## Appendix 2: Examination Criteria

Statistics:

You are required to:

1. Summarise your dataset clearly, using relevant descriptive statistics and appropriate plots. These should be carefully motivated and justified, and clearly presented. You should critically analyse your findings, in addition to including the necessary Python code, output and plots in the report. You are required to plot at least three graphs. [0-35]

*Summarised datasets and graphs can be found throughout the entirely of the report and within the Jupyter notebooks.*

1. Use two discrete distributions (Binomial and/or Poisson) in order to explain/identify some information about your dataset. You must explain your reasoning and the techniques you have used. Visualise your data and explain what happens with the large samples in these cases. You must work with Python and your mathematical reasoning must be documented in your report. [0-30]

*Please see sections 2.10 and 3.7, reasoning can often be found in the Notebook as I was struggling with the reports word count unfortunately.*

1. Use Normal distribution to explain or identify some information about your dataset. [0-20]

Please see Section 1.8

1. Explain the importance of the distributions used in point 3 and 4 in your analysis. Justify the choice of the variables and explain if the variables used for the discrete distributions could be used as normal distribution in this case. [0-15]

*Please see sections 2.10 and 3.7, again within the Notebooks as I was unable to discuss within the word count.*

Data preparation and Visualisation:

1. You must perform appropriate EDA on your dataset, rationalizing and detailing why you chose the specific methods and what insight you gained. [0-20]

*See Jupyter notebooks for instances of this, I have attempted to investigate key summary statistics throughout the notebooks, regarding each dataset.*

1. You must also rationalise justify and detail all the methods used to prepare the data for ML. [0-30]

*Rationale often discussed throughout report and within Jupyter notebooks*

1. Appropriate visualizations must be used to engender insight into the dataset and to illustrate your final insights gained in your analysis. [0-20]

*Numerous graphical visualisations can be found throughout the report*

1. All design and implementation of your visualizations must be justified and detailed in full, referring to Tufts Principles [0-30]

*Often included in the Jupyter notebooks as I wanted to refrain from exceeding the word count. I make reference to this in the Markdowns entitled ‘Alignment with Tufte’s Principles’.*

Machine learning for Data Analytics:

1. Explain which project management framework (CRISP-DM, KDD or SEMMA) is required for a data science project. Discuss and justify with real-life scenarios. Provide an explanation of why you chose a supervised, unsupervised, or semi-supervised machine learning technique for the dataset you used for ML modelling. [0 - 20]

*See CRISP-DM Section at the beginning of the report*

1. Machine learning models have a wide range of uses, including prediction, classification, and clustering. It is advised that you assess several approaches (at least two), choose appropriate hyperparameters for the optimal outcomes of Machine Learning models using an approach of hyperparameter tunning, such as GridSearchCV or RandomizedSearchCV. [0 - 30]

*Used numerous times in Sections 1.7, 2.12, 2.13 and 2.14*

1. Show the results of two or more ML modelling comparisons in a table or graph format. Review and critically examine the machine learning models' performance based on the selected metric for supervised, unsupervised, and semi-supervised approaches. [0 - 30]

*See Sections 1.7, 2.12, 2.13 and 2.14*

1. Demonstrate the similarities and differences between your Machine Learning modelling results using the tables or visualizations. Provide a report along with an explanation and interpretation of the relevance and effectiveness of your findings. [0 - 20]

*See Sections 1.7, 2.12, 2.13 and 2.14. Discussion in report and within the notebooks.*

Programming:

1. The project must be explored programmatically, this means that you must implement suitable Python tools (code and/or libraries) to complete the analysis required. All of this is to be implemented in a Jupyter Notebook. Your codebook should be properly annotated. The project documentation must include sound justifications and explanation of your code choices (code quality standards should also be applied). [0-50]

*See entirety of report and notebooks*

1. Briefly discuss your use of aspects of various programming paradigms in the development of your project. For example, this may include (but is not limited to) how they influenced your design decisions or how they helped you solve problems. Note that marks may not be awarded if the discussion does not involve your specific project. [0-50]

*See entirety of report and notebooks*