

# Linear Regression Project Proposal

Josh Jingtian Wang

2/4/2022

## Question/Need:

The former general manager of the Houston Rockets, Daryl Morey, was the first GM to apply analytics to basketball. His basketball philosophy (dubbed “Moreyball”) favors three-point field goals and lay-ups over mid-range jumpers. It is evident that “Moreyball” has swept over the league in the past five years with most teams adopting the three pointer-dominant strategy, but it is inconclusive whether “Moreyball” has contributed to team wins. With basketball being one of the biggest sports in the world, it would be important to look at if “Moreyball” is actually “money”.

## Data Description:

The primary data will be scraped from <https://www.basketball-reference.com/> (and <https://www.nba.com/stats/> if needed).

All input variables will be per-game: 3pt shots attempted of the winner, 2pt layups attempted of the winner, 2pt jumpshots attempted of the winner, 3pt shots attempted of the loser, 2pt layups attempted of the loser, 2pt jumpshots attempted of the loser, and possibly more.

The output variable will be the winning margin of the game.

A linear regression will be performed on every season from 199? to 2021. Then a time series analysis will be performed on the co-efficients to see if teams have adapted to Moreyball over the years.

## Tools:

Data scraping will be done via BeautifulSoup (Selenium if needed). Data cleaning will be done via Python’s pandas module. Linear Regression will be done via Python’s statsmodel module. Data visualization will be done with Python’s matplotlib and seaborn modules.

## MVP Goal:

Data scraping. EDA on the trend of 3pt and 2pt shots attempted over the past two decades.