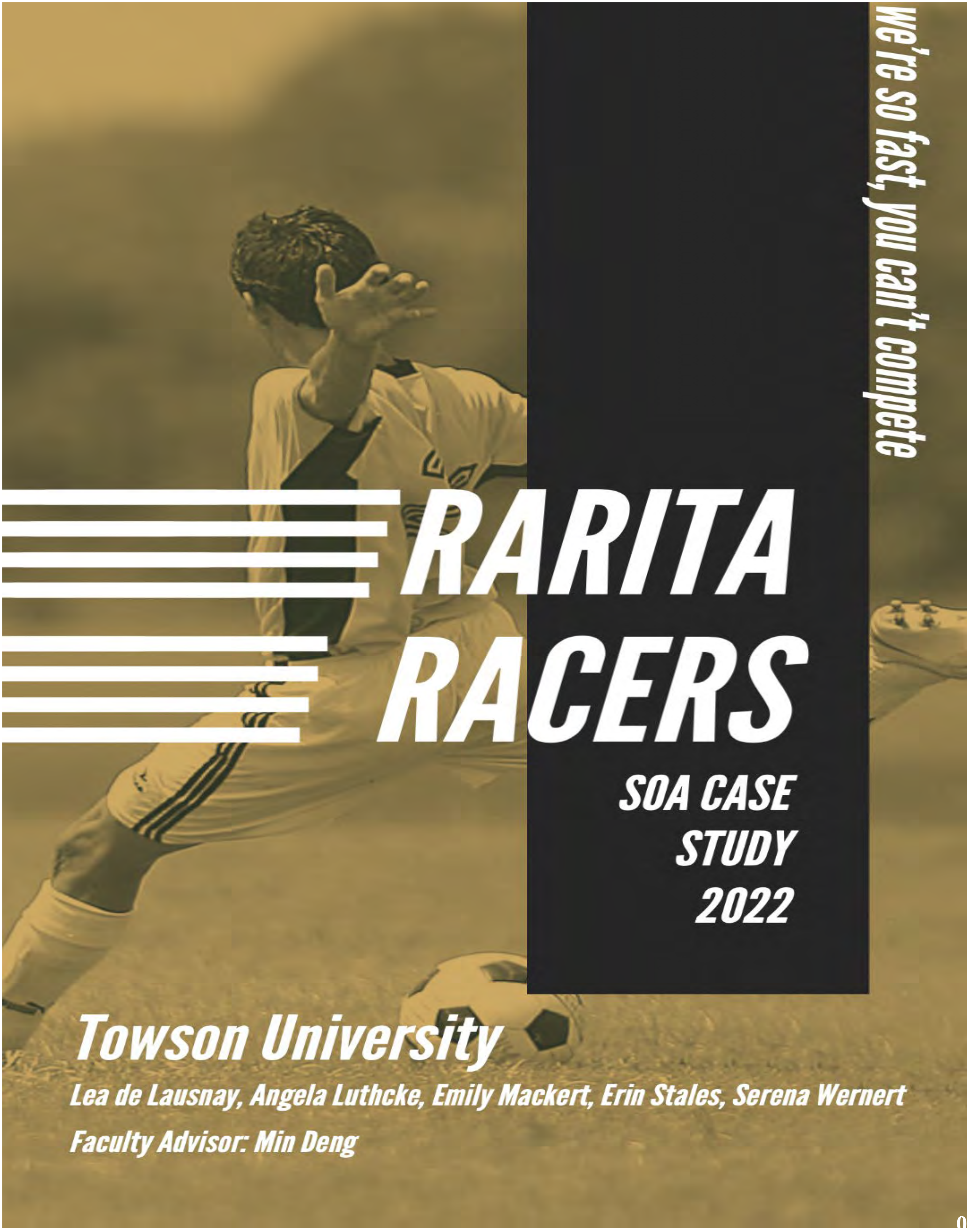


we're so fast, you can't compete

RARITA RACERS

SOA CASE STUDY 2022

Towson University

Lea de Lausnay, Angela Luthcke, Emily Mackert, Erin Stales, Serena Wernert

Faculty Advisor: Min Deng

Table of Contents

1. Executive Summary

Rarita is not currently involved with International Football; in the interest of reaping economic and political benefits, the nation seeks to create a national soccer team. Ultimately, Rarita would like for the national soccer team to qualify for the international FSA league within five years.

Three regions comprise the nation: the East, Central, and West. We are also given economic data for Rarita and 21 other countries. With this information, we compare Rarita's economic standing with those of other countries in the FSA league. Additional data provide statistics on player demographics and performance, which is analyzed to select an optimal team for Rarita. The data provided was complex due to missing and inconsistent data. These limitations, though a burden on the accuracy of projected models, allow us to come to reasonable assumptions, which become the framework for team selection.

Initial investigation of the provided information along with additional research led to analysis of risk and risk mitigation. These risks play a significant role in the economic goals and decisions. From here, we analyzed the data using principal component analysis, Excel, and R code to create models that forecast future trends in both the economy and revenue/expense allocations. Using general linear models, we found that from 2021 to 2031, revenues can be expected to increase by 25.7%, while expenses can be expected to increase by 71.68%.

Using tournament data, we modeled rank using a zero-truncated poisson model. From this model, we created two team rosters with different cost levels. Then we ran the low cost roster through our model to project the outcome of the next tournament. We found that this team would rank first in the tournament, and we conclude that it will likely continue to improve in the coming years with correct planning.

2. Introduction

2.1 Background

Rarita has been overlooked within the FSA League due to lack of both organization and competitiveness. We have been hired by the Commissioner of Sport for Rarita, Hammessi Bayes, to form a competitive national team for Rarita. The objective for this team is to achieve success within ten years by positively influencing Rarita economically, politically, and competitively.

Despite not having a national team, Rarita has had many successful soccer players play for different winning teams of the FSA. Rarita's players have previously been divided between East, Central, and West provinces of the nation. Our official team includes a mixture of players from different nations and provinces of Rarita.

We have been given an initial amount of 995 million Doubloons, Rarita's national currency. Moving forward we utilize non-governmental funding sources through sponsorship and investments.

2.2 Objective

We observe countries that won the FSA experience a great financial impact to their economy. To experience the same economic influence, we consider data limitations, assumptions, and risks when also trying to achieve our goal. Our objective is to rank within the top ten FSA members within five years and win the FSA within ten years.

3. Data Limitations and Assumptions

Incomplete data and limited historical data require data assumptions, which impacts overall analysis, including projections over 10 years. In the following section we discuss missing data and the considerations to reform these situations.

3.1 Data and Data Limitation

Data Sheet	Description	Usage
Player Data	Contains league, tournament, and salary data for players in 2020 and 2021.	Main use of the data is seen in team creation and competitive analysis.
Economic Data	Contains economic demographic data, spot rates, and inflation/deflation rates for Rarita as well as GDP for other countries.	Main use from the impact of the national soccer team on the economy and government in Rarita.
Team Management Data	Contains revenue, expense, attendance, and social media data for national teams.	Main use in projection of costs, budgets, and revenues for the national team.

In multivariate analysis, complete data is very important for constructing an accurate and representative model - player data was incomplete and some values lacked reasonability. Estimation of missing values is thus limited to accuracy and reach of our model.

Player data is limited. Historical data is vital for projections - we are provided only two years of player data for league, tournament, and salary. The data is also inconsistent. This increases uncertainty when projecting player performance over ten years, as requested by the client.

There is a lack of breakdown in revenue and expenses. Ambiguity surrounding costs, where revenue is allocated, and the current resources Rarita already has during the implementation stage causes less precise estimations of expenses and profit.

Economic data is not available for 2021. This can affect our predictions since it does not take into account the impact of the pandemic.

There is data for inflation rates for the last 20 years but only ten years of data for GDP, population, and gross income for Rarita. For other countries, we only have five years (2016-2020) of data. This is a relatively small sample size to trend rates and economic factors for ten years.

3.2 Preliminary Data Analysis

An initial pass over the data was conducted to identify missing data and patterns that may exist. We also detailed our team selection criteria based on tournament national teams makeup. Any missing data was imputed through the HMISC process (See Appendix A).

3.3 Assumptions

Assumptions are identified to help develop accurate results. We group assumptions into four categories: tournament, player, data, and other.

3.3.1 Tournament Assumptions

Tournament-related assumptions include that players may compete representing any national team, despite citizenship status, and Rarita is the only additional team for the upcoming tournament. These are assumed because players in the data wouldn’t be provided if they weren’t available for choice, and there is no information regarding competition for players in the data.

3.3.2 Player Assumptions

Player-related assumptions include that if there is a salary for a player in 2020, but not for 2021, then that player is a free-agent, meaning it is not necessary to pay the extra 10% of their salary when recruiting them.

We assume that players from the league player data are the only players available for choice because tournament data exists from specific tournament(s), while league data shows active categories.

We assume that players may play for the country’s team starting at age 16 because it is Rarita’s policy.

3.3.3 Data Assumptions

Data-related assumptions include that social media and attendance data (located with revenues and expense data) are from 2020. The revenues and expense data is from 2016-2020. Hence, the social media and attendance data is from the most recent year.

We assume that some data needs to be adjusted, such as negative statistics being converted to zero due to lack of logical sense. For example, it would be impossible to score a negative number of goals.

3.3.4 Other Assumptions

Other assumptions include that the Rarita Racers have a stadium previously built because, based on the allotted budget to the team, there is not enough money to cover player salaries, team expenses, etc., along with building a stadium and the associated expenses.

The date of our official team formation is 1/1/2022. There is no information regarding the official team formation date, so we chose the first day of the next year, allowing the remainder of the current year to analyze data and go through necessary procedures.

4. Risk and Risk Mitigation

The development of a national football team comes with risks that must be recognized to achieve our goal. We produced a risk analysis to summarize key risks below and reported additional risk through a risk categorization and definition tool (see Appendix B)

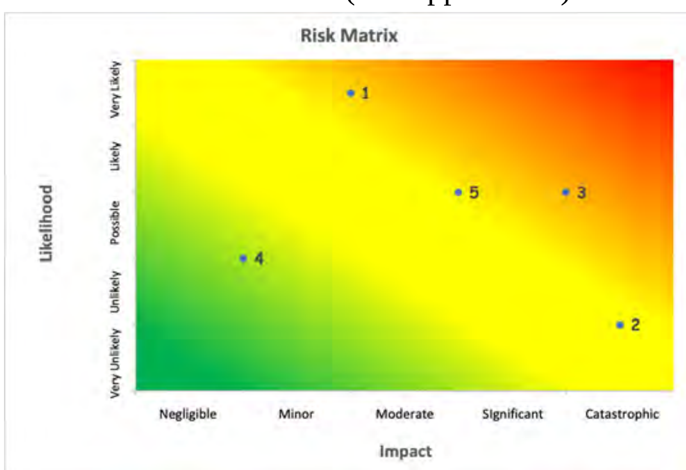


Figure 1: Risk Matrix

There are five key risks, which are assessed in the above risk matrix:

1. **Player Injury:** Majority of player injuries are minor and last less than a month. To mitigate this risk, we analyze player age, past injuries, and invest in trainers to prevent injury and optimize recovery.

2. **Natural and Man-made Disasters:** Unexpected natural disasters or health related pandemics could prevent players from competing or fans from attending games. Ending a season early could dramatically impact revenues and decrease profits.

3. **Unexpected player performance:** We could overestimate or underestimate the skill level of a player, which could obstruct the ability to rank in the FSA. We must spend more time studying the players before official recruitment.

4. **Changes in player contract:** We must account for increased salaries per renegotiations, increased experience, and/or leasing players. It could be necessary to alter our team to minimize expenses. We need to look at past data to understand common contract changes along with salary.

5. **Changes in inflation:** Catastrophic events could impact inflation rates, leading to higher or lower rates than predicted. Changes can cause different profit margins and affect economic standing. To mitigate this risk, we save as much money as possible to plan for the best and worst case scenario.

5. Team Selection Criteria

5.1 Analysis Methodology

The aim of the team selection process is to create rosters optimizing performance statistics within budget. Initial constraints were based on the national teams' rosters from the 2020-2021 tournament. The driving forces include: number of players per position, age, and salary cap. Using the interquartile range we determined 6-10 defense, 4-8 forwards, 1-2 goalkeepers, and 10-12 midfield, along with an age range of 25-30.

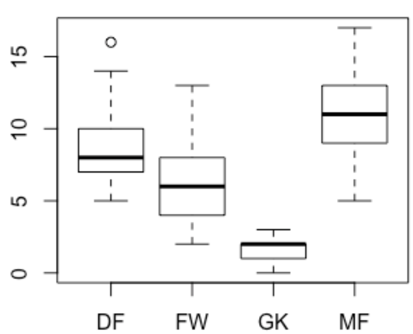


Figure 2: Number of Players by Position Box Plot

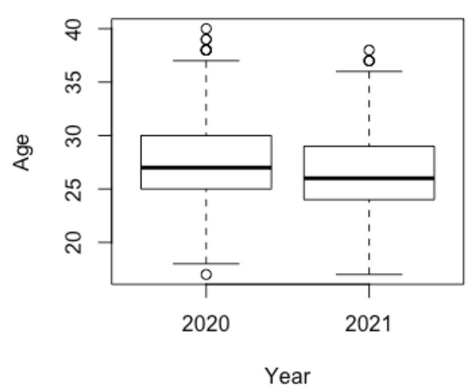


Figure 3: Range for Age Box Plot

5.2 Team Program Design

5.2a Principal Component Analysis for Multicollinearity

We examine correlations between all player parameters. We split the data into 2 groups: goalie data and player data. It is evident that the data is highly correlated and parameter reduction would be beneficial.

Figure 4: Player Parameters Correlation Plot

Figure 5: Goalie Player Parameters Correlation Plot

Principle component analysis (PCA) is well-suited for correlated data and effectively collapses a set of variables into fewer uncorrelated linear combinations of original parameters [11]. PCA utilizes orthogonal transformation to adapt a wide range of potentially correlated variables during the observation into a list of linearly uncorrelated variables [6]. Accounting for most of the variation, we set a threshold of 80% for new component creation.

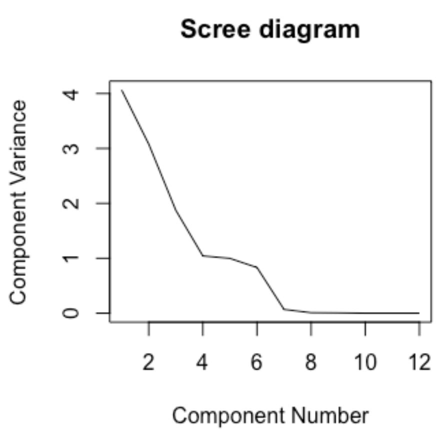


Figure 6: Goalie Scree Diagram

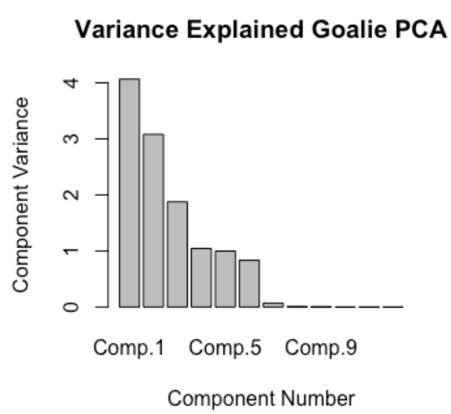


Figure 7: Goalie Variance Diagram

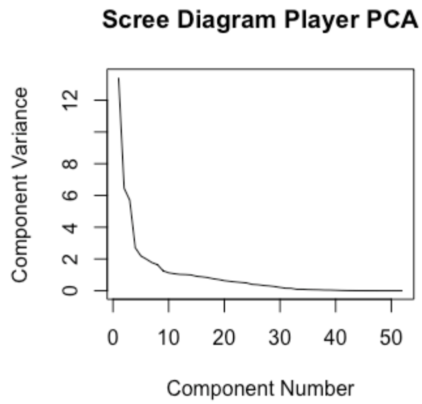


Figure 8: Player Scree Diagram

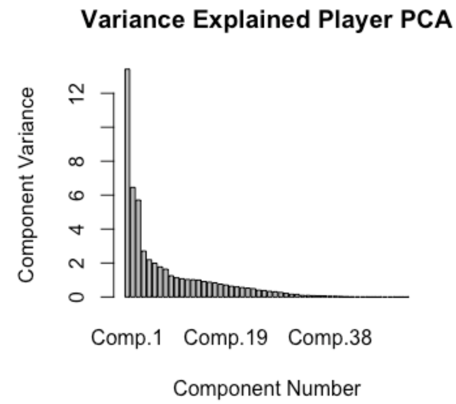


Figure 9: Player Variance Diagram

As seen above, we can create new variables. Goalie and player data now consists of 4 and 12 components, respectively. This has established essential indicators for selecting the best players. (See Appendix C for PCA output)

5.2b Model Creation with Zero-Truncated Poisson

Using 2021 tournament data, we modeled rank as a discrete variable by aggregating the team's principal components as dependent variables and assigning rank as independent variable. The discrete models that were considered include: negative binomial regression, poisson regression, and zero-truncated negative binomial and poisson regression.

In order to rank and select a model, we analyzed the AIC, Log Likelihood, and degrees of freedom for each model. Additionally, we examined multicollinearity through VIF and finalized the models to contain only significant variables with $\alpha = .05$ (See Appendix C). We proceeded with zero-truncated poisson.

Goalie:

	AIC	**LL**	**df**
Neg Binomial	130.8678	-60.43392	4
Poisson	128.8677	-60.43385	5
ZT Poisson	128.319	-61.15949	3
ZT Neg Binomial	130.719	-60.35949	5

Player Model:

	AIC	LL	df
Neg Binomial	152.3081	-68.15405	8
Poisson	151.2846	-68.6423	7
ZT Poisson	151.0514	-68.52568	7
ZT Neg Binomial	152.0238	-68.01192	8

Goalie Model:

$$log(\lambda) = .69029 + 1.23911\ G1 + .13980\ G2$$

Player Model:

$$log(\lambda) = 2.46859 - .48251C2 - .52429C4 - .53898C8 + .31129C9 - .68218C10 + .78098C14$$

5.2c Optimization for Final Rooster selection

We ran optimization based on the criteria in section 5.1 to create two team rosters: low cost, utilizing 20% of budget, and high cost, utilizing 50% of budget.

	Player	Pos	PredictRank	Salary
1	S. Nachum	DF	1.693744	1969000
2	U. Kaahwa	DF	1.570861	7953000
3	C. Nalwadda	DF	1.510875	4543000
4	H. Azizi	DF	1.449559	5870000
5	B. Lindberg	DF	1.299351	25096500
6	R. Bogere	DF	1.272461	18628500
7	D. Lehner	MF	1.265096	6730000
8	H. KoroÅ...Âjec	DF	1.216709	38852000
9	G. Vidal	FW	1.178760	5362500
10	K. Driciru	MF	1.157814	36674000
11	C. Ji	MF	1.138882	34347500
12	A. Perez	FW	1.127549	7660000
13	U. Katushabe	MF	1.116438	30525000
14	K. Kazlo?	FW	1.077029	5920000
15	W. Mbaziira	DF	1.072074	4000000
16	L. Tarigan	FW	1.058373	6400000
17	P. Otoo	MF	1.035250	31086000
18	T. Kamugisha	GK	1.033467	4196500
19	A. Khainza	GK	1.028298	1639000
20	J. Namirembe	MF	1.027846	8330000
21	S. Kisakye	DF	1.016940	42262000
22	X. Takagi	DF	1.006038	4810000
23	J. Bah	MF	1.003421	20394000
24	R. Nabwire	FW	1.000159	28380000
25	P. Martin	FW	1.000000	33110000
26	K. Adong	FW	1.000000	28006000
27	F. Akongo	MF	1.000000	1958000

Figure 10: Low end budget roster

	Player	Pos	PredictRank	Salary
1	Q. Lange	DF	1.913618	5698000
2	D. Mo	DF	1.912058	3663000
3	R. Hamed	MF	1.862431	99000
4	U. Sun	MF	1.798094	4158000
5	E. Vasav	DF	1.755284	6479000
6	A. Afful	MF	1.708537	1738000
7	S. Nachum	DF	1.693744	1969000
8	H. Nuwahereza	FW	1.681158	2090000
9	H. Makumbi	FW	1.598193	7430000
10	U. Kaahwa	DF	1.570861	7953000
11	C. Nalwadda	DF	1.510875	4543000
12	H. Azizi	DF	1.449559	5870000
13	A. Katusiime	MF	1.378948	9559000
14	I. Saha	FW	1.353650	8380000
15	R. Bogere	DF	1.272461	18628500
16	D. Lehner	MF	1.265096	6730000
17	G. Vidal	FW	1.178760	5362500
18	A. Perez	FW	1.127549	7660000
19	K. Kazlo?	FW	1.077029	5920000
20	W. Mbaziira	DF	1.072074	4000000
21	L. Tarigan	FW	1.058373	6400000
22	A. Khainza	GK	1.028298	1639000
23	J. Namirembe	MF	1.027846	8330000
24	X. Takagi	DF	1.006038	4810000
25	J. Bah	MF	1.003421	20394000
26	K. Adong	FW	1.000000	28006000
27	F. Akongo	MF	1.000000	1958000

Figure 11: High end budget roster

5.3 2022 Tournament Ranking

After adding our low end roster through the model against the current nations, we would place first for the 2022 tournament.

	Nation	GoalieRank	PlayerRank	OverallRank
1	Bernepamar	9.383393	6.622725	8.003059
2	Byasier Pujan	10.759125	10.448585	10.603855
3	Djipines	11.837429	15.360758	13.599093
4	Eastern Niasland	23.161872	16.078114	19.619993
5	Eastern Sleboube	19.352589	18.599662	18.976126
6	Esia	10.882507	9.022471	9.952489
7	Galamily	6.919535	11.146225	9.032880
8	Giumle Lizeibon	8.530426	16.356289	12.443358
9	Greri Landmoslands	13.773121	10.015144	11.894132
10	Ledian	28.927056	17.366136	23.146596
11	Leoneku Guidisia	13.951785	18.766845	16.359315
12	Manlisgamncent	11.352275	8.133093	9.742684
13	Mico	8.710719	2.565182	5.637951
14	New Uwi	20.631267	18.096439	19.363853
15	Nganion	3.841291	7.337166	5.589229
16	Ngoque Blicri	15.825693	22.776385	19.301039
17	Nkasland Cronestan	16.308407	12.850262	14.579335
18	People's Land of Maneau	3.367994	6.306295	4.837144
19	Quewenia	9.944772	10.015102	9.979937
20	Sobianitedrucy	5.359708	4.508951	4.934330
21	Southern Ristan	6.021551	7.240236	6.630893
22	Varijitri Isles	23.430698	26.924989	25.177844
23	Xikong	8.726784	10.273458	9.500121
24	Rarita	1.012716	1.002616	1.007666

Figure 12: 2022 Tournament standings using ZT Poisson Regression

6. Economic Impact

6.1 Funding Sources

We are given a lump sum of 995,000,000 Doubloons (∂) to fund Rarita's national team. We explored non-governmental funding ideas to increase our team's revenues.

Type of Advertising	**Justification**	**Rarita's Revenue**
Luxury Box Seating	Sports Stadiums sell luxury box seating to companies and high paying customers to attract an international audience. As teams gain popularity within tournaments, they have an increase in revenue (Shapiro et al., 2017).	We assume Rarita's stadium will sell 3-year contracts for 65 boxes, 17 home games, and a base rate of ∂2567.25 per game and box. Every 3 years the price will increase by ∂1026.90.

Type of Advertising	Justification	Rarita's Revenue
Rarita's Stadium Name	It's common for stadiums to enter 20 year contracts for millions of dollars (Paden, J., 2021). We used stadium naming rights revenue data for sports organizations to predict Rarita's yearly funding from the company (ESPN Internet Ventures., 2021).	We assume Rarita will enter into a 20 year contract for ∂55485255.42, guaranteeing a yearly revenue of ∂2774262.77.
Jersey Advertising	On average, teams made 27916430.00 euros in a year (Gaines, C., 2012).	Will make ∂31852646.63 from shirt sponsors.
Local Vendor Partners	Rarita's stadium will rent vendor spaces to local restaurants of Rarita to sell food. Other stadiums found "...fully equipped concession is expected to cost about $6,000 to $45,000…" (Joy Nwokoro, 2021). We would undercharge a flat fee of $5000 to only cover utility costs and consider the profit as community outreach.	We assume Rarita's stadium will have 20 vendors, charging ∂41076 for a yearly contract.

6.2 Rarita's Expenses and Revenues

We only considered the nations who participated in the 2020 FSA tournament to predict Rarita's historical data on a FSA Tournament Basis. We use this information to forecast the revenues and expenses for 10 years.

We are given expenses and GDP for every nation from 2016-2020, excluding Eastern Sleboube. Taking every nation's total expenses, we created a general linear model for each year using R studio (see Appendix G1). We predicted Rarita's expenses per capita for 2017-2020 using Rarita's GDP. Taking these predicted expenses, we forecasted Rarita's expenses for 10 years (see Appendix G2). We conducted sensitivity analyses on these forecasted expenses (Appendix G2).

To calculate Rarita's revenues on a tournament basis, we used a general linear model with independent variables of nations GDP, Social Media, and Average League Attendance for each year between 2016-2020. Afterwards, we forecasted the revenues per capita for 2021-2031 and used the forecasted population data to calculate the revenues in Doubloons. After adding our other funding sources, we converted our revenues back to per capita (see Appendix G3 and G4).

7. Implementation and Analysis Considerations

(template.net)

8. Conclusion and Recommendations

After extensive analysis and in-depth modeling, we conclude that the optimal team consists of the players that appear on the rosters in section 5.2 Team Program Design. These players were chosen based on performance and rank analysis. We assessed any possible correlations between variables that indicate the effectiveness of player performance, such as play time and salary. We addressed these concerns through principal component analysis.

8.1 Further Considerations

After a year of tournament play, we advise reevaluating the model fit. With more historic data, the model selection process could be improved to calculate more relevant standings for ten years.

We undervalued funding sources because we are a new team. In the future, we can gain more funding as recognition by other nations increases. We should reassess funding sources annually to optimize profit.

In the future, other variables may be used to project revenue and expenses. We have forecasted all economic data (see Appendix F) to analyze effects of any overlooked correlations.

References

[1] Alwell, K. (2020). Analyzing Competitive Balance in Professional Sport.

[2] ESPN Internet Ventures. (2021, September 29). *Sports Business*. ESPN. Retrieved March 10, 2022, from https://www.espn.com/sportsbusiness/s/stadiumnames.html

[3] *Free design templates | adobe creative cloud express*. (n.d.). Retrieved March 17, 2022, from https://www.adobe.com/express/discover/templates

[4] Gaines, C. (2012, July 25). *Ads on sports jerseys might be worth more than you think, and that's a good thing*. Business Insider. Retrieved March 10, 2022, from https://www.businessinsider.com/ads-on-sports-jerseys-might-be-worth-more-than-you-think-and-thats-a-good-thing-2012-7

[5] Joy Nwokoro, J. (2021, May 26). *Estimated Cost of Building a Concession Stand in 2022*. ProfitableVenture. Retrieved March 10, 2022, from https://www.profitableventure.com/cost-build-concession-stand/

[6] Li, W., & Sang, G. (2018, August). Principal Component Analysis on Football Competitions via Linear Model in China Football Association Super League Tournament. In *2018 International Conference on Management, Economics, Education and Social Sciences (MEESS 2018)* (pp. 57-61). Atlantis Press.

[7] Lopez, M. J., Matthews, G. J., & Baumer, B. S. (2018). How often does the best team win? A unified approach to understanding randomness in North American sport. The Annals of Applied Statistics, 12(4), 2483-2516.

[8] Paden, J. (2021, December 19). *The business of naming stadiums*. AmadorValleyToday. Retrieved March 17, 2022, from https://www.amadorvalleytoday.org/17148/sports/the-business-of-naming-stadiums/

[9] Pérez-Toledano, M. Á., Rodriguez, F. J., García-Rubio, J., & Ibañez, S. J. (2019). Players' selection for basketball teams, through Performance Index Rating, using multiobjective evolutionary algorithms. *PloS one*, *14*(9), e0221258.

[10] Pettinger, T. (2018, January 8). *Effects of a falling inflation rate*. Economics Help. Retrieved February 22, 2022, from https://www.economicshelp.org/blog/357/inflation/effects-of-a-falling-inflation-rate/#:~:text=A%20falling%20rate%20of%20inflation%20means%20that%20prices%20will%20be,competitive%20increasing%20exports%20and%20growth

[11] Shah, P., & Patel, M. N. (2018). Ranking the cricket captains using principal component

analysis. *Int. J. Physiol. Nutr. Phys. Educ, 3*(2), 477-483.

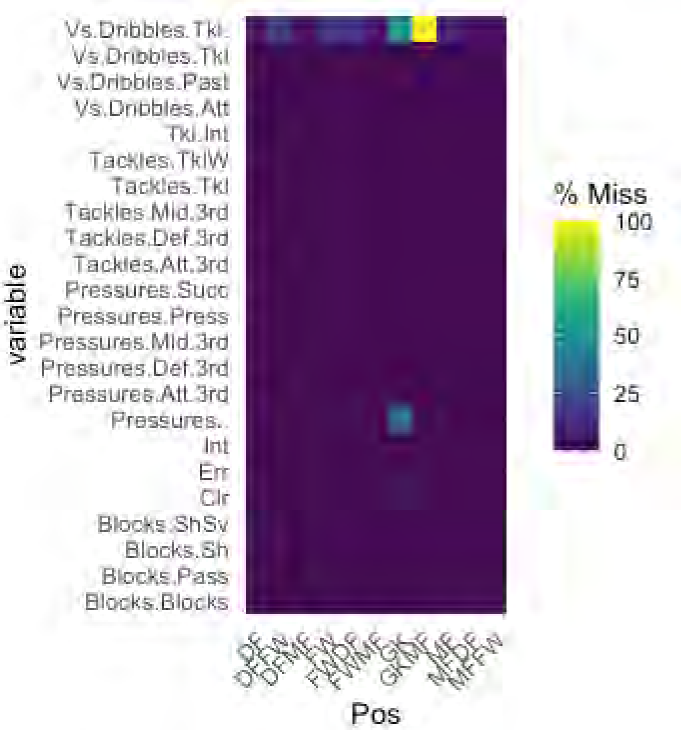
[12] Shapiro, S. L., DeSchriver, T., & Rascher, D. A. (2017, February 28). *Factors affecting the price of luxury suites in major North American Sports Facilities*. SSRN. Retrieved March 10, 2022, from https://ssrn.com/abstract=2924159

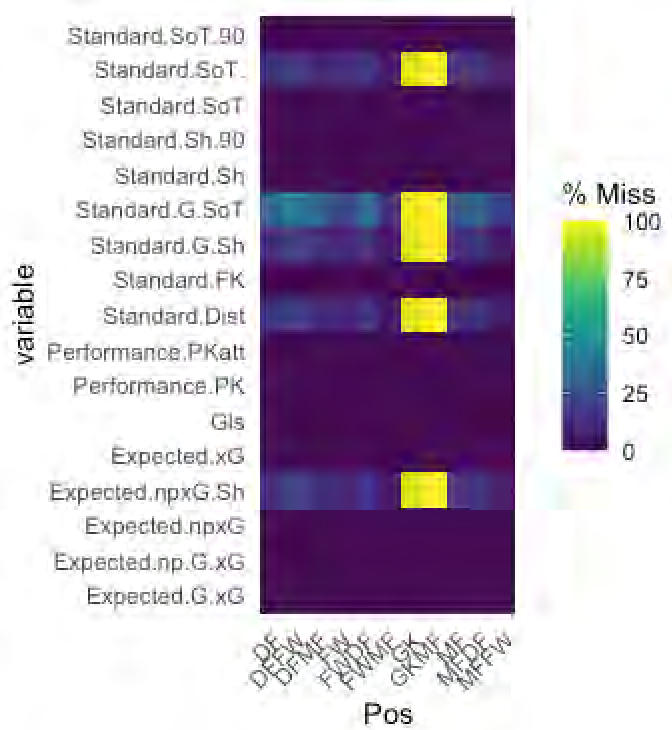
[13] Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J., ... & Grossniklaus, M. (2017). How to make sense of team sport data: From acquisition to data modeling and research aspects. Data, 2(1), 2.

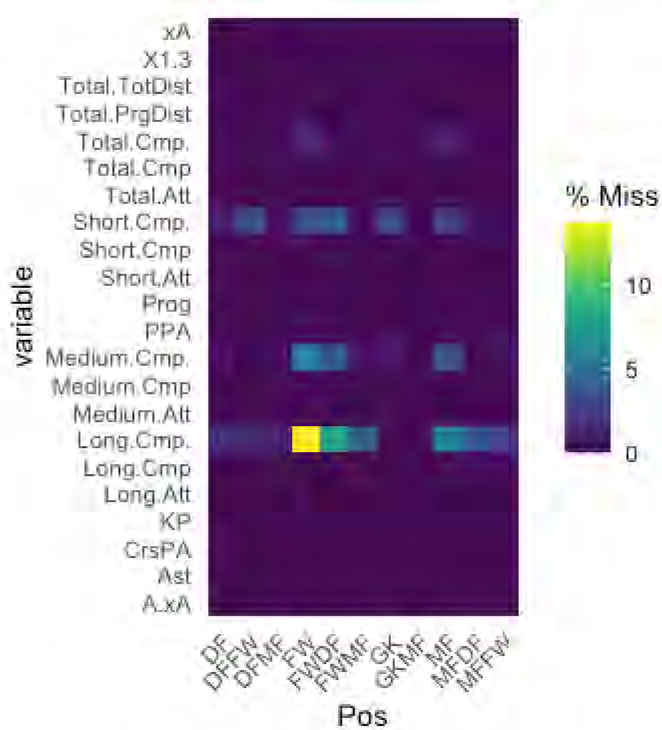
Appendices

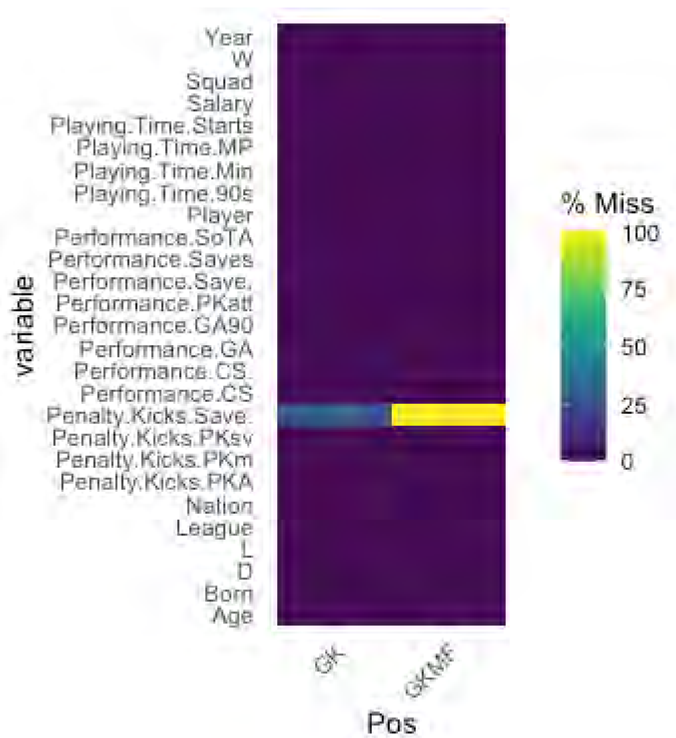
Appendix A: Data Limitations and Assumptions

A1: Missing Data by Player Position









A2: Imputing Missing Data

```
Multiple Imputation using Bootstrap and PMM

aregImpute(formula = ~Standard.SoT. + Standard.G.Sh + Standard.Dist +
    Standard.G.SoT + Expected.npxG.Sh + Total.Cmp. + Short.Cmp. +
    Medium.Cmp. + Long.Cmp. + Vs.Dribbles.Tkl. + Pressures..,
    data = df, n.impute = 5)

n: 5554        p: 11    Imputations: 5        nk: 3

Number of NAs:
   Standard.SoT.    Standard.G.Sh    Standard.Dist    Standard.G.SoT Expected.npxG.Sh       Total.Cmp.
            1031             1031             1031              1768             1031               31
      Short.Cmp.      Medium.Cmp.       Long.Cmp. Vs.Dribbles.Tkl.      Pressures..
              89              105              246               653              195

                 type d.f.
Standard.SoT.       s    2
Standard.G.Sh       s    2
Standard.Dist       s    2
Standard.G.SoT      s    2
Expected.npxG.Sh    s    2
Total.Cmp.          s    2
Short.Cmp.          s    2
Medium.Cmp.         s    2
Long.Cmp.           s    2
Vs.Dribbles.Tkl.    s    2
Pressures..         s    1

Transformation of Target Variables Forced to be Linear

R-squares for Predicting Non-Missing Values for Each Variable
Using Last Imputations of Predictors
   Standard.SoT.    Standard.G.Sh    Standard.Dist    Standard.G.SoT Expected.npxG.Sh       Total.Cmp.
           0.345            0.618            0.276             0.554            0.292            0.811
      Short.Cmp.      Medium.Cmp.       Long.Cmp. Vs.Dribbles.Tkl.      Pressures..
           0.554            0.529            0.510             0.223            0.181
```

A3: Major League Soccer Descriptive Statistics

Shooting

Gls		*Sh*		*SoT*		*SoT%*	
Mean	**1.18795**	**Mean**	**10.4508**	**Mean**	**3.68281**	**Mean**	**31.3690**
Standard Deviation	**2.07928**	**Standard Deviation**	**12.6946**	**Standard Deviation**	**5.20682**	**Standard Deviation**	**23.2653**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**14**	**Maximum**	**74**	**Maximum**	**34**	**Maximum**	**100**

Sh/90		*SoT/90*		*G/Sh*		*G/SoT*	
Mean	**1.19395**	**Mean**	**0.43195**	**Mean**	**0.09005**	**Mean**	**0.27453**
Standard Deviation	**1.36594**	**Standard Deviation**	**0.79470**	**Standard Deviation**	**0.14154**	**Standard Deviation**	**0.28191**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**12.27**	**Maximum**	**11.25**	**Maximum**	**1**	**Maximum**	**1**

FK		*PK*		*PKatt*		*xG*	
Mean	**0.41409**	**Mean**	**0.08516**	**Mean**	**0.11453**	**Mean**	**1.19427**
Standard Deviation	**1.313518327**	**Standard Deviation**	**0.418435823**	**Standard Deviation**	**0.483653497**	**Standard Deviation**	**1.788397107**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**12**	**Maximum**	**5**	**Maximum**	**6**	**Maximum**	**11.8**

npxG		*npxG/Sh*		*G-xG*		*np:G-xG*	
Mean	**1.104845815**	**Mean**	**0.096557377**	**Mean**	**-0.006314244**	**Mean**	**-0.0020558**
Standard Deviation	**1.618384275**	**Standard Deviation**	**0.063662452**	**Standard Deviation**	**0.928600187**	**Standard Deviation**	**0.925867734**
Minimum	**0**	**Minimum**	**0.01**	**Minimum**	**-4.8**	**Minimum**	**-4.8**

Maximum	**11**	**Maximum**	**0.44**	**Maximum**	**4.5**	**Maximum**	**4.5**

Passing

Cmp		***Att***		***Cmp%***		***TotDist***		***PrgDist***	
Mean	**338.242**	**Mean**	**421.033**	**Mean**	**77.9397**	**Mean**	**6831.44**	**Mean**	**2290.05**
Standard Deviation	**293.320**	**Standard Deviation**	**352.941**	**Standard Deviation**	**10.6011**	**Standard Deviation**	**6215.16**	**Standard Deviation**	**2331.92**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**1298**	**Maximum**	**1580**	**Maximum**	**100**	**Maximum**	**30835**	**Maximum**	**12689**

Cmp		***Att***		***Cmp%***		***Cmp***		***Att***	
Mean	**129.556**	**Mean**	**147.588**	**Mean**	**87.3687**	**Mean**	**147.883**	**Mean**	**170.643**
Standard Deviation	**114.841**	**Standard Deviation**	**129.147**	**Standard Deviation**	**8.70798**	**Standard Deviation**	**138.671**	**Standard Deviation**	**154.756**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**608**	**Maximum**	**699**	**Maximum**	**100**	**Maximum**	**681**	**Maximum**	**723**

Cmp%		***Cmp***		***Att***		***Cmp%***		***Ast***	
Mean	**84.2705**	**Mean**	**55.571**	**Mean**	**88.0337**	**Mean**	**61.3495**	**Mean**	**0.84287**
Standard Deviation	**11.3706**	**Standard Deviation**	**59.8667**	**Standard Deviation**	**92.0533**	**Standard Deviation**	**18.004**	**Standard Deviation**	**1.3817**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**100**	**Maximum**	**331**	**Maximum**	**488**	**Maximum**	**100**	**Maximum**	**9**

xA		***A-xA***		***KP***		***44564***		***PPA***	
Mean	**0.8187**	**Mean**	**0.02408**	**Mean**	**7.6989**	**Mean**	**25.97**	**Mean**	**6.6828**

Standard Deviation	**1.1896**	**Standard Deviation**	**0.75999**	**Standard Deviation**	**10.072**	**Standard Deviation**	**27.916**	**Standard Deviation**	**9.4177**
Minimum	**0**	**Minimum**	**-2.6**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**8.1**	**Maximum**	**3.8**	**Maximum**	**66**	**Maximum**	**148**	**Maximum**	**82**

CrsPA		***Prog***	
Mean	**1.773**	**Mean**	**28.967**
Standard Deviation	**3.148**	**Standard Deviation**	**30.827**
Minimum	**0**	**Minimum**	**0**
Maximum	**26**	**Maximum**	**209**

Defense

Tkl		***TklW***		***Def 3rd***		***Mid 3rd***		***Att 3rd***	
Mean	**15.474**	**Mean**	**9.7635**	**Mean**	**7.6622**	**Mean**	**5.790**	**Mean**	**2.0220**
Standard Deviation	**15.491**	**Standard Deviation**	**10.071**	**Standard Deviation**	**8.8174**	**Standard Deviation**	**6.2205**	**Standard Deviation**	**2.5863**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**97**	**Maximum**	**64**	**Maximum**	**49**	**Maximum**	**37**	**Maximum**	**17**

Tkl		***Att***		***Tkl%***		***Past***		***Press***	
Mean	**5.0396**	**Mean**	**13.914**	**Mean**	**35.029**	**Mean**	**8.8751**	**Mean**	**120.95**
Standard Deviation	**5.6318**	**Standard Deviation**	**13.881**	**Standard Deviation**	**22.934**	**Standard Deviation**	**9.1022**	**Standard Deviation**	**104.96**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**29**	**Maximum**	**81**	**Maximum**	**100**	**Maximum**	**52**	**Maximum**	**598**

Succ		***%***		***Def 3rd***		***Mid 3rd***		***Att 3rd***	

Mean	**36.587**	**Mean**	**29.717**	**Mean**	**41.324**	**Mean**	**52.795**	**Mean**	**26.835**
Standard Deviation	**32.648**	**Standard Deviation**	**13.134**	**Standard Deviation**	**41.692**	**Standard Deviation**	**49.880**	**Standard Deviation**	**31.232**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**186**	**Maximum**	**100**	**Maximum**	**267**	**Maximum**	**291**	**Maximum**	**186**

Blocks		*Sh*		*ShSv*		*Pass*		*Int*	
Mean	**13.565**	**Mean**	**2.9236**	**Mean**	**0.0543**	**Mean**	**10.641**	**Mean**	**5.2790**
Standard Deviation	**12.55946691**	**Standard Deviation**	**4.411553909**	**Standard Deviation**	**0.233230891**	**Standard Deviation**	**10.15907126**	**Standard Deviation**	**5.965709578**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**70**	**Maximum**	**30**	**Maximum**	**2**	**Maximum**	**62**	**Maximum**	**38**

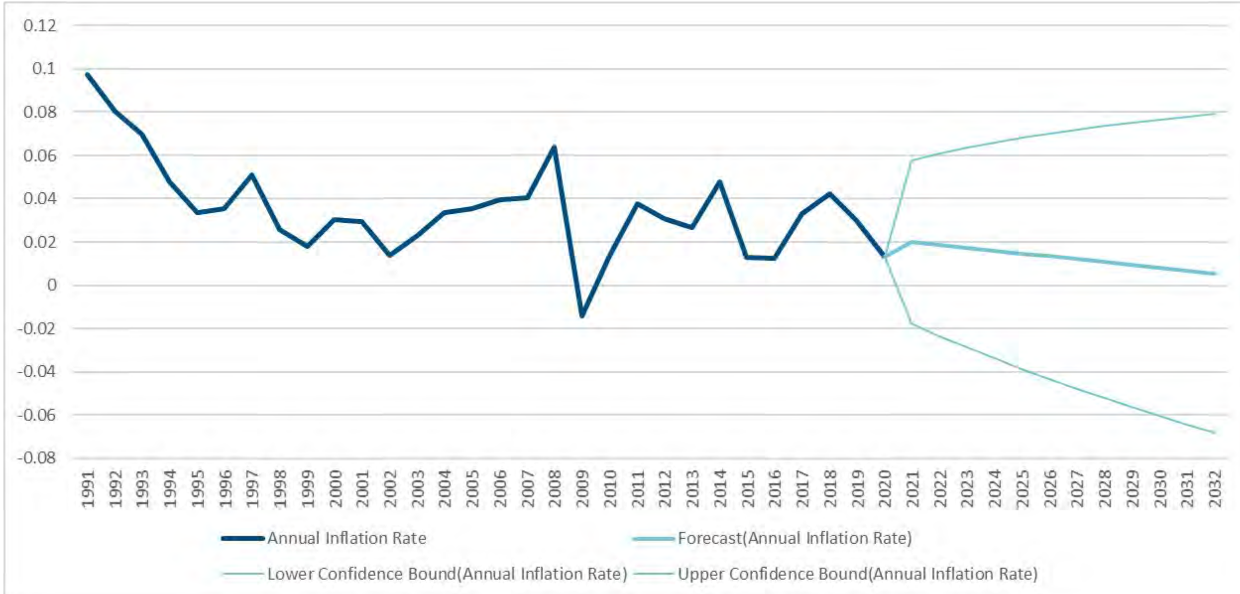
Tkl+Int		*Clr*		*Err*	
Mean	**20.753**	**Mean**	**18.32**	**Mean**	**0.2114**
Standard Deviation	**20.542**	**Standard Deviation**	**26.807**	**Standard Deviation**	**0.5389**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**116**	**Maximum**	**149**	**Maximum**	**4**

Goal Keeping

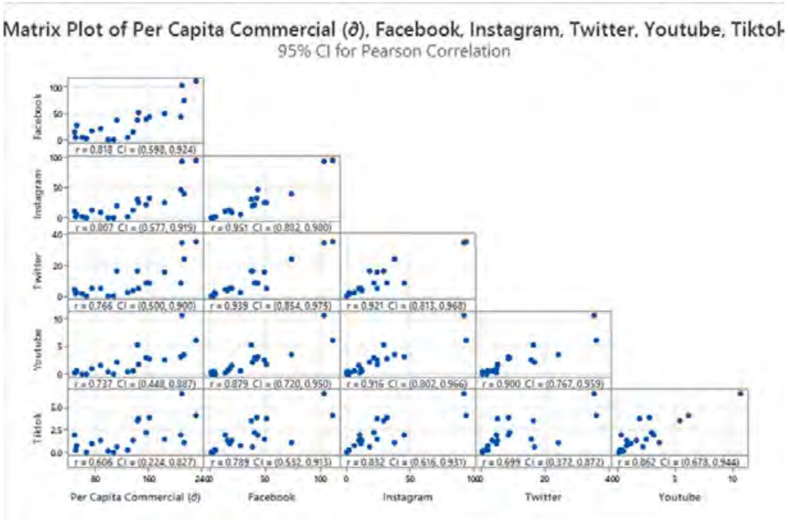
MP		*Starts*		*Min*		*90s*	
Mean	**10.3684**	**Mean**	**10.24561**	**Mean**	**922.070**	**Mean**	**10.24561**
Standard Deviation	**7.56078**	**Standard Deviation**	**7.63282**	**Standard Deviation**	**684.352584**	**Standard Deviation**	**7.60584**
Minimum	**1**	**Minimum**	**1**	**Minimum**	**90**	**Minimum**	**1**

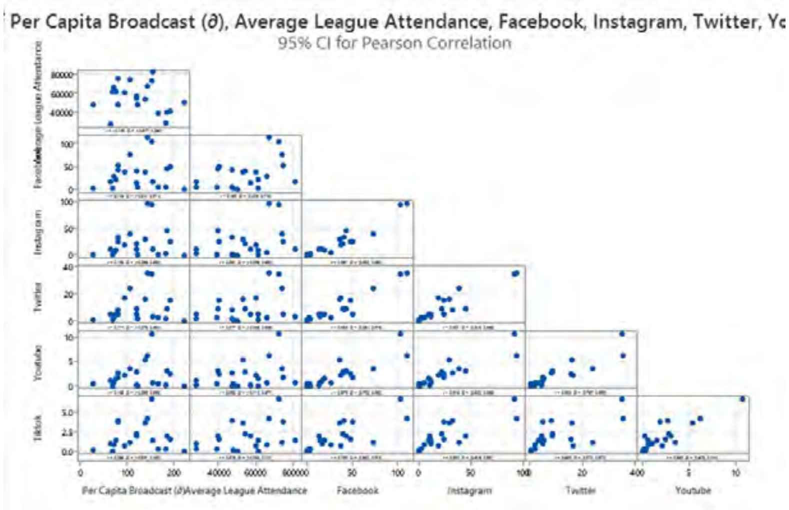
Maximum	**23**	**Maximum**	**23**	**Maximum**	**2070**	**Maximum**	**23**
GA		***GA90***		***SoTA***		***Saves***	
Mean	**14.6491**	**Mean**	**1.62245**	**Mean**	**44**	**Mean**	**30.3333**
Standard Deviation	**10.634**	**Standard Deviation**	**0.7441**	**Standard Deviation**	**32.1330**	**Standard Deviation**	**23.0**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**1**	**Minimum**	**1**
Maximum	**40**	**Maximum**	**3.43**	**Maximum**	**116**	**Maximum**	**79**
Save%		***W***		***D***		***L***	
Mean	**66.9824**	**Mean**	**3.91228**	**Mean**	**2.42105**	**Mean**	**3.91228**
Standard Deviation	**12.920**	**Standard Deviation**	**3.6561**	**Standard Deviation**	**2.42713**	**Standard Deviation**	**3.12981**
Minimum	**20**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**100**	**Maximum**	**13**	**Maximum**	**9**	**Maximum**	**13**
CS		***CS%***		***PKatt***		***PKA***	
Mean	**2.6315**	**Mean**	**22.3842**	**Mean**	**1.3859**	**Mean**	**1.017**
Standard Deviation	**2.525806**	**Standard Deviation**	**21.209**	**Standard Deviation**	**1.42370**	**Standard Deviation**	**1.1416**
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**	**Minimum**	**0**
Maximum	**9**	**Maximum**	**100**	**Maximum**	**5**	**Maximum**	**4**
PKsv		***PKm***		***Save%***			
Mean	**0.24561**	**Mean**	**0.122807**	**Mean**	**19.39722**		
Standard Deviation	**0.509926527**	**Standard Deviation**	**0.331133089**	**Standard Deviation**	**33.18296022**		
Minimum	**0**	**Minimum**	**0**	**Minimum**	**0**		
Maximum	**2**	**Maximum**	**1**	**Maximum**	**100**		

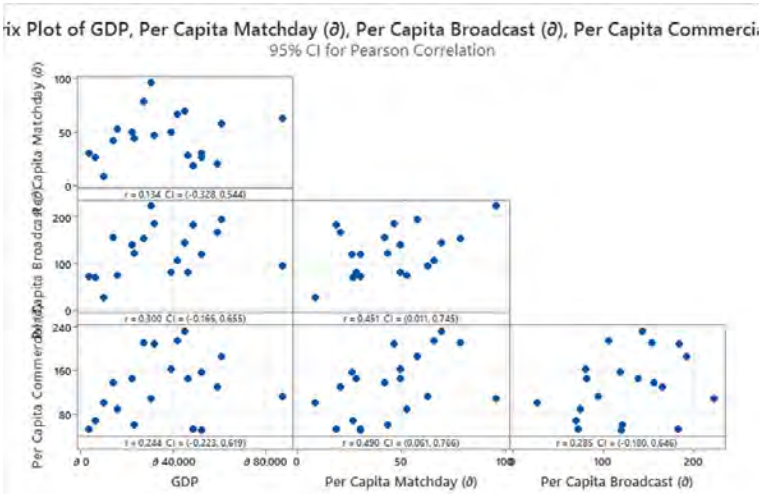
A4: Forecasted Inflation Rates



A5: Revenue Correlation Matrices







Appendix B: Risk and Risk Mitigation

B1: Risk Considerations and Development Tool

Risk Category	Risk Subcategory	Risk Division	Risk	Explanation of Risk
Operational	Technology	Tech Develop not performing as expected	Data Security	Advancements in technology for modeling player performance are not accurate; expectations are not matching outcome
Operational	Human Resources	Talent Management	Recruiting	Ability to develop new talent each year to keep the team competitive
Operational	Human Resources	Talent Management	Player Injuries	Injury or ailments that prevent players from participating in games, can affect outcome of games
Operational	Natural Disasters	Operations	Unexpected Natural or man-made disasters	Natural disasters or health related pandemics could affect arenas and players from competing and fans attending
Strategic	Execution	Management	Player Performance	Under/overestimating player performance can affect probability of winning FSA
Strategic	Execution	Strategic	Standings	Incorrectly predicting tournament standings
Financial	Market	Talent Management	Player Contracts	Accounting for increased salaries per renegotiations, increased experience, and/or leasing players
Financial	Market	External	Changes to inflation/deflation	Unexpected changes in the economy affecting expenses and revenues
Financial	Politics	Internal	Controversial Topics	Potential controversial topics affecting matchday attendance, stockholders, or sponsorships
Financial	Market	External	Investments	Changes to sponsorship and/or stock investments
Financial	Market	Internal	Revenue & Expenses	Incorrectly projecting future revenue and expenses

Appendix C: Team Program Design

C1: Criteria Selection

2020 Player Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
DF	55	7.691	2.26	4	6	9	13
DFFW	55	0.182	0.389	0	0	0	1
DFMF	55	0.982	0.972	0	0	1.5	4
FW	55	4.745	2.351	2	3	6	12
FWDF	55	0.036	0.189	0	0	0	1

2021 Player Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
DF	24	6.542	1.641	4	5	7.25	10
DFFW	24	0.25	0.442	0	0	0.25	1
DFMF	24	0.625	0.77	0	0	1	2
FW	24	3.833	1.659	1	2	5	7
FWDF	24	0.333	0.565	0	0	1	2

Age of Player Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Year: 2020							
Age	1527	27.558	3.973	17	25	30	40
Year: 2021							
Age	488	26.584	3.976	17	24	29	38

C2: R PCA

```
#PCA for Goalie Data
scaled.gdf<-scale(gdf)
gf_pca<-princomp(x=scaled.gdf)
print(summary(gf_pca), loadings = TRUE)

g1<-cgoalie[,c(1,2,18,19)]
g2<-gf_pca$scores[,c(1:5)]
gdata<-cbind(g1,g2) #pca for goalie data

#PCA for Player Data
pdf<-mdf[,c(4:56)]
```

scaled.pdf<-scale(pdf)
df_pca<-princomp(x = scaled.pdf)
print(summary(df_pca), loadings = TRUE)

d1<-mdf[,c(1:3,57,58)]
d2<-df_pca$scores[,c(1:14)]
data<-cbind(d1,d2)

#correlation plots
ggcorr(Xs)

C3: PCA Output

Goalie Output:

```
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Comp.5     Comp.6      Comp.7
Standard deviation     2.0158389 1.7550333 1.3694415 1.02144299 0.99918091 0.91212046 0.262191482
Proportion of Variance 0.3392598 0.2571529 0.1565697 0.08710619 0.08335066 0.06945846 0.005739287
Cumulative Proportion  0.3392598 0.5964127 0.7529825 0.84008864 0.92343930 0.99289776 0.998637052
                             Comp.8       Comp.9      Comp.10      Comp.11      Comp.12
Standard deviation     0.0973791870 0.0772512844 2.096751e-02 2.064109e-02 3.007483e-03
Proportion of Variance 0.0007916862 0.0004982327 3.670411e-05 3.557019e-05 7.551392e-07
Cumulative Proportion  0.9994287379 0.9999269706 9.999637e-01 9.999992e-01 1.000000e+00

Loadings:
                    Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11
Age                                        0.425  0.860  0.280                                    
Playing Time MP     -0.486 -0.104                                                    0.229   0.835
Playing Time Starts -0.486 -0.103                                                   -0.863        
Playing Time Min    -0.486 -0.103                                                    0.305  -0.380
Playing Time 90s    -0.486 -0.103                                                    0.329  -0.393
Performance GA       0.120 -0.485         0.342 -0.222  0.215        -0.615  0.399                
Performance GA90     0.118 -0.488         0.333 -0.223  0.209         0.737                       
Performance SoTA           -0.439  0.418 -0.209                      -0.225 -0.719                
Performance Saves          -0.302  0.385 -0.575  0.301                0.165  0.556                
Performance CS              0.245        -0.307 -0.151  0.903                                     
Performance PKatt    0.109 -0.271 -0.566 -0.253  0.140         0.712                              
Penalty Kicks PKA    0.105 -0.254 -0.583 -0.242  0.147        -0.699        -0.116                
                    Comp.12
Age                        
Playing Time MP            
Playing Time Starts        
Playing Time Min      0.716
Playing Time 90s     -0.698
Performance GA             
Performance GA90           
Performance SoTA           
Performance Saves          
Performance CS             
Performance PKatt          
Penalty Kicks PKA          
```

Player Output:

```
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Comp.5     Comp.6     Comp.7     Comp.8     Comp.9    Comp.10
Standard deviation     3.6630235 2.5407317 2.3894982 1.64593241 1.48062481 1.41024431 1.33069657 1.28031998 1.11954659 1.06271969
Proportion of Variance 0.2580762 0.1241613 0.1098201 0.05210658 0.04216563 0.03825227 0.03405859 0.03152867 0.02410754 0.02172231
Cumulative Proportion  0.2580762 0.3822375 0.4920576 0.54416419 0.58632981 0.62458209 0.65864067 0.69016934 0.71427688 0.73599919
                         Comp.11    Comp.12    Comp.13    Comp.14   Comp.15    Comp.16    Comp.17    Comp.18    Comp.19    Comp.20
Standard deviation     1.0355275 1.01499567 1.00903164 0.99768960 0.9538248 0.93526015 0.91097199 0.86882328 0.83928809 0.79616368
Proportion of Variance 0.0206249 0.01981513 0.01958295 0.01914518 0.0174987 0.01682416 0.01596168 0.01451882 0.01354848 0.01219195
Cumulative Proportion  0.7566241 0.77643922 0.79602217 0.81516735 0.8326660 0.84949021 0.86545189 0.87997071 0.89351919 0.90571114
                          Comp.21    Comp.22    Comp.23     Comp.24     Comp.25     Comp.26    Comp.27    Comp.28     Comp.29
Standard deviation     0.76826264 0.74618578 0.72274644 0.699887691 0.633372272 0.615165746 0.57637341 0.55419151 0.515100954
Proportion of Variance 0.01135241 0.01070933 0.01004709 0.009421613 0.007715901 0.007278683 0.00638964 0.00590729 0.005103325
Cumulative Proportion  0.91706355 0.92777289 0.93781998 0.947241593 0.954957494 0.962236177 0.96862582 0.97453311 0.979636432
                           Comp.30     Comp.31     Comp.32     Comp.33     Comp.34     Comp.35    Comp.36     Comp.37     Comp.38
Standard deviation     0.456995203 0.400076620 0.389162371 0.299610835 0.298616705 0.272041126 0.25739893 0.243981766 0.210842518
Proportion of Variance 0.004016907 0.003078611 0.002912931 0.001726568 0.001715129 0.001423435 0.00127433 0.001144941 0.000855037
Cumulative Proportion  0.983653340 0.986731951 0.989644883 0.991371450 0.993086579 0.994510014 0.99578434 0.996929286 0.997784323
                            Comp.39      Comp.40      Comp.41      Comp.42      Comp.43      Comp.44      Comp.45      Comp.46
Standard deviation     0.2044146414 0.1720869236 0.1310739367 0.0987888011 0.0819374670 5.959486e-02 5.271643e-02 0.0410090807
Proportion of Variance 0.0008036974 0.0005695925 0.0003304466 0.0001877085 0.0001291319 6.831029e-05 5.345157e-05 0.0000323466
Cumulative Proportion  0.9985880201 0.9991576127 0.9994880592 0.9996757677 0.9998048997 9.998732e-01 9.999267e-01 0.9999590081
                            Comp.47      Comp.48      Comp.49      Comp.50      Comp.51      Comp.52
Standard deviation     3.709768e-02 2.025772e-02 1.189256e-02 1.002040e-02 9.397788e-03 3.801644e-03
Proportion of Variance 2.647049e-05 7.893134e-06 2.720314e-06 1.931251e-06 1.698712e-06 2.779787e-07
Cumulative Proportion  9.999855e-01 9.999934e-01 9.999961e-01 9.999980e-01 9.999997e-01 1.000000e+00

Loadings:
                  Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15
Age                                           0.124                0.125  0.137          0.491   0.204   0.345   0.150   0.175
90s                                    0.118  0.111                       0.116          0.520           0.340
Gls                             -0.245  0.431
Standard Sh                     -0.331         0.197               -0.132
Standard SoT                    -0.323  0.158  0.102               -0.169
Standard Sh/90                  -0.289         0.202               -0.144                                -0.125          -0.110
Standard SoT/90                 -0.288  0.130  0.118               -0.178                                                -0.123
Performance PK                                 0.151                      -0.233  -0.138   0.367  -0.107  -0.401           0.516
Performance PKatt                              0.102        -0.211  0.612
Standard FK                                    0.103        -0.193  0.609                                -0.108
Expected xG                     -0.320         0.247                                      -0.128           0.134
Expected npxG                   -0.313         0.228               -0.101                 -0.144           0.171
Expected G-xG                           0.522 -0.285
Expected np:G-xG                        0.521 -0.285
Total Cmp          0.266                             -0.120
Total Att          0.269
Total TotDist      0.261                             -0.175
Total PrgDist      0.240                             -0.218
Short Cmp          0.263
                  Comp.16 Comp.17 Comp.18 Comp.19 Comp.20 Comp.21 Comp.22 Comp.23 Comp.24 Comp.25 Comp.26 Comp.27 Comp.28 Comp.29
Age                                 0.658   0.101
90s                         0.138  -0.695           0.144
Gls                                         0.165  -0.163  -0.114
Standard Sh                                         0.196  -0.301   0.225
Standard SoT                               -0.189           0.331  -0.210
Standard Sh/90                             -0.149   0.375  -0.411   0.319
Standard SoT/90   -0.105                   -0.341   0.176   0.425  -0.293
Performance PK    -0.139  -0.395            0.305
Performance PKatt                                                                                          0.697
Standard FK                                -0.122                                                         -0.702
Expected xG        0.107                    0.208  -0.290
Expected npxG      0.127                    0.260  -0.342
Expected G-xG
Expected np:G-xG
Total Cmp                                                          -0.115
Total Att
Total TotDist
Total PrgDist                                                       0.298
Short Cmp                                                          -0.245
```

C4: R Code Player Rank Modeling

```
#aggregating goalie/team
team_goal<-mgoal[,-c(1,3,4)]%>% group_by(Nation) %>% summarise_all("mean")
team_play<-mplayers[,-c(1,3,4,5)]%>% group_by(Nation) %>% summarise_all("mean") #all 24

#adding rank as variable
team_goal$rank<-c(8,15,16,23,19,14,7,10,11,18,17,13,4,20,3,21,22,2,5,1,6,24,12)
team_play$rank<-c(8,15,16,9,23,19,14,7,10,11,18,17,13,4,20,3,21,22,2,5,1,6,24,12)

#Poisson Regression Goalie ----
Xgr<-team_goal[,-c(1)]
fitgoalie<-glm(rank ~ ., data=Xgr, family="poisson")
summary(fitgoalie)

#Multicolin
imcdiag(fitgoalie)

#New Model Goalie
newfitgoalie<-glm(rank ~ G1+G3+G5, data=Xgr, family="poisson")
summary(newfitgoalie)

newfitgoalienb<-glm.nb(rank ~G1+G3+G5, data=Xgr)
summary(newfitgoalienb) #neg bin regression

fitgoalieztp <- zerotrunc(rank ~ G1+G3+G5, data=Xgr, dist="poisson")
summary(fitgoalieztp)

fitgoalieztnb <- zerotrunc(rank ~ G1+G3+G5, data=Xgr, dist="negbin")
summary(fitgoalieztnb)

#getitng AIC values
AIC(newfitgoalie)
AIC(newfitgoalienb)
AIC(fitgoalieztp)
AIC(fitgoalieztnb)

#loglikelihoodvalue
logLik(newfitgoalie)
logLik(newfitgoalienb)
logLik(fitgoalieztp)
logLik(fitgoalieztnb)
```

```
#Poisson Regression Players ----
Xpr<-team_play[,-(1)]
#old pass regress
fitp<-glm(rank ~ ., data=Xpr, family="poisson")
summary(fitp)

#Multicolin
imcdiag(fitp)

#New Models Pass
newfitp<-glm(rank ~ C2+C4+C8+C9+C10+C14, data=Xpr, family="poisson")
summary(newfitp)

imcdiag(newfitp)

newfitpnb<-glm.nb(rank ~ C2+C4+C8+C9+C10+C14, data=Xpr)
summary(newfitpnb) #neg bin regression

fitpztp <- zerotrunc(rank ~ C2+C4+C8+C9+C10+C14, data = Xpr, dist="poisson")
summary(fitpztp)

fitpztnb <- zerotrunc(rank ~ C2+C4+C8+C9+C10+C14, data = Xpr, dist="negbin")
summary(fitpztnb)

#getitng AIC values
AIC(newfitp)
AIC(newfitpnb)
AIC(fitpztp)
AIC(fitpztnb)

#loglikelihoodvalue
logLik(newfitp)
logLik(newfitpnb)
logLik(fitpztp)
logLik(fitpztnb)

#FINAL MODELS ----
#Players
fitpztp <- zerotrunc(rank ~ C2+C4+C8+C9+C10+C14, data = Xpr, dist="poisson")
```

summary(fitpztp)

#Goalies
fitgoalieztp <- zerotrunc(rank ~ G1+G3+G5, data=Xgr, dist="poisson")
summary(fitgoalieztp)

#Predicting for test data
testdatagoal$predictrank<-predict(fitgoalieztp, testdatagoal, type="response")
testdataplay$predictrank<-predict(fitpztp, testdataplay, type="response")

C5: Model Selection Output

Goalie Model:

```
Call:
imcdiag(mod = newfitgoalie)


All Individual Multicollinearity Diagnostics Result

      VIF    TOL     Wi     Fi Leamer    CVIF Klein   IND1   IND2
G1 1.2518 0.7988 2.5183 5.2884 0.8938 -2.5342     0 0.0799 1.3267
G3 1.2864 0.7773 2.8645 6.0154 0.8817 -2.6043     0 0.0777 1.4685
G5 1.0321 0.9689 0.3205 0.6731 0.9844 -2.0893     0 0.0969 0.2048

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test

G3 , G5 , coefficient(s) are non-significant may be due to multicollinearity

R-square of y on all x: 0.8151

* use method argument to check which regressors may be the reason of collinearity
===================================
Call:
zerotrunc(formula = rank ~ G1 + G5, data = Xgr, dist = "poisson")

Deviance residuals:
      Min        1Q    Median        3Q       Max
-2.703663 -0.624870 -0.008854  0.846694  1.337434

Coefficients (truncated poisson with log link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.69029    0.24658   2.799  0.00512 **
G1           1.23911    0.14999   8.262  < 2e-16 ***
G5           0.13980    0.08253   1.694  0.09027 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Number of iterations in BFGS optimization: 5
Log-likelihood: -61.16 on 3 Df
```

Player Model:

```
Call:
imcdiag(mod = newfitp)


All Individual Multicollinearity Diagnostics Result

       VIF    TOL     Wi     Fi Leamer   CVIF Klein   IND1   IND2
C2  1.1293 0.8855 0.4654 0.6141 0.9410 1.6069     0 0.2460 0.9099
C4  1.1703 0.8545 0.6130 0.8089 0.9244 1.6653     0 0.2374 1.1565
C8  1.2245 0.8167 0.8082 1.0664 0.9037 1.7425     0 0.2268 1.4573
C9  1.2154 0.8228 0.7753 1.0229 0.9071 1.7294     0 0.2286 1.4084
C10 1.1274 0.8870 0.4585 0.6050 0.9418 1.6042     0 0.2464 0.8979
C14 1.0219 0.9786 0.0787 0.1038 0.9892 1.4541     0 0.2718 0.1700

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test

C4 , C8 , C9 , C10 , C14 , coefficient(s) are non-significant may be due to multicollinearity

R-square of y on all x: 0.6644

* use method argument to check which regressors may be the reason of collinearity
===================================


Call:
zerotrunc(formula = rank ~ C2 + C4 + C8 + C9 + C10 + C14, data = Xpr, dist = "poisson")

Deviance residuals:
    Min      1Q  Median      3Q     Max
-2.4292 -1.2521  0.1218  0.6227  2.3148

Coefficients (truncated poisson with log link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.46859    0.08003  30.845  < 2e-16 ***
C2          -0.48251    0.08349  -5.779  7.5e-09 ***
C4          -0.52429    0.17443  -3.006  0.00265 **
C8          -0.53898    0.18066  -2.983  0.00285 **
C9           0.31129    0.17724   1.756  0.07903 .
C10         -0.68218    0.35413  -1.926  0.05406 .
C14          0.78098    0.24074   3.244  0.00118 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Number of iterations in BFGS optimization: 11
Log-likelihood: -68.53 on 7 Df
```

C6: R Code Optimization for Team Selection

```
#Salary Cap
lowend<-7370370*27 #20% of budget
highend<- 16574250*27 #50% of budget

#Optimization - Low End
obj = DF$PredictRank
con = rbind(t(model.matrix(~ Pos + 0, DF)), rep(1,nrow(DF)), DF$Salary)
dir = c(">=",">=","<=",">=","==","<=")
```

```
rhs = c(10,6,1,8,27,7370370*27)
resultlow = lp("min", obj, con, dir, rhs, all.bin = TRUE)

DF$included<-resultlow$solution
teamlow<-split(DF,DF$included)

#Team - high
obj = DF$PredictRank
con = rbind(t(model.matrix(~ Pos + 0, DF)), rep(1,nrow(DF)), DF$Salary)
dir = c(">=",">=","<=",">=","==","<=")
rhs = c(10,6,2,8,27,16574250*27)
resulthigh = lp("min", obj, con, dir, rhs, all.bin = TRUE)

DF$includedhigh<-resulthigh$solution
teamhigh<-split(DF,DF$includedhigh)
```

Appendix D: Literature Review

Gaines, C. (2012, July 25). *Ads on sports jerseys might be worth more than you think, and that's a good thing*. Business Insider. Retrieved March 10, 2022, from https://www.businessinsider.com/ads-on-sports-jerseys-might-be-worth-more-than-you-think-and-thats-a-good-thing-2012-7

Sport jersey sponsorships dramatically improve a team's profits and allow a change in money allocation. Based on the top English soccer teams, around 31 million is paid to these teams to have an advertisement on their shirt. This money can pay for a stronger or more talented roster or to decrease prices for the fans.

Joy Nwokoro, J. (2021, May 26). *Estimated Cost of Building a Concession Stand in 2022*. ProfitableVenture. Retrieved March 10, 2022, from https://www.profitableventure.com/cost-build-concession-stand/

One way to increase profit margins is to build concession stands throughout stadiums. The article clearly presented the necessary factors to help determine the cost of the stand. This will help assist in the process of expanding our revenues. From the concession stands, there can be as much as 80 percent profit, depending on what is sold and other costs.

Paden, J. (2021, December 19). *The business of naming stadiums*. AmadorValleyToday. Retrieved March 17, 2022, from https://www.amadorvalleytoday.org/17148/sports/the-business-of-naming-stadiums/

For over a century, companies have been willing to pay millions of dollars to have a stadium named after them. This promotes advertisement for the company along with boosting the revenues for the team. Across all sports, the cost of naming stadiums remains consistent. On average, around 200 million dollars is spent on the naming rights.

Pérez-Toledano, M. Á., Rodriguez, F. J., García-Rubio, J., & Ibañez, S. J. (2019). Players' selection for basketball teams, through Performance Index Rating, using multiobjective evolutionary algorithms. *PloS one*, *14*(9), e0221258.

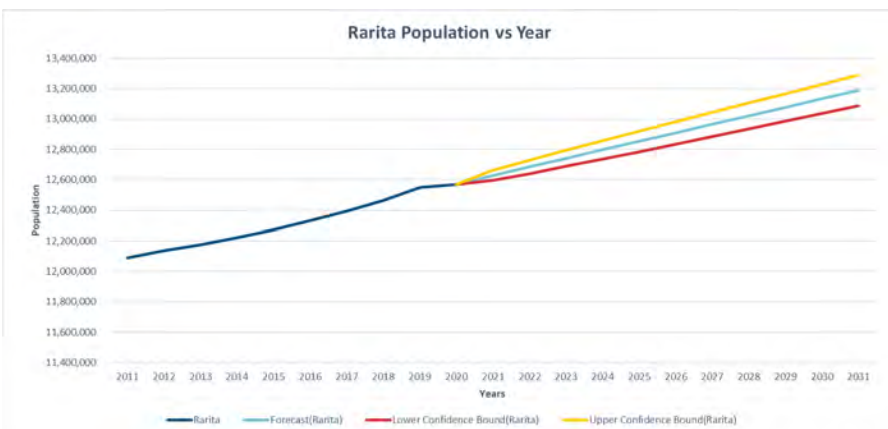
Two main goals of roster creation are to minimize the financial costs and to maximize the expected performance of the players selected. Create a valuation metric that is some sort of reference indicator of valuable players. Team valuation index is the sum of all players' normalized valuation indexes. This valuation metric is an objective performance indicator.

Pettinger, T. (2018, January 8). *Effects of a falling inflation rate*. Economics Help. Retrieved February 22, 2022, from https://www.economicshelp.org/blog/357/inflation/effects-of-a-falling-inflation-rate/#:~:text=A%20falling%20rate%20of%20inflation%20means%20that%20prices%20will%20be,competitive%20increasing%20exports%20and%20growth

There are many different effects of a falling interest rate, but it is not always detrimental. Some of the benefits include increased competitiveness, increased real wages, and improved return for investors. However, it can also cause a decrease in the GDP, deflationary pressure, along with higher unemployment.

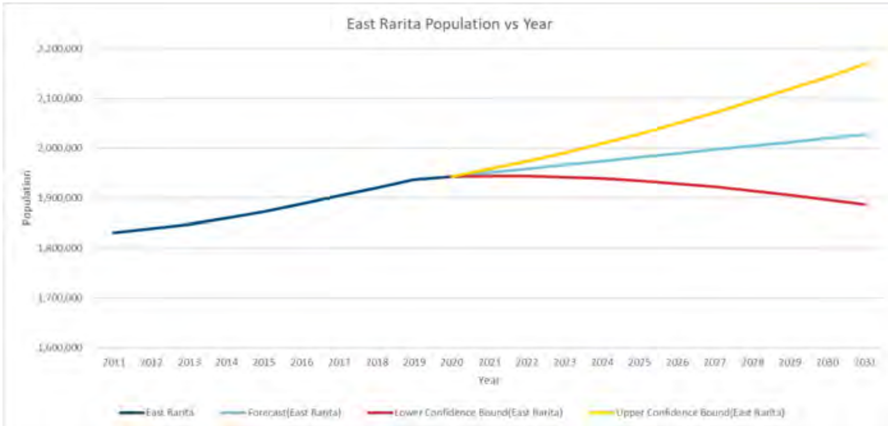
Appendix E: Further Considerations

Population Forecasting:



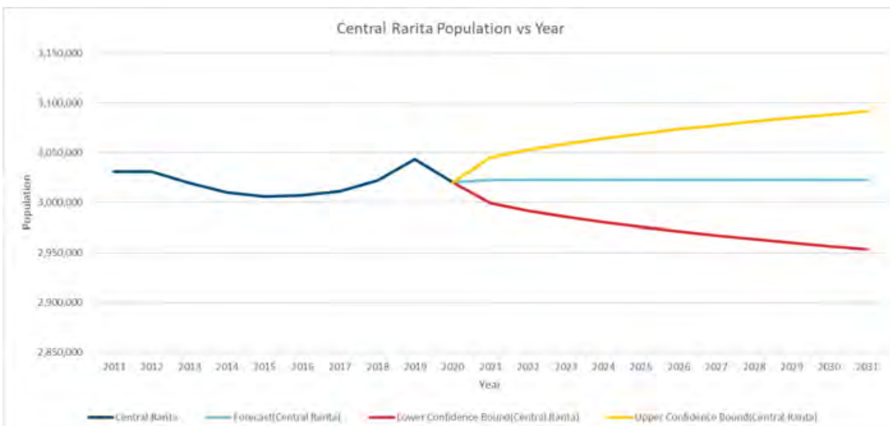
Graph E1: Rarita Population vs Year

Shows the population trend for Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



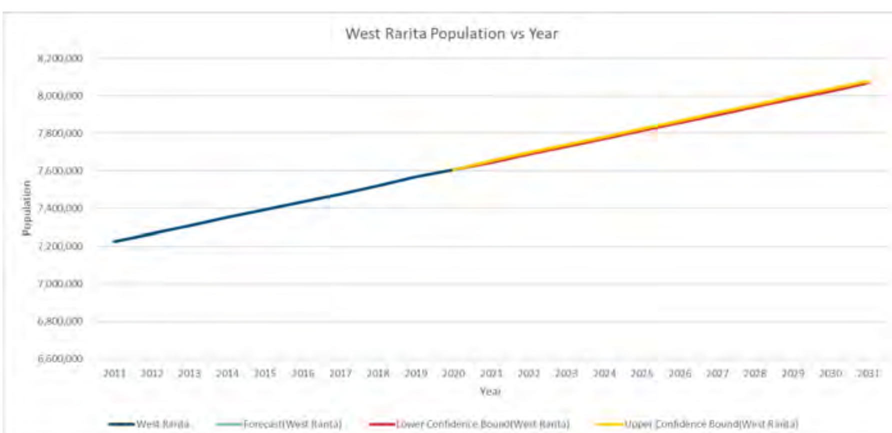
<u>Graph E2: East Rarita Population vs Year</u>

Shows the population trend for East Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



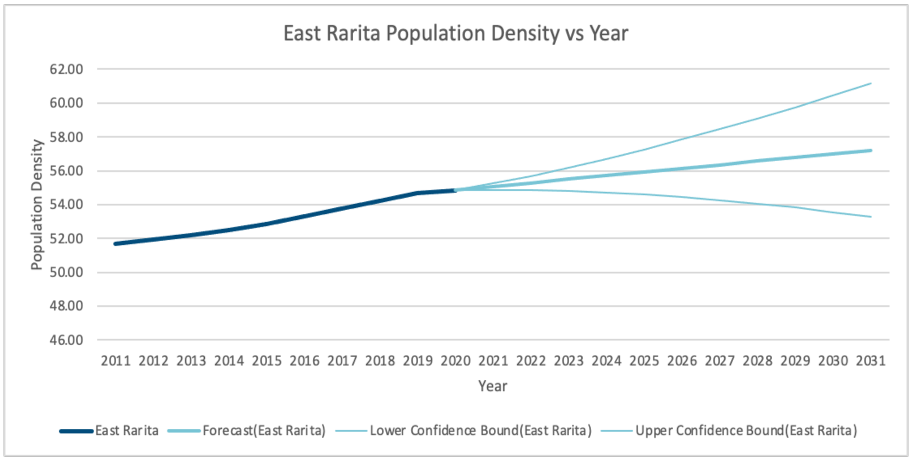
<u>Graph E3: Central Rarita Population vs Year</u>

Shows the population trend for Central Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



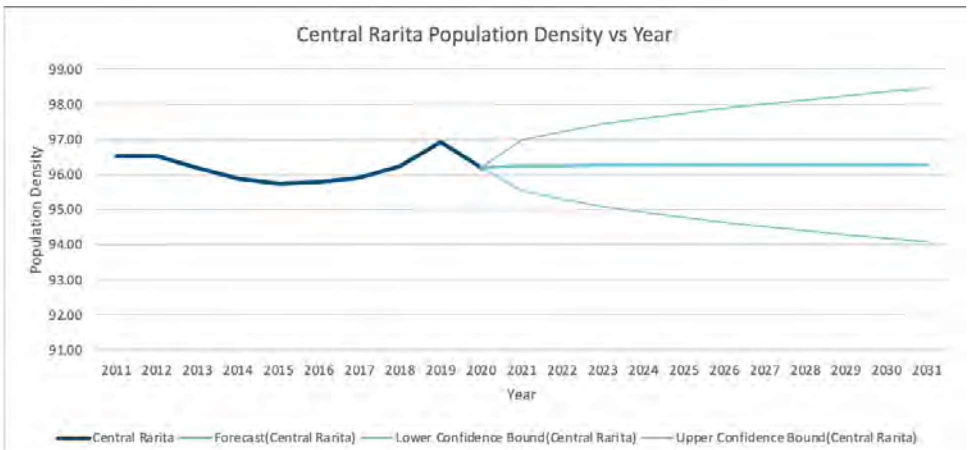
<u>Graph E4: West Rarita Population vs Year</u>

Shows the population trend for West Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



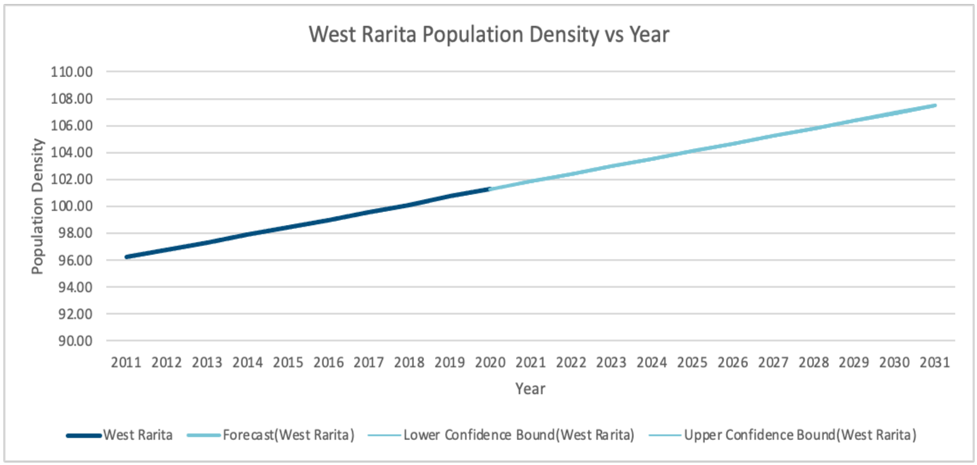
Graph E5: East Rarita Population Density vs Year

Shows the population density (people/ km^2) trend for East Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



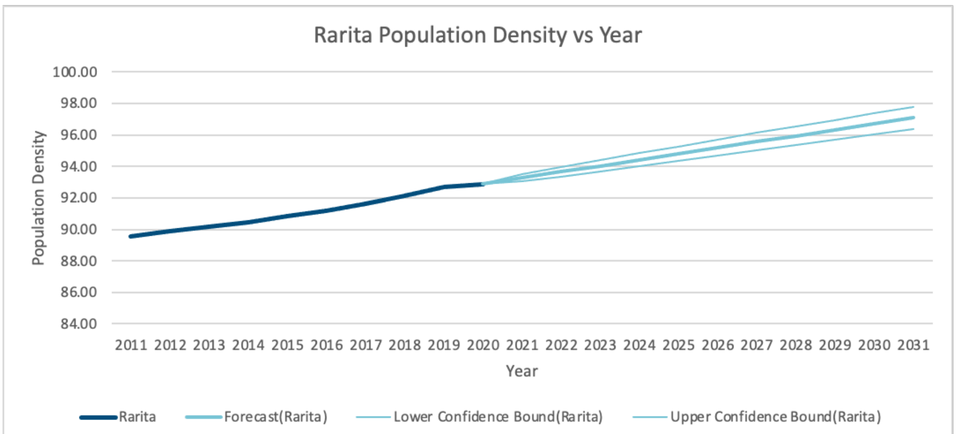
Graph E6: Central Rarita Population Density vs Year

Shows the population density (people/ km^2) trend for Central Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



Graph E7: West Rarita Population Density vs Year

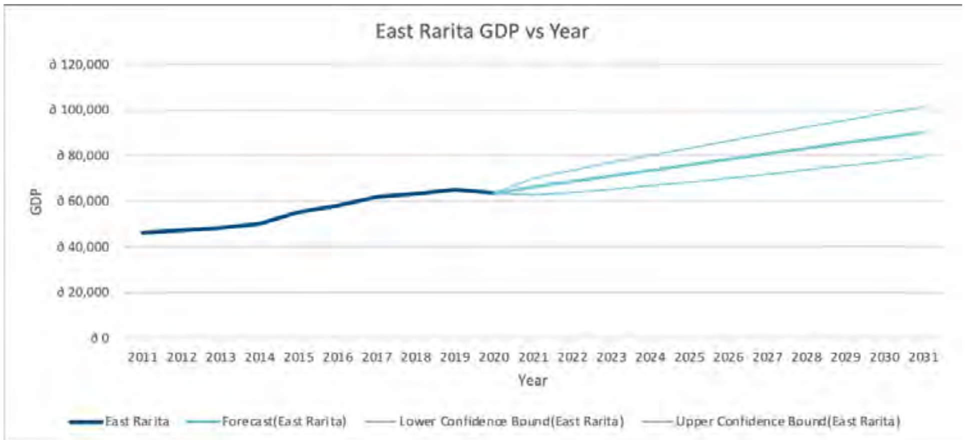
Shows the population density (people/ km^2) trend for West Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



Graph E8: Rarita Population Density vs Year

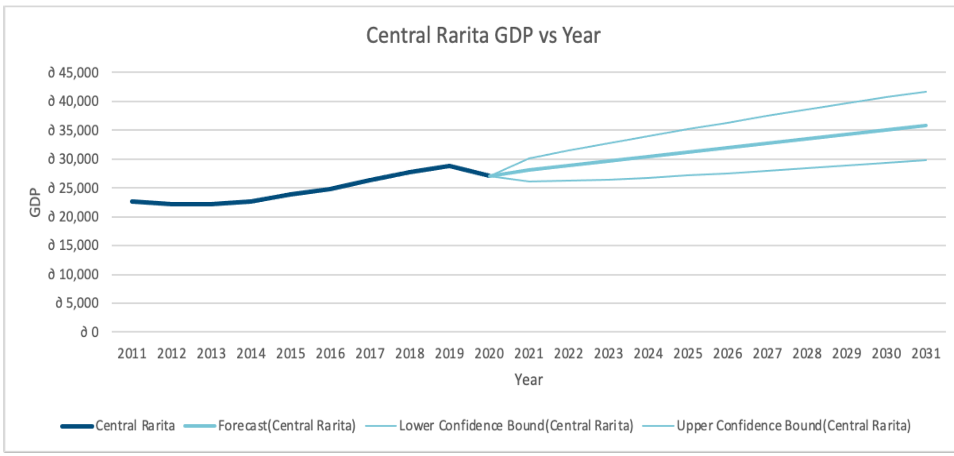
Shows the population density (people/ km^2) trend for Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.

GDP Forecasting:



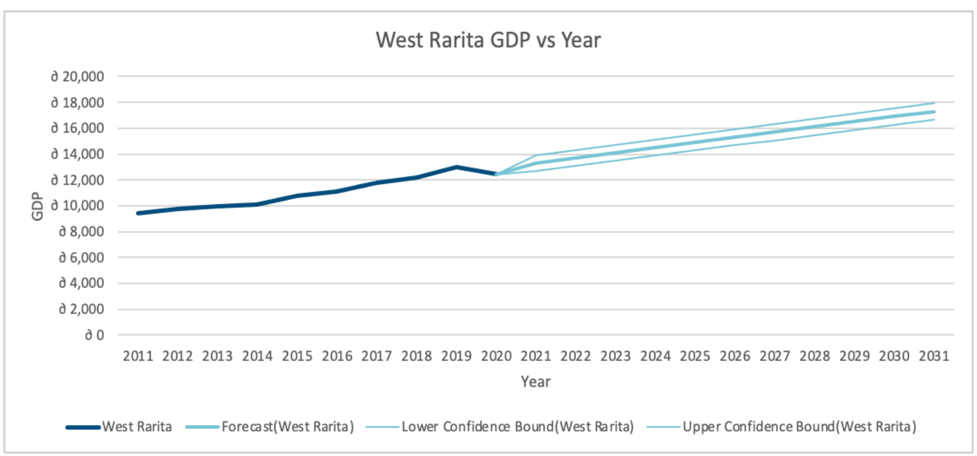
Graph E9: East Rarita GDP vs Year

Shows the GDP per capita trend for East Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



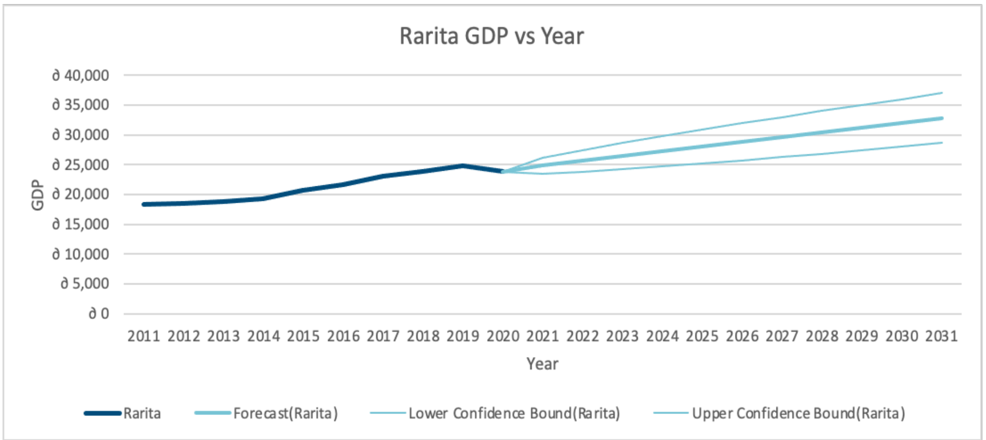
Graph E10: Central Rarita GDP vs Year

Shows the GDP per capita trend for Central Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



Graph E11: West Rarita GDP vs Year

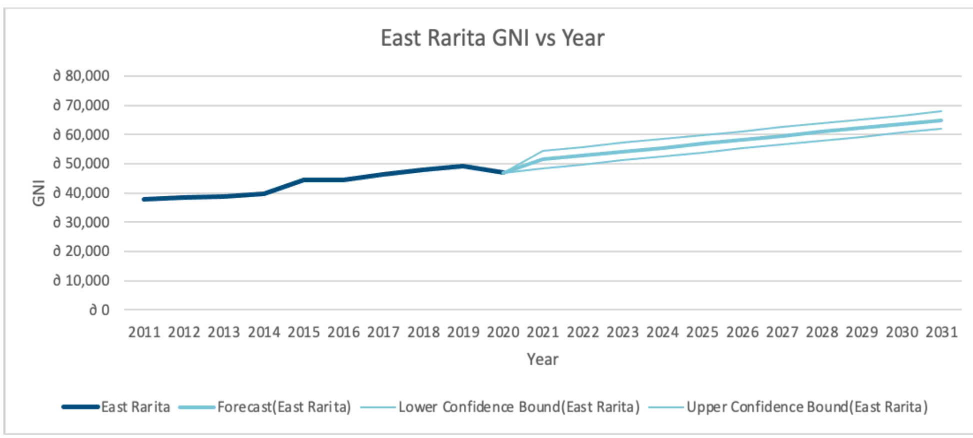
Shows the GDP per capita trend for West Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



Graph E12: Rarita GDP vs Year

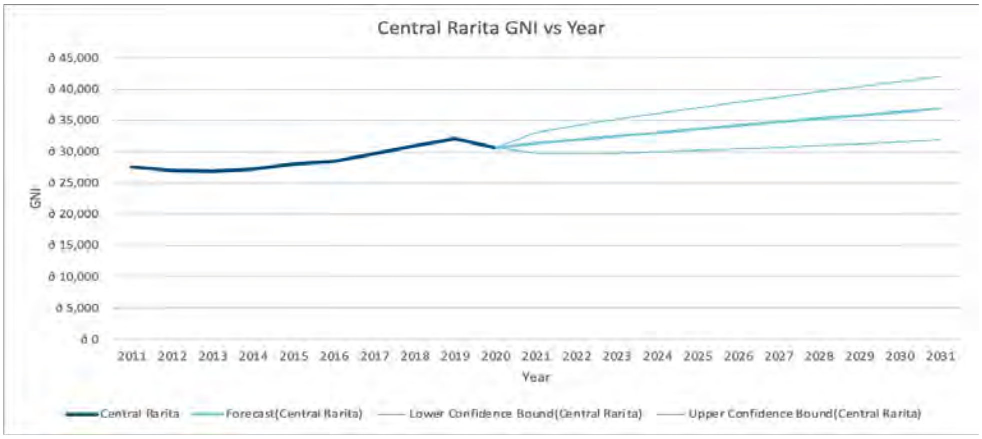
Shows the GDP per capita trend for Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.

GNI Forecasting:



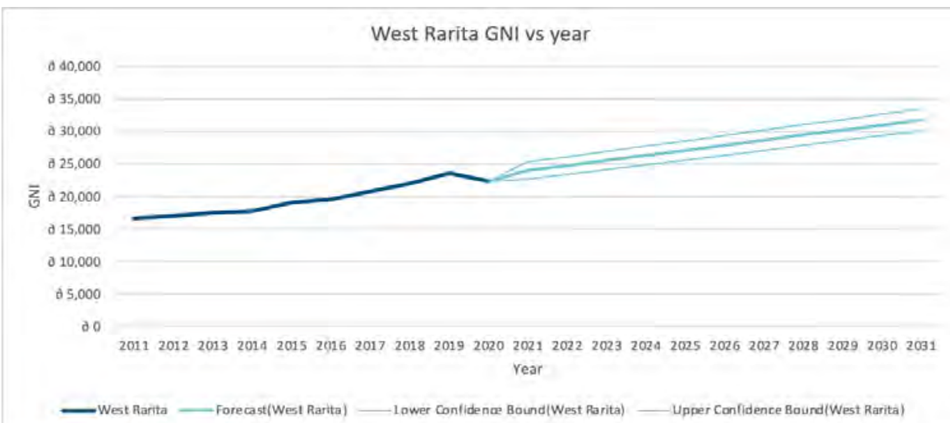
Graph E13: East Rarita GNI vs Year

Shows the GNI per capita trend for East Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



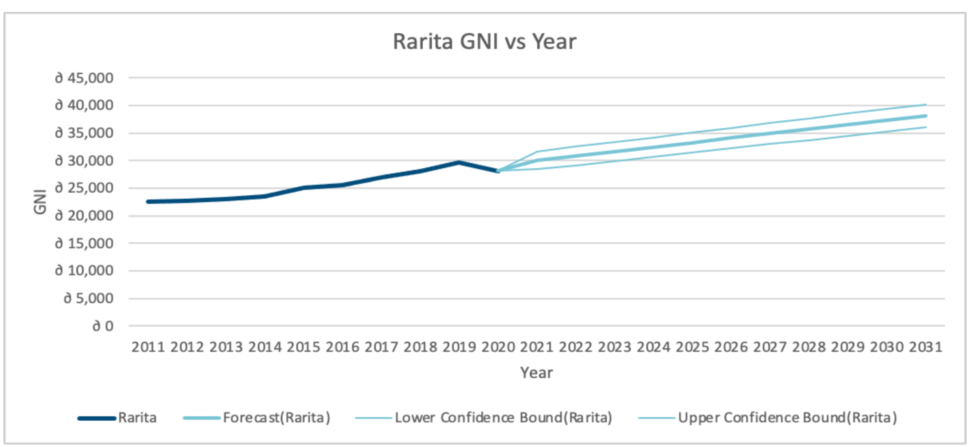
Graph E14: Central Rarita GNI vs Year

Shows the GNI per capita trend for Central Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



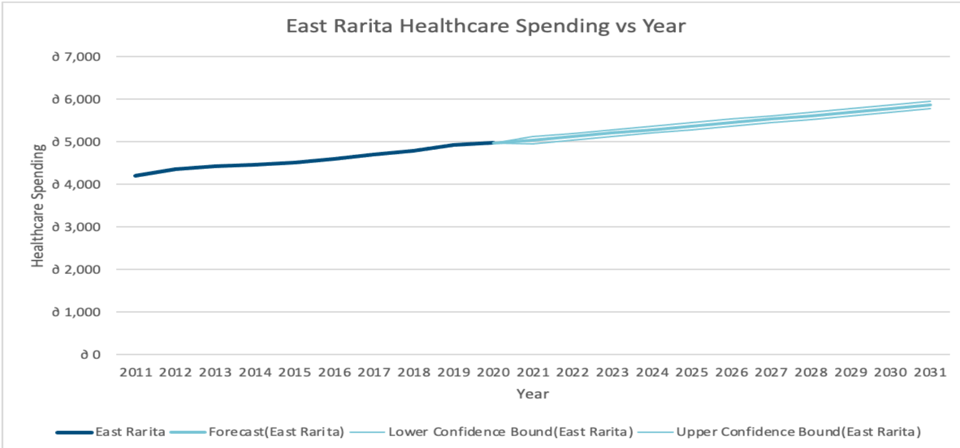
Graph E15: West Rarita GNI vs Year

Shows the GNI per capita trend for West Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



Graph E16: Rarita GNI vs Year

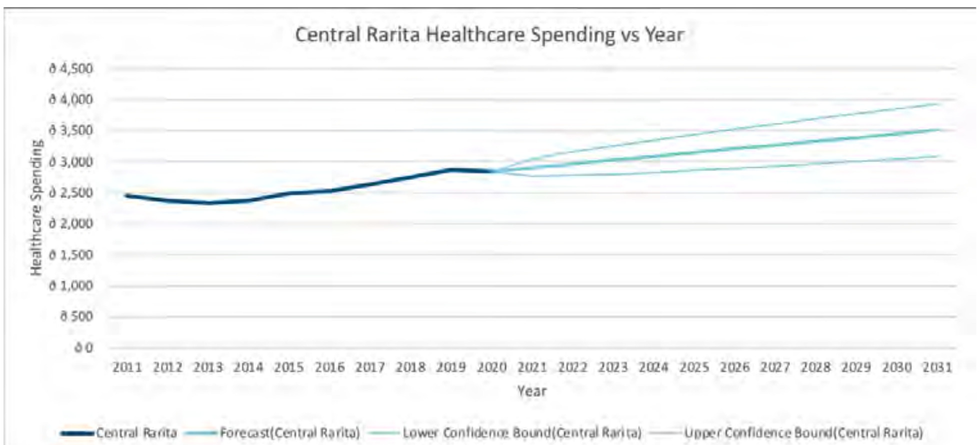
Shows the GNI per capita trend for Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



Healthcare Spending Forecasting:

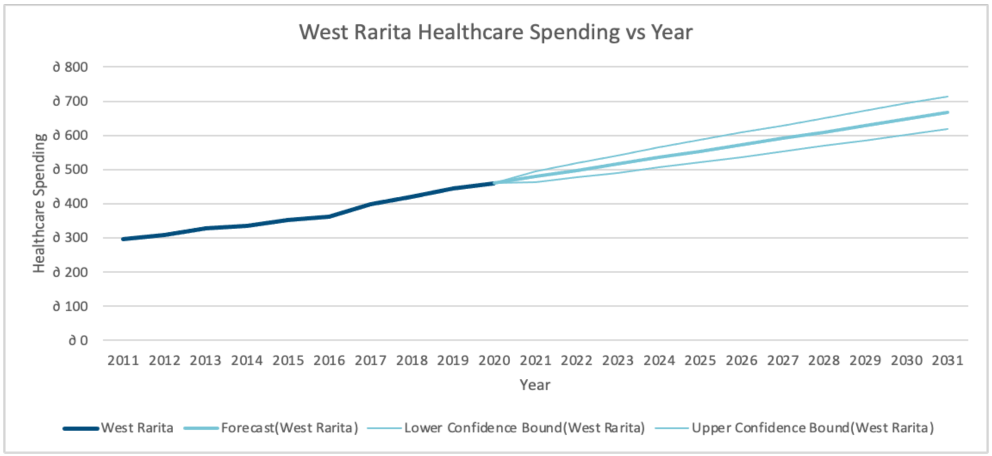
Graph E17: East Rarita Healthcare Spending vs Year

Shows the healthcare spending per capita trend for East Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.

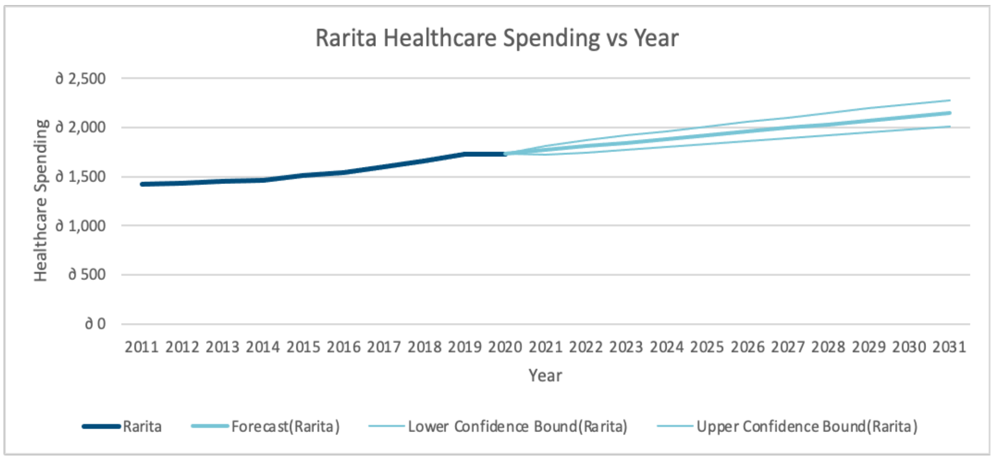


Graph E18: Central Rarita Healthcare Spending vs Year

Shows the healthcare spending per capita trend for Central Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.

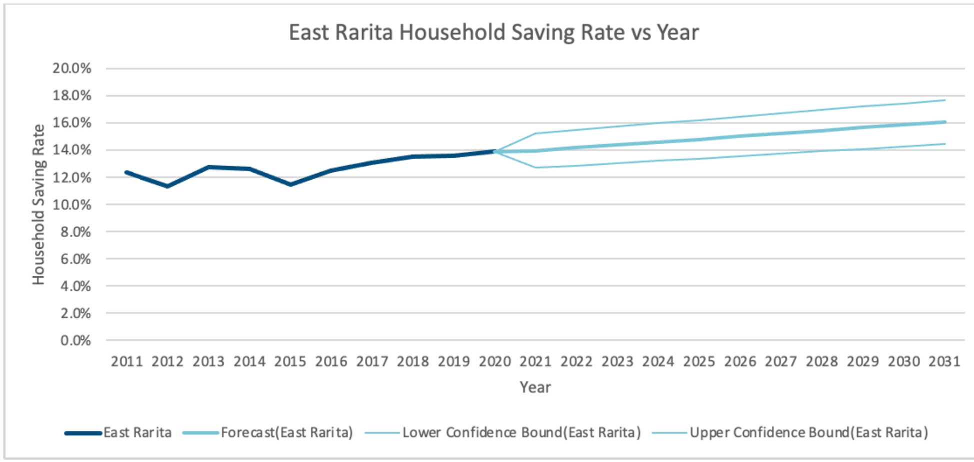


Graph E19: West Rarita Healthcare Spending vs Year
Shows the healthcare spending per capita trend for West Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.

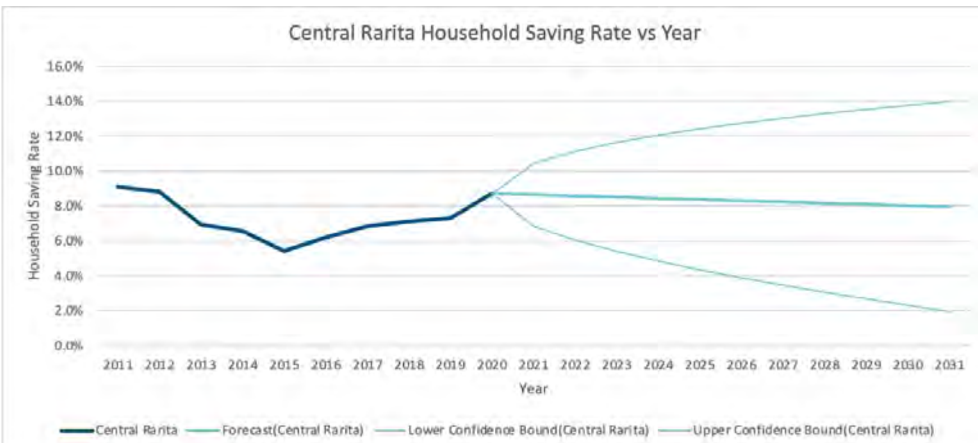


Graph E20: Rarita Healthcare Spending vs Year
Shows the healthcare spending per capita trend for Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.

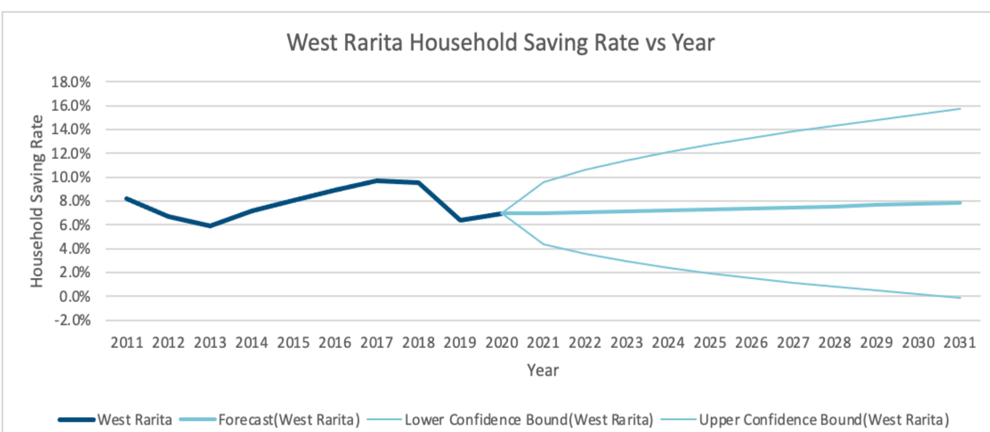
Household Saving Rate Forecasting:



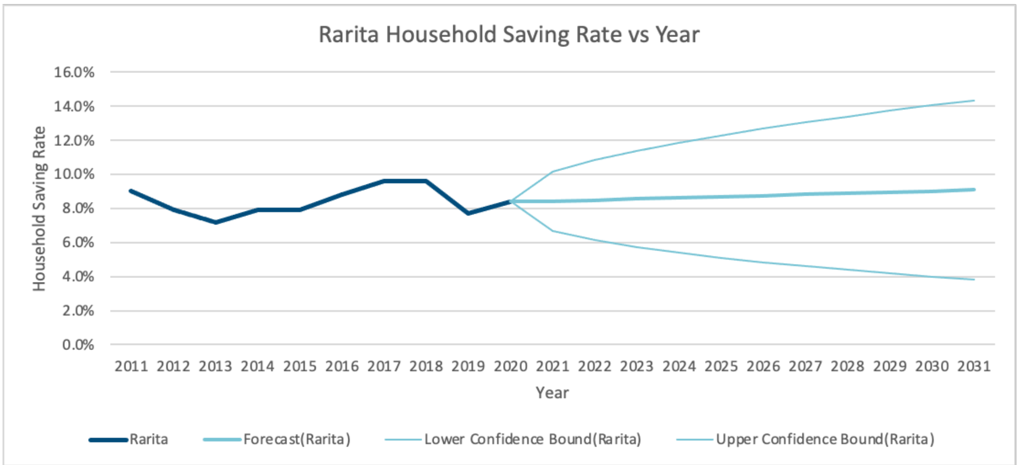
Graph E21: East Rarita Household Saving Rate vs Year
Shows the household saving rate trend for East Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



Graph E22: Central Rarita Household Saving Rate vs Year
Shows the household saving rate trend for Central Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



Graph E23: West Rarita Household Saving Rate vs Year
Shows the household saving rate trend for West Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.



Graph E24: Rarita Household Saving Rate vs Year
Shows the household saving rate trend for Rarita from 2011 to 2020 and forecasts the future population until 2031. Depicts the 95% confidence interval.

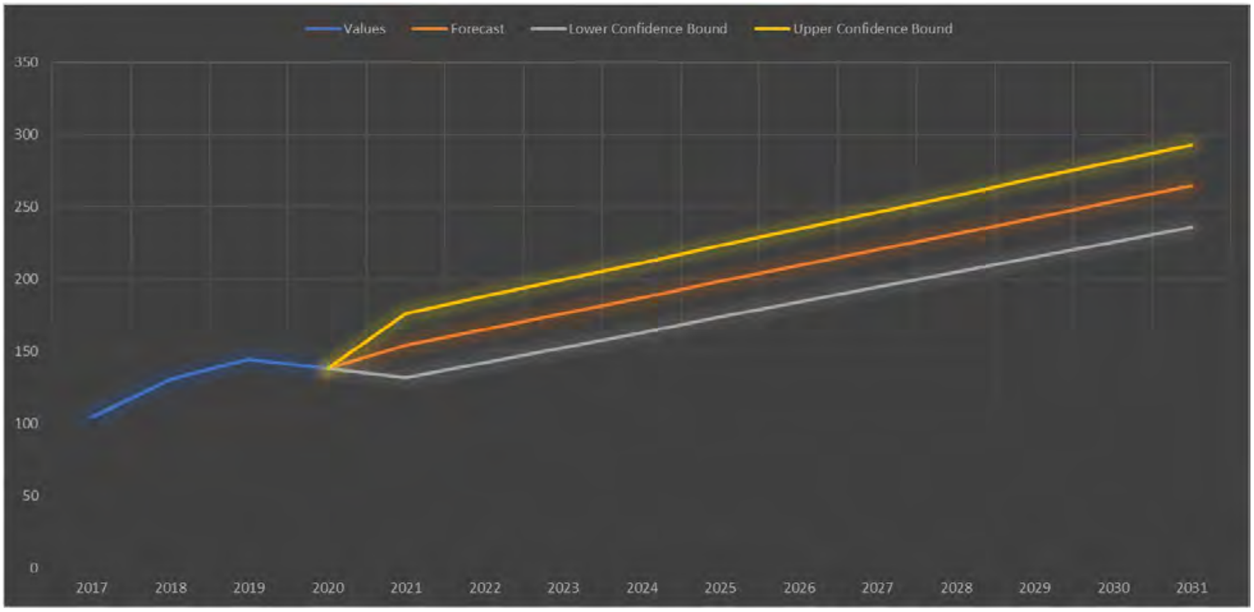
Appendix F: Expenses and Revenues

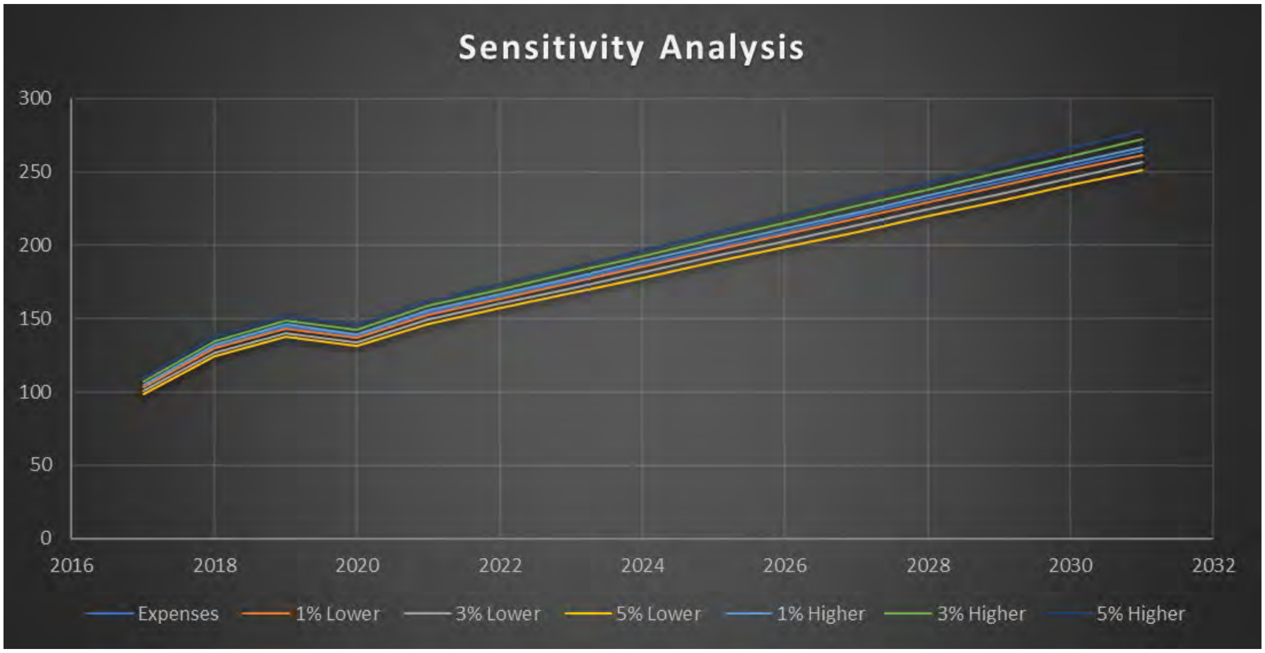
F1: Expenses R Code

```
data <-please.work
View(please.work)
datanew<-raritas.expenses.and.gdp
View(raritas.expenses.and.gdp)

T1 <-glm(X2020.Total.Expense~ X2020 - 1,data=subset(data))
summary(T1)

T2 <-glm(X2019.Total.Expense~ X2019 - 1,data=subset(data))
summary(T2)

T3 <-glm(X2018.Total.Expense~ X2018 - 1,data=subset(data))
summary(T3)

T4 <-glm(X2017.Total.Expense~ X2017 - 1,data=subset(data))
```

```
summary(T4)

T5 <-glm(X2016.Total.Expense~ X2016 - 1,data=subset(data))
summary(T5)

predOut <- predict(object=T1, newdata=datanew, type="response")
print(predOut)

predOut <- predict(object=T2, newdata=datanew, type="response")
print(predOut)

predOut <- predict(object=T3, newdata=datanew, type="response")
print(predOut)

predOut <- predict(object=T4, newdata=datanew, type="response")
print(predOut)

testM5 <- raritas.expenses.and.gdp[c(8)]
View(testM5)
predOut <- predict(object=T5, newdata=testM5, type="response")
print(predOut)
```

F2: Forecasted Expenses



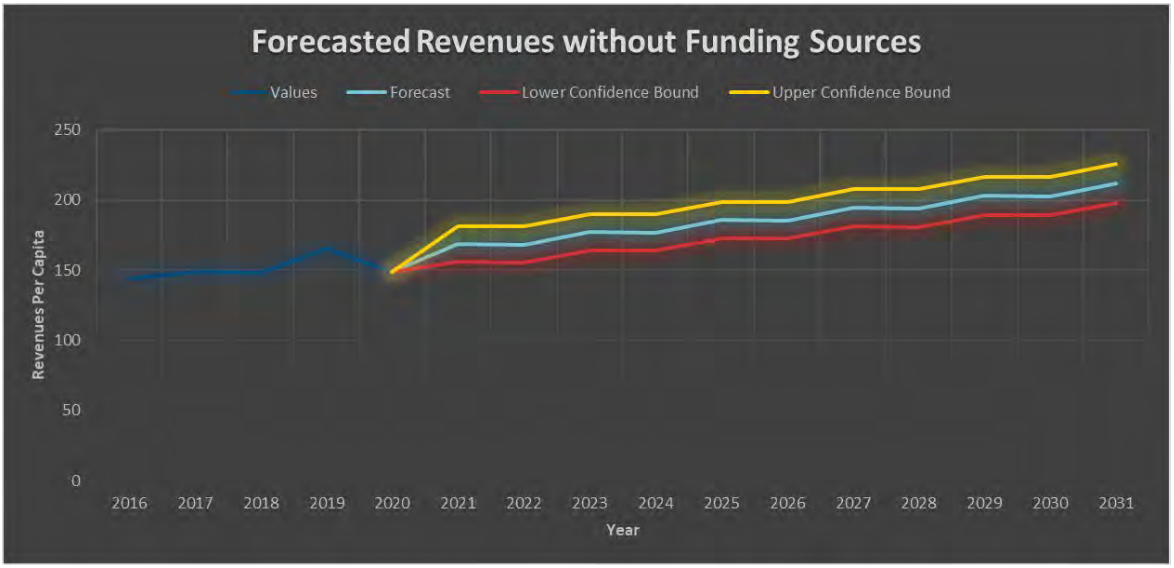


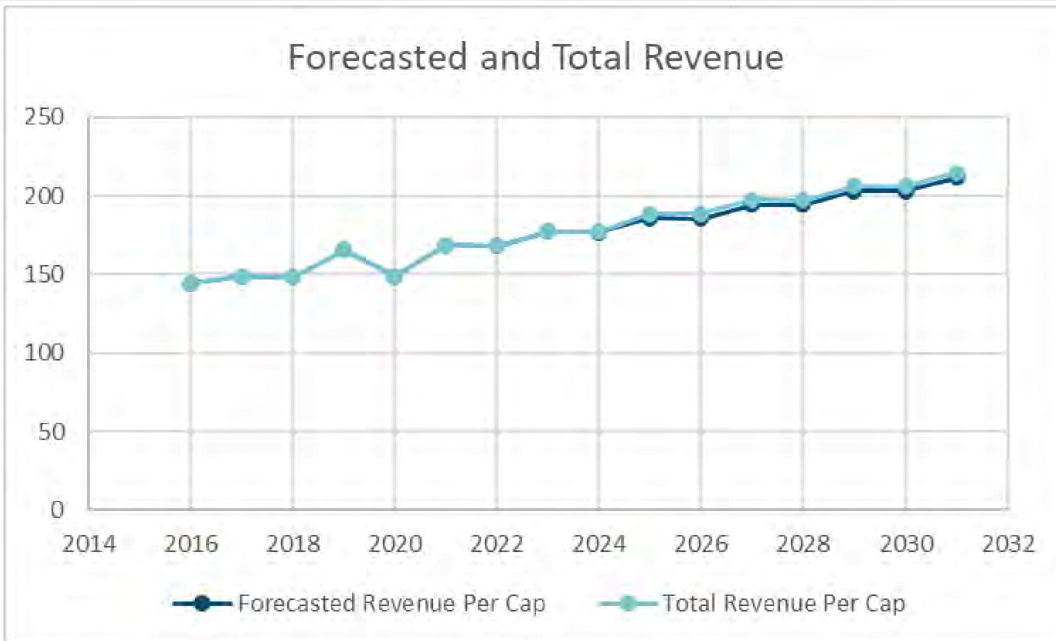
F3: Revenues R Code

```
data <- LastChance
View(data)

T1 <-glm(X2020TotalRevenue ~ X2020LeagueAttend + X2020TotalSocialMedia + GDP2020 -
1,data=subset(data))
summary(T1)

T2 <-glm(X2019TotalRevenue ~ X2019LeagueAttend + X2019TotalSocialMedia + GDP2019 -
1,data=subset(data))
summary(T2)

T3 <-glm(X2018TotalRevenue ~ X2018LeagueAttend + X2018TotalSocialMedia + GDP2018 -
1,data=subset(data))
summary(T3)

T4 <-glm(X2017TotalRevenue ~ X2017LeagueAttend + X2017TotalSocialMedia + GDP2017 -
1,data=subset(data))
summary(T4)
```

```
T5 <-glm(X2016TotalRevenue ~ X2016LeagueAttend + X2016TotalSocialMedia + GDP2016 -
1,data=subset(data))
summary(T5)

rarita2020 <- RaritaLCinput[c(7, 12, 21)]
View(rarita2020)
predOut2020 <- predict(object= T1, newdata = rarita2020, type = "response")
print(predOut2020)

rarita2019 <- RaritaLCinput[c(8, 13, 20)]
View(rarita2019)
predOut2019 <- predict(object= T2, newdata = rarita2019, type = "response")
print(predOut2019)

rarita2018 <- RaritaLCinput[c(9, 14, 19)]
View(rarita2018)
predOut2018 <- predict(object= T3, newdata = rarita2018, type = "response")
print(predOut2018)


rarita2017 <- RaritaLCinput[c(10, 15, 18)]
View(rarita2017)
predOut2017 <- predict(object= T4, newdata = rarita2017, type = "response")
print(predOut2017)

rarita2016 <- RaritaLCinput[c(11, 16, 17)]
View(rarita2016)
predOut2016 <- predict(object= T5, newdata = rarita2016, type = "response")
print(predOut2016)
```

F4: Forecasted Revenues


Forecasted Revenues without Funding Sources
Values
Forecast
Lower Confidence Bound
Upper Confidence Bound
Revenues Per Capita
250
200
150
100
50
0
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
Year


Forecasted and Total Revenue
250
200
150
100
50
0
2014
2016
2018
2020
2022
2024
2026
2028
2030
2032
Forecasted Revenue Per Cap
Total Revenue Per Cap