**1. Student bi-weekly performance summary**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Adm. No. | Name | No. of hours present | Progress1 | Remarks |
| 1. 2007476 | Johnnie | 8 | A | - |
| 2. 2112790 | Jayden | 8 | B | - |
| 3. 2112802 | Wee Loon | 8 | A | - |

1 State whether: A=On Schedule B=Ahead Schedule for no. of days C=Behind Schedule for no. of days

**2. Weekly Scrum**

|  |  |
| --- | --- |
| Week No: 4-5 Date: 10/11/2022 | |
| Member Name 1: | **Johnnie (Data Engineer)** |
| Last week’s Progress | * User Stories: Research on our users’ background to know what problems they may face. * User Stories: Compiled user stories card 4 and 5 * Database setup: Reviewed, edited and tested SQL scripts to make sure importing of data works all the time. |
| This week deliverables | * Implemented ETL pipeline: * Feature engineering * Data imputation * Merging * Inserting cleansed data into DB & creating cleansed schema relations |
| Obstacles | * Making sure the new features are useful * Make sure redundance in transformation actually works. For example, the removal of duplicated values. |
| Member Name 2: | **Jayden (SQL Developer)** |
| Last week’s Progress | * Product Backlog: Allocated number of man hours our team will spend for each product (task) * Product Backlog: Created a list of products that we need to complete * Research on Scrum framework for effective teamwork |
| This week deliverables | * Imported raw data into SQL database using SQL scripts * Developed 3 SQL Complex Queries on the raw data |
| Obstacles | * Had to account for differing data types and creating a auto-increment for primary key in raw data import, as well as foreign keys * Had to build complex subqueries to join all 3 tables together |
| Member Name 3: | **Wee Loon (Data Analyst)** |
| Last week’s Progress | * Database Setup: Designed SQL schema and created their tables * Database Setup: Wrote SQL scripts to import raw data into MSSQL |
| This week deliverables | * Imported data from SQL into python for initial EDA * Conducted data visualisation in python to gain understanding on data. Made visuals on drivers, car model, rating, etc. |
| Obstacles | * Visualising a huge dataset, had to decide carefully which diagrams/plots to use * Generating plot on sensor data takes very long due to immense amount of data. |