School of Computing and Information Systems
The University of Melbourne
COMP20008, Elements of Data Processing, Semester 1, 2025

Assignment 2 - Group Contract

Group Name: W08G4
Workshop: Monday 10am
Tutor: Hasti Samadi

Group Members:

- Jonathan Bian - jonathanbian@student.unimelb.edu.au
- Sandiv Dathunarachchi - sdathunarach@student.unimelb.edu.au
- Tianhao (Toby) Tang - tianhaot1@student.unimelb.edu.au
- Qinyi (Evanz) ZHENG - qinyi.zheng.1@student.unimelb.edu.au

Research Question:
Question: Which combination of driver, vehicle, and environmental factors most significantly influences road accident severity in Victoria, and how accurately can we predict whether an accident will result in a serious injury or fatality?
We will utilize the Victorian road crash datasets (Accident, Person, Vehicle) given. The goals are to identify key factors correlating with accident severity (focusing on Serious Injury/Fatality vs. Other Injury/Non-Injury), build and compare at least 2 supervised machine learning models to predict this severity, and use clustering to identify high-risk accident profiles.
Objective: To gain a deeper understanding of the factors contributing to severe road accidents in Victoria and to develop predictive capabilities that could potentially inform road safety interventions
Methods: Data merging, cleaning, preprocessing (handling missing values, encoding categorical variables, feature engineering e.g., vehicle age), correlation analysis (statistical tests, visualizations), supervised learning (e.g., Logistic Regression, Decision Trees for classification) K-Means clustering, data visualization,

Project Overview:
We will first load and merge the provided Accident, Person, and Vehicle datasets. A lot of effort will be put into cleaning and preparing this data which includes dealing with missing values, encoding categorical features, possibly creating new features such as vehicle age or time-of-day categories. We will carry out a correlation analysis to find relationships between different factors (driver demographics, vehicle characteristics, environmental conditions like light/road geometry/speed zone) and how severe an accident is. Next, we will apply at least two supervised learning models and check their efficiency in predicting serious accidents (Severe Injury/Fatality). To identify unique groups or profiles of accidents/vehicles, K-Means clustering is going to be used based on chosen characteristics.

Roles and Responsibilities:

Member Name	Role and Responsibilities
Jonathan	Data cleaning, report, overall wrapping up
Sandiv	Analytical visualization and slides making
Toby	Data cleaning and correlation coefficient analysis
Evanz	Clustering visualization and slides making

Communication Plan:
Regular group calls via Zoom

Meeting Schedule:
1-2 times a week via Zoom preferably on Tuesdays and Saturdays

Decision-Making Process:
Decisions will be made by majority vote (3 out of 4)

Work Plan and Timeline:

- **Week 9 (Apr 28 - May 4):** Finalize Contract, Set up code Repo. Initial Data Loading/Merging/Exploration (Sandiv). **Group Contract Due May 2**.
- **Week 10 (May 5 - May 11):** Intensive Data Cleaning/Preprocessing (Sandiv). Initial Correlation Analysis & EDA (Toby). Feature Selection discussion (All). SL/Clustering data prep (Jonathan/Evanz).
- **Week 11 (May 12 - May 18):** Complete Correlation (Toby), SL Modelling & Evaluation (Jonathan), Clustering Implementation & Interpretation (Evanz). Draft Report sections (All). **Code & Report Due Fri May 16**. *Internal code review before submission.*
- **Week 12 (May 19 - May 25):** Draft Presentation Slides based on report. Practice presentation timing. **Slides Due Fri May 23**.
- **Week 13 (May 26 - May 30): Oral Presentations**. Final practice sessions. **Team Evaluation Due Fri May 30**.

Code of Conduct:

- Communicate respectfully.
- Participate actively in meetings and contribute fairly to tasks.
- Meet internal deadlines agreed upon by the group to avoid delaying others.
- Offer help or ask for it if falling behind.
- Provide constructive feedback on code and report drafts.
- Uphold academic integrity standards; cite if needed including AI use

Disagreements or non-responsiveness:

1. **Communication Attempt:** Initially, the concerned group members will attempt to communicate directly with the individual in question to address the issue and seek resolution.

2. **Mediation:** If direct communication fails to resolve the disagreement or non-responsiveness, the matter will be brought to the attention of the tutor of the workshop for mediation. The tutor will facilitate a discussion to find a mutually acceptable solution.
3. **Escalation:** If the issue remains unresolved after mediation, it will be escalated to the head tutor or subject coordinator for further intervention and resolution.

It is expected that all group members will engage in this process in good faith and with a commitment to resolving conflicts constructively for the benefit of the project and the team as a whole.

Signature:
By signing below, each group member acknowledges their commitment to adhere to the terms outlined in this contract.

Student 1: Yuchao Bian	Date:02/05/2025
Student 2: Sandiv Dathunarachchi	Date: 02/05/2025
Student 3: Tianhao Tang	Date: 02/05/2025
Student 4: Qinyi Zheng	Date: 02/05/2025

This contract is intended to establish clear expectations and promote effective collaboration among group members throughout the duration of the data analytics project. Any amendments to this contract should be discussed and agreed upon by all group members.