Reflective Document

Student Id:

## Part 1

**Describe your approach to this part of the assessment**

My approach to this part of the assessment involved deploying various concepts and techniques learned during class sessions. I carefully reviewed all the provided learning materials multiple times to ensure a deep understanding. I started by critically analyzing the requirements of each task and breaking them down into smaller, manageable components. This systematic breakdown enabled me to address each task individually while maintaining an integrated approach to ensure coherence among the tasks. My primary focus was on implementing functions in Python as per the requirements, which included adhering strictly to the expected outputs. Additionally, I expanded my learning resources by referring to external materials like W3Schools and YouTube tutorials. These resources provided supplementary knowledge that helped me address challenges in the coding process. This multi-faceted approach helped me build confidence and ensure the solutions aligned closely with the illustrative examples provided in the brief.

**Did you find it easy or difficult? Why?**

I found this part quite challenging. One major difficulty stemmed from the requirement to define all functions without relying on imported libraries. This restriction pushed me to fully understand the underlying logic behind Python functions rather than depending on pre-built modules. While it was mentally taxing, successfully overcoming challenges brought immense joy and a sense of accomplishment. Another difficulty arose from the strict formatting requirements, such as replicating the exact format of the results as illustrated in the brief. For instance, in Task 1, I initially struggled to ensure the column data appeared as integers rather than strings. In Task 5, I encountered issues with incorporating a specific character (\*) between columns. However, I found certain tasks, like calculating the Pearson Correlation Coefficient for two data lists, less daunting due to my mathematics background. This understanding made interpreting formulas and implementing them in Python relatively straightforward.

**What problems did you encounter?**

I encountered several problems during this part of the assessment. Key issues included:

1. Formatting Results: In Task 1, the column data initially appeared as strings instead of integers. In Task 5, my code ran correctly but failed to display the character (\*) between columns as required.

2. Static Data in Functions: After running my code in Task 1, I noticed that the elements for the header ‘age’ remained static while the headers changed dynamically. This error propagated to Tasks 4 and 5, affecting their outputs.

3. Avoiding Rounding Errors: While calculating the Pearson Correlation Coefficient without libraries, I had to ensure precision to avoid rounding errors or incorrect answers.

**How did you overcome them?**

To resolve these issues: 1. \*\*Formatting Results:\*\* I addressed the issue of column data appearing as strings by converting the data type. Initially, converting to integers led to errors with decimal values, so I changed the data type to float, which worked for both integers and decimals. For Task 5, I used the join() method, assigning the character ‘\*’ to join row elements ("\*".join(row)). 2. \*\*Static Data in Functions:\*\* I sought advice from my coursemates, who pointed out the error in my function’s logic. With their guidance, I adjusted my approach, ensuring the elements changed dynamically along with the headers. 3. \*\*Avoiding Rounding Errors:\*\* I validated input data by converting all elements to floats and implemented safeguards to handle edge cases, such as zero denominators.

**Identify any strengths/weaknesses of the approach you took**

\*\*Weaknesses:\*\* - I struggled with creating functions to execute the tasks due to a limited programming background. - My over-reliance on external resources occasionally slowed my progress. \*\*Strengths:\*\* - Collaboration with peers and tutors provided valuable insights. - Persistence and a willingness to revisit errors until resolved strengthened my problem-solving skills.

**How could this approach be improved?**

The approach could be improved by dedicating more time to practicing Python programming, especially function creation, to build a stronger foundation. Additionally, breaking tasks into smaller, testable units would streamline the debugging process.

**Suggest an alternative approach you could have taken**

An alternative approach would have been to leverage built-in functions and libraries, if permitted, to simplify tasks and focus more on understanding the logic behind the code.

## Part 2

**Describe your approach to this part of the assessment**

My approach to this part was shaped by the structure of the questions, which emphasized data exploration and analysis. The tasks provided an opportunity to delve deeper into data handling and visualization techniques in Python, encouraging further research and practice. This part consisted of three tasks: 1. \*\*Exploring Data and Identifying Patterns:\*\* - I examined the datasets (Part\_2a and Part\_2b) and observed a relationship between them. Part\_2a contained student demographics, while Part\_2b detailed student activities. I merged the datasets to facilitate analysis. - I decided to investigate the relationship between click events and student scores to identify patterns. Using Pandas for data handling and Matplotlib and Seaborn for visualization, I plotted a histogram of click events. This revealed a right-skewed distribution, indicating a higher frequency of students with lower click events. - To present actionable insights, I created a new variable, ‘Engagement Level,’ categorizing click events into ‘Low,’ ‘Moderate,’ ‘High,’ and ‘Very High’ levels. This classification helped highlight trends that could aid in improving student performance. 2. \*\*Identifying and Removing Outliers:\*\* - Initially, I used the Z-Score method to detect outliers but found it unsuitable for the skewed dataset. After further research, I opted for the Interquartile Range (IQR) method, which better handled the dataset’s distribution. 3. \*\*Hypothesis Testing:\*\* - I began by formulating null and alternative hypotheses to explore the relationship between click events and scores. - Based on research, I selected the Pearson Correlation Coefficient for the analysis.

**Did you find it easy or difficult? Why?**

Initially, I found this part challenging due to unfamiliarity with certain techniques, such as detecting and removing outliers. For instance, using the Z-Score method without understanding its limitations led to errors. However, deeper research helped me identify the more appropriate IQR method for skewed datasets. Hypothesis testing also posed difficulties, especially in determining the suitable test for two numerical variables. Reviewing module materials and seeking guidance from tutors provided clarity and confidence to complete this task.

**What problems did you encounter?**

1. \*\*Detecting and Removing Outliers:\*\* - The Z-Score method did not work well with the skewed dataset, leading to incorrect results. 2. \*\*Hypothesis Testing:\*\* - I struggled to identify the most appropriate test for analyzing the relationship between click events and scores. 3. \*\*Visualization Challenges:\*\* - Ensuring clear and interpretable visualizations required multiple iterations.

**How did you overcome them?**

1. \*\*Detecting and Removing Outliers:\*\* - I switched to the IQR method after researching its suitability for skewed datasets. 2. \*\*Hypothesis Testing:\*\* - I reviewed statistical inference materials and sought guidance from my tutor, which clarified the approach. 3. \*\*Visualization Challenges:\*\* - I iteratively refined plots to ensure clarity and relevance, using scatter plots and point plots to highlight relationships.

**Identify any strengths/weaknesses of the approach you took**

\*\*Weaknesses:\*\* - Self-doubt often led to excessive validation of results. - Procrastination limited the time available for thorough exploration. \*\*Strengths:\*\* - Creative thinking enabled me to present insights in a meaningful way. - Collaboration and leveraging diverse resources enhanced my learning experience

.**How could this approach be improved?**

This approach could be improved by incorporating step-by-step visualizations after each analysis stage, ensuring continuous validation and understanding of the data.

**Suggest an alternative approach you could have taken**

An alternative approach would involve using regression analysis for hypothesis testing, providing additional insights into the relationship between click events and scores.