**Part 1:**

The primary objective of this project is to evaluate potential biases in NBA 2K25 player ratings by identifying overrated and underrated players using machine learning techniques. The analysis will determine whether certain attributes disproportionately influence overall ratings and assess archetypes for players based on their given information.

**Part 2:**

* **Source:** Kaggle (Publicly available, ethical for use)
* **Size:** 526 players, 105 columns
* **Contents:**
  + Player Information: Name, team, nationality, height, weight, salary
  + Position & Archetype: Primary positions, predefined 2K archetypes.
  + Attribute Ratings: Shooting, defense, athleticism, playmaking, etc.

From the data set, the main variables of interest are the ratings, physical attributes, and position. NBA 2k uses archetypes, but there are 113 different archetypes, so they lose meaning for trying to group players when there are so many. To process the data, unnecessary columns like name and college were dropped. Height and wingspan were recorded in (ft’in”) format as well as cm format. Only the cm formatted variables were kept in order to have them easily stored as numeric data types.

**Part 3:**

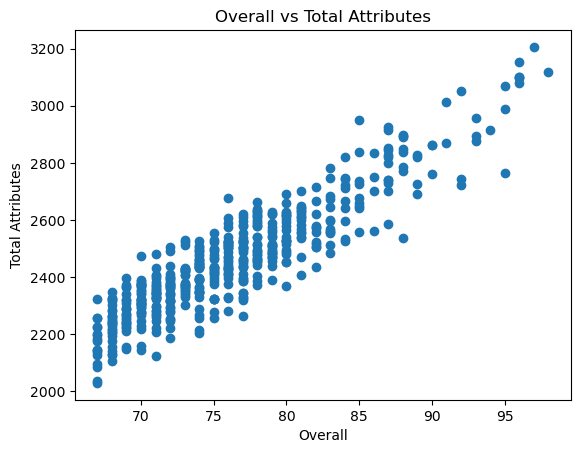
Summaries for all variables are contained below. Since the data is from a videogame, missing values are not an issue for the analysis and the sample size is consistent.

The correlation matrix of numeric features shows some high correlations between variables. Considering the large number of variables, it will be important to work on feature selection and feature importance to create a parsimonious and viable model that does not have issues with multicollinearity.

A close-up of a graph

AI-generated content may be incorrect.

**Part 4:**

A graph of a number of ratings

AI-generated content may be incorrect.

The scatter plot of attribute sum and overall rating shows that there is a generally linear trend between the two variables, however there is a fair amount of variance in total attributes across the overall ratings, indicating that some other factors must be contributing to the values beyond simple rating addition. The density plot shows the distribution of overall ratings and indicates right-skew in the overall ratings, where only a few players are achieving high overalls, which introduces interest into what differentiates those players from the rest of the group.

**Part 5:**

To this point, there have been no significant challenges in the data work. I anticipate challenges in feature selection and ensuring model assumptions are adequately met. I also anticipate having some difficulty with the dimensionality, so PCA will likely be helpful in reducing the dimensionality of the data to a more interpretable and useful format.

**Descriptive Statistics:**

*Descriptives are in the following format for numeric features:*

variable: count, mean, sd, min, 25%, median, 75%, max

position: 526, {PF: 107}, {SG: 106}, {PG: 105}, {SF: 104},{C: 104}

team: 525, 30 distinct (1 for every NBA team)

height\_cm: 526, 199.76, 8.02, 173, 193, 201, 206, 226

weight\_kg: 526, 95.56, 10.51, 72, 89, 95, 104, 137

wingspan\_cm: 526, 210.21, 9.33, 180, 203, 211, 216, 244

season\_salary: 399, 1.26e+07, 1.28e+07, 6.0e+05, 2.995020e+06, 7.723000e+06, 1.713000e+07, 5.576122e+07

years\_in\_the\_nba: 525, 5.79, 4.31, 1, 2, 5, 8, 22

overall: 526, 76.27, 6.44, 67, 71, 75.5, 80, 98

group\_outside\_scoring: 526, 73.16, 9.08, 43, 67, 73, 80, 93

close\_shot: 526, 81.05, 10.15, 53, 73, 81, 90, 99

mid\_range\_shot: 526, 74.51, 10.72, 25, 69, 74, 80, 97

three\_point\_shot: 526, 74.50, 12.73, 25, 72, 77, 82, 99

free\_throw: 526, 75.84, 10.14, 36, 70, 77, 83, 94

shot\_iq: 526, 66.28, 14.05, 25, 55, 70, 75, 98

offensive\_consistency: 526, 66.32, 19.98, 25, 50, 70, 80, 98

group\_athleticism: 526, 76.27, 4.79, 59, 73, 76, 79, 92

speed: 526, 73.38, 11.71, 25, 68, 75, 82, 98

agility: 526, 74.55, 11.44, 30, 68, 76, 83, 96

strength: 526, 60.27, 13.18, 30, 52, 60, 69, 96

vertical: 526, 75.60, 9.17, 40, 70, 75, 82, 99

stamina: 526, 86.77, 5.30, 70, 83.25, 87, 90, 99

hustle: 526, 82.19, 9.46, 35, 75, 85, 90, 99

overall\_durability: 526, 81.03, 3.26, 60, 80, 80, 82, 98

group\_inside\_scoring: 526, 62.45, 9.65, 39, 56, 62, 68, 93

layup: 526, 77.09, 8.41, 47, 71, 77, 83, 98

standing\_dunk: 526, 50.86, 20.79, 25, 35, 45, 70, 95

driving\_dunk: 526, 69.96, 13.90, 25, 65, 75, 80, 97

post\_hook: 526, 50.75, 16.47, 25, 38, 49, 62, 97

post\_fade: 526, 54.17, 13.53, 25, 44, 53, 62, 96

post\_control: 526, 54.02, 16.28, 25, 42, 52, 65, 98

draw\_foul: 526, 58.52, 16.82, 25, 45, 55, 70, 98

hands: 526, 83.56, 8.58, 37, 79, 84, 90, 98

group\_playmaking: 526, 63.96, 12.64, 32, 56, 64, 74, 96

pass\_accuracy: 526, 65.83, 13.91, 25, 59, 68, 75, 98

ball\_handle: 526, 69.58, 14.74, 30, 62, 71, 79, 99

speed\_with\_ball: 526, 65.93, 15.60, 25, 57, 69, 77, 97

pass\_iq: 526, 68.11, 13.02, 30, 60, 68, 78, 98

pass\_vision: 526, 50.28, 16.97, 26, 35, 45, 67, 98

group\_defense: 526, 60.27, 9.53, 39, 54, 59, 66, 87

interior\_defense: 526, 57.30, 15.43, 25, 46, 58, 69, 95

perimeter\_defense: 526, 65.40, 13.38, 25, 57, 66, 75, 95

steal: 526, 51.61, 16.09, 25, 40, 46, 60, 98

block: 526, 56.03, 16.18, 25, 45, 55, 67, 99

help\_defense\_iq: 526, 68.74, 10.94, 45, 60, 68, 78, 95

pass\_perception: 526, 63.82, 12.56, 35, 54, 62, 74, 97

defensive\_consistency: 526, 59.01, 16.36, 30, 45, 55, 73.75, 95

group\_rebounding: 526, 56.24, 14.25, 31, 45, 54, 65.75, 94

offensive\_rebound: 526, 50.83, 17.07, 25, 38, 46, 61.75, 97

defensive\_rebound: 526, 61.13, 14.31, 33, 49.25, 60, 70.75, 97

intangibles: 526, 59.20, 17.64, 25, 50, 60, 73.75, 98

potential: 526, 80.20, 5.73, 70, 76.25, 79, 83, 99

total\_attributes: 526, 2464.21, 193.89, 2030, 2329, 2438, 2558.75, 3206