



Sports Performance Analysis

Presented By : Sakshi A

Agenda

- Objective
- Dataset Overview
- Visualizations
- Correlation & Regression
- Clustering & Dashboard
- Project Outcome

Objective :

This project, titled *Sports Performance Analysis*, focuses on evaluating player and team performance using statistical data. The primary objective is to identify patterns, understand strengths and weaknesses, and suggest tactical improvements. It integrates Python for data processing, Power BI for dashboard creation, and machine learning techniques for pattern discovery. The project combines real-time metrics like goals, speed, and assists with analytical tools to deliver insightful visualizations. Overall, it serves as a comprehensive case of applying data science in sports analytics.

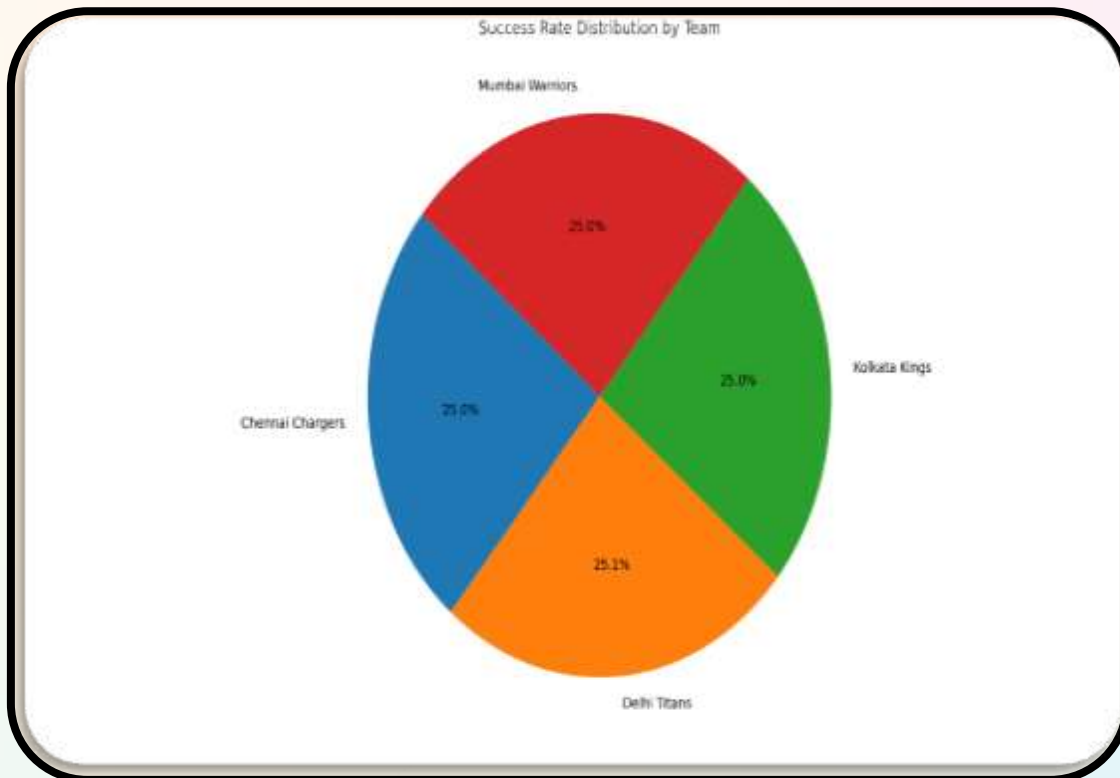
Dataset Overview :

The dataset contains over 10,000 rows of synthetic sports data, covering various sports such as **Football, Cricket, and Basketball**. Each entry includes detailed player statistics like goals, assists, speed, success rate, minutes played, and rating. Additionally, it captures team-based performance data like matches won or lost. This wide range of features enables comprehensive analysis at both individual and team levels. It is structured to reflect realistic sports scenarios, allowing meaningful machine learning and visualization outcomes.

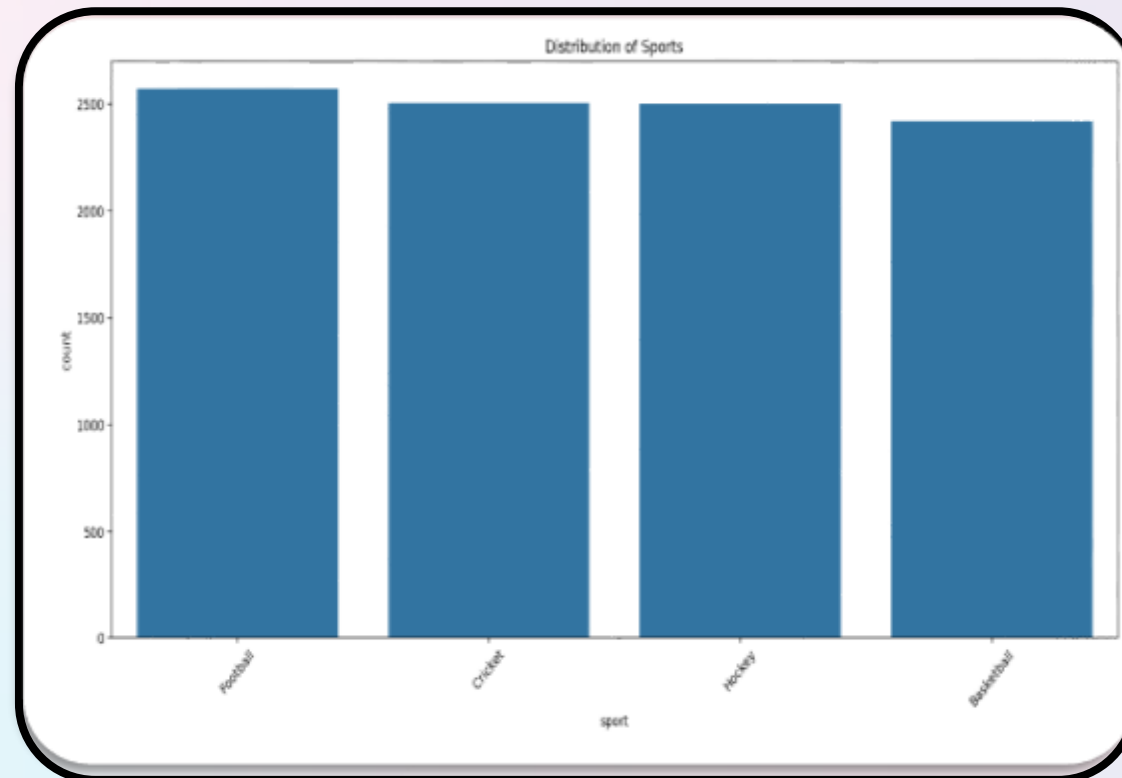
Visualizations



This pie chart illustrates **how success is distributed across different teams** by calculating the average success rate of each team. It offers a quick comparison of overall team efficiency. This visualization helps identify which teams consistently perform better and which may need strategic improvements.

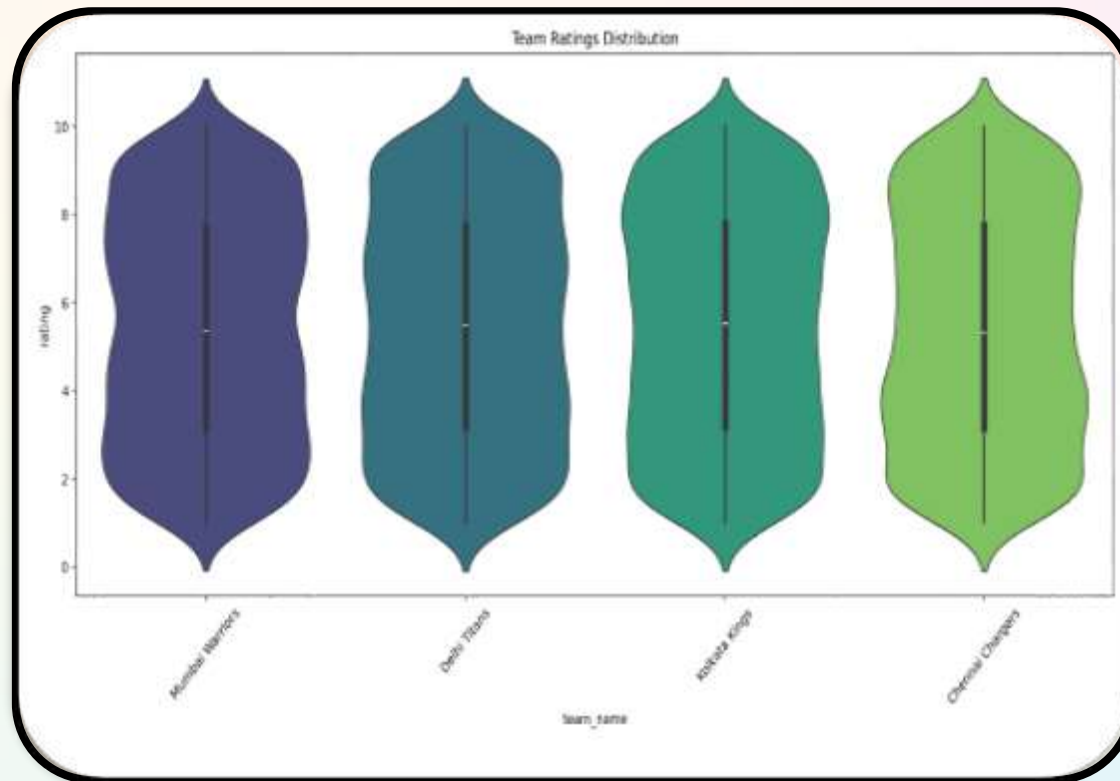


This count plot displays **how many entries belong to each sport, giving an idea of the dataset's distribution**. Sports like Football and Cricket appear more frequently, indicating that the analysis might be more robust for these categories. This visualization helps in assessing data balance before applying clustering or regression.

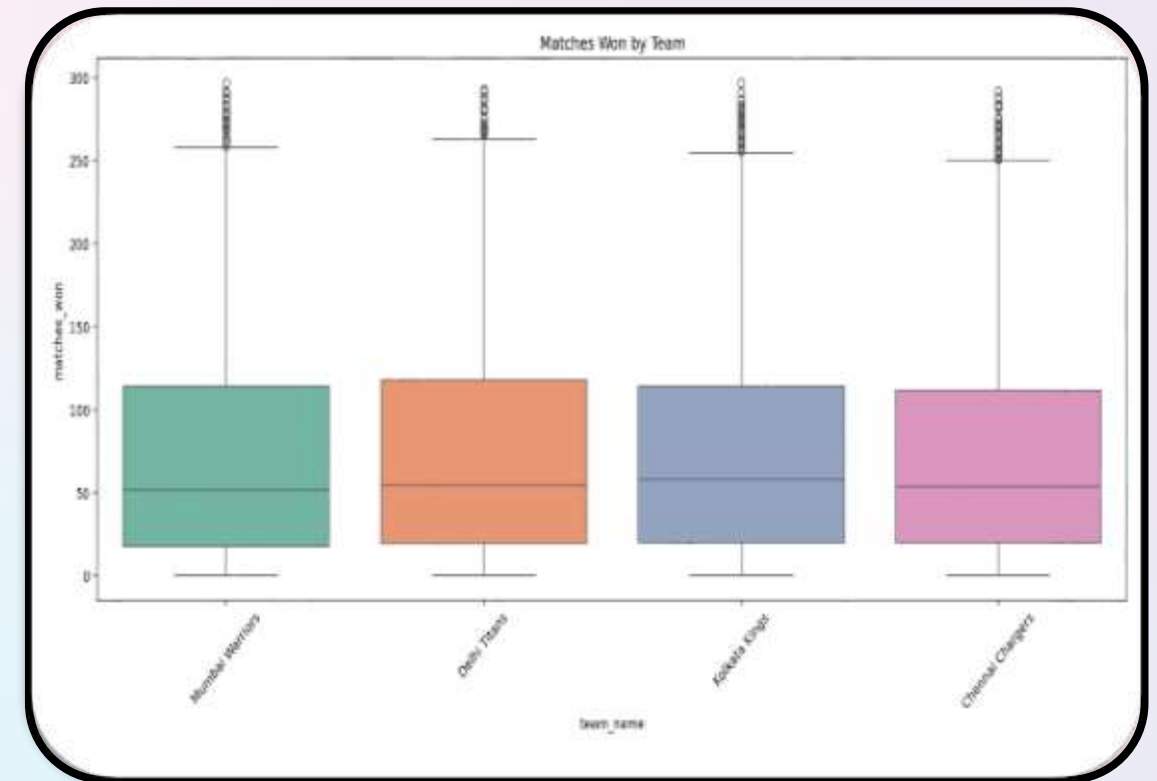


Distribution of Success Rate & Sport

The violin plot shows the **distribution of player ratings within each team**. It reveals both the spread and density of ratings, helping us understand if a team has a consistent skill level or is highly varied.

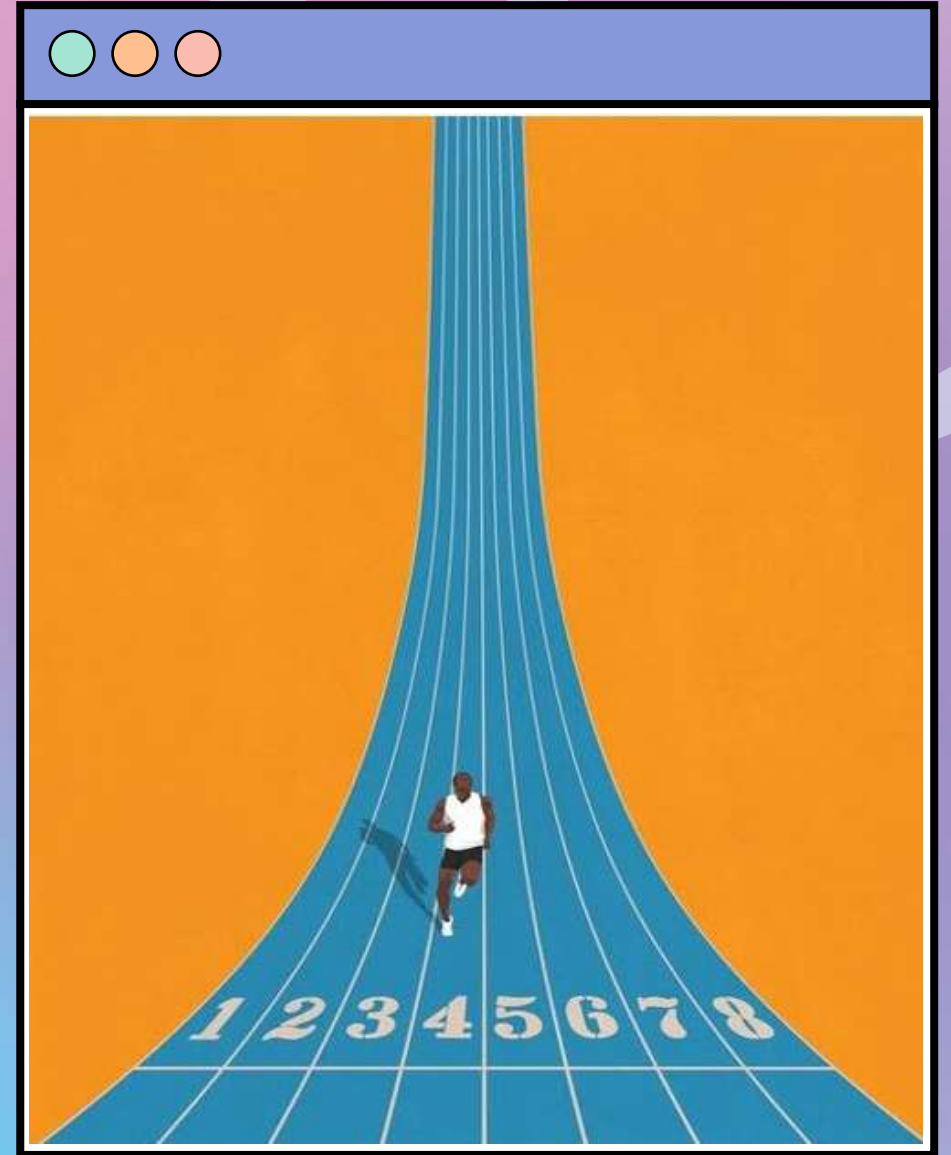


This box plot **compares the number of matches won by each team**, highlighting central tendencies and variability. Teams with high median values are consistent winners, while those with lower medians may struggle with performance.



Rate Distribution & Matches Won

Correlation & Regression



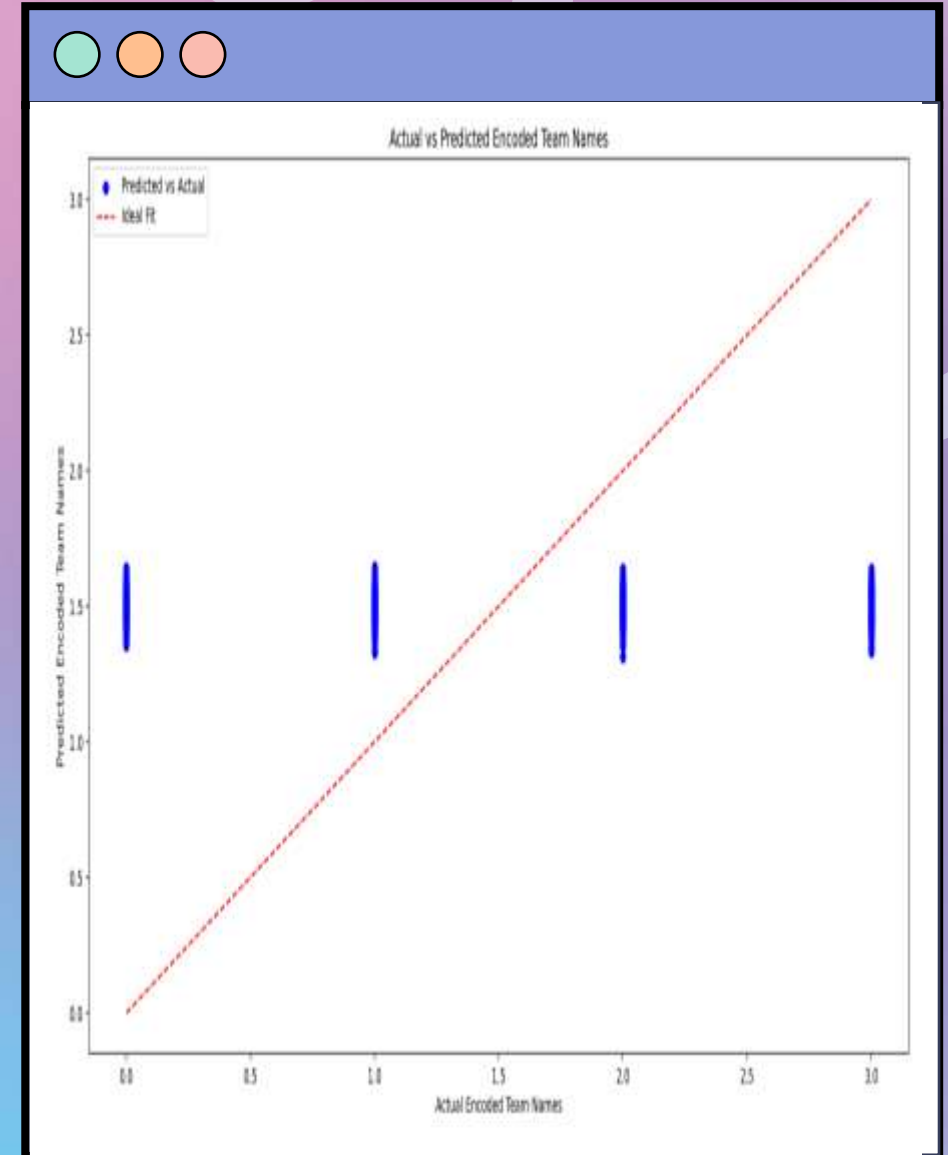
START



6/9/2025

This plot attempts to predict team names using linear regression by **encoding them numerically**. Although team names are categorical and not ideal for regression, the visualization helps demonstrate how features relate to classification. Points close to the red line indicate accurate predictions, while a spread shows inconsistency. This model primarily serves to demonstrate the predictive ability of numerical features. **It underlines the need for classification models when predicting non-numeric targets like team names.**

Linear Regression



Clustering & Dashboard

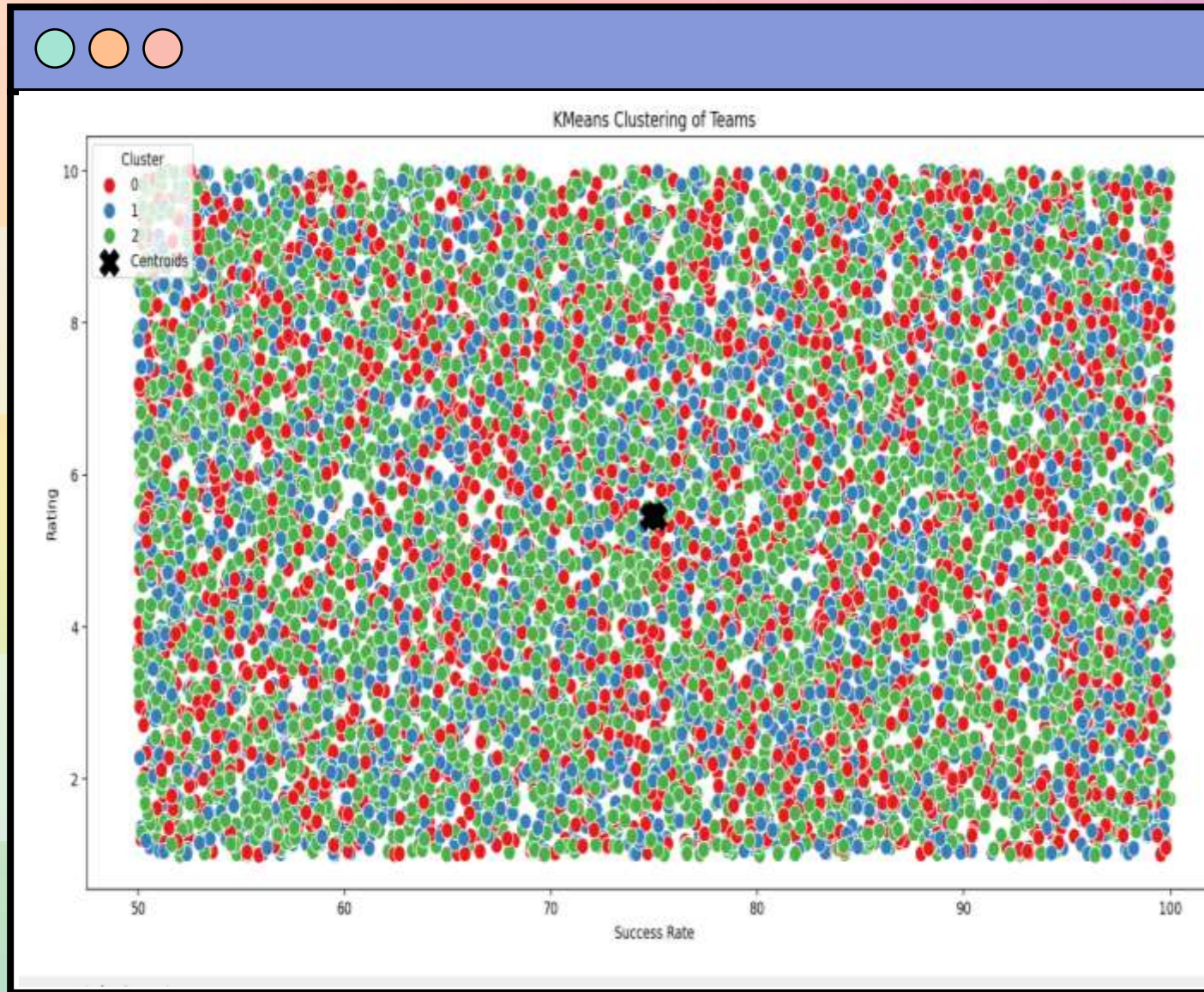


START



6/9/2025

10



Clustering was conducted using the **K-Means algorithm** to group similar-performing players or teams. The model used features like **success rate, rating, matches won, and lost** to assign cluster labels. Each cluster represents a distinct performance segment – for instance, high scorers, balanced players, or underperformers.

Sports Performance Analysis Dashboard

Team_name

All

Player_name

All

Sport

All

Sum of matches_won and Sum of success_rate by speed

376✓

Goal: 241.00 (+56.02%)

Sum of minutes_played and Sum of score by matches_played

1741✓

Goal: 1422 (+22.43%)

Sum of assists and Sum of goals by team_name

12134✓

Goal: 3006 (+303.66%)

Sum of matches_lost and Sum of saves by team_result

252880✓

Goal: 17426 (+1351.16%)

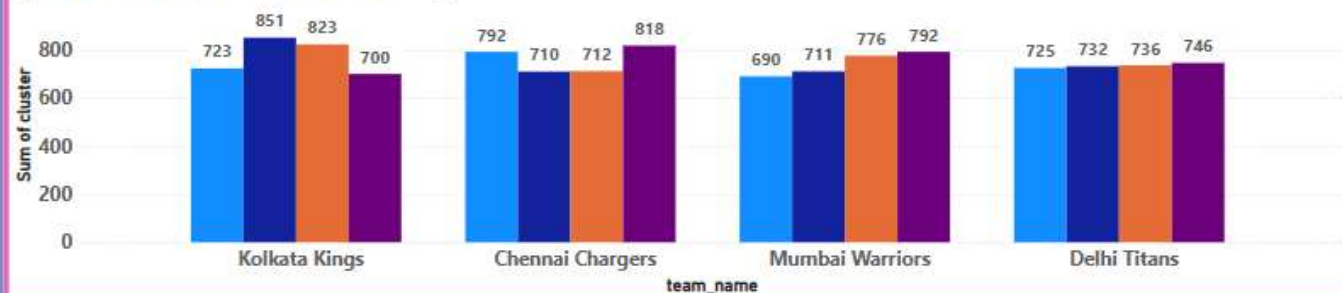
Sum of matches_won and Sum of score by injuries and team_name

• Sum of matches_won • Sum of score



Sum of cluster by team_name and sport

sport • Basketball • Cricket • Football • Hockey



Interactive Filters for Team, Player, and Sport enable focused analysis and real-time visual updates.

KPI Cards display key metrics:

- Matches Won & Success Rate by player speed Minutes Played & Score.
- Assists & Goals by Team.
- Matches Lost & Saves

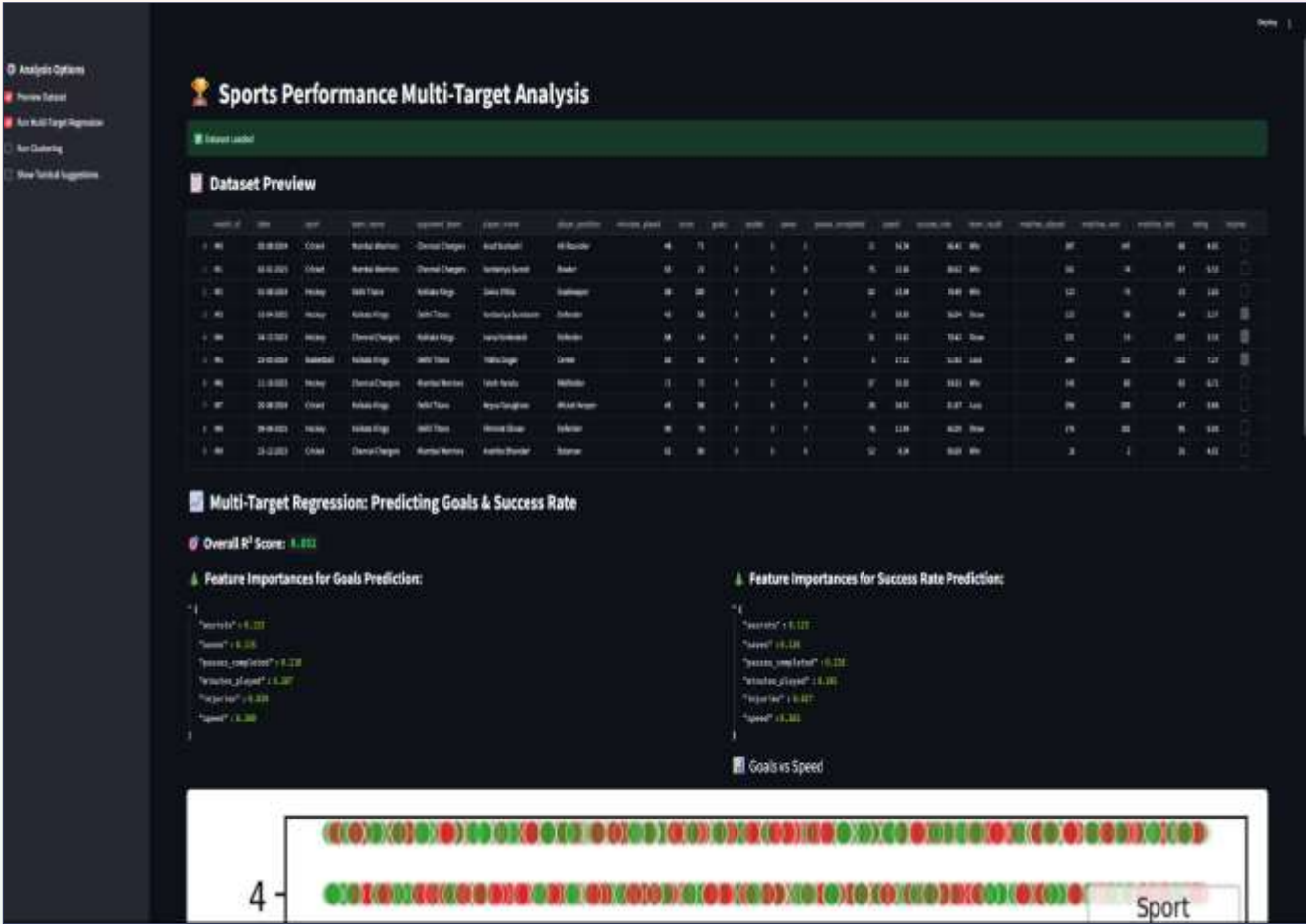
Injury Analysis (Line Chart) compares team performance with and without injured players.

Clustering by Team & Sport (Bar Chart) shows performance grouping using cluster values.


Performance Insights:

- High speed and consistent minutes correlate with better team results.
- Assists have a strong influence on overall success.
- Injuries negatively affect both matches won and total score.
- Cluster distribution helps identify elite vs average teams.


Streamlit App: Sports Performance Analyzer



This interactive web application was built using **Streamlit** to analyze sports performance data. The app allows users to predict two key performance metrics **goals** and **success rate** using multi-output regression powered by a Random Forest model. Features such as **assists**, **passes completed**, **minutes played**, **injuries**, **saves**, and **speed** are used to train the model, and the app displays **R² scores** and feature importances for both targets. A **KMeans clustering** section groups players based on performance attributes to help identify high-potential or underperforming players. The app also offers **tactical suggestions**, listing top scorers and efficient players based on calculated stats. Through its clean interface and integrated charts (scatterplots, bar graphs), the app provides valuable insights without requiring any coding making it ideal for coaches, analysts, and data science enthusiasts alike.



Project Outcome :

- 
- This project successfully combined data visualization, machine learning, and domain knowledge to analyze sports performance. A Power BI dashboard offered intuitive insights, while clustering and regression revealed deeper patterns. We identified key performance drivers and provided tactical recommendations for both individual players and teams. The outcomes demonstrate the power of data science in transforming raw statistics into meaningful strategies. It shows that sports analytics is not just about numbers — it's about better performance through insights.

THANK YOU