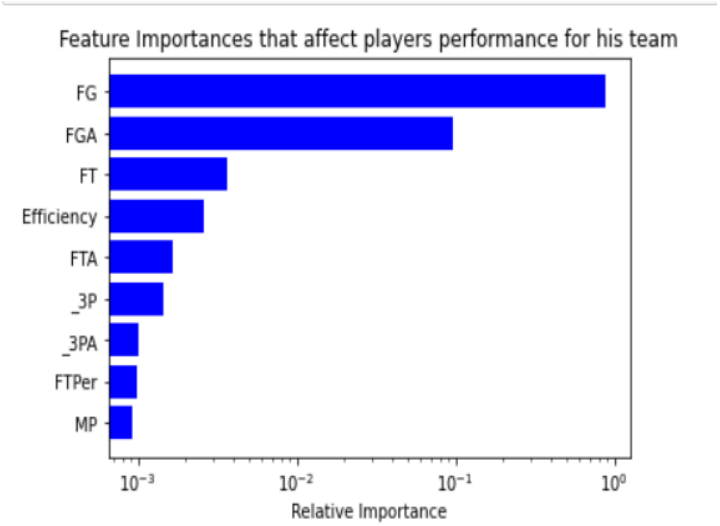
# Deliverable 3

1. Visualization tool chosen and why:
    a. For our visualization tool, we decided to build a data dashboard through Data Studio. The reason for a dashboard is because we felt it would be the most beneficial to our end user. A team manager, owner, coach, and even fan can look at a data dashboard and easily find the answers they need. This tool allows an end user to have everything they need right in front of them at once. It takes the data they may not understand and turns it into actionable insights they can understand in an easier, more efficient way.
2. Visualizations/charts considered initially and why:
    a. Some visualizations and charts we considered initially include tables, pie charts, bar charts, and scatter plots. We planned to put all of these charts and visualizations in a dashboard with a data studio to produce an easy way for our end user to visualize the answers to our proposed questions. The tables would be used to indicate player rank and team rank. Pie charts are useful in our efforts to visualize minutes played per player, age distribution of top performing teams, and field goal percentage of the top players. Scatter plots and bar charts are also useful in visualizing this data. For our attempt to predict future NBA stars, we considered a dot plot with a trend line to indicate performance increase over time of specific players.
3. Summary and Conclusions
    a. Previous steps and any changes made since: Throughout the entirety of the project process, we utilized GCP. We imported our dataset into a storage bucket, cleaned and pre-processed data using DataPrep, merged datasets to explore physical attributes of players and how they may impact performance, worked

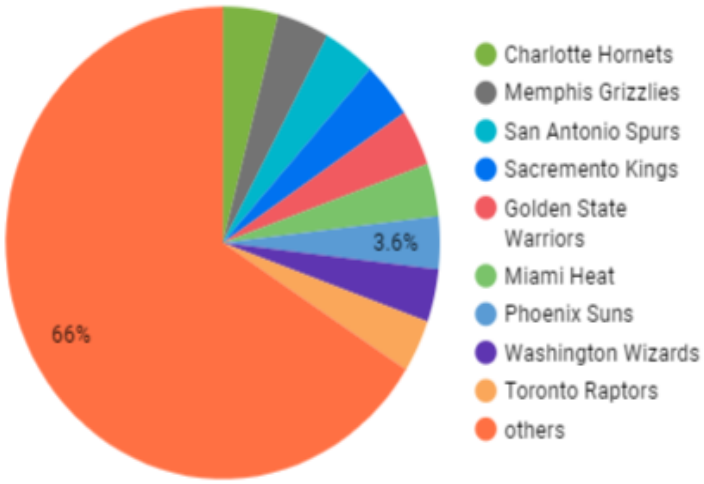
within a python notebook to further cleanse datasets, and ran queries using both BigQuery and tools within Jupyter notebooks. Our project was mainly focused on descriptive analysis, but we did attempt to utilize Pycaret to work with predictive modeling. We wanted to predict the performance of a player at a certain time in the future, a prediction that would be useful to a team owner or manager wanting to know if they should keep a certain player or give a higher salary on the player's new contract.

b. Visualizations and how they aid in answering questions proposed:

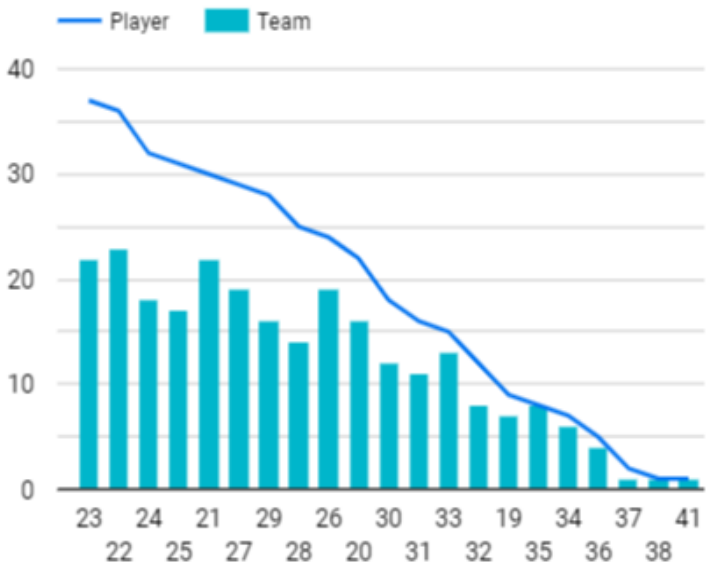
1. What variables affect a player's performance/value to the team?
   The variables that influence a player's performance or value on a team are retrieved using a random regressor model with target values points and rank. The relevance of the traits is then visualized




2. What is the age distribution of the top five performing teams?
   Using a pie chart to filter out the top teams in Data Studio.

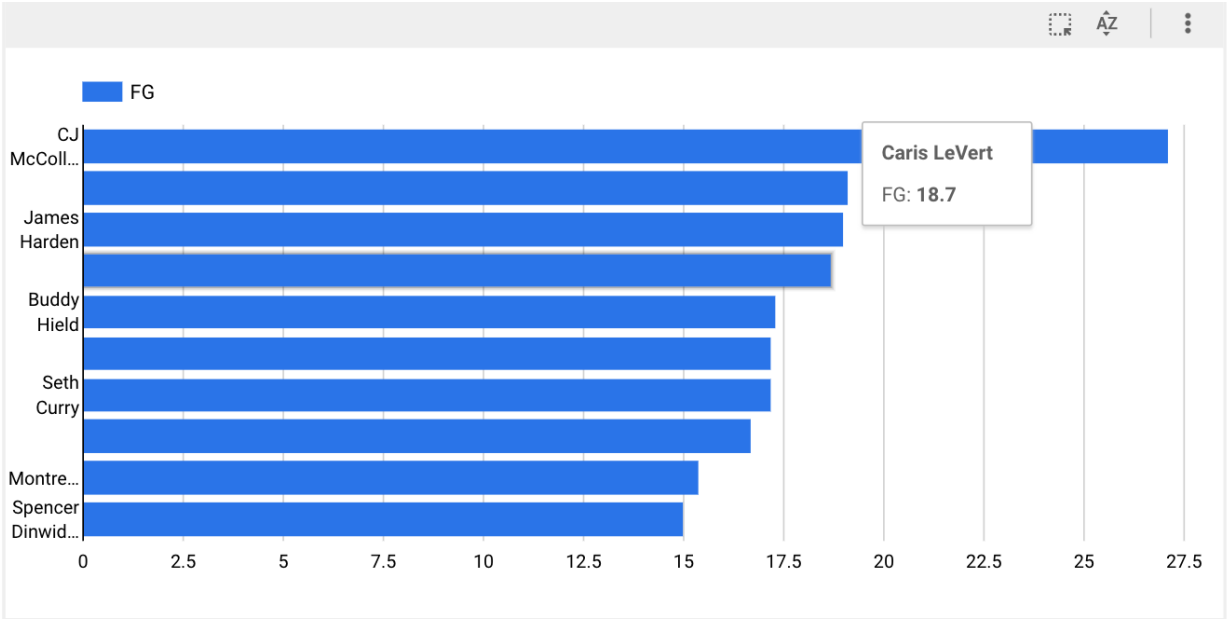



Where players are metrics, we showed a mix of bar and line chart versus Team and age distribution

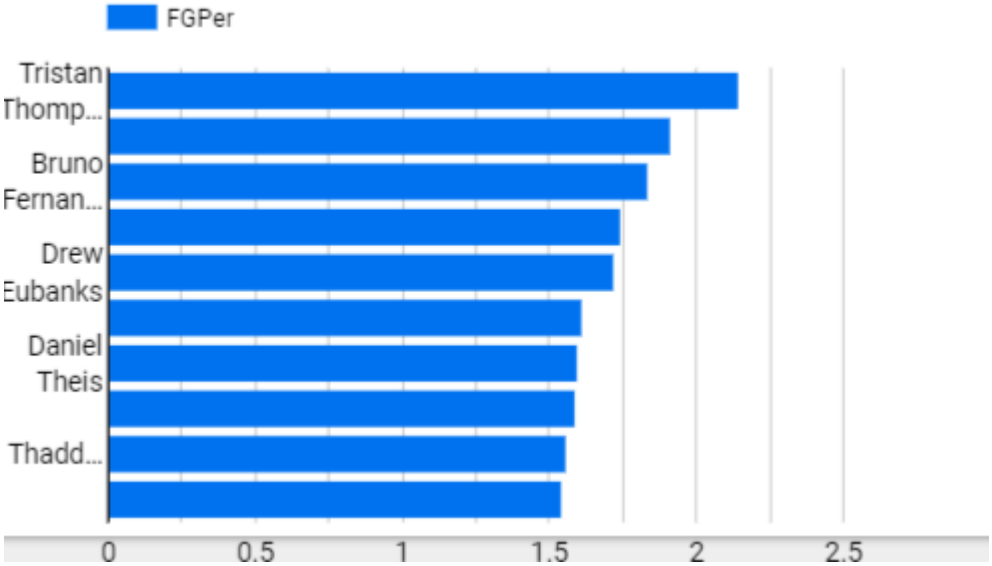



3. Minutes played per game by players?

Here is a bar chart indicating the ten players with the highest average minutes played per game.




4. Field goal percentage of the top ten players?
   In Data Studio, we've formatted a new field called FGPer (field goal percentage). And the FGPer is plotted players in descending order against a horizontal bar graph.




5. Predicting Future NBA Superstars?
   For future NBA Superstars, We've created new fields for every position in Boolean datatype. And formatted them using the performance measures of players like points, age and efficiency using case function. Tables are plotted Top players for every position as shown.

| Position_c / Record Count | | |
|---|---|---|
| Player | false | true |
| Danuel House... | 4 | - |
| Alize Johnson | 4 | - |
| Wenyen Gabriel | 4 | - |
| Tristan Thom... | - | 4 |
| Ish Smith | 3 | - |
| Aaron Holiday | 3 | - |
| Rajon Rondo | 3 | - |
| Jevon Carter | 3 | - |
| Seth Curry | 2 | |

| Position_SG / Record Count | | |
|---|---|---|
| Player | false | true |
| Danuel House... | 4 | - |
| Alize Johnson | 4 | - |
| Wenyen Gabriel | 4 | - |
| Tristan Thom... | 4 | - |
| Ish Smith | 3 | - |
| Aaron Holiday | 3 | - |
| Rajon Rondo | 3 | - |
| Jevon Carter | 3 | - |
| Norman Powell | 2 | |

| Position_PG / Record Count | | |
|---|---|---|
| Player | false | true |
| Danuel House... | 4 | - |
| Alize Johnson | 4 | - |
| Wenyen Gabriel | 4 | - |
| Tristan Thom... | 4 | - |
| Seth Curry | 3 | - |
| Bryn Forbes | 3 | - |
| CJ McCollum | 3 | - |
| Norman Powell | 3 | - |
| Buddy Hield | 2 | |

| Position_PF / Record Count | | |
|---|---|---|
| Player | false | true |
| Danuel House... | 4 | - |
| Tristan Thom... | 4 | - |
| Alize Johnson | - | 4 |
| Wenyen Gabriel | - | 4 |
| Ish Smith | 3 | - |
| Aaron Holiday | 3 | - |
| Rajon Rondo | 3 | - |
| Jevon Carter | 3 | - |
| Seth Curry | 2 | |

| Postion_SF / Record Count | | |
|---|---|---|
| Player | false | true |
| Danuel House... | - | 4 |
| Tristan Thom... | 4 | - |
| Alize Johnson | 4 | - |
| Wenyen Gabriel | 4 | - |
| Ish Smith | 3 | - |
| Aaron Holiday | 3 | - |
| Rajon Rondo | 3 | - |
| Jevon Carter | 3 | - |
| Seth Curry | 2 | |

c. ETL Process: Extraction - We retrieved NBA player stats data from kaggle.com and loaded it into the storage bucket. Transformation - We pre-processed, cleaned and merged the dataset by using dataprep and Jupyter notebooks. Loading - We loaded the formatted data into a storage bucket for running queries using BigQuery and for building dashboards using data studio.

d. Conclusions made to the problem statement: Based on the descriptive analytics we have done throughout this project, our end user can easily gather all information about their players(rookies or veterans) and make informed decisions about their teams' future. With the visualizations and dashboard, decisions can be made on whether a certain player is an asset to the team in question when working on contract negotiations. Our analysis is not only beneficial to team stakeholders, but also fans and those interested in certain teams who would want to look deeper into player statistics and how they impact the teams.

More visualizations from our dashboard: