Final Report for Spring 2022

TEAM 23: FIXTURE

Project Title: Fixture: Fantasy Soccer Optimizer

Team Members: Gabriel Frydman (SSE), gfrydman@seas.upenn.edu
Prerit Kothari (SSE), preritk@wharton.upenn.edu
Mary Passarella (CIS), marypass@seas.upenn.edu
Claes Sjoborg (SSE), claes@seas.upenn.edu
Alexa Spagnola (EE), lex@seas.upenn.edu
(Please note that this is an Inter-Departmental Senior Design Project)

Advisor: Hamed Hassani, hassani@seas.upenn.edu

Date: Apr 26, 2022

Table of Contents

I. Cover Page **1**

II. Executive Summary **4**

III. Overview of Project **5**

IV. Technical Description **5**
- Specifications and Requirement 5
- Iterations 5
- Societal, Environments, or economic considerations 6
- Technical description and approach 6
- Final status of the project and test results 9
- Overall Evaluation 9
- Conclusion 10

V. Self-Learning **10**
- Self-Learning 11
- Useful Courses 11

VI. Ethical and Professional Responsibilities **12**
- Ethical Responsibilities 12
- Professional Responsibilities 12

VII. Meetings **12**

VIII. Reflection on Milestones **13**

IX. Discussion of Teamwork **13**
- Coordination 13
- Interdepartmental Senior Design Project 13

X. Budget & Documentation **14**

XI. Standards and Compliance **14**

XII. Work Done Since Last Semester **14**
- Technical Work and Progress 14
- Business Plan 15
  - Need and Value Proposition 15
  - Stakeholders 15
  - Opportunity for Problem Area 16
  - Market Size Estimation & Segmentation 16
  - Revenue Model 17
  - Cost & Competition 17
  - Intellectual Property 18

XIII. Discussion and Conclusion **18**

XII. Appendices **19**

Appendix A. Linear Regression Example 19
Appendix B. Excerpt from code 19
Appendix C. Website Homepage 20
Appendix D. Website Info Page 20
Appendix E. Website Player Similarity Page 21
Appendix F. Website Log in/Sign Up Page 21
Appendix G. Website Favorites/Social Media Page 21
Appendix H. Original Budget Plan 22

II. Executive Summary

Last semester, we set out our objective to create a tool that can help fantasy soccer league players unlock maximum potential of their constrained budget and have the highest likelihood of winning the prize money. In order to do so, we had multiple goals for the fall and the spring semester. In the fall semester, we were able to create a minimum viable product (i.e. the underlying model) using in-depth player statistics to optimize budget choices made by our user based on player's value at a given time. In the spring semester, we were able to successfully create a tool that helps users make an optimal choice based on performance statistics and future-centric (expectation-based) metrics. In the spring semester, we also developed a user-friendly UI that recommends players based on a given input of positions and budget. Our ultimate goal is to have a dynamic model factoring in only active players. Not only did we aim to solve the problem of information asymmetry in the market, but also level the playing field between fantasy players who have a lot of experience versus fantasy players who are just getting into managing a team. We tested our model at the end of the semester by comparing it to industry standards and using an objective metric of change in points that we developed ourselves.

This report is a comprehensive document going over all the technical, ethical and business descriptions of our project. We also touch upon the progress we made over this semester and when and how we achieved our milestones. We started our work this semester by developing a complete timeline on how we plan to achieve our and then divided responsibilities amongst each team member. In order to refine the model we built in the fall semester, we incorporated data from both *Statsbomb* and FBref. After successfully cleaning and wrangling the data, we ran multiple linear regressions. The key stakeholders (beyond this project team) for this project were fantasy soccer aficionados, both expert and hobbyist, as users, soccer players currently on Premier League teams, owners of the databases from which we pull player statistics. Given that the Premier League has been facing certain racism when it comes to player performance, we were hoping our model coulf, in whatever measure, have a social impact and let people see an objective analysis of players through statistics instead of the color of their skin. We adhered to certain engineering standards for quality assurance such as finding a dataset with as minimal inherent bias as possible was paramount.

Finally, as mentioned above, our testing procedures included comparing performance with other existing predictive models and computing a change in total points due to a suggested change over a set period of time. This report also touches upon the budget and costs of developing the project, as well as a potential business plan to scale and monetize our operations by considering the overall market.

III. Overview of Project

When it comes to fantasy sports, the goal is to maximize the points you can earn by picking a roster of players that is bound to win matches. At a high level, Fixture is a tool to help those interested in making a fantasy soccer roster that optimizes both player value and budget. Using the Fixture website, users can select the position they are interested in receiving a player recommendation for and our model will suggest the player(s) with the best value. Fixture levels the playing field between those with a high level of understanding of fantasy soccer and player value and those who are new to fantasy soccer or sports betting as a whole. Within our website, we also created log in options so users can create accounts and save their favorites. It also allows the users to directly share any news or information from the website directly on their social media handles if they wish to do so. Fixture aims to be the best new tool in the fantasy soccer market, giving fantasy soccer aficionados and newcomers alike the opportunity to maximize points and create the best roster available.

IV. Technical Description

A. *Specifications and Requirement*

- *Python*, specifically using *Google Colab* as our IDE, has been our exclusive development environment. Here we were able to perform all the normal Python functions as well as utilize SQL functionality and import any necessary Python packages/libraries such as pandas and numpy.
- *JavaScript, HTML/CSS, and Microsoft FluentUI* were the main tools used to develop the front-end component (our website) and UI. Microsoft FluentUI is a set of React components that helps create a streamlined and cohesive look to a website.
- *GPU/CPU Performance* is a constraint to consider. When we initially build our model on data from one year, this is not as important. However, when we begin to factor in data from the last 5-10 years, time to run statistical operations on massive datasets will become a time constraint.
- We had next to no costs associated with our project, since all the data we used for making our model was publicly available or requested from sources free of charge
- The main regulation we had to take into consideration as we launched our project/website was users' privacy so we did not collect user data even though that can be a possibility in the future to scale the project

B. Iterations

- *Alternative Solutions* - Whilst we landed on creating a points prediction model using a combination of statistical techniques, we also considered a more traditional optimization approach. This would consist of constructing an objective function designed to maximize player points and then introduce budget and other constraints (e.g. chemistry, such that the outputted a recommendation that would

best fit the conditions specific to the user's situation. However, we decided that we wanted complexity in our model so would need to use more complex statistical techniques in order to compete with the industry standard.

- *Model Iterations* - In the fall, we set out to build the most technically complex and accurate model we could using RNN, LSTM. Speaking to subject matter experts, head of data science at FanDuel, we realized explainability is equally as important. By then Incorporating feedback from our fellow classmates and TAs, we realized an educational tool would be extremely valuable
- Website Iterations:
    - Version 0.1.0
        - MVP; only worked for Goalkeepers
    - Version 0.1.1
        - Separate by position or generate a full roster, sign in/out, shareable to FB/Twitter
    - Version 0.1.2 (Current)
        - Version 0.1.1 + player similarity, cohesive learning tool for fantasy soccer

C. Societal, Environments, or economic considerations

- The main implication we had to consider for this was the accuracy of our project. We made it clear through a disclaimer to our users that this website is only a basis for recommending players and in no way do we guarantee that our predictions will always be accurate. There were no environmental factors related to our project. Since the project is currently a free service, we do not have any economic considerations to make at the moment.
- Another socio-economic factor we considered was ensuring that our tool does not promote gambling. We make it clear through a disclaimer that we are not a betting website and no fantasy betting can be done through our portal. We have simply developed a tool to help people enjoy fantasy soccerl more and become better at it

D. Technical description and approach

- Our end goal is to output a player that is best suited for a person's fantasy soccer team. This recommendation is based on the expected points that this player will achieve on that week as well as opposition strength and budget constraints. Instead of training our model on just stats like shots, shots on target, assists etc, we wanted to incorporate underlying statistics, like Expected Goals/Assists Per 90, as well in order to more accurately predict performance over time. An overview of our solution can
  be seen below.

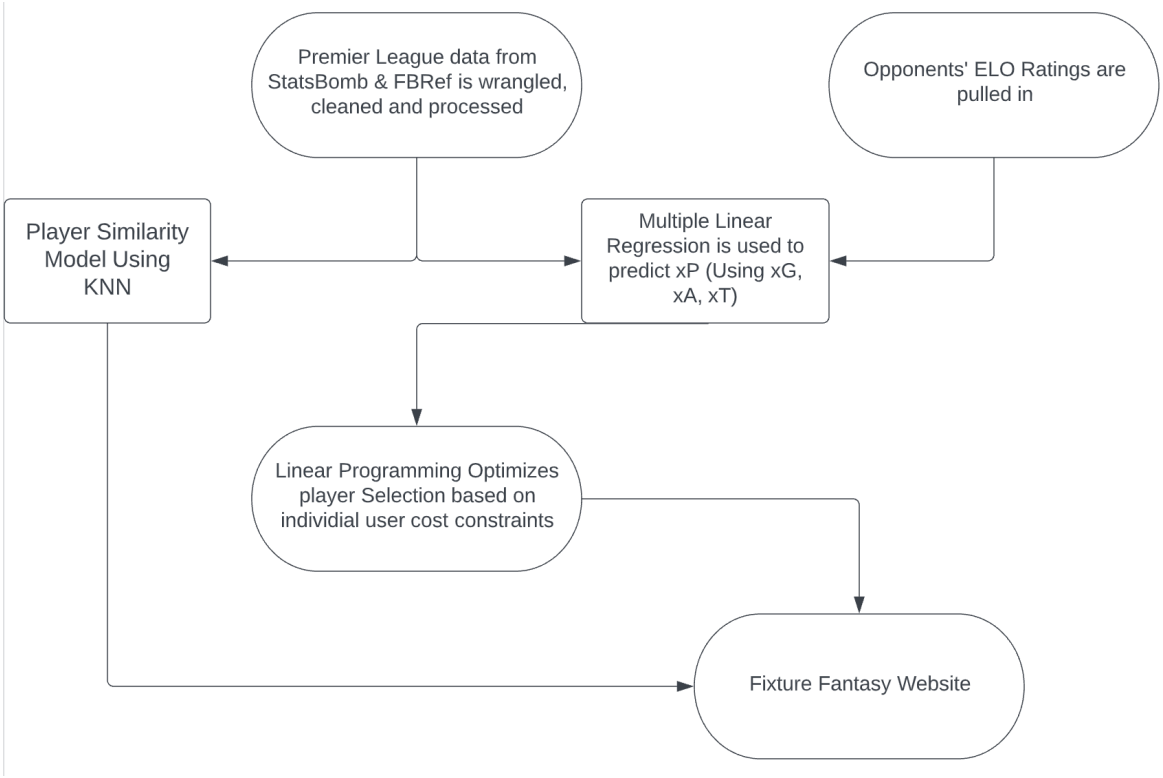


Fig 1. Overview of Solution

- We started by collecting data from multiple reliable sources (STATBOMB, FBREF, Fantasy Premier League). Substantial Data Wrangling was required in order to make sure the data was in the correct format for Linear Regression. This included combining several datasets, dropping unnecessary columns, and filtering on specific positions. For example, below is an excerpt from our Goalkeepers (GK) dataset. It shows some of the metrics that were used to evaluate expected points on a match-by-match basis.

	name	position	team	xP	assists	bonus	bps	clean_sheets	creativity	element
3	Aaron Ramsdale	GK	Sheffield Utd	2.5	0	0	12	0	0.0	483
19	Alex McCarthy	GK	Southampton	1.8	0	0	17	0	0.0	363
23	Alisson Ramses Becker	GK	Liverpool	5.5	0	0	9	0	0.0	252
51	Bernd Leno	GK	Arsenal	3.5	0	1	29	1	0.0	8
155	Hugo Lloris	GK	Spurs	3.8	0	0	15	0	0.0	383

Fig 2. Cleaned GK Data

- Continuing with the same example of goalkeepers, we ran a simple linear regression model on the goalkeeper data.
  We used an industry standard normalization formula for different Goalkeeper activities:
  (goals conceded * -1.37) + (saves * 0.26) + (touches * 0.05) + (minutes * 0.07) + 8.50

- Below is a visualized result of our model:

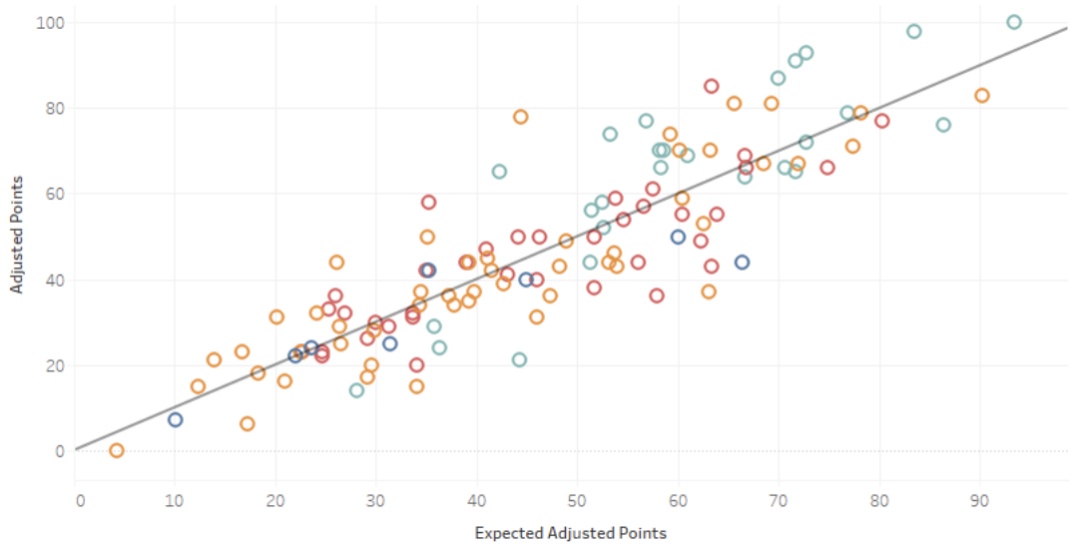
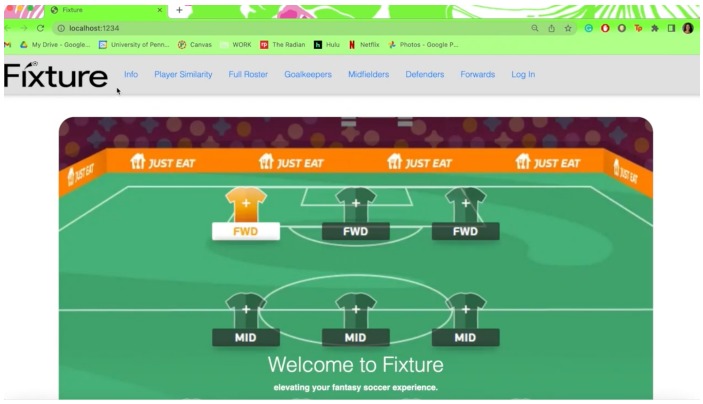


Fig 3. GoalKeeper Expected points model

E. Final status of the project and test results

- The image below shows the final user interface of our website with the different pages displaying in the header



- In *Appendices C - G,* the user interface of each page within the website can be observed. In our final product, we have been able to develop a very simple to use interface that both experts and novices can easily interact with and create the best possible team for themselves.
- Our test results indicated that our model suggested the optimal player more than 90% of the times based on the users' requirements. We tested our tool by comparing it to the industry standards and the choices of expert players given a certain budget constraint and player position.

F. Overall Evaluation

- When we evaluate our project ourselves over the course of the two semesters, we feel delighted about the progress we made and the milestones we were able to achieve. Not only did we deliver on a fully functional website with a complex back-end, but also learn a lot in our journey. We also received very positive and constructive feedback along the way, some of which is mentioned below
- We received many pieces of useful feedback this semester from Students, Department Professors, and real-world practitioners. From Students, we were told that we could think about different leagues from all over the world when making fantasy recommendations where we had previously only thought to look at the English Premier League. From Professors, we were able to really evaluate the technical needs of our project and get feedback on which models and methods we should be looking at given the data we have and output we want to create. We were fortunate to be able to meet with the VP of Data Science at FanDuel Sports Keith Goldner. He offered us a lot of helpful advice about defining the scope of our project and which parameters matter more in different settings. For example, in a fantasy setting, when playing a Tournament, you want to maximize expected player value whereas in Head-To-Head play you want to maximize overall upside.
- One piece of feedback that our team disagreed with concerned the complexity of our initial model. Whilst we were told to design a complex neural network for our model, we were contrastingly advised by Keith Goldner to start with a simple model and then build from there. This has been incredibly helpful advice as a neural network might have yielded better overall results, by starting with simple Regression, we were able to identify key predictors moving forward.

G. Conclusion

Our goals of the project were:

- Solving the asymmetry of information in fantasy sports
- Helping beginners who find it difficult to make a team that makes sense according to the statistics and chemistry of players
- We wanted to build a tool that can help level the playing field
- Want people to enjoy fantasy soccer more and provide a financial incentive for more people to participate

In conclusion, when we look at the tool we finally developed, we believe that we were

able to achieve the objectives we defined at the beginning of the year. We calculated an adjusted R^2 value of 0.757 meaning whilst not displaying a 1:1 relationship, there is a strong correlation between expected goals conceded, saves, touches, and minutes to be a strong predictor of FPL points. Interestingly, there is quite a distribution of value categories with many goalkeepers (in the top right hand corner) having overperformed against their expected values. We also interestingly found that between expensive and cheap players, there is little variation compared to their expected performance.

We also conducted different forms of more in depth analysis using neural networks that increased the complexity and accuracy of the model. One of the biggest takeaways from the MVP we put together is that our data is time indexed, in our case by gameweek, and therefore was suitable for a recurrent neural network. We hope to explore a range of recurrent neural networks to gain a more in-depth understanding of player performance and build a more accurate model if we plan to scale this as a startup. Specifically, we will be looking into using Long Short Term Neural Networks and Hidden Markov Models if we want to be able to monetize our tool.

V. Self-Learning

A. Self-Learning

Whilst some members of the team were familiar with the statistical concepts used in this project, we still required help, which we got from real-world practitioners and Professors, to decide what kind of approach to take and which one was best. Handling real-world data is obviously very different from datasets used in classes. This meant learning how to deal with unclean data (data with missing information, incomplete columns, column names that are hard to identify) and cleaning it appropriately. We also had to do research on how to construct a website and develop the front end of our project as well. We learned a lot from professor Hassani about more complex techniques like Long Short Term networks. For the actual website development, we had to get more familiar with Microsoft Fluent UI and React + Javascript. We also had to learn how to employ new libraries we were not familiar with in Python through Colab which we needed to build the back-end.

B. Useful Courses

Some helpful classes for the model aspect of our project include ESE305: Foundations of Data Science, ESE402: Statistics of Data Science, CIS545: Big Data Analytics. ESE305 and ESE402 taught us the math underlying the statistical methodologies we employ. CIS545 taught us how to use Python to harness the well-known methods and implement them on a large scale to make meaningful predictions. Another class that helped us optimize a budget was ESE504: Optimization Theory.

VI. Ethical and Professional Responsibilities

A. Ethical Responsibilities

Our project presented certain ethical and professional responsibilities that we needed to address. First, since our users might be making changes that influence their fantasy rankings, they might win or lose money based on how well our suggested player performs. With this, there comes an ethical consideration of what happens if our suggestion does not perform as well as we predicted and the user makes those changes. They might hold our algorithm or team responsible for losing a particular bet or game against other fantasy players. We dealt with this by giving a warning to every user as soon as they land on our website that our suggestions are just predictions made on statistics and fixtures, and that they are themselves responsible for any changes they make on their team.

B. Professional Responsibilities

Second, to maintain user privacy, we did not attach any additional tools to our website that can help us track user activity. If we plan to do that at any point in the future, we will get consent from individual users so that we can try to improve our tool for their benefit. Third, since we might have a login feature for those users who plan to have subscription based plans, we need to ensure that their data stays private, encrypted and secured. We would not like anyone, even within our own team, to be able to read usernames and passwords, which means that all this information was encrypted. We also have professional responsibilities of making sure our security for data accessibility is strong enough that no one can get access to this data.

VII. Meetings

We set up bi-weekly meetings with our advisor, Professor Hassani, to hold ourselves accountable in meeting our designated milestones over the course of the fall semester as well as to receive feedback and keep us on track with our project. Our meeting with FanDuel VP of Data Science Keith Goldner provided us with real-world insights into where we should focus the scope of our project and how to prioritize our assortment of data points. For instance, we confirmed that using expected goals rather than actual goals would result in a more accurate model. We also gained clarity on what the current standard is for predictive analytics in fantasy sports, and motivated us to endeavor to utilize stacking, which would take into account how various players perform when interacting with one another rather than on an isolated basis. As for the members of the team itself, we all met every week at the very minimum to work on the project. While some of the meetings were scheduled physically, some were just on zoom to touch base regarding progress and discuss any new ideas or iterations.

VIII. Reflection on Milestones

The milestone assignments were a great way to keep the team on schedule while also giving us opportunities to receive valuable feedback. Our initial milestones for the project at the beginning of Spring were:

1. Building a simple user interface through a website
2. Building the player similarity model for all four types of positions (we only did this for goalkeepers in the Fall semester)
3. Linear programming for cost optimization with the budget as a constraint
4. Building a full roster player similarity model using KNN
5. Connecting the models with the user interface we build
6. Writing and attaching disclaimers where needed to make sure we are ethically and professionally being responsible
7. Use sentiment analysis through Twitter to remover biases from the value of player and achieve an overall accuracy of at least 80% in the final testing

Reflecting on our milestones, we were able to achieve all of them except the sentiment analysis through Twitter. This is because we achieved approximately 90% accuracy in our prediction even without the sentiment analysis, and realized that adding this complexity would not improve our tool by much. In fact, there was a possibility that it might reduce accuracy since sentiment analysis might provide conflicting outputs when compared to the output of the linear programming, and so we decided to not incorporate it.

IX. Discussion of Teamwork

A. Coordination

Our team learned how to successfully navigate working together virtually. We regularly speak on Zoom calls to address tasks, discuss subdivision of work, and hold one another accountable. With a unique blend of personalities and skill sets, we have found our team to be well-rounded and synergistic.

B. Interdepartmental Senior Design Project

Gabe and Claes gravitated towards tackling the data collection and data cleaning tasks throughout the semester while working on the backend. Prerit, as an M&T student, led the development of potential business applications and worked on linear optimization. Mary, a CIS student, developed the front-end and UI for Fixture, and Alexa, an electrical engineering student, worked on documenting our progress, coordinating with and helping all team members. All team members worked equally on writing reports and making presentations. Unfortunately, Prerit was stuck in India due to Covid and other personal reasons, so he could not be physically present for the presentations. Teammates also adjusted and filled in gaps wherever help was

needed. Our team has learned to communicate and ask for assistance from other team members when necessary, designating a particular person to take ownership of a task and submit the required materials for grading.

X. Budget & Documentation

Appendix H shows a review of our original planned budget from Spring Semester. This total budget was $300. Much to the delight of our team and even more so SEAS, our final expenditure turned out to be $0. This is because we carefully chose the tools we wanted to employ to develop our website, and chose to only use data available for free online or at request. For back-end and front-end development, we only used software tools that were available to Penn students free of cost, hence developing all of Fixture just through our efforts and no expenditure. By the end of the semester, we simply did not have enough time to get the website hosted, which is where we would have incurred costs for AWS or a similar platform to host.

XI. Standards and Compliance

We complied with all engineering standards throughout the development of the project. Our final product is at par with other competitors in the industry and just like them adheres to the required engineering standards. We briefly reviewed IEEE-SA, ASTM, AAMI, ISO 9001, and NSSN to make sure we were not violating any compliances before we started development and once we finished. We used open source datasets from industry leading soccer data collectors (FBREF & STATSBOMB) and had to take into account when new players join teams and track injuries to see which players are available each week.

XII. Work Done Since Last Semester

A. Technical Work and Progress

All the technical work we did can be found under the reflection on milestones section where we discuss our 7 milestones and the extent to which each was achieved. We developed a successful model of Fixture that can help fantasy soccer players and attached a very simple but interactive user interface to it through our website. The UIUX can be seen in Appendices C-G. We also successfully completed all the required documentations, presentations and assignments for the Spring Semester well before the deadlines while simultaneously developing our final product. As an M&T requirement, we also developed a business plan related to our project this semester. As per Sangeeta's request, we have attached our business plan in this final report as a subsection of the work done since last semester. Please find this below.

B. Business Plan

● Need and Value Proposition

Every year Fantasy sports lose a ton of potential revenue since new fantasy players find it tough to navigate the sport from a management standpoint. Even though the fantasy industry is approximately $60- $70 billion right now with a greater than 10% compounded annual growth rate, it has a potential of being a much larger industry if players could easily enter any sport and start being competitive in organizing their team. Therefore, our project objective is to create a tool that can help fantasy soccer league players make the most of their budget and have the highest likelihood of winning the prize money. When it comes to fantasy sports, the goal is to maximize the points you can earn by picking a roster of players that is bound to win matches. At a high level, Fixture is a tool to help those interested in making a fantasy soccer roster that optimizes both player value and budget. Using the Fixture website, users can select the position they are interested in receiving a player recommendation for and our model will suggest the player(s) with the best value. Fixture levels the playing field between those with a high level of understanding of fantasy soccer and player value and those who are new to fantasy soccer or sports betting as a whole. This is the value we aim to create for our users.

● Stakeholders

The key stakeholders (beyond this project team) for this project are fantasy soccer aficionados, both expert and hobbyists (as users), soccer players currently on Premier League teams, owners of the databases from which we pull player statistics. Given that the Premier League has been facing certain racism when it comes to player performance, we are hoping our model can, in whatever measure, have a social impact and let people see an objective analysis of players through statistics instead of the color of their skin.

● Opportunity for Problem Area

Current tools have a very complicated interface that only experts are easily able to navigate, and tend to dissuade new players from getting into the game. We plan to keep our interface as simple as possible for easy interaction. In addition, having an information tab that can be educational for new players can give us an edge over competitors who only try to attract experienced fantasy players. In the long term, there is also an opportunity to convert into a fantasy platform where rosters can be created and bets can be made between different users.

- ## Market Size Estimation & Segmentation

We believe we can segment our market by current skill in fantasy of fantasy players and the league they watch and play fantasy for. According to fsga.org, there were 59 million fantasy players in 2017 with a 2 million YoY growth. This means the estimate for 2022 is 69 million players. From where we can segment the market by premier league players which is the league our tool focuses on, and we assume that should be about 50% of players since premier is the most watched league in the world. This gives us 34.5 million premier league fantasy players just in the US. Now, we can segment by skill. Our target market would be amateurs, which we assume account for a quarter of the total players, and that gives a total market size of 8.6 million players which we can easily target. Therefore, even if we get half of the market share in this specific segment, we can start making a revenue of $4.3 million a month with a $1 subscription fee per month. This does not even take into account the money we can generate through advertisement on our platform and data mining.

Below, you can see a snapshot of the current market size:

Figure: Global Fantasy Market Size

As we can see, there are bets being placed of about $3 billion by first-time better only in the US. Even if a small percentage of these bets are being made in Soccer, that gives us a market big enough to create our own space in. A revenue of $4m, as calculated above, would only account for about 1% of the market share in our target market and thus sounds like an achievable objective.

- ## Revenue Model

We thought of two different ways in which we can start monetizing from this project. Through these two options we could potentially generate revenue and eventually become a profitable business. The first route is through a subscription/one-time

payment to use the service and the other route is to make the service free initially and generate revenue through a big user base. The two approaches are explained in detail below:

1. The first model is a simple financial approach of generating revenue directly from the end-user. If we can build a model that consistently adds value for the user (initially can be used for free) and if our model outperforms other models in the market, users should be willing to pay for the service eventually as they can make higher returns through the bets they have placed on a fantasy group. Whether this approach should be a one-time payment method or a subscription method will depend on further market research and the general willingness to pay for a user. This can be called the freemium approach.
2. The second approach, which is harder to achieve unless the user base is big enough, is to generate revenue through other businesses while the users get to use the service for free. The two main sources of revenue in this case would be other businesses who would like to run ads if the website has a constant traffic of users, and other businesses or sports agencies who would like to purchase data in order to find out which players are gaining popularity in order to boost their own revenues through that information. This can be called the data-centric approach.

For both the approaches, it is imperative that our model adds value for the user. Otherwise, we would never be able to gain attraction or a big enough user base.

- ## Cost & Competition

In terms of development costs, we did not have any costs for creating the tool and our service itself. Even though our initial budget was $300, we developed the final website at absolutely no cost. Our real costs in scaling a project like this would be the salaries incurred for hiring engineers to improve and maintain our model and marketing costs for acquiring more and more users. We will need to raise a round of seed funding of around $1,000,000 to have a runway of around one year. This would include hiring 3-4 engineers other than the members of our team and 1-2 marketing specialists. Additional rounds of funding will depend on when we are able to start generating revenue, which we expect to take at least 6 months after launch.

There is some competition in this space from companies like FantasyPro's Draft Wizard and My Playbook, FantasyFootballer's Ultimate draft kit, and CSWR's Custom Ranking Sheet. This means that we will have a customer acquisition cost (CAC) associated with our startup because of marketing costs and switching costs. If we are able to choose the right channels of marketing and attain a CAC that is lower than the subscription fee we plan to charge, this venture can easily become profitable and investors would be likely to get a very high ROI on their investments.

- ### Intellectual Property

  Except the actual model we develop in the back-end, which should be encrypted and fully secure from access to others, there is no intellectual property we have in our project unless we plan to mine user data (after getting their consent) and create knowledge graphs from that data.

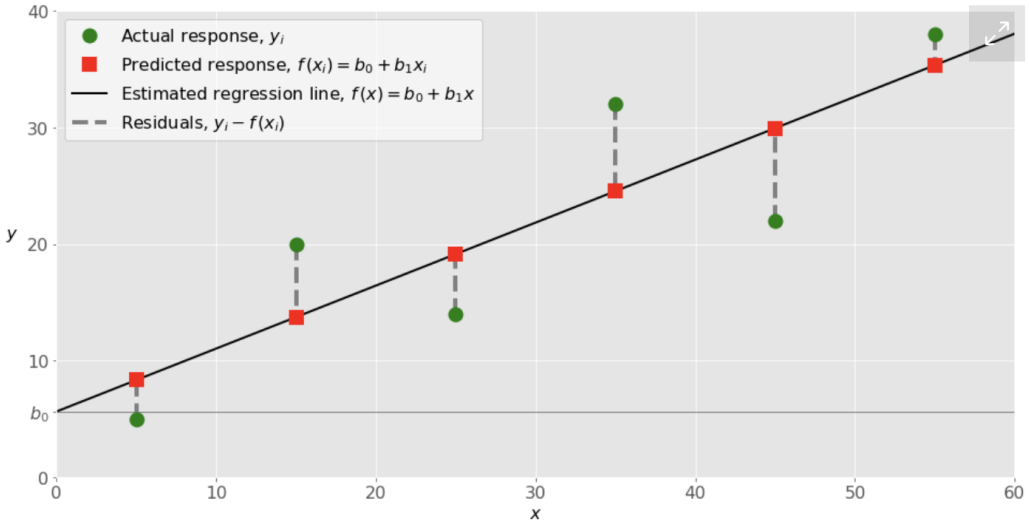
XIII. Discussion and Conclusion

Overall, we believe we had two very successful semesters in developing a fully functional website that is very much in alignment with the initial objectives we had decided at the beginning of the year. Starting with the ideation stage, we were able to develop our project to a working model with a simple but sleek interface by effectively working together. By dividing work to suit each team member's individual strengths, choosing a topic with real world impact that we are passionate about and by consistently working on our project milestones throughout the semester, we have been able to bring our vision to reality. Some of the challenges we faced were incorporating neural networks, integrating our model with AWS, and using sentiment analysis to understand a player's "hype". We learned a lot while trying to navigate these challenges. We also learned that it is very important to define the primary and secondary drivers of any project that we undertake in the future. No project can optimize cost, scope of the project, and time to execute all at the same time. This is because each one comes at a tradeoff from the other. For example, if the team makes the scope of the project really vast, then it will take a lot more time to finish it. Therefore, we learned that it is important to define which is the primary driver and secondary driver for a project out of scope, time and budget. Given our restricted budget and time, we were not able to fully incorporate sentiment analysis in our project or get the website hosted using AWS, and therefore we will remember this as a lesson. We would like to give credit to the Director of Data Science at FanDuel for communicating with us at multiple points of time and helping us structure our model. We would also like to thank all the professors, advisors and TAs associated with our project and senior design since we are grateful for getting this opportunity.

XII. Appendices

Appendix A. Linear Regression Example

The following figure illustrates simple linear regression:



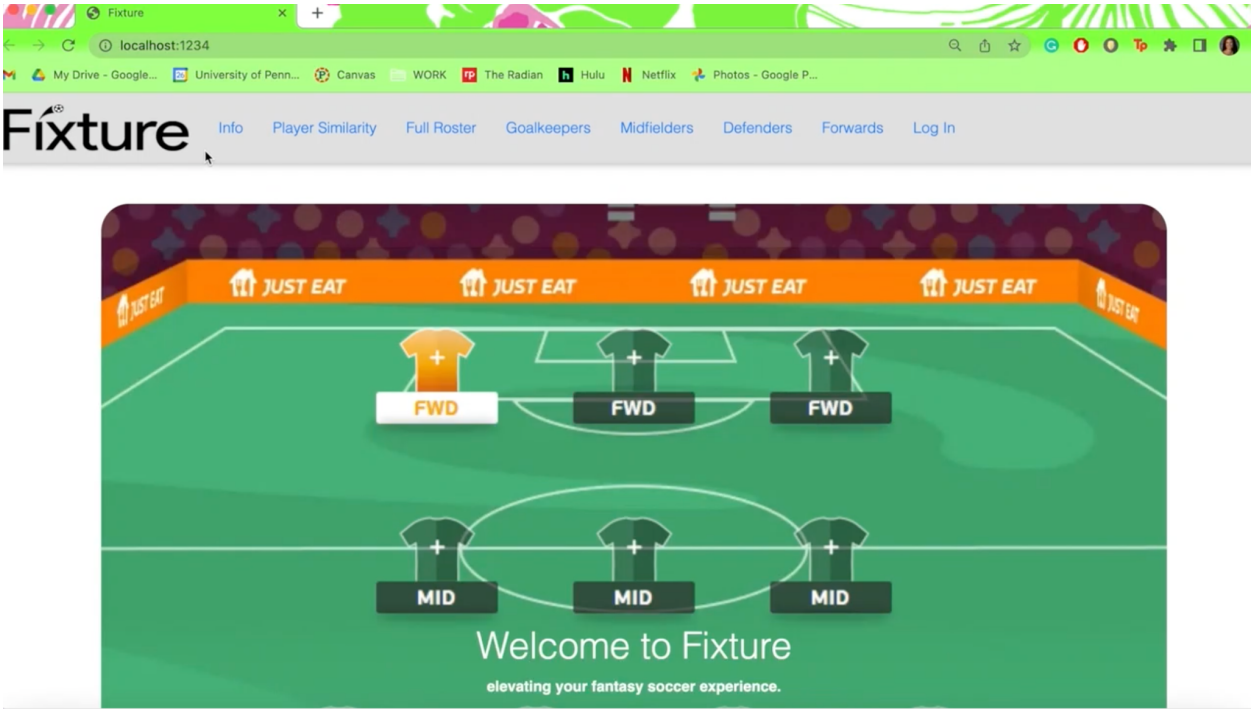
Example of simple linear regression

Appendix B. Excerpt from code

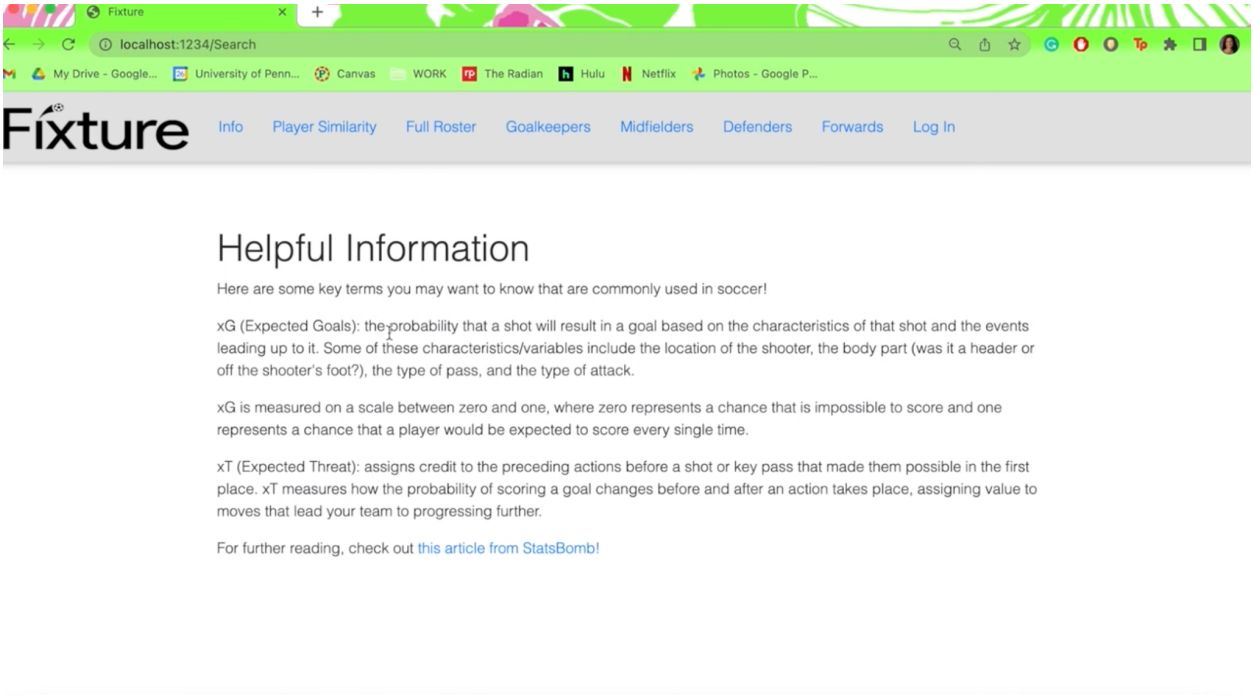
Adding opposition ratings to dataset for goalkeepers.

```
#Adding Elo ratings to df

#GW 1-5
conditions_wl_5 = [(gks['opponent_team'] == 'Arsenal') & (gks['GW'] == 1),(gks['opponent_team'] == 'Aston Villa') & (gks['GW'] == 1),(gks['opponent_team'] == 'Brighton
                ,(gks['opponent_team'] == 'Arsenal') & (gks['GW'] == 2),(gks['opponent_team'] == 'Aston Villa') & (gks['GW'] == 2),(gks['opponent_team'] == 'Brighton')
                ,(gks['opponent_team'] == 'Arsenal') & (gks['GW'] == 3),(gks['opponent_team'] == 'Aston Villa') & (gks['GW'] == 3),(gks['opponent_team'] == 'Brighton')
                ,(gks['opponent_team'] == 'Arsenal') & (gks['GW'] == 4),(gks['opponent_team'] == 'Aston Villa') & (gks['GW'] == 4),(gks['opponent_team'] == 'Brighton')
                ,(gks['opponent_team'] == 'Arsenal') & (gks['GW'] == 5),(gks['opponent_team'] == 'Aston Villa') & (gks['GW'] == 5),(gks['opponent_team'] == 'Brighton')

values_wl_5 = [*eng_matchday_1_elos['Elo'], *eng_matchday_2_elos['Elo'],*eng_matchday_3_elos['Elo'],*eng_matchday_4_elos['Elo'],*eng_matchday_5_elos['Elo']]

gks['Elo'] = np.select(conditions_wl_5, values_wl_5)
```

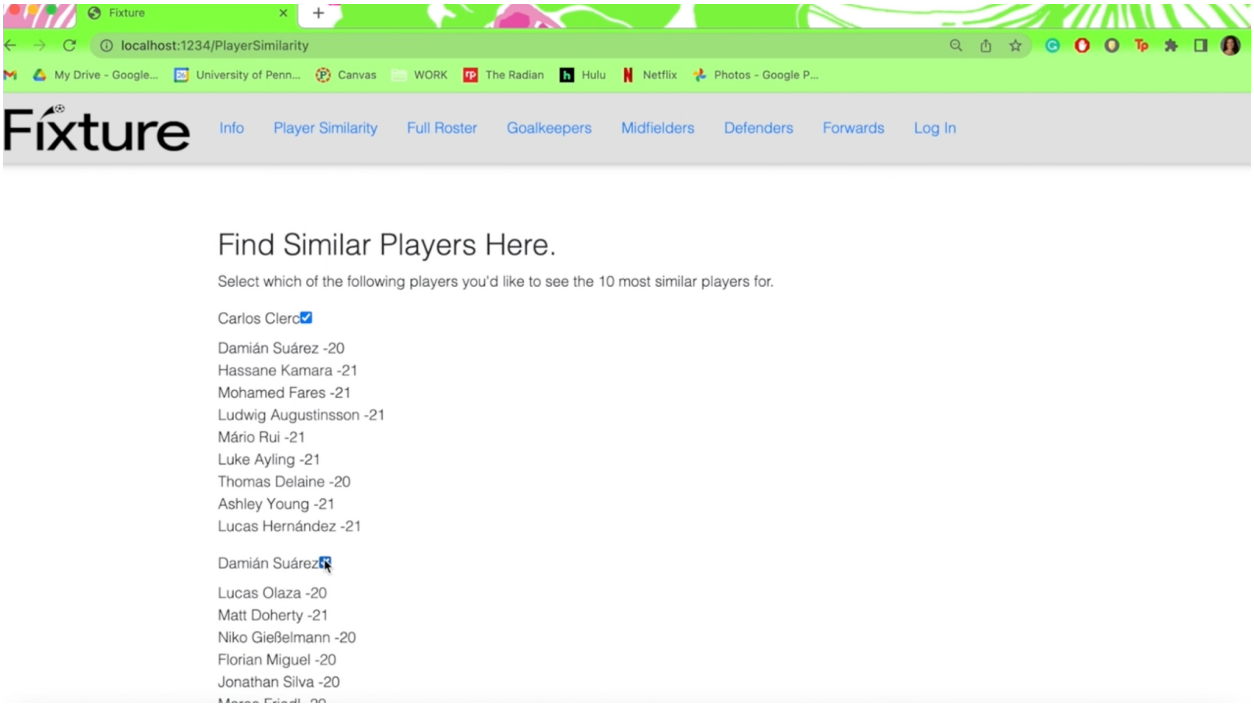
Appendix C. Website Homepage



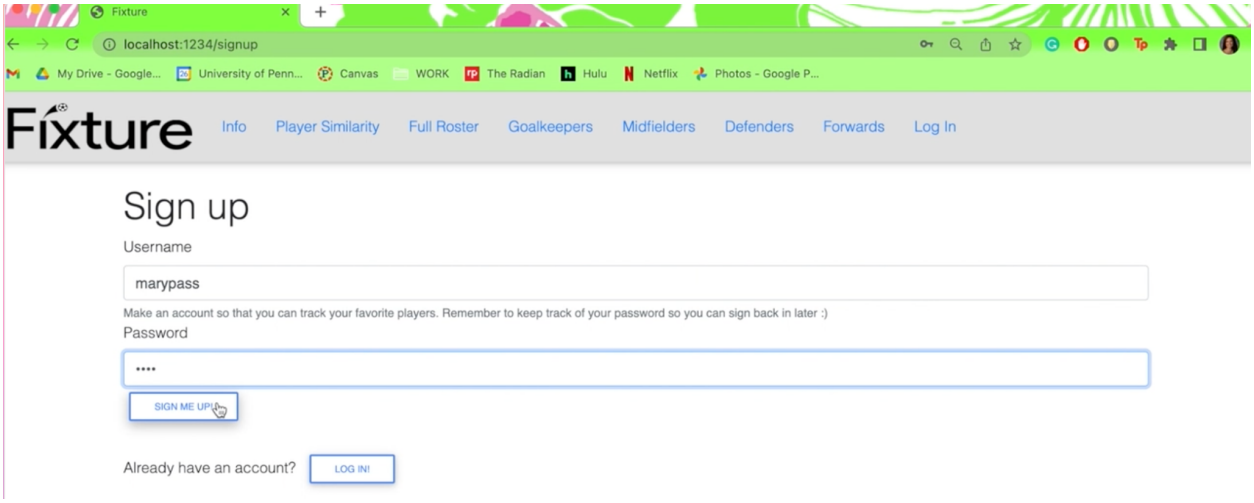
Appendix D. Website Info Page



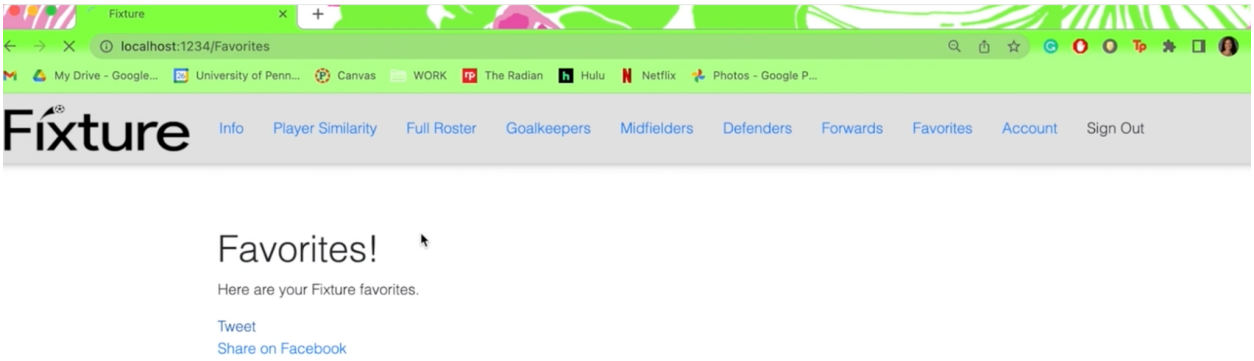
Appendix E. Website Player Similarity Page



Appendix F. Website Log in/Sign Up Page



Appendix G. Website Favorites/Social Media Page



Appendix H. Original Budget Plan

TEAM 23 BUDGET

Description	Estimate unit cost	Quantity	Source of funding	Subtotal
AWS Software Tools	$150	1	ESE	$150
Data for Analytics (Online)	$100	1	ESE	$100
Domain	$50	1	ESE	$50
Software Usage	$50	1	ESE	$50
Total				$350