Faculty of Engineering and Information Technology
School of Computer Science

42047 – Data Processing Using Python

AUTUMN 2022
ASSIGNMENT Part-B: Data Analysis and Visualization

Marking Guide for Part-B (60 pts):

Below is the marking guide for **Part-B: Code/IPython Notebook for Data Analysis and Visualization** of the assessment. It is designed to allow you to get a Pass grade with minimal effort while still demonstrating that you understand the core principles of Python Programming and exploratory data analysis, to get a Distinction with reasonable effort, and to get a High Distinction with solid effort, and 100% with considerable effort. It is recommended that you pay attention to the grade distribution and work towards your own skill level.

Task	Items	Max Points
IPython Notebook Design	Includes appropriate sections for each stage of data exploration, appropriate comments/description/intent for each section, etc.	**5**
Code Quality	Includes proper indenting and white spacing, helpful comments and meaningful class/method/property/field/variable names. Use of proper Python programming syntax, exception handling etc.	**5**
Data Preparation	- Techniques used - Use of appropriate methods to explore data - Use of statistical methods (mean, median, etc.) with interpretation - Use of appropriate plotting for understanding relevant attributes to use - Adding attributes if needed to find answer to the business question. (e.g.: StartDate and EndDate are two existing attributes, a new attribute Duration was created using them)	**15**
Missing Value exploration	- Use of appropriate data visualization techniques to identify missing values - Appropriate actions taken if required - Interpretation if any	**10**
Outlier identification	- Use of appropriate visualization techniques to identify outliers. - Appropriate actions taken to handle them - Interpretation if any.	**10**
Data Visualization	- Use of appropriate data visualization technique based on relevant attributes type. - Use of plots/Charts such as Box, Pair, Bar, Pie, etc. - Interpretation of the plots	**15**
	Total	**60**