I started out this project by thinking about what steps I would need to take. My initial thoughts on what steps to take were to first determine what data I would need, collect that data, use machine learning techniques to cluster on that data, and then lastly use those clusters to figure out what the best lineups are. The columns that I used for my data were:

USG% - Usage Percentage, a measure for how often they are used

TOV% - Turnover Percentage, a measure for how often they turn the ball over

STL% - Steal Percentage, a measure for how efficiently they force turnovers

AST% - Assist Percentage, a measure for how efficiently they are at creating assists

ORB% - Offensive Rebounding Percentage, a measure for how efficiently they get offensive rebounds

DRB% - Defensive Rebounding Percentage, a measure for how efficiently they get defensive rebounds

PFP36 – Personal Fouls per 36 Minutes, a measure for how often they foul on the floor

FTR – Free Throw Rate, a measure for how often they get to the line per field goal attempt

FT% - Free Throw Percentage, a measure for how well they shoot from the line

RIM FGR – Rim Field Goal Rate, a measure for how many of the player’s shots come between 0 and 3 feet

RIM FG% - Rim Field Goal Percentage, a measure for how efficiently the player shoots between 0 and 3 feet

MR FGR – Mid-Range Field Goal Rate, a measure for how many of the player’s shots come between 3 and 23 feet

MR FG% - Mid-Range Field Goal Percentage, a measure for how efficiently the player shoots between 3 and 23 feet

3PAR – 3-Point Attempt Rate, a measure of how many of the player’s shots come from 3-point range

3P% - 3-Point Percentage, a measure of how efficiently the player shoots from 3-point range

RIM DFGAP36 – Rim Defensive Field Goals per 36 Minutes, a measure for how often a player contests shots between 0 and 6 feet

RIM DFG% - Rim Defensive Field Goal Percentage, a measure for how opposing players shoot at the rim when this player contests

I compiled all the lineup data and all the player data except for the last two columns by scraping off basketball-reference using Python’s Beautiful Soup package, only considering players who have played more than 100 minutes this season. I obtained the defensive field goal statistics by simply copying and pasting off the defensive player tracking portion of the NBA’s website, as the NBA’s website is difficult to scrape data from since the front end of the website appears to be designed using Angular JS. I joined the data together using Python’s Pandas data frame, and used Python’s Scikit-learn package, a machine learning package, to perform clustering on that data. I used Scikit-learn to preprocess and scale the data, and then reduce the number of dimensions from 17 to 6, which I found to be the ideal number of components to use for my data. I then used Scikit-learn to create 9 clusters, where 9 was the ideal number of clusters to use for my data. I found the clusters to be:

0 – Versatile Swingmen (77 players: DeMarre Carroll, Kent Bazemore, Danny Green)

1 – Rebounders (40 Players: Andre Drummond, DeAndre Jordan, Tristan Thompson)

2 – Stretch Big Men (43 Players: Kevin Love, Kristaps Porzingis, Channing Frye)

3 – Penetrating Passers (47 Players: Ricky Rubio, Rajon Rondo, Manu Ginobili)

4 – Benchwarmers (19 Players: Brian Roberts, Cameron Payne, Paul Pierce)

5 – Playmaking Stars (32 Players: LeBron James, Kawhi Leonard, Stephen Curry)

6 – Scoring Wings (83 Players: Carmelo Anthony, Devin Booker, Klay Thompson)

7 – Wing Defenders (23 Players: Trevor Ariza, Andre Iguodala, Andre Roberson)

8 – Inside Scorers (38 Players: Zach Randolph, Nikola Vucevic, Shaun Livingston)

After filtering by POSS > 257 (Lineups with > 257 POSS make up about 76% of total POSS), I obtained the top combinations of players below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Combo | POSS | DPOSS | PTS | DPTS | ORTG | DRTG | Net RTG |
| 0-0-5-5-8 | 314 | 321 | 422 | 341 | 134.39 | 106.23 | 28.16 |
| 0-1-5-5-6 | 2206 | 2201 | 2679 | 2253 | 121.44 | 102.36 | 19.08 |
| 1-5-5-6-7 | 1541 | 1546 | 1824 | 1590 | 118.36 | 102.85 | 15.51 |
| 0-0-2-5-5 | 558 | 546 | 662 | 565 | 118.64 | 103.48 | 15.16 |
| 0-1-5-5-8 | 350 | 362 | 404 | 366 | 115.43 | 101.1 | 14.33 |

0 0 5 5 8

2 Versatile Swingmen, 2 Playmaking Stars, 1 Inside Scorer

Lineups:

Lowry (5) – DeRozan (5) – Carroll (0) – Patterson (0) – Valanciunas (8)

Grant (0) – Wade (5) – Butler (5) – Mirotic (0) – Lopez (8)

Grant (0) – Wade (5) – Butler (5) – Mirotic (0) – Gibson (8)

Grant (0) – Wade (5) – Butler (5) – Valentine (0) – Gibson (8)

Notes: Toronto lineup played almost all the possessions, this group is ranked so highly because the former Toronto starting lineup was very effective

0 1 5 5 6

1 Versatile Swingmen, 1 Rebounder, 2 Playmaking Stars, 1 Scoring Wing

Lineups:

Steph (5) – Klay/Clark(6) – Durant (5) – Draymond (0) – Zaza/JaVale/Looney (1)

Lowry (5) – DeRozan (5) – Powell/Ross (6)– Carroll (0) – Poeltl/Nogueira (1)

Irving (5) – Shumpert/Korver (0) – LBJ (5) – Derrick Williams (6) – Thompson (1)

Wade (5) – Butler (5) – McDermott(6) – Valentine/Mirotic (0) – Felicio (1)

Notes: Golden State lineups make up about half of the possessions for this type of lineup, and account for most of the effectiveness of this type of lineup, but other lineups of this type are still fairly effective

1 5 5 6 7

1 Rebounder, 2 Playmaking Stars, 1 Scoring Wing, 1 Wing Defender

Lineups:

Steph (5) – Klay/Clark (6) – McCaw/Iggy (7) – Durant (5) – Zaza/JaVale/Looney (1)

Paul (5) – Jamal/Redick(6) – Luc Richard/Alan Anderson (7) – Griffin (5) – DeAndre (1)

Schroder (5) – Hardaway JR (6) – Sefolosha/Bembry (7) – Millsap (5) – Dwight (1)

Notes: Bulk of the possessions in this lineup come from Clipper’s starting lineup which is a +15 lineup.