**SEGMENTING AND CLASSIFYING THE BEST STRIKERS**

**INTRODUCTION**

In the world of football, strikers play a pivotal role in deciding the fate of matches and championships. Identifying the best strikers among a pool of talent involves a comprehensive analysis of various factors ranging from performance metrics to personal attributes.

This project, takes a deep look into a dataset containing information on **500 strikers**, aiming to uncover patterns, insights, and classifications that distinguish top-performing strikers from the rest.

**PROJECT DESCRIPTION**

This project involves utilizing **data analytics techniques** to explore and understand the characteristics and performance metrics of strikers. By employing **descriptive statistics, data visualization, feature engineering, and machine learning algorithms**, we identify the key attributes that contribute to a striker's success and classify them based on their performance.

**PURPOSE**

The primary goal is to provide a **systematic framework** for analyzing and categorizing strikers based on their performance metrics and personal attributes. This analysis helps **coaches, scouts, and football analysts** make informed decisions in **team selection, recruitment, and strategic planning**.

**DATASET DESCRIPTION**

The dataset comprises variables related to **500 strikers**, including demographic information and performance metrics. Key variables include **Nationality, Footedness, Goals Scored, Assists, Shot Accuracy, Dribbling Success, Hold-up Play, Defensive Contribution, and Consistency**.

**ANALYSIS AND KEY FINDINGS**

## 1. Descriptive Statistics & Key Metrics

* **Maximum Goals Scored by an Individual Striker:** 34
* **Portion of Right-footed Strikers:** 53.40%
* **Nationality with the Highest Average Goals Scored:** Brazil and Spain
* **Average Conversion Rate for Left-footed Players:** 0.198
* **Number of Left-footed Players from France:** 42

## 2. Statistical Analysis

* **Correlation Between Hold-up Play and Consistency Score:** 0.147 (Weak Positive Correlation)
* **Shapiro-Wilk Test for Consistency Score:** p-value = 0.451 (**Normally Distributed**)
* **Levene’s Test for ANOVA (Homoscedasticity Assumption):** p-value = 0.8083 (**Heteroscedasticity Accepted**)

## 3. Regression Analysis

* **Beta Value for Hold-up Play:** A small positive effect (**0.0015**) – Each unit increase in Hold-up Play leads to a negligible increase in Consistency.

## 4. Clustering Analysis

* **Average Total Contribution Score per Cluster:**
  + **Cluster 0:** 104.72
  + **Cluster 1:** 126.36
  + **Cluster 1 represents the best-performing strikers.**

## 5. Machine Learning Model (Logistic Regression)

* **Accuracy Score:** 100%
* **Confusion Matrix:**
  + **44 Regular Strikers Predicted Correctly**
  + **0 Best Strikers Predicted Incorrectly**

**CONCLUSION**

Through a comprehensive analysis of the dataset, we've gained valuable insights into the characteristics and performance metrics of strikers. By **segmenting and classifying strikers**, we provide a framework for identifying **top-performing strikers** and predicting their performance type.

This project serves as a valuable resource for **football professionals and analysts**, aiding in **talent identification, team selection, and strategic planning**.