**NBA Player Production Using Hierarchical Clustering**

**Section 1: Importance of Player Valuation:**

Player valuation in the NBA is a crucial step in determining contract and trade offers for players. Teams often lose players and need to replace a certain level of production with another, so signing players to reasonable contracts while maximizing their contribution to winning is the goal. If a team can replace the production of a player with someone who may be undervalued in the market, it would be a positive decision for an NBA team. While the fit of a player on an NBA team within the team’s system and tactics is also a necessary consideration when signing a player, a possible deficiency in the market can be discovered when looking at a player’s level of production compared to other, similar players.

To demonstrate a method to compare players of similar levels of production, hierarchical agglomerative clustering is used to group players into similar clusters. Hierarchical clustering is a statistical methodology that groups observations into a chosen number of clusters, where observations with similar characteristics are placed into the same cluster. This allows for the grouping of similar observations, and in this case, allows for players with similar levels of box score production to be grouped together.

**Section 2: Methodology used:**

As stated, the methodology used in this report is hierarchical clustering, which uses an iterative process to break down the data set into groups of similar observations. The version of hierarchical clustering used in this report is agglomerative clustering, where each observation starts as its own cluster, and over time begins to join with similar observations to create fewer overall clusters that contain more values. In hierarchical clustering, the process continues until all observations merge into a single cluster, known as the root. However, the merging process is typically stopped before reaching this final cluster, once the desired number of clusters is achieved.

The data set is a collection of all 239 NBA players that averaged more than ten minutes per game in the 2023-2024 regular season, and all data is collected from the website Basketball Reference. The data set contains a set of basic box score statistics that are chosen to represent the players production level during the season. The chosen variables are used for clustering for each player. The variables used to cluster each player include three-point shots attempted per game, three-point percentage, two-point shots attempted per game, two-point percentage, free throws attempted per game, free throw percentage, rebounds per game, assists per game, steals per game, blocks per game, turnovers per game, personal fouls per game, and points per game. The theory behind the chosen variables is that a player’s type can be represented by various statistics of both quantity and quality. Quantity statistics include shot selection (3-point, 2-point, and free throw statistics), rebounds, assists, and defensive metrics. Quality can be assessed through the quantity of points, rebounds, and assists, as well as the efficiency of their shots, measured by shooting percentages.

**Section 3: Results of Clustering:**

After running the cluster analysis with the chosen variables, the total agglomerative clustering coefficient was 98% in the initial results as seen in Figure 1. After looking at the Dendrogram presented in figure 2, the number of clusters was set to 10, which allowed for the final collection of clusters. The final collection of clusters was then used to separate players into the ten clusters, which still indicates coherent where the number of players in each cluster can be seen below:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 30 | 3 | 11 | 21 | 5 | 17 | 41 | 42 | 57 | 12 |

While some of the clusters may be smaller comparably than others, these are players who do have very few substitutes and levels of production. For example, cluster 2 contains Giannis Antetokounmpo, Luka Doncic, and Shai-Gilgeous Alexander, who are three of four players who finished top four in voting last season, and cluster 5 contains a group of five very productive big men in multiple statistical categories including Nikola Jokic, Alperen Sengun, and Anthony Davis. These players are not as easily replaceable, meaning that there are not as many players in their cluster. This demonstrates how even though some clusters are much smaller than others, the division of players into clusters generally appears to be accurate. The players in similar clusters can be compared with one another to determine possible

**Section 4: Conclusions and Future Work:**

The model did an effective job at dividing NBA players into different clusters to look at the production of individual players in comparison with one another. A possible decision in the analysis would be to omit or add new variables in analyzing a player’s value. An example of this could be to put more of an emphasis on defensive contribution in variable choice, as initially only steals and blocks are included as defensive metrics, neither of which are always effective at conveying a defender’s actual skill/ utility within a team defense. Another possible step in future work for this type of modeling would be comparing player contracts to other players within the same cluster. Looking at how a player is currently valued in the market in comparison to players of similar production could help a team evaluate how good individual options are from a monetary perspective, as well as possible substitutes for that player. Players in the same cluster may have wildly different contract arrangements for various reasons, but when compared on an individual level to other players in the cluster there is room for insight that could allow for positive value for an organization.

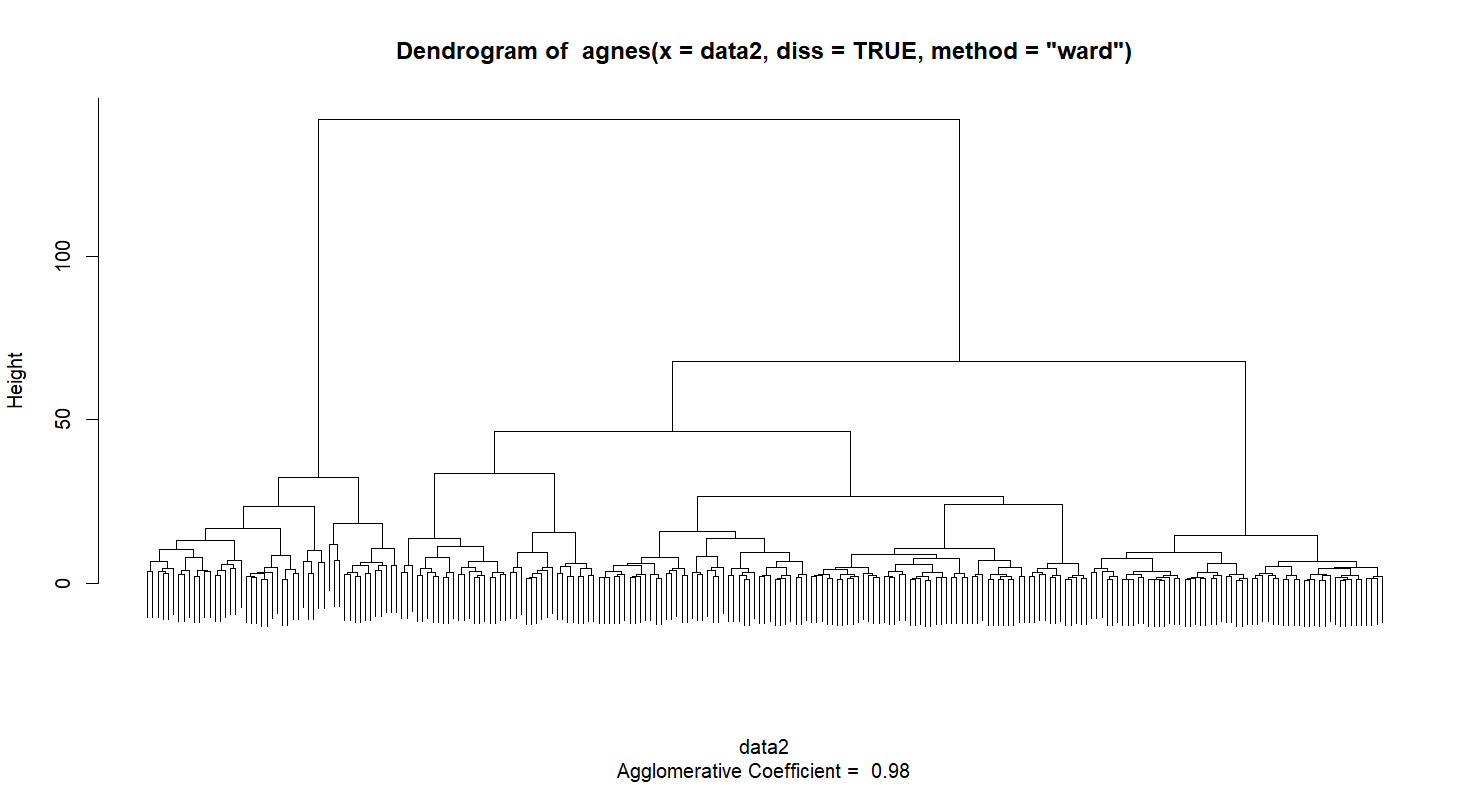
**Appendix**

**Figure 1: Results of Agglomerative Clustering**

**A screenshot of a computer

Description automatically generated**

**Figure 2: Dendrogram of Cluster Analysis**

****

**References**

“2023-24 NBA Player Stats: Per Game.” *Basketball-Reference.com*, 2024, [www.basketball-](http://www.basketball-)

reference.com/leagues/NBA\_2024\_per\_game.html.

“Kia MVP Voting Results.” National Basketball Association, 2024.