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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Adm. No. | Name | No. of hours present | Progress1 | Remarks |
| 1. 2222329 | Goh Rui Zhuo | 8 | A | - |
| 2. 2222291 | Toh Kien Yu | 8 | A | - |
| 3.2235189 | Wang ZiXin | 8 | A | - |

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

**2. Bi-Weekly Scrum**

|  |  |
| --- | --- |
| Week No: 3-5 Date: 30 Oct – 3 Nov  Next Week 6 Nov– 10 Nov | |
| Member Name 1: | **Goh Rui Zhuo** |
| Last 2 weeks’ Progress | * Created ERD diagram and deciding on the proper database design.   + Utilised the metadata, and the info of the database columns.   + Database contained accurate primary key and foreign key. * Created SQL database with creation of table of sensors, drivers, and safety labels.   + Utilised SQL Server to create database.   + Created insert statement with bulk insert.   + Inserted the full dataset. * Connected to python pyodbc to get the dataset into Jupyter Notebook * Product Backlog created and refined |
| This 2 weeks deliverables | * Data Preprocessing on dataset to be analyse further.   + Outlier testing   + Check for logical data. * Exploratory Data Analysis with matplotlib   + Analyse on each dataset   + Merge dataset and analyse together. * Create two SQL query to analyse on the data. * Target to have derived columns. * Fully analyse the data |
| Obstacles | Initial struggle to import dataset from root directory.  Unable to create ETL pipeline initially but manage to connect it after that |
| Member Name 2: | **Toh Kien Yu** |
| Last 2 weeks’ Progress | * Creation of ERD diagram   + Designed database design   + Created the Database Tables Script in SQL * Adding of User Stories.   + Took account into the needs and requirements of different user.   + Specified the acceptance and confirmation * ETL Pipeline to obtain data   + Imported data into a Data Frame from SQL Database using Python * Product Backlog   + Listed down backlog items   + Listed down the effort required (man hours) |
| These 2 weeks deliverables | * Data cleaning   + Impute missing values in the dataset (Decide on the logic to impute these missing values)   + Drop extreme outliers and values from the dataset. * Exploratory data analysis on dataset   + Merge datasets for analysis between tables   + Create visualizations on sensors, drivers and safety label dataset * Add markdowns and analysis on charts * Create SQL queries for users to generate meaningful insights   + Create a query to analyse the danger percentage of different trip durations. * Refine Product Backlog |
| Obstacles | * Faced an issue where I was indecisive of whether imputing columns based on the mean value group by booking ID or forward/backward for different columns as it might not be practical for some columns such as seconds. |
| Member Name 3: | **Wang ZiXin** |
| Last 2 weeks’ Progress | * Creation of ERD diagram   + Check the data type and relation type * Creation of user story   + Acceptance criteria and conformation * Data preprocessing   + Remove the unwanted column from dataset |
| This 2 weeks’ deliverable | * EDA   + Creating of visual charts (histograms, BarCharts, scatterplot) * Creating SQL queries   + Create a query to find relationship between age group of drivers and rating |
| Obstacles | * Face difficulties when creating charts using safety dataset as it is too big |