Dataset links:

all_seasons.csv: https://www.kaggle.com/datasets/justinas/nba-players-data
games.csv: https://www.kaggle.com/datasets/nathanlauga/nba-games?resource=download

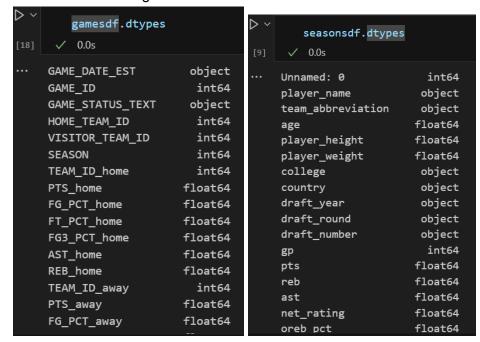
Both are csv

Entries for all seasons.csv: 12844

o Columns for all seasons.csv: 22

Entries for games.csv: 26651

Columns for games.csv: 21



- game.csv has a different format for the season column than the all_seasons.csv
- The game.csv needs to have the game date reformatted to be consistent with the years of all seasons.csv
- The all_seasons.csv can be cleaned on the I, J, K columns where "Undrafted" is a value among all the other int values in the column.
- There do not seem to be missing values, but there are inconsistent data formats between the two datasets (i.e. season column)
- Merging with seasons column after cleaning
- We are uniquely positioned to work with these files because Harsh is a big basketball fan. Aaron and Minji are not very familiar with basketball, so they could help gather an unexpected insight.