THE ULTIMATE MARKET PREDICTOR

FEL1-Team 9: Audric Yap, Joshua Chin, Gabriel Lim

TABLE OF CONTENTS

PROBLEM STATEMENT

Predict the current market value of football players to better understand what drives the value of players, using available personal and game statistics

I

EDA

THE DATASET

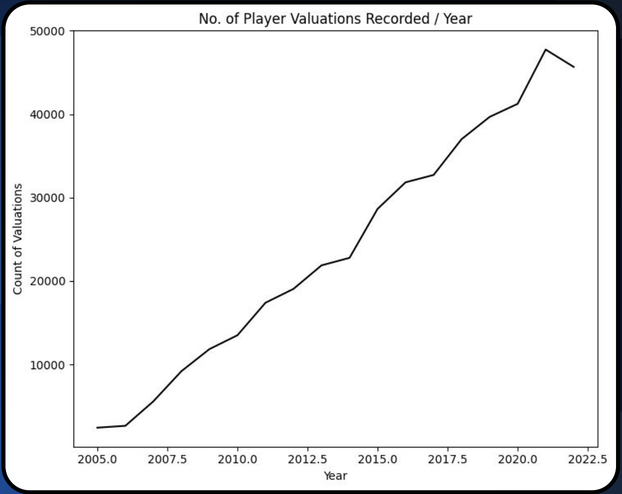
- Dataset obtained from **Kaggle**
- **Scraped from the TransferMarkt** website for its reliability and consistency
- Contains **detailed information** on player and game statistics, valuations and more

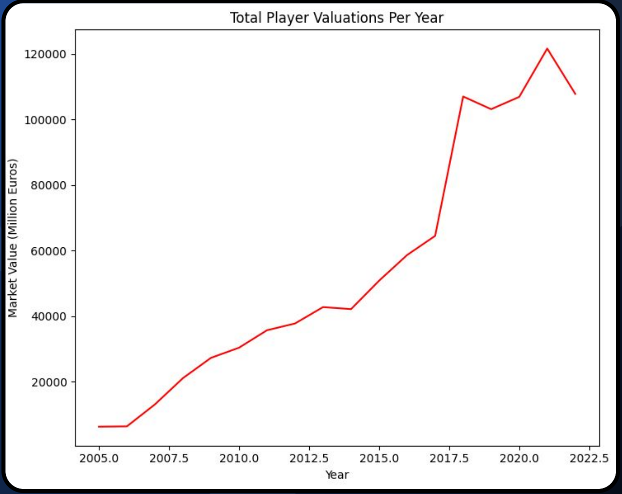
NO. OF MARKET VALUATIONS

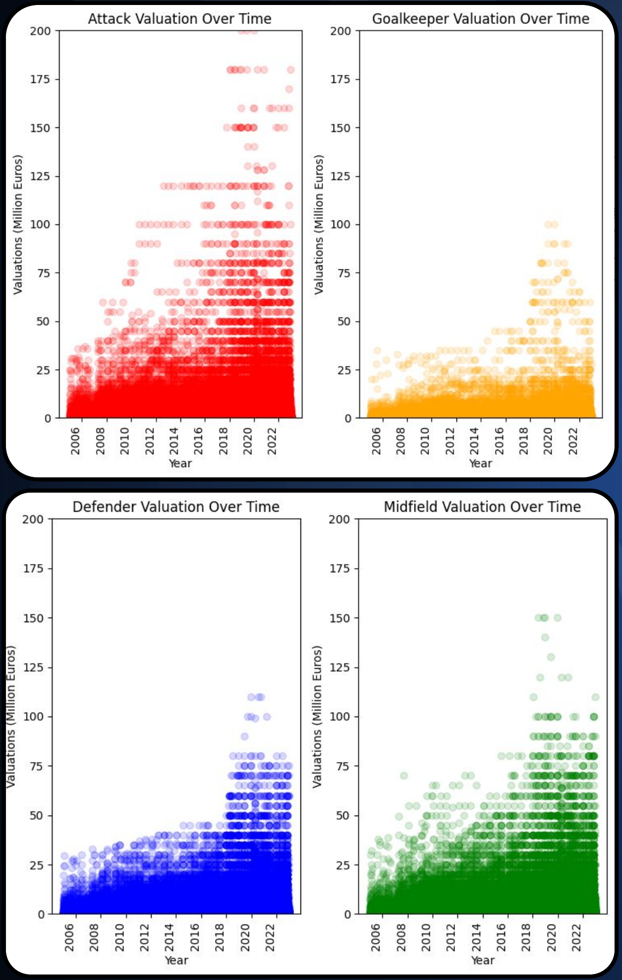
- **Number of player valuations increases** consistently over time

- **Big spike** in the sum of player valuations past 2017, followed by an **inconsistent rise** till current day

- This could be attributed to a multitude of factors such as **sudden rising stars and the COVID-19 pandemic,** as well as **inflation**



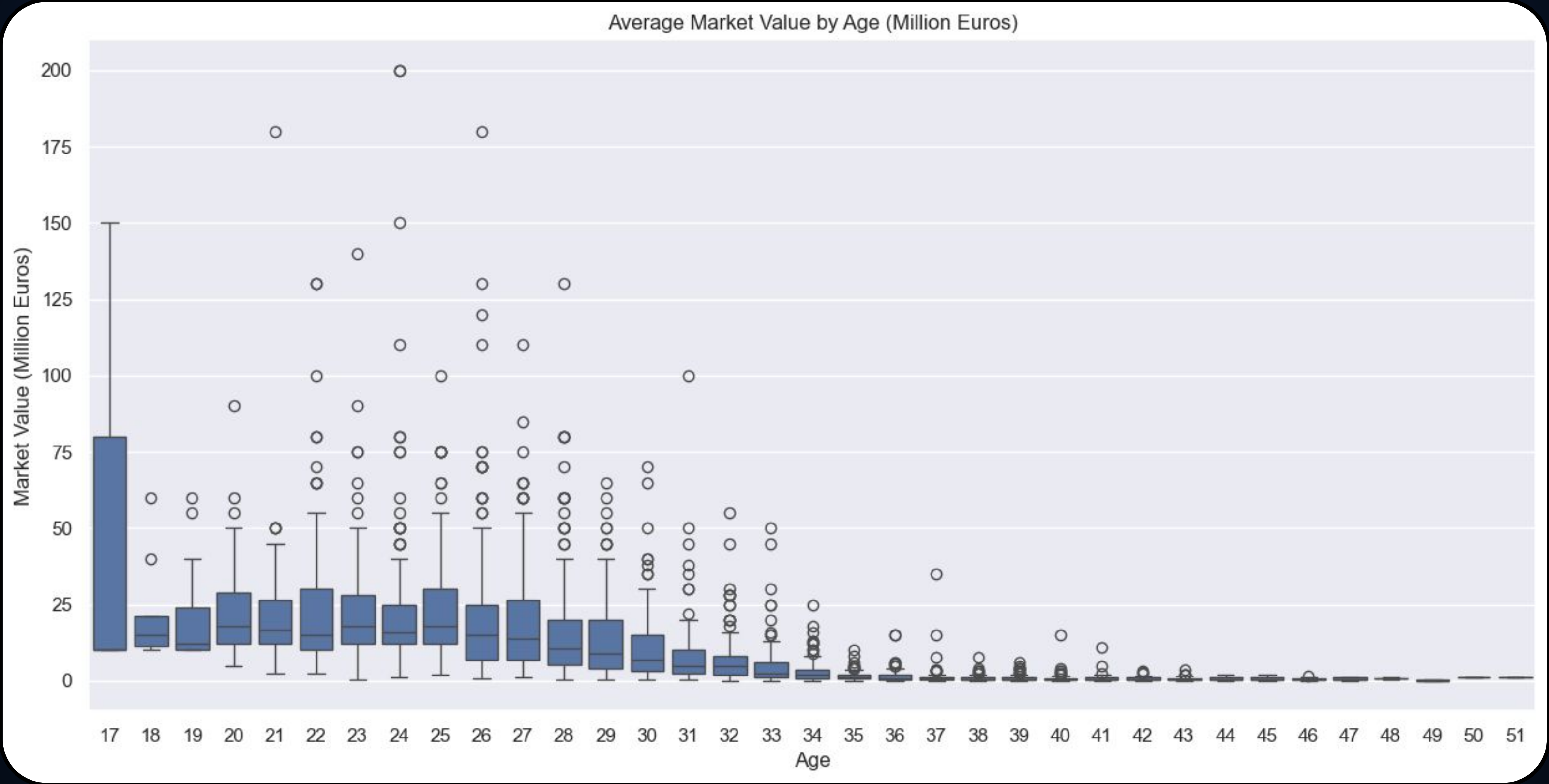




MARKET VALUATIONS BASED ON POSITION

- Generally as time progresses, the players' valuations in **all positions increases**, particularly during **2018 onwards**
- **Attackers seem to be valued more**, followed by Midfielders, Defenders and lastly Goalkeepers
- This is **reflective of real-world scenarios:**
    - Vinicius Jr., a world-class Attacker, is worth **€200 Million**
    - William Saliba, a world-class Defender, is contrastingly worth only **€80 Million**

MARKET VALUATIONS BASED ON AGE



II

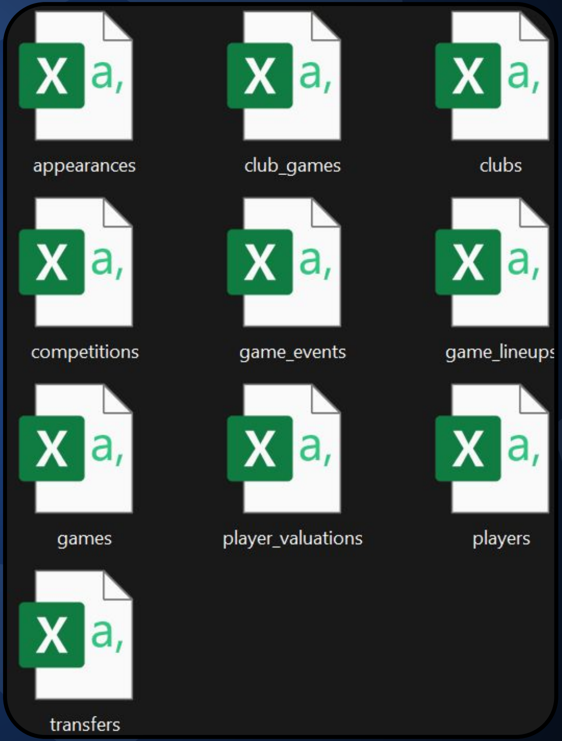
DATA PROCESSING

MERGING THE DATA

Data is split between multiple .csv files. We would need to **merge them together** to one data frame for **easier training**

After **cutting out unimportant data**, we decided to merge the following .csv files:

- players.csv
- appearances.csv
- games.csv
- competitions.csv



FEATURE ENGINEERING

- Mapped each player's league competition to a **ranking based on UEFA coefficients**
- **Compiled game statistics** for all players from 2020 to 2023:
  - Games Played
  - Minutes Played
  - Goals and Assists (Individual and Team)
  - Yellow and Red Cards
- Obtained the **current age** of players from based on current day
- **OneHotEncoded player positions** for more meaningful analysis

FINAL DATA FOR TRAINING

PERSONAL STATS

The current age and height of players

POSITION

The players' preferred playing positions

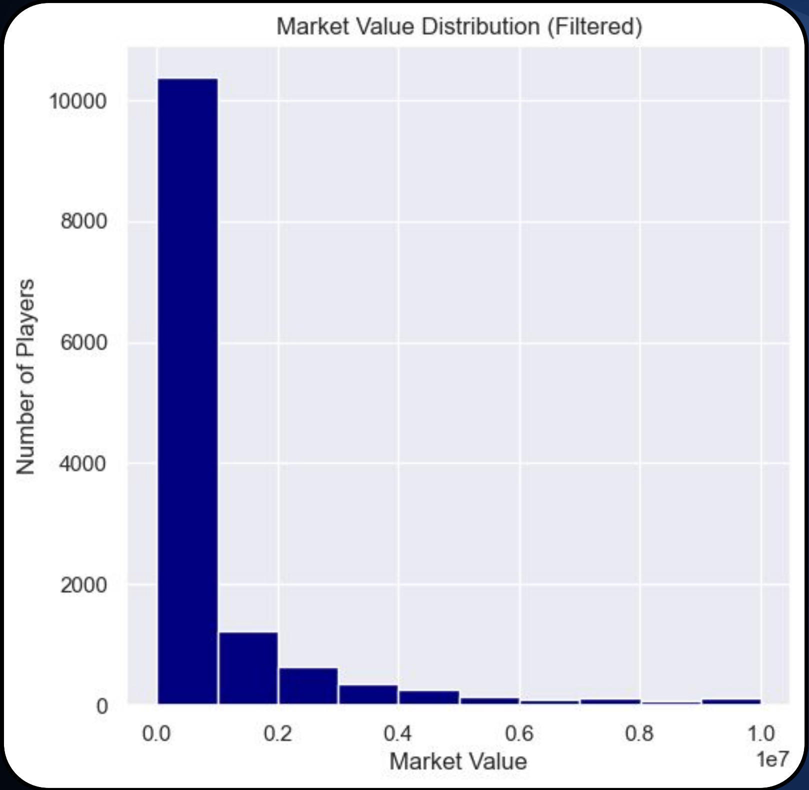
GAME STATS

Game statistics from 2020 season till 2023 season, as well as competition ranks

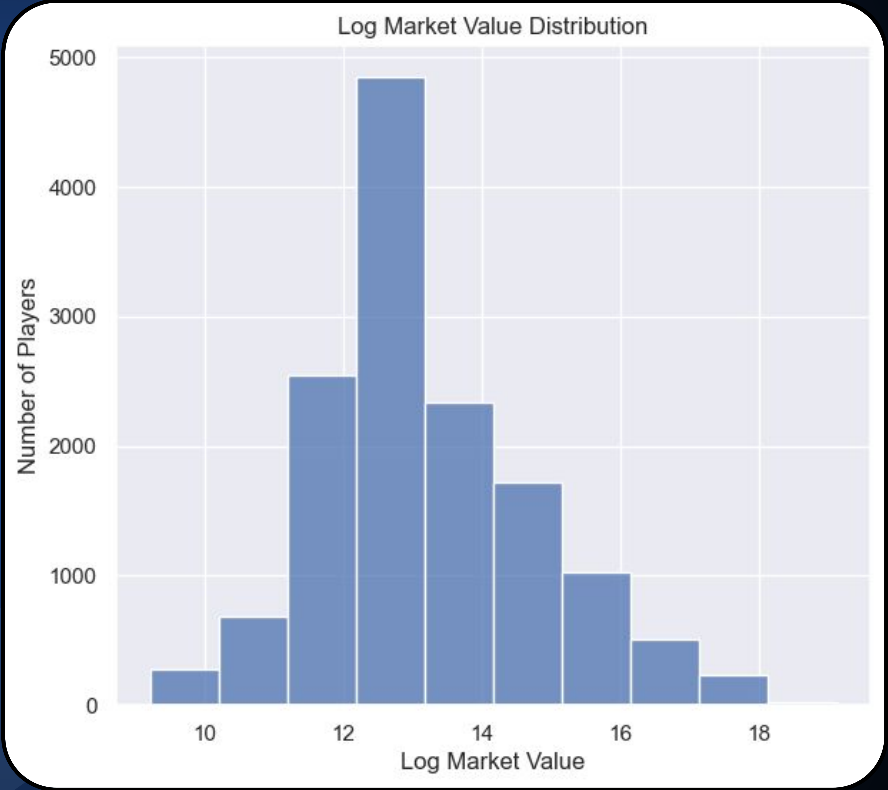
REGION

The regions in which players are born in

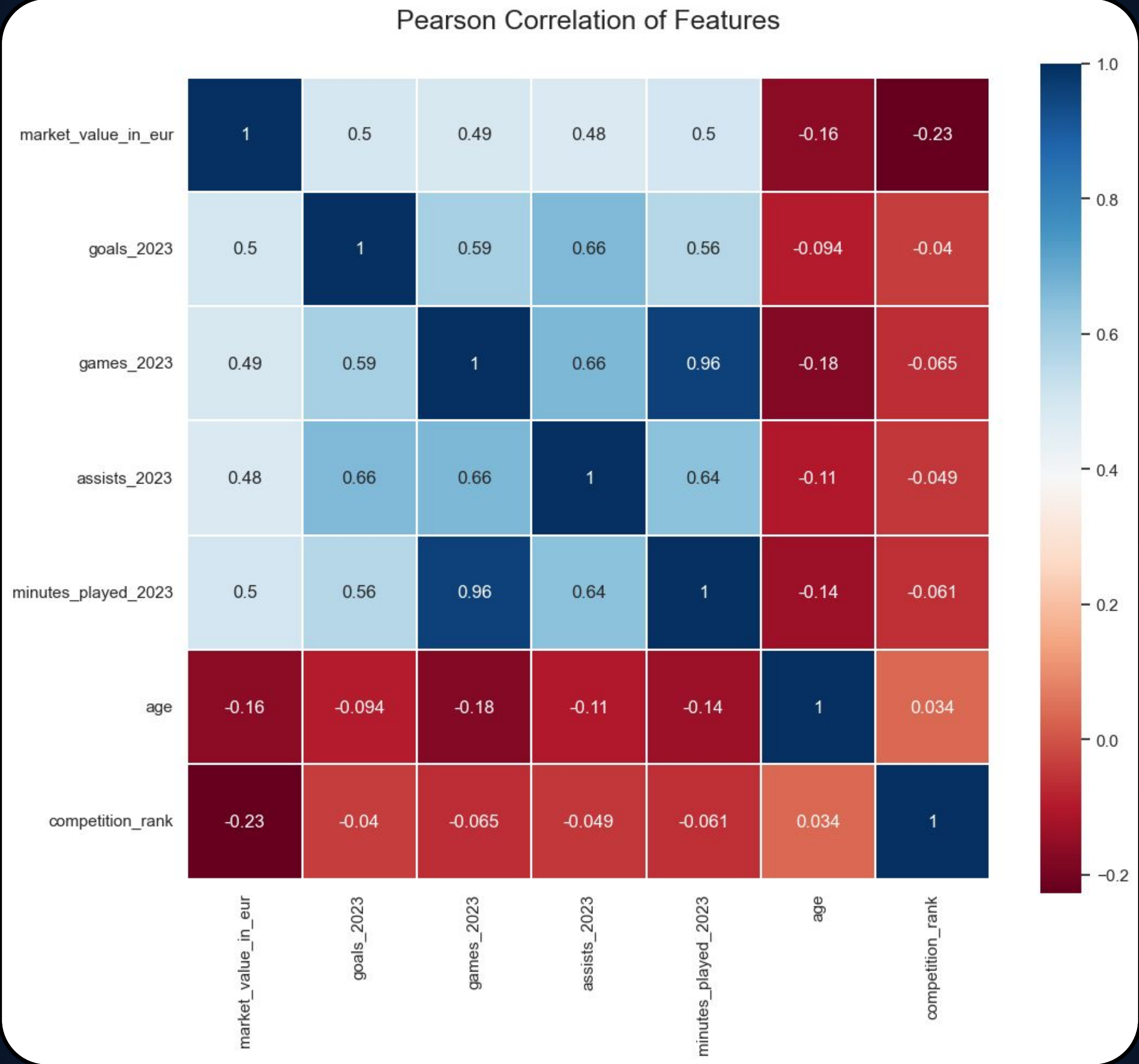
TRANSFORMING MARKET VALUE “y”



Before Log



After Log


Pearson Correlation of Features
market_value_in_eur
goals_2023
games_2023
assists_2023
minutes_played_2023
age
competition_rank
1 0.5 0.49 0.48 0.5 -0.16 -0.23
0.5 1 0.59 0.66 0.56 -0.094 -0.04
0.49 0.59 1 0.66 0.96 -0.18 -0.065
0.48 0.66 0.66 1 0.64 -0.11 -0.049
0.5 0.56 0.96 0.64 1 -0.14 -0.061
-0.16 -0.094 -0.18 -0.11 -0.14 1 0.034
-0.23 -0.04 -0.065 -0.049 -0.061 0.034 1
1.0
0.8
0.6
0.4
0.2
0.0
-0.2

III

MACHINE LEARNING

METRIC IMPORTANCE

In the football transfer market, the focus is primarily on **interpretability as well as reliability**. As such, our metric focuses will be as such:

- **Primary Metric: Mean Absolute Error (MAE)**
  - Chosen as football market values are expressed in **real-world currency**, in this case Euros, so decision makers like club analysts or agents are **easily able to understand** the average deviation between predicted and actual market values

- **Secondary Metric: R^2**
  - Chosen so that the model is able to **explain variability in market values**, thus increasing its **reliability** in capturing market trends

TESTED MODELS

LINEAR REGRESSION

Fits a **linear model** with coefficients to **minimize the residual sum of squares** between the observed targets in the dataset

RANDOM FOREST REGRESSOR

Fits decision tree regressors on various sub-samples and uses averaging to **improve the predictive accuracy** and **control over-fitting**

ELASTIC NET

Uses the **penalties** from both the **lasso and ridge** techniques to **regularize** regression models.

XGBOOST REGRESSOR

Builds an **ensemble of decision trees**, where **each tree is trained** to make predictions based on a subset of the available data

BASELINE MODEL RESULTS

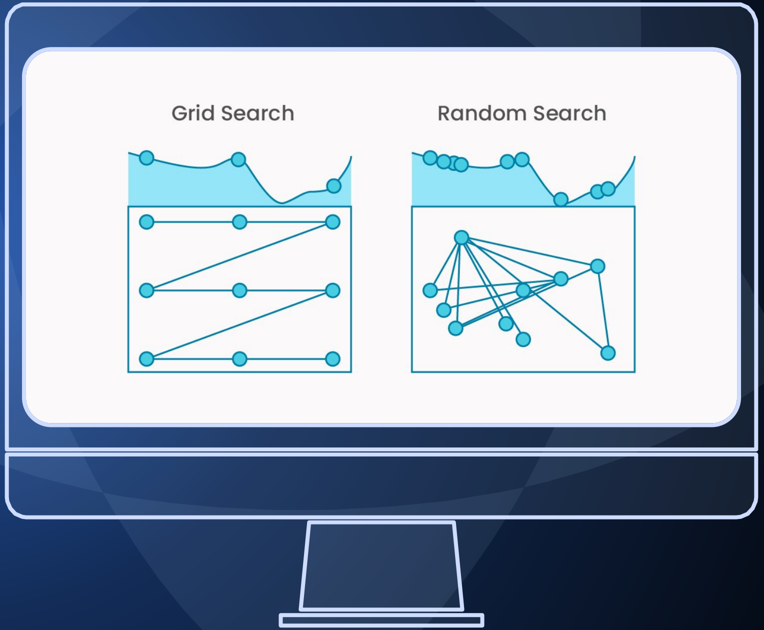
Model	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	R^2 Score
Linear Regression	0.7873	1.0353	0.5948
Elastic Net	0.8272	1.1457	0.5516
Random Forest	0.6981	0.8513	0.6668
XGBoost	0.6548	0.7546	0.7047

HYPER-PARAMETER TUNING

Given that **XGBoost** is currently the **best model**, having MAE and MSE closest to 0 and R^2 closest to 1, we want to tune it using **GridSearchCV**:

Here are our chosen **optimal hyper-parameters**:

- colsample_bytree: 0.6
- learning_rate: 0.03
- max_depth: 6,
- n_estimators: 500
- subsample: 0.9



XGBOOST POST-TUNING

MAE	MSE	R^2
0.641	0.730	0.714
-0.013	-0.024	+0.010

IV

INSIGHTS

INTERPRETING THE RESULTS

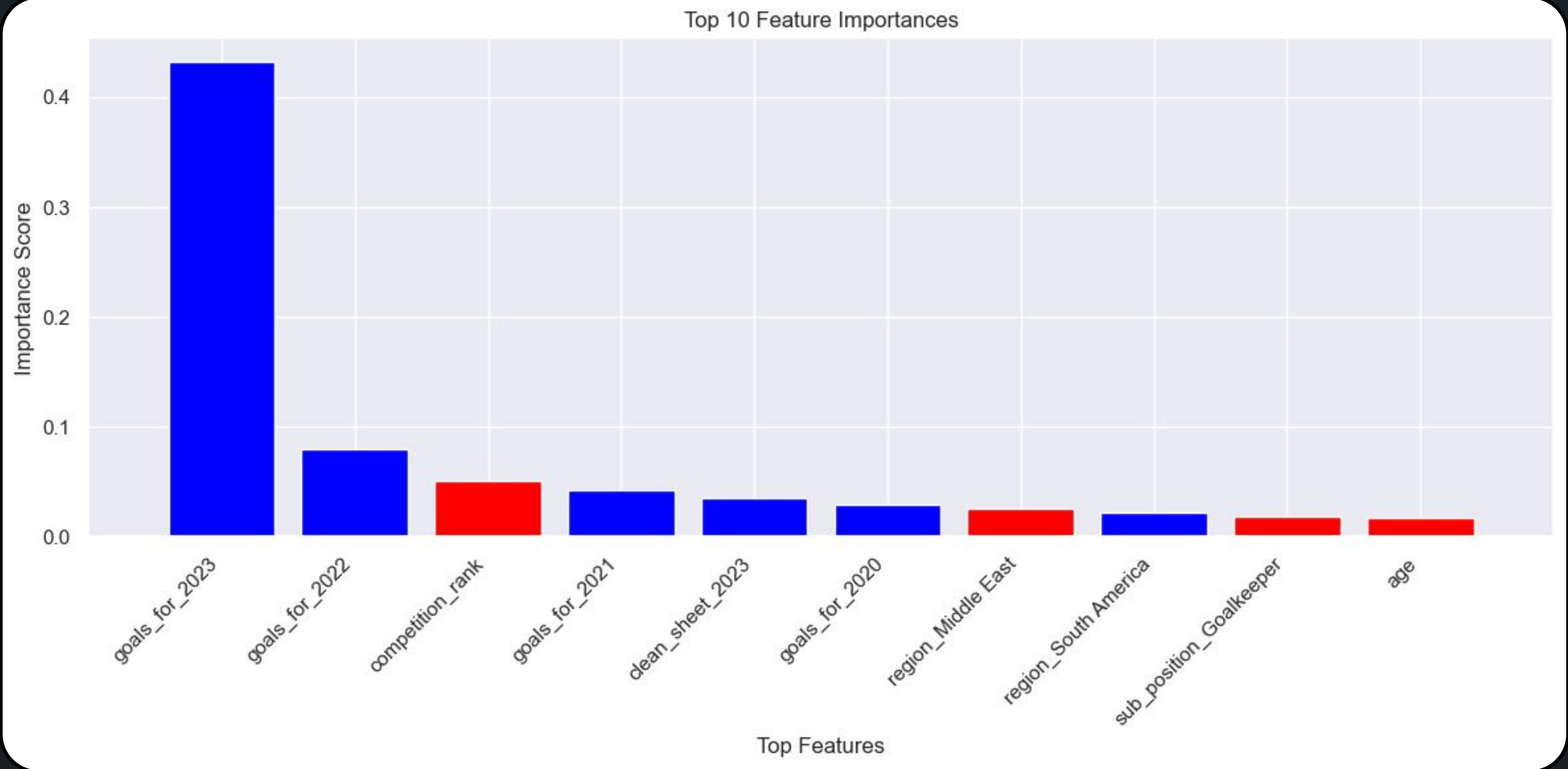
From our XGBoost Model, we obtained a **MAE of 0.643**

- Recomputing MAE using the original scale, we found that the MAE is around **€1.14 Million**
- On average, the model predicts market values with an **absolute error of €1.14 Million**, so it is **relatively accurate** for predicting market values in a domain where values can range widely up to tens of millions

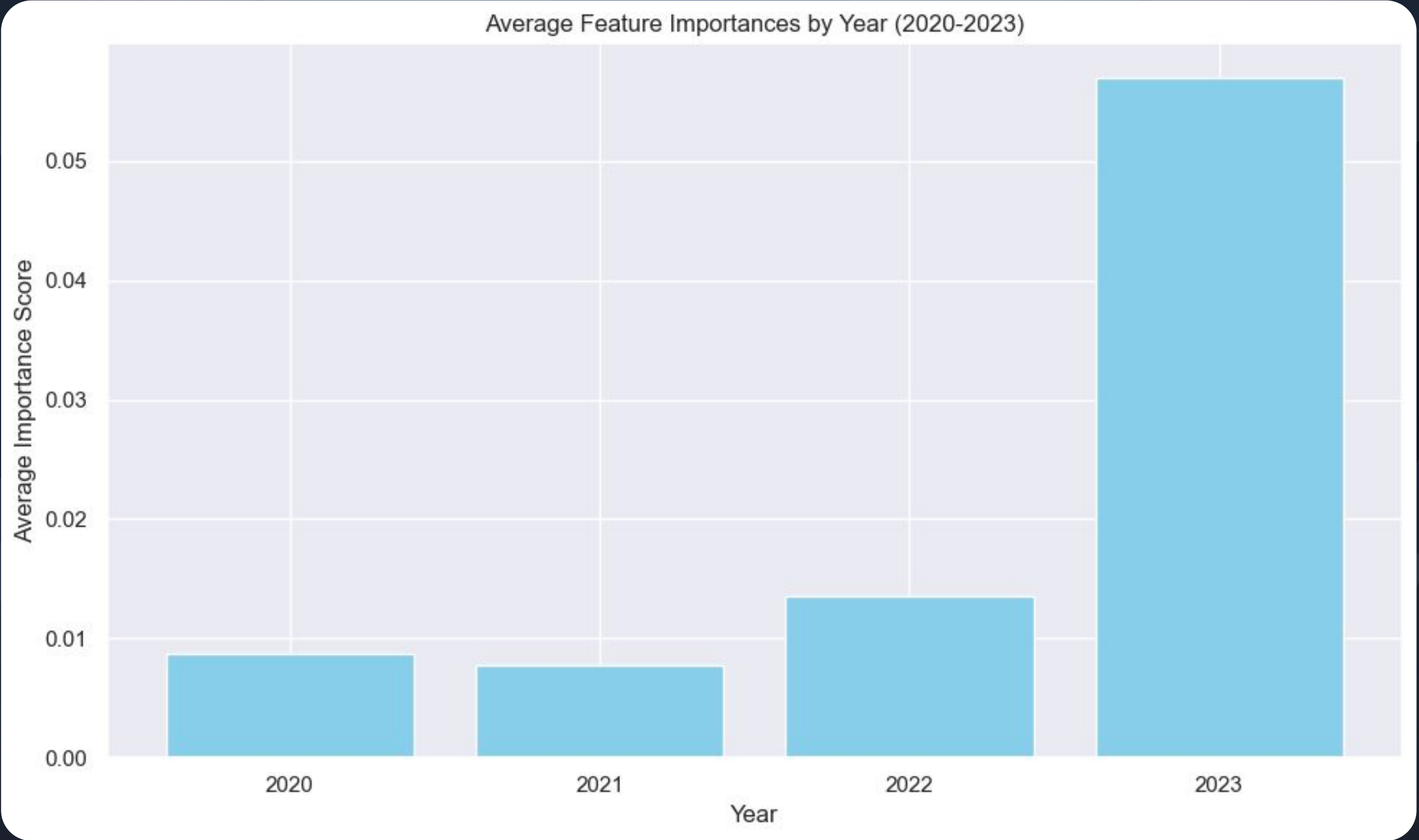
Our model also obtained an **R^2 score of 0.714**

- Explains **71.4% of variability** in football players' market values
- This is a **good result** as other **external factors like club/player sentiments that is not captured** in the model can be attributed to this result

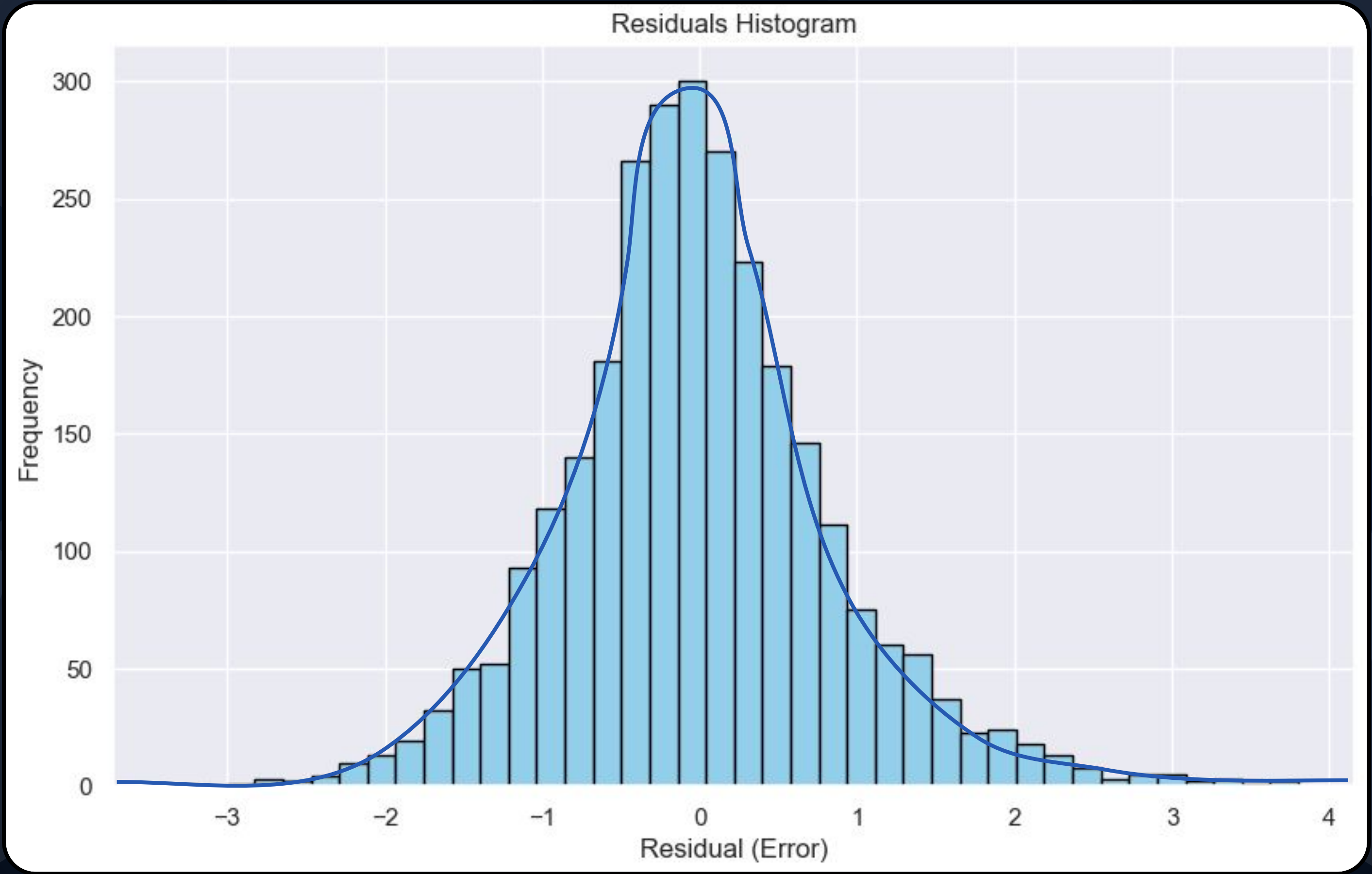
XGBOOST TOP FEATURE IMPORTANCE



XGBOOST FEATURE IMPORTANCE BY YEAR



XGBOOST ERROR RESIDUALS



CONCLUSION

- Our model is able to predict, with a **low error margin**, the current market prices of football players based on their **past and current game statistics and personal traits**

- The model also has a **variability of 71.4%** of the market captured, allowing it to **pick up on market trends** reliably and easily

- This displays the **robustness** of our model in predicting football market values, which is a **useful tool for any football club** looking to make a player investment

THANK YOU