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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Adm. No. | Name | No. of hours present | Progress1 | Remarks |
| 1. 2122883 | Chia Wee Leong | 10 | A | - |
| 2. 2122924 | Ryan Ueda Teo Shao Ming | 10 | A | - |
| 3. 2123165 | Nimmagadda Vineesh | 10 | 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: 1-2 Date: 13/1/2023 | |
| Member Name 1: | **Wee Leong** |
| Last week’s Progress | * NIL |
| This and next week’s deliverables | * Create pipeline between SQL Server and Python * Combined rows from safety, driver and sensor into one dataframe. * Aggregated records with duplicated driverID based on mean, median, min and max values. * Placed the aggregated records onto new columns |
| Obstacles | * There was a large amount of data to aggregate, so the team had to decide on how to aggregate the data such that information loss is minimized. * After the records were aggregated, the number of columns increased greatly. Thus, the team had to figure out what type of feature selection needs to be done to reduce the number of columns. |
| Member Name 2: | **Ryan** |
| Last week’s Progress | * NIL |
| This and next week’s deliverables | * Performed EDA on combined data obtained from Wee Leong * Performed feature importance analysis using RandomForestClassifier * Performed outlier detection using One-Class SVM. |
| Obstacles | * There were too many features on the aggregated data. Thus, each feature’s importance was miniscule, and it was hard to extract meaningful insights. * There was little to no improvement in model scores when outlier detection was done, which made the team worry if the implementation was done wrongly. |
| Member Name 3: | **Vineesh** |
| Last week’s Progress | * NIL |
| This and next week’s deliverables | * Performed PCA on reduced data obtained through feature importance * Performed outlier detection using Isolation Forest. * Performed outlier detection using One-Class SVM |
| Obstacles | * It was challenging to determine the number of principal components to extract for PCA, as the 3 methods of determining the number of PCs to extract were different. * It was hard to extract meaningful insights from each principal component. * There was little to no improvement in model scores when outlier detection was done, which made the team worry if the implementation was done wrongly. |