DASU20: Group 5 Final Report

George Kontos	Tijn La Vos	Milda Virbickaitė	Ian Zhang
1531409	1757946	1749498	1706543
g.condos@student.tue.nl	m.b.l.vos@student.tue.nl	m.virbickaite@student.tue.nl	i.j.zhang@student.tue.nl

Intro

This assignment is for the course “Data acquisition and visualization through embodied sensors.” We used the lessons learned in Assignment 1 (creating the profile of a runner) and Assignment 2 (analyzing real time data), to complete Assignment 3 which was to create a motivational profile for the group of runners through the data showing their results from Eindhoven half-marathon and to create a visualization of the progress of one of the runners for the rest of the team to keep track of. The personal and identifiable data of runners needs to be handled in a way no privacy rules are broken and the work complies with the code of ethics.

Assignment 1

Bib number:	Finish position:	Finish time:	Average speed:	Pace:	Team:	Gender:
6241	5132	2:45:23	7.65 km/h	7:50 min/km	N/A	male

Physical appearance

- short shocks
- athletic sneakers
- athletic poly-material T-shirt
- loose poly-material shorts
- analog watch
- bulkier/stocky build

Behavioral observations

- beginner
- began at a fast pace (around 6:27 min/km) only to drastically drop in pace between the 5 km mark to the 10 km mark
- after the 10 km mark the pace slows down consistently from 7:49 min/km to 8:41 min/km

Bib number:	Finish position:	Finish time:	Average speed:	Pace:	Team:	Gender:
6407	5143	2:51:41	7.37 km/h	8:08 min/km	N/A	female

Physical appearance

- short shocks
- athletic sneakers
- poly-material T-shirt
- loose poly-material shorts
- fitness smartwatch
- wireless headphones
- average/heavier build
- tied back hair

Behavioral observations

- beginner
- exponential drop in pace
- starting from 7:27 min/km at the 5 km mark to 9:04 min/km at the 21.1km mark

Figure 1: Profile of 5 half-marathon runners.

Reflection of Assignment 1

In Assignment 1, we were taught about how to make observations about runners via publicly found data and visual information. Especially what kind of details can be extracted from minute details, such as what

the attire of the runner could imply. However, in this assignment it was not clear to use Vos's profiles so that was accidentally omitted in this assignment.

Assignment 2

1. **Accessing the data interpretation software and applications for assignment 3**

Data Interpretation Software Pros and Cons

Excel

Pros: The program can easily open data to make Graphs/Tables/Charts. It also displays all the data that was collected so that one can parse through every detail.

Cons: The program will present all data sets in a table format requiring the user to make formulas to clean the data. Excel also has no ability to provide geographical context to one's data. One will need to calculate data such as speed, distance, time (since it gives real time rather than recording time).

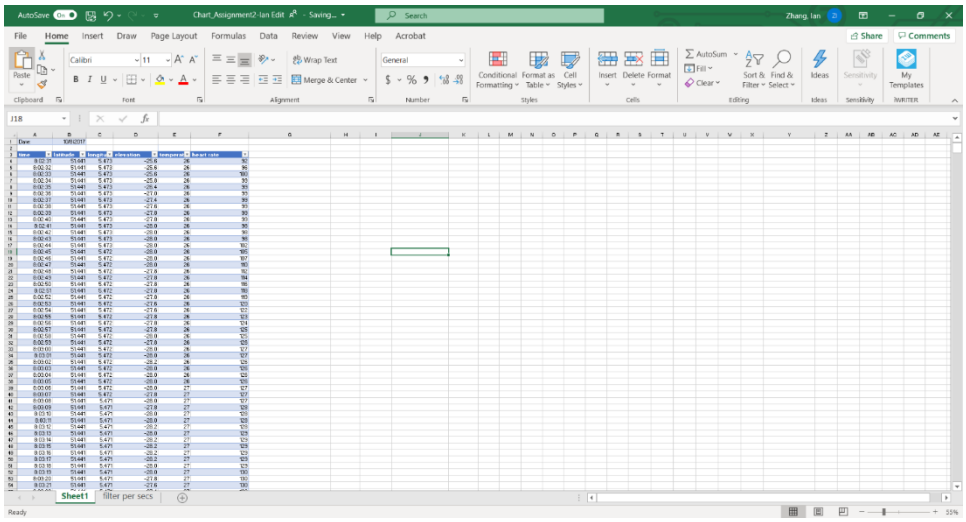


Figure 2: Excel Import of GPX Data

Python

Pros: Useful when handling large volume datasets. Python is focused on simplicity therefore also making the overview of datasets easily readable. Does not need a lot of code to program a visual overview.

Cons: The learning curve to understand the coding language for python takes more time/experience when in comparison to using excel. It's harder to link specific tables from the dataset to create the query. It feels like selecting data from an SQL database. Overall, the python interface is not as easy to work with as excel if one does not have a good understanding of the coding language.

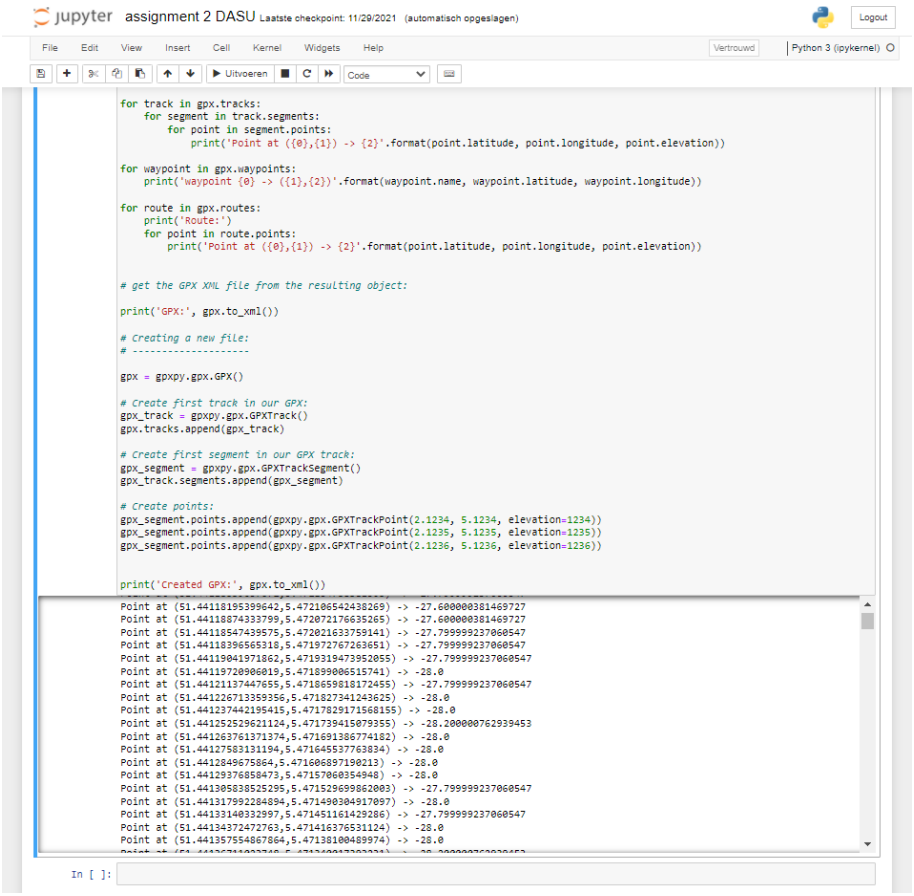


Figure 3: Python Coding and output from GPX Data

GPX.Studio

Pros: The data can be opened in a browser meaning that no external programs or programming is required. On top of that, GPX.Studio presents the data, which usually needs to be run through formulas, in a user friendly/graphically pleasing/easy to understand way. In other words, GPX.Studio has some UI/UX design incorporated with it. Included with the data is the geographical context for the data which could give more clues and information about the run.

Cons: Because the program is so optimized, it is harder to find specific pieces of data (a specific time/distance). The optimization also excludes statistics the designers found unnecessary, such as heart rate or temperature. Overall, to have this clean interfacing and design, sacrifices to the data were made which could be vital for the third assignment.

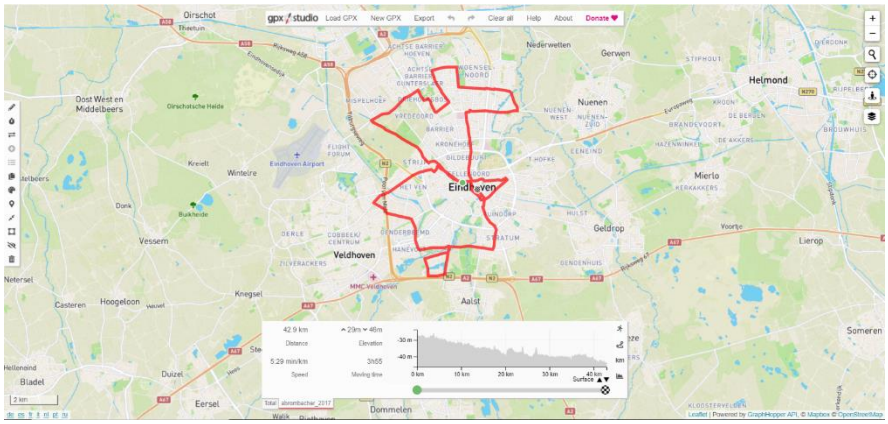


Figure 4: GPX Studio visualization of data

2. **What data contains a GPX file? How is this data structured? How can I use this for assignment 3?**

A GPX file contains, Time (World clock time), Latitude, Longitude, Elevation, Temperature, Heart Rate. The data is structured as an extensive list where all contained information is recorded. For assignment 3, the GPX file will give us good insight on how the runner's body is reacting to the work-out, general information about the session, as well as the data required to prove why we might suggest a specific fitness regime.

Analysis of GPX data
It was found that there was a minor difference between tracking every second and every thirty seconds which led us to reduce the data points every thirty seconds. We also concluded that the main pieces of data we will be using for our suggestions would be the time, longitude, latitude, as well as heart rate (as we cannot seem to prove that temperature influenced the run as well as elevation).

General Running stats - Speed: 10.8 km/h, 3-hour 55 minutes run.

Heart Rate

Throughout the run, the heart rate ranged from 127 beats per minute (BPM) to 169 BPM.
Based off the trendline, there is a general decrease in heart rate from the start (165 BPM) to the end (145 BPM).
From the start to the 15th minute (~3.04km), the heart rate begins to rise. Between the 15th minute (~3.04km) to the 91st minute (~17.5 km), the runner's heart rate is above the trendline going as high as 169 BPM and approximating around 160 BPM. Between the 91st minute and the 193rd minute (~34.29 km), the heart rate is below the trendline jumping at sharp intervals between 125 BPM to 157 BPM. After the 193rd minute, the heart rate is above the trendline until the 231.5th minute (~42.03 km). Ending the run with a lower than trendline heart rate between 148 BPM to 139 BPM.

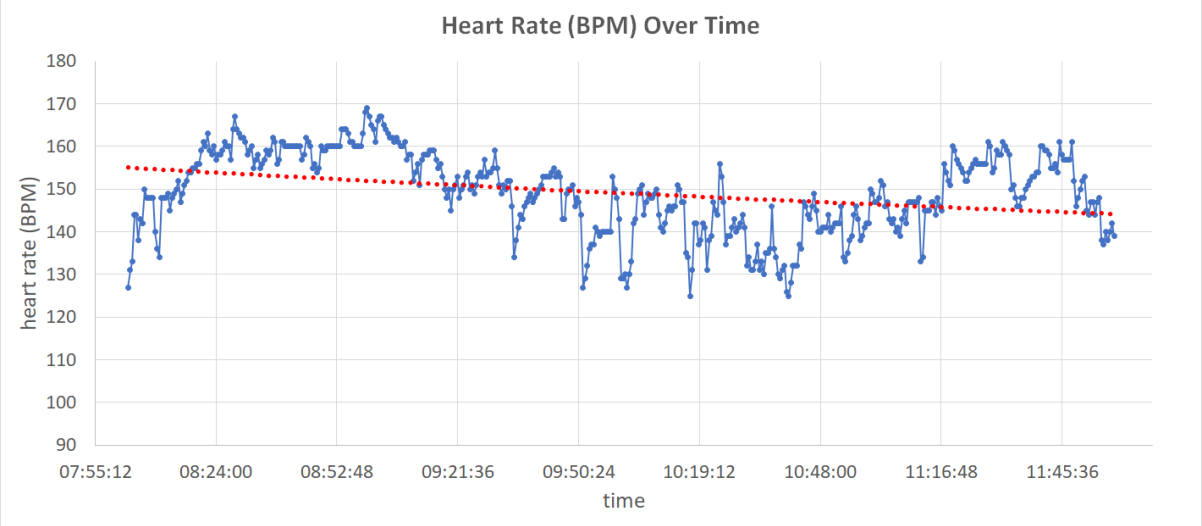


Figure 5: Heart Rate over run

Speed and Pace

Throughout the run, the speed deviates from 6.48 km/h (which was logged at the 54.5th minute mark or 10.56 km) to 21.6 km/h (which was logged at the 21st minute mark or 4.23 km).
The run average around 14.4 km/h from start to finish where the runner generally alternating between 10.8 km/h to 18 km/h.

Pace/Speed at milestones:
5km: 5:08 min./km, or ~11.7 km/h, 10km: 5:19 min/km, or ~11.3 km/h, 15km: 5:27 min/km, or ~11 km/h, 20km: 5:19 min/km, or ~11.3 km/h, 25km: 5:16 min/km, or ~11.4 km/h, 30km: 5:30 min/km, or ~10.9 km/h, 35km: 5:36 min/km, or ~10.7 km/h, 40km: 5:53 min/km, or ~10.2km/h, 42.19km: 6:11 min/km, or ~9.7km/h

Side note: We are not sure if we did our pace calculation correctly. For example, if we were to determine the pace at the 30 km mark, would the calculation be time/30km or time/5km (as the pace is calculated between the 25-30km marks). Now, we have it time/total distance rather than time between two km markers.

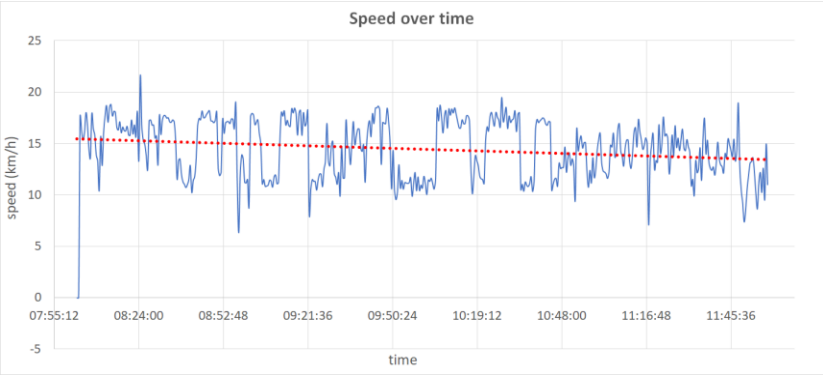


Figure 6: Speed over time

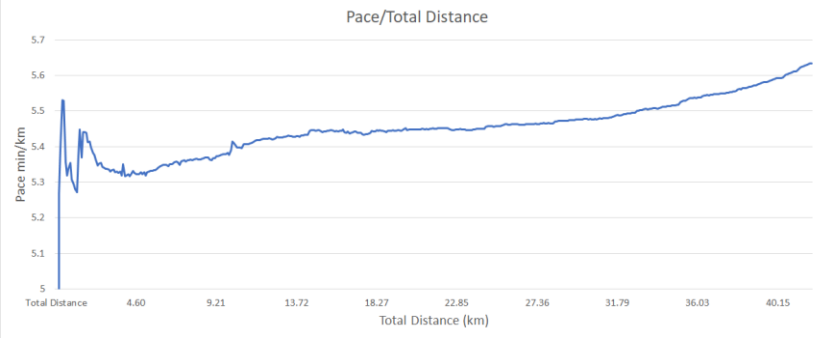


Figure 7: Pace over total distance

3. Observations and suggestions

Overall, it looks like an average to above average run when compared to runners in the 2021 Eindhoven marathon (this runner is about 6 minutes slower than the 763rd /1526 finisher). From the running data as well as Vos' typologies of runners, we assume that the runner is an individual fitness runner as well as an individual competitive runner based off of the time (3 hours and 55 minutes in comparison to the 763rd/1526 runner whose time was 3:49:56), average speed (10.8 km/h compared to the 763rd runner whose average speed was 11.01 km/h), heart rate (Ranging from 127-169 Bpm), as well as the 2nd part of the run where the runner is doing intervals of high and low speed runs [5]. The runner is assumed to be an individual runner due to the lack of data proving that they could have done the marathon, or training, in a group [10].

By observing the average speed, time, and heart rate of the run, we can make the conclusion that the runner has trained for the marathon and would be considered at least an individual fitness runner, they are not a top athlete as they are sitting about middle of the pack in time and speed. Using this observation in connection to Reiss' theory of motivation [8], this runner has a fervent desire for physical activity, curiosity on how the runner can push their limits, independence in terms of the need of control over oneself, and power as running a marathon in a decent time requires a strong will. These desires are what leads us to conclude that this runner can fit the profile of an individual fitness runner.

If you were to observe the speed as well as the heart rate over the course of the run, then one could assume that the runner can also be a competitive runner even though their time might not be as good as a top runner. The competitive runner profile can be seen in the middle of the run, between the 6 km mark to the 30 km mark where the runner is running 1-3 km intervals of ~18 km/h to ~10 km/h, as well as the start of the run where the runner is pushing their heart rate to above the recommended limit of 160 Bpm at the 4.96 km mark to 12.65 km mark [3]. The interval running makes us believe that the runner has a desire to perform at a higher rate, yet they might lack the possible endurance to maintain that speed. Once again, using Reiss' theory of motivation, we assume that this runner also had the desire of vengeance (or the need to compete). When we begin to look closer into the GPX data, the running pattern as well as heart rate allows us to believe that this runner has more motivations than an individual fitness runner making us apply the individual competitive runner profile to this individual.

After looking through the data of the runner, there are a few points of improvement that we would like to suggest improving performance and longevity in the sport. It is observed from the run that during the middle section of the run (from ~6.63km mark to the ~29.41km mark) the runner begins to run at intervals of 1-3km between the speeds of ~18 km/h to ~10 km/h. This lack of consistency in the run can be attributed to the initial burst of speed which can be seen between the start and 6.63km mark. This

could be a product of the runner's competitiveness, curiosity in how far they can push themselves, as well as excitement. An area of improvement for a better run will be tracking the speed of the runner, especially during the beginning of the run. Instead of using a lot of energy at the start, we would recommend starting at a slower speed and then slowly build up to the desired speed later into the race in order to conserve energy for the entire marathon [6]. The goal of the runner in this case is consistency, therefore instead of a downward trendline in speed (which can be observed in the graph) we would like to see a slow increase to plateau in the trendline. Running at a consistent speed instead of 1-3km intervals of high and low speed. Another area of improvement we would recommend for improvement in longevity is the tracking of heart rate. In Jane Chertoff's 2018 article "Running heart rate: What's safe and what's too high?", it was found that for runners 20-40 years old, the healthy range is between 120-160 Bpm [3]. As this runner begins the run by working at a higher rate, the Bpm for the first part of run goes above the recommended heart rate. Like the first point, but more on the concern of longevity in the sport, we would recommend the runner to be mindful of their heart rate, especially at the beginning of the run. There is a fear that stressing the heart can not only be harmful to the performance of the run, but also in the longevity of the runner as the heart will be overworked. To improve performance as well as longevity, we would recommend the runner to have a shift in perspective of their goals during the run. Instead of pushing themselves to run at their limit throughout the whole marathon, they should focus on controlling their speed and work rate, allowing the runner to performance to be more consistent and allowing the runner to develop their endurance and speed rather than pushing themselves to the limit leading to injury.

Reflection of Assignment 2

In this assignment, we learned how to clean, analyze, and create visuals from a large amount of data gathered from a smart watch which converted it into a GPX file. In Assignment 2, we also made up for the missed explanation of Reiss's theory of desires as well as Vos's motivational profiles in Assignment 1. These theoretical understandings as well as practical understanding of data analysis and filtering will become important when going through Assignment 3.

(More data and information collected for Assignment 2 can be found in Appendix A)

Assignment 3

1. **Profiling the Half-Marathon Team DASU20**

Individual Runners

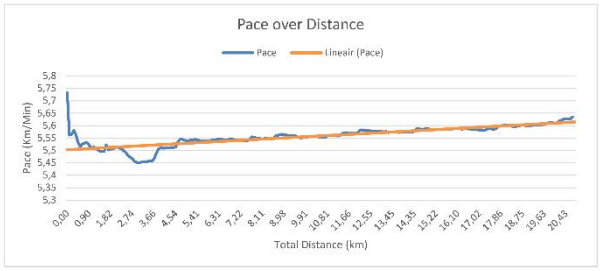
Runner 1

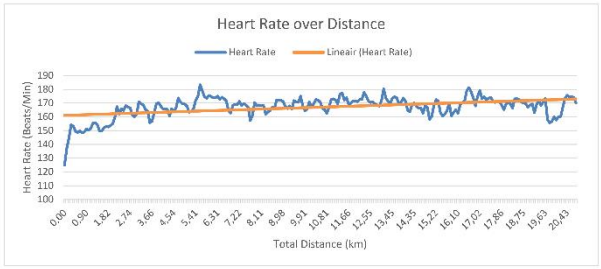
RUNNER 1

Bib number :	Runner Overall Position :	Runner Gender Position:	Runner time :	Pace :	Speed :
#40307	2743/5155	2362/3828	1:56:52	5.32 min/km	10.83 km/h

Runner Accessories:

- **Smartwatch**
- **Sponsored Synthetic-Fabric Running Shirt**
- **Synthetic-Fabric Running Shorts**
- **Ankle Socks**
- **Running Shoes**





Runner Motivational Profile: The runner is a Competitive Social Runner based on their pace in comparison to the other runners in the group as well as in the general half-marathon grouping. The runner is running at a competitive pace; however, they are not running on their own as they finished their run around the same time as Runner 4. The fast beginning shows the competitiveness of the runner; however, they are not using up all their energy in the first half of the run and are able to pick up the pace before the finish. The determination to finish at the higher spot and better time is strong as the pace at the last part of the run is much higher than in the middle. This also leads us to believe the runner is quite experienced as they can keep a reserve for the finish throughout the marathon.

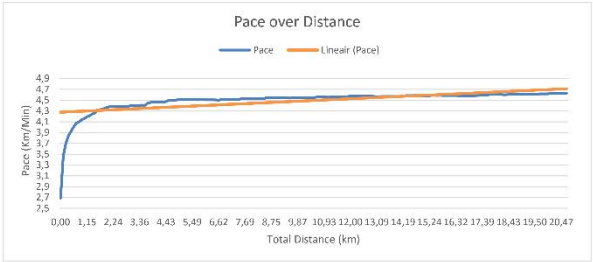
Runner 2

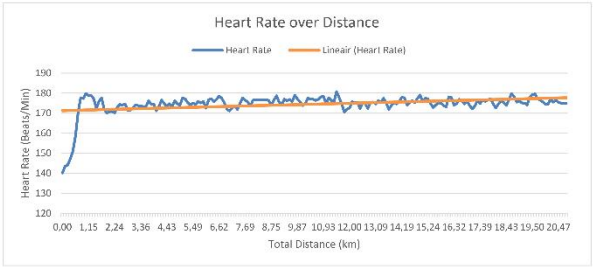
RUNNER 2

Bib number :	Runner Overall Position :	Runner Gender Position:	Runner time :	Pace :	Speed :
#40310	645/5155	609/3828	01:35:57	04:33 min/km	13.19 km/h

Runner Accessories:

- **Fitbit**
- **Headband, sunglasses**
- **Nike running sneakers**
- **Running shorts**
- **TUE shirt**





Runner Motivational Profile: Social Competitive/Fitness runner. The time of this runner is the highest in the team. It is clearly visible from the pace graphs throughout the run that they are in exceptionally good shape. The statistics are very consistent, although differently from some other runners, the pace does not increase in the last part, so the runner did not have saved-up energy for a “strong finish” (relative to the rest of the run, not the whole half-marathon population). The runner does not wear a lot of extra running gadgets but based on the result has done a lot of practice beforehand, which could be an argument for assigning them to the individual fitness runner motivation profile, however, there is not enough data about their training style to determine it.

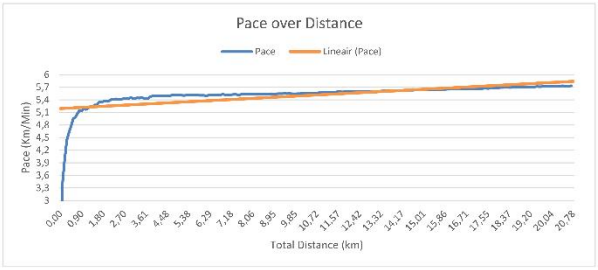
Runner 3

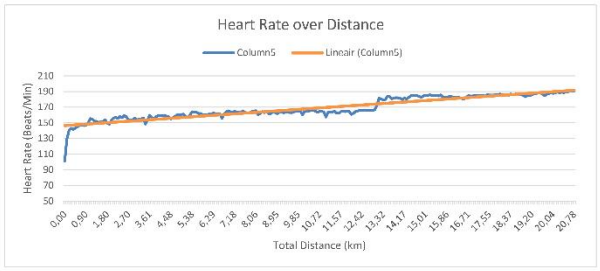
RUNNER 3

Bib number :	Runner Overall Position :	Runner Gender Position:	Runner time :	Pace :	Speed :
#40309	3402/5155	582/1327	01:59:32	05:40 min/km	10.59 km/h

Runner Accessories:

- **Hoortrunners shirt**
- **Running shorts**
- **Sneakers**
- **Garmin smart watch**





Runner Motivational Profile: Social competitive/fitness runner, belonging to a runner's club as indicated by the shirt. The run is quite average within the general population (however worst in the group), with a very steady two-thirds of the half-marathon, where the pace was kept consistent, however as the heart rate increased and energy levels fell, the pace gradually fell in the last half. This can lead one to believe there could be improvements in training to keep enough energy for a good finish.

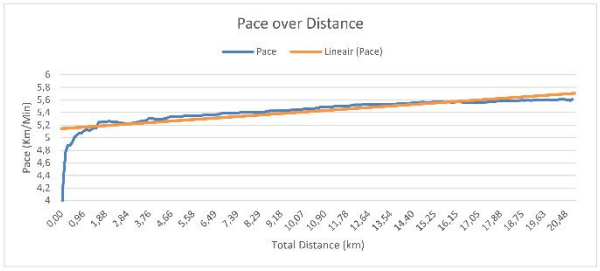
Runner 4

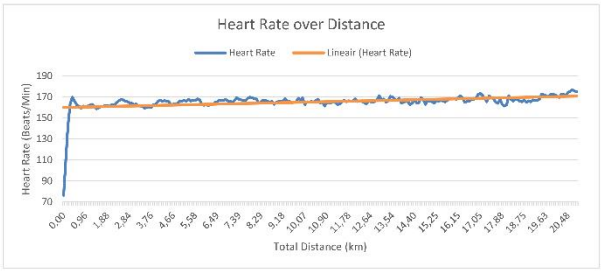
RUNNER 4

Bib number :	Runner Overall Position :	Runner Gender Position:	Runner time :	Pace :	Speed :
#40308	3186/5155	2684/3828	1:56:33	05:31 min/km	10.86 km/h

Runner Accessories:

- **Running shorts, running vest with hydration pack**
- **Garmin smart watch**
- **Sponsored Synthetic-Fabric Running Shirt**
- **CamelBak Backpack with water hose**





Runner Motivational Profile: Social competitive runner, seen as experienced based off their clothing/accessories. The pace of this runner was gradually decreasing throughout the run as the pace built up during the first 5 kilometers could not be sustained. During the middle part of the run, the pace stabilizes, seen as well in the heart rate graph the measurements are the most consistent compared to the beginning or end of the run, where the heart rate changes more drastically. This very stable part of the run shows the runner is experienced and able to keep pace, however the competitive rush in the beginning used up a lot of energy and led to a gradual decline in their pace.

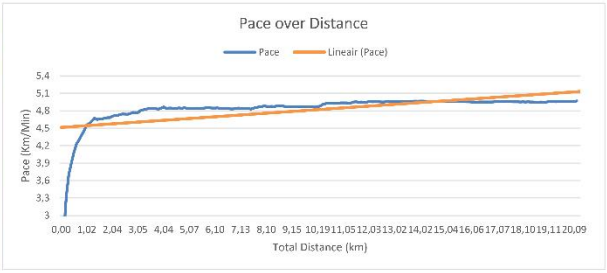
Runner 5

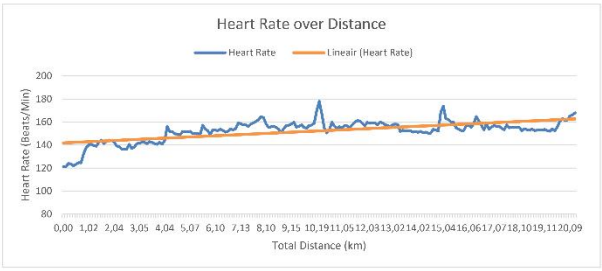
RUNNER 5

Bib number :	Runner Overall Position :	Runner Gender Position:	Runner time :	Pace :	Speed :
#40306	1146/5155	1063/3828	01:43:24	04:54 min / km	12.24 km/h

Runner Accessories:

- **Running leggings**
- **Synthetic fitness tank top**
- **Fitbit tracker on the hand**





Runner Motivational Profile: Social competitive/fitness runner. This runner has a comparingly stable – above-average run, which leads to believe they have a high motivation and experience of running not only with a team, but individually as well. Their pace is one of the most consistent in the team and overall, changing only very minimally. All 4 quarters of the race were completed in similar lengths of time, at the end of the run the pace was almost the same as in the beginning, showing a great preparation of the runner. The heart rate gradually increases as expected and is overall consistent with a few jumps.

Motivational team profile

This group consists of runners of different running styles and performance, this gives it away that it is not a team of elite sporters who usually train together with sporters who are of similar levels. It could be assumed that all the runners have started practicing consistently running separately and did not all train together from the beginning. Belonging to the team in a half marathon and the fact that some runners belong to running clubs, makes it possible to attribute these runners to a type of social runner, as described in Vos' et. al. *From problem to solution: Developing a personalized smartphone application for recreational runners following a three-step design approach* [10]. Social runners find motivation through running together, and social contact is an important part of the activity to them. It satisfies their basic desires for the need of friends and social standing with a need to exercise. Runners of this type require more external encouragement than individual runners, that is why sharing each other's progress and discussing the ways to improve among a group can be an extremely rewarding approach to reach better results in the sport.

It is difficult to determine if all the runners in the group run for social purposes since the information about their individual running habits is not available. While some can clearly be assigned to a runners' club by their clothing, others could also be individual runners if they only joined the team for a marathon and they practice the sport individually and do not hold social contact as their main driver. For example, Runner 2 and 5, who are noticeably ahead of the rest of the group, could be assumed as an individual

competitive runner who does not need social encouragement to train and push themselves. Unlike the rest of the group who finished in a 3-minute window between each other, these two runners ran at their own pace to push themselves on the day of the race leaving the other three runners behind. Runners 1, 3, and 4 could then be assumed as the social runners as they most likely ran together during the race itself.

Next to that, the group can be classified as competitive as they have participated in a half-marathon, however there is an argument for some of them also being fitness runners who run not just to reach the best result possible at the race but for their health benefits and staying in a good physical shape. This could be argued since while the places in the marathon were not the highest, the pace of most runners from the group was impressively consistent throughout the half marathon and the heart rate graphs show they were not pushed to a harmful limit. From this it can be deducted that the runners care about their wellbeing and train to run not only as fast as possible but also safely to not overbear their bodies. In this case, the basic desires of independence, order and physical activity would prevail against Vengeance – the need to compete, and Power – the need for influence of will.

2. Interpreting and Cleaning the Training GPX Data

For assignment 3 it was initially decided to import the GPX files into python and calculate the additional metrics, such as speed and pace, there. This gave the opportunity to filter the data differently. Instead of keeping the data of every 30th second, we could filter the datasets with a gaussian filter. After all data were created in python they would be exported to excel where the graphs would be made with the addition of linear trend lines, something automatically made in excel.

However, this process had issues that had to be addressed. Most importantly, there were time stamps where cadence data were missing, creating gaps in the graphs as seen in Figure 8. This was fixed by interpolating the datasets. Three interpolation methods were tested: linear, pad and polynomial. There was no apparent difference in any of our datasets, so the simpler linear method was thus applied. Additionally, the python code was calculating seven kilometres extra in the total distance, so the more reliable excel formula, that was used in the previous assignment was also used in assignment 3. Lastly, in excel, in order to keep every 30th second data, every data time that is not decided perfectly with 30 seconds was left out. Even though this works, a simpler way was used. This way is python pandas command resample which also offered the additional value to start resampling the data from the actual initial time of the dataset instead of the first that rounds up perfectly with 30 seconds.

So, the final process is as follows. GPX file is imported in python - dataset is resampled every 30 seconds with origin the first time stamp of the dataset (using pandas.DataFrame.resample) - dataset is linearly interpolated (using pandas.DataFrame.interpolate) - dataset is exported as excel file (using pandas.DataFrame.to_excel) - in excel Distance between Time points (=ACOS(COS(RADIANS(90-latitude_0)) * COS(RADIANS(90-latitude_1)) + SIN(RADIANS(90-latitude_0)) * SIN(RADIANS(90-latitude_1)) * COS(RADIANS(longitude_0-longitude_1))) * 6371), Speed (=(total_distance/elapsed_time)*60) and Pace (=elapsed_time/total_distance) are calculated - graphs of heart rate, cadence, speed and pace over elapsed time and distance are made, including the linear trend line of each graph.

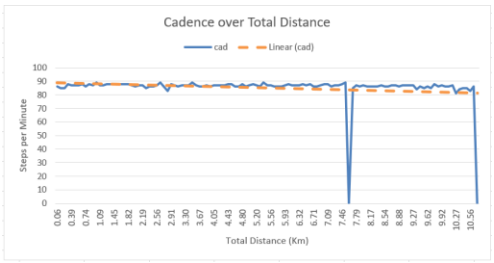


Figure 8a

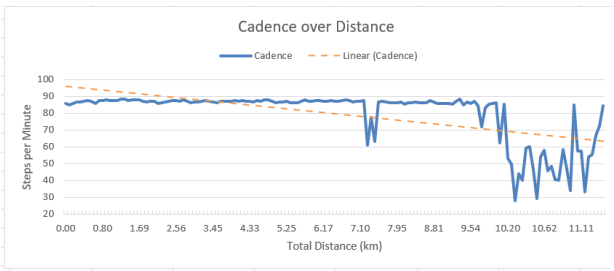


Figure 8b

Figure 8: Examples of what data looks like before interpolation (Figure 8a) and after interpolation (Figure 8b)

3. Visualizer Reasoning

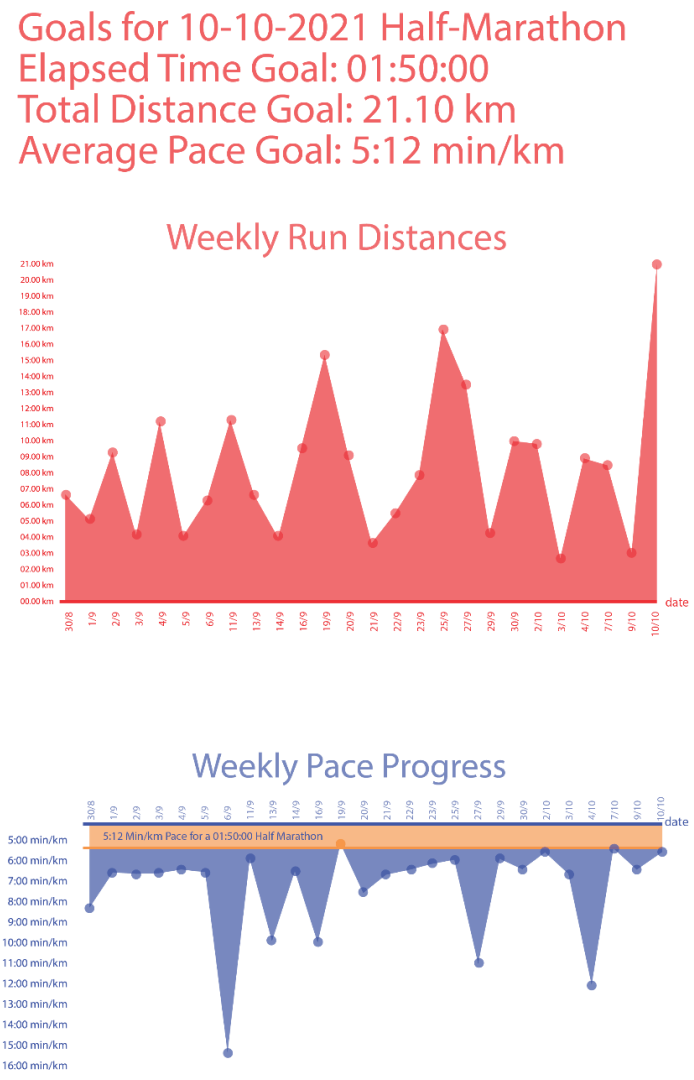


Figure 9: A general outlook of the progress of the runner between runs.

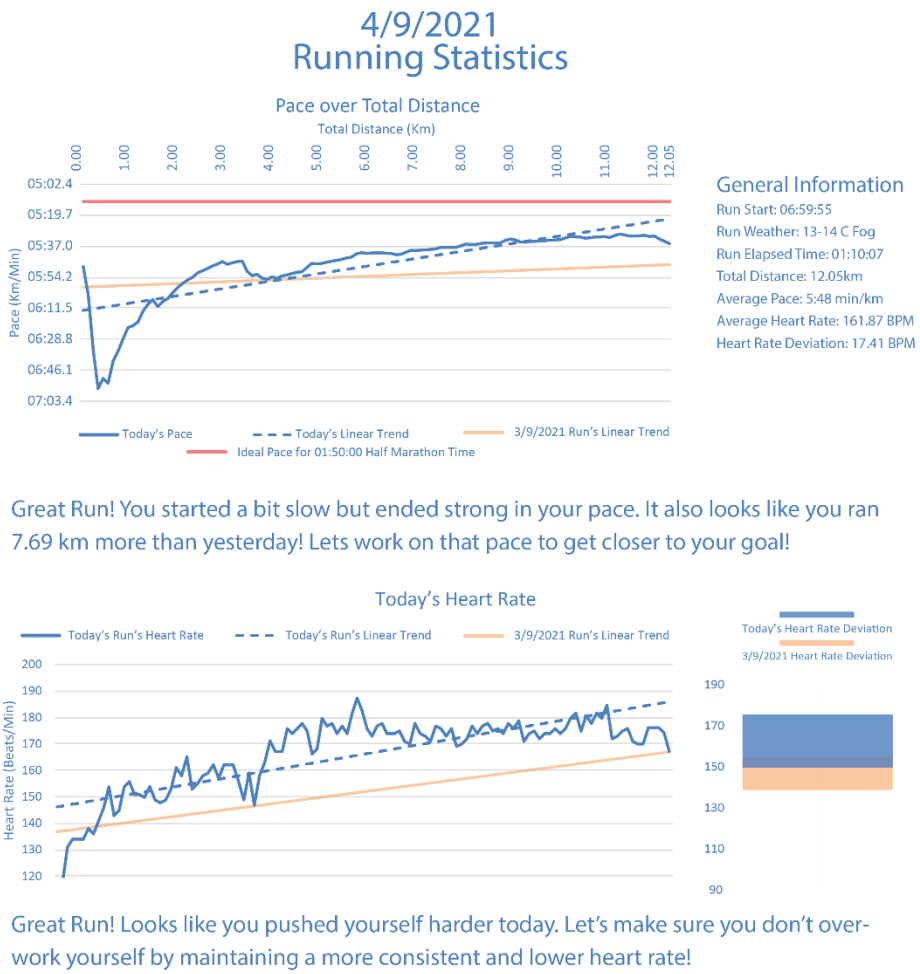


Figure 10: A detailed outlook of a specific run, in relation to previous runs and running goal.

The visualizer is a tool which will allow Runner 1's team to help keep track and motivate Runner 1 to improve on their half-marathon training. The visualizer is meant to provide information to non-professional runners which requires a respect of privacy as well as a simplification of data to efficiently inform the non-professional runners of one's progress. However, this visualizer also appeals to the competitive runner as well as social runner. As the idea for improvement is based on progress towards a runner's goal, this will appeal to the competitive runner profile who has a need for vengeance, power, physical activity, and independence [8]. If one is a social runner, then the ability to comment on the data, as seen with the text boxes, will allow team members (or whoever the runner decides to share their information with) to give encouragement as well as feedback on the training. This feature will appeal to those who have the basic desires of curiosity, physical activity, social contact, and tranquility depending on the experience of the runner [8]. An additional point for this visualizer is to encourage competitiveness not through the competition of each other, even though that can be a great motivator especially for a social runner profile, but through competition of themselves. In other words, trying to push oneself as far as their body can go without causing injury. This is because there is a range of experience and conditioning within the team which means every runner's experience and progress will be different. So

having the visualizer focus on the goals of a single runner can help concentrate advice in relation to the abilities of that runner rather than giving advice which can give the runner the risk injury.

In this visualizer, it starts with a screen which displays the goal of Runner 1 (which is set by Runner 1), a tracker of the runner's distances, as well as a daily progress report of the runner's pace (Figure 9). The runner's goal would be set by the runner who chooses the distance they are training for as well as the time they would like to finish that distance in. From those two inputs, the ideal pace will then be set for the runner to hit. The goal would then be for the runner to try to meet that pace and run consistently at that pace or faster during their training.

The following graph displays the runner's distances per run. This graph shows the runner, as well as their team, what kind of training the runner is going through (if they did a speed run, long run, or recovery run) [4]. To get a better idea of what kind of run the runner went through, the last graph shows the average pace of each run. This graph should be used in tandem with the distance graph to track the runner's progress in both endurance and speed. Figure 9 will provide a general glossary overlook of Runner 1's progress in their training.

Figure 10 could be accessed when one selects a specific run to get more specific data. On this page, what can be seen is the detailed pace of the run (as well as a trendline of the run, trendline of the previous run, and the ideal pace to hit the target time), general information of the run (When the runner started the run, weather during the run, elapsed time of the run, total distance, average pace, average heart rate, and heart rate deviation), heart rate during the run (as well as trendline of the run and trendline of the previous run), and heart rate deviation.

A more detailed pace of the run is provided to give more info about how the run progressed. For example, the one in Figure 10 could be seen as a good run because the runner over time kept incrementally improving their pace ending faster than they began which is something that is encouraged in a half marathon [1]. In this graph, one can also see the pace trendline from a previous run as well as the pace goal in order to achieve Runner 1's ideal finishing time. This is to track the progress of the runner from their previous run as well as in comparison to their final goal.

The general information section gives more information about the run itself. Such as the starting time, weather, elapsed time, total distance, average pace, average hear rate, and heart rate deviation. The starting time and weather (for the visualizer, the information for the weather of that day was found using [9]) were found important as they can give the context for the run (if they usually ran at night and changed it to a morning run, or if the weather was good or bad). This data would be taken from when the gpx data to give more accurate information about the runner's location and time of run. It is on purpose to only show the temperature and weather condition of the run, and omit the actual location of the run, in order to respect the privacy of Runner 1. However, it was included to provide more context for a run (such as if it were harsh weather, it could explain why the runner ran a shorter distance that day). The elapsed time, distance, and average pace are general information which gives context clues on what kind of run it was (speed run, long run, or recovery run). Heart rate and heart rate deviation is data which is more used to track the general health of the runner. Making sure that they are not overworking themselves during the run, however also tracking if they are properly training their heart for the half-marathon.

The Heart-Rate graph as well as box plot of the heart rate deviation is also included to keep track of the runner's condition during their run. A more detailed look at the progression of the runner's heart rate was found important as it can provide information about how the body is reacting when the runner is pushing themselves. The trendline for their previous run's heart rate can also be seen to give the runner an idea of how their current run compared to their previous run. As a half-marathon is a practice in endurance, the

heart rate deviation is also displayed in a box plot to try to improve the runner's ability to maintain a consistent heart rate. If there is a large deviation in heart rate, that means the runner could be over-exerting themselves during the run and should re-evaluate their training regime. If the heart rate is consistent, within a 10% deviation of their max heart rate [2], then it could be concluded that the run was at a good pace. Any larger deviation would then lead to questions about the intensity of the training as well as the possibility of injury from the training. A boxplot of the previous run's heart rate deviation is also provided to give more context on the progression of the runner's training. Overall, these two graphs are meant to keep track of the condition of the runner making sure that the runner is training at a productive and safe pace.

The two text boxes will give messages and tips to the runner based off their current run, in comparison to previous runs as well as their goal. It is also in this section where the rest of the team can type in more messages of encouragement as well as training advice for Runner 1.

The visualizer is built on a series of layers where if one wants to find more information about a run they could. We understand that the runners might not be a high-level athlete which is why we omitted information such as cadence. Speed was omitted as it felt like redundant information when the visualizer is measuring pace. Other information such as elevation and specific longitude and latitude coordinates were omitted in order to protect the privacy of the runner. The addition of comments helps the user immediately see if whether the run was good or not, since for a non-expert in the field the graph might not be immediately clear if they only wish to quickly check on the peer's progress. These choices hopefully give the visualizer the balance of enough information for one who is a casual yet dedicated and motivated runner rather than an elite athlete performing at a top competitive level.

Altogether, the visuals should communicate to Runner 1 and their team the general progress of Runner 1 in comparison to their goal (Figure 9). If there is the need for more detailed information, that information would be available through selecting a specific run where then one can find a more in-depth analysis of that specific run, in comparison to previous runs and their goal. With these two pages, we hope to find a balance of providing enough information for advice and the tracking of progress while respecting the privacy of the runner.

(Codes for the data used in the visualization can be found in Appendix B. The visuals were created in Adobe Illustrator)

F.A.I.R.

To comply with the principals for research objects applied to human and machine-driven activities, [11] we need to check whether the (meta) data is handled the right way and is suitable for the use:

To be **findable**:

F1 - (meta)data are assigned a globally unique and eternally persistent identifier - Each of the datasets have a special identifiable number: Runner 1-5, Activity 1- 25.

F2 - data are described with rich metadata – All types of data are classified and named accordingly, with descriptions relative to the research

F3 - metadata clearly and explicitly include the identifier of the data it describes – the top row clearly states what data is represented below, with the units of measurement included.

F4. (meta)data are registered or indexed in a searchable resource – the tables of data are formed using the same order and format, with time and distance ascending.

To be **accessible**:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol - GPX files can be retrieved by a variety of commonly used programs.

A1.1 the protocol is open, free, and universally implementable - programs used for access and analysis of data are free and available for intended audience – Excel, Python, GPX studio.

A1.2 the protocol allows for an authentication and authorization procedure, where necessary - no special authentication or authorization is necessary for this format.

A2. metadata are accessible, even when the data are no longer available - Metadata is listed in this report page: 7

To be **interoperable**:

I1. (meta)data use a formal, accessible, shared, and universally applicable language for knowledge representation – metadata is presented in a formal language, familiar Excel tables and graphs, use of Python code – universal coding language.

I2. (meta)data use vocabularies that follow FAIR principles - names for the metadata are dedicated in a way it is not possible to mistake or misinterpret the data

I3. (meta)data include qualified references to other (meta)data - one column of data contains references to another relevant set and it is easy to trace of data (formulas in excel for example).

To be **re-usable**:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes – all the relevant data that is not captured by sensors is added through formulas (such as speed and pace), graphs and observations are added to the report.

R1.1. (meta)data are released with a clear and accessible data usage license – the data is retrieved through the university course page and is available only to the students and assessors.

R1.2. (meta)data are associated with detailed provenance – the origin of the data is described in the report and only used for the intended purpose.

R1.3. (meta)data meet domain-relevant community standards – the data has been accessed and analyzed with tools regularly used for such and similar tasks.

Conclusion

Through this assignment we developed a methodology for designing products and services based on data analysis. We started with an investigation into the users - in this case runners - to figure out what kind of visualization would fit their profile. Eventually we assigned a social competitive profile to the group of runners, with a fitness motivation as they can also be assumed to exercise for health benefits. For this profile, we created a visualization that can be the most useful to reach the runners' goals – overall and specific runs portrayed through graphs for maximum insights. The individually sets the goals and provides running data to keep track of the progress while a comment section is available to help keep the runners social and connected with each other. During the design process of visualization as well as the analysis of given data, we constantly checked our decisions with the F.A.I.R. principles and the Ethical code of conduct. Since this project is made for educational purposes and will not be published publicly,

the ethical rules are not as strict and some identification is possible to include, such as runner number and results to help the assessor grasp the information faster.

One of the biggest challenges to be able to review the data was to clean and interpolate it without losing insightful information. Choosing which data is valuable and what is excessive and unnecessary is a core decision as the implementation of visuals directly depends on it. For further improvement of the visual system, it is possible to look deeper into UI psychology and its influences as well as what kind of devices the program could be adapted for. In addition, more insights into runners and user testing would be required to create such improvements. The value of this process and the result can add to further improvements of designing support systems for sporters as these systems can be applied to many kinds of sports and types of sporters that may have not yet found a suitable program to track their progress.

References

[1] A.CE-C.PT Christine Luff. 2021. What affects time for running a half-marathon? (February 2021). Retrieved January 20, 2022 from https://www.verywellfit.com/what-is-a-good-time-for-running-a-half-marathon-2910898

[2] Heather Mayer Irvine. 2019. Everything you need to know about heart-rate training. (February 2019). Retrieved January 20, 2022 from https://www.runnersworld.com/uk/training/beginners/a760176/heart-rate-training-the-basics/

[3] Jane Chertoff. 2018. Running heart rate: What's safe and what's too high? (September 2018). Retrieved December 15, 2021 from https://www.healthline.com/health/running-heart-rate

[4] Nike. 2022. Half-marathon training plan. (2022). Retrieved January 20, 2022 from https://www.nike.com/running/half-marathon-training-plan

[5] MYLAPS. MYLAPS Sporthive event results. Retrieved December 15, 2021 from http://results.sporthive.com/events/6850745558983373568

[6] Parker Northrup says: Tomb Says: and world_runner says: 2017. Marathon race strategy: The perfect strategy for the marathon. (September 2017). Retrieved December 15, 2021 from https://runnersconnect.net/coach-corner/marathon-race-strategy/

[7] Python Software Foundation. 2022. Gpxcsv. (2022). Retrieved January 22, 2022 from https://pypi.org/project/gpxcsv/

[8] Steven Reiss. 2002. Who am I?: The 16 basic desires that motivate our behavior and define our personalities, New York, NY: Berkeley Books280.

[9] Time and Date AS. 2022. Past weather in Eindhoven, Netherlands - September 2021. (2022). Retrieved January 20, 2022 from https://www.timeanddate.com/weather/netherlands/eindhoven/historic?month=9&year=2021

[10] Steven Vos, Mark Janssen, Jos Goudsmit, Coen Lauwerijssen, and Aarnout Brombacher. 2016. From problem to solution: Developing a personalized smartphone application for recreational runners following a three-step design approach. Procedia Engineering 147 (2016), 799–805. DOI:http://dx.doi.org/10.1016/j.proeng.2016.06.311

[11] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Appendix A

Additional Resources

- Jens Jakob Andersen. 1970. Marathon statistics 2019 worldwide (research). (August 1970). Retrieved December 14, 2021 from https://runrepeat.com/research-marathon-performance-across-nations

Average marathon Stats (UK)

This data is somewhat irrelevant since the Netherlands produce, on average, the fastest marathon runners in the world.

- Jane McGuire. 2020. How does the average runner pace a marathon? (October 2020). Retrieved December 14, 2021 from https://www.runnersworld.com/uk/training/marathon/a27787958/average-marathon-finish-time/)
- Pace: 6.43 minutes per kilometer, Speed: 9 km/h for men. Pace: 7.26 minutes per kilometer, Speed: 8.1 km/h for women
- Total Time: Men (UK)– 4:23:27, Women (UK) - 5:00:27

Male Runners Pace (UK):

5K: 5:14 min/km, or 11.45 km/h

10K: 5:16 min/km, or 11.4 km/h

15K: 5:18 min/km, or 11.32 km/h

20K: 5:37 min/km, or 10.69 km/h

25K: 5:48 min/km, or 10.34 km/h

30K: 5:55 min/km, or 10.14 km/h

35K: 6:22 min/km, or 9.42 km/h

40K: 6:30 min/km, or 9.24 km/h

Female Runners Pace (UK)

5K: 5:51 min/km, or 10.26 km/h

10K: 5:55 min/km, or 10.14 km/h

15K: 6:14 min/km, or 9.63 km/h

20K: 6:21 min/km, or 9.46 km/h

25K: 6:32 min/km, or 9.19 km/h

30K: 6:51 min/km, or 8.75 km/h

35K: 7:02 min/km, or 8.54 km/h

40K: 7:04 min/km, or 8.49 km/h

Running Positions in Eindhoven Marathon

Results 2021 from (http://results.sporthive.com/events/6850745558983373568)

- 1st place (Total) Sex: Male, Speed: 20 km/h, Pace: 3 min/km, Time: 2:06:32
- Middle of the pack (Total 763) Sex: Male, Speed 11.01 km/h, Pace 5:27 min/km, Time, 3:49:56
    - 5km: 5:16 min/km, or 11.4 km/h

      10km: 5:13 min/km, or 11.5 km/h

      15km: 5:13 min/km, or 11.5 km/h

      20 km: 5:03 min/km, or 11.9 km/h

      25km: 5:19 min/km, or 11.3 km/h

      30km: 5:24 min/km, or 11.1 km/h

      35km: 5:53 min/km, or 10.2 km/h

      40km: 6:04 min/km, or 9.9 km/h

      42.19km: 5:53 min/km, or 10.2 km/h

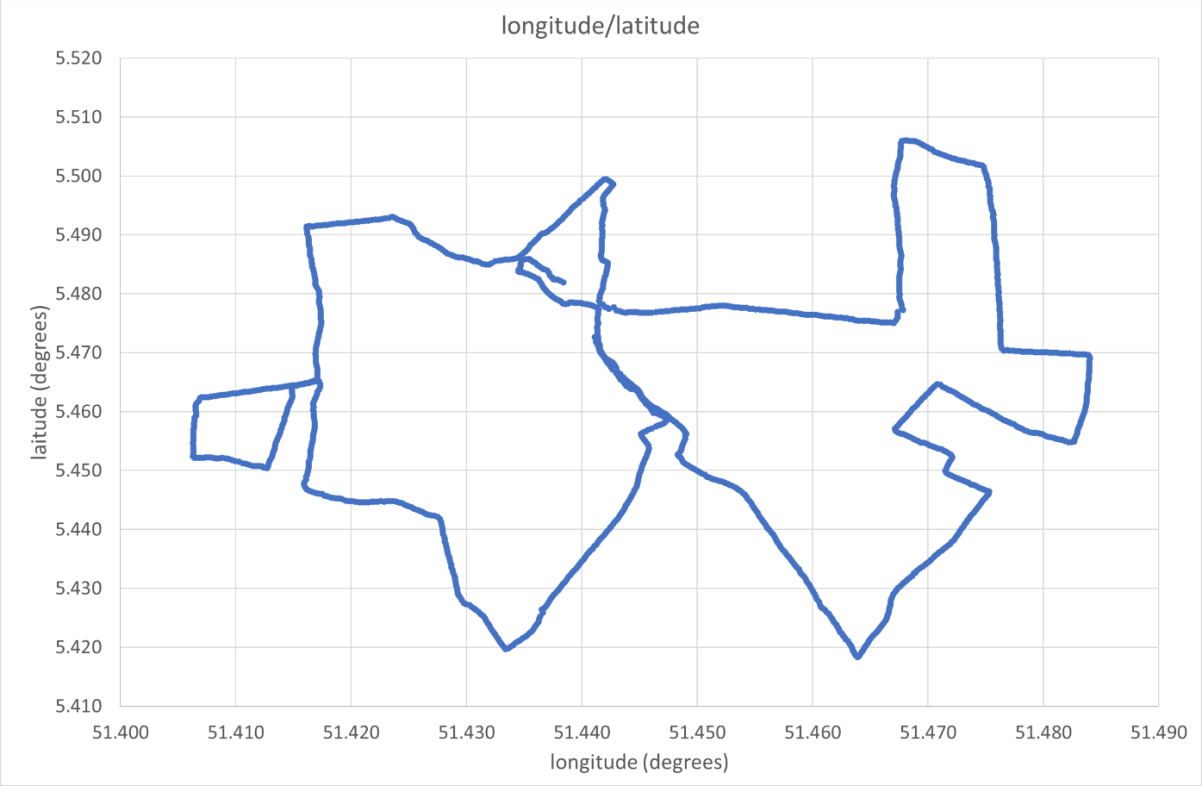
1. Last Place (Total 1526) Sex: Female, Speed 6.92 km/h, Pace: 8:40 min/km, Time: 6:09:51

Running positions in Amsterdam Marathon

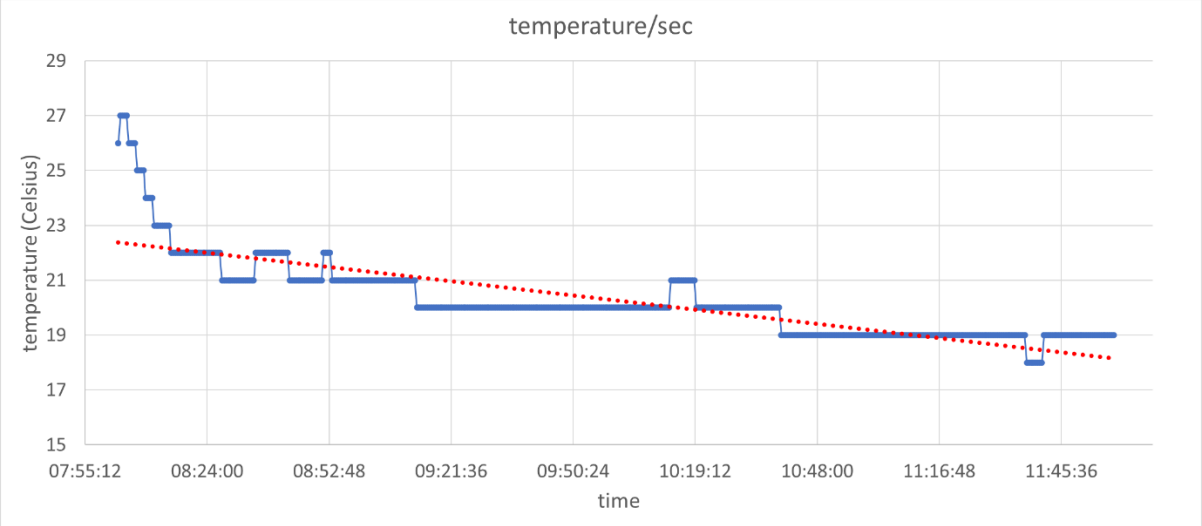
Results 2021 from (https://results.sporthive.com/events/6853741552290238720/races/479865):

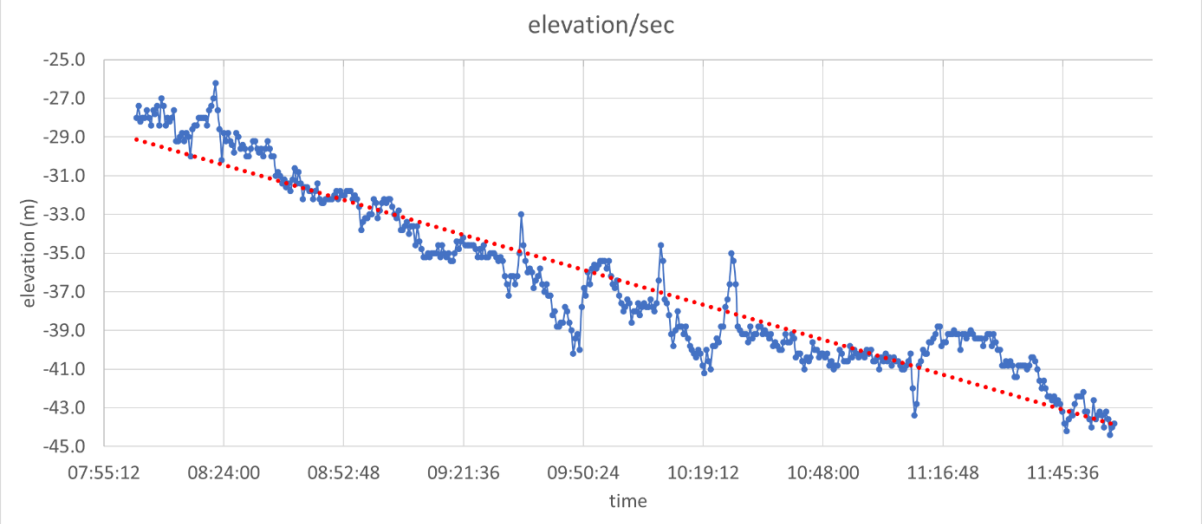
- 1st place Speed 20.48 km/h, Pace 02:56 min/km, Time: 02:03:38
- 40th place Speed 17.74 km/h, Pace 03:23 min/km, Time: 02:22:42
- 2880th place Speed 11.19 km/h, Pace 05:22 min/km, Time: 03:55:00
- 6612th place Speed 6.24 km/h, Pace 09:37 min/km, Time 06:45:58

Graphs from Excel


longitude/latitude
5.520
5.510
5.500
5.490
5.480
5.470
5.460
5.450
5.440
5.430
5.420
5.410
laitude (degrees)
51.400
51.410
51.420
51.430
51.440
51.450
51.460
51.470
51.480
51.490
longitude (degrees)


Speed over Distance
25
20
15
10
5
0
Speed(km/H)
Total Distance
4.602
9.213
13.720
18.270
22.848
27.359
31.791
36.026
40.145
Distance Ran (km)





Appendix B

Python code used to interpolate data and create graphs as well as visuals [7].

```
import numpy as np  # import auxiliary library, typical idiom
import pandas as pd  # import the Pandas library, typical idiom
import openpyxl as Workbook
from statsmodels.distributions.empirical_distribution import ECDF
from scipy.interpolate import interp1d
from scipy.ndimage.filters import gaussian_filter1d
from scipy import stats
```

```
# next command ensures that plots appear inside the notebook
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()  # set Seaborn defaults
plt.rcParams['figure.figsize'] = 10, 5  # default hor./vert. size of plots, in inches
plt.rcParams['lines.markeredgewidth'] = 1  # to fix issue with seaborn box plots; needed after import seaborn

# source: https://pypi.org/project/gpxcsv/
pip install gpxcsv

from gpxcsv import gpxtolist
gpx_list = gpxtolist('datasets/activity_7415960856.gpx')

import pandas as pd
df = pd.DataFrame(gpx_list)

#format: '%Y-%m-%dT%H:%M:%S.%fZ' or '%Y-%m-%dT%H:%M:%SZ'
df['time_simple'] = pd.to_datetime(df['time'], format='%Y-%m-%dT%H:%M:%S.%fZ')

df.head()

df2 = df.set_index('time_simple').resample('30s', origin='2021-09-02 18:24:18').mean()
Df2

df2.replace(0, np.NaN, inplace=True)
df3 = df2.interpolate()
Df3
```

```
#df2.to_excel (r'runner_(Runner number)_res.xlsx', index = True, header=True)

df3.to_excel (r'activity_(Run Number).xlsx', index = True, header=True)
```