FOOTBALL PLAYER RATING PREDICTION

Guided By-
Prof. Suhas Joshi

PA 68 Akshit Langeh
PB 75 Shubh Sharma
PA 58 Prathamesh Joshi
PB 07 Aditya Jaiswal

Problem Statement

- The problem of evaluating the performance of a football player is attracting a lot of companies and scientific communities, thanks to the availability of a massive data being generated during a match.
- Thus the football player rating predictor is a consolidated metric that can be used to not only provide the present but also the future potential ratings of a player.

Scope

- Selecting new talents is unfeasible for Player Scouts as it is too much time consuming and there is a major need for data analysis.
- Thus data-driven performance scores could help in selecting a small subset of the best players who meet specific constraints or show some pattern in their performance.
- This allows scouts and clubs to analyse a larger set of players and saving considerable time and economic resources while broadening scouting operations and career opportunities of the talented players.
- Not only in football but also various fantasy football uses such evaluations to compare performance of professional players. Due to availability of a massive data being generated during a match training dataset to our needs is feasible.
- Also this model is not limited to only football any other sport with similar database.

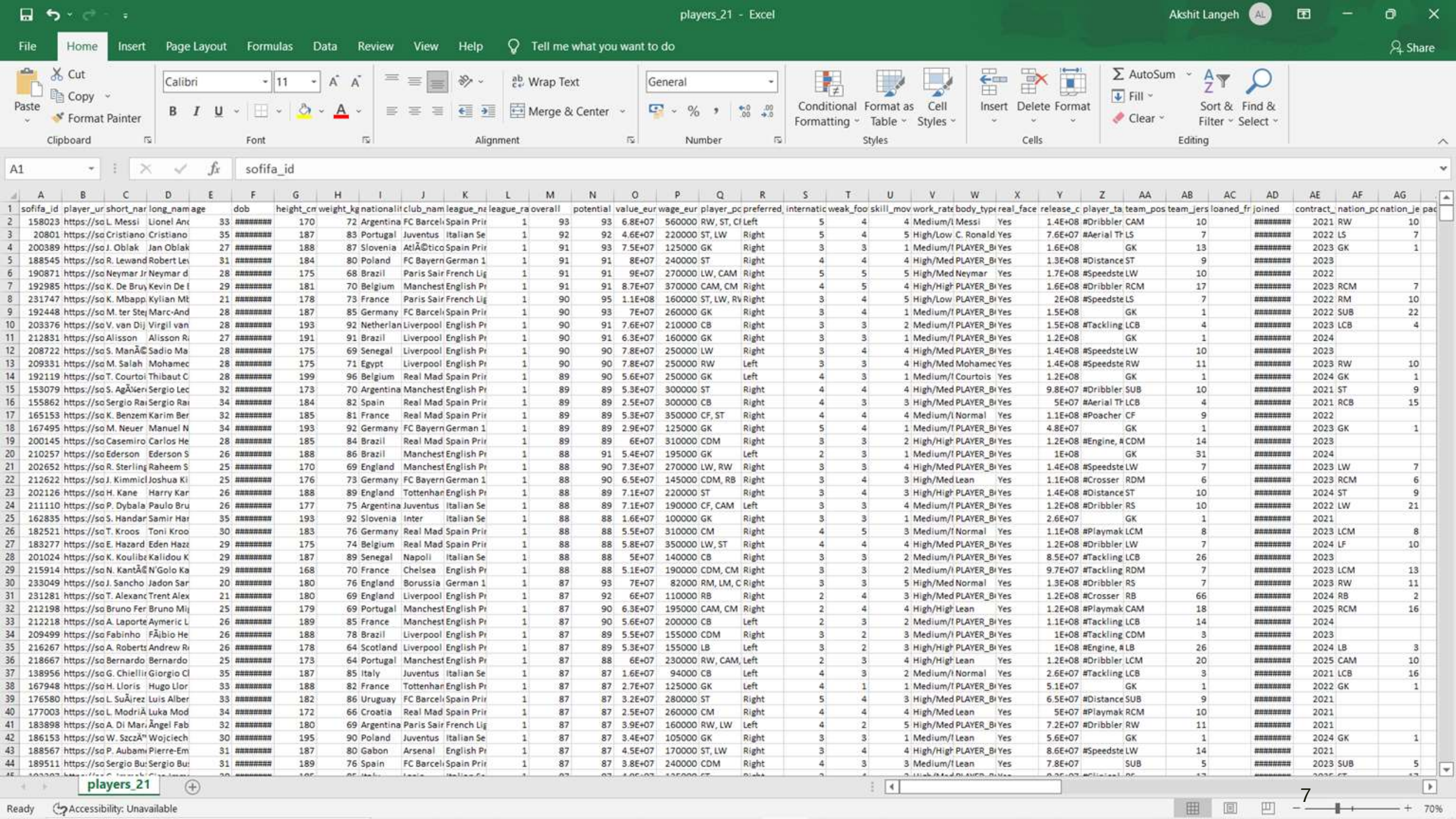
Literature Survey

Sr. No	Title of the paper	Year of Publication/Place of publication	Algorithm/ Methodology used	Advantages/ highlight of Paper	Gap of Paper
1	PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach	25 Jan 2019 / ITALY	Linear Support Vector Machine and Clustering	1)detailed analysis of multi dimensional data.	The data is becoming biased because it is being generated match wise.
2	Influences of player nationality, playing position, and height on relative age effects at women's under-17 FIFA World Cup	August 2012	Chi-square analysis	1) RAEs do exist in adult women's youth soccer. 2) Age Is taken into consideration.	It is difficult to assume RAEs in professional football.
3	A football player rating system	October 2020	ELO Algorithm	1) The ELO Algorithm is an adaptive rating system used in various individual sports. 2) It has been expanded in accordance with the requirements for team sports	This Algorithm is made for individual sports analysis.

4	On the Development of a Soccer Player Performance Rating System for the English Premier League	August 2012	EA Sports PPI	1) Takes many factors into consideration. 2) Points Sharing 3) Appearance Points 4) Assists Points	Complexities arise with bigger data and more precise methodology.
5	A methodology for football players selection problem based on multi-measurements criteria analysis.	December 2017	Multi Criteria Decision Making.	1) Groups of players are made for better evaluation. 2) TOPSIS method for selection	Grouping will highlight the best group, individual performances cant be evaluated.

About Dataset

- The datasets include player data after end of a season (in our case season 2019-2020) and is taken from Kaggle.
- There are about 100+ attributes present in the dataset that includes almost all the details about all the football players active in that season.
- Data like player position and also the role of player in the team plus player attributes with statistics as Attacking, Skills, Defense, Mentality, GK Skills, etc.
- Also in addition player's personal data like Nationality, Club, Date Of Birth, Wage, Salary, etc. has been provided along with the URL of the scraped player
- Hence from this dataset we can train a model to not only predict player's overall rating but also the potential that player carries.
- With this dataset is being used to train the model, each member of our group has also decided to create a sample database consisting of players that are yet to be professional, so that we can predict how much potential such players carry in order to be a transfer target for the scouts.



sofifa_id	player_ur	short_nar	long_nam	age	dob	height_cn	weight_kg	nationalit	club_nam	league_na	league_ra	overall	potential	value_eur	wage_eur	player_pc	preferred_	internatic	weak_foo	skill_mov	work_rate	body_type	real_face	release_c	player_ta	team_pos	team_jers	loaned_fr	joined	contract_	nation_pc	nation_je	pac
158023	https://so	L. Messi	Lionel And	33	########	170	72	Argentina	FC Barcel	Spain Prir	1	93	93	6.8E+07	560000	RW, ST, CF	Left	5	4	4	Medium/l	Messi	Yes	1.4E+08	#Dribbler	CAM	10		########	2021	RW	10	
20801	https://so	Cristiano	Cristiano	35	########	187	83	Portugal	Juventus	Italian Se	1	92	92	4.6E+07	220000	ST, LW	Right	5	4	5	High/Low	C. Ronald	Yes	7.6E+07	#Aerial Th	LS	7		########	2022	LS	7	
200389	https://so	J. Oblak	Jan Oblak	27	########	188	87	Slovenia	AtlÃ©tico	Spain Prir	1	91	93	7.5E+07	125000	GK	Right	3	3	1	Medium/l	PLAYER_BI	Yes	1.6E+08		GK	13		########	2023	GK	1	
188545	https://so	R. Lewand	Robert Le	31	########	184	80	Poland	FC Bayern	German 1	1	91	91	8E+07	240000	ST	Right	4	4	4	High/Med	PLAYER_BI	Yes	1.3E+08	#Distance	ST	9		########	2023			
190871	https://so	Neymar Jr	Neymar d	28	########	175	68	Brazil	Paris Sair	French Lig	1	91	91	9E+07	270000	LW, CAM	Right	5	5	5	High/Med	Neymar	Yes	1.7E+08	#Speedste	LW	10		########	2022			
192985	https://so	K. De Bru	Kevin De B	29	########	181	70	Belgium	Manchest	English Pr	1	91	91	8.7E+07	370000	CAM, CM	Right	4	5	4	High/High	PLAYER_BI	Yes	1.6E+08	#Dribbler	RCM	17		########	2023	RCM	7	
231747	https://so	K. Mbapp	Kylian Mb	21	########	178	73	France	Paris Sair	French Lig	1	90	95	1.1E+08	160000	ST, LW, RV	Right	3	4	5	High/Low	PLAYER_BI	Yes	2E+08	#Speedste	LS	7		########	2022	RM	10	
192448	https://so	M. ter Ste	Marc-And	28	########	187	85	Germany	FC Barcel	Spain Prir	1	90	93	7E+07	260000	GK	Right	3	4	1	Medium/l	PLAYER_BI	Yes	1.5E+08		GK	1		########	2022	SUB	22	
203376	https://so	V. van Dij	Virgil van	28	########	193	92	Netherlan	Liverpool	English Pr	1	90	91	7.6E+07	210000	CB	Right	3	3	2	Medium/l	PLAYER_BI	Yes	1.5E+08	#Tackling	LCB	4		########	2023	LCB	4	
212831	https://so	Alisson	Alisson Ra	27	########	191	91	Brazil	Liverpool	English Pr	1	90	91	6.3E+07	160000	GK	Right	3	3	1	Medium/l	PLAYER_BI	Yes	1.2E+08		GK	1		########	2024			
208722	https://so	S. ManÃ©	Sadio Ma	28	########	175	69	Senegal	Liverpool	English Pr	1	90	90	7.8E+07	250000	LW	Right	3	4	4	High/Med	PLAYER_BI	Yes	1.4E+08	#Speedste	LW	10		########	2023			
209331	https://so	M. Salah	Mohamec	28	########	175	71	Egypt	Liverpool	English Pr	1	90	90	7.8E+07	250000	RW	Left	3	3	4	High/Med	Mohamec	Yes	1.4E+08	#Speedste	RW	11		########	2023	RW	10	
192119	https://so	T. Courtoi	Thibaut C	28	########	199	96	Belgium	Real Mad	Spain Prir	1	89	90	5.6E+07	250000	GK	Left	4	3	1	Medium/l	Courtois	Yes	1.2E+08		GK	1		########	2024	GK	1	
153079	https://so	S. AgÃ¼er	Sergio Lec	32	########	173	70	Argentina	Manchest	English Pr	1	89	89	5.3E+07	300000	ST	Right	4	4	4	High/Med	PLAYER_BI	Yes	9.8E+07	#Dribbler	SUB	10		########	2021	ST	9	
155862	https://so	Sergio Ra	Sergio Ra	34	########	184	82	Spain	Real Mad	Spain Prir	1	89	89	2.5E+07	300000	CB	Right	4	3	3	High/Med	PLAYER_BI	Yes	5E+07	#Aerial Th	LCB	4		########	2021	RCB	15	
165153	https://so	K. Benzem	Karim Ber	32	########	185	81	France	Real Mad	Spain Prir	1	89	89	5.3E+07	350000	CF, ST	Right	4	4	4	Medium/l	Normal	Yes	1.1E+08	#Poacher	CF	9		########	2022			
167495	https://so	M. Neuer	Manuel N	34	########	193	92	Germany	FC Bayern	German 1	1	89	89	2.9E+07	125000	GK	Right	5	4	1	Medium/l	PLAYER_BI	Yes	4.8E+07		GK	1		########	2023	GK	1	
200145	https://so	Casemiro	Carlos He	28	########	185	84	Brazil	Real Mad	Spain Prir	1	89	89	6E+07	310000	CDM	Right	3	3	2	High/High	PLAYER_BI	Yes	1.2E+08	#Engine, #	CDM	14		########	2023			
210257	https://so	Ederson	Ederson S	26	########	188	86	Brazil	Manchest	English Pr	1	88	91	5.4E+07	195000	GK	Left	2	3	1	Medium/l	PLAYER_BI	Yes	1E+08		GK	31		########	2024			
202652	https://so	R. Sterling	Raheem S	25	########	170	69	England	Manchest	English Pr	1	88	90	7.3E+07	270000	LW, RW	Right	3	3	4	High/Med	PLAYER_BI	Yes	1.4E+08	#Speedste	LW	7		########	2023	LW	7	
212622	https://so	J. Kimmicl	Joshua Ki	25	########	176	73	Germany	FC Bayern	German 1	1	88	90	6.5E+07	145000	CDM, RB	Right	3	4	3	High/Med	Lean	Yes	1.1E+08	#Crosser	RDM	6		########	2023	RCM	6	
202126	https://so	H. Kane	Harry Kar	26	########	188	89	England	Tottenhar	English Pr	1	88	89	7.1E+07	220000	ST	Right	3	4	3	High/High	PLAYER_BI	Yes	1.4E+08	#Distance	ST	10		########	2024	ST	9	
211110	https://so	P. Dybala	Paulo Bru	26	########	177	75	Argentina	Juventus	Italian Se	1	88	89	7.1E+07	190000	CF, CAM	Left	3	3	4	Medium/l	PLAYER_BI	Yes	1.2E+08	#Dribbler	RS	10		########	2022	LW	21	
162835	https://so	S. Handar	Samir Har	35	########	193	92	Slovenia	Inter	Italian Se	1	88	88	1.6E+07	100000	GK	Right	3	3	1	Medium/l	PLAYER_BI	Yes	2.6E+07		GK	1		########	2021			
182521	https://so	T. Kroos	Toni Kroo	30	########	183	76	Germany	Real Mad	Spain Prir	1	88	88	5.5E+07	310000	CM	Right	4	5	3	Medium/l	Normal	Yes	1.1E+08	#Playmak	LCM	8		########	2023	LCM	8	
183277	https://so	E. Hazard	Eden Haza	29	########	175	74	Belgium	Real Mad	Spain Prir	1	88	88	5.8E+07	350000	LW, ST	Right	4	4	4	High/Med	PLAYER_BI	Yes	1.2E+08	#Dribbler	LW	7		########	2024	LF	10	
201024	https://so	K. Koulibe	Kalidou K	29	########	187	89	Senegal	Napoli	Italian Se	1	88	88	5E+07	140000	CB	Right	3	3	2	Medium/l	PLAYER_BI	Yes	8.5E+07	#Tackling	LCB	26		########	2023			
215914	https://so	N. KantÃ©	N'Golo Ka	29	########	168	70	France	Chelsea	English Pr	1	88	88	5.1E+07	190000	CDM, CM	Right	3	3	2	Medium/l	PLAYER_BI	Yes	9.7E+07	#Tackling	RDM	7		########	2023	LCM	13	
233049	https://so	J. Sancho	Jadon San	20	########	180	76	England	Borussia	German 1	1	87	93	7E+07	82000	RM, LM, C	Right	3	3	5	High/Med	Normal	Yes	1.3E+08	#Dribbler	RS	7		########	2023	RW	11	
231281	https://so	T. Alexanc	Trent Alex	21	########	180	69	England	Liverpool	English Pr	1	87	92	6E+07	110000	RB	Right	2	4	3	High/Med	PLAYER_BI	Yes	1.2E+08	#Crosser	RB	66		########	2024	RB	2	
212198	https://so	Bruno Fer	Bruno Mi	25	########	179	69	Portugal	Manchest	English Pr	1	87	90	6.3E+07	195000	CAM, CM	Right	2	4	4	High/High	Lean	Yes	1.2E+08	#Playmak	CAM	18		########	2025	RCM	16	
212218	https://so	A. Laporte	Aymeric L	26	########	189	85	France	Manchest	English Pr	1	87	90	5.6E+07	200000	CB	Left	2	3	2	Medium/l	PLAYER_BI	Yes	1.1E+08	#Tackling	LCB	14		########	2024			
209499	https://so	Fabinho	FÃ¡bio He	26	########	188	78	Brazil	Liverpool	English Pr	1	87	89	5.5E+07	155000	CDM	Right	3	2	3	Medium/l	PLAYER_BI	Yes	1E+08	#Tackling	CDM	3		########	2023			
216267	https://so	A. Roberts	Andrew R	26	########	178	64	Scotland	Liverpool	English Pr	1	87	89	5.3E+07	155000	LB	Left	3	2	3	High/High	PLAYER_BI	Yes	1E+08	#Engine, #	LB	26		########	2024	LB	3	
218667	https://so	Bernardo	Bernardo	25	########	173	64	Portugal	Manchest	English Pr	1	87	88	6E+07	230000	RW, CAM,	Left	2	3	4	High/High	Lean	Yes	1.2E+08	#Dribbler	LCM	20		########	2025	CAM	10	
138956	https://so	G. Chiellir	Giorgio Cl	35	########	187	85	Italy	Juventus	Italian Se	1	87	87	1.6E+07	94000	CB	Left	4	3	2	Medium/l	Normal	Yes	2.6E+07	#Tackling	LCB	3		########	2021	LCB	16	
167948	https://so	H. Lloris	Hugo Llor	33	########	188	82	France	Tottenhar	English Pr	1	87	87	2.7E+07	125000	GK	Left	4	1	1	Medium/l	PLAYER_BI	Yes	5.1E+07		GK	1		########	2022	GK	1	
176580	https://so	L. SuÃ¡rez	Luis Alber	33	########	182	86	Uruguay	FC Barcel	Spain Prir	1	87	87	3.2E+07	280000	ST	Right	5	4	3	High/Med	PLAYER_BI	Yes	6.5E+07	#Distance	SUB	9		########	2021			
177003	https://so	L. ModriÄ	Luka Mod	34	########	172	66	Croatia	Real Mad	Spain Prir	1	87	87	2.5E+07	260000	CM	Right	4	4	4	High/Med	Lean	Yes	5E+07	#Playmak	RCM	10		########	2021			
183898	https://so	A. Di Mar	Ángel Fab	32	########	180	69	Argentina	Paris Sair	French Lig	1	87	87	3.9E+07	160000	RW, LW	Left	4	2	5	High/Med	PLAYER_BI	Yes	7.2E+07	#Dribbler	RW	11		########	2021			
186153	https://so	W. SzczÃ™	Wojciech	30	########	195	90	Poland	Juventus	Italian Se	1	87	87	3.4E+07	105000	GK	Right	3	3	1	Medium/l	Lean	Yes	5.6E+07		GK	1		########	2024	GK	1	
188567	https://so	P. Aubam	Pierre-Em	31	########	187	80	Gabon	Arsenal	English Pr	1	87	87	4.5E+07	170000	ST, LW	Right	3	4	4	High/High	PLAYER_BI	Yes	8.6E+07	#Speedste	LW	14		########	2021			
189511	https://so	Sergio Bu	Sergio Bu	31	########	189	76	Spain	FC Barcel	Spain Prir	1	87	87	3.8E+07	240000	CDM	Right	4	3	3	Medium/l	Lean	Yes	7.8E+07		SUB	5		########	2023	SUB	5	

Abstract

- Player rating prediction system is completely a data-driven model. To train the model we are using dataset from Kaggle which consist of about 106 columns and is obtained from EA sports player ratings. When trained, the model will be applied to another dataset which contains players that are yet to play at professional level thus providing us with their overall rating, their potential and also the position they will play in.
- Firstly various data pre-processing techniques will be used to make data fit for the regression and classification algorithms. Data Reduction techniques need to be used to select a subset of attributes from the total attribute while data cleaning techniques to fill up the missing values and data transformation techniques to normalize and encode the data values.
- After the data is processed then the features that are suitable for overall rating, potential rating and position prediction will be selected. The data so obtained will be trained using algorithms such as multiple linear regression, polynomial linear regression, random forest etc. and for position we need to use classification algorithms like KNN, decision tree, naïve bayes etc.
- To calculate the accuracies of the various algorithms being used we will make use of R-square (regression) and confusion matrix (classification) such that appropriate model can be selected.
- Data visualisation of the above dataset to provide insights and thus helping in analysis of player.

Implementation

Preprocessing and Testing - https://colab.research.google.com/drive/1zMwHRC_W7RkGD5PGcr46IaG8N0_uSQIZ

Test Case -

https://colab.research.google.com/drive/1TiTJBfSlQ0xHHvokjGCrHUc_BO2adW2Z

Observations

Algorithm Used	Accuracy Score
K Neighbours	0.470726
SVC	0.495726
Naïve Bayes	0.347222
Decision Tree	0.357265
Random Forest	0.477778

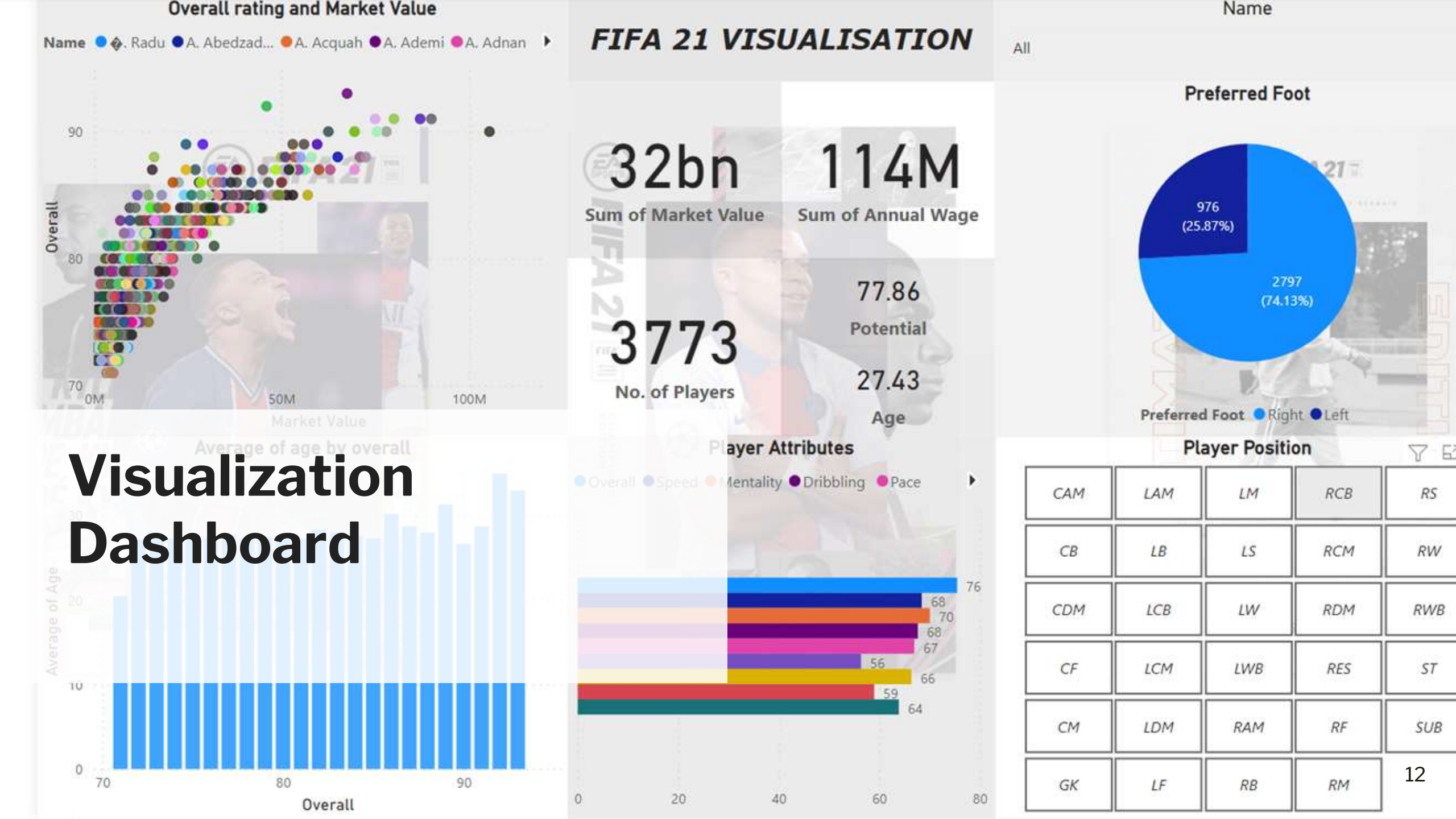
^Dependent Variable – 'player_positions' (classification)

Algorithm Used	Accuracy Score
Multiple Regression	0.90
Polynomial Regression	0.973
SVR	0.971
Decision Tree	0.912
Random Forest	0.966

^Dependent Variable – 'Overall' (regression)

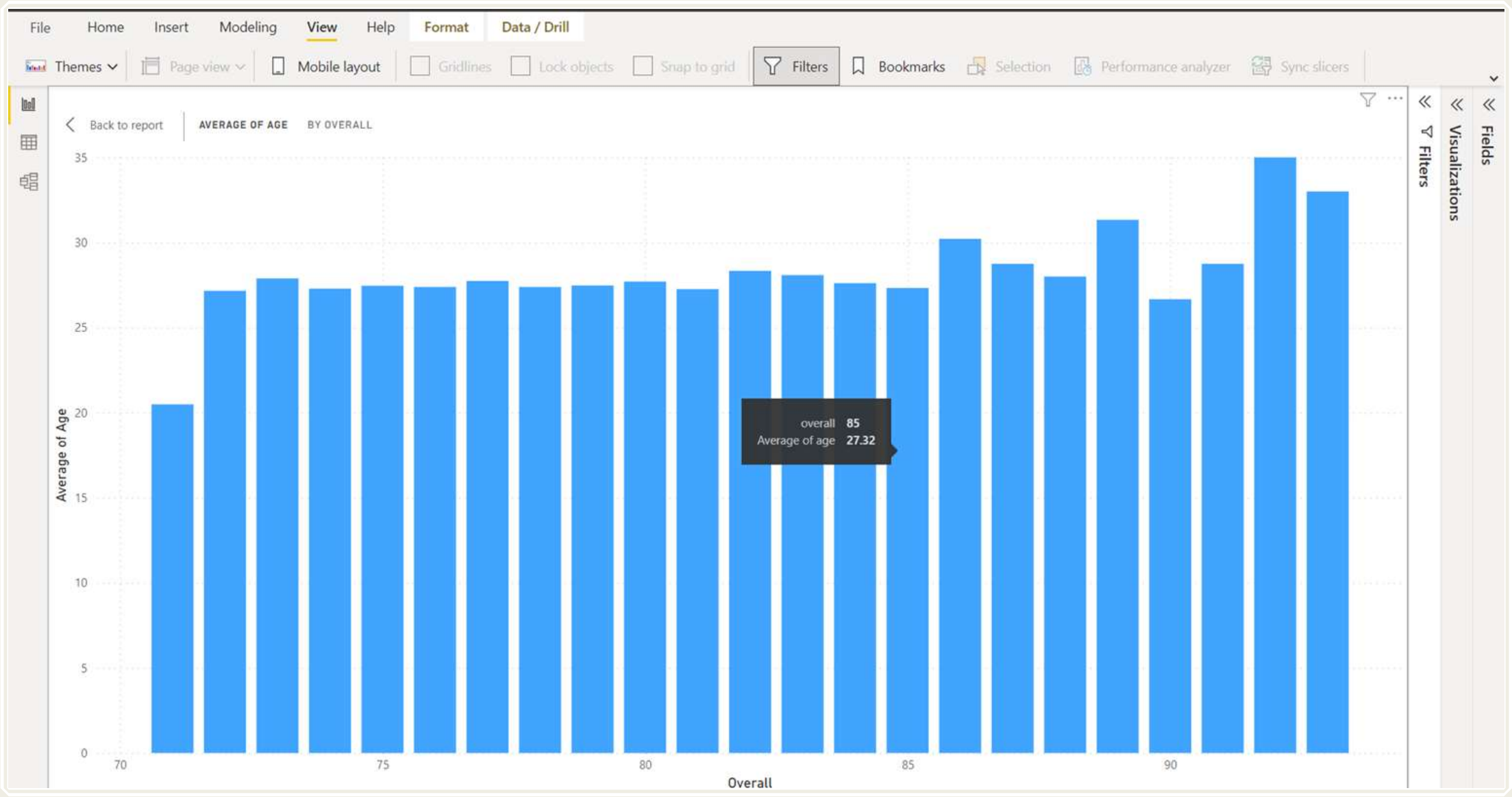
Algorithm Used	Accuracy Score
Multiple Regression	0.896
Polynomial Regression	0.973
SVR	0.971
Decision Tree	0.912
Random Forest	0.966

^Dependent Variable – 'Potential' (regression)

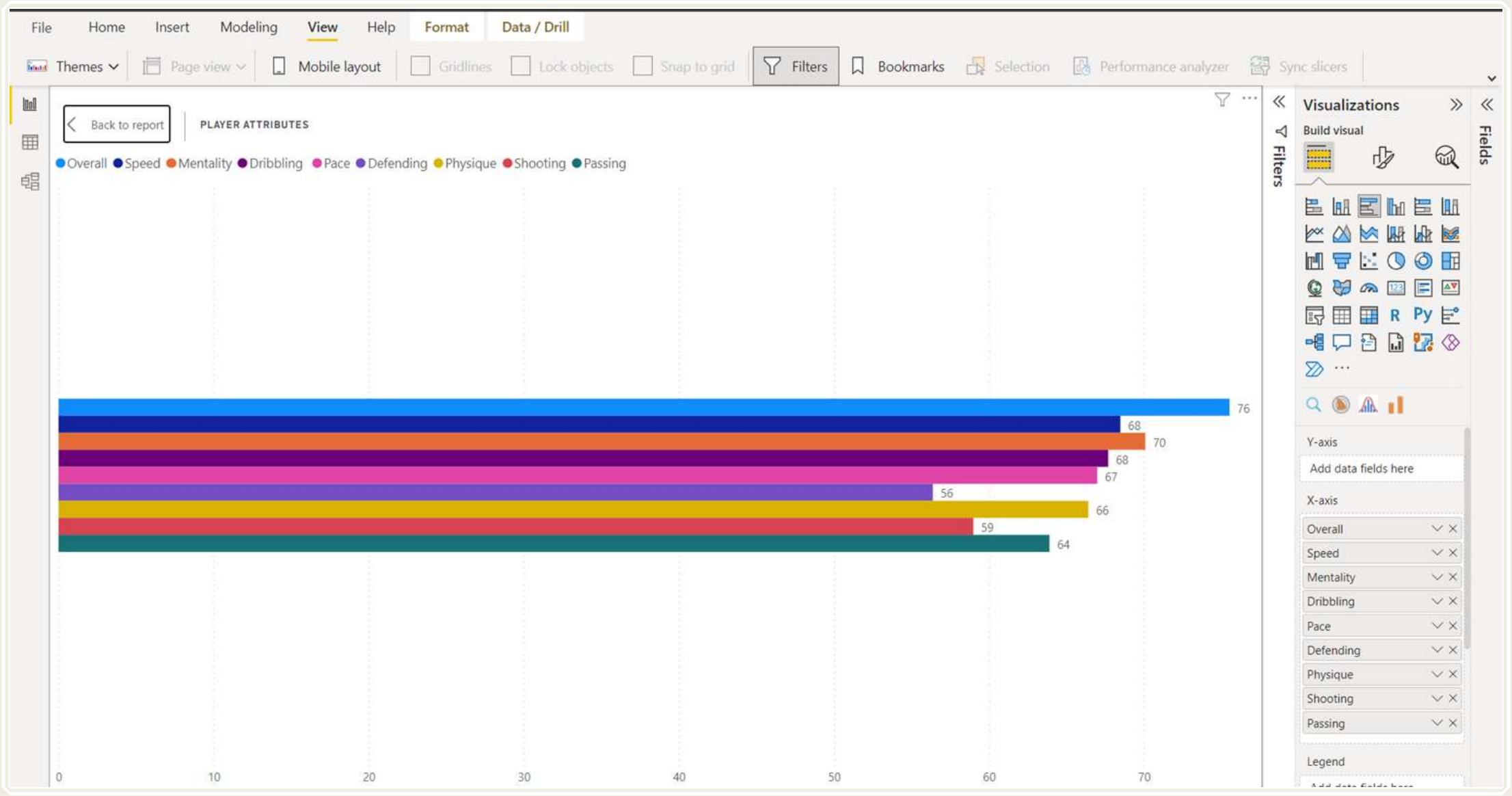


Visualization Dashboard

Player Overall rating vs Market Value

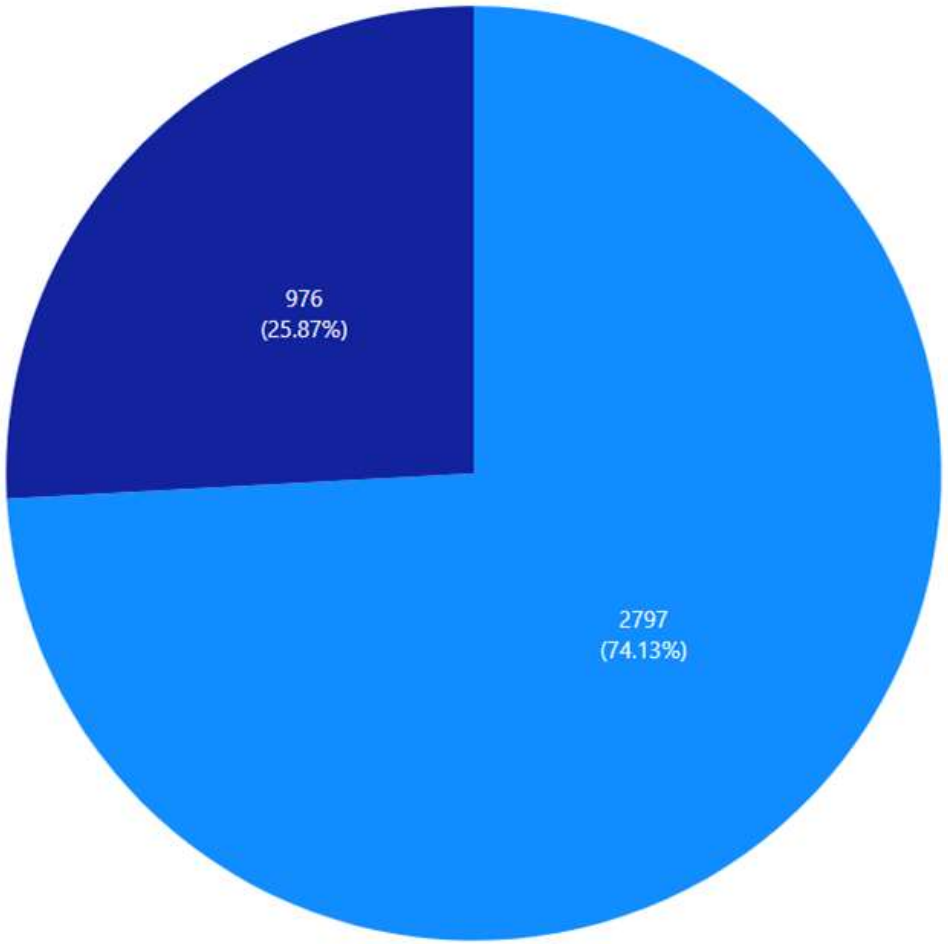


Average Age vs Overall Rating



Players Attributes

Pie chart visualization of distribution of preferred foot among Football Players



Novelty

- The researches that uses the EA rating dataset is mostly used in video game engine but with our model the same can be applied to real world challenges by giving various scouts and sport related communities an opportunity to choose wisely among a huge array of non professional players available.
- This model can be used to find how much potential does a young player carries and which roles in a team he can fulfil. The model once trained will be applied to the dataset consisting of non professional players and by virtue of which various features of the player can be calculated like the position or role of the player in a real or a fantasy team.
- Another feature of the model is its flexibility as instead of football we can train it on other dataset that belong to other sport.

References

- https://arxiv.org/pdf/1802.04987.pdf
- https://rstudio-pubs-static.s3.amazonaws.com/524786_4d3da4afc6bd406c8f720572b460cc3e.html
- https://towardsdatascience.com/predicting-premier-league-odds-from-ea-player-bfdb52597392
- https://www.kaggle.com/
- https://medium.com/analytics-vidhya/are-footballers-playing-in-their-best-positons-using-football-manager-dataset-a5346605541