## Mercedes Benz Project

The **objective** of the project is to develop a ML algorithm that can accurately predict the time a car will spend on the test bench based on the vehicle configuration

The following steps are carried out before doing the modeling and prediction:-

1. Data acquisition and loading

2. Data preprocessing and Exploratory Data analysis

3. Missing value imputation, checking duplicates

4. Outlier Analysis

5. Target variable analysis- normality, plots and histograms

6. Univariate analysis

7. Bivariate analysis

8. Categorical columns analysis & checking the unique values in train and test dataset

9. Label encoding of categorical columns

10. Checking variance of the features, with Variance thresholder

11. Dimensionality reduction-PCA

12. Residual analysis to check linearity, equal variance and normality of residuals

ML algorithms and predictions:-

12. Linear Regression- test r2 score -5.050270899339964e+17

13. Linear Regression with PCA-test r2 score -0.9233975941396504

14. Random Forest Regression-test r2 score 0.49605679165794636

15. Random forest regression with parameter tuning-test r2 score 0.566504334178492 (The bias variance tradeoff was better after the parameter tuning)

16. KNN Regressor- test r2 score 0.3150917474026599

17. KNN Regressor with parameter tuning-test r2 score 0.4447419476928036

18. Adaboost Regressor – test r2 score -0.26341362500933085

19. Gradient Boosting Regressor-test r2 score 0.5652914004198634

20. XGBoost Regressor

Training Score for XGboost Regressor ------------------------------

Mean squared Error 66.45753043603483

r2 score 0.5918529577268289

Testing Score for XGboost Regressor -------------------------------

Mean squared Error 66.77907151393237

r2 score 0.5675887570022307

21.XGBoost Regressor with tuning test r2 score-0.5672818660280121