""" This is a small exercice that I have created for a workshop that I've done to my colleagues in the company"""

Application on Linear Regression : NBA

Goal : -Become familiar with the use and the interpretation of a data frame  
 -Become familiar with the Linear Regression method  
 -Calculate SSE, RMSE, R2  
 -Make Predictions

Part 1

* Load train data “NBA\_train.csv” using read.csv( ) function. Name your variable NBA   
  *Our data contains stats from all teams in season since 1980*
* Explore your data using str( ), summary( )  
  *#Some variables start with X cause R doesn’t accept that variables start with a number*
* How many times a team won 11 games? list these teams? (use table & subset function)
* How many times a team with 43 wins didn’t make it to the playoffs (use table function)   
  How many wins a team need (at least) to be in the playoffs?
* Add point difference variable to our data, Call it PtsDiff
* Create a plot of teams wins according to PtsDiff using plot()  
  *#So we can see that linear regression will be a good fit to our model*
* Create ur linear regression model using lm( dep ~ indep1 +indep2+.. , data=..), name it WinsReg
* Explore it using summary()
* What is our linear regression equation? How many games a team with 200 PtsDiff will win?
* Using abline() function show the regression line on your plot
* How many points difference do we need to be in the playoffs?

Part 2

* Create a linear model to predict points scored using X2PA + X3PA + FTA + AST + ORB + DRB + TOV + STL + BLK, name it PointsReg
* Which variables are significant?
* Calculate the sum of squared errors SSE (You can access to the errors vector using PointsReg$residuals). Is it interpretable?
* Calculate the root mean squared errors RMSE (it’s like the average error we make) .  
  (use nrow(NBA) to get the number of observations)
* What is the less significant variable in our model?   
  Create PointsReg2 without the use of this variable.   
  Compare the R squared value
* Create the best possible model and compare it to the initial one that uses all the variables

Part 3 – Predictions for 2012/2013

* Load the test set, name your variable NBA\_test
* Make predictions ( for season 2012/2013) using predict( model ,newdata= dataframe ) and the testing
* Calculate SSE, SST , R2