1. **Describe the dataset. (ex: The number of microaneurysms found in a patient's eye and whether or not they have diabetic retinopathy; or Lending Tree loan data, including who defaulted and who paid off their loan )**

Data set revolves around the NBA, focuses on team statistics. Things like standard counting stats (PTS, 3P%) will be used along with Advanced stats (TS%, Per 48, 100 possessions).

1. **How many records does the dataset have?**

With 20 years of data (2000 - 2020) and 30 teams each year that gives us at least 600 records.

1. **How many features does the dataset have? List or describe a few of them.**

There are 30 features:

PTS – Raw average of points scored by a team per game.

3P% - Raw average of 3-point percentage by a team per game.

TS% - Average true shooting by team per game.

PACE – Average pace of a team per game.

1. **What can you try to predict in this dataset? (ex: We can use the number of microaneurysms measured in the patient's eye to predict whether or not they have diabetic retinopathy; or We can try using the features, including age, income, home ownership status, etc, to predict whether or not someone will default on their loan.)**

This dataset is being used to predict if a team will make the playoffs based on current team stats, using previous years’ team data in the prediction model. We plan on having two models with the only difference being, the second model will use per 100 possession data instead of the standard raw team data. We’ll use this to compare our findings between the two models.

**5. Is this a labeled dataset, appropriate for a supervised learning classification problem? (In other words, if you are trying to predict whether or not someone has a disease, does your dataset contain whether or not each record has the disease?)**

This is a labeled dataset using supervised learning classification. We are trying to predict if a team makes the playoffs based on team stats (standard, 100 possessions, and per 48 mins). The label is ‘Playoffs’ with a yes or no.

1. **Provide a link to the dataset, if there is one. If you are getting your data from somewhere other than a link, where are you getting it from?**

We’re pulling data using an NBA API from the official NBA stats page. We’ll pull both the standard raw team data and advanced team data from the NBA stats site.

<https://stats.nba.com/teams/traditional/?sort=W_PCT&dir=-1&Season=2019-20&SeasonType=Regular%20Season&PerMode=Per100Possessions>

<https://github.com/swar/nba_api>