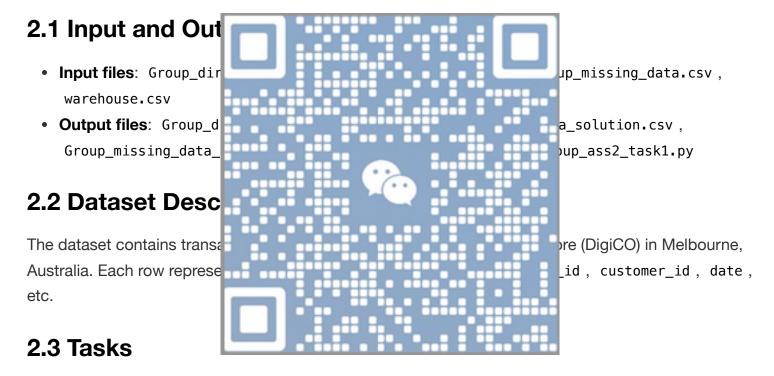# FIT5196 - S2 - 2024 Assessment 2

1. Introduction

This is a group assessment worth 40% of the total mark for FIT5196. It consists of three tasks related to data analysis and manipulation.

2. Task 1: Data Cleansing (50%)

2.1 Input and Out

- **Input files**: `Group_dir`   `up_missing_data.csv`, `warehouse.csv`
- **Output files**: `Group_d`   `a_solution.csv`, `Group_missing_data_`   `up_ass2_task1.py`

2.2 Dataset Desc

The dataset contains transa   re (DigiCO) in Melbourne, Australia. Each row represe   `_id`, `customer_id`, `date`, etc.

2.3 Tasks

1. **Detect and fix errors in** `_dirty_data.csv`
2. **Impute the missing values in** `_missing_data.csv`
3. **Detect and remove outlier rows in** `_outlier_data.csv` (w.r.t. the `delivery_charges` attribute only)

2.4 Methodology

The `group_id_ass2_task1.ipynb` should demonstrate the methodology to achieve correct results. This includes using appropriate Python functions for input, process, and output, and presenting the solution in an efficient and proper way.
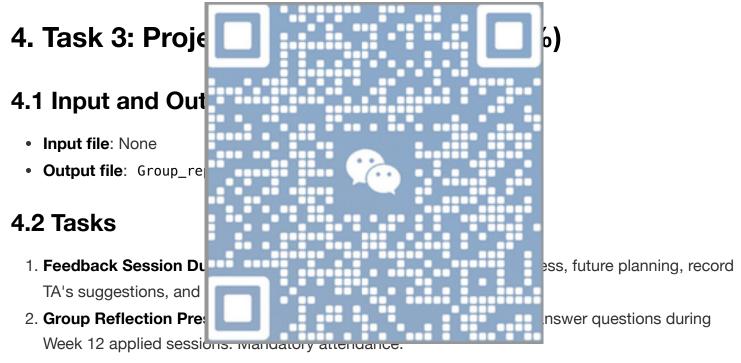
3. Task 2: Data Reshaping (15%)

3.1 Input and Output Files

- **Input file**: `suburb_info.xlsx`
- **Output file**: `Group_ass2_task2.ipynb`

3.2 Task Description

Study the effect of different normalisation/transformation methods on columns `number_of_houses`, `number_of_units`, `population`, `aus_born_perc`, `median_income`, `median_house_price` to prepare data for a linear regression model to predict `median_house_price`.

4. Task 3: Proje⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛)

4.1 Input and Out⬛⬛⬛⬛⬛⬛

- **Input file**: None
- **Output file**: `Group_re`⬛⬛⬛⬛

4.2 Tasks

1. **Feedback Session Du**⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛ess, future planning, record TA's suggestions, and ⬛⬛⬛⬛⬛⬛⬛⬛⬛
2. **Group Reflection Pre**⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛nswer questions during Week 12 applied sessions. Mandatory attendance.
3. **Reflective Report**: Provide a report based on feedback, tailored solutions, and any related findings.

5. Submission Requirements

- Submit 6 files: `Group_dirty_data_solution.csv`, `Group_missing_data_solution.csv`, `Group_outlier_data_solution.csv`, `Group_ass2_task1.ipynb`, `Group_ass2_task1.py`, `Group_ass2_task2.ipynb`, `Group_report.pdf`
- Zip all files into `Group_ass2.zip`
- Follow file naming standards and ensure files are parsable and readable.

6. Appendix

- Instructions for generating `.py` files from notebooks.
- Submission checklist details.