



Football Scouting with Code

Michał Stobiecki, Robert Matthews
and Dylan Iteka

Problems and Goals



Scraping

```
df = pd.DataFrame()

df["name"] = player_name
df["age"] = player_age
df["nationality"] = player_nationality
df["club"] = player_club
df["position"] = player_position
df["value_in_millions_of_euros"] = list(map(lambda x: x.translate(str.maketrans('', '')), player_value))
df["matches"] = player_matches
df["goals"] = player_goals
df["own_goals"] = player_own_goals
df["assists"] = player_assists
df["yellow_cards"] = player_yellow_cards
df["2_yellow_cards_in_1_match"] = player_2_yellow_cards
df["red_cards"] = player_red_cards
df["matches_when_player_came_from_bench"] = player_comes_from_bench
df["matches_when_player_played_but_then_left_the_pitch"] = player_benched
df["country_association"] = player_country_association

df.to_csv("transfermarkt.csv")
```

Playing data about players and valuation

```
df_league = pd.DataFrame()

df_league["league"] = list_league_name
df_league["club"] = list_teams

df_league.to_csv("club_league_merge.csv")
```

Various leagues and club names

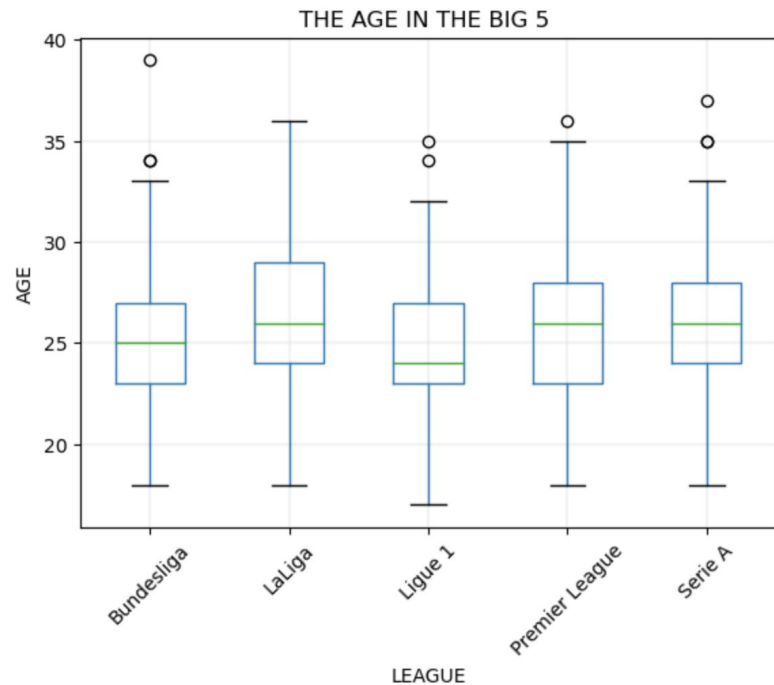
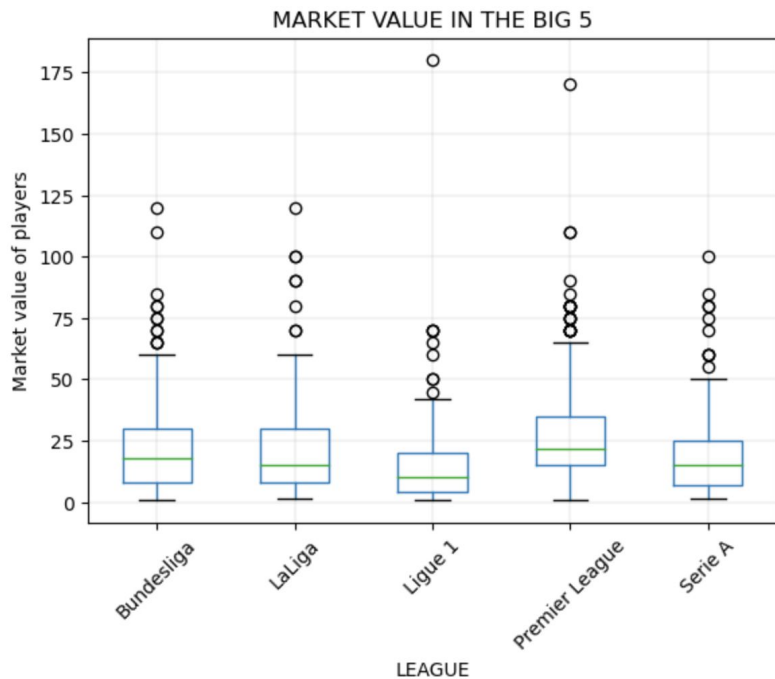
```
df_CL = pd.DataFrame()
df_CL["club"] = list_teams_UCL

df_CL["club_was_in_last_2_seasons_in_UCL"] = 1

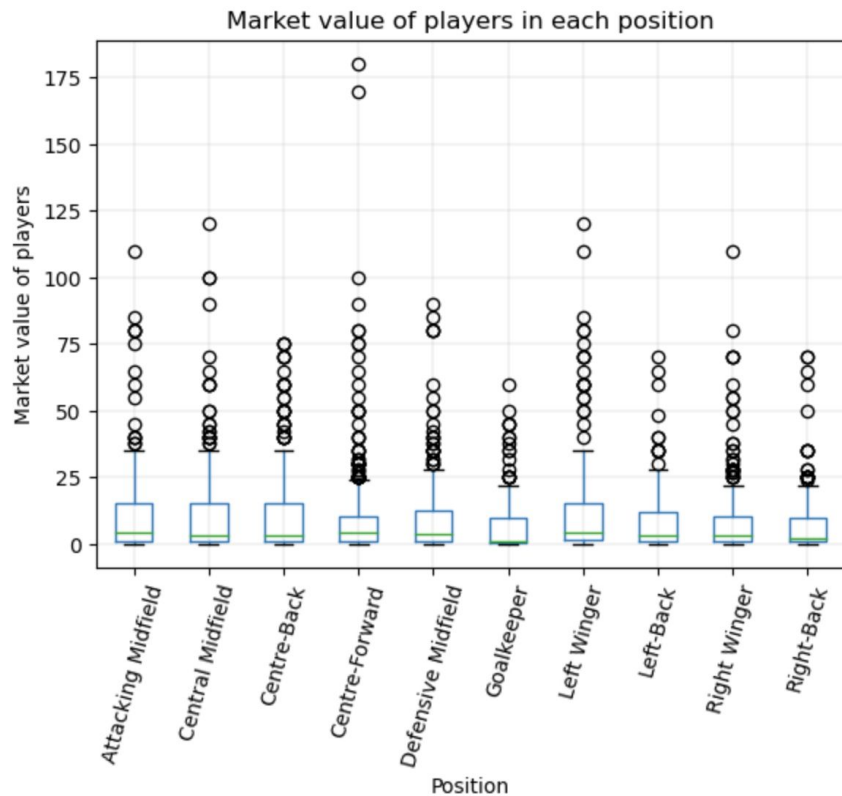
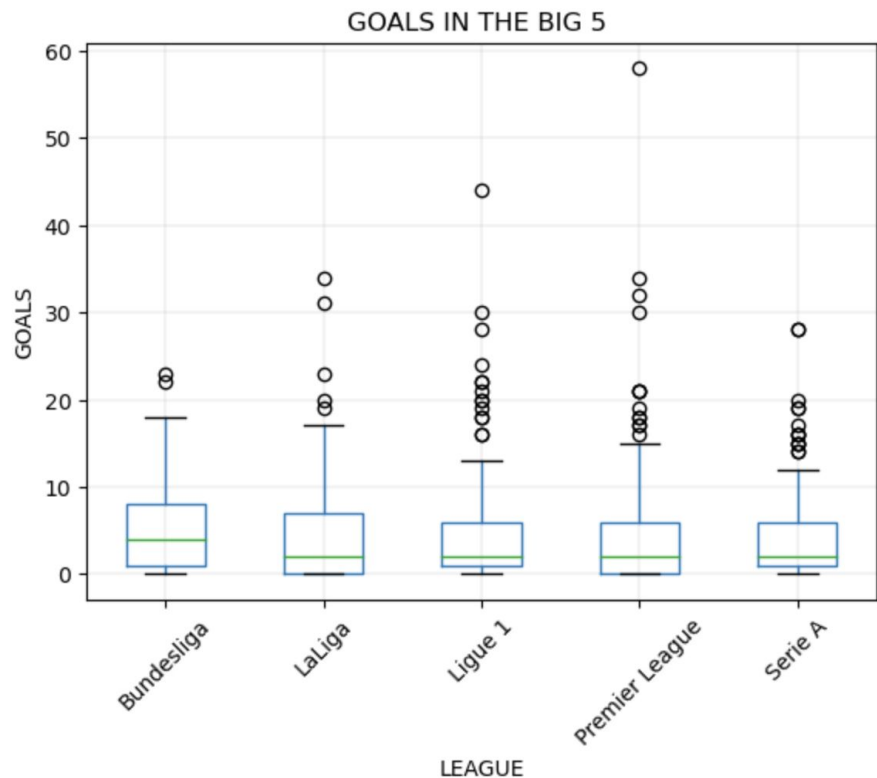
df_CL.to_csv("club_UCL_merge.csv")
```

Champions league club names

Visualisation by Leagues (pt. 1)

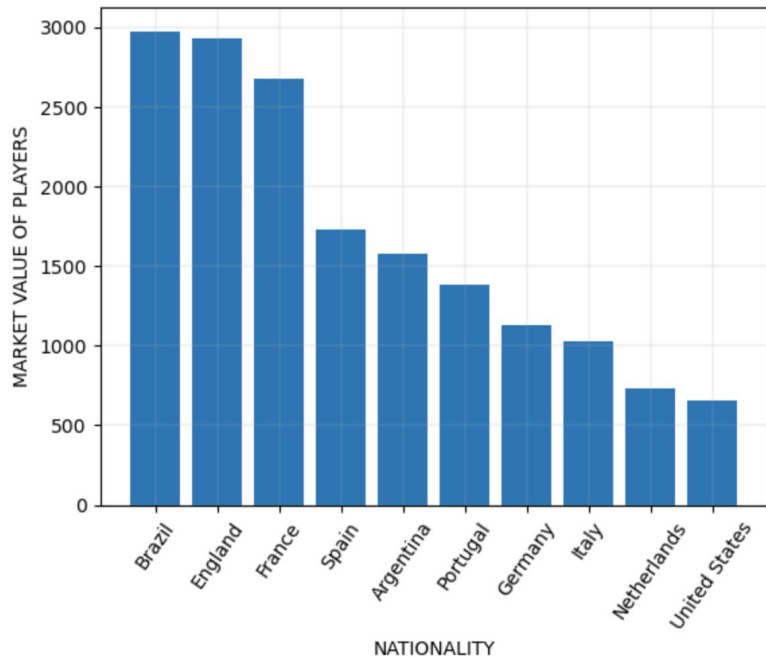


Visualisation by Leagues (pt. 2)

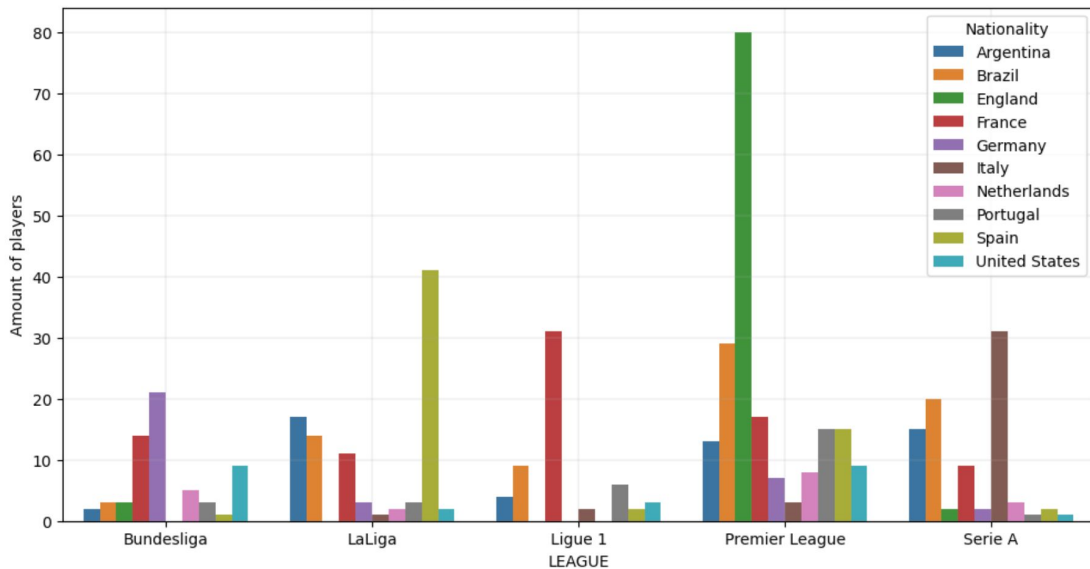


Visualisation by Nationality

TOP 10 MOST EXPENSIVE NATIONALITIES



NUMBER OF PLAYERS OF EACH NATIONALITY IN THE BIG 5



Unsupervised Learning to Link Similar Players

Stats used:

"matches", "goals",
"assists",
"goals_per_match",
"assists_per_match",
"age"

Condition:

Playing in UCL club

19 matches
4 goals
28 years

20 matches
4 goals
28 years

Better value?

#5 Lucas Ocampos
Sevilla FC
LaLiga
League level: First Tier
Joined: Jul 5, 2019
Contract expires: Jun 30, 2024
€12.00m
Last update: Mar 21, 2023

PLAYER OF THE 2022/23 SEASON VOTING
VOTE NOW!

Statistics
Appearances: 19
Goals: 4
Assists: -
Yellow Cards: 7
Second Yellows: -
Red cards: -
Starting eleven: 46%
Minutes: 54%
Goal participation: 12%

NEXT MATCHES
International Friendlies - 8. Matchday
Thursday, 06/15/2023 - 2:00 PM
Argentina vs Australia

NATIONAL TEAM CAREER
National team Debut
17 Argentina Oct 9, 2019 10 2
Argentina U17 - 8 3

#7 Georges-Kevin N'Koudou
Besiktas
Super Lig
League level: First Tier
Joined: Aug 22, 2019
Contract expires: Jun 30, 2023
€4.20m
Last update: Mar 14, 2023

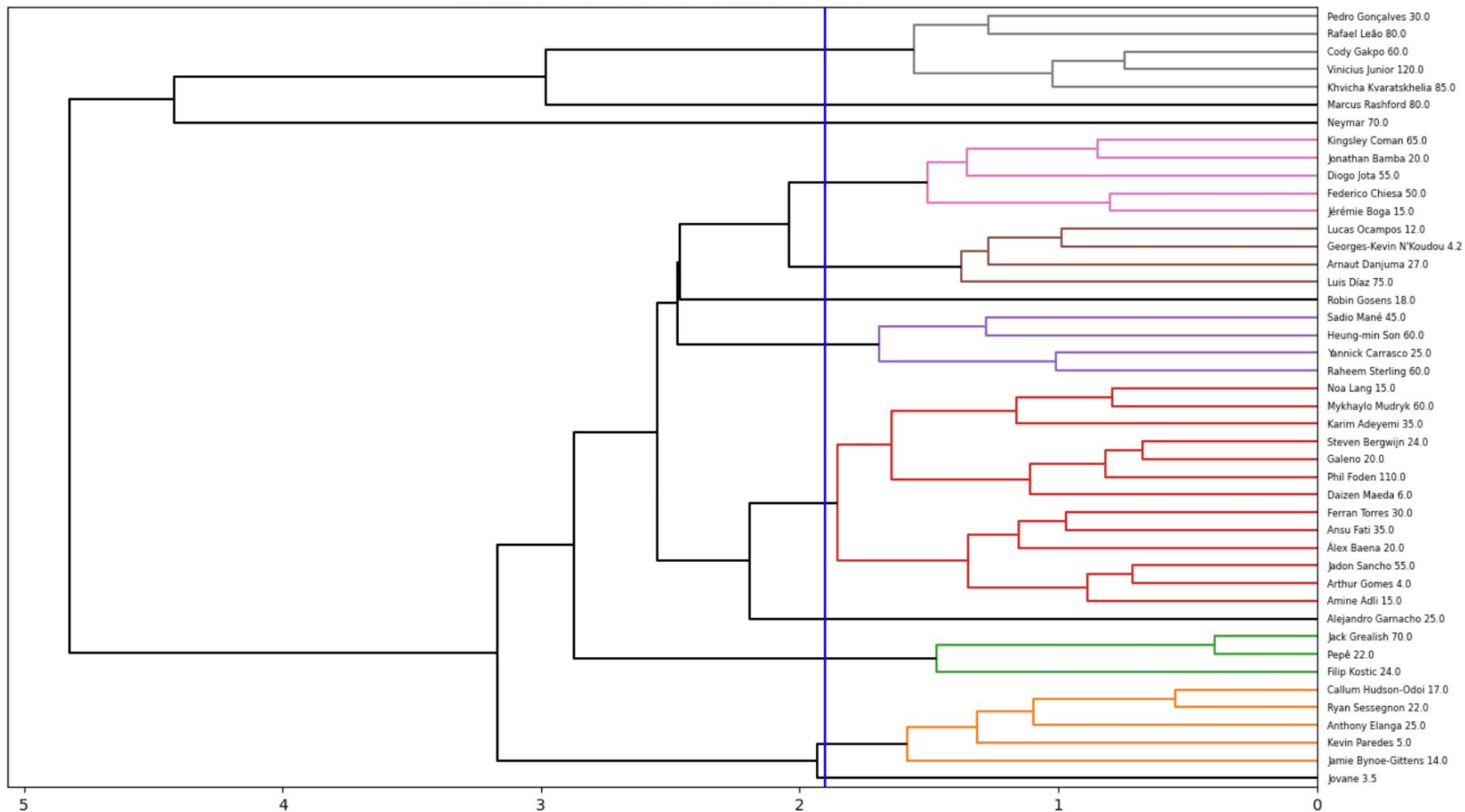
PLAYER OF THE 2022/23 SEASON VOTING
VOTE NOW!

Statistics
Appearances: 20
Goals: 4
Assists: -
Yellow Cards: 2
Second Yellows: -
Red cards: -
Starting eleven: 27%
Minutes: 35%
Goal participation: 5%

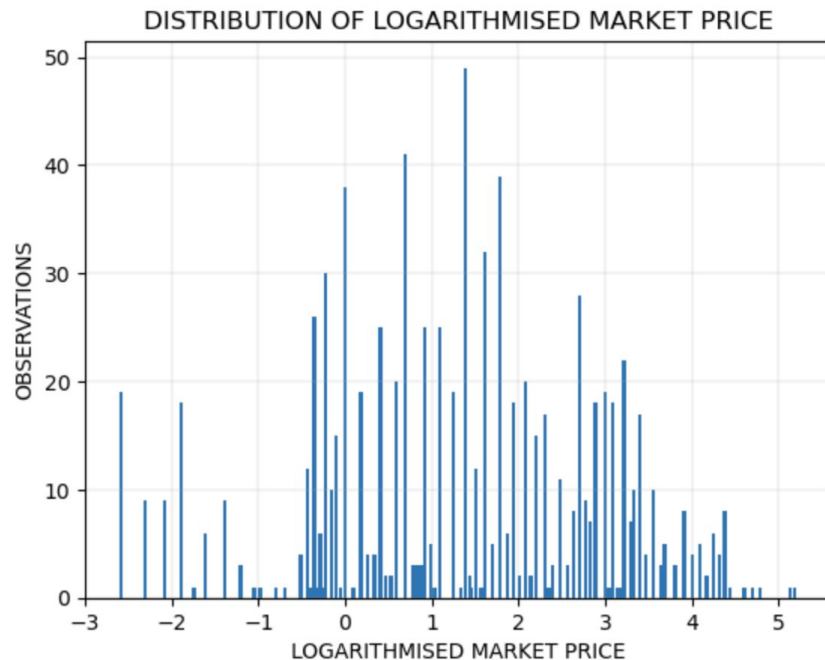
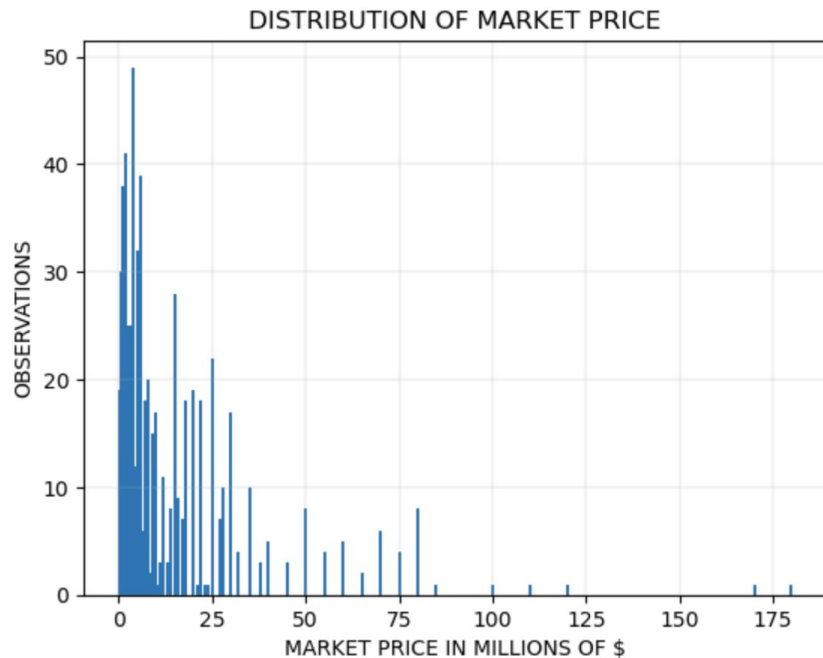
NEXT MATCHES
Super Lig - 38. Matchday
Wednesday, 06/07/2023 - 7:00 PM
Besiktas vs Konyaspor

NATIONAL TEAM CAREER
National team Debut

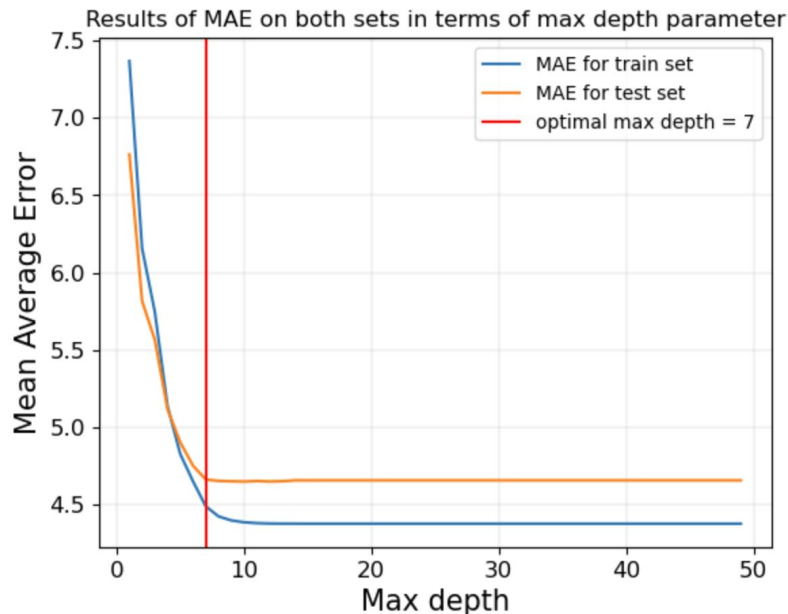
DENDROGRAM FOR LEFT WINGERS



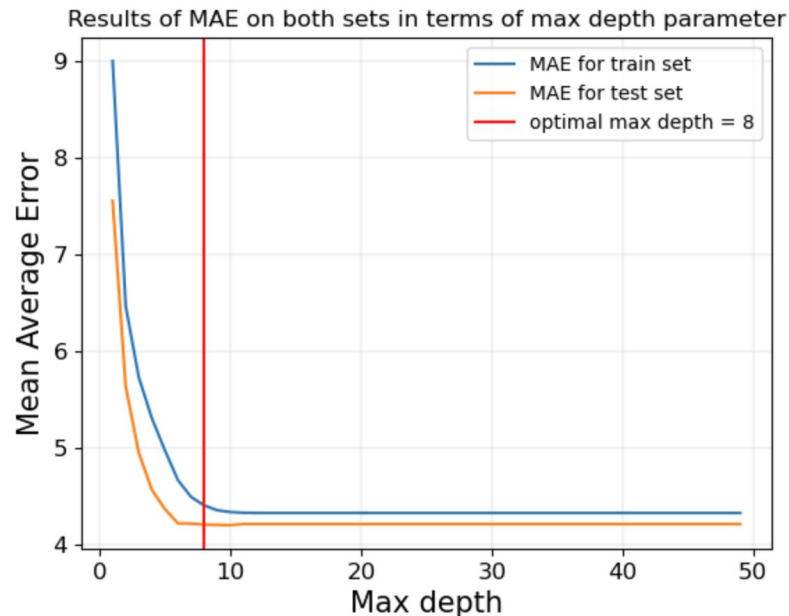
Supervised Learning - Distribution (Log?)



Distribution without changes vs logarithmised distribution



MAE on train set is equal to 4.49
MAE on test set is equal to 4.66



MAE on train set is equal to 4.41
MAE on test set is equal to 4.21

Logarithmic transformation improved quality of predictions.
The logarithmised distribution has lower MAE error on both sets.

Conclusions

- Ballpark value estimations possible with ML
- Potential to find bargains that are for some reason underpriced
- Tool could be useful for big clubs to validate their search

