

Fifa21 Project

EDA:

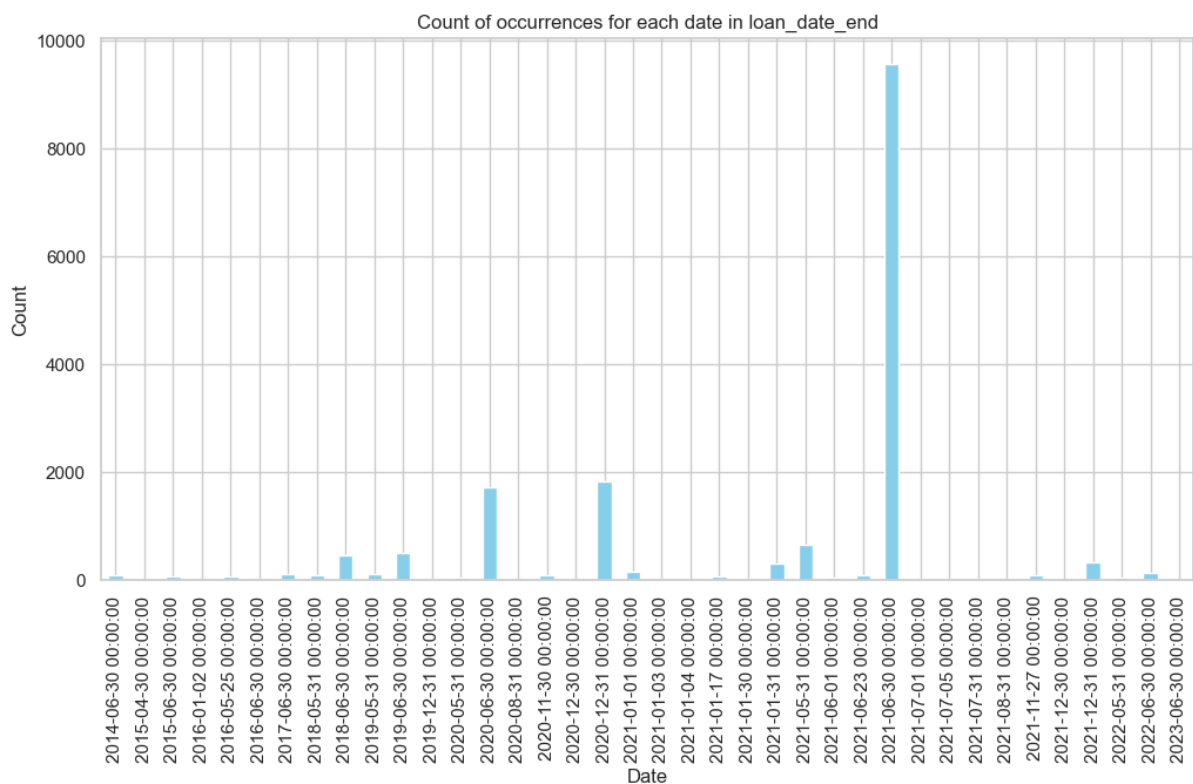
What we did with this project is to do a specific EDA and manage to answer several outcomes.

The data initially was a raw database and in the first instance we noticed that we had some data cleaning to do. After some cleaning we also decided to drop some columns that were useless for our outcomes.

Once we reached the final version of the cleaning we started to work on the outcomes.

Outcomes:

Our first outcome was the count of the loan date of all the players to determine when its the best time to start looking for new players.



As you can see, the best date to start negotiating with the players was at the end of the season, on June 30th.

Next, we started working on the market value of the players.

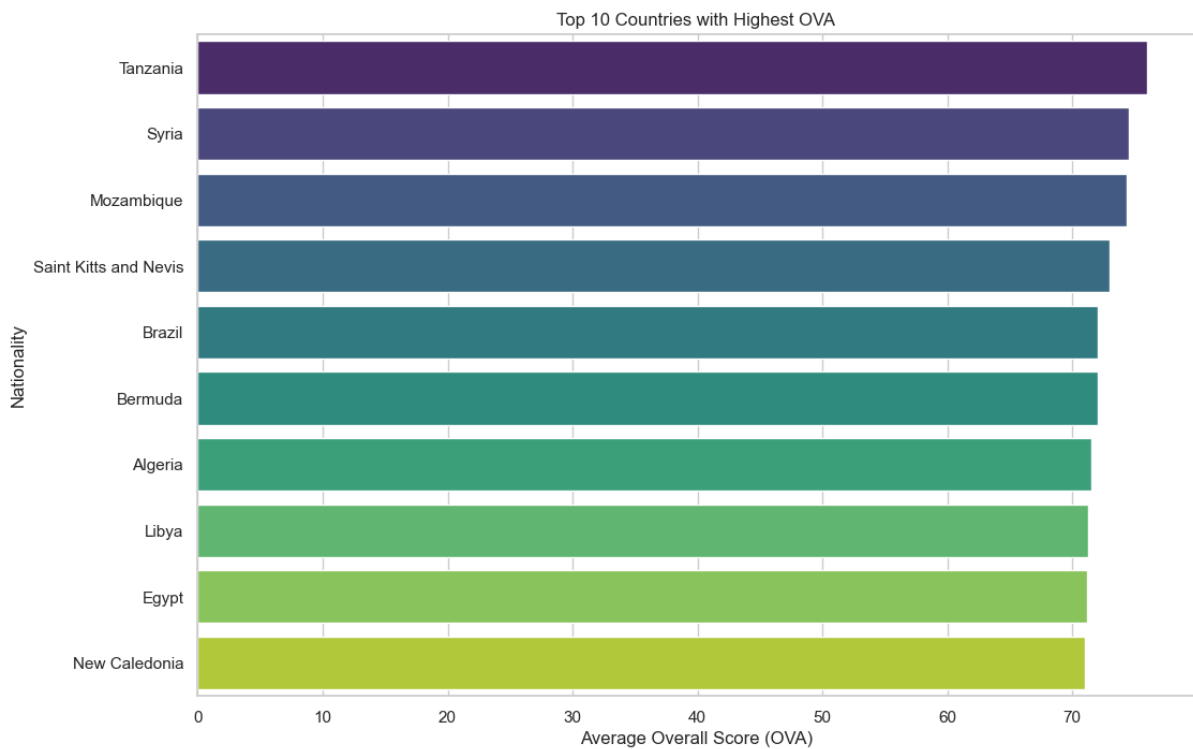
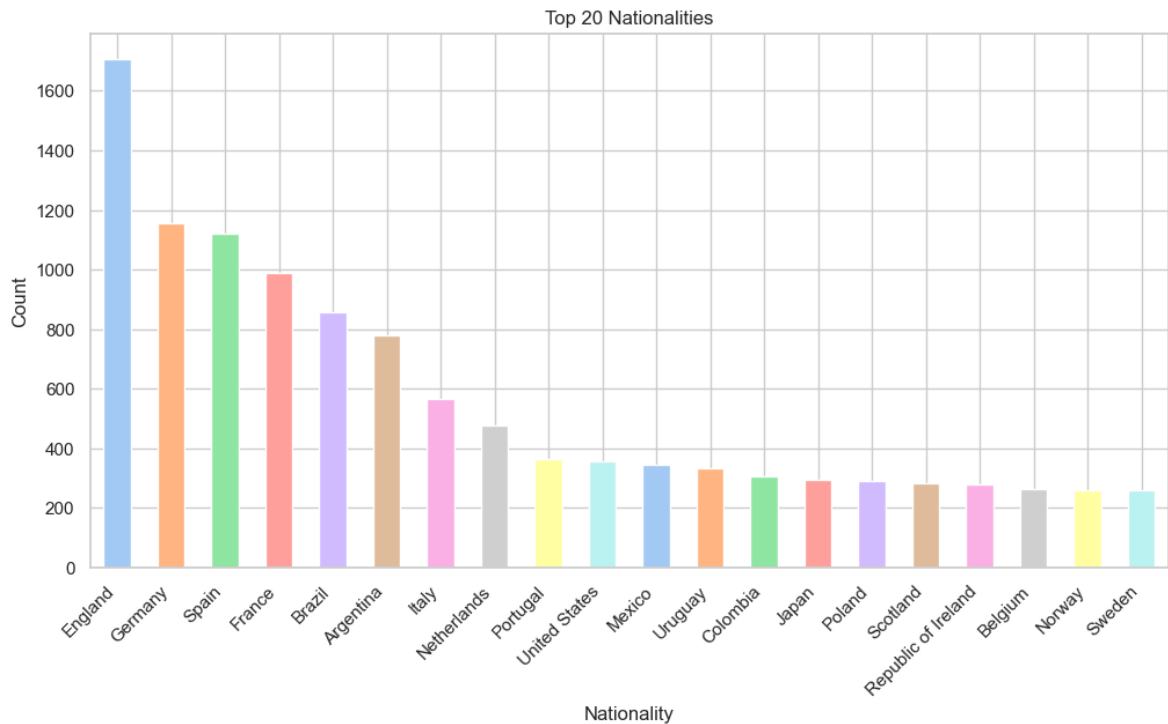
	name	value_in_millions
9331	K. Mbappé	105.5
2650	Neymar Jr	90.0
2871	K. De Bruyne	87.0
2287	R. Lewandowski	80.0
5030	S. Mané	78.0
5109	M. Salah	78.0
4166	V. van Dijk	75.5
3665	J. Oblak	75.0
4049	R. Sterling	72.5
3961	H. Kane	71.0

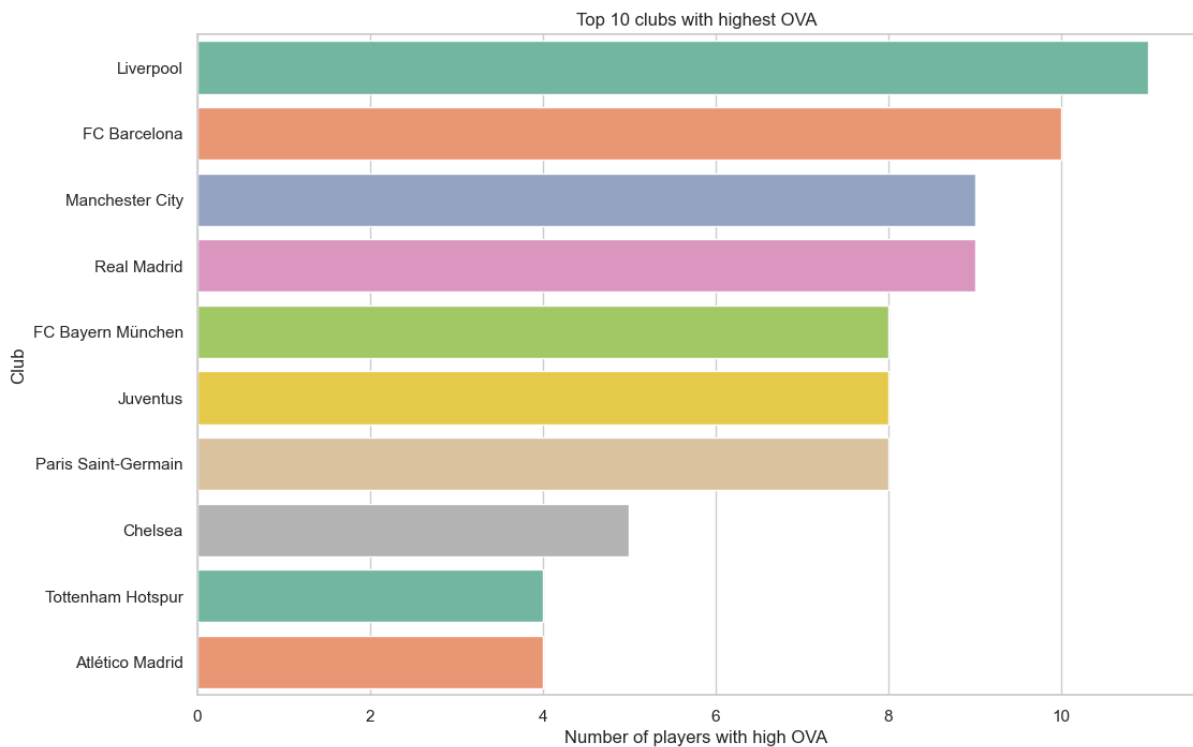
With a simple code we showed the best 10 players of the market in descending order based on the value, before that we cleaned the data and edited the value_in_millions to adapt it to a numerical value. Before that we had categorial values because of the “M” “K” and “€” so we dropped it in order to show the outcome correctly.

Just before that we created the mean, standard and the min/max of some values to determine the info describing some important values:

index	count	mean	std	min	25%	50%	75%	max
Age	17125.0	25.3	4.9	16.0	21.0	25.0	29.0	53.0
Overall Rating	17125.0	67.0	6.9	38.0	62.0	67.0	72.0	93.0
Best Overall	17125.0	67.9	6.6	42.0	64.0	68.0	72.0	93.0
Potential	17125.0	72.5	5.8	47.0	69.0	72.0	76.0	95.0
Value (in millions)	17125.0	2.6	5.4	0.0	0.375	0.8	2.4	105.5

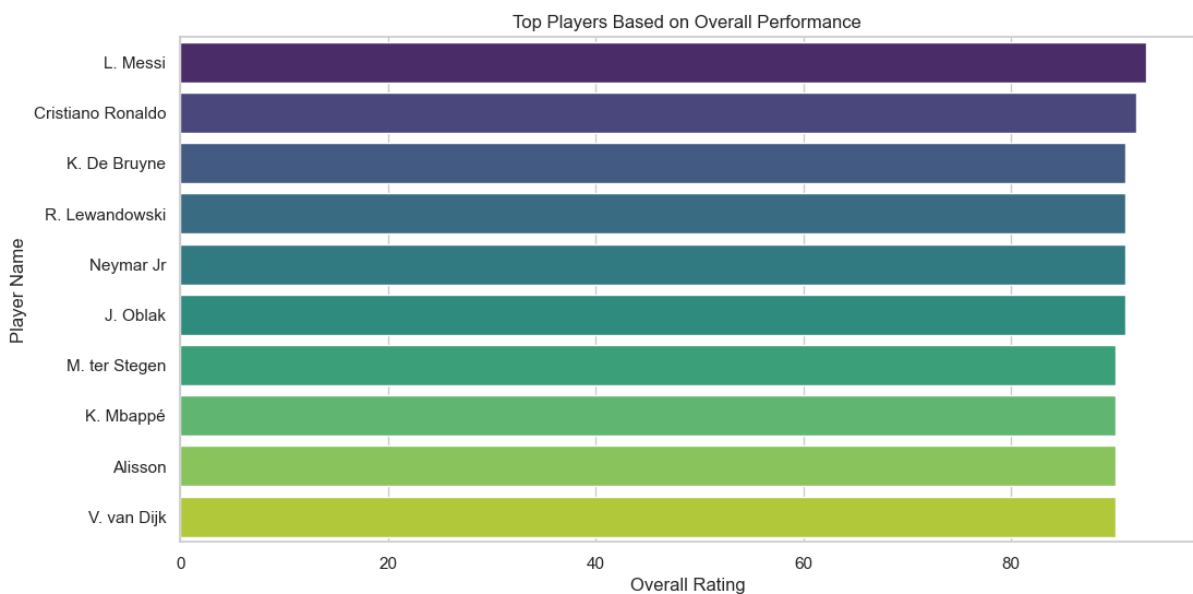
We created a new outcome about the nationwide of all the players, with more info, like the top 20 nationalities, the top 10 countries with highest OVA (overall) and the top 10 clubs with highest OVA in the game.





In this first instance of the date we focused first on the teams, right before that we focused on the players to have an insight of some of the outcomes.

Our first outcome was to reach the best 10 players based on the OVA.



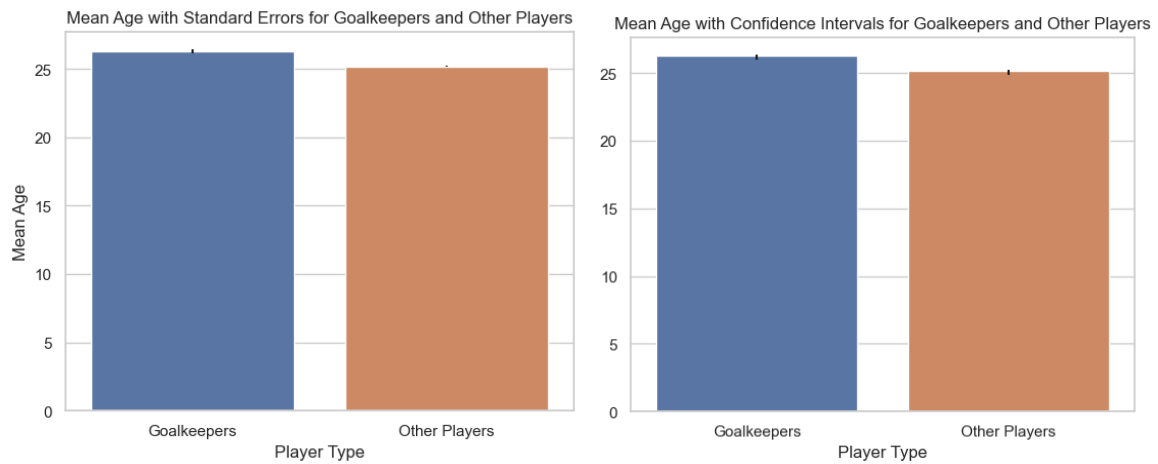
After the first outcome we decided to start working on some of the top 10 roles in the game, divided on:

- Best defenders
- Best scorers
- Best dribblers
- Best goalkeepers
- Best vision
- Best athletes

Showing all of them in different plots:



We included some outcomes that we thought it was interesting to add. Like some outcomes including the goalkeepers and comparing it with the other players.



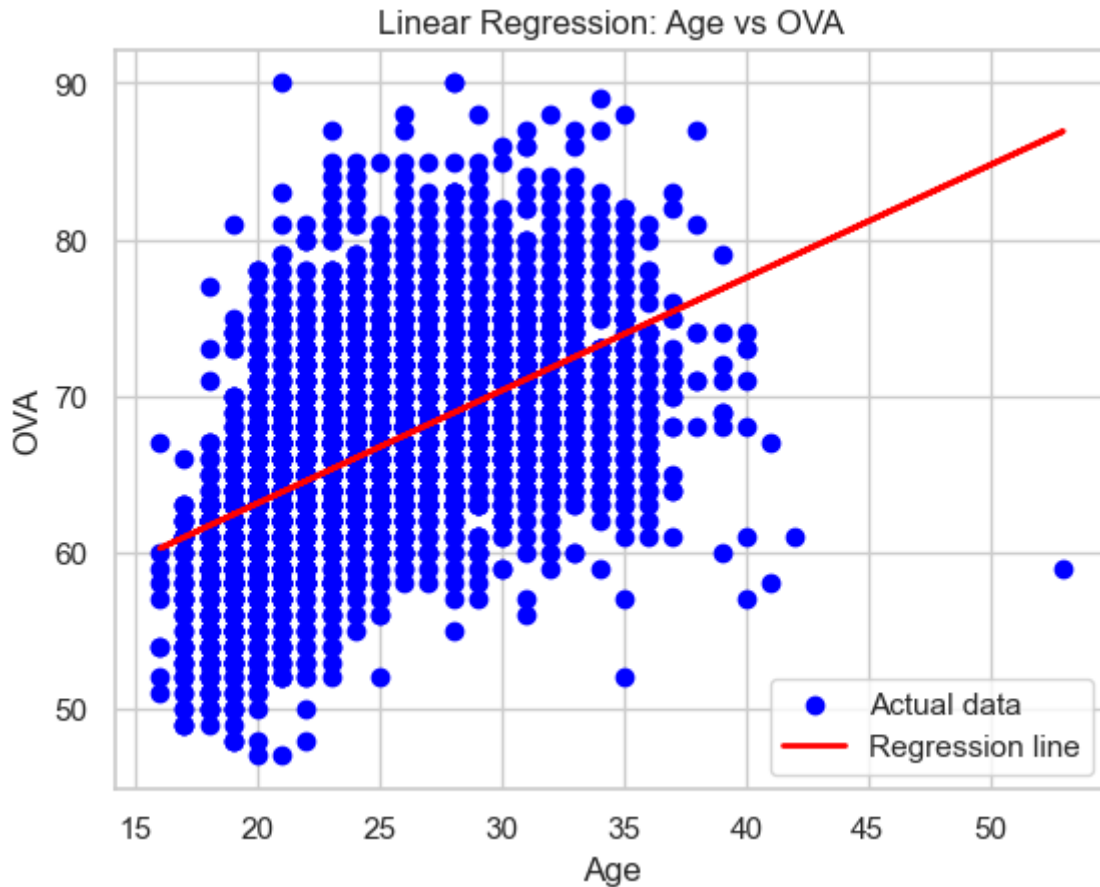
Mean Age for Goalkeepers: 26.28
Confidence Interval for Goalkeepers: (26.00, 26.57)

Mean Age for Other Players: 25.17
Confidence Interval for Other Players: (25.10, 25.25)

And some others really interesting like the best 11 players for a perfect team in the game:

	ova	name	position	attacking	defending	vision	power	finishing	mentality
2 2 8 7	91	R. Lewandowski	ST	423	96	79.0	420	94	391
6 6 2	89	S. Agüero	ST	411	83	83.0	403	94	341
5 0 3 0	90	S. Mané	LW	410	122	85.0	406	90	358
5 1 0 9	90	M. Salah	RW	392	122	84.0	393	91	376
1 7 5 3	88	T. Kroos	CM	397	205	90.0	355	76	378
3 6 3 0	89	Casemiro	CDM	349	259	80.0	437	64	396
6 3 8 7	87	A. Robertson	LB	328	248	79.0	365	57	378
9 2 1 2	87	Alexander-Arnold	RB	365	241	84.0	370	56	371
4 1 6 6	90	V. van Dijk	CB	316	272	65.0	402	52	347
6 8 1	89	Sergio Ramos	CB	374	263	71.0	402	65	414
3 6 6 5	91	J. Oblak	GK	95	57	65.0	268	11	140

R-squared: 0.26



OLS Regression Results

```
=====
=
Dep. Variable:          ova    R-squared:
0.267
Model:                  OLS    Adj. R-squared:
0.267
Method:                  Least Squares    F-statistic:
6250.
Date:                    Sat, 06 Jan 2024    Prob (F-statistic):
0.00
Time:                    11:56:28    Log-Likelihood:
-54623.
No. Observations:        17125    AIC:
1.093e+05
Df Residuals:            17123    BIC:
1.093e+05
Df Model:                 1
Covariance Type:         nonrobust
=====
=
               coef      std err          t      P>|t|      [0.025
0.975]
-----
-
const          48.8155      0.234      208.674      0.000      48.357
49.274
```

```

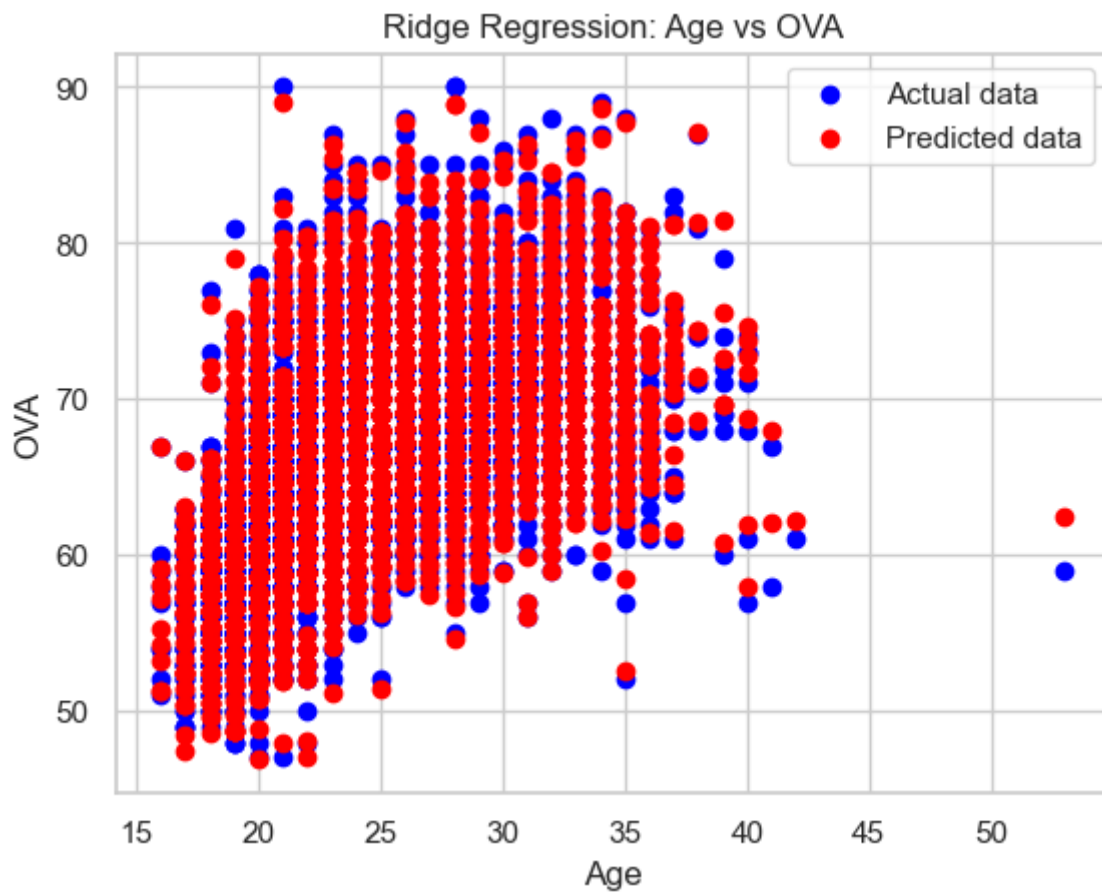
age          0.7181      0.009      79.055      0.000      0.700
0.736
=====
=
Omnibus:                278.506      Durbin-Watson:
1.636
Prob(Omnibus):          0.000      Jarque-Bera (JB):
348.398
Skew:                   0.238      Prob(JB):
2.22e-76
Kurtosis:              3.511      Cond. No.
134.
=====
=

```

After this analysis we decided to create another machine learning on Spearman method just to compare the data:

Mean Squared Error: 1.05

R-squared: 0.98



Mean Squared Error: 1.05
R-squared: 0.98
Coefficients: [0.08314047 0.96810844]
Intercept: -0.86

OLS Regression Results

```
=====
=
Dep. Variable:          ova    R-squared:
1.000
Model:                OLS    Adj. R-squared:
1.000
Method:              Least Squares    F-statistic:
1.201e+25
Date:                Sat, 06 Jan 2024    Prob (F-statistic):
0.00
Time:                11:56:29    Log-Likelihood:
3.8706e+05
No. Observations:    17125    AIC:
-7.740e+05
Df Residuals:        17075    BIC:
-7.736e+05
Df Model:              49
Covariance Type:      nonrobust
=====
==
```

	coef	std err	t	P> t	[0.025	0.975]
const	-1.51e-14	8.58e-12	-0.002	0.999	-1.68e-11	1.68e-11
id	-1.653e-16	1.26e-17	-13.170	0.000	-1.9e-16	-1.41e-16
age	8.882e-16	1.49e-13	0.006	0.995	-2.91e-13	2.93e-13
bov	-7.327e-15	3.12e-13	-0.023	0.981	-6.19e-13	6.04e-13
pot	1.0000	3.06e-13	3.27e+12	0.000	1.000	1.000
growth	-1.0000	3.09e-13	-3.24e+12	0.000	-1.000	-1.000
attacking	-1.283e-13	5.49e-14	-2.337	0.019	-2.36e-13	-2.07e-14