

**[+4] Analyze your data and write up your key findings. The findings should be about 0.5-1p for each motivating question and any other interesting findings.**

- **For qualitative data where you cannot easily remember the details of the results, a thematic analysis is required. include your codebook.**

Question 1: Does this popup/app educate people about data privacy in a dynamic way?

- Follow ups: Is the information well organized? Is it relevant and critical information? Does it allow users to choose the depth of knowledge they can receive?

Key findings:

- Many people noted that the text provided by the popup was accessible to all levels of knowledge and was easy to read and understand
  - There were requests for minor changes to be made, such as reordering the elements and simplifying challenging terminology
  - Some people also felt like the “one-liner” first sentence was too vague and needed more specific details
- The users felt that the popup had content that was important and relevant to them
- They felt like they gained important knowledge they maybe didn’t have before
  - However also felt like they wanted it to be grounded by direct quotes or references back to the original text
- Many users liked that the scale system provided a quick visual analysis of the policy, but they needed clarification on what each rating actually meant for the policy
  - EG, what does “good” versus “bad” mean the company is “good” or “bad at” in their data handling?
  - They also noted that the color scale might not be accessible to all users, but noted that the text rating was also below which was good
  - People liked how the scale allowed users to choose their level of depth, if they just wanted to look at the scale and nothing else they would still get the gist of the policy
- Users felt like the depth of knowledge the popup achieved was reasonable, and that the information provided were key features of the policy that they would want to know about.
  - However, many people also noted that they haven’t actually read the privacy policies of companies so they don’t necessarily know what else to look for
  - Several users also mentioned that they might want to see more than just the three sections we outlined

Thematic Analysis

Group	Code	Definition	Example Excerpt
-------	------	------------	-----------------

	Undefined metrics	The use of a scale or rating system where the points on the scale were not clearly defined	“The main thing would be making the scale at the bottom more clear.”
	Levels of depth	Users appreciated that there were three options for understanding the privacy policy, and they felt like it helped maintain engagement	<p>“rating scale, and "Learn More" buttons for further details make it user-friendly and informative”</p> <p>“colorbar/rating was good for a high level overview”</p> <p>“The rating. It was very clear if it is good or not”</p>
	Fact checked	Many users noted that they wanted some kind of reference back to the original policy so that they could 1) see more detail or 2) fact check the LLM	<p>“I’d like you to add some references and more data”</p> <p>“ i wish that there was a direct quote from the privacy policy”</p>
	Informative	Most users found the popup to be informative and said that they learned something new that they felt was relevant and beneficial for them to know	

Question 2: Do the people care? Or, is the system designed in the best possible way to get people to use the app?

- When asked what they liked about the popup, many participants said that it was a good idea and they thought the design was well done
- Users said that they appreciated the levels of depth provided, and that they were more likely to use it because they could just look at the scale to see the ratings without having to actually read through the summaries
- Some users also mentioned that they were surprised that medical and financial data was a category and it made them consider learning more since they didn’t know that was something they should be concerned about companies having
- Some users noted that they would not change their habits for using instagram or other large social media sites

- However they still liked the popup and were glad to have that information provided
- many also noted that they would primarily use this on websites of less consequence
  - EG, sites that they would choose to not use if they knew it was bad versus sites like instagram that they would choose to use anyway
- People noted that there is a limit to how much the LLM's can actually synthesize from the policy since it's just summarizing a policy that is usually very obscure and unspecific so it might not be that beneficial
- Any users noted that people might be more likely to use the system if bullet points were used instead of sentences, or if key words were highlighted/bolded

**For any quantitative data, submit a script for analysis + your data. Recommendation: Create a notebook for your analysis. Someone should be able to run your notebook to reproduce your results.**

During the evaluation, we measured the amount of time users took interacting with key parts of the extension. Namely, we measured time spent reading each learn more section and time spent on the app. With a python script, we calculated several different statistics:

1. General Stats (mean, median, standard deviation, etc.)
2. Pearson Correlation
3. Outliers
4. Clustering
5. Regression

Here are a couple of the interesting insights we found from these metrics:

- On average, users spent the most time reading through the first learn more section and less time on each subsequent section
  - **24%** less time spent on the third learn more section than the first
  - Reasoning: Users likely become more lazy to read through each section with as much attention to detail, resulting in less time spent on later sections
- There is a high correlation between all of the time measurements (all > 0.79)
- The highest correlation coefficients were found between the time spent on the learn-more sections and the total time spent on the app
  - Users who spend more time on a learn-more section spend more time on the app (intuitive conclusion)
- **No outliers** were found in evaluation samples for any variable

- outliers were classified as measurements that were  $> 2.5$  z-scores away from the mean
- Conclusion: Our users spent similar amounts of time on each section of the extension and in total
  - We might want to evaluate on a wider range of participants and then check for outliers
  - On the other hand, the data suggest that users take a standard amount of time on each section of the app
- Clusters were mostly formed based on total amount of time spent on the app
  - Most users spent  $< 1$  min on the app  $\Rightarrow$  cluster 1
  - Users with  $> 1$  min on the app  $\Rightarrow$  cluster 2
- **$R^2 = 0.94$**  for a linear regression on Total Time on App
  - Conclusion: This is a very good  $R^2$  value, suggesting that time spent before scrolling and time spent on each learn more section can be used to predict total time spent on the app very well
  - This conclusion makes sense because total time spent on the app is essentially the total time spent on the app (some users keep looking around afterwards)