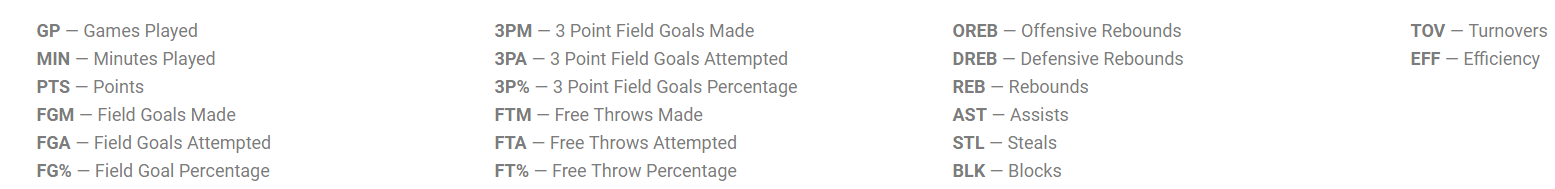
**Project Title:** Examining Points Scored in the NBA

**Project Abstract:**

* **Situation:** Examining points scored by each player in the NBA (2018-2019) season.
* **Task:** The overall motive of this project is to examine the top points scored per game for the 2018-2019 season and see whether there is any correlation between the number of games played (GP) and the time spent in game (MIN).
* **Action:** I’m going to examine the data set on stats.nba.com in order to figure out if there is any overall correlation between the number of points earned per game to the number of games played / time spent within each game.
* **Results:** After examining all the data, I will be able to deduct whether or not the highest point earners per game also spent the most amount of time per game as well as look at whether or not the data is consistent for each player / only had one game where they scored infinite points (outlier).

<https://stats.nba.com/leaders/?Season=2018-19&SeasonType=Regular%20Season>

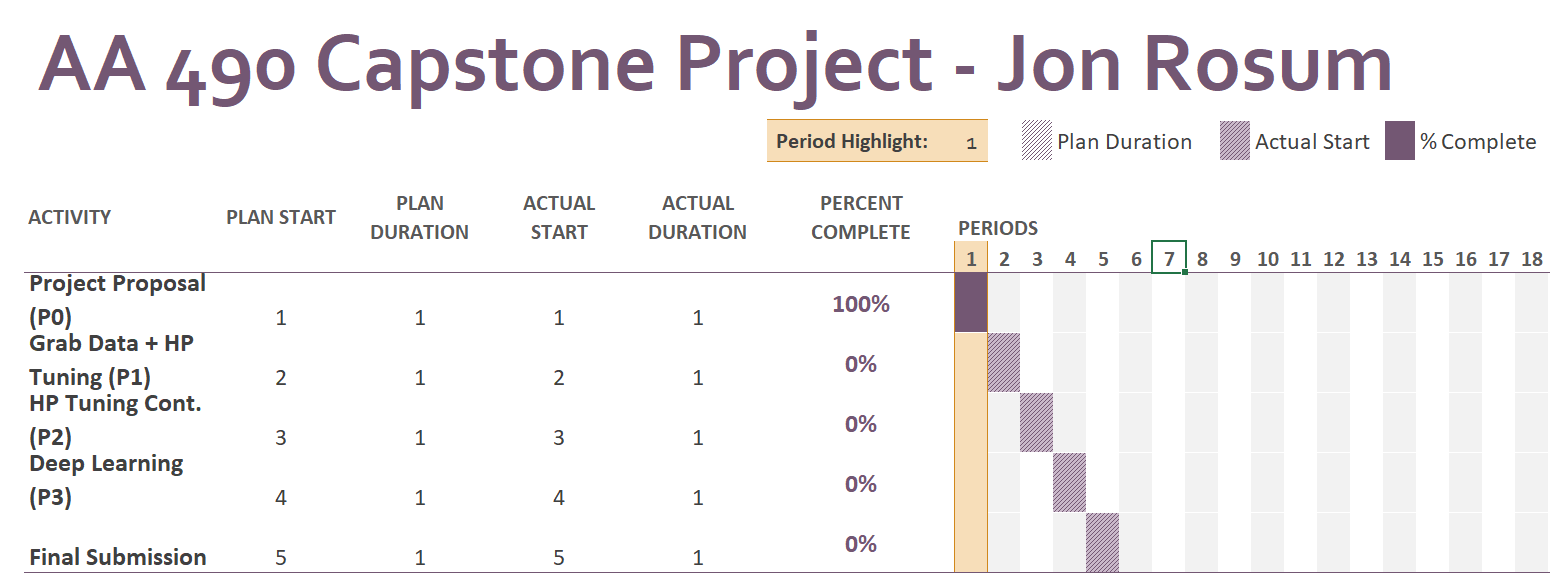


**Maching Learning Algorithm + Pipeline Steps:**

The machine learning algorithm strategy that I’m considering the most is Supervised Learning. The input data (called training data) that has a known label or result that we are examining is PTS (Points) and whether there is a correlation between GP (Games Played) and MIN (Minutes Played). The analysis within this project will be repeated until there is a desired level of accuracy reached. The algorithms that I will use are Logistical Regression and Back Propagation Neural Network. Logistical Regression will be a good tool in order to define any correlation that might exist within these variables. I will start with normalizing the data set and then running several logistical regressions + neural networks in order to maximize accuracy.

**Metrics Used to Consider Success:**

The most important metric that is going to be used in order to evaluate success is going to be Classification Accuracy. Within this analysis we want to ensure that the accuracy is high in our findings to prevent misclassification from occurring. The other metric that can be used that correlations with Classification Accuracy is Log Loss. When the Log Loss of the analysis is quite low that means that the accuracy is quite high and makes it so that analysis was effective.



**Team Members:** Jonathan Rosum, Jrosum@bryant.edu