**SEHS4697 Big Data and Cloud Analytics**

**Group Project (2024/25 Semester 2)**

**Group ID: XX (4 - 5 students in each group)**

|  |  |  |
| --- | --- | --- |
| **Student Name** | **Student ID** | **Signature** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Project Title: Project Name Here**

**Abstract: A brief abstract of the proposed research problem and choice of dataset, no more than 250 words.**

**Table of Contents:** **Include page numbers.**

|  |  |
| --- | --- |
| 1. Introduction & Problem Definition | Page x |
| 1. Data Summary, Data Pre-processing & Initial Insights | Page x |
| 1. Data Analysis | Page x |
| 1. Discussion | Page x |
| 1. References | Page x |
| 1. Appendix | Page x |
| 1. Declaration of Members' Contributions | Page x |

* YOU NEED TO REMOVE THE TEXT IN RED BEFORE SUBMITTING YOUR FINAL REPORT.
* Limit your report to a maximum of 10 pages, excluding the cover page and the appendix. (Times News Roman 11 point, single line spacing)
* To ensure fair judgment, Please remove any identifying information after cover about participants, professors, or institutions, including names, logos, and study details.
* Submit the report cover page by **24 FEB**.
* Submit this report (60%), with a PPT (20%) and a 15-minute video presentation (20%) before **15 April.** Late submission can result in penalty. Submission will not be accepted 24 hours after the deadline.
* Bonuses (up to 15%) will be awarded for groups **successfully** registering and participating in local or international competitions.

**I. Introduction & Problem Definition**

In this section, you can briefing explain the research problems and their background information, e.g., research motivation and objectives, research methods and technical introduction, supporting data and effectiveness. Pay attention to the following points:

• You research problem should be able to answer by using the dataset.

• You should focus on a small set of problems, say, within 5 questions.

• You should justify why the problems are important. What are the values of solving those problems? E.g. bringing in more customers, increase the profit, etc.

• Your problems should be able to assess by measurable quantities. E.g. how much more customers will be retained?

**II. Data Summary, Data Pre-processing & Initial Insights**

In this section, you need to include the following points.

* Source of your data, e.g., real data set from some popular public data platforms: [*UCI Machine Learning Repository*](https://archive.ics.uci.edu/dataset/563/iranian+churn+dataset)*,* [*CMU StatLib*](https://lib.stat.cmu.edu/datasets/)*,* [*Kaggle*](https://www.kaggle.com/)*,* [*Data.world*](https://data.world/)*.* A rich dataset will allow you to try out multiple analysis techniques.
* Are there any issues with your data? E.g. outliers, missing values, invalid values and etc. If there are issues, how do you deal with the issues?
* Summary of the data you plan to use by tables and charts
* Initial insights you got from the data summary relevant to the problem you plan to solve
* Construction of new variables from the data and why are they needed

**III. Data Analysis**

In this section, you will perform analysis on your dataset using methods learnt in our classes. Your analysis should be relevant to the problem you raised in Section I.

**IV. Discussion**

In this section, you should evaluate the result produced in Section III and discuss if the problem raised in Section I is solved. You should also describe the challenges faced when analysing the data and suggestions for future research directions and improvements.

**V. References**

You need to include a list of references using APA style.

**VI. Appendix**

Materials such as source code and computer software output if relevant, should be put in this appendix for reference.

**VII. Declaration of Members' Contributions**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group Member** | **Data preparation** | **Research** | **…** | **…** | **Report Writing** | **Total** |
| Student 1 | 5% | 5% | … | … | 5% | 25% |
| Student 2 | 10% | 0% | … | … | 5% | 25% |
| Student 3 | 5% | 5% | … | … | 10% | 25% |
| Student 4 | 0% | 15% | … | … | 5% | 25% |
| Total | 20% | 25% | … | … | 25% | **100%** |

All group members are expected to contribute equally to the project. If the contributions are uneven among group members, individual marks will be adjusted based on their contributions. Each group has to submit a declaration of contributions by all members in the group. Below is an example of the declaration. The columns consist of the task breakdown of your project. You need to break your project into at least 4 different tasks (columns in the table). The total contribution of all members should be equal to 100%.