**Project 1 - Historical NBA Analysis**

**Questions**

**We have a dataset of NBA stats for players and team from 1947 till 2017. For this project, we will only be considering the data from 1986 till 2016 as most of these parameters were not calculated till mid ‘70s.**

* Can the Team Four Factor Rating of Basketball help in accurately determining the team’s win/success?

\* Four Factors are –

- Shooting – Effective Field Goal % - eFG% - (40%)

- Turnover – Turnover Rate % - TOV% - (25%)

- Rebound – Offensive Rebound Rate % - ORB% - (20%)

- Free Throw – Free Throw Rate – FT% - (15%)

\* Team Four Factor Rating Formula = ((0.4\*eFG%) – (0.25\*TOV%) + (0.2\*ORB%) + (0.15\*FT%))

* We will be using data of 20 years from 1986 till 2017
* We can first plot individual scatter plots for win% against each of these factors to see how they correlate with it.
* Then we can have Team Four Factor Rating on the x axis and win% of the team on the y axis
* We will have scatter plot for win% against Team Four Factor Rating for a selected team
* We will plot a line regression which will determine how Team Four Factor Rating has a correlation with win% of the team
* Based on our line coefficient equation (y = mx + b), we can then predict win% for 2017 data
* We can then check the actual win% of the team from the 2017 data and determine the accuracy of our prediction

**How does age of the player affect PER (Player Efficiency Rating)?**

* Player’s age will be on the x axis and PER will be on the y axis
* We will have Scatter plot for PER against the age
* We will plot a linear regression to determine correlation between player’s efficiency and age
* Based on our line coefficient equation (y = mx + b), we can then predict player’s efficiency based on their age
* We can then plot a comparison line graph for a selected player’s PER for their entire career along with line graph for average PER
* We can visually check how selected player’s PER was compared to average PER

**How does a player’s move (Top 10 only based on overall points) affect the new team’s winning percentage?**

* Based on the “Points” field, we can sort our Top 10 Players of the selected years
* We can pick any player from these Top 10 and we will check the column when their team changes and for which year
* Based on this data, we will find the team win% for the previous year and compare both to see if the team performed better or worse

**We can find the top 50 or 100 players in the dataset and create a subset of just their data.**

* Once this is created, we will run a regression containing MP (minutes played) as the dependent variable and other players’ statistics as independent variables as well as dummy variables for every year using a by function (this is a R function that we will figure out in Python) in order to run this regression on each individual player.
* Based on this we will be able to create a time series for each player using the coefficients recorded for each individual year.
* Once this data is visualized, we will be able to track players’ performance over each year and see which players were best in which year.

**Methodology**

* Timeframe
  + 1867 - 2017
* Data Points
  + Four factors
  + PER (Player Efficiency Rating)
  + Age
  + Winning Percentage
  + Players move
  + Points
  + Minutes Played
* Sources
  + <https://www.kaggle.com/drgilermo/nba-players-stats?select=Seasons_Stats.csv>
  + <https://data.world/gmoney/nba-team-records-by-year/workspace/file?filename=Historical+NBA+Performance.xlsx>
  + <https://www.sportsgamblingpodcast.com/2020/04/20/nba-most-valuable-statistic/>
  + <https://pypi.org/project/nba-api/>

**Team Roles**

* Kate
  + Analytics Outline/ Data / Analysis
* Jay
  + Analytics Outline/ Data / Analysis
* Dennis
  + Analytics Outline/ Data / Analysis
* Byrant
  + Analytics Outline/ Data / Analysis
* Sam
  + Analytics Outline/ Data / Analysis