**1. Data**

The NBA player stats dataset contains season stats for every NBA player from 1950-2017, including games played, games started, and advanced stats such as offensive and defensive box plus minus (OBPM and DBPM) and offensive and defensive win shares (OWS and DWS), which serve as catch-all statistics to quantify overall offensive and defensive performance respectively (henceforth collectively referred to as collective statistics). The games played and games started data can be used to define “starter” (started more than 50% of games played in the season\*), “reserve” (played in at least 35% of the team’s games that season\*\*) and “bench” (did not qualify as a “starter” or “reserve”) designations. The statistical profiles (preferring offense or defense) can then be calculated between starters and reserves as an estimate of the prioritization of offense or defense (“bench” designation is omitted because the small sample size for the individual players often leads to greatly skewed results). The player information data includes name, team name, and position for each year played, allowing analysis of differences in this preference of offense or defense between teams, between position groups, and trends over the years. Given the advanced stats desired for this type of analysis, the data analyzed needs to be limited to years since 1982 because earlier years did not record the statistics necessary to calculate OBPM, DBPM, OWS, and DWS, or did not track games started.

The NBA season records from every year dataset contains team records for each team for each year, as well as an indication of whether that team qualified for the playoffs. By joining this dataset to the NBA player stats dataset, further analysis can be performed looking at differences in OBPM, DBPM, OWS, and DWS for playoff-qualifying and non-playoff-qualifying teams. This allows us to quantify whether the preference of offense or defense statistically improves a team’s chances of qualifying for the playoffs.

\*cutoff for “starter” designation according to qualifications for 6th Man of the Year Award qualifications: to qualify for the 6th Man of the Year award, a player must start fewer than 50% of his team’s games that season

\*\*cutoff for “reserve” designation according to qualifications for NBA RookieLeaders: to qualify for NBA League Leader, a player must play in 70% of his team’s games that season (58 games in an 82 game season); minimums for NBA Rookie Leaders is half of the overall requirement

<https://www.kaggle.com/drgilermo/nba-players-stats>

<https://www.kaggle.com/boonpalipatana/nba-season-records-from-every-year>

**2. Research Design**

The research being conducted is trying to answer the question of whether offense or defense is more valued and/or more valuable to NBA teams during the regular season. OBPM and OWS are used to quantify a player’s offensive performance, and DBPM and DWS are used to quantify a player’s defensive performance. Each player's OBPM, OWS, DBPM, and DWS are normalized by subtracting the team average for each respective statistical category for that season (e.g. Stephen Curry’s 2017 OBPM will be normalized by subtracting the average OBPM of all players on the Golden State Warriors in the 2017 season). This normalization is used to test team preference versus availability (e.g. starting a player with a 0 OBPM and a 0 DBPM can show that the team values offense if the average OBPM is -2 and the average DBPM is +2, since they are choosing to play someone with a higher OBPM even though players with a higher DBPM are available, or it can show the opposite if the average OBPM is +2 and the average DBPM is -2). Team averages are only calculated across starters and reserves; bench players are excluded from all calculations.

The first null hypothesis to be tested is that there is no statistically significant difference between the collective statistics of starters and the collective statistics of reserves. The alternative hypothesis is that there is a statistically significant difference (statistical significance defined as ɑ=0.05). This will be tested by analyzing the normalized collective statistics of players designated “starters” and players designated “reserves” using the appropriate statistical test given the distribution.

The second null hypothesis to be tested is that there is no statistically significant difference between the difference in collective statistics of the starters and reserves for playoff teams and the difference in collective statistics of the starters and reserves for non-playoff teams. The alternative hypothesis is that there is a statistically significant difference (statistical significance defined as ɑ=0.05). This will be measured by calculating the difference in the averages of the collective statistics for starters and reserves for each team in each year (meaning the individual player statistics will be condensed down to a single measure for each team each year). Then the appropriate statistical test given the distribution will be used to measure the difference in these differences between playoff-qualifying and non-playoff qualifying teams.

The third null hypothesis to be tested is that there is no statistically significant difference between the difference in collective statistics of the starters and reserves for each team in 1982-1987 and the difference in collective statistics of the starters and reserves for each team in 2012-2017. The alternative hypothesis is that there is a statistically significant difference (statistical significance defined as ɑ=0.05). This will be measured by calculating the difference in the averages of the collective statistics for starters and reserves for each team in each year (meaning the individual player statistics will be condensed down to a single measure for each team each year). Then the appropriate statistical test given the distribution and the dependent nature of the data to measure the difference in these differences between 1982-1987 and 2012-2017. Multiple years are combined to increase sample size.

The fourth null hypothesis to be tested is that there is no statistically significant difference between the collective statistics for each position group (point guard, shooting guard, small forward, power forward, center). The alternative hypothesis is that there is a statistically significant difference (statistical significance defined as ɑ=0.05). This will be tested by using the appropriate statistical test for multivariate analysis to measure the difference between the normalized collective statistics between the five position groups.

**3. Audience**

This analysis is geared towards NBA coaches and GMs to identify what kinds of players to play during games and target in free agency, trades, and the draft. If playoff-qualifying teams showed a statistically significant difference in prioritization of offensive or defensive players over non-playoff-qualifying teams, then those are the types of players who should be targeted and utilized most. Additionally, if there are statistically significant differences between position groups, the list of players to target can be further narrowed. To most effectively use this data for such position-specific purposes would require further analysis of offensive vs defensive ability affecting playoff qualification on a per-position basis.