



Submission Guidelines:

- Complete **BOTH** tasks mentioned below using the corresponding data sets provided.
- Expected Output:
 - Jupyter NoteBook / Google Codelab (separate book for each task)
 - Key findings document for the Analysis done in the notebook.
- If it is determined that the project has been clearly copied from any source, the candidate will be disqualified

• **Submission Date Time: Saturday - 07th Sep - 10 AM**

- **Task 1: Exploratory Data Analysis (visualizations) and Insights**

(Data to be used : Test1_1.csv, Test1_2.csv, Test1_3.csv)

1. Explain the preprocessing steps taken in detail (deletion/addition of columns/rows etc)
2. Conduct a comprehensive EDA using all the above provided data sets accompanied by visualizations to provide an in-depth analysis of the dataset. Additionally, note down key takeaways.

- **Task 2: Predictive modelling**

(Data to be used : Test2.csv)

1. Classification ML Model for Crash Descriptor Prediction:
 - Build a predictive ML model of your choice to predict the variable **crash descriptor**, using the provided dataset (Test3.csv)
 - Document the process of model selection, including the rationale behind the chosen model and any preprocessing steps applied to the data.
 - Evaluate the model performance with metrics such as accuracy, precision, recall, f1 score.

2. Time Series Forecasting for Monthly Crash Counts:

- Build a time series model, to forecast the number of crashes that occur month wise, using the provided dataset (Test3.csv)
- Document the process of model selection, including the rationale behind the chosen model and any preprocessing steps applied to the data.
- Evaluate the model performances by getting the MSE, MAE, RMSE along with the mean of the test set.