denied (0: approved; 1: denied). |
| 9 | Location | The city where the customer lives (Johannesburg, Cape Town, or Durban). |
| 10 | Risk\_score | The customer’s credit risk score. |

For this activity submission, you will be required to create three worksheets and then combine these to create a single dashboard. Questions 1 to 3 will require you to create one worksheet per question, while Question 4 will ask you to create the dashboard.

For each question, follow the instructions to create the relevant worksheet or dashboard, then export the results as an image and paste it in under the relevant question. You are also required to respond to questions about the accompanying visual. Both your written responses and your images will be graded.

##### Question 1

The board is interested in the risk distribution of their customer base, and management has asked you to visualise this by creating a histogram of the “Risk score” variable. Change the x-axis title to “Frequency” and the y-axis title to “Risk score”. Rename this worksheet “Distribution of risk”. If desired, you can modify the graph by changing the colour of the bars or the size and font of the worksheet title.

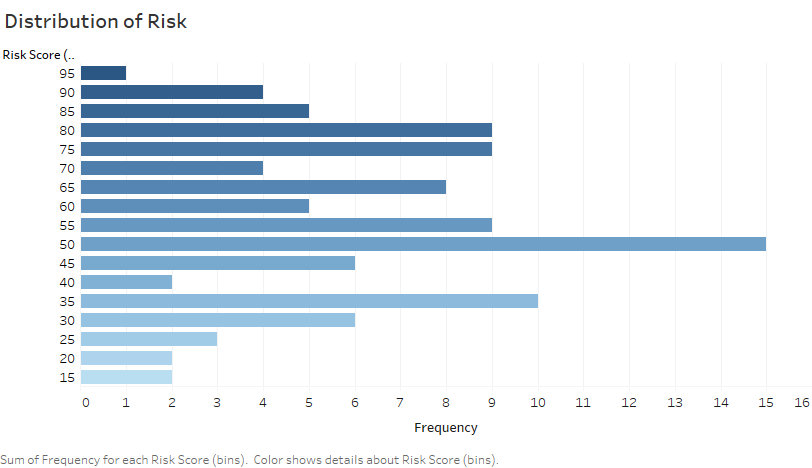
Why do you think the board asked you to use a histogram here? What value can be extracted from the histogram you have created? (Max. 150 words)

Start writing here:

As a loan providing company, it is important to understand which customer(s) are likely to pay back loans in a timely manner. This will help the company know how much you may qualify for and whether you will qualify, based on your credit score. Such information protects the company financially from potential ‘bad payers’ and helps the company judge the financial risk the client may pose.

From the histogram, it can be deduced that 1 % of the customer base have a very high-risk score of more than 95. The higher the credit risk score, the higher the risk the customer will pose to the company. Fifteen percent of the customers (15 %) of customer base, have an average risk score of between 46 and 50. These customers are the least predictable in how they will handle their loan repayments. Only 2 % of the customers are most likely to be very consistent at repayments. It is important to note that the majority of the customers fall above the 50 point credit risk score.

Paste your worksheet image here:



##### Question 2

The board is interested in understanding whether any relationship exists between the customer’s household income and their current credit card debt. To do this, they have asked you to create a worksheet with a scatter plot showing this relationship. Income should be on the x-axis and credit card debt (Cred debt) should be on the y-axis.

In order to view individual data points on the plot, be sure to change each variable to a “Dimension”, as opposed to a “Measure”, which is automatically assigned by Tableau. Rename this worksheet “Relationship between income and credit card debt”.

If desired, you can modify the graph by changing the colour and size of the data points or the size and font of the worksheet title. You can also right-click on the graph and add a trend line to gain more information and statistics on the relationship between income and credit card debt.

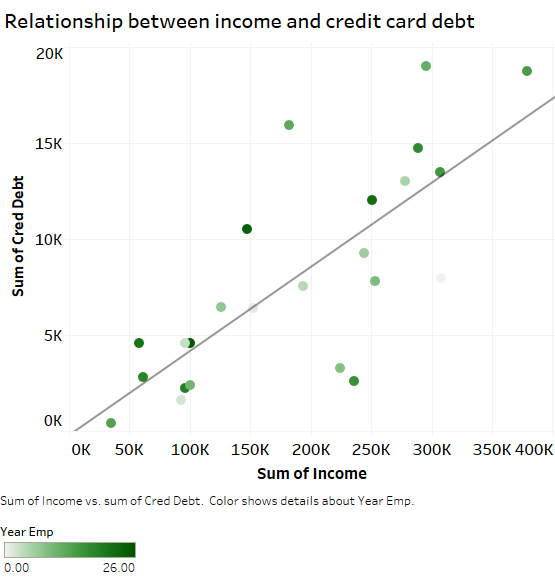
Why would the board ask you to create this graph?

(Max. 150 words)

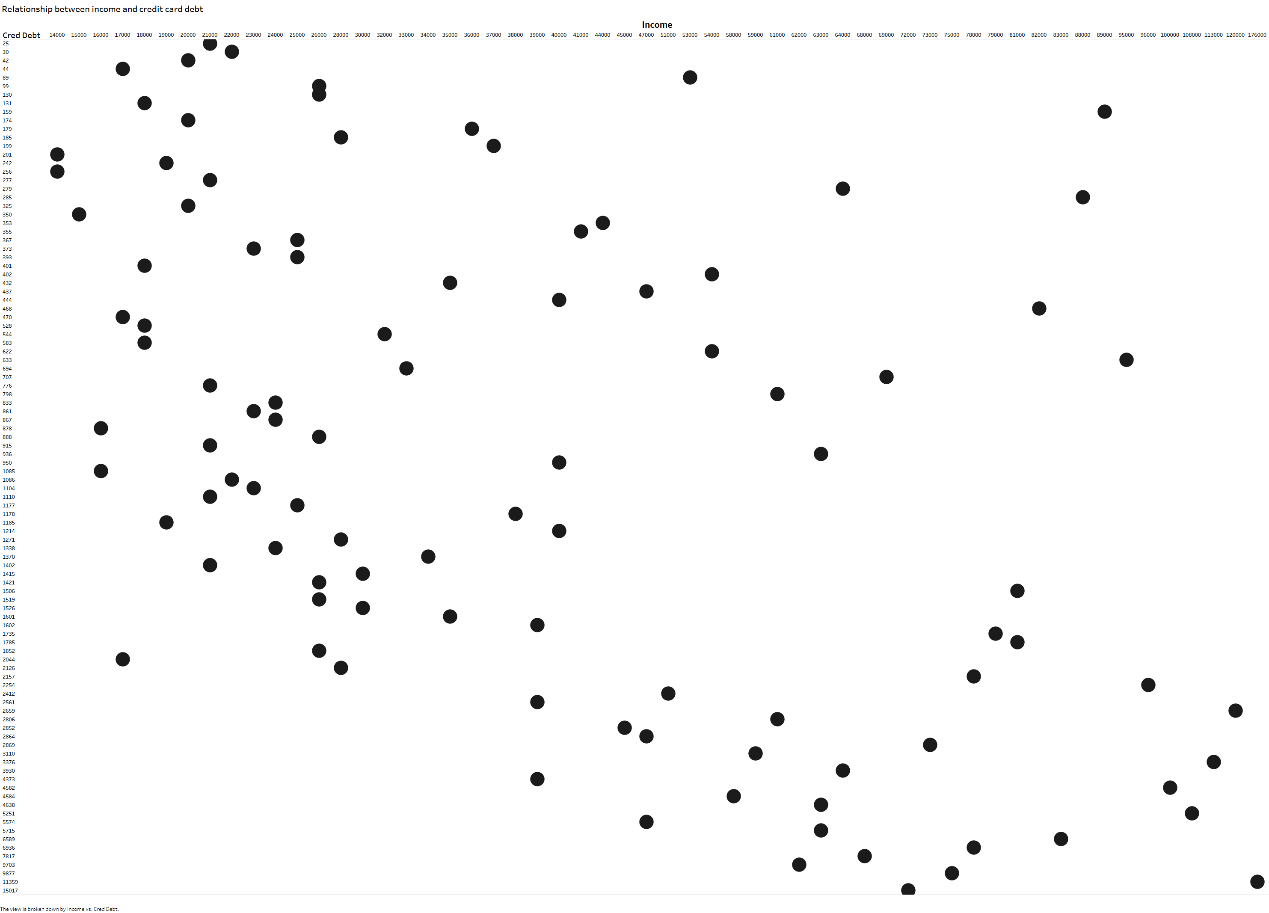
Start writing here:

This graph would be to determine whether there is any correlation of credit debt and the household income. This is to determine whether loans should be given based on the household income (should there be a correlation) or not.

Paste your worksheet image here:



Graph for the individual data points I presented below. This is not visually the best representation of the data. It was not possible to add a trend line to this plot even though both variables were placed on “dimension” instead of measure.



##### Question 3

The final worksheet the board wants you to create is a bar graph showing the average risk score of customers based on their education level. Education should appear on the x-axis and the **average** risk score should appear on the y-axis.

You will notice that Tableau automatically sorts the education level alphabetically, not ordinally. This can be modified by right-clicking on the “Education” variable at the top of the worksheet, selecting “Sort”, and then selecting “Manual”.

You should also remember that Tableau automatically assigns each measure as the “SUM”. This will need to be changed to allow the “Risk score” variable to measure the average. Rename this worksheet “Education level and risk”. If desired, you can modify the graph by changing the colour of the bars or the size and font of the worksheet title.

Why might the board be interested in analysing a customer’s risk score based on their education level? What is the graph inferring?

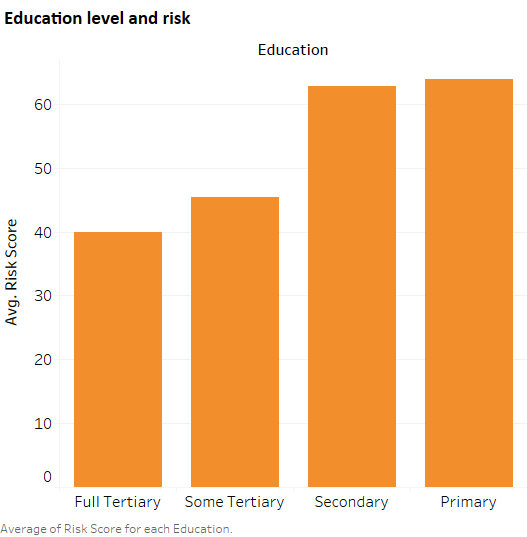
(Max. 150 words)

Start writing here:

Education is a key contributor to the knowledge of a customer. It is expected that the highly educated would make better financial decisions as opposed to the less educated and hence be of less risk to the company. Investigating the relationship between education and credit risk score will also help determine how the company should make the decision of giving loans in such a way to reduce the risk to themselves.

The graph suggests that the people with only lower education levels (primary and secondary schooling) pose a higher risk to the company because they have high average credit risk ratings. The most educated people seem to indeed make better financial decisions when it comes to paying back their loans as they have a lower average credit risk rating.

Paste your worksheet image here:



##### Question 4

The board has now asked you to compile the worksheets you have created into a single dashboard. When creating the dashboard, consider the following specifications provided by the board:

* Ensure that all three worksheets are represented on the dashboard.
* The bars of the histogram and bar chart should be in different colours to prevent confusion at what is being viewed. Blue is suggested for the histogram and orange for the bar graph.

Finally, the board has asked you to provide a brief summary of the key findings the dashboard illustrates.

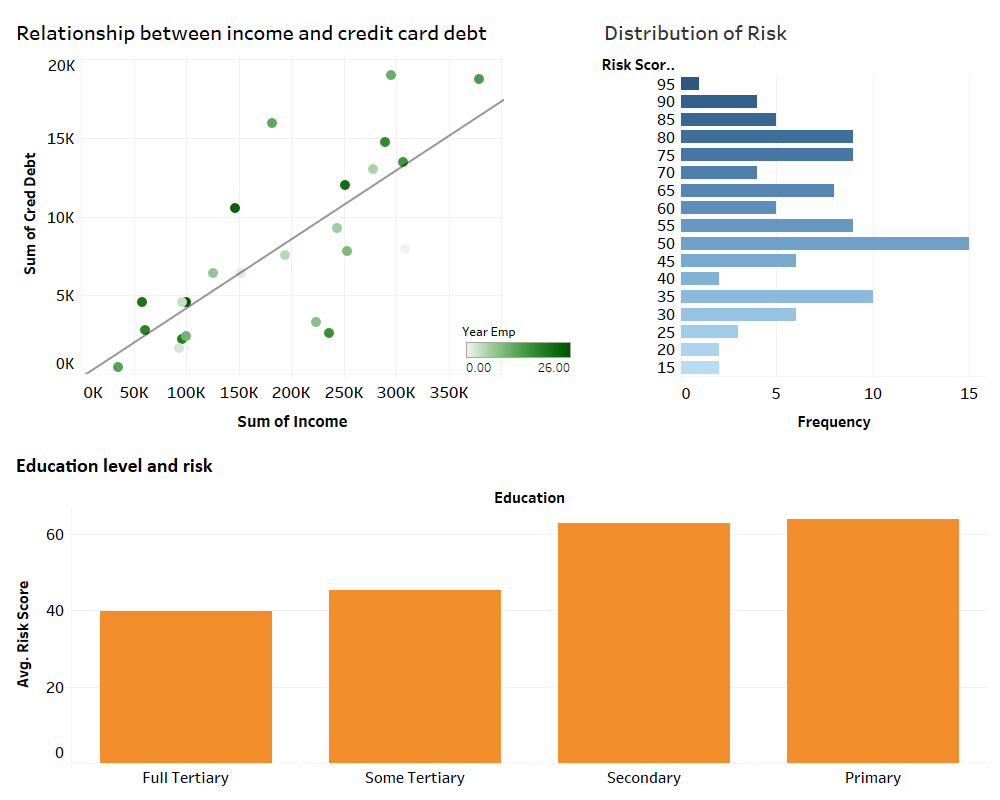
(Max. 150 words)

Start writing here:

The data shows that there is a high distribution of people who take longer paying back their loans. It also shows that there’s a negative correlation between average risk score and level of education. The two graphs depicting the income and credit debt both show a positive correlation between the income and the credit card debt. People who get paid more will take up more debt and are the clients who likely source out loans/ credit. This is despite the number of years they have been employed.

In conclusion: The less educated take up less credit debt however will generally take longer than the ‘more educated ‘ to pay it back, while the educated may pay the debts back on time, but they may take up a lot more credit and have a lot more debt (assuming those more educated earn more, this may require more analysis to confirm)).

Paste your dashboard image here:



#### 4. Rubric

The following rubric will be used to grade your answers:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Unsatisfactory (0)** | **Limited (2)** | **Accomplished (3)** | **Exceptional (5)** |
| **Question 1**  *The correct graph has been created and it adheres to the instructions given. The graph contains the correct variables and the written response addresses the question.* | No submission.  OR  The response does not address the question. | The image supplied depicts the incorrect graph, or the incorrect variables are used.  AND  The written response does not address the question adequately. | The image supplied depicts the correct graph using the correct variables, but not all instructions are adhered to.  AND  The written response addresses the question adequately. | The image supplied depicts the correct graph using the correct variables, and all instructions are adhered to.  AND  The written response addresses the question exceptionally. |
| **Question 2**  *The correct graph has been created and it adheres to the instructions given. The graph contains the correct variables and the written response addresses the question.* | No submission.  OR  The response does not address the question. | The image supplied depicts the incorrect graph, or the incorrect variables are used.  AND  The written response does not address the question adequately. | The image supplied depicts the correct graph using the correct variables, but not all instructions are adhered to.  AND  The written response addresses the question adequately. | The image supplied depicts the correct graph using the correct variables, and all instructions are adhered to.  AND  The written response addresses the question exceptionally. |
| **Question 3**  *The correct graph has been created and it adheres to the instructions given. The graph contains the correct variables and the written response addresses the question.* | No submission.  OR  The response does not address the question. | The image supplied depicts the incorrect graph, or the incorrect variables are used.  AND  The written response does not address the question adequately. | The image supplied depicts the correct graph using the correct variables, but not all instructions are adhered to.  AND  The written response addresses the question adequately. | The image supplied depicts the correct graph using the correct variables, and all instructions are adhered to.  AND  The written response addresses the question exceptionally. |
| **Question 4**  *The dashboard displays correctly, and the overall instructions given for its creation have been adhered to. The written response provides a summary of key findings.* | No submission.  OR  The response does not address the question. | The dashboard does not display correctly, and instructions have not been adhered to adequately.  AND  The written response does not provide a summary. | The dashboard displays correctly, and instructions have partly been adhered to.  AND  The written response provides an adequate summary. | The dashboard displays correctly, and the instructions have been adhered to fully.  AND  The written response provides an excellent summary. |

**Total:** 20 marks