**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** |  |
| **Assessment Title:** |  |
| **Lecturer Name:** |  |
| **Student Full Name:** |  |
| **Student Number:** |  |
| **Assessment Due Date:** |  |
| **Date of Submission:** |  |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

**Introduction**

**Statistics**

*Descriptive Statistics*

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 three graphs. [0-35]

**Distributions**

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]

**Normal Distribution**

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

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.

**Programming in Statistics:**

**Data preparation and Visualization :** (Graded out of 100)

**EDA**

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]**

**Data Preparation**

* **Data processing**
* **Data cleaning**
* **Data splitting**
* **Feature selecetions**

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

**Data Visualization**

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

4. All design and implementation of your visualizations must be justified and detailed in full. **[0-30]**

**Programming:**

**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 modeling. **[0 - 20]**
2. 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]**
3. Show the results of two or more ML modeling 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]**
4. 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]**

**Programming: :** (Graded out of 100)

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]**

**Please recall that simply performing the analyses is a requirement to achieve a grade of PASS. Critical analysis and independent research are required for higher marks.**

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]**

**CA1 NOTE DO NOT ZIP YOUR SUBMISSION FILES, ALL FILES MUST BE SUBMITTED INDIVIDUALLY**

**Submissions that are suspected of plagiarism and/or inclusion of AI (CHATGBT, BARD etc…) Generated content will be referred to the college authorities.**

**Note ALL Students are required to use Git for any Assignments that they are working on.**

This means that ALL changes must be committed to Git during your assignment. (Not just a single commit at the end!) This is to allow you to display your incremental progress throughout the assessments, give you practice for your capstone/thesis, allows you to create an online portfolio that can be used to showcase your work and to ensure that there are no problems with final uploads (as all your work will be available on GitHub). It is expected that there will be a minimum of 10 commits (with many of you making very many more). You may Only use your CCT email for your git account, private/work email-based accounts will not be accepted. You must also include ALL your lecturer's CCT emails as a collaborator on your account.

**Submission Requirements**

* All assessment submissions must meet the minimum requirements listed below. Failure to do so may have implications for the mark awarded.
* All assessment submissions must:
  + 4000 (+/- 10%) words in report (not including code, code comments, titles, references or citations)
  + Report submission MUST be a word document only (No PDF’s!);
  + Code in a Jupyter Notebook file only but may be referenced in the word document.
  + GITHUB Link
  + Be submitted by the deadline date specified or be subject to late submission penalties
  + Be submitted via Moodle upload
  + Use [Harvard Referencing](http://40.115.124.2/sp/subjects/guide.php?subject=harvardref) when citing third party material
  + Be the student’s own work.
  + Include the CCT assessment cover page.

**Additional Information**

* Lecturers are not required to review draft assessment submissions. This may be offered at the lecturer’s discretion.
* In accordance with CCT policy, feedback to learners may be provided in written, audio or video format and can be provided as individual learner feedback, small group feedback or whole class feedback.
* Results and feedback will only be issued when assessments have been marked and moderated / reviewed by a second examiner.
* Additional feedback may be requested by *contacting the appropriate lecturer,* Additional feedback may be provided as individual, small group or whole class feedback. Lecturers are not obliged to respond to email requests for additional feedback where this is not the specified process or to respond to further requests for feedback following the additional feedback.
* Following receipt of feedback, where a student believes there has been an error in the marks or feedback received, they should avail of the recheck and review process and should not attempt to get a revised mark / feedback by directly approaching the lecturer. Lecturers are not authorised to amend published marks outside of the recheck and review process or the Board of Examiners process.
* Students are advised that disagreement with an academic judgement is not grounds for review.
* For additional support with academic writing and referencing students are advised to contact the CCT Library Service
* For additional support with subject matter content students are advised to contact the [CCT Student Mentoring Academy](https://moodle.cct.ie/mod/forum/view.php?id=55148)
* For additional support with IT subject content, students are advised to access the [CCT Support Hub](https://moodle.cct.ie/course/view.php?id=1861).