# Data Preperation

The preparation of data was taken care of data carefully where the following steps were taken:

* Firstly the columns that came unwontedly were removed line the team\_id – as the values were more of less same
* Then came the initial correlation step were the the relations of various columns were taken with the target column
* The for the columns were the missing values resembled categorical value there I placed it with “Unknown” value
* For the numerial columns – they were imputed with there mean values

# EDA

For EDA please refer to the power point presentation

# Model Building

For model building I made various model starting with the most basic binary classification algo-Logistic Relation.

For there I moved on to more advanced like the RandomForest amongst others. But ultimately steeled on LightGBM for it gave the best result amongst present.

After selecting a base model I started with the experimentation process in which the best parameters were to be chosen.

# Conclusion

The following are the conclusion:

* Shot\_basics and range\_of\_shot were the most important columns – they were determined with the help of heatmap and feature importance after generating the model
* The most influential parameters in the lgbm model were the tree leaf numbers and the tree depth.
* The lightgbm and GXBoost model closely resembled the accuracy score given.
* The missing values played a huge role in determining the accuracy score improvement.
* There was a mismatch in the instructions given to remove the test data and the file that was to be submitted.