

School of Computing and Information Systems
The University of Melbourne
COMP20008, Elements of Data Processing, Semester 2, 2025

Assignment 2 - Group Contract

Group Name: W02G4

Workshop: Tuesday 9:00 am

Tutor: Di Wu

Group Members:

- Xuan Zhang – xuazhang11@student.unimelb.edu.au
- Wenshuo Cai - wenshuo.cai@student.unimelb.edu.au
- Bochong Yuan – Bochongyuan@student.unimelb.edu.au
- Gaoling Li – kehuili@student.unimelb.edu.au

Research Question:

Briefly describe the scope and goals of your data analysis. Outline the problem statement or objective you aim to address and the methods you plan to use for analysis.

Can machine learning techniques (clustering techniques, supervised ML method) identify distinct household travel behaviour profiles?

Project Overview:

Outline the methods and strategies you are going to use to complete the assignment.

1. Try to use PCA during the data preprocessing.
2. Test clustering techniques (K-means, Hierarchical Clustering).
3. Try all the supervised models (regression, k-NN, decision tree) that we've learned to identify the relation within the dataset.
4. Comparing all the methods, analyse and find the "best" method.

Roles and Responsibilities:

Specify the roles each group member will undertake throughout the project. Define the tasks, responsibilities, and contributions expected from each member. Ensure that roles are distributed fairly and leverage each member's strengths.

Member Name	Role and Responsibilities
Wenshuo Cai	<ul style="list-style-type: none">• Coordinate project timeline, task assignments, and group meetings.• Collect and clean the household travel dataset (handle missing values, outliers).• Perform data preprocessing: feature scaling, encoding categorical variables.• Implement PCA for dimensionality reduction and share processed data with team.

Xuan Zhang	<ul style="list-style-type: none"> • Design and run clustering methods: K-Means, Hierarchical Clustering (including dendrograms, elbow/silhouette analysis for optimal cluster number). • Interpret cluster profiles to identify distinct travel behaviour groups. • Provide cluster labels for downstream supervised models.
Bochong Yuan	<ul style="list-style-type: none"> • Apply supervised ML models: Regression, k-Nearest Neighbours, Decision Tree. • Use cluster labels (from Student B) or other target variables to explore predictive relationships. • Perform hyperparameter tuning (cross-validation) and evaluate model performance (accuracy, RMSE, etc.).
Gaoling Li	<ul style="list-style-type: none"> • Compare results from PCA, clustering, and supervised models to identify the “best” approach based on accuracy, interpretability, and insights. • Create visualizations (heatmaps, PCA plots, decision tree diagrams).

Communication Plan:

Establish a communication strategy to ensure effective collaboration within the group. Specify how and when you will communicate (e.g., regular meetings, communication platforms), and establish guidelines for timely responses.

Platforms:

- Primary: Wechat
- Secondary: Email for formal updates or when files cannot be shared via Teams.
- Meeting Updates: Meeting notes will be posted in a shared Google Doc within 24 hours of each meeting.
- Response Time: Team members will reply to messages within 24 hours on weekdays and 48 hours on weekends unless prior notice is given.
- Collaboration Tools: GitHub (for code version control), Google Drive (for report drafts, slides, datasets).

Meeting Schedule:

Outline the frequency, duration, and preferred times for group meetings. Set expectations for attendance, punctuality, and participation during meetings.

- **Frequency: Weekly meetings** every **Wednesday at 4:00–5:00 PM** (1 hour).
- **Location:** University study room or Teams video call if remote.
- **Additional Meetings:** Extra sessions scheduled during key milestones (e.g., just before report submission).
- **Expectations:**
 - Punctual attendance is required.
 - Members should review the agenda and any assigned tasks beforehand.
 - Notify the group at least **24 hours in advance** if unable to attend.

Decision-Making Process:

Define the process for making decisions within the group. Establish methods for resolving conflicts or disagreements in a constructive manner.

- **Consensus First:** Decisions will be made by **open discussion and consensus** whenever possible.
- **Majority Vote:** If consensus cannot be reached, a **simple majority vote (3 out of 4)** will determine the outcome.
- **Tie-Breaker:** If a tie occurs, the **Project Lead** (Student A) will make the final call after considering all viewpoints.

Work Plan and Timeline:

Provide a tentative timeline for key project milestones, including data preparation, analysis, report drafting, and presentation preparation. Allocate time for review and revisions.

Milestone	Target Completion	Responsible Lead
Data collection & cleaning	Week 2	Wenshuo Cai
PCA preprocessing	Week 3	Wenshuo Cai
Clustering (K-Means, Hierarchical)	Week 4	Xuan Zhang
Supervised models (Regression, k-NN, Decision Tree)	Week 5	Bochong Yuan
Model comparison & evaluation	Week 6	Gaoling Li
Draft report & visualizations	Week 7	All
Internal review & revisions	Week 8	All
Final report submission	Week 9	All
Presentation slides & rehearsal	Week 9	All

Code of Conduct:

Set expectations for professional behaviour, including respect for differing opinions, active participation, and accountability for tasks. Address how to handle underperforming team members.

- **Respect:** Treat all members courteously and value diverse opinions.
- **Accountability:** Complete assigned tasks by agreed deadlines.
- **Participation:** Actively contribute to discussions, coding, and report writing.
- **Quality Standards:** Commit to clear documentation, reproducible code, and professional writing.
- **Underperformance:**
 - Step 1: Direct, constructive feedback from teammates.
 - Step 2: Offer support (e.g., redistributing tasks) if workload is an issue.
 - Step 3: Follow the “Disagreements or Non-Responsiveness” procedure if problems persist.

Disagreements or non-responsiveness:

Set what will happen in the event of disagreements among group members or instances of non-responsiveness from a team member.

A possible process could be as follows:

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: Xuan Zhang Date: 25/09/2025

Student 2: Wenshuo Cai Date: 25/09/2025

Student 3: Bochong Yuan Date: 25/09/2025

Student 4: Gaoling Li Date: 25/09/2025

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.