**Basketball Stats**

**Personal Project by Bryan Delas Penas**

Dates: 12/17/2020 – Current Date

**Currently on Phase 1**

**Description:**

This project will be broken down into three main phases and an optional four phase. By using GitHub as a version control, it allows for seamless version updates. The phases can be broken down into the following:

**Phase 1: Python Scraping**

By using beautifulsoup4, we can scrap data from basketball-reference.com. We are scraping data from 1980 to 2020, as this is when the three-point shot was added in basketball. This phase can be broken into four main directories.

1. *Basketball-Stats\\Output*
2. *Basketball-Stats\\python\_scrapers*
3. *Basketball-Stats\\Run\_Scrapers*
4. *Basketball-Stats\\Test\_Scrapers*

*Basketball\\Output:* Contains the output of the web scraper in a common separated file (csv). It is broken down into two sub-directories:

*Basketball-Stats\\Output*

1. *\\Season\_Stats*
2. *\\Team\_Stats*

*Basketball-Stats\\Output\\Team\_Stats* contains subdirectories that are named based on the seasons. It ranges from 1980 to 2020. Inside of these subdirectories is the following

*Basketball-Stats\\Output\\Team\_Stats*

1. *\\Pergame*
2. *\\Perposs*
3. *\\Team\_Roster*
4. *\\Total*
5. Per-game is the average stats based on the 82-game season.
6. Per-poss is the average stats based on several possessions
7. Team\_Roster is the players that are in a certain team.
8. Total is the total amount per stat

**Phase 2: Creating the Database**

**Phase 3: Predicting MVP**

***Pre-Planning:***

By using a random forest, and using the database from the following phase, we can create a program in Jupiter Notebook that trains and predicts the season MVP based on certain features.