**CST3340 Coursework 2**

Coursework 2

**a. Overview**

This section contains the material for coursework 2. It is worth 60% of the final grade for this module.

**b. Important Dates**

You must ensure you meet the deadlines for each element.

|  |  |  |
| --- | --- | --- |
| **Element** | **Type** | **Due Date** |
| Individual CW2 Report  40% of the final module grade | Formative Feedback | During Labs in Week 23: Week beginning: 22nd March 2021 |
| Final Report | Week 24: Midnight Friday 16th April 2021 |
| Oral Presentation for CW2  20% of the final module grade | Presentation | During allocated time slot: Week beginning 19th April 2021 |
| Summative Feedback | | June 2021 |

**c. Plagiarism**

Plagiarism in data analysis is easy to spot. Please be aware that the penalties are severe, and your final degree award classification is at risk.

**d. Coursework Content**

**Coursework 2 should be done individually.**

**e. Where to submit**

Each student should submit an individual report of no more than **2400-word**. Please do not include your name or student Id as your submission will be marked anonymously. You are also expected to attend a **5-minute** oral presentation**. The presentation is compulsory, students who do not attend the presentation will fail this coursework.**

All submission should be via the submission link in your learning environment.

Do not handwritten assessed coursework directly to your tutor, and do not submit it by email to your tutor. Coursework which is not submitted via unihub will not be accepted.

**f. Coursework Requirements:**

**You are required to analyse a large data set of your choice, which has been agreed with your module tutor:**

Your project may use any combination of data analysis techniques, data-mining algorithms and software that has been covered in the module. You may also apply them to any aspect(s) of the dataset for knowledge discovery.

You should cover the areas indicated below and your findings should be presented in the form of a report no more than **2400 word**. You will also be expected to give a **5-minute** oral presentation. **The presentation is compulsory, students who do not attend will fail this coursework.**

Please see below the aspects that you should consider:

**Individual Report**

* Data Analysis and Visualisation (35 marks)
  + Initial analysis of the data using visualisation techniques within Tableau (use diagrams/graphs to highlight important patterns/findings).
  + Discussion and interpretation of result.
  + Discussion of overall trends and patterns observed.
* Selection of Data Mining Algorithm (10 marks)
  + Select one data mining algorithm suitable for further analysis of your data.
  + Clearly justify your choice, with reference to the visualisation analysis carried out.
* Data Pre-processing (10 marks)
  + Identify your input and class variables, if relevant (i.e. which variable are you going to consider for your class variables).
  + Identify and resolve any anomalies in the data (i.e. missing values, outliers etc.).
  + Carry out any appropriate pre-processing/transformations to the data set.
* Data Mining (25 marks)
  + Use the chosen data mining algorithm for further analysis of your pre-processed data set.
  + Clearly discuss the implementation of the data mining algorithm.
  + Discuss and interpret the results.
* Data Ethics (10 marks)
  + A discussion of data ethical issues related to the analysis and use of business data.
* Conclusion (10 marks)
  + A discussion of the overall visualisation results (e.g. What were the important findings? Summary of overall trends and patterns).
  + A discussion of the data mining results (e.g. How well did the model fit your data?).
  + A discussion of the business intelligence that can be obtained from these results.
* **Oral Presentation** (100 marks)
  + This **5-minute** oral presentation will allow you to discuss your analysis and results.