# DS 3000 – Dataset

## Project Topic Idea: <How NBA Players perform relative to their contracts.*>*

Of the project ideas you submitted in the previous deliverable, select the one you want to work on throughout this project. In making your decision, please refer to the feedback you have on the previous deliverable. If all ideas were found plausible, feel free to select the one that you liked the best **as a team**. You will be working on this project idea during the remainder of the semester.

Complete this section based on the previous deliverable. After deciding on your topic idea, simply copy and paste same information from the Topic Proposals document here.

### Problem Statement

* Describe the problem you would like to tackle.
  + We would like to understand how players in the NBA are performing relative to their contract and find the contract that the player is actually worth.
* What is the topic of your project?
  + NBA players’ performance relative to their salary, and what they are actually worth.
* What do you want to learn about it?
  + What players are outperforming their contract values, and what players are getting overpaid for their performance.

### Significance of the Problem

* Why is it important to tackle this problem in your project?
  + In the NBA, especially with a salary cap, it is important to utilize the money that you give to players wisely. In order to win the championship, teams must not give out large contracts to players that do not actually deserve it. You must have players that perform at or above their contract value, in order to be successful in the league.
* In what ways could the insights from this project be useful?
  + It could help NBA GM’s and players’ agents understand what the player is worth, when negotiating contracts. As well as, show teams what players should be traded, as they are not performing up to the value of their contract.

### Dataset(s)

* **When searching for potential datasets, you can refer to the Datasets folder on Canvas to search potential dataset repositories. This has been available on Canvas since the early weeks of the semester.**
* Describe where you obtained your data. Provide a link to the original source.
  + The data was obtained in two locations. Most of it came from basketball-reference.com and we also got data from Celtics Hub (who got their data from various places). The links were:
    - <https://www.celticshub.com/2017/12/07/nba-player-salaries-1991-2017/>
    - <https://www.basketball-reference.com/leagues/NBA_2013_advanced.html>
    - <https://www.basketball-reference.com/leagues/NBA_2014_advanced.html>
    - <https://www.basketball-reference.com/leagues/NBA_2015_advanced.html>
    - <https://www.basketball-reference.com/leagues/NBA_2013_per_game.html>
    - <https://www.basketball-reference.com/leagues/NBA_2014_per_game.html>
    - <https://www.basketball-reference.com/leagues/NBA_2015_per_game.html>
    - <https://www.basketball-reference.com/contracts/salary-cap-history.html>
* This should be the dataset(s) you are using and should correspond to the attached dataset.

Dataset File

Download or scrape your data from the source you identified above. Save your dataset as a CSV file. The first row of the file should contain variable names.

**Your dataset should have at least 1000 rows, corresponding to samples/records, and 10 columns, corresponding to features and target variables. This is the bare minimum. The more, the better!**

Describe your variables below (add more rows if necessary):

|  |  |  |
| --- | --- | --- |
| **Variable name in file** | **Description** | **Feature/ Outcome** |
| Pos | Position | Feature |
| Age | Player Age | Feature |
| G | Games Played | Feature |
| Gs | Games Started | Feature |
| MP | Minutes Played | Feature |
| FG | Field Goals Per Game | Feature |
| FGA | Field Goals Attempted Per Game | Feature |
| 3P | Three Points Made Per Game | Feature |
| 3PA | Three Points Attempted Per Game | Feature |
| 2P | Two Points Made Per Game | Feature |
| 2PA | Two Points Attempted Per Game | Feature |
| FT | Free Throws Made Per Game | Feature |
| FTA | Free Throws Attempted Per Game | Feature |
| ORB | Offensive Rebounds Per Game | Feature |
| DRB | Defensive Rebounds Per Game | Feature |
| TRB | Total Rebounds Per Game | Feature |
| AST | Assists Per Game | Feature |
| STL | Steals Per Game | Feature |
| BLK | Blocks Per Game | Feature |
| TOV | Turnovers Per Game | Feature |
| PF | Personal Fouls Per Game | Feature |
| PTS | Points Per Game | Feature |
| PER | Player Efficiency Rating | Feature |
| TS% | True Shooting % | Feature |
| 3Par | Three Point Attempt Rate | Feature |
| FTr | Free Throw Rate | Feature |
| ORB% | Offensive Rebound % | Feature |
| DRB% | Defensive Rebound % | Feature |
| TRB% | Total Rebound % | Feature |
| AST% | Assist % | Feature |
| STL% | Steal % | Feature |
| BLK% | Block % | Feature |
| TOV% | Turnover % | Feature |
| USG% | Usage % | Feature |
| OWS | Offensive Win Shares | Feature |
| DWS | Defensive Win Shares | Feature |
| WS | Win Shares (Total) | Feature |
| OBPM | Offensive Box Plus/Minus | Feature |
| DBPM | Defensive Box Plus/Minus | Feature |
| BPM | Box Plus/Minus | Feature |
| VORP | Value Over Replacement Player | Feature |
| Year | Year which Season occurred | Feature |
| UID | Unique User ID | Feature |
| Salary | Amount earned by player | Feature |
| Cap% | % of team’s salary capacity space | Outcome |
| In the Feature/Outcome column, indicate whether the variable is a feature or outcome variable. You need to have at least one outcome variable, with several feature variables. |  |  |
|  | | |

**Based on what we discussed regarding machine learning (Week 07 Day 02), does your dataset include a set of feature variables and one outcome variable that you can use for a supervised machine learning task? Please explain. You need to meet this requirement and show us you understand that you are required to use a predictive model in your project.**

Yes, our dataset does include a set of feature variables and one outcome variable that we can use for supervised learning. Since a player’s stats correlates to how much money they make, their stats will likely allow a machine learning algorithm to classify players that it has not seen in training.

**Further info on submitting the dataset:**

Submit a CSV file, or multiple files, containing your data. If the dataset is too large, you can upload it to Github or any other online repository, and provide a public link.

If you have scraped your data, you should also submit a Jupyter Notebook containing your Python code used to scrape the data. Please be reasonable and comment your code out whenever it makes sense to do so.