Enhanced mood tracking using a smartphone app

Gaétan Moisson-Franckhauser
MSc in Digital Media Engineering
s192765@student.dtu.dk

Tanguy Navez
MSc in Digital Media Engineering
s192702@student.dtu.dk

ABSTRACT

Many applications exist to track the evolution of mood. These applications often lack consistency at the level of the Preparation and Action stages, and offer little flexibility to take advantage of the data acquired. This involves improving the design of the inputs and visualization, by evaluating the performance of several models through a qualitative study carried out with potential users. In this perspective of maximizing the acquisition of data in order to benefit from it, without annoying the user, the best frequency of input requests has been deduced from an experiment. Finally, a new personalized recommendation feature allowing the user to go further in the Action phase was tested. We present our findings in this paper, in the light of which we have developed a MVP.

KEYWORDS

mood, self-tracking, recommendations, smartphone app

1 INTRODUCTION

Analysis

The mood is a difficult variable to track; ever-changing, depending on both external and internal factors, and not always triggered by a specific event but by a collection of factors. Mental disorders can affect men and women in different ways, with for instance anxiety and depression being more common among women, and certain types of disorders are even exclusive to a gender [5]. It is also needed to differentiate how mental disorders can affect different age groups, with around one-quarter of children affected by anxiety disorders, depression and attention deficit hyperactivity disorder (ADHD) [6]. All these mental conditions are linked to mood disorders, which are commonly tracked through clinical interviews and paper surveys. However, these methods lack precision regarding the mood evolution, due to recall bias and compliance issues, as well as they are limited on the quantity of data they can summarize [4].

The industry and researchers are now using new technologies in order to give a better measurement of the mood and mental health: the need for this kind of tools has lead to an emergence of mobile apps to track the user mood, with hundreds of prototypes available on the market [3]. Supporting self-management of mental health and giving the opportunity to track one's mood evolution through this kind of apps can help individuals monitoring their well-being, and gives new insights on potential conditions. Diving into the mood-tracking apps currently available on the market reveals that most of them support thoroughly the Collection and Reflection stages, but have an incomplete support for the Preparation and Action stages, resulting in a lack of effective use of the data gathered and a lack of concrete goals for the users using them [1]. From this perspective, we were interested in studying how to design an app that would offer new functions for the Preparation and Action stages, with more flexibility on the data gathered and better insights for the user with personal recommendations.

Background and Prototype

Our objective, in order to extend the support of the Preparation stage, was to give the opportunity to the user to choose what data would be gathered and used by the app in the mood-tracking process, and get a better understanding as to how the tracking would be done. By integrating other apps, we wanted to allow the user to be able to use other data than just the mood input, and have this data automatically collected and processed by the mood-tracking app for the Reflection and Action stages.

To extend the support of the Action stage, our idea was to expand on the personal recommendations feature: using the additional data provided by integrated apps, coupled with a machine learning algorithm, we would be able to provide tailor-made advice to the user.

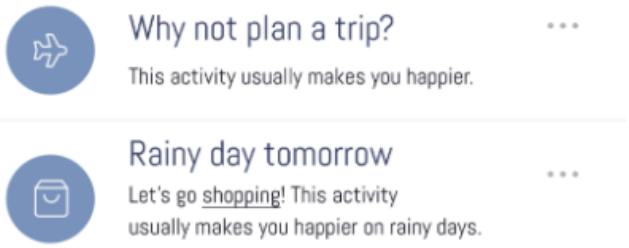


Figure 1: The two examples of recommendations shown to participants

In order to test these new features that would extend the Preparation and Action stages, we created a prototype for a mood-tracking app, simply called Mood. The original mock-up for the app was created on Figma [7], a tool that allowed us to test different designs for the final app. A user study was then conducted in order to evaluate the best designs for

Type	Description	
Evaluation	Type	Inputs
	Form	Input designs Visualization designs Recommendation examples
Experiment	Type	Features
	Within subject	IV: Notification frequency DV: Input frequency

Table 1: Methods summary

the input and visualization parts of the app. This user study is detailed in the following sections and Appendix 6.

For the input, we compared the solutions offered by other mood-tracking apps on the market, and noted that using the phone notifications was the best way to get the user to input data regularly. We also considered that notifications can be considered invasive if repeated to often; thus we needed to find the ideal default frequency for the notifications, in order to get regular mood input without affecting the user. This is the object of our experiment on input frequency.

Finally, for the personal recommendations, we had to take into account that the use of integrated apps data might lead for example to a lack of trust regarding privacy by the app users [2]. With multiple potential points of view on this issue, we decided to conduct a user study to gather potential concerns regarding this feature.

2 METHODS

We conducted an experiment to determine the frequency of input requests, as well as evaluations on input and visualization designs. The implementation of a recommendations feature was also addressed through a short evaluation. A summary of the different methods is displayed in table 1.

User study (Appendix 6):

This study is a three in one evaluation. Its objective is to test the input and visualization designs, as well as to conduct a pre-study on the value of implementing personal recommendations. In the best case, we would have liked to conduct this study directly with the users by letting them test our different prototypes in order to collect their impressions live without having to ask directed questions. Because of the current viral situation limitations that don't allow us to easily meet live users, we decided to use a Google Form. We thought we would get more answers to our form by sending a longer one instead of three different.

The questions on the form are mostly questions that invite the participant to indicate an order of preferences between different design proposals and then ask the participant to justify his choices and propose his ideas. We have limited the size of the form so the participant can complete it in about ten minutes. Our sample for the design evaluation was composed of 13 participants, all in the 20-30 age range; for the evaluation of the personal recommendations concept we only got feedback from 7 out of these 13 people, the questions related to this feature having been added to the Google Form later.

For the notification design, four different features were tested:

- Incorporating a slider in the notification, allowing the user to give a mood input without having to open the app.
- Adding a number indicator on the slider, to display the precise input the user will submit.
- Initializing the slider position on the last input given by the user.
- Displaying a written information with the last input value and time.

Each step of the form was made of a choice between two designs: one with a feature from this list, one without said feature. An additional field let the individuals justify their choice.

For the visualization design, four designs were tested (figure 3):

- a design inspired by a bar plot with numerical information on top
- a line plot design
- a mix of a bar plot and mood rates plot
- a radar chart

The form asked the users to select the design they considered as the best for the visualization purpose described, then justify their choice by comparing the different designs.

Finally, the part of the survey regarding the personal recommendations featured two examples of recommendations that the app could provide. The set of three questions that followed questioned the individual on how they feel about the examples given, and the idea of personal recommendation created through app integration.

Notifications and inputs frequency experiment:

In order to assess the right frequency of notifications asking for mood input from the user, we conducted an experiment consisting in comparing the number of inputs from the user to the frequency of notifications received. Several notification frequencies were tested over periods of 5 days. We thought that these 5 days periods were a good compromise between the time available for this project and the constraint of having enough time for a semblance of habit to develop in the participant. We first tested a notification frequency of one notification per hour, and proceeded to move on to less notifications per day as described in the Results section.

For this purpose, we created a test application running on Android. The application is connected to a server and manages the registration and connections of the users, so that their mood inputs can be saved. Coding our own application allowed us to change the frequency of notifications easily so that we were able to test several of them. Because of technical constraints (downloading the application directly via an APK without using the app-store) and the difficulty of finding people willing to test themselves over 25 days, the experiment was conducted only on ourselves.

3 RESULTS

Evaluation of input design:

The user study regarding the design of the input showed no difference in opinion depending on the gender. Results regarding the age can't be qualified as the maximum age difference was only of 4 years.

Summarizing how the answers classified the 4 features tested, the following ranking is obtained (from "most needed" to "least needed"):

- Input directly in the notification
- Number indicator
- Default position of the slider on last input value
- Last input value and time written

The vast majority of the answers classified the "Input in the notification" as being the most needed addition (85% of them). It was also classified a better choice than "Input in the app" by 92% of the interviewees. The respondents qualified the input in the notification as *"way quicker and intuitive"*, *"practical and faster* [than the input in the app]*"* and *"a better call to action"*. Overall, the most important reasons mentioned in favour of this option are quickness, easiness of use and intuitiveness. The only interviewee who voted against the input in the notification mentioned that they did not allow apps to display notifications, considering them *"too invasive"*.

The number indicator, showing the exact rate input about to be given by the user, was considered as a better option than the absence of a number indicator by 62% of the respondents, and considered as the most needed feature by 2 respondents. Those in favour of the number indicator notably described it as *"better to have a quantitative value for the user and thus compare with previous evaluations"*, *"more precise"* and *"more clear"*. Those against the presence of a number indicator consider its absence as "more refined" from a design perspective, and half of them mention that the don't want/need/should put a number on their mood, as they consider it as *"subjective"*.

Lastly, for the "last input information", only the results regarding the "written last input" show a tendency: 69% of the interviewees consider that the design of the notification is better without a written last input information, reaching a percentage of 80% among the respondents who considered that there shouldn't't be a number indicator. Regarding the default position of the slider in the notification, 54% of the interviewees preferred a default position set at 50 (middle of the slider) instead of the last input. This opinion is mostly held among those who chose the presence of a number indicator as the best design option (62,5%), but the margin is thin.

To summarize these results, Figure 2 showcases the ideal design according to the survey results: a slider in the notification, with a number indicator, a default position at 50 and without a written "last input information".

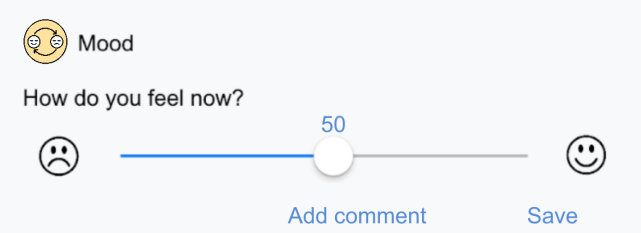


Figure 2: Ideal input design, according to the survey

Evaluation of visualization design:

The user study regarding the design of the visualization showed a small difference in opinion depending on the gender: Design C was considered the best design within males interviewees (50% of them voted for it), whereas Design B was considered the best design by female interviewees (60% of them voted for it). Summing all the votes, Designs B and C are considered the best ones with five votes each, while Design A gathered two votes and Design D one.

Asking the surveyed persons to rate the different designs on a scale from 1 ("Not suited for the visualization") to 5 ("Perfect for the visualization") led to the validation percentages showed in Figure 2: none of the designs got a validation rate of more than 65%, with Design B getting the best rates (4 interviewees out of 13 rated this design 5/5).

Design	A	B	C	D
Validation %	48%	65%	62%	27%

Table 2: Visualization designs validation, according to the survey

Design B, the line plot design shown in Figure 3 was the most validated design by the interviewees. It was described as *"easier to read and understand"*, thanks to how it shows the *"evolution over time"*. It is also considered *"completely intuitive"* by the interviewees who selected this design as the most suited for the visualization. Suggestions given to make the design better were mostly linked to how the design was presented in the form (on a white background, and with small

texts); the only one linked to the design itself pointed out that smileys could be used. An interviewee pointed out that having a smooth line wouldn't be possible without enough input from the user, thus making the mock-up of Design B unreasonable.

Design C, the mix of a bar plot and mood rates plot, was described in the survey answers as *"clearer"* and *"less complicated"* than the other designs. An answer suggested making the average across weeks optional, by replacing it with the average per day. Some alternatives to represent hourly data were also suggested, such as *"a mix between bar diagram for a daily average and the curves of design B for hourly input instead of the dots"*.

Design A (the design inspired by a bar plot with numerical information on top) was described as *"clear"* and *"easier to visualize the event and to rank the days"*, while Design D (the radar chart) was described as *"easier to compare your results during the current week with your average well-being"*. An interviewee pointed out that the non-linearity of Design D was *"troubling"*.

Some additional comments suggested to mix some of the designs presented, namely B and C on one hand, and C and D on the other hand.

Figure 3: Ideal visualization design, according to the survey

Notifications and inputs frequency experiment:

The results of the experiment are displayed in the Table 3 and Figure 4. The initial test of hourly notifications gave us an average number of inputs per participant below 4/day (in addition to being very annoying for the participants). It allowed us to refine our scope for the following 5 days sessions to frequencies between 1 and 4 notifications per day.

There is a big difference in the number of inputs in a day when notification frequency is increased, so there was indeed a choice to make on the frequency of notifications in order to maximize the number of input without being too annoying for the user.

	Frequency	Number of inputs		
		Participants		
		Gaétan	Tanguy	Mean
	1/hour	3,21	3,8	3,50
	4/day	2,9	3,21	3,06
	3/day	2,4	3,0	2,70
	2/day	1,8	2,2	2,0
	1/day	1,0	1,6	1,3
Mean		2,26	2,76	
SD		0,88	0,86	

Table 3: Notification and input frequency experiment results

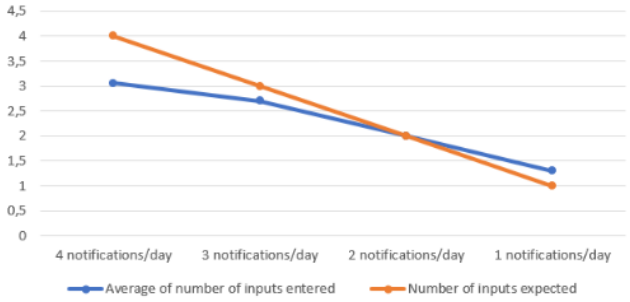


Figure 4: Number of inputs against number of notifications

We have observed that we go more often on the application by ourselves to rate our mood when we receive few notifications (1 or 2 per day). For 3 and 4 notifications per day, we skip some. For 4 notifications per day, we responded to an average of 3 notifications. For 3 notifications per day, the average is still close to 3. It therefore seems that the best compromise between annoyance and collection of a maximum of mood data is reached for 3 notifications per day.

From this experience, we got an unexpected learning. At each input, the user is also asked for a message if he wants to keep track of the events that have influenced the rating of his well-being. During the experience, we realized that writing a message justifying the state of our mood took time, and we were at some point lazy about it. Some activities came back often (sports, studying, ...). It could be interesting to let the user select between some basic choices of activities instead of writing a message. This would also allow to propose additional filters for visualization.

Evaluation of personal recommendation feature:

As expected, the recommendation feature was of interest for mostly of users (83 %).

Among the three questions we asked, one was concerning the potential danger of certain recommendations. We gave to the participants the example of a recommendation leading to shopping (figure 1), which for us could cause a problem if it was given to someone who is addicted to it. Participants in the survey all raised the importance of paying attention to this system of recommendations, but their fear were mainly concerning the usage we could make of the user data, and not the misuse of its recommendations by a person who may have addiction problems

Another point raised is the risk of the user getting caught in a routine by always receiving the same kind of recommendation. The user should be allowed to block a recommendation that comes back too often.

Participants were also asked what topics in the recommendations might be of interest to them. They proposed the same ideas that we had in mind (music, weather, location, ...) which reinforces us in the idea of the interest of the recommendations. A very interesting idea was also raised by on of them: *"Maybe if two or more friends use your app, and it suggests the same activity to both of them, you could implement a way of communication, or you could suggest they do the activity together".*

4 DISCUSSION AND FUTURE WORK

User study:

Even though the design evaluations regarding the input and the visualization gave interesting results as of which designs were preferred by the interviewees, the most interesting findings were the ideals for the users: efficiency and simplicity for the input, clarity and simplicity for the visualization. It is to be noted that the visualizations provided were all based on a weekly visualization; comments pointing out that a line plot wasn't suited for such a short period were on point. The validation of the designs B and C can be considered as high enough for a mock-up, whereas Design D (the radar chart) was discarded, considered too confusing for the mood-rating visualization. A future design evaluation for the visualization should take into account the difference between different time period visualization, and a good proposal could be our Design C for the weekly basis, and Design B for the monthly/yearly basis.

Concerning the personal recommendations, we were unable to implement this evaluation over the short period that we had, but ideally, we had envisioned another more precise evaluation. It would have consisted in a form sent to users of the application for at least one month, asking them to rate and give their feelings on personal recommendations that the app could have created for them. It appears from the results that the economic model needs to be reflected upon: the greatest fear of the users is that their personal data collected will be transmitted to other companies. The question of financing such an application (subscription, advertising, ...) should be studied further.

Notifications and inputs frequency experiment:

There is a large gap between the averages of the number of inputs per participant per day, and we would expect it to increase with a larger sample. If 3 notifications/day could be taken as a good baseline, we may have found a lower result with a larger sample. Thus, it remains important that we give the user the possibility to choose the frequency of notifications, but we can set it to 3 per day by default.

Several limitations are to be considered with regards to the modalities of the experience:

- Type of the experiment: Since we conducted all the experiment on the same users, it is possible that a habit is formed with a previously tested notification frequency and has an influence on the following session.
- Time: We had a short time to develop this experience, so we chose to work over periods of 5 days. This periods can be a little too short for a real habit to develop in the participant.
- Population: We conducted the experiment on both of us. Thus, the following limitations must be taken into account:
  - The sample is very small
  - The sample is not diversified in terms of population or gender
  - We probably have an more interest in the application than a basic user that probably pushes us to respond more often to notifications.

Another factor that should be tested is the time at which notifications pop on the phone, as it can influence the ability of the user to answer them. Keeping the previous points in minds, it would be interesting to carry out a more complete study on a larger panel of testers, who are not responsible for the application, over a longer period of time and by crossing test conditions.

5 CONCLUSION

Our user evaluations helped us shape the ideal designs for the input and visualization of a mood-rating app, while confirming the interest of implementing a personal recommendations system. Our user experiment showed a best frequency of 3 notifications per day by default to request user input. Overall, the results obtained were not unanimous, and our findings point towards a personalized proposal for the user. From these results, we were able to create a MVP, presented in the Appendices.

REFERENCES

[1] Lesley Chan Vivian Pham Yunan Chen Kai Zheng Clara Caldeira, Yu Chen. 2018. *Mobile apps for mood tracking: an analysis of features and user reviews.* Retrieved April 29, 2020 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977660/

[2] Juan Orjuela Elizabeth Fife. 2012. *The Privacy Calculus: Mobile Apps and User Perceptions of Privacy and Security.* Retrieved April 29, 2020 from https://journals.sagepub.com/doi/10.5772/51645?icid=int.sj-full-text.similar-articles.2

[3] Mac Kozal. 2019. *Why did I design and build a mood tracking app? — a UX case study.* Retrieved April 29, 2020 from https://uxdesign.cc/ux-case-study-1d741a0d5eee

[4] Gin S Malhi, Amber Hamilton, Grace Morris, Zola Mannie, Pritha Das, and Tim Outhred. 2017. The promise of digital mood tracking technologies: are we heading on the right track? *Evidence-Based Mental Health* 20, 4 (2017), 102–107. https://doi.org/10.1136/eb-2017-102757 arXiv:https://ebmh.bmj.com/content/20/4/102.full.pdf

[5] The National Institute of Mental Health. 2019. *Women and Mental Health.* Retrieved April 29, 2020 from https://www.nimh.nih.gov/health/topics/women-and-mental-health/index.shtml

[6] MD Roxanne Dryden-Edwards. 2018. *Mental Illness in Children.* Retrieved April 29, 2020 from https://www.medicinenet.com/mental_illness_in_children/article.htm

[7] Gaétan Moisson-Franckhauser Tanguy Navez. 2020. *Figma - Mood Prototype.* Retrieved May 12, 2020 from https://www.figma.com/proto/WsdMRYBLQvI4CIIA0WKLc0/Mood_Prototype

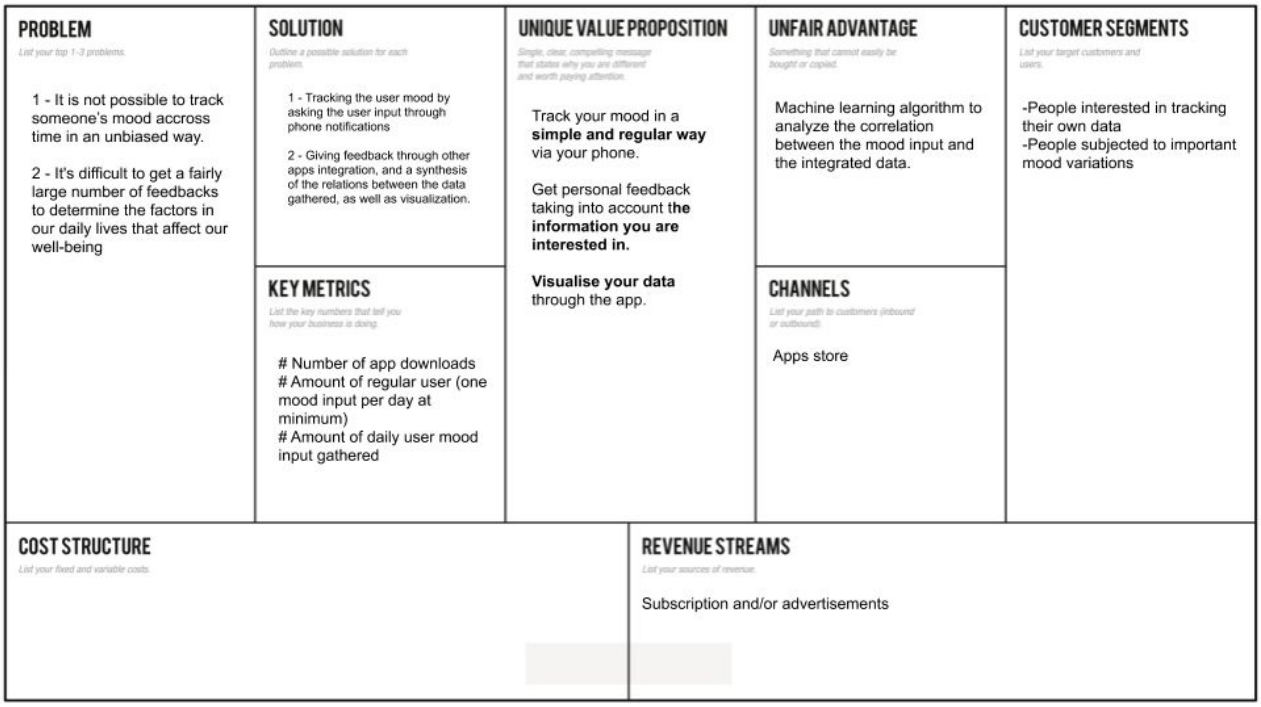
A APPENDICES

(1) Lean Canvas
(2) User story map
(3) Presentation of the test application
(4) Architecture diagram
(5) Links to the prototype
(6) User evaluation description

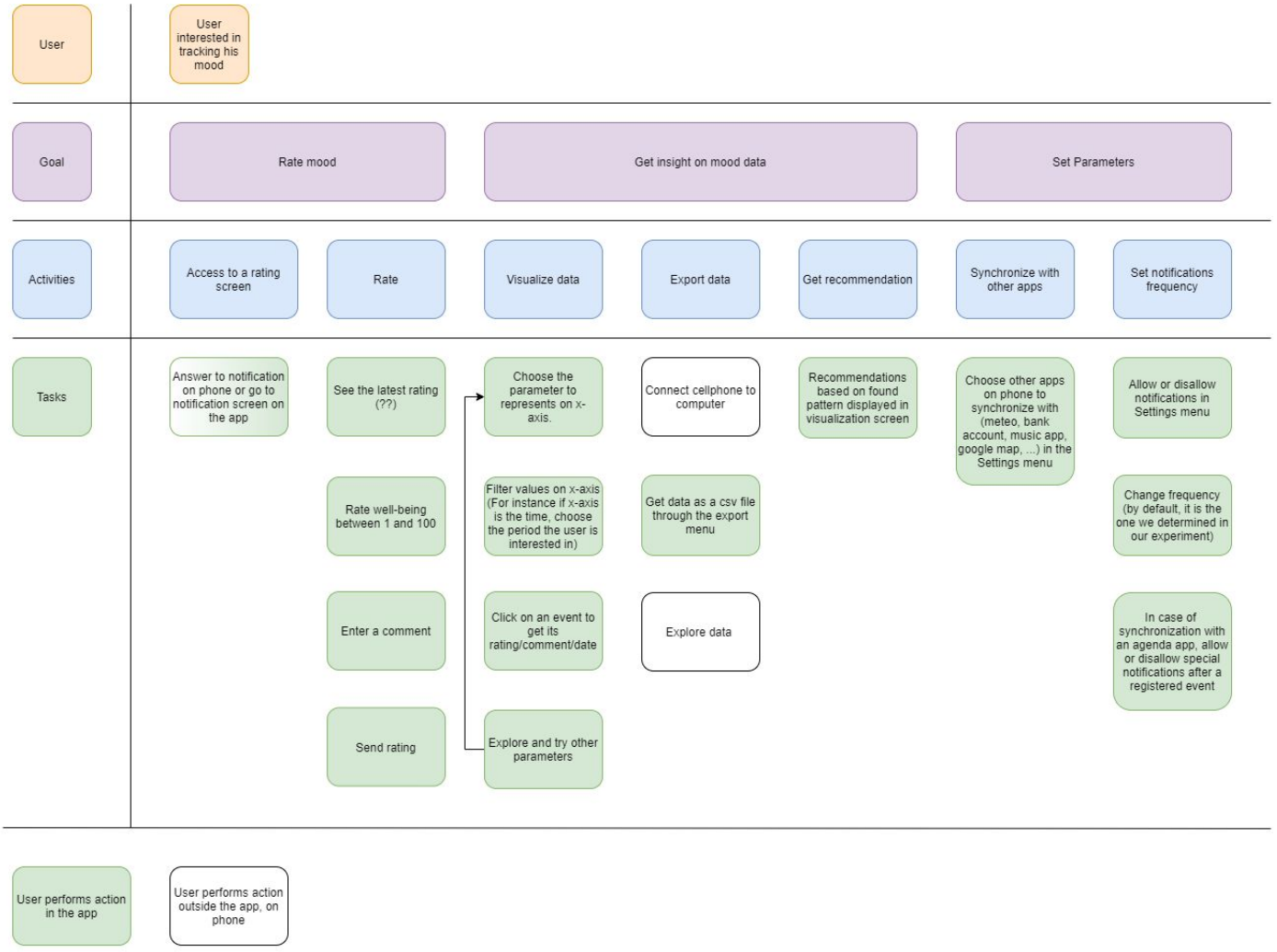
B WORK DISTRIBUTION

The workload was evenly distributed among the two authors. Tanguy mainly specialized his work on the UX appendices, the prototype creation and the user experiment, while Gaétan mainly worked on the background study, the user evaluation material and the analysis of its results. Both were equally involved in the redaction of this report.

Appendix 1: Lean canvas



Appendix 2: User story map



Note: We are not interested in the login/register goals here

Appendix 3: Presentation of the test application

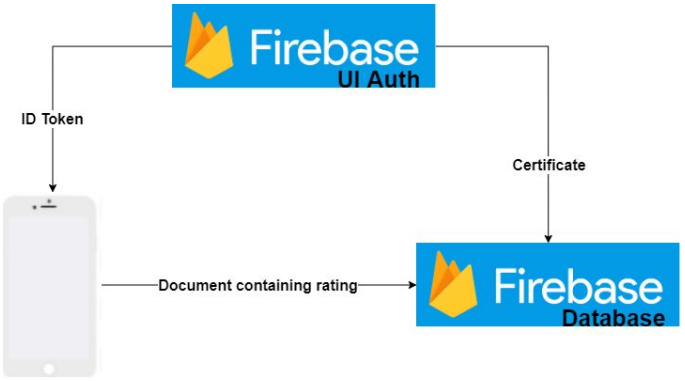
We realized a test application whose objective is to manage the inputs of the user.

Used Technology:
The approach we have chosen is that of Compiled App Development. We used the React Native language. Our motivations in this choice are the fact that the code works on all platforms and the ease of access to native device features (notifications). Our development platform is Expo which allows us to easily test our application on the phone in real time.
For the backend, we use firebase which integrates easily with expo projects, and offers a lot of functionality to easily manage the logging/registering part in the cloud without having to use a real server.

Main Features:

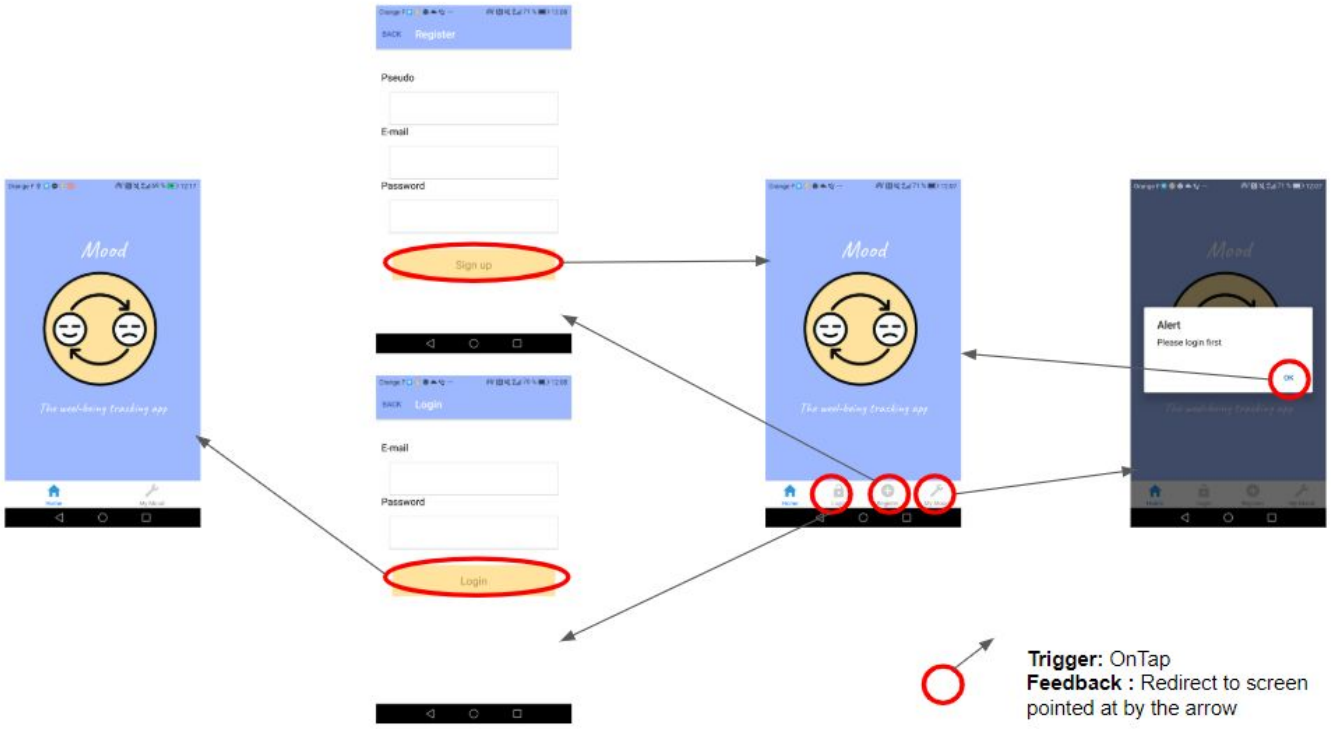
- **Home Screen:** From here, the user can access all screens available to him (he cannot access his Dashboard if he has not logged in for instance).
- **Login and Registration Screens:** These are not the most important features but it is necessary to link the inputs to a particular user to be able to retrieve them. The app is connected with a firebase server to handle it.
- **Dashboard Screen:** It allows the user to send inputs (rating, comment). It creates a new document attached to the user in a firebase database for each new input.
- **Notifications:** The application sends notifications on a regular basis (frequency we determine prior to export). Clicking on a notification sends the user back to the Dashboard screen where he will be able to rate his mood. Currently, we just use phone time, so we don't need to set up any backend to handle pushing notifications.

Architecture diagram:
Here is the actual diagram we used for our tests. The architecture diagram as conceived in the final version is in Appendix 4.

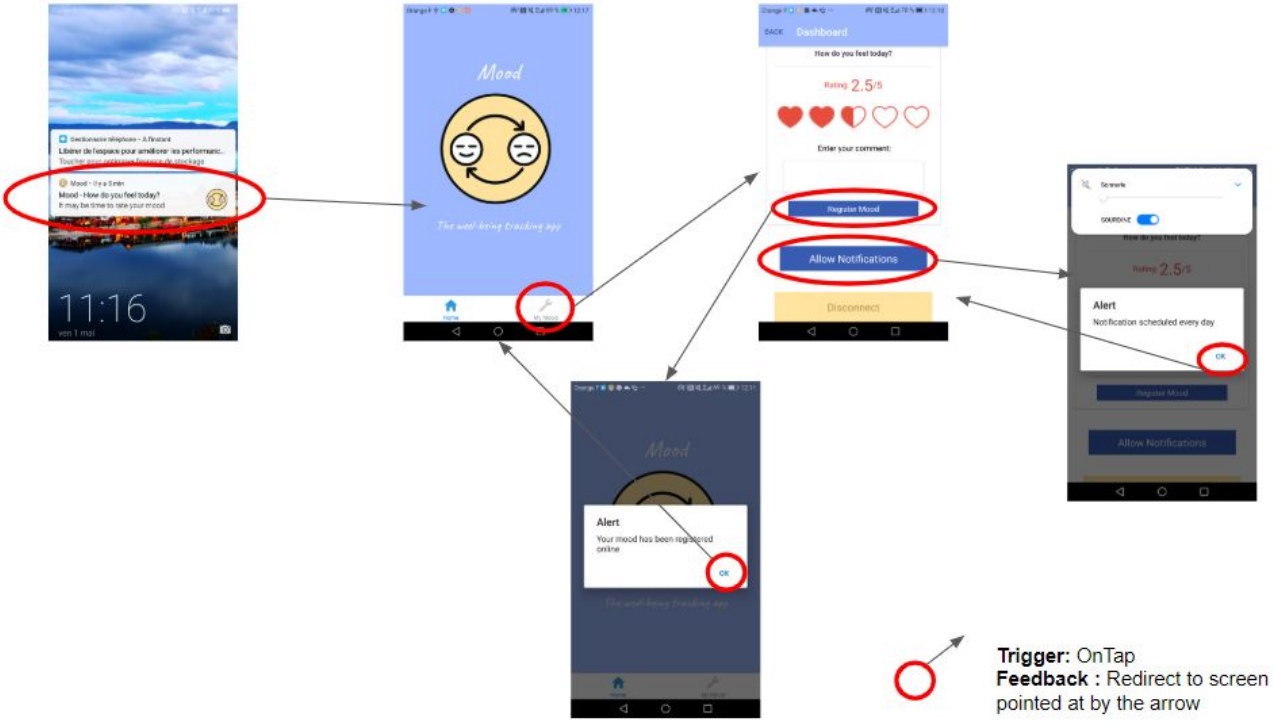


Screenshots of the test application:
Keep in mind that this is an application designed to help us test the inputs, and that it does not look like the final product as we imagined it (Appendix 5)

Registering and Login



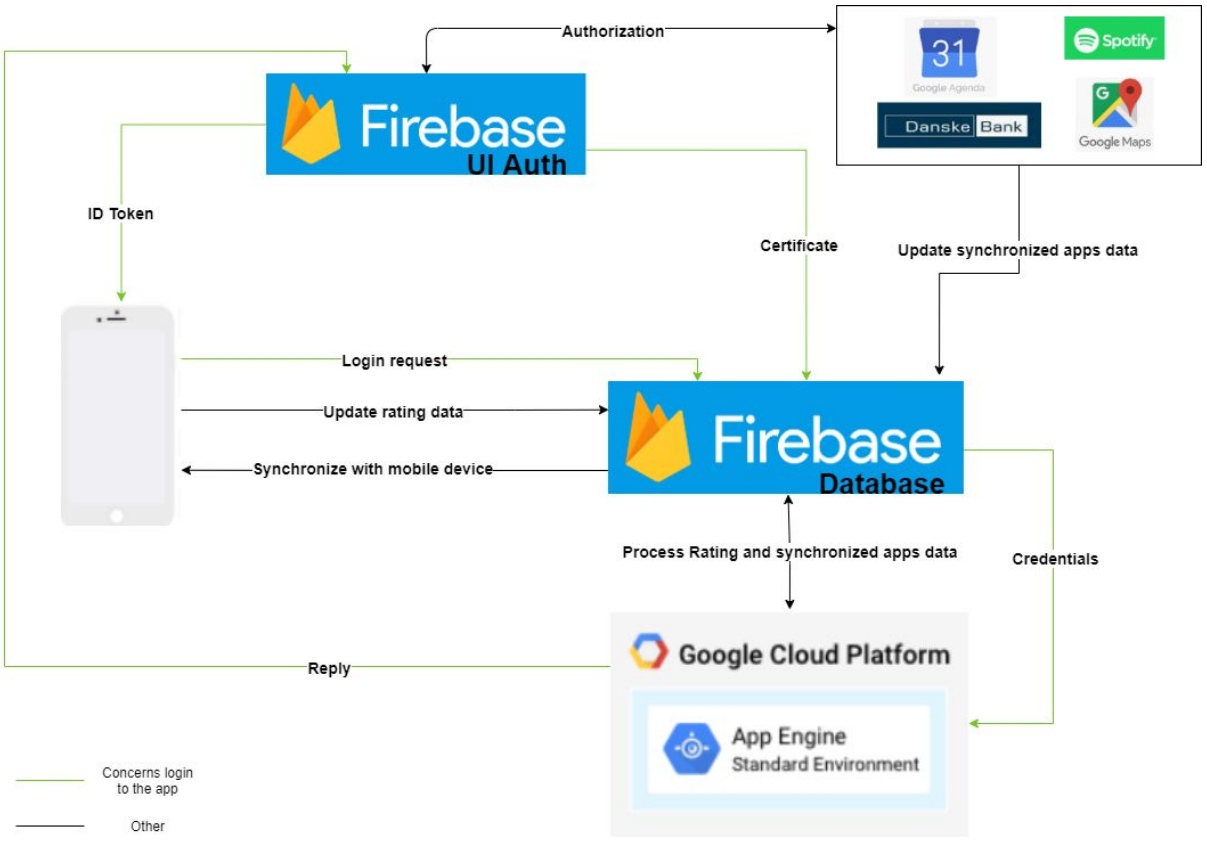
Notification and Rating



Link to GitHub repository: https://github.com/CubicOtter/Mood

The APK is attached to this report.

Appendix 4: Architecture diagram



This is the final architecture diagram, as we have imagined it. The currently implemented architecture diagram can be seen in Appendix 3.

Appendix 5: Prototype

Link to the Figma prototype:
https://www.figma.com/proto/WsdMRYBLQvI4CIIA0WKLc0/Mood_Prototype?node-id=133%3A212&scaling=min-zoom

Video presentation of the prototype:
https://www.youtube.com/watch?v=QDm9jbeQgHs

Appendix 6: The User Study Evaluation

Link to the user evaluation: https://forms.gle/JmwaNw3ZuQSWBZhh8

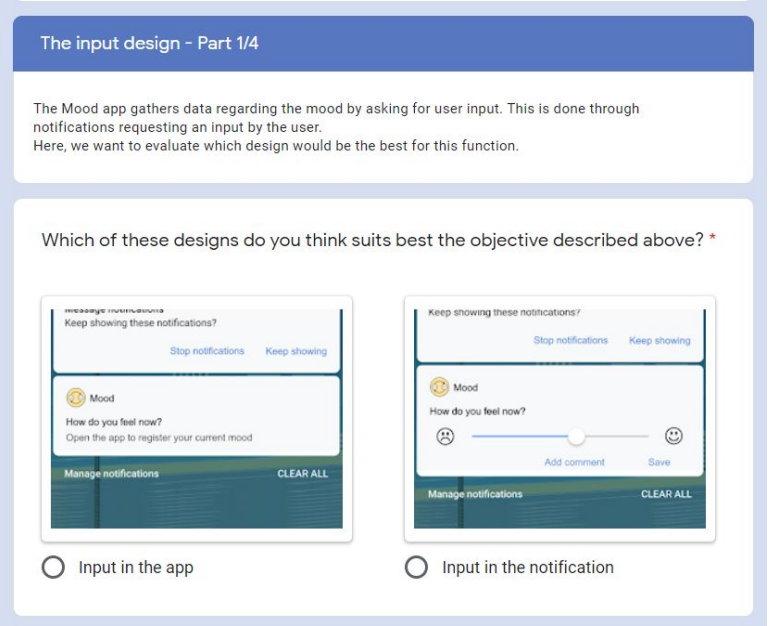
First part: Input evaluation

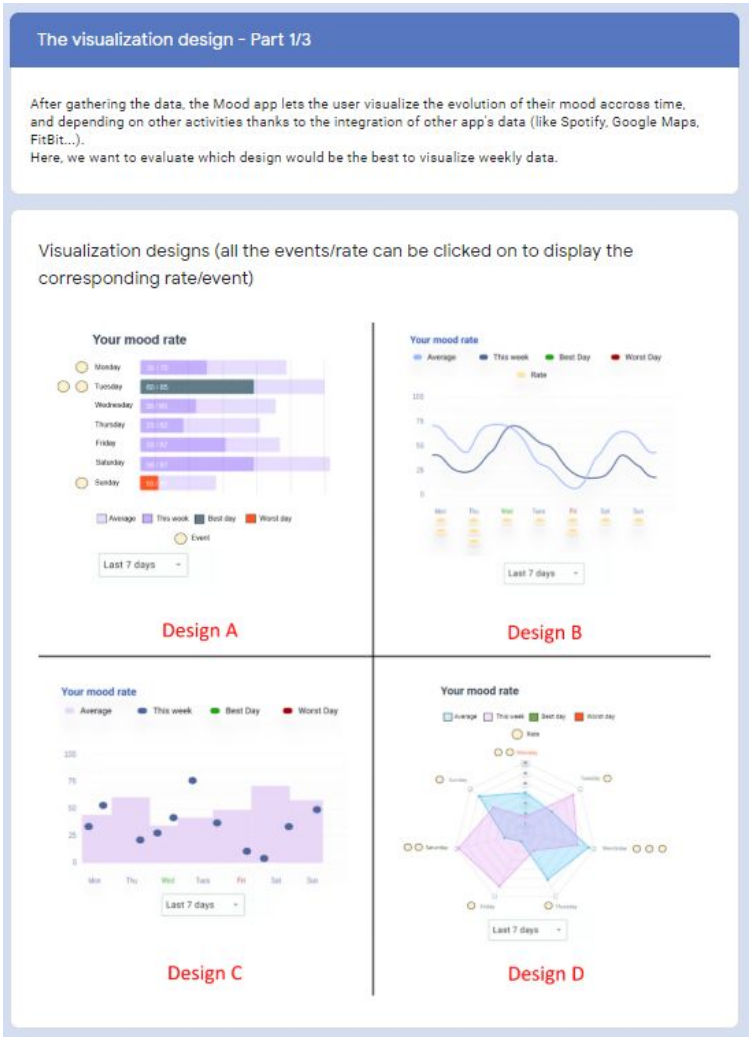
At each step of the evaluation, the interviewee is asked to chose between two design choices: one with a specific feature, and one without this feature. On the following section, the interviewee is prompted to explain their choice.

The four following features are tested :

- Input directly in the notification
- Number indicator
- Default position of the slider on last input value
- Last input value and time written

Finally, the interviewee is asked to rank the 4 features from most needed to least needed.





Second part: Visualization evaluation

The second part of the form questions the interviewee regarding different designs for the visualization design.

Firstly, the interviewee has to select which design they consider as the most fitting the mood-rating display objective.

Then, they are asked to justify their choice, and to suggest how to make the selected visualization design better.

Finally, the interviewee is asked to rate each visualization design in terms of how well it displays weekly mood data, on a scale from 1 to 5. The interviewee is also presented with a field to add a comment regarding the different designs and how to make them better.

Third part: Personal recommendations evaluation

The last part of the form is based on open questions asked to the interviewee. They are precedented by two examples of potential personal recommendations made by the app.

The goal of these three questions is to gather the practical and ethical point of view of the interviewee regarding the personal recommendation feature, with open questions to let them state their personal opinion.