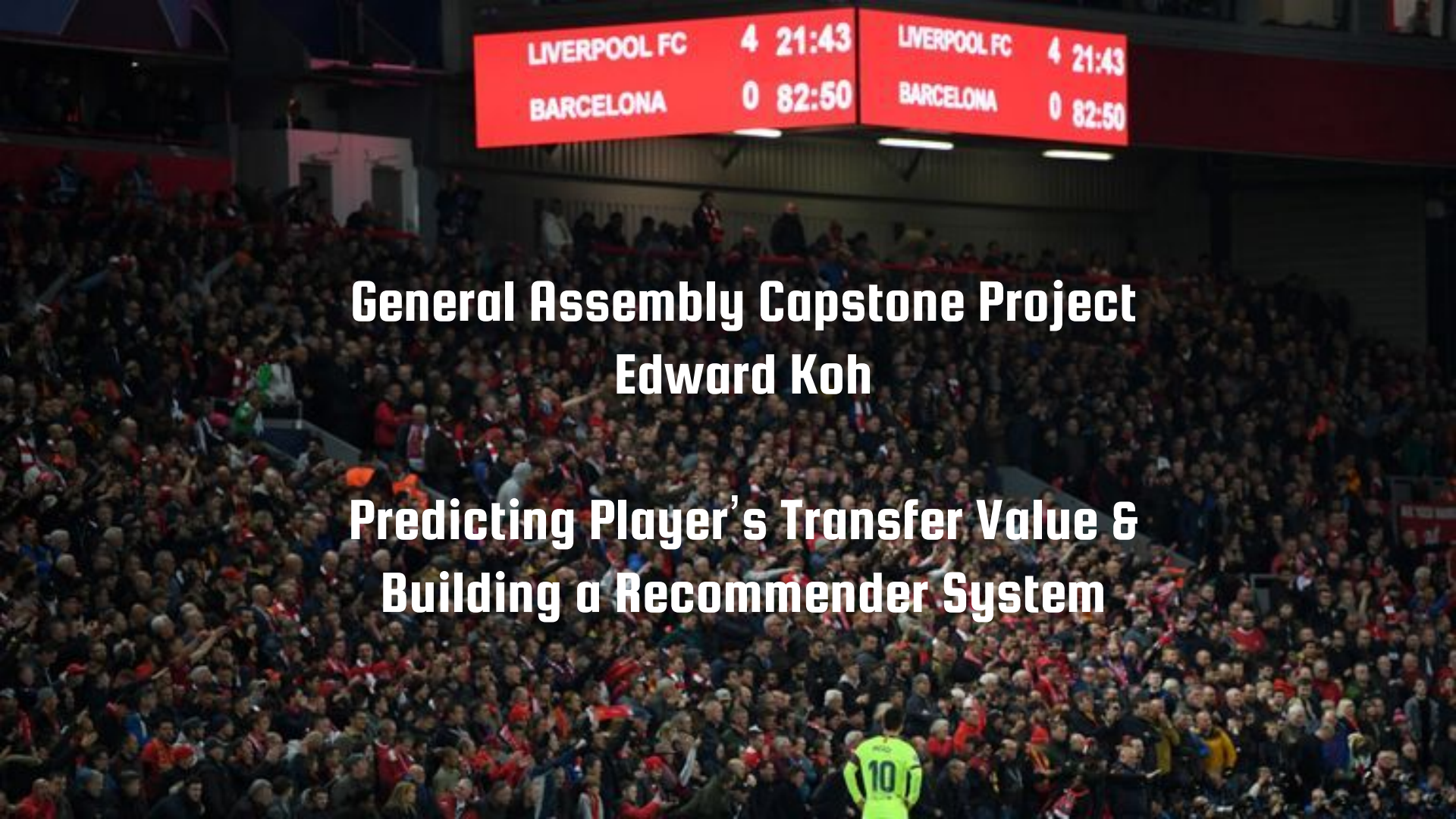


General Assembly Capstone Project

Edward Koh

Predicting Player's Transfer Value & Building a Recommender System

TABLE OF CONTENTS

01

Introduction

INTRODUCTION

Football is a world renowned sport with 4 billion fans worldwide.

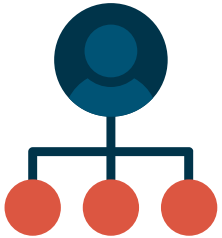
- Especially popular in Europe, South America
- Countries have their respective leagues and competitions.
- Football clubs are able to purchase players from another club by paying a transfer fee.
- Dataset: Fifa19 data (Kaggle)
- Dimensions: 18207 rows, 89 columns

PROBLEM STATEMENT

Football Clubs have been overpaying for player transfers, along with making impulse purchases on players which do not suit the club. This leads to massive financial consequences should the new player fail to perform up to expectations.

Regression Model

Creating a regression model which correctly predicts a player's transfer value based on various features.



Recommender System

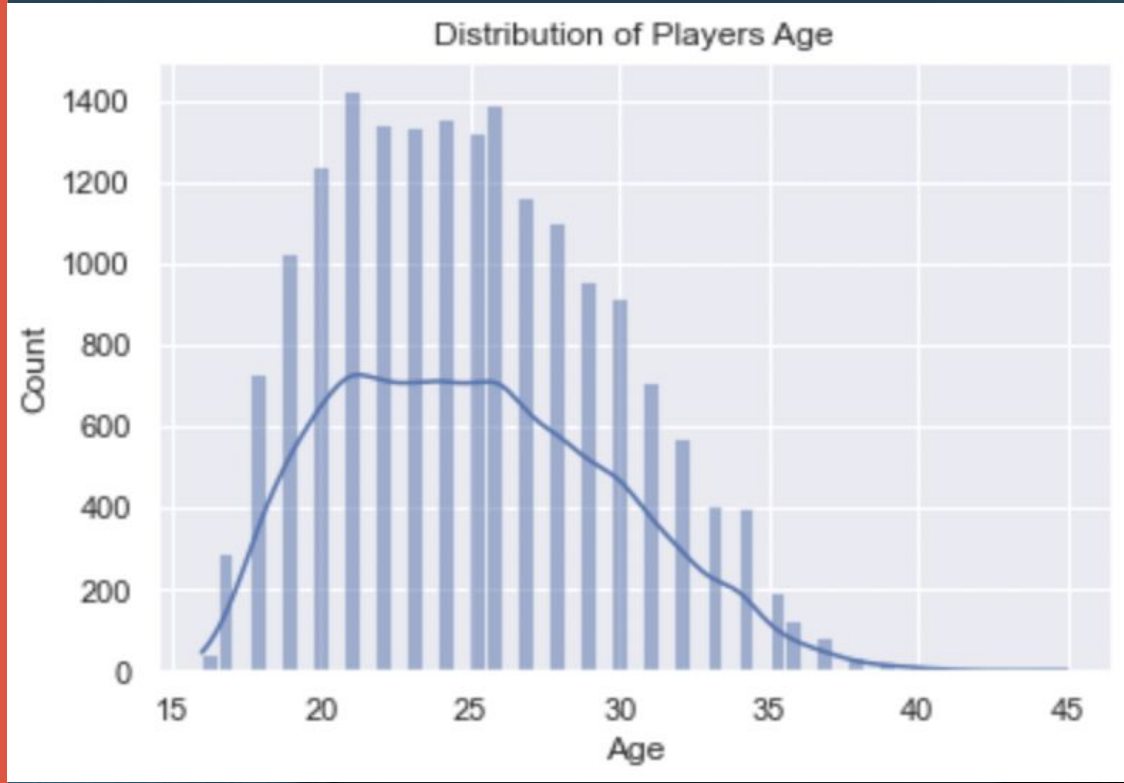
Building a recommender system to to search for similar players to replace departing stars.

02
Exploratory Data Analysis

Age

Mean: 25.12 years old
Median: 25 years old
Mode: 21 years old

Youngest: 16 years old
Oldest: 45 years old

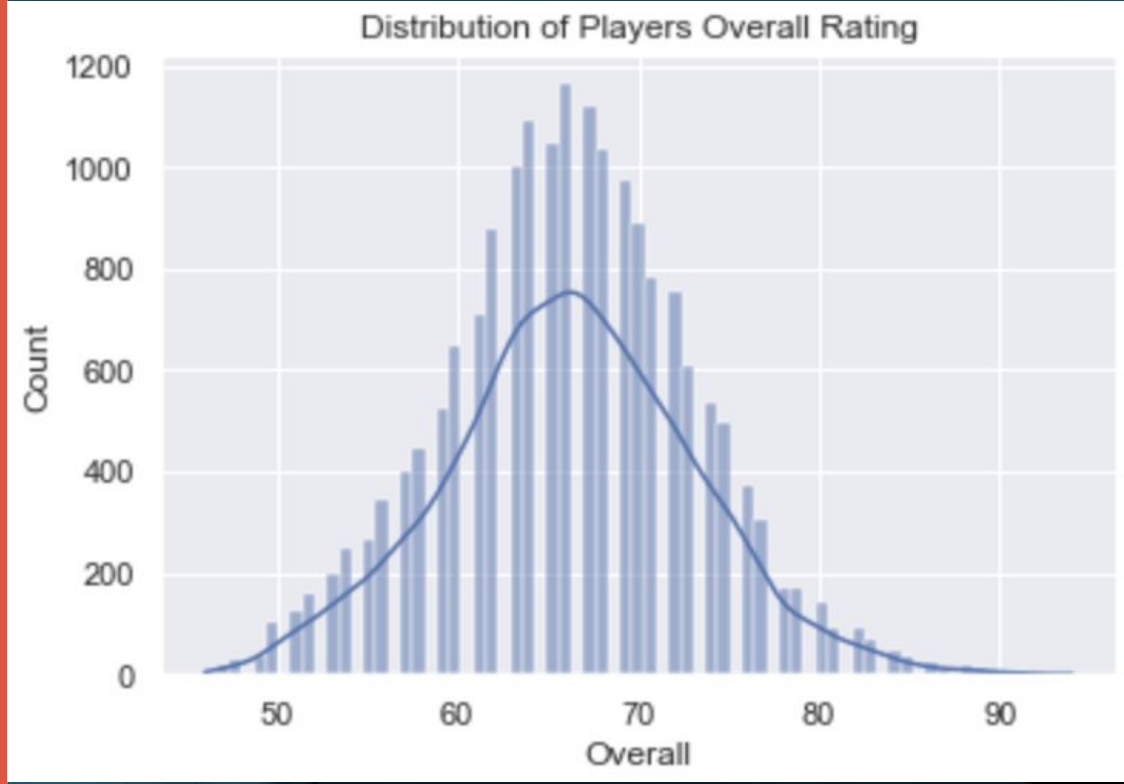


Overall

Mean, Median & Mode: 66

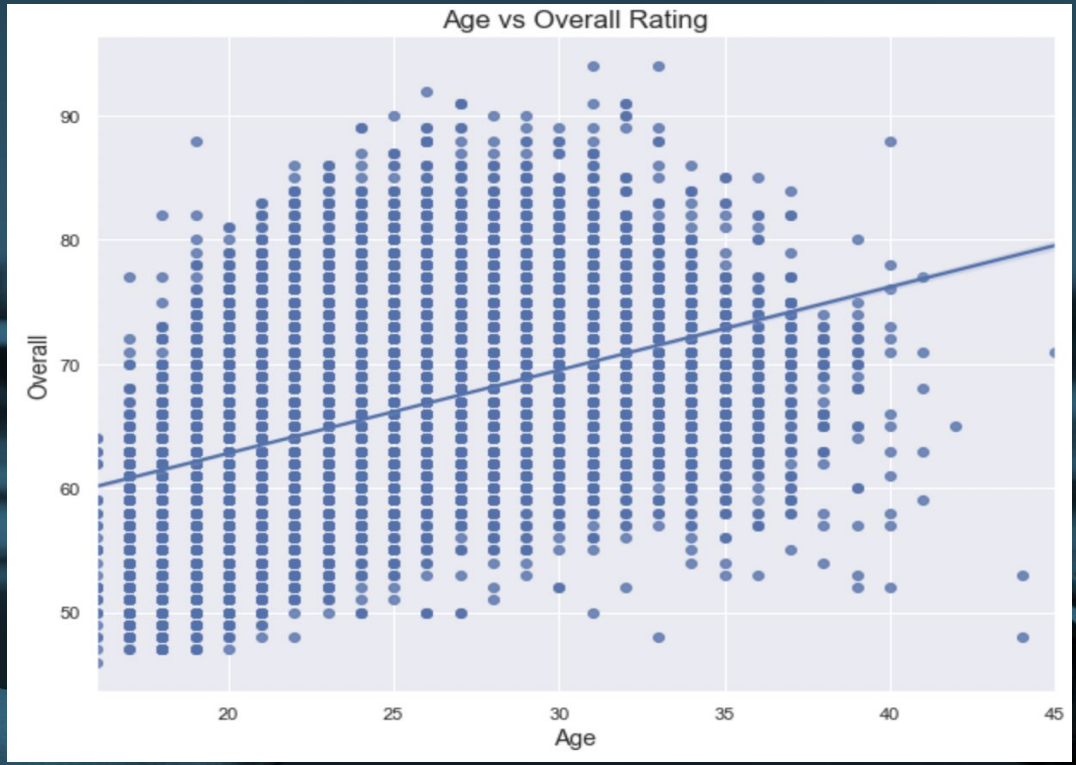
Highest: 94

Lowest: 46



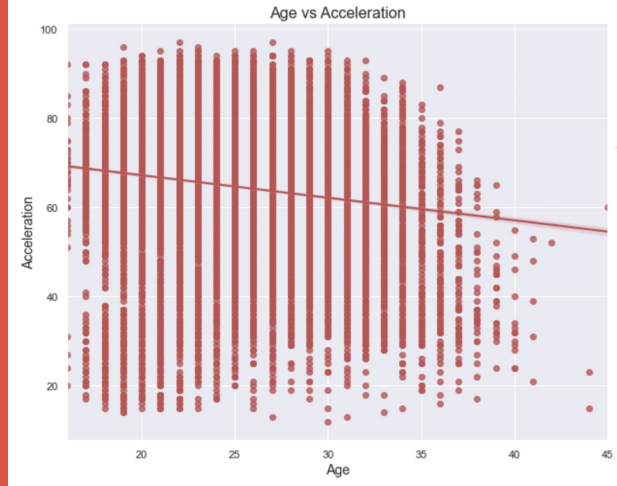
Overall vs Age

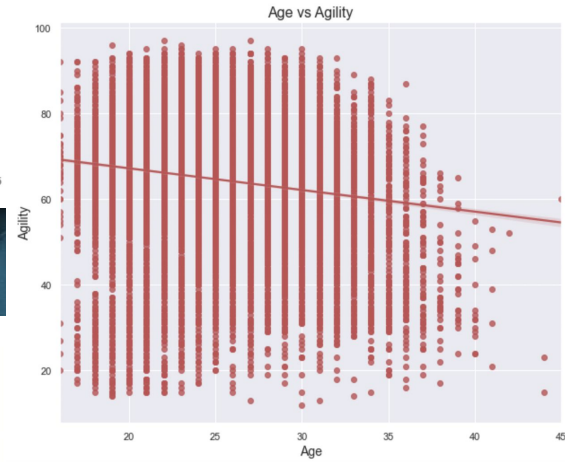
Player's overall score increase as they age.



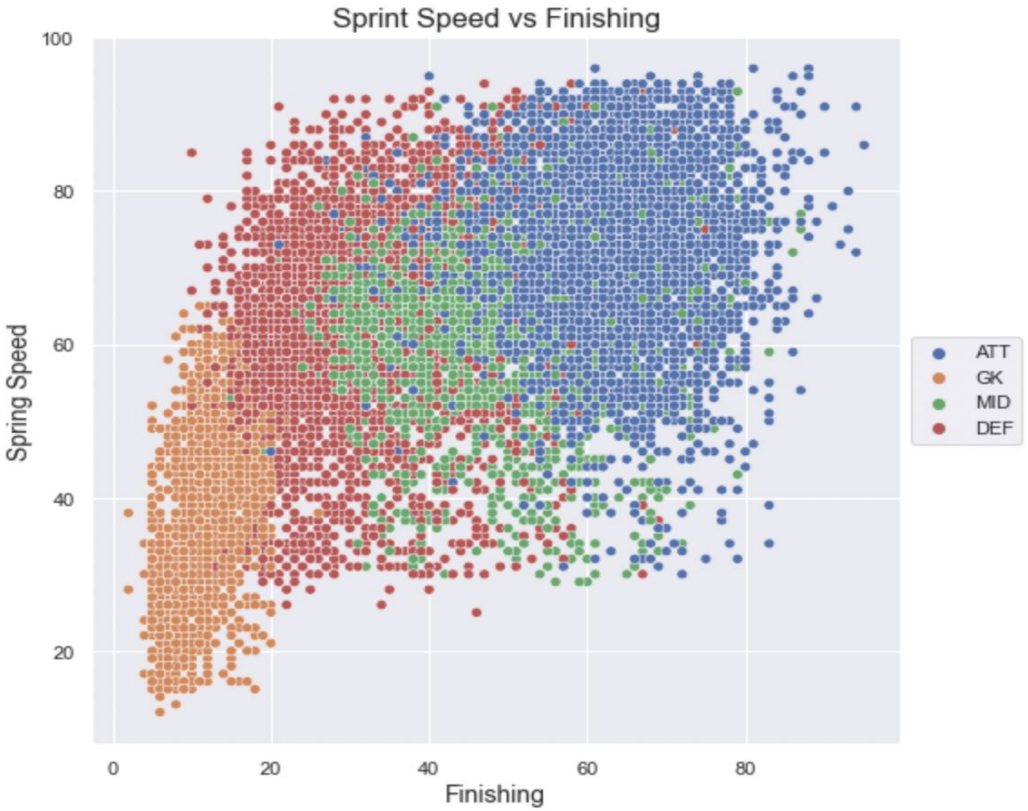
Speed, Acceleration & Agility vs Age

Sprint Speed, Acceleration and Agility decreases over time as players age.





Position

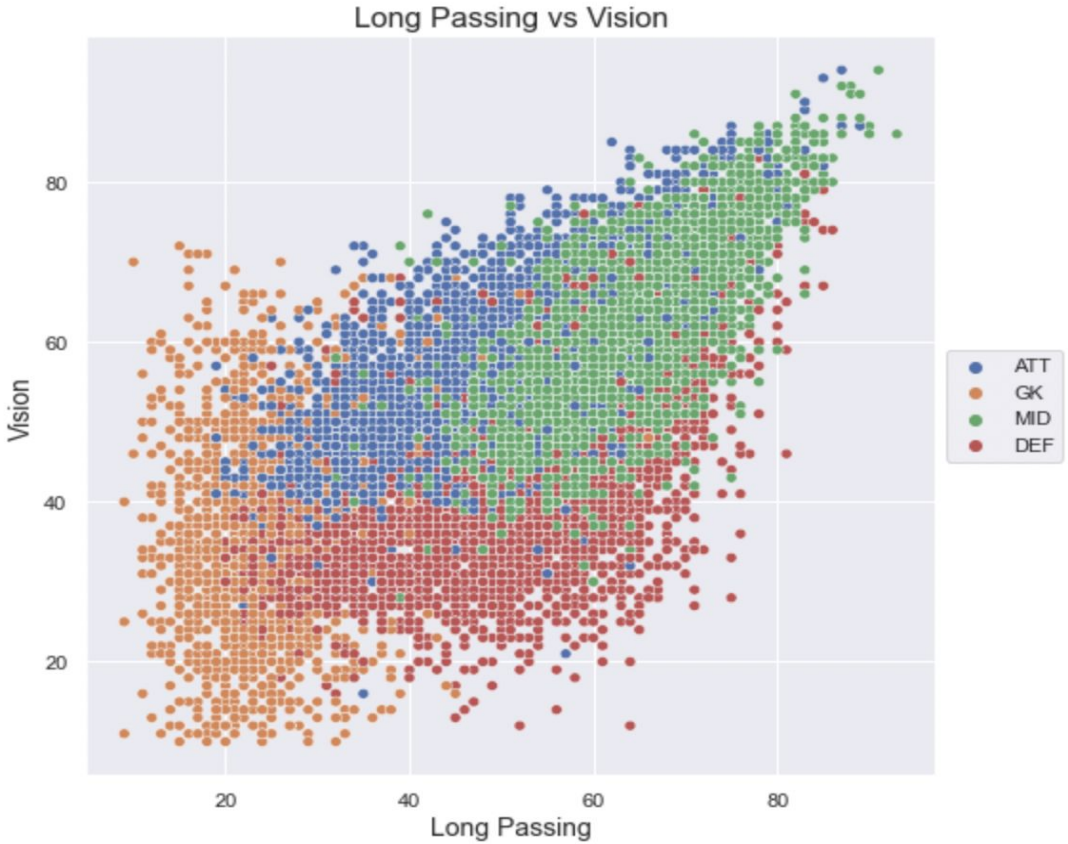

Sprint Speed vs Finishing
Spring Speed
Finishing
100
80
60
40
20
0
20
40
60
80
ATT
GK
MID
DEF


Centre forward
Centre forward
Left midfield
Centre midfield
Centre midfield
Right midfield
Left-back
Centre-back
Centre-back
Right-back
Goalkeeper

Position


Interception vs Aggression
Aggression
Interception
0
20
40
60
80
20
40
60
80
ATT
GK
MID
DEF

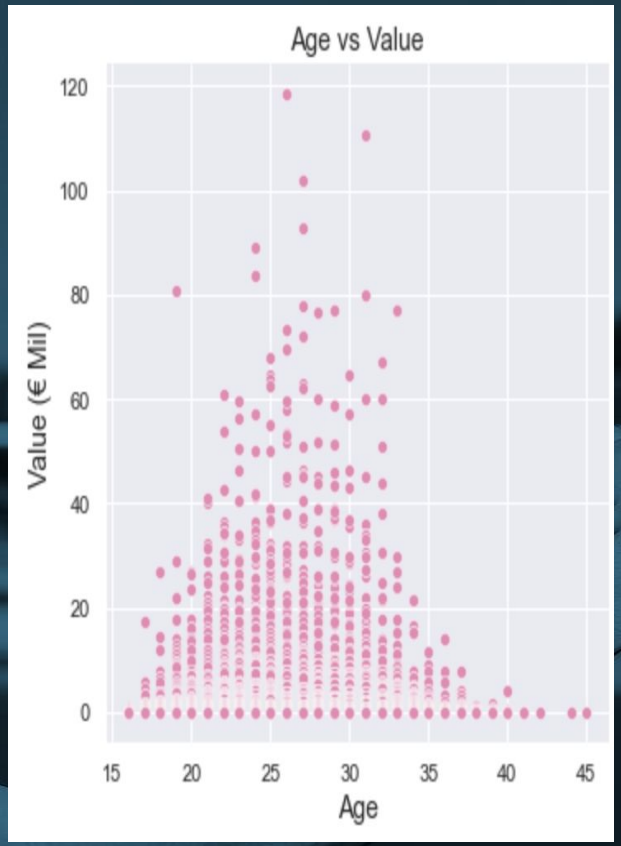
Position

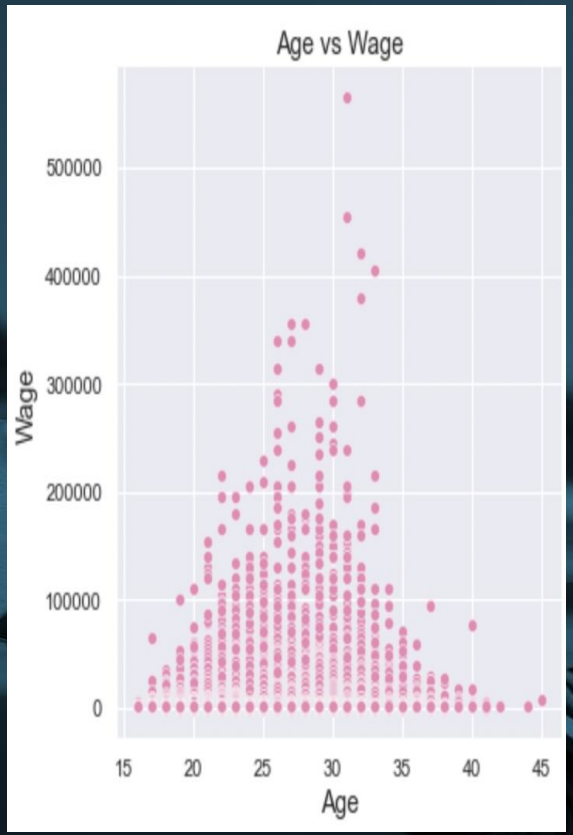

Long Passing vs Vision
Vision
Long Passing
20
40
60
80
ATT
GK
MID
DEF

Wage, Value vs Age

Players peak at 27-31 years old in general.

During that period, they are at the top of the game thus their wages and valuation are at a high.

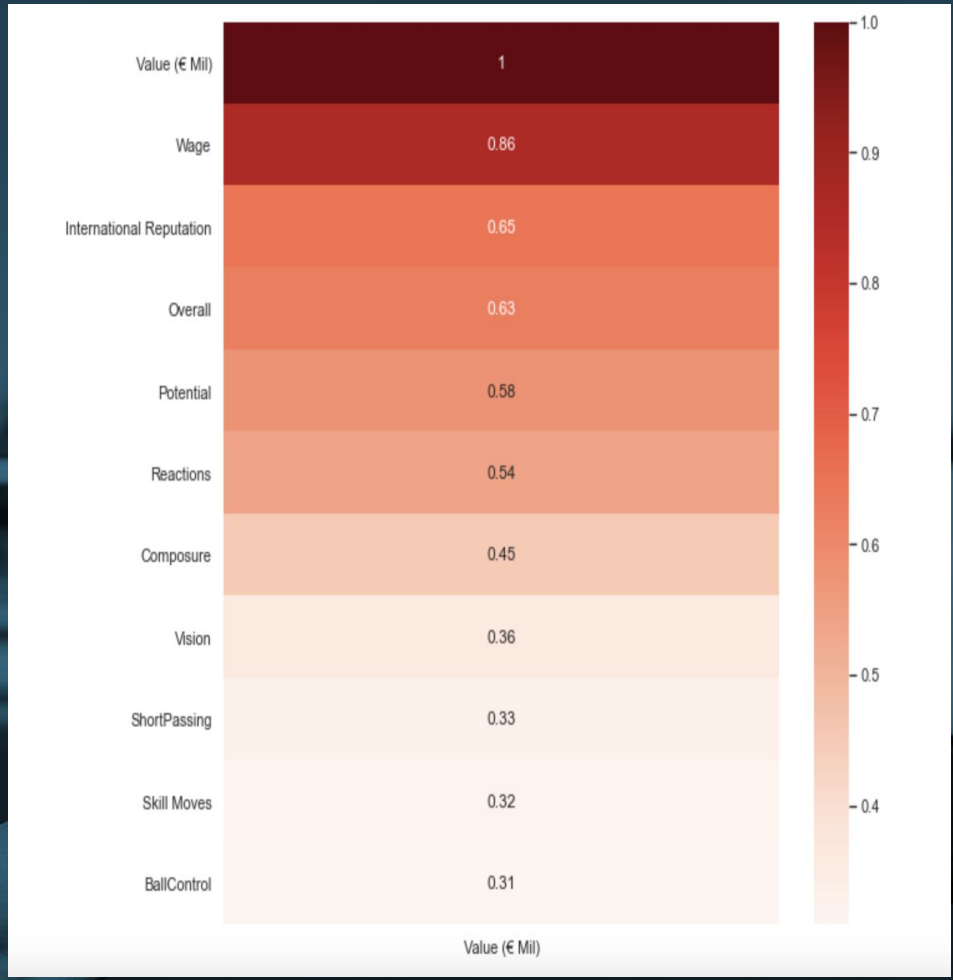




Top 10 Attributes

Top 3 features with highest correlation to a player's value:

1) Wage
2) International Reputation
3) Overall




Fly
Emirates
EA SPORTS
FIFA19
FIFA
OFFICIAL LICENSED PRODUCT

03

Modeling

Models Used

1) Linear Regression

2) Lasso Regression

3) Ridge Regression

4) Random Forest Regression

FIRST TEST: RAW DATASET

CROSS VAL SCORE

	Linear Regression	Lasso Regression	Ridge Regression	Random Forest
Cross Validation Score	-2.931560e+25	0.843481	0.842346	0.971896

ACCURACY & VARIANCE

	Lasso Regression	Ridge Regression	Random Forest
Train Score	0.877622	0.873287	0.996732
Accuracy Score	0.834774	0.834733	0.980449
Variance	0.042847	0.038554	0.016284

SECOND TEST: CLEANED DATASET

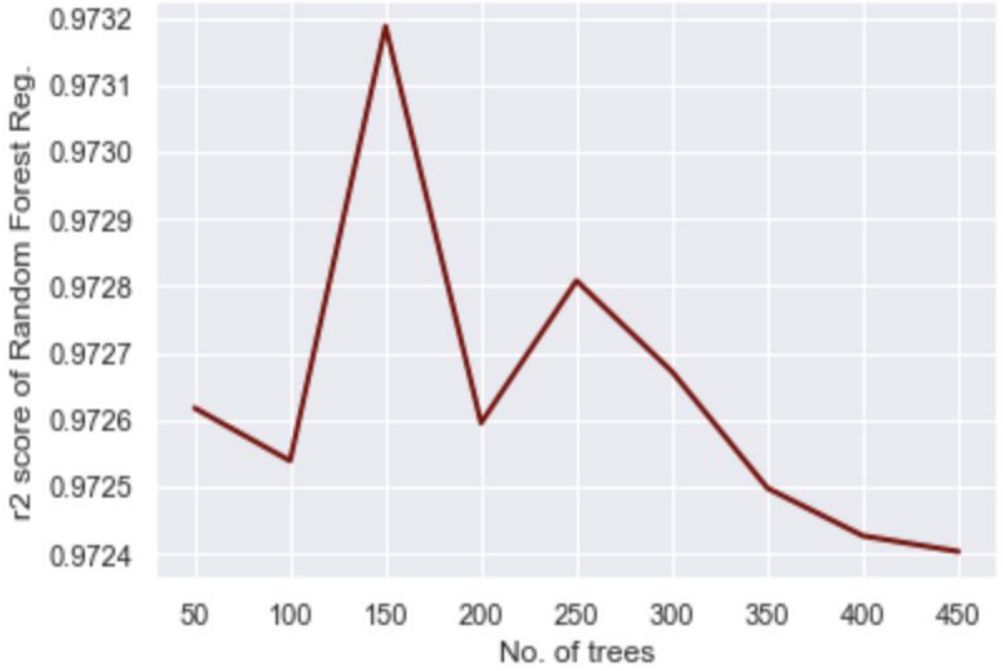
CROSS VAL SCORE

	Linear Regression	Lasso Regression	Ridge Regression	Random Forest
Cross Validation Score	-6.955243e+24	0.811363	0.810203	0.960073

ACCURACY & VARIANCE

	Lasso Regression	Ridge Regression	Random Forest
Train Score	0.840692	0.839766	0.994509
Accuracy Score	0.841983	0.843025	0.972539
Variance	0.001291	0.003259	0.021970

Getting Best Parameters



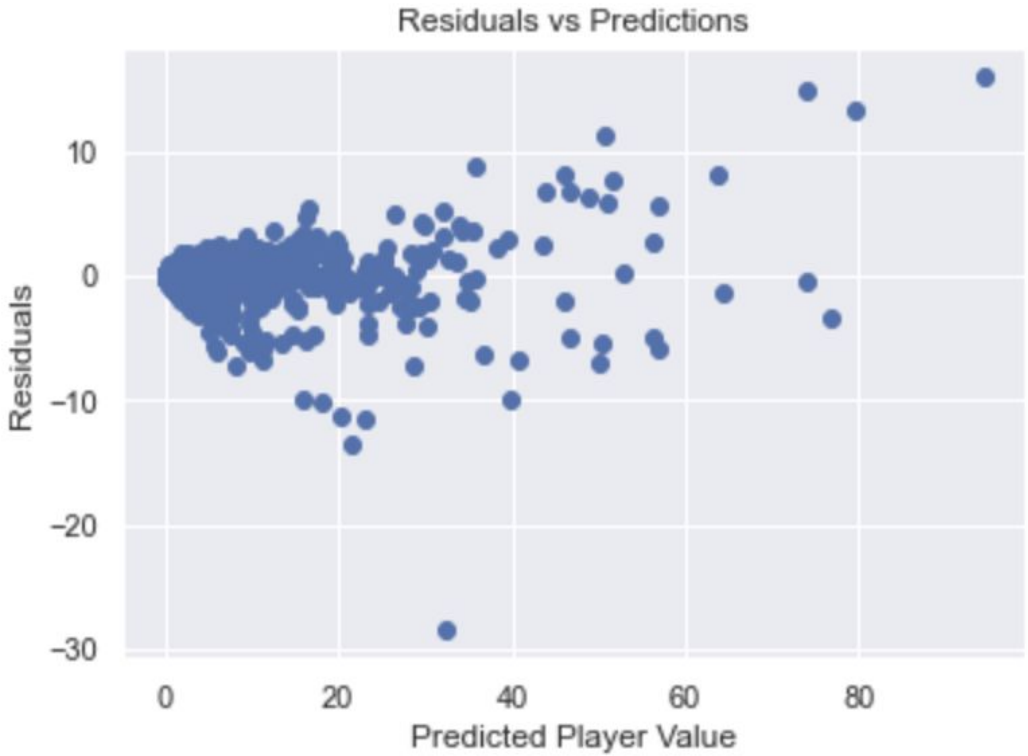
Final Model

	Random Forest Regression
Train Score	0.994581
Accuracy Score	0.973188
Variance	0.021393
Cross Validation Score	0.960449

Predictions

	ID	Real Values	Predicted Values
0	229692	1.20	1.18
1	239605	0.27	0.20
2	214981	0.40	0.41
3	189235	0.78	0.67
4	204497	9.00	8.70
5	237631	4.90	4.74
6	244538	0.32	0.35
7	189839	0.35	0.48
8	226416	0.62	0.40
9	217758	0.82	0.82

Residuals


Residuals vs Predictions
Residuals
10
0
−10
−20
−30
0
20
40
60
80
Predicted Player Value

Feature Importance

Overall is the most important feature considered by the Random Forest Regressor in determining the target (Player's Value).

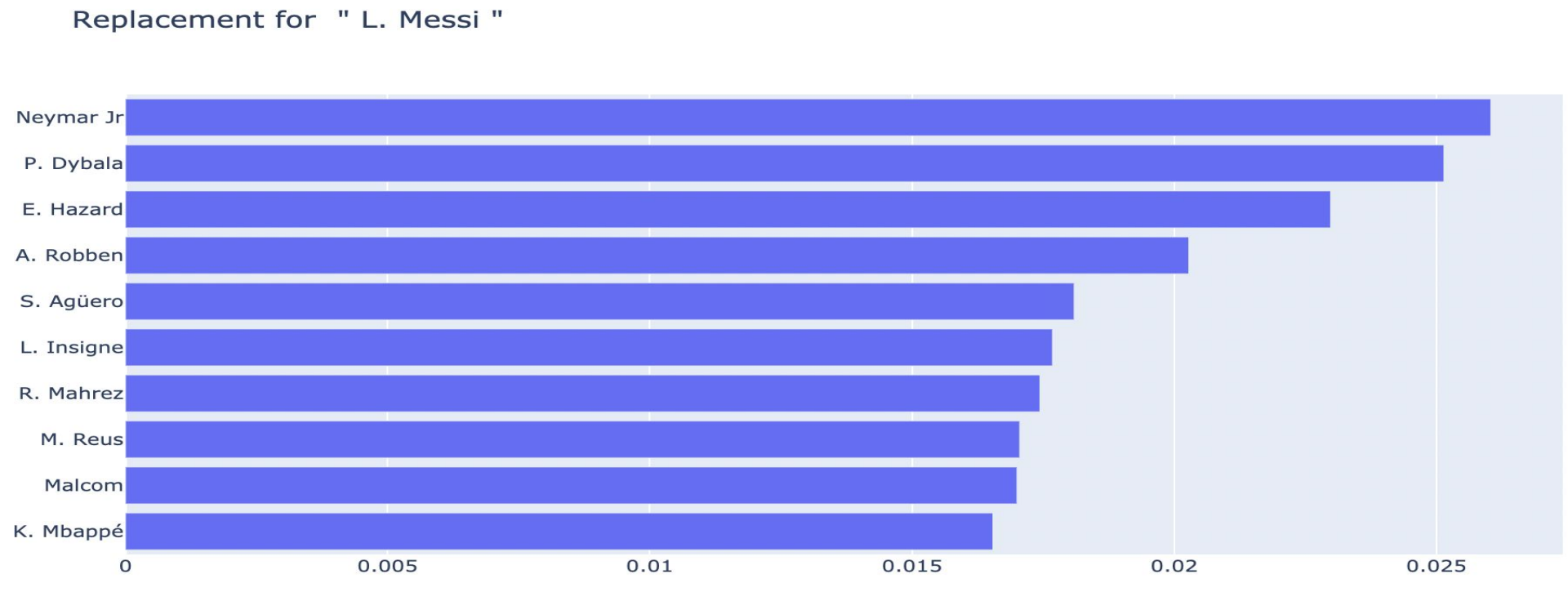
	Feature	Importance
0	Overall	0.84710
1	Potential	0.07731
2	Finishing	0.00892
3	Composure	0.00836
4	Reactions	0.00797
5	Wage	0.00639
6	Positioning	0.00403
7	Stamina	0.00392
8	LongShots	0.00391
9	BallControl	0.00386

04

Recommender System

GAFC Player Scouter ver 1.38

Testing the Recommender System



GAFC Player Scouter Ver I.38 Scout Report

	Club	Name	Age	Nationality	Main Position	Overall	Potential	Preferred Foot	Weak Foot	Wage	Value (€ Mil)
0	Paris Saint-Germain	Neymar Jr	26	Brazil	WING	92	93	1	5	290000	118
1	Juventus	P. Dybala	24	Argentina	ST	89	94	0	3	205000	89
2	Chelsea	E. Hazard	27	Belgium	ST	91	91	1	4	340000	93
3	FC Bayern München	A. Robben	34	Netherlands	WING	84	84	0	2	110000	16
4	Manchester City	S. Agüero	30	Argentina	ST	89	89	1	4	300000	64
5	Napoli	L. Insigne	27	Italy	WING	88	88	1	3	165000	62
6	Manchester City	R. Mahrez	27	Algeria	WING	85	85	0	4	205000	40
7	Borussia Dortmund	M. Reus	29	Germany	WING	86	86	1	4	100000	44
8	FC Barcelona	Malcom	21	Brazil	WING	82	89	0	4	140000	32
9	Paris Saint-Germain	K. Mbappé	19	France	WING	88	95	1	4	100000	81

05

Conclusion

Summary

With accuracy as the success metric for the final model, we are able to predict a players' value based on the features and also which features are most important in determining the forecast.

In the perspective of a financial view of a football club, this allows us to draft up better transfer plans and keep to our budget. It also lower the possibility of overpaying for a particular player thus putting greater financial strain on the transfer budget and club financials.

Future Work

1) Try the model out on FIFA21 dataset when it is available.

2) Try out more variations of the dataset to see if other models can produce a better performance than Random Forest Regression.

3) Create a flask API web page to get comments about the recommender system's effectiveness and improve from there.

The End