Names: Reid Whitson and Luke Latiolait

Project Name: ?

\_\_\_\_\_\_\_\_(name of project) will take all of the players in the NBA over recent seasons and their associated stats and then given a new player, we will predict how many points they will score. We are using a dataset from kaggle that has all of these data points. This dataset is very extensive and we will be taking a few of the columns which we think have the greatest impact on the score. As is, the dataset size is 24,700 x 53; we hope to narrow down the data to more recent years (1990 onwards) and narrow down  the columns to statistics that we see as having more direct impacts on players points per game (PPG).

Narrowing down the data will be a large part of this project, and is part of the pre-processing phase. I anticipate that we may run into issues when deciding what is most important with which attributes to keep and which to get rid of. This is very important as it really defines what are deciding factors in out classifier. If we choose bad attributes, we could get very inaccurate results. We are planning on doing a KNN classifier and then comparing it to the results of sci-kit learn and discussing the differences. The reason why we choose this is because we both wish to learn sci-kit learn and believe this is a good way for us to dive into the library and analyze the results between it and a homegrown tests.

Things to clarify

1. KNN?
2. Do we want to do something between college and NBA? I think it would be cool