

Vitality platform

Municipality Eindhoven

Research theme: Webscraping &
Matchmaking

Semester name: DDBL



DATASTIC



Project Description

Context

In this day and age citizens expect increasingly more transparency, accessibility and responsive services from their government (Gemeente Eindhoven). This can only happen by responding to citizens' wants and needs. The municipality of Eindhoven wants to create a platform that makes it as easy as possible to find the sport & facility activities that match the citizens' specific preferences. The platform will be used for all citizens of the city. To get the most out of the platform, there has to be some research on the current mismatches. Therefore, did the municipality created target groups to start customer journeys with. In this project can't all customer journeys for all the target groups be looked at. Besides serving every citizen in the city, there also are two defined specific target groups where in this project can take a deep dive into their customer journey. These defined target groups are expats and youth from 12 years.

Municipality Eindhoven decided to choose expats/internationals as a target group because they are very much skilled to use digital tools but are not familiar with the Dutch sports system. A tool that gives them quick and easy insight into their sport needs, could therefore be very impactful for them. The government hopes people will do sports during their entire lifetime. The municipality does know that doing sports at a young age improves the chances that people will do sports when they're adults. Therefore, the second target group is the youth. This project will focus on children of 12 years and older, because it turns out that many children stop doing sports in this category.

That's why municipality Eindhoven has started to research user needs of the citizens of Eindhoven. The municipality hired Bureau Moeilijke Dingen to research the user needs of expats and in what way this can be met. They came up with a concept of an app where the citizens of Eindhoven can easily find information about sporting events (Gemeente Eindhoven). To make this app and integrate this they made a project plan where they work in phases. In phase 1 all the sport offers will be brought in scope and there will be looked at the matchmaking aspect of the app. Also, does phase 1 consist of law and ethic. In phase 2 there will be looked at the central platform with integrated platforms while also maintaining control of privacy and with transparent AI. The last phase will consist of the development of the digital economy. This project will focus on the first phase.

Main research question

How can municipality Eindhoven create a platform with integrated intern and extern data for citizens of Eindhoven to find and fulfil their sport needs with matchmaking?

Sub questions

1. What are the end users' requirements for a sport platform?
2. How can sport activities that exist on different websites be combined?
3. What is the most suitable matchmaking technique to show user's sport preferences?
4. How can the chosen matchmaking technique be implemented to display the user's sport preferences?
5. How can the matchmaking be ethically responsible and transparent?

Results

The result of this project leads to the proof of concept of the first step of the initial municipality's project plan. In the past six months the Datastic group has researched, developed, tested and validated the initial concepts that were advised by Bureau Moeilijke Dingen such as obtaining information from web using web scraping techniques, matchmaking possibilities, law and ethics investigation. In this chapter the results of above concepts will be discussed as well as other Portfolio assets of Datastic.

Portfolio assets:

- Project Plan
- Research Document
- WebScraping prototype
- Matchmaking prototype
- Law and Ethics investigation

The municipality asked Datastic to create a prototype based on the web scraping and matchmaking part of the platform of vitality. Datastic divided this project in separate phases where each phase had their own outcome.

Firstly, during the first phase end-users liked to see the description, timetable, place, location, recent activity, age, experience level and price for each activity within the application. In the second phase of the project, the web scraping took place. During this phase Datastic figured out that almost all of the websites didn't contain open activities. There is no content to scrape and the HTML structure for each website diverse a lot compared to others. Also, the content that end-users wanted to see in the application, is not available at the activities. Datastic advice to do one of the following things:

- Create your own websites where associations or commercial companies can put their activity into a form. This means every activity will be putted into the Municipality's site in the same structure, so it is easy usable for their platform (this will reduce the costs) and every needed content can be asked to the organization.
- Contact the organizations about your plans and perform research if they are willing to update their websites with the content the municipality likes to scrape.

In the third phase matchmaking took place. There were a few options available to do the matchmaking with, like machine learning and a rule engine. Datastic chose for a rule engine because the machine learning wasn't possible with the available information. Datastic advice the municipality to keep using the rule engine at the first phase of the platform. This rule engine fits perfectly to the desired swiping solution the municipality wants. In a later stage of the platform machine learning could be added by the municipality which can offer activities to the users based on their previous preferences. The municipality must consider which solution they will use because of the complexity of the model.

During the phases also the ethical part was taken in mind by Datastic. Based on different core values a couple of target groups were determined. These target groups all have different expatiations about the application. Datastic advice the municipality to use these expectations when they are developing the platform. Also take in mind to look further then the used core values.

Methodology

Sub question	Research strategy	Method
1. What are the end users' requirements for a sport platform?	<ul style="list-style-type: none"> - Field - Workshop 	<ul style="list-style-type: none"> - Interview <ul style="list-style-type: none"> o Interviews held with people in the focus group - Focus group <ul style="list-style-type: none"> o Conclusion from interview is discussed with the focus group - Requirement prioritization <ul style="list-style-type: none"> o With the conclusion from the interviews and discussions a list of requirements is made using the MoSCow method
2. How can sport activities that exist on Eindhoven Sport website be combined?	<ul style="list-style-type: none"> - Library - Workshop 	<ul style="list-style-type: none"> - Literature study <ul style="list-style-type: none"> o Datastic did research on languages and platform that can be used for webscraping. - Best, good & bad practices <ul style="list-style-type: none"> o Datastic did research on the best way to webscrap and what common mistakes are made with webscraping. - Expert Interview <ul style="list-style-type: none"> o Held expert interviews with the people from the work field who have a better understanding about webscraping. - Prototyping <ul style="list-style-type: none"> o A prototype was made to webscrape the website's that the municipality had choicen, with conclusion that most websites are unusable.
3. What is the most suitable matchmaking technique to show user's sport preferences?	<ul style="list-style-type: none"> - Library - Steppingst one - Lab 	<ul style="list-style-type: none"> - Literature study <ul style="list-style-type: none"> o Datastic did research on machine learning and rule engines. For machine learning there was also looked at clustering, K means algorithm and decision threes. - Best, good & bad practices <ul style="list-style-type: none"> o Datastic did research on the best way to make a rule engine and what common mistakes are made with making a rule enigne. - Data analysis <ul style="list-style-type: none"> o Because no information could be webscraped, a file with mock up data was made. - Persona <ul style="list-style-type: none"> o Datastic designed a persona to create reliable and realistic representations of the key audience segments for reference.
4. How can the chosen matchmaking technique be	<ul style="list-style-type: none"> - Workshop 	<ul style="list-style-type: none"> - Prototyping <ul style="list-style-type: none"> o Datastic made a rule engine using a score system. - Ideation

<p>implemented to display the user's sport preferences?</p>		<ul style="list-style-type: none"> ○ Future steps that need to be made are data storage and collaborative filtering. This is to start using machine learning.
<p>5. How can the matchmaking be ethically responsible and transparent?</p>	<ul style="list-style-type: none"> - Workshop - Showroom 	<ul style="list-style-type: none"> - Brainstorm <ul style="list-style-type: none"> ○ The ethical responsibility from the vitality platform is discussed with the help of a counselling ethics discussion with use of a brainstorm. - Product review <ul style="list-style-type: none"> ○ With the TICT Tool there was looked at the value Inclusivity, which was one of the values that was found to be important in the brainstorm session. - Requirement prioritization <ul style="list-style-type: none"> ○ With the brainstorm session and the TICT Tool it emerged the top three values are important for this project.

Handover

resultsboxing

	Title	Day	Time	Locations
2	Wednesday Kickboxing &	Wednesday		Sporthal 3A
4	Friday Introduction t	Friday	21:30-23:00	Dojo

Figure 1 Results webscraping

Sport Providers:	SportProvider ...	Location
31	Keebler LLC ...	891 Del Sol Parkway
32	Jacobi-Borer ...	42 Mockingbird Pass
33	McKenzie-Rosenbaum ...	63932 Lake View Hill
34	Veum and Sons ...	78 Moulton Park
35	Leannon-Kautzer ...	87 Reinke Trail
36	Metz, Price and Swift ...	7 Duke Avenue
37	Will-Krajcik ...	06545 Charing Cross Pass
38	Morar Group ...	85 Sachtjen Parkway
39	Gislason, Bernier and Hudson ...	77219 Morningstar Park
40	Labadie LLC ...	06 Northport Parkway
41	Legros, Glover and Johnston ...	8573 Oriole Pass
42	Hagenes and Sons ...	393 Fulton Plaza
43	Terry Inc ...	1502 Elmside Terrace

Figure 2 Results Rule engine



Figure 3 TICT-Tool Municipality