|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
|  | | DSA-210 FINAL REPORT | | |  | |
|  |  | | | | |  |
|  | | | Burak Bartu Yılmaz |  | | |
|  | | | 32448How to Defend Lebron James **SUPERVISED BY**  **Özgür Asar** |  | | |

**SABANCI UNIVERSITY**

## ****Introduction****

LeBron James is widely regarded as one of the greatest and most versatile players in NBA history. Many players — and LeBron himself — claim that he is “unguardable.”This project seeks to explore that claim by leveraging real-world data and analytical techniques. By examining detailed shot data from the 2022–2023 NBA season, this study aims to identify patterns in LeBron’s shooting behavior and propose effective, data-driven defensive strategies.Lets watch Lebron James’s claim in this video <https://youtube.com/shorts/ZUIvBXH5VZs?si=1lp92MUwXEllGOEh>

Then let's watch the opinion of Giannis Antetokounmpo who is one of the best active players in the NBA on how LeBron can be guarded <https://youtube.com/shorts/9tMBa7IWHw0?si=U-DOLzH40cFf6HRF>

In this project, we aim to find out whether their claims are valid or not.

## ****Data Collection and Cleaning****

The dataset used includes LeBron James’ 2022–2023 shot attempts and contains features such as:

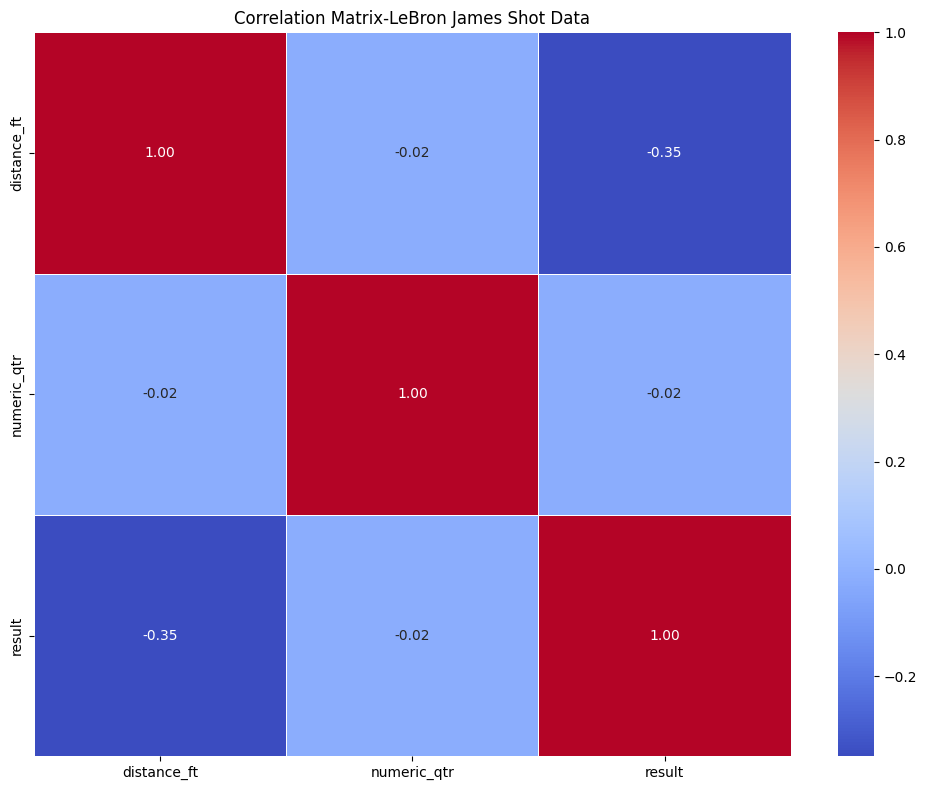
* Shot distance (distance\_ft)
* Horizontal and vertical court positions (left, top)
* Result of the shot (made/missed)
* Game context: quarter, time remaining, etc.

**Cleaning steps:**

* Dropped irrelevant columns (lead, color, opponent, team)
* Removed OT data due to limited volume
* Converted shot result to binary (1 = made, 0 = missed)
* Derived new features:
  + **Zone** (paint, mid-range, 3-point)
  + **Side** (left, right, middle)
  + **Shooting area** (combination of side and zone)

## ****Exploratory Data Analysis (EDA)****

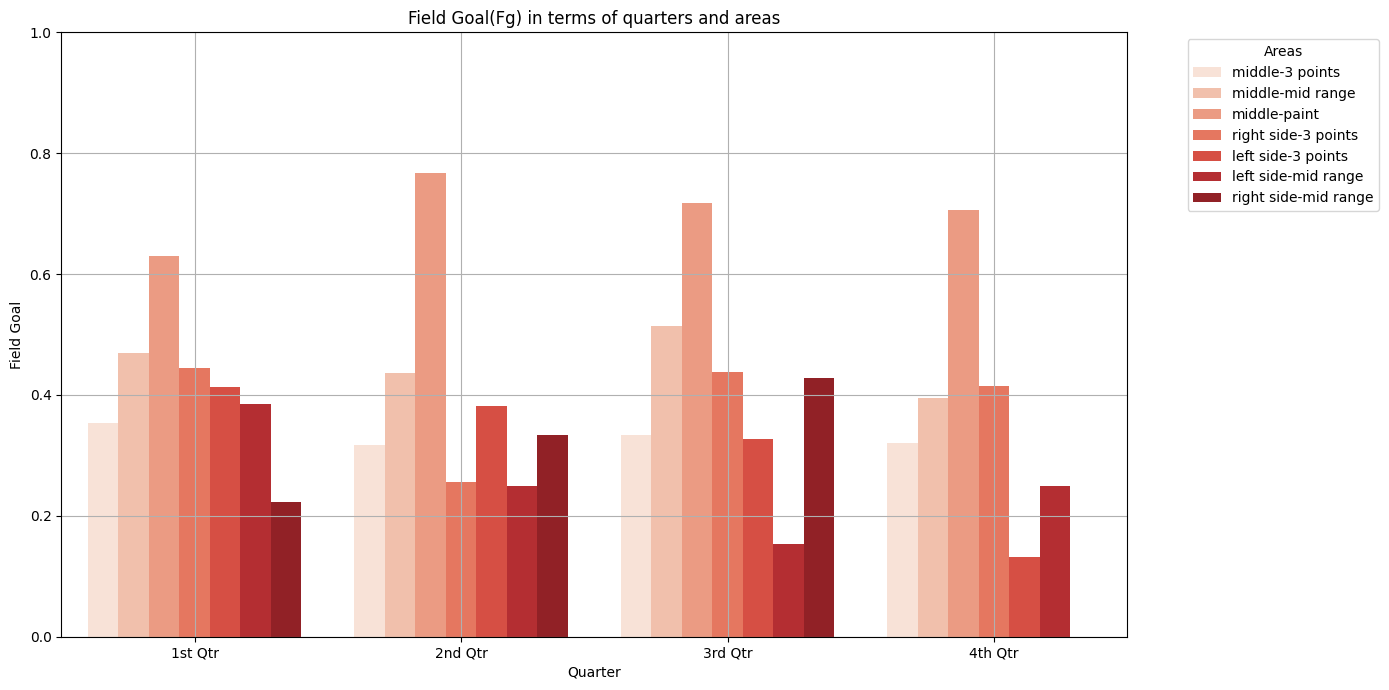
### 1. **Correlation Matrix**



* Weak negative correlation between distance\_ft and result: as distance increases, accuracy slightly decreases.
* Quarter of the game (qtr) has minimal impact on accuracy.

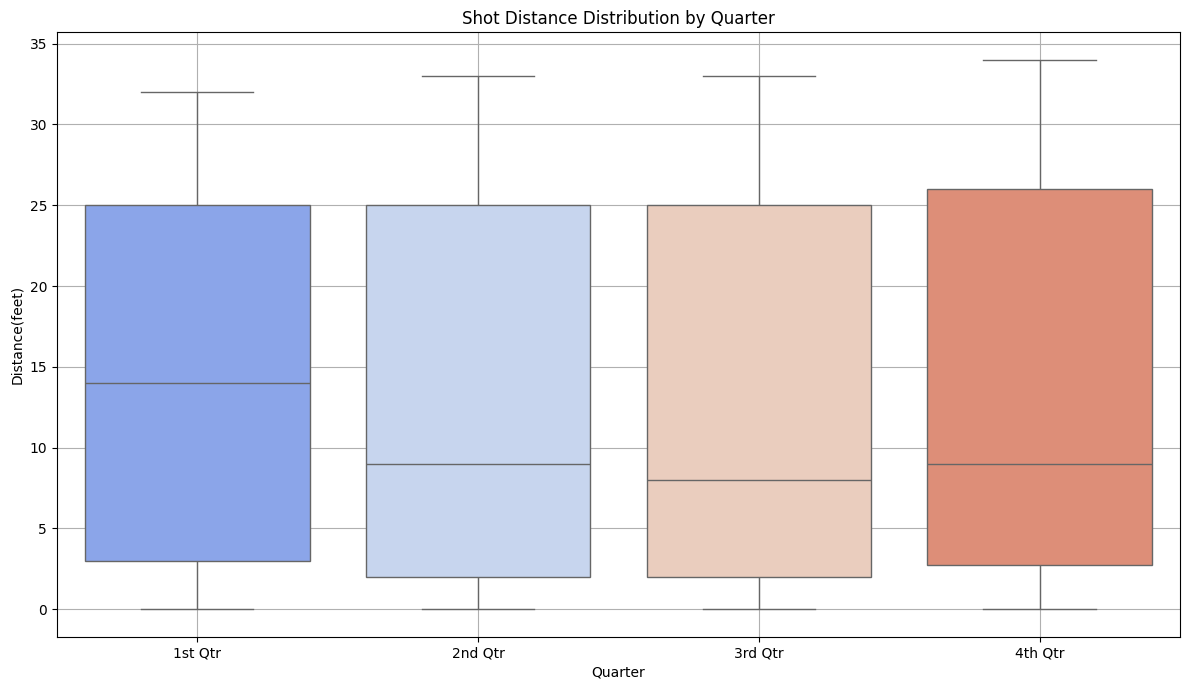
### 2. **Visual Insights**

**Barplot**:

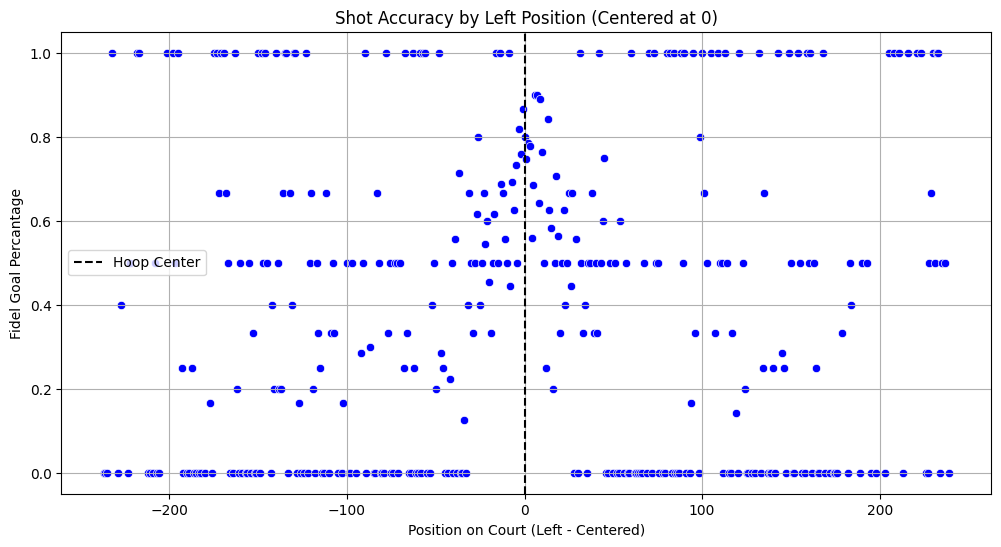


LeBron is most efficient in the paint regardless of quarter. His 3-point efficiency fluctuates by game context.

* **Boxplot**:



* The quarter with the highest shot distance value is the first quarter. This suggests that LeBron attempts long-distance shots early in the game to get into rhythm and improve his three-point shooting.
* Shot distances vary more in 4th quarters, suggesting strategic shot selection under pressure.
* **Scatterplot**:

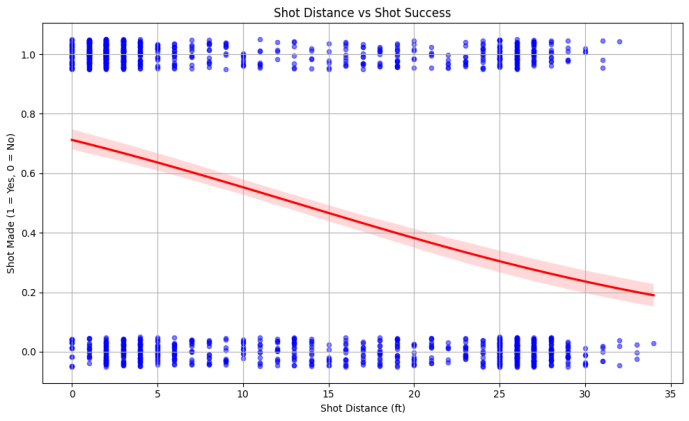
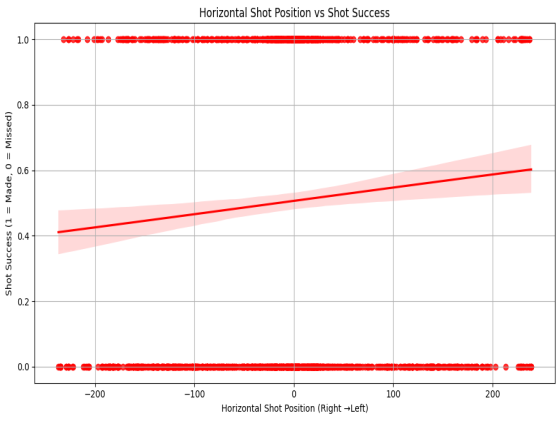


* Shooting accuracy shows variation across court width. A centered left = 0 position marks the hoop line; performance varies left-to-right.

## ****Statistical Modeling****

### Regression Analysis:

**Logistic Regression:**

* **Logistic Regression** indicates that both shot distance and horizontal position have statistically significant (but modest) impact on shot success.For example when shot distance increases,shot success decreases.Also the visualised data shows that LeBron is more successful with shots taken from the left side compared to those taken from the right side.
* **Visual regression plots** show a clear decline in accuracy as distance increases and some drop-off when shots are taken from outer side angles.

## ****Hypothesis Testing****

Using Pearson correlation:

* **Shot distance vs result**: statistically significant negative correlation (longer shots = lower success).Also when we apply hypothesis test in my code output is

Correlation: -0.34887740598914646

P-Value: 3.340776344976172e-44

There is a significant relationship.

Our P-value<0.05.As a result, H0:Distance have effect on LeBron’s shooting efficiency. Ha:Distance does not have effect on Lebron's shooting efficiency. We accept H0 between these two hypothesis

* **Horizontal shot location vs result**: statistically significant but weaker negative correlation (off-center shots = slightly less accurate). Also when we apply hypothesis test in my code output is

Correlation: 0.07422028318817266

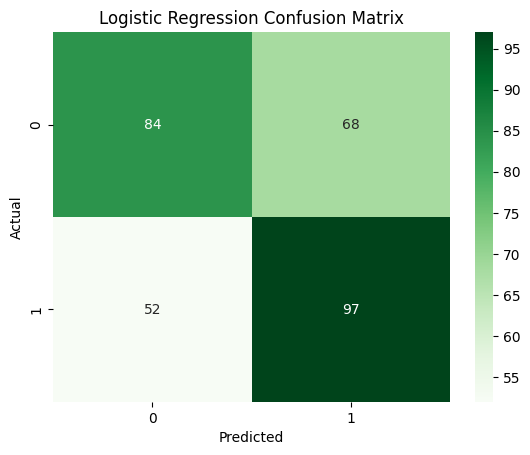
P-Value: 0.0040138596223554986

Significant relationship.

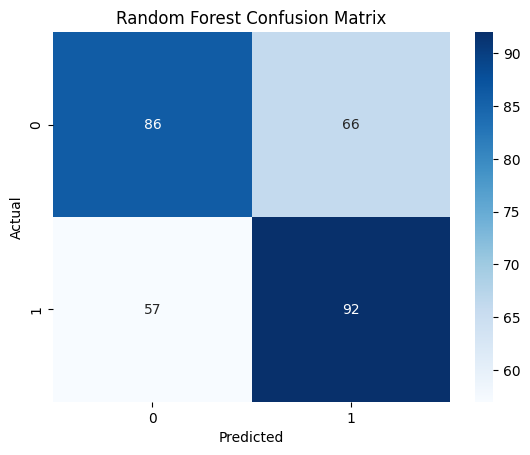
Our P-value<0.05.Therefore H0:Horizontal position have effect on LeBron’s shooting efficiency Ha:Horizontal position does not have effect on LeBron’s shooting efficiency

## ****Machine Learning****

**Logistic Regression:**

  
This model achieved about 69% accuracy. It is simple and easy to interpret, making it straightforward to understand how features influence the prediction.

**Random Forest Classifier:**

  
This model reached around 77% accuracy. It can capture complex and nonlinear interactions between variables such as shot angle, distance, and timing.

The Random Forest outperformed Logistic Regression, indicating that nonlinear relationships play an important role in predicting whether a shot will be made or missed.

## ****İmportant Findings****

* **Most efficient zone**: Paint area
* **Least efficient**: Right-side, long 3-point attempts
* **Left-side shots** showed slightly better performance overall
* Defensive efforts should **force LeBron into low-efficiency zones**, especially:
  + **Long right-side 3s**
  + **Off-balance mid-range shots away from center**

## "How can players guard LeBron?

Based on this analysis:

* Force LeBron **away from the paint** into deeper zones
* **Shade right** on defense, encouraging him to shoot from that side
* Apply more pressure in **late-game situations**, where he tends to take lower-percentage shots
* Consider using defenders who can challenge mid-range shots without fouling

## ****Conclusion****

Although Lebron James is one of the best player in the basketball history,he is **not unguardable**. Lebron James’versitality and effectiveness are inevitable .However this project demonstrates that with the right data, defensive weaknesses and inefficiencies can be identified. A team employing these insights can develop smarter matchups and strategic adjustments to reduce his overall effectiveness.

## F****uture Work****

* Incorporate **player matchups** and **defender distance** for deeper analysis
* Extend the model to **multiple seasons** for consistency checks
* Integrate real-time tracking data for motion-aware defense predictions