**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** | Chia Yao An, Kendrick |
| **Your Name** | Sng Tian Hao, Keith |
| **Your Student ID** | 1273381F |
| **Your Class** | C |

# 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 | X |  |  |  |  |
| Analysis | X |  |  |  |  |

# 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.

|  |
| --- |
| * Used 10 datasets for a comprehensive analysis of Singapore’s labour market with 8 graphs produced * Emulated typical data analysis workflow in a business context (ETL for data ingestion, querying from database, manipulation & cleaning, followed by analysis of dataset and visualization) * Utilised tools outside the scope of this module (API calling, PCA) * Clean notebook with markdown annotations for clarity and logical flow |

# 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.

|  |
| --- |
| * Went beyond basic requirements of CA2 to ensure analysis is comprehensive and narrative is holistic/coherent (e.g., 3 datasets and 4 graphs is not enough to analyse the labour market, so I created more based on what’s necessary for a compelling storyline) * As above in Question 2 |

# 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.

|  |
| --- |
| 7.5. I think there are better methods and ways to visualise or analyse data from another angle which I need to pick-up. For example, how can I look at multiple leave types and compare across many different industries beyond simple descriptive visualisations – maybe unsupervised learning methods to cluster the different groupings would be a better analysis technique to elucidate interesting findings.  In addition, great data analysts are not just people with good technical skillsets. They are also equipped with a deep understanding of the business context. This comes with experience in the industry/field they are in, which will help tease out more meaningful insights from analysis, and this is a level which I strive to achieve in the long-term. |

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 like both aspects and I think I’m stronger in the programming portion because I often forget about the details of statistics because it’s very theoretical. Regardless, I’m eager to expand my theoretical depth to deepen my understanding of predictive modelling in the next semester. |

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 analytical thinking gained from the two CAs really challenged me to come up with a coherent narrative for analysis. It’s simple to analyse datasets in silos, but very difficult to piece together a convincing and meaningful story to answer a question with business/real-world implications.  Besides that, technical skills from visualisation libraries, pandas, numpy (I neglected this because I often used pandas which was more straightforward), MySQL and MongoDB will definitely help me in the workplace for data cleaning, extraction, and analysis. |

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

|  |
| --- |
| Date-time manipulation in python which will probably be a common use case in the workplace.  Statistical testing in python would be helpful as well. Self-learn on datacamp on by googling. |

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