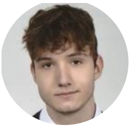
x x Capstone Project 2025

1.FC Köln
Transfer Predictions

Meet our Team

Kevin Drost

Rishvanyas Premanand



Balint Farkas

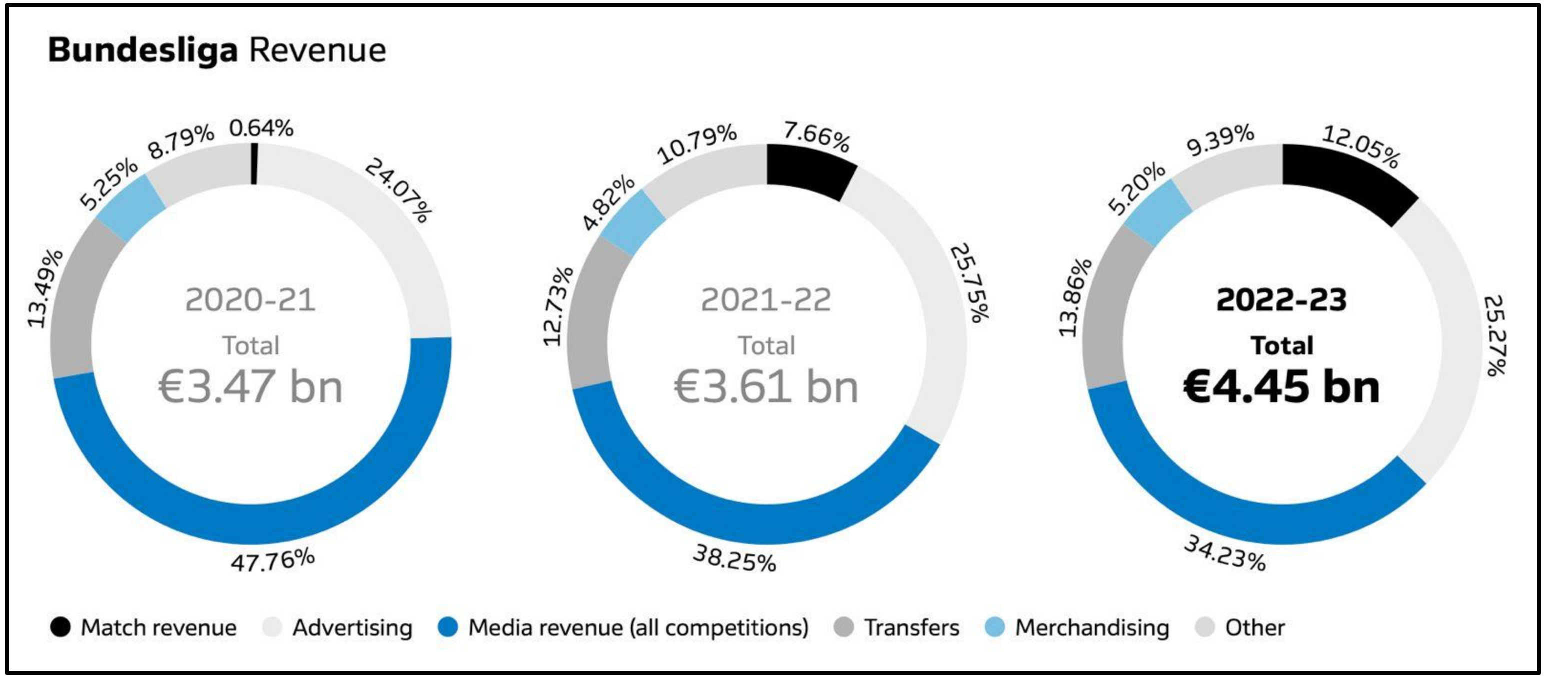
Laurenz Flender

1.FC Köln – Historic Club in Turbulent Times

2023/24 Season:

- 140 thousand members
- €159M in revenue and €11.8M profit
- The equity ratio climbed to 28% (mid-range among Bundesliga)
- Transfer Ban

Bundesliga Revenue



Source: The 2024 Economic Report (DFL)

Potential Benefits - Winning with Data

Make scouting more efficient and reliable

Expand to global markets

Create more transfer revenue

Good transfers can lead to better performances of the team

Motivation and Project Goal

Successful transfers are key to **team performance** and **financial sustainability**

Predict if a **new signing** will succeed in **their first season**

A player is successful if they **play at least 50%** of all possible minutes in the season

Our Precise Goal of the Project

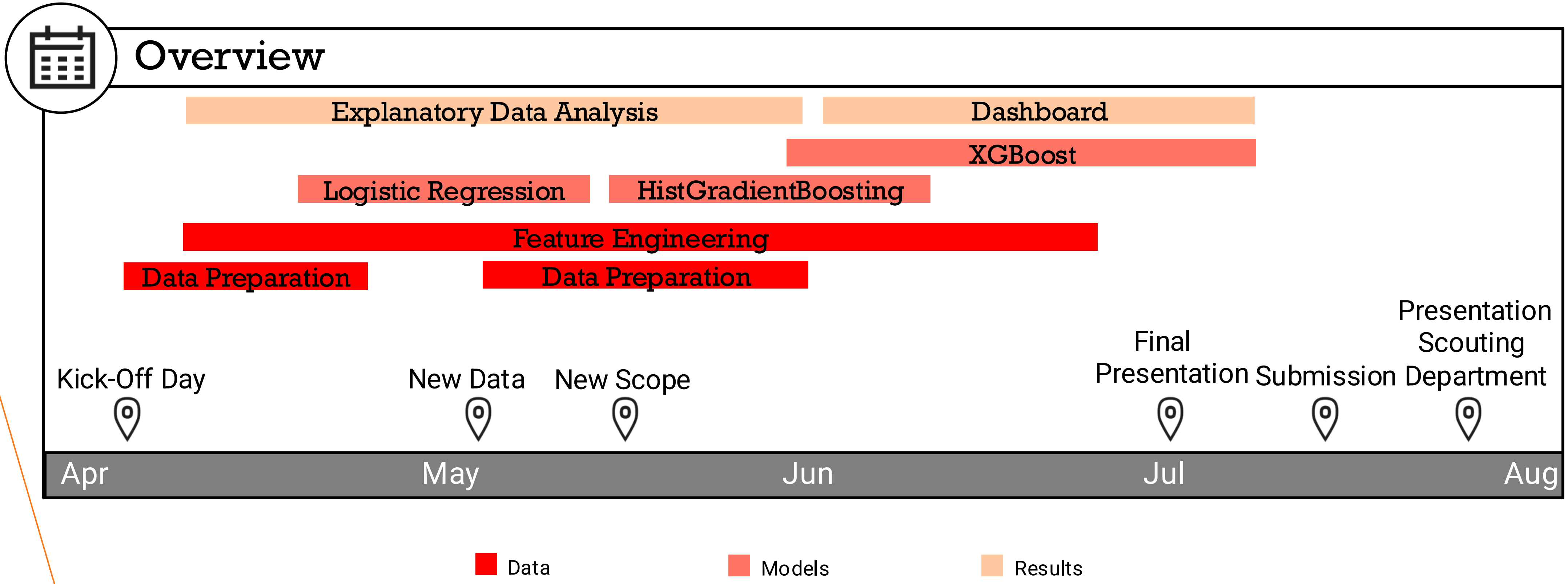
Create a predictive model to gauge success probability (quantified in playing percentage) of any given player to the Bundesliga (or top 5 league)

Our final estimate will then be implemented in a player database by FC Cologne

Allows the scouting department to have another perspective that is not merely the statistics of each player alone

Timeline

Overview

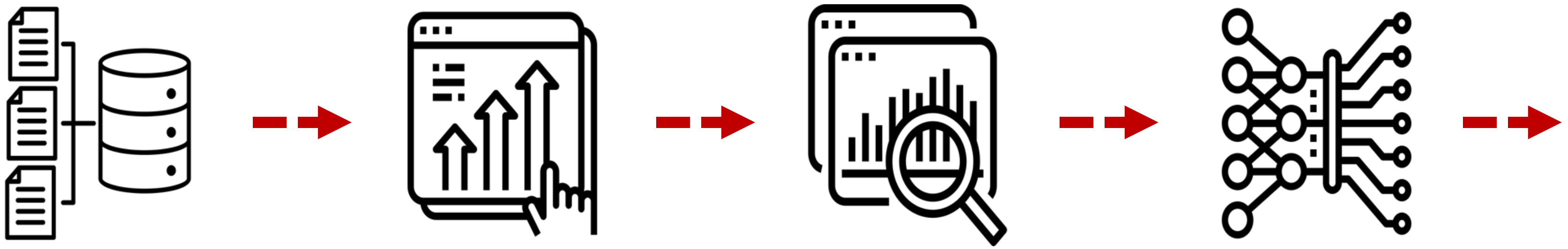


Method

1. Data Preparation

3. Conducting EDA

5. Interactive Dashboard



2. Feature Engineering

4. Modeling

Data Preparation

- Source: Transfermarkt.de
- Period: 2019 -2024
- Observations: ~ 70000 transfers

Dropping Unnecessary Columns + **Adding Features** + **Binary Indicator of Success** + **Keeping Only Last Transfer Of A Player**

Feature Engineering

Features were engineered from 4 key areas:

Transfer Data + **Player Profiles** + **Market Values** + **Performance**

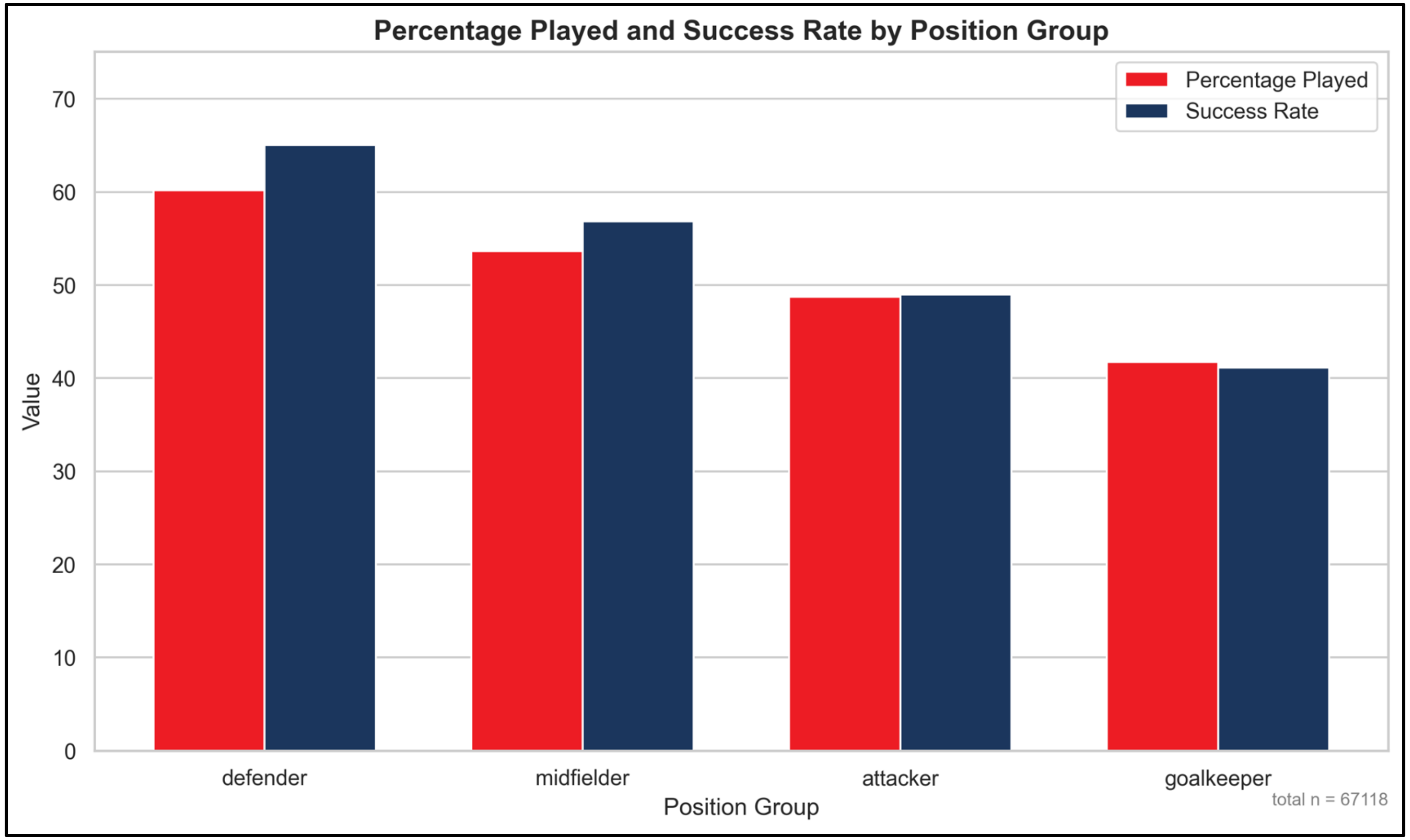
Feature Engineering

Added additional features:

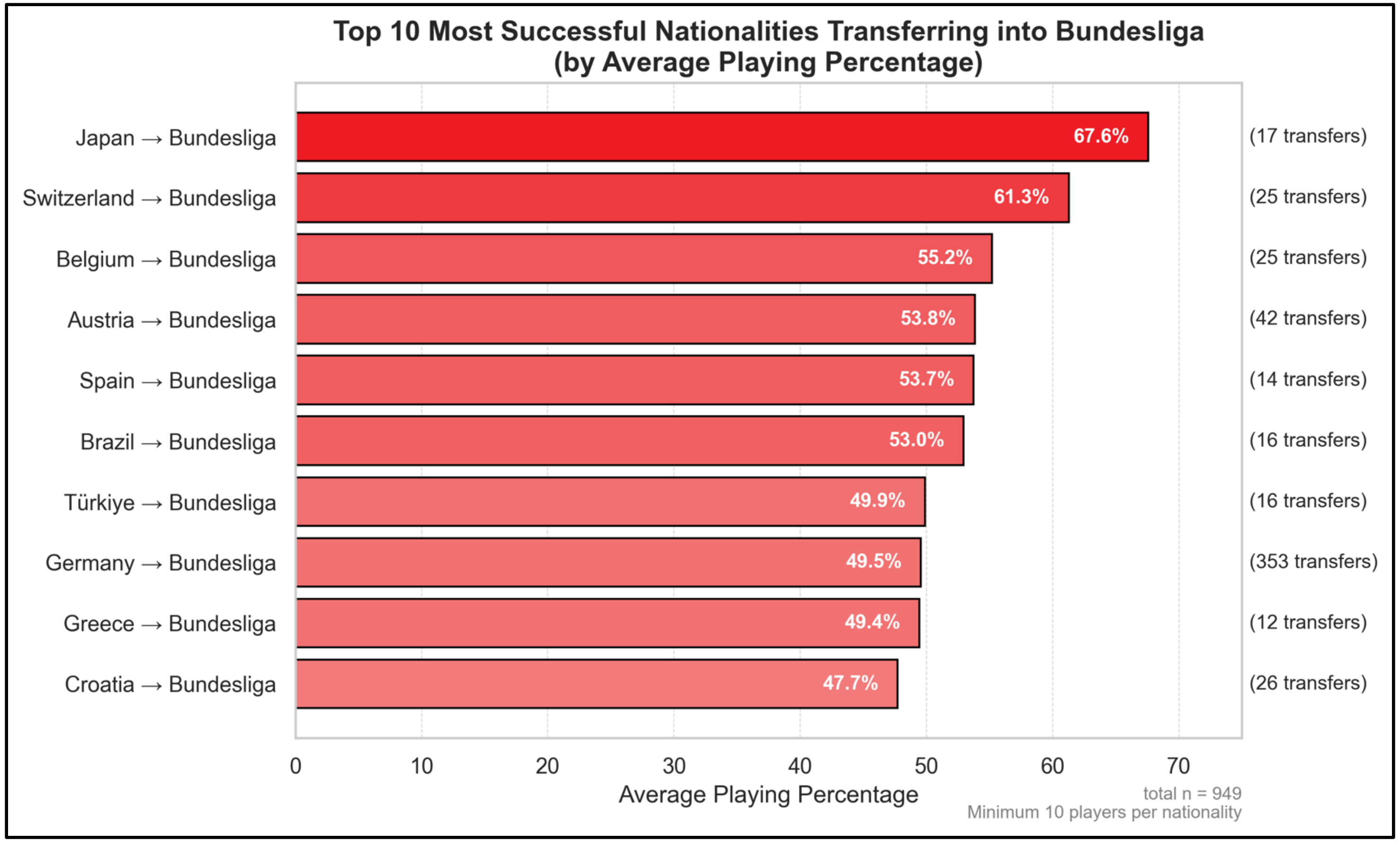
- Scorer (Goals + Assists) + Clean Sheets
  → Grouping Clean Sheets and Scorers
- Fee to Value Ratio (Under / Overpay)
- Foreign Transfer (Cultural differences / Language barriers)
- Value per Age and Value Age Product (Interactions between player age and value)
- Team Market Value Ratio (To Team Market Value / From Team Market Value)

Result: Improved feature quality and model accuracy

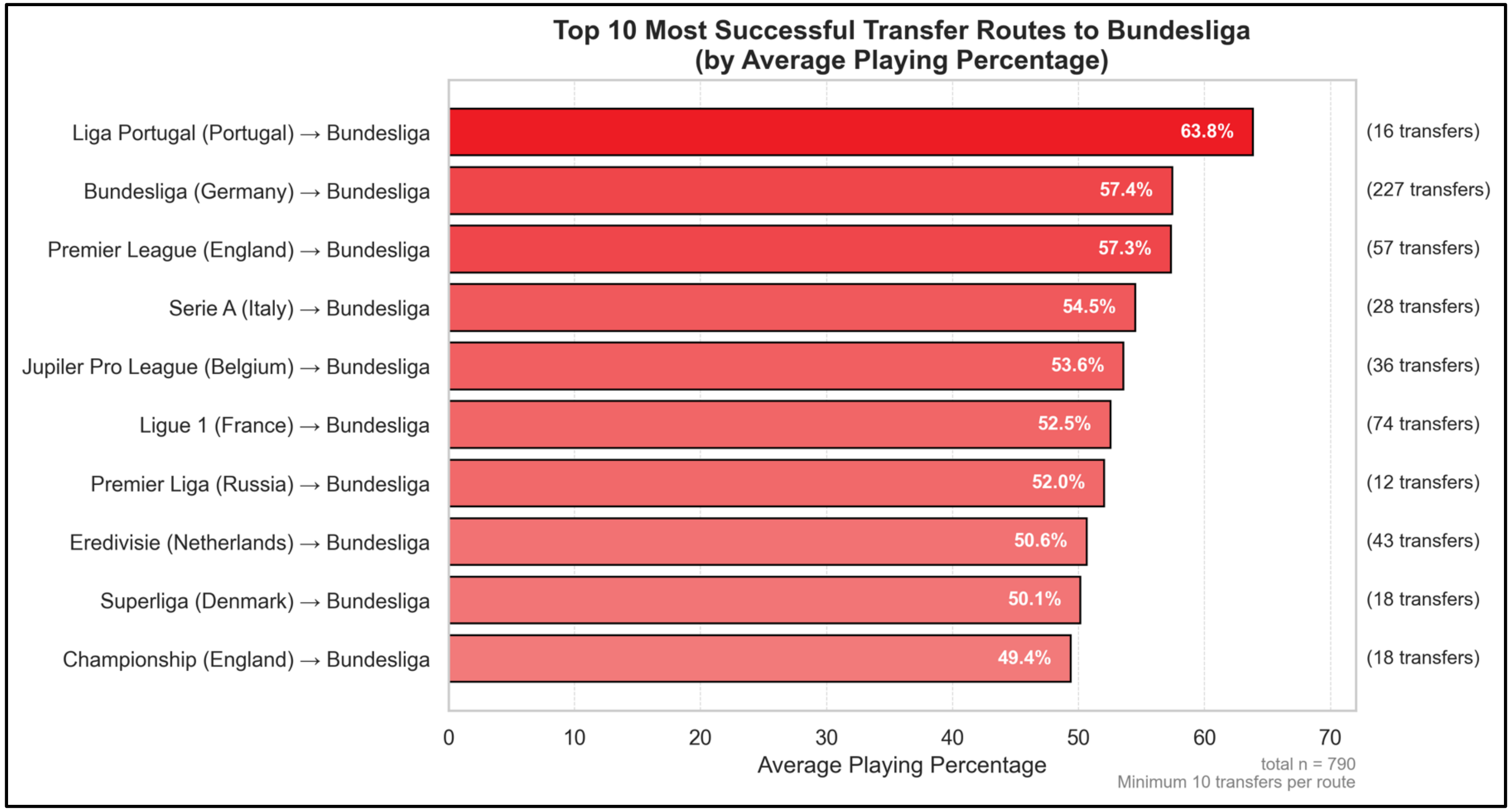
Successful Playing Positions



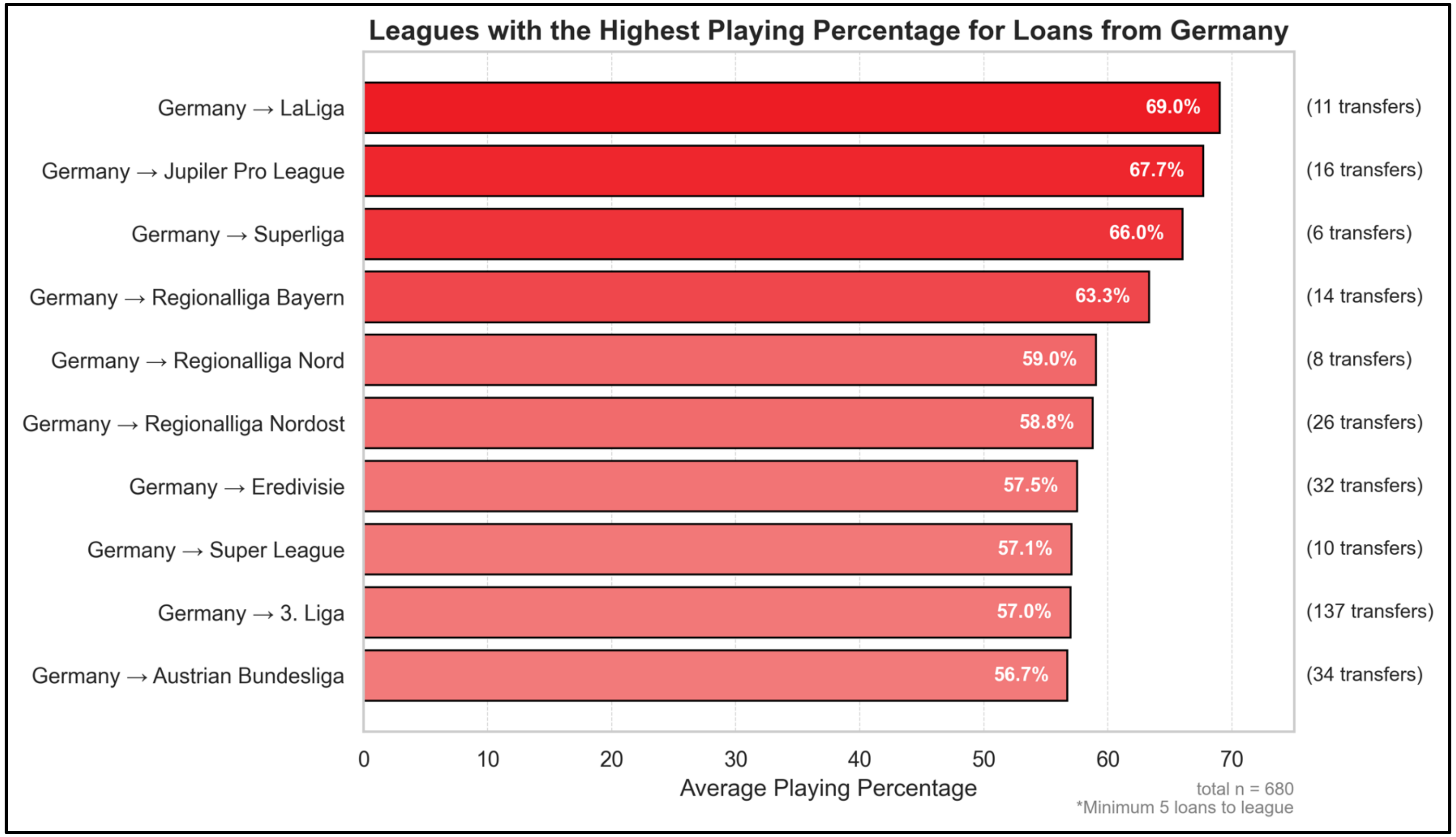
Successful Nationalities in Bundesliga



Successful Source Leagues



Promising Loan Destinations



Logistic Regression Model

Basic logistic regression model with target variable
→ Success (>50% playing time)

	Precision	Recall	F1-Score
0 (no success)	0.61	0.45	0.52
1 (success)	0.68	0.80	0.73
Accuracy			0.66

Project Scope Change

Evaluating based on 50% threshold:

- Limited interpretation
- Abstract cut-off criteria for Success
- Weaker predictive power

Our solution: Predict the playing percentage first, make bins after

Boosted Trees: Histogram Gradient Boosting, XGBoost Regression

- Results with more information behind it
- Able to compare similar transfers
- Possibility to see more easily where the model is weak

HistGradientBoosting: Performance

Performance:

- **R^2**: 0.226, **RMSE**: 25.95

Top 3 features:

- Percentage of minutes played before the transfer
- Assists per minute before the transfer
- Origin competition area (country of the previous league)

Interpretation:

- The model struggles with extreme predictions (very low or very high % played)
- It tends to predict toward the mean, e.g. ~40–60%, even when the ground truth is near 0% or 100%
- This is reflected in the predicted distribution line (see plot), which underrepresents the extremes of the true distribution

HistGradientBoosting

Better at identifying no success (higher specificity/precision for failures) but worse at predicting success compared to logistic regression

	Precision	Recall	F1-Score
0 (no success)	0,74	0,68	0,71
1 (success)	0,61	0,67	0,64
Accuracy			0,68

HistGradientBoosting: Insights

Histogram-based GBDTs apply feature binning, which:

- Reduces noise and smooths variance in input features
- Encourages conservative split decisions, biasing the model slightly toward central outcomes
- This makes the model more robust and generalizable, but it also explains its hesitation to predict extreme cases

Be aware of regularization effects

- Model avoids extreme predictions (e.G., 0% or 100%) → favors conservative estimates

Final Model: XGBoost

01

XGBoost chosen for handling complex, mixed-type football transfer data using discrete bins

02

RMSE used as loss function; model trained with 5-fold cross-validation

03

Hyperparameters tuned via Bayesian optimization

04

Final model: 1083 estimators, learning rate 0.0190, max depth 6

05

LightGBM used alongside XGBoost for feature interpretation with SHAP

Final XGBoost Model Results

Performance:

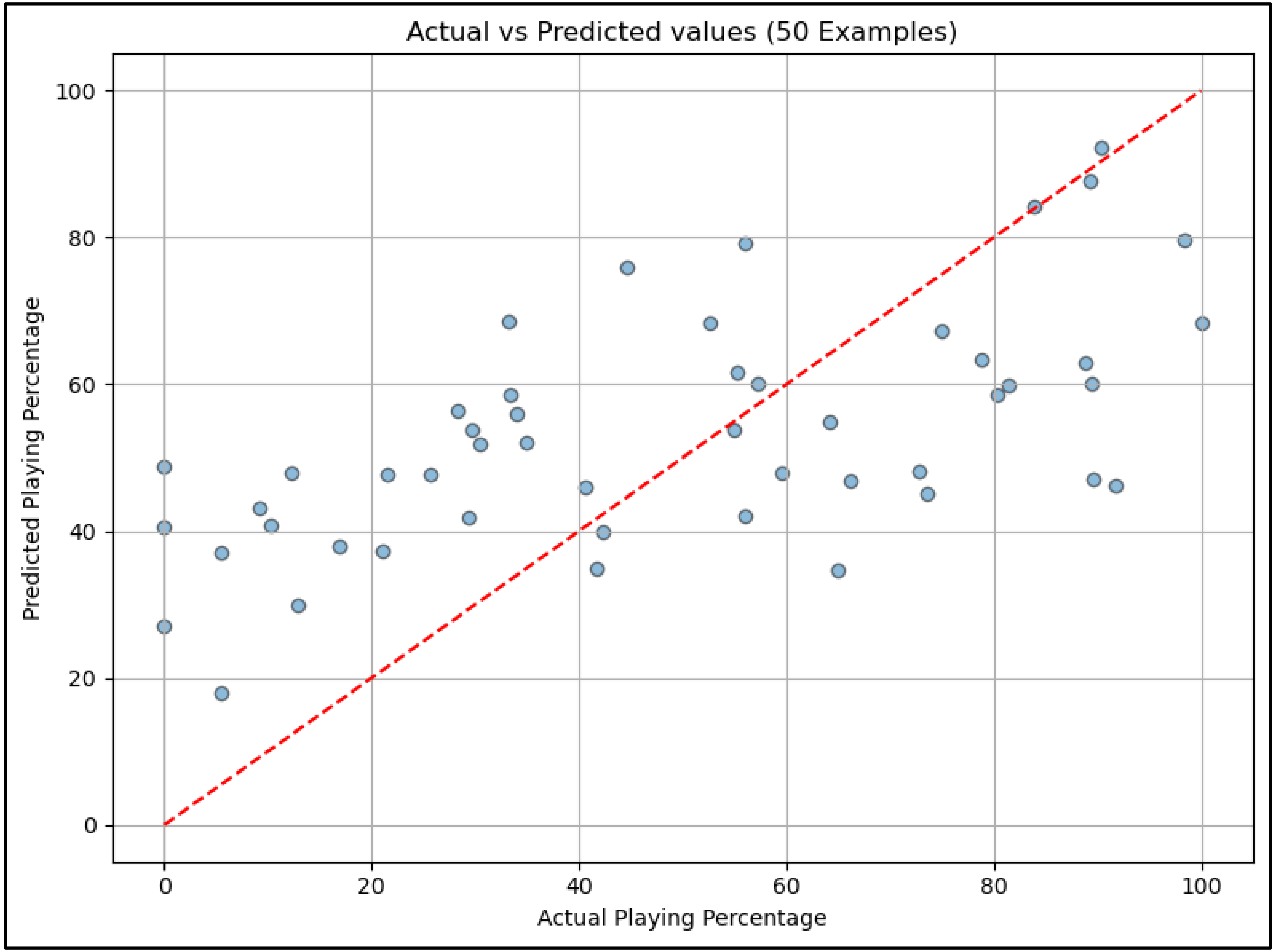
- **R^2**: 0.276, **RMSE**: 25.89

Top 3 features:

- To competition area
- To team market value
- Percentage played before

Interpretation:

- The model still struggles with extreme predictions, however less than the previous models
- Generally predicts the direction of deviation from the mean well, the magnitude less


Actual vs Predicted values (50 Examples)
Predicted Playing Percentage
Actual Playing Percentage
0
20
40
60
80
100

Insights

Variable	How it works
Age	Increases logarithmically, peaks at 33
Main Position	Reinforces the EDA
Target Team Overall Market Value	Large clubs have bigger squad depth, harder to be a starter
Loan	players on loan play more, players coming back from loan play less
Percentage of minutes played last season	Increases exponentially, key players play a lot more in their future club too
Market Value of Player	Increases exponentially, almost no change from €60M upwards

Our 4 Do’s and Don’ts for Transfer Decisions

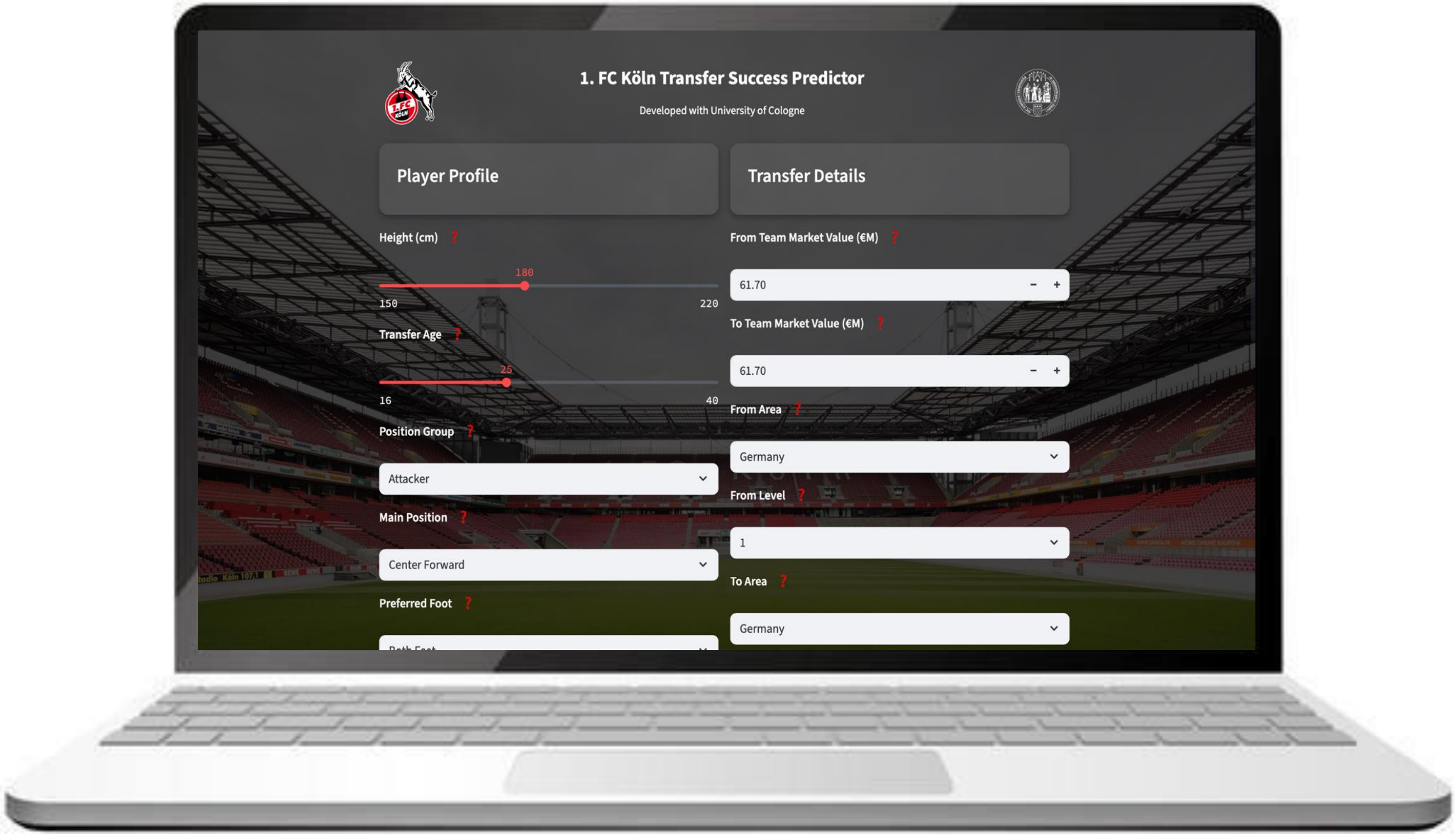
Do loan young players (18–22 years) to support development

Do sign players who played a lot of minutes in previous seasons

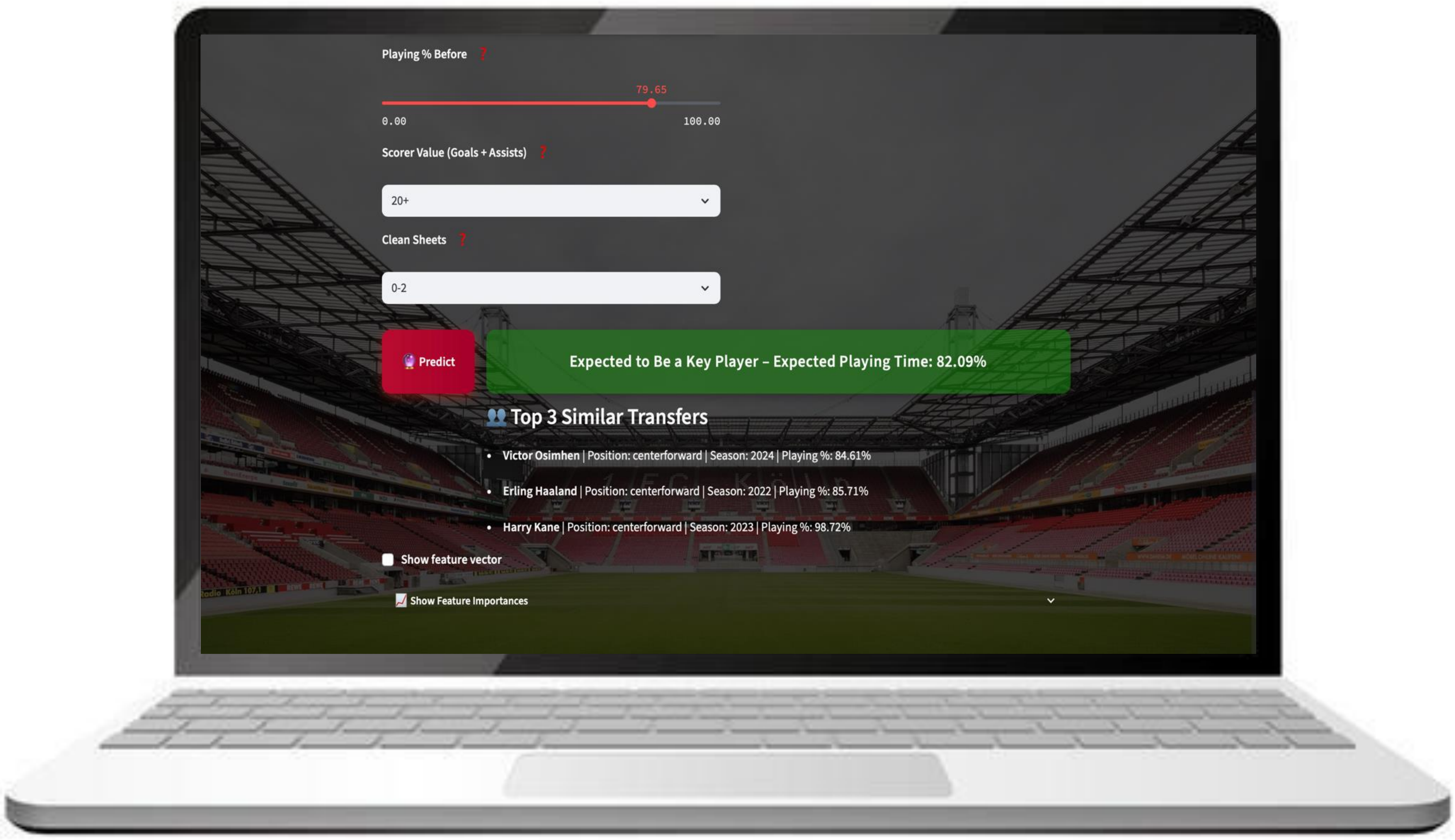
Don’t loan out players over 22 years old, as it offers limited benefit

Don't sign older, experienced players if you are a low market value team

Dashboard


1. FC Köln Transfer Success Predictor
Developed with University of Cologne
Player Profile
Transfer Details
Height (cm)
180
150
220
Transfer Age
25
16
40
Position Group
Attacker
Main Position
Center Forward
Preferred Foot
From Team Market Value (€M)
61.70
To Team Market Value (€M)
61.70
From Area
Germany
From Level
1
To Area
Germany

Dashboard

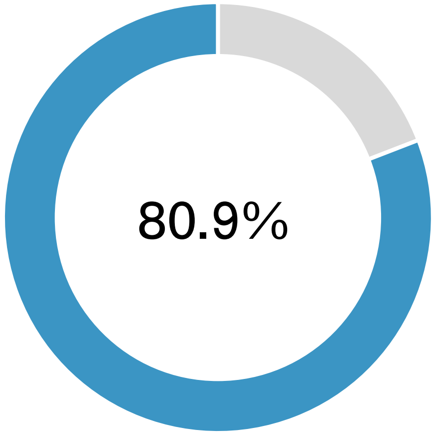

Playing % Before
79.65
0.00
100.00
Scorer Value (Goals + Assists)
20+
Clean Sheets
0-2
Predict
Expected to Be a Key Player – Expected Playing Time: 82.09%
Top 3 Similar Transfers
Victor Osimhen | Position: centerforward | Season: 2024 | Playing %: 84.61%
Erling Haaland | Position: centerforward | Season: 2022 | Playing %: 85.71%
Harry Kane | Position: centerforward | Season: 2023 | Playing %: 98.72%
Show feature vector
Show Feature Importances

Model Examples – Nico Schlotterbeck

Our estimate:

Predicted Playing Percentage: 74.64%

Actual playing percentage:

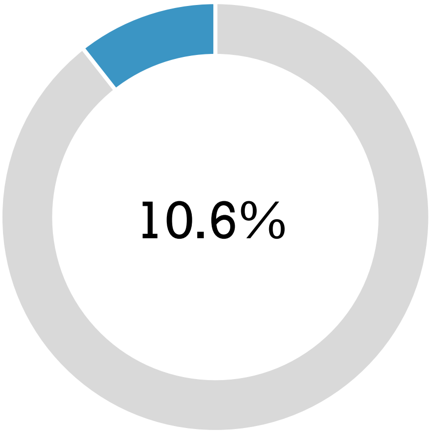


Model Examples – Álvaro Odriozola

Our estimate:

Predicted Playing Percentage: 43.70%

Actual playing percentage:



Business Impact

HELPS 1. FC KÖLN IMPROVE SCOUTING AND TRANSFER PLANNING

FILTERS THE TRANSFER MARKET TOWARD BETTER-FITTING PLAYERS

HELPS TO AVOID COSTLY SIGNINGS THAT FAIL TO INTEGRATE

MAKES COMPLEX PLAYER PROFILES EASY TO QUICKLY ASSESS AT A GLANCE

SUPPORTS LONG-TERM PERFORMANCE AND FINANCIAL SUSTAINABILITY

Conclusion & Outlook

Use the model with the football fundamentals in mind:

Similar players model for real life comparisons

Other possible variables for a better model:

More performance data

Team specific data of wants and needs

Thank You and Check Out Our Dashboard!

Sources

- https://fc.de/aktuelles/news/naechster-schritt-zur-wirtschaftlichen-erholung
- https://www.transfermarkt.de/1-fc-koln/startseite/verein/3
- https://www.espn.com/soccer/story/_/id/37502437/german-fa-announces-distribution-plan-464-billion-tv-cash-windfall
- https://en.wikipedia.org/wiki/1._FC_Köln
- https://media.dfl.de/sites/3/2024/03/6DB5
  ChT2B_DFL_Wirtschaftsreport_2024_EN.pdf

Group Contribution

Group Member	Contribution
Balint Farkas (7430884)	Equal contribution
Rishvanyas Premanand (7434745)	Equal contribution
Kevin Drost (7434810)	Equal contribution
Laurenz Flender (7430935)	Equal contribution