Football Analytics: Strategy Builder

Feasibility Report

Aadith Krishna - 121106867

Conor O Donoghue - 121109516

Nikhil Mohan Joshi – 121104673

Philson Manarcad - 121106376

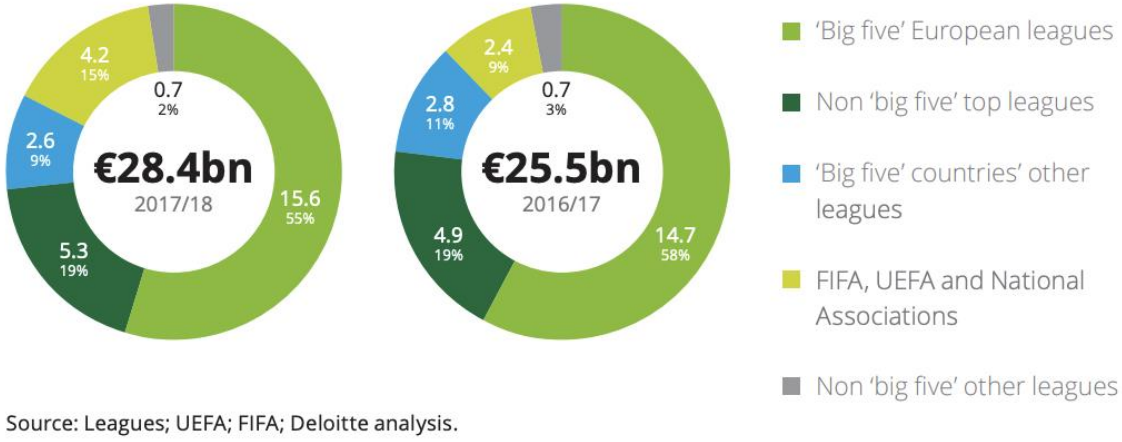
Thomas Donal Murphy - 116337076

Yash Joshi - 121106471

Executive Summary

Football is associated with passion and intuition. Unanimously, it is one of the most viewed and profitable sports across the globe; case and point UEFA Euro 2020. The cumulative global audience for the live match event crossed 5.23 billion in 2020 making UEFA Euro 2020 one of the most viewed moments of all time (UEFA EURO 2020, 2021). Deloitte's Annual Review of Football Finance 2019 stated the overall size of the European football market, in terms of revenue had reached €28.4 billion in 2017/18 (Annual Review of Football Finance, 2019).

Chart 1: European football market size – 2016/17 and 2017/18 (€ billion)



Source: Leagues; UEFA; FIFA; Deloitte analysis.

Technology has enhanced sports consumption, and has revolutionized performance management among football clubs. However, the rise of technology has also increased the ever-growing competition among those clubs. Not only do they need to be on top of their game but also fulfil the expectations of their many stakeholders which include fans, players, shareholders, and sports media. To achieve the same, football clubs are tapping into the essence of data analytics to harness actionable insights for improving their overall performance metrices. Although a question arises – How do they tap into this essence and go towards profitability and simultaneously improve their sporting performance? Enter, football strategy tools.

Strategy tools leverage advanced statistics and infuse huge data gathered from various sources such as IoT sensors and GPS devices which track player and ball movements in real time through optical tracking. This data is further processed and delivered as insights through intuitive visualization. For example, a heat map showing the intensity of ball possession or analysis of team data as a network in which different nodes represent players and the lines between these nodes represent their interaction. Visualization of such data allows a coach to identify and test the passages and strategy of play which further provides deeper insights useful for creating effective strategies.

We, as a group, would like to propose our strategy tool aimed towards lower tier clubs to deliver a one stop solution to analyse their teams and opponents through a data-driven approach and further assist their strategy decisions. We have created this feasibility report to better illustrate our problem statement and why existing solutions have not been successful. Talking about the products and services we aim to offer, the report also showcases a market analysis performed by us for identifying market trends, projected market growth, and market evolution. Furthermore, a brief introduction to our technical solution proposed is elucidated highlighting how we acquire the data essential for analysis and finally, what technologies and frameworks we would use to construct the user interface for the proposed strategy tool.

The Problem

Over the previous two decades, data analytics has become more important in many aspects of our life, including business, healthcare, media, and sports. Football was supposed to be resistant to this tendency until a few years ago. Early adopters in the main football leagues are already flourishing as a result of the competitive edge that data analytics investments are beginning to provide: Liverpool, Brentford and AZ Alkmaar are just a few of several successful case studies. Clubs who do not aim to hop on the analytics bandwagon, in our opinion, risk being left behind. We still have the difficulty of being able to analyse the data. If football teams, for example, are inundated with massive amounts of data but lack the technical expertise to assess and extract useful information, data becomes essentially useless. It's like presenting a complete set of stock, currency, and commodity prices, ratios, and indications to someone who is unfamiliar with financial markets: the data alone does not make that person an infallible trader. Football clubs need data to make informed judgments, but they also require analytics to make sense of it. (Soccerment, 2020)

In order for all of the clubs to have a strategic advantage at this time, they must examine the facts on how they play against their opponents. Teams generally, analyse their self-data and try to find a solution to optimize their gameplay and their own style of playing. This can also improve the overall performance of the team, but for having competitive advantage over opponents, club also has to analyse and study the opponent team's data and then strategize the gameplay for the particular match to be played.

Currently, the data analytics solutions given to the clubs focusses on developing and nurturing their own players and teams, so as to improve their performance. Some solutions concentrate on scouting and transfers, while some talk about analysis on injury prevention and post-match analysis, but the need is of having a consolidated solution which analyse the opponent teams' variety of data and compares with our club teams' data.

So, what we are proposing here is a comprehensive solution that contains all the above -mentioned features, for the football teams to take data driven decisions based on information

gathered from the matches they played. With the tool, the teams make amendments to their tactics to improve the gameplay that will help in achieving competitive advantage over the opponents. The insights generated from the tool with the help of match & player level data can be used by managers and support staff to tweak the strategy that suits their squad or recruit new players that will improve the team.

Product/Service

In the current scenario, all football clubs are heavily investing on data analytics to help them get competitive advantage over their opponents. For clubs that are in lower tiers, their operating cost is low to invest on such expensive solutions and personals. But it is important for them to have access to a solution that help them in using data to take decisions to improve their team.

In this project, the team is focused on developing a tool that would assist the managers and coaches of the football clubs in using the data generated by their players during matches to build meaningful insights and take strategic decisions based on them. The football strategy tool we are developing consists of multiple views/pages that aid different types of decision making for the managers.

Data

There are 2 main data sources required for this Project. First one being **Event level data** which has been made available by **Statsbomb for the English Women's Soccer League** for public use. Each match includes, on average, over 3,000 events including shots, passes, dribbles, defensive pressures, duels, saves, clearances, ball recoveries, interceptions, fouls, etc. A single event contains an unmatched level of detail to describe its unique characteristics, such as pitch location, body part, pattern of play (open play, set piece, counter, etc), duration, and outcome (StatsBomb | FBref.com, 2022). Since the scope of this project is to focus on the use of data rather than data generation, the group will be using this data for the analysis discussed in the **Team View** section. However, this event level data can be generated by local football clubs as well which can be then plugged into this solution and used for strategizing.

Second one is the **Player level** information which can be scraped from **fbref.com**. This data will be used for evaluating player performance and comparison discussed in **Player View** section.

High Level Product Details

The proposed solution intends to assists the football teams in strategizing for their upcoming matches against the opponents with the help of the following features:

Team View

In this part of the solution, the manager of the team will be able to search for any team that is playing in their league and compare with their own team. It enables them to understand the performance metrics of the clubs, key players, similar players in both the teams, the strategy, gameplay metrics etc. that will help them to learn about the opponents to setup the team against them and get a strategical upper hand. The table below shows the different featured incorporated into the team view of the tool.

FEATURE	DEFINITION
EXPECTED THREAT (XT)	Define the strong areas in the pitch and key players in areas
WIN PROBABILITY	Compute the win probability based on historical match information
PASS MAPS	Based on passing information to understand influence of players on the pitch
PLAYER & TEAM RANKING	Use performance metrics to compare the player statistics & team statistics and rank them with respect to the league

Expected Threat

Expected Threat (xT) is a derived metric computed with the help of match level information of team to understand the threat of teams in different areas of the pitch. It helps to define the strong areas of a team along with the key players in those areas. With the help of an expected threat diagram, the managers will be able to setup his playing eleven based on the strengths and weaknesses of the opponents (Expected threat -socceraction 1.2.1 documentation, 2022)

Win Probability

When the manager is comparing his team with the opponent team, with the help of historical match information and other match related metrics, the win probability of his team is predicted. It helps the manager to understand how strong the opposition is based on the information about the team in the past. We use a concept called expected Goal (xG) which is computed based on probability of scoring a goal from different areas in the pitch. The expected goals model

combined with other match related metrics are used for this prediction of match outcome of future matches. (Herbinet, 2018)

Pass Maps

Pass Maps 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

Ranking Players & Teams

While comparing the teams, managers will be able to see the group of players that are similar in terms of playing style or stats from both teams. Along with that the ranking of either teams based on various match performance metrics are shown to understand how the teams are positioned in the league.

Player View

This part of the solution helps the managers to understand in detail about the performance of the players in the league based on metrics depending upon the position they play in. This information helps the managers to understand how good a player is compared to other players in the league. The table below shows the different featured incorporated into the player view of the tool.

FEATURE	DEFINITION
PLAYER PERFORMANCE METRICS	Visualise the performance metrics of players based on the position of the player to measure the quality
PLAYER GROUPS & COMPARISON	Rank players based on statistics/metrics with the help of scoring system to identify top performing players

Player Performance Metrics

Player performance is measured with then 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.

Player Groups

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

Primarily, **Python** is the main programming language which would suffice most of the needs for the solution, especially the backend. As for the front end, we intend to use a JavaScript powered library called **React JS** for constructing our frontend and user interface touch point. Our platform would be web-based application and React allows 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.

In case of data storage, we may leverage S3 or database to store the data depending on the data size 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**

Market Analysis

Market Trends

Trends or key issues affecting the market

Recent technological advances in the areas of IoT and wearable devices, combined with an increase in revenue in sports has meant that areas of sport science have become somewhat of a priority for sports teams in most disciplines. It is estimated that the sports technology market is valued at over $40 billion. (Comer, 2022) This has brought huge interest from tech firms to race to bring in the next product that will take the game to the next level. We feel that our product has the potential to be an essential for every team, not only for the top-level teams such as Barcelona's, Bayern Munich's, and Liverpool's of football, but the lower-level teams such as Cork City and Cobh Ramblers where financial margins are so tight the difference between finishing the season 1 point higher or lower will have a great financial impact into the viability of the club. Sports science has developed at a rapid pace in the last 10 years from basic video analysis to wearable heart rate and GPS monitors and now microchipped smart sliotars. Our product will have huge potential for growth for sports across disciplines beyond football.

Projected market growth

To best examine our projected market growth, it is best to look at the financial projections for the first year of operating at full capacity. Below are our estimates on our quarterly performance.

Selling & Buying Goods	Q1	Q2	Q3	Q4	Total
Selling price (€)	1000	1000	1000	1000	
Attrition Rate (%)	25	25	50	35	
New Suscribers (Sell units)	10	50	25	100	185
Overhead Expenses					
Fixed overheads (€)	1000	1000	2000	1000	5000
Variable overheads (%)	10	20	5	10	
Overheads paid (%)	110	80	100	60	
Funding					
Opening Share Capital €	300	600	900	1200	1500
Opening Reserves €	6700	14073	58898	123800	184366
New Share Capital €	300	300	300	300	1200
Opening Loans O/S €	0	750	500	250	0
New Loans in €	1000	0	0	0	1000
Loan Principal Repaid €	250	250	250	250	1000
Loan interest Paid €	7	7	7	7	28
Opening Bank Balance €	1000	9643	57706	120379	184282
Fixed Assets					
Opening Net Book Value €	6000	5580	5112	4871	4744
Acquire new assets €	200	100	300	400	1000
Sell assets €	0	0	0	0	0
P/L on Disposals €	0	0	0	0	
Depreciation Rate %	10	10	10	10	

Projected Profit/(Loss)					
Account	Q1	Q2	Q3	Q4	Total
Sales	10000	58000	71000	159000	298000
Overheads	2000	12600	5550	16900	37050
Trading Profit/(Loss)	8000	45400	65450	142100	260950
Depreciation	620	568	541	527	2256
Interest	7	7	7	7	28
Profit(Loss) on sale of asset	0	0	0	0	0
Net Profit/(Loss)	7373	44825	64902	141566	258666
Cash Flow					
Opening Balance	1000	9643	57706	120379	184282
Cash-in	11300	58300	71300	159300	300200
Cash-out	2657	10237	8627	14397	36758
Closing balance	9643	57706	120379	265282	447724
Effective Sales Calculation					
Quarter 1 Sign-ups	10	8	8	8	34
Quarter 2 Sign-ups		50	38	38	126
Quarter 3 Sign-ups			25	13	38
Quarter 4 Sign-ups				100	100
Total Sales Units	10	58	71	159	298
Total Sales Value	10000	58000	71000	159000	298000

In order to promote product adoption, initial period will involve providing free trials to the clubs for garnering visibility and feedback collection which will drive future enhancements. As clubs realise the benefits of the product, the competitive disadvantage of not having a strategy tool will be then the driving force for other clubs to adopt the product.

Competitor Analysis

Over the last 10 to 15 years football analytics has grown massively. Third party vendors who originally collected data for fans and media are now selling data to football clubs (The Numbers Game, 2018). Through conducting competitor analysis multiple potential competitors were found in the football analytics sector.

Video Analytics Software

Majority of football analytics software today revolves around video analysis with companies such as Coach Paint – Tracab, Play – Metrica Sport, Spiideo and Isports Analysis. These applications allow for detailed video analysis with the ability to add "telestration" (arrow tags etc) icons to a video and break a match down into video clips easily for presenting to players and coaches. Much of the software is backed by AI to allow for data tracking of individual players with metrics such as distance travelled, and top speeds being tracked. Please see the table below which highlights the strengths and weaknesses of these potential competitors.

Wearable Technology Vendors

Wearable technology has become widely used in all sports with small GPS devices being attached to players clothes while training and playing. These devices very accurately track player movement metrics such as distances travelled, and top speed reached as well as heart rate and body temperature. Some vendors such as Catapult provide software where the data collected can be stored and extracted. These devices are a primary source of data Collection.

Statsbomb

Statsbomb are an established data collection vendor supplying sports teams in American football and soccer with match level data for its customers. Statsbomb provide a wide array of flexible and customisable suite of data visualisation platforms which helps teams extract insights quickly. This helps teams with player recruitment, performance analysis and opposition scouting (Statsbomb, 2022). Please see the table below which highlights the strengths and weaknesses of Statsbomb.

Please see the table below which highlights the strengths and weaknesses of these competitors.

Competitor Name	Competitor Strengths	Competitor Weaknesses	Competitive Advantage of Our Idea
Play - Metrica Sport (Video and Data Analysis Tool)	• Breaks down match videos into clips. • Easy to use telestration tools for visualisation, backed by AI. • Allows for easy presentation for team analysis. • Automated data tracking compatible with multiple video medias	• Requires manual human intervention for data analysis and decision making. • Looks at passed events and does not predict for the future.	• Predictive analytics techniques for decision making. E.g., Expected threat calculation • No further manual data analysis technique required once application features are run. • Provides metrics for comparing players and teams instead of non-quantifiable interpretations of video. • Automatically extracts value from collected data for analytics teams.
Coach Paint (Illustration and Tagging Tool)	• Easy to use telestration tools for visualisation, backed by AI. • Automated player and data tracking. • Customizable GUI for presentation.	• Requires manual human intervention for visualisation tagging. • Looks at passed events and does not predict for the future.	
Catapult - GPS Tracker (wearable Technology)	• Data collection device for player movement tracking. • Extremely reliable and accurate	• Requires manual human intervention for data analysis and decision making. • Looks at passed events and does not predict for the future. • Requires purchase of hardware for each player. • No access to opposition data.	
Statsbomb	• Flexible and customisable suite of data visualisation platforms • Predictive analytics functionality (expected goals).	• Expensive software that only top-level clubs can afford.	• Expected threat which takes into account a series of events leading up to a goal and not just the position the goal was scored from (Expected Goal).

Barrier to Entry

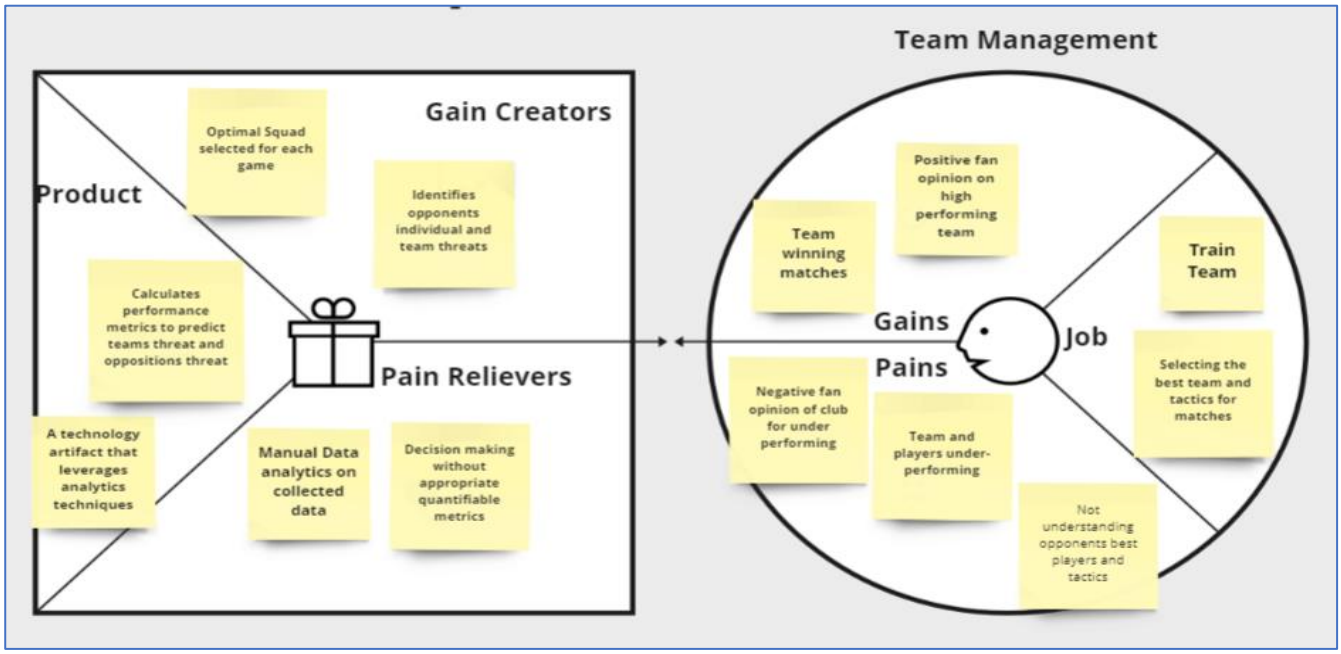
When it comes to software start-ups there is a high risk of failure with 19% failing because of competitors outperforming them (Griffith, 2014). Therefore, there is a risk of competitors stealing market share when a new solution is put into the market. However, according to Andrew Rapacke a registered patent attorney and managing partner of Rapacke Law Group (Rapacke, 2019) the only full proof way of protecting a software solution is by taking a defensive or offensive patent approach, provisional patent approach, trade secrets approach or trademark approach. This is a

legal process and can be a very long process with high costs. Therefore, this group feels that the best method for ensuring competitiveness in the market is by creating a reliable, and innovative solution to performance analysis that is not currently offered in the market. Introducing this first and continually evaluating its performance to make better solutions, would ensure competitive advantage in the long term.

From our competitor analysis it was clear that current most software solutions (video analysis and wearable technology) require manual intervention by humans in order to analyse data and evaluate player and competitor performance. This results in software solutions that are labour intensive and limited by a club's technical ability. Our solution provides these clubs with predictive analytics functionality without the need for sports clubs to have on-site expertise in the area. This will quickly enable teams to make decisions based on the numbers. Furthermore, manual data analytics of collected data will not be required, with quantifiable metrics being automatically generated for teams. This is a huge advantage for teams seeking a new analytics solution to compete with opposition teams in their leagues.

The key competitive advantage we have over Statsbomb is our expected threat functionality that has been discussed in the product section of this report. Furthermore, our product is aimed at smaller clubs to get quick analytic insights on match events and will be affordable to small clubs with small budgets. The expected threat will allow clubs to focus on the events leading up to a goal in order to determine where your own and competitor goals are coming from. This allows clubs to understand the events leading up to a goal instead of what position of the pitch a goal is most likely to occur.

Customer Value Proposition



References

UEFA.com. 2021. UEFA EURO 2020. [online] Available at: <https://www.uefa.com/insideuefa/about-uefa/news/026d-132519672495-56a014558e80-1000--uefa-euro-2020-impresses-with-5-2-billion-cumulative-global-liv/> [Accessed 31 March 2022].

Www2.deloitte.com. 2019. Annual Review of Football Finance. [online] Available at: <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/sports-business-group/deloitte-uk-annual-review-of-football-finance-2019.pdf> [Accessed 31 March 2022].

Soccerment, 2020. THE GROWING IMPORTANCE OF FOOTBALL ANALYTICS. [Online] Available at: https://soccerment.com/the-importance-of-football-analytics/

Socceraction.readthedocs.io. 2022. Expected threat -socceraction 1.2.1 documentation. [online] Available at: <https://socceraction.readthedocs.io/en/latest/documentation/xT.html>

Herbinet, C., 2018.Predicting Football Results Using Machine Learning Techniques. Post Graduate. Imperial College London.

FBref.com. 2022. StatsBomb | FBref.com. [online] Available at: <https://fbref.com/en/statsbomb/> [Accessed 12 January 2022].

Comer, Y., 2022. The Competitive Edge - How Analytics are Driving Success in Sport - Cois Coiribe. [online] Cois Coiribe. Available at: <https://impact.nuigalway.ie/ai-and-creativity/the-competitive-edge-how-analytics-are-driving-success-in-sport/> [Accessed 29 March 2022].

(The Numbers Game, 2018) Youtube.com. 2018. *The Numbers Game | How Data Is Changing Football | Documentary*. [online] Available at: <https://www.youtube.com/watch?v=ILcXH_4rwr4> [Accessed 26 March 2022].

(Statsbomb, 2022) We are Data Champions | StatsBomb. 2022. *Sports teams | Sports data solutions for analysts | StatsBomb*. [online] Available at: <https://statsbomb.com/who-we-help/sports-teams/> [Accessed 26 March 2022].

(Griffith, 2014) Griffith, E., 2014. *Startups are failing because they make products no one wants*. [online] Fortune. Available at: <https://fortune.com/2014/09/25/why-startups-fail-according-to-their-founders/> [Accessed 26 March 2022].

Rapacke, A., 2019. *Startups: How to Stop Bigger Competitors from Stealing Your Ideas - The Rapacke Law Group*. [online] The Rapacke Law Group. Available at: <https://arapackelaw.com/startups/startups-how-to-stop-bigger-competitors-from-stealing-your-ideas/> [Accessed 26 March 2022].