

MODELLING FOOTBALL PLAYERS VALUATION

December 15th, 2023

Marco Wong



2018

€222 MILLION



PROBLEM

In football's hyper competitive world, clubs face significant financial risks due to inaccurate player valuations.

SOLUTION

Utilize last season's player statistics to accurately determine player valuations.



Player Name



Age:

Team:

Height:

Matches Played:

Passes Completed:

Goal Creating Actions:

Aerial Won%:



€ X mil

DATASETS

Player Statistics

2021 – 2022 Season

2689 rows x 124 columns

2022 – 2023 Season

2921 rows x 143 columns

Features

Player Valuation

Transfermarkt

Player Valuation (20 years)

440643 rows x 9 columns

Players

30302 rows x 23 columns

Train Target

Transfer Fees

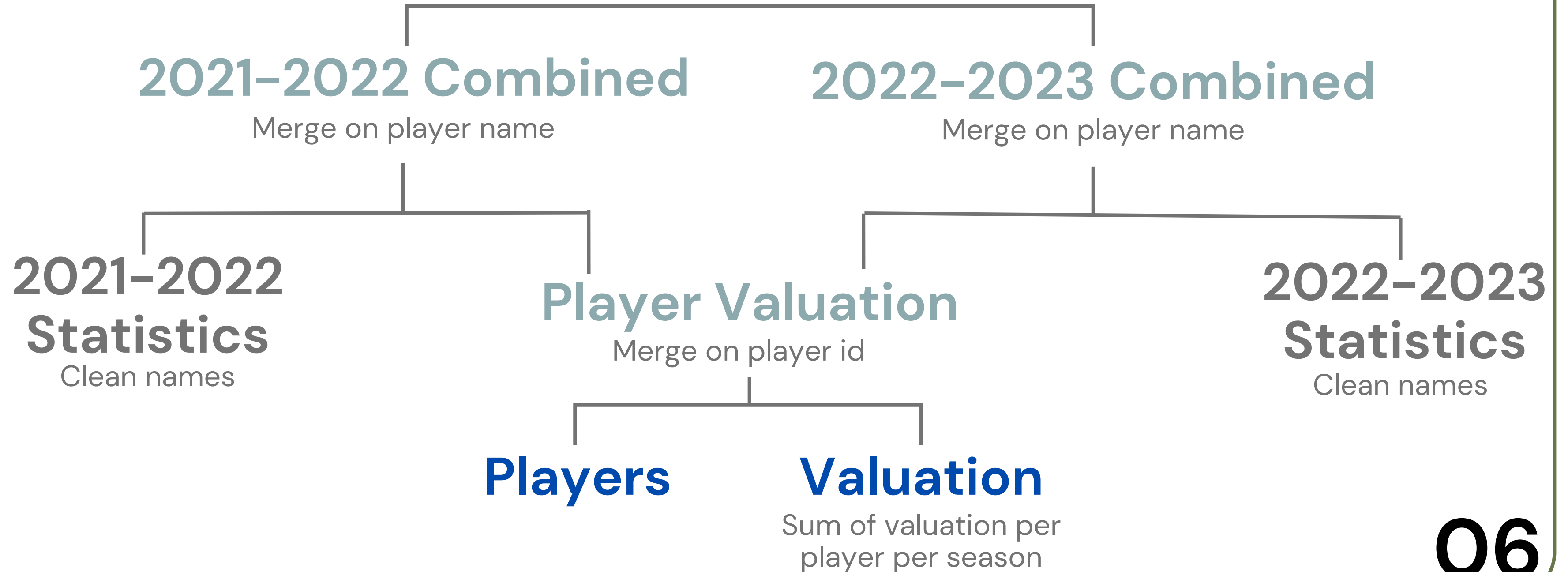
Transfermarkt

To be scraped

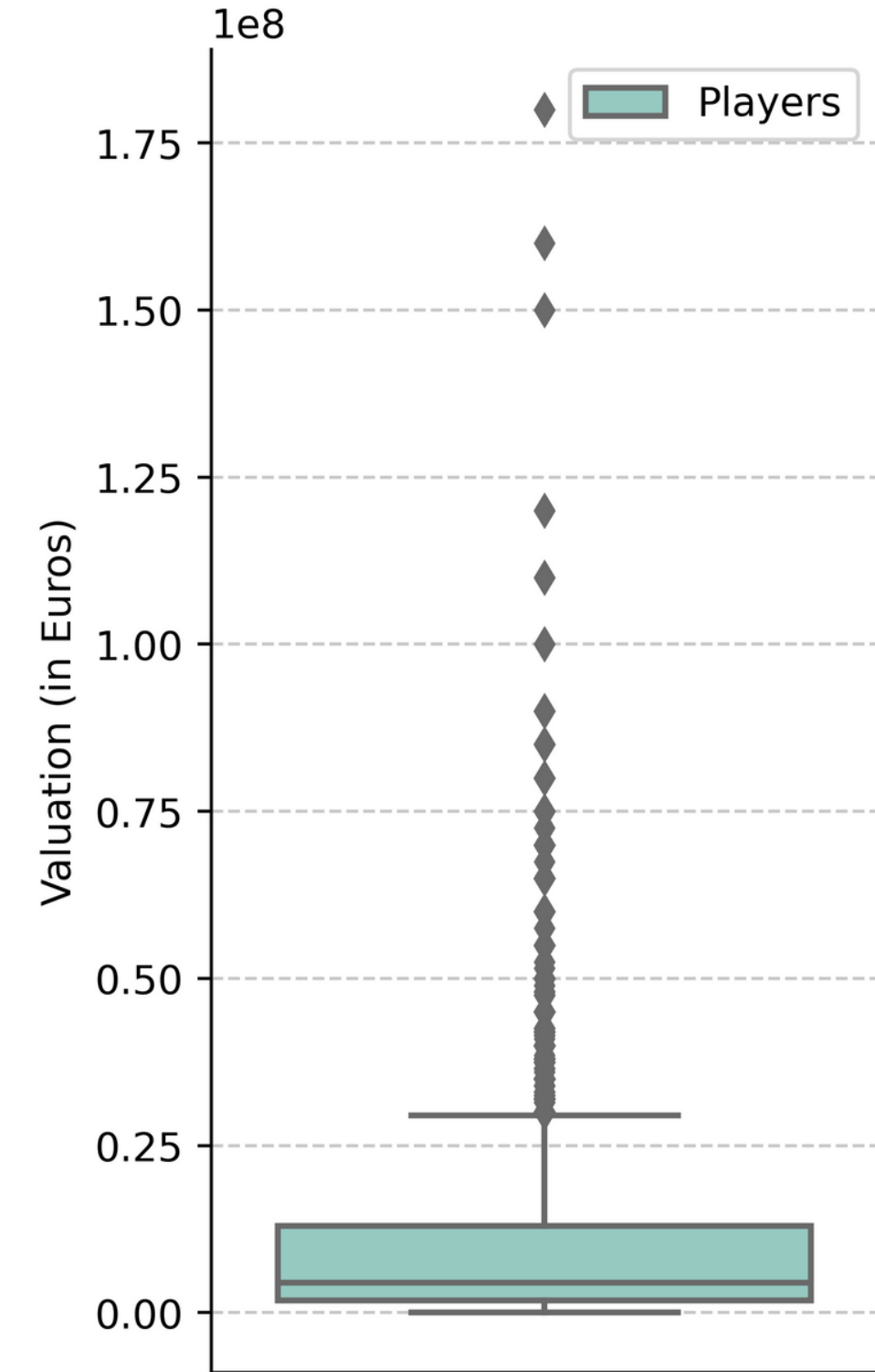
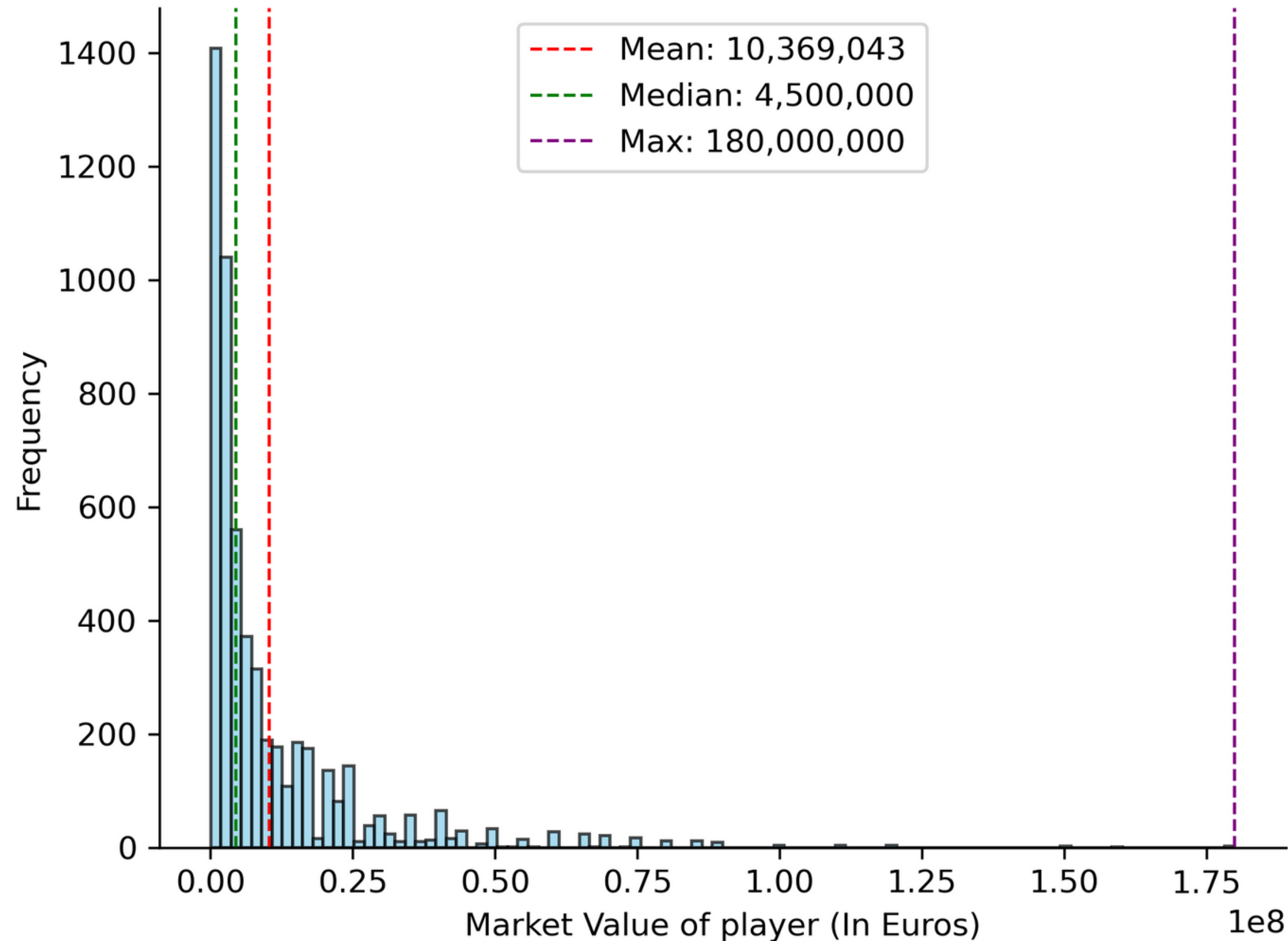
Test Target

PREPROCESSING

Final DataFrame



DISTRIBUTION OF VALUATION



BASELINE MODEL

Linear Regression

R SQUARED

Mean Cross Validation Score:

0.375

HIGH CORRELATION



GcaDef: Defensive actions that lead to a goal

Comp_Premier League: Player in Premier League

90's: Minutes played divided by 90

Pos_FW: Position of player is Forward

MEAN ABSOLUTE ERROR

Mean Cross Validation Score:

7,661,307

LOW CORRELATION

GCA: Goal creating action

Comp_Ligue 1: Player in Ligue 1

TKL: Number of players tackled

BLKPass: Number of times blocking a pass by standing in its path

NEXT STEPS

FEATURE SELECTION

Reduce feature set by correlation analysis and backwards feature selection

DATA SCRAPE

Scrape data for transfer fees in 2022 and 2023

OPTIMIZATION

Tune hyperparameters for optimal model

PRODUCT PREPARATION

Create pipelines and provide documentation



THANK YOU

Have any questions?

<https://github.com/MarcoWong96/Capstone-Project.git>

