**Cairo University – Faculty of Engineering MTHS204 0 Advanced Probability and Statistics Fall 2024**

****

**Team Name: In the Net Submitted to: Prof. Maha Amin Hassanein**

|  |  |
| --- | --- |
| **Ahmed Medhat Abdelaziz** | **4230142** |
| **Abdelrahman Mohamed Abdelfattah** | **1220065** |
| **Abdelrahman Mohamed Kamal** | **1220255** |
| **Fayrouz Ahmed Magdy Moharam** | **1220136** |
| **Lujina Shawky Ahmed Shehata** | **1220138** |
| **Omar Waleed Elsayed Abo-Elnasr** | **1220070** |
| **Mazen Mohamed Sayed Khalifa (Leader)** | **1220018** |

**Table of contents:**

[Abstract 3](#_TOC_250005)

[Introduction 3](#_TOC_250004)

[General Methodology 4](#_TOC_250003)

Statical Questions and numerical Results 6

Question 1 7

Question 2 9

Question 3 11

Question 4 13

Question 5 15

[Conclusion 18](#_TOC_250002)

[Appendix 18](#_TOC_250001)

[References 19](#_TOC_250000)

# Abstract:

Penalties in football are high-stakes events that often decide the outcome of matches, presenting unique tactical and psychological challenges. This research investigates the factors influencing penalty success by analyzing data from six seasons. A multi-disciplinary approach was employed, combining statistical analysis, psychological evaluation, and contextual modeling to explore the interplay of physical attributes, mental resilience, and external influences.

Key findings reveal that the goalkeeper’s positioning and strategy, the penalty taker’s psychological state, and situational factors such as match importance and crowd dynamics significantly impact outcomes. The study highlights the complexity of these interactions and the difficulty in isolating individual influences.

The conclusions emphasize the importance of tailored training regimes for penalty takers and goalkeepers, as well as the role of psychological preparation in high-pressure scenarios. This work provides actionable insights for coaches and analysts, aiming to refine penalty strategies and enhance performance in critical moments.

# Introduction:

Football penalties represent pivotal moments in the game, often determining the outcome of matches and becoming a focal point for tactical and psychological considerations. These high- stakes situations demand exceptional skill and composure, as they test not only the physical capabilities of players but also their mental resilience. Key factors influencing penalty outcomes include the goalkeeper’s reaction time and positioning, the penalty taker’s technique and psychological state, and external elements such as crowd influence and match context.

This study delves deeply into these aspects by analyzing penalty data over six seasons, aiming to uncover the multifaceted factors that contribute to success. By doing so, it provides a nuanced understanding of how physical, psychological, and contextual variables interact to shape outcomes in penalty situations.

## Motivation

Understanding penalties is crucial for enhancing performance and decision-making in football. This research offers valuable insights for improving penalty strategies, selecting the most reliable penalty takers, and refining training methodologies. Moreover, it sheds light on the psychological pressures and environmental dynamics that players face during these critical moments, offering a broader perspective on human behavior under stress**.**

## Objectives

1. Identify the primary factors that influence penalty success.
2. Explore the interplay between physical attributes, psychological preparedness, and contextual influences.
3. Develop actionable recommendations for coaches, players, and analysts to optimize penalty-taking and defending strategies.

## Challenges

Solving the complexities of football penalties involves navigating several challenges:

1. **Interplay of Variables**: The interaction between physical, psychological, and contextual factors is intricate and dynamic, making it difficult to isolate the impact of individual elements.
2. **Limited Data Context**: Penalty outcomes are influenced by unique match situations such as scorelines, player fatigue, and weather conditions, which are often challenging to quantify comprehensively.
3. **Psychological Pressure**: The mental state of players under high-pressure scenarios varies significantly, introducing unpredictable variables that are hard to measure or simulate accurately.
4. **Goalkeeper Influence**: Goalkeepers employ diverse strategies, such as feints or positional adjustments, to unnerve penalty takers, adding another layer of complexity.
5. **External Influences**: Factors such as crowd noise, match location, and referee decisions can significantly affect outcomes but are difficult to control or replicate in studies.

# General Methodology: -

## Dataset Overview:

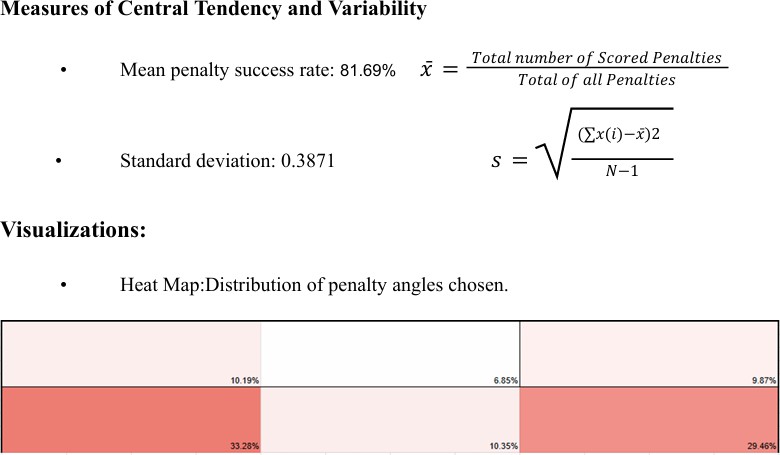
* + **Source:** Six seasons of penalty data collected manually. Each penalty was reviewed by video replay to obtain the necessary information.
  + **Sample Size:** 628 penalties.

## Key Variables:

1. **Player Information:** Includes the name, region, and footedness of the penalty taker.
2. **Match Context:** Captures details such as match location, time remaining, match condition, and team category.
3. **Penalty Characteristics:** Encompasses the angle chosen for the shot and the penalty outcome (goal, saved, missed).
4. **Goalkeeper Information:** Covers goalkeeper-specific data such as name and height.

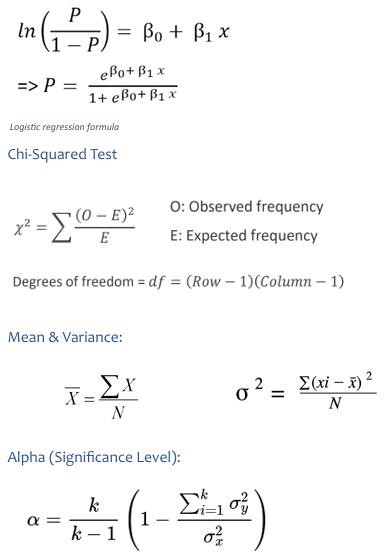
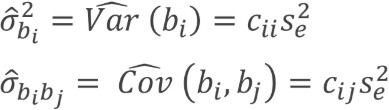
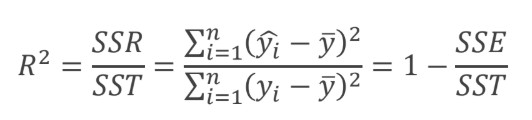
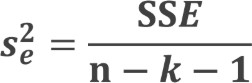
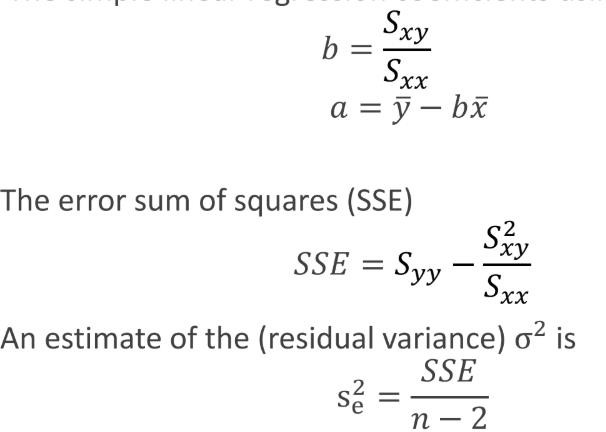
## Procedure:

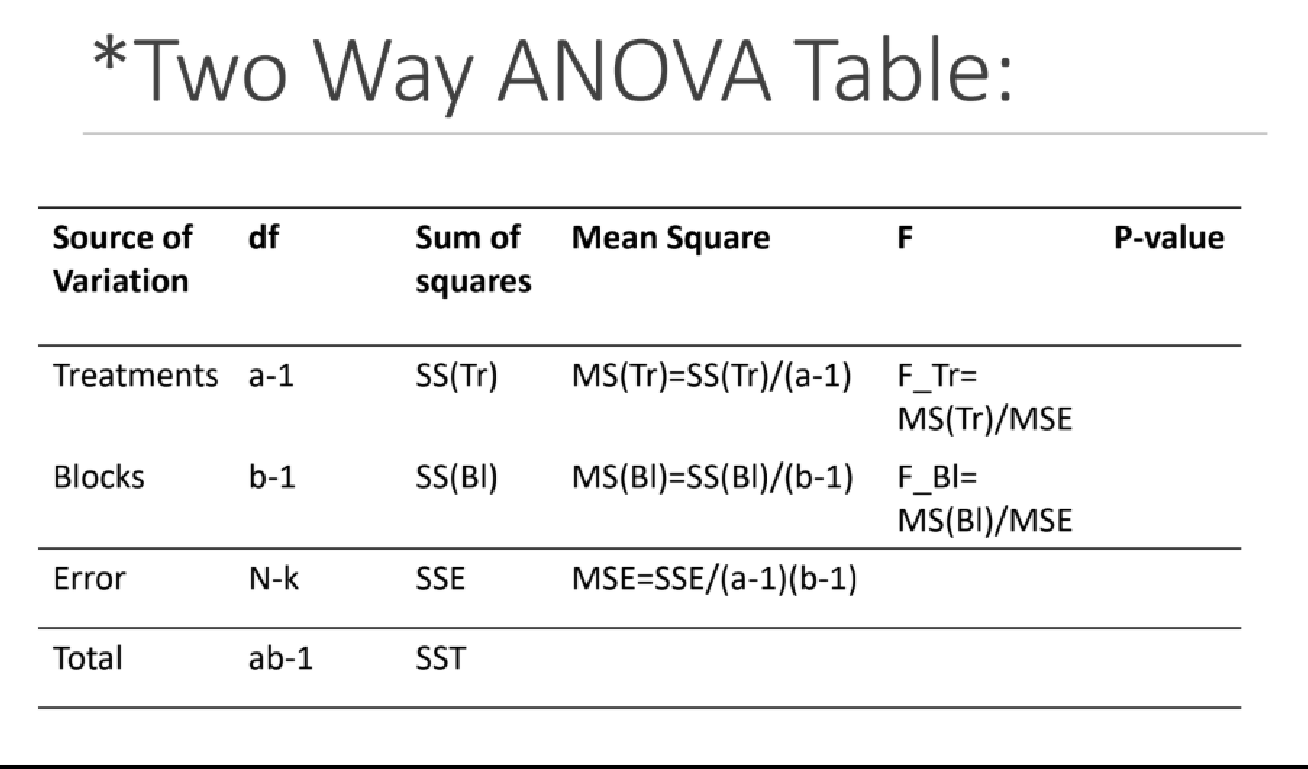
1. **Data Collection:** Video replay was used to meticulously record each penalty’s details across six football seasons.
2. **Data Preprocessing:** The data was cleaned and standardized to ensure consistency across variables. Missing values were imputed where applicable.
3. **Variable Categorization:** Variables were categorized into physical, psychological, and contextual factors for detailed analysis.
4. **Model Development:** Statistical models and machine learning algorithms were developed to identify patterns and relationships between variables.





# General rules used:



****

**Statical Questions: -**

**Question 1: - Does the Relationship Between Goalkeeper Height and Penalty Success Vary with the Shot Angle?**

Methodology: We performed a **multivariate regression analysis** to assess how goalkeeper height and penalty shot angle impact penalty success. Analyze interaction effects by including a term for height × angle.

|  |  |
| --- | --- |
| Dependent Variable: | Penalty outcome (binary). (e.g., 1 = scored, 0 = missed). |
| Independent Variables: | Goalkeeper height (numeric). |
| dummy variables | penalty angles (e.g., right top, left bottom). |
| R² = | 1.56%: The model explains only a small portion of the variability in penalty success rates. |
| Significance F = | 0.565 suggesting that the combination of goalkeeper height  and penalty angles does not reliably predict penalty success in this dataset. |

Scatter-plot

|  |  |
| --- | --- |
| Left- Bottom Angle: | Left -Top Angle: |
| Middle – Bottom Angle: | Middle Top Angle: |

|  |  |
| --- | --- |
| Right – Bottom Angle: | Right – Top Angle: |

Hypothesis Testing:

|  |  |
| --- | --- |
| Null Hypothesis (H₀): | Goalkeeper height and penalty shot angles do not significantly affect penalty success. |
| Alternative Hypothesis (H₁): | Goalkeeper height and penalty shot angles significantly affect penalty success, with potential interaction effects. |
| Hypothesis Result: | we fail to reject the null hypothesis as the overall model and individual predictors were not significant. |
| Implications: | penalty success is likely influenced by factors other than goalkeeper height and shot angle. |

Limitations: -

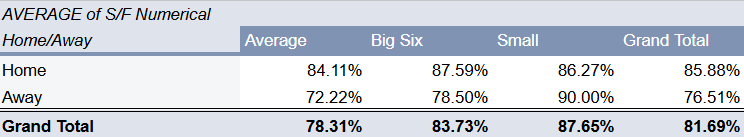
|  |  |
| --- | --- |
| Low R² Value: | limited explanatory power, highlighting the need for additional variables |
| Linear Model Assumptions | outcomes may not follow a simple linear relationship with  goalkeeper height and angle, potentially requiring non-linear models |
| Unexplored Factors | Key variables like penalty taker experience, match pressure, and goalkeeper positioning were not included |

Conclusion: -While goalkeeper height and penalty angles show minimal influence on penalty success in this dataset, this study provides valuable insights into the limitations of these factors. The findings underline the complexity of penalty outcomes and the need for further research incorporating additional variables and advanced analytical technique conclude to 3 sentences.

Methodology: ANOVA Analysis to determine whether penalty success rates are influenced by:

**Question 2: How does match location (home/away) impact penalty success rates for Big Six, average, and small teams?**

1. Match location (home vs. away).



1. Team category (Big Six, Average, Small).

|  |  |
| --- | --- |
| alpha (significance level) | 0.05 |

Hypothesis Testing:

1. Match Location

|  |  |
| --- | --- |
| Null Hypothesis (H₀): | The mean penalty success rate is the same for home and away matches. Match location has no effect on penalty success rates. |
| Alternative Hypothesis (H₁): | The mean penalty success rate is different for home and away matches. Match location has an effect on penalty success rates. |

1. Team Category

|  |  |
| --- | --- |
| Null Hypothesis (H₀): | The mean penalty success rate is the same across all team categories (Big Six, Average, Small). Team category has no effect on penalty success rates. |
| Alternative Hypothesis (H₁): | The mean penalty success rate differs between team categories. Team category has an effect on penalty success rates. |

Results (1):

|  |  |
| --- | --- |
| F-value | 1.4299812 |
| P-value | 0.3543170863 |

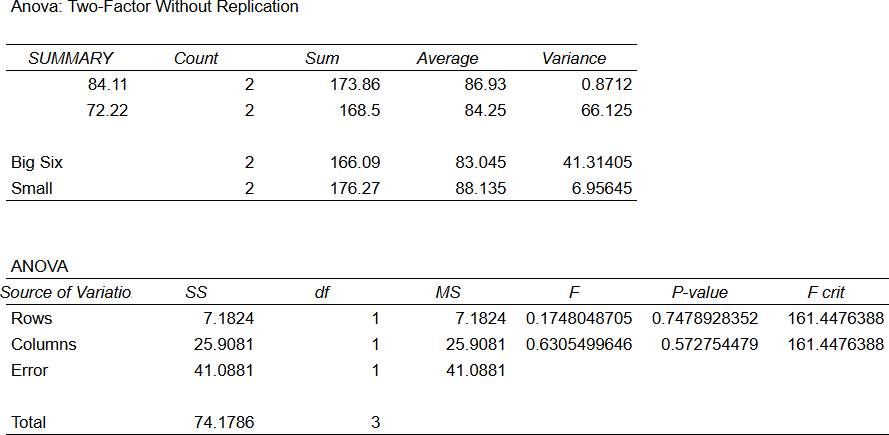
|  |  |
| --- | --- |
| Interpretation | The p-value (p > 0.05) indicates that the difference in penalty success rates between home and away matches is not statistically significant. |
| Hypothesis Result: | We fail to reject the null hypothesis |
| Implications: | Data indicates it is not enough to claim that Match Location has a significant effect on penalty success rates |

Results (2):

|  |  |
| --- | --- |
| F-value | 1.433269707 |
| P-value | 0.4109696501 |
| Interpretation | The p-value (p > 0.05) indicates that the difference in penalty success rates among Big Six, Average, and small teams are not statistically significant. |
| Hypothesis Result: | We fail to reject the null hypothesis |
| Implications: | Data indicates it is not enough to claim that team category has a significant effect on penalty success rates. |

The ANOVA results suggest:

1. Match location: does not significantly affect penalty success rates.
2. Team categories: does not significantly affect penalty success rates.



**Question3: How do location, education, and remote work affect automation risk?**

**Methodology:** We performed a **Multiple Linear Regression Analysis** to examine how location, required education, and remote work ratio (%) influence automation risk in jobs. Dummy variables were created for each category of **Location** and **Required Education**, excluding one category as the reference group to avoid multicollinearity.

|  |  |
| --- | --- |
| Dependent: | Automation Risk (%) (numeric) |
| Independent: | Remote Work Ratio (%) (numeric)  Location (categorical, dummy variables)  Required Education (categorical, dummy variables) |

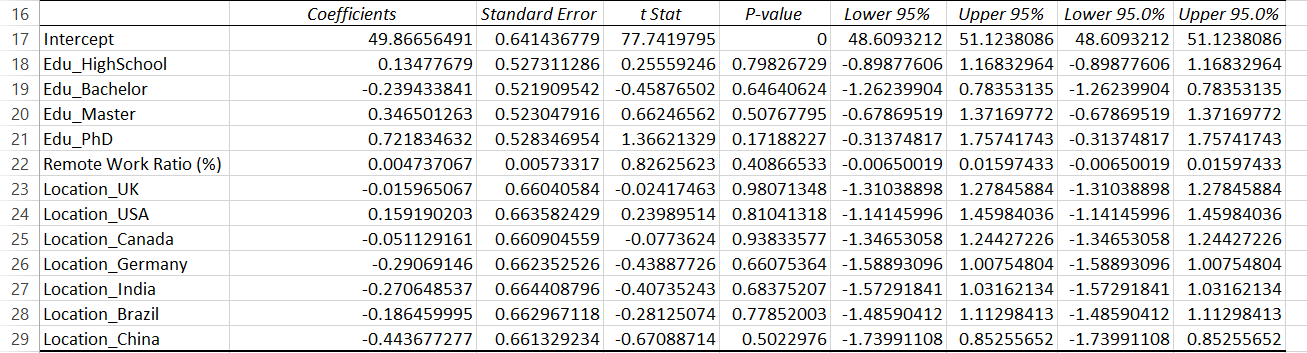
* **Confidence level is 95%**
* **1 is for presence, 0 for absence**

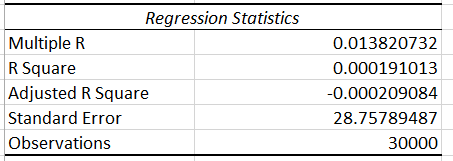
**The regression equation was:**

Automation Risk=β0+β1(Remote Work Ratio)+β2...+βn+ϵ

|  |  |
| --- | --- |
| Null Hypothesis (H₀): | Location, required education, and remote work ratio have no significant effect on automation risk. |
| Alternative Hypothesis (H₁): | At least one of these predictors has a significant effect on automation risk. |

**Results**





**Model statistics:**

* R² = 0.000 → Model explains almost none of the variation. {Poor fit}
* Adjusted R² = -0.000
* F-statistic p-value = 0.929 → Model not significant.

**Interpretation**

* None of the predictors have a statistically significant effect (all p-values > 0.05).
* Remote Work Ratio (%) has a tiny positive coefficient (+0.0047), meaning a 1% increase in remote work changes automation risk by 0.0047%, which is negligible.
* Differences between locations and education levels are also minimal and not significant.

**Assumption Checks**

* **Linearity:** No major violations detected.
* **Multicollinearity:** Not an issue (dummy variable encoding avoids redundancy).
* **Normality of residuals:** Residuals roughly follow a normal distribution.

**Plots**

**Remote work ratio**

* Since we converted the location and education variables into dummy variables, therefore it is either 1 or 0, present or absent respectively.

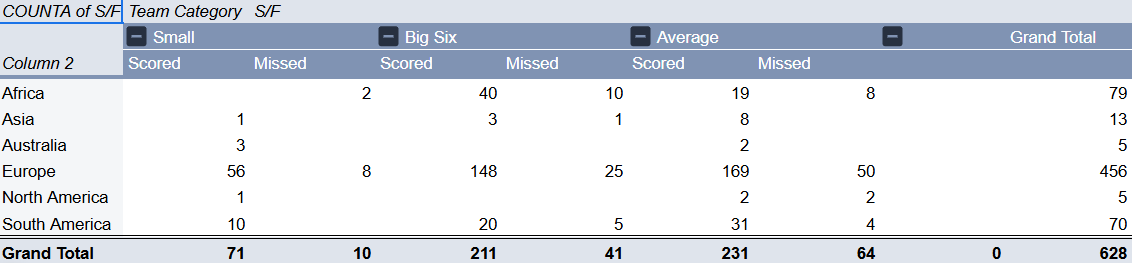
**Conclusion**

The multiple linear regression analysis shows that Location, Required Education, and Remote Work Ratio (%) do not significantly predict Automation Risk, so we failed to reject the null hypothesis. The model’s explanatory power is negligible (R² = 0.000) and no predictor had a p-value < 0.05. Other variables not included in this model likely have a stronger influence on automation risk.

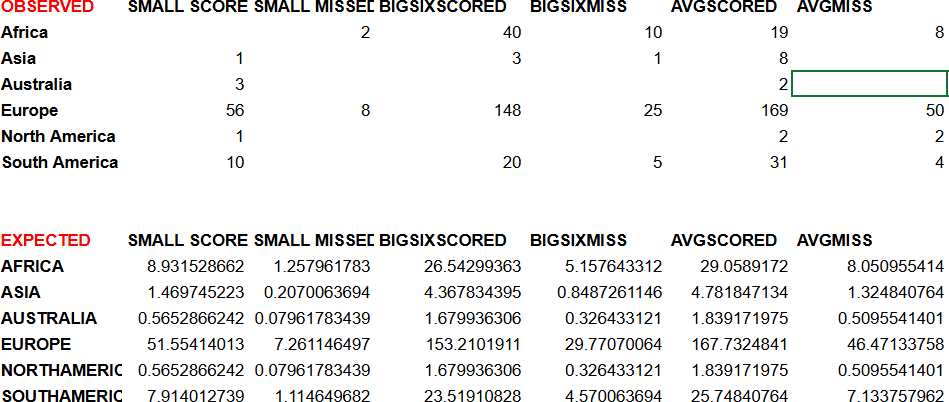
**Question4: How do penalty success rates differ across players from different countries, and does this depend on their team’s reputation (Big Six, average, and small teams)?**

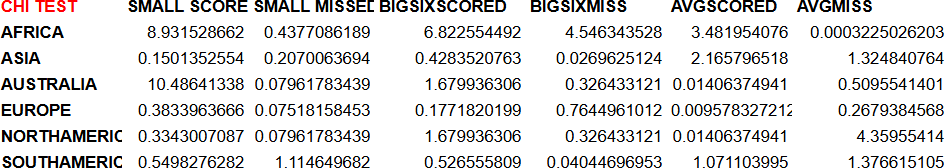
Methodology: Chi-Square Test for Independence

1. Player’s region and team’s categories (**Big Six, average, and small teams**).
2. Dependent variable: Penalty success rate. Application Steps: -
3. Group players based on their countries/regions (e.g., Europe, South America)
4. Create a contingency table of success rates by region and team categories.



1. Perform a chi-square test to assess independence.



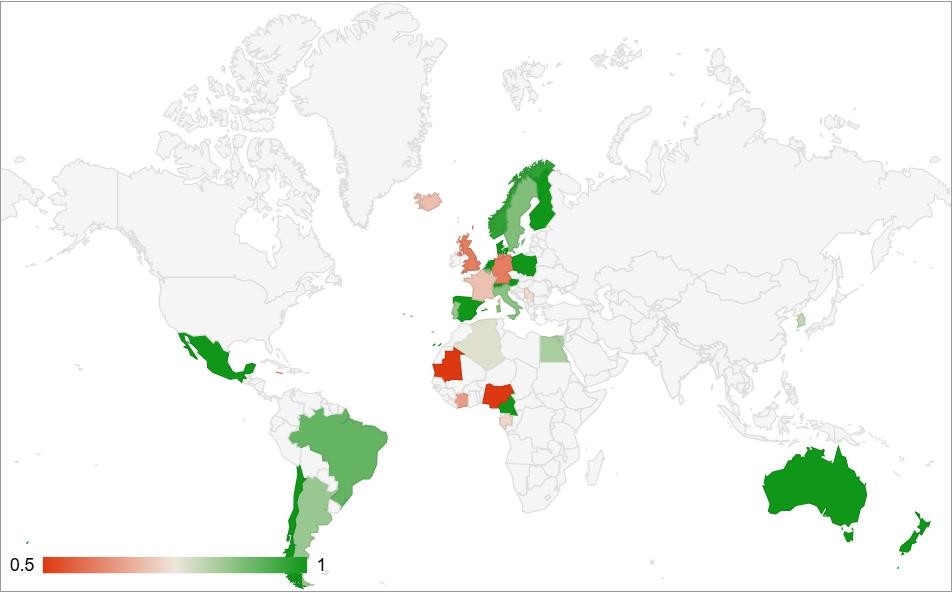


Results: -

|  |  |
| --- | --- |
| P-value | 0.05% |
| Significance Level (α) | 5% (0.05) |
| Degrees of freedom | 25 |
| Test Value | 54.77 |
| Critical Value | 37.65 |

Hypothesis Testing:

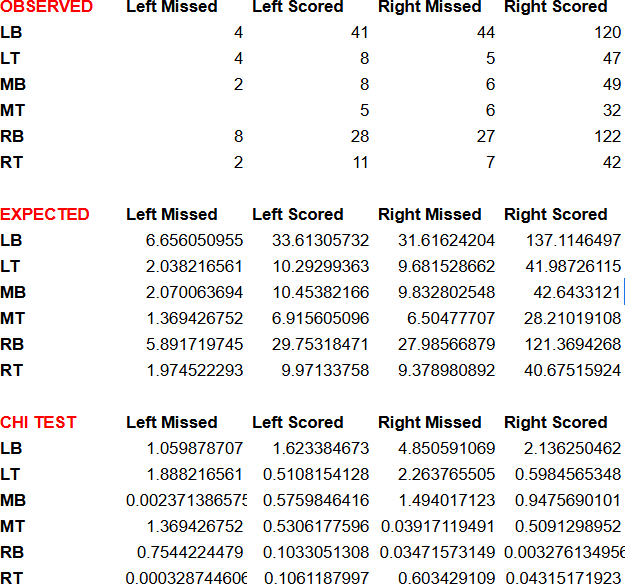
|  |  |
| --- | --- |
| Null Hypothesis (H₀): | Penalty Success rate will not change upon changing the team league and region |
| Alternative Hypothesis (H₁): | Penalty success rate will change upon changing the team league and region |
| Hypothesis Result: | Test value > Critical value meaning, it’s in the rejection region, reject H0 |
| Interpretation | penalty success rates vary significantly across countries, influenced by the team’s reputation. |



Methodology: A Chi-Square Test for Independence was conducted to determine whether penalty success rates (scored vs. missed) depend on:

**Question 5: How do penalty success rates differ across players from different penalty angles and footedness (left-footed vs. right-footed)?**

1. Penalty angle (Left Bottom, Left Top, Middle Bottom, Middle Top, Right Bottom, Right Top) 2-Footedness (Left-footed vs. Right-footed)



Results:

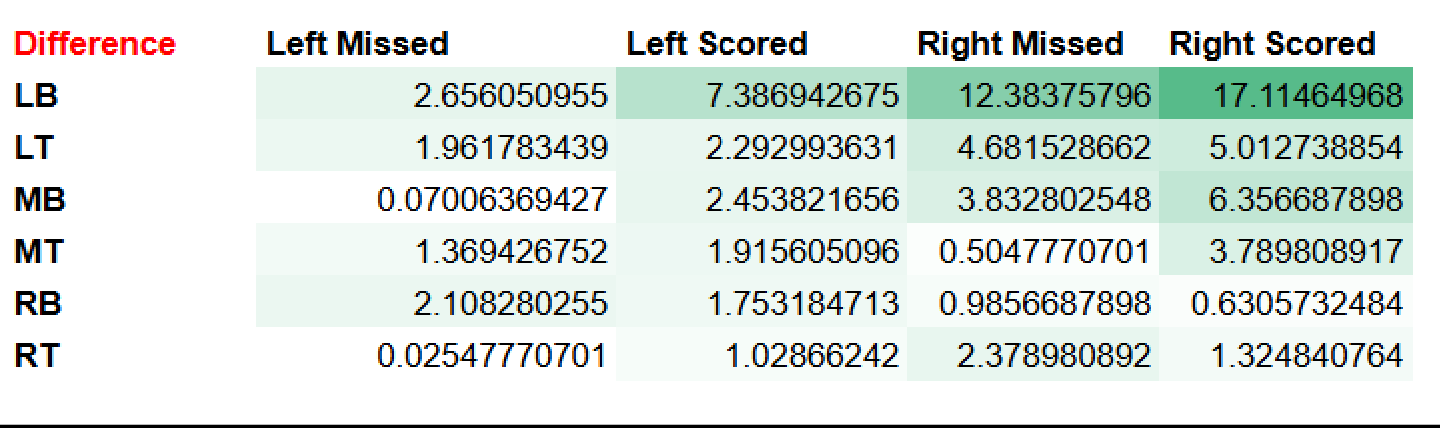
|  |  |
| --- | --- |
| Chi-Square test statistic | 22.048 |
| P-value | 0.1065 |
| Significance Level (α) | 5% (0.05) |
| Degrees of freedom | 15 |
| Test Value | 22.05 |
| Critical Value | 24.996 |

Hypothesis Testing:

|  |  |
| --- | --- |
| Null Hypothesis (H₀): | Penalty success rate is independent of the penalty angle and footedness.  There is no relationship between the penalty angle and footedness with penalty success (scored/missed). |
| Alternative Hypothesis (H₁): | Penalty success rate is dependent on the penalty angle and footedness.  Penalty success varies significantly based on the penalty angle and footedness. |
| Hypothesis Result: | P-value is greater than 0.05, we fail to reject the null hypothesis. |
| Interpretation | Observed frequencies of penalty outcomes (scored or missed) for various angles and footedness do not deviate significantly from the expected frequencies, meaning the two variables  (penalty angle and footedness) do not influence the success rates of penalties in this analysis. |

Conclusion:

Based on the Chi-Square test, penalty success rates are not significantly influenced by the penalty angle or footedness of the player in this dataset. Therefore, there is insufficient evidence to suggest a relationship between penalty success and the angle/footedness variables.



## Conclusion:

This study thoroughly explores the factors influencing penalty success in football, offering several critical insights. It finds that goalkeeper height has minimal interaction with penalty angles, and match location shows no significant effect on success rates across team categories, challenging assumptions about home advantage. Psychological pressures, such as time remaining and match conditions, also appear to have limited impact, emphasizing the primacy of player skill and preparation. However, significant variations in success rates across players from different countries, influenced by their teams’ reputations, highlight the role of cultural and contextual dynamics. Furthermore, penalty success at specific angles is not linked to footedness, suggesting technical precision outweighs biomechanical advantages.

These findings underscore the complexity of penalty scenarios, driven by an intricate interplay of physical, psychological, and contextual factors. The insights are highly practical: coaches can design tailored training regimens and optimize player selection, analysts can enhance predictive models for strategy development, and teams can address cultural influences to improve performance. Future research should investigate additional variables, such as goalkeeper tactics and decision-making under pressure, to deepen understanding. Overall, this study offers valuable guidance for refining strategies, enhancing player preparation, and improving outcomes in crucial match situations**.**

## Appendix:

Tools:

## Excel sheet

**Google Sheets**: Utilized for data collection, statistical analysis, and visualizations. Google Sheets was instrumental in creating pivot tables and charts for analyzing the dataset

**XL Miner Analysis Tool Pak:** Add-on in google sheets

Data Collection The dataset was collected manually. The process involved:

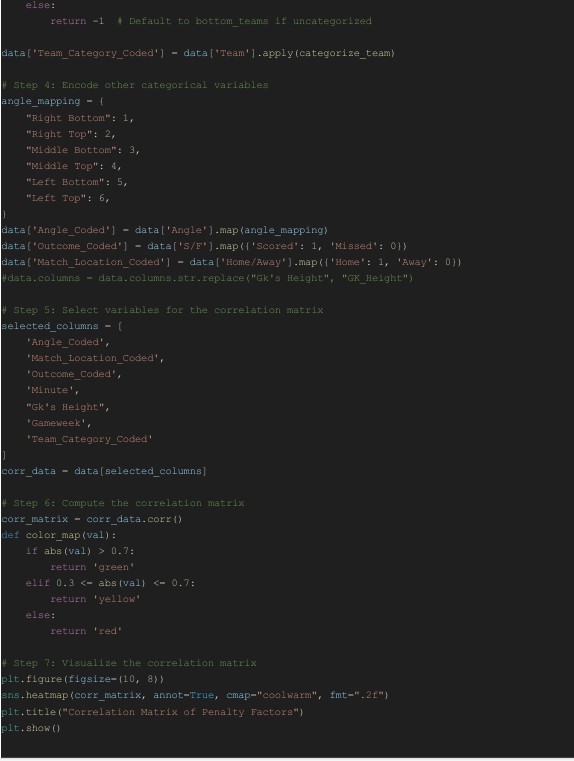
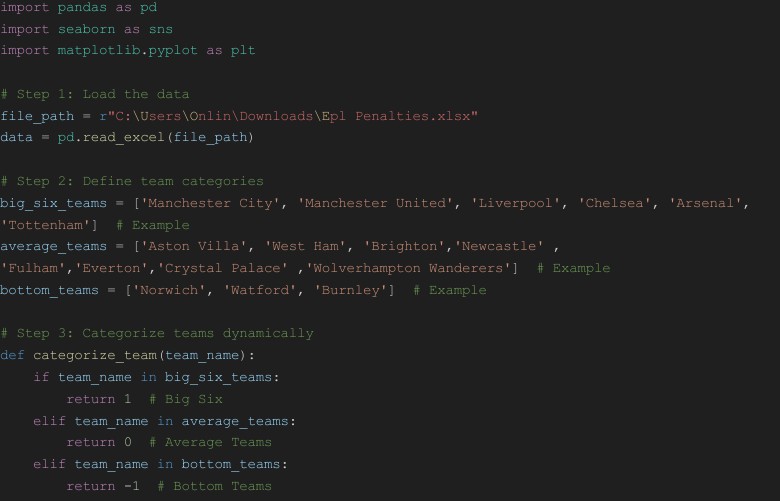
1. Gathering initial basic data about penalties from Transfer market.
2. Watching penalty videos one by one to record the required detailed information.
3. This labor-intensive process spanned approximately 40 hours.

## Statistical Analysis

Google Sheets was used for:

* + Performing statistical analysis with pivot tables.
  + Generating charts to visualize the dataset.

Python Code: Python Script: A Python script was used to enhance the statistical analysis of the dataset. This script performed advanced operations, such as calculating correlations and generating heatmaps to visualize relationships between variables



## References:

Dataset: Self-collected data, accessible via this [https://docs.google.com/spreadsheets/d/1oLWVA7EH5PCjrjFCl67KtWl2mU0imrDCJ4t5rCtED](https://docs.google.com/spreadsheets/d/1oLWVA7EH5PCjrjFCl67KtWl2mU0imrDCJ4t5rCtED9g/edit?usp=sharing) [9g/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1oLWVA7EH5PCjrjFCl67KtWl2mU0imrDCJ4t5rCtED9g/edit?usp=sharing)

Transfer Market: [https://www.transfermarkt.com/premier-](https://www.transfermarkt.com/premier-league/topErhalteneElfmeter/wettbewerb/GB1) [league/topErhalteneElfmeter/wettbewerb/GB1](https://www.transfermarkt.com/premier-league/topErhalteneElfmeter/wettbewerb/GB1)