Predicting Market Value - English Premier League Players

Market Value

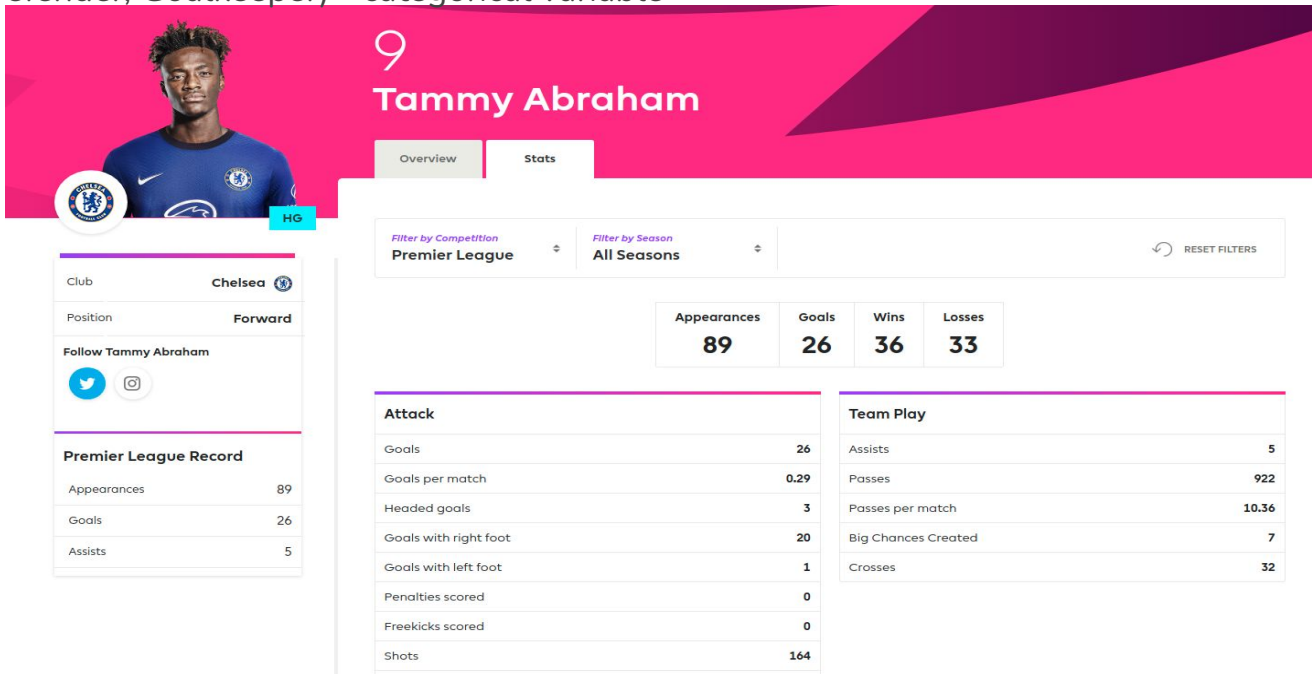
- Important for Clubs to successfully manage transfer market
- Market value of a soccer player is crucial → Potential Transfer Cost, Salary($$$)
- Determined by several components(age, # of goals, # of passes...etc)

Objective

- Predict market value of players in current English Premier League season based on cumulative stats recorded from 1992 ~ 2021.
- Scrape from EPL + Transfermarkt.com

Features Included

- Appearance
- Position(Forward, Midfielder, Defender, Goalkeeper) - categorical variable
- Goals
- Assists
- Passes
- Passes per match
- Fouls
- Yellow cards
- Red cards



Attack

Goals	26
Goals per match	0.29
Headed goals	3
Goals with right foot	20
Goals with left foot	1
Penalties scored	0
Freekicks scored	0
Shots	164
Shots on target	68
Shooting accuracy %	41%
Hit woodwork	3
Big chances missed	30

Defence

Tackles	25
Blocked shots	31
Interceptions	12
Clearances	55
Headed Clearance	45

Player A

Attack

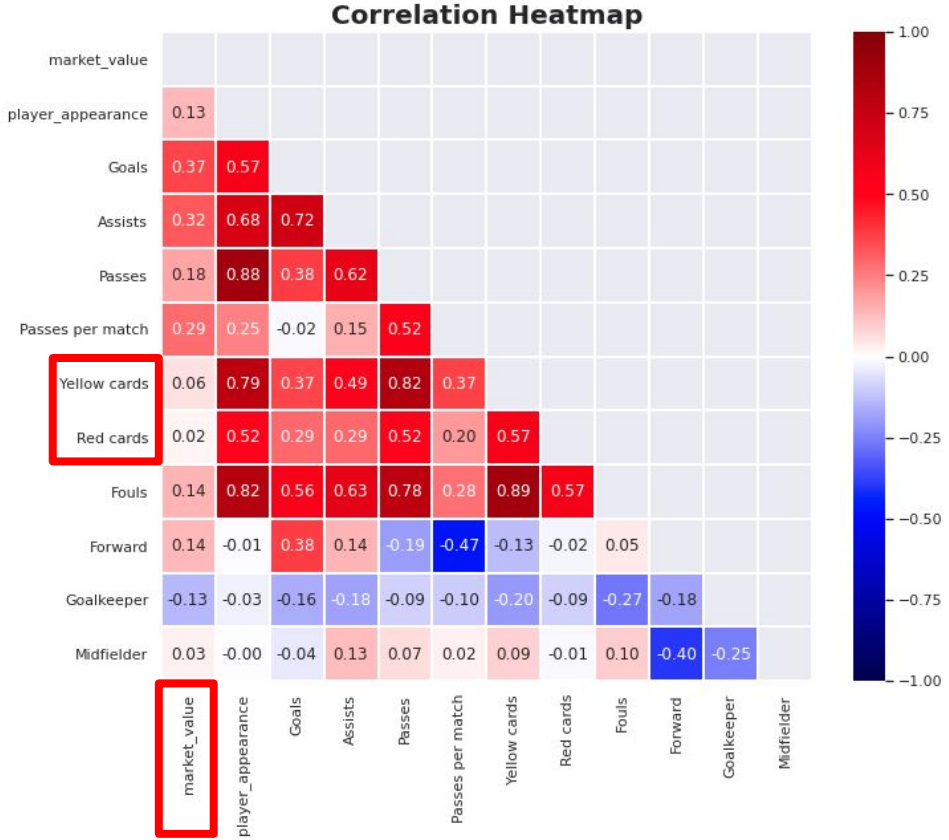
Goals	0
Headed goals	0
Goals with right foot	0
Goals with left foot	0
Hit woodwork	0

Player B

Defence

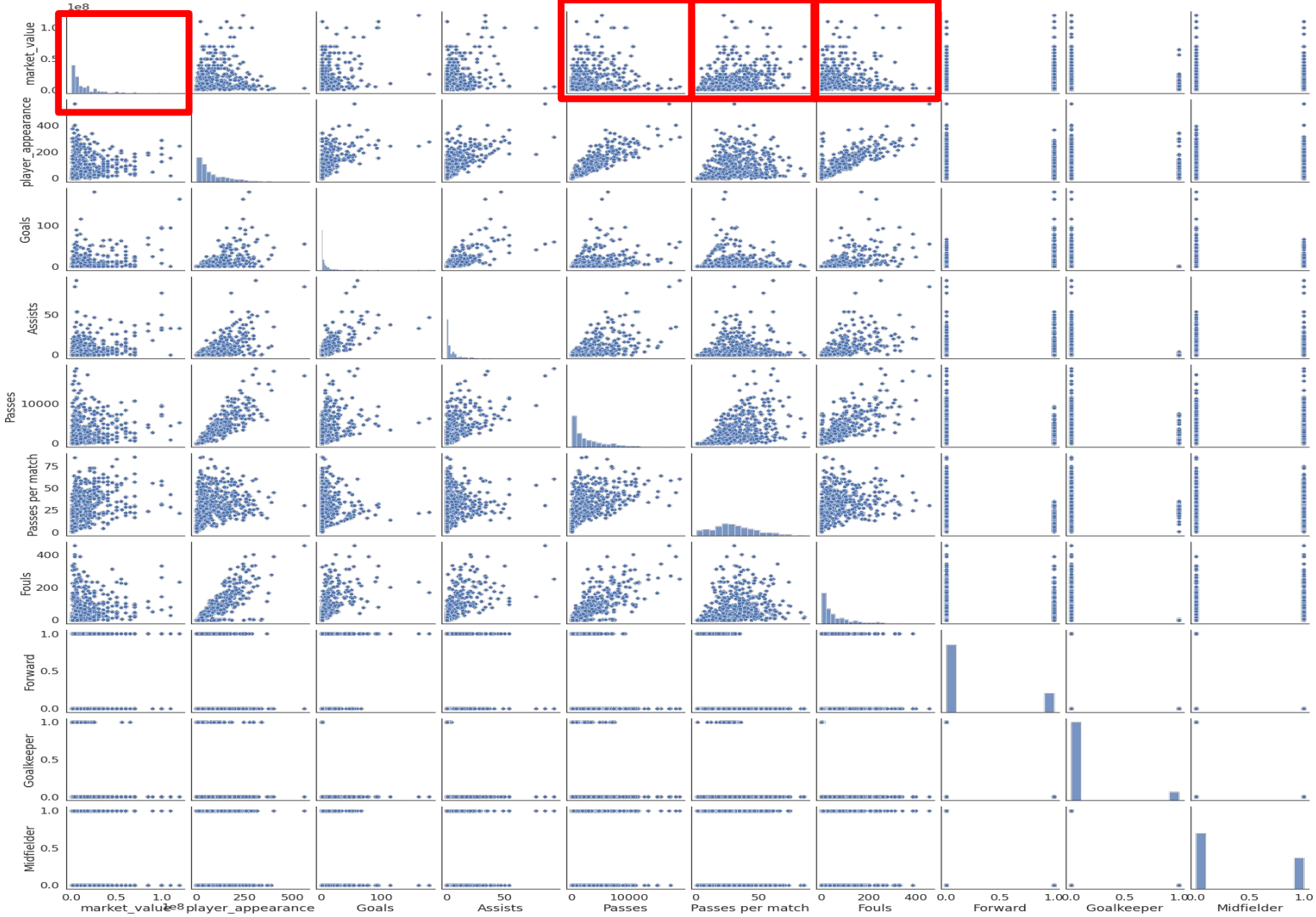
Clean sheets	9
Goals Conceded	35
Tackles	34
Tackle success %	56%
Last man tackles	0
Blocked shots	2
Interceptions	39
Clearances	174
Headed Clearance	89
Clearances off line	0
Recoveries	125

EDA to select/remove certain features



EDA Pair Plot Analysis

- Market value highly skewed
- Distribution of Market Value for certain features show slight hint of exponential



OLS Regression Results

Dep. Variable:	Q('market_value')	**R-squared:**	0.313
Model:	OLS	**Adj. R-squared:**	0.303
Method:	Least Squares	**F-statistic:**	29.19
Date:	Thu, 13 May 2021	**Prob (F-statistic):**	6.26e-42
Time:	12:24:48	**Log-Likelihood:**	-10550.
No. Observations:	586	**AIC:**	2.112e+04
Df Residuals:	576	**BIC:**	2.116e+04
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-5.827e+06	2.65e+06	-2.195	0.029	-1.1e+07	-6.14e+05
player_appearance	-3.608e+04	2.46e+04	-1.464	0.144	-8.45e+04	1.23e+04
Goals	3.842e+05	6.45e+04	5.956	0.000	2.58e+05	5.11e+05
Assists	2.482e+05	1.09e+05	2.278	0.023	3.42e+04	4.62e+05
Passes	454.7049	650.566	0.699	0.485	-823.065	1732.475
Q('Passes per match')	5.574e+05	6.52e+04	8.556	0.000	4.29e+05	6.85e+05
Fouls	-5.768e+04	1.72e+04	-3.350	0.001	-9.15e+04	-2.39e+04
Forward	1.352e+07	2.49e+06	5.424	0.000	8.62e+06	1.84e+07
Goalkeeper	1.741e+06	2.67e+06	0.651	0.515	-3.51e+06	6.99e+06
Midfielder	6.247e+06	1.78e+06	3.511	0.000	2.75e+06	9.74e+06

Omnibus:	158.155	**Durbin-Watson:**	1.947
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	562.523
Skew:	1.225	**Prob(JB):**	7.07e-123
Kurtosis:	7.128	**Cond. No.**	2.48e+04

OLS Regression Results

Dep. Variable:	Q('Log market_value')	**R-squared:**	0.319
Model:	OLS	**Adj. R-squared:**	0.308
Method:	Least Squares	**F-statistic:**	29.94
Date:	Thu, 13 May 2021	**Prob (F-statistic):**	6.78e-43
Time:	12:32:02	**Log-Likelihood:**	-911.80
No. Observations:	586	**AIC:**	1844.
Df Residuals:	576	**BIC:**	1887.
Df Model:	9		
Covariance Type:	nonrobust		

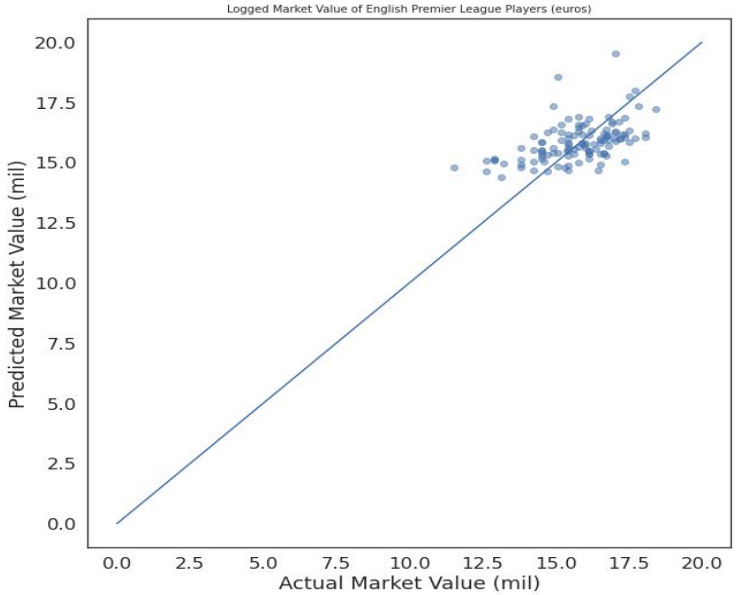
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	13.7081	0.191	71.796	0.000	13.333	14.083
player_appearance	0.0052	0.002	2.913	0.004	0.002	0.009
Goals	0.0131	0.005	2.824	0.005	0.004	0.022
Assists	0.0086	0.008	1.101	0.271	-0.007	0.024
Passes	-0.0001	4.68e-05	-2.925	0.004	-0.000	-4.5e-05
Q('Passes per match')	0.0582	0.005	12.416	0.000	0.049	0.067
Fouls	-0.0038	0.001	-3.069	0.002	-0.006	-0.001
Forward	1.0630	0.179	5.930	0.000	0.711	1.415
Goalkeeper	-0.2679	0.192	-1.393	0.164	-0.646	0.110
Midfielder	0.5265	0.128	4.114	0.000	0.275	0.778

Omnibus:	16.682	**Durbin-Watson:**	2.098
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	17.685
Skew:	-0.424	**Prob(JB):**	0.000144
Kurtosis:	2.935	**Cond. No.**	2.48e+04

Regression Modeling

1. Linear Regression
2. Ridge Regression
3. Lasso Regression

Cross Validation Test on Training Data	R-Squared Value
Linear Regression	0.293709
Ridge Regression	0.293776
Lasso Regression	0.239135

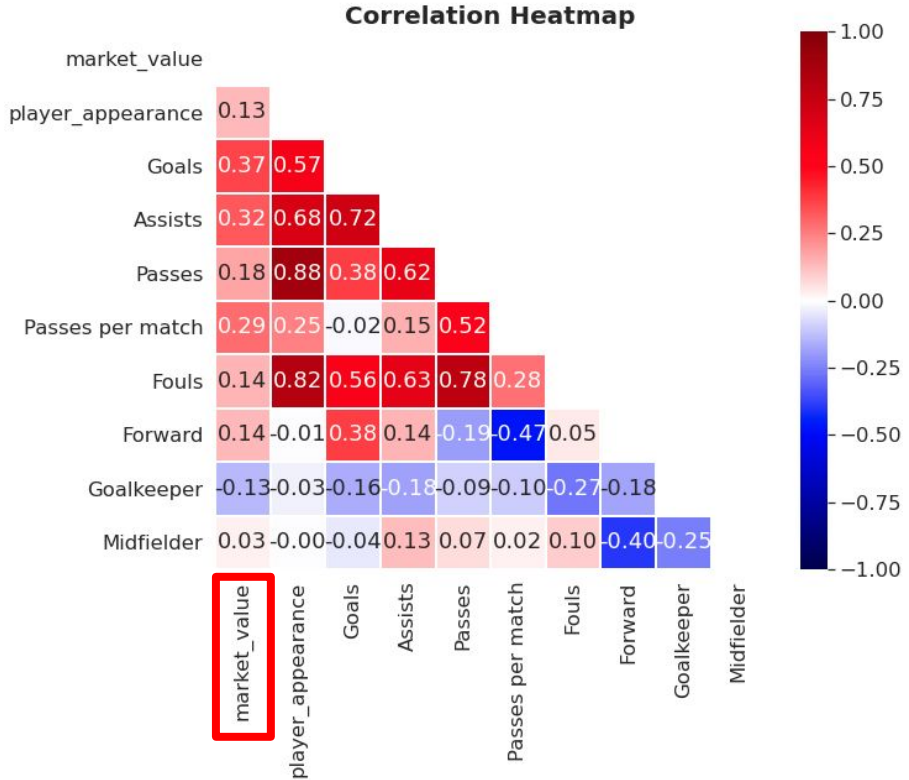


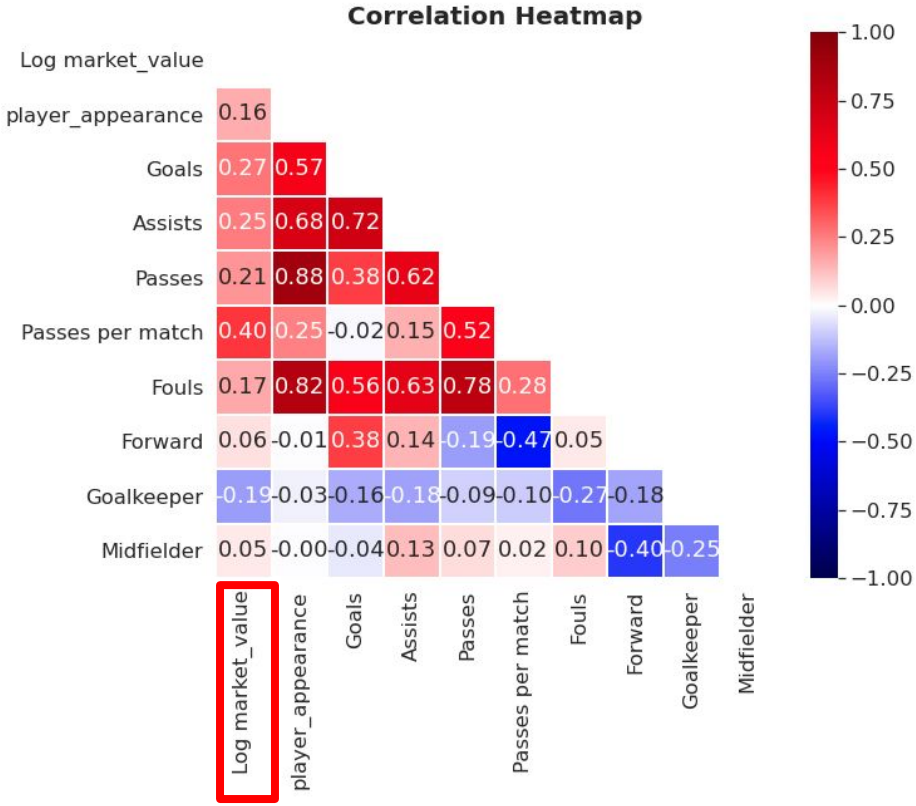
Result of Test set through chosen model : Ridge

R-Squared Value	0.234976
MAE	0.929720

Future Work/Improvements

- Consider data collection from other source
- Look closely into polynomial regression
- Increase dataset


Correlation Heatmap
market_value
player_appearance 0.13
Goals 0.37 0.57
Assists 0.32 0.68 0.72
Passes 0.18 0.88 0.38 0.62
Passes per match 0.29 0.25 -0.02 0.15 0.52
Fouls 0.14 0.82 0.56 0.63 0.78 0.28
Forward 0.14 -0.01 0.38 0.14 -0.19 -0.47 0.05
Goalkeeper -0.13 -0.03 -0.16 -0.18 -0.09 -0.10 -0.27 -0.18
Midfielder 0.03 -0.00 -0.04 0.13 0.07 0.02 0.10 -0.40 -0.25
market_value
player_appearance
Goals
Assists
Passes
Passes per match
Fouls
Forward
Goalkeeper
Midfielder
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Correlation Heatmap
Log market_value
player_appearance 0.16
Goals 0.27 0.57
Assists 0.25 0.68 0.72
Passes 0.21 0.88 0.38 0.62
Passes per match 0.40 0.25 -0.02 0.15 0.52
Fouls 0.17 0.82 0.56 0.63 0.78 0.28
Forward 0.06 -0.01 0.38 0.14 -0.19 -0.47 0.05
Goalkeeper -0.19 -0.03 -0.16 -0.18 -0.09 -0.10 -0.27 -0.18
Midfielder 0.05 -0.00 -0.04 0.13 0.07 0.02 0.10 -0.40 -0.25
Log market_value
player_appearance
Goals
Assists
Passes
Passes per match
Fouls
Forward
Goalkeeper
Midfielder
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00