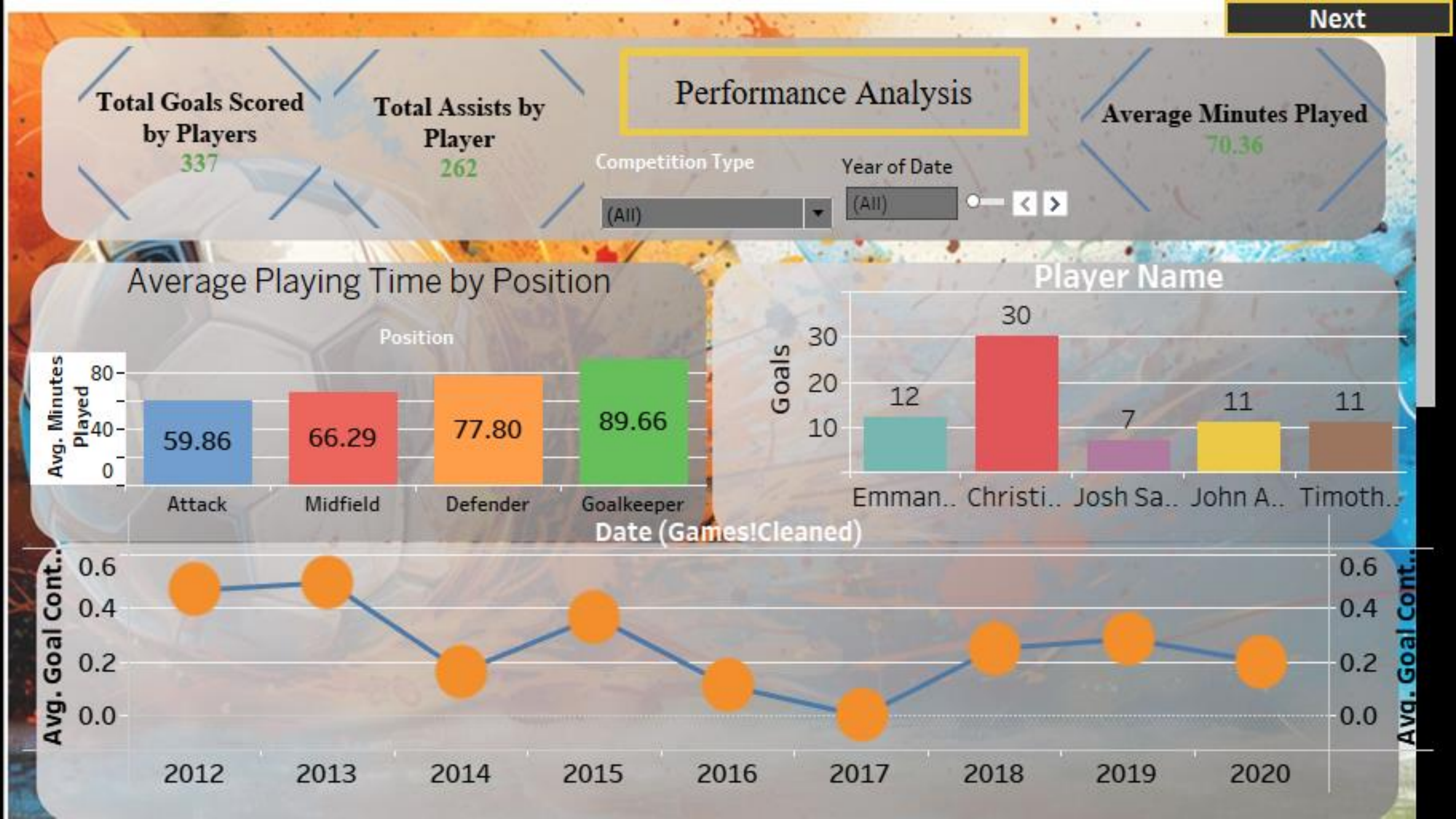
Comprehensive Football Analysis:
Insights & Strategic Recommendations

By

Kalaimani Muthu

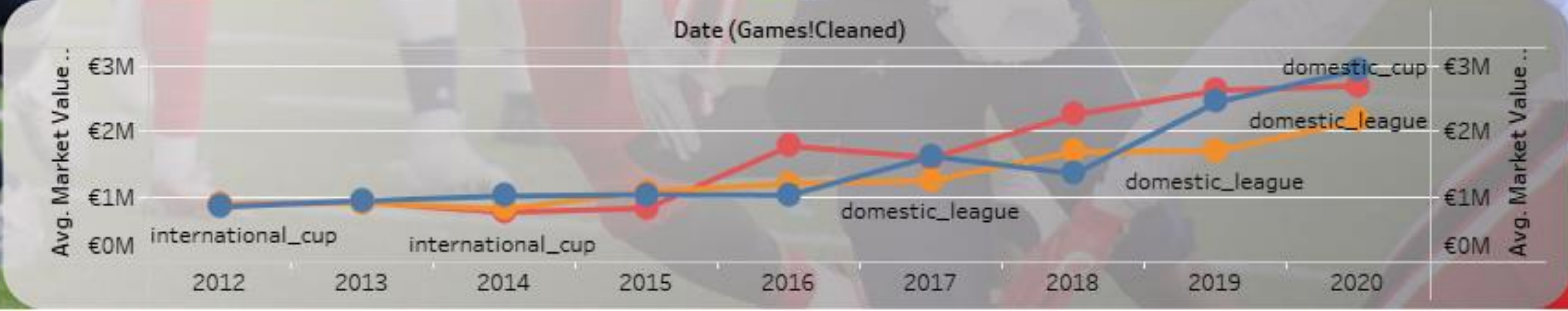

Next
Performance Analysis
Total Goals Scored by Players
337
Total Assists by Player
262
Competition Type
(All)
Year of Date
(All)
Average Minutes Played
70.36
Average Playing Time by Position
Position
Avg. Minutes Played
0
40
80
59.86
66.29
77.80
89.66
Attack
Midfield
Defender
Goalkeeper
Player Name
Goals
10
20
30
12
30
7
11
11
Emman..
Christi..
Josh Sa..
John A..
Timoth..
Date (Games!Cleaned)
Avg. Goal Cont..
0.0
0.2
0.4
0.6
2012
2013
2014
2015
2016
2017
2018
2019
2020

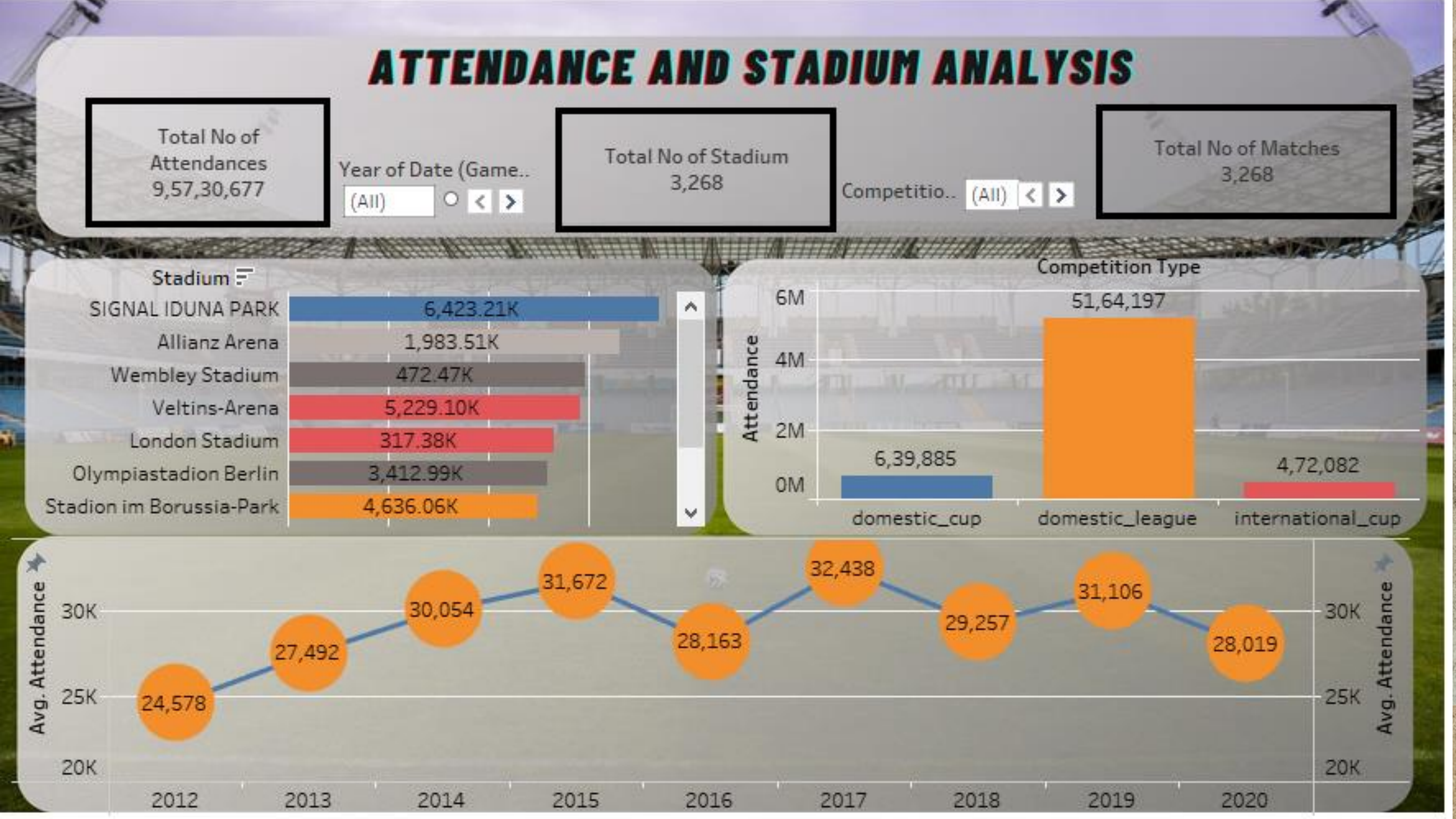
PLAYER PROFILE AND MARKET VALUE ANALYSIS

Position: Attack

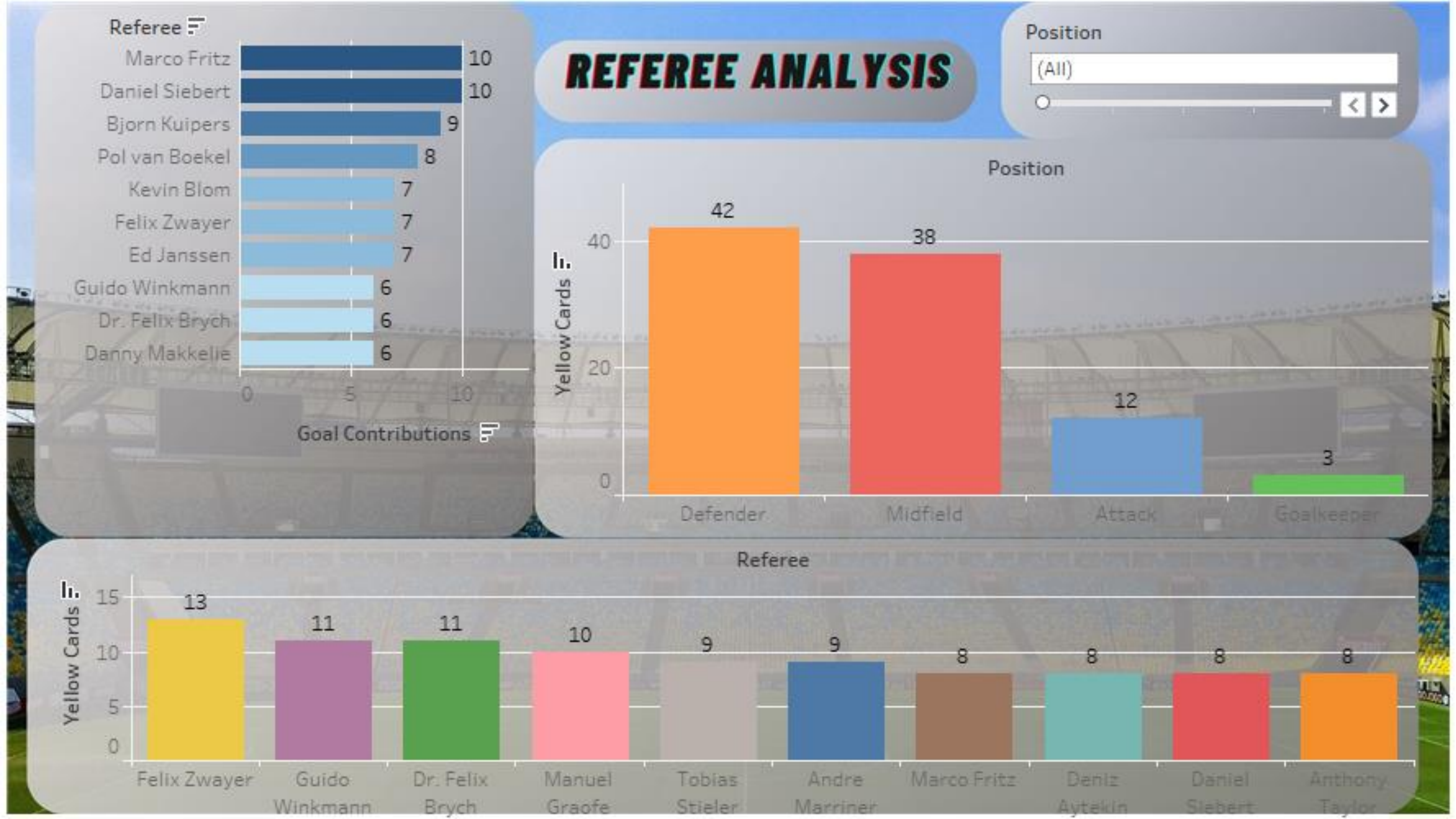
Country Of Birth: (All)

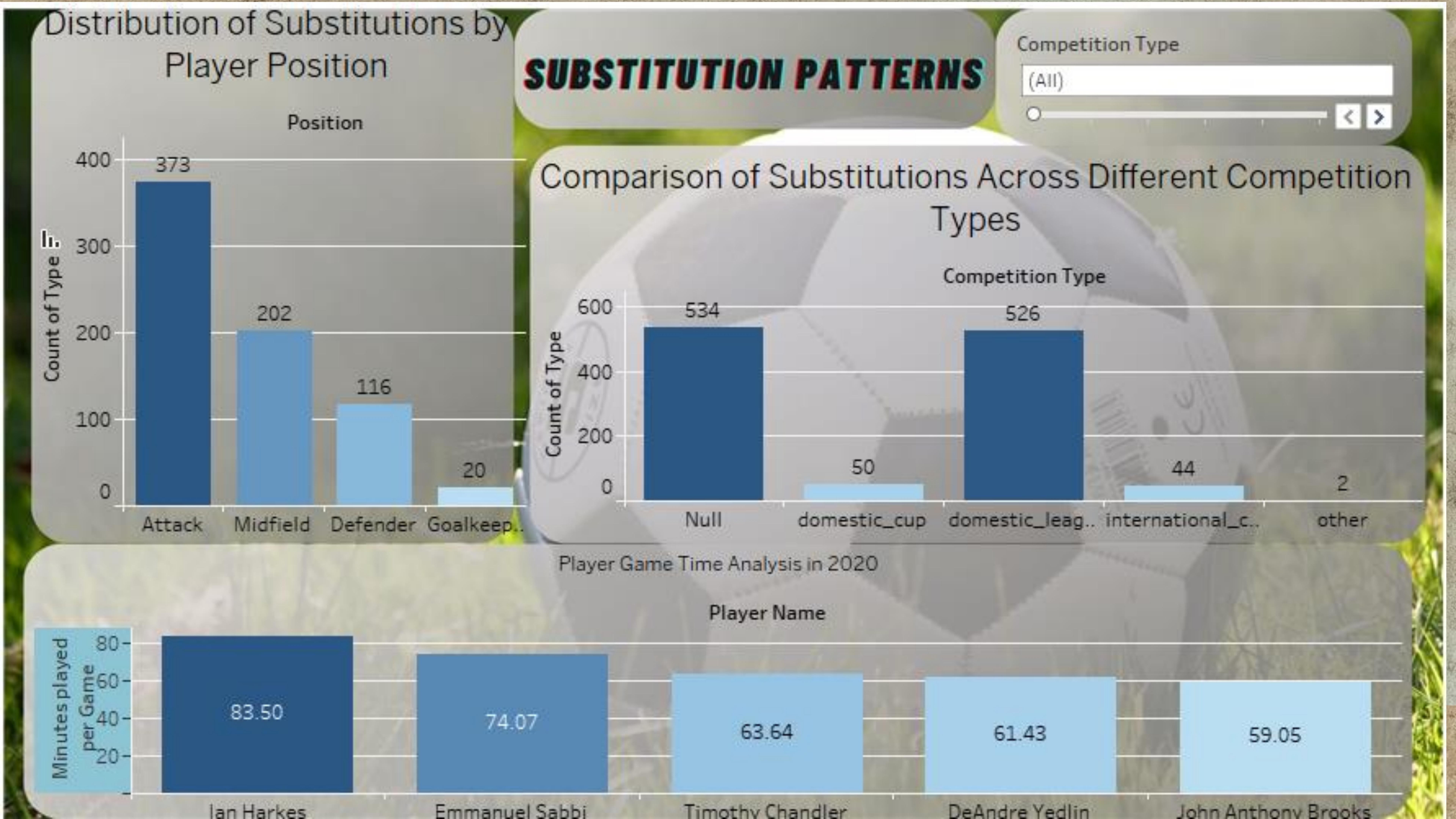
Year of Date: 2012





ATTENDANCE AND STADIUM ANALYSIS
Total No of Attendances
9,57,30,677
Year of Date (Game..
(All)
Total No of Stadium
3,268
Competitio..
(All)
Total No of Matches
3,268
Stadium
SIGNAL IDUNA PARK
6,423.21K
Allianz Arena
1,983.51K
Wembley Stadium
472.47K
Veltins-Arena
5,229.10K
London Stadium
317.38K
Olympiastadion Berlin
3,412.99K
Stadion im Borussia-Park
4,636.06K
Competition Type
Attendance
6M
4M
2M
0M
6,39,885
51,64,197
4,72,082
domestic_cup
domestic_league
international_cup
Avg. Attendance
30K
25K
20K
24,578
27,492
30,054
31,672
28,163
32,438
29,257
31,106
28,019
2012
2013
2014
2015
2016
2017
2018
2019
2020

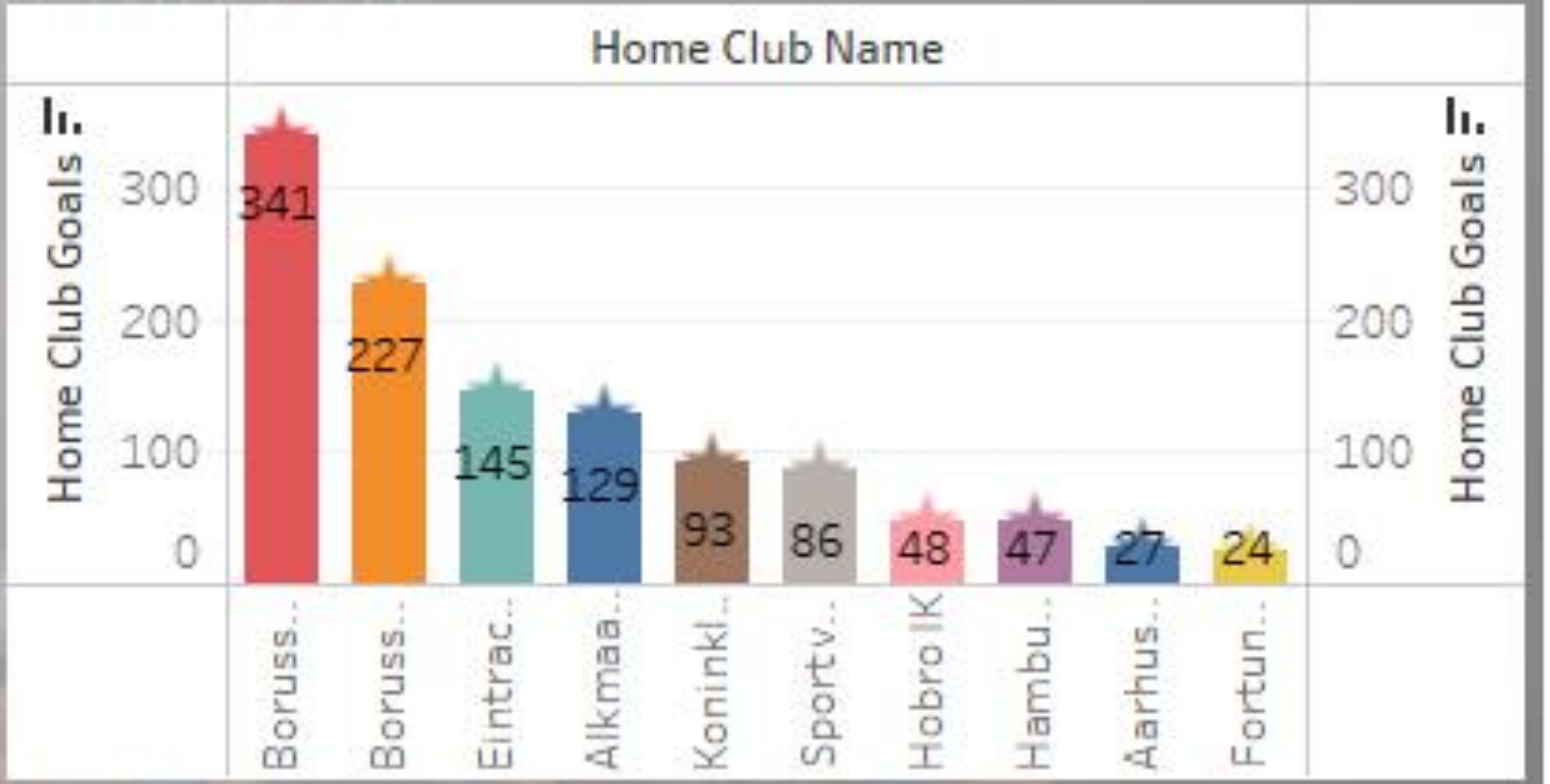

REFEREE ANALYSIS
Position
(All)
Referee
Marco Fritz 10
Daniel Siebert 10
Bjorn Kuipers 9
Pol van Boekel 8
Kevin Blom 7
Felix Zwayer 7
Ed Janssen 7
Guido Winkmann 6
Dr. Felix Brych 6
Danny Makkelie 6
0
5
10
Goal Contributions
Position
Yellow Cards
0
20
40
42
38
12
3
Defender
Midfield
Attack
Goalkeeper
Referee
Yellow Cards
0
5
10
15
13
11
11
10
9
9
8
8
8
8
Felix Zwayer
Guido Winkmann
Dr. Felix Brych
Manuel Graofe
Tobias Stieler
Andre Marriner
Marco Fritz
Deniz Aytekin
Daniel Siebert
Anthony Taylor



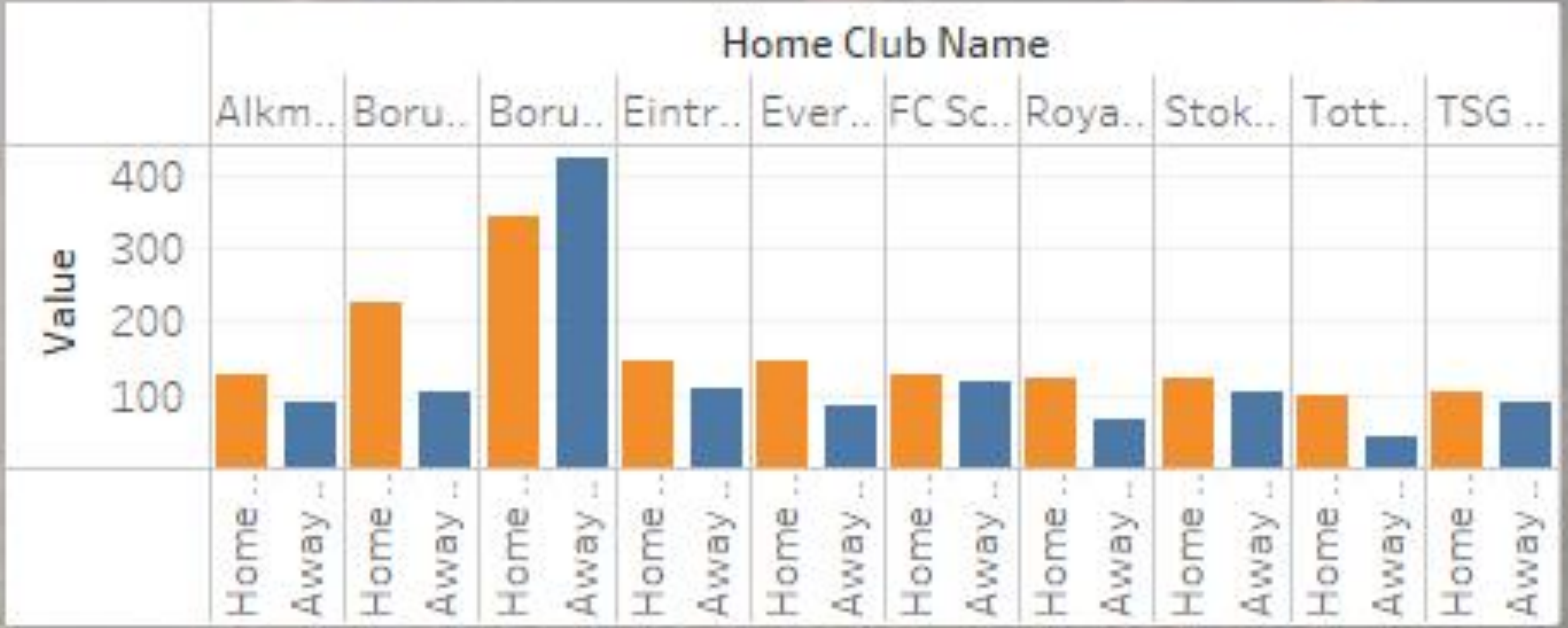
SUBSTITUTION PATTERNS
Distribution of Substitutions by Player Position
Position
Count of Type
400
300
200
100
0
373
202
116
20
Attack
Midfield
Defender
Goalkeep..
Competition Type
(All)
Comparison of Substitutions Across Different Competition Types
Competition Type
Count of Type
600
400
200
0
534
50
526
44
2
Null
domestic_cup
domestic_leag..
international_c..
other
Player Game Time Analysis in 2020
Player Name
Minutes played per Game
80
60
40
20
83.50
74.07
63.64
61.43
59.05
Ian Harkes
Emmanuel Sabbi
Timothy Chandler
DeAndre Yedlin
John Anthony Brooks

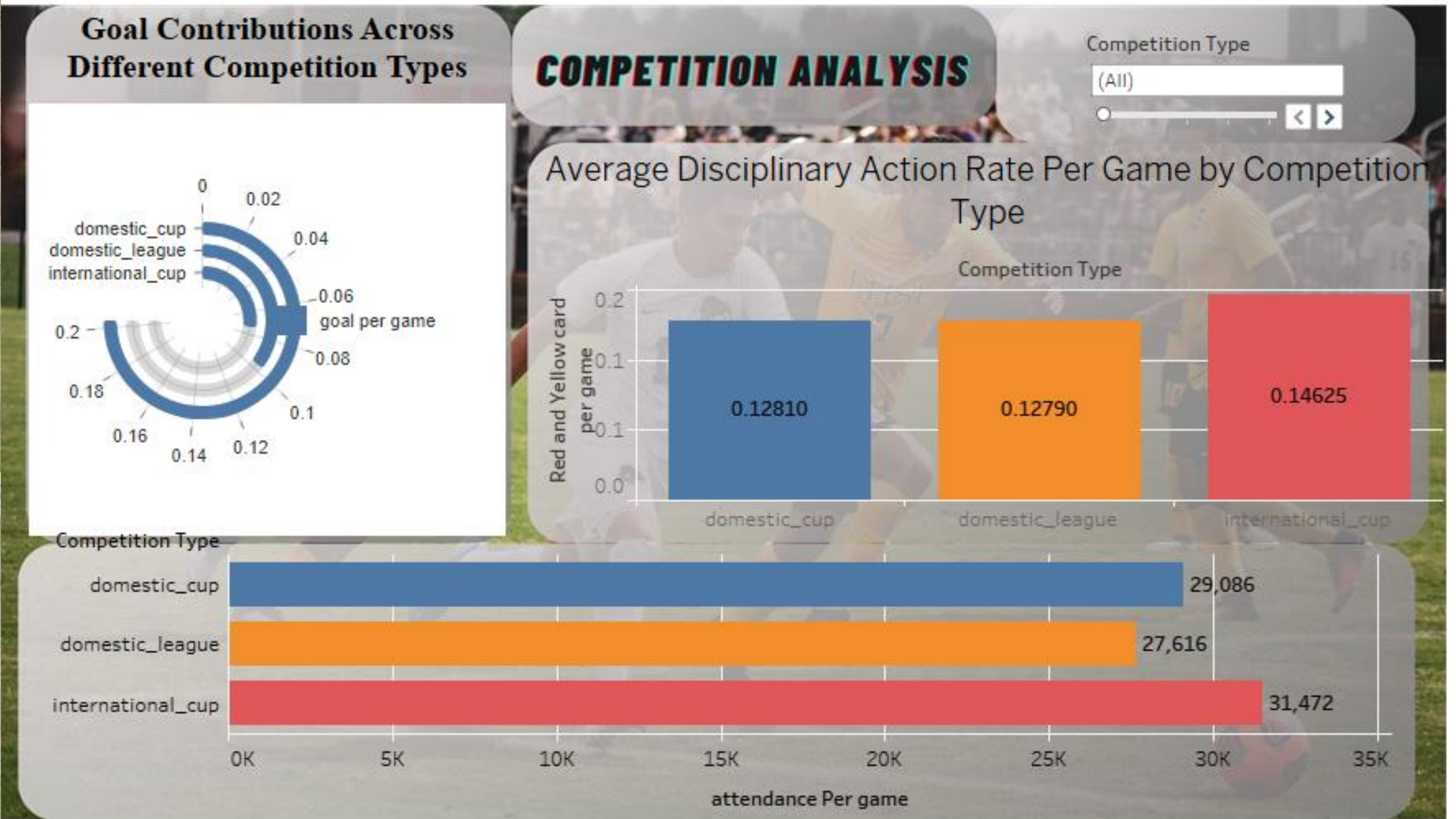
Home Game Performance Under Different Managers

Goal Scoring Proficiency at Home Matches

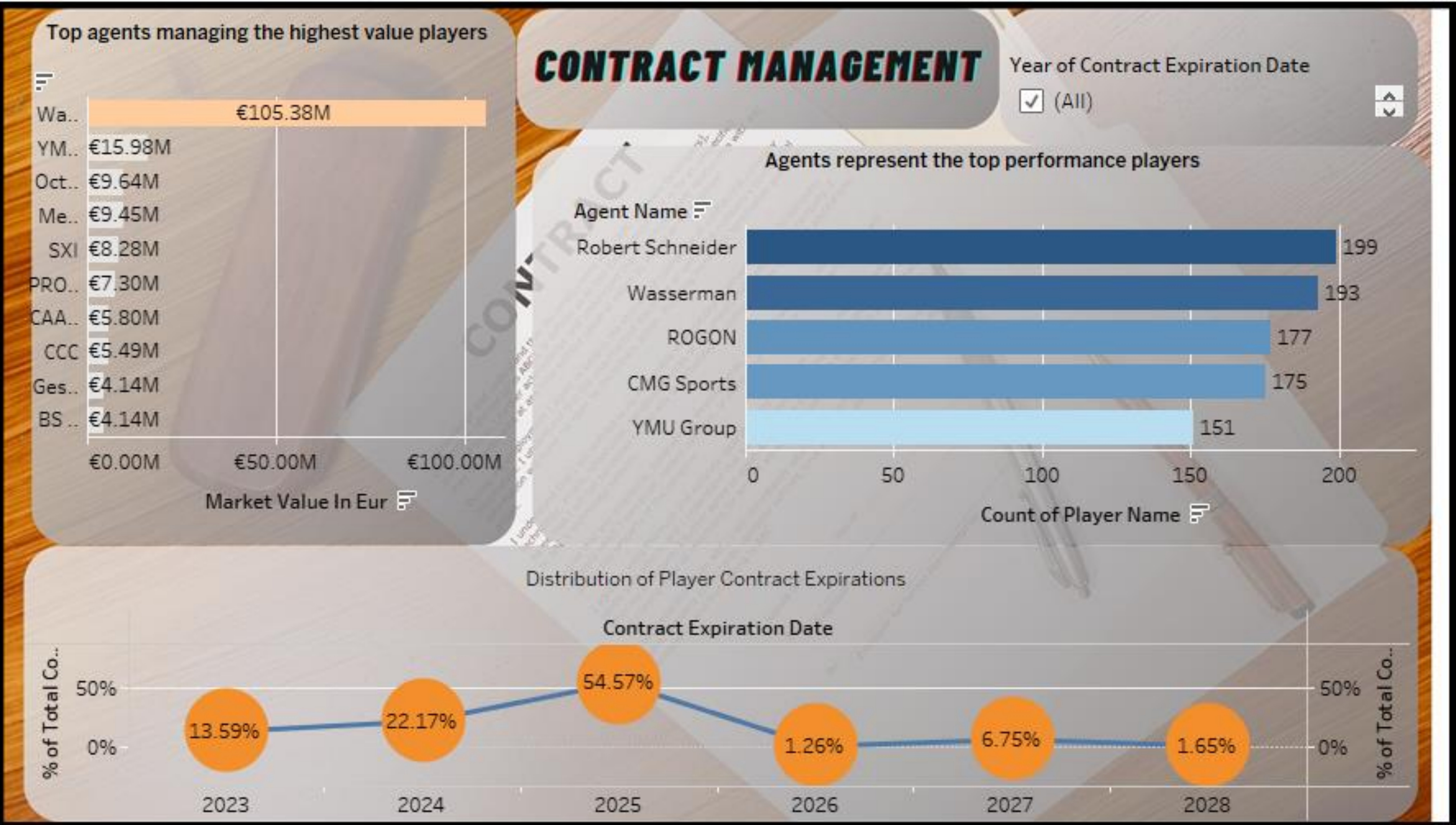


Goal Scoring Proficiency at Qway Matches





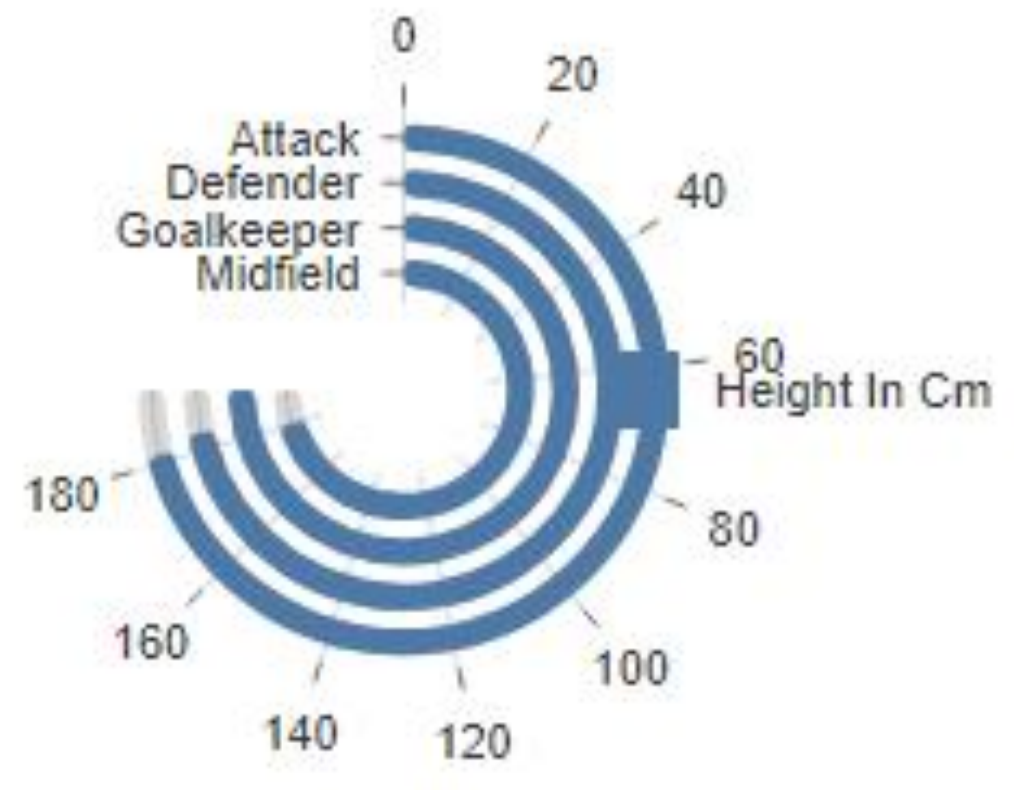
Goal Contributions Across Different Competition Types
COMPETITION ANALYSIS
Competition Type
(All)
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2
domestic_cup
domestic_league
international_cup
goal per game
Average Disciplinary Action Rate Per Game by Competition Type
Competition Type
Red and Yellow card per game
0.2
0.1
0.1
0.0
0.12810
0.12790
0.14625
domestic_cup
domestic_league
international_cup
Competition Type
domestic_cup
domestic_league
international_cup
29,086
27,616
31,472
0K
5K
10K
15K
20K
25K
30K
35K
attendance Per game

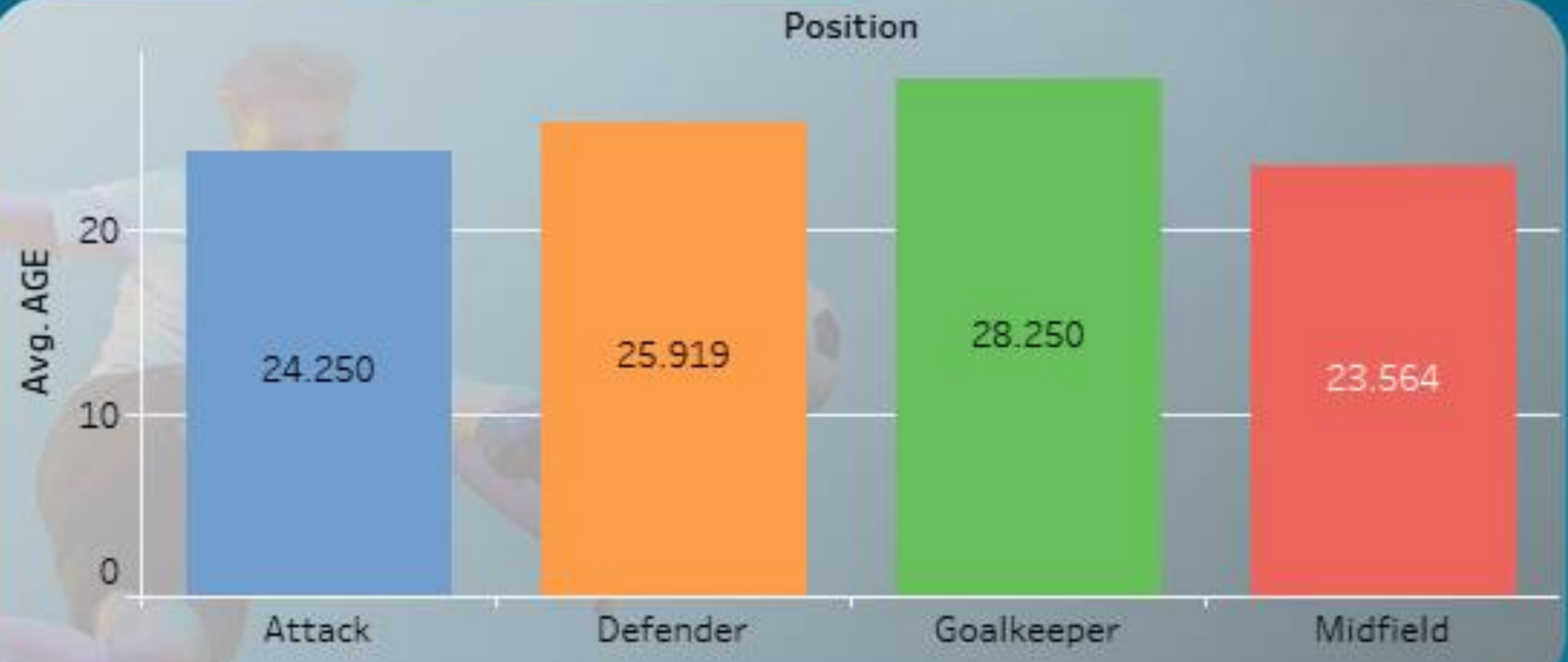

CONTRACT MANAGEMENT
Top agents managing the highest value players
Wa.. €105.38M
YM.. €15.98M
Oct.. €9.64M
Me.. €9.45M
SXI €8.28M
PRO.. €7.30M
CAA.. €5.80M
CCC €5.49M
Ges.. €4.14M
BS .. €4.14M
€0.00M
€50.00M
€100.00M
Market Value In Eur
Year of Contract Expiration Date
(All)
Agents represent the top performance players
Agent Name
Robert Schneider 199
Wasserman 193
ROGON 177
CMG Sports 175
YMU Group 151
0
50
100
150
200
Count of Player Name
Distribution of Player Contract Expirations
Contract Expiration Date
% of Total Co..
50%
0%
13.59%
22.17%
54.57%
1.26%
6.75%
1.65%
2023
2024
2025
2026
2027
2028

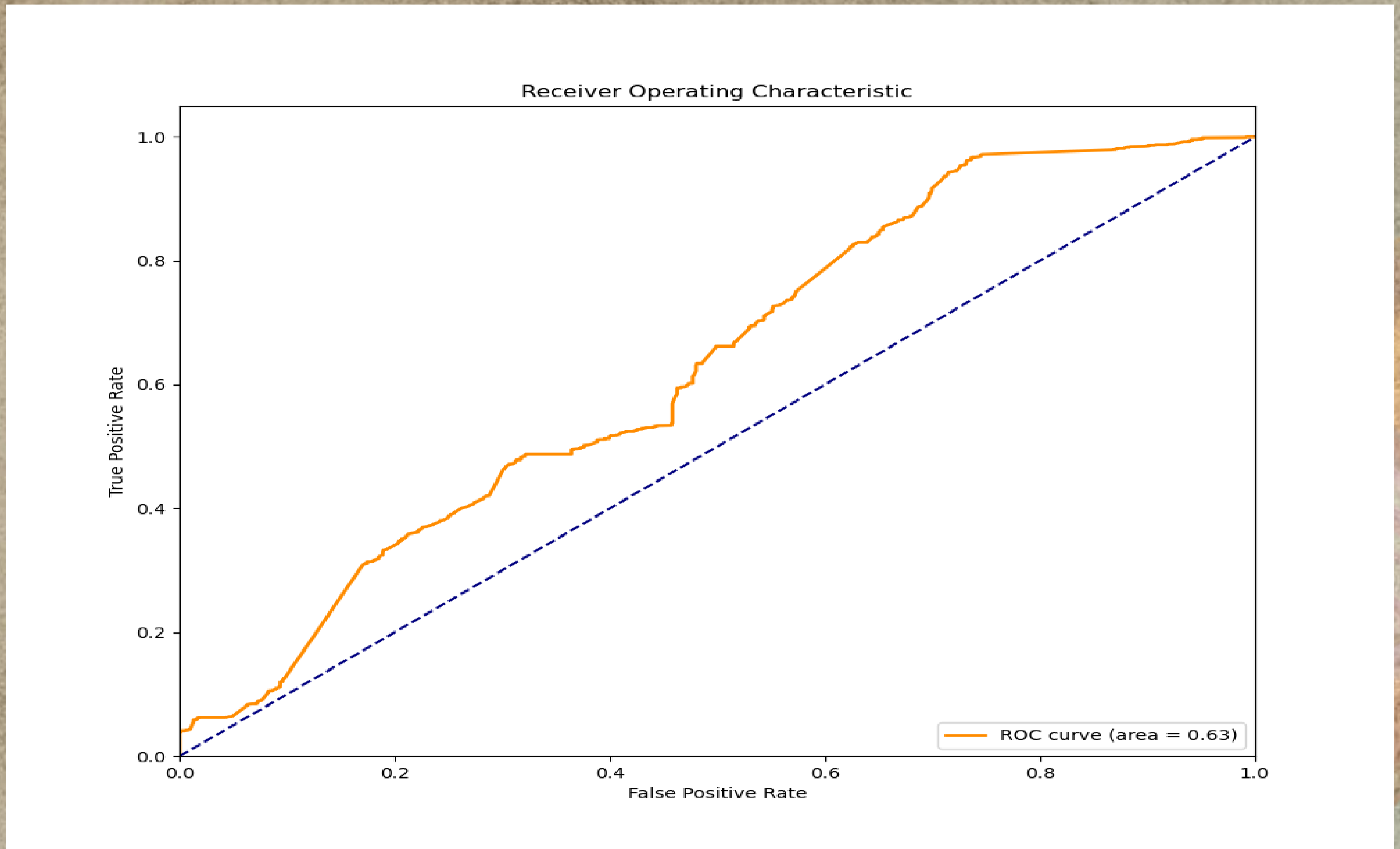
PLAYER ATTRIBUTES AND DEMOGRAPHICS

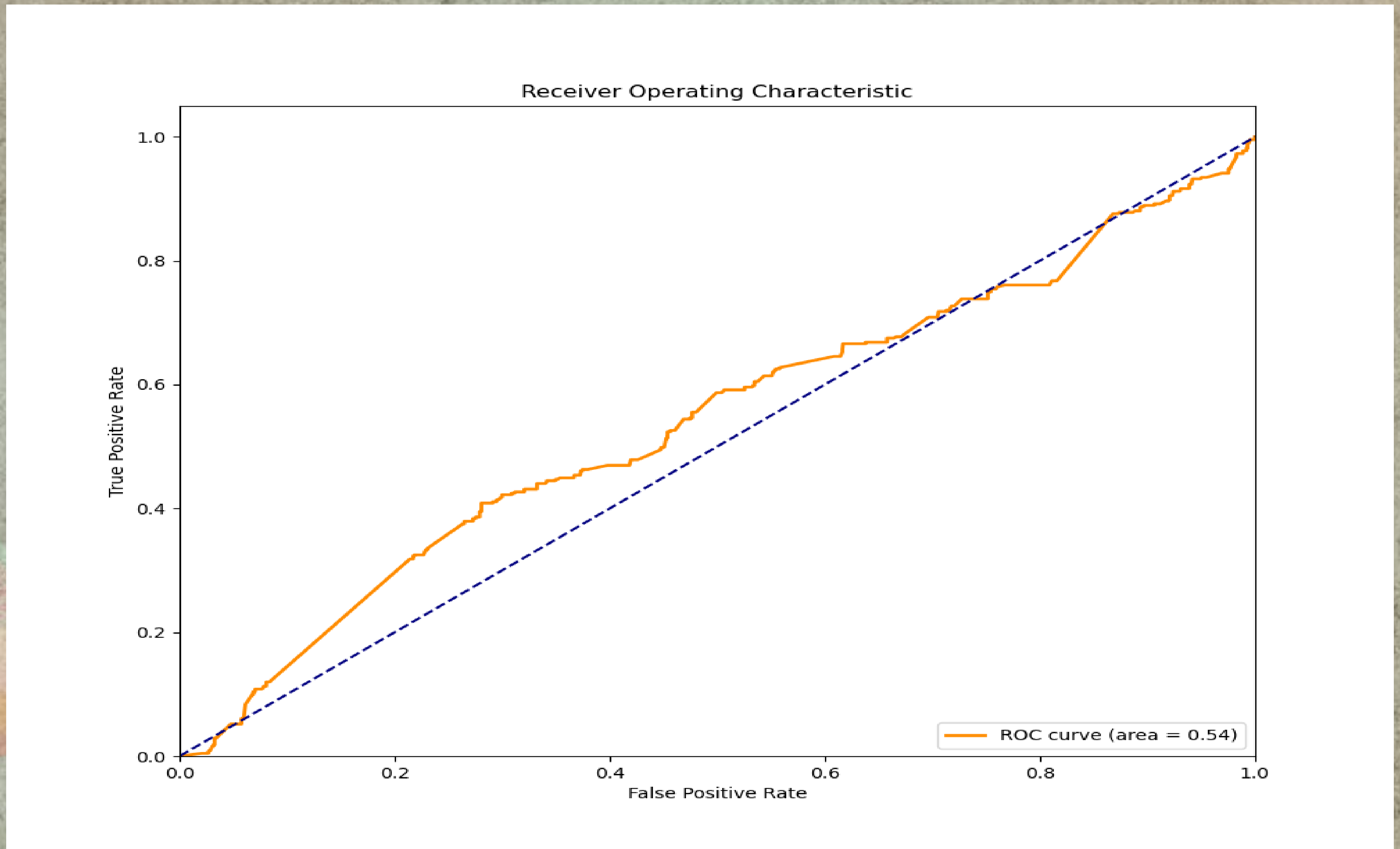
Country Of Birth

(All)









Model Analyses Summary

Logistic Regression (Predicting Starting Lineup from Goals Scored)

- **ROC Curve (Chart 1): Shows moderate predictive ability with an AUC of 0.627.**

Model Analyses Summary

Logistic Regression (Predicting Yellow Card from Previous Cards)

- **ROC Curve (Chart 2): Slight improvement in prediction with AUC of 0.642.**

Simple Linear Regression (Predicting Market Value from Scoring Frequency)

Model Performance:

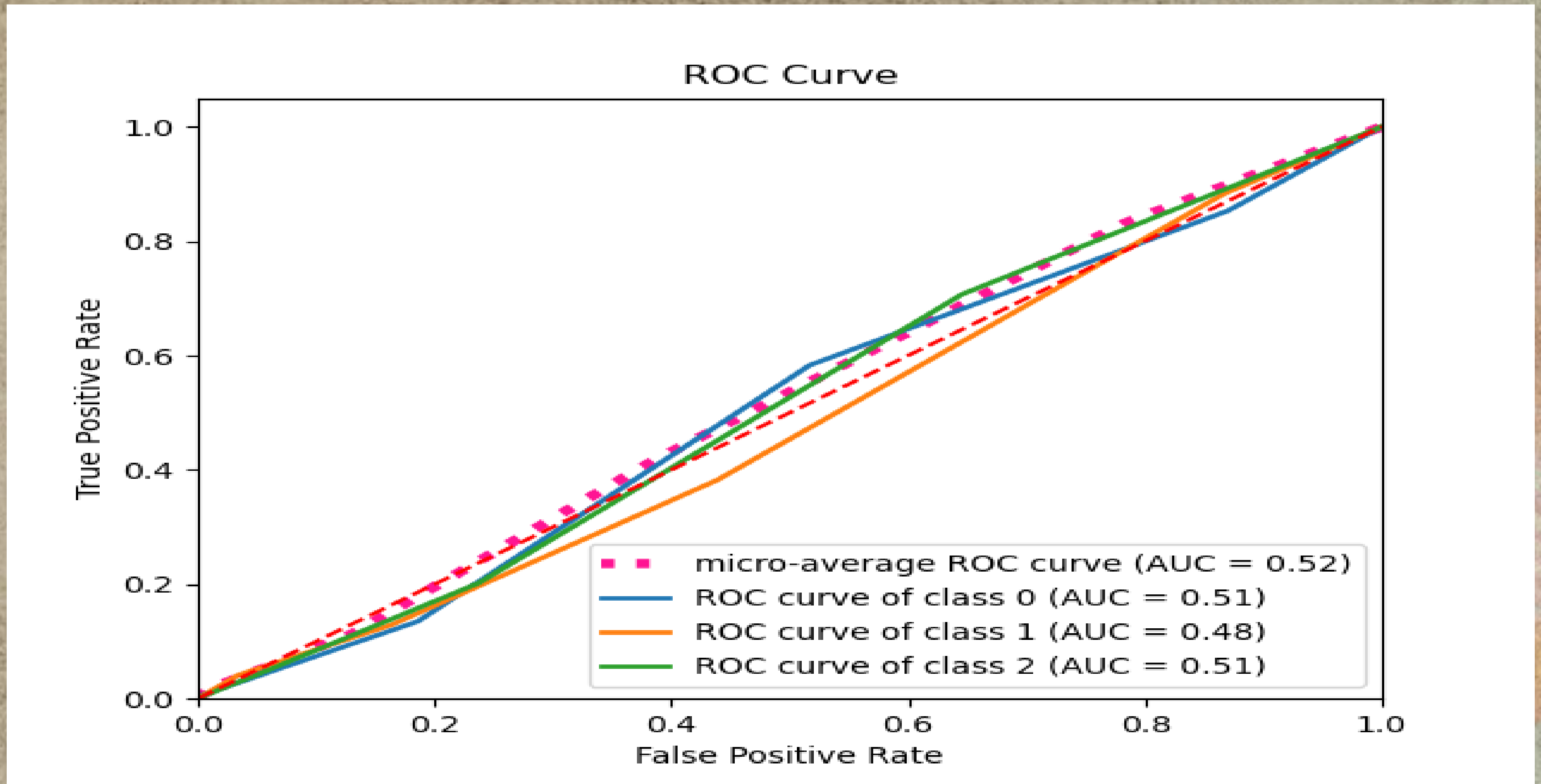
- Intercept: 0.0083
- Coefficient for Scoring Frequency: 0.02749
- R^2 Value: 0.00224 (Very low, indicating that scoring frequency alone does not explain variations in market value effectively)
- Errors: Mean Squared Error of 0.957, Mean Absolute Error of 0.779, Root Mean Squared Error of 0.978

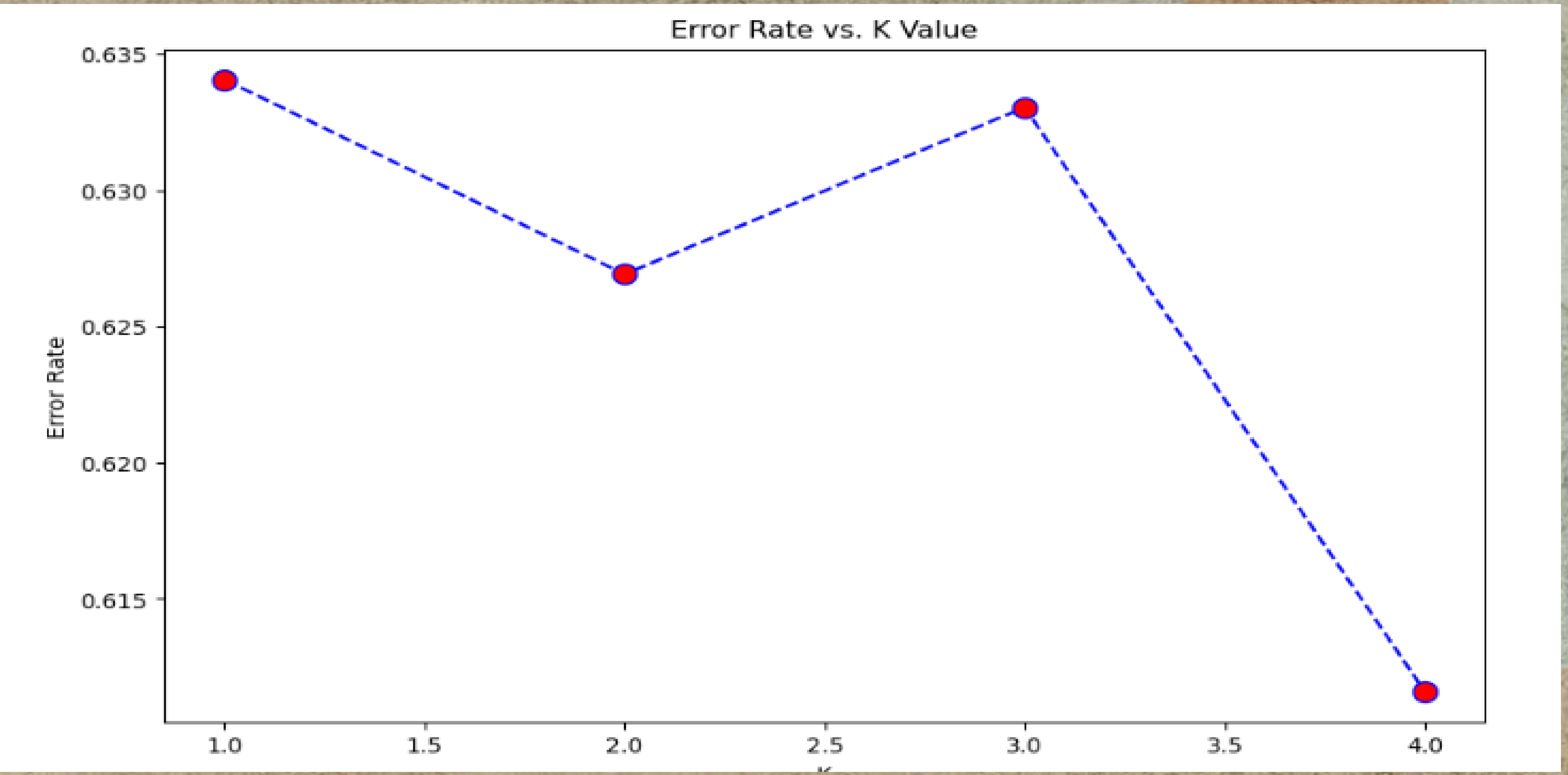
Multiple Linear Regression (Predicting Market Value from Various Features)

Model Performance:

- Intercept: 0.0032
- Coefficients: Height (-0.00747), Minutes Played (0.11474), Goals (-0.16384), Other feature (0.01855)
- R^2 Value: 0.04407 (Low, suggesting that these features together explain only a small portion of the variance in market value)
- Errors: Mean Squared Error of 0.927, Mean Absolute Error of 0.748, Root Mean Squared Error of 0.963

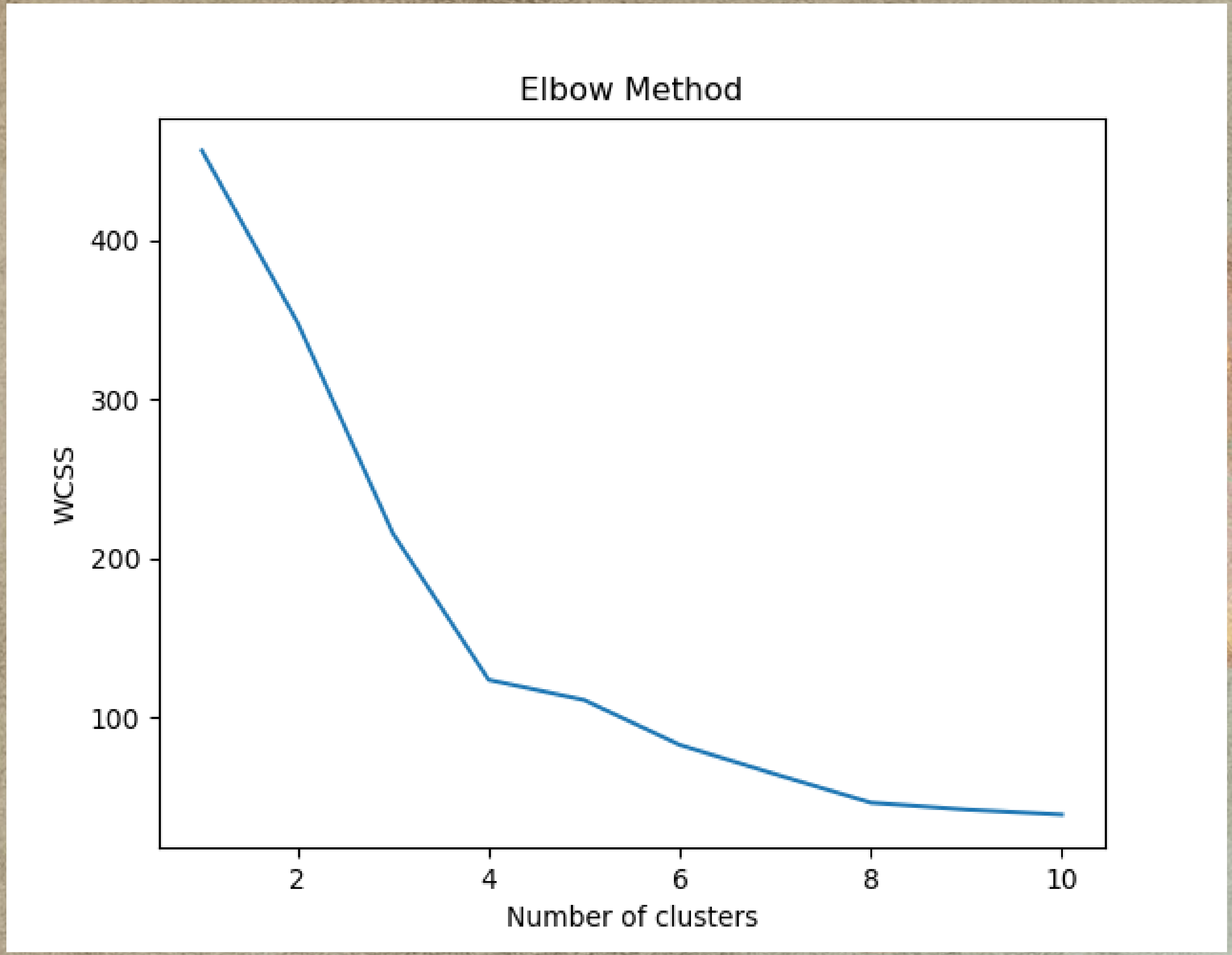
KNN Classification (Attendance Categories)

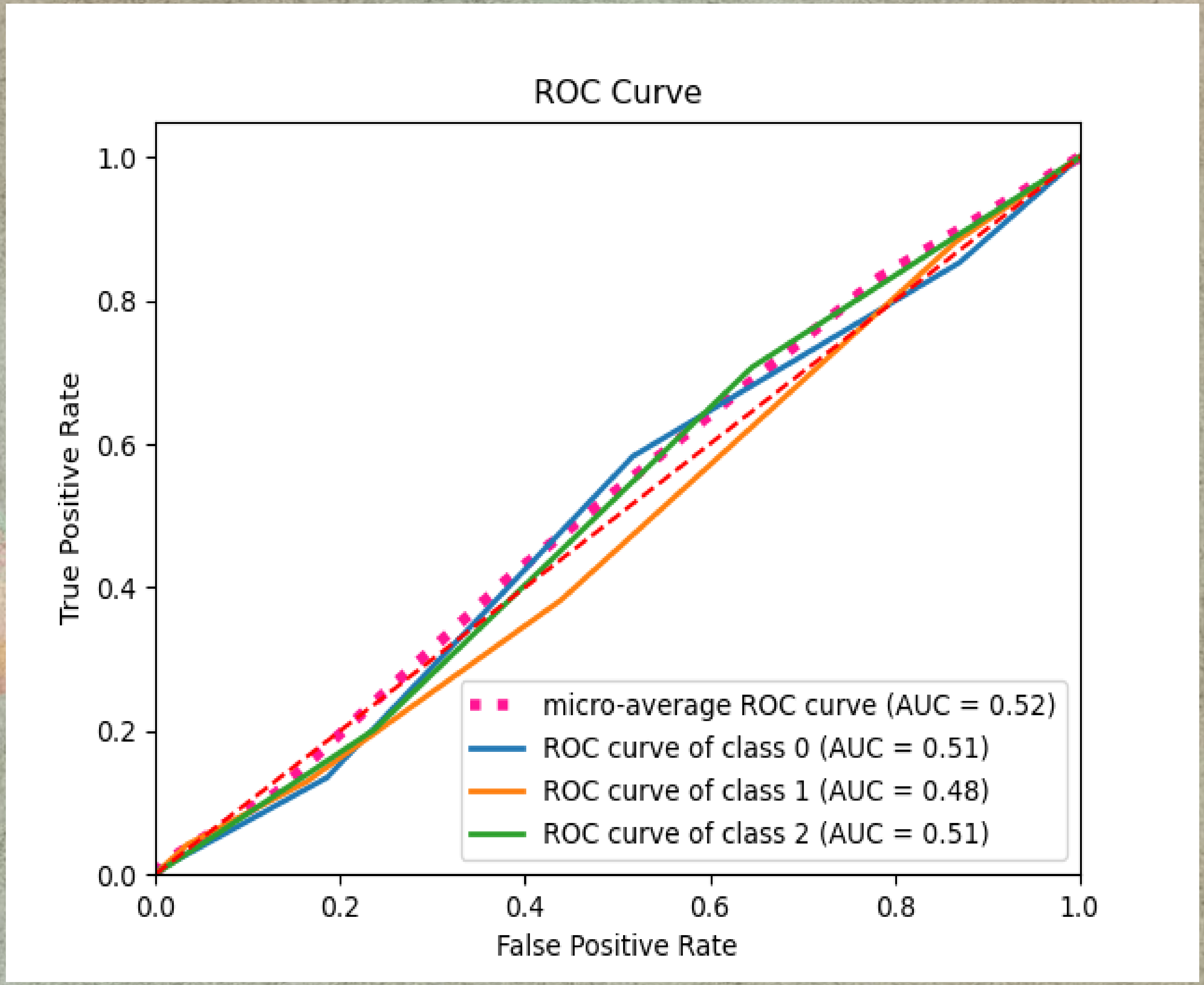




Minimum error: 0.6388888888888888 at K = 2
Optimal K = 2
Classifier n_neighbors: 2
Sample predictions: [0. 2. 1. 0. 0. 1. 0. 2. 1. 1.]
Accuracy: 0.3611111111111111
Recall: 0.3611111111111111
F1 Score: 0.3611111111111111
Precision: 0.3611111111111111
classification_report:

	precision	recall	f1-score	support
0.0	0.37	0.64	0.47	163
1.0	0.39	0.26	0.31	170
2.0	0.18	0.06	0.09	99
accuracy			0.36	432
macro avg	0.31	0.32	0.29	432
weighted avg	0.33	0.36	0.32	432





Kmean

MSE: 0.92696700261963999

RMSE: 0.9627912559945899

R^2 score: 0.04407255379719721

The average silhouette score is : 0.4727793028522318

Key Findings and Insights

- **Performance Analysis**:
  - Predictive accuracy for starting lineup based on goals scored is modest (AUC = 0.63).
- **Attendance Analysis**:
  - Classification accuracy for predicting match attendance categories is low (38.84%).
- **Player Demographics**:
  - Physical attributes such as height and footedness reveal distinct player groups, aiding in personalized training strategies.

Recommendations Table:

Focus Area	Current Metric	Target Improvement	Expected Benefit
Model Prediction Accuracy	AUC: 0.63	+59% to AUC: 1.00	Significantly enhanced predictive reliability
Attendance Prediction	Accuracy: 38.84%	+61.16% to 100%	Highly targeted marketing and operational planning
Player Cluster Utilization	Silhouette Score: Low	Improve score by 50%	Refined player role identification and team dynamics

Business Conclusions

- Average playing time varies by position, with goalkeepers at 89.66 minutes and attackers at 59.77 minutes.
- Market value insights reveal midfielders at the highest (€86.29M) and goalkeepers at the lowest (€23.85M).
- Total attendances reached over 9.5 million across all matches analyzed.

- **Strategic Recommendations:**
  - Utilize position-specific training regimens based on playing time insights.
  - Reassess market valuation frameworks, especially for goalkeepers.

Machine Learning Models:

1. **Logistic Regression:**
   1. **Starting Lineup Prediction:** Accuracy was 64%, with precision varying significantly between predicting starters (63%) and non-starters (94%).
   2. **Yellow Card Prediction:** High precision (87%) but extremely low sensitivity for predicting actual yellow card occurrences.
   3. **Recommendations:** Integrate more game-specific data points (like opposition and match stakes) to enhance predictive accuracy.
2. **Linear Regression:**
   1. **Market Value from Scoring Frequency:** Extremely low explanatory power ($R^2 = 0.002$).
   2. **Market Value from Multiple Features:** Slightly better with an R^2 of 0.044.
   3. **Recommendations:** Broaden the feature set to include non-performance metrics like media presence and sponsorship value.
3. **KNN Classification:**
   1. **Attendance Categorization:** Low overall accuracy (36.1%) with the best performance at K=2.
   2. **Recommendations:** Experiment with other classification methods that might capture complex patterns more effectively, such as Random Forests or Gradient Boosting Machines.
4. **K-Means Clustering:**
   1. **Player Grouping:** Identified clusters based on physical attributes but lacked strong distinctiveness.

Thank You