Predicting Player's Value : FIFA19

Made by - Akash Kumar

Importance of Value

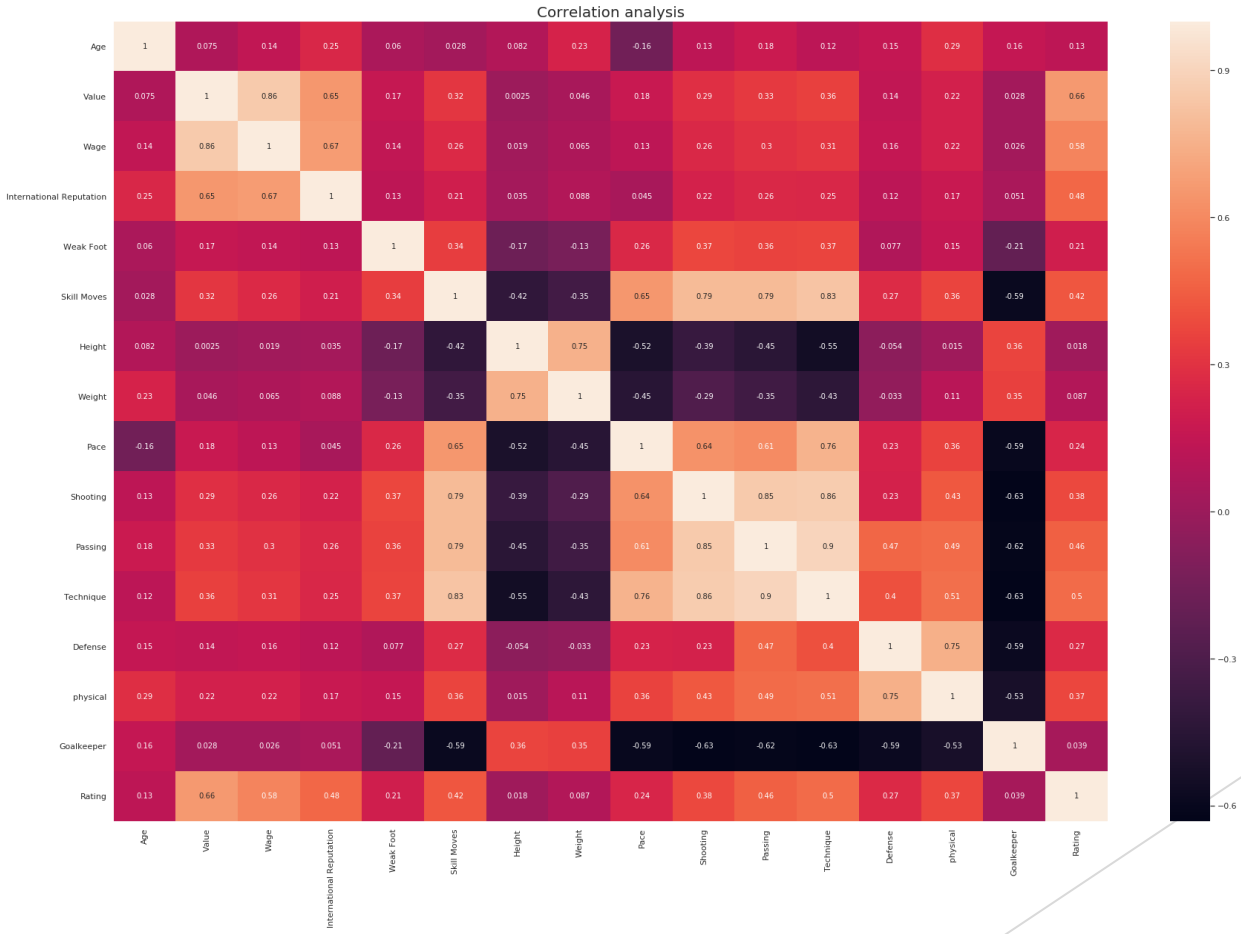
- Value determine the position of player's in the market.
- A player's value is everything as it gives us a complete background picture of the player's performance as well as his popularity among his fans.
- Any data that might contain the player's value could be used to predict the player's reputation in the market.
- This project aims to bring out that reputation through value prediction.
- Any club might want to know his player's as well as other club's players market value for competitive advantage and could also bring new emerging talent into the market.
- Others might also be interested are fans and other business people who are in sports industry

Data Acquisition and Cleaning

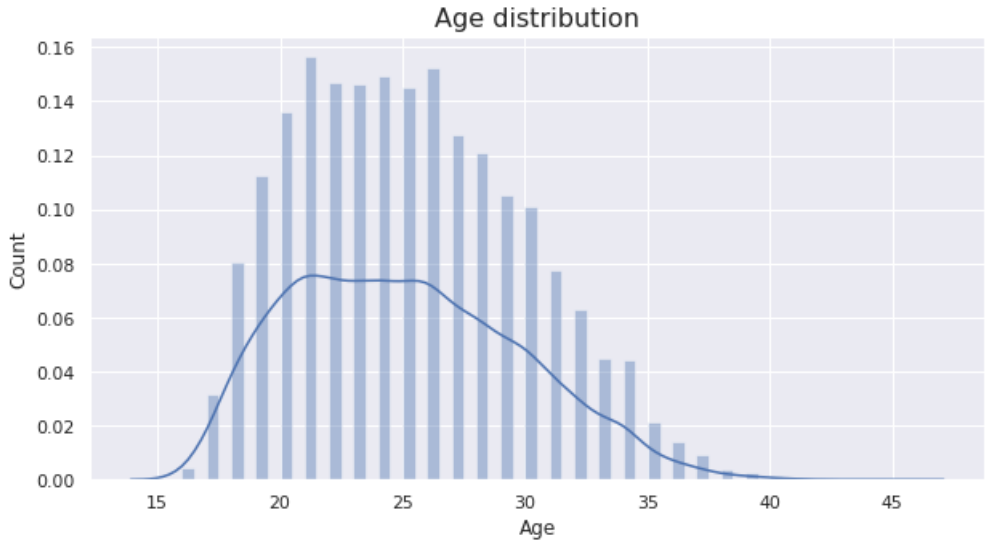
- All the player's data containing his name, photo, player's position etc. can be found in the csv file of Kaggle's FIFA19 dataset.
- Raw data contains 18207 rows and 89 columns and missing data with redundant attributes.
- Dropped unnecessary columns and filled missing values
- Aggregated similar columns leading to dimensionality reduction.
- Final data after cleaning contains 18207 rows and 23 columns

Correlation Analysis

- Heat map showing the correlation between the numerical features.
- Strikers have higher value than other players.

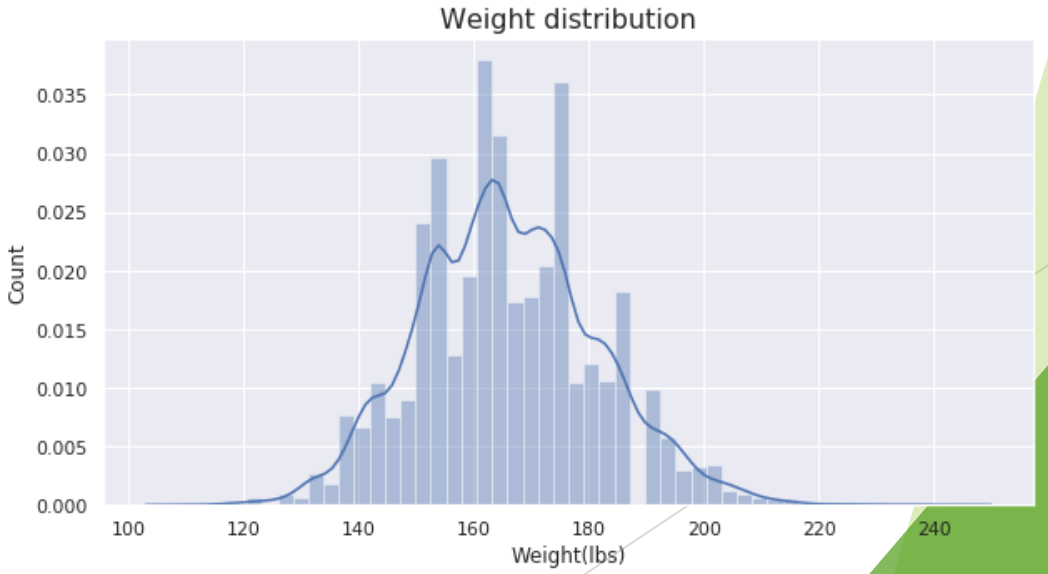


Weight and Age distribution

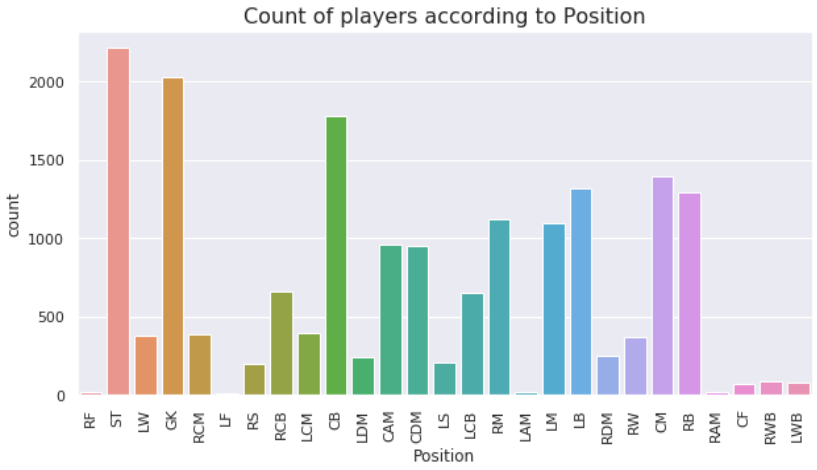


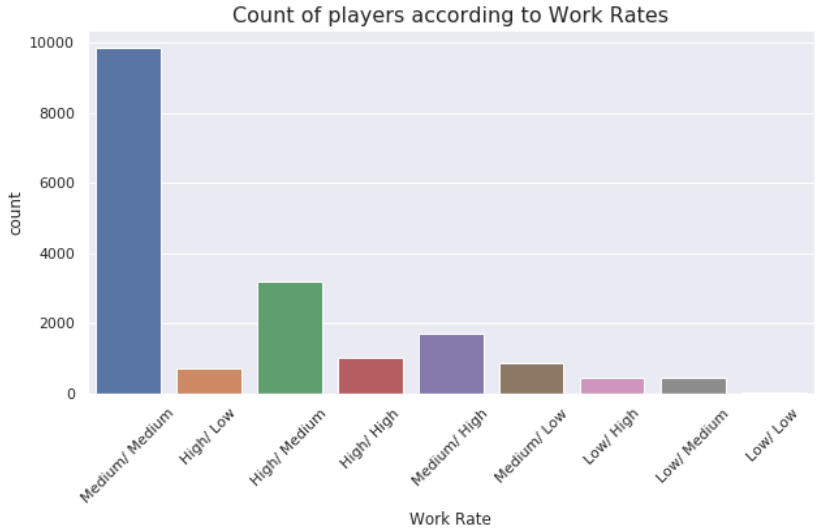
- Young players have more advantage in the market

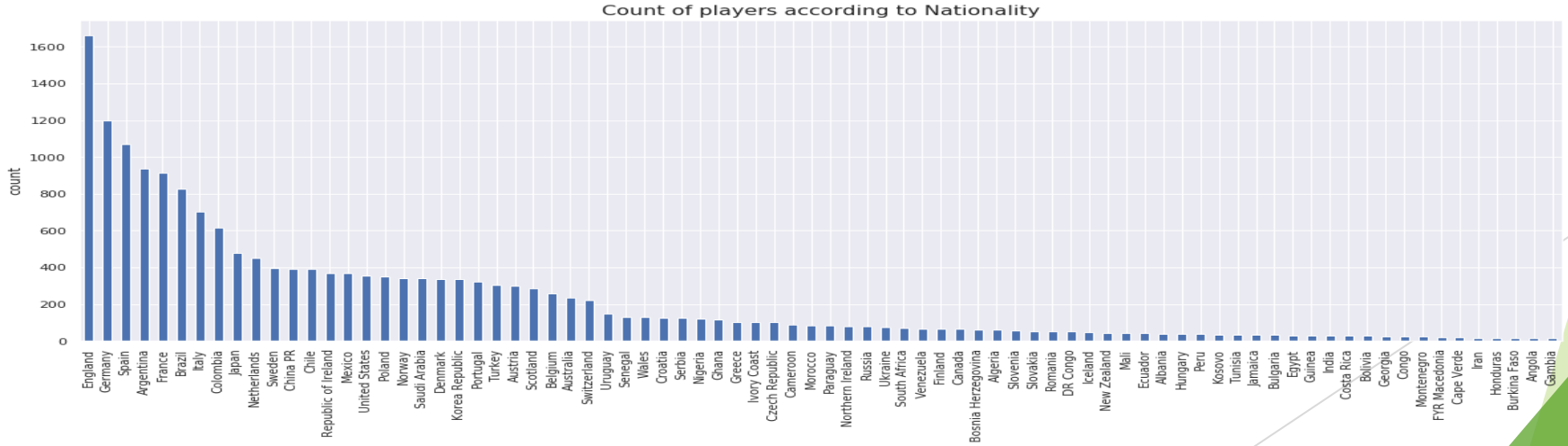
- Fit players tend to survive more in the market



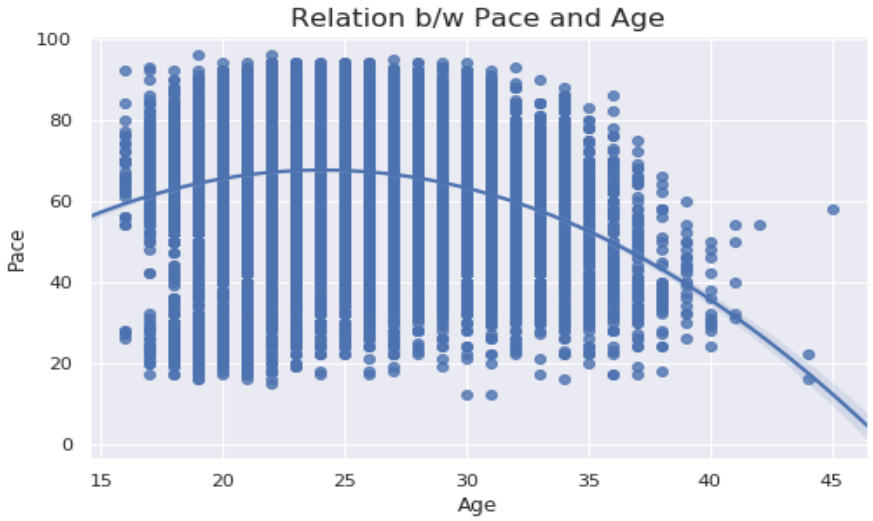
Count Plots


Count of players according to Position
count
Position
RF
ST
LW
GK
RCM
LF
RS
RCB
LCM
CB
LDM
CAM
CDM
LS
LCB
RM
LAM
LM
LB
RDM
RW
CM
RB
RAM
CF
RWB
LWB

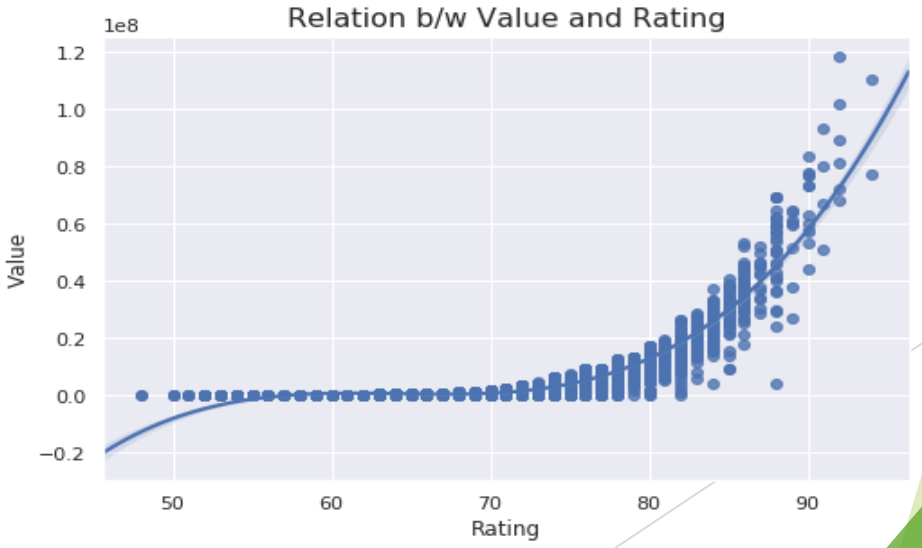

Count of players according to Work Rates
count
Work Rate
Medium/ Medium
High/ Low
High/ Medium
High/ High
Medium/ High
Medium/ Low
Low/ High
Low/ Medium
Low/ Low


Count of players according to Nationality
count
England
Germany
Spain
Argentina
France
Brazil
Italy
Colombia
Japan
Netherlands
Sweden
China PR
Chile
Republic of Ireland
Mexico
United States
Poland
Norway
Saudi Arabia
Denmark
Korea Republic
Portugal
Turkey
Austria
Scotland
Belgium
Australia
Switzerland
Uruguay
Senegal
Wales
Croatia
Serbia
Nigeria
Ghana
Greece
Ivory Coast
Czech Republic
Cameroon
Morocco
Paraguay
Northern Ireland
Russia
Ukraine
South Africa
Venezuela
Finland
Canada
Bosnia Herzegovina
Algeria
Slovenia
Slovakia
Romania
DR Congo
Iceland
New Zealand
Mali
Ecuador
Albania
Hungary
Peru
Kosovo
Tunisia
Jamaica
Bulgaria
Egypt
Guinea
India
Costa Rica
Bolivia
Georgia
Congo
Montenegro
FYR Macedonia
Cape Verde
Iran
Honduras
Burkina Faso
Angola
Gambia

Regression Plots

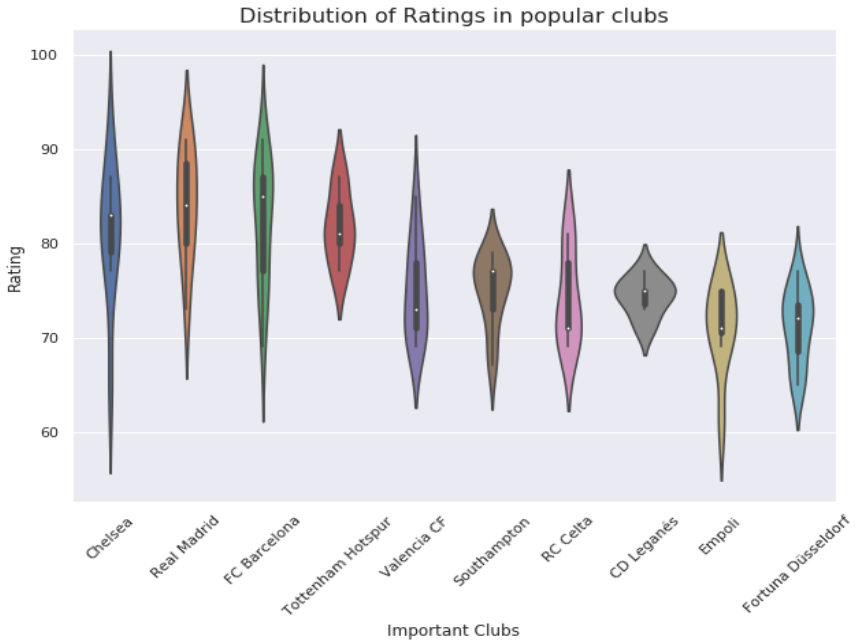


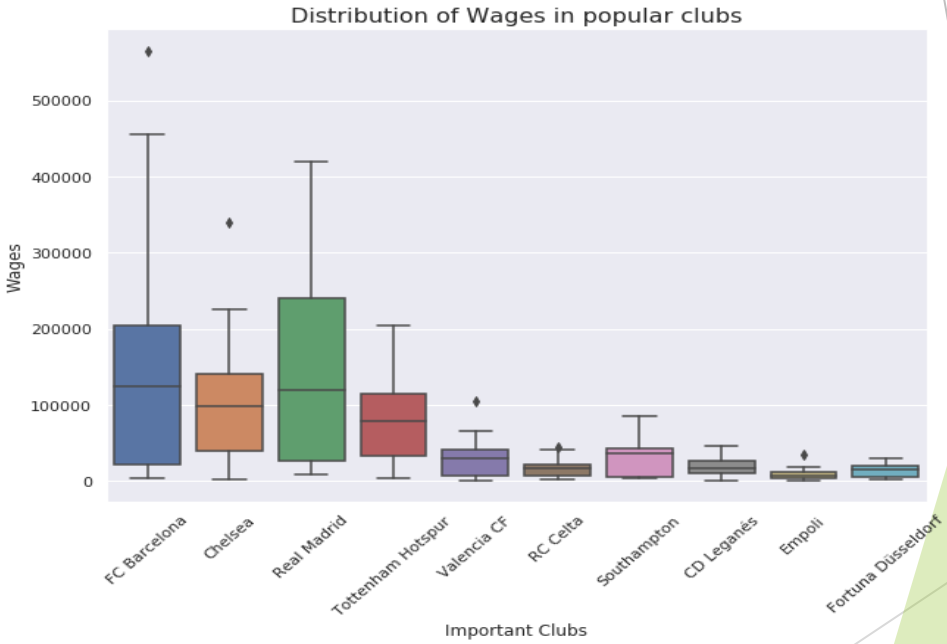
- Greater the age lesser the pace of the player



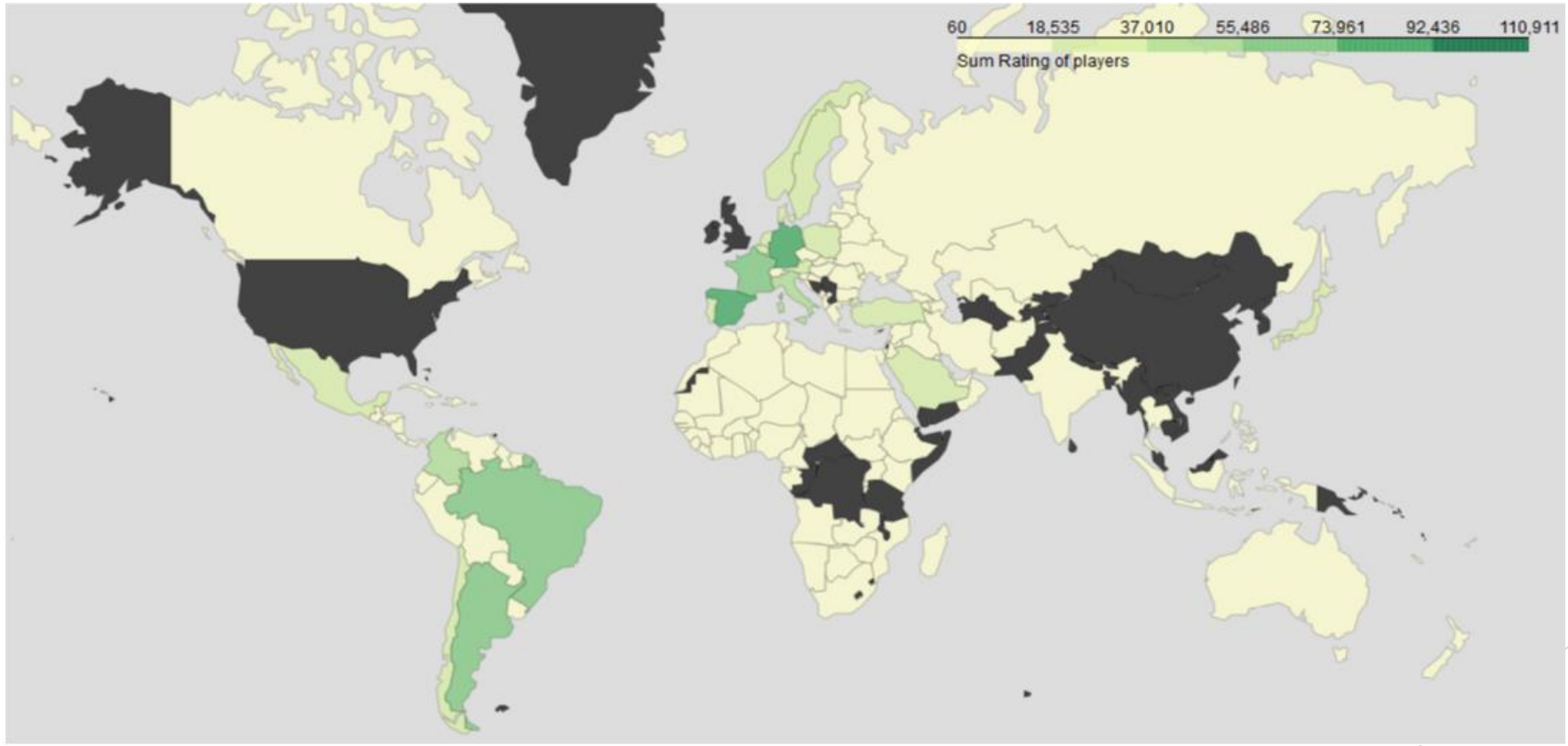
- Higher rated player have higher value in the market

Popular club analysis

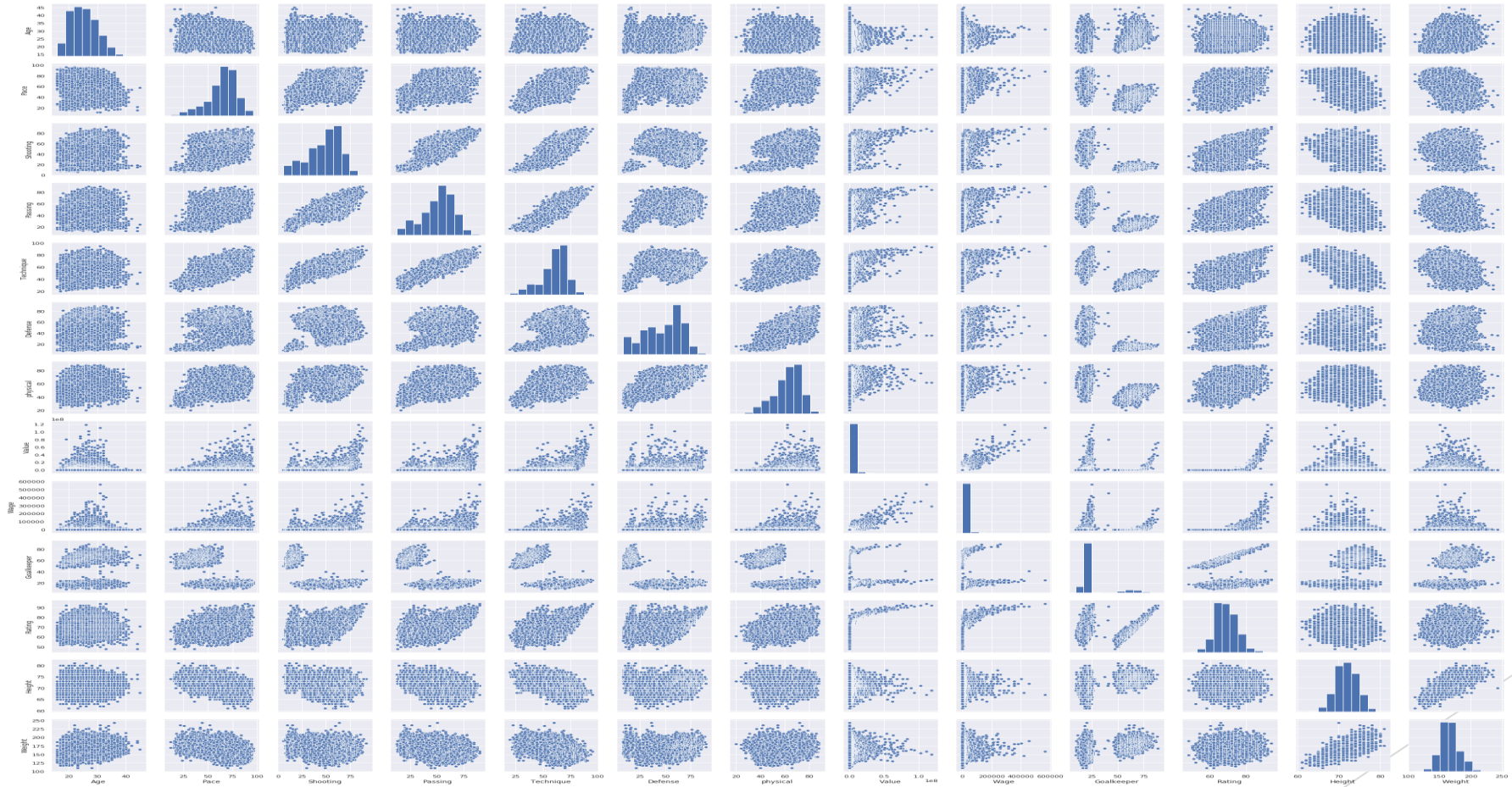




Map of highest rating

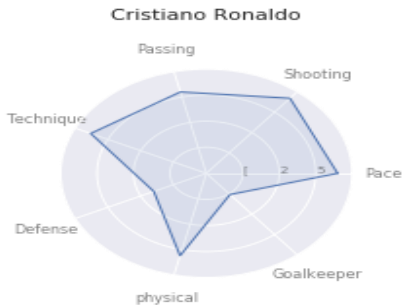


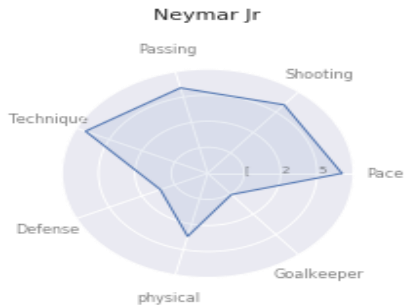
Pair Plot important features



The Top 11


L. Messi
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique

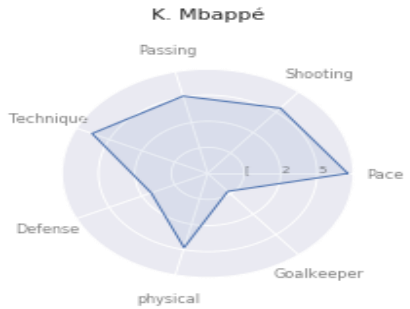

Cristiano Ronaldo
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique

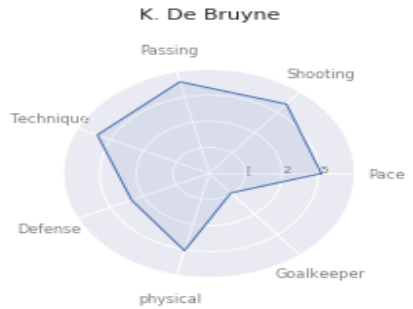

Neymar Jr
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique


J. Oblak
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique


De Gea
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique


P. Dybala
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique


K. Mbappé
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique


K. De Bruyne
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique


L. Suárez
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique

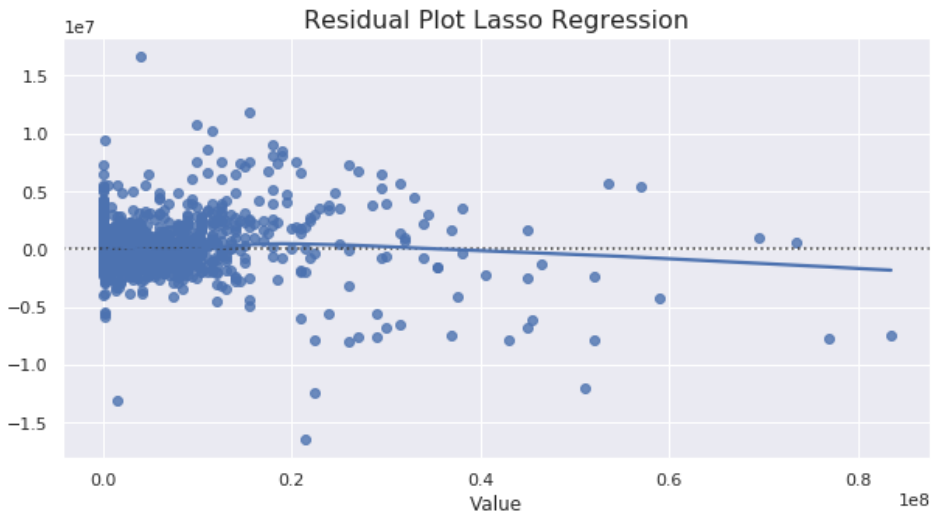

L. Modrić
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique


E. Hazard
Passing
Shooting
Pace
Goalkeeper
physical
Defense
Technique

Modelling

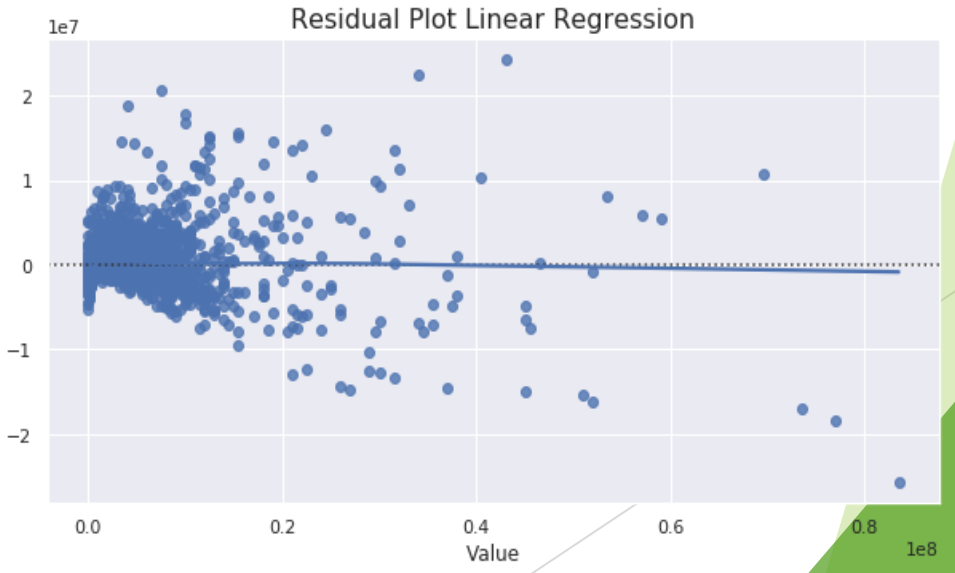
- Formation of training and target sets.
- Conversion of position into numerical features
- Scaling of data for modelling the train set
- Employing Linear, Lasso, Ridge and Random Forest Regression

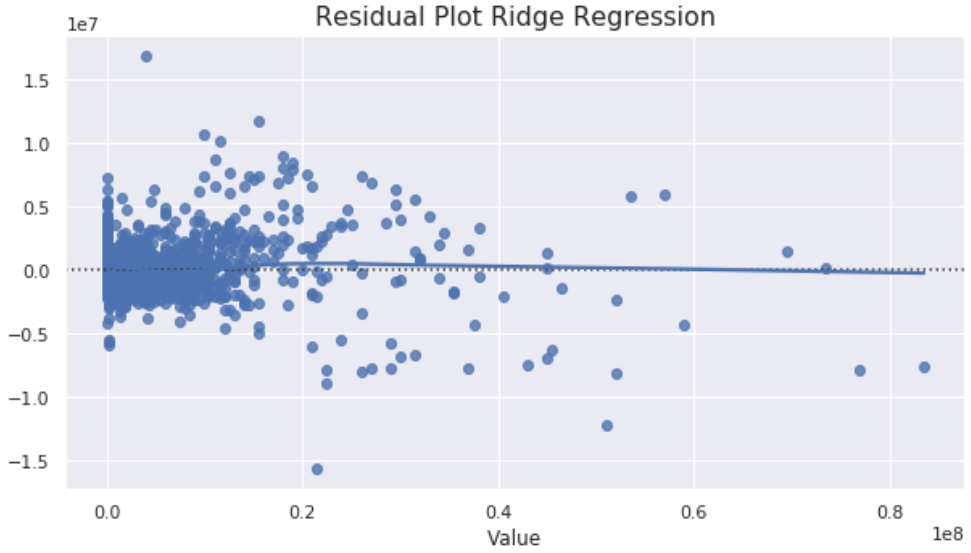
Regression models



- Lasso regression with polynomial feature of degree two
- R2 score is 0.93

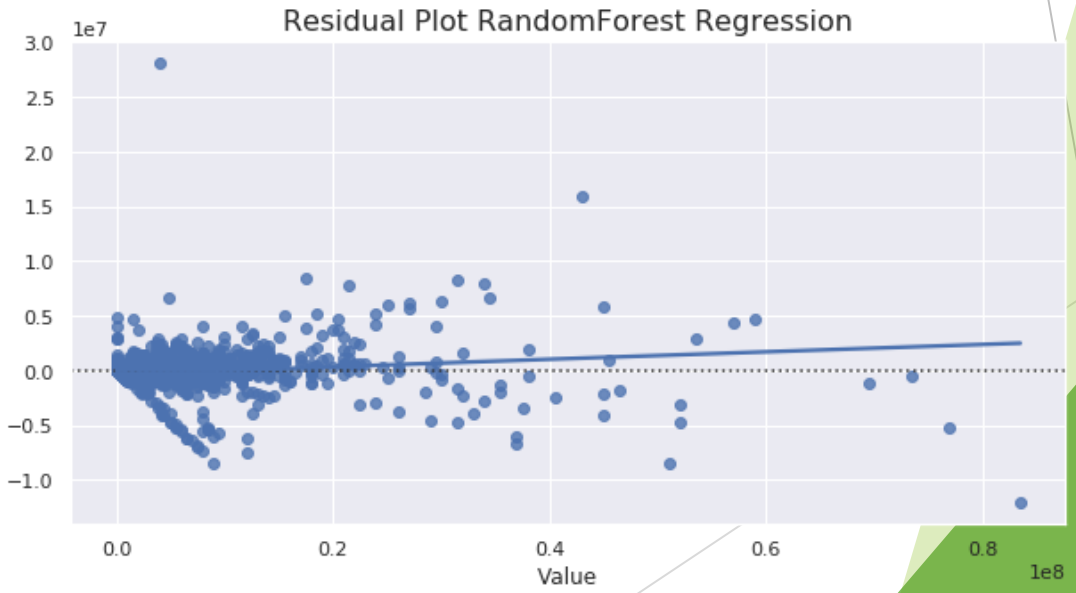
- Simple linear Regression without polynomial feature.
- R2 score is 0.77





- Ridge regression with polynomial feature of degree two
- R2 score is 0.94

- Random Forest regressor
- R2 score is 0.96



Result and Conclusions

Regression Models	R2 Scores
Linear Regression	0.77
Lasso Regression	0.93
Ridge Regression	0.94
Random Forest Regressor	0.96

- The models performed well on the test data but it could be much better by taking in account the outliers because there were many player's having very less value and some having very high value.
- These models could also be used to predict the ratings of the value according to their skills.
- The clubs could use these to predict their player's value in the market and to solve many business problems.