

## Playing Moneyball in the NBA

---

In this recitation we will apply some of the ideas from Moneyball to data from the National Basketball Association-- that is, the NBA.

1. Read in the data from file  
[https://storage.googleapis.com/dimensionless/Analytics/NBA\\_train.csv](https://storage.googleapis.com/dimensionless/Analytics/NBA_train.csv)
2. Save the data in dataframe "NBA"
3. Explore it a little bit using the "str" command.
4. We have 835 observations of 20 variables.
  - SeasonEnd: Year the season ended.
  - Team: name of the team.
  - Playoffs (binary variable): Whether or not a team made it to the playoffs that year.
  - W: Number of regular season wins.
  - PTS: points scored during the regular season.
  - oppPTS: opponent points scored during the regular season.
  - FG: number of successful field goals.
  - FGA: number of field goal attempts.

PS:- 'A' behind the name of any variable means, number of attempts such as FG and FGA ; FT and FTA

- X3P:- Number of 3 pointers scored
- X2P:- Number of 2 pointers scored
- X2PA:- Number of 2 pointers attempted
- X3PA:- Number of 3 pointers attempted

PS: If we start any variable with a number, R will automatically convert it to start with a character by adding an "X" in front of that. Our csv file didn't have any variable with X. It had 2P, 3P etc.


- ORB: Offensive rebounds.
  - DRB: Defensive rebounds.
  - AST: Assists.
  - STL: Steals.
  - BLK: Blocks.
  - TOV: Turnovers
5. The goal of a basketball team is similar to that of a baseball team, making the playoffs. So how many games does a team need to win in order to make the playoffs?
  6. In basketball, games are won by scoring more points than the other team. Can we use the difference between points scored and points allowed throughout the regular season in order to predict the number of games that a team will win?  
  
(Hint: - First add a variable that is the difference between points scored and points allowed. Let's call this NBA\$PTSdiff.)
  7. Make a scatter plot to see if it looks like there's a linear relationship.
  8. Create a linear regression model to predict the number of wins. Use PTSdiff as your independent variable, and W for wins as the dependent variable. Call this 'WinsReg'.
  9. Look at the summary of WinsReg.
  10. Write down the regression equation that you have found.
  11. What should be the points difference, if we want Wins to be greater than or equal to 42.


## Linear Regression with common basketball statistics

Our dependent variable would now be PTS, and our independent variables would be some of the common basketball statistics that we have in our data set.

# Dimensionless Technologies Private Limited

Visit us at: [www.dimensionless.in](http://www.dimensionless.in)

 - [info@dimensionless.in](mailto:info@dimensionless.in)

 - 9923170071, 8108094992



Eg:- the number of two-point field goal attempts, the number of three-point field goal attempts, offensive rebounds, defensive rebounds, assists, steals, blocks, turnovers, free throw attempts.

1. Build a regression model using X2PA, X3PA, FTA, AST, ORB, DRB, TOV, STL, BLK and name it PointsReg
2. Look at the summary of PointsReg
3. Compute RMSE.
4. Calculate the average number of points in a season and compare it with RMSE.

## Improving the model

Since not all the variables were significant. Let's see if we can remove some of the insignificant variables one at a time.

1. Which variable we should remove first and why?
2. Remove that variable, and create a new model 'PointsReg2'. Check the  $R^2$
3. Remove another one of the insignificant variables and create a new model 'PointsReg3'. Check the  $R^2$
4. Remove another insignificant variable and create a new model 'PointsReg4'. Check the  $R^2$
5. Also calculate the RMSE for this model.

## Making Predictions

We'll try to make predictions for the 2012-2013 season.

1. Load the dataset "NBA\_test.csv" in a dataframe NBA\_test.
2. Predict the number of points scored and save it in PointsPrediction.
3. Compute "out-of-sample" R-square.
4. Also compute RMSE.