**SCHOOL OF COMPUTING (SOC)**

**IT8701 Introduction to Programming for Data Science**

**Self Reflection (CA2)**

|  |
| --- |
| **Instructions:**   1. Submit this together with your other deliverables at Polymall “Assignments->CA2” folder 2. Name your file “YourStudentID-YourName-YourLecturer.docx” |

|  |  |
| --- | --- |
| **Your Lecturer’s Name** | Chan Chiew Jin |
| **Your Name** | Muhammad Iylia Bin Mohd Hutta |
| **Your Student ID** | P7474841 |
| **Your Class** | NSDDA1/CE/2220/4 |

# QUESTION 1: RATE THE EFFORTS AND COMPETENCY THAT IS DEMONSTRATED IN THIS ASSIGNMENT

Tick in the column that best describes the efforts, technical competency and depth of data analysis that is demonstrated in this assignment.

Justify your rating in the second and third questions below

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **WAY Above Average** | **Above Average** | **Average** | **Below Average** | **Way Below Average** |
| Coding |  | √ |  |  |  |
| Analysis |  | √ |  |  |  |

# QUESTION 2: JUSTIFICATION FOR RATING GIVEN FOR CODING

Please provide evidence that you have met the requirements (AVERAGE) or if you think your submission is above average or even above average, state details of what you have done here so that your lecturer does not miss out the efforts you have put in for this assignment. For CA2, the basic requirements are to produce 4 different graphs with at least 3 datasets.

|  |
| --- |
| I used a total of 9 datasets for this assignment, produced 5 different graphs and also used the sqlalchemy library to store data into a table on a MySQL database and then to retrieve the data using a SQL query for use in Python.  Within the SQL code, I used Common Table Expressions (CTE) and a join to get a table and the data I needed from two different tables that were stored in a MySQL database.  For Python, I put data together from 9 different datasets and joined them together after checking for missing data and standardising datatypes across the necessary columns. For plotting graphs, I used seaborn as much as I could so that I could explore more visually appealing charts. I also produced a distribution chart using joyplot but did not use it in the final analyses for the slides. For supervised machine learning with linear regression, ANOVA and post-hoc t-tests, I learnt how to do the necessary analysis with scikit-learn. |

# QUESTION 3: JUSTIFICATION FOR RATING GIVEN FOR DATA ANALYSIS

Please provide evidence that you have met the requirements (AVERAGE) or if you think your submission is above average or even above average, state details of what you have done here so that your lecturer does not miss out the efforts you have put in for this assignment.

|  |
| --- |
| For this assignment, I attempted to go beyond simple descriptive analytics and went into inferential analytics where I ran a supervised machine learning model with linear regression and ran an ANOVA with post-hoc t-tests. Prior to running the ANOVA and t-tests, I also made the necessary analyses for tests of normality of the data and also equality of variances. Where data was non-normal, I ran the Kruskal-Wallis H test, which accounts for non-normal data. Otherwise, I used the parametric ANOVA with ordinary least squares and linear modelling. |

# QUESTION 4: YOUR FUTURE PLANS

How do you rate your programming competency with data analysis tasks after completing this assignment? Give yourself a rating from 0 to 10.

|  |
| --- |
| I would rate myself a 7 as I still feel that I am not yet fast enough to clean data and produce these analysis, which is the demands of my job. |

After finishing the PDC1 of your Specialist Diploma, which do you think you prefer or is stronger at? The Statistics or Programming portion? How has this realisation affected your mindset of a Data Science job? Do you enjoy a Data Science role that mainly involves application of lots of statistical concepts (improving predictive algorithms for instance) or one that requires a lot of programming (e.g. code to acquire or clean data) or perhaps both equally excite you? 😊

|  |
| --- |
| I feel that I am equally strong at both as I already have a strong statistics background from my statistics and research emphases as a Psychology undergraduate. My strength at programming only increased recently as this course has given me ample practice such that I am able to use it for my job now, which means more practice as well. I would say that both equally excite me very much! |

Are there any useful skills that you gained from this module? Share how you think the skills you learnt from this module can be applied in your current job or in a future career / job change.

|  |
| --- |
| The programming skills are superb as I am now the go-to developer on my team at work where I write Python and recently even JavaScript on Google Apps Script scripts to automate many tasks we do at my job. I have also taken over scripts from our ex-developer to maintain them for my team and other teams within my department. The skills I have used for scripting and automation have contributed to reducing 40-80% of our man-hours across several projects and teams.  However, I would say that my skills now are thanks to the foundation that this course has provided. Many of my other learning came from experiential coding and coding on my own whilst learning on the go. |

What was not taught in this module, but you wish to learn? How do you plan to learn these missing skills?

|  |
| --- |
| I think teaching SQL queries and the language more in-depth would be more helpful. I already have knowledge and certifications in SQL so I was able to grasp that portion of the module pretty fast. However, complete beginners would definitely struggle a lot as I’ve seen in most of my classmates.  As for me, I plan to learn more about NoSQL databases as I understand that it could greatly increase the scope of my skills and potentially lead to more exciting developer opportunities. |

**-- End of Self-Reflection --**