REAL-WORLD ANALYTICS PROJECT (RWAP)

Submitted to:
Prof. Ashok Harnal
Prof. Amarnath Mitra

Submitted by:
- Vasu Bansal(045055)
- Rishi Goswami(045045)
- Shalini Chauhan(045051)
- Ritika Grover(321156)

About Data

This dataset provides detailed information about professional football players as recorded in FIFA . Each row represents a player, with data on their biographical details, club information, FIFA ratings, attributes, special traits, and market value.

Key Aspects:

- Player Info: Name, age, nationality, height, weight.
- Club Details: Club name, league, positions, contract info.
- FIFA Ratings: Overall rating, potential, specific skill ratings.
- Attributes: Preferred foot, skill moves, work rate.
- Special Traits: Player's style of play, unique abilities.
- Market Value: Player's market value, wage, release clause.

Overall, this dataset provides a comprehensive profile of players as modeled in FIFA, useful for analyzing their performance, market value, and playing styles.

01 **DATA SHAPE : (161584,101)**

02 **Categorical Variables: 8**
Non-Categorical Variables: 93

03 **8 variables have missing values**

Problem Statement: A Football Club,aims to leverage data analysis and machine learning to compete effectively against top clubs. They seek to identify key player skills, optimal positions, and team composition to ensure each player performs efficiently and contributes to a strong, cohesive team.

Analytical Dashboard Objectives

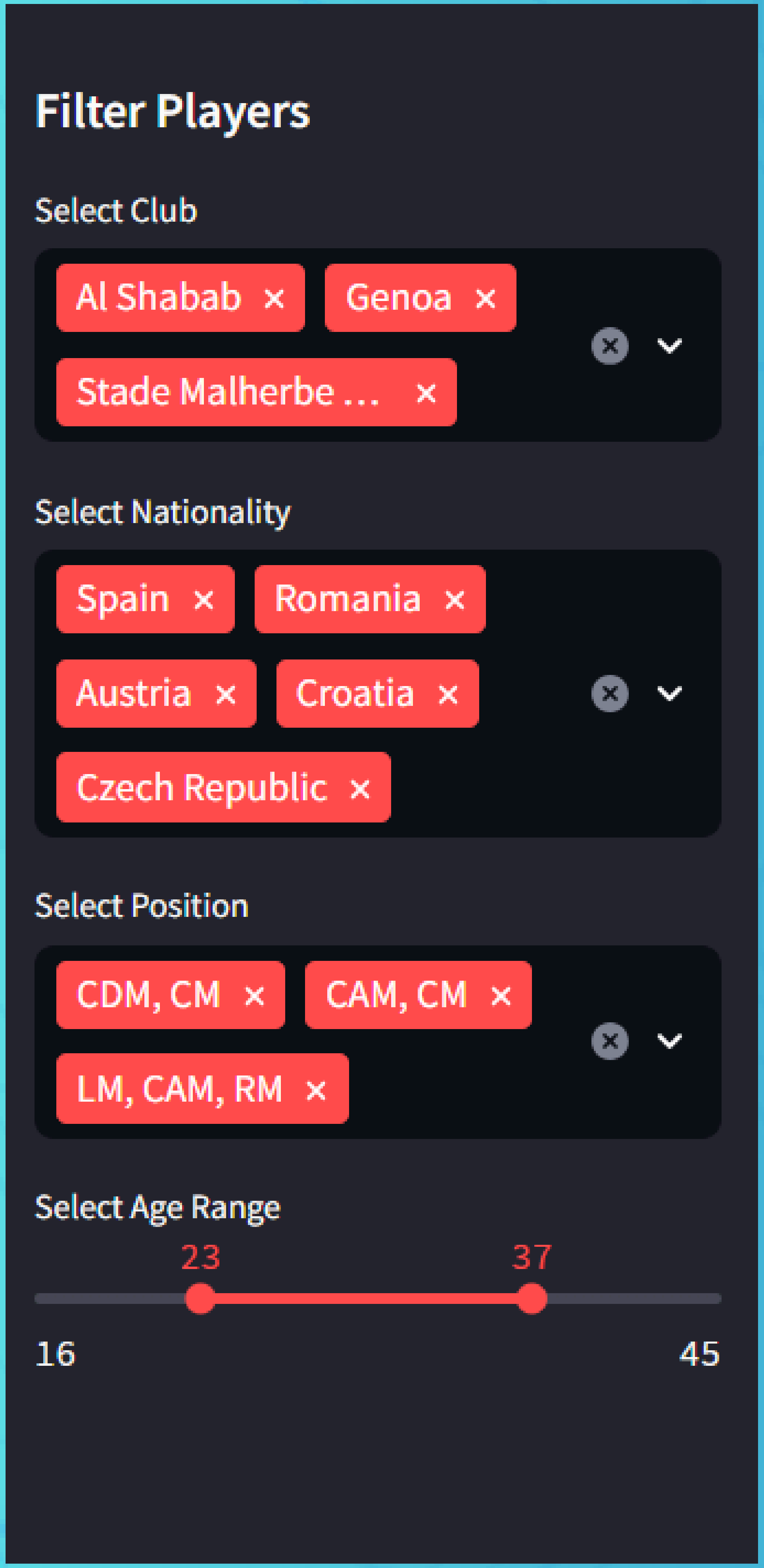
1. Helping the club board know the best players in the different Clubs.
2. Helping them Understand their competitor's Clubs.
3. Knowing the skills that need to be in their players.
4. Helping them put the players in their suitable Position.
5. Identifying Key Factors for Player Development
6. Optimizing Player Acquisition Strategy
7. Analyzing how certain skills affect player wages

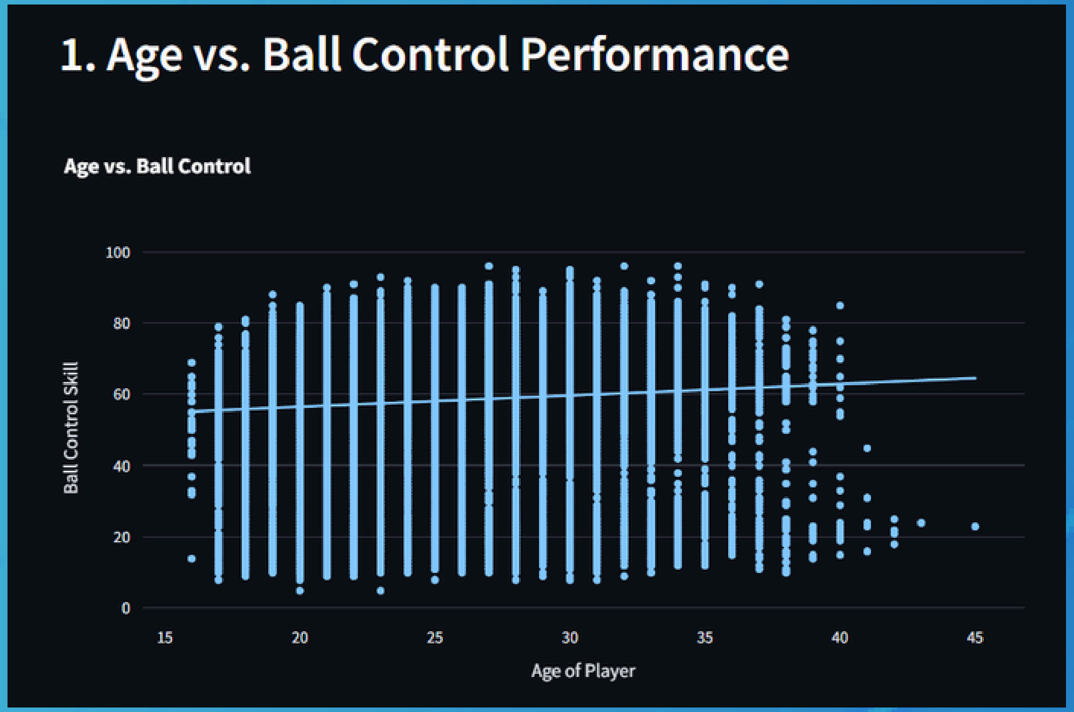
ML Model Objectives

1-Help the club assign players to roles where they are most likely to excel and contribute effectively to the team by predicting the position of players using Supervised Learning Classification Algorithms (LR,DT,KNN,RF,SVM) and comparing model results fo finding best model

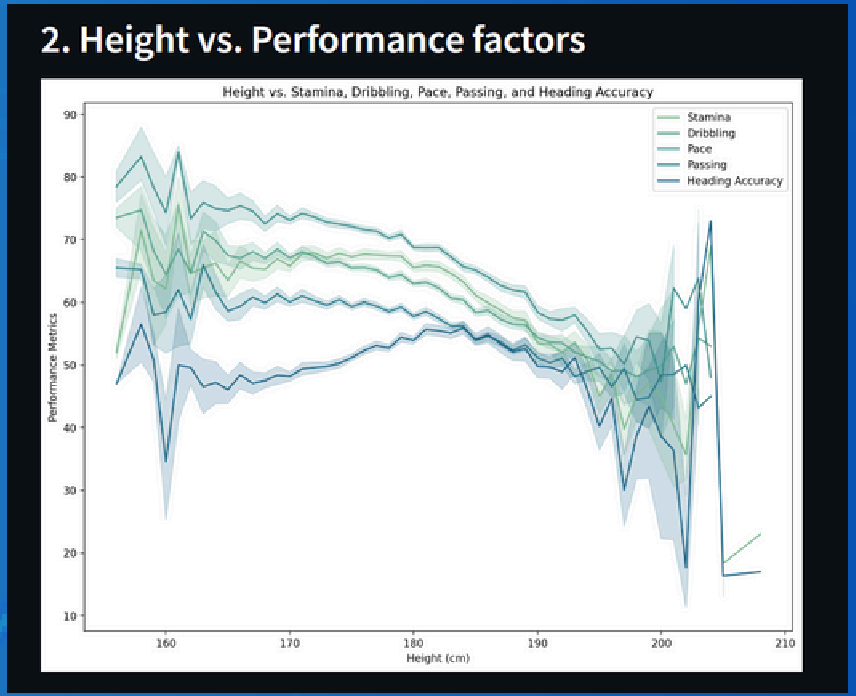
2- Feature analysis to find important features for helping the club to make policies which can help players perform better.

DASHBOARD ANALYSIS

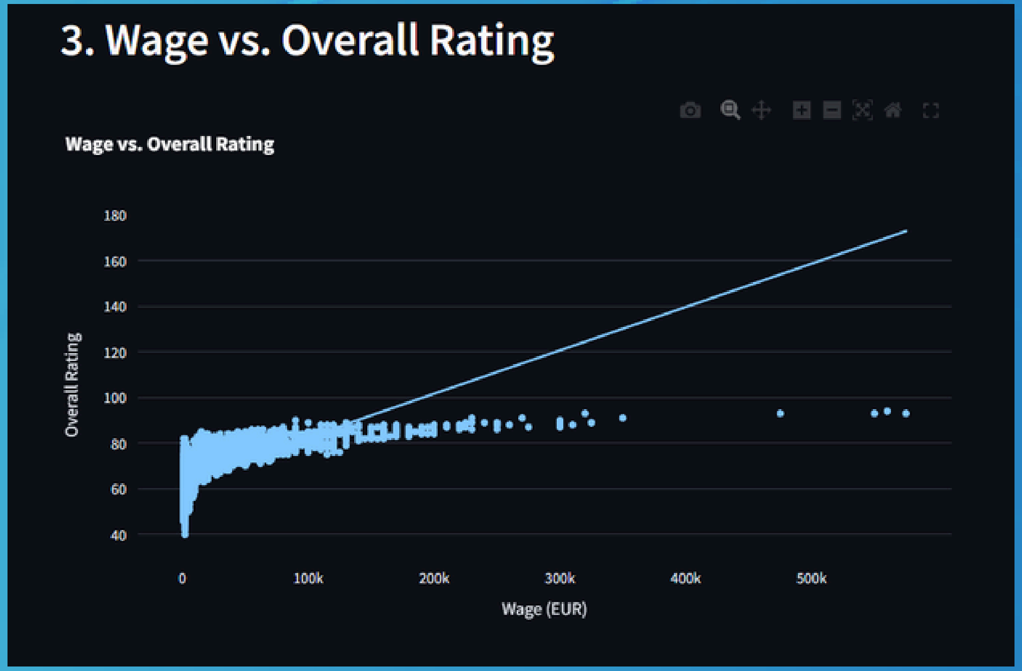




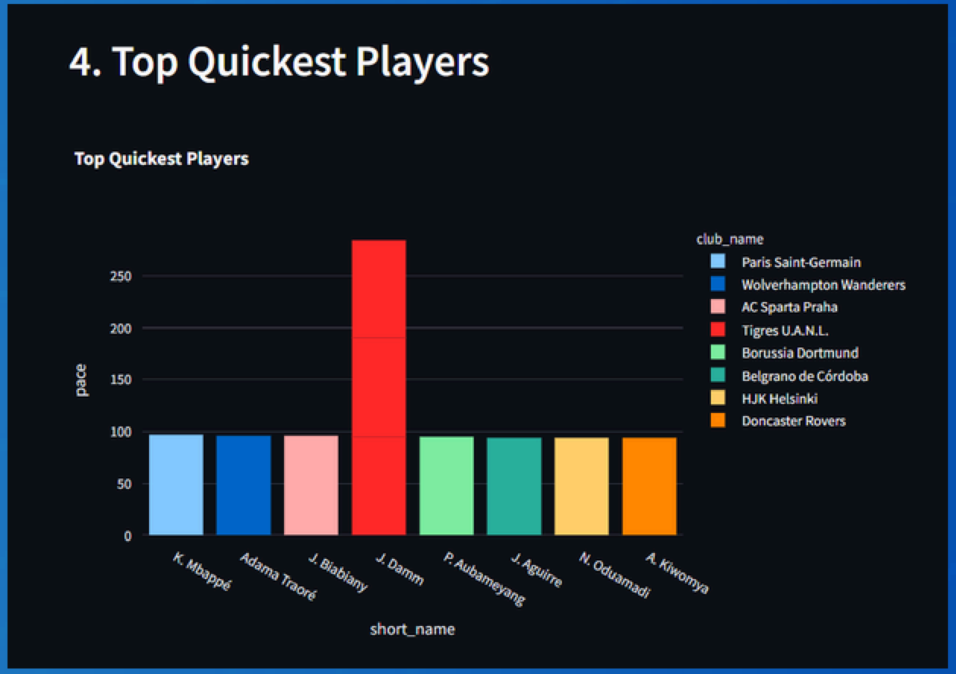
Does the Age of the Player Affect on his Ball Control Performance?



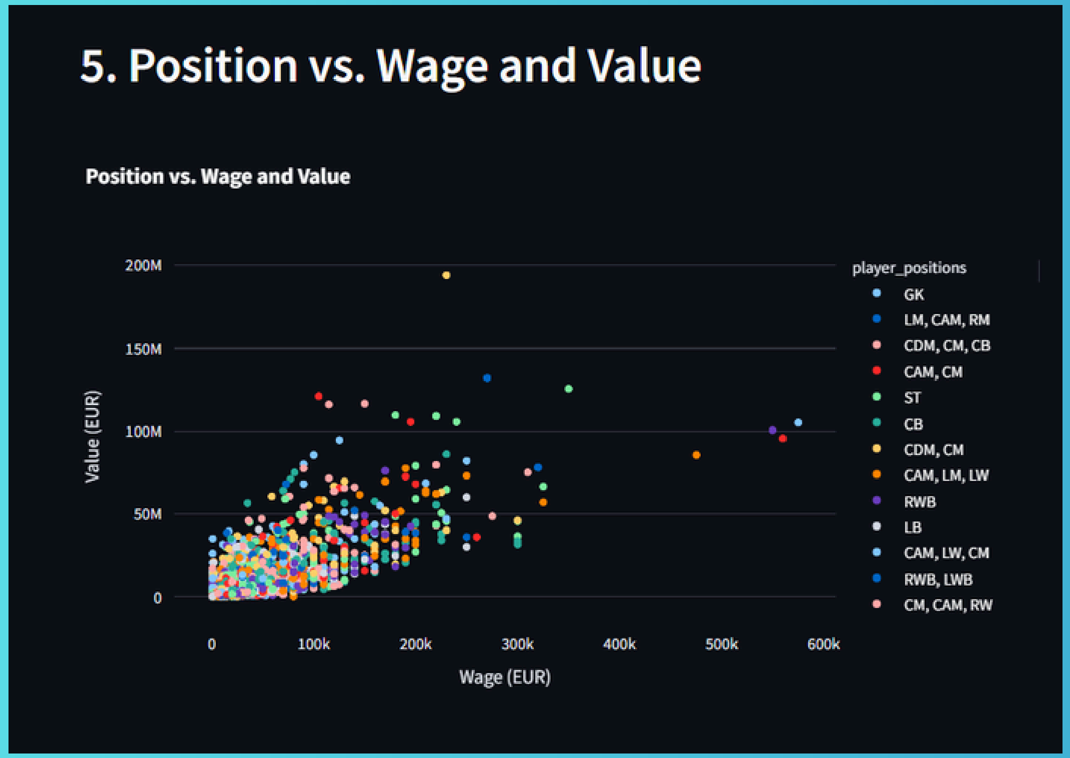
How does Height affects factors like stamina, dribbling, pace, passing and Heading-Accuracy



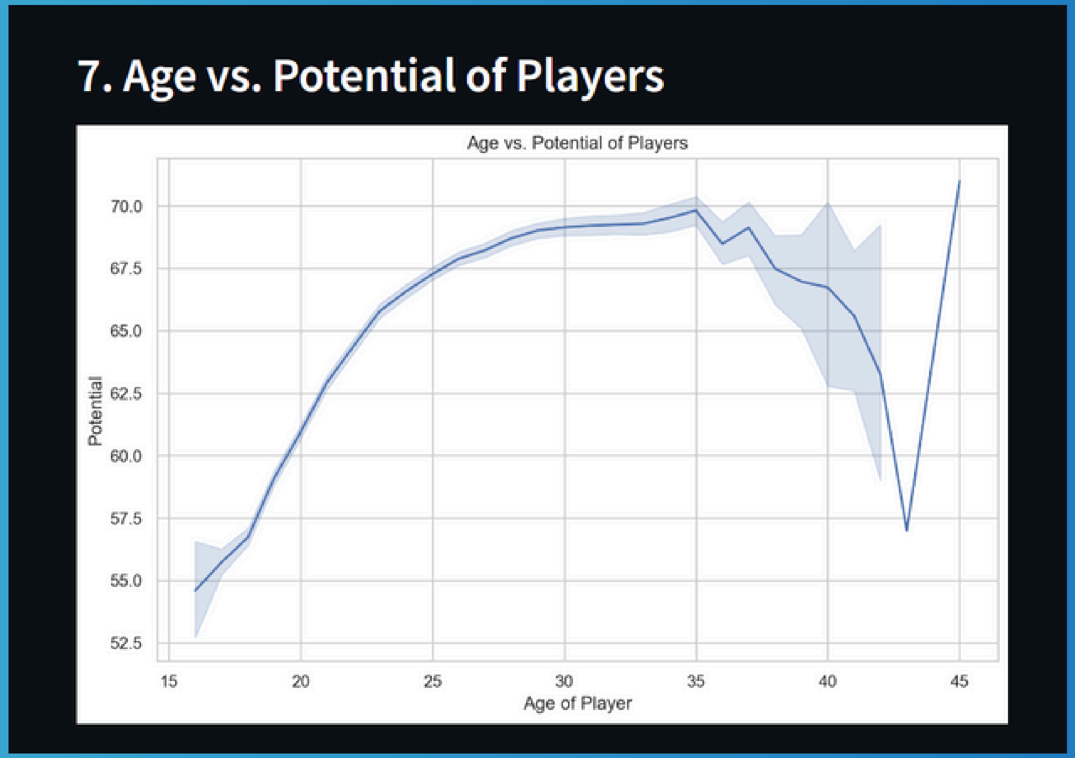
Whether there is a relation between Wage and Overall Rating of the Players



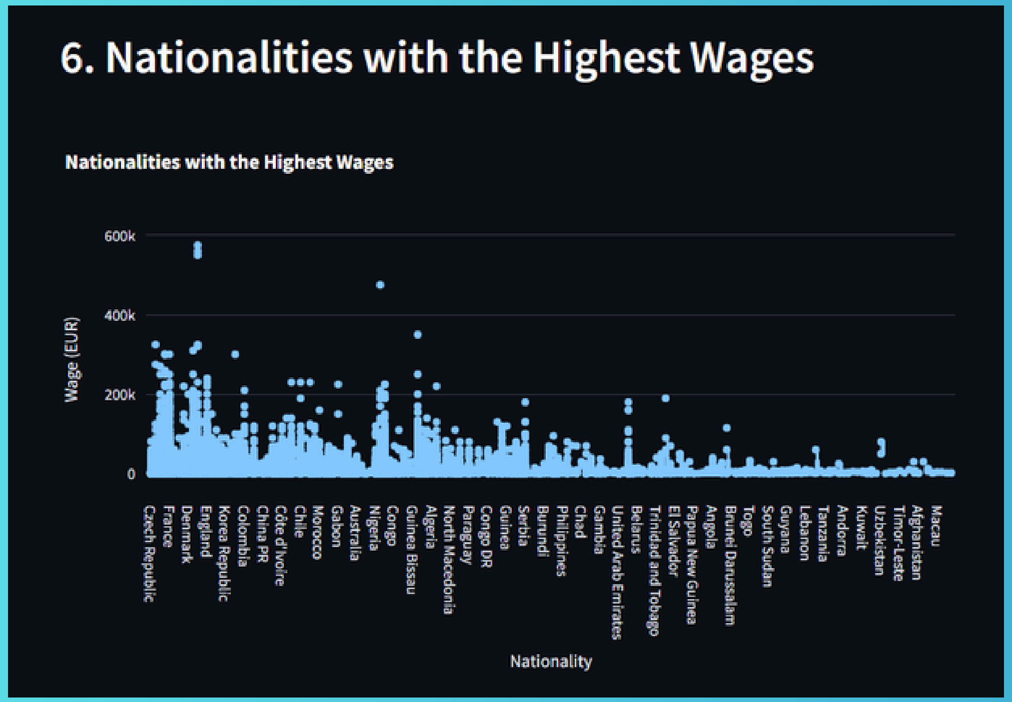
Observing the Top Quickest Players from different Clubs



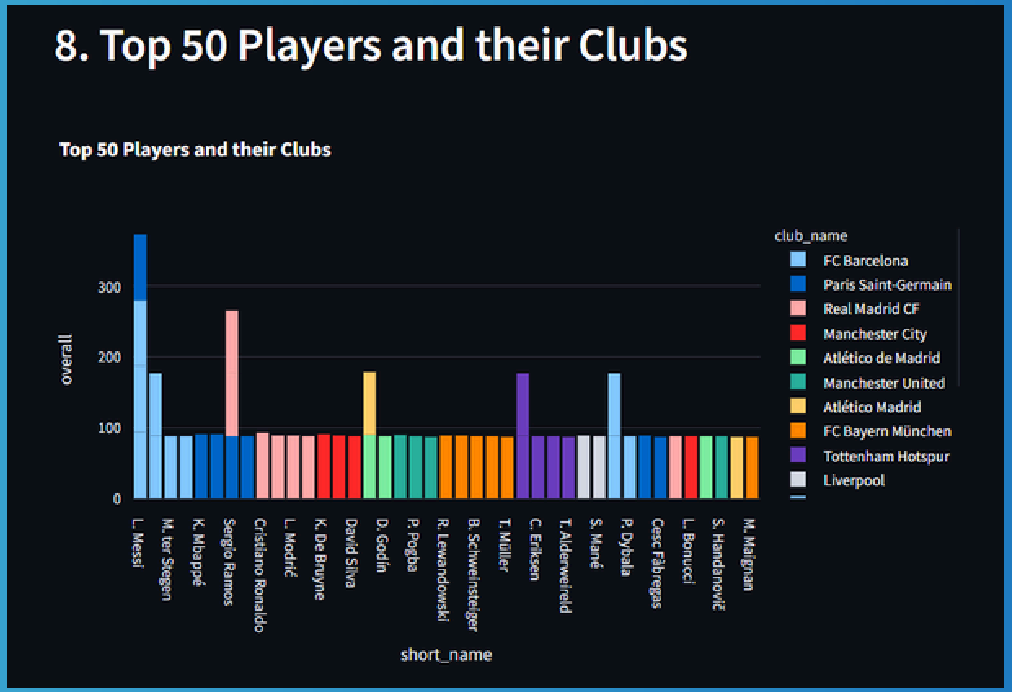
Determining if their is a relation between the Position of the Player and his Wage and Value



Show the effect of the Age on the Potential of the Players

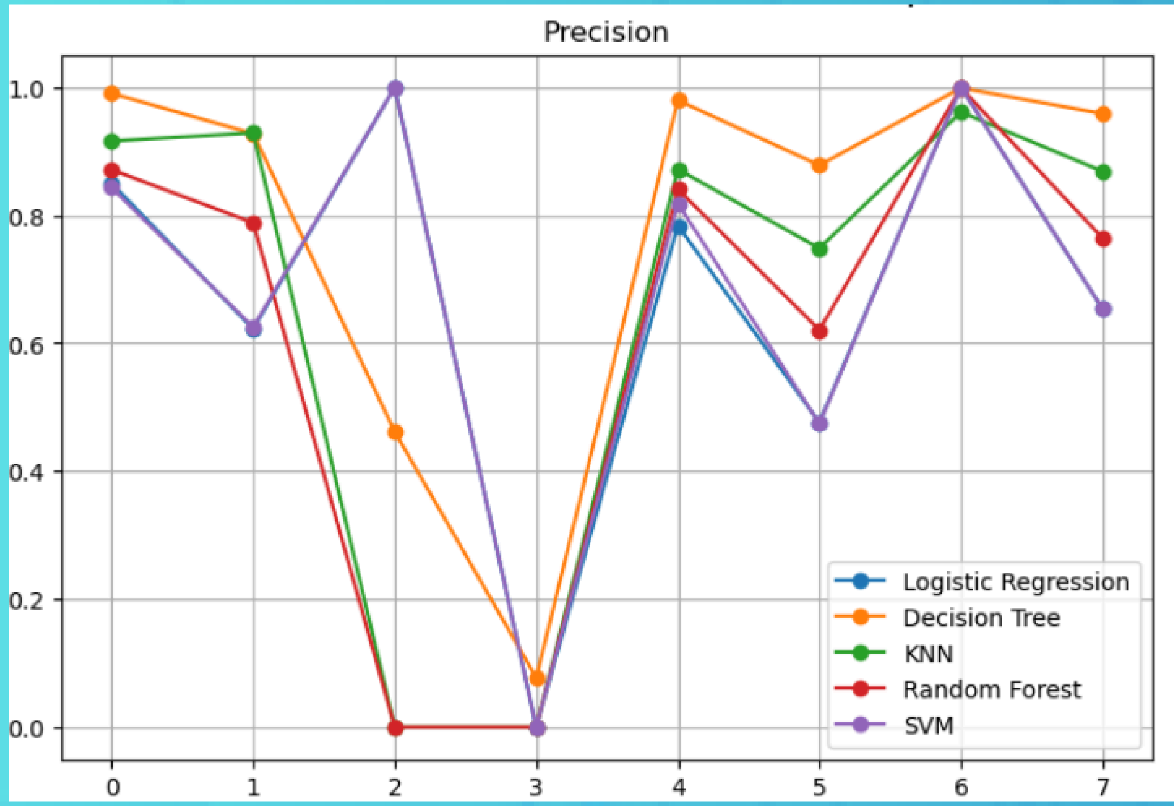


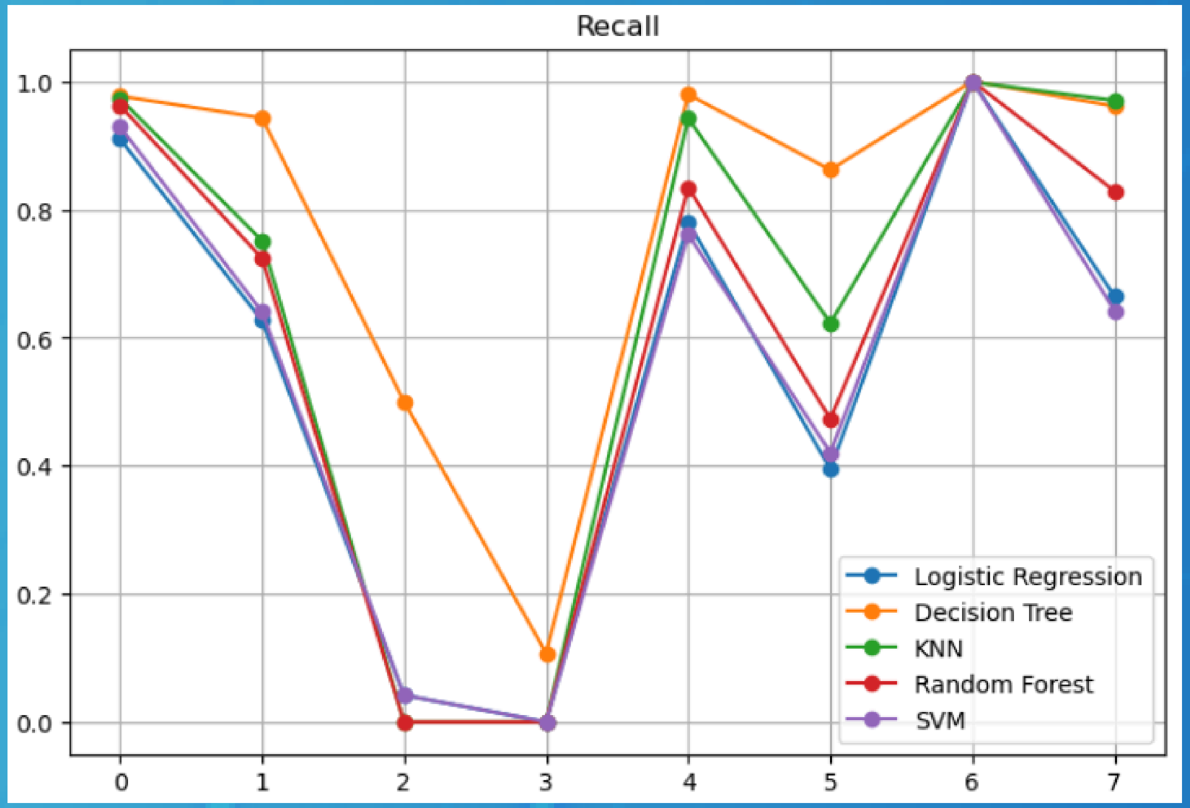
Nationality of the Players that got the highest Wages

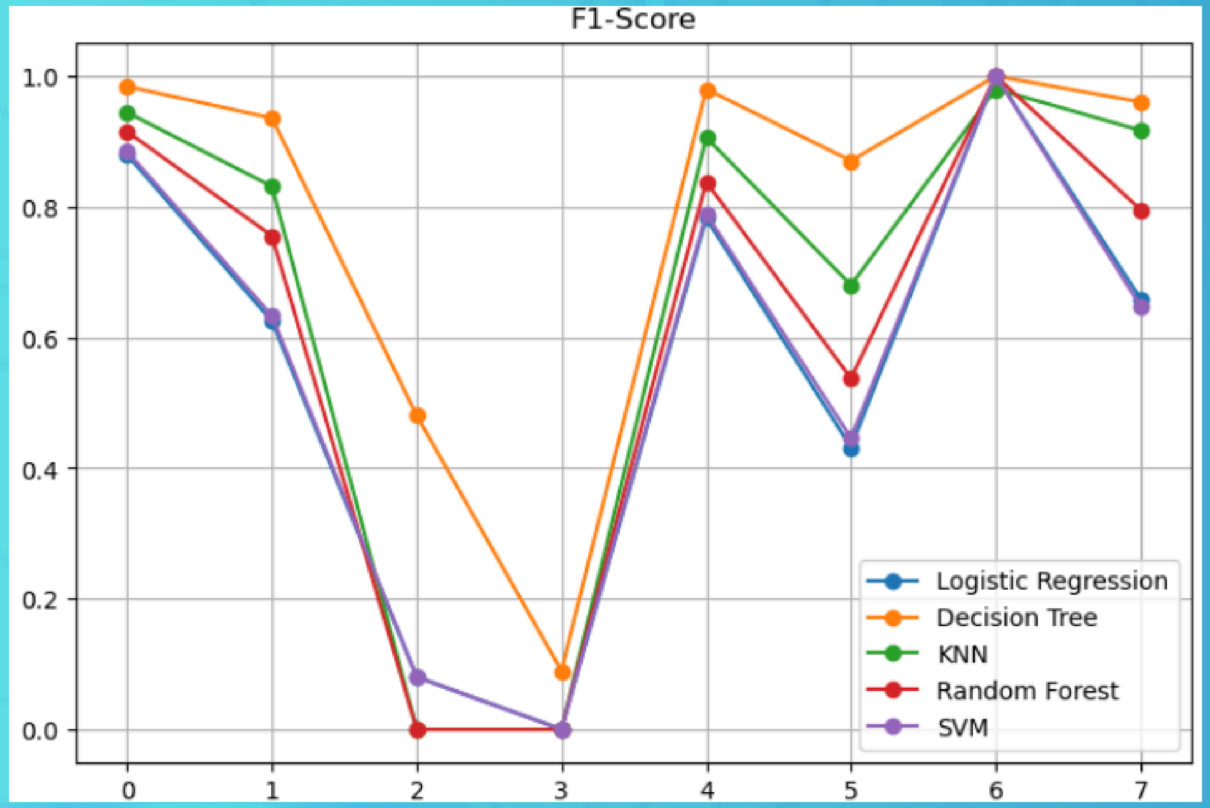


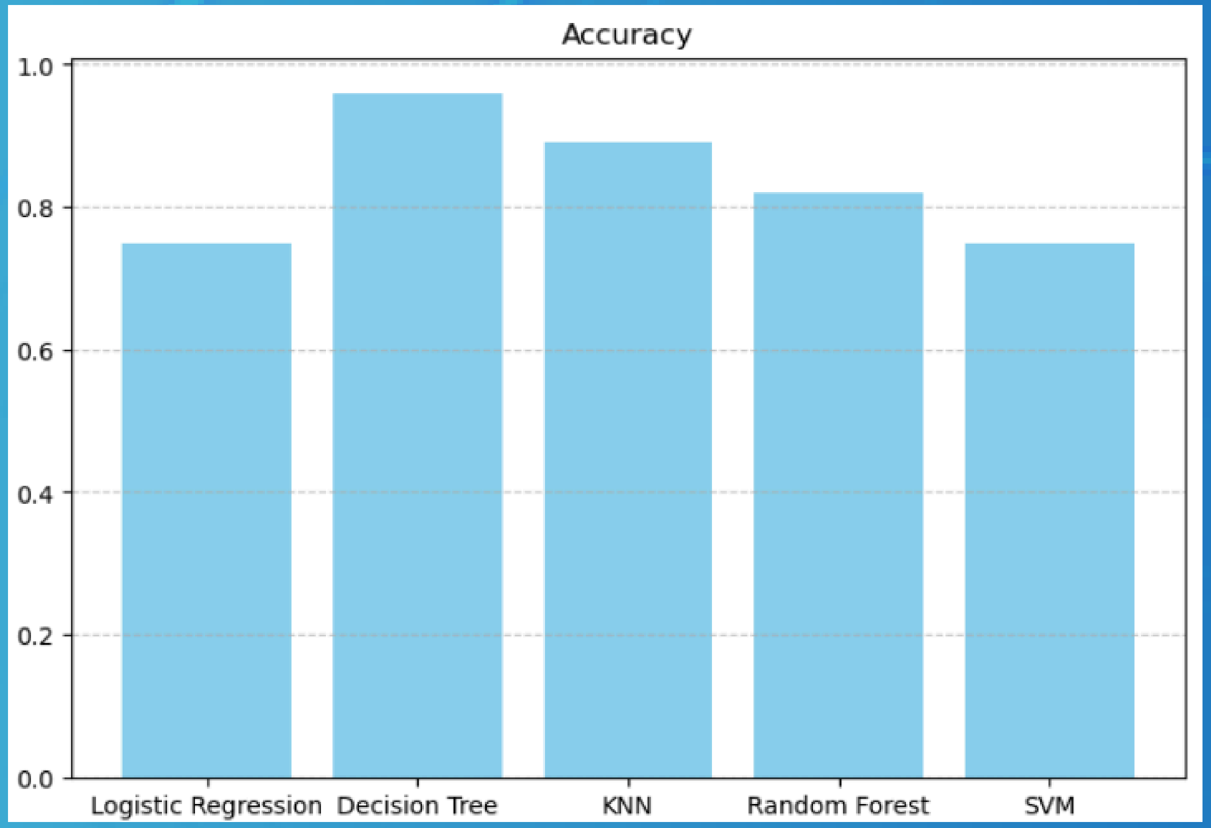
View the Top 50 Players and their Clubs

ML MODELS









FEATURE ANALYSIS

```
|--- defending_sliding_tackle_mmnorm <= 0.51
|   |--- goalkeeping_diving_mmnorm <= 0.42
|   |   |--- player_positions_code <= 207.50
|   |   |   |--- player_positions_code <= 173.50
|   |   |   |   |--- truncated branch of depth 20
|   |   |   |--- player_positions_code >  173.50
|   |   |   |   |--- truncated branch of depth 7
|   |   |--- player_positions_code >  207.50
|   |   |   |--- player_positions_code <= 672.00
|   |   |   |   |--- truncated branch of depth 14
|   |   |   |--- player_positions_code >  672.00
|   |   |   |   |--- truncated branch of depth 17
|   |--- goalkeeping_diving_mmnorm >  0.42
|   |   |--- class: 6
|--- defending_sliding_tackle_mmnorm >  0.51
|   |--- mentality_vision_mmnorm <= 0.53
|   |   |--- player_positions_code <= 93.00
|   |   |   |--- defending_standing_tackle_mmnorm <= 0.46
|   |   |   |   |--- class: 7
|   |   |   |--- defending_standing_tackle_mmnorm >  0.46
|   |   |   |   |--- class: 0
|   |   |--- player_positions_code >  93.00
|   |   |   |--- player_positions_code <= 103.50
|   |   |   |   |--- class: 1
|   |   |   |--- player_positions_code >  103.50
|   |   |   |   |--- truncated branch of depth 13
|   |--- mentality_vision_mmnorm >  0.53
|   |   |--- player_positions_code <= 306.50
|   |   |   |--- player_positions_code <= 125.00
|   |   |   |   |--- truncated branch of depth 16
|   |   |   |--- player_positions_code >  125.00
```

Imp Features with thresh hold values

Observations
Analytical Dashboard

- Age and Ball Control: As a player's age increases, their ball control tends to decrease.
- Height and Performance Attributes: Increased height is associated with a decrease in attributes such as stamina, dribbling, pace, and passing. Conversely, increased height correlates with improved heading ability.
- Overall Rating and Wage: Higher overall ratings are generally associated with higher player wages.
- Positions and Wages: Players in positions such as Left Midfielder (LM), Right Midfielder (RM), Right Back (RB), Left Wing Back (LWB), and Right Wing Back (RWB) tend to have lower wages.
- Positions and Performance Values: Players occupying positions like Left Back (LB), Right Back (RB), Left Wing Back (LWB), Right Wing Back (RWB), Center Forward (CF), and Right Winger (RW) generally exhibit lower performance values.
- Player Nationalities and Wages: Players from France, England, and Denmak typically command higher wages.
- Age and Potential: As a player's age increases, their potential tends to decrease.

Above observations are based on all 82 variables of the data including all clubs, all nationalities, all positions and cage from 16 to 45

Observations ML Model

- **Precision:**Decision Tree shows the highest precision overall.
- KNN and Random Forest also perform well but have some low precision in specific classes.
- Lowest Precision:Logistic Regression and SVM struggle with certain classes, particularly those with small support values.

- **Macro Average:** Decision Tree: Highest (0.79)
- KNN and Random Forest: Moderate (0.66 and 0.60 respectively)
- Logistic Regression and SVM: Lower (0.56 and 0.55)
- **Weighted Average:** Decision Tree: Highest (0.96)
- KNN and Random Forest: Moderate (0.89 and 0.81)
- Logistic Regression and SVM: Lower (0.74 and 0.75)

Accuracy

- Decision Tree: 0.96 (Highest accuracy)
- KNN Classifier: 0.89
- Random Forest: 0.82
- Logistic Regression: 0.75
- SVM: 0.75

Conclusion: The Decision Tree is the best-performing model among the ones compared, with consistent and high metrics across the board.

- **Recall:** Decision Tree consistently achieves high recall, particularly excelling across most classes.
- KNN also performs well but is less consistent than the Decision Tree.
- Lowest Recall: Logistic Regression, SVM, and Random Forest have low recall for some classes, particularly those with fewer instances.

- **F1-Score:** Decision Tree again leads with the highest F1-scores across most classes.
- KNN also shows solid F1-scores but with some variability.
- Lowest F1-Score:Logistic Regression, SVM, and Random Forest exhibit lower F1-scores in specific classes, especially where precision or recall are low.

Managerial Insights analytical dashboard

1. Player Positioning Insights: Display player positions across classes 0, 1, 4, 6, and 7 on the dashboard, based on Decision Tree insights, to quickly identify optimal player roles. Coaches can directly use this visualization to assign players to positions where they are likely to excel.
2. Performance-Based Wage Adjustments: Integrate a dashboard view showing wage alignment with performance metrics, particularly for positions like Left Back (LB) and Right Back (RB) that traditionally exhibit lower performance.HR and financial teams can use this to justify wage structures and make informed decisions on contract renewals.
3. Competitor Club Analysis: Implement a comparative dashboard that benchmarks player skills against those of competitor clubs, identifying areas where your team is ahead or needs improvement.Management can use this feature to develop targeted strategies to outperform competitors.
4. Wage Insights by Nationality:Provide a dashboard breakdown of average player wages by nationality, focusing on France, England, and Denmark, where wage expectations are higher. This enables better planning during contract negotiations and budget allocations for player acquisitions.
5. Age and Potential Trends: Show a trend analysis on the dashboard linking player age with potential, allowing easy identification of players nearing their peak or decline. This assists in planning long-term player development strategies and determining when to invest in younger talent.

Managerial Insights
ML Model & Feature analysis

1. Utilize the Decision Tree model for player position predictions because it provides the highest accuracy (0.96) and consistently strong performance across precision, recall, and F1-scores.
2. Prioritize training in defending and goalkeeping skills as these features (defending_sliding_tackle_mmnorm, goalkeeping_diving_mmnorm) are critical in determining player roles according to the Decision Tree analysis.
3. Assign taller players to heading-intensive roles and shorter, agile players to speed-reliant positions, leveraging height-related feature insights for optimal player performance.
4. Continuously update and refine the Decision Tree model with new data to maintain its effectiveness in player position predictions and adapting to evolving player dynamics.
5. Use the Decision Tree's high precision and recall to identify and enhance key skills in players that align with successful positions, maximizing their on-field effectiveness.

THANK YOU