## Baseball

Baseball is a bat-and-ball sport in which two opposing teams, each consisting of nine players, take turns batting and fielding. The game begins when a pitcher from the fielding team throws a ball that a batter from the batting team attempts to hit with a bat. The offensive team's (batting team) goal is to hit the ball into the field of play, allowing its players to advance counter-clockwise around four bases and score "runs." The defensive team's (fielding team's) goal is to keep hitters from becoming runners and runners from moving around the bases. When a runner legally moves around the bases in order and hits the home plate, a run is scored (the place where the player started as a batter).

The winning team is the one that scores the most runs by the end of the game.

The batting team's first goal is to have a player safely reach first base.

If a batter reaches first base without being ruled "out," he or she might attempt to move to the next base as a runner, either immediately or during his or her teammates' turns at bat. The fielding team seeks to prevent runs by getting hitters or runners "out," or removing them from the game. Both the pitcher and the fielders have strategies for getting the batters out. The opposing teams alternate batting and fielding turn, with the batting team's turn ending once the fielding team registers three outs.

An inning is one turn of batting for each team. A game normally consists of nine innings, with the team scoring the most runs at the end of the game-winning. Extra innings are frequently played if the score is tied after nine innings. Although most games end in the ninth inning, baseball does not have a game clock.

## Why Baseball Sports Analytics Project?

The Project can be used to analyse baseball player's performance data

## Dataset

The discipline of sports analytics is exploding. To study the performance of players and teams, owners, coaches, and fans use a variety of statistical metrics and models. The examination of annual statistics on hitting averages for individual players in the sport of baseball provides a basic example. The sample utilised here comes from the Lahman Baseball Database and contains 4535 rows of data for a select selection of players from 1960 to 2004.

## Catboost

Gradient Boosting is a machine learning approach that is commonly used for classification and regression issues. It's simple to use and works well with both heterogeneous and small data sets. It effectively turns a group of many weak learners into strong learner. Yandex developed CatBoost, or Categorical Boosting, an open-source boosting library. CatBoost can be used in ranking, recommendation systems, forecasting, and even personal assistants, in addition to regression and classification.

Advantages of CatBoost:

* On several datasets, superior quality when compared to other GBDT libraries.
* Prediction speed is the fastest in the class.
* Both numerical and category features are supported.
* Out-of-the-box GPU and multi-GPU support for training.
* Tools for visualisation are offered.
* With Apache Spark and the CLI, distributed training can be done quickly and consistently.

## Understanding Code

[Code]

[Code explanation]

## Short Description

Baseball is a bat-and-ball game played between two opposing teams, typically of nine players each, that take turns batting and fielding.

The project aims to predict total runs scored by baseball players.

The dataset utilised here was collected from the Lahman Baseball Database and contains 4535 rows of data for a select sample of players from 1960 to 2004.

'r2\_score' has been used to check the model's performance.

## Keywords

baseball sports analytics, baseball, play, sports, entertainment, players,

game, ground, ball, sport, batting, bat, ball, runs, bowling, fielding, teams, bases, teammates, running, fans, catboost, gradient boosting, boosting, machine learning, algorithm, prediction, supervised learning, structured data, artificial intelligence