Football Players Value Prediction

Burak Arslan

Introduction

Is it possible to determine a player's value

by his statistics ?

Methodology

1. Data Fetch & Data process
   a) Web Scraping
      I. 2019/2020 Season Statistics
      II. Football Manager 2021 Game Player Values
      III. Transfermarkt.de Player Values
2. Exploratory Data Analysis
   a) Feature Tuning
3. OLS & Regression Models
4. Prediction

Exploratory Data Analysis

Results

All Players

R^2 score : 0.69

Defenders

R^2 score : 0.68

Midfielders

R^2 score : 0.79

Forwards
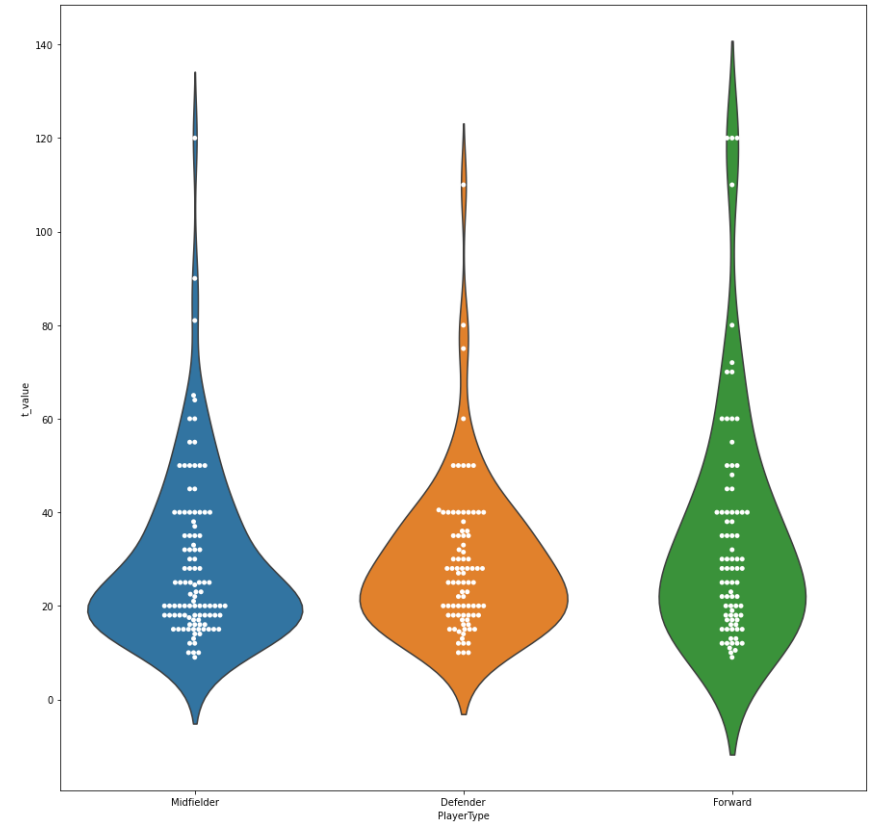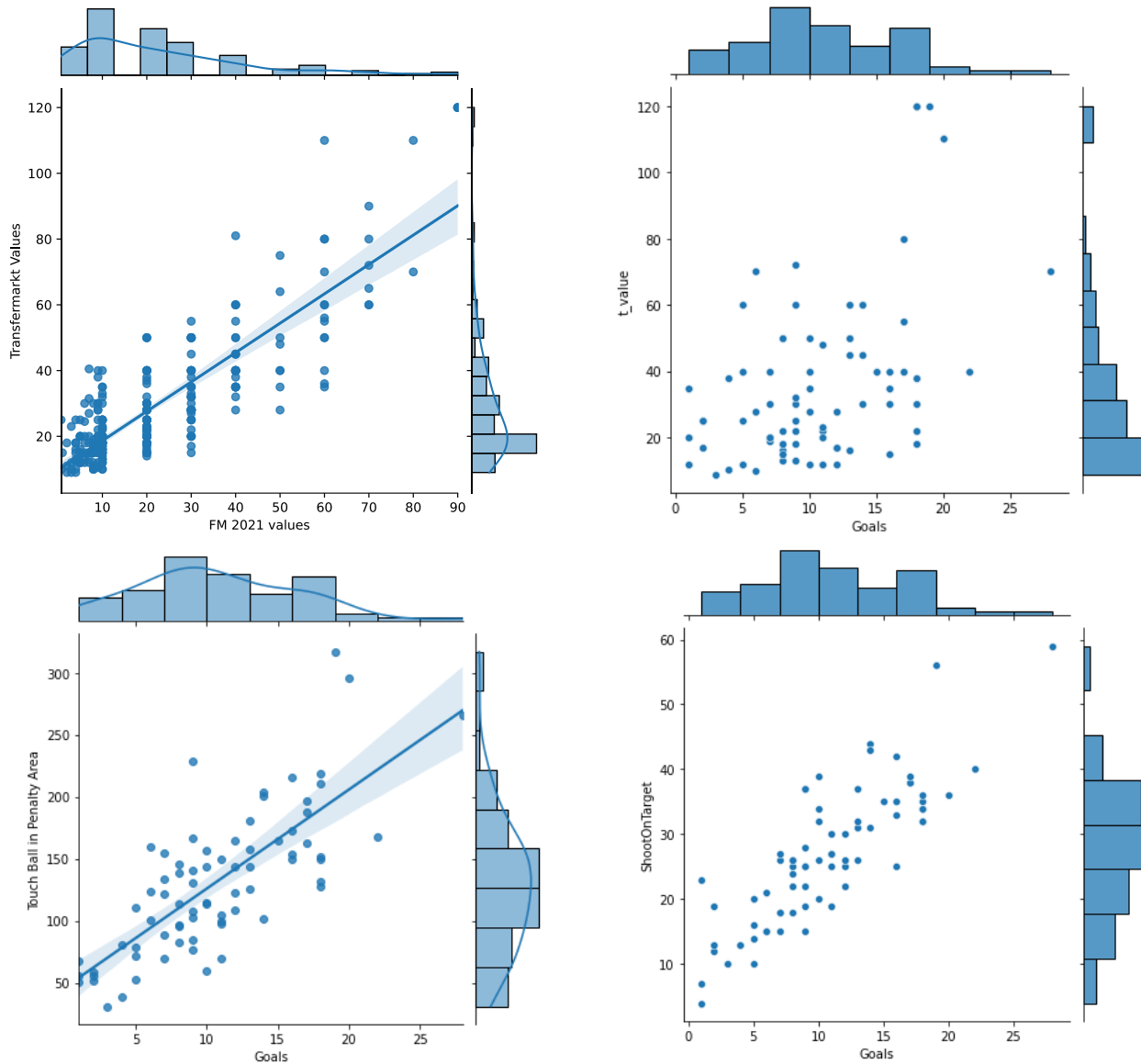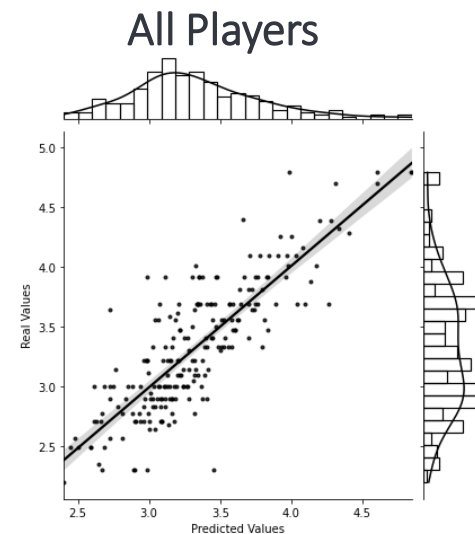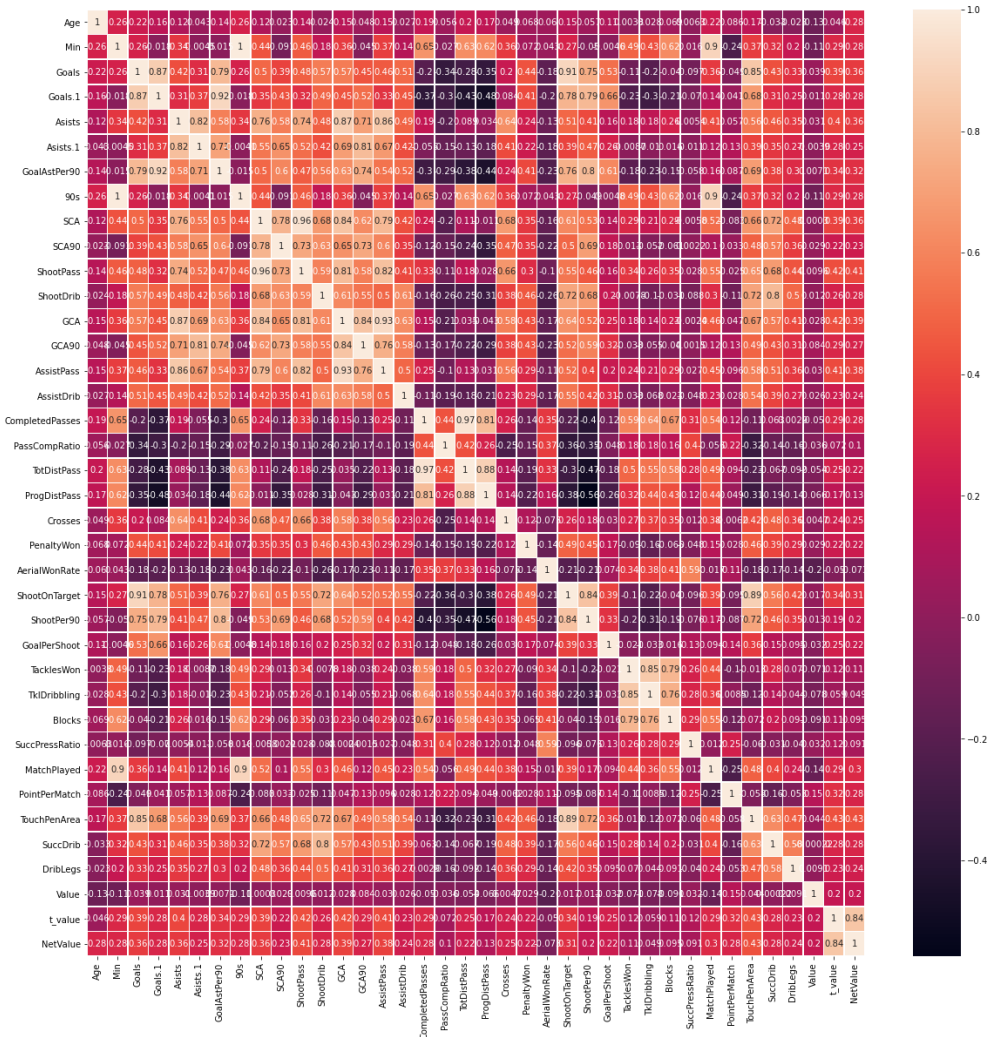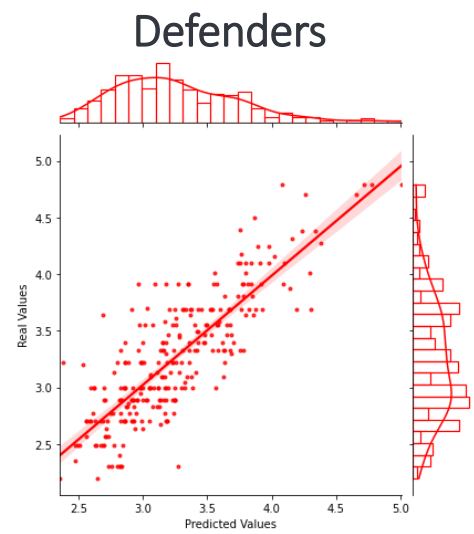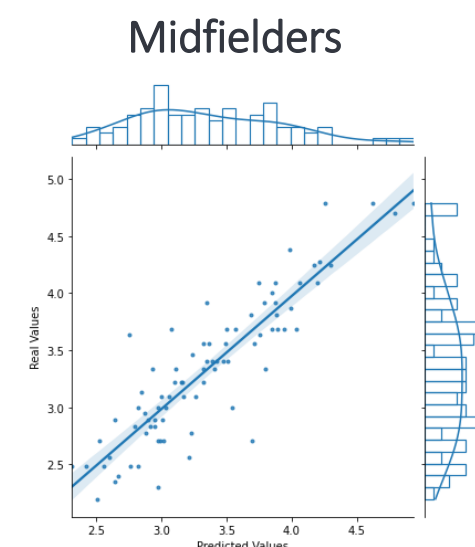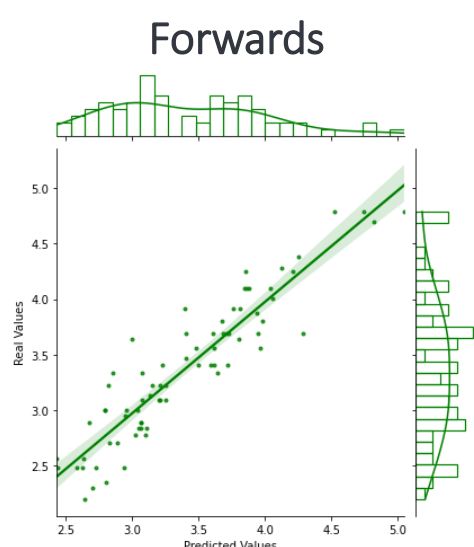
R^2 score : 0.86

Results

New Features for Better Fitting

1- Agegroup instead of Age
2- Winner : (Point per Match Played ^ 2 * Played Match count)
3- Goal Change Creator instead of Assist, Shoot on Target, Dribbling lead to assist, pass lead to assist, Penalty won, Successful pressing at 3rd Area
4- Features that has high correlation between each other had beed removed.

Features that were converted to categorical variable into dummy/indicator variables :
1- League
2- Player Position

And many features had been removed to overcome overfitting !

Players that were substitute mostly, also had been removed.

Natural logarithm of Transfermarkt values

OLS Regression Results (left)

Dep. Variable:	v_value	R-squared (uncentered):	0.865
Model:	OLS	Adj. R-squared (uncentered):	0.847
Method:	Least Squares	F-statistic:	46.68
Date:	Sun, 10 Jan 2021	Prob (F-statistic):	3.95e-91
Time:	13:14:21	Log-Likelihood:	-1157.9
No. Observations:	289	AIC:	2386.
Df Residuals:	254	BIC:	2514.
Df Model:	35		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Age	-1.1072	0.291	-3.801	0.000	-1.681	-0.534
Min	-0.6481	0.346	-1.874	0.062	-1.329	0.033
Goals	1.6613	0.714	2.326	0.021	0.255	3.068
Goals.1	-141.0973	218.693	-0.645	0.519	-571.780	289.585
Asists	1.5685	1.160	1.352	0.178	-0.716	3.853
Asists.1	-137.4894	218.400	-0.630	0.530	-567.596	292.617
GoalAstPer90	131.2717	218.570	0.601	0.549	-299.168	561.711
90s	58.1142	31.115	1.868	0.063	-3.161	119.390
SCA	-0.1565	0.161	-0.974	0.331	-0.473	0.160
SCA90	0.3174	2.215	0.143	0.886	-4.045	4.680
ShootPass	0.2867	0.196	1.460	0.145	-0.100	0.673
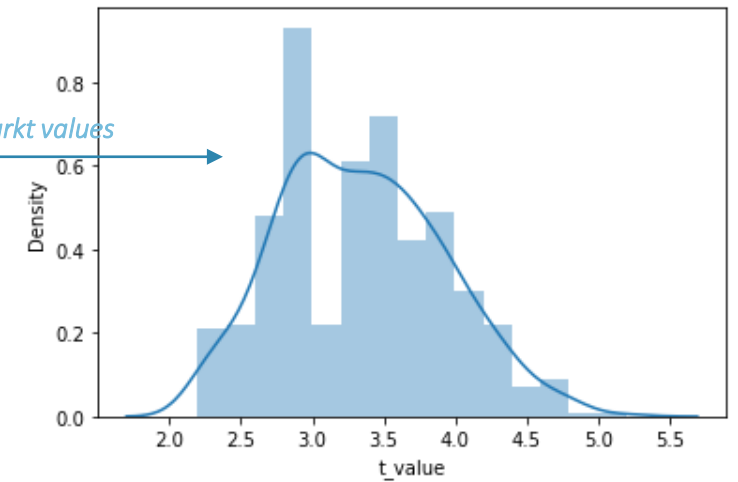
OLS Regression Results (right)

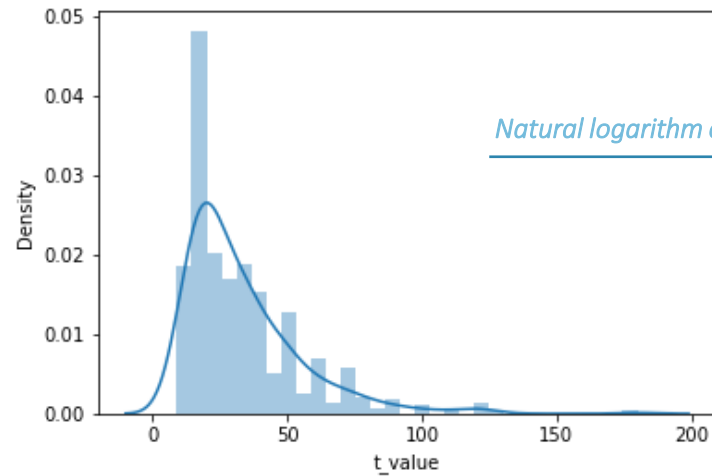| Dep. Variable:	v_value	R-squared (uncentered):	0.984
Model:	OLS	Adj. R-squared (uncentered):	0.982
Method:	Least Squares	F-statistic:	442.1
Date:	Sun, 10 Jan 2021	Prob (F-statistic):	3.79e-207
Time:	13:19:11	Log-Likelihood:	-158.64
No. Observations:	289	AIC:	387.3
Df Residuals:	254	BIC:	515.6
Df Model:	35		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Age	-0.0106	0.009	-1.158	0.248	-0.029	0.007
Min	-0.0217	0.011	-1.996	0.047	-0.043	-0.000
Goals	0.0399	0.022	1.773	0.077	-0.004	0.084
Goals.1	-4.9898	6.890	-0.724	0.470	-18.558	8.579
Asists	0.0098	0.037	0.269	0.788	-0.062	0.082
Asists.1	-4.4356	6.881	-0.645	0.520	-17.986	9.115
GoalAstPer90	4.5283	6.886	0.658	0.511	-9.032	18.089
90s	1.9621	0.980	2.002	0.046	0.032	3.893
SCA	0.0010	0.005	0.206	0.837	-0.009	0.011

Conclusions

- For more accuracy, models must build on each player type.
- For more good results, we need more data, not more features.
- The statistics is not a fair enough to determine players value.
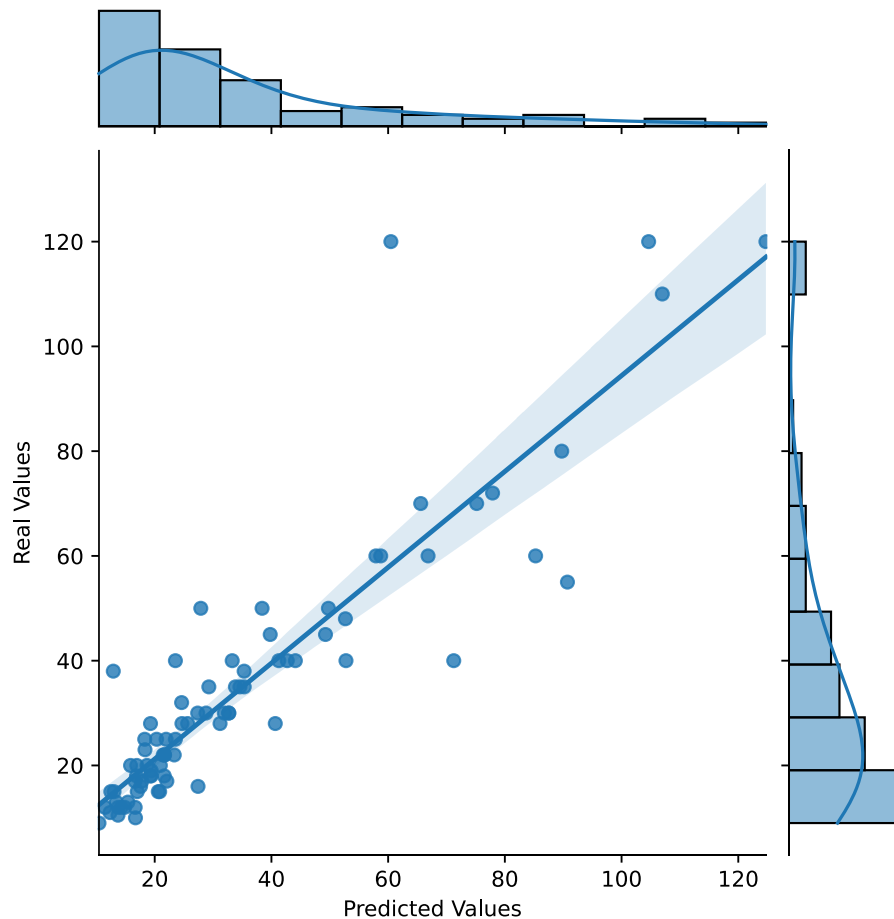
PlayerName	Transfermarkt Value	Prediction	Difference
Mohamed Salah	120.0	155.318826	35.318826
Harry Kane	120.0	92.070160	-27.929840
Sadio Mané	120.0	114.180262	-5.819738
Raheem Sterling	110.0	123.729914	13.729914
Marcus Rashford	80.0	70.329963	-9.670037
Roberto Firmino	72.0	61.710520	-10.289480
Bernardo Silva	70.0	46.966574	-23.033426
Richarlison	60.0	46.549537	-13.450463
Christian Pulisic	60.0	48.336902	-11.663098
Phil Foden	60.0	56.552633	-3.447367
Gabriel Jesus	60.0	47.389728	-12.610272
Anthony Martial	55.0	57.795870	2.795870
Mason Greenwood	50.0	45.297065	-4.702935
Jack Grealish	50.0	43.081864	-6.918136
Riyad Mahrez	48.0	51.430697	3.430697
Dominic Calvert-Lewin	45.0	53.338864	8.338864
Nicolas Pépé	40.0	36.905394	-3.094606
Pierre-Emerick Aubameyang	40.0	39.929488	-0.070512
Diogo Jota	40.0	41.584178	1.584178
Raúl Jiménez	40.0	72.393873	32.393873
Tammy Abraham	40.0	51.913139	11.913139
Sergio Agüero	35.0	37.190653	2.190653
Alexandre Lacazette	35.0	32.565067	-2.434933
Callum Hudson-Odoi	35.0	52.801773	17.801773
Sébastien Haller	30.0	37.046302	7.046302
Moise Kean	25.0	25.875327	0.875327

Future Work

Data Enrichment Options & Challenges :

- Substitute players : Missing Stats.
- Team members may affect each other's statistics.
- Prone to injury.
- Last year on contract.
- Player home country popularity.
- Can play big matches.
- Player leadership.
- Work rate & ethics.
- Web search popularity.
- The economic strength and prestige of the club.