# MODULE 2 UNIT 2

## Activity submission

Learning outcomes:

LO4: Practise creating dashboards and stories in Tableau.

LO5: Analyse the insights extracted from data.

### 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 M2U2 Activity Submission** – **e.g. Lilly Smith M2U2 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, “Paste the calculated fields formula you created here” and “Paste the visualisations you created here” under each question. Please insert the relevant formulae and visualisations there.

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

4. You will be required to include visualisations that you have made in Tableau in this activity. To do this, export them as an image file and paste them into this document where indicated. Tableau has a page that details how to export your Tableau view as an image file.

5. 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. Return to the Module 2 downloads folder and save a copy of the data file before attempting to answer the following 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 4

Question 2 4

Question 3 4

Question 4 8

**TOTAL 20**

#### 3. Questions

In Module 1, you were introduced to the company Speedy Loans, a financial services company that provides lines of credit to its customer base. As a new data analyst at the company, you were tasked with creating three worksheets and a single dashboard that visualised some basic analyses using Tableau.

Upper management has now asked you to report on a few additional variables that could provide insight into their customer base, specifically how credit risk varies by customer location and education in South Africa. They believe this information is relevant, since Cape Town and Johannesburg are known to have higher livings costs than Durban, for example, and education levels can have a profound influence on how people manage their finances.

Upon scanning through the data set, you notice that the variables management hopes to visualise are absent. As a result, you decide to use calculated fields in Tableau to create new variables.

For this activity submission, you will be required to use calculated fields to create three unique variables that can be visualised in Tableau. You will then combine these worksheets into dashboards and create a visual story for the final question, in which you will provide a narrative of what the data shows. Each question has specific instructions on what is expected. After creating the new variables, you will be required to copy the calculated field formula into your answer, along with the visualisations of the results.

Note:

The questions included in this activity submission depend on the data contained in the data set UCT BAN M2U2 Activity submission data available in the Module 2 downloads folder. Ensure that you have downloaded and connected to this data in Tableau before answering the questions. You will be required to include visualisations that you have made in Tableau in this activity. To do this, export them as an image file and paste them into this document. Tableau has a page that details how to export your Tableau view as an image file.

##### Question 1

The company board is interested in determining whether there is any correlation (or relationship) between a customer’s income and their risk score based on where they live and on their education level.

To do this, use the calculated fields **correlation function** to create a new variable called “Correlation (Income:Risk)”. This new variable can then be plotted against “Location” and “Education” in two separate worksheets using a bar graph for each visualisation.

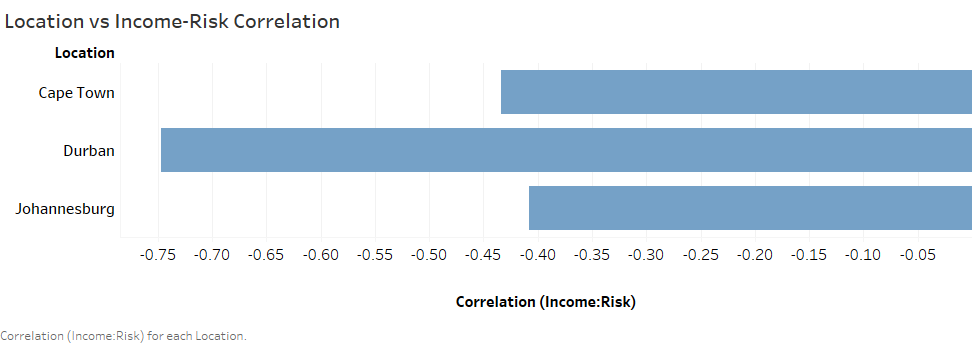
“Correlation (Income:Risk)” should be on the x-axis and “Location” or “Education” should be on the y-axis. Ensure that the two graphs you create are in different colours and that you provide an appropriate title for each worksheet. You should also take care to order education by level from lowest to highest: primary, secondary, some tertiary, and full tertiary.

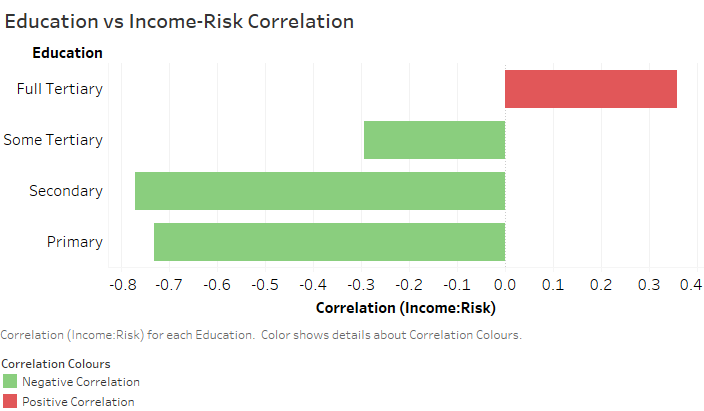
Paste the calculated fields formula you created here:

CORR( [Income],[Risk Score])

IF [Correlation (Income:Risk)] > 0 THEN 'Positive Correlation' ELSE 'Negative Correlation' END

Paste the visualisations you created here:





##### Question 2

Next, the board has asked you to calculate the **credit card** **debt-to-income ratio** in order to decide whether customers are spending too much on credit relative to their income to offer them a loan. The company considers a credit card debt-to-income ratio below a 15% threshold acceptable.

To do this, use a calculated field to create a new variable using the “Cred­\_debt” and “Income” variables. While there is no ratio function in Tableau, refer to the lesson in Unit 1 to remind yourself of the formula. When using this formula, be sure to use the average credit card debt and average income in your calculation.

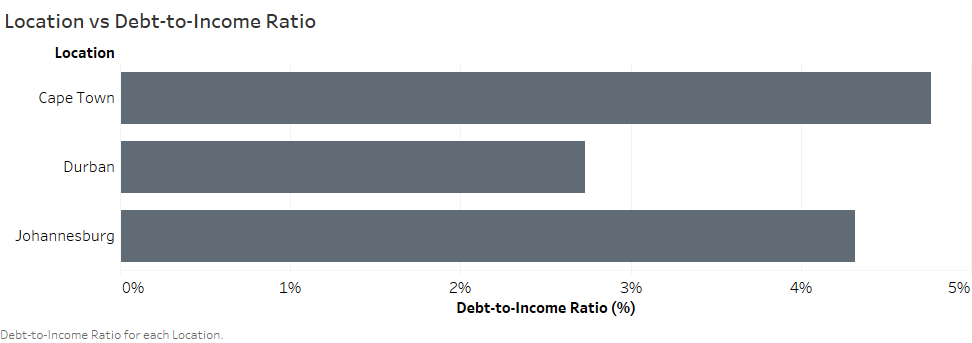
Label this new variable “Debt-to-Income Ratio” and plot it against “Location” and “Education” in two separate worksheets using bar graphs. In order to change the debt-to-income ratio to a percentage, right-click on the “Debt-to-Income Ratio” variable in the “Measures” pane, select “Default Properties”, then select “Number format…”. Change the number format to percentage with zero decimal places.

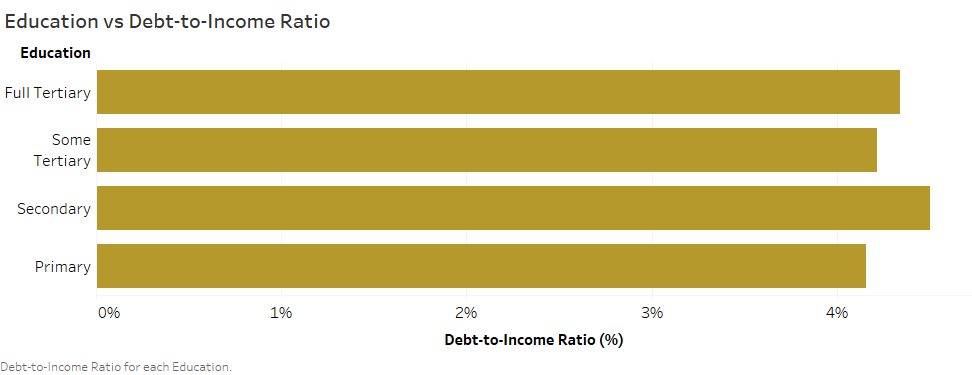
For these visualisations, “Debt-to-Income Ratio” should be on the x-axis and “Location” or “Education” should be on the y-axis. Ensure that the two graphs you create are in different colours and that you provide an appropriate title for each worksheet. You should also take care to order education by level from lowest to highest: primary, secondary, some tertiary, and full tertiary.

Paste the calculated fields formula you created here:

AVG([Cred Debt])/AVG([Income])

Paste the visualisations you created here:





#### Question 3

The board has also informed you that customers with credit card debt exceeding R10,000 are typically considered high-risk, whereas customers with any amount of debt below this are considered low-risk. You have been asked to create a variable to differentiate customers as low-risk and high-risk based on the aforementioned threshold and investigate how this differs according to location and education.

To do this, use a logical calculated fields function, such as an **If** **statement**. Using calculated fields, create a variable that categorises customers in the following way: **if** credit card debt is above R10,000, the customer is categorised as “HIGH-RISK”, **else** they are categorised as “LOW-RISK”. Name this variable “Low- vs. High-risk”. Refer to the notes in Unit 1 if you need a reminder of how to create logical calculated fields.

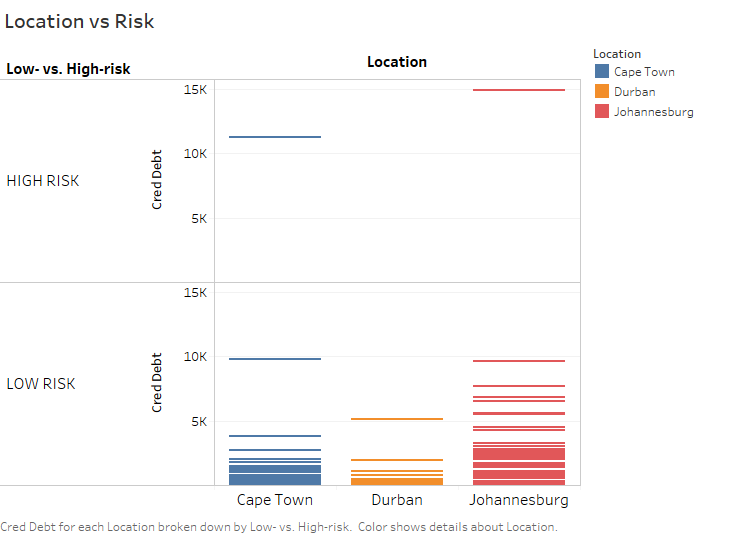
You will notice that Tableau creates a “Dimensions” variable, since it has generated a categorical variable. Drag this variable into the “Rows” field and plot it against “Location” or “Education” in the columns field. Next, drag “Cred\_debt” into the rows field and change the “Cred\_debt” variable from a measure to a dimension. By changing the variable to a dimension, Tableau increases the level of detail in the visualisation so that you can view each customer as a separate data point on the graph.

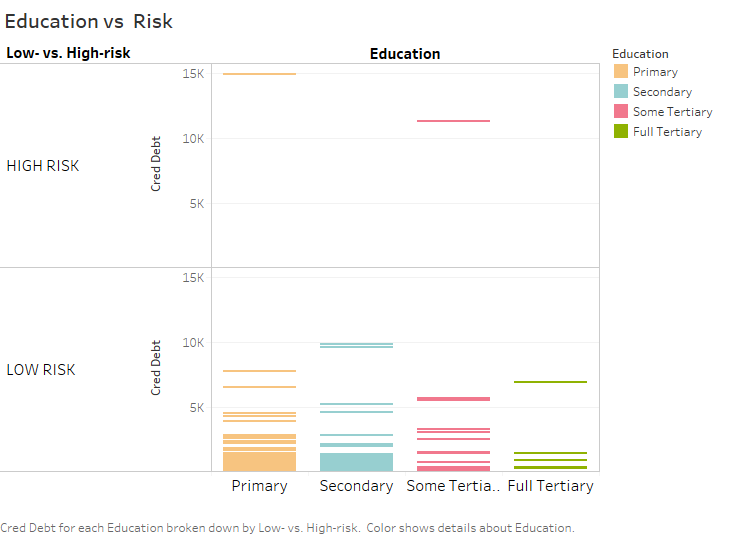
If desired, you can change the colour of each location or education category by dragging the “Location” or “Education” variable into the “Marks” pane and dropping it onto the “Color” field. This will automatically adjust the colours for you. For the graph plotted against “Education”, remember to order the data from lowest level of education to highest, as you did previously, and ensure that you provide an appropriate title for each worksheet.

Paste the calculated fields formula you created here:

IF [Cred Debt] > 10000 THEN 'HIGH RISK' ELSE 'LOW RISK' END

Paste the visualisations you created here:





#### Question 4

Finally, the board would like you to combine the results of the analyses into a visual story that you will present to the team. They have given you the following instructions:

* Create two dashboards to organise the results of the analyses in the worksheets.
* Combine the dashboards into a visual story containing two story points (or slides), adding the appropriate captions and an appropriate title for the story.
* Provide a written summary of the story that highlights the main findings of the analyses. In this write-up, draw conclusions about how customer risk varies based on location and education level. Consider whether the data indicates any education level or customer location as being indicative of greater or lower risk.

(Max. 250 words)

Paste the story points you created here:

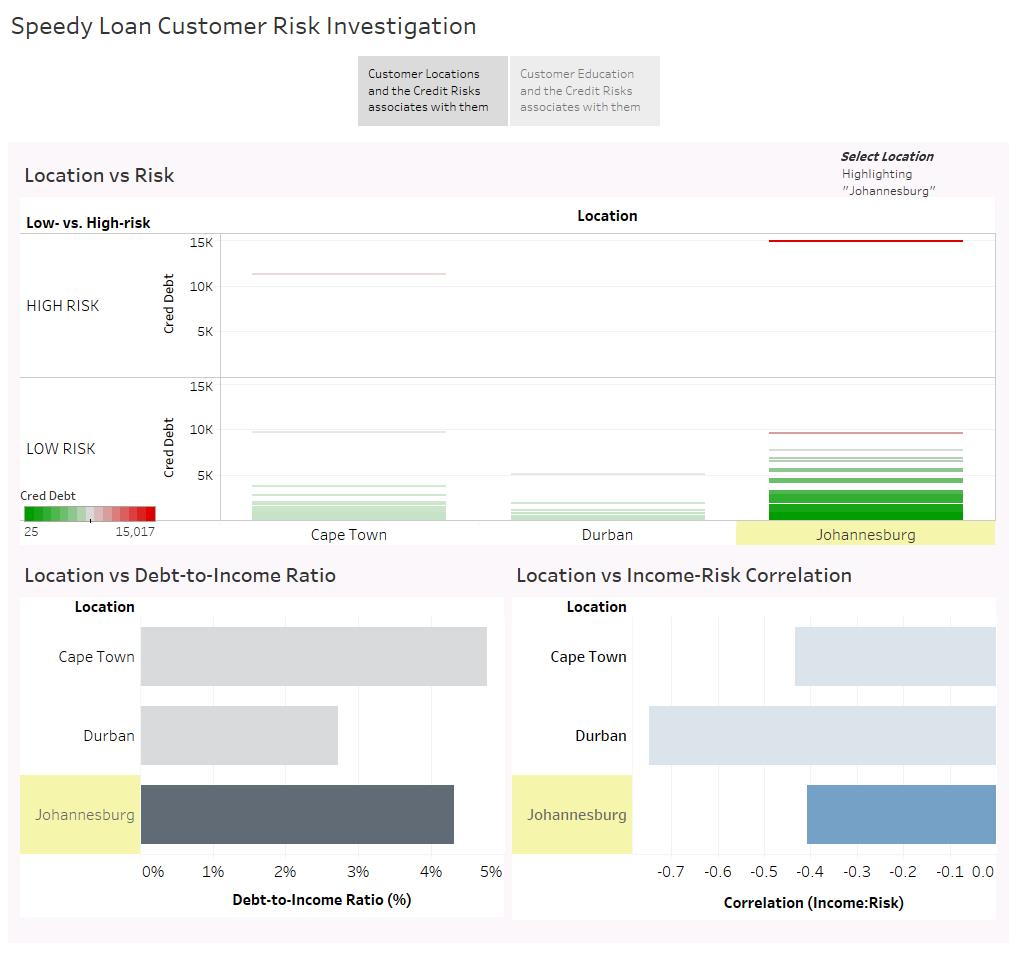


Figure 1: Slide 1 of Story

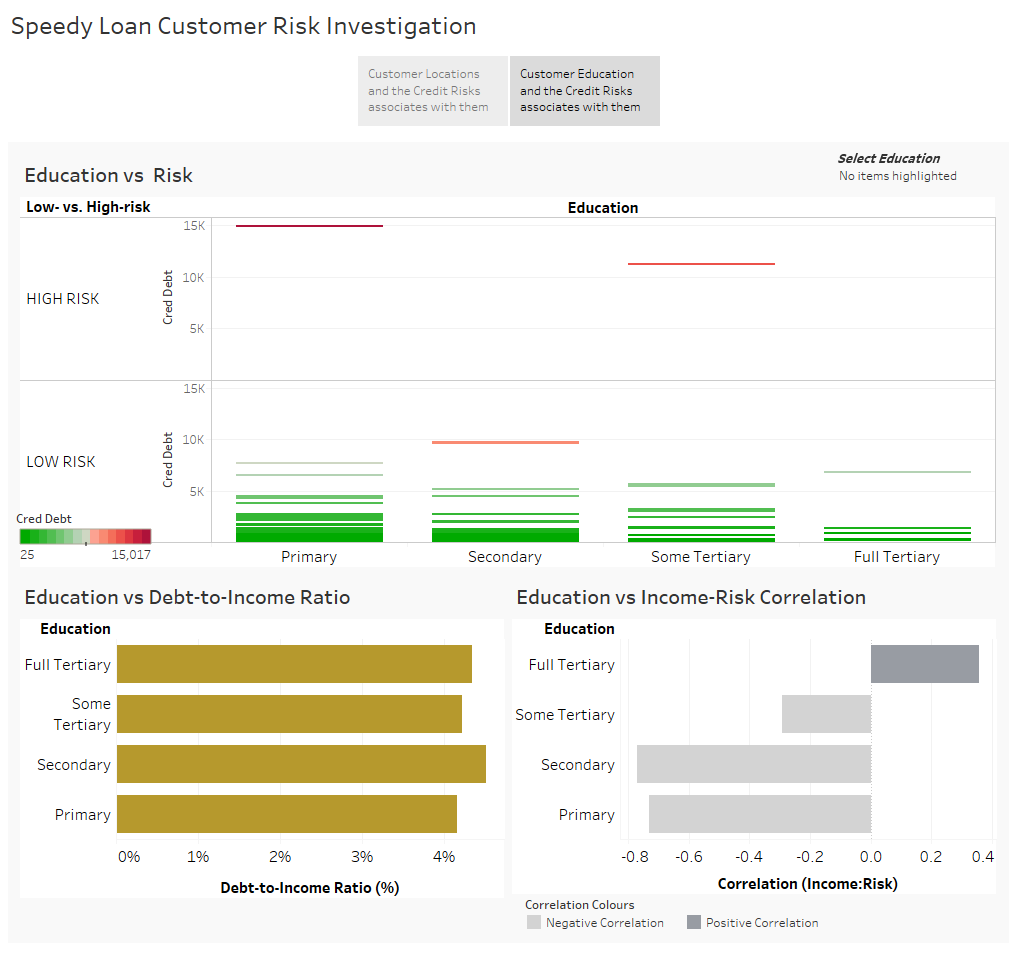


Figure 2: Slide 2 of Story

Provide a written narrative of the story here:

All locations show a negative correlation between income and risk. Durban shows a greater inverse relationship between these two variables. Meaning, for Durban customers, the inverse relationship between income and credit risk is stronger than the other cities. Thus, you would find more people in Durban who are likely to have a lower credit risk with a higher income and vice versa.

The *inverse* relationship between income and risk generally decreases the more a customer is educated, however, at tertiary level there is a positive yet weak correlation between the customers income and the risk they pose. Customers with secondary school education have the strongest correlation between the variables and are likely to be lower risks at high income profiles.

With reference to location, all the customers are below the 15% debt-to-income threshold and are not spending much on credit relative to their income, this percentage is between 4% and 5% with Cape Town and Johannesburg. Durban customers have a significantly low debt-to-income ratio compared to the other two cities.

There is not much difference with the debt-to-income ratio with education level and the ratio is similar to that given with respect to location (between 4 % and 5%).

Johannesburg and Cape Town each have a customer who is categorized as a high-risk customer due to amount of debt they have. Analysing both graphs, the customers from the cities have Primary and Some Tertiary education level respectively.

#### 4. Rubric

The following rubric will be used to grade your answers:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Unsatisfactory** | **Limited** | **Accomplished** | **Exceptional** |
| **Question 1**  *The calculated field is used correctly in Tableau with the correct variable(s). The correct graphs are provided and the instructions for generating the graphs are adhered to.* | No submission.  OR  The response does not address the question. (0) | The calculated field is used incorrectly in Tableau, or the incorrect variable(s) are used.  AND  The incorrect graphs are provided. (1) | The calculated field is used correctly in Tableau, but the incorrect variable(s) are used.  AND  The correct graphs are provided, but the instructions are not fully adhered to. (3) | The calculated field is used correctly in Tableau with the correct variable(s).  AND  The correct graphs are provided and the instructions are fully adhered to. (4) |
| **Question 2**  *The calculated field is used correctly in Tableau with the correct variable(s). The correct graphs are provided and the instructions for generating the graphs are adhered to.* | No submission.  OR  The response does not address the question. (0) | The calculated field is used incorrectly in Tableau, or the incorrect variable(s) are used.  AND  The incorrect graphs are provided. (1) | The calculated field is used correctly in Tableau, but the incorrect variable(s) are used.  AND  The correct graphs are provided, but the instructions are not fully adhered to. (3) | The calculated field is used correctly in Tableau with the correct variable(s).  AND  The correct graphs are provided, and the instructions are fully adhered to. (4) |
| **Question 3**  *The calculated field is used correctly in Tableau with the correct variable(s). The correct graphs are provided and the instructions for generating the graphs are adhered to.* | No submission.  OR  The response does not address the question. (0) | The calculated field is used incorrectly in Tableau, or the incorrect variable(s) are used.  AND  The incorrect graphs are provided. (1) | The calculated field is used correctly in Tableau, but the incorrect variable(s) are used.  AND  The correct graphs are provided, but the instructions are not fully adhered to. (3) | The calculated field is used correctly in Tableau with the correct variable(s).  AND  The correct graphs are provided and the instructions are fully adhered to. (4) |
| **Question 4**  *The correct Tableau story points are provided with appropriate titles and captions. The written analysis of the results is clear and adequate.* | No submission.  OR  The response does not address the question. (0) | Only one correct Tableau story point is provided and the written response does not adequately address the results of the analysis.  OR  The Tableau story points are provided with appropriate titles and captions, but no written response is provided. (3) | The correct Tableau story points are provided, but no titles or captions are provided.  AND  The written response only partially addresses the results of the analysis. (6) | The correct Tableau story points are provided with appropriate titles and captions.  AND  The written response fully and clearly addresses the results of the analysis. (8) |

**Total:** 20 marks