Instructions for BUS graduate class Term Paper Manuscript

Jongwook Woo

Department of Information Systems, California State University

Los Angeles

Tel. 323-343-2916, Fax. 323-343--5209

e-mail : jwoo5@caltstatela.edu

Group 1

#### Loc Bryan Truong,  Anthony Cardoza, Jose Eichenseer, Tro Haboian, John McCarty, Vinh Nguyen

**Abstract:** The purpose of this paper is to analyze the dataset from the National Basketball Association (NBA). The dataset contains data from the past five regular seasons statistics of points, rebounds, assists, three-point percentage, and overall highest ranked players in the NBA. The story consists of a calculated statistical analysis of the data for the top 25 players within a five-year period. Using the cloud conception tool SAP Analytics Cloud (SAC) we will process the data into charts, so they are ready to be compared and forecasted. Once created, we will run a forecast on the current regular season, to see if we can make predictions of the current regular season top 25 players. The research is meant to test and create forecasts to see how accurate our predictions can be, given the data we have in hand.

**1. Introduction**

In this case, we will be using SAP Analytics Cloud to create datasets, dynamic stats, and forecasted charts on NBA top 25 ranked players in the past 5 years. The data contains stats of the top 25 ranked players for the past five years, including but not limited to field goal percentage, 3 point percentage, rebounds, assists, points per game, and average points per regular season. We will use the data sets from 2017 through 2022 regular season statistics we have on the top 25 ranked players. We will import the data sets to SAP Cloud Analytics so we can create visualization and clear understanding of each data set. We will use what we have learned from the labs throughout the semester to create dataset models and charts with predictions so we can compare to the current regular season stats and see how accurate we can be based on the information we have in hand. Once we have imported all the data set we will create a chart with predictions and make comparisons on how accurate we can get to the current season standings. Keep in mind there are plenty of other variable information we can add to get more accurate predictions such as player injuries, sickness, rest time for games, etc. These will all change the prediction outcomes as we are adding more datasets to our charts and calculations. Our research is meant to test how accurately we can make predictions and see how athletes will perform.

**2. Related Work**

The market for NBA stat predictions is heavily saturated, with each entry having a different dataset being interpreted by the predictions that are created. We will explain our predictions and compare them to what is out there on the market, while having a goal of matching Vegas odds[[1]](#footnote-0) as closely as we can. Vegas odds is the biggest dataset we can find on making predictions.

[1]

Create data sets based on NBA.com stats as these are the most pure stats we can import from the web, with no special conditions. This gets updated hourly by The National Basketball association.

[2]

Once we have created our own datasets and forecasts based on the stats we want to use, we will compare the forecasts to other related sources we found on the web, such as Sportsline.com and Fantasydata.com. We will see how closely we can get our predictions to these two popular websites.

[3]

Finally, once we have completed the steps above, our goal will be to make our dataset comparable to Vegas odds. Vegas odds are the most popular odds to make predictions on player stats as they allow you to gamble on the odds they provide, meaning they would want to get the most accurate number they can to not lose income on games.

**3. General Instructions**

The data was gathered from NBA.com, which is controlled by the National Basketball Association, making this data source the most accurate information available. The data is updated consistently even with live updates on players and team stats. These stats go all the way back to when the NBA began. We will use the data of the top 25 ranked players over the last 5 years of play. There were 9 concurrent players within the last 5 years, making them our final dataset. The total size of the data will be a little over 70KB of data in Excel format.

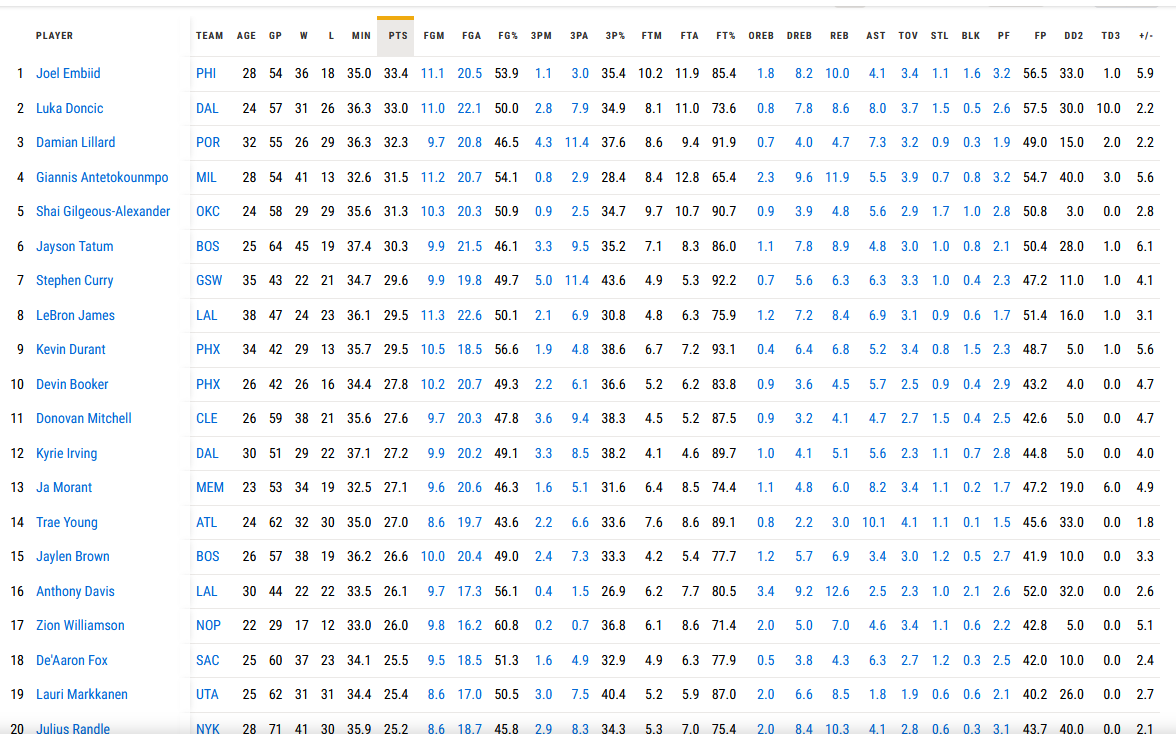
Table 1 - Data Size

| Data Sets | Size (Total 70KB) |
| --- | --- |
| 2017 - 18 Regular Season | 6KB |
| 2018 - 19 Regular Season | 6KB |
| 2019 - 20 Regular Season | 6KB |
| 2020 - 21 Regular Season | 6KB |
| 2021 - 22 Regular Season | 6KB |
| Players in the top 25 for the past 5 years | 11KB |
| 5 year stats of top 25 players | 30KB |

**4. Implementing Data**

The raw data set was imported from NBA.com and we can use this data set in any way we choose. On NBA.com there are over 20 separate statistical groups we can choose from. For our data we decided to implement all the data groups and use what we need. The reason we did this was to allow us to implement more data without having to create more models and importing more excel datasets. The below table shows our imported data set.

Table 2 - NBA.com Stats



**5. Processing Data**

Once we had created our data sheets there was some clean up we needed to do in SAP Cloud Analytics. First, we used our title as the header of the data. Then, we need to swap some of the measures to measurements and others to generic such as the name of the player; this will allow us to easily pick our measures when we are creating the chart and making forecasts. This step needed to be repeated for each data set we added. In total,we have five spreadsheets, one for each regular season stats of the top 25 players, one sheet for the top 25 players within the past 5 years, and one sheet of concurrent players in the top 25 ranked for the past 5 years concurrently. In the end we are left with 9 players that were consistently in the top 25 within the last 5 years. We will make our predictions on these players for the current season statistics. It is important to note that our data sets do not take injuries into consideration. Once we had all the datasets imported and cleaned up it was time to make our dynamic charts to make visualization clear and understandable.

**6. Analysis and Visualization**

Once our data sets were imported and set up correctly to make visualization understandable, we began to create our story. Our goal was to make the information easy and clear to allow any user to interpret the information with no difficulty. We have created a dynamic table for each separate year, to allow users to go back to a certain year in time and see the performance of a specific player. We then combined all five years data and filtered out non concurrent players. With the new data set and concurrent players we created a time series chart to make predictions based on the information we had imported.

**6.1 Points, Rebounds, Assists, 3 point %**

For each yearwe had imported, we wanted to make sure that we can easily see average points, rebounds, assists, and 3-point percentage on any given player. We created a chart for each regular season year, starting from the 2017 regular season and ending on the 2022 regular season. Our dynamic charts will update each player's stats depending on who is selected. For example, when we select Anthony Davis stats for the 2017 season we see his Points, Rebounds, Assists, and 3 Point percentage for that year. If we wanted to look at Damian Lillards stats for the 2021 regular season we would select the page with 2021 - 2022 Top 25 Stats and select Damian Lillard in the table below, and we will see his stats for that given year.

Table 3 - SAP

2017 - 2018 - Anthony Davis

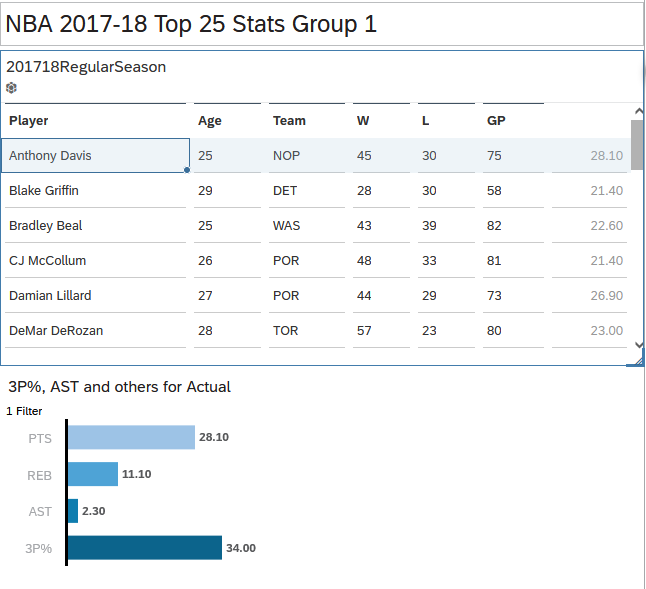
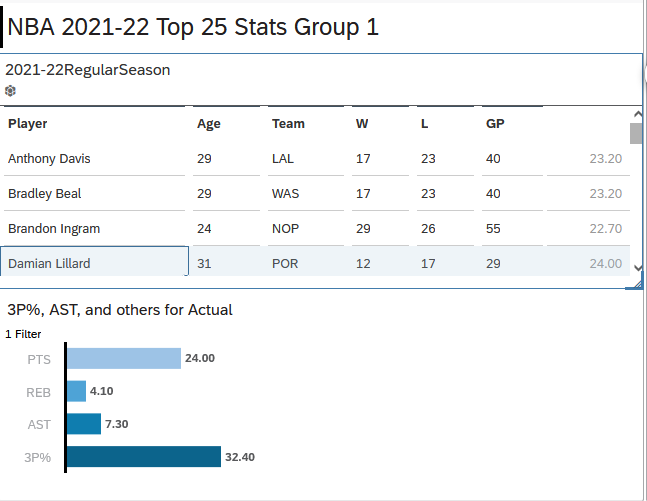


Table 3.1 SAP

2021 - 2022 - Damian Lillard



**6.2 Top 25 Ranked Players Consistently Forecast**

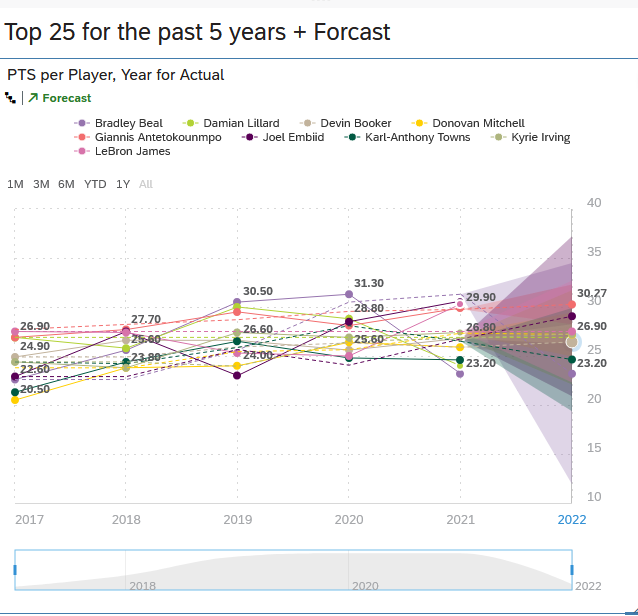
Here we will import our data into a time series chart to allow us to make forecasts on our current data. When we create a forecast we can compare them to the current regular season stats, since these were not included in our statistics. For example, when we made our predictions, the forecast told us that Bradly Beal should be averaging 23.20 points, with an upper confidence bound of 34.48 and a lower confidence bound of 11.92. When we look at his stats for the current regular season which is almost over, we see he is averaging 23.10 points. making our prediction extremely accurate. We see an accurate trend with most players with the worst prediction being off by four points average for Joel Embiid. Based on the very minimal statistics we have plugged in, we were very surprised to see how accurate we can get the predictions. This would lead us to adding more data points to see if we can fix some of the miscalculations aggregates of four points.

Forecast Time Series chart

Table 4 - Comparison

| Player | Upper Confidence | Lower Confidence | Average | Current |
| --- | --- | --- | --- | --- |
| Bradley Beal | 34.48 | 11.92 | 23.20 | 23.10 |
| Damian Lillard | 31.70 | 22.10 | 26.90 | 32.23 |
| Devin Booker | 28.05 | 24.82 | 26.44 | 27.00 |
| Donovan Mitchell | 29.44 | 24.96 | 27.20 | 27.5 |
| Giannis Antetokounmpo | 32.29 | 28.25 | 30.27 | 31.30 |
| Joel Embiid | 37.22 | 20.89 | 29.05 | 33.00 |
| Karl-Anthony Towns | 29.90 | 19.37 | 24.64 | 20.80 |
| Kyrie Irving | 28.25 | 26.24 | 27.24 | 27.00 |
| Lebron James | 32.70 | 22.30 | 27.50 | 29.50 |

Table 4.1 - Forecast



**6.3 Analysis**

As we have mentioned, we see that our forecast is fairly accurate based on the datasets it is taking into consideration. We have created predictions on the top 25 consistently ranked players and we can see that some are falling off for the year, and we see others are outperforming their last year's performance. Based on the information we can start to add more data sets and more niche statistics such as blow outs, min-rests, and injuries. We can see if our predictions will get more accurate based on the information, or if the data sets will make it skew more to one direction.

**7. Conclusion**

In conclusion, we can start making predictions based on our datasets above and implement general game knowledge to our forecast. Take Bradley Beal for example. Our forecast and the actual current season stats are extremely close - we can make a prediction that he will not be in the top 25 ranked players this year. We can also tell who is outperforming their personal stats from last year: Joel Embiid and Giannis Antetokounmpo are leaning more towards their upper confidence bound of the forecast thus making them outperform their stats from last year and securing their spot in the top 25 ranked players[[2]](#footnote-1). I think a future goal of our group would be to implement this on a game to game level basis, giving up stats on certain games rather than overall averages. Implementing this on a larger scale of players will also help us make more predictions, such as if a certain player outperforms versus another certain player or vice versa, if a player plays worse when up against a certain opponent. These variables will only help our predictions as we are offering more information and taking these datasets into consideration as well.

### References

[1] National Basketball Association. “Players Traditional: Stats.” *Players Traditional | Stats | NBA.com*, <https://www.nba.com/stats/players/traditional?Season=2021-22>

[2] Sports Line. “NBA Player Projections for Single Game and DFS.” *SportsLine*, <https://www.sportsline.com/nba/expert-projections/simulation/>

[3] Fantasy Data. “NBA Projections 2023.” *FantasyData*, <https://fantasydata.com/nba/fantasy-basketball-projections>

[4] Vegas Odds “NBA Las Vegas Odds (2023): NBA Basketball Scores, Lines, Odds, News.” Edited by Vegas Odds, *Vegas Odds*, 31 Aug. 2022, <https://www.vegasodds.com/nba/>

1. An ultimate [resource](https://www.vegasodds.com/about/) and home for all things odds. [↑](#footnote-ref-0)
2. Resulting in being the 1st or 2nd ranked player in the league. [↑](#footnote-ref-1)