

PSV Eindhoven





RESEARCH REPORT PSV

Company	PSV
Company supervisor	
School	Fontys ICT Hogeschool Minor Data Driven Business Lab
Content coach	
Process coach	Mr. Rob Verhoeven
Students	Gergana Agorasteva Remco Bisschops Femke Boogerd Nikita Gavrilov Yaniek Martens Lieke Nijs Victor Plesciuc Ilya Tsakunov Aleksandr Vereshchagin



DOCUMENT HISTORY

REVISIONS

Version	Status	Date	Changes
0.1	Draft	02-11-2021	Subqestion 1
0.2	Draft	01-12-2021	Subquestion 2 + introduction
0.3	Draft	23-12-2021	Subquestion 3 + 4
0.4	Draft	12-02-2021	Conclusion + advice and recommendations

APPROVAL

This document has the following approvals:

Version	Date Approval	Name	Function	Paraaf

DISTRIBUTIE

This document is sent to:

Version	Date approval	Name	Function
0.1		R. Verhoeven	Mentor
0.2		J. Verbossen	Client
0.4		J. van der Straten	Content coach



TABEL OF CONTENTS

Document history	2
1. Introduction	5
2. Data sources.....	6
2.1. <i>What data sources could be useful to enrich the fan experience based on the requirements?</i>	<i>6</i>
2.1.1. Data analytics	6
2.1.2. Literature study	7
2.1.3. Best Good and Bad practices.....	8
2.1.4. Survey	10
2.2. <i>conclusion.....</i>	<i>11</i>
3. User requirements	13
3.1. <i>How will the available data be used to enrich the fan experience of the stadium?.....</i>	<i>13</i>
3.1.1. Interview	13
3.1.2. Literature study	13
3.1.3. Business case exploration.....	15
3.1.4. Requirement prioritization	16
3.2. <i>conclusion.....</i>	<i>17</i>
4. Combining data	18
4.1. <i>How can the available data sources be combined?</i>	<i>18</i>
4.1.1. Literature Study	18
4.1.2. Community research	19
4.1.3. IT architecture Sketching.....	20
4.2. <i>conclusion.....</i>	<i>21</i>
5. Connecting data TO UI.....	22
5.1. <i>How can the combined data be connected to the interface of the application?</i>	<i>22</i>
5.1.1. Document analysis	22
5.1.2. Expert interview	23
5.1.3. Community research	24
5.2. <i>conclusion.....</i>	<i>25</i>
6. Conclusion.....	26
7. Advise & Recommendations.....	27
8. References	28



9. Appendix	30
9.1. <i>What is Fan Experience</i>	30
9.1.2 Interviews	30
9.2. <i>survey results psv</i>	36
9.3. <i>Data sources</i>	39
9.3.1. Ortec.....	39
9.3.2. Sorama	40
9.3.3. Food & Beverage	42
9.3.4. Entrance data.....	43
9.3.5. Social media data.....	45
9.3.6. Other sports.....	48
9.3.8 Interview with Roy smits.....	54
9.3. <i>Project plan</i>	55



1. INTRODUCTION

Fontys Hogescholen HBO ICT & Business semester 7 consists of a Minor, which is Data Driven Business (DDBL). During the Minor an investigation is conducted at a company with an IT issue. This issue is being explored by a group of students called Datastic. The knowledge that the students have is combined and put into practice during the research. The project will take place from August 30 to February 3.

The client for the project will be PSV. This project will take place at the marketing and communication department. PSV is currently creating a Brainport Experience Box within the stadium where supporters can experience the match by a different way and wants to improve the fan experience in the Brainport Experience Box.

The issue that PSV Eindhoven is now facing is that the data collected comes from a variety of sources. The customer needs all the data to be combined, cleaned, and prepared so that it can be presented in the app more easily.

The goal of this project is to deliver a proof of concept with an advice on how to improve the visitors experience in the Brainport Experience Box. In order to do this, data will be researched on how it can make this improvement happen. After that, the data will be collected and cleaned whereafter it can be used to investigate and improve the experience. The ultimate goal is to create an application for visitors of the Brainport Experience Box. This will be done by future student teams.

The project consists of a main question and sub-questions to reach the project goal. The main question is: 'How can existing data be gathered, cleaned, and combined with new data based on the user requirements within the PSV Experience Box to enrich the fan experience?' There are several sub questions made in order to answer the main question:

1. What data could be useful to enrich the fan experience based on the user requirements?
2. How will the available data be used to enrich the fan experience of the stadium?
3. How can the available data sources be combined in a way that it is more manageable?
4. How can the combined data be connected to the interface of the application?

This document has been prepared to capture all relevant basic information and principles of the project so that it can be properly managed. It aims to define the project on the one hand and, on the other hand, to serve as a basis for management and enable assessment of the project's success.

Chapters 2 to 6 describe the process and implementation of the research for each sub-question. Finally, chapter 7 follows in which the conclusion and recommendations are formulated.

2. DATA SOURCES

2.1. WHAT DATA SOURCES COULD BE USEFUL TO ENRICH THE FAN EXPERIENCE BASED ON THE REQUIREMENTS?



Sub question one will be answered in this chapter. To answer the question, Datastic used an approach based on the ICT research methods. The process of the methods can be seen in figure 1.

Figure 1 Research methods question 1

2.1.1. DATA ANALYTICS

PSV uses different types of data to improve fan experience. Some of data is provided by partners of PSV, some is internal. This chapter describes current available data sources, that could be potentially used in PSV Experience Box. In figure 2 underneath the data sources are shown. Then each data source will be explained.

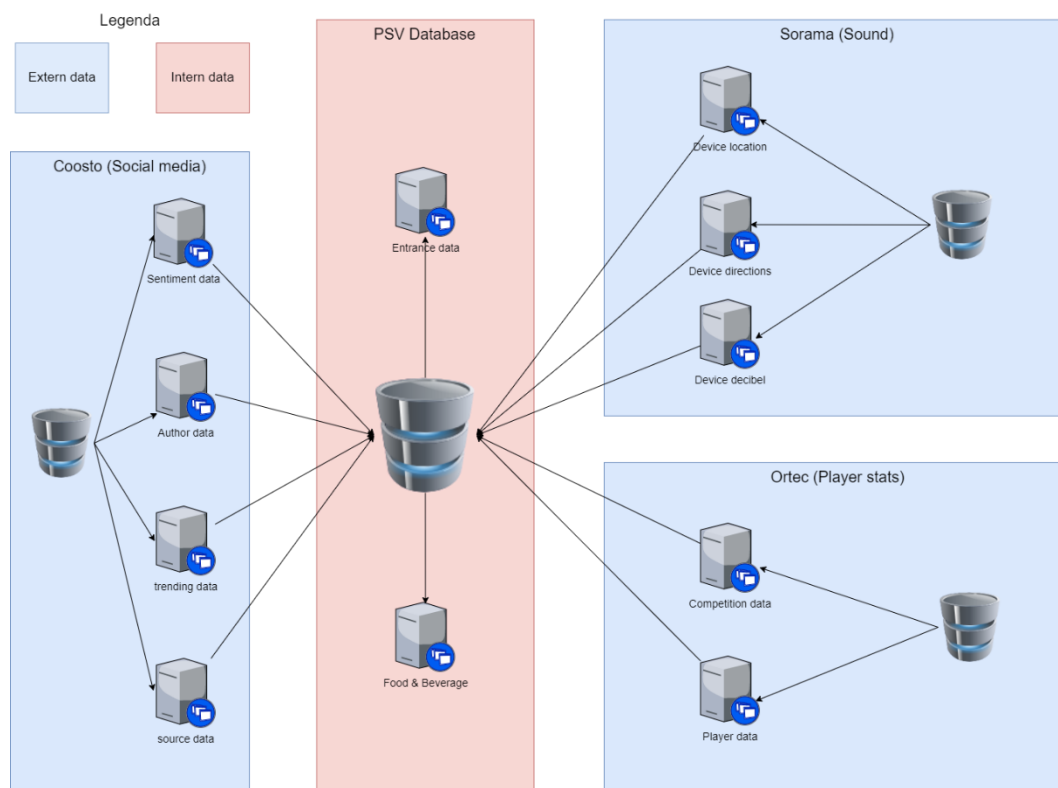


Figure 2 Data flow overview

To have a brief overview of all data sources in one place, a Data Dictionary was created. Data Dictionary is a collection of names, definitions, and attributes about data sources that are being used



within a project. It is not a documentation for data sources but could be useful for new users. The example of Data Dictionary is below. The more detailed description of each data source will be discussed further.

Table 1 data dictionary

Ordinal position	Column name	Type	Is_nullable	Description	Relevant	Note
1	SourceID	int	YES	ID of source	YES	
2	TransactionID	float	NO	ID of transaction	NO	
3	TransactionHistoryID	nvarchar	YES	NULL values	NO	
4	OccasionID	nvarchar	YES	ID of occasion	NO	
5	BeneficiaryID	nvarchar	YES	ID of Beneficiary	NO	personal info
6	CardNumber	float	YES	Number of card	NO	personal info
7	AccessDateTime	datetime2	NO	Time of entrance in format Y-M-D h-m-s	YES	
8	PointofAccess	nvarchar	NO	ID of entrance point	YES	
9	GroupID	nvarchar	YES	ID of Group	NO	
10	OccasionSeason	nvarchar	NO	Number of season (e.g. 19-20)	NO	
11	OccasionStartDate	datetime2	NO	Start of season date	NO	
12	OccasionCompetition	nvarchar	NO	E.g. League	NO	

The first data source is held by Ortec and provides stats of players and matches. This information could help teams to improve their performance, explore the statistics of their opponents, monitor their own statistics and rank players based on a requested playing style or philosophy. Detailed reports are generated on their online portal. So, to summarize, Ortec provides data for analysing sports performance.

The second data source is sound-related. The provider is Sorama, which is high-tech company from Eindhoven. PSV will integrate Sorama’s devices at stadium. Combination of devices and software can trigger anomalies and classify specific sounds. Live sound events can be projected over a map, giving users more insight into sound.

The next data come from internal PSV source, which is Food & Beverage data. PSV provided historical data of total sales, sales per hour and data about the most popular positions. However, there are some issues related to this data source: irrelevant information for fans and no real-time data.

Talking about internal data sources, PSV has another one, which is Entrance data. It shows the average entry in minutes before the match. Again, as the previous source, it is not real-time data.

The last source is social media data provided by Coosto. Different data, such as number of posts’, last posts, posts’ sentiment, trending topics, sources of posts, is gathered from popular social networks. Gathered data is visualized in dashboards.

For an overall overview and the data dictionaries, see chapter 9.3., Data sources.

2.1.2. LITERATURE STUDY

Now the available data is known, Datastic had to research the meaning of Fan Experience to gather the right data. To determine what fan experience exactly is, Datastic performed research. The research is done of two kinds of forms. First, an investigating literature study that is available online is done. Secondly, fans that are visiting the stadium were interviewed.



Sports fans want to see their team doing well. But at the same time, they are also consumers, who have high expectations of any commercial organizations they interact with. Therefore, if clubs want to succeed nowadays, they need to improve fan experience they provide.

Providing a positive fan experience is about focusing on the end-to-end experience that fans receive. It starts long before a matchday (e.g., purchasing tickets) and should last long after it.

Fan experience covers every step – from visiting website to social media interaction, phone calls, marketing activity, and of course everything that happens during matchday, e.g., parking, food& beverage services, safety and so on (adam@olivetreecreative.co.uk, 2018). This means that fan experience is a very broad topic and every human as another expectation of it.

For the interviews that are held with seasonal pass holders at PSV (Appendix 9.1.2), there were a couple of answers given by them. Most of them say that going to the stadium, watching the match with friends and a drink combined with having a positive result, are nice things to have as fan experience. If one of those things is missing, for example a positive result, the fan experience will decrease but still exists. Another thing that is mentioned is nice music and lights. For example, when you enter the stadium, players walk on the pitch, there is a kind of ambiance what is resulting in a nice fan experience.

2.1.3. BEST GOOD AND BAD PRACTICES

Most professional sports have been looking into the adoption of data to increase the fan experience for remote/in-person viewing of games. Approaches to this are surprisingly similar across sports and most use cases are based on the next data sources.

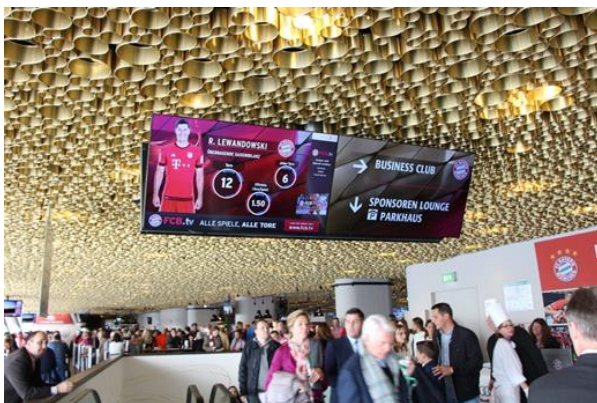


Figure 3 TV at venue

Teams gather data about the players to showcase information about them during games and provide more insight to viewers. This is easily done for remote viewers via the tv screen or streaming service used but also added to stadiums or venues by adding TVs on location or creating phone applications that are accessible either freely or subscription based. (AR to sports)



The replay functionality also seems to play a vital role in most data-adoption use-cases. This has been made accessible in venues in the same way as the player data and is becoming more and more used all around different sports. In baseball, football, and basketball a few teams have been successful in showing instant replays for players in the venues but still struggle with issues like delay and buffering since they lack 5g technology (most cases). Wolfsburg, a German football team claims to have fixed this issue (Wolfsburg).



Figure 4 5G in stadium

A basketball team LA Clippers, leaders in data adoption and AR, successfully managed to show player data using AR technologies in the instant replays' functionalities and even provide a subscription-based product, CourtVision. CourtVision is intended to revolutionize the sports watching experience, with the help of data. They provide replay functionality that, with AR showcases a vast amount of insight into the players, teams and even coaches. Their data storage of choice is Amazon Web Services (AWS). (CourtVision)

Features showcasing interesting data about teams in general also play a big part in these products.



Figure 5 basketball game LA Clippers

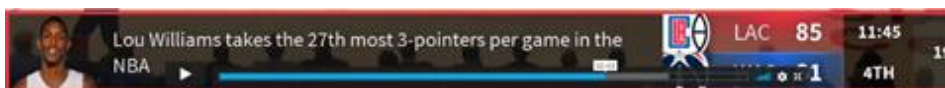


Figure 6 additional data regarding the game

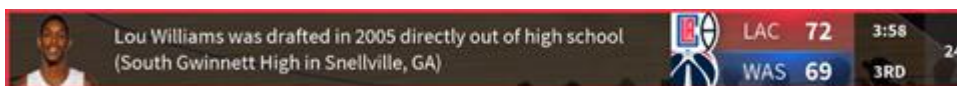


Figure 7 additional data regarding the player

Other data sources haven't been focused on extensively due to the much bigger interest in the previous mentioned data, but teams also provide publicly accessible data such as weather data, social media data, transport data and others.

For an overall overview and the data dictionaries, see chapter 9.2.6., Other sports.

2.1.4. SURVEY

A survey was set out under visitors of the Brainport Experience Box. An extensive analysis can be found in appendix 9.2.

Seventeen responses were received which is a response rate of 43,6%. This means a reliability level of 87% and a margin of error of 14%.

The rates respondents give to watching the game have been analyzed. There has been looked into the rates watching on tv, in the stadium and into the experience box. If a respondent never watched on tv or in the stadium they did not rate the experience. As shown in Figure 8, the stadium experience has the highest mean score, after which the Experience Box scores the second highest. Watching the game on tv scores the lowest. From this numbers, the following remarks could be made:

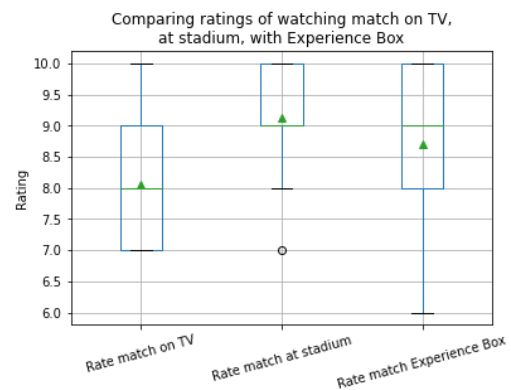


Figure 8 boxplot experience ratings

- The highest rating has "Match at stadium" with mean value 9.13. "Match at PSV Experience Box" has lower mean value - 8.71. And "Match on TV" has the lowest mean rating - 8.06.
- Fans will probably prefer watching match at stadium with or without PSV Experience Box rather than watching match on TV.
- However, the lowest given rate (6) got "Match at PSV Experience Box". For other 2 groups the lowest given rate is 7. Moreover, rate 7 is recognized as outlier for "Match at stadium".
- 75% of respondents' rates for "Match at stadium" are in range from 9 to 10 but it also worth mentioning, that "Match at stadium" is missing 2 values, so the total number of responds for this column is 15.

There was also looked if the expectations from the experience box were met. As shown in Figure 9, for 82.4% of the respondents the expectations were met. For 17.6% it was not met. There were a variety of reasons why this was the case.



Figure 9 expectations Experience Box

The last question that was asked was what visitors of the Experience Box would like to see. There were six options given and the respondents could rank them from one to six. Number one got the most points and number six the least. In Figure 10 the overall scores are visualized. Data that would be appreciated is the player's and social media data. The respondents valued the weather and traffic-and public transport data the least.

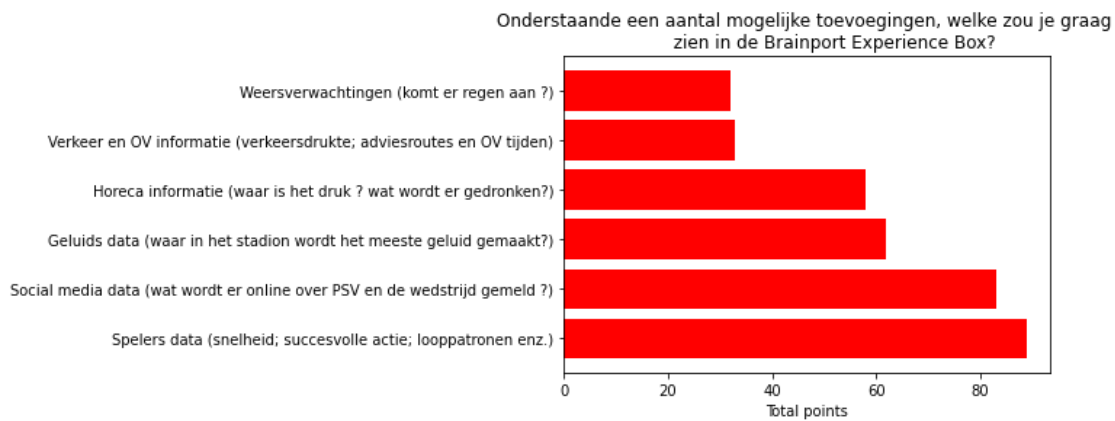


Figure 10 additions Experience Box

As was already mentioned, PSV and Datastic made a survey to reinforce the research. The target group for that research was fans, who have already visited PSV Experience Box. However, to improve project results, PSV may later make a second survey, this time for all PSV fans. One of the ways to do that is to send mailing to season card holders, another way is to conduct a survey right in front of the stadium.

Besides that, one more solution exists. Almost all football clubs have fans community, and these communities often have online resources where they communicate. It could be social media, youtube channels, websites etc. PSV is not an exception.

The biggest PSV fan community is "PSV Fans". Besides the website, they have accounts in all popular social media: Instagram, Facebook and Twitter (number of followers is as for 01.12.2021).

<https://www.psvfans.nl/>

<https://www.facebook.com/psvfans1913> (26 000 followers)

https://twitter.com/psv_fansnl (10200 followers)

<https://www.instagram.com/psvfansnl/> (2359 followers)



Figure 11 logo PSV fans website

As fans could be also considered as customers of PSV, it is essential to listen to their opinions and wishes. Mentioned fans community could be potentially used as a distribution channel. In our case, a survey can be made and published on 'PSV Fans' by agreement with the owner of this website.

The second website for PSV fans is <https://psv.supporters.nl>. However, it is a website on the domain supportert.nl, so similar websites exist (e.g. ajax.supporters.nl). Therefore, it is a formulaic website and unlikely useful for project scope.

2.2. CONCLUSION

Fan experience is a broad topic that could mean something else to every fan visiting the stadium. This requires a broad approach to enrich the fan experience. A commonly used addition are (TV) screens in the stadium or



venue. To be able to have a replay function on all the additional screens, it is important to have a good internet connection.

From all the different data sources that are available, the player data from Ortec, the social media data from Coosto and the sound data from Sorama have the most added value regarding the fan experience. All these data sources are external. Besides that, the internal data regarding food & beverage and entrance data are also ranked high. These data sources could be useful to enrich the fan experience.

3. USER REQUIREMENTS

3.1. HOW WILL THE AVAILABLE DATA BE USED TO ENRICH THE FAN EXPERIENCE OF THE STADIUM?

Sub question two will be answered in this chapter. The process of the methods Dastatic used to answer the question can be seen in Figure 12.

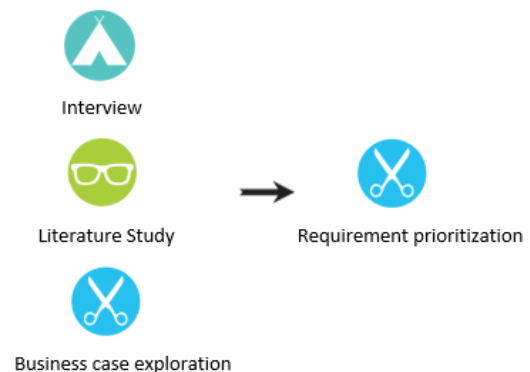


Figure 12 Research methods question 2

3.1.1. INTERVIEW

To get a better understanding of the needs of PSV fans, five interviews were held. J. Kamp gave some advice to do approximately 5/6 interviews and compare the results. The results must match for two third. Otherwise, more interviews must be done to achieve this result. By having five interviews, complete and reliable research is compiled, because all interviewees gave approximately the same answers.

Structured interviews were chosen to obtain detailed information about the problem and the wishes. In this way it is easy to compare the answers because every interviewee gets the same questions.

The first outcome of the interview shows that four of the five interviewees travel by car to the stadium and therefore do not find the transport data useful. We can conclude that this data does not have a high priority

The second outcome is that three of the five interviewees specifically would like to see player data. Besides that, 2 of the five say they would like to see social media and sound data as well. The conclusion is that the most wanted data is player, sound and social media data.

In addition to the wanted data. If it is shown on a tablet all of them say that it would increase their fan experience a little. But on the other hand, they also say they won't use the tablet during a match.

For the full interviews, see appendix 9.1.

3.1.2. LITERATURE STUDY

In this chapter a brief explanation about customer requirements and needs will be provided. Customers have the power to influence the success of one company. But before diving into the requirements of the PSV fans, let's answer the question what exactly are customer needs and how can they be identified?



A customer needs motivation that leads to purchasing a product or service. The necessity drives the customer's making choice. Customers' needs are frequently viewed as an opportunity to resolve or offer surplus value back to the initial reason. The first stage in determining customer requirements is to conduct a customer needs investigation that helps the company determine consumer needs. Football teams rely on relationship marketing to create and sustain long-term relationships with fans.

After summarizing the results of all interviews, the student group discovered that the main requirements of the customers/fans are information and statistics about the players and their performance; social media, sound, and food and beverage data; and an instant video replay function, so that even if the fans miss a moment from the match, they can immediately review it on the tablets.

There are several different types of requirements and placing the fan needs in the right category is essential for further investigation. The main types of requirements are:

- Functional Requirements

A functional requirement is a task that must be completed to give operational capability (or satisfy an operational requirement). Some operational and support-related functional needs can be deduced from the required operational competence. Others are frequently the outcome of meticulous systems engineering. Most systems must accomplish eight generic roles over their life cycle, according to systems engineering experience: development, production, verification, deployment, training, operations, support, and disposal. The eight basic system functions are referred to as such. Each must typically be considered when determining all a system's functional needs.

- Performance Requirements

A performance requirement is a specification of the extent to which a function must be carried out, which is usually expressed in terms of quantity, accuracy, coverage, timeliness, or preparedness. Other performance criteria generally necessitate extensive systems engineering.

- System Technical Requirements

As an outcome, both allocated and derived needs are generated.

- Allocated Requirements

These are requirements that flow directly from the system requirements to the system elements.

Requirements that are derived from the design solution (and so are sometimes called design requirements). Internal interface limitations between system pieces are among them.

- Specifications

Based on the requirements that were summarized in the upper paragraph a conclusion can be made that they are categorized as specifications. The general goal of a specification is to establish a foundation for acquiring a product or service that meets a specified demand. A specification, by definition, establishes boundaries and hence removes objects that fall outside of those parameters.

(Breschi, 2022)



3.1.3. BUSINESS CASE EXPLORATION

In this subchapter there are described data volumes, known costs and risks concerning the delivery of PSV's product.

Sorama data

Initially, it has a huge volume of approximately 20GB and is stored in a folder system with text files. After being processed by Datastic, it is an Excel file, with about 8MB and 100.000 rows. The possibility of retrieving data using a websocket has not been explored by Datastic yet, but it can help retrieve live data and potentially lower the size that this data has.

Ortec data

This data is retrieved using an API and a python script. It is stored in the local SQL database, and it has ten tables. Considering this data will be stored by Ortec itself PSV can concern themselves with retrieving only the data that is valuable, thus lowering the size needed for storage.

Coosto data

Using an API and a python script, Datastic retrieved one table that shows the sentiment revolving the #PSV tag in social media environments. It has 93 rows, and it is stored in the local SQL database. To access Coosto API regularly a 50€ subscription-based fee is required.

Entrance data

Datastic received this data from PSV. It has 1447 rows, and it is stored in the local SQL database.

Food and beverage

Datastic received this data from PSV. It has 396 rows and it is stored in the local SQL database.

Risks

PSV needs to consider risks when trying to deliver this product to the fans. Failing to deliver the promised quality for the app statistics can lower the amount of people who would want to have access to it. This is also the case if PSV would delay promised features added on the tablets. Another risk would be attracting less interest than expected and thus creating a product that does not appeal to many fans. As an innovative project, PSV should make sure that investing in it is justified, and the project will prove to be of a profit for the company.



3.1.4. REQUIREMENT PRIORITIZATION

In this subchapter there will be described how Datastic applied the MoSCoW prioritization method to sort out the most important aspects for the project from the final user perspective.

To gather all the requirements, Datastic interviewed seasonal pass holders for their interests. Also, fans that have been used the Experience Box are asked for their thoughts. These interests are transferred to requirements and based on the MoSCoW method they will be prioritized.

Pros of using this prioritization framework:

- It's good for involving stakeholders without a technical background in the product prioritization process.
- Quick, easy, and intuitive way of communicating priorities to the team and the customers.
- It allows you to think about resource allocation when you classify your features and requirements into each bucket.

Cons of using this prioritization framework:

- It's tempting for teams and stakeholders to overestimate the number of Must-Have features.
- It's an exercise in formulating release criteria more than a prioritization method.

The MoSCoW method is prioritizing the requirements in different groups. Underneath every group is explained (Productplan.com, 2021):

- Must Haves; As the name suggests, this category consists of initiatives that are “musts” for the team
- Should Haves; They are essential to the product, project, or release, but they are not vital. If left out, the product or project still functions. However, the initiatives may add significant value.
- Could haves; initiatives are not necessary to the core function of the product. However, compared with “should-have” initiatives, they have a much smaller impact on the outcome if left out.
- Won't Haves; The category can manage expectations about what the team will not include in a specific release (or another timeframe you're prioritizing).

Underneath every requirement is placed into the framework, this framework created a clear view about the priority's Datastic must take:

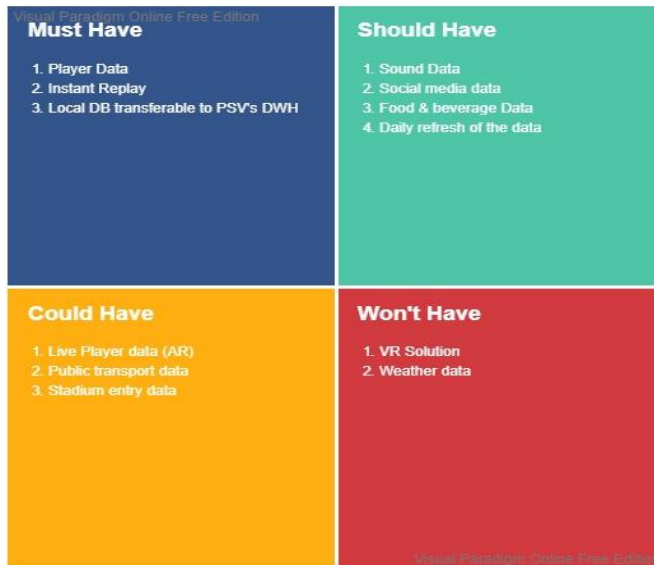


Figure 13 MoSCoW requirement framework

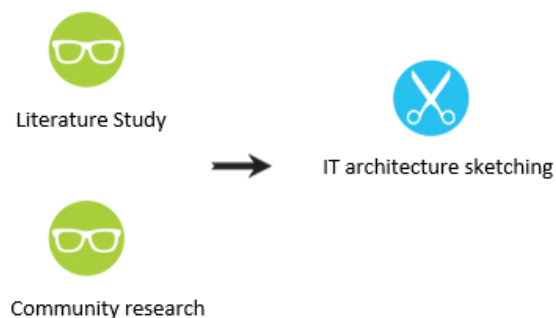
3.2. CONCLUSION

The requirements for the experience box show that customers are open for new and interesting sources of information that can help them experience the moment to the most and will give them more inside to the game. When looking at using the different data sources, a few risks should be considered. Next to that is it important to keep the expected costs and time schedule in mind. The MoSCoW requirement framework can be concluded that not all the data that is available should be used because it will not enrich the fan experience.

4. COMBINING DATA

4.1. HOW CAN THE AVAILABLE DATA SOURCES BE COMBINED?

Sub question two will be answered in this chapter. The process of the methods Datastic used to answer the question can be seen in Figure 14.



4.1.1. LITERATURE STUDY

To basically understand this kind of process, Datastic should look to ETL. This means Extract, Transform & load. This is the dominant paradigm for efficiently getting data from multiple sources into a single location where it can be used for self-service queries and data analytics. ETL consists of three sub-processes (Tobin, 2020):

- Extract: Data is first getting extracted from its source locations, these sources may be files, websites, software applications and relation/non-relational databases.
- Transform: The extracted data will be transformed to make it suitable for your purposes. Think about adding & deleting rows, joining sources together and converting data in one format to another.
- Load: Finally, the transformed data will be loaded into the target location, which is a database/data warehouse most of the time.

The transforming process is the most important part of combining data. A couple of steps must be made to ensure that the data you are saving is correct.

Step 1: Data cleaning

This involves deleting information that is old, inaccurate, duplicate or out of date. As well as detecting errors and typos.

Step 2: Data reconciliation

With data reconciliation the identification, integration and standardization of different data records that refer to the same entity will be made.

Step 3: Data summarization

This will add new data records by performing operations on existing records. This step could be ignored if it's not the case into your dataset.

Step 4: Data filtering

Data filtering ignores irrelevant information by selecting only necessary rows, columns, and fields from a larger dataset.

Step 5: Data aggregation

By aggregating the data, it will be combined from multiple sources so it can be presented in a more digestible, understandable format.



When Datomic is going to combine the data sources, they will investigate the steps that were researched. By doing this, following this order, and completing every step, they will try to get the data combined in the best possible way.

4.1.2. COMMUNITY RESEARCH

Combining or blending data sources is a part of ETL-process, which was described above. However, there is another approach - ELT. It is generally the same process, except that 2nd and 3rd step are swapped. One the providers of SaS ELT platform is Panoply (panoply.io). According to the community of Panoply (blog.panoply.io) the ELT approach involves extracting data, and then storing or loading it into a data warehouse. From there, you can combine or transform data to solve business needs. One



Figure 15 Panoply logo

important distinction is that data blending requires a primary and secondary data source. The primary data source is the original source that you pull information from. This is typically raw data like sales information for an individual or a team. The secondary source is data that you extract from one or more primary sources for research and analysis. The secondary source operates completely independently of the primary source. By using primary and secondary data sources, you can separate the data and avoid losing or altering information during the blending phase. (ETL vs ELT: The Difference is in the How, 2021)

It's easy to confuse *data blending* with *data joining* because these two concepts are similar. The key difference is that data joining only allows you to combine data from a single source. For example, you may have data from two SQL databases that you need to combine into a single location. In this case, you could join the databases together into one set. You could also join data from different Excel sheets.

Data blending first requires setting up a blended data source. First, the data needs to be acquired and find the information you need across different locations. This may include social media, spreadsheets, and tables, among other things.

Next step is to combine and load data to a destination for storage (e.g., a data warehouse). The third step is to clean the data. This involves:

- correcting data entry errors
- deleting unnecessary data
- and redesigning the data set into a better format

4.1.3. IT ARCHITECTURE SKETCHING

In Figure 16 the IT architecture is shown. The different primary sources will be combined and with help of a Python script put into the MS SQL database. For both the Coosto and Ortec data base, an API is used to make the connection. In the figure is shown which tables can be found in these databases and how often it is updated. For the connection with the Sorama data, a web socket is used. This web socket will make a direct connection to the interface without going to the PSV database first. Next to the external data, internal data is also used. This data is put directly into the PSV data source from the primary source, without help of a Python script.

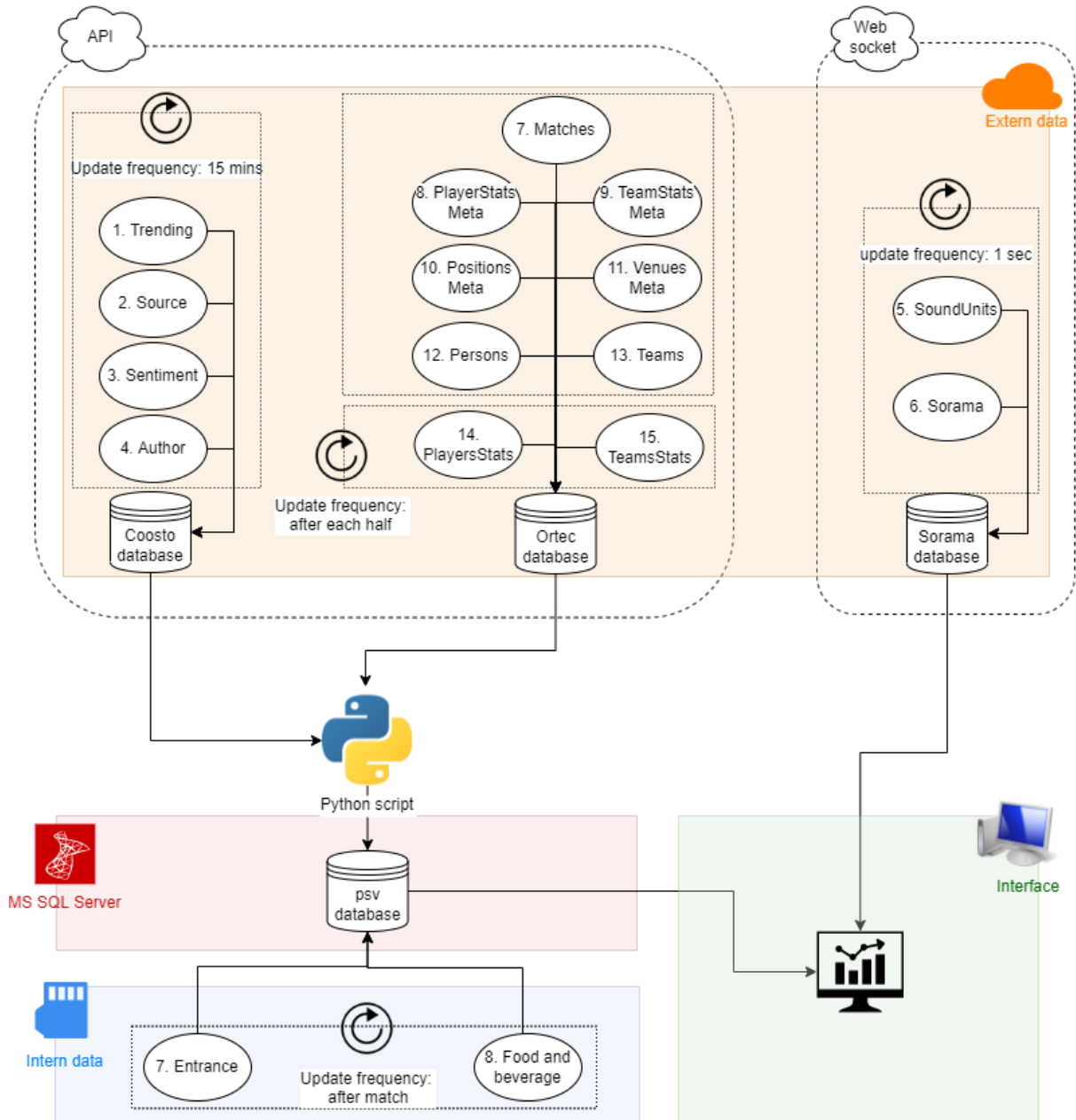


Figure 16 IT Architecture



4.2. CONCLUSION

Before the data can be used for the interface, a transforming process consisting of five steps needs to be made. When these steps are made, the data can be combined in the most optimal way.

For the combining of blending stage, the ETL-process can be used. The data that is used will come from the primary source and will be altered in the secondary source in order to avoid losing or altering information from the primary source. Datastic will put every data source in a MS SQL data base. This database is the secondary data source. This will be used for the user interface.

5. CONNECTING DATA TO UI

5.1. HOW CAN THE COMBINED DATA BE CONNECTED TO THE INTERFACE OF THE APPLICATION?

Sub question two will be answered in this chapter. The process of the methods Datastic used to answer the question can be seen in Figure 17.

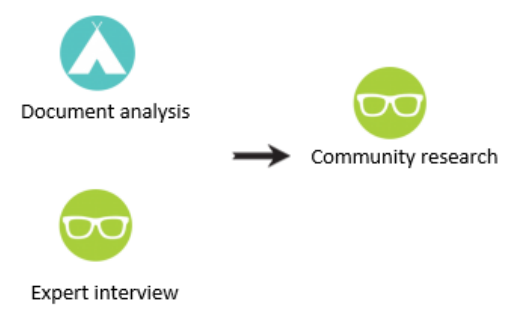


Figure 17 Research methods question 4

5.1.1. DOCUMENT ANALYSIS

A project group from the minor Digital Experience Design. made a design for the application that eventually will be used. The application is divided in three different pages, namely livestream, dashboard, and food & drink.

On the livestream page, the game is being streamed. In this way, visitors are able to watch the game from other angles. The data dashboard shows different information about the competition, the team, the different players, and statistics about the current match. For this, the data from Ortec is used. As shown in the IT architecture, this data will be updated after each half. The look of this page is show in Figure 18.

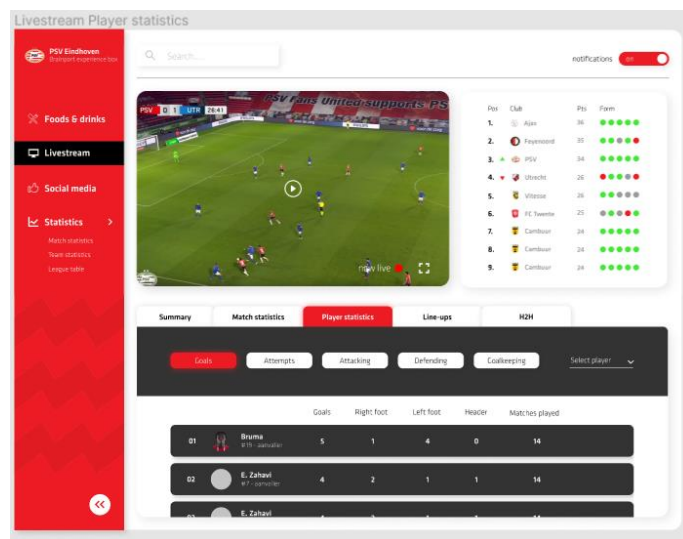


Figure 18 player statistics

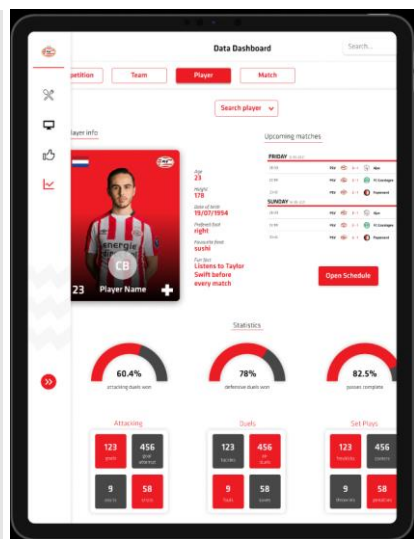


Figure 19 livestream and player statistics

Lastly, the food and beverage page show the different food and beverages that can be ordered in-seat. In Figure 20 is shown how ordering in the app looks like. This regards internal data.

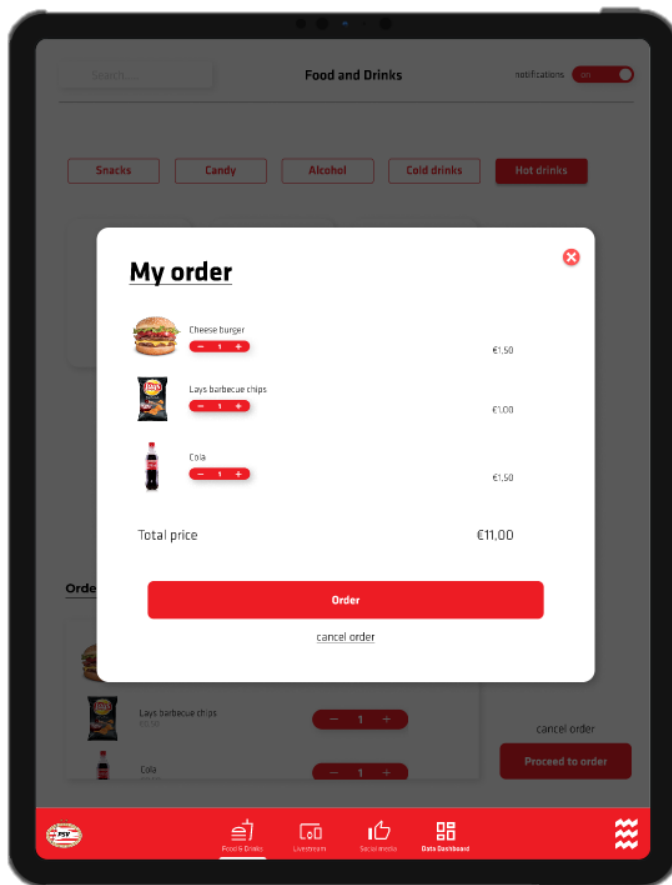


Figure 20 food & beverage page

5.1.2. EXPERT INTERVIEW

To give more weight to the research, an expert interview about combining various sources of data was conducted with a person from PSV – Roy Smits. Because the interview questions were conducted by Datastic, they were created with the group's needs in mind in connection with the research question.

The interview that was done has resulted in some input and ideas on how to combine all the data sources that are provided to the project group and information for the insides of the current environment of PSV. The interview resulted in pointing out that PSV has a data warehouse where once a day the data from the sport Alliance warehouse is dumped into. For this warehouse, an SQL-server is used in an own private cloud environment of C-mac but migrating to Microsoft Azure is considered.

A daily dump from the Sport Alliance warehouse to the PSV warehouse is made. The link between the Sport Alliance and their own warehouse is made automatically and an advantage of this is that no extra tables need to be made in their warehouse. That is why the football club imports it into their own data warehouse. The data in the warehouse is kept for a lifetime. However, the player data is too large and varied to store it all, so not all the data is stored.

The whole interview can be found in appendix 9.3.8

5.1.3. COMMUNITY RESEARCH

During this research multiple data sources were investigated. To fulfill the assignment, the student group is researching different types of opportunities to combine the available data and one of these ways is by conducting community research that can give insides about the topic.

After some investigation about the topic in trusted online communities, it came out that one of the ways to combine data is by data merging. Data merging is the process of combining two or more data sets into a single data set. When there is raw data saved in multiple files, worksheets, or datasets that need to be examined all at once, this technique is required.

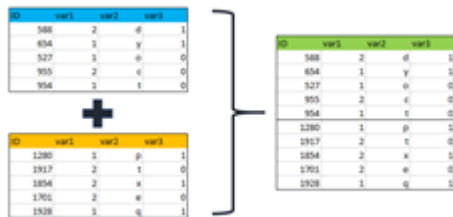


Figure 21 merging new cases

There are several types of data merging. Merging in new cases assumes that the variables in the two files you are merging are the same in nature. This also assumes that the IDs for each case are different. If you have a variable in one file that does not have a match in the other, then missing data (blank values) may be inserted for those rows that do not have data

Another type is merging in new variables. Merging in new variables requires the IDs for each case in the two files to be the same, but the variable names should be different. you are adding new variables with information for each existing case in your data file. As with merging new cases where not all variables are present, the same thing applies if you merge into new variables where some cases are missing – these should simply be given blank values. It could also happen that you have a new file with both new cases and new variables. The approach then will depend on the software that is used for a merger. If the software cannot handle merging both variables and cases at the same time, then consider first merging in only the new variables for the existing sample, and then append the new cases across all variables as a second step to the merger.

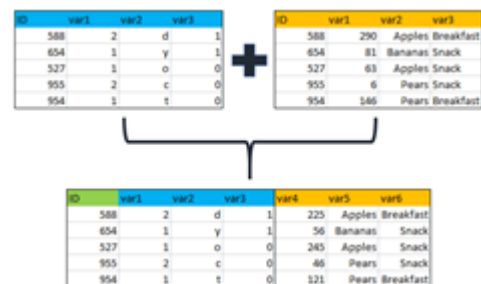


Figure 22 merging new variables

Another method of merging is to augment the data with information from other sources. If new data needs to be added to an existing table, for example, a look-up code needs to be added as the identifier and to enter the values into a new variable in our data file. The formula needed for this is =VLOOKUP (). The look-up code should be unique inside the file holding the extra data, but the same value may appear several times within the file that is being improved.

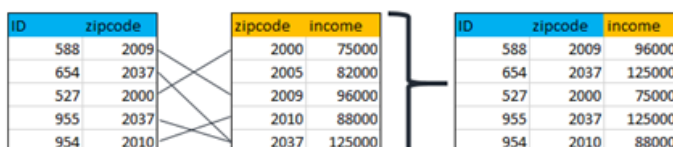


Figure 23 augment data from other sources



Combining diverse data sources opens new potential to expand the utility of each data source while also posing challenges in terms of the approach to be utilized to extract information from these sources.

(Engdahl, 2020) & (Kayfitz, 2019)

5.2. CONCLUSION

The user interface is designed by another group of students. The data sources that are provided by PSV and Datastic are used to optimize the fan experience within the application. Both internal and external data is used. When looking at the secondary data source, currently an SQL-server is used. In the future, PSV could switch to Microsoft Azure. The merging of data can be done in several different ways. For PSV, it is recommended to look at the type of data that needs to be merged and chose a type of technique based on that.



6. CONCLUSION

In conclusion, Datastic researched the possibility of using five different data sources to improve the fan experience within the PSV stadium. These data sources could be used to improve the PSV app that fans can access from the Experience Box tablets. Datastic also investigated prioritizing the types of data used by asking regular season card holders for PSV, conducted a survey, community research and other practices to find out what the user needs are. Using real-time data played an important role into how to store and to process the data.

The data sources that Datastic investigated are entrance data, which is data regarding information about people who attend the games in the stadium, food and beverage data, regarding the products sold within the stadium, Coosto data, regarding social media data, Ortec, regarding players and teams and Sorama data, regarding sound in the stadium. Entrance data is the only data source that could not be processed in real-time, but one day after the event. Datastic did not research data sources that proved uninteresting to fans.

The research concludes that these data sources play a major role in enriching the fan experience not only in football, but all the other sports. But one data source is not that important as the other one. During our interviews we asked fans to rate data use cases to see which ones they appreciate more. The rating is as follows, from low to high priority: Weather data, Traffic and public transport data, food and beverage data, Sorama sound data, Coosto social media data, Ortec player data.

However, while fans were enthusiastic about the use cases of these data sources, they tend to be reluctant about using the tablets during games saying it causes distraction. There is clearly a need for a balance between quantity of data shown and the quality of data shown. The way in which data should be represented should cause interest and should incentivize talks and interaction between fans.

Storing this data was made using local databases, which proved to be quite bothersome because of the volume of the data stored, its complexity and the fact that real-time data should be used. Thus, Datastic concluded that in the future, the adoption of a cloud service for data management would be advised.



7. ADVISE & RECOMMENDATIONS

During the project Datastic has researched various data sources and how to integrate them into a more accessible, organized database structure. Datastic will give an advice based on the data sources. At the end an advice will follow about implementing the code into the PSV environment.

Data sources

Entrance data: Currently, there is no real-time data that Datastic can use because the data is received one day after the event took place. However, it is possible to receive the data real time. Based on the survey and interviews Datastic advice to not use this data in the dashboard, because of no interest of the fans.

First to discuss is the data source Food and beverage. The interview and survey results show that fans do not place a high interest for this type of data. However, if it would be possible to see where the shortest queue is at the stadiums' bars, they would find it very useful. Currently this is not possible because the data shows us the number of products sold, not the amount of people queueing up to buy the products. Datastic considers this to be of a lower priority than working with the other data sources.

Secondly the advice for the data source of Coosto. Based on the survey results, the social media data has a high place in the interest of the fans, however based on the interviews with regular season holders the consensus was unanimously for not checking the tablets during the game. Datastic advises to include social media data in the tablet's dashboards', but to ensure that fans are incentivised to look at them during games, more research is needed.

Then the advice of the most valuable data source from Ortec will be described. Player data is the most used data source not only for football, but all other sports as well. The format of the data is complex, having multiple tables that are closely related to each other. Other sports, basketball, formula 1, have developed their data processing much more than in football and represent good examples of how to showcase this type of data in real-time to the fans. Because of this and the research conducted Datastic concludes that this is a vital source that PSV should use in their dashboards.

Finally, the advice for the Sorama data source. Before being processed, the data received by Datastic was about 23GBs stored locally in a file format. This doesn't allow us to process real-time data and show it to PSV's viewers. Alternatively, it is possible to use a WebSocket to retrieve real-time data from Sorama, this data is much easier to work with since it is already partly processed and could provide a solution to showing accurate, real-time visualizations.

A WebSocket is a way of communicating between a client and a server through a TCP handshake connection. This can be done using C #, Python, JavaScript and others. Datastic did not create this script due to time constraints. If the connection between the WebSocket and Tableau must be made to create visualizations in real-time, this can be done using an enterprise data pipeline, like Alooma.

Data Cloud Service

PSV should consider using a cloud storage service. Working with dynamic data that is coming together in real-time can be quite bothersome, use a lot of disk space and make it hard to process locally stored data. Also, since the project hasn't matured yet, it can be viewed as an opportunity to adopt the use of a cloud service and build the backbone of the PSV's data organization from scratch. There are a few advantages when using cloud services namely, easily accessible data, security, recovery and possibly price. Three of the most used cloud services that Datastic recommends are: Microsoft Azure, Amazon Web Services, Google Cloud.



8. REFERENCES

- Brainport Experience Box . (n.d.). Retrieved from brainporteindhoven.com:
<https://brainporteindhoven.com/psv/nl/trots/experiencebox>
- What is pestle analysis. (n.d.). Retrieved from <https://pestleanalysis.com/what-is-pestle-analysis/>
- RISK MITIGATION PLANNING. (n.d.). Retrieved from <https://www.mitre.org/publications/systems-engineering-guide/acquisition-systems-engineering/risk-management/risk-mitigation-planning-implementation-and-progress-monitoring>
- PSV Eindhoven. (2021, 09 09). *PSV - Organisatie*. Retrieved from PSV.nl.
- Brainport Eindhoven. (2021, 09 20). *Wat is Brainport Eindhoven*. Retrieved from Brainport Eindhoven:
<https://brainporteindhoven.com/nl/ontdek/wat-is-brainport-eindhoven>
- Koersen, W. (2018, 06 21). *Web Scraping, Regular Expressions, and Data Visualization: Doing it all in Python*. Retrieved from Towardsdatascience: <https://towardsdatascience.com/web-scraping-regular-expressions-and-data-visualization-doing-it-all-in-python-37a1aade7924>
- Summit. (n.d.). *Reading Data from the Web: Web Scraping & Regular Expressions*. Retrieved from Summit:
<https://www.summet.com/dmsi/html/readingTheWeb.html>
- MHerman. (2012, 05 10). *Web Scraping with Regular Expressions*. Retrieved from MHerman:
<https://mherman.org/blog/web-scraping-with-regular-expressions/>
- Octoparse. (n.d.). *Using This RegEx Tool to Match HTML Tags*. Retrieved from Octoparse:
<https://www.octoparse.com/blog/using-regular-expression-to-match-html#>
- Techopedia. (2017, 03 23). *Parse*. Retrieved from techopedia:
<https://www.techopedia.com/definition/3853/parse>
- Developer. (2021, 09 20). *Parse - MDN Web Docs Glossary: Definitions of Web-related terms | MDN*. Retrieved from Developer: <https://developer.mozilla.org/en-US/docs/Glossary/Parse?retiredLocale=nl>
- html5rocks. (n.d.). *How Browsers Work: Behind the scenes of modern web browsers - HTML5 Rocks*. Retrieved from html5rocks:
https://www.html5rocks.com/en/tutorials/internals/howbrowserswork/#HTML_Parser
- Ortec sports. (2020). *Business Case Performance* . Retrieved from www.ortecsports.com:
<https://ortecsports.com/wp-content/uploads/2020/03/Business-Case-Performance-PSV.pdf>
- adam@olivetreecreative.co.uk. (2018, 05 15). *Why is everyone talking about 'fan experience'?* Retrieved from rapidretail: <https://rapidretail.co.uk/talking-about-fan-experience/>
- Productplan.com. (2021, 09 11). *MoSCoW Prioritization*. Retrieved from Productplan:
<https://www.productplan.com/glossary/moscow-prioritization/>
- CourtVision. (n.d.). <https://www.clipperscourtvision.com/>. Retrieved from <https://www.clipperscourtvision.com/>
- AR to sports. (n.d.). Retrieved from <https://digital.hbs.edu/data-and-analysis/bringing-the-ar-cloud-to-sports/>



Tobin, D. (2020, 07 08). *How to Get Data from Multiple Sources*. Retrieved from xplenty: <https://www.xplenty.com/blog/get-data-from-multiple-sources/>

Breschi, A. (2022, 01 12). *16 Types of Customer Needs (and How to Solve for Them)*. Retrieved from Blog Hubspot: <https://blog.hubspot.com/service/customer-needs>

Engdahl, M. (2020, 12 7). *What is data merging?* Retrieved from Displayr: <https://www.displayr.com/what-is-data-merging/>

Kayfitz, A. (2019, 5 2). *How to Merge Data from Multiple Sources into a Single Platform*. . Retrieved from Practical Ecommerce: <https://www.practicalecommerce.com/how-to-merge-data-from-multiple-sources-into-a-single-platform>

ETL vs ELT: The Difference is in the How. (2021, 07 27). Retrieved from Panoply: <https://blog.panoply.io/etl-vs-elt-the-difference-is-in-the-how>

(n.d.). <https://www.dfl.de/en/news/kick-off-for-5g-latest-mobile-communications-technology-activated-in-the-first-bundesliga-stadium/>. Retrieved from <https://www.dfl.de/en/news/kick-off-for-5g-latest-mobile-communications-technology-activated-in-the-first-bundesliga-stadium/>



9. APPENDIX

9.1. WHAT IS FAN EXPERIENCE

Before the sub question 2 can be answered, which is: "What other data could be useful to enrich the fan experience based on the requirements?", It is important to know the meaning of Fan Experience. Datastic did this based on interview, which can be seen underneath.

9.1.2 INTERVIEWS

Interview Bas:

What is your age?

25

What is your gender?

Man

Do you live in or outside Eindhoven?

Buiten

Do you travel to the stadium by public transport?

Ja, trein

In your opinion, what do you think fan experience is?

De volle beleving van naar de wedstrijd toelevens, zowel buiten als in het stadion, om vervolgens onder het genot van n drankje de wedstrijd te kijken.

Do you know the Brainport Experience Box in the PSV Stadium?

Nee.

What additional information would you like to see during a match? (at the moment we have weather data, public transport data, Player data, Food & Beverage, Social media data, Sound data)

Meer uitgebreide speler data, deze is vaak zichtbaar na de wedstrijd maar zal deze ook tijdens de wedstrijd willen zien.

When information is available on a tablet, do you think the tablet will be used during the match?

Nee, op dit moment denk ik van niet. Ik geniet van het traditioneel voetbal kijken met een biertje.

If you will be able to see the player performance data on your screen, do you think it would increase your fan experience?

Ja, in mijn geval wel. Ik kijk tijdens de wedstrijd vaak op mijn telefoon voor statistieken en als dit op het scherm getoond wordt, hoeft dat niet meer.



Interview Dirk:

What is your age?

25

What is your gender?

Man

Do you live in or outside Eindhoven?

Outside, Someren

Do you travel to the stadium by public transport?

Nee altijd met de auto

In your opinion, what do you think fan experience is?

Een ultieme nieuwe beleving om naar voetbal te kijken. Ik denk dat dit een ruim begrip is. In mijn geval is het de beleving van naar het stadion toe gaan, bier drinken, wedstrijd kijken en tevreden zijn met het resultaat.

Do you know the Brainport Experience Box in the PSV Stadium?

Ja via Remco omdat hij daar op school iets voor moet doen, daarvoor had ik er nog nooit van gehoord.

What additional information would you like to see during a match? (at the moment we have weather data, public transport data, Player data, Food & Beverage, Social media data, Sound data)

Ik denk dat die genoeg is heb verder geen voorkeur

When information is available on a tablet, do you think the tablet will be used during the match?

Ik denk op sommige momenten wel als je snel een goal of een overtreding terug wil zien

If you will be able to see the player performance data on your screen, do you think it would increase your fan experience?

Met het terugkijken van beelden wel en ook speler statistieken zou interessant zijn.



Interview Cas:

What is your age?

26

What is your gender?

Man

Do you live in or outside Eindhoven?

Buiten Eindhoven

Do you travel to the stadium by public transport?

Nee

In your opinion, what do you think fan experience is?

Ik denk dat veel fans de ervaring leuk vinden in het stadion. Er is altijd een feestje en veel gezelligheid.

Do you know the Brainport Experience Box in the PSV Stadium?

Wel eens wat van voorbij zien komen op social media.

What additional information would you like to see during a match? (at the moment we have weather data, public transport data, Player data, Food & Beverage, Social media data, Sound data)

Ik denk dat dit al genoeg is.

When information is available on a tablet, do you think the tablet will be used during the match?

Ik ben er zelf niet echt een fan van. Ik zou het wel snappen als mensen dit leuk vinden.

If you will be able to see the player performance data on your screen, do you think it would increase your fan experience?

Misschien een beetje.



Interview Roy:

What is your age?

25

What is your gender?

Male

Do you live in or outside Eindhoven?

Outside, Asten.

Do you travel to the stadium by public transport?

We travel with the car.

In your opinion, what do you think fan experience is?

Good football and some nice music and lights.

Do you know the Brainport Experience Box in the PSV Stadium?

Yes, heard of it!

What additional information would you like to see during a match? (at the moment we have weather data, public transport data, Player data, Food & Beverage, Social media data, Sound data)

Sound, social media and player data is fine!

When information is available on a tablet, do you think the tablet will be used during the match?

No don't think so, Maybe mid game or when there is an injury

If you will be able to see the player performance data on your screen, do you think it would increase your fan experience?

It would be fun!



Interview Joep:

What is your age?

26

What is your gender?

Man

Do you live in or outside Eindhoven?

Outside

Do you travel to the stadium by public transport?

Nee, eigen vervoer, auto

In your opinion, what do you think fan experience is?

Naar mijn mening is het alleen maar beter geworden. Er is meer sfeer in het stadion als van tevoren. Ik geef het een 9/10

Do you know the Brainport Experience Box in the PSV Stadium?

Nee, ken ik niet

What additional information would you like to see during a match? (at the moment we have weather data, public transport data, Player data, Food & Beverage, Social media data, Sound data)

Op het moment weet ik geen aanvulling, zodra er iets in me op komt geef ik het nog door. Ik neem aan dat er prijzen bij staan bij Food & Beverage. Misschien kun je op de tablet aangeven of het druk is om drank te halen of op het moment rustig.

When information is available on a tablet, do you think the tablet will be used during the match?

Ik persoonlijk zal er niet veel gebruik van maken.

If you will be able to see the player performance data on your screen, do you think it would increase your fan experience?

Ik denk het wel, Ik ben altijd benieuwd naar statistieken.



Interview psv fan experience Bart

1. *What is your age?*

46

2. *What is your gender?*

Man

3. *Do you live in or outside Eindhoven?*

Inside

4. *Do you travel to the stadium by public transport?*

I travel by bike

5. *In your opinion, what do you think fan experience is?*

Having fun with your friends looking at sports. Seeing my friends happens less often than previously, so seeing them at a match becomes more important. It is not only the game itself, but also going with them in the city centre, grabbing a bite before the game etc. It is more than what happens on the field.

6. *Do you know the Brainport Experience Box in the PSV Stadium?*

My personal opinion is that it would distract me from my friends and the game. If I would go to the experience box with my friends, I don't think I would use the tablet because it would be too much distraction.

7. *What additional information would you like to see during a match? (at the moment we have weather data, public transport data, Player data, Food & Beverage, Social media data, Sound data)*

I would like to see the food & beverage data so that I know if it is crowded or not. I also would like a replay button; in that case I can replay a goal if I miss it. I don't think I would look into the player data, but I don't know what I miss if it's not there. Same goes for the social media data.

I think that I would use the in-seat ordering if I can order it on my phone. It is a better idea than a boy walking with a beer backpack.

8. *When information is available on a tablet, do you think the tablet will be used during the match?*

I don't think I want that. This is a really personal opinion because I don't want to be distracted. If I am in a social environment, I don't want any distractions from either my phone or a tablet.

9. *If you will be able to see the player performance data on your screen, do you think it would increase your fan experience?*

I think that it is possible, to the extent that if we have the data available, it would be a nice discussion to have before the match starts. During the game I would not use it. The data would improve my fan experience previous to the game, but not during the game.

9.2. SURVEY RESULTS PSV

A survey was set out under visitors of the Brainport Experience Box. Seventeen responses were received which is a response rate of 43,6%. This means a reliability level of 87% and a margin of error of 14%.

For the analysis of the answers, Python is used in combination with Jupiter Notebook. Since the survey was sent to the respondents in Dutch, some of the answer option will be displayed in Dutch. First of all, the distribution of the respondents' ages has been described. This is showed in Figure 24. There can be seen that almost half of the respondents is twenty years or younger.

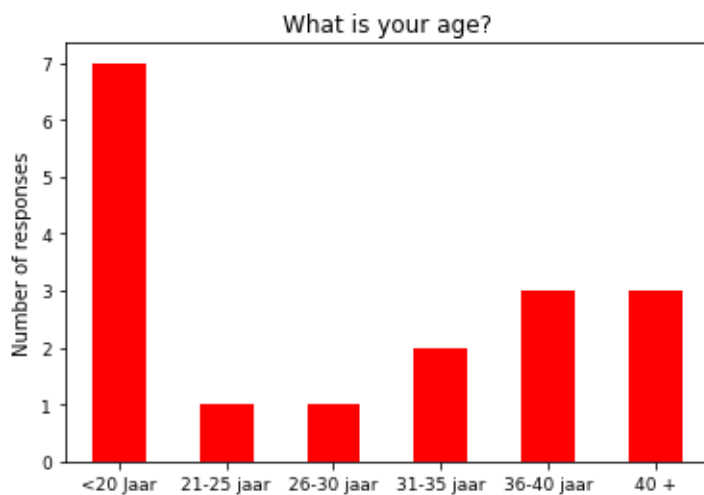


Figure 24 age of respondents

Next, the gender distribution is shown in Figure 25. More than 70% of the respondents were female.

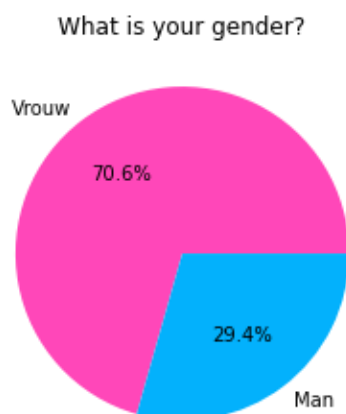


Figure 25 gender of respondents



Furthermore, the rates respondents give to watching the game have been analyzed. There has been looked into the rates watching on tv, in the stadium and into the experience box. If a respondent never watched on tv or in the stadium they did not rate the experience. As shown in Figure 26, the stadium experience has the highest mean score, after which the Experience Box scores the second highest. Watching the game on tv scores the lowest. From this numbers, the following remarks could be made:

- 1) The highest rating has "Match at stadium" with mean value 9.13. "Match at PSV Experience Box" has lower mean value - 8.71. And "Match on TV" has the lowest mean rating - 8.06.
- 2) Fans will probably prefer watching match at stadium with or without PSV Experience Box rather than watching match on TV.
- 3) However, the lowest given rate (6) got "Match at PSV Experience Box". For other 2 groups the lowest given rate is 7. Moreover, rate 7 is recognized as outlier for "Match at stadium".
- 4) 50% of respondents' rates for "Match at stadium" are in range from 9 to 10 but it also worth mentioning, that "Match at stadium" is missing 2 values, so the total respond for this column is 15.

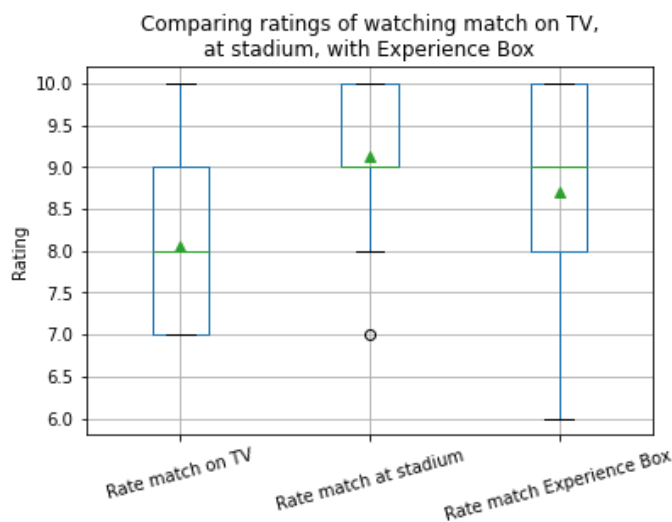


Figure 26 mean experience ratings

There was also looked if the expectations from the experience box were met. As shown in Figure 27, for 82.4% of the respondents the expectations were met. For 17.6% it was not met. There were a variety of reasons why this was the case.



Figure 27 expectations Experience Box



The last question that was asked was what visitors of the Experience Box would like to see. There were six options given and the respondents could rank them one till six. Number one got the most points and number six the least. In Figure 10 the overall scores are visualized. Data that would be appreciated is the player and social media data. The respondents valued the weather and traffic- and public transport data the least.

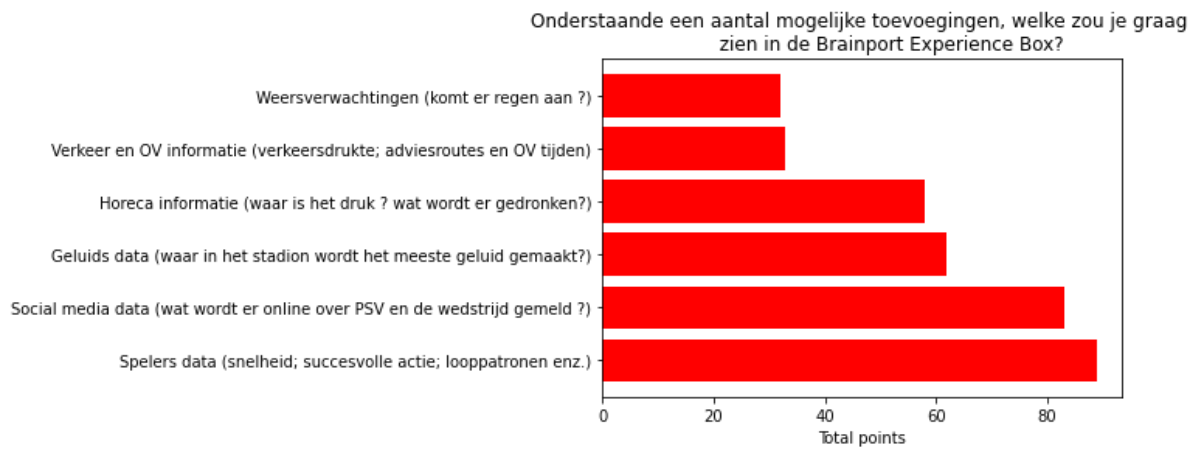


Figure 28 additions Experience Box



9.3. DATA SOURCES

PSV holds different types of data sources. These current existing data sources within PSV are explained in the following paragraphs. This is important to get a better understanding of the data that is accessible.

9.3.1. ORTEC

Ortec sports is a company that provides information about football players and whole teams that helps these teams improve their performance, explore the statistics of their opponents, monitor their own statistics and rank players based on a requested playing style or philosophy. Ortec is using/working with Effectiveness in Action software (EiA) that gains the best performance analysis of players – collecting data on the positions, actions and effectiveness using the EIA Video Analysis System. Detailed reports are generated from the data on their online Pro-Portal - a unique combination of the actions of the players in a tactic tree and the mathematical overlay. The goal/mission of Ortec is to contribute to further improving the sporting results and experiences of the teams they work with (Ortec sports, 2020).

The following data is available:

The table below shows the data that is from Ortec. There are four data sources with this format. The table describes which data is analyzed. The four data sources are regarding player and keeper data. Ortec also keeps track of match data, which will be explained later.

- AwayKeeperStatistics
- AwayPlayerStatistics
- HomeKeeperStatistics
- HomePlayerStatistics

Table 1 Ortec data players

Ordinal position	Column name	Type	Is_nullable	Description	Necessary	Note
1	PersonID	int	NO	ID of player/keeper	YES	
2	StatisticID	int	NO	ID of source	YES	
3	StatisticName	nvarchar	YES	Name of statistical subject	YES	
4	StatisticValue	int	YES	Value/score of statistical subject	YES	

The table below shows the data from a match. The columns describe the data. There are 2 sources with this format

- AwayTeamStatsTotal
- HomeTeamStatsTotal

Table 2 Ortec data match

Ordinal position	Column name	Type	Is_nullable	Description	Necessary	Note
1	StatisticID	int	NO	ID of source	YES	
2	StatisticName	nvarchar	YES	Name of statistical subject	YES	
3	StatisticValue	int	YES	Value/score of statistical subject	YES	

So, we can conclude that Ortec creates software for analyzing sports performance. During matches, each player and team's action is video-linked and given an objective score based on coaching criteria.



As a result, customers may take athlete tracking and training, as well as talent identification and development.

9.3.2. SORAMA

Sorama is an internationally successful high-tech company from Eindhoven that specializes in visualizing, interpreting and making sound functionally applicable. PSV will integrate various forms of this in the Philips Stadium. With this, the club wants to make existing systems in the field of light, sound and CCTV smarter and improve the guest and fan experience during matches and events. “We are very happy with this collaboration”, says Frans Janssen, Commercial Director of PSV. “Its potential is great and fits within our vision of the stadium of the future. We want to lead the way with products that promote the football experience. Sorama’s applications do that. This collaboration fits in perfectly with our vision of continuing to innovate,” says Janssen.

Sorama’s L642 platform is a next-generation acoustic monitoring sensor designed for integration into smart cities to help understand sources of sound and alleviate noise pollution – which can have a damaging effect on humans. Offering scalability and easy integration, the sensor platform aims to monitor and learn where, when and how noise is created. The onboard software and AI can even trigger anomalies and classify specific sounds. Live sound events can be projected over a map using the Sorama Sound Surface, giving users more insight into sound.

File AllUnitscsv		
Column	Description	Example
unit	Serial number of unit	100010002
x	X coordinate of unit, mm	-51681
y	Y coordinate of unit, mm	-67710
z	Z coordinate of unit, mm	25000
cableExitSide	Orientation of unit (East, West, North, South)	E
normalx	Normal vector (direction of unit)	0
normaly		0
normalz		-1
Folder beamformPoints		
File for every unit - json format		
Attribute	Meaning	
x	Distance between unit and beamformPoints, metres	
y		
z		
File with measurements - json format		
Attribute	Meaning	
soundSurfaceld	ID of direction	
timestamp	time incl. milliseconds	
values	db value for every beamformpoint of the corresponding unit	

Table 4 Sorama data description

Solutions

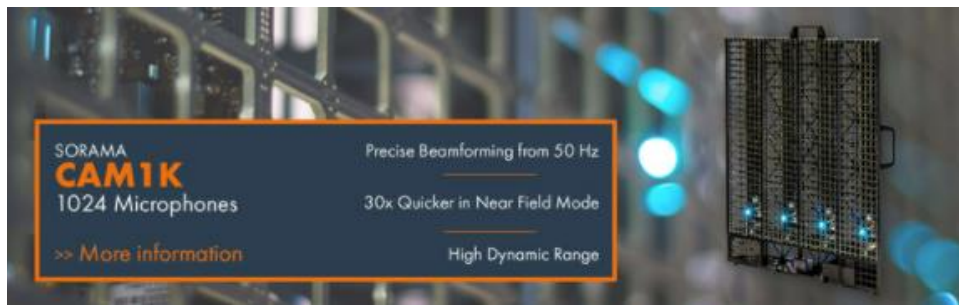


Fig1. Cam1K



Fig2. Cam64



Fig3. L642



Fig4. ScanA

References:

<https://bits-chips.nl/artikel/sorama-assists-psv-with-the-goal-of-enhancing-fan-experience/>

9.3.3. FOOD & BEVERAGE

The number of factors that may have impact on fan experience are varying. Among the most important elements of fan experience is Food & Beverage services. Fans mostly face long lines and poor/slow serviceduring the match. In other words, this field has quiet big scope for improvements.

To analyse the current situation on a PSV stadium and to find out the possible solutions for improving fan experience in Food & Beverage area the data are needed. PSV provided historical data of total sales, sales per hour and aslo data about the most popular positions.

However, provided data is not so informative for fans as it is mostly related to sales. The most relevant information fro fans could be part with “Sales per hour” as capacity of stand could be extracted from it. Also 1-hour range is probably not suitable because football match lasts approximately 90 minutes. Therefore, this range could be adjusted to smaller one, e.g. 15 or 30 minutes.

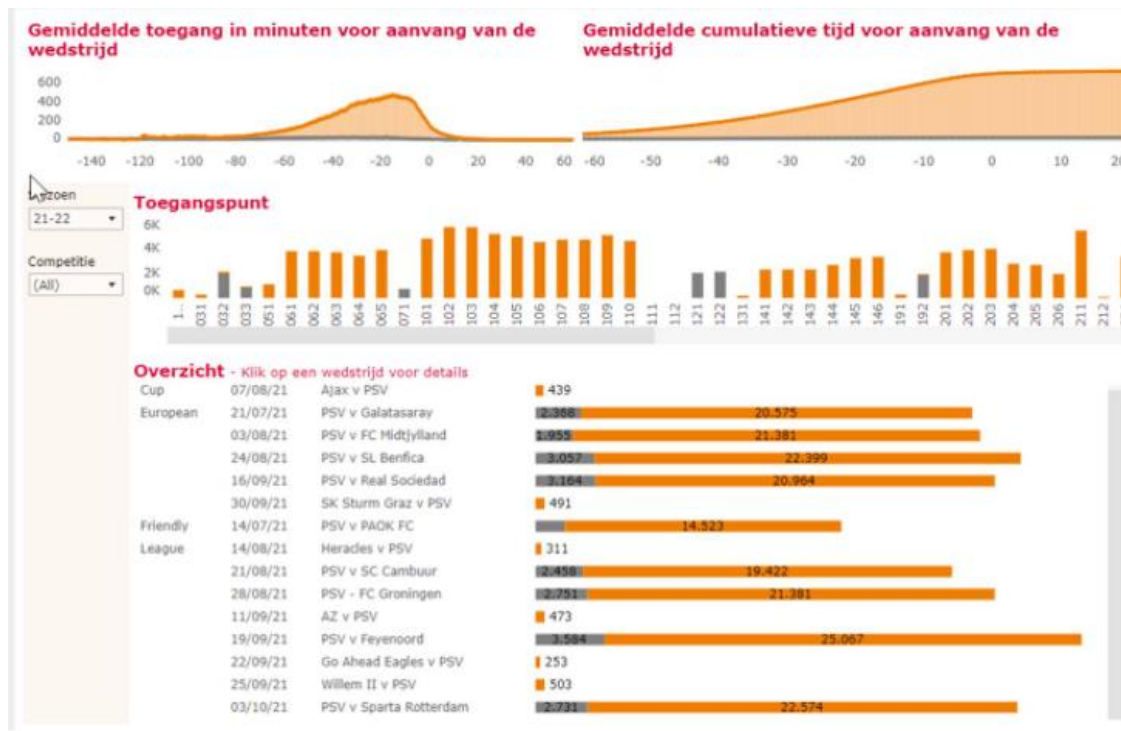
Ordinal position	Column name	Type	Is_nullable	Description	Necessary	Note
1	TransactionGuld	nvarchar	YES	ID of transaction	NO	
2	SourceID	smallint	YES	ID of intern source	NO	
3	OrderID	int	NO	ID of order	NO	
4	TransactionID	time	YES	ID of transaction (in format of time h:m:s.ms)	NO	
5	ProductID	int	YES	ID of product	YES	
6	TransactionPurchaseDate	time	YES	Date in format d-m-y h-m-s ms	YES	
7	TransactionChannel	nvarchar	YES	e.g. TopEvents	NO	
8	TransactionLocation	nvarchar	YES	ID (name) of counter	YES	
9	OrderValue	int	YES	Total order price	NO	Sales info, irrelevant for fans
10	TransactionPrice	int	YES	Total transaction price	NO	
11	ProductPrice	int	YES	Price per unit	NO	
12	ProductUnits	tinyint	YES	Number of units	NO	
13	CardNumber	float	YES	Card number if customer has one	NO	Same values
14	GroupID	int	YES	Grouping by cards	NO	
15	PurchaserGroupID	int	YES		NO	
16	ProductName	nvarchar	YES	Full product name	YES	
17	ProductDescription2	nvarchar	YES	NULL values only	NO	
18	ProductDescription1	nvarchar	YES	Same as ProductName	NO	
19	ProductCategory	nvarchar	YES	Category of product: food, alcohol, non-alcohol	YES	
20	Transday	date	YES	Day of transaction (without time)	YES	

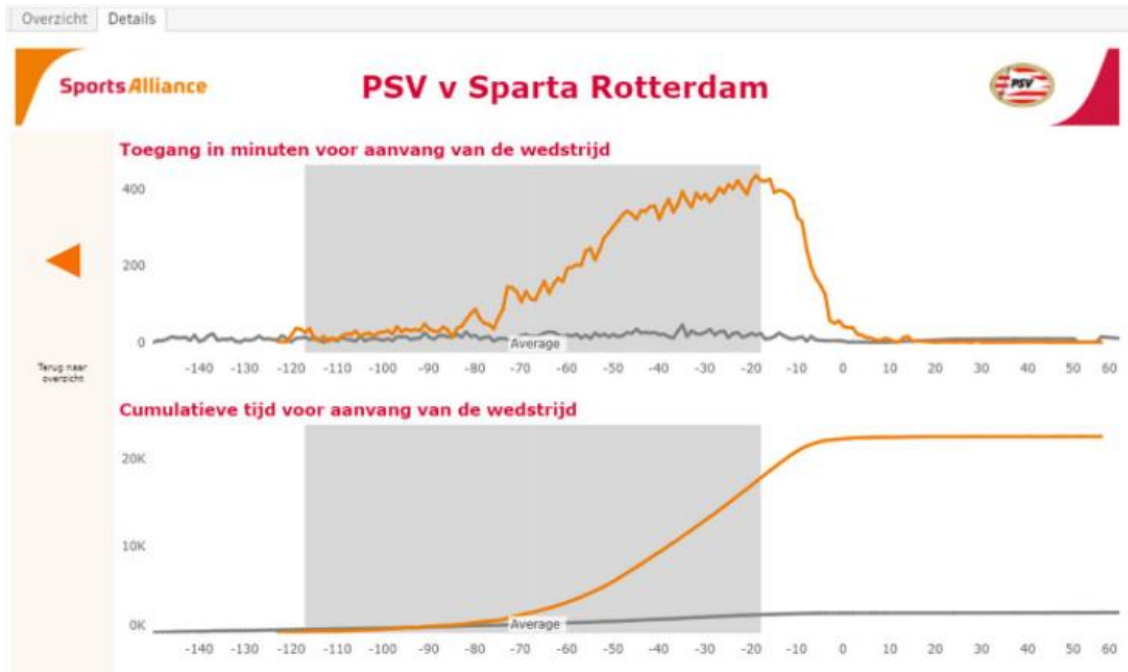
Table 5 Food & Beverage data description



Ordinal position	Column name	Type	Is_nullable	Description	Necessary	Note
1	SourceID	int	YES	ID of source	YES	
2	TransactionID	float	NO	ID of transaction	NO	
3	TransactionHistoryID	nvarchar	YES	NULL values	NO	
4	OccasionID	nvarchar	YES	ID of occasion	NO	
5	BeneficiaryID	nvarchar	YES	ID of Beneficiary	NO	personal info
6	CardNumber	float	YES	Number of card	NO	personal info
7	AccessDateTime	datetime2	NO	Time of entrance in format Y-M-D h-m-s	YES	
8	PointofAccess	nvarchar	NO	ID of entrance point	YES	
9	GroupID	nvarchar	YES	ID of Group	NO	
10	OccasionSeason	nvarchar	NO	Number of season (e.g. 19-20)	NO	
11	OccasionStartDate	datetime2	NO	Start of season date	NO	
12	OccasionCompetition	nvarchar	NO	E.g. League	NO	

Table 5 Entrance data description





9.3.5. SOCIAL MEDIA DATA

In the social media dashboard for PSV are a couple of visuals shown. In this part of research Datastic is describing which kind of visuals are shown and how they can be used to show in the Brainport Experience Box. Please keep in mind that every visual is based on a period of 24 hours.

After this, Datastic received an API where it is possible to extract data from Coosto Social media data. Datastic managed to extract data with API and python. The data will be shown in the tables underneath. The columns will make clear how the data is organized and looks like.

Sources:

Underneath the source data is shown. This data can give an insight on which social media platform is talked about the most about PSV.



Tabel 3 social media sources

Ordinal position	Column name	Type	Is_nullable	Description	Necessary	Note
1	id	int	NO	ID of sentiment	YES	
2	Name	nvarchar	YES	name of the sentiment	YES	
3	Frequency	int	YES	number of posts	YES	
4	Overall_sentiment	int	YES	average of negative, positive, neutral sentiment	NO	
5	Positive_sentiment	int	YES	number of positive posts	NO	
6	Negative_sentiment	int	YES	number of negative posts	NO	
7	Neutral_sentiment	int	YES	number of neutral posts	NO	

Sentiment:

Underneath the sentimental data is shown. How many positive messages are written and how many negative. This could be very useful to see during the match. Maybe the viewers at home experience a different match then the people within the stadium.

Tabel 4 social media sentiment

Ordinal position	Column name	Type	Is_nullable	Description	Necessary	Note
1	id	int	NO	ID of sentiment	YES	
2	Date	date	YES	date of the sentiment	YES	
3	Frequency	int	YES	number of posts	YES	
4	Overall_sentiment	int	YES	average of negative, positive, neutral sentiment	NO	
5	Positive_sentiment	int	YES	number of positive posts	YES	
6	Negative_sentiment	int	YES	number of negative posts	YES	
7	Neutral_sentiment	int	YES	number of neutral posts	NO	

Trending Topics:

Trending topics can be shown during the match to visitors of the box, the visitors could see if they missed something out of these topics.

Tabel 5 social media trending

Ordinal position	Column name	Type	Is_nullable	Description	Necessary	Note
1	id	int	NO	ID of source	YES	
2	Topic	nvarchar	YES	topic of trending name	YES	
3	Score	int	YES	number of time the name is mentioned	YES	

Authors:

The author data can give insight in who is talking about PSV on social media platforms.



Tabel 6 social media author

Ordinal position	Column name	Type	Is_nullable	Description	Necessary	Note
1	id	int	NO	ID of source	YES	
2	Freq	int	YES	Number of posts	YES	
3	Sent	int	YES	average of negative, positive, neutral sentiment	NO	
4	SentP	int	YES	number of positive posts	NO	
5	SentN	int	YES	number of negative posts	NO	
6	Sent0	int	YES	number of neutral posts	NO	
7	Influence	float	YES	score of influence author has	YES	
8	Gender	nvarchar	YES	gender of author	NO	
9	Followers	int	YES	number of followers author	NO	
10	Reactions	int	YES	number of reactions on post	NO	

Sources:

<https://www.oracle.com/industries/food-beverage/sports-entertainment/>

<https://www.forbes.com/sites/sap/2013/04/10/sports-and-analytics-fan-experience-matters/?sh=4811845d5fd3>

<https://rapidretail.co.uk/talking-about-fan-experience/>

<https://www.oracle.com/a/ocom/docs/dc/em/fan-experience-report-fb.pdf>

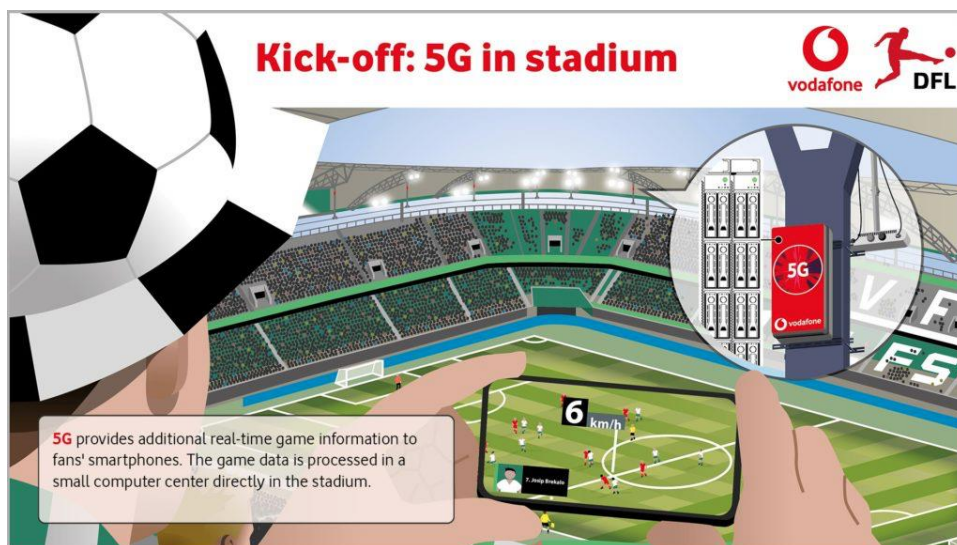
FC Wolfsburg Project

Source: <https://www.dfl.de/en/news/kick-off-for-5g-latest-mobile-communications-technology-activated-in-the-first-bundesliga-stadium/>

FC Wolfsburg from Germany is trying to improve fan experience of visiting stadium by implementing modern technologies, such as 5G. In partnership with Vodafone, the DFL Deutsche Fußball Liga and French software company, Wolfsburg introduced mobile app in September 2019.

This real-time app provides match statistics, players' individual values on fans' mobile phones. The app processes the information in visual form, producing live graphics, stats and analysis that was before only in TV match translations. AR technology which is also included in app allows to display statistics of the particular player by directing smartphone on a field.

A lot of data is being collected during the match, so 5G technology is responsible for information processing. 5G devices are installed directly in the stadium. This makes possible to collect all the data and to deliver it to fans' mobile phones in real time. For the 2019/20 season, Wolfsburg's Volkswagen Arena was the first to have a 5G network.



Basketball and the adoption of data

Like every other professional sport data analytics started playing a huge role on the basketball court. There are multiple approaches on how to implement technology into making the experience of watching an NBA/WNBA game, but everyone agrees that data is too important nowadays and is trying to gather as much of it as possible for future usage.

Recently, the adoption of cameras at training facilities and game court arenas has seen a big increase and most of the bigger NBA teams have systems that can record every movement of both the ball and all 10



players 25 times per second. Data collected about the players can be used in endless ways, most popular being: rest of players(analysing performance and fatigue factors), drafting players, scoring analysis and matchups.

An innovative idea came from Kirk Goldsberry to create a heat map of the court that shows the accuracy of shooting three pointers that could show and help players improve their shooting performance while giving fans insight into the current state of a player/team.

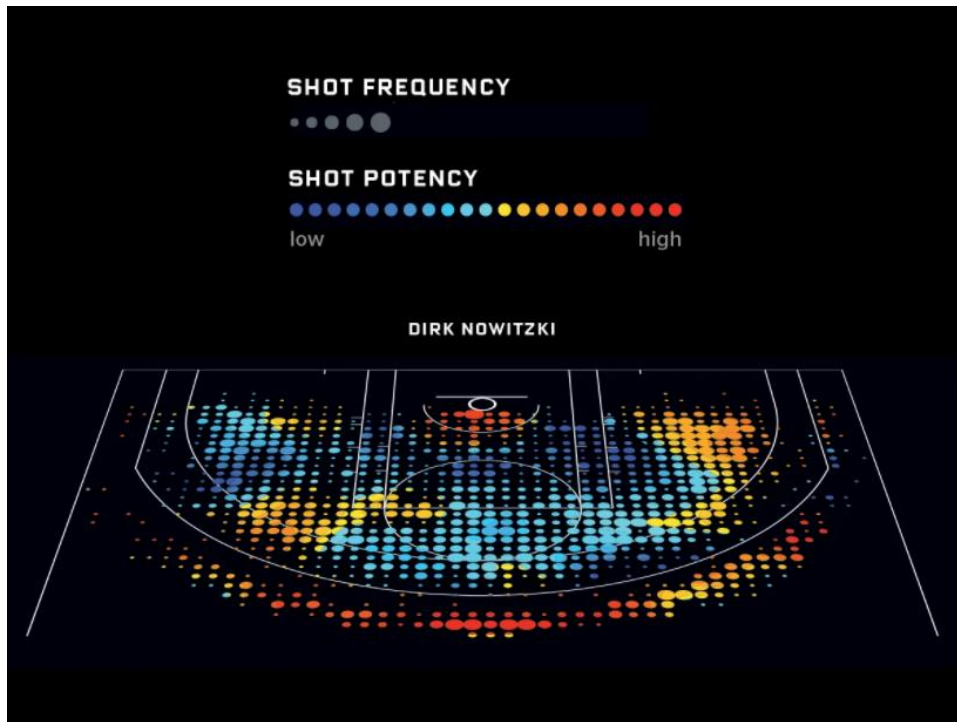


Figure 1. Dirk Nowitzki heatmap

Figure 2. Ray Allen heatmap

AR/VR and Big Data have also seen a massive increase in adoption although they have yet to be used publicly by fans. This is because the projects are still in their infancy. Two of the more developed projects are SPORTSXR and CourtVision.

By incorporating advanced sensing technology, computer vision, machine learning, and data visualization, SportsXR and CourtVision are taking sports decision-making to the next level.

Currently, both projects are prototypes that are being tested on phone/tablets applications.



Figure 3. SportsXR



Figure 4. SportsXR



Figure 5. CourtVision

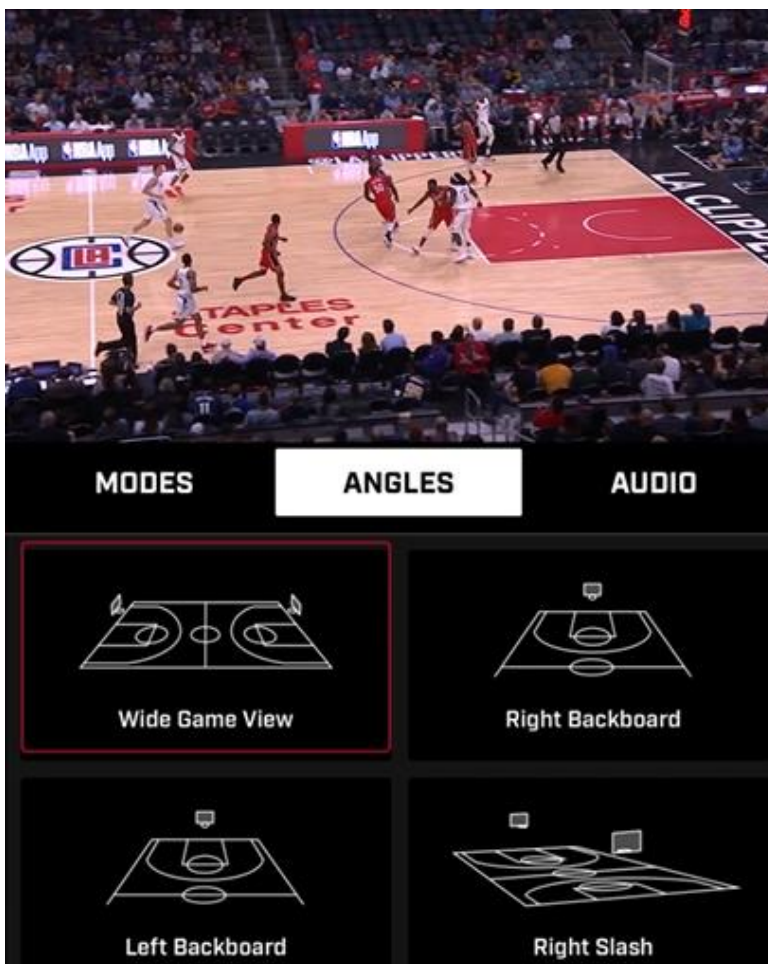


Figure 6. CourtVision

CourtVision also provides interesting facts about the players during the games that appear as little notifications on the screens people watch the game on. This could be very useful in our case since notifications like this could appear on the PSV tablets.

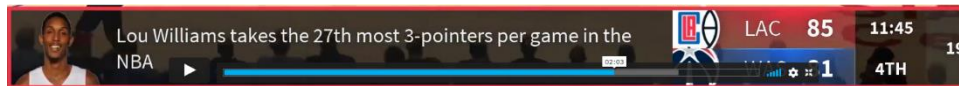


Figure 7. CourtVision

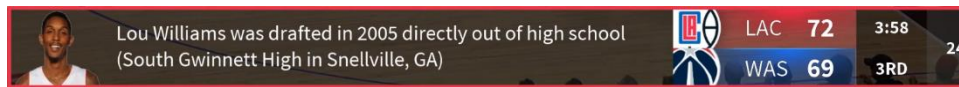


Figure 8. CourtVision

Both projects have instant replay functionalities and possibilities for sharing clips via Facebook/Twitter or other social platforms.

Sources:

<https://onlinedsa.merrimack.edu/nba-analytics-changing-basketball/>

<https://www.wired.com/2014/10/faster-higher-stronger/>

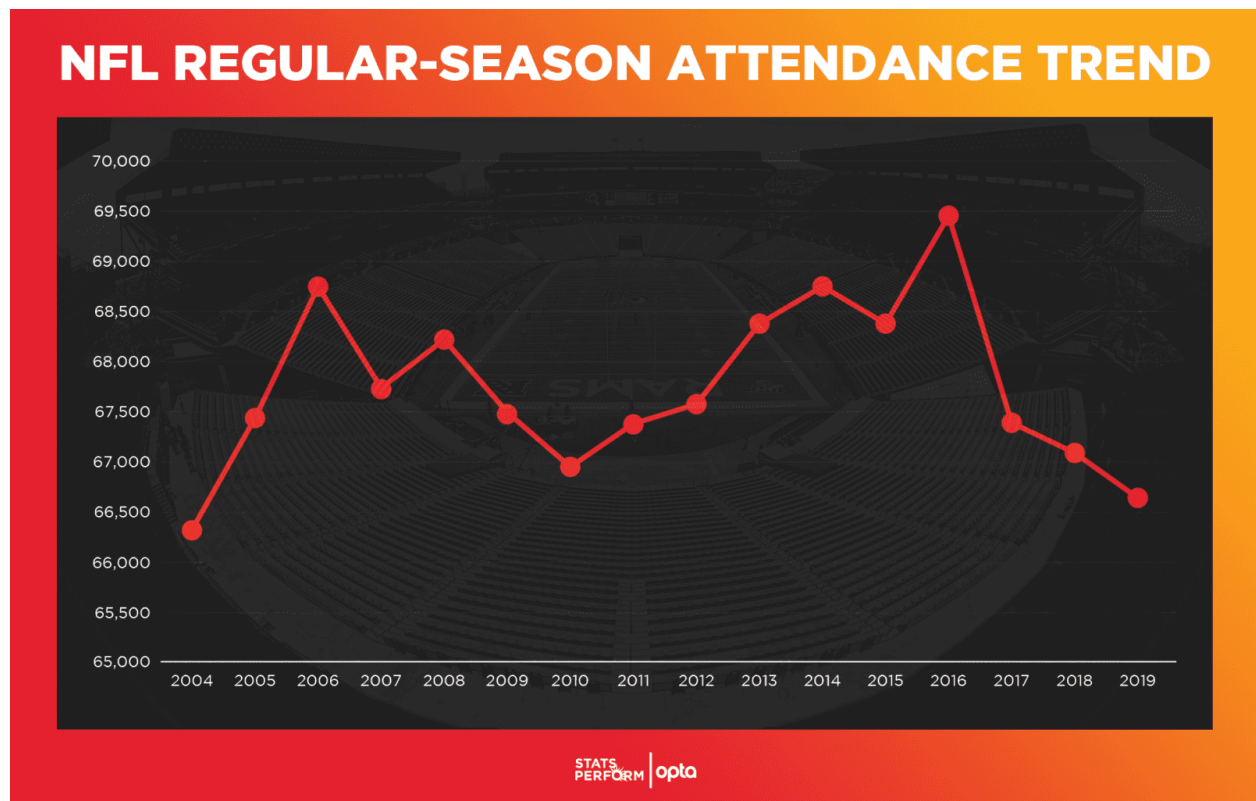
<https://cmoinsider.pl/en/vr-and-ar-big-data-on-nba-floors/>

<https://digital.hbs.edu/data-and-analysis/bringing-the-ar-cloud-to-sports/>



DATA CAN TRANSFORM THE FAN EXPERIENCE WITHIN STADIUMS

For the whole home game attendance for professional sports like ice hockey, baseball, basketball and soccer, the attendance declined for more than 10% over the past 10 years. Underneath in the image a trend within the NFL is shown:

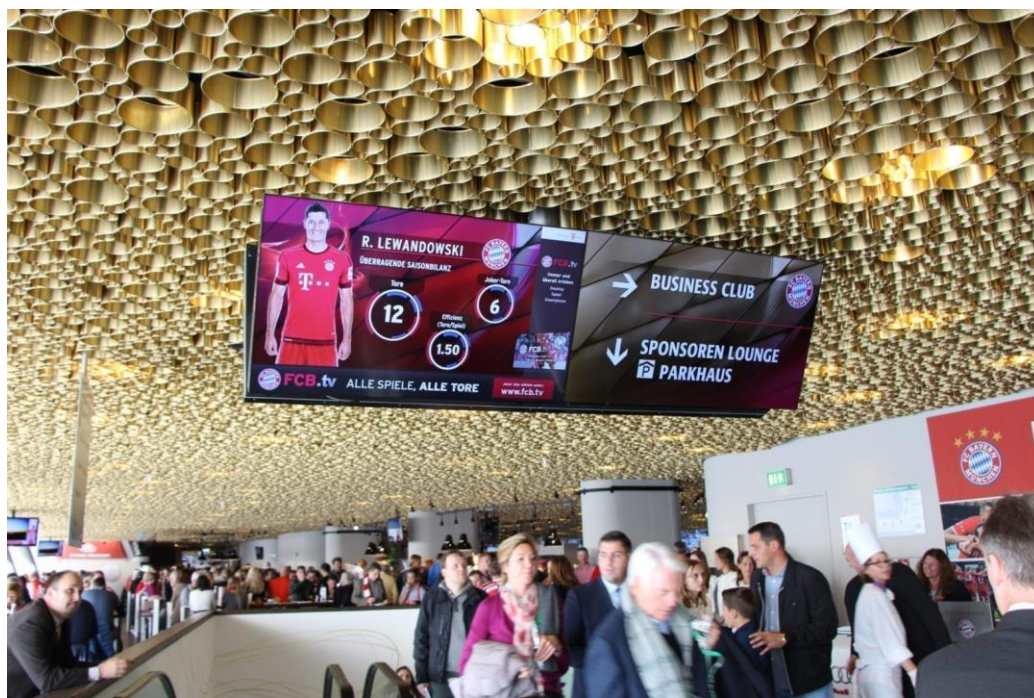


There is a declining line since 2016. What could be the reason for this? It all has to do with fan engagement. Nowadays watching a sports match within the stadium, the in-person fan experience doesn't compare favourably to watching a game sitting in you living room. People can watch any game from anywhere and like the comfort of their own home, eating own food & drinks.

The fans currently watching the games at home on a spectacular, high-definition big screen. The use of second screening is rife, 80% of sports viewers say they use a computer or smartphone while watching live sports on TV. These devices are used to search for players stats, visualizations, live scores and social media.

This means that fans require better reasons to leave their comfort home. Research proved that 45% of the premium fans (who often are in the stadium or have a season pass) would pay more money for a better in-person experience.

As example, at Bayern Munich, they were looking to modernize their Allianz Arena football stadium to engage fans trough a more connected stadium experience. They added 1200 screen around the stadium with match content that engaged fans through real-time match statistics (see image below).



Source: <https://www.statsperform.com/resource/how-data-can-transform-the-fan-experience-in-sports-stadiums/>

9.3.8 INTERVIEW WITH ROY SMITS

What is the environment of the data warehouse?

Until a couple of years ago, we did not have a warehouse. We received a data warehouse solution from Sport Alliance. Data was less important for us then now. It was a low level of data usage.

When Yoeri came to PSV we started to use more and more data from different sources. At that point we really started to use the data warehouse where it was meant for.

After a while we wanted more out of the data and started to use it for presenting with for example Tableau. We faced some problems here while it was difficult to add more data or tables to the Sport Alliance server. We couldn't present the data we wanted on Tableau due to these problems.

We created an own data warehouse where once a day the data from the sport Alliance warehouse is dumped into. For this warehouse we use an SQL-server in our own private cloud environment of C-mac. You can easily maintain and update that server. We are looking to maybe migrate it to Microsoft Azure.

Do you have an architecture picture?

Currently, we are making a daily dump from the Sport Alliance warehouse to our own warehouse, so there are no direct applications connected to our warehouse. They are all connected to the Sport Alliance.



In the picture, the yellow arrows are all the transactional- and personal data from Dynamics. This will be put into the Sport Alliance warehouse. They are making sure that the data is going into the warehouse and is presented correctly. An advantage of this is that we don't have to do that, we don't have to hire people etc. for this. A disadvantage is that we cannot make extra tables in their warehouse. That is why we import it to our own data warehouse. The link between Sport Alliance and our own is made automatically.

To test if it is better to use something in our own SQL-database I think it is best to test in a self-created database.

How do they use the data warehouse now?

That is a difficult question. The data in this DWH is stored for a lifetime. That is not complied with the privacy. In our case it is sustainable to keep it that way.

However, if we look at player data, it is too big and much to store it all, so we chose not to store all the data. We do not have the capacity to store everything. The time framing here is different.

We do not have any reason right now to delete any data from our customers.

What is the time schedule of the incoming data?

The incoming data in our own DWH is once a day. For the SA DWH it depends on the different application.

How often should it be refreshed?

For us it is refreshed automatically. For SA they also prefer to refresh it automatically. If this really isn't possible, they, do it manually.

Are we allowed to add new tables to the data warehouse?

It is not possible in the sport Alliance, but it is possible in our own warehouse.

For now, it is not possible to add any tables. If you want to test it, you should test it in your own environment. But of course, if it is necessary in a later state, it may be possible to do it in our own DWH.

How can we get access the data? Can that be arranged asap?

I will give you access to some of the data. I will ask yoei and when I have it I will let you know.

For Sorama, we are setting up a VPN-connection. Maybe there is also some test data available.

PSV Eindhoven



DATASTIC

9.3. PROJECT PLAN





PROJECT PLAN PSV

Company	PSV
Company supervisor	
School	Fontys ICT Hogeschool Minor Data Driven Business Lab
Content coach	
Process coach	Mr. Rob Verhoeven
Students	Gergana Agorastva Remco Bisschops Femke Boogerd Nikita Gavrilov Yaniek Martens Lieke Nijs Victor Plesciuc Ilya Tsakunov



LIST OF TERMS

Term	Explanation
PoC	Proof of Concept
Ict research methods	This method provides support for structuring research activities for each research question



MANAGEMENT SUMMARY

OVERVIEW

The company PSV Eindhoven are a Dutch sports club that are in the top division. As innovators, PSV is looking into creating a much more interactive environment for spectators to watch their games. The technologies used for this to happen can range from Virtual Realities, Augmented Realities And even phone/tablet applications with Live-replay functionalities.

Datastic is a company formed by nine Fontys students with different educational backgrounds. Alongside mentorship provided by Fontys and PSV, Datastic will provide research into ways of using available data sources related to PSV in order to improve the viewer's experience of a game, what technologies could be used and how to Improve the already existing PSV application.

PROBLEM

Current problem – information offered in the experience box now is limited.

- The information to which stadium spectators currently have access is too limited, which affects people's experience and involvement with the game.

Potential problems

- Finding live data sources
- Data management

SOLUTION

Datastic company is supposed to collect, clear, aggregate the data from different sources for further usage in application.

HIGHLIGHTS

- Different sources of information were provided
- Datastic will only deliver a proof of concept (PoC) of application and advice about this
- Datastic will be collaborating with another team responsible for frontend part of application
- Target audience: football fans, PSV clients, visitors of PSV Stadium
- Responsibility of Datastic: collecting, clearing and aggregating the data
- Team consists of 9 developers + process coach
- Communication mostly via MS Teams, weekly meetings on Mondays
- Deadline: January 15th

KEYS TO SUCCESS

- Regular progress checks
- Stakeholders are committed
- Timeline for project
- Risks are being mitigated
- Scope is realistic



Table of contents

Document history	2
1. Introduction	5
2. Data sources	6
2.1. <i>What data sources could be useful to enrich the fan experience based on the requirements?</i>	6
2.1.1. Data analytics	6
2.1.2. Literature study	7
2.1.3. Best Good and Bad practices.....	8
2.1.4. Survey.....	10
2.2. <i>conclusion</i>	11
3. User requirements	13
3.1. <i>How will the available data be used to enrich the fan experience of the stadium?</i>	13
3.1.1. Interview	13
3.1.2. Literature study	13
3.1.3. Business case exploration.....	15
3.1.4. Requirement prioritization	16
3.2. <i>conclusion</i>	17
4. Combining data	18
4.1. <i>How can the available data sources be combined?</i>	18
4.1.1. Literature Study.....	18
4.1.2. Community research	19
4.1.3. IT architecture Sketching.....	20
4.2. <i>conclusion</i>	21
5. Connecting data TO UI	22
5.1. <i>How can the combined data be connected to the interface of the application?</i>	22
5.1.1. Document analysis	22
5.1.2. Expert interview	23
5.1.3. Community research	24
5.2. <i>conclusion</i>	25
6. conclusion	26
7. Advise & Recommendations	27
8. Bibliography	Fout! Bladwijzer niet gedefinieerd.



9. Appendix	30
9.1. <i>What is Fan Experience</i>	30
8.1.1 Literature study.....	Fout! Bladwijzer niet gedefinieerd.
8.1.2 Interviews	30
9.2. <i>Data sources</i>	39
9.2.1. Ortec.....	39
9.2.2. Sorama	40
9.2.3. Food & Beverage	42
9.2.4. Entrance data.....	43
9.2.5. Social media data.....	45
9.3. <i>Project plan</i>	55



Introduction

In the next five months, Datistic will help PSV improve the fan experience in the Brainport Experience box. This experience box will investigate how data can be used to improve the experience of a football match for the fans.

In this project plan, PSV will first be described as a football team and organization. After that, the definition of the project is explained where the project goal, scope, research question and plan of action are mentioned. This will lead to the expected deliverables. After this, there will be zoomed in on the project management. Here, the communication, possible risks and planning are described.



Context and background

1.1.

The company PSV Eindhoven is established in 1913. The term PSV means Philips Sport Vereniging. A couple of Philips employees founded the club. They started as a Dutch sports club from Eindhoven and grown to one of the biggest sport clubs in the Netherlands. Their football team is playing in the Eredivisie. This is the top tier division of the Netherlands. Its history contains two golden eras revolving around the UEFA Cup victory in 1978 and the European Cup in 1987-88. The team has won the Eredivisie 24 times and the KNVB Cup nine times (PSV Eindhoven, 2021).

The stadium of PSV is in the city center of Eindhoven, which also contains the offices. They also have a training facility named “De Herdgang”. Throughout the years PSV has become more than only a football club. It became a company. They earn their revenue by selling tickets for matches at the stadium, having sponsor deals, developing young players, and selling them (PSV Eindhoven, 2021).

This project will take place at the marketing and communication department of PSV. They are currently creating a Brainport Experience Box within the stadium where supporters can experience the match by a different way. Brainport is an innovative eco-system in Brabant. With a strong technical manufacturing industry and a unique cooperation between companies, it bundles all those knowledges to create a strong organization (Brainport Eindhoven, 2021).

They will get a tablet during the match where a lot of information will be displayed like player stats, stadium noise data, public transport information and more. The purpose of the Brainport Experience Box is to use the strength of the ecosystem Brainport Eindhoven to develop and test ideas and technologies to come to new smart stadium fan experiences and fan engagement solutions. Example given: concepts of mobility, availability, crowd management, liability & sustainability but also safety and digitization (Brainport Experience Box , sd).

2.2 Problem Analysis

The issue that PSV Eindhoven is now facing is that the data collected comes from a variety of sources. The customer needs all the data to be combined, cleaned, and prepared so that it can be presented in the app more easily.

Some visitors now have access to tablets, which only display a restricted amount of information for the time being, limiting people's experience and involvement with the game.

The concept is that visitors will have access to tablets on which they will be able to utilize an app that displays various sorts of information that have already been discussed in the preceding paragraphs.

The primary problem that could arise throughout the project is a difficulty with data management - in terms of time, preparing the data, converting it to the correct format, and so on. The data that will be received will come from a variety of sources, making it difficult to incorporate it into a single app.

Summarized, the current problem is that information offered in the experience box now is limited.

2.3. Project definition

In chapter three, the project will be defined. The project goal will be stated in 3.1, where after the intended project results and the scope will be described.

2.4. Project goal



The upcoming months until January, a proof of concept for PSV will be created. This concept contains all aspects on how to gather, clean and combine existing and new data for future usage within the PSV Experience Box.

2.5. Intended project result

At the end of the project, the goal is to deliver a proof of concept with an advice on how to improve the visitors experience in the Brainport Experience Box. In order to do this, data will be researched to how and what data can make this improvement happen. After that, the data will be collected and cleaned whereafter it can be used to investigate and improve the experience. The ultimate goal is to create an application for visitors of the Brainport Experience Box. This will be done by future student teams.

2.6. Scope

A scope will be defined to make clear what will and will not take place during the project.

- Datastic will not deliver a fully working application, we will only deliver a proof of concept and advice about this.
- The target audience will be the client of PSV and the visitors from the Brainport Experience Box at the PSV stadium.
- The project group will focus on the departments of data, digital marketing (not for us but important for the project)
- During the project, data will be collected, look at useful data to add to existing data, cleaning data and combine the data.
- The project has to be completed during the semester from September until January.

2.7. research questions

Research question

How can existing data be gathered, cleaned, and combined with new data based on the user requirements within the PSV Experience Box to enrich the fan experience?

Sub questions

1. What data could be useful to enrich the fan experience based on the user requirements?
2. How will the available data be used to enrich the fan experience of the stadium?
3. How can the available data sources be combined in a way that it is more manageable?
4. How can the combined data be connected to the interface of the application?



2.8. Plan of action

The table below shows the approach used during the research. The approach is based on the ict research methods. This method provides support for structuring research activities for each research question.

Tabel 7 Approach

Subquestion	Research strategy	Methods	How?
1. What data sources could be useful to enrich the fan experience based on the requirements?	<ul style="list-style-type: none"> Library Lab Field 	<ul style="list-style-type: none"> Literature Study Best good and bad practices Data analytics Survey 	<ul style="list-style-type: none"> Do research on what fan experience is. Research on how other data could be useful by looking at best, good and bad practices Analyze the data we got provided from PSV. Send survey to Experience Box customers to get information on what data they would like to see.
2. How will the available data be used to enrich the fan experience of the stadium?	<ul style="list-style-type: none"> Field Workshop 	<ul style="list-style-type: none"> Interview Business case exploration Literature study Requirement prioritization 	<ul style="list-style-type: none"> Interview stakeholders on how they want to use the data Analyze the most viable options or scenarios Do literature study on how to set up requirements based on interviews. Make a MoSCoW table of possible solution
3. How can the available data sources be combined?	<ul style="list-style-type: none"> Library Workshop 	<ul style="list-style-type: none"> Literature study Community research IT architecture sketching 	<ul style="list-style-type: none"> Research on how to combine data Search for the problem in trusted online communities Sketching the high-level architecture
4. How can the combined data be connected to the interface of the application?	<ul style="list-style-type: none"> Library Field 	<ul style="list-style-type: none"> Document analysis Community research Expert interview 	<ul style="list-style-type: none"> See what the interface of the app consists of Search for the problem in trusted online communities Do an interview with an expert on how to connect the data to the interface of the application



2.9. Deliverables

Product 1: Proof of Concept (PoC)

The key aim is to create a concept of backend part for web application, where the data from different sources are combined. Cleaned, analyzed and visualized data is expected as an output of application. That application will be used on tablets from the visitors at the Experience Box.

Since the project is a cooperation between two student groups, as part of the minor DDBL, Datastic is mainly required to produce proof of concept and guidance on it rather than a completely functional application.

In other words, to deliver a document with specifications for all conceivable sources, how they function, how to pull data out, and information on how to combine data from all sources. And, in the best-case scenario, if everything goes according to plan and after collaborating with the second group, a functioning prototype may be provided.

Product 2: Advice about PoC - Information about the possible usage of data

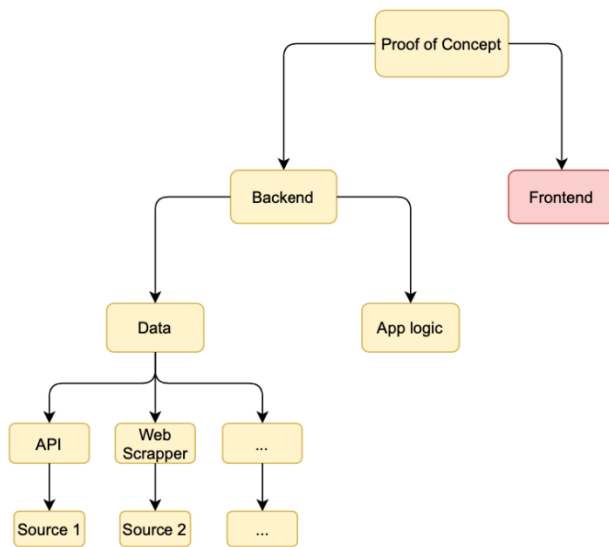
The second product could be considered as a part of Product 1 and includes the document with advice about possible ways of using data from the application.

Product 3 - Project plan

A project plan is essential at the start of every project. It serves as the foundation for the entire project, establishing rules for the entire duration of work. It also includes some of the most important parts, such as a plan of action, an explanation of the deliverables, a project goal, various analyses, and so on - all the activities, tasks, and resources that will be used to complete the project.

2.10. Product decomposition structure

The main product of this project from Datastic company is a Proof of Concept. PoC includes web application and information about how the data could be used. Web application is simply divided into frontend and backend parts. The first part is the responsibility of another company. However, as was said above, we will be collaborating with this company. Backend part consists of data and app logic. App logic is represented by scripts. Data are collected from different sources by using different tools.



Figuur 1 Decomposition structure



2.11. Project management

2.11.1. Preconditions

Predicates that must always be true prior to the execution of the project:

1. Team members have a broad idea of the project's scope
2. Team members have defined roles within the company
3. Team members and stakeholders plan achievable goals in the given timeframe
4. Team members and stakeholders are aware of the project risks and constraints

2.11.2. project organisation

Tabel 8 Project organization

Student	Role
Lieke Nijs	Project Leader
Agorasteva Gergana	Note Taker
Femke Boogerd	Chairman
Victor Plesciuc	Team Member
Remco Bisschops	Team Member
Ilya Tsakunov	Team Member
Nikita Gavrilov	Team Member
Yaniek Martens	Team Member
Sasha Vereshchagin	Team Member

Project Leader - manages the team, makes sure the deadlines are met and the deliverables are sufficient

Chairman – meeting coordinator, spokesperson for the group

Note Taker – takes organized, legible and typed notes every meeting

2.12. communication and consultation

The communication between PSV and Datastic will be made mainly through Microsoft Teams and Microsoft Outlook. Physical meetings will only be arranged if the safety COVID-19 measures are properly met.

Meetings Planning

For the duration of this project, recurring meetings will take place every Monday at 10:00-11:00 AM.

2.13. risk management

This chapter is devoted to Risk Management. In context of this project, we assigned 6 risk categories. 5 of them went from the PESTLE Analysis. PESTLE stands for Political, Economic, Social, Technological, Legal and Environmental factors. This analysis allows to form an impression of the factors that might impact the product



(What is pestle analysis, n.d.). One of the elements (environmental risk) was replaced with financial category as non-relevant for this project.

Tabel 9 Legend category risks

Category	
P	Politic
E	Economic
S	Social
T	Technological
L	Legal
F	Financial

Typical Risk Management is usually considered in 2 aspects: probability and impact. Both aspects were divided into 3 levels: Low, Medium, High.

Tabel 10 Legend probability risks

Probability	
L	Low
M	Medium
H	High

Tabel 11 Legend impact risks

Impact	
L	Low
M	Medium
H	High

For each risk were identified probability and impact and depending on these indicators risk level was determined.

Tabel 12 Legend risk levels

Risk Levels	
Low Probability X Low Impact	Low
Low Probability X Medium Impact	Low
Medium Probability X Low Impact	Low
Medium Probability X Medium Impact	Medium
Low Probability X High Impact	Medium



High Probability X Low Impact	Medium
Medium Probability X High Impact	High
High Probability X Medium Impact	High
High Probability X High Impact	High

The most common ways of risk mitigation are shown in a table below. (RISK MITIGATION PLANNING, n.d.)

Tabel 13 Legend Risk mitigation

Risk Mitigation	
Av	Avoidance
Acc	Acceptation
S	Sharing
R	Reduction
T	Transfer

Based on initial information about the project, the following risks were defined. The list below is not final and will be expanded as new information will be received.

Tabel 14 Risk management table

ID	Cat.	Risk	Prob.	Impact	Level	Mitigation	Actions
R1	L	Data leakage	M	H	H	Av	Improving privacy
R2	F	Exceeding the project budget	M	H	H	R	Planning in advance
R3	S	New wave of pandemic	M	H	H	Acc	Preparing based on previous experience
R4	S	Lack of specialists	L	M	M	Av	Cooperation with educational institutes
R5	L	Stealing an idea by competitors	L	H	M	Av	Making patents, Regular market research

The first risk is data leakage. If personal data is used, the risk level is very high. This should be avoided by improving the privacy and making sure the probability will lower from medium till low. The second risk is exceeding the project budget. When the budget is exceeded, the project cannot be executed properly so this should be avoided. This can be done by planning in advance. The third one is a new pandemic wave. This does have an impact on the project. However, there is already a hybrid between online working and physical meetings. In case of a new pandemic wave, this will go more to online working. Another important part and a



risk is lack of specialists. This could stop the development of project for a while. However, we consider this risk as low probable as PSV has partnership with Brainport Eindhoven - a place, where the concentration of perspective specialists is relatively high. The next risk is again part of legal category. As PSV Experience Box could be considered as innovative product, it may probably be an attractive target for the competitors, who can release the same product in a shorter time, which could jeopardize the whole project. Based on this, a regular market research should be done. In addition, making patents on all developed products is recommended.

2.14. global planning

In Figure X the planning for the project is shown. This planning will show every deliverable that is mentioned with the start and end date next to it. During the project we are supposed to finish the deliverable before every deadline.

PSV - Brainport Experience Box

		Project Start Date	06/09/2021			Display Week	1
		Project Lead	Lieke				
WBS	TASK	START	END	DAYS	days completed	% DONE	WORK DAYS
1	Projectplan	06-09-21	27-09-21	22	0	0%	16
1.1	Introduction	06-09-21	17-09-21	12	12	100%	10
1.2	Context & Background	06-09-21	17-09-21	12	12	100%	10
1.3	Project Definition	06-09-21	17-09-21	12	12	100%	10
1.4	Project Management	06-09-21	17-09-21	12	0	0%	10
1.5	Verwerken feedback docent	17-09-21	25-09-21	9	0	0%	6
1.6	Verwerken feedback opdrachtgever	27-09-21	03-10-21	7	0	0%	5
2	Research report	20-09-21	14-11-21	56	0	0%	40
2.1	Introduction	01-11-21	07-11-21	7	0	0%	5
2.2	Subquestion 1	20-09-21	31-10-21	42	0	0%	30
2.3	Subquestion 2	20-09-21	31-10-21	42	0	0%	30
2.4	Subquestion 3	20-09-21	31-10-21	42	0	0%	30
2.5	Subquestion 4	20-09-21	31-10-21	42	0	0%	30
2.6	Conclusie	01-11-21	14-11-21	14	0	0%	10
2.7	Discussie & Advies	01-11-21	14-11-21	14	0	0%	10
2.8	Feedback Fontys	15-11-21	21-11-21	7	0	0%	5
2.9	Feedback PSV	22-11-21	28-11-21	7	0	0%	5
3	Proof of Concept	15-11-21	23-01-22	70	0	0%	50
						0%	
						0%	

Figuur 2 Project planning



Bibliography

- Brainport Experience Box . (n.d.). Retrieved from brainporteindhoven.com:
<https://brainporteindhoven.com/psv/nl/trots/experiencebox>
- What is pestle analysis. (n.d.). Retrieved from <https://pestleanalysis.com/what-is-pestle-analysis/>
- RISK MITIGATION PLANNING. (n.d.). Retrieved from <https://www.mitre.org/publications/systems-engineering-guide/acquisition-systems-engineering/risk-management/risk-mitigation-planning-implementation-and-progress-monitoring>
- PSV Eindhoven. (2021, 09 09). *PSV - Organisatie*. Retrieved from PSV.nl.
- Brainport Eindhoven. (2021, 09 20). *Wat is Brainport Eindhoven*. Retrieved from Brainport Eindhoven:
<https://brainporteindhoven.com/nl/ontdek/wat-is-brainport-eindhoven>
- Koersen, W. (2018, 06 21). *Web Scraping, Regular Expressions, and Data Visualization: Doing it all in Python*. Retrieved from Towardsdatascience: <https://towardsdatascience.com/web-scraping-regular-expressions-and-data-visualization-doing-it-all-in-python-37a1aade7924>
- Summit. (n.d.). *Reading Data from the Web: Web Scraping & Regular Expressions*. Retrieved from Summit:
<https://www.summet.com/dmsi/html/readingTheWeb.html>
- MHerman. (2012, 05 10). *Web Scraping with Regular Expressions*. Retrieved from MHerman:
<https://mherman.org/blog/web-scraping-with-regular-expressions/>
- Octoparse. (n.d.). *Using This RegEx Tool to Match HTML Tags*. Retrieved from Octoparse:
<https://www.octoparse.com/blog/using-regular-expression-to-match-html#>
- Techopedia. (2017, 03 23). *Parse*. Retrieved from techopedia:
<https://www.techopedia.com/definition/3853/parse>
- Developer. (2021, 09 20). *Parse - MDN Web Docs Glossary: Definitions of Web-related terms | MDN*. Retrieved from Developer: <https://developer.mozilla.org/en-US/docs/Glossary/Parse?retiredLocale=nl>
- html5rocks. (n.d.). *How Browsers Work: Behind the scenes of modern web browsers - HTML5 Rocks*. Retrieved from html5rocks:
https://www.html5rocks.com/en/tutorials/internals/howbrowserswork/#HTML_Parser
- Ortec sports. (2020). *Business Case Performance* . Retrieved from www.ortecsports.com:
<https://ortecsports.com/wp-content/uploads/2020/03/Business-Case-Performance-PSV.pdf>
- adam@olivetreecreative.co.uk. (2018, 05 15). *Why is everyone talking about 'fan experience'?* Retrieved from rapidretail: <https://rapidretail.co.uk/talking-about-fan-experience/>
- Productplan.com. (2021, 09 11). *MoSCoW Prioritization*. Retrieved from Productplan:
<https://www.productplan.com/glossary/moscow-prioritization/>
- CourtVision. (n.d.). <https://www.clipperscourtvision.com/>. Retrieved from <https://www.clipperscourtvision.com/>
- AR to sports. (n.d.). Retrieved from <https://digital.hbs.edu/data-and-analysis/bringing-the-ar-cloud-to-sports/>



Tobin, D. (2020, 07 08). *How to Get Data from Multiple Sources*. Retrieved from xplenty: <https://www.xplenty.com/blog/get-data-from-multiple-sources/>

Breschi, A. (2022, 01 12). *16 Types of Customer Needs (and How to Solve for Them)*. Retrieved from Blog Hubspot: <https://blog.hubspot.com/service/customer-needs>

Engdahl, M. (2020, 12 7). *What is data merging?* Retrieved from Displayr: <https://www.displayr.com/what-is-data-merging/>

Kayfitz, A. (2019, 5 2). *How to Merge Data from Multiple Sources into a Single Platform*. . Retrieved from Practical Ecommerce: <https://www.practicalecommerce.com/how-to-merge-data-from-multiple-sources-into-a-single-platform>

ETL vs ELT: The Difference is in the How. (2021, 07 27). Retrieved from Panoply: <https://blog.panoply.io/etl-vs-elt-the-difference-is-in-the-how>

(n.d.). <https://www.dfl.de/en/news/kick-off-for-5g-latest-mobile-communications-technology-activated-in-the-first-bundesliga-stadium/>. Retrieved from <https://www.dfl.de/en/news/kick-off-for-5g-latest-mobile-communications-technology-activated-in-the-first-bundesliga-stadium/>



2.15. appendix 1 - psv stadium meeting

Femke: Yeah. So the first question was, is there a demand for an experienced box from the customers? Or is it just something you would like to do?

PSV: Yeah. So there's not a demand for fans who say, we want experienced box, but we do get some signals. For instance, there are fans who are asking for is it possible to do in seat delivery of catering? So some of the things we want to deploy really are designed or or created by us and some things you want to test.

Femke: That demand is it only the delivery or is it also for more information?

PSV: So yeah, to yeah, there are some fans who saying, can we have more data on the screens, for instance, we get those kinds of requests, but it's not those thousands of supporters who request that.

Femke: But it's not like they ask for “No, I'm not happy now how it goes”?

PSV: No, this project is to maybe try to be, uh, before the question is there, maybe to already have it or maybe to create a demand.

Femke: So, there aren't any requirements from customers

PSV: No, not yet.

Lieke: This is actually based on what you want to measure to the customer.

PSV: Yeah, and I think a lot of times with innovation maybe are trying to create something where it's not yet a demand for, a real demand.



Femke: Um yeah. And what is the current situation with the experience ? What is there already possible?

PSV: There are three possibilities – on seat delivery, you get some ortec data on players. So player and performance data and can you see video and review, we can play back moments of the match.

Lieke: How is that data presented?

PSV: Yeah. So now it's just a simple front end application from ortec, but we already know that's not good enough. It show only limited information.

Femke: Does everyone see the same thing on their screens, or can you choose for yourself what you want to see?

PSV: No, you just get what is there. You can't choose yourself, its quite limited.

Speaker1: Yeah, OK. And how do people that already went to the experience box, how did they experience it?

PSV: Yeah, they think it's nice, but to be honest and liked it. But they say, they want some more interactivity. They had good experience, but not yet on the data part. It can be much better.

Femke: So, it's not yet what you want it to be.



PSV: No, no, no. That's also why we have this group looking at more on the data and another group who will also look at what should the application look like.

Lieke: Is that group already known?

PSV: They didn't start yet. I will ask when they start. But maybe I think timing wise, it's not that bad because first we have to look at the data. And I think maybe the last step is which type of graph are we going to show? Yeah, that's important. But uh, first question is, can we get the data? And then also, can we create something ourselves? Because now we only have an application, but we want to know from ortec, how is your database looking and can we put it there immediately?

Femke: Is the data that is already there available. Can we get the data?

PSV: I asked them last week to give us more information about, uh, can show us what kind of database is it, uh, open. Can we get API connection or. So I understand that you need those insights, but I'm waiting on it. But I expect more information this week. But the only thing I could give you already is the social data. The second link worked, I think.

Femke: Yeah, and we made a scope about the project and kind of research questions we wanted to see if you think about it.

Lieke: Another question: since when is the experience box there?

PSV: Since this season and we had two ideas. First, we thought it has to be totally finished before we're going to start or we said we can just start and by looking what's happening, to work towards a good end product because also it's a waste to keep them (the seats) empty.



Lieke: What are all the options you can be invited to the seats?

PSV: We do an activation on social media and by email, so there's bit of activation is from brain boards. So we do a quiz, I can send it to you, and you need to get some answers and then you can win two tickets for the Experience box. If you leave behind your data of course.

Femke: And is it possible to maybe ask them, the people who had already been to the experience box, to ask them for how did you experience it? What are your thoughts?

PSV: I will check if they already do it, and otherwise we should start it by now. I. You can ask them what they think. I will check what's already there and otherwise I think it's a good point that we start doing it.

Lieke: And how many times if the box filled up, is it every match?

PSV: Every home match, it's always filled up. There are always people who want to go to a match.

Victor: On video, I saw you had VR glasses there. What for?

PSV: Oh, no, we don't have them, but it's more to show how can this be the future of watching football. It's just to build up the image towards where is the consumption of football going. Will there be a time when VR or AR will be included? It's a question mark, but we don't have the VR glasses. Also there's also a project around 5G. So I think within a couple of weeks we will also have 5G and that's also again, an enabler for some techniques.



Remko: Is it only for the box and for the World Stadium?

PSV: Uh, the end goal is to have it for the entire stadium. Yeah, but that will not go live immediately because you also have to have something that needs 5G. Yeah.

Femke: Yeah, I already went to all the questions that we had, so I think we can discuss the scope first

PSV: I think It's good. I totally agree. So the project will not be finished, uh, by you guys, but it's more an investigation. Maybe draw how it should look like and maybe a first working prototype. But I agree that it's, probably not doable to to have it completely finished.

Remko: And what are your research questions? Um, it was a question in the slides, but we tweak it at a bit. So the main question will be how can we enrich the fan experience of the stadium visitors using multiple data sources? Because the other question was, I think for the whole project, uh, with the digital marketing part also involved. And then we have sub questions. Uh, what data sources are available and what other data could be useful to enrich the fan experience of the stadium? So that's the first part. We are going to look first at what data, you have available. So the six data source in the slides and then we can think about the more possibilities ourselves. For the second part, how will the available data be used to enrich the fan experience of the stadium? So then we are going to look, uh, how can this data be used to enrich the experience?

PSV: Yeah. So how can you use the data? Yeah.



Remko: Uh, the third thin, is how come the available data sources be combined? How can you combine everything together and put it in a database or something like that? And for the last one was how can the combined data be connected to the interface of the application? So for that part, we are going to, uh, discuss with the other group. So that's how we thought.

PSV: I think it's good. Um, maybe we came up with four sources and we can all take data out for that. So that's good. Uh, but it's also depending on the second group, if they say, yeah, all those four types of data need to be in one app or tool. That makes it more difficult than if they say no, it's no problem that the weather is in a different section than the player data. But you are depending on also what they think is necessary for the end user. So I think this is a good setup. So indeed, what kinds of sources are there? Which ones do we now think are relevant? Maybe that can also change that, but we have to, take our best guess for now and then take the next question - How is that data accessible, for instance? And then you have to think kind of get out the data in real time or not, because I know, for instance, we have food and beverage data, but I can only access it now, the day after and then it's not relevant for that box because they are in the stadium at the match. Or maybe they can look at what was consumed two weeks ago, but that's not that interesting. So, yeah, it's a it's the kind of data interesting how the source looking like, can we connected and in which way - is it in real time because you need it fast if you want to display something about the match?