

```
In [1]: # Libraries
import pandas as pd
```

PSTAT 100 Project plan report

This is a guide to preparing your project plan. It functions both as a guide to the work you'll need to do and as a guide to preparing the deliverable. You can use it as a template to draft the plan report; if so, please remove the text explanations of each section.

While you may find it useful initially to follow the outline given, you do not need to adhere to it exactly -- you're free to organize your submission in the way that seems most natural to you. However, please do keep the high-level sections, so that your report includes the following headers:

- 1. Background
- 2. Data description
- 3. Initial exporations
- 4. Planned work

Your report does not need to be long. It should be about 2-4 pages, and may not be much longer than this template once you replace the guiding text with your own work.

Group information

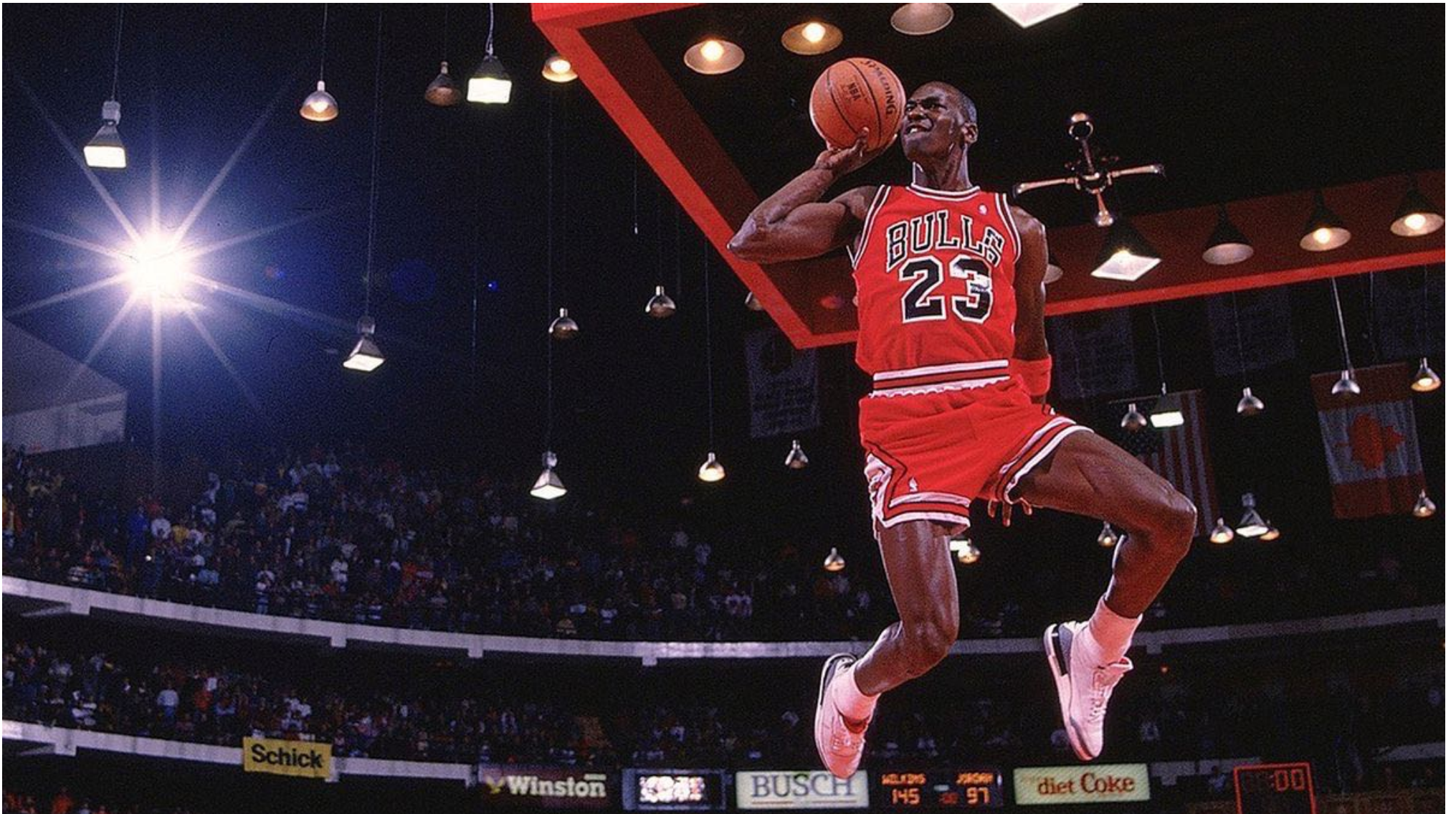
Group members: Haiting Huang, Matthew James Brenna, Ermin Pinjic, Haorui Zhang

Contributions:

- 1. Matthew prepared the data description and tidying the dataset.
- 2. Haorui worked on tidying the dataset and initial exploration.
- 3. Ermin worked on data background and data description.
- 4. Haiting finished planned work and initial exploration.

0. Background

NBA, the one place where the most talented basketball players from all over the world play together -- or against each other, in order to accomplish one goal; to win their team a championship. As players continue through the regular season, their game statistics are recorded and averaged out after the 82 games are finished. If a team is able to get into the playoffs, their statistics will continue to be recorded during the playoffs.\ Gathering data from a reliable NBA website, there a several variables which can be drawn out. In particular: the amount of minutes played, field goal percentage, usage percentage, offensive box plus/minus, win shares, age, team, and season will be analyzed.\ The reason this data is being presented is to determine how a players statistics are affected in the regular season compared to playing playoff basketball, where the stakes of winning are much higher. Furthermore, we can also determine if a specific team is more dependent on a certain player during the playoffs compared to the regular season.



1. Data description

General description The data are observations of NBA player performance during the regular season and playoffs of one NBA season, from 2017-2021.

Source The data came from [basketball-reference.com](https://www.basketball-reference.com). Their website is part of a larger group that collects data for sports such as basketball, football, baseball, etc.

Collection methods The data on [basketball-reference.com](https://www.basketball-reference.com) is provided by SportRadar, which is the official statistics provider of the NBA. This data is presumably collected by watching and recording the events of NBA games.

Sampling design and scope of inference The population is all players from 2017-2021 who played in both the regular season and playoffs that season. There is no scope of inference for this data (it is not generalizable) because it is an exhaustive list of all NBA players who played in both the regular season and playoffs for these years, not a sample.

Units and Observations

Each observational unit is one NBA player's information and statistics from the regular season and playoffs from one NBA season.

Variable Descriptions

| Name | | Variable description | Type | Units of measurement | |
|-----------------|---|----------------------|--|---|--|
| Player | Name of player and code from initial website | Character | | n/a | |
| Age | Age of player during season | Numeric | | Age | |
| Tm | Team the player played for | Character | | n/a | |
| MP_regseason | Minutes played per game (regular season) | Numeric | | Minutes per game | |
| FG%_regseason | Percentage of Field Goals Made (regular season) | Numeric | | Field goals made/Field goals attempted_ | |
| USG%_regseason | Usage Percentage (regular season) | Numeric | Percentage of team's plays that the player "ended" with a Field Goal Attempt, Free Throw Attempt, or Turnover | | |
| WS/48_regseason | Win Shares per 48 minutes (regular season) | Numeric | read https://www.basketball-reference.com/about/ws.html (https://www.basketball-reference.com/about/ws.html) | | |
| OBPM_regseason | Offensive Box Plus-Minus (regular season) | Numeric | read https://www.basketball-reference.com/about/bpm2.html (https://www.basketball-reference.com/about/bpm2.html) | | |
| MP_playoffs | Minutes played per game (playoffs) | Numeric | | Minutes per game | |
| FG%_playoffs | Percentage of Field Goals Made (playoffs) | Numeric | | Field goals made/Field goals attempted_ | |
| USG%_playoffs | Usage Percentage (playoffs) | Numeric | Percentage of team's plays that the player "ended" with a Field Goal Attempt, Free Throw Attempt, or Turnover | | |
| WS/48_playoffs | Win Shares per 48 minutes (playoffs) | Numeric | read https://www.basketball-reference.com/about/ws.html (https://www.basketball-reference.com/about/ws.html) | | |
| OBPM_playoffs | Offensive Box Plus-Minus (playoffs) | Numeric | read https://www.basketball-reference.com/about/bpm2.html (https://www.basketball-reference.com/about/bpm2.html) | | |

```
In [2]: # Load tidied data and print rows
data = pd.read_csv('data.csv')
data.head()
```

Out[2]:

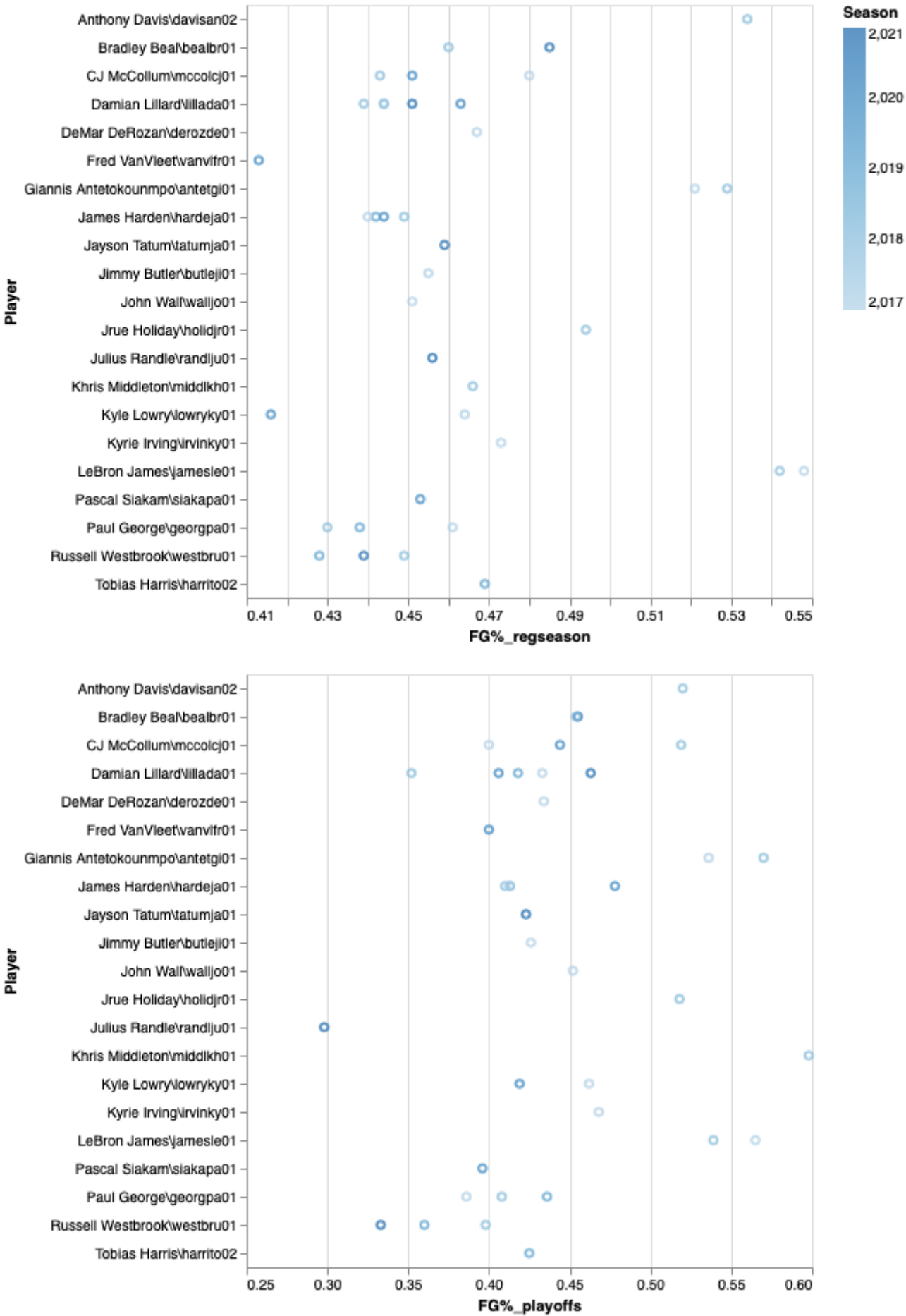
| | Player | Age | Tm | MP_regseason | FG%_regseason | USG%_regseason | WS/48_regseason | OBPM_regseason | MP_playoffs | FG%_pl |
|---|-----------------------------|-----|-----|--------------|---------------|----------------|-----------------|----------------|-------------|--------|
| 0 | Bogdan Bogdanović\bogdabo01 | 28 | ATL | 29.7 | 0.473 | 21.3 | 0.142 | 3.3 | 33.2 | |
| 1 | Clint Capela\capelca01 | 26 | ATL | 30.1 | 0.594 | 19.9 | 0.207 | 2.7 | 31.6 | |
| 2 | John Collins\collijo01 | 23 | ATL | 29.3 | 0.556 | 22.2 | 0.174 | 2.5 | 32.0 | |
| 3 | Kris Dunn\dunnkr01 | 26 | ATL | 11.3 | 0.083 | 16.2 | -0.202 | -13.7 | 6.6 | |
| 4 | Bruno Fernando\fernabr01 | 22 | ATL | 6.8 | 0.409 | 14.4 | -0.019 | -6.1 | 2.0 | |

2. Initial explorations

Dimensions: 1063 x 14

Missing values: No.

| | Name | min | max | mean | var |
|--|-----------------|--------|-------|---------|----------|
| | Age | 19 | 40 | 26.62 | 17.94 |
| | MP_regseason | 3.0 | 37.8 | 21.4 | 79.23 |
| | FG%_regseason | 0.083 | 0.833 | 0.46 | 0.00648 |
| | USG%_regseason | 5.9 | 41.7 | 18.7 | 32.391 |
| | WS/48_regseason | -0.246 | 0.459 | 0.107 | 0.0047 |
| | OBPM_regseason | -19.4 | 9.4 | -0.54 | 9.065 |
| | MP_playoffs | 0 | 43 | 19.47 | 152.14 |
| | FG%_playoffs | 0 | 1 | 0.433 | 0.02686 |
| | USG%_playoffs | 0 | 61.5 | 18.37 | 65.31 |
| | WS/48_playoffs | -1.045 | 1.801 | 0.065 | 0.02652 |
| | OBPM_playoffs | -35.8 | 39 | -0.7384 | 30.85847 |



This is players that play more than 35 minutes for both regular season and playoffs, so they are all star players for their teams. It is a good way to analyze whether star players play better or worse for playoffs and how it can influence the teams' performance.

3. Planned work

Questions

Please propose two focused questions that you plan to explore.

1. The difference between a player's performance in the regular season and the playoffs is judged by his MP_regseason, FG%_regseason, USG%_regseason, OBPM_regseason, and WS/48_regseason. Then we can find out which players' performance became significantly better/worse in the playoffs compared to the regular season.
2. The difference between a player's performance in the 2017-2021 regular season and the playoffs is judged by his MP_regseason, FG%_regseason, USG%_regseason, OBPM_regseason, and WS/48_regseason. Then identify the best player, who may have been good in the regular season and playoffs in recent years, or who has been improving.

Proposed approaches

For each question, please describe an idea or two about how you might approach the question.

- 1. *We need to know the teams that made the playoffs this year and the key players on that team to give us the best analysis. And we need to collect their respective data from NBA's official website or related websites (MP_regseason, FG% _REGseason, USG% _REGseason, OBPM_regseason, And WS/48_regseason.) to make a chart to analyze their performance.*
- 2. *We need to get a better picture of the teams that have made the playoffs in recent years and the players that have played so that we can compare their growth or decline in recent years. And the non-starters who are not in the rotation for the most part in the playoffs, and the guys who haven't played in the playoffs for a couple of seasons, I think, are not referenced, and in the end, just like the first question, We need to collect their MP_regseason, FG%_regseason, USG% _REGseason, OBPM_regseason, and WS/48_regseason for more accurate analysis.*

Submission Checklist

- 1. Save file to confirm all changes are on disk
- 2. Run *Kernel > Restart & Run All* to execute all code from top to bottom
- 3. Save file again to write any new output to disk
- 4. Select *File > Download as > HTML*.
- 5. Open in Google Chrome and print to PDF on A3 paper in portrait orientation.
- 6. Submit to Gradescope