FIFA 2019 PLAYER DATASET

GROUP 18

PRITHVI GUDE

PHANINDRA PANTHAGANI

MOHANA KRISHNA JALADI

AKHIL CHELIKANI

ASHEESH NARESH CHAND

ASHWIN KONDAPALLI

Table of Contents

1. INTRODUCTION

1.1 BACKGROUND

Whenever a player is placed on the transfer market, his market value plays a key role in the transfer negotiations. Market value of the player is basically the transfer fee estimate of the player estimated by football experts. If a player has no transfer history, then his market value cannot be estimated. This is one of the limitations of this approach.

1.2 OUR SUGGESTION

This limitation can be overcome using a data driven approach, using a dataset comprising of around 18000 players we ran multiple regressions using Forward, Backward and Stepwise selection with AIC as the criteria for getting the best model with the best predictors to predict a market value and the regression results suggested that data driven approach is a much more reliable method of estimating market values as opposed to the traditional method of estimating market values.

1.3 DATA CLEANING

1. Some of the players had zero market values in the dataset, this was because these players had no Transfer history and hence, they have a market value estimate of zero. These were only 240 out of 18000 observations which is only 1.35 % of the Data.

2. The Dataset contained the column work rate which contained both defensive and attacking work rates, we split the column into defensive and attacking work rates

3. The market value column in the dataset contained values in millions and k's and we converted all the values in the column to k's

4. Also, we Created a new column for Position, as for players who were not goalkeepers Positions like center back, left center back and right center back had the same position ratings. So, we decided to consider them as a single position.

   - Center backs, right center backs and left center backs can be classified as center backs
   - Left backs and right backs can be classified as full backs
   - Right wing backs and left wing backs can be classified as wing backs

- Left attacking midfielders, right attacking midfielders and center attacking midfielders can be classified as attacking midfielders
- Left center midfielders, right center midfielders and center midfielders can be classified as center midfielders
- Left midfielders and right midfielders can be classified as wide midfielders
- Right defensive midfielders left defensive midfielders and center defensive midfielders can be classified as defensive midfielders
- Center forwards, left forwards, right forwards can be classified as center forwards
- Left strikers, right strikers and strikers can be classified as strikers
- Left wingers, right wingers can be classified as wingers
- The rest of the players are classified as goalkeepers

1.4 QUESTIONS WE DEALT WITH IN THE EDA

1. How does the age affect the market value of the players? Is there a trend in the market values of the players by age what is the age at which the player reaches his highest market value?
2. Are players being discriminated by nationality?
3. Are players being discriminated based on positions?
4. How do the Attacking and Defensive work rates affect the market value? If there is a Difference, then how is it distributed across different work rates?
5. If the club is looking for a player to use him in a specific position, what are the important skills that the manager should be looking for that position? (and)
6. How is Overall Rating calculated? What are the skills and the attributes involved in overall rating calculation?

1.5 MAKING POTENTIAL RATING A CATEGORICAL VARIABLE

The Dataset Has a column called Potential (rating on a scale of 1 to 100) which is basically an indicator of how good a player can be. Instead of considering the column as a continuous variable It is better if we divide the Potential rating into 4 classes.

- The 1st class contains players having potential greater than 90, These players have the potential to be special like Messi or Ronaldo

- The 2nd class contains players having potential rating between 85 to 89, These players are the exciting prospects
- The 3rd class contains players having potential rating between 80 to 84, These players are showing great potential
- The rest of the players are classified as class 4

2. EXPLORATORY DATA ANALYSIS

2.1 DISTRIBUTION OF PLAYERS BASED ON INTERNATIONAL REPUTATION

International_Reputation	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	16532	91.04	16532	91.04
2	1261	6.94	17793	97.98
3	309	1.70	18102	99.69
4	51	0.28	18153	99.97
5	6	0.03	18159	100.00
Frequency Missing = 48				

Players with International reputation of 1 are the most in this data. Also, there are only 6 players with International Reputation of 5.

2.2 SCATTERPLOT MATRIX - INTERPRETATION

Scatterplot Matrix for Data
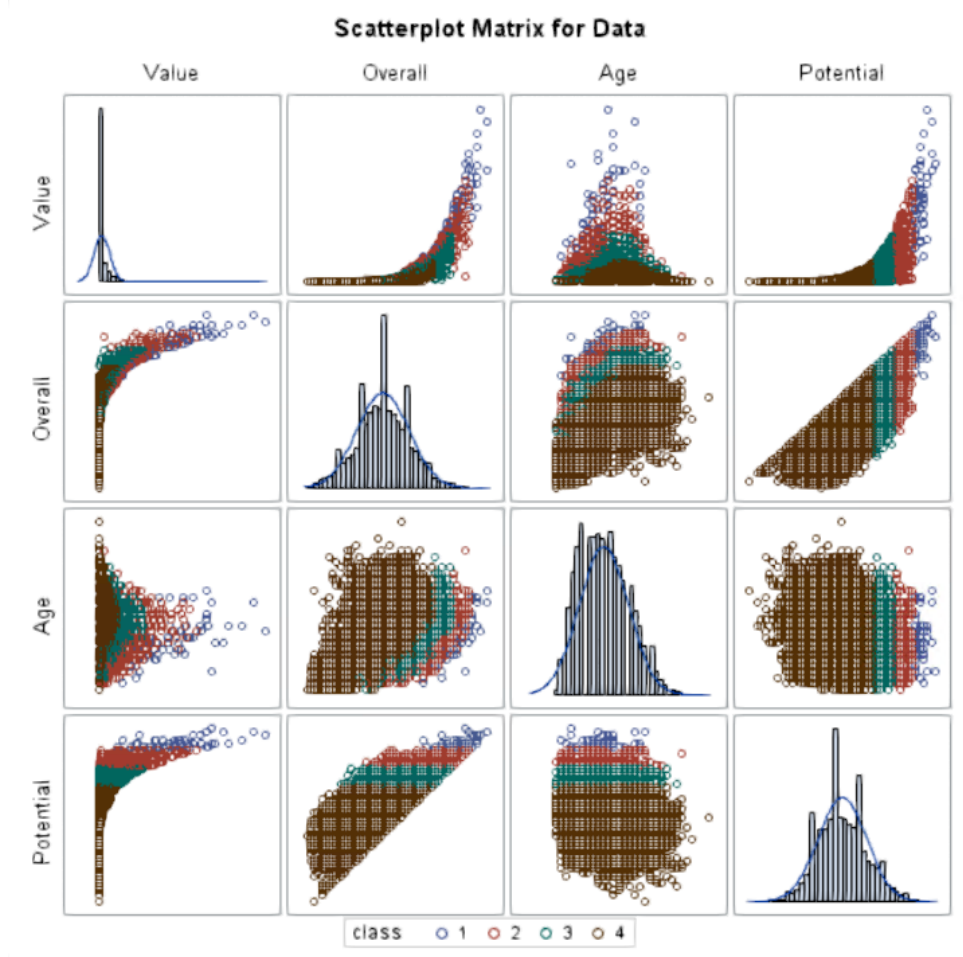
We can see from the Value vs Overall Scatterplot that as the Overall rating increases, the Market Value of the player increases Exponentially.

This kind of a relationship is seen in the Value vs Potential scatterplot but if we look closer we can observe that there are many players with higher potential Rating whose overall rating is lower, suggesting the presence of the players who are underrated.

2.3 EFFECT OF AGE ON VALUE
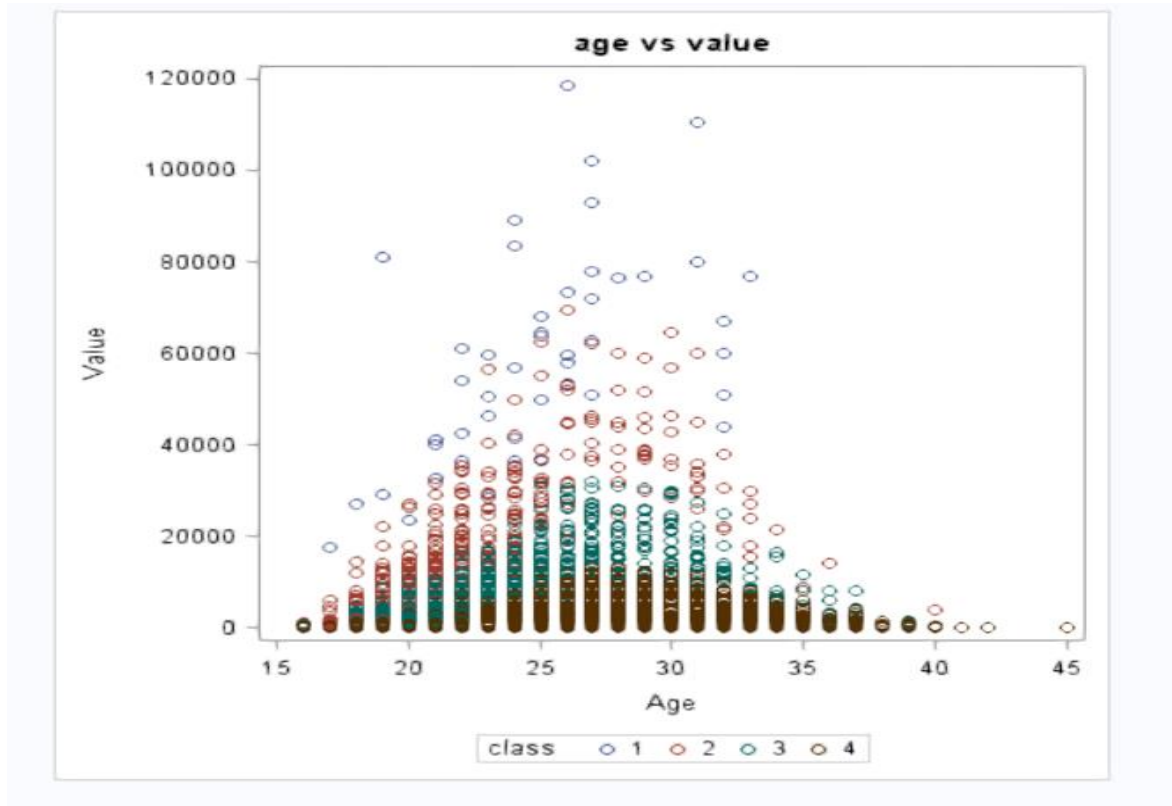
As we can see from the above plot, there seems to be more players with higher market values in the middle as opposed to the Extremes. So, there is an indication that the value of the players initially increases with age and after it reaches a peak it decreases. However, we cannot confirm this from the scatter plot as there is too much variation in the data.

So, in order to capture the trend of value by age we took a class by class approach to it as we figured that different classes might have different trends.

The goal of the below analysis is not to get accurate estimates but to get the best fit to the data taking only age to see the trend. So, we choose an appropriate degree for age so that we get best R square

Below are the results for the regressions run for taking value as the dependent variable and age as the independent variable.

Regression Results for Class 1 for R square with degree 3 gives the best R square

R-Square	Coeff Var	Root MSE	value Mean
0.433213	30.86442	17941.49	58130.00

Regression Results for Class 2 for R square with degree 3 gives the best R square

R-Square	Coeff Var	Root MSE	value Mean
0.625049	41.89666	8828.737	21072.65

Regression Results for Class 3 for R square with degree 3 gives the best R square

R-Square	Coeff Var	Root MSE	value Mean
0.652334	46.42752	3874.275	8344.781

For Class 4 we could not see any trend as there was too much variation in the data

Fit plots for Class 1, 2 and 3:
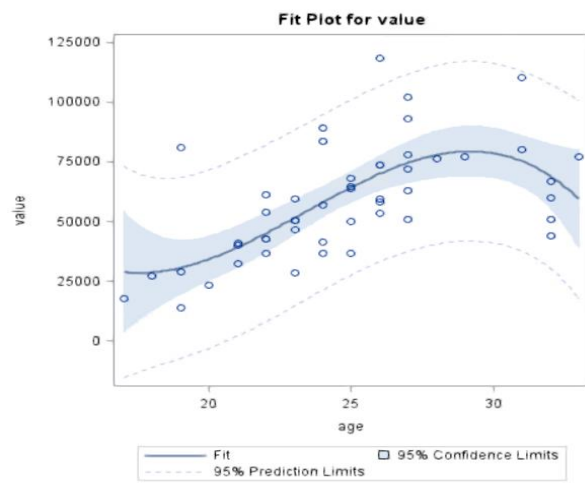
Class 1   **Class 2**
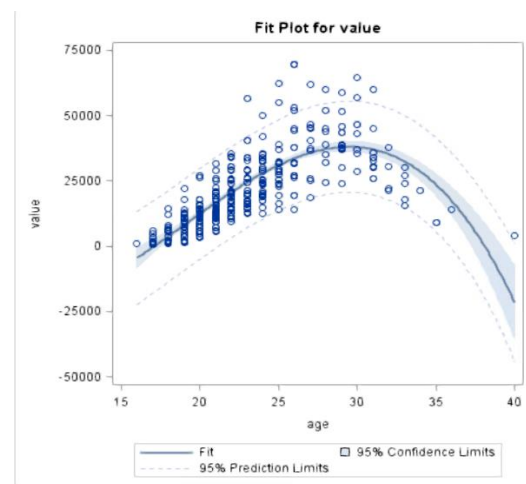
Class 3
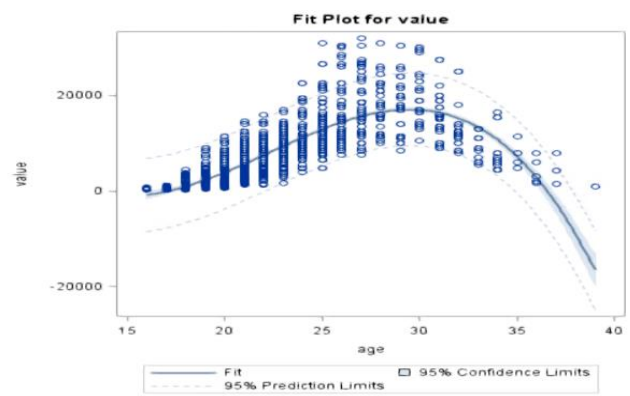
As we can see from the above fit plots, as the age increases the player gains more experience and his value increases and he reaches a peak point where his value is maximum. After that, the

value of the player decreases with increasing age (as the player grows older after his prime his quality of performance decreases) for class 1, class 2 and class 3 players.

The players regardless of whether they belong to class 1, class 2 or class 3 reach their peak market value around the age of 30.

2.4 ARE PLAYERS BEING DISCRIMINATED BASED ON NATIONALITY?

Some of the Countries have very few players compared to the other countries, for example Botswana has only a single player. So, to find out whether there is any discrimination based on nationality we took the top 10 countries with the highest player counts.

These countries are England, Germany, Spain, Argentina, France, Brazil, Italy, Colombia and Netherlands.

So, we ran a Regression using Value as the dependent variable and Nationality as the independent variable while also taking other variables into consideration such as International Reputation, weak foot, skill moves, Overall Rating, Class of the player, Attacking Work Rate and Defensive Work Rate. Reference is taken as England as it has the greatest number of players.

Below is the Output of the Regression model

| Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	39558.54901	B	1767.972681	22.38	<.0001
Nationality Argentina	-339.41769	B	117.440701	-2.89	0.0039
Nationality Brazil	-336.19677	B	126.300596	-2.66	0.0078
Nationality Colombia	-59.47765	B	132.440657	-0.45	0.6534
Nationality France	-413.39195	B	116.795113	-3.54	0.0004
Nationality Germany	-109.05775	B	106.731517	-1.02	0.3069
Nationality Italy	-585.53077	B	127.408031	-4.60	<.0001
Nationality Japan	-35.22983	B	144.495725	-0.24	0.8074
Nationality Spain	-44.65079	B	113.815874	-0.39	0.6948
Nationality England	0.00000	B	.	.	.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Overall	268.30870		6.411357	41.85	<.0001
International_Reputa 1	-49396.45983	B	1649.290760	-29.95	<.0001
International_Reputa 2	-46956.30230	B	1648.443724	-28.49	<.0001
International_Reputa 3	-40834.25204	B	1646.816889	-24.80	<.0001
International_Reputa 4	-34736.73194	B	1688.017447	-20.58	<.0001
International_Reputa 5	0.00000	B	.	.	.
Attacking_Work_Rate High	281.80791	B	75.856088	3.72	0.0002
Attacking_Work_Rate Low	-602.70722	B	144.783773	-4.16	<.0001
Attacking_Work_Rate Medium	0.00000	B	.	.	.
Defensive_Work_Rate High	138.34347	B	88.173537	1.57	0.1167
Defensive_Work_Rate Low	-220.11222	B	112.165520	-1.96	0.0498
Defensive_Work_Rate Medium	0.00000	B	.	.	.
Skill_Moves 1	-6625.64973	B	593.638613	-11.16	<.0001
Skill_Moves 2	-6343.30855	B	588.564519	-10.78	<.0001
Skill_Moves 3	-6587.62025	B	585.901170	-11.24	<.0001
Skill_Moves 4	-4016.76185	B	591.334437	-6.79	<.0001
Skill_Moves 5	0.00000	B	.	.	.
class 1	37042.35179	B	510.657974	72.54	<.0001
class 2	11600.03233	B	209.318015	55.42	<.0001
class 3	3102.21421	B	114.053831	27.20	<.0001
class 4	0.00000	B	.	.	.

The reference was taken as England as it has the most number players in the Dataset.

So, from the estimates for Nationality for the top 10 countries we have chosen we can say with 95% confidence that players from Brazil, France, Italy and Argentina have significantly lesser market values than England considering all the other factors remain the same.

2.5 ARE PLAYERS BEING DISCRIMINATED BASED ON POSITIONS?

Distribution of Players based on Position:

Position	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Attacking Midfielder	988	5.52	988	5.52
Centre Back	3038	16.97	4026	22.48
Centre Forward	105	0.59	4131	23.07
Centre Midfielder	2153	12.02	6284	35.09
Defensive Midfielder	1421	7.94	7705	43.03
Full Back	2572	14.36	10277	57.39
Goal Keeper	1989	11.11	12266	68.50
Striker	2537	14.17	14803	82.67
Wide Midfielder	2200	12.29	17003	94.95
Wing Back	165	0.92	17168	95.87
Winger	739	4.13	17907	100.00
Frequency Missing = 48				

Position Centre Back has the highest number of players. The number of players across each position is randomly distributed in this dataset.

Distribution of players based on International Reputation:

Regression Model with Market value on (Position | Overall Age Age^2 International_Reputation) after removal of outliers with Cookd's method. *848 outliers were found.*

Market_value = β_0 + β_1 (Position*Overall) + β_2 Age + β_3 Age^2 + β_4 International_Reputation

Below table shows us the model estimates and statistics with Position Centre Back and International_Reputation 1 used as reference bases.

Dependent	Parameter		Estimate	Biased	Standard Error	t Value	Pr > |t|
Value	Intercept		-15236.17824	1	394.358329	-38.64	<.0001
Value	Position	Attacking Midfielder	-5548.02543	1	601.706104	-9.22	<.0001
Value	Position	Centre Forward	-31633.37624	1	4185.427717	-7.56	<.0001
Value	Position	Centre Midfielder	-1192.17049	1	414.466376	-2.88	0.0040
Value	Position	Defensive Midfielder	-938.41756	1	531.556930	-1.77	0.0775
Value	Position	Full Back	2544.16877	1	424.031081	6.00	<.0001
Value	Position	Goal Keeper	6148.90671	1	414.266059	14.84	<.0001
Value	Position	Striker	-1427.07201	1	413.644915	-3.45	0.0006
Value	Position	Wide Midfielder	-4200.29396	1	447.687192	-9.38	<.0001
Value	Position	Wing Back	753.13319	1	1474.120906	0.51	0.6094
Value	Position	Winger	-10741.92939	1	779.234549	-13.79	<.0001
Value	Position	Centre Back	0.00000	1	.	.	.
Value	Overall		317.58535	1	4.536413	70.01	<.0001
Value	Overall*Position	Attacking Midfielder	85.89149	1	8.975693	9.57	<.0001
Value	Overall*Position	Centre Forward	483.50418	1	63.916432	7.56	<.0001
Value	Overall*Position	Centre Midfielder	24.41607	1	6.270658	3.89	<.0001
Value	Overall*Position	Defensive Midfielder	13.32010	1	7.922846	1.68	0.0927
Value	Overall*Position	Full Back	-40.05904	1	6.400657	-6.26	<.0001
Value	Overall*Position	Goal Keeper	-92.78880	1	6.327123	-14.67	<.0001
Value	Overall*Position	Striker	26.14967	1	6.238994	4.19	<.0001
Value	Overall*Position	Wide Midfielder	67.34409	1	6.729647	10.01	<.0001
Value	Overall*Position	Wing Back	-11.86611	1	22.564254	-0.53	0.5990
Value	Overall*Position	Winger	162.85583	1	11.668566	13.96	<.0001
Value	Overall*Position	Centre Back	0.00000	1	.	.	.
Value	Age		-154.74507	0	24.722025	-6.26	<.0001
Value	age_2		-0.61553	0	0.467023	-1.32	0.1875
Value	International_Reputa 2		2200.66081	1	49.500701	44.46	<.0001
Value	International_Reputa 3		15196.06773	1	210.347953	72.24	<.0001
Value	International_Reputa 4		31062.18922	1	821.450344	37.81	<.0001
Value	International_Reputa 1		0.00000	1	.	.	.

The p-values for positions Defensive Midfielder and Wing Back seem to be insignificant and hence their beta values are zero. This means that the Market value of players in this position is equivalent to Market Value of a Centre Back position player when the all the other attributes are held constant.

Due to the interaction existing between Overall rating and position, Market Value of a player with respect to positions can be compared at some constant levels of overall rating.

Distribution of Overall Rating
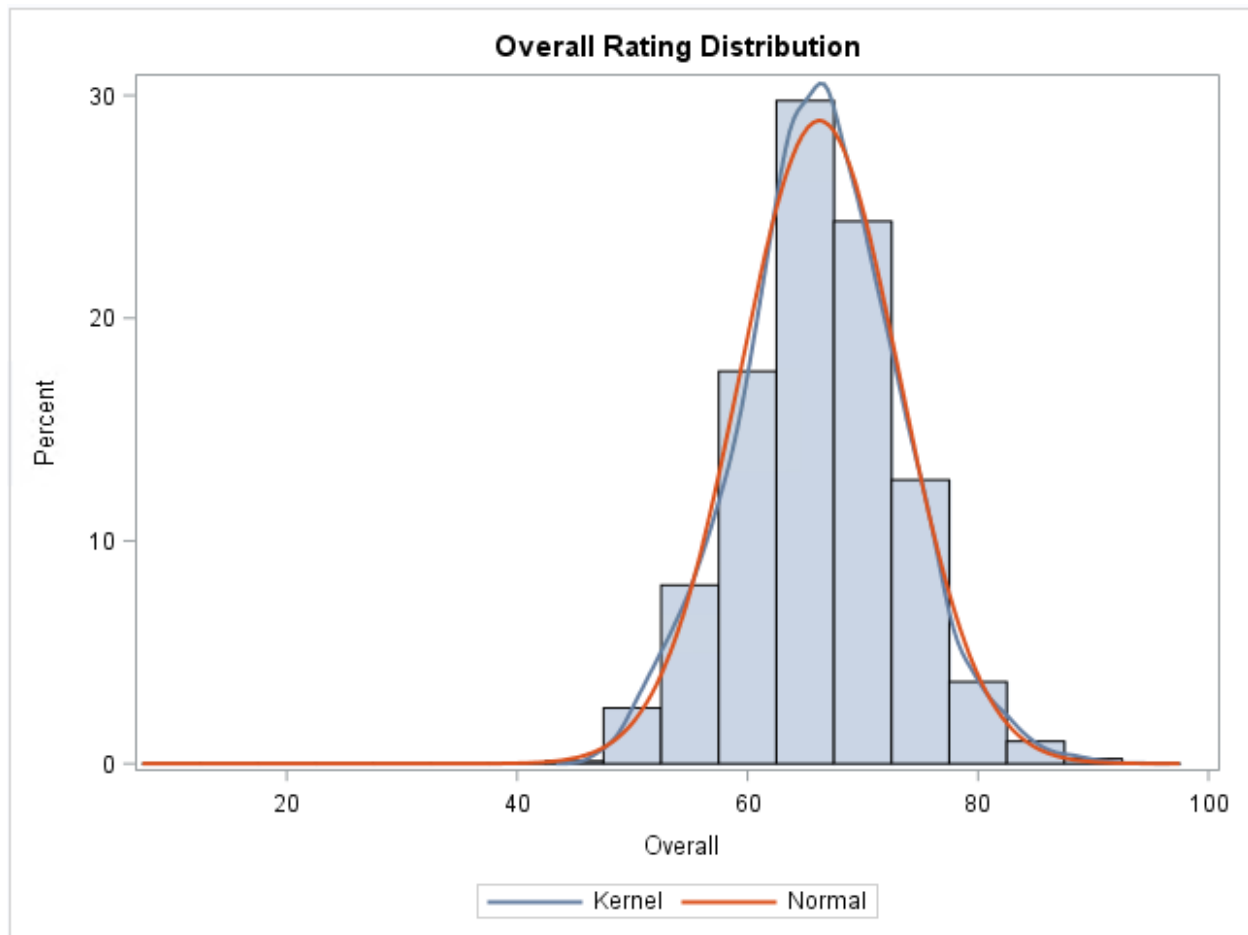
Below are the Market Values calculated for each position at constant overall rating with Centre Back as the reference base.

Position	Value_90	Value_85	Value_80	Value_75	Value_70	Value_65	Value_60	Value_55
Attacking Midfielder	2182	1753	1323	894	464	35	-395	-824
Centre Forward	11882	9464	7047	4629	2212	-206	-2623	-5041
Centre Midfielder	1005	883	761	639	517	395	273	151
Full Back	-1061	-861	-661	-460	-6	-60	-139	341
Goal Keeper	-2202	-1738	-1274	-810	-260	118	141	1046
Striker	926	796	665	534	-346	273	582	11
Wide Midfielder	1861	1524	1187	851	403	177	142	-496
Winger	3915	3101	2287	1472	514	-156	-160	-1785
Centre Back	0	0	0	0	0	0	0	0
Defensive Midfielder	0	0	0	0	0	0	0	0
Wing Back	0	0	0	0	0	0	0	0

Value_90 – Market Value at overall rating 90

Value_85 – Market Value at overall rating 85

Value_80 – Market Value at overall rating 80

Value_75 – Market Value at overall rating 75

Value_70 – Market Value at overall rating 70

Value_65 – Market Value at overall rating 65

Value_60 – Market Value at overall rating 60

Value_55 – Market Value at overall rating 55

Value based on position and overall
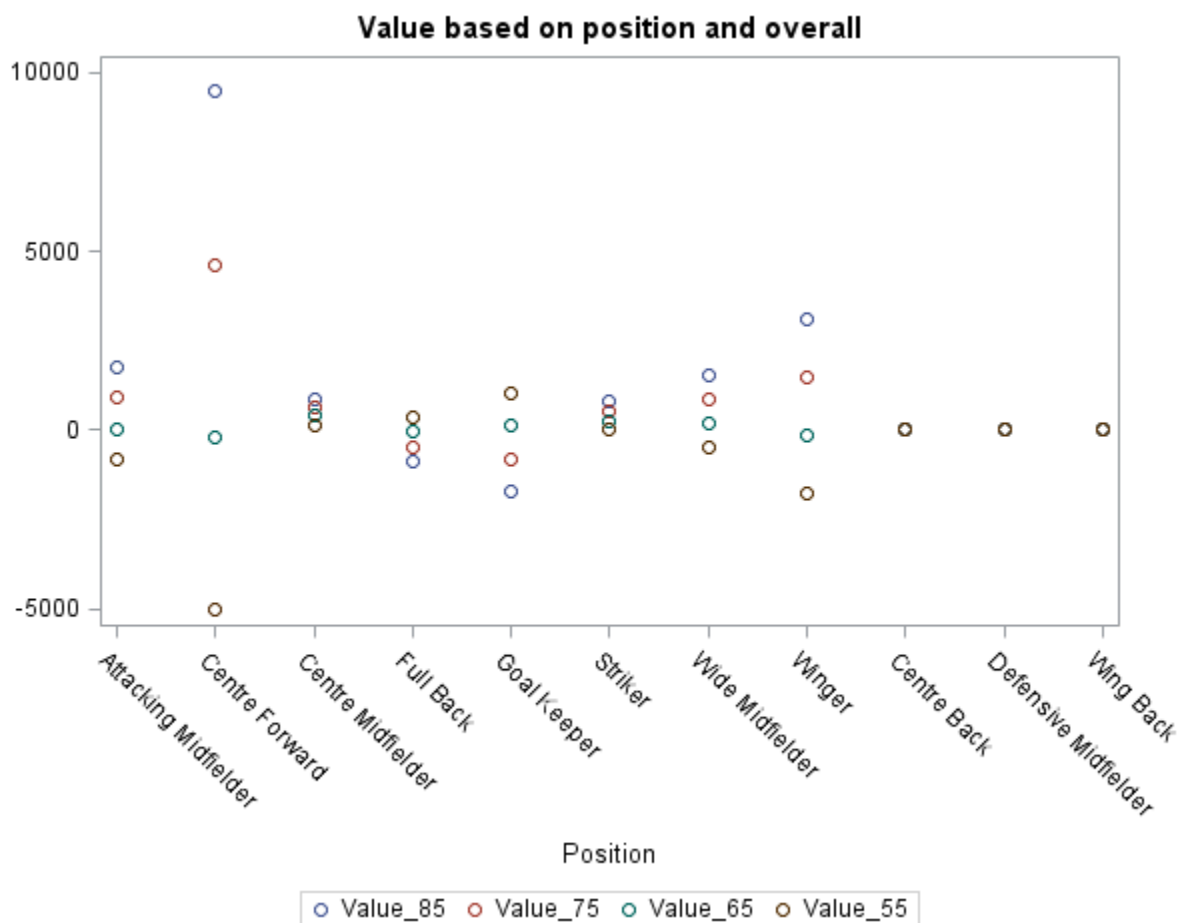
Centre Forward:

As per our dataset, Centre forward with overall 85 is highly paid, almost $10000k more than the player with Centre Back position when the other attributes are held constant. They are the least paid players when their overall rating is below 60

Winger:

Winger position is the second most paid after Centre Forward when the overall is 85 plus. Like for Centre Forward, their value decreases with an increase in overall rating.

Full Back and Goal Keeper:

The value of the players with these positions increases when their overall rating decreases, i.e., Players with overall rating of 55 are more paid than the players with an overall rating of 85.

Attacking Midfielder and Wide Midfielder:

Both these positions have almost the same market values at different overall rating levels. **Centre Midfielder and Striker** have same market values at different overall rating levels. Their value is only slightly higher than the value of the players occupying Centre Back, Defensive Midfielder and Wing Back positions.

From these results, we can conclude that there is an existence of discrimination in market value among the players occupying different positions.

2.6 EFFECT OF ATTACKING AND DEFENSIVE WORK RATES ON MARKET VALUE

1. How does The Market Value of the players change With Attacking Work Rate?

So, we have taken the same 10 countries (The countries with highest player counts) and ran a regression model using Value as the dependent variable, Attacking Work Rate as the Independent Variable and also taking other variables into consideration such as International Reputation, Overall rating, Class of the player, Defensive Work Rate, Weak Foot, Skill Moves. Using the means Statement, we conducted the Tukey's test to see if there is any difference in the Market Values of the players based on The Attacking Work Rate

Below is the Output from Tukey's Test

Comparisons significant at the 0.05 level are indicated by ***.				
Attacking_Work_Rate Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
High - Medium	2028.74	1930.01	2127.47	***
High - Low	2421.75	2211.63	2631.86	***
Medium - High	-2028.74	-2127.47	-1930.01	***
Medium - Low	393.01	193.18	592.83	***
Low - High	-2421.75	-2631.86	-2211.63	***
Low - Medium	-393.01	-592.83	-193.18	***

We can conclude with 95 % confidence that as the Players with higher Attacking Work Rate have higher market values when all the other factors remain the same. Players with higher Attacking work rate are usually on the attacking side of the field and usually these are the players who play a more important part in scoring goals. So, they have greater market values compared to the other players.

2. How does The Market Value of the players change With Defensive Work Rate?

We used the same regression model as we have used in the previous case, but this time we used the means statement with Defensive Work Rate and conducted the Tukey's test to see if there is any difference in market values based on defensive work rate. Below is the output of the Tukey's test for Defensive Work Rate

Comparisons significant at the 0.05 level are indicated by ***.			
Defensive_Work_Rate Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
High - Low	712.43	532.27	892.59	***
High - Medium	1397.37	1281.61	1513.13	***
Low - High	-712.43	-892.59	-532.27	***
Low - Medium	684.94	529.57	840.31	***
Medium - High	-1397.37	-1513.13	-1281.61	***
Medium - Low	-684.94	-840.31	-529.57	***

We can conclude that with 95 % confidence that players with high defensive work rate have higher market values than players with low and medium defensive work rates. However, players with medium defensive work rates have lesser market values compared to the players with low defensive work rate.

Players with low defensive work rate are generally have high attacking work rate and occupy positions like striker, winger etc. who play a more important role in scoring goals and hence have higher market values than players with medium defensive work rate who particularly occupy the midfield/defense positions.

What we can infer from the above results is that the specialists I.e. players with high attacking work rate or players with high defensive work are paid higher compared to the midfielders

2.7 POSITION BASED RATING

If the club is looking for a player to use him in a specific position, what are the important skills that the manager should be looking for that position? (and) How is Overall Rating calculated? What are the skills and the attributes involved in overall rating calculation?

Regression Model with overall on position and all the skill ratings and their interaction terms with Centre Back position taken as the reference base. The model gave us a R-squared of 97% who is a significant indicator of how overall is being calculated. Significant p-values were filtered and below are some of the insights taken from model estimates.

Parameter	Estimate
Intercept	8.211874214
Position Centre Forward	-3.809432369
Position Centre Midfielder	-6.536381420
Position Defensive Midfielder	-3.165076909
Position Full Back	2.701386153
Position Goal Keeper	-7.416414783
Position Striker	-0.910212175
Position Centre Back	0.000000000

Vision:

Vision*Position	Attacking Midfielder	0.103359	<.0001
Vision*Position	Centre Midfielder	0.088741	<.0001
Vision*Position	Defensive Midfielder	0.042336	<.0001
Vision*Position	Wide Midfielder	0.040986	<.0001
Vision*Position	Winger	0.044625	<.0001
Vision*Position	Centre Back	0	

Vision is significant for players occupying Midfielder positions.

Composure:

Composure		0.021761	<.0001
Composure*Position	Centre Midfielder	0.017286	0.0035
Composure*Position	Defensive Midfielder	0.019548	0.0044
Composure*Position	Goal Keeper	-0.02034	<.0001
Composure*Position	Winger	-0.02185	0.0268
Composure*Position	Centre Back	0	

Positioning:

Positioning*Position Attacking Midfielder	0.080688	<.0001
Positioning*Position Centre Forward	0.166406	<.0001
Positioning*Position Centre Midfielder	0.033234	<.0001
Positioning*Position Full Back	0.008598	0.0498
Positioning*Position Striker	0.138855	<.0001
Positioning*Position Wide Midfielder	0.08833	<.0001
Positioning*Position Winger	0.110459	<.0001
Positioning*Position Centre Back	0	

Positioning skill is highly weighed for the players occupying forward positions.

Penalties

Penalties*Position	Centre Forward	0.048259	0.013
Penalties*Position	Centre Back	0	

Penalties is significant only the Centre forward players. This maybe because the penalties are mostly taken by the center forward players.

Shot Power

ShotPower*Position	Full Back	0.01078	0.0062
ShotPower*Position	Striker	0.128393	<.0001
ShotPower*Position	Wide Midfielder	0.012944	0.014
ShotPower*Position	Centre Back	0	

Shot Power is highly weighed for the Striker position.

Finishing

Finishing*Position	Attacking Midfielder	0.064543	<.0001
Finishing*Position	Centre Forward	0.101515	0.007
Finishing*Position	Centre Midfielder	0.018293	0.0004
Finishing*Position	Full Back	0.010428	0.0252
Finishing*Position	Striker	0.168262	<.0001
Finishing*Position	Wide Midfielder	0.044056	<.0001
Finishing*Position	Wing Back	0.033512	0.0224
Finishing*Position	Winger	0.092491	<.0001
Finishing*Position	Centre Back	0	

Finishing is highly weighed for forward position players.

Curving

Curve*Position	Wing Back	0.035271	0.0071
Curve*Position	Centre Back	0	

Curving skill is significant only for the Wing Back position.

Volleys

Volleys*Position	Centre Midfielder	-0.00979	0.035
Volleys*Position	Full Back	-0.01008	0.0181
Volleys*Position	Wide Midfielder	-0.01463	0.0025
Volleys*Position	Centre Back	0	

Crossing

Crossing		-0.00819	0.0009
Crossing*Position	Attacking Midfielder	0.027156	0.0001
Crossing*Position	Centre Midfielder	0.016411	0.0009
Crossing*Position	Defensive Midfielder	0.012533	0.016
Crossing*Position	Full Back	0.094973	<.0001
Crossing*Position	Striker	0.02536	<.0001
Crossing*Position	Wide Midfielder	0.106618	<.0001
Crossing*Position	Wing Back	0.122282	<.0001
Crossing*Position	Winger	0.064054	<.0001
Crossing*Position	Centre Back	0	

Crossing is highly weighed for players occupying wide field positions.

ShortPassing

ShortPassin*Position Attacking Midfielder	0.105892	<.0001
ShortPassin*Position Centre Midfielder	0.14643	<.0001
ShortPassin*Position Defensive Midfielder	0.114389	<.0001
ShortPassin*Position Full Back	0.039404	<.0001
ShortPassin*Position Goal Keeper	-0.04545	<.0001
ShortPassin*Position Wide Midfielder	0.076528	<.0001
ShortPassin*Position Wing Back	0.145795	<.0001
ShortPassin*Position Winger	0.060526	<.0001
ShortPassin*Position Centre Back	0	

ShortPassing is highly weighed for Midfielder and Wing Back players.

LongPassing

LongPassing*Position Centre Midfielder	0.072661	<.0001
LongPassing*Position Defensive Midfielder	0.092145	<.0001
LongPassing*Position Wing Back	-0.06743	0.0046
LongPassing*Position Centre Back	0	

Dribbling

Dribbling		0.007787	0.0102
Dribbling*Position	Attacking Midfielder	0.173132	<.0001
Dribbling*Position	Centre Forward	0.186469	<.0001
Dribbling*Position	Centre Midfielder	0.055155	<.0001
Dribbling*Position	Defensive Midfielder	0.016894	0.0211
Dribbling*Position	Full Back	0.017256	0.0022
Dribbling*Position	Striker	0.125436	<.0001
Dribbling*Position	Wide Midfielder	0.195822	<.0001
Dribbling*Position	Winger	0.20056	<.0001
Dribbling*Position	Centre Back	0	

Dribbling seems be highly weighed for both Forward and Midfielder players.

BallControl

BallControl*Position Attacking Midfielder	0.071172	<.0001
BallControl*Position Centre Midfielder	0.14079	<.0001
BallControl*Position Defensive Midfielder	0.091681	<.0001
BallControl*Position Full Back	0.042601	<.0001
BallControl*Position Goal Keeper	-0.0423	<.0001
BallControl*Position Striker	0.061563	<.0001
BallControl*Position Wide Midfielder	0.096735	<.0001
BallControl*Position Winger	0.105338	<.0001
BallControl*Position Centre Back	0	

GoalKeeper Skills

GKDiving*Position	Defensive Midfielder	0.025549	0.0436
GKDiving*Position	Goal Keeper	0.217343	<.0001
GKDiving*Position	Wing Back	0.072273	0.0335
GKDiving*Position	Centre Back	0	
GKHandling*Position	Centre Forward	0.091984	0.042
GKHandling*Position	Goal Keeper	0.214178	<.0001
GKHandling*Position	Centre Back	0	
GKKicking*Position	Goal Keeper	0.050404	<.0001
GKKicking*Position	Winger	0.032057	0.0432
GKKicking*Position	Centre Back	0	
GKPositioni*Position	Goal Keeper	0.217328	<.0001
GKPositioni*Position	Winger	0.054113	0.0013
GKPositioni*Position	Centre Back	0	
GKReflexes*Position	Centre Forward	-0.0914	0.0356
GKReflexes*Position	Goal Keeper	0.208486	<.0001
GKReflexes*Position	Centre Back	0	

GoalKeeper skills are highly weighed to GoalKeeper only. Skills like Marking, StandingTackle, SlidingTackle, Interceptions, Aggression, LongShots, HeadingAccuracy are significant to all the positions.

Below table shows us the important skills that the manager should be looking for in a player for each position

Position	Positioning	Vision	Composure	Penalties	Shot Power	Curving	Volleys	Crossing	ShortPassing	LongPassing	Dribbling	BallControl	Finishing
Centre Forward	✓			✓							✓		✓
Striker	✓				✓						✓		✓
Winger	✓											✓	✓
Centre Midfielder		✓	✓				✓		✓	✓		✓	
Attacking Midfielder		✓							✓		✓		
Defensive Midfielder			✓						✓	✓			
Wide Midfielder							✓	✓			✓		
Full Back							✓	✓					
Wing Back						✓		✓	✓				

Based on the model estimates which seem to be obvious to the real world and the highly significant R-sqaured (97%), we can conclude that the overall variable has position and all the other rating attributes involved in it.

3. MARKET VALUE PREDICTION (MVP) MODEL:

Market value was predicted with some of the top predictors and the base model below had an adjusted R^2 of 0.84. We will use forward, backward and stepwise in further steps to determine the best model which also has interaction terms with the best AIC value.

MVP model with top predictors

The GLM Procedure

Dependent Variable: Value

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	24	449607183819	18733632659	2852.47	<.0001
Error	13010	85443251236	6567505.8598		
Corrected Total	13034	535050435055			

R-Square	Coeff Var	Root MSE	Value Mean
0.840308	78.10763	2562.715	3281.004

3.1 CHRONOLOGY OF VARIABLES

1. We created a new categorical variable called classPotential which takes the values:
   a. 1 if Potential>90
   b. 2 if 85<=Potential<=89
   c. 3 if 80<=Potential<=84
   d. 4 if Potential<80.
2. International Reputation 5 is the highest class of Players like Messi and Ronaldo. International Reputation 1 are the players from less famous countries.
3. Attacking Work Rate and Defensive work rate are variables which take the variables High, Medium and Low depending on how well the player attacks or defends.
4. We split the whole data into 11 different positions for example combing ST, LST and CF into one position called Striker.

3.2 PARAMETER ESTIMATES

| Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	-35441.66207	B	456.852205	-77.58	<.0001
Age	-132.23479		6.232452	-21.22	<.0001
Position Attacking Midfielder	568.38986	B	113.131923	5.02	<.0001
Position Centre Back	4.32927	B	85.056899	0.05	0.9594
Position Centre Forward	1819.63611	B	299.846111	6.07	<.0001
Position Centre Midfielder	552.31299	B	90.660103	6.09	<.0001
Position Defensive Midfielder	30.03566	B	99.671657	0.30	0.7632
Position Goal Keeper	-213.58843	B	99.801772	-2.14	0.0324
Position Striker	585.96840	B	87.462058	6.70	<.0001
Position Wide Midfielder	597.19438	B	88.685507	6.73	<.0001
Position Wing Back	332.15419	B	252.582160	1.32	0.1885
Position Winger	587.75874	B	125.189953	4.69	<.0001
Position Full Back	0.00000	B	.	.	.
Overall	583.53404		6.557370	88.99	<.0001
International_Reputa 2	1163.35725	B	88.829529	13.10	<.0001
International_Reputa 3	6864.86578	B	170.868251	40.18	<.0001
International_Reputa 4	14200.87263	B	389.282261	36.48	<.0001
International_Reputa 5	30950.19678	B	1087.542503	28.46	<.0001
International_Reputa 1	0.00000	B	.	.	.
Weak_Foot	32.37596		35.825660	0.90	0.3662
classPotential 1	36516.77701	B	404.751463	90.22	<.0001
classPotential 2	9562.38942	B	161.004431	59.39	<.0001
classPotential 3	1888.08856	B	89.250292	21.15	<.0001
classPotential 4	0.00000	B	.	.	.
Attacking_Work_Rate High	135.43417	B	116.511825	1.16	0.2451
Attacking_Work_Rate Medium	-1.65698	B	107.255372	-0.02	0.9877
Attacking_Work_Rate Low	0.00000	B	.	.	.
Defensive_Work_Rate High	459.11507	B	97.060564	4.73	<.0001
Defensive_Work_Rate Medium	390.19704	B	83.658880	4.66	<.0001
Defensive_Work_Rate Low	0.00000	B	.	.	.

Interpretation of Parameter Estimates:

- If age increases by 1 year, the Market Value reduces by -$132.23
- If the Overall of a player increases by 1, the Market Value increases by $563.53
- Market Value of a Player with International Reputation 5 is the highest. The players with International Reputation=5 have a Market value $30950.19 greater than the players with International Reputation 1.
- The highest classpotential=1 players whose Potential>1 have market Value $35516.77 more than least ClassPotential(4) players whose potential is less than 80.
- Also, we see that Centre forward has higher market value about $1819.64 more than Full back player.

3.3 BEST MARKET VALUE PREDICTION (MVP) MODEL USING STEPWISE SELECTION AIC AS CRITERIA

Now let's use stepwise selection to find out the best model using adjusted R^2 as criteria. We will use the variables obtained above as the initial set of predictors

Best MVP model using Stepwise Selection with AdjR^2 as criteria

The GLMSELECT Procedure

Stepwise Selection Summary

Step	Effect Entered	Effect Removed	Number Effects In	Number Parms In	AIC	ASE	Test ASE
0	Intercept		1	1	192751.650	40236698.0	44266438.3
1	classPotential		2	4	183216.311	16094672.9	14863763.5
2	Overall		3	5	176478.716	8425471.4	8044186.7
3	Overall*classPotenti		4	8	169067.687	4132877.8	4065410.7
4	Skill_Moves		5	12	168163.949	3786402.0	3764478.4
5	Overall*Skill_Moves		6	16	166367.045	3183828.8	3498697.3
6	Age		7	17	165394.997	2899513.4	3243573.9
7	Age*classPotential		8	20	164403.766	2634715.0	3104912.8
8	International_Reputa		9	24	163856.988	2498017.9	3008061.4
9	Overall*Internationa		10	28	161512.115	1992803.1	2681183.5
10	Age*International_Re		11	32	160694.665	1840930.4	2348270.9
11	Internati*classPoten		12	41	159954.298	1711623.4	2181994.1
12	Position		13	50	159631.557	1656521.0	2118020.3
13	Skill_Mov*Internatio		14	63	159323.710	1604254.3	2350511.3
14	Skill_Mov*classPoten		15	75	159047.741	1558699.5	2761612.3
15	Age*Skill_Moves		16	79	158942.138	1541787.1	2794237.8
16	Age*Overall		17	80	158915.523	1537555.9	2790936.7
17	Defensive_Work_Rate		18	82	158910.383	1536207.1	2800340.9
18	Defensive*Internatio		19	88	158729.828	1508060.9	3162620.0
19	Skill_Mov*Defensive_		20	94	158669.655	1497644.5	3135023.6
20	Overall*Defensive_Wo		21	96	158629.043	1491242.0	3113956.4
21	Defensive*classPoten		22	102	158587.045	1483529.0	3166815.6
22	Age*Defensive_Work_R		23	104	158572.056	1480826.1	3163204.6
23	Weak_Foot		24	105	158569.614	1480194.4	3169835.4
24	Weak_Foot*Internatio		25	108	158502.636	1469857.0	3212401.6
25	Overall*Weak_Foot		26	109	158475.796	1465791.7	3222396.6
26	Weak_Foot*Skill_Move		27	113	158465.976	1463285.5	3316307.2
27	Preferred_Foot		28	114	158467.907	1463275.7	3315263.3
28		Preferred_Foot	27	113	158465.976*	1463285.5	3316307.2

*Optimal Value of Criterion

Stepwise selection stopped because the sequence of effect additions and removals is cycling.

Stepwise selection stopped because the sequence of effect additions and removals is cycling.

Coefficient Progression for Value
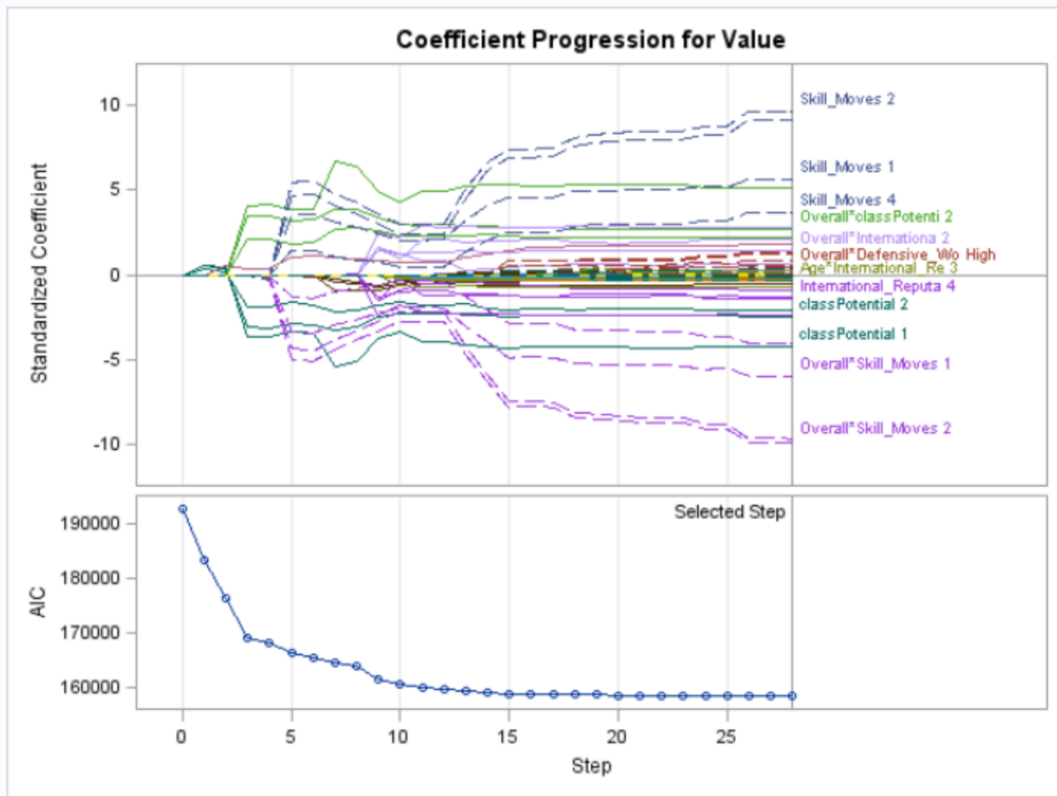
Best MVP model using Stepwise Selection with AIC as criteria

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 28).

Effects: Intercept Position Age classPotential Age*classPotential International_Reputa Age*International_Re Internati*classPoten Overall Age*Overall Overall*classPotenti Overall*Internationa Skill_Moves Age*Skill_Moves Skill_Mov*classPoten Skill_Mov*Internatio Overall*Skill_Moves Weak_Foot Weak_Foot*Internatio Overall*Weak_Foot Weak_Foot*Skill_Move Defensive_Work_Rate Age*Defensive_Work_R Defensive*classPoten Defensive*Internatio Overall*Defensive_Wo Skill_Mov*Defensive_

Note: The p-values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally liberal.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	112	4.037475E11	3604888786	2436.82	<.0001
Error	10300	15237191743	1479339		
Corrected Total	10412	4.189847E11			

|	
Root MSE	1216.28081
Dependent Mean	3289.32248
R-Square	0.9636
Adj R-Sq	0.9632
AIC	158466
AICC	158469
SBC	148870
ASE (Train)	1463285
ASE (Test)	3316307

3.4 BEST MARKET VALUE PREDICTION (MVP) MODEL USING FORWARD SELECTION WITH AIC AS CRITERIA

		Forward Selection Summary				
Step	Effect Entered	Number Effects In	Number Parms In	AIC	A SE	Test A SE
0	Intercept	1	1	192751.650	40236698.0	44266438.3
1	classPotential	2	4	183216.311	16094672.9	14863763.5
2	Overall	3	5	176478.716	8425471.4	8044186.7
3	Overall*classPotenti	4	8	169067.687	4132877.8	4065410.7
4	Skill_Moves	5	12	168163.949	3786402.0	3764478.4
5	Overall*Skill_Moves	6	16	166367.045	3183828.8	3498697.3
6	Age	7	17	165394.997	2899513.4	3243573.9
7	Age*classPotential	8	20	164403.766	2634715.0	3104912.8
8	International_Reputa	9	24	163856.988	2498017.9	3008061.4
9	Overall*Internationa	10	28	161512.115	1992803.1	2681183.5
10	Age*International_Re	11	32	160694.665	1840930.4	2348270.9
11	Internati*classPoten	12	41	159954.298	1711623.4	2181994.1
12	Position	13	50	159631.557	1656521.0	2118020.3
13	Skill_Mov*Internatio	14	63	159323.710	1604254.3	2350511.3
14	Skill_Mov*classPoten	15	75	159047.741	1558699.5	2761612.3
15	Age*Skill_Moves	16	79	158942.138	1541787.1	2794237.8
16	Age*Overall	17	80	158915.523	1537555.9	2790936.7
17	Defensive_Work_Rate	18	82	158910.383	1536207.1	2800340.9
18	Defensive*Internatio	19	88	158729.828	1508060.9	3162620.0
19	Skill_Mov*Defensive_	20	94	158669.655	1497644.5	3135023.6
20	Overall*Defensive_Wo	21	96	158629.043	1491242.0	3113956.4
21	Defensive*classPoten	22	102	158587.045	1483529.0	3166815.6
22	Age*Defensive_Work_R	23	104	158572.056	1480826.1	3163204.6
23	Weak_Foot	24	105	158569.614	1480194.4	3169835.4
24	Weak_Foot*Internatio	25	108	158502.636	1469857.0	3212401.6
25	Overall*Weak_Foot	26	109	158475.796	1465791.7	3222396.6
26	Weak_Foot*Skill_Move	27	113	158465.976	1463285.5	3316307.2
27	Preferred_Foot	28	114	158467.907	1463275.7	3315263.3
28	Preferred*Internatio	29	117	158447.238	1459532.9	3374643.4
29	Preferred*classPoten	30	120	158440.110	1457694.0	3313072.5
30	Weak_Foot*Preferred_	31	121	158437.005	1456979.5	3321914.4
31	Overall*Preferred_Fo	32	122	158435.997	1456558.7	3335784.0
32	Weak_Foot*classPoten	33	125	158435.897*	1455705.6	3309127.7
		*Optimal Value of Criterion				

Selection stopped at a local minimum of the AIC criterion.

Stop Details			
Candidate For	Effect	Candidate AIC	Compare AIC
Entry	Age*Preferred_Foot	158437.667 >	158435.897

Coefficient Progression for Value
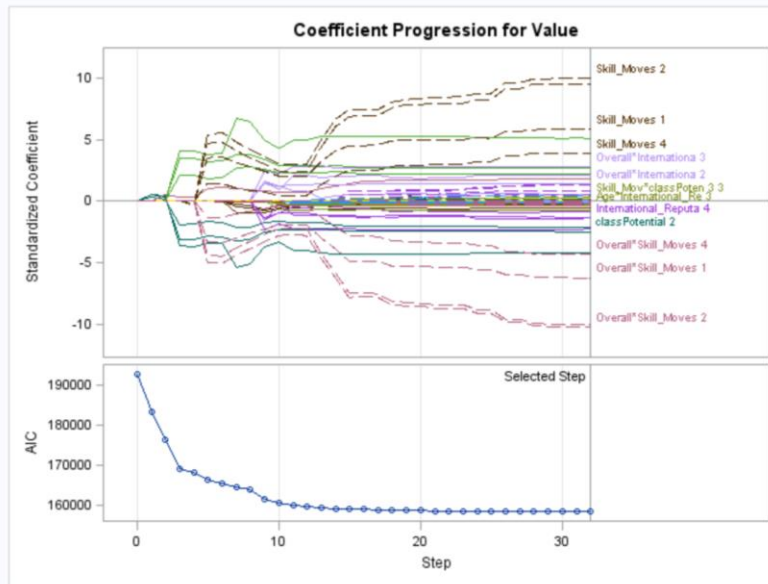
Best MVP model using Forward Selection with AIC as criteria

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 32).

Effects: Intercept Position Age classPotential Age*classPotential International_Reputa Age*International_Re Internati*classPoten Overall Age*Overall Overall*classPotenti Overall*Internationa Preferred_Foot Preferred*classPoten Preferred*Internatio Overall*Preferred_Fo Skill_Moves Age*Skill_Moves Skill_Mov*classPoten Skill_Mov*Internatio Overall*Skill_Moves Weak_Foot Weak_Foot*classPoten Weak_Foot*Internatio Overall*Weak_Foot Weak_Foot*Preferred_ Weak_Foot*Skill_Move Defensive_Work_Rate Age*Defensive_Work_R Defensive*classPoten Defensive*Internatio Overall*Defensive_Wo Skill_Mov*Defensive_

Note: The p-values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally liberal.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	124	4.038265E11	3256665106	2210.32	<.0001
Error	10288	15158262627	1473393		
Corrected Total	10412	4.189847E11			

Root MSE	1213.83383
Dependent Mean	3289.32248
R-Square	0.9638
Adj R-Sq	0.9634
AIC	158436
AICC	158439
SBC	148927
ASE (Train)	1455706
ASE (Test)	3309128

3.5 BEST MARKET VALUE PREDICTION (MVP) MODEL USING BACKWARD SELECTION WITH AIC AS CRITERIA

Backward selection was done and it converged at a lot less 10 features compared to other selection criteria in the model using AIC as optimal criteria.

Backward Selection Summary						
Step	Effect Removed	Number Effects In	Number Parms In	AIC	ASE	Test ASE
0		47	165	158314.087	1427764.85	4702350.43
1	Skill_Mov*Preferred_	46	161	158306.817	1427864.96	4711857.57
2	Attacking*Defensive_	45	157	158300.378	1428079.08	4706868.13
3	Weak_Foot*Defensive_	44	155	158296.415	1428084.11	4706651.36
4	Weak_Foot*Attacking_	43	153	158292.648	1428116.05	4706344.89
5	Preferred*Attacking_	42	151	158289.997	1428301.16	4707247.94
6	Preferred*Defensive_	41	149	158287.240	1428471.58	4702343.61
7	Age*Attacking_Work_R	40	147	158284.884	1428697.09	4705665.10
8	Weak_Foot*classPoten	39	144	158282.774	1429230.92	4681220.65
9	Age*Weak_Foot	38	143	158280.779	1429231.72	4681256.38
10	Age*Preferred_Foot	37	142	158279.073*	1429272.00	4686416.84
* Optimal Value of Criterion						

Selection stopped at a local minimum of the AIC criterion.

Stop Details				
Candidate For	Effect	Candidate AIC		Compare AIC
Removal	Weak_Foot*Preferred_	158281.028	>	158279.073

Stop Details

Candidate For	Effect	Candidate AIC		Compare AIC
Removal	Weak_Foot*Preferred_	158281.028	>	158279.073

Coefficient Progression for Value
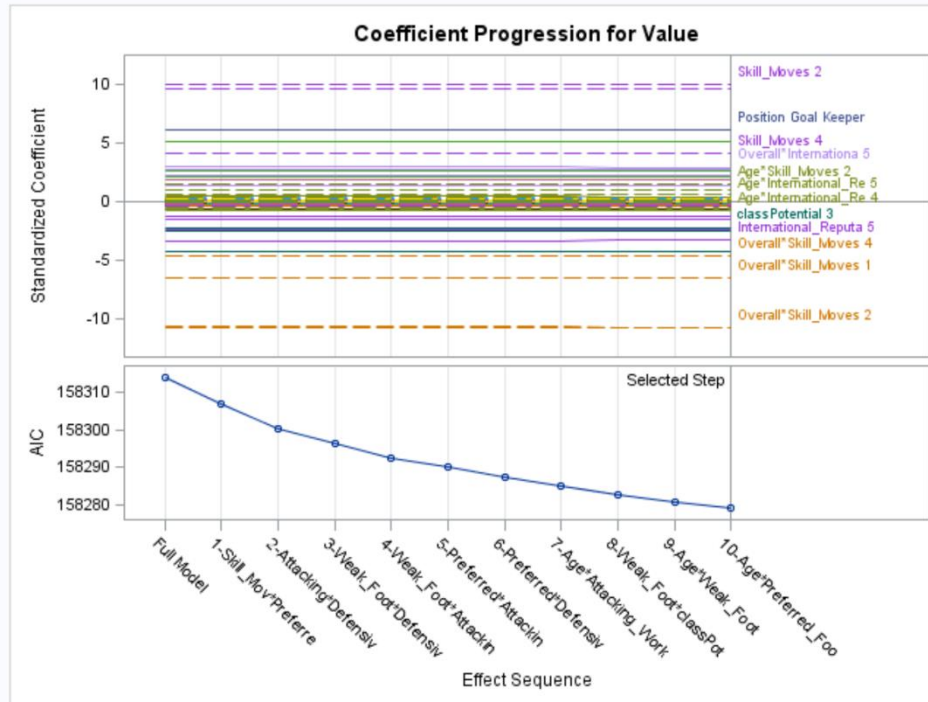
Best MVP model using Backward Selection with AIC as criteria

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 10).

Effects: Intercept Position Age classPotential Age*classPotential International_Reputa Age*International_Re Internati*classPoten Overall Age*Overall Overall*classPotenti Overall*Internationa Preferred_Foot Preferred*classPoten Preferred*Internatio Overall*Preferred_Fo Skill_Moves Age*Skill_Moves Skill_Mov*classPoten Skill_Mov*Internatio Overall*Skill_Moves Weak_Foot Weak_Foot*Internatio Overall*Weak_Foot Weak_Foot*Preferred_ Weak_Foot*Skill_Move Attacking_Work_Rate Attacking*classPoten Attacking*Internatio Overall*Attacking_Wo Skill_Mov*Attacking_ Defensive_Work_Rate Age*Defensive_Work_R Defensive*classPoten Defensive*Internatio Overall*Defensive_Wo Skill_Mov*Defensive_

Note: The p-values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally liberal.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	141	4.041017E11	2865969691	1977.85	<.0001
Error	10271	14883009367	1449032		
Corrected Total	10412	4.189847E11			

Root MSE	1203.75752
Dependent Mean	3289.32248
R-Square	0.9645
Adj R-Sq	0.9640
AIC	158279
AICC	158283
SBC	148894
ASE (Train)	1429272
ASE (Test)	4686417

3.6 BEST MVP MODEL COMPARISON USING AIC AS CRITERIA

Criteria	AIC	Adj-Rsq
Stepwise	158466	0.9632
Forward	158436	0.9634
Backward	158279	0.9640

Backward Selection Model came out to be the best model with an AIC as low as 158279 and an adjusted R-square of 0.9640

The best model had the following terms or effects

Effects: Intercept, Position Age classPotential Age*classPotential International_Reputa

Age*International_Re Internati*classPoten Overall Age*Overall

Overall*classPotenti Overall*Internationa Preferred_Foot

Preferred*classPoten Preferred*Internatio Overall*Preferred_Fo Skill_Moves

Age*Skill_Moves Skill_Mov*classPoten Skill_Mov*Internatio

Overall*Skill_Moves Weak_Foot Weak_Foot*Internatio Overall*Weak_Foot

Weak_Foot*Preferred_ Weak_Foot*Skill_Move Attacking_Work_Rate

Attacking*classPoten Attacking*Internatio Overall*Attacking_Wo

Skill_Mov*Attacking_ Defensive_Work_Rate Age*Defensive_Work_R

Defensive*classPoten Defensive*Internatio Overall*Defensive_Wo

Skill_Mov*Defensive_

3.7 BEST YOUNGEST PLAYER

- We created a new dataset of age below 22 years as fifabelow22.Using the best model obtained above, we predicted the Market Value of players and found out the best young player who had the highest market value.
- The best youngest player came out to be **Kylian Mbappe** have the following characteristics

Attributes	Player
Name	K. Mbappé
Age	19
Nationality	France
Overall	88
Potential	95
classPotential	1
Club	Paris Saint-Germain
Value	81000
p_Value	88264.25
Wage	€100K
Special	2118
International_Reputation	3

- Note that he had classPotential=1, one of the few young players to achieve such a high potential.

3.8 TOP 10 TEAMS AND THEIR BIAS IN MARKET VALUES

The top 10 teams from the UEFA website came out to be as below :

1. FC Bayern München
2. Borussia Dortmund
3. Liverpool
4. FC Barcelona
5. Juventus
6. Paris Saint-Germain
7. Manchester United
8. Manchester City
9. Chelsea
10. Real Madrid

Using the best model in backward Selection, we predicted the Overall Market value of the top 10 teams.

Club	Actual totalclub Value (in K$)	Predicted totalclub Value (in K$)	Bias Between Actual & Predicted Values (in K$)	Overpriced/ Underpriced
Borussia Dortmund	226945	247550.182	-20605.18196	UnderPriced
Manchester United	453050	450329.1688	2720.831245	OverPriced
Chelsea	492510	470160.1411	22349.85889	OverPriced
Liverpool	514580	490011.7077	24568.29231	OverPriced
Juventus	522475	524059.9515	-1584.951474	UnderPriced
Paris Saint-Germain	524525	505435.0605	19089.93947	OverPriced
FC Bayern München	553640	539988.4905	13651.50949	OverPriced
Manchester City	558555	549198.241	9356.758985	OverPriced
FC Barcelona	698500	692845.1586	5654.841351	OverPriced
Real Madrid	727600	727268.3072	331.6928309	OverPriced

Inference:

- We see that most of the top 10 club players highly OverPriced their players and the highest OverPrice is for Liverpool with a whooping overprice of $24568.29 followed by Chelsea ($22349.86) and Paris Saint-Germain ($19089.94) in 2nd and 3rd place for overpricing.
- Borussia Dortmund and Juventus played lesser to their players than their Predicted Values.

3.9 ANALYSIS FOR TOP 10 BALLON D'OR PLAYERS

Name	Actual_Value	Predicted_Value	Actual-Predicted Value Bias	Over/ Under/ Exactly Priced
L. Messi	110500	110500	1.00E-09	Slightly Overpriced
K. De Bruyne	102000	99052.09577	2947.904228	Over Priced
E. Hazard	93000	92673.73842	326.2615839	Over Priced
H. Kane	83500	79342.18095	4157.819045	Over Priced
K. Mbappé	81000	88264.25139	-7264.251391	Under Priced
A. Griezmann	78000	76292.29383	1707.706173	Over Priced
Cristiano Ronaldo	77000	77000	-1.90E-07	Slightly Underpriced
M. Salah	69500	53169.09745	16330.90255	Over Priced
L. Modric	67000	72159.38732	-5159.38732	Under Priced
R. Varane	50000	44786.52618	5213.473822	Over Priced

Inference:

- Leo Messi and Cristiano Ronaldo were slightly over and underpriced according to their Market Value

- Emerging players like Kylian Mbappe(-$7264.25139) and Luka Modric(-$5159.38732) were underpriced.

- Other Top 10 players were overpriced. Mohammed Salah, the Egyptian player was highly Overpriced ($16330.90) followed by Raphael Varane($5213.47) and Harry Kane($4157.82).

4. MANAGERIAL APPLICATIONS & STRATEGIES

In the real world, buying and selling of football players takes place during the transfer window, based on the bids a player receives from a club. In our analysis, we have shown that a player's market value is dependent on many factors like international reputation, age, nationality, positions, attacking and defense work rates.

Using our analysis, a manager can decide on how much he can bid on a player, based on his current attributes and future potential. Also, when the manager receives a bid for his player, he can decide whether the bid is underpriced or overpriced and sell the player at profits for the club.

Using our analysis

- we have shown the bias between the market values of top 10 clubs (ranked by UEFA)
- we have shown the bias between the top 10 ballon d'or players (ranked based on 2018 ballon d'or awards)
- we have found the youngest player (age under 22) with the highest market value.

On top of this, we can also recommend the following through our analysis

- What are the skills, a manager needs to look for, while buying a player for a positon
- We have identified the trend of market vakue with age, and have understood that player value peaks at the age of 30.

5. REFERENCES

1) For Creating new variable called ClassPotential :

   https://fifacareermodetips.com/guides/understanding-potential/

2) Top 10 Ballondor Players 2018:

   https://www.goal.com/en-us/lists/luka-modric-and-the-2018-ballon-dor-final-rankings/g4abmah6vk6u1qvmfansvvf5y#g74clktf0a36104y6e8k4r7ob