RARITA FSA LEAGUE REPORT 2022

By ML Group

Lan Yu, Haoyang Wang, Xuan Zhang, Mengtong Bi, Hengzhan Zheng

From University of New South Wales

CONTENTS

1. OBJECTIVES

This report aims to help Rarita form a competitive national football team, which is expected to boost the country's economy. We select 21 appropriate players with relatively high competitiveness as the basis for further analysis of the team operation and future economic impacts. We also disclose assumptions embedded in model construction, present a cost-benefit analysis and discuss relevant risks and their corresponding mitigation methods. Combining with data limitations, those are to ensure more comprehensive understandings for the committee on the uncertainty of the analysis provided in the report. We include technical references and details of model constructions in Appendix.

2. TEAM SELECTION AND IMPLEMENTATION PLAN

2.1 CRITERIA

We used correlation matrix and machine learning models to objectively filter useful performance features of players based on tournament ranking data. We calculated a weighted average rank for each individual player using the performance features selected by their importance in the constructed models. More details of discussions on methodologies, steps, and results of linear, shrinkage and ensemble model construction are presented in Appendix 7.1.1.

We defined the competitiveness and potentials of players by the overall rank and the selected performance features summarized in Table 1. Importantly, players in shooting positions are valued on their overall abilities to make shots on target and transferring those shots to goals. Players in passing positions are valued on their overall ability to pass the ball and complete crosses to assist in shooting. Similarly, players in defense positions are valued on their capability to make tackles and dribbling to effectively block the ball. Goalkeepers are valued on their ability to save the balls shot in. Those are the overarching abilities vital for players in each. The weights of each selected performance features are outlined in Appendix 7.1.5.

Position	Performance Features					
Shooting	Minutes play divided by 90	Goals	Shots on target	Average distance from goal of all shots taken	Penalty kicks attempted	Expected goals
Passing	Minutes play divided by 90	Passes completed	Passes completed (15-30 yards)	Completed crosses into the 18-yard box		
Defense	Minutes play divided by 90	Tackles in defensive 1/3	Number of times dribbled past plus number of tackles	Number of times dribbled past plus number of tackles	Interceptions	
Goalkeeping	Minutes play divided by 90	Shot on target against	Win	Lose	Penalty kick missed	

Table 1 The selected player performance features by position

From Figure 1 and Figures 6-8 in Appendix 7.1.1, correlation matrixes show other performance features, that are not selected, are also included in player valuation and selection to some extents.

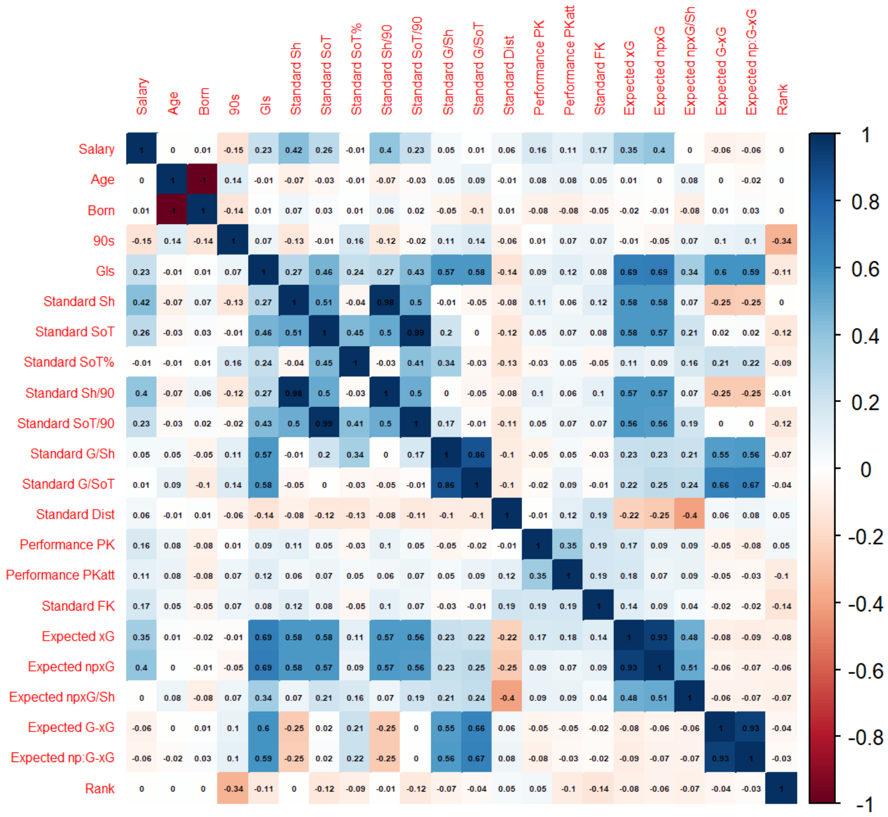


Figure 1 Correlation matrix for shooting position in 2021 tournament data

2.2 SUCCESS PROBABILITY

Implementing Lasso to rate the players by shooting, passing, defense and goalkeeping, we then formulated rating for countries participated in 2020 and 2021's tournament, see Table 2 below.

Country	2020 Shooting	2020 Passing	2020 Defense	2020 GK	2020 Rank	2021 Shooting	2021 Passing	2021 Defense	2021 GK	2021 Rank
Dosqaly	497.6909447	NA	NA	0	1	352.7391181	368.7065116	339.2875229	0	9
Nganion	602.9305957	NA	NA	18.888865	2	343.5117469	364.4834319	335.187419	5.002325	3
Sobianitedrucy	545.6341944	NA	NA	3.1946654	3	418.9245177	443.6850534	413.1619734	16.74678	1
Southern Ristan	425.9418704	NA	NA	35.120696	4	438.0590393	466.0807912	432.7364057	8.106167	6
Greri Landmoslands	458.7482082	NA	NA	22.583477	5	313.7789297	332.4690186	300.6298718	12.95198	11
Byasier Pujan	413.5738402	NA	NA	22.66867	6	377.0644905	399.9741279	364.4606279	9.802903	15
Mico	406.2811211	NA	NA	28.504033	7	346.5528926	374.6103966	339.8820948	8.336276	4
People's Land of Maneau	499.645499	NA	NA	8.0967486	8	327.3537824	352.0169003	323.0488368	3.655158	2
Esia	441.6195018	NA	NA	51.208352	9	328.9346742	334.254998	314.0220832	11.33364	14
Nkasland Cronestan	418.4391839	NA	NA	28.386417	10	344.2313123	372.4114677	343.313735	17.16669	22
Quewenia	436.4847453	NA	NA	12.719965	11	318.9494623	350.0490739	318.6668538	10.89765	5
Manlisgamncent	482.673003	NA	NA	12.70662	12	352.028291	378.1666523	350.8445894	11.8179	13
Xikong	443.4501447	NA	NA	12.978442	13	323.5404157	363.8442823	319.5829647	10.9761	12
Bernepamar	466.0538313	NA	NA	37.66031	14	346.5981681	365.5242285	335.3957843	10.97127	8
Unicorporated Tiagascar	611.7893164	NA	NA	54.57402	15	NA	NA	NA	NA	
Cuandbo	526.6910053	NA	NA	59.222303	16	NA	NA	NA	NA	
Galamily	774.0424344	NA	NA	45.858055		358.8697115	387.4702346	343.7859986	9.759904	7
Giumle Lizeibon	395.3225141	NA	NA	22.340215		377.8991037	398.0344018	358.5324658	10.65672	10
Djipines	470.4373743	NA	NA	33.569843		306.2395699	330.8195404	309.9440756	15.03445	16
Leoneku Guidisia	454.0665117	NA	NA	10.571307		357.9588129	384.4503233	358.3378698	15.94862	17
Ledian	594.3507672	NA	NA	50.182247		362.4345456	361.8233316	356.9842438	16.89463	18
Eastern Sleboube	494.5653531	NA	NA	55.804125		361.0312433	380.72917	355.4607929	34.9586	19
New Uwi	409.4448828	NA	NA	25.503422		310.6862737	338.1483588	315.7564057	15.99429	20
Ngoque Blicri	468.0636084	NA	NA	28.897346		388.5754263	407.1035275	381.5906411	16.07632	21
Eastern Niasland	457.5867988	NA	NA	25.990522		363.4272856	382.5180061	362.2149803	19.54734	23
Varijitri Isles	538.7145351	NA	NA	29.913856		327.1959021	349.9411894	324.1571647	22.09936	24

Table 2 Rating Table for Participants in 2020 and 2021 Tournament. Top 10 in both 2020 and 2021 are in yellow; Ranking Top 10 either in 2020 or 2021 are in green; Never ranked in Top 10 are in Red.

In order to select a competitive team, we must refer to the corresponding indicators of the yellow teams from Table 2. Applied linear regression (formula attached below) on ranks, we found out that Goalkeeping played a significant role when matching.

$$Rank = 23.37557 - 0.04489 * Shooting - 0.21452 * Passing + 0.22477 * Defense + 0.57595 * Goalkeeping$$

Based on the criteria of being competitive, we constructed the optimization to select team members. Table 3 shows the optimal team selection. By neural network, this lineup had probability of success being competitive within range between 0.105 to 0.124.

2.3 10-YEAR STRATEGY

2.3.1 TEAM SELECTION

We selected players based on selection criteria in Section 2.1 and expected to implement a 4-2-4 flexible strategy consisting of 4 forwards, 2 midfields, 4 defense and 1 goalkeeper shown in Table 3. Table 3. To implement the flexible strategy successfully, at least one forward position and defense position can also play midfield, so the 4-2-4 team structure can easily convert to 3-4-3 or other competitive structures dependent on opponent's strategy. Assuming average retirement age for all football players is around 35 years old, players with current age above 25 are subject to replacement.

No	Player	Nation	Pos	Squad	Age	Minutes play divided by 90
1	L. Ndyanabo	Imaar Vircoand	FWMF	Festive Governors	33	33.54
2	Y. Manjate	Byasier Pujan	FW	Unaccountable Foxes	26	32.55
3	L. De Wit	Greri Landmoslands	FW	Fighting Wave	35	31.19
4	R. Nkosi	Sobianitedrucy	FW	Fighting Clippers	29	31.86
5	K. Chisi	Imaar Vircoand	MF	Mean Wolves	26	35.54
6	F. Lee	Sobianitedrucy	MF	Marvelous Patriots	22	31.91
7	J. Okullo	Esia	DF	Weak Chargers	23	20
8	Y. Twinomugisha	Janmico	DF	Big Foxes	28	28.92
9	C. Kawooya	Republic of Denand L	MFDF	Solemn Cougars	23	24.13
10	C. Amoding	Iverde	DF	Supreme Janes	20	0.08
11	T. Kamugisha	Lefghau	GK	Marvelous Coyotes	26	35.92

Table 3 Rarita National Team Member List

We selected substitute players using the same weighted average ranking with an additional criterion of age shown in Table 4.

No	Player	Nation	Pos	Squad	Age	Minutes play divided by 90
1	B. Male	Nganion	FWMF	Horrible Bison	23	26.73
2	V. Golob	Sobianitedrucy	FWMF	Marvelous Patriots	25	27.45
3	H. Nsamba	Byasier Pujan	FW	Horrible Storm	25	17.81
4	H. Robert	Dosqaly	FW	Flawless Cows	21	26.42
5	O. Wanjala	Rarita	MF	Black Coyotes	23	35.72
6	D. Sigauke	Sobianitedrucy	MF	Sugar Bengals	23	32.12
7	J. Okullo	Esia	DF	Weak Chargers	23	20
8	C. Kawooya	Republic of Denand L	MFDF	Solemn Cougars	23	24.13
9	K. Shibata	Rarita	DF	Black Coyotes	29	0.41
10	I. Kumari	Dosqaly	GK	Somber Stallions	25	37.93

Table 4 Substitute Team Member List

2.3.2 SOURCES OF REVENUE AND ADDITIONAL FUNDING

The sources of revenue for Rarita's national football team can be classified in three categories, namely Matchday, Broadcast and Commercial. The detailed descriptions of the three categories are displayed in Appendix 7.2.4.

All three sources of revenue significantly rely on brand development to attract new supporters and improve fans base. The growth of revenue can be achieved by both improving rankings in competitions and setting up effective and practical commercialization strategies that are outlined in Section 2.3.3.

Assuming all strategies are successfully practiced, the relevant profit and loss can be estimate with present value of 21335.49 million as shown in Table 5. The model assumptions are discussed in Section 4.2 and Appendix 7.2.2 and methodology is outlined in Appendix 7.2.1. From the projection, except the initial funding of 995 million doubloons, non-governmental funding is needed at the end of 2023. The funding source of 350 million doubloons can be achieved by multiple long-term extensive commercial sponsorship or transferring team ownership to large corporates or constructing global membership, similar to Manchester United and FC Barcelona (Krabbenbos, 2013). Other sources of non-government funding are listed in Appendix 7.2.3.

Timeline in year	0	1	2	3	4	5	6	7	8	9	10
Profit & Loss (in millions)/Year	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031
Total Expense		**ð 679.02**	**ð 1,033.79**	**ð 1,485.58**	**ð 1,971.07**	**ð 2,746.24**	**ð 3,921.26**	**ð 4,997.24**	**ð 5,504.39**	**ð 6,171.88**	**ð 7,027.13**
Staff cost		ð 563.28	ð 821.47	ð 1,148.74	ð 987.25	ð 1,438.59	ð 2,862.85	ð 3,645.83	ð 3,078.01	ð 3,459.89	ð 3,940.02
Other expense		ð 115.74	ð 212.33	ð 336.84	ð 983.81	ð 1,307.66	ð 1,058.41	ð 1,351.41	ð 2,426.38	ð 2,711.99	ð 3,087.11
Total Revenue		**ð 165.08**	**ð 830.68**	**ð 1,697.25**	**ð 2,933.50**	**ð 4,299.88**	**ð 5,920.87**	**ð 8,448.31**	**ð 9,595.03**	**ð 10,692.46**	**ð 11,735.80**
Matchday		ð 64.21	ð 163.67	ð 288.53	ð 464.85	ð 645.92	ð 893.06	ð 1,165.22	ð 1,322.33	ð 1,476.87	ð 1,628.23
Broadcast		ð 246.79	ð 534.40	ð 902.07	ð 1,410.30	ð 1,975.15	ð 2,595.32	ð 3,379.33	ð 3,065.96	ð 3,424.87	ð 3,777.17
Commercial		-ð 145.93	ð 132.61	ð 506.64	ð 1,058.34	ð 1,678.82	ð 2,432.49	ð 3,903.77	ð 5,206.74	ð 5,790.71	ð 6,330.41
Other Revenue		ð 4.95	ð 47.38	ð 106.71	ð 189.54	ð 299.13	ð 404.60	ð 524.44	ð 596.67	ð 661.62	ð 718.93
Overall profit		**-ð 508.99**	**-ð 155.73**	**ð 318.38**	**ð 1,151.97**	**ð 1,852.77**	**ð 2,404.21**	**ð 3,975.52**	**ð 4,687.31**	**ð 5,182.19**	**ð 5,427.61**
Funding	**ð 995.00**		**ð 350.00**								
Ending balance	ð 995.00	ð 487.24	ð 682.66	ð 1,002.31	ð 2,154.39	ð 4,014.50	ð 6,451.67	ð 10,495.23	ð 15,307.36	ð 20,708.34	ð 26,470.57
Cost coverage	ð 879.26	ð 274.91	ð 345.82	ð 18.50	ð 846.74	ð 2,956.09	ð 5,100.26	ð 8,068.85	ð 12,595.37	ð 17,621.23	
PV of Profit		-ð 508.44	-ð 155.55	ð 313.65	ð 1,118.37	ð 1,763.64	ð 2,238.42	ð 3,606.89	ð 4,112.40	ð 4,397.82	ð 4,448.29
Cumulative PV		-ð 508.44	-ð 664.00	-ð 350.34	ð 768.03	ð 2,531.66	ð 4,770.09	ð 8,376.98	ð 12,489.38	ð 16,887.20	ð 21,335.49

Table 5 Direct 10-Year Profit & Loss for building the national football team

2.3.3 KEY STRATEGIES

For the pursuit of world-wide competitiveness, the focus of Rarita's national football team should be on strategies driving management, brand development, growing commercial revenue and other revenues, improving overall popularity and positions in tournaments (INTERNAL ANALYSIS OF CHELSEA FOOTBALL CLUB, 2022).

As long-term competitive success is built on strategic consistency across all levels of operations, the operation group should be multi-level to establish a durable chain of command (PANNES, 2020). An efficient management ensures smooth operation and goal achievements, which is described in Appendix 7.2.5.

The brand will be developed and managed consistently through sporting success in various tournaments and within league competitions, global branding and global fan base attraction. Commercial innovation should be treated as the core to trigger revenue growth and are encouraged throughout the organization. Developing a football-based computer or mobile game is a great innovation to boost revenue and attract boarder range of audience.

To maximize revenues, expansion of stadium and invention of creative activities are also very important. Launching a new streaming platform for a subscription fee can increase fan engagement to increase matchday and broadcast revenues.

The long-term revenue is composed from 50% commercial, 35% broadcast and 15% matchday. This composition is part of commercialization strategy. Supporting data is derived from 2021 Deloitte's Football Money League report shown in Figure 2.

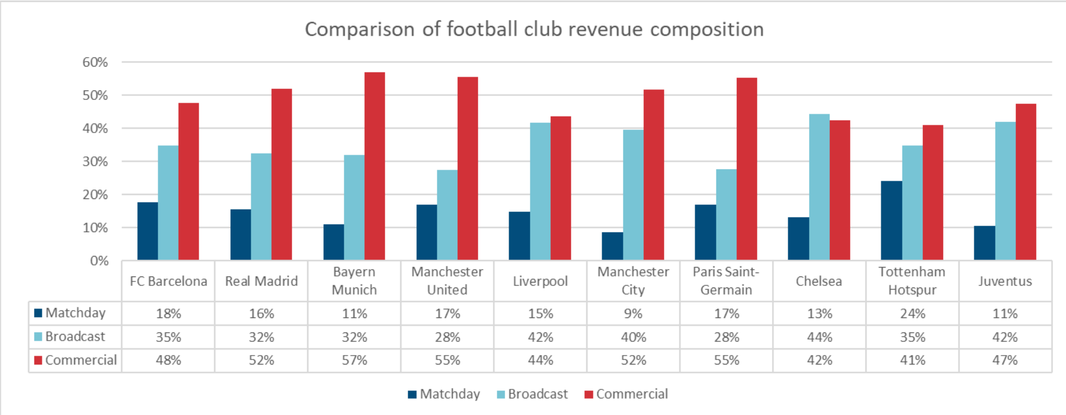


Figure 2 2020 Revenue composition of top 10 Deloitte Football Money League

2.4 KEY MONITORING METRICS

The strategy above is monitored by the three metrics below, other useful metrics are outlined in Appendix 7.2.4.

- **Capital injections** are investments into a company owns a football team. This metric reflects profitability status of football teams and should be reported annually.
- **League Average** are usually based on popularity, TV views, the presence of top players, social media buzz generated from league games, and the success of clubs from these leagues in various continental tournaments. Teams with high league averages would gain attention from fans and the media, leading to higher profitability. League average is reported on an annual basis. Standard deviation of league average monitors stability of performance.
- **Operating Surplus** is the difference between revenue and expenditure. A higher surplus means higher profitability. Operating Surplus is reported annually.

3. ECONOMIC IMPACTS

3.1 DIRECT IMPACTS

After projecting the future cash flow, with assumed inflation rate of 3%, we calculated the IRR of Rarita's football league to be 7%. We believe building a national football team provides acceptable level of return. However, the NPV only turns positive in year 9/10, indicating high volatility of the project. There are significant negative cash flows during the first few years of team establishment as shown in Table 6.

	Cash Flow	NPV Til Year n	
2021	-995000000	-995000000	Inflation rate
2022	-1215683932	-2175275662	3%
2023	-692107137	-2827653018	IRR
2024	-792811870	-3553188189	7%
2025	-199009264	-3730005342	
2026	133659438.6	-3614709536	
2027	59126368.2	-3565192134	
2028	1193062688	-2595122990	
2029	1492994770	-1416539131	
2030	1657342564	-146324059	
2031	1656028020	1085916314	

Table 6 NPV Calculation based on the cash flow model

3.1.1 SHORT-TERM

An outstanding football team performance has positive impact in tourism, retailing, accommodation and employment sectors for the nation. For example, according to Dubai Sports Council's report, Croatia winning second place in the world cup experienced a 250% increase in visitors on the day of the final match. Similarly, in France, after the 2018 world cup, there was a 40% increase in sales of TV and 20% recovery rate in French restaurants (Jr, 2019). Even though Rarita national football team may generate negative returns in short term, the team's outstanding performance could potentially boost several other sectors for the nation, thus boost Rarita's overall GDP indirectly.

3.1.2 LONG-TERM

Establishing a football team with good ranking can convert citizens into football fans. The feel-good effect on the citizens could increase their sense of pride and happiness resulting in higher willingness to consume. The average willing-to-pay of citizens was increased from 4.26 dollar per person to 10 dollar per person after the World Cup (Liu, 2013). In long term, with increasing competitiveness of national football team, the improved international perception can boost Rarita's international trade and investment and potentially bring consistent future increase in tourism revenue.

3.2 INTANGIBLE EFFECTS

Football is more than a sport. The Social Return On Investment (SROI) model considers some positive social and economic impacts so that it helps individual associations make an evidence-based case for increased government investment in football (Campelli, 2022). The main contributions from intangible effect are health and reduced crime rates, which are elaborated in Appendix 7.3.1.

3.3 REGIONAL IMPACTS

Based on analysis of economic dataset in Appendix 7.3, we assumed East Rarita has better economic condition and an older population comparing to West Rarita. Tables 7-10 show the correlations between the profit generated from football teams and the Rarita's economic indices by provinces. Conclusively, football performance has the highest impacts on East Rarita.

	GDP	Income	Population	Population Density	Household spending	Household saving	Profit
GDP	1						
Income	0.99368	1					
Population	0.89031	0.8892039	1				
Population Density	0.88538	0.8852357	0.99989684	1			
Household spending	0.92339	0.9183522	0.996469809	0.99531588	1		
Household saving	-0.4515	-0.514243	-0.58467288	-0.593878812	-0.555162932	1	
Profit	-0.1983	-0.222165	-0.53272798	-0.536749841	-0.479090345	0.160313427	1

Table 7 Correlation table of football activity profit and economic indices for Rarita

	GDP	Income	Population	Population Density	Household spending	Household saving	Profit
GDP	1						
Income	0.94144	1					
Population	0.8924	0.7673314	1				
Population Density	0.89315	0.7686835	0.999994419	1			
Household spending	0.85738	0.7295866	0.994172146	0.993864487	1		
Household saving	0.90414	0.7499243	0.965626862	0.966049504	0.935639727	1	
Profit	-0.2055	-0.092102	-0.54153354	-0.542075107	-0.532938469	-0.559035537	1

Table 8 Correlation table of football activity profit and economic indices for East Rarita

	GDP	Income	Population	Population Density	Household spending	Household saving	Profit
GDP	1						
Income	0.99599	1					
Population	0.923222	0.931358	1				
Population Density	0.921236	0.930063	0.9999655	1			
Household spending	0.905186	0.93418	0.845423	0.846450362	1		
Household saving	0.483115	0.535196	0.3540206	0.357232847	0.788022463	1	
Profit	-0.13627	-0.21032	-0.064141	-0.068832735	-0.432211256	-0.695198809	1

Table 9 Correlation table of football activity profit and economic indices for Central Rarita

	GDP	Income	Population	Population Density	Household spending	Household saving	Profit
GDP	1						
Income	0.99766	1					
Population	0.88228	0.8754324	1				
Population Density	0.88265	0.8756727	0.99999762	1			
Household spending	0.91522	0.907659	0.9902079	0.990282123	1		
Household saving	-0.7312	-0.717177	-0.76874875	-0.769336908	-0.709395897	1	
Profit	-0.2389	-0.265537	-0.59489634	-0.593375682	-0.520123913	0.3357413	1

Table 10 Correlation table of football activity profit and economic indices for West Rarita

In the next ten year, we propose a few stadiums will be built in East Rarita, meaning a huge amount of expenses. This can result in significant negative NPV during first few years but bring continuous matchday income. West Rarita may also share the benefits from East Rarita.

4. ASSUMPTIONS

4.1 TEAM CONSTRUCTION

- All player performance features are measured and recorded reasonably and accurately. The relative values of performance features not the values themselves show the competitiveness of players.
- The positions described in four letters are assumed to be the same. The reason for inconsistency in position abbreviation is salary change in the same year.

4.2 PROFIT-LOSS ANALYSIS

We separate the revenue and expense growth in three stages, namely next year, short-term and long-term as shown in Table 11. Other assumptions used and justification of major assumptions are outlined in Appendix 7.2.2.

Key assumptions	Values
With national team built	
2022 total expense growth rate	40%
2022 total revenue growth rate	15%
Short-term total expense growth rate	20%
Song-term total revenue growth rate	25%
Long-term total expense growth rate	11%
Long-term total revenue growth rate	10%
Without national team built	
Total expense growth rate	9%
Total revenue growth rate	8%
Total expense growth rate variance	1.38%
Total revenue growth rate variance	1.03%

Table 11 Key assumptions for Profit & Loss cash flow model

Assuming successfully implemented the 10-year strategies, the revenue composition gradually approaches the optimal composition outlined in Section 2.3.3. Table 12 displays the annual assumptions used for cash flow model construction.

Timeline in year	0	1	2	3	4	5	6	7	8	9	10
Year	**2021**	**2022**	**2023**	**2024**	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**	**2031**
Assumptions	**With national team built**										
Lending player proportion	35%	40%	45%	50%	55%	60%	60%	60%	60%	60%	60%
Total expense growth	10%	40%	21%	21%	24%	19%	10%	27%	6%	11%	14%
Total revenue growth	5%	15%	24%	24%	21%	26%	33%	19%	14%	10%	8%
Expense proportion											
Staff	65%	70%	70%	70%	60%	60%	70%	70%	60%	60%	60%
Other	35%	30%	30%	30%	40%	40%	30%	30%	40%	40%	40%
Revenue proportion											
Matchday	15.60%	18%	18%	17%	17%	16%	16%	15%	15%	15%	15%
Broadcast	39%	45%	45%	44%	44%	43%	42%	40%	35%	35%	35%
Commericial	45.40%	37%	38%	39%	40%	41%	42%	45%	50%	50%	50%
Other income	3.0%	3%	3.5%	4.0%	4.5%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%
	Without national team built										
Total expense growth	10%	7.0%	11.2%	11.3%	19.7%	7.0%	-15.4%	25.3%	1.3%	8.3%	13.6%
Total revenue growth	5%	9.7%	6.1%	6.0%	-1.2%	9.7%	29.0%	-6.0%	14.6%	8.6%	4.1%
Expense proportion											
Staff	65%	66%	66%	66%	66%	66%	66%	66%	66%	66%	66%
Other	35%	0.34	34%	34%	34%	34%	34%	34%	34%	34%	34%
Revenue proportion											
Matchday	16%	17%	17%	17%	17%	17%	17%	17%	17%	17%	17%
Broadcast	39%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%
Commericial	45%	43%	43%	43%	43%	43%	43%	43%	43%	43%	43%
Other income	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%

Table 12 Assumptions by year

The revenue and expense compositions for the scenario without national team built are stable and compatible with historical trend shown in Figure 3.

Figure 3 RFL historical revenue and expense composition

All the key assumptions in the cash flow model are subject to change due to external factors and sensitivity analysis is conducted in Section 5.1.

5. RISK AND MITIGATION

5.1 SENSITIVITY TEST OF KEY ASSUMPTIONS

From Figure 4, the NPV of direct profit of building the national football team is significantly subject to changes of the first-year, short-term and long-term total expense and revenue growth, because of the compound effects of money. Diminishing marginal effects are evident from increasing revenue growth rate and decreasing expense growth rate. Importantly, the operation and strategic risks can strongly deteriorate the short-term revenue growth rate while keeping expense growth rate high. To avoid the undesirable financial outcomes for Rarita, short-term growth rates in expense must be controlled while promoting revenue growth.

NPV Change in millions for 2022 total revenue and expense growth rate											
Revenue/ Expense	-40%	-30%	-20%	-10%	0%	10%	20%	30%	40%	50%	60%
-40%	9,722	5,238	755	-3,729	-8,213	-12,697	-17,181	-21,664	-26,148	-30,632	-35,116
-30%	18,356	13,872	9,388	4,904	420	-4,063	-8,547	-13,031	-17,515	-21,999	-26,482
-20%	26,989	22,505	18,021	13,538	9,054	4,570	86	-4,398	-8,881	-13,365	-17,849
-10%	35,622	31,139	26,655	22,171	17,687	13,203	8,720	4,236	-248	-4,732	-9,216
0%	44,256	39,772	35,288	30,804	26,321	21,837	17,353	12,869	8,385	3,902	-582
10%	52,889	48,405	43,922	39,438	34,954	30,470	25,986	21,503	17,019	12,535	8,051
20%	61,523	57,039	52,555	48,071	43,587	39,104	34,620	30,136	25,652	21,168	16,685
30%	70,156	65,672	61,188	56,705	52,221	47,737	43,253	38,769	34,286	29,802	25,318
40%	78,789	74,306	69,822	65,338	60,854	56,370	51,887	47,403	42,919	38,435	33,951
50%	87,423	82,939	78,455	73,971	69,488	65,004	60,520	56,036	51,552	47,069	42,585

NPV Change in millions for short-term total revenue and expense growth rate													
Revenue/ Expense	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%	25%	30%	35%	40%
-30%	-16,408	-19,054	-22,350	-26,437	-31,481	-37,672	-45,230	-54,404	-65,479	-78,778	-94,662	-113,537	-135,855
-25%	-14,241	-16,886	-20,182	-24,270	-29,314	-35,505	-43,062	-52,236	-63,312	-76,611	-92,495	-111,369	-133,687
-20%	-11,512	-14,158	-17,453	-21,541	-26,585	-32,776	-40,333	-49,507	-60,583	-73,882	-89,766	-108,640	-130,958
-15%	-8,086	-10,732	-14,028	-18,115	-23,159	-29,350	-36,908	-46,082	-57,157	-70,456	-86,340	-105,215	-127,533
-10%	-3,803	-6,449	-9,745	-13,832	-18,876	-25,067	-32,625	-41,799	-52,875	-66,173	-82,057	-100,932	-123,250
-5%	1,526	-1,120	-4,416	-8,504	-13,547	-19,738	-27,296	-36,470	-47,546	-60,845	-76,729	-95,603	-117,921
0%	8,120	5,474	2,178	-1,909	-6,953	-13,144	-20,701	-29,875	-40,951	-54,250	-70,134	-89,009	-111,326
5%	16,235	13,589	10,293	6,206	1,162	-5,029	-12,586	-21,761	-32,836	-46,135	-62,019	-80,894	-103,212
10%	26,163	23,517	20,221	16,134	11,090	4,899	-2,658	-11,832	-22,908	-36,207	-52,091	-70,966	-93,283
15%	38,239	35,593	32,297	28,210	23,166	16,975	9,418	244	-10,832	-24,131	-40,015	-58,890	-81,207
20%	52,844	50,198	46,902	42,814	37,771	31,580	24,022	14,848	3,772	-9,527	-25,411	-44,285	-66,603
25%	70,407	67,761	64,465	60,378	55,334	49,143	41,585	32,411	21,335	8,037	-7,847	-26,722	-49,040
30%	91,414	88,768	85,472	81,384	76,341	70,150	62,592	53,418	42,342	29,044	13,160	-5,715	-28,033
35%	116,407	113,761	110,466	106,378	101,334	95,143	87,586	78,412	67,336	54,037	38,153	19,279	-3,039

NPV Change in millions for long-term total revenue and expense growth rate							
Revenue/ Expense	-15%	-10%	-5%	0%	5%	10%	15%
-15%	14,671	12,829	10,852	8,735	6,472	4,059	1,489
-10%	17,767	15,924	13,947	11,830	9,568	7,154	4,585
-5%	21,083	19,241	17,264	15,147	12,885	10,471	7,901
0%	24,630	22,788	20,811	18,694	16,431	14,018	11,448
5%	28,416	26,573	24,596	22,479	20,217	17,803	15,234
10%	32,449	30,607	28,630	26,513	24,250	21,837	19,267
15%	36,739	34,896	32,919	30,803	28,540	26,127	23,557

Figure 4 Sensitivity tests for total revenue and expense growth rate

Recommended ranges of key assumptions are shown in Table 13. The worst scenario can involve more frequent pandemic outbreaks, lasting recession and failure to implement strategies, leading to higher expense growth and lower revenue growth. This is contrasted with best and practical scenario with better consumer expectations and more disposable income.

Assumption Range	Worst scenario	Best&Practical Scenario
First year total expense growth	40%	30%
First year total revenue growth	6%	20%
Short-term total expense growth	20%	15%
Short-term total revenue growth	20%	30%
Long-term total expense growth	12%	9%
Long-term total revenue growth	8%	12%
Total expense growth	10%	8%
Total revenue growth	7%	10%
Total expense variance	2%	1%
Total revenue variance	1%	1%
NPV at Year 10	0	58,383,567,431

Table 13 Key assumption range

5.2 QUANTIFIABLE RISKS

High incidence of injuries on football player could pose a burden on Rarita's new team due to absences of player and recovery costs. According to Owoeye, VanderWey and Pike (2020), the incidence of injuries in male professional adult has an overall exposure of 2.5-9.4 injuries/1000h, and during games, the exposure has even higher risk of injuries as shown in Table 14. Most injuries occur during the initial and final 15 minutes, indicating inappropriate warm-up and fatigue are significant factors of injuries. To mitigate this risk, suitable warn-up would be introduced, for example, neuromuscular training (NMT) warm-up programs. Equipment choices is also critical, including appropriate shoes.

Type of exposure	Category of participation		
	Male elite youth	Male professional adult	Female youth and adult
Overall (range)	2.0–19.4 injuries/1000 h	2.5–9.4 injuries/1000 h	
Game (range)	9.5–48.7 injuries/1000 h	8.7–65.9 injuries/1000 h	12.5–30.3 injuries/1000 h
Practice (range)	3.7–11.4 injuries/1000 h	1.4–5.8 injuries/1000 h	1.2–3.8 injuries/1000 h

Table 14 Incidence of injuries in soccer

Another quantifiable risk is extreme epidemic risk outlined in Appendix 7.4.1 as another global pandemic is unlikely to occur in the next ten years.

5.3 QUALITATIVE RISKS

1. **Ethical issues and reputation risks:** With the strong desires to win, football players may choose to dope. This can compromise the integrity of competition and can bring the reputation of the whole country into a scandal.
   - Mitigation: Avoid to signing players with criminal record: Team should balance financial and reputational considerations with the possibility of achieving a higher league position and promotion.
2. **Political risks:** The tension between the host countries and neighbor countries or other countries can cause damages to football athletes, including body injuries, kidnapping, and murder of athletes.
   - Mitigation: Look for domestic players meeting the selection criteria, because they are more of a known quantity on and off the field and less upheaval is required.

Other mitigation methods are outlined in Appendix 7.4.2

5.4 RISK RANKING

1. Ethical and reputation risks
2. Healthcare risks
3. Political risks

The two key metrics to rank risks are frequency and severity. By common sense, the healthcare risk has the highest frequency of incidence followed by ethical and reputation risks. However, a strike on reputation would directly result in the significant deduction on future revenue compared to loss of revenue due to injuries of athletes. Hence, it ranked the first followed by healthcare risks. Political risk has lowest ranking due to its extremely low frequency. There are countable numbers of international events affecting the football team.

6. DATA LIMITATIONS

1. The 'Tournament Passing' and 'Tournament Defense' data in 2020 is missing. As player performance features are selected by variable significance, the unavailability of 2020 tournament data disallows the split of training and test datasets on aggregate team level. This can lead to not optimal model chosen to support the process of selecting players.
2. Tournament and league data only involves the past two consecutive years of 2020-2021. No links between the improvement of ranking and past revenue growth and expense growth can be explored due to limited data. This results in higher dependence in assumptions that are the sources of uncertainty.

Limitations of missing data and negative values are discussed in Appendix 7.5.

7. CONCLUSION

We propose weighted averages of rankings based on different features of players by positions and estimate the team's competitiveness in probabilities. The 10-year cash flow model is built upon the commercial strategy to generate 7% IRR and PV of 21335.49 million using valid assumptions. We predict constructing a national football team can have good impacts on Rarita's economy from multiple aspects including tourism, retailing, employment and willingness to consume. Those should affect East Rarita the most. We discussed ethical, reputation, healthcare and political risks and their mitigation methods. Furthermore, our analysis is subject to missing and insufficient data limitations, changes of assumptions and future economic conditions. Those factors bring uncertainty to the strategies and profits suggested.

7. APPENDIX

7.1 MODEL CONSTRUCTION FOR PLAYER SELECTION

7.1.1 GENERAL METHODOLOGY FOR PLAYER SELECTION

Without an overall score and ranking for each player in each position, the only objective indicator to measure successfulness is the tournament ranking results at national team level.

Salary is arguably an objective measure of competitiveness and potential as players' pay can vary significantly based on player experience and the league or team they belong to. Hence, we did not select salary as our model output but more object ranking.

Due to missing value of 2020 tournament data, we only used 2021 tournament data to construct predictive models. This made split of training and testing datasets impossible on aggregate team levels due to the inaccuracy of model outputs when using a small training dataset.

The modelling steps are outlined below:

1. We plotted correlation matrix to explore the linear relationship between tournament ranking and all player performance features. This step aims to reduce the collinearity problems within variables and reduce modelling noise by selecting the variables with more influential powers on the final ranking.
2. We then aggregate player level data to national team level by addition or taking averages of the individual performance variables.
3. We conducted exploratory data analysis to visualize the distinctive impacts of the selected variables on the independent variable, namely the tournament ranking results.
4. We used linear, shrinkage and Ensemble models with relatively non-highly-correlated variables as inputs to quantify the relationships between the performance features and output of tournament ranking. Without the use of training and testing datasets, the summary model statistics of R squared and other criterions are used for model selection.
5. We selected performance features and assigned weights based on the importance and predictive powers of each variable in the constructed models.
6. In the League dataset, we ranked each performance features selected to eliminate the effects of large value gaps. Next, we calculated an overall rank for each player based on weighted average rank of those selected performance features.
7. Finally, we selected fourteen players with the highest ranks with filtered positions and substitute players are also selected based on their ranking and age.

From Figure 5, except 90s, Total Cmp, Short Cmp, Short Att, Total Att, Total Cmp%, Total TotDist have the highest correlation to tournament rankings.

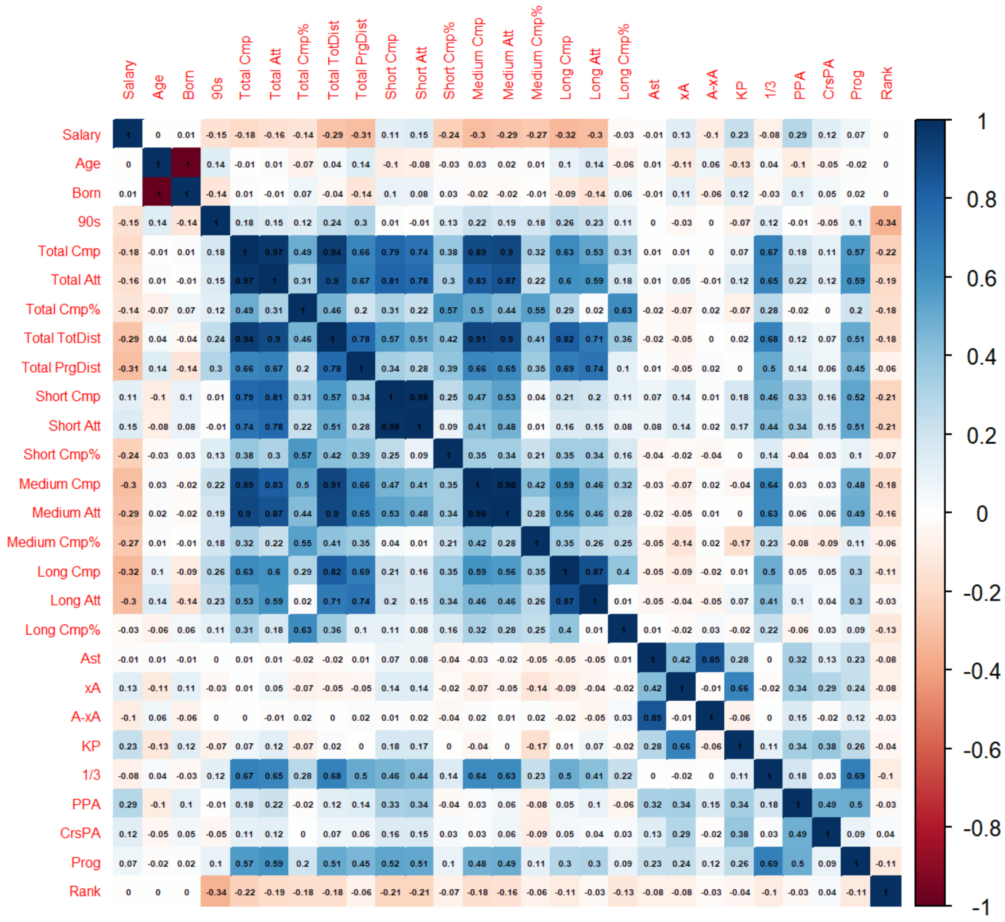


Figure 5 Correlation matrix for passing position in 2021 tournament data

From Figure 6, except playing Time 90s and its similar measurements, Performance GA, performance SoTA, Performance Saves, L, W, Penalty Kicks PKm have the highest correlation to tournament rankings.

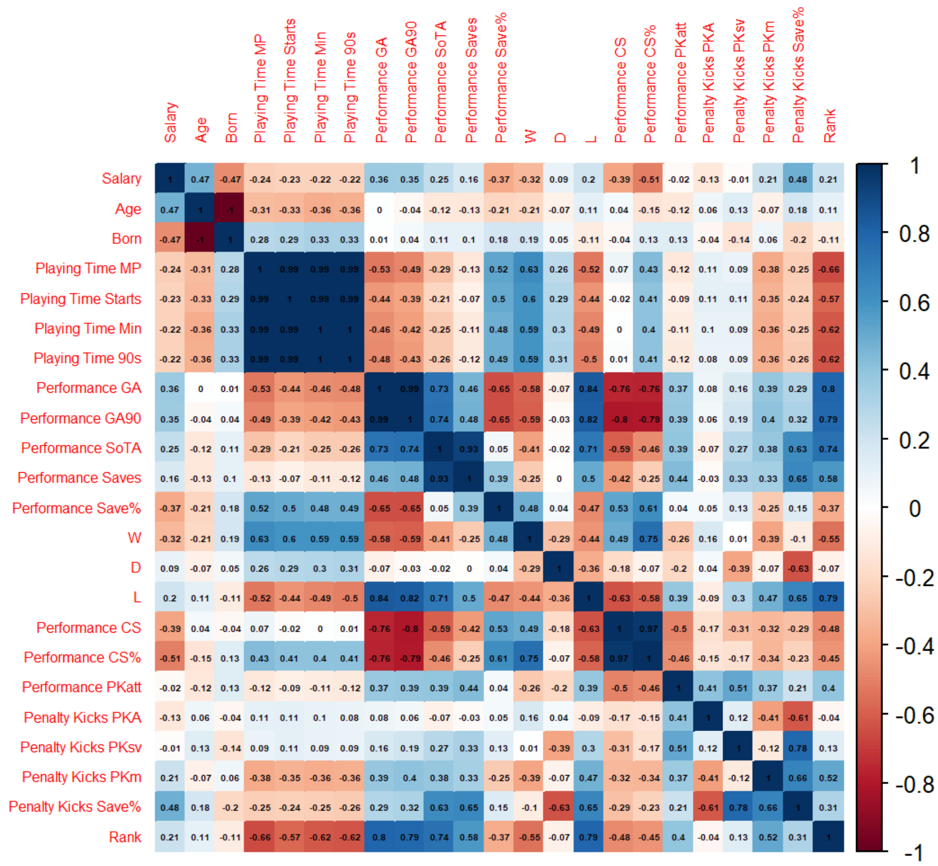


Figure 7 Correlation matrix for defense position in 2021 tournament data

From Figure 7, except 90s, Tackles Def 3rd, Tackles Tkl, Pressures Def 3rd, Tkl+Int have the highest correlation to tournament rankings.

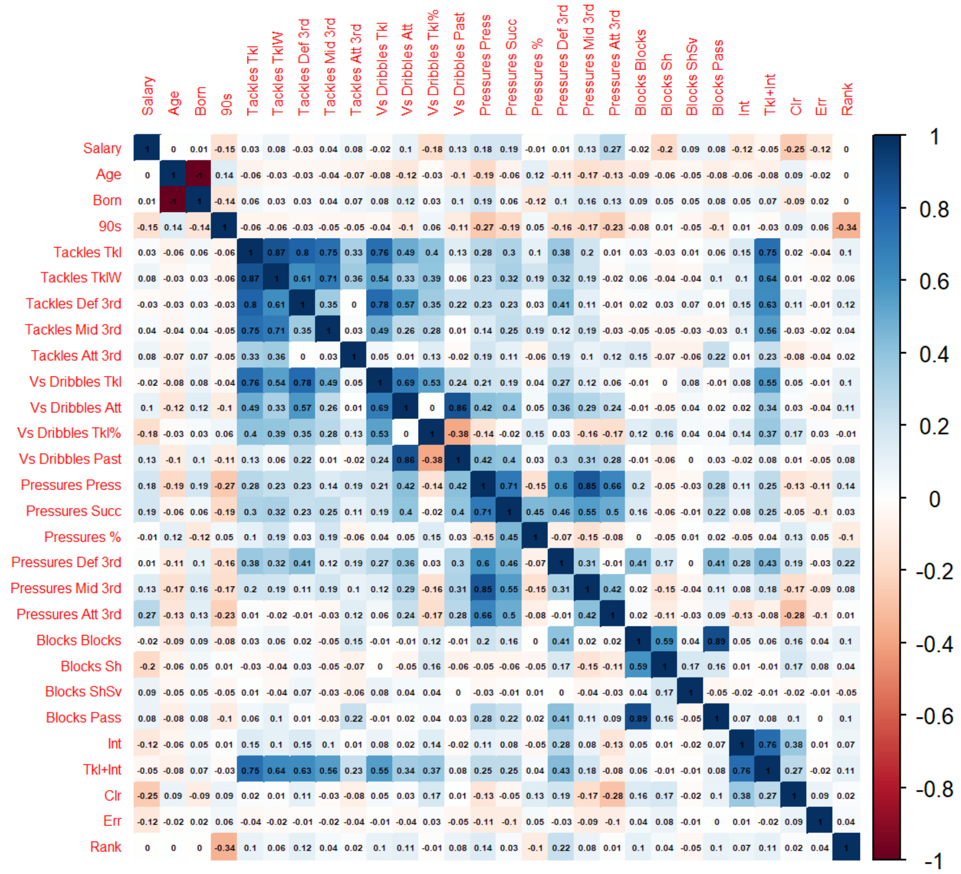


Figure 7 Correlation matrix for goalkeeping position in 2021 tournament data

7.1.2 LINEAR AND SHRINKAGE MODELS

We adopted linear models and backwards stepwise feature selection to select features by importance and predictive powers.

From Table 15, for shooting data, offensive positions such as forwards and midfields, 90s, Standard Dist, Performance Pkatt, Expected xG,Gls, Standard SoT are the most important factors based on p-value and significance tests. Backwards stepwise selection emphasized on the top four important features.

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.565e+01  5.770e+00   4.445 2.03e-05 ***
PosFW               -7.214e+00  2.525e+00  -2.857 0.005074 **
PosFWDF             -1.159e+01  3.032e+00  -3.824 0.000214 ***
PosFWMF             -9.349e+00  2.435e+00  -3.840 0.000202 ***
PosGK               -1.045e+01  2.533e+00  -4.127 6.98e-05 ***
PosMF               -6.648e+00  1.818e+00  -3.657 0.000386 ***
PosMFDF             -1.028e+01  2.803e+00  -3.669 0.000371 ***
Salary              -2.933e-08  2.852e-08  -1.028 0.305992
`90s`               -9.931e-01  1.419e-01  -6.998 1.85e-10 ***
Gls                 -3.208e+00  2.334e+00  -1.375 0.171951
`Standard Sh`        2.072e-01  1.618e-01   1.281 0.202771
`Standard SoT`      -4.911e-01  4.498e-01  -1.092 0.277210
`Standard Dist`      6.827e-02  2.858e-02   2.389 0.018531 *
`Performance PK`    -1.245e+00  3.609e+00  -0.345 0.730822
`Performance PKatt` -6.506e+00  3.634e+00  -1.790 0.076025 .
`Standard FK`       -1.036e+00  1.630e+00  -0.635 0.526539
`Expected xG`        3.552e+00  2.468e+00   1.439 0.152761
`Expected G-xG`      2.340e+00  2.607e+00   0.897 0.371415
Age                  4.483e-02  1.751e-01   0.256 0.798404
`Standard SoT%`      4.680e-03  2.519e-02   0.186 0.852919
`Standard G/Sh`     -2.938e-01  5.946e+00  -0.049 0.960677
`Standard G/SoT`     6.627e-01  2.682e+00   0.247 0.805292
`Expected npxG/Sh`  -7.507e+00  8.645e+00  -0.868 0.387014
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.252 on 115 degrees of freedom
Multiple R-squared:  0.5038,	Adjusted R-squared:  0.4089
F-statistic: 5.308 on 22 and 115 DF,  p-value: 1.07e-09
```

Linear model fitted with 2021 tournament shooting data

```
Step:  AIC=461.07
Rank ~ Pos + `90s` + `Standard Dist` + `Performance PKatt`

                      Df Sum of Sq    RSS    AIC
<none>                             3372.5 461.07
- `Performance PKatt`  1     77.06 3449.6 462.19
- `Standard Dist`      1    222.81 3595.3 467.90
- Pos                  6    893.44 4265.9 481.50
- `90s`                1   2471.58 5844.1 534.94

Call:
lm(formula = Rank ~ Pos + `90s` + `Standard Dist` + `Performance PKatt`,
    data = get(tour_2021_team[1])[, -1])

Coefficients:
        (Intercept)                PosFW              PosFWDF              PosFWMF
           27.58627             -8.93989            -12.73566            -10.68890
              PosGK                PosMF              PosMFDF                `90s`
          -11.01577             -7.30327            -11.51990             -1.10396
    `Standard Dist`  `Performance PKatt`
            0.05165             -5.50371
```

Backwards stepwise feature selection for 2021 tournament shooting data

Table 15 Linear model and backwards stepwise selection for shooting data

From Table 16, for passing data, except positions, 90s, Total Cmp% and Total PrgDist are the most significant features based on p-values. Whereas, backwards stepwise selection suggests Total Cmp as an alternative feature replacing Total Cmp%.

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)       2.416e+01  8.797e+00   2.746  0.00705 ** 
PosFW            -5.346e+00  3.893e+00  -1.373  0.17245    
PosFWDF          -9.837e+00  3.830e+00  -2.569  0.01155 *  
PosFWMF          -7.416e+00  3.643e+00  -2.036  0.04416 *  
PosGK            -1.193e+01  3.492e+00  -3.416  0.00089 ***
PosMF            -2.645e+00  3.282e+00  -0.806  0.42211    
PosMFDF          -8.531e+00  3.443e+00  -2.477  0.01476 *  
Salary            3.337e-08  1.997e-08   1.671  0.09756 .  
`90s`            -1.038e+00  1.397e-01  -7.429 2.51e-11 ***
`Total Cmp`      -5.304e-02  2.388e-01  -0.222  0.82468    
`Total PrgDist`   5.399e-03  3.795e-03   1.423  0.15767    
`Short Cmp`       4.669e-03  2.554e-01   0.018  0.98545    
`Medium Cmp`      4.762e-02  2.379e-01   0.200  0.84172    
`Long Cmp`       -1.176e-01  2.823e-01  -0.417  0.67767    
`Long Att`        4.128e-02  1.279e-01   0.323  0.74749    
Ast              -2.864e+00  2.891e+00  -0.991  0.32410    
xA               -3.840e-01  2.926e+00  -0.131  0.89582    
`A-xA`            2.286e+00  2.889e+00   0.791  0.43048    
KP                1.059e-01  3.123e-01   0.339  0.73523    
PPA               1.218e-01  3.070e-01   0.397  0.69231    
CrsPA             8.888e-01  6.802e-01   1.307  0.19401    
Prog              1.210e-01  1.707e-01   0.709  0.47985    
Age               1.280e-03  1.823e-01   0.007  0.99441    
`Total Cmp%`     -1.400e-01  9.388e-02  -1.491  0.13883    
`Short Cmp%`      4.939e-02  7.311e-02   0.676  0.50074    
`Medium Cmp%`     6.031e-02  6.398e-02   0.943  0.34790    
`Long Cmp%`       4.224e-02  3.281e-02   1.288  0.20061    
`1/3`            -1.141e-01  3.586e-01  -0.318  0.75088    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.26 on 110 degrees of freedom
Multiple R-squared:  0.524,	Adjusted R-squared:  0.4071 
F-statistic: 4.484 on 27 and 110 DF,  p-value: 9.944e-09
```

Linear model fitted with 2021 tournament passing data

```
Step:  AIC=460.65
Rank ~ Pos + Salary + `90s` + `Total Cmp` + `Total PrgDist` + 
    CrsPA

                  Df Sum of Sq    RSS    AIC
<none>                         3266.3 460.65
- CrsPA            1     59.40 3325.7 461.14
- `Total PrgDist`  1    129.01 3395.3 464.00
- Salary           1    139.31 3405.6 464.42
- `Total Cmp`      1    143.22 3409.5 464.58
- Pos              6    628.29 3894.6 472.93
- `90s`            1   1889.52 5155.8 521.65

Call:
lm(formula = Rank ~ Pos + Salary + `90s` + `Total Cmp` + `Total PrgDist` + 
    CrsPA, data = get(tour_2021_team[2])[, -1])

Coefficients:
    (Intercept)            PosFW          PosFWDF          PosFWMF            PosGK            PosMF  
      2.531e+01       -6.993e+00       -1.039e+01       -8.294e+00       -1.087e+01       -3.221e+00  
        PosMFDF           Salary            `90s`      `Total Cmp`  `Total PrgDist`            CrsPA  
     -9.635e+00        3.685e-08       -1.086e+00       -3.793e-02        7.078e-03        7.394e-01  
```

Backwards stepwise feature selection for 2021 tournament passing data

Table 16 Linear model and backwards stepwise selection for passing data

From Table 17, for defense data, except positions and 90s, Tackles Def 3rd, Tackles Att 3rd, Pressures Succ are the most significant features based on p-values. Whereas, backwards stepwise selection suggests the importance of Vs Dribbles Att, Int variables.

```
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)           9.638e+00  8.778e+00   1.098   0.2748    
PosFW                -4.883e+00  4.149e+00  -1.177   0.2419    
PosFWDF              -9.356e+00  4.260e+00  -2.196   0.0303 *  
PosFWMF              -7.948e+00  3.855e+00  -2.062   0.0418 *  
PosGK                -6.633e+00  3.963e+00  -1.674   0.0972 .  
PosMF                -5.625e+00  3.097e+00  -1.816   0.0723 .  
PosMFDF              -9.670e+00  3.827e+00  -2.527   0.0130 *  
Salary               -8.532e-09  2.699e-08  -0.316   0.7526    
Born                  1.087e-01  1.068e-01   1.019   0.3108    
`90s`                -1.275e+00  1.619e-01  -7.873 3.66e-12 ***
`Tackles Tkl`        -1.346e+00  3.210e+00  -0.419   0.6758    
`Tackles TklW`       -1.245e-01  4.173e-01  -0.298   0.7660    
`Tackles Def 3rd`     4.648e+00  2.441e+00   1.904   0.0597 .  
`Tackles Mid 3rd`     4.542e+00  2.409e+00   1.885   0.0622 .  
`Tackles Att 3rd`     5.075e+00  2.524e+00   2.011   0.0469 *  
`Vs Dribbles Tkl`    -2.405e+00  2.809e+00  -0.856   0.3940    
`Vs Dribbles Att`     2.172e+00  2.774e+00   0.783   0.4353    
`Vs Dribbles Past`   -1.698e+00  2.753e+00  -0.617   0.5388    
`Pressures Press`     3.317e+00  2.382e+00   1.392   0.1668    
`Pressures Succ`     -1.768e-01  1.048e-01  -1.687   0.0945 .  
`Pressures Def 3rd`  -3.294e+00  2.389e+00  -1.379   0.1709    
`Pressures Mid 3rd`  -3.320e+00  2.387e+00  -1.391   0.1674    
`Pressures Att 3rd`  -3.388e+00  2.399e+00  -1.412   0.1610    
`Blocks Blocks`       3.224e+00  2.564e+00   1.257   0.2114    
`Blocks Sh`          -3.383e+00  2.644e+00  -1.279   0.2036    
`Blocks ShSv`         1.600e+00  1.866e+00   0.857   0.3934    
`Blocks Pass`        -2.727e+00  2.543e+00  -1.072   0.2860    
Int                   3.358e+00  3.106e+00   1.081   0.2822    
`Tkl+Int`            -3.150e+00  3.086e+00  -1.021   0.3098    
Clr                  -1.387e-01  1.327e-01  -1.045   0.2983    
Err                   1.242e+00  2.405e+00   0.517   0.6066    
Year                 -1.064e-01  1.052e-01  -1.011   0.3144    
Age                   4.607e-01  2.654e-01   1.736   0.0856 .  
`Vs Dribbles Tkl%`    6.676e-02  2.819e-02   2.368   0.0197 *  
`Pressures %`         2.159e-03  2.728e-02   0.079   0.9371    
---
 Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Residual standard error: 5.063 on 103 degrees of freedom
 Multiple R-squared:  0.5871,	Adjusted R-squared:  0.4508 
 F-statistic: 4.307 on 34 and 103 DF,  p-value: 5.341e-09
```

Linear model fitted with 2021 tournament defense data

```
Step:  AIC=457.83
Rank ~ Pos + Born + `90s` + `Tackles Def 3rd` + `Tackles Mid 3rd` + 
    `Tackles Att 3rd` + `Vs Dribbles Att` + `Pressures Succ` + 
    `Blocks Blocks` + Int + `Tkl+Int` + Age + `Vs Dribbles Tkl%`

                    Df Sum of Sq    RSS    AIC
<none>                           2891.5 457.83
- Age                1     44.02 2935.5 457.92
- Born               1     76.16 2967.7 459.42
- `Tackles Def 3rd`  1     93.44 2985.0 460.22
- `Tackles Mid 3rd`  1     99.15 2990.7 460.49
- Pos                6    326.07 3217.6 460.58
- `Vs Dribbles Att`  1    103.54 2995.0 460.69
- `Tkl+Int`          1    109.42 3000.9 460.96
- `Blocks Blocks`    1    110.57 3002.1 461.01
- `Vs Dribbles Tkl%` 1    115.38 3006.9 461.23
- Int                1    117.62 3009.1 461.34
- `Pressures Succ`   1    119.77 3011.3 461.44
- `Tackles Att 3rd`  1    129.20 3020.7 461.87
- `90s`              1   1990.58 4882.1 528.12

Call:
lm(formula = Rank ~ Pos + Born + `90s` + `Tackles Def 3rd` + 
    `Tackles Mid 3rd` + `Tackles Att 3rd` + `Vs Dribbles Att` + 
    `Pressures Succ` + `Blocks Blocks` + Int + `Tkl+Int` + Age + 
    `Vs Dribbles Tkl%`, data = get(tour_2021_team[3])[, -1])

Coefficients:
      (Intercept)                PosFW              PosFWDF              PosFWMF                PosGK  
       15.9632419           -5.0775728           -8.6890709           -6.9126668           -5.9672588  
            PosMF              PosMFDF                 Born                `90s`    `Tackles Def 3rd`  
       -4.4594185           -9.1105696            0.0005311           -1.2115606            3.8865440  
`Tackles Mid 3rd`    `Tackles Att 3rd`    `Vs Dribbles Att`     `Pressures Succ`      `Blocks Blocks`  
        4.0208207            4.7046617            0.4068428           -0.1550157            0.3304055  
              Int            `Tkl+Int`                  Age   `Vs Dribbles Tkl%`  
        4.4065330           -4.2000098            0.2268438            0.0522161  
```

Backwards stepwise feature selection for 2021 tournament defense data

Table 17 Linear model and backwards stepwise selection for passing data

There are no linear models for goalkeeping data due to the limited number of observations.

We adopted shrinkage techniques to determine a small subset of variables with the strongest impact. Lasso is used to shrink the coefficients towards zero by applying L1 penalty. Ridge cannot assist in feature selection as the coefficients will not be reduced to zero but a very small number.

From Table 18, the coefficients of many variables are reduced to zero. For shooting data, lasso suggests the importance of 90s, Gls, Standard Sh, Standard SoT, Standard Dist, Performance Pkatt and offensive positions, which complies with the linear model and backwards stepwise selection. This is similar to the passing and defense data.

	s0
(Intercept)	16.76513201
Salary	.
Age	.
90s	-0.89265906
Gls	-0.34856191
Standard Sh	0.12427526
Standard SoT	-0.23524306
Standard SoT%	.
Standard G/Sh	.
Standard G/SoT	.
Standard Dist	0.04078991
Performance PK	.
Performance PKatt	-5.04653590
Standard FK	-0.60625122
Expected xG	.
Expected npxG/Sh	-4.99611068
Expected G-xG	-0.21554227
PosDF	7.78355913
PosFW	1.10987359
PosFWDF	-1.20268584
PosFWMF	.
PosGK	-0.30713234
PosMF	1.84749495
PosMFDF	-0.02618392

	s0
(Intercept)	2.028739e+01
Salary	2.314379e-08
Age	.
90s	-8.744078e-01
Total Cmp	.
Total Cmp%	-4.765096e-02
Total PrgDist	.
Short Cmp	.
Short Cmp%	.
Medium Cmp	.
Medium Cmp%	.
Long Cmp	.
Long Att	.
Long Cmp%	.
Ast	-2.532163e-01
xA	-2.377104e-01
A-xA	.
KP	.
1/3	.
PPA	.
CrsPA	6.235007e-01
Prog	.
PosDF	8.392075e+00
PosFW	.
PosFWDF	-1.272232e+00
PosFWMF	.
PosGK	-2.743664e-01
PosMF	3.089492e+00
PosMFDF	.

(Intercept)	1.171052e+01
Born	8.301948e-06
90s	-9.959669e-01
Tackles Tkl	.
Tackles TklW	.
Tackles Def 3rd	.
Tackles Mid 3rd	.
Tackles Att 3rd	3.801551e-01
Vs Dribbles Tkl	.
Vs Dribbles Att	.
Vs Dribbles Past	3.657117e-01
Pressures Press	.
Pressures Succ	-8.389793e-02
Pressures Def 3rd	4.325683e-02
Pressures Mid 3rd	3.288074e-03
Pressures Att 3rd	.
Blocks Blocks	1.000040e-01
Blocks Sh	.
Blocks ShSv	.
Blocks Pass	2.114702e-01
Int	1.272014e-01
Tkl+Int	.
Clr	.
Err	1.214071e+00
Year	1.195228e-04
Age	1.680811e-01
Vs Dribbles Tkl%	4.023295e-02
Pressures %	.
PosDF	4.836937e+00
PosFW	8.017550e-01
PosFWDF	-2.134699e+00
PosFWMF	-1.175476e+00
PosGK	.
PosMF	8.202123e-02
PosMFDF	-2.124210e+00

Table 18 Coefficients of Lasso for shooting, passing and defense positions

The combination of correlation matrix, linear models, backwards stepwise selection, lasso and Ensemble models discussed below generate the optimal feature selections that should be used to select players.

7.1.3 ENSEMBLE MODELS

According to Table 19, for shooting data, 90s, Standard_SoT, Gls, Standard_Sh and Performance_PKatt are the most significant variables.

```
> rf_shooting

Call:
 randomForest(formula = Rank ~ ., data = tour_2021_team_shooting[,      -1], importance = TRUE)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 5

          Mean of squared residuals: 35.01021
                    % Var explained: 24.44
```

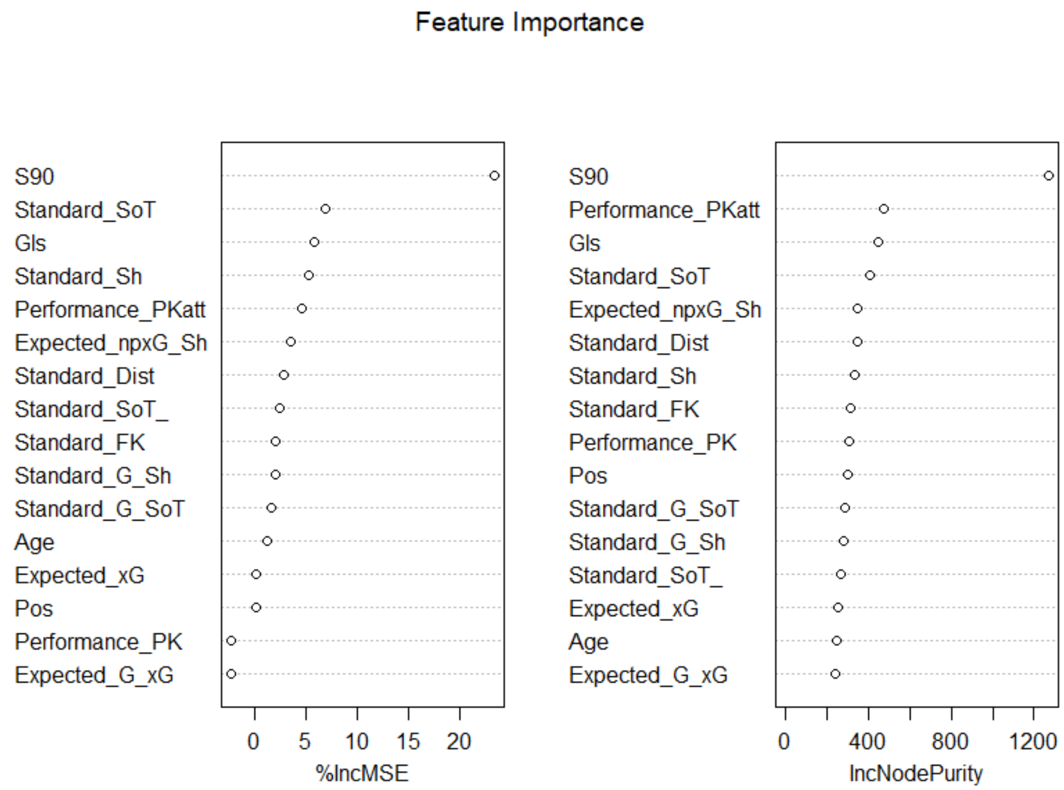


Table 19 Random Forest model and feature importance for shooting data

From Table 20, for passing data, 90s, Short_Cmp, Medium_Cmp, Ast are relatively important. Additionally, Total_PrgDist and Total_Cmp are indicated as significant factors.

```
> rf_passing

Call:
 randomForest(formula = Rank ~ ., data = tour_2021_team_passing[,      -1], importance = TRUE)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 7

          Mean of squared residuals: 37.82581
                    % Var explained: 18.36
```

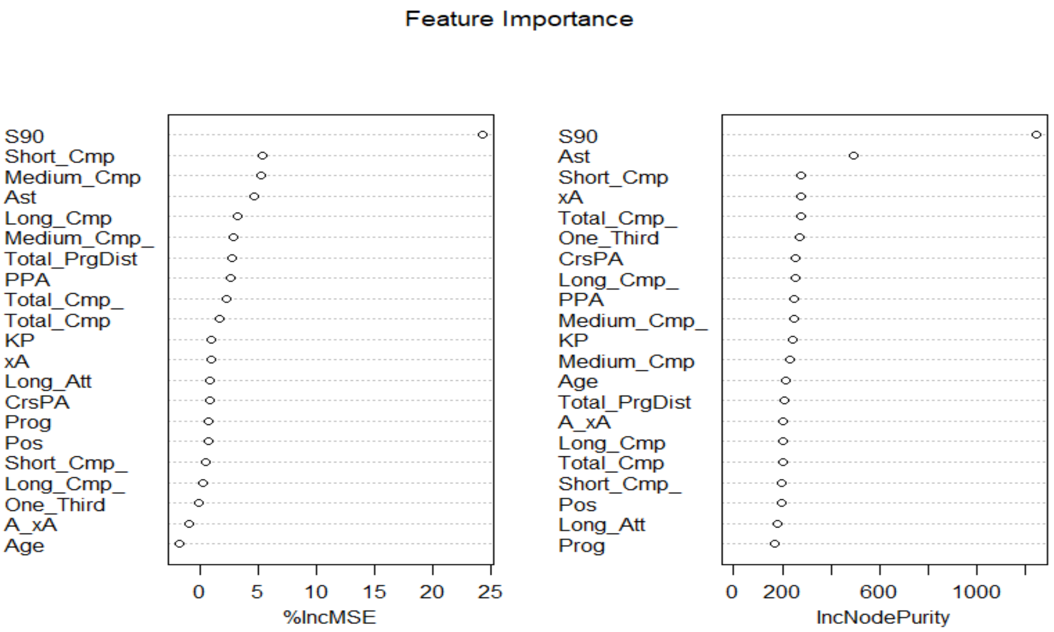


Table 20 Random Forest model and feature importance for passing data

From Table 21, for defense data, 90s, Pressures_Succ, Born, Vs_Dribbles_Att, Tackles Mid 3rd and Tackles Def 3rd are relatively significant factors.

```
> rf_defense

Call:
 randomForest(formula = Rank ~ ., data = tour_2021_team_defense[,      -1], importance = TRUE) 
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 9

          Mean of squared residuals: 33.93544
                    % Var explained: 26.76
```

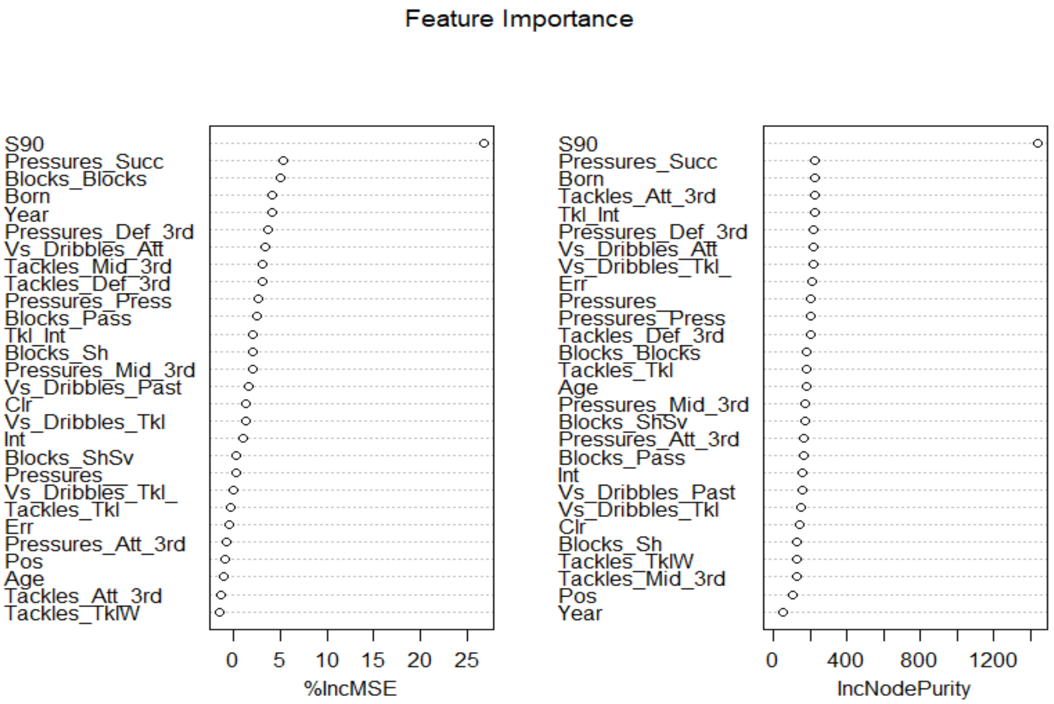


Table 21 Random Forest model and feature importance for defense data

As is shown in Table 22, Playing_Time_90s, L, Performance_SoTA, Penalty_Kicks_PKm, W, Performance GA and Performance_Saves are indicated as important variables.

```
> rf_goalkeeping

Call:
 randomForest(formula = Rank ~ ., data = tour_2021_team_goalkeeping[,      -1], importance = TRUE)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 5

          Mean of squared residuals: 11.93112
                    % Var explained: 75.87
```

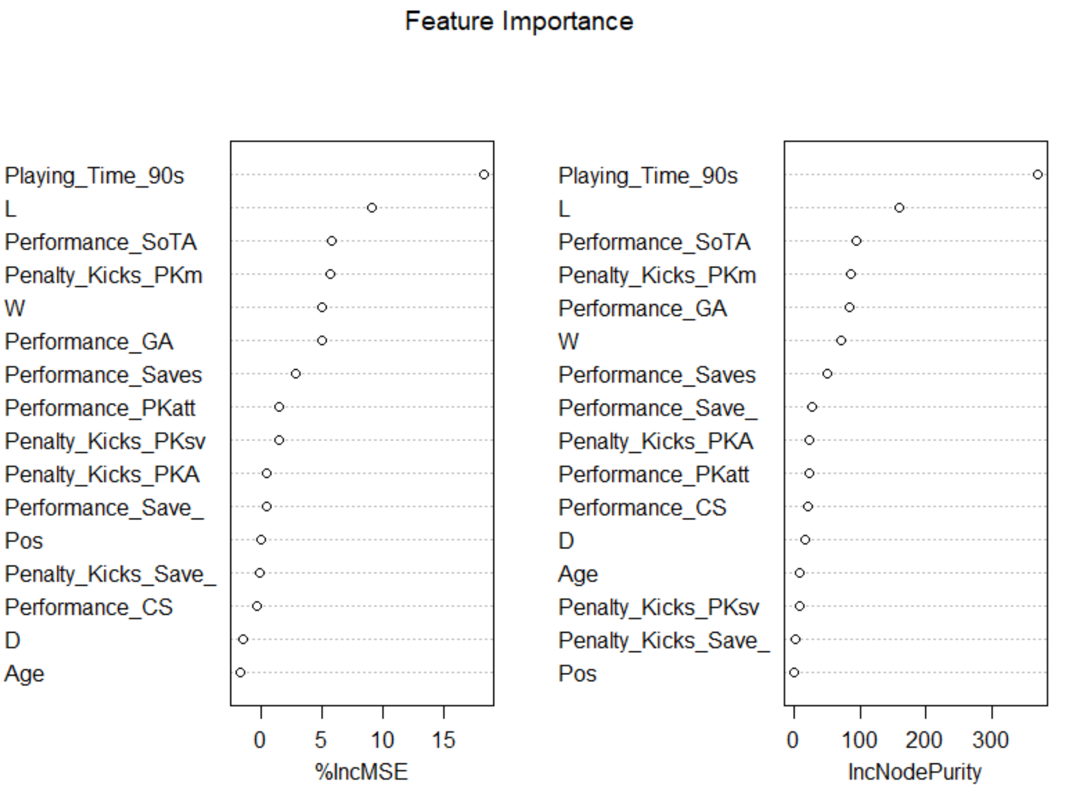


Table 22 Random Forest model and feature importance for goalkeeping data

The important variables indicated in the feature importance figures above for each dataset are then considered to assist with final variable weights determinations.

7.1.4 EXPLORATORY DATA ANALYSIS

We conducted exploratory data analysis to confirm divergence between predictors and the output. Since there are significant number of predictors, only a few significant predictors from the above models are selected and plotted in Figures 8-10.

Interestingly, remarkable differences are visible by different positions. Especially, offensive positions including forwards and midfields have higher means than other defensive positions. The difference in color represents different rankings. The shapes of boxes are the ranges of values of goals for each team. Due to the notable difference in rankings, goals can be seen as a relatively predictive variable.

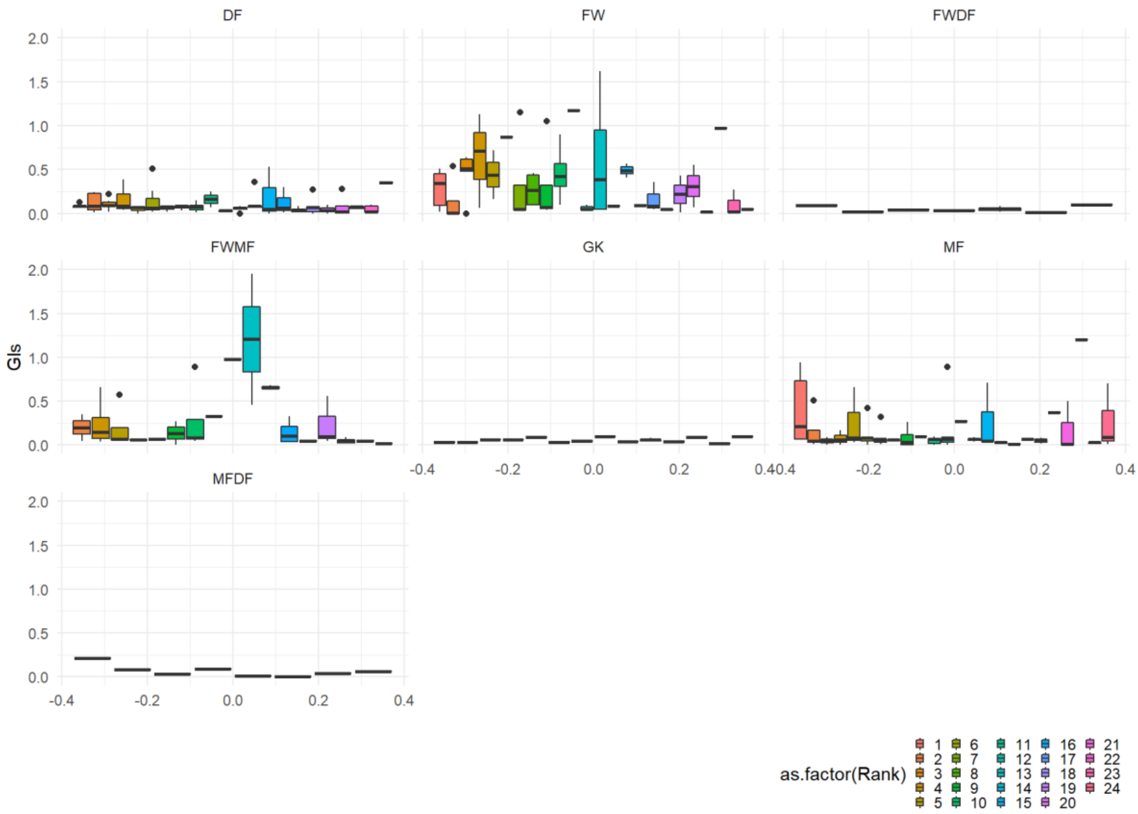


Figure 8 Boxplot of goals by position and ranking

From Figure 9, diverged density plots and range of x-axis are the proofs of significant difference in team performance for teams with different ranking.

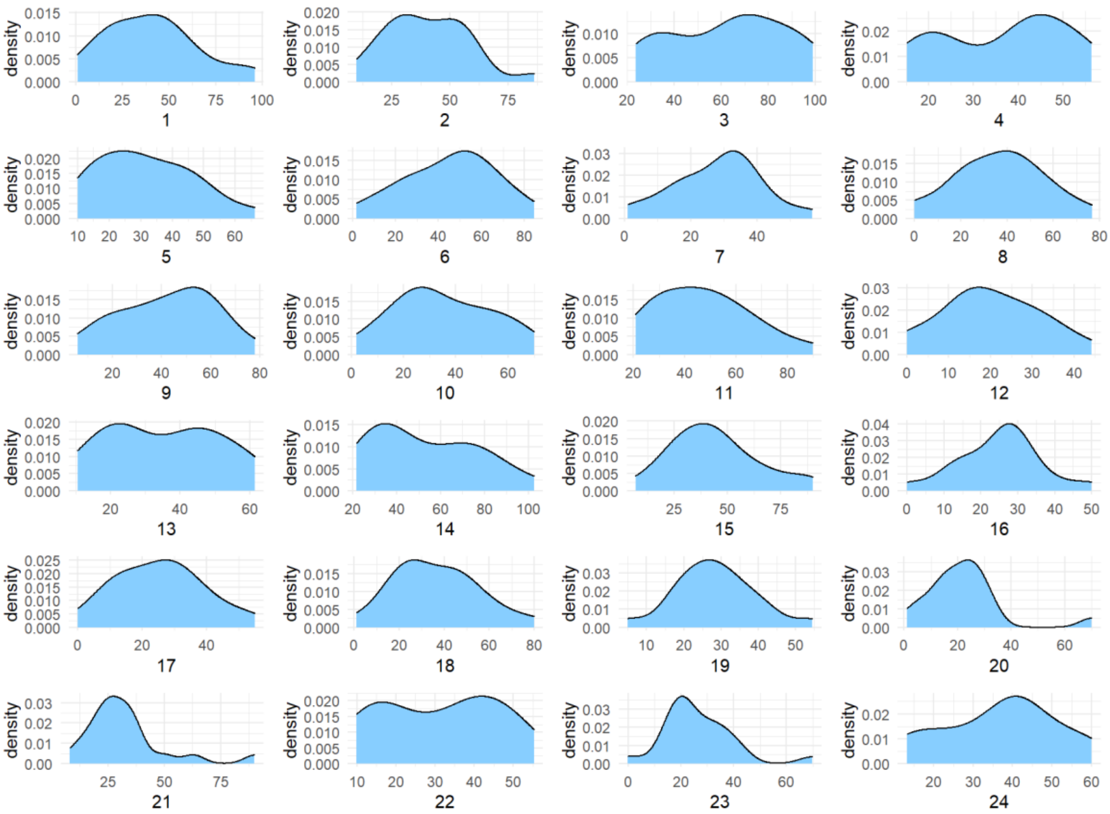


Figure 9 Frequency plot of passes completed by rank

Predictively, teams with higher 2021 tournament ranking have less loses and more wins. This is showed in the increasing trend of average loses with increasing rankings in Figure 10.

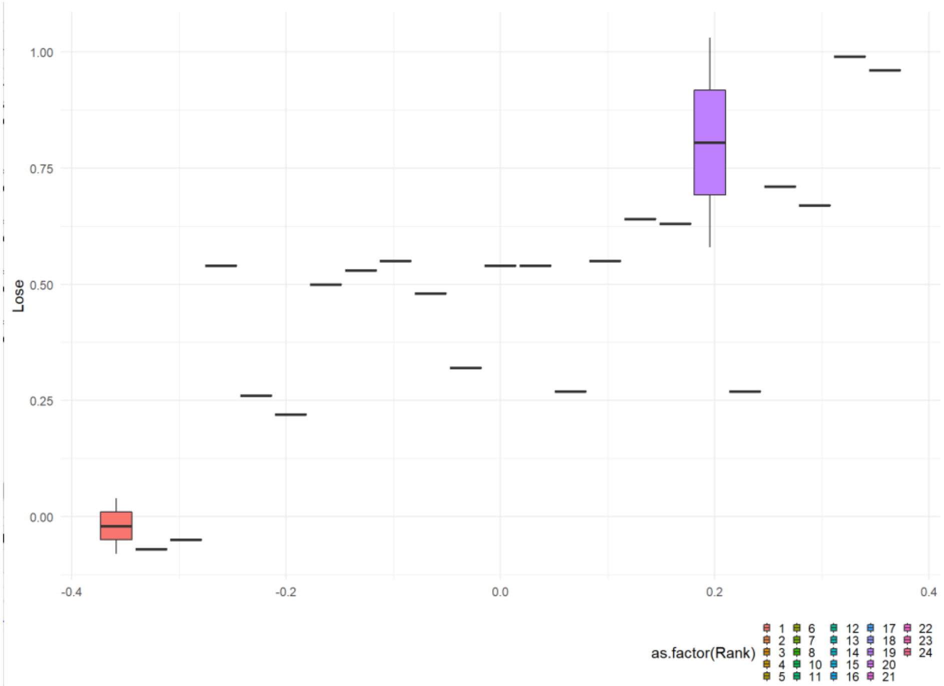


Figure 10 Boxplot of lose by ranking

All those plots confirm their predictive powers in ranking, which support the below conclusions.

7.1.5 PERFORMANCE FEATURERS SELECTED AND THEIR WEIGHTS

Position	Performance Features					
Shooting	Minutes play divided by 90	Goals	Shots on target	Average distance from goal of all shots taken	Penalty kicks attempted	Expected goals
Weights	15%	25%	20%	15%	15%	10%
Passing	Minutes play divided by 90	Passes completed	Passes completed (15-30 yards)	Completed crosses into the 18-yard box		
Weights	10%	35%	25%	30%		
Defense	Minutes play divided by 90	Tackles in defensive 1/3	Number of times dribbled past plus number of tackles	Number of times dribbled past plus number of tackles	Interceptions	
Weights	10%	30%	25%	20%	15%	
Goalkeeping	Minutes play divided by 90	Shot on target against	Win	Lose	Penalty kick missed	
Weights	10%	25%	20%	30%	15%	

Table 23 Performance features and corresponding weights

7.1.6 METHODOLOGY FOR TEAM CONSTRUCTION AND PROBABILITY CALCULATION

Applied Lasso to rate players both in tournaments and leagues, each player had a comprehensive quantified indices in shooting, passing, defense and goalkeeping, which was then summed up to calculate teams' performances in these four areas. After that we used data in Table 2 to generate linear models to determine the weights of each index impacting team's performances. It turned out that passing outshined others to contribute most to high ranking, followed by shooting. Under 10% level of significance, F-test's p-value of 0.005762 indicated the whole model's significance. Probability of being successfully competitive was treated as an optimization problem, which was subject to the weights of each index we just obtained, range of each index within being top 10, and Rarita players' individual performance.

7.2 10-YEAR STRATEGY AND PROFIT & LOSS ANALYSIS

Timeline in year	-1	0	1	2	3	4	5	6	7	8	9	10
Year	**2020**	**2021**	**2022**	**2023**	**2024**	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**	**2031**
Profit & Loss (in millions)						**With national team built**						
Total Expense	**∂ 1,868.95**	**∂ 2,055.85**	**∂ 2,878.19**	**∂ 3,479.03**	**∂ 4,206.48**	**∂ 5,227.53**	**∂ 6,232.21**	**∂ 6,870.39**	**∂ 8,691.29**	**∂ 9,248.08**	**∂ 10,225.82**	**∂ 11,631.47**
Staff cost	∂ 1,234.95	∂ 1,336.30	∂ 2,014.73	∂ 2,435.32	∂ 2,944.53	∂ 3,136.52	∂ 3,739.33	∂ 4,809.28	∂ 6,083.90	∂ 5,548.85	∂ 6,135.49	∂ 6,978.88
Other expense	∂ 634.00	∂ 719.55	∂ 863.46	∂ 1,043.71	∂ 1,261.94	∂ 2,091.01	∂ 2,492.89	∂ 2,061.12	∂ 2,607.39	∂ 3,699.23	∂ 4,090.33	∂ 4,652.59
Total Revenue	**∂ 2,993.92**	**∂ 3,143.62**	**∂ 3,615.16**	**∂ 4,491.66**	**∂ 5,579.37**	**∂ 6,768.70**	**∂ 8,506.43**	**∂ 11,348.67**	**∂ 13,549.65**	**∂ 15,441.13**	**∂ 17,042.17**	**∂ 18,342.99**
Matchday	∂ 309.59	∂ 490.40	∂ 650.73	∂ 786.04	∂ 948.49	∂ 1,116.84	∂ 1,361.03	∂ 1,815.79	∂ 2,032.45	∂ 2,316.17	∂ 2,556.32	∂ 2,751.45
Broadcast	∂ 797.41	∂ 1,226.01	∂ 1,626.82	∂ 1,998.79	∂ 2,454.92	∂ 2,944.38	∂ 3,657.77	∂ 4,766.44	∂ 5,419.86	∂ 5,404.40	∂ 5,964.76	∂ 6,420.05
Commercial	∂ 943.46	∂ 1,427.20	∂ 1,337.61	∂ 1,706.83	∂ 2,175.96	∂ 2,707.48	∂ 3,487.64	∂ 4,766.44	∂ 6,097.34	∂ 7,720.57	∂ 8,521.08	∂ 9,171.50
Other Revenue	∂ 943.46	∂ 94.31	∂ 108.45	∂ 157.21	∂ 223.17	∂ 304.59	∂ 425.32	∂ 567.43	∂ 677.48	∂ 772.06	∂ 852.11	∂ 917.15
Overall profit	**∂ 1,124.97**	**∂ 1,182.08**	**∂ 845.43**	**∂ 1,169.83**	**∂ 1,596.07**	**∂ 1,845.77**	**∂ 2,699.54**	**∂ 5,045.71**	**∂ 5,535.85**	**∂ 6,965.10**	**∂ 7,668.45**	**∂ 7,628.68**
						Without national team built						
Total Expense	**∂ 1,868.95**	**∂ 2,055.85**	**∂ 2,199.17**	**∂ 2,445.24**	**∂ 2,720.90**	**∂ 3,256.46**	**∂ 3,485.97**	**∂ 2,949.13**	**∂ 3,694.05**	**∂ 3,743.69**	**∂ 4,053.94**	**∂ 4,604.33**
Staff cost	∂ 1,234.95	∂ 1,336.30	∂ 1,451.45	∂ 1,613.86	∂ 1,795.79	∂ 2,149.26	∂ 2,300.74	∂ 1,946.43	∂ 2,438.07	∂ 2,470.84	∂ 2,675.60	∂ 3,038.86
Other expense	∂ 634.00	∂ 719.55	∂ 747.72	∂ 831.38	∂ 925.10	∂ 1,107.20	∂ 1,185.23	∂ 1,002.71	∂ 1,255.98	∂ 1,272.86	∂ 1,378.34	∂ 1,565.47
Total Revenue	**∂ 2,993.92**	**∂ 3,143.62**	**∂ 3,450.08**	**∂ 3,660.98**	**∂ 3,882.12**	**∂ 3,835.20**	**∂ 4,206.55**	**∂ 5,427.80**	**∂ 5,101.34**	**∂ 5,846.10**	**∂ 6,349.71**	**∂ 6,607.19**
Matchday	∂ 309.59	∂ 490.40	∂ 586.51	∂ 622.37	∂ 659.96	∂ 651.98	∂ 715.11	∂ 922.73	∂ 867.23	∂ 993.84	∂ 1,079.45	∂ 1,123.22
Broadcast	∂ 797.41	∂ 1,226.01	∂ 1,380.03	∂ 1,464.39	∂ 1,552.85	∂ 1,534.08	∂ 1,682.62	∂ 2,171.12	∂ 2,040.54	∂ 2,338.44	∂ 2,539.88	∂ 2,642.88
Commercial	∂ 943.46	∂ 1,427.20	∂ 1,483.54	∂ 1,574.22	∂ 1,669.31	∂ 1,649.14	∂ 1,808.82	∂ 2,333.95	∂ 2,193.58	∂ 2,513.82	∂ 2,730.37	∂ 2,841.09
Other Revenue	∂ 943.46	∂ 94.31	∂ 103.50	∂ 109.83	∂ 116.46	∂ 115.06	∂ 126.20	∂ 162.83	∂ 153.04	∂ 175.38	∂ 190.49	∂ 198.22
Overall profit	**∂ 1,124.97**	**∂ 1,182.08**	**∂ 1,354.42**	**∂ 1,325.57**	**∂ 1,277.69**	**∂ 693.80**	**∂ 846.77**	**∂ 2,641.50**	**∂ 1,560.33**	**∂ 2,277.79**	**∂ 2,486.26**	**∂ 2,201.07**

P&L for two scenarios with and without the national team built

Timeline in year	0	1	2	3	4	5	6	7	8	9	10
Profit & Loss (in millions)/Year	**2021**	**2022**	**2023**	**2024**	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**	**2031**
Total Expense		**∂ 679.02**	**∂ 1,033.79**	**∂ 1,485.58**	**∂ 1,971.07**	**∂ 2,746.24**	**∂ 3,921.26**	**∂ 4,997.24**	**∂ 5,504.39**	**∂ 6,171.88**	**∂ 7,027.13**
Staff cost		∂ 563.28	∂ 821.47	∂ 1,148.74	∂ 987.25	∂ 1,438.59	∂ 2,862.85	∂ 3,645.83	∂ 3,078.01	∂ 3,459.89	∂ 3,940.02
Other expense		∂ 115.74	∂ 212.33	∂ 336.84	∂ 983.81	∂ 1,307.66	∂ 1,058.41	∂ 1,351.41	∂ 2,426.38	∂ 2,711.99	∂ 3,087.11
Total Revenue		**∂ 165.08**	**∂ 830.68**	**∂ 1,697.25**	**∂ 2,933.50**	**∂ 4,299.88**	**∂ 5,920.87**	**∂ 8,448.31**	**∂ 9,595.03**	**∂ 10,692.46**	**∂ 11,735.80**
Matchday		∂ 64.21	∂ 163.67	∂ 288.53	∂ 464.85	∂ 645.92	∂ 893.06	∂ 1,165.22	∂ 1,322.33	∂ 1,476.87	∂ 1,628.23
Broadcast		∂ 246.79	∂ 534.40	∂ 902.07	∂ 1,410.30	∂ 1,975.15	∂ 2,595.32	∂ 3,379.33	∂ 3,065.96	∂ 3,424.87	∂ 3,777.17
Commercial		-∂ 145.93	∂ 132.61	∂ 506.64	∂ 1,058.34	∂ 1,678.82	∂ 2,432.49	∂ 3,903.77	∂ 5,206.74	∂ 5,790.71	∂ 6,330.41
Other Revenue		∂ 4.95	∂ 47.38	∂ 106.71	∂ 189.54	∂ 299.13	∂ 404.60	∂ 524.44	∂ 596.67	∂ 661.62	∂ 718.93
Overall profit		**-∂ 508.99**	**-∂ 155.73**	**∂ 318.38**	**∂ 1,151.97**	**∂ 1,852.77**	**∂ 2,404.21**	**∂ 3,975.52**	**∂ 4,687.31**	**∂ 5,182.19**	**∂ 5,427.61**
Funding	**∂ 995.00**		**∂ 350.00**								
Ending balance	∂ 995.00	∂ 487.24	∂ 682.66	∂ 1,002.31	∂ 2,154.39	∂ 4,014.50	∂ 6,451.67	∂ 10,495.23	∂ 15,307.36	∂ 20,708.34	∂ 26,470.57
Cost coverage	∂ 879.26	∂ 274.91	∂ 345.82	∂ 18.50	∂ 846.74	∂ 2,956.09	∂ 5,100.26	∂ 8,068.85	∂ 12,595.37	∂ 17,621.23	
PV of Profit		-∂ 508.44	-∂ 155.55	∂ 313.65	∂ 1,118.37	∂ 1,763.64	∂ 2,238.42	∂ 3,606.89	∂ 4,112.40	∂ 4,397.82	∂ 4,448.29
Cumulative PV		-∂ 508.44	-∂ 664.00	-∂ 350.34	∂ 768.03	∂ 2,531.66	∂ 4,770.09	∂ 8,376.98	∂ 12,489.38	∂ 16,887.20	∂ 21,335.49

P&L for direct impacts of the national team

Table 24 Profit and loss analysis

7.2.1 CASH FLOW MODEL METHODOLOGY

The given data presents revenue and expense per capita for all Rarita football teams. The population of 2020 is known for Rarita in the economic data. The product of 2020 population and revenue per capita is the total revenue earn by all Rarita football teams. This also applied to total expense. We used the total revenue and expense at the end of 2020 as starting points for future revenue and expense prediction.

Next, we predicted the total expense growth rate and revenue growth rate for each year for 2021 and for future 10 years. We expected the funding of 995 million doubloons to occur at the end of 2021. By analyzing the historical trend of total revenue and expense growth rate from 2016 to 2019 as shown in Figure 11, a reasonable growth rate for 2021 is determined with the slow pace of economic recession considered. 2020 data is excluded due to the COVID-19 impacts, that is unlikely to occur in the near future. Assumption values are displayed in Section 4.2. we predicted the total revenue and expense using the previous year figure timing the growth rate.

The proportion assumptions of total revenue and total expense do not affect the overall profit as they are calculated using the total revenue (expense) timing the corresponding proportion assumption.

Two scenarios are calculated as the direct revenue and expense of one national team is difficult to quantify, however, by subtraction of with and without national team-built scenarios, the direct impacts can be evaluated.

We calculated the cost coverage at the end of year balance with additional funding subtracting subsequent year other expense that is assumed to happen at the beginning of year. If the cost coverage value is less than zero, then it is assumed to be not sufficient funds for the national football to operate. The amount of funding is decided in this way.

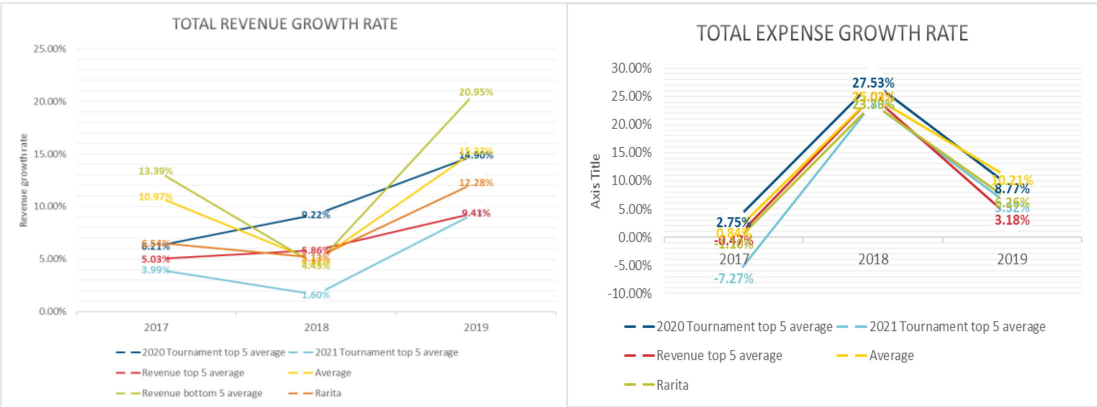


Figure 11 Rarita total revenue and expense growth rate comparison

7.2.2 OTHER ASSUMPTIONS

1. All cash flows occur at the end of year expect other expense.
2. The team will successfully meet the competitive criterium in 5 and 10 years respectively.
3. Volatility of total expense and total revenue growth rate after the national team built is significantly lower than before.
4. Three-stage separation model: There will still be residual impacts of COVID-19 pandemic on global economy in 2022 but minimal impacts for 2022 onwards. For 2023-2027, the team will be subject to multiple changes including player composition, revenue and expense composition due to successful implement of strategies and achieving higher rankings in the FSA and other potential changes in short-term. At this stage, the team is expected to experience revenue boost and expense management. After Year 7, the team is expected to be mature in team operation and should be subject to long-term expense and revenue growth rate.
5. Total expense and revenue growth rates follow a normal distribution with mean and standard deviations calculated from historical data.

6. Future statistics for RFL are assumed to keep the trends analyzed from the historical data provided. The mean and variance calculated using 2016-2020 data with the trend shown in Figure 12, justifying the assumptions used for without-national-team-built scenario.

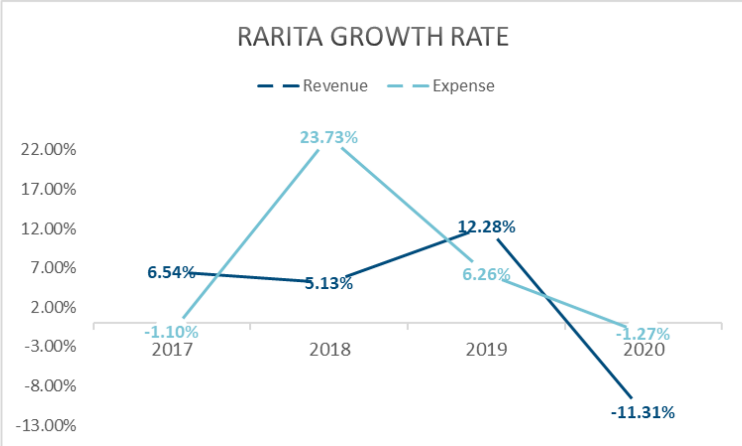


Figure 12 Rarita revenue and expense growth rate

7. Anomalies in historical revenue and expense trends imply strategic shift of football leagues or new establishment of new teams. Nations with lower tournament rankings established their national teams later than nations with higher rankings. Based on historical analysis, the nation of 'Eastern Sleboube' is assumed to establish its national team around 2019-2020. Its growth rates in revenue and expense are used as reference. From Figure 13, when COVID-19 hit all other national football clubs, 'Eastern Sleboube' shows a diverged trend with total revenue growth of 32%. A possible reason for this irregular growth is that the nation recently built a national team. This can be further confirmed with lacking of 2020 tournament ranking and a relatively low ranking of 19 (out of 21) in 2021 tournament results.

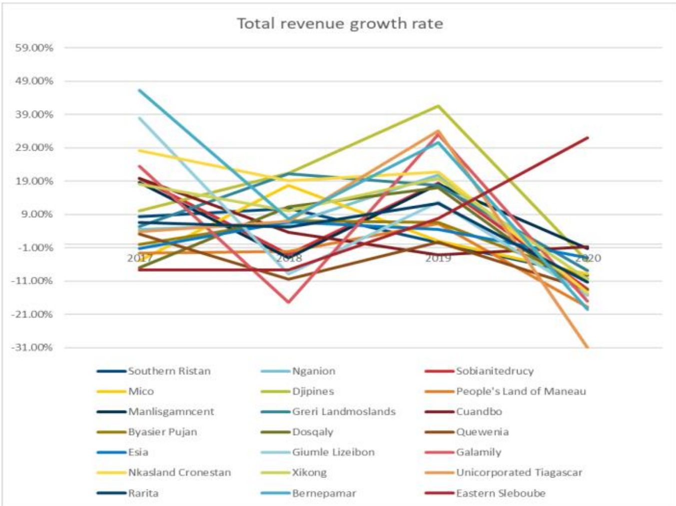


Figure 13 Total revenue growth rate for all nations

The long-term total expense and revenue growth rates of 11% and 10% respectively can be justified by the boxplot of 3 or 4-year compound annual growth rate (CAGR) for all nations in Figure 14, with the assumption of mean conversion in the long-term. The long-term growth is also supported in Deloitte analysis shown in Figure 15.

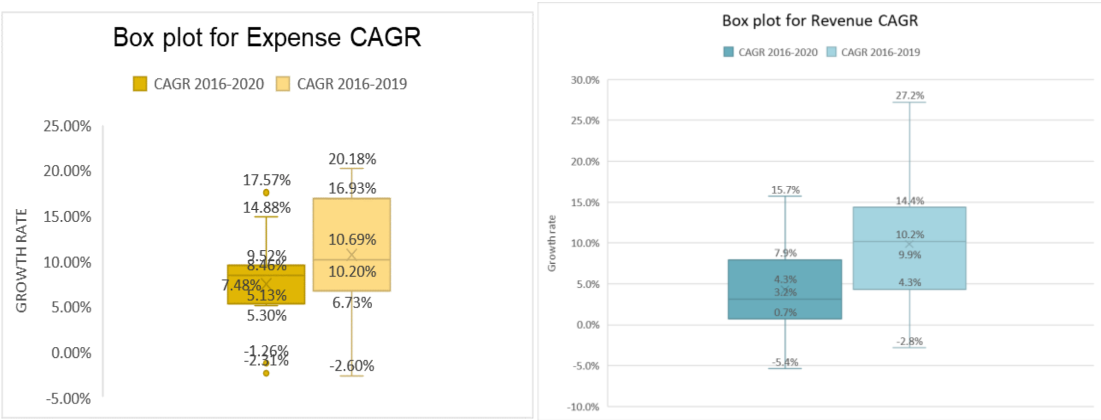


Figure 14 Boxplots for expense and revenue CAGR

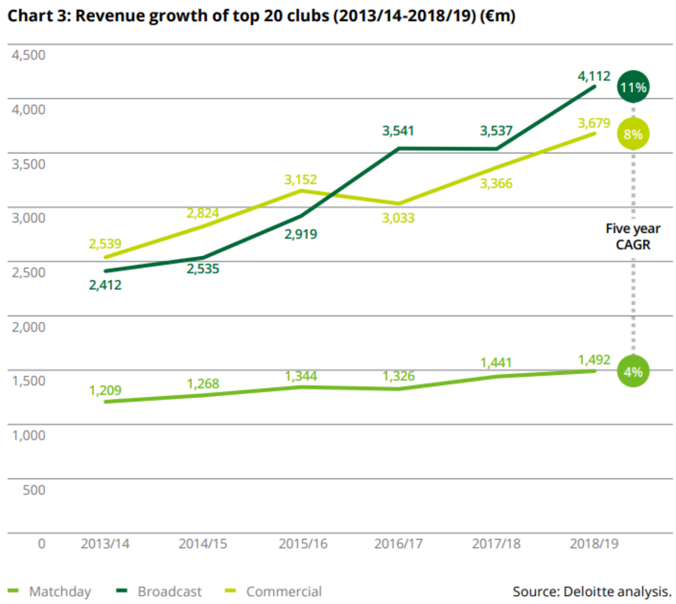


Figure 15 From Deloitte Football Money League 2020 Report

From Section 2.3.1, all players selected are from foreign country, justifying for higher staff cost in first three years as shown in section 4.2. With the assumption of frequent acquisition and transfer of players and changes in operation teams and management, older players with current age above 25 will be replaced by better or equivalent younger players, justifying for higher staff cost in Year 6 and 7.

7.2.3 NON-GOVERNMENT FUNDING SOURCES

- **Grants:** two major types of grants from the non-government sector includes club grants and foundation or corporate grants. Club grants are usually community development funding programs coordinated by local councils. Local companies make foundation or corporate grants to show their responsibility to the community. However, the corporate fund shrinks over time.
- **Sponsorship:** Football sponsors provide funds to football teams to purchase essential assets including team kits, equipment, training facilities, or travel. As exchanges, football teams would need to advertise for the sponsors to help build their reputation and influence.
- **Retail, merchandising, apparel & product licensing:** the strength of the football team as a brand can be leveraged to supplement as a source of funding. Selling clothes and other licensed products featured by the football brand can collect funds from the retail branch.

7.2.4 THREE TYPES OF REVENUE SOURCES

- **Matchday revenue** composed largely of ticket sales at local stadiums. As more successful clubs in various competition have higher brand image and many supporters, they are likely to generate more matchday revenue (How Do Football Clubs Make Money?, 2022). Stadium tours and loaning out the stadium for filming and other proposes is another source of matchday revenue during the off-season.

- **Broadcast revenue consists of TV deals as** foreign countries will buy the rights to broadcast live games (Asika, 2017). League has the revenue distribution rights to clubs, similar to La Liga, clubs with better ranking are more likely to take a larger share of broadcast revenue (Sarkar, 2016).
- **Commercial revenue** depends on degrees of mechanization including sales of all jerseys, hats, scarves, jackets and badges. Establishing large contracts with sponsors is another major source of commercial revenue, which can also be a source of non-governmental funding dependent on the size of funding injection.

7.2.5 10-YEAR STRATEGY FROM THE MANAGEMENT SIDE

A sporting director (SD) is solely responsible for building and maintaining team competitiveness to ensure optimal player acquisition and professional assessments of players' in-game performances. While, sales team should focus on sponsorship development, ticket sales and commercialization strategies. Client servicing team should adopt a client-centric approach to maintain and effectively use the customer relationship management (CRM) systems and identifying and satisfying customers' needs. Marketing team should be in charge of brand management and enhancement. Those functions are essential to pursue more growth opportunities and retain and acquire undervalued talents.

7.2.6 OTHER STRATEGY MONITORING METRICS

- **Market capitalization:** Market capitalization refers to the total value of all a company's shares of stock. Market capitalization is reported on an annual basis.
- **Growth of net debt:** Net debt determines how well a company can pay its debts when they were due. Hence, a stable, lower growth of net debt is preferred. Growth rate is monitored on an annual basis.
- **Net debt to revenue ratio:** The net debt to revenue ratio measures the amount of debt to overall revenue. A lower and stable net debt to revenue ratio means good financial health. The net debt to revenue ratio is reported on an annual basis.
- **Equity turnover:** Equity turnover consists of the proportion of the football team's revenue to its shareholder equity. Higher rates mean managements have efficiently used funds and be able to make more revenue. It represents the amount of return from each dollar shareholders' equity. Equity turnover ratio is reported on an annual basis.
- **Budgeted Revenues:** A revenue that is expected to be achieved in a year.
- **Average players' market value:** A player's market value can be a metric showing how valuable the player is to the team and club as an asset. A higher market value of players in a team means that the team is more popular. A higher market value means a higher commercial potential, better ticket selling, and a better resource of sponsorship. Average players' market value is reported on an annual basis.

7.3 ECONOMIC ANALYSIS

7.3.1 INTANGIBLE EFFECTS

7.3.1.1 HEALTH

Sports can not only enhance the physical strength of body but build up the mental health. In Germany, 4.9 billion dollars of the 5.6 billion dollars of the health savings was generated by playing football came from subjective well-being (Campelli, 2022). Therefore, with the new football team established, Rarita's governments would potentially reduce relevant healthcare expense.

7.3.1.2 REDUCED CRIME RATE

Playing sports can reduce the risk of an individual turning to crime from 52.5% to 37%. This is because people have greater propensity to be employed rather than getting money illegally (Campelli, 2022).

For Rarita, building the national team could possibly assist with the nation's crime rate control, which is essential for the country's development.

7.3.2 ECONOMIC ANALYSIS

GROSS NATIONAL INCOME (GNI) PER CAPITA

Year	East Rarita	Central Rarita	West Rarita	Rarita	Year	East Rarita	Central Rarita	West Rarita
2011	∂ 37,890	∂ 27,534	∂ 16,652	∂ 22,596	2011	31.37%	37.74%	54.42%
2012	∂ 38,347	∂ 26,957	∂ 17,096	∂ 22,778	2012	31.37%	36.34%	55.26%
2013	∂ 38,662	∂ 26,806	∂ 17,509	∂ 23,026	2013	31.24%	35.39%	55.94%
2014	∂ 39,588	∂ 27,230	∂ 17,819	∂ 23,449	2014	31.27%	34.82%	55.65%
2015	∂ 44,427	∂ 27,950	∂ 19,082	∂ 25,121	2015	32.63%	32.96%	55.35%
2016	∂ 44,416	∂ 28,439	∂ 19,615	∂ 25,565	2016	31.42%	32.04%	54.64%
2017	∂ 46,270	∂ 29,667	∂ 20,870	∂ 26,912	2017	30.86%	31.28%	54.63%
2018	∂ 47,989	∂ 30,964	∂ 21,976	∂ 28,164	2018	31.05%	31.52%	55.67%
2019	∂ 49,322	∂ 32,042	∂ 23,614	∂ 29,625	2019	30.59%	31.23%	57.25%
2020	∂ 46,830	∂ 30,615	∂ 22,383	∂ 28,140	2020	30.34%	30.83%	56.76%

Table 25 Gross domestic product data

GROSS DOMESTIC PRODUCT (GDP) PER CAPITA

Year	East Rarita	Central Rarita	West Rarita	Rarita	Year	East Rarita	Central Rarita	West Rarita
2011	∂ 46,119	∂ 22,581	∂ 9,445	∂ 18,292	2011	38.18%	30.95%	30.87%
2012	∂ 47,214	∂ 22,190	∂ 9,733	∂ 18,523	2012	38.62%	29.92%	31.46%
2013	∂ 48,159	∂ 22,123	∂ 9,977	∂ 18,785	2013	38.91%	29.21%	31.88%
2014	∂ 49,897	∂ 22,646	∂ 10,127	∂ 19,260	2014	39.41%	28.96%	31.63%
2015	∂ 55,404	∂ 23,866	∂ 10,741	∂ 20,770	2015	40.70%	28.15%	31.16%
2016	∂ 58,175	∂ 24,817	∂ 11,086	∂ 21,646	2016	41.16%	27.96%	30.88%
2017	∂ 62,042	∂ 26,405	∂ 11,759	∂ 23,047	2017	41.38%	27.84%	30.78%
2018	∂ 63,406	∂ 27,687	∂ 12,155	∂ 23,820	2018	41.02%	28.18%	30.79%
2019	∂ 65,046	∂ 28,839	∂ 13,013	∂ 24,880	2019	40.34%	28.11%	31.55%
2020	∂ 63,534	∂ 27,080	∂ 12,451	∂ 23,863	2020	41.16%	27.27%	31.57%

Table 26 Gross national income data

According to Table 25 and 26, it is obvious that East Rarita has the highest average GDP and gross income, indicating better economic condition, while West Rarita has the lowest indices. The percentages of gross GDP and income in each province has also been calculated. However, it is shown that West Rarita has the highest proportion of gross income, and the proportion of Gross GDP is not as low. This is possibly due to the high population in West Rarita.

HEALTHCARE SPENDING PER CAPITA

Year	East Rarita	Central Rarita	West Rarita	Rarita
2011	∂ 4,203	∂ 2,447	∂ 296	∂ 1,427
2012	∂ 4,367	∂ 2,367	∂ 308	∂ 1,437
2013	∂ 4,434	∂ 2,334	∂ 329	∂ 1,449
2014	∂ 4,458	∂ 2,375	∂ 335	∂ 1,465
2015	∂ 4,510	∂ 2,487	∂ 352	∂ 1,509
2016	∂ 4,604	∂ 2,534	∂ 362	∂ 1,541
2017	∂ 4,699	∂ 2,639	∂ 398	∂ 1,604
2018	∂ 4,787	∂ 2,747	∂ 420	∂ 1,657
2019	∂ 4,932	∂ 2,870	∂ 445	∂ 1,725
2020	∂ 4,979	∂ 2,839	∂ 460	∂ 1,730

Table 27 Healthcare spending data

HOUSEHOLD SAVINGS RATE

Year	East Rarita	Central Rarita	West Rarita	Rarita
2011	12.4%	9.1%	8.2%	9.0%
2012	11.4%	8.8%	6.7%	7.9%
2013	12.7%	6.9%	5.9%	7.2%
2014	12.6%	6.6%	7.2%	7.9%
2015	11.5%	5.4%	8.0%	7.9%
2016	12.5%	6.2%	8.9%	8.8%
2017	13.1%	6.9%	9.7%	9.6%
2018	13.5%	7.1%	9.6%	9.6%
2019	13.6%	7.3%	6.4%	7.7%
2020	13.9%	8.7%	6.9%	8.4%

Table 28 Household saving data

Based on Healthcare Spending and Household Saving data, it is obvious that population in East Rarita have higher healthcare spending and household saving. Considering the saving pattern across different age group, it could be assumed that East Rarita has relatively elder age group, while population in West Rarita could potentially be younger.

Overall, it could be concluded that among the 3 provinces, East Rarita has the best economic condition, while West Rarita has relatively low performance.

7.4 OTHER RELATED RISK AND MITIGATION METHODS

7.4.1 EXTREME EPIDEMIC RISK

The probability of extreme epidemics in any year is clarified as 2% by Marani, Katul, Pan and Parolari (2021). During 2020-2021, under the effect of Covid-19, the total revenue downturn by year is 11% for the 'Big Five' leagues ((Annual Review of Football Finance 2021 | Deloitte UK, 2022)), and the Enterprise value dropped by 15% for the 32 most prominent European football clubs (Sartori 2021). Certain pandemics would impose huge pressures on Rarita's football league, while the probability of certain events is relatively low. However, during the beginning of team formation in 2022, there exists the impact of Covid-19. To control and further mitigate the risk, rescheduling match calendars, more consistent cost control and optimization of governance could be helpful (Sartori 2021).

7.4.2 MITIGATION FOR ETHICAL AND REPUTATIONAL RISKS

Mitigation: Thorough due diligence before making the deal: There have been numerous examples of footballers re-entering the professional game after serving a jail sentence. A team with such players may hurt the goodwill of the team. It may violate the agreements with sponsors. It may also hurt the relationship with fans.

Mitigation: Set strict internal rules, and increase the severity of the punishment for doping to enormously big: Many of the companies that might sponsor football clubs are neither registered nor known in the nation and do not target local consumers. Dealing with unknown counterparties carries inherent risks; when a club agrees to a 'blind' deal with a partner, the greater the potential exposure to reputational and financial damage.

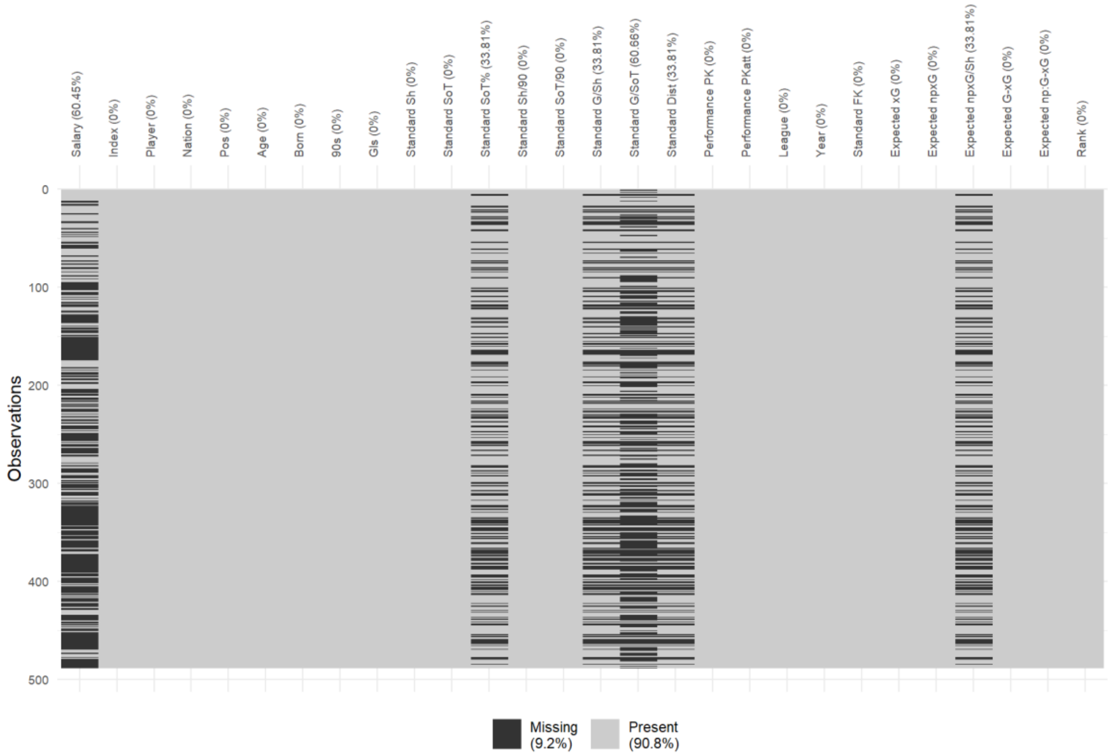
7.4.3 MITIGATION FOR POLITICAL RISKS

Mitigation: Apply border control at a government level: Most football teams owns 1/5 overseas players. However, overseas players may bring risks. Overseas players usually get a higher salary than domestic players. If the overseas player cannot perform well, the market value of that player goes down. It would be hard for the club to transfer the player out, and the club still needs to pay for the salary. In addition, under certain political risks, sanctions may be carried on players with certain nationalities.

7.5 OTHER DATA LIMITATIONS

7.5.1 MISSING VALUE

There are no data entries in the passing and defense sheet for 2020 tournament. This data limit restricts the model using 2021 to train and test datasets only. There are also some missing values across columns in the dataset as shown in Figures 16 and 17. If data is missing in this way, we choose MICE algorithm to simulate data for missing value. This may decrease the accuracy of the model.



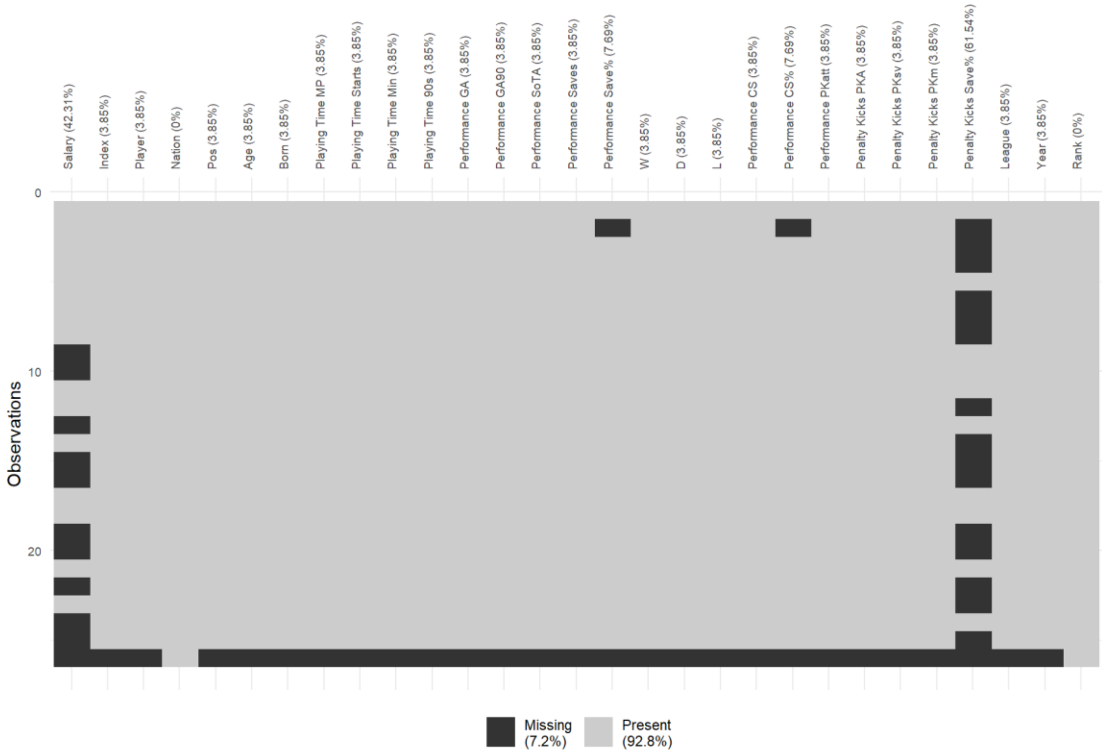


Figure 16 From Deloitte Football Money League 2020 Report

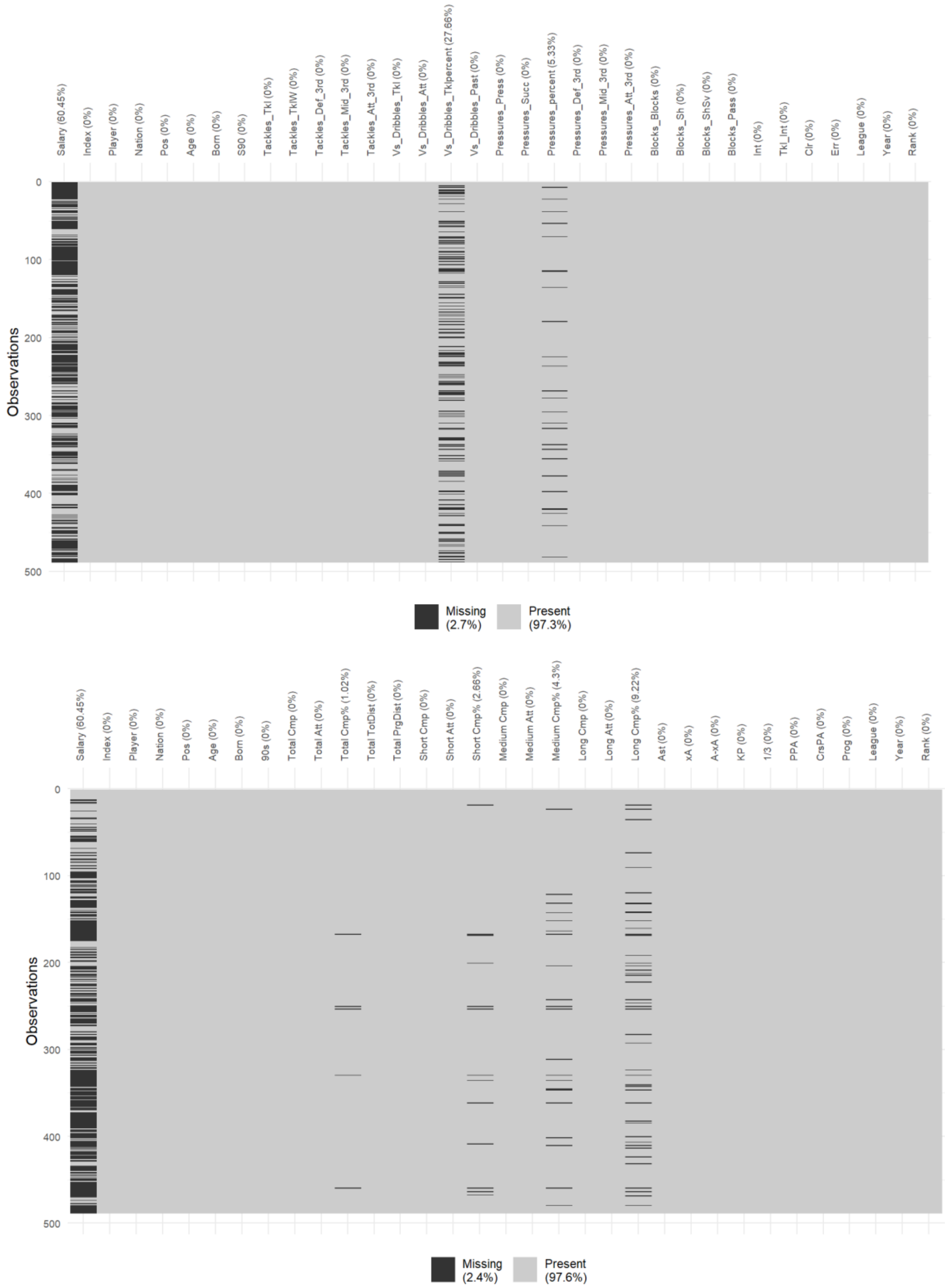


Figure 17 From Deloitte Football Money League 2020 Report

```
> summary(comp.lm)

Call:
lm(formula = Rank ~ Shooting + Passing + Defense + GK, data = comp)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.8988  -3.4933   0.1643   3.3601   7.3735 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) 23.37557   13.16148   1.776  0.09174 . 
Shooting    -0.04489    0.20208  -0.222  0.82656   
Passing     -0.21452    0.15447  -1.389  0.18098   
Defense      0.22477    0.22238   1.011  0.32484   
GK           0.57595    0.18257   3.155  0.00522 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.401 on 19 degrees of freedom
Multiple R-squared:  0.5181,	Adjusted R-squared:  0.4166 
F-statistic: 5.107 on 4 and 19 DF,  p-value: 0.005762
```

Figure 18 Weights of shooting, passing, defense and goalkeeping for team performance rating

7.5.2 NEGATIVE VALUE

There are negative data entries across some columns. For example, there is negative value for number of goals scored which does not make sense in common sense. We noticed that the negative values approach to zero closely with the lower bound of -0.1 and as we valued the ranking not the numbers, so we do not replace negative values. This may affect the model slightly.

7.6 R CODE

```
library(dplyr)
library(corrplot)
library(readxl)
library(stringr)
library(ggplot2)
library(naniar)
library(mice)
library(glmnet)
library(data.table)
library(MASS)
library(tidyr)
library(tidyverse)
library(gridExtra)
library(VIM)
library(e1071)
library(glmnet)
library(caret)
library(ROSE)
library(formatR)
library(kableExtra)
```

```
#import data
import_path="C:/Users/lenovo/Desktop/Control cycle/SOA challenge/"
tournament_shooting <- read_excel(paste0(import_path,"player-data.xlsx"),
                    sheet = "Tournament Shooting", range = "A18:AA2033",
                    col_types = c("numeric","text","text", "text", "text",
                              "numeric", "numeric", "numeric",
                              "numeric", "numeric", "numeric",
                              "numeric", "numeric", "numeric",
                              "numeric", "numeric", "numeric",
                              "numeric", "numeric", "text", "numeric",
                              "numeric", "numeric", "numeric",
                              "numeric", "numeric", "numeric"))

tournament_passing <- read_excel(paste0(import_path,"player-data.xlsx"),
                         sheet = "Tournament Passing", range = "A14:AF502")

tournament_defense <- read_excel(paste0(import_path,"player-data.xlsx"),
                         sheet = "Tournament Defense", range = "A12:AG500")

tournament_goalkeeping <- read_excel(paste0(import_path,"player-data.xlsx"),
                           sheet = "Tournament Goalkeeping", range = "A13:AB142")
X2021_rank <- read_excel(paste0(import_path,"player-data.xlsx"),
            sheet = "Tournament Results", range = "E11:F35")
colnames(X2021_rank)=c('Rank','Nation')



#check the data issues for the imported datasets
df.name=c('shooting','passing','defense','goalkeeping')
dfnames=paste0('tournament_',df.name)
for (data.name in dfnames){
  print(paste0(data.name,":"))
  print(summary(get(data.name)))
}

#filtering data for 2021 tournament
tour_goalkeeping_2021=tournament_goalkeeping%>%filter(Year=='2021')
tour_shooting_2021=tournament_shooting%>%filter(Year=='2021')

# combine the position dataset to tournament result
df2021=c('tour_shooting_2021','tournament_passing','tournament_defense','tour_goalkeeping_2021
')
for ( data.name in df2021){
```

```
  assign(data.name,get(data.name)%>%full_join(X2021_rank,by='Nation'))
}

#classify the data type
col.type=paste0("col.type_",df.name)
num_col=paste0("num_col_",df.name)
char_col=paste0("char_col_",df.name)
for ( i in 1:4){
  assign(col.type[i],as_tibble(sapply(get(df2021[i]), class)))
  assign(num_col[i],colnames(get(df2021[i]))[get(col.type[i])=='numeric'])
  assign(char_col[i],colnames(get(df2021[i]))[get(col.type[i])!='numeric'])
}

#correlation plot
for ( i in 1:4){

p=corrplot(cor(get(df2021[i])[,setdiff(get(num_col[i]),'Year')],use='pairwise.complete.obs'),tl.cex=0.5,
method = "color",number.cex=0.3,addCoef.col = "black")
  print(p)
}



#missing value calculation
miss_df=paste0("miss_",df.name)
Na_df=paste0("Na_",df.name)
neg_df=paste0("neg_",df.name)
data.issue=paste0("data.issue.",df.name)
for ( i in 1:4){
  assign(miss_df[i],as.matrix(colSums(is.na(get(df2021[i])[,get(num_col[i])]))))
  #% of missing values
  assign(Na_df[i],data.frame(round(get(miss_df[i])/nrow(get(df2021[i]))*100,3)))
  #% of negative values
  assign(neg_df[i],sapply(1:length(get(num_col[i])),function(j)
round(sum(get(df2021[i])[get(num_col[i])[j]]<0, na.rm=TRUE)/dim(get(df2021[i]))[1]*100,3)))
  #data.issue %summary
  assign(data.issue[i],cbind("Neg_percent"=get(neg_df[i]),"Na_percent"=get(Na_df[i])))
  #missing value plots
  p=vis_miss(get(df2021[i]))+ggtitle(paste("Fig",i,": Missing values",df2021[i]))+theme(axis.text.x =
element_text(angle = 90),axis.text = element_text(size = 7))
  print(p)
  g=md.pattern(get(df2021[i]),rotate.names = T)
  print(g)
}
```

```
vis_miss(get(df2021[3]))+ggtitle(paste("Fig",i,": Missing values",df2021[3]))+theme(axis.text.x =
element_text(angle = 90),axis.text = element_text(size = 7))
# Filling missing data on Pos & Nation
tournament_defense[is.na(tournament_defense$Pos),]$Pos=tournament_shooting[tournament_sho
oting$Player==tournament_defense[is.na(tournament_defense$Pos),]$Player,]$Pos

#aggr(tournament_shooting, col=c('navyblue','red'), numbers=TRUE, sortVars=TRUE,
#     labels=names(tournament_shooting), cex.axis=.2, gap=1,
#     ylab=c("Histogram of missing data","Pattern"))




#missing data issue
#orginal column names
orig_colnames=paste0("orig_colnames_",df.name)
for ( i in 1:4){
  assign(orig_colnames[i],colnames(get(df2021[i])))
}
colnames(tour_shooting_2021)[colnames(tour_shooting_2021)=="90s"]="S90"
colnames(tour_shooting_2021)=gsub(" |/|:|-|%","_",colnames(tour_shooting_2021))

colnames(tournament_passing)[colnames(tournament_passing)=="90s"]="S90"
colnames(tournament_passing)=gsub(" |/|:|-","_",colnames(tournament_passing))
colnames(tournament_passing)[colnames(tournament_passing)=="1_3"]="one_over3"
colnames(tournament_passing)=gsub("%","percent",colnames(tournament_passing))

colnames(tournament_defense)[colnames(tournament_defense)=="90s"]="S90"
colnames(tournament_defense)=gsub(" |/|:|-","_",colnames(tournament_defense))

colnames(tournament_defense)[colnames(tournament_defense)=="Tkl+Int"]="Tkl_Int"
colnames(tournament_defense)=gsub("%","percent",colnames(tournament_defense))

#solution 1: imputate data
tour_shooting_2021_i=mice(tour_shooting_2021,meth='rf',maxit=50,seed=500)
tour_shooting_2021_i=complete(tour_shooting_2021_i,1)
summary(tour_shooting_2021_i)

tournament_passing_i=mice(tournament_passing,meth='rf',maxit=50,seed=500)
tournament_passing_i=complete(tournament_passing_i,1)

tournament_defense_i=mice(tournament_defense,meth='rf',maxit=50,seed=500)
```

```
tournament_defense_i=complete(tournament_defense_i,1)


#check imputation
colnames(tour_shooting_2021_i)=orig_colnames_shooting
colnames(tournament_passing_i)=orig_colnames_passing
colnames(tournament_defense_i)=orig_colnames_defense

unique(tour_shooting_2021_i$Pos)
tour_shooting_2021_i[tour_shooting_2021_i$Pos=='MFFW',]$Pos='FWMF'
tour_shooting_2021_i[tour_shooting_2021_i$Pos=='DFFW',]$Pos='FWDF'
tour_shooting_2021_i[tour_shooting_2021_i$Pos=='DFMF',]$Pos='MFDF'

unique(tournament_passing_i$Pos)
tournament_passing_i[tournament_passing_i$Pos=='MFFW',]$Pos='FWMF'
tournament_passing_i[tournament_passing_i$Pos=='DFFW',]$Pos='FWDF'
tournament_passing_i[tournament_passing_i$Pos=='DFMF',]$Pos='MFDF'
unique(tournament_defense_i$Pos)
tournament_defense_i[tournament_defense_i$Pos=='MFFW',]$Pos='FWMF'
tournament_defense_i[tournament_defense_i$Pos=='DFFW',]$Pos='FWDF'
tournament_defense_i[tournament_defense_i$Pos=='DFMF',]$Pos='MFDF'
unique(tour_goalkeeping_2021$Pos)
tour_goalkeeping_2021=tour_goalkeeping_2021[tour_goalkeeping_2021$Pos=='GK',]

tour_goalkeeping_2021[tour_goalkeeping_2021$Pos!='GK',]$Pos='GK'
#exploratory plots
plot.freqpoly.shooting=lapply(unique(tournament_passing_i$Pos), function(x) ggplot(data =
tournament_passing_i%>%filter(grepl(x, Pos)), mapping = aes(x = 'Total Cmp' ))+
                      geom_freqpoly(binwidth = 0.05)+theme_minimal()+xlab(x))
do.call(grid.arrange, c(plot.freqpoly.shooting, ncol = 2, nrow = 5))

#by rank
plot.bar.shooting=lapply(unique(tournament_passing_i$Rank), function(x) ggplot(data =
tournament_passing_i%>%filter(Rank==x), mapping = aes(x =`Total Cmp`))+
                    geom_density(fill='skyblue1')+theme_minimal()+xlab(x))
do.call(grid.arrange, c(plot.bar.shooting, ncol = 4, nrow = 6))

#boxplot in loop
for (j in setdiff(num_col_shooting,c('Born','Year','Rank'))){
g=ggplot(data = tour_shooting_2021_i, mapping =
aes(y=tour_shooting_2021_i[[j]],fill=as.factor(Rank)))+
  geom_boxplot()+theme_minimal()+ylab(j)+
facet_wrap(vars(Pos))+theme(legend.position="bottom",legend.justification="right")+
theme(legend.key.size = unit(0.2, 'cm'))
```

```
print(g)
}
ggplot(data = tour_shooting_2021_i, mapping =
aes(y=tour_shooting_2021_i$Gls,fill=as.factor(Rank)))+
  geom_boxplot()+theme_minimal()+ylab('Gls')+ylim(0,3)+
facet_wrap(vars(Pos))+theme(legend.position="bottom",legend.justification="right")+
theme(legend.key.size = unit(0.2, 'cm'))


ggplot(data = tour_shooting_2021_i, mapping =
aes(y=tour_shooting_2021_i$Standard_SoT,fill=as.factor(Rank)))+
  geom_boxplot()+theme_minimal()+ylab("Standard SoT")+ylim(0,5)+
facet_wrap(vars(Pos))+theme(legend.position="bottom",legend.justification="right")+
theme(legend.key.size = unit(0.2, 'cm'))


ggplot(data =tournament_defense_i, mapping = aes(y=tournament_defense_i$`Vs Dribbles
Att`,fill=as.factor(Rank)))+
  geom_boxplot()+theme_minimal()+ylab("Vs Dribbles
Att")+ylim(0,5)+theme(legend.position="bottom",legend.justification="right")+
theme(legend.key.size = unit(0.2, 'cm'))


ggplot(data =tour_goalkeeping_2021, mapping =
aes(y=tour_goalkeeping_2021$L,fill=as.factor(Rank)))+

geom_boxplot()+theme_minimal()+ylab("Lose")+theme(legend.position="bottom",legend.justificatio
n="right")+  theme(legend.key.size = unit(0.2, 'cm'))


'''
+ facet_grid(cols = vars(fl))
+ facet_wrap(vars(fl))
+ facet_grid(rows = vars(year), cols = vars(fl))
'''


#convert player-level to team level
num_col_shooting=setdiff(num_col_shooting,c('Nation','Born','Standard Sh/90','Standard
SoT/90','Expected npxG','Expected np:G-xG','Year'))
num_col_passing=setdiff(num_col_passing,c('Nation','Born','Total Att','Total TotDist','Short
Att','Medium Att','Year'))
num_col_goalkeeping=setdiff(num_col_goalkeeping,c('Nation','Born','Year','Playing Time MP','Playing
Time Starts','Playing Time Min','Performance GA90','Performance CS%'))


num_col_mean=paste0("num_col_mean_",df.name)
num_col_sum=paste0("num_col_sum_",df.name)
tour_2021_sum=paste0("tour_2021_sum_",df.name)
tour_2021_mean=paste0("tour_2021_mean_",df.name)
tour_2021_team=paste0("tour_2021_team_",df.name)
df2021_i=c('tour_shooting_2021_i','tournament_passing_i','tournament_defense_i','tour_goalkeepin
g_2021')
```

```
for ( i in 1:4){
  assign(num_col_mean[i],c('Age','Rank',get(num_col[i])[grepl('%|/',get(num_col[i]))]))
  assign(num_col_sum[i],setdiff(get(num_col[i]),get(num_col_mean[i])))
  assign(tour_2021_sum[i],get(df2021_i[i])%>%group_by(as.factor(Nation),as.factor(Pos))%>%
         summarise_at(get(num_col_sum[i]),sum,na.rm=T))
  assign(tour_2021_mean[i],get(df2021_i[i])%>%group_by(as.factor(Nation),as.factor(Pos))%>%
         summarise_at(get(num_col_mean[i]),mean,na.rm=T))
  setnames(get(tour_2021_sum[i]),old=c("as.factor(Nation)","as.factor(Pos)"),new=c("Nation","Pos"))

setnames(get(tour_2021_mean[i]),old=c("as.factor(Nation)","as.factor(Pos)"),new=c("Nation","Pos"))
  assign(tour_2021_team[i],get(tour_2021_sum[i])%>%full_join(
   get(tour_2021_mean[i]),by=c('Nation','Pos')))
}



#linear model
linear_model=paste0("linear_model",df.name)
'''
for ( i in 1:4){
  assign(linear_model,lm(Rank~.,data=get(tour_2021_team[i])[,-1]))
  print(paste0(linear_model[i],":"))
  print(summary(get(linear_model[i])))
}
'''



linear_shoot=lm(Rank~.,data=get(tour_2021_team[1])[,-1])
summary(linear_shoot)
stepAIC(linear_shoot,direction = 'backward')

linear_pass=lm(Rank~.,data=get(tour_2021_team[2])[,-1])
summary(linear_pass)
stepAIC(linear_pass,direction = 'backward')

linear_defense=lm(Rank~.,data=get(tour_2021_team[3])[,-1])
summary(linear_defense)
stepAIC(linear_defense,direction = 'backward')



#lasso
x.shoot=cbind(data.matrix(get(tour_2021_team[1])[,num_col_shooting]),
model.matrix( ~ Pos-1, get(tour_2021_team[1])) )
```

```
cv_shoot <- cv.glmnet(x.shoot[ ,!(colnames(x.shoot) == "Rank")], get(tour_2021_team[1])[["Rank"]],
alpha = 1)
best_lambda <- cv_shoot$lambda.min
plot(cv_shoot)
best_shoot<- glmnet(x.shoot[ ,!(colnames(x.shoot) == "Rank")], get(tour_2021_team[1])[["Rank"]],
alpha = 1, lambda = best_lambda)
coef(best_shoot)
##passing
x.pass=cbind(data.matrix(get(tour_2021_team[2])[,num_col_passing]),
        model.matrix( ~ Pos-1, get(tour_2021_team[2])) )

cv_pass <- cv.glmnet(x.pass[ ,!(colnames(x.pass) == "Rank")], get(tour_2021_team[2])[["Rank"]], alpha
= 1)
best_lambda <- cv_pass$lambda.min
plot(cv_pass)
best_pass<- glmnet(x.pass[ ,!(colnames(x.pass) == "Rank")], get(tour_2021_team[2])[["Rank"]], alpha
= 1, lambda = best_lambda)
coef(best_pass)
##defense
tour_2021_team_defense=tour_2021_team_defense[,-3]
x.defense=cbind(data.matrix(get(tour_2021_team[3])[,-c(1:2)]),
          model.matrix( ~ Pos-1, get(tour_2021_team[3])) )
cv_defense <- cv.glmnet(x.defense[ ,!(colnames(x.defense) == "Rank")],
get(tour_2021_team[3])[["Rank"]], alpha = 1)
best_lambda <- cv_defense$lambda.min
plot(cv_defense)
best_defense<- glmnet(x.defense[ ,!(colnames(x.defense) == "Rank")],
get(tour_2021_team[3])[["Rank"]], alpha = 1, lambda = best_lambda)
coef(best_defense)
##goalkeeping
x.goalkeep=data.matrix(get(tour_2021_team[4])[,-c(1:2)])

cv_goalkeep <- cv.glmnet(x.goalkeep[ ,!(colnames(x.goalkeep) == "Rank")],
get(tour_2021_team[4])[["Rank"]], alpha = 1)
best_lambda <- cv_goalkeep$lambda.min
plot(cv_goalkeep)
best_goalkeep<- glmnet(x.goalkeep[ ,!(colnames(x.goalkeep) == "Rank")],
get(tour_2021_team[4])[["Rank"]], alpha = 1, lambda = best_lambda)
coef(best_goalkeep)

#ridge
##shooting
cv_shoot_r <- cv.glmnet(x.shoot, get(tour_2021_team[1])[["Rank"]], alpha = 0)
best_lambda_r  <- cv_shoot_r $lambda.min
plot(cv_shoot_r )
```

```
best_shoot_r <- glmnet(x.shoot, get(tour_2021_team[1])[["Rank"]], alpha = 0, lambda =
best_lambda_r )
coef(best_shoot_r)
##passing
cv_pass_r  <- cv.glmnet(x.pass, get(tour_2021_team[2])[["Rank"]], alpha = 0)
best_lambda_r  <- cv_pass_r $lambda.min
plot(cv_pass_r )
best_pass_r <- glmnet(x.pass, get(tour_2021_team[2])[["Rank"]], alpha = 0, lambda = best_lambda_r )
coef(best_pass_r )
##defense
cv_defense_r  <- cv.glmnet(x.defense, get(tour_2021_team[3])[["Rank"]], alpha = 0)
best_lambda_r  <- cv_defense_r $lambda.min
plot(cv_defense_r )
best_defense_r <- glmnet(x.defense, get(tour_2021_team[3])[["Rank"]], alpha = 0, lambda =
best_lambda_r )
coef(best_defense_r )
##goalkeeping
cv_goalkeep_r  <- cv.glmnet(x.goalkeep, get(tour_2021_team[4])[["Rank"]], alpha = 0)
best_lambda_r  <- cv_goalkeep_r $lambda.min
plot(cv_goalkeep_r )
best_goalkeep_r <- glmnet(x.goalkeep,  get(tour_2021_team[4])[["Rank"]], alpha = 0, lambda =
best_lambda_r )
coef(best_goalkeep_r )
# Random Forest
colnames(tour_2021_team_shooting)[colnames(tour_2021_team_shooting)=="90s"]="S90"
colnames(tour_2021_team_shooting)=gsub(" |/|:|-|%","_",colnames(tour_2021_team_shooting))
rf <- randomForest(Rank~.,tour_2021_team_shooting[,-1],
          importance=TRUE)
varImpPlot(rf, main = "Feature Importance")
rf_shooting <- rf
rf_shooting

colnames(tour_2021_team_passing)[colnames(tour_2021_team_passing)=="90s"]="S90"
colnames(tour_2021_team_passing)[colnames(tour_2021_team_passing)=="1_3"]="One_Third"
colnames(tour_2021_team_passing)=gsub(" |/|:|-|%","_",colnames(tour_2021_team_passing))
rf_passing <- randomForest(Rank~.,tour_2021_team_passing[,-1],
          importance=TRUE)
varImpPlot(rf_passing, main = "Feature Importance")
rf_passing

colnames(tour_2021_team_defense)[colnames(tour_2021_team_defense)=="90s"]="S90"
colnames(tour_2021_team_defense)[colnames(tour_2021_team_defense)=="Tkl+Int"]="Tkl_Int"
colnames(tour_2021_team_defense)=gsub(" |/|:|-|%","_",colnames(tour_2021_team_defense))
```

```
rf_defense <- randomForest(Rank~.,tour_2021_team_defense[,-1],
          importance=TRUE)
varImpPlot(rf_defense, main = "Feature Importance")
rf_defense

colSums(is.na(tour_2021_team_goalkeeping))
tour_2021_team_goalkeeping$Pos[is.na(tour_2021_team_goalkeeping$Pos)] <- "GK"
tour_2021_team_goalkeeping<-tour_2021_team_goalkeeping[-24,]
tour_2021_team_goalkeeping[is.na(tour_2021_team_goalkeeping)] <- 0
colnames(tour_2021_team_goalkeeping)=gsub(" |/|:|-
|%","_",colnames(tour_2021_team_goalkeeping))
rf_goalkeeping <- randomForest(Rank~.,tour_2021_team_goalkeeping[,-1],
              importance=TRUE)
varImpPlot(rf_goalkeeping, main = "Feature Importance")
rf_goalkeeping

#Construct Team Performances
comp=read.table(file.choose(),header=T)
attach(comp)
comp.lm=lm(Rank~Shooting+Passing+Defense+GK,data=comp)
summary(comp.lm)
```

8. BIBLIOGRAPHY

1. ABC Assignment help. 2022. *INTERNAL ANALYSIS OF CHELSEA FOOTBALL CLUB*. [online] Available at: <https://www.abcassignmenthelp.com/internal-analysis-of-chelsea-football-club> [Accessed 24 March 2022].
2. Asika, K., 2017. *How does a football club make its money?*. [online] Quora. Available at: <https://www.quora.com/How-does-a-football-club-make-its-money> [Accessed 24 March 2022].COMMERCIALIZATION STRATEGIES IN FOOTBALL
3. Campelli, M., 2022. *Quantifying the economic and social impact of football – Sustainability Report*. [online] Sustainabilityreport.com. Available at: <https://sustainabilityreport.com/2021/01/28/quantifying-the-economic-and-social-impact-of-football/> [Accessed 24 March 2022].
4. Deloitte United Kingdom. 2021. *Football Money League 2021| Deloitte UK*. [online] Available at: < https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html> [Accessed 24 March 2022].
5. Deloitte United Kingdom. 2022. *Football Money League 2022| Deloitte UK*. [online] Available at: < https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html> [Accessed 24 March 2022].
6. Deloitte United Kingdom. 2022. *Annual Review of Football Finance 2021 | Deloitte UK*. [online] Available at: <https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/annual-review-of-football-finance.html> [Accessed 19 March 2022].
7. Football-stadiums.co.uk. 2022. *How Do Football Clubs Make Money?*. [online] Available at: <https://www.football-stadiums.co.uk/articles/how-do-football-clubs-make-money/> [Accessed 24 March 2022].
8. Insurancebusinessmag.com. 2022. *Insuring red card risks at World Cup tournaments*. [online] Available at: <https://www.insurancebusinessmag.com/us/risk-management/news/insuring-red-card-risks-at-world-cup-tournaments-104751.aspx> [Accessed 24 March 2022].

9. Jr, B., 2019. *Revealed: the economic benefits of football*. [online] Arabian Business. Available at: <https://www.arabianbusiness.com/gcc/saudi-arabia/saudi-arabia-sport/410639-revealed-the-economic-benefits-of-football> [Accessed 24 March 2022].
10. Krabbenbos, T., 2013. *COMMERCIALIZATION STRATEGIES IN FOOTBALL*. pp.24-40.
11. Liu, Y., 2013. *Assessing the Long-term Economic Impacts of the World Cup as Mega-sport Event | The People, Ideas, and Things (PIT) Journal*. [online] Pitjournal.unc.edu. Available at: <http://pitjournal.unc.edu/article/assessing-long-term-economic-impacts-world-cup-mega-sport-event> [Accessed 24 March 2022].
12. Marani, M., Katul, G., Pan, W. and Parolari, A., 2021. Intensity and frequency of extreme novel epidemics. *Proceedings of the National Academy of Sciences*, 118(35).
13. Ohio University. 2022. *Athletic directors must strategically manage ethical issues in sports. Discover how Ohio University's Master of Athletic Administration prepares graduates to run successful athletic programs.*. [online] Available at: <https://onlinemasters.ohio.edu/blog/ethical-issues-in-sports/> [Accessed 24 March 2022].
14. Owoeye, O., VanderWey, M. and Pike, I., 2020. Reducing Injuries in Soccer (Football): an Umbrella Review of Best Evidence Across the Epidemiological Framework for Prevention. *Sports Medicine - Open*, 6(1).
15. Sarkar, N., 2016. *How do football clubs make money?*. [online] Sportskeeda.com. Available at: <https://www.sportskeeda.com/football/how-do-football-clubs-make-money> [Accessed 24 March 2022].
16. Sartori, A., 2021. *Football Clubs' Valuation: The European Elite 2021*. [online] Available at: <https://home.kpmg/hu/en/home/insights/2021/05/football-clubs-valuation-the-european-elite-2021.html> [Accessed 19 March 2022].
17. Stats Perform. 2022. *The challenges faced by Football League Clubs in devising and executing an overseas player identification and recruitment strategy - Stats Perform*. [online] Available at: <https://www.statsperform.com/resource/the-challenges-faced-by-football-league-clubs-in-devising-and-executing-an-overseas-player-identification-and-recruitment-strategy/> [Accessed 24 March 2022].
18. Sports Shorts. 2022. *Football Behind Bars: Reputation Management Issues for Clubs*. [online] Available at: <https://www.sports.legal/2017/03/football-behind-bars-reputation-management-issues-for-clubs/> [Accessed 24 March 2022].
19. Whelan, A., 2022. *How Can Football Clubs Mitigate Reputational Risks When Evaluating Sponsors?*. [online] Duff & Phelps. Available at: <https://www.kroll.com/en/insights/publications/sport/football-clubs-mitigate-reputational-risks> [Accessed 24 March 2022].