Analysis: How is Shooting Distance Impactful on Shooting Accuracy

Sean Fisher

1 Introduction

This project aims to analyze the relationship between shooting distance and shooting accuracy. By examining data from various shooting scenarios, we seek to understand how the distance from the target influences a player's success rate. Through visualization techniques, we will highlight trends and patterns that may emerge as shooting distances change. Ultimately, our goal is to provide clear insights into how distance impacts shooting performance, which can inform strategies and improve player decision-making.

2 Preprocessing

I started with a dataset from Kaggle containing all recorded shots from the 2023 NBA season, capturing every make and miss from every game. To begin the data cleaning process, I created a player shots variable, which provided the total shot count for each player throughout the season. Using this variable, I calculated both the average shot distance and shooting percentage for each player. These metrics were then consolidated into a new DataFrame, forming the foundation for further analysis. Later in the project, I did the same technique with 2022 data. I would then combine both 2022 and 2023 datasets.

3 Data and Methods

3.1 Visualization and Statistical Analysis

After the preprocessing of my data, here is a visualization of the relationship between shooting distance and accuracy. Some players are labeled who ended up shooting over 50 percent while having 1300 or more shots. The bubbles grow or shrink based of the amount of shots that players bubble took.

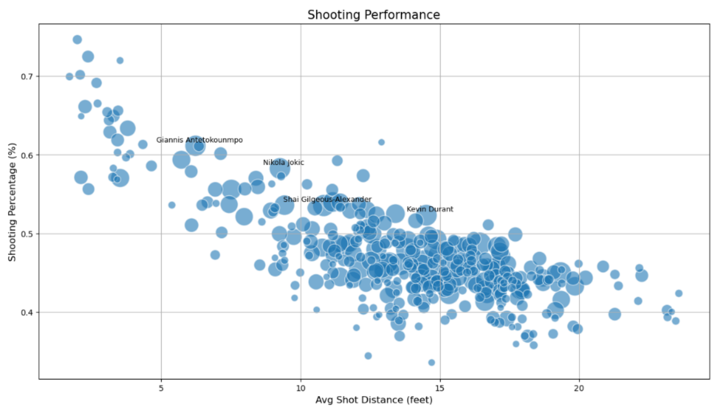

Figure 1: Shooting Performance for 2023 Season

Performing some statistical analysis I found the correlation between shooting distance and accuracy had a -0.77 percent effect. Which indicates a significant negative effect distances plays on accuracy. Here is a linear regression model I performed with the data showing this relationship.

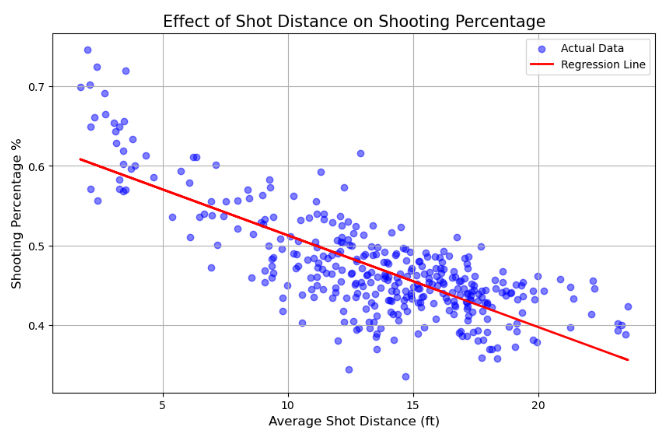

Figure 2: Linear Regression Model

3.2 Clustering

In this part, I aimed to use K-Means clustering to group together similar playing styles for players. With my prediction variables being shooting percentage and average shooting distance. I used Elbow Method in order to find how many clusters is recommended to use in my model, which is shown before.

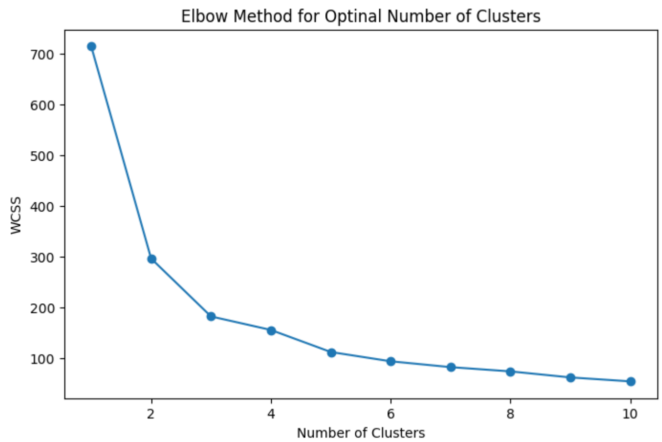

Figure 3: Elbow Method for Clustering

Between 3 and 4 clusters, WCSS has a significant drop off in its decreasing volatility, which indicates to us that we cluster our data into three different groups. Performing the cluster and using the same visuaization technique we used from earlier, we can see how the clustering was performed.

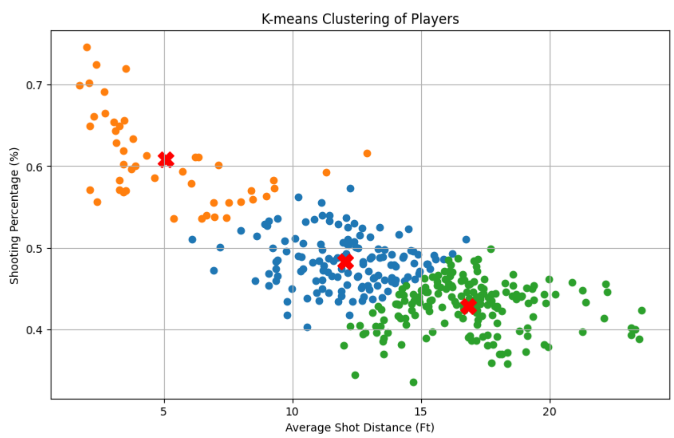

Figure 4: Clustering Data

To evaluate the performance of the clustering model, I used the Silhouette Score, which measures how well-separated the clusters are. The score ranges from -1 to 1, where 1 indicates well-defined clusters, 0 suggests overlapping clusters, and negative values imply that points may have been assigned to the wrong cluster. Our model achieved a Silhouette Score of 0.409, indicating that the clusters are moderately well-defined. Considering that the model uses only two variables, this score reflects a reasonable level of clustering performance.

Below we can see the top 5 performers from each cluster and notice they all play similar positions.

Cluster	Top 5 Players
0	Shai Gilgeous-Alexander, Kevin Durant, Pascal Siakam, LeBron James, Kawhi Leonard
1	Nikola Jokic, Giannis Antetokounmpo, Anthony Davis, Zion Williamson, Domantas Sabonis
2	Luka Doncic, Paul George, Michael Porter Jr., Devin Vassell, Tyrese Haliburton

3.3 2024 Shooting Accuracy Predictions

I took a dataset from 2022 and with 2023 made predictions for the up and coming year. I analyzed the differences between the two years first, to see if there was a change in performance between the two years. With that data I performed another linear regression. Here is a chart of the top five scorers predicted for next year.

Rank	Player Name	Predicted Shooting Percentage
1	Luka Doncic	0.483
2	Jalen Brunson	0.483
3	Anthony Edwards	0.482
4	De'Aaron Fox	0.482
5	Shai Gilgeous-Alexander	0.481

All the other values are stored in a variable I added to the cleaned dataset.

4 Conclusion and Future Work

Shooting accuracy and shooting distance demonstrate a strong correlation, and these two variables alone provide valuable insight into a player's potential performance. However, many other factors can influence a player's success. For example, the type of shot—whether it's a jump shot, layup, or three-pointer—can significantly affect outcomes. Additionally, situational factors, such as a player's performance under pressure in clutch moments, also play a crucial role. My predictions may face challenges, as players can experience rapid improvement during their rookie seasons, or their performance may decline with age. There is still much to explore, and I am excited to dive deeper into these complexities.