Football Analytics: Strategy Builder

Alpha Prototype Report

Group 8

Aadith Krishna

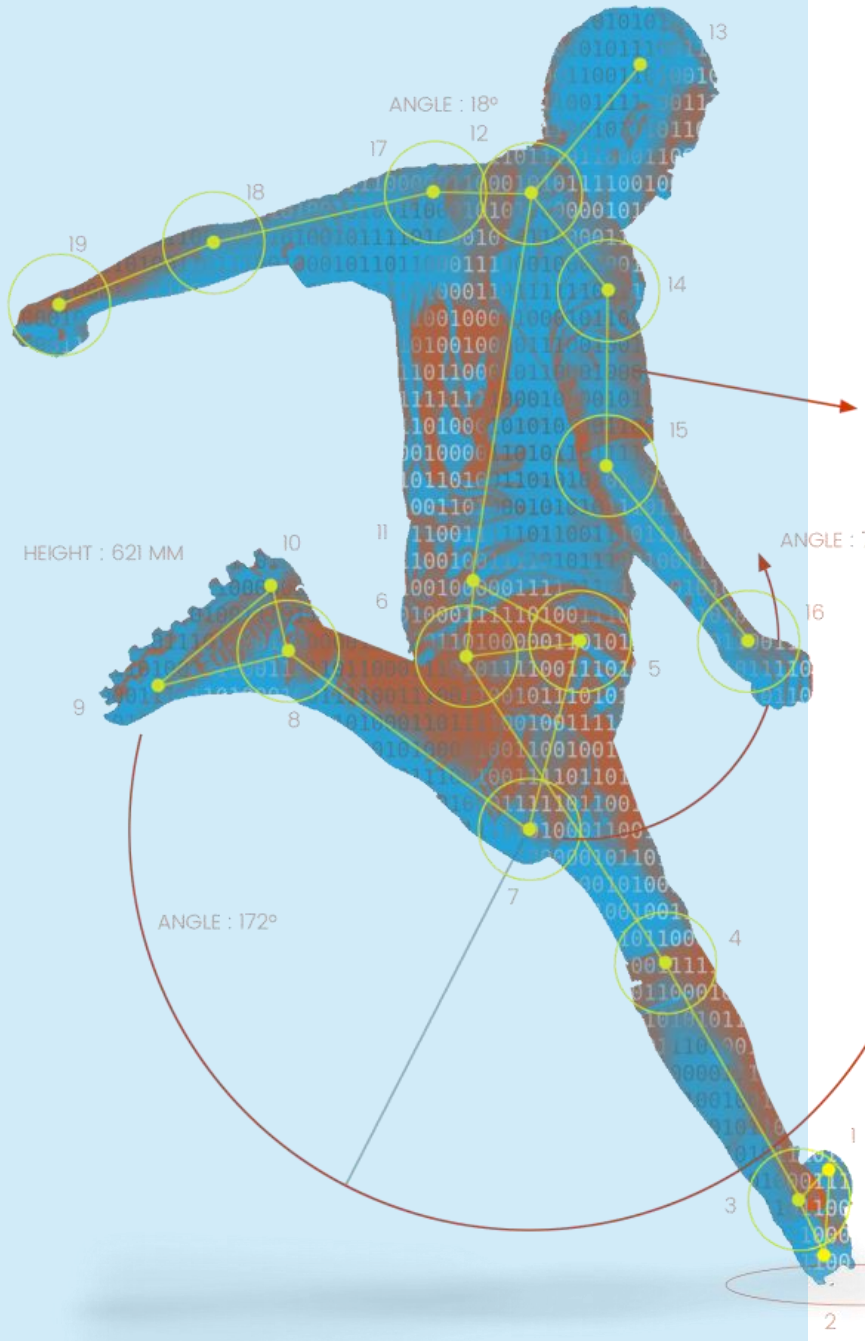
Conor O Donoghue

Nikhil Mohan Joshi

Philson Manarcad

Thomas Donal Murphy

Yash Joshi



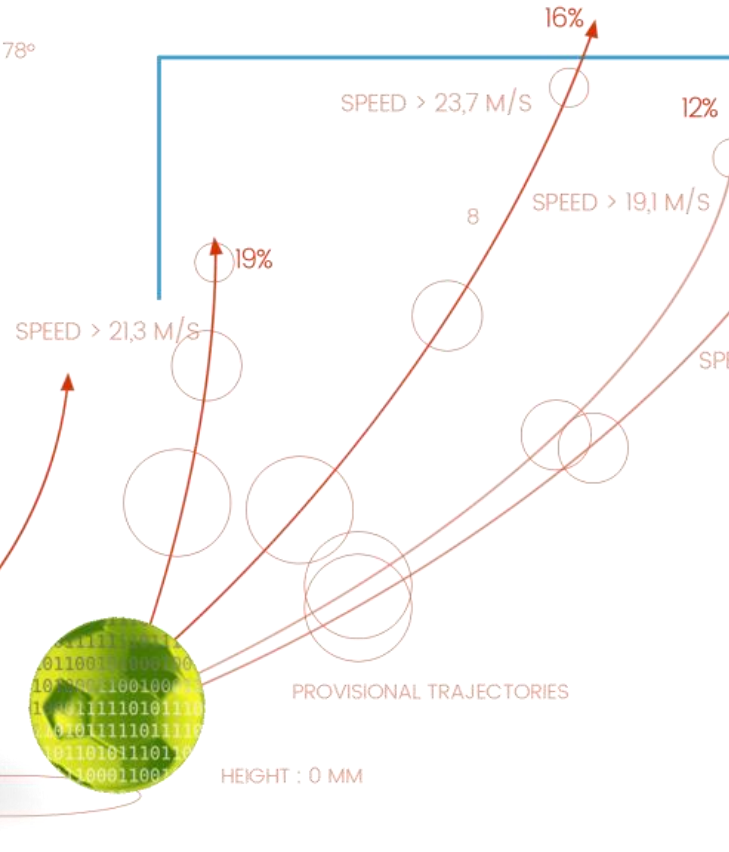
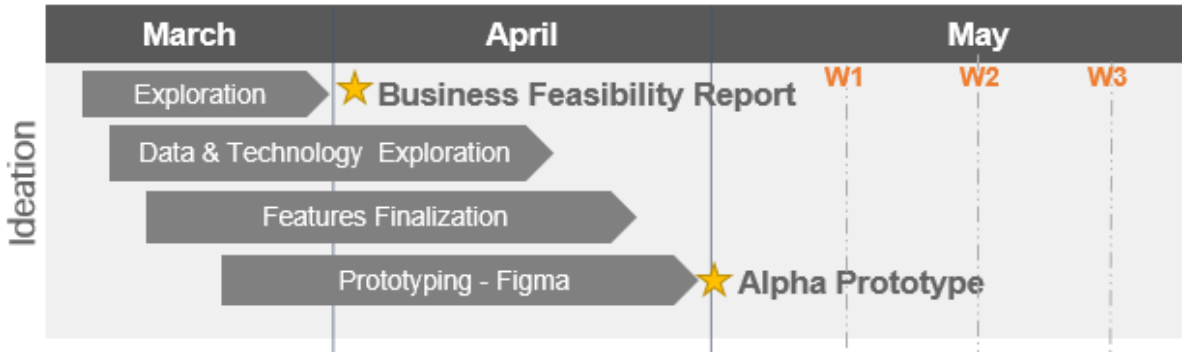


Table of Contents

Cover

The project has been divided into four phases. **Ideation**, **Backend Development**, **Front End Development**, and the **Business Plan** are the steps in the end-to-end cycle. Until now, the team has concentrated on the Ideation phase. Based on what to offer to the lead customer (*in this case the Low Tier football Clubs),* we have been able to finalize the Final Idea. The team also spent time evaluating the specific features/metrics that should be included in the final product, as well as the data and technology that should be used to implement them.

Since the product being developed is focused on delivering information/insights to the consumer, a thorough investigation into how to present them in the product was undertaken, which was then transformed into a working prototype in Figma. The prototyping was carried out in accordance with the Design Thinking Principles learned in the first semester.



Ideation Phase in Project RoadMap

The prototype that's been produced is capable of demonstrating how the proposed product would seem and how users may use it to strategize for games. The Upcoming section will go over a full description of each view and the features inside it.

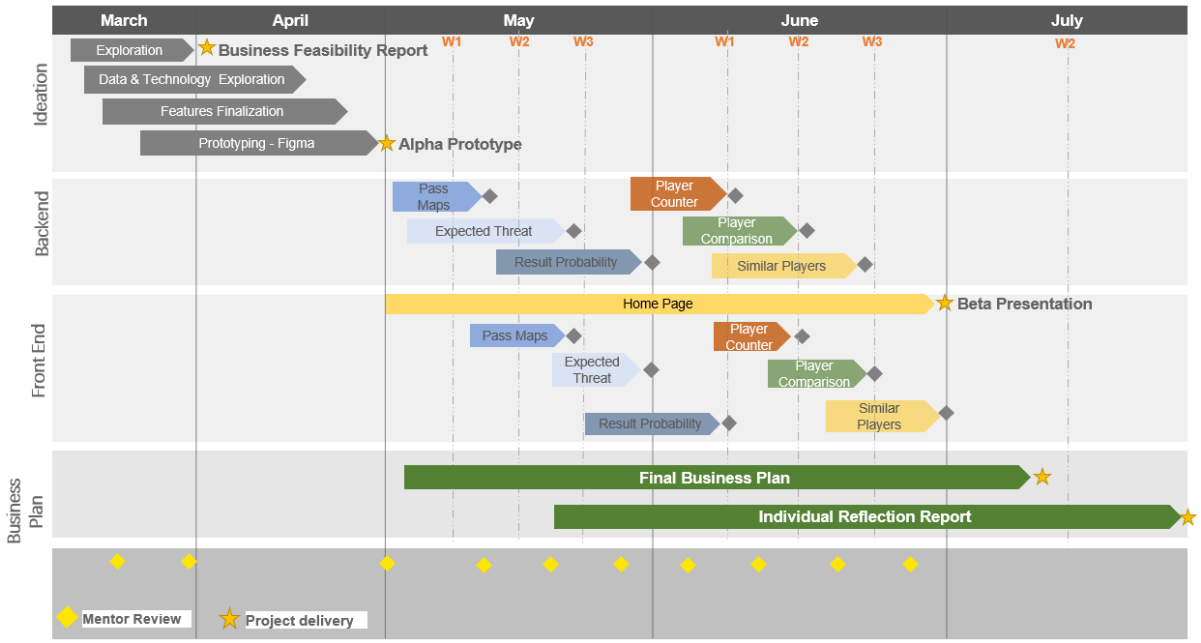
One of the elements of this Product is the use of Expected Threat to measure pitch dominance/coverage by each side, which is one of the challenges encountered. The Competitors were discovered to be utilizing an Expected Goal model during the feasibility study's competitor analysis. Expected Threat is a more advanced model that considers a succession of occurrences rather than just one (*which is the basis of Expected Goals*). The computation of expected threat has been complicated, and the team had to sift through a large number of research articles to figure it out.

Additionally, Data Acquisition for the Team View section has been a challenge because of the data points that need to be generated per match. Therefore, this has been mitigated by using the data made publicly available by STATSBOMB for the English Women's Soccer League for public use. The data will be suited to our need as the clubs in the dataset resemble a great deal of similarity to the Low Tier Clubs who are our lead customers. Shots, passes, dribbles, defensive pressures, duels, saves, clearances, ball recoveries, interceptions, fouls, and other occurrences average over 3,000 each match.

A single event may be described in unprecedented depth, including pitch position, body part, play type (open play, set piece, counter, etc.), duration, and outcome (StatsBomb | FBref.com, 2022). The team will utilize this data for the analysis outlined in the Team View section because the purpose of this project is to focus on data usage rather than data development. Local football teams may also create event-level data, which can subsequently be fed into this solution and utilized for strategizing.

Project Action Plan

Now that the Ideation Phase is almost through, the next 2 months will be focussed on the Execution Phase. The Team will work on the Backend and Front-End development in the next phase in conjunction. Because each feature in the product is a separate deliverable, the team has planned to finish a feature's backend first, then hand it over for front-end development to provide ample time for front-end development. There will be a total of six discrete deliverables for both the Backend and Front End, which put together will be our final product. Below is the Project Roadmap that the Team has agreed upon. The Roadmap involves regular mentor review meetups for the next couple of months.



Design Materials

This section of the report discusses the design thinking tools used to identify the requirements, stakeholders, and possible technology solution for our football analytics strategy builder tool. Design thinking tools are vital tools for any system designers to have in their arsenal to influence the process, production and expression of design ideas. The design and consulting firm IDEO's president highlighted that design thinking should be seen as "a human-centred approach to innovation that draws from the designer's toolkit to integrate the needs of people, the possibilities of technology, and the requirements for business success" (Fugere, 2015).

Requirements Gathering

To identify our requirements, three User Personas and an Empathy Map were used. This allowed us to understand the behaviours and needs of our potential users.

User Personas

User Personas were developed from the perspective of a football manager, trainer and player going through a football season. This helps us understand the type of individuals we were designing for and prevented us as designers, creating self-referential (designers designing for themselves) and elastic user design (stakeholders designing features for their own convenience) for our technology solution (Adobe XD Ideas, 2021).

Persona Template

Name John Caulfield
Age 57
Location Galway
Technical Level Basic
Job Manager of Galway United

Biography

John is the manager of Galway United, a club in the second tier of Irish football. Former manager of UCC and Cork City FC, where he won the FAI Premier Division title along with an FAI cup and Presidents Cup.

Motivations

Johns budget for sports science and analytics is extremely limited and almost non existent due to financial constraints of the second tier of Irish football. John wants to win promotion to the Premier Division this season

Frustrations

John does not have the financial means to create and utilize any form of analytical scouting on his opponents in his bid for promotion.

John is not highly technical and the idea of mathematical probabilities to estimate threats and player fatigue is completely beyond his comprehension.

Their ideal experience

A system where he can simply select an opposition team and get all the information in a meaningful and easy to comprehend way but most importantly in a cost effective way.

Quote

"We don't take outside investment and that's fine. Do I envy other clubs for getting that investment? No. But it does change the goalposts."

John Caulfield speaking about financial constraints when managing Cork City FC

Persona Template

Name Brian McCormick
Age 43
Location Dublin
Technical Level Intermediate
Job Defensive strategy Coach for Shamrock Rovers F.C

Draw your persona

Biography

Provide some details about your persona

Brian is a retired footballer who is now the defensive strategy coach for Shamrock Rovers F.C. Brian has qualifications in various sports coaching courses and has recently enrolled for a part time course in sports analytics.

Motivations

Brian is seeking a better way of understanding opponents play styles and key threats in order to instruct the manager on better defensive strategies. Brian is seeking better methods to traditional information collecting and interpretation of opponent playing styles.

Frustrations

What's is stopping them from using the product or service? What annoys them?

Brian is frustrated with current practices in his football club for identifying opponents weaknesses. Traditional observational methods are used from manual interpretation of match video tapes. This is a long process where it is easy to miss key information.

Their ideal experience

Features and content that will provide the ideal user experience

Ideally Brian would have an easy system that can help communicate player weaknesses to his collogues and players. This would help Brian to communicate key areas of weakness and strengths so Brian could train players in the weeks leading up to a match and inform his manager on the best player options available for that match

Quote

Provide a sample quote

"I have heard of the potential of sports analytics software and am curious to see the possibilities it could offer us for more detailed preparation on the training ground"

Persona Template

Name Cian Murphy
Age 21
Location Cork
Technical Level Basic
Job Soccer Player for Cork City FC

Draw your persona

Biography

Provide some details about your persona

Cian Murphy is a young aspiring striker playing for Cork City football in the SEE Airtricity League Premier Division. Cian has recently broken into the starting team in the past 2 seasons.

Motivations

Cian is constantly seeking new ways to improve his strengths and weaknesses and understand how he can best help his team. Cian is hoping his team perform well so they can compete for trophies at the end of the season.

Frustrations

Cian gets frustrated when he cant identify weaknesses in an opponents setup to put him in a position to score. Cian is annoyed when his team cant gain possession of the ball in a game. Cian also is annoyed when there is no pre game analysis of opponents like large premier league clubs are constantly conducting.

Their ideal experience

Features and content that will provide the ideal user experience

Cian would love a way for himself and his teammates to have a process in place for identifying opponent weaknesses that can help his team overturn the ball so he and his team have a better chance of scoring a goal.

Quote

Provide a sample quote

"It's frustrating not having an easy way of understand other teams weaknesses and strengths in our league. We could work better as a team if such a system was in place"

Left to Right:

Persona 1: Football Manager

Persona 2: Coaching Staff

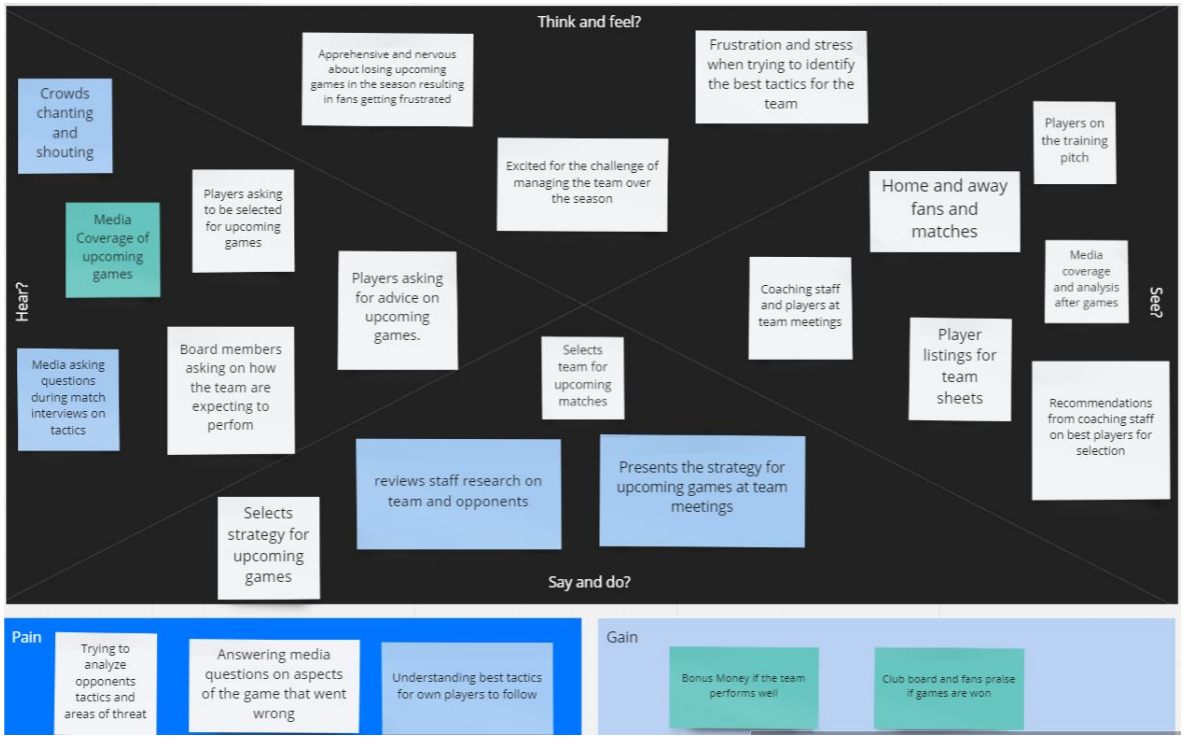
Persona 3: Player

User Personas Findings

- Financial constraints are resulting in football analytics software being overlooked. Current software is considered too expensive.
- The technical nature of sports analytics and statistics is pushing clubs away from analytics software that may be beneficial to teams.
- There is an interest in sports analytics among clubs highlighting that a market for affordable software solutions exists.
- Methods for identifying opponents strengths and weaknesses are traditional and timely.
- Better methods for communicating opponents weaknesses are highly desired.
- An easy and quick way to identify ways a team can identify threats in an oppositions makeup and understand how they can counter act them is highly sought after.

User Empathy Map

The user Empathy Map was developed from the perspective of a football manager, our lead customer. This allowed us to put ourselves in a football mangers' shoes and understand what they think, feel, hear, say and do over the course of a footballing season. It also identified possible pain points in current managerial processes to identify opportunities for innovation. "Empathy Maps can be used to describe personas. The Empathy Maps goal is to create a degree of empathy with the user so the product developing team starts to understand more deeply the users and become more aware of their real needs" - (Ferrira et al, 2015)

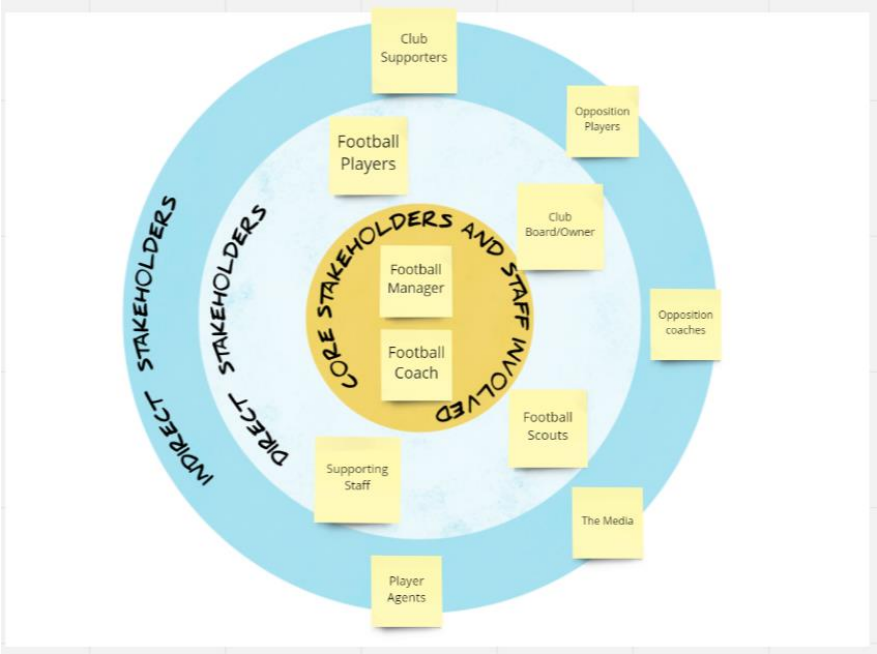


Empathy Map Findings

- A manager needs to be able to make decisions on and answer the difficult questions of what tactics are to be chosen and why. This decision will be judged by fans, the media and by the clubs board all stakeholders in the managers decision. these stakeholders ultimately decide the satisfaction levels with the manager.
- A manager needs to be able to communicate tactics effectively to teams and explain the why on these tactics in a way that everyone understands at team meetings.
- If a manager identifies the right tactics to play his satisfaction levels and performance bonus is more likely to be high, a key motivation for a manager.
- A key frustration for a manager is not understanding opponents tactics and threats.

Understanding Stakeholders

We then identified the key stakeholders in our solution to prioritise the identified requirements identified in the previous section. We used a stakeholder map to identify the key parties that would be affected in some way by our proposed solution.

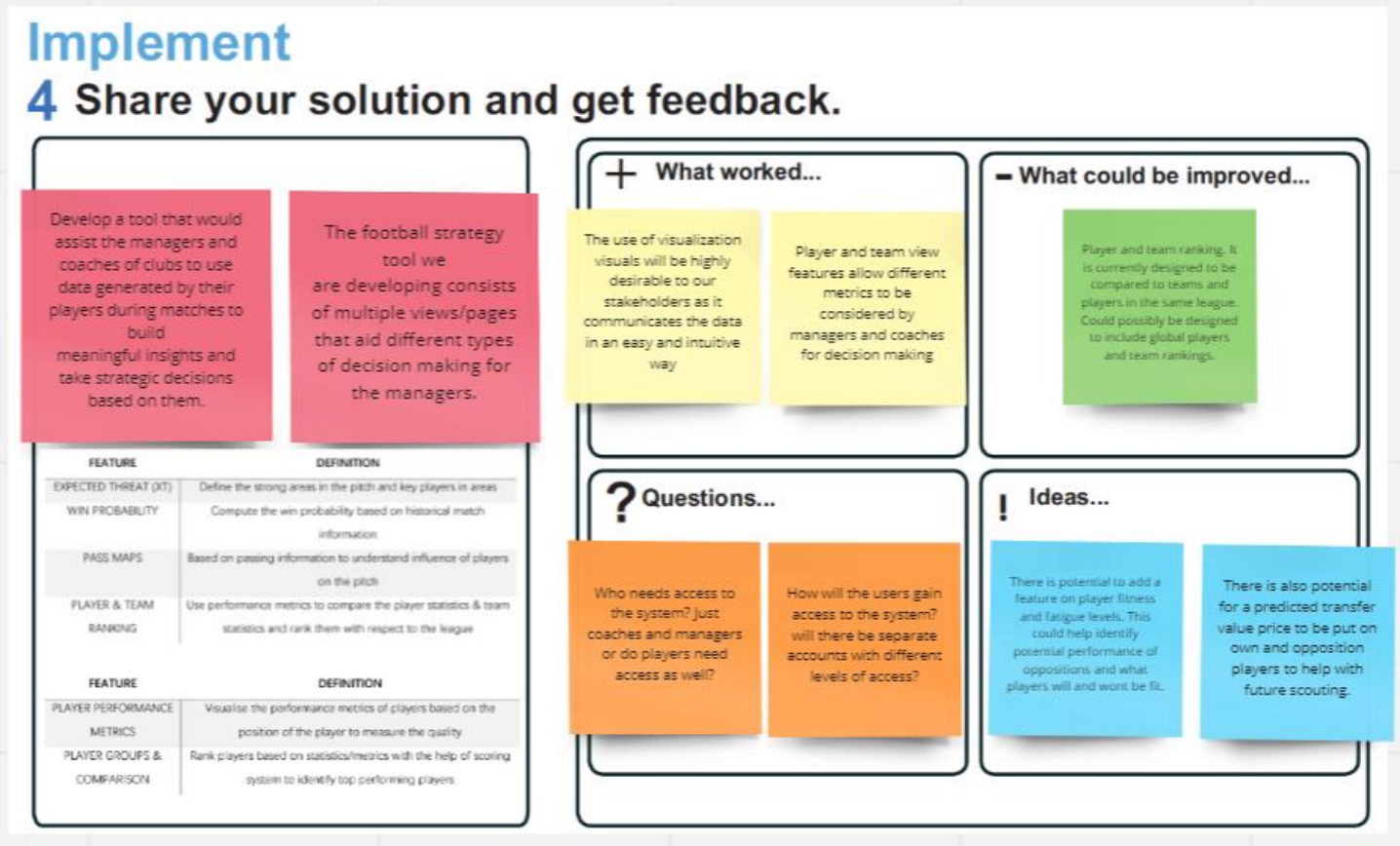


The main purpose of the stakeholder map being used was to get a visual representation of all the people who could influence our project and how they could be connected and interact with it (Savina, 2020).

- Core stakeholders are Football Managers & coaching staff as they require the analytics strategy builder to help conduct their daily tasks of making decisions on team tactics.
- Direct stakeholders have an invested interest in the solution as the decisions managers and coaches make based on systems recommendations will directly affect the way their work.
- Indirect stakeholders will be affected indirectly by the use of the system.

Idea Construction & Critique

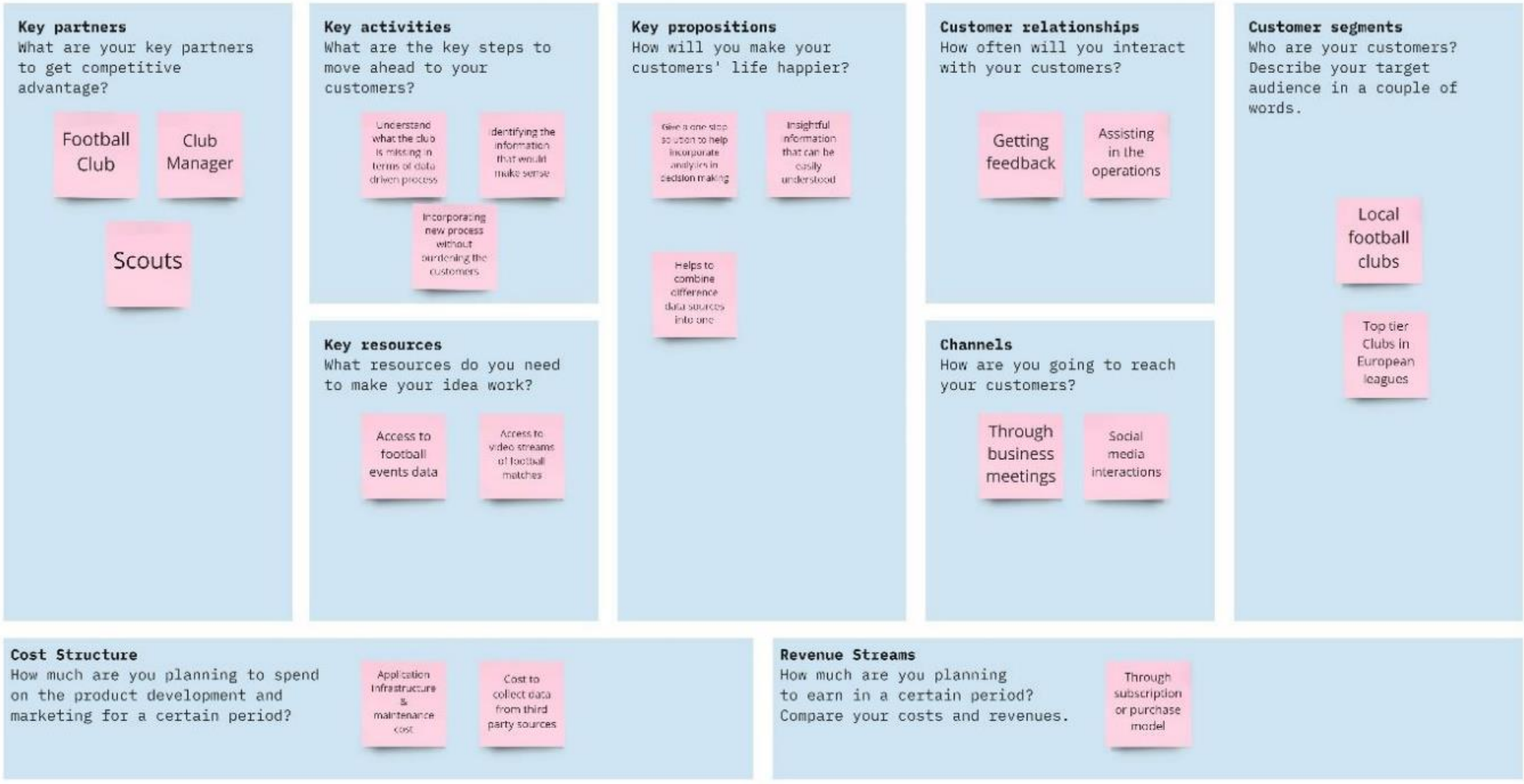
Based on our requirements gathering and stakeholder investigation. We compiled the following analytics solution for our problem and designed decided on the features it should contain in order to meet these requirements. The ideas were discussed among the project team and what worked well was discussed as well as possible areas to consider and to expand into in the future.



Business Model Canvas

In order to condense our idea in a way that can be easily communicated and understood we decided to develop a business model canvas. This is an essential tool that designers should have in their arsenal for defining and communicating a business concept. This one page document will go through the fundamentals of our analytics solution for Football Clubs and lays forth the idea in a logical manner.

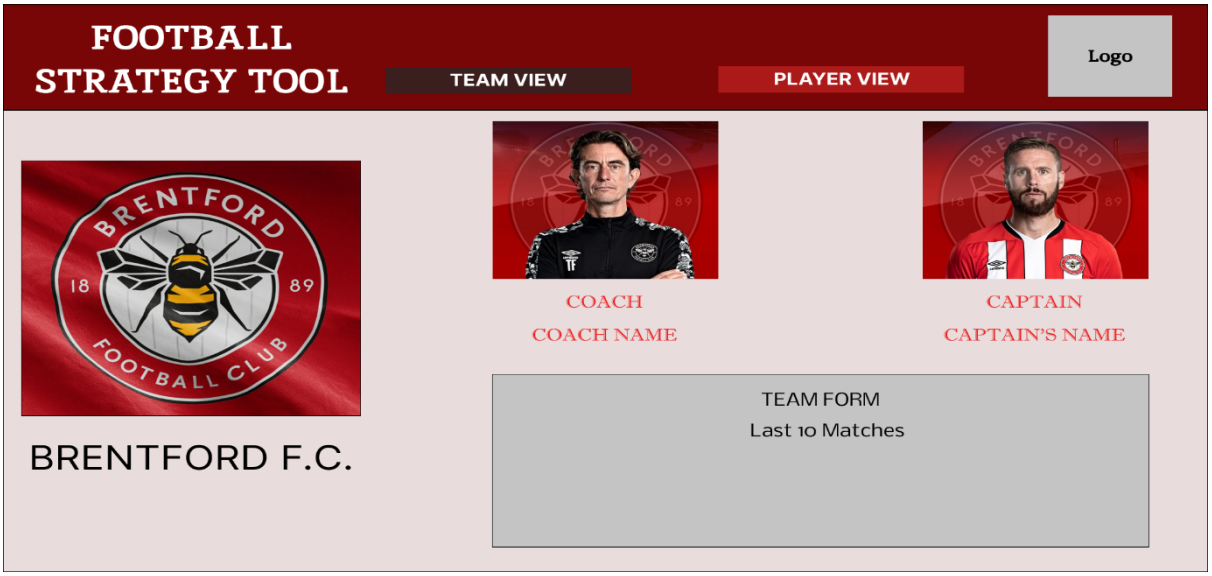
The Business Model Canvas



Prototype

Home Page

The homepage gives the details like the form of the team, coach and captain. From the homepage user can redirect himself to the Team View and Player View sections of the tool.

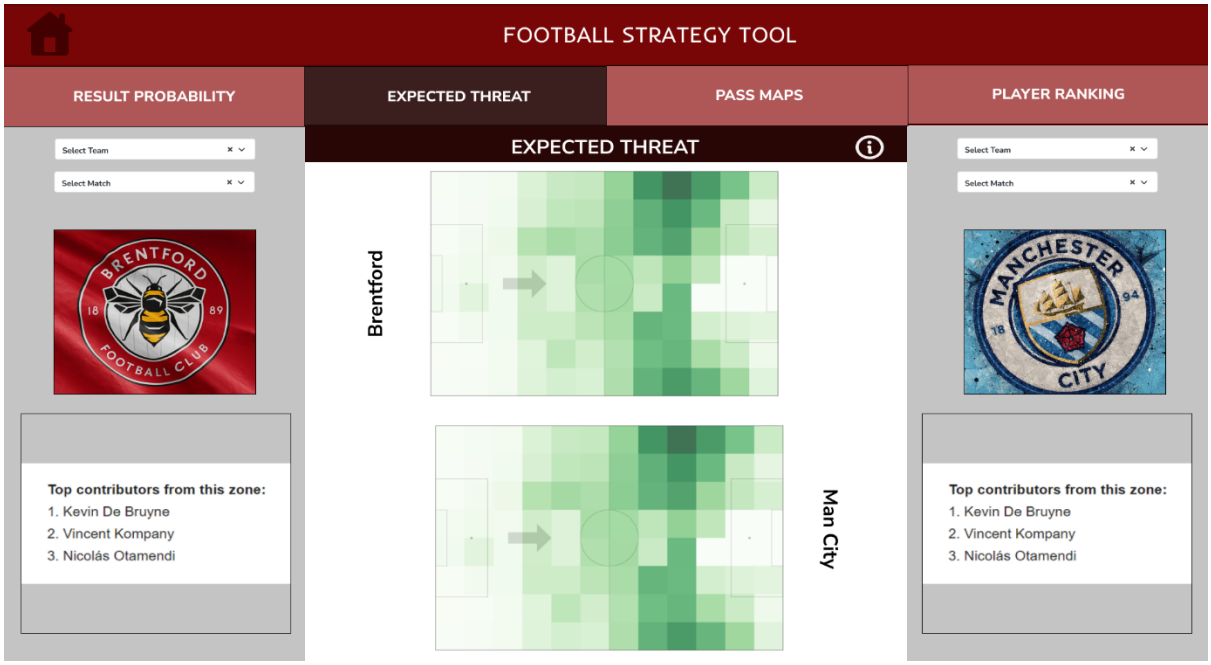


Team View

Results Probability

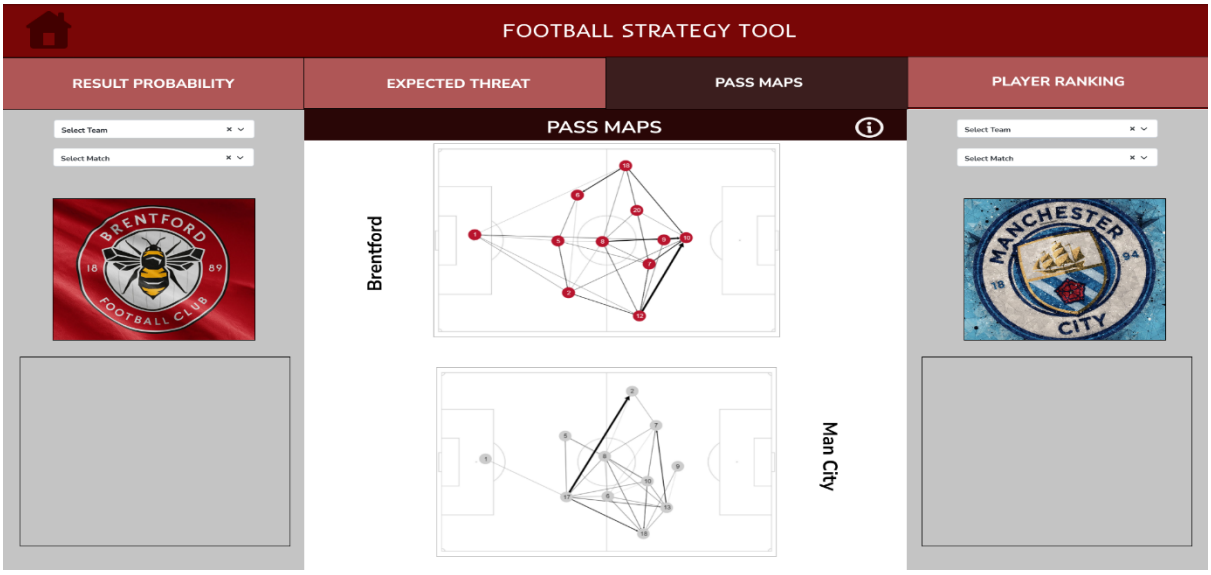
The first section of the Team View gives the result probability of the match between the selected home and away teams. It is computed based on the xG (expected goals) metrics of the historic matches played between the clubs.

Expected Threat



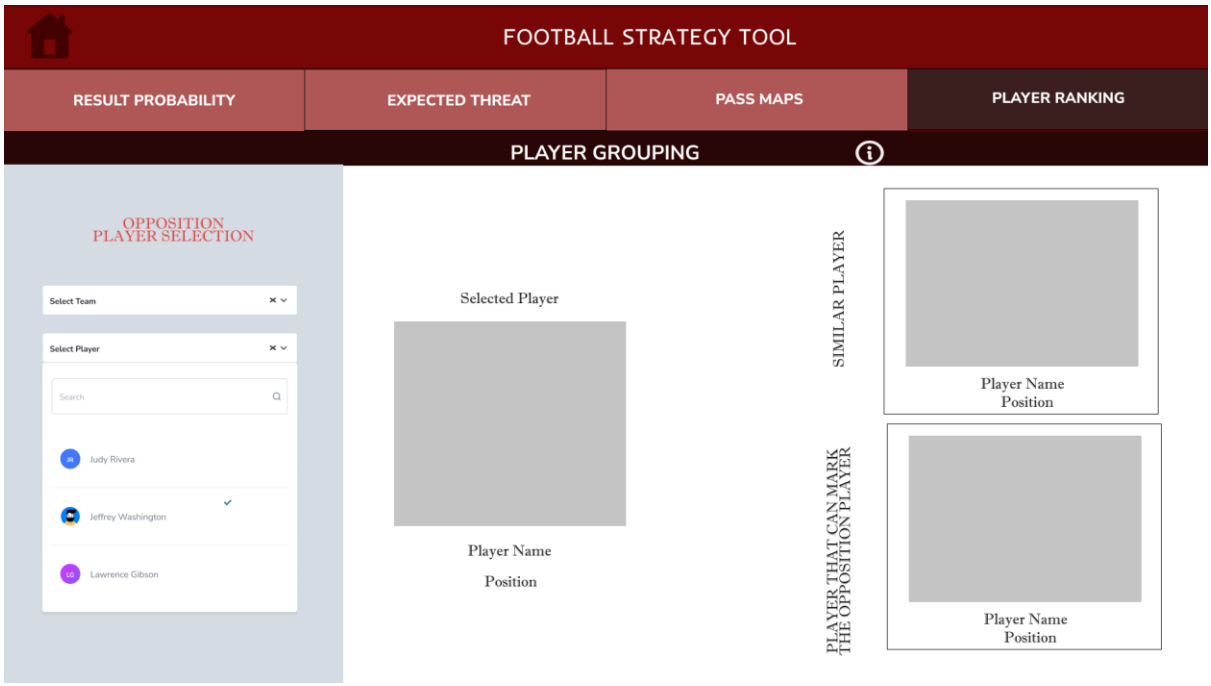
Second part of the Team View shows the expected threat of each team over the different areas of the pitch. It gives the top contributors (player of the team) to the threat in the particular zone selected.

Pass Map



Pass Maps are shown in the next section which are diagrams build based on the player pass information to teammates. Similar types of passes are clustered together to build a network graph that connects all the players of the team. It helps to understand the most influential players in the team for different possessional set ups.

Player Grouping



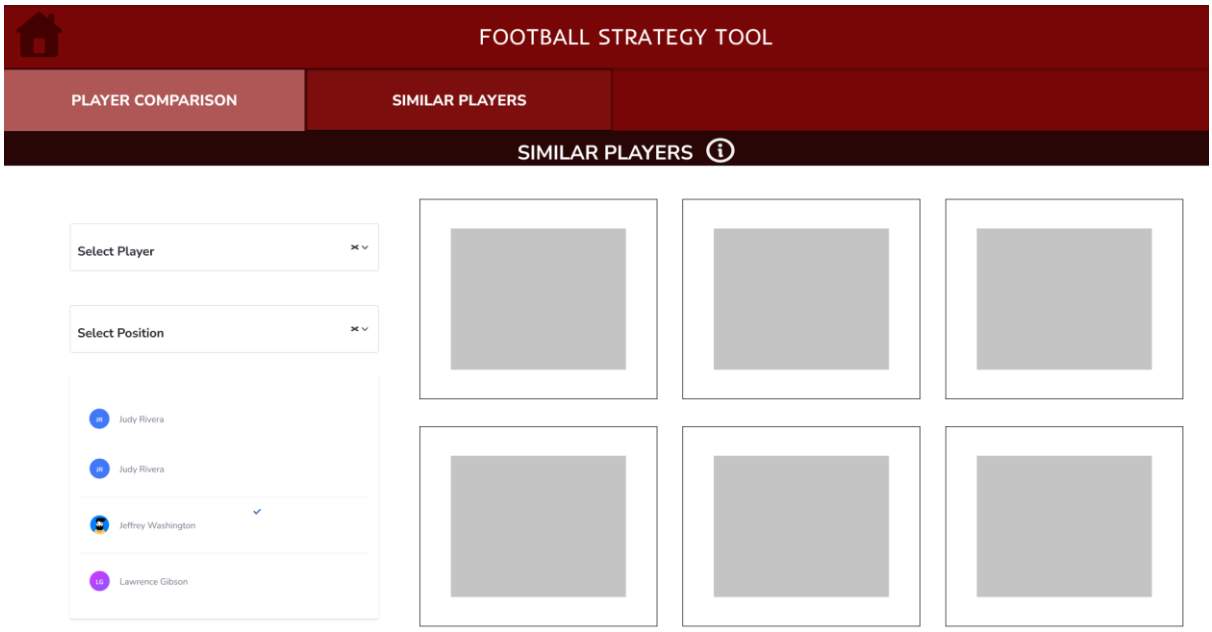
In the next part, the managers will be able to see the players that are similar in terms of playing style of the selected opponent player. It also gives the best player in our squad who can be used to counter or mark the opponent to nullify the impact.

Player View

Player Comparison

Performance of two players is measured and compared with help of various metrics that is defined based on the positions the players play. These metrics are represented with the help of charts and diagrams for easier understanding. Managers can use this information to identify the key attributes of their star players which will help them set up the team to maximise their potential as well as understand the strengths and weaknesses of the opponent players.

Player Similarity – For Scouting



Ranking the players based on their in-game statistics and clustering them into similar groups to identify the players who are similar in terms of performance. This information could be helpful for managers in identifying the opponent players that resemble the playing style of their players that help them to prepare well when facing them. The ranking information can be used as means of scouting the players who can play in the team.

Technology Stack

In its essence our product is data-centric, as well as visually appealing. Our technology stack is categorized as follows:

Backend: Python for constructing our backend logic and controlling the workflow of our web application

Web Frontend: React JS and Node JS for the frontend architecture along with suitable graph libraries such as Chart.js, d3-statsbombevents, etc.

Database: Amazon S3 for object storage service

DevOps: AWS CodeCommit for version controlling and AWS Cloud for deploying and hosting

As our product would be a web-based application, React JS would allow a faster prototyping functionality with cleaner abstraction, well-defined lifecycle, and a clear component-based codebase structuring. Furthermore, React is agnostic of what goes on at the server-level since interacts with HTTP APIs regardless of the language used for creating the backend; which in our case is Python. The python-based APIs would use standardization such as REST which is enough for establishing compatibility with React. For data storage and pipeline, we intend to use Amazon S3 or database to store the data we acquire. Any other computational (IDE for Python) requirements can be sufficed with the help of AWS Cloud. For the backend, following are the Python libraries that will be put into use.

Web Scraping to pull the data from websites: BeautifulSoup4, Selenium

Exploratory Data Analysis - Matplotlib, seaborn

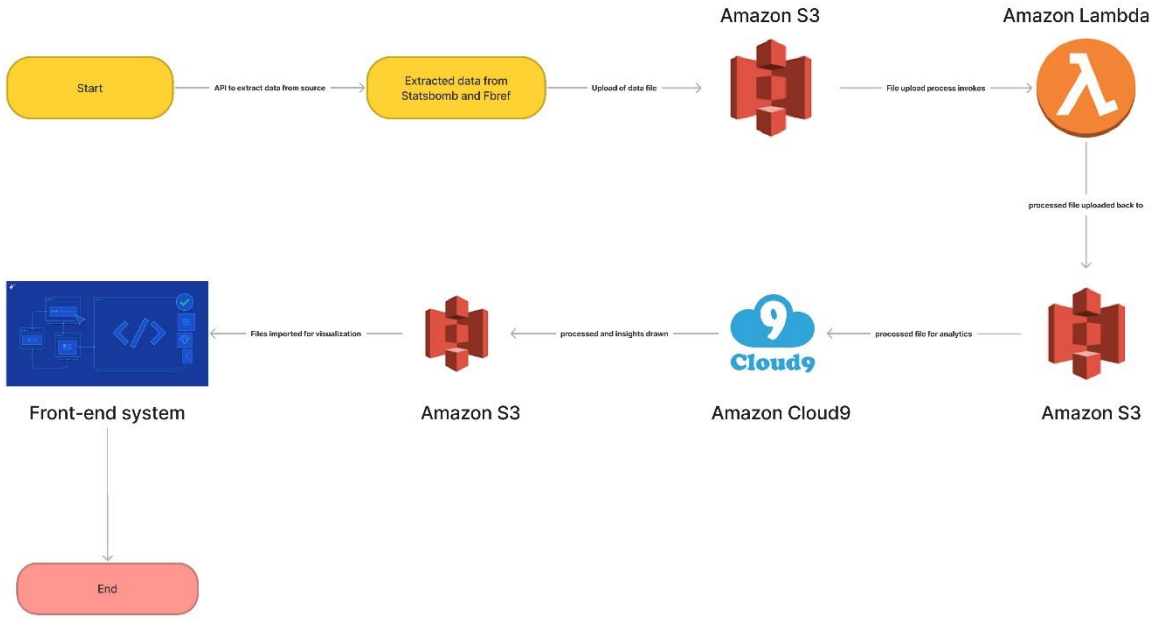
For Clustering and Classification: Scikit – Learn

Deep Learning algorithm(optional) – TensorFlow

Platform – AWS Cloud (S3, RDS, Cloud9, Lambda)

Front End – React JS, Material UI, Redux for state management, Chart Js, and d3-statsmbombevents

Data Pipeline Architecture



Extracted Data -> Amazon S3

Statsbomb has provided some great open data which can be extracted using the API provided as it is the simpler, quicker and easier way to access the open data. Statsbomb has provided an API to extract the open data with the use of library statsbombpy. Using this library, we can extract data from Statsbomb of events in a specific match from specific season of a specific competition. Fbref provides open-source data set of players report, we can just import the information of a specific player from their website into a csv file. We can then store these extracted data into Amazon S3 environment. You can simply exchange data stored in Amazon Simple Storage Service (Amazon S3) with numerous apps. Each application, however, has its own set of needs and may demand a different representation of the data.

Amazon S3 -> Lambda function -> Amazon S3

AWS Lambda is an Amazon Web Services serverless computing service (AWS). AWS Lambda users write functions, which are self-contained programs written in one of the supported languages and runtimes, and upload them to AWS Lambda, which then executes them in a fast and flexible way. Lambda functions may be used to do everything from providing web pages to processing data streams to using APIs and connecting with other AWS services. S3 Object Lambda is a feature that lets you write your own code to treat data from S3 before sending it to an application. S3 Object Lambda integrates with current apps and utilizes AWS Lambda functions to handle and alter data as it is pulled from S3. Once a file is uploaded to the S3 bucket, the lambda function will be synchronously run in line with a typical S3 GET request and begin processing the file's content; once done, the lambda function will upload the processed file back to the S3 bucket.

Amazon S3 -> AWS Cloud9 environment -> Amazon S3

AWS Cloud9 is a browser-based integrated development environment (IDE) that allows you to write, execute, and debug code. A code editor, debugger, and terminal are all included. Cloud9 is pre-installed with key tools for major programming languages such as JavaScript, Python, PHP, and more, so you don't have to install files or setup your development workstation to get started.

Cloud9 also offers a smooth development experience for serverless apps, allowing you to simply specify resources, debug, and switch between local and remote execution. You can share your development environment with your team fast using Cloud9, allowing you to pair program and track each other's inputs in real time.

Objects in an Amazon S3 bucket will be downloaded to a folder in your AWS Cloud9 environment from the AWS Cloud. After that, you may use analytics in Python to program and draw conclusions from this data file. Once this file has been processed, it may be returned to the Amazon S3 bucket. The upload to Amazon S3 from AWS Cloud9 will be done using the Toolkit interface or a command to upload a file to a bucket. Both methods will upload a file from the AWS Cloud9 environment to the AWS Cloud and save it as an S3 object. A file can be uploaded to a bucket or a folder that organizes the contents of that bucket.

Amazon S3 -> Front-end system

This pre-processed and analysed final data set file will be picked up and utilized in the integration to the front-end system for visualization purposes.

References

- Fugere, K., 2015. The Importance Of Design Thinking For Big Data Startups. [Blog] *TechCrunch*, Available at: <https://techcrunch.com/2015/07/21/the-importance-of-design-thinking-for-big-data-startups/?guccounter=1> [Accessed 20 April 2022].
- (Adobe XD Ideas, 2021) Ideas. 2021. What Are User Personas and Why Are They Important? | Adobe XD Ideas. [online] Available at: [Accessed 20 April, 2022]
- (Ferrira et al, 2015) Ferreira, B., Silva, W., Oliveria E., and Conte, T. 2015. Designing Personas with Empathy Map. USES Research Group, Instituto de Computação - IComp Universidade Federal do Amazonas (UFAM) 2015
- (Savina, 2020) Savina, A., 2020. Complete Stakeholder Mapping Guide. [Blog] *Miro Blog*, Available at: <https://miro.com/blog/stakeholder-mapping/#:~:text=Stakeholder%20mapping%20is%20the%20visual,and%20how%20they%20are%20connected.> [Accessed 20 April 2022].