Reflective Report

Section A: UFCFVQ-15-M_Programming_Task_1

The section on defining functions was initially challenging for me as it marked a significant milestone in my coding journey. The course specifications seemed overwhelming at first, and I struggled to find suitable learning resources. Many online tutorials required importing Python libraries, which was not allowed as per the course guidelines. To overcome this, I focused on the first two weeks after the assessments released on learning, which provided practical lessons on defining functions, reading CSV data without external libraries, and calculating the Kendall Tau Rank Correlation Coefficient. Online learning platforms also played a crucial role in my understanding.

The difficulty increased when I had to align my results with the format in the appendix section. Initially uncertain about the required similarity to the examples in the appendix, I sought clarification from the tutor during our last session. Were the tutor mentioned that the results didn't have to match exactly, the output format similarity sufficed.

My main weakness was the delayed start due to my lack of initial knowledge, particularly regarding defining functions to perform calculations. However, through dedicated learning efforts and practical applications, I acquired the necessary skills.

I acknowledge that my code might be lengthy, and I aspire to improve efficiency. I believe that with continued practice and exposure to more advanced coding techniques, I will enhance my ability to write concise and effective code. This experience has highlighted the importance of ongoing learning and practice in my coding development.

Section B: UFCFVQ-15-M_Programming_Task_2

Section two presented a more manageable challenge for me as I had the essential knowledge to tackle the task. Prior to the assessment release, I devoted time to revisiting the weekly assessment tasks on the Jupyter notebook. During the question-solving process, I noticed that my work lacked clear comments and interpretations for each code chunk. To address this issue, I

consulted the "Week 11 Data Science Process" Jupyter notebook in the Programming for Data Science section on UWE Blackboard. This notebook served as a reference guide, helping me understand where to add comments, interpretations, and how to structure my code.

One notable challenge I encountered was in the identification and removal of outliers. Since my code did not detect outliers, I opted for a visualization method to identify them. Subsequently, I removed outliers using the IQR method. However, I later discovered that some outliers persisted in my dataset. Detecting this issue promptly allowed me to reevaluate the outlier removal process and apply the IQR method selectively to columns with extreme outliers.

Another difficulty arose in calculating hypotheses in Python. While I had prior experience with hypothesis calculations using R software, translating this process into Python was a novel challenge. Determining the appropriate approach was crucial, and I opted to use the Pearson correlation coefficient to calculate the p-value for my datasets. This approach proved to be more straightforward and effective compared to other methods I had considered.

Despite these challenges, I found Week 11 Data Science Process Jupyter notebook as a reference notebook, which provided guidance on code organization, hypothesis formulation, and outlier detection. Moving forward, I aim to enhance my proficiency in Python, especially in hypothesis testing, and continue refining my coding practices for greater efficiency.

Reference

Week 11 Data Science Process Jupyter notebook

Kendall rank correlation coefficient - Wikipedia

Python Tutorial (w3schools.com)

Courses - DataCamp Learn