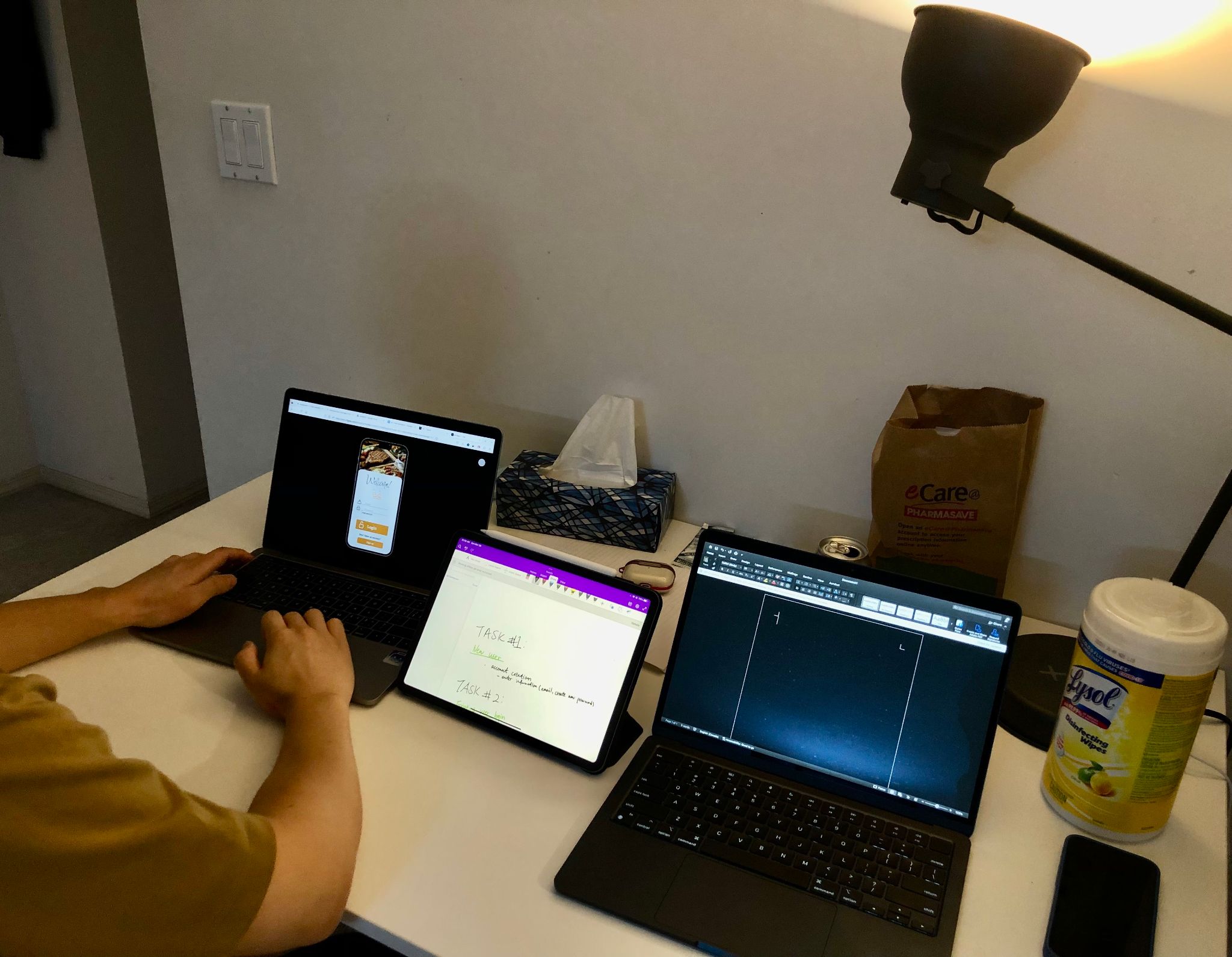
# Study Description

Our study employed both remote and in-person usability testing, with six participants in total. Five of these participants were college students, while the sixth was a recent graduate who also worked part-time as a delivery driver. To analyze the results of our study, we used thematic analysis as our primary method. We chose this approach because it allowed us to collect direct opinions from users on specific features, which we could then record as quotes for analysis. We began by categorizing the data into quantitative and qualitative types. For quantitative data, we recorded the time it took each participant to complete a task, the number of questions they asked while performing a single task, and the frequency with which they made errors due to unclear design elements. By conducting this analysis, we were hoping to gain valuable insights into the usability of our product.

## Setup of in-person and remote sessions:



# 

**For the in-person usability test** we had the participants sit down with a team member and set them up with a laptop with our prototype. The team member has a timer which keeps track of the time taken by the participants to complete each task respectively, at the same time our team would also take down notes(i.e. quotes, operational errors and suggestions)

**For the remote study,** we opted for unmoderated testing, which allowed our participants to take the test at their convenience and pace. We provided them the tasks to complete, and the results were recorded and analyzed later. Although in-person usability tests provided us with more in-depth insights and personalized feedback, the unmoderated approach is more cost-effective and efficient for our remote study.

Here is our tasks for our participants to complete:

[Task 1](https://docs.google.com/document/d/1c9WzTNthhXBWur79QMfWVqSWddG51lxIThT7958WQXs/edit?usp=sharing)

[Task 2](https://docs.google.com/document/d/1LweNOKZ1o6CsMqdCreY5nqzI2r__Te7NntMc-7b35hY/edit?usp=sharing)

[Taks 3](https://docs.google.com/document/d/1NH2aasvD3SuxrYECSb2xNoCMGn7QUuoHM9agdGoqIFo/edit?usp=sharing)

[Task 4](https://docs.google.com/document/d/1m0YqmMxIxguzAWnnoX0ss-4lwhH2et7g9i4yxYISU-Y/edit?usp=sharing)

[Taks 5](https://docs.google.com/document/d/1DRrDgndmY90_Dm_iHc1zURX-_bfgZ0CuXaXPz36ql78/edit?usp=sharing)

[Task 6](https://docs.google.com/document/d/1j3v_WbpVu3R4aXaWuFb1zzzAIgxINSSPX5pf54gHGHQ/edit?usp=sharing)

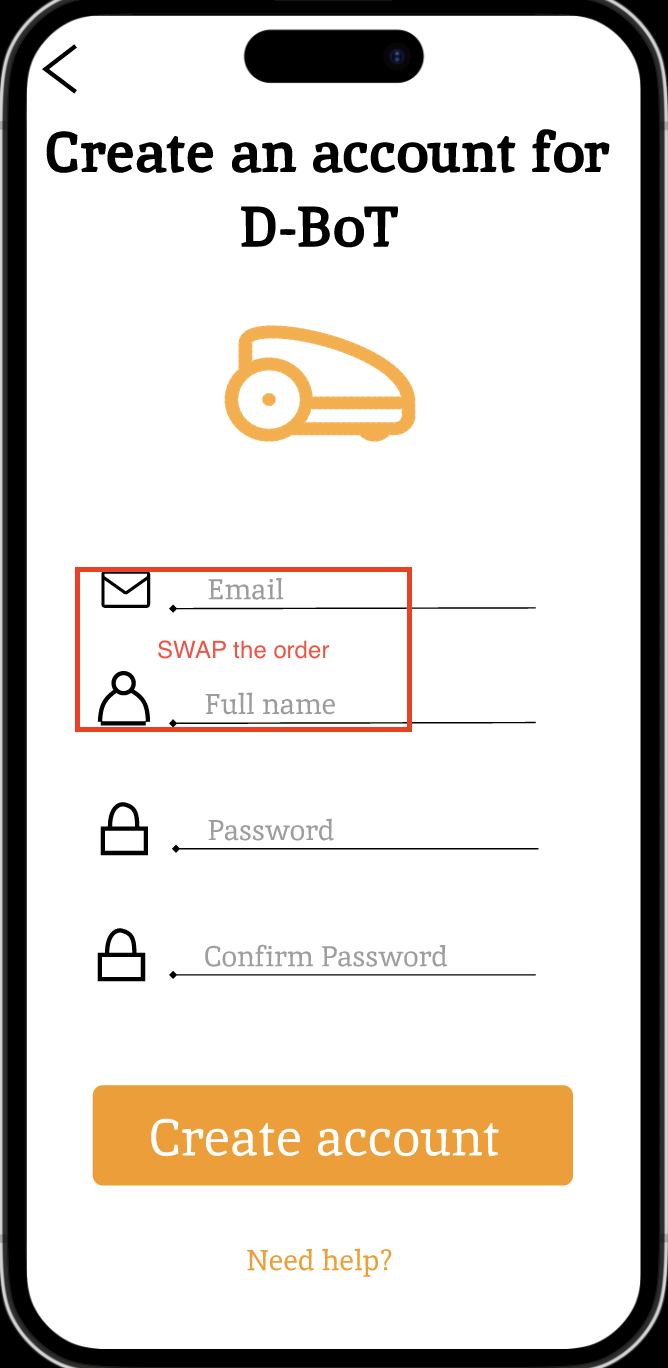
[Task 7](https://docs.google.com/document/d/13gLepsXYM67VQJYf7OJ_TM2oINgks3xfAGglr0XAY-c/edit?usp=sharing)

# Results

According to the feedback and suggestions provided by participants. We have integrated a series of useful improvement ideas. Below are the things that may have caused participants confusion during the testing phase. We have aggregated the quotes of participants, visuals to the correlated interfaces and our solutions if any.

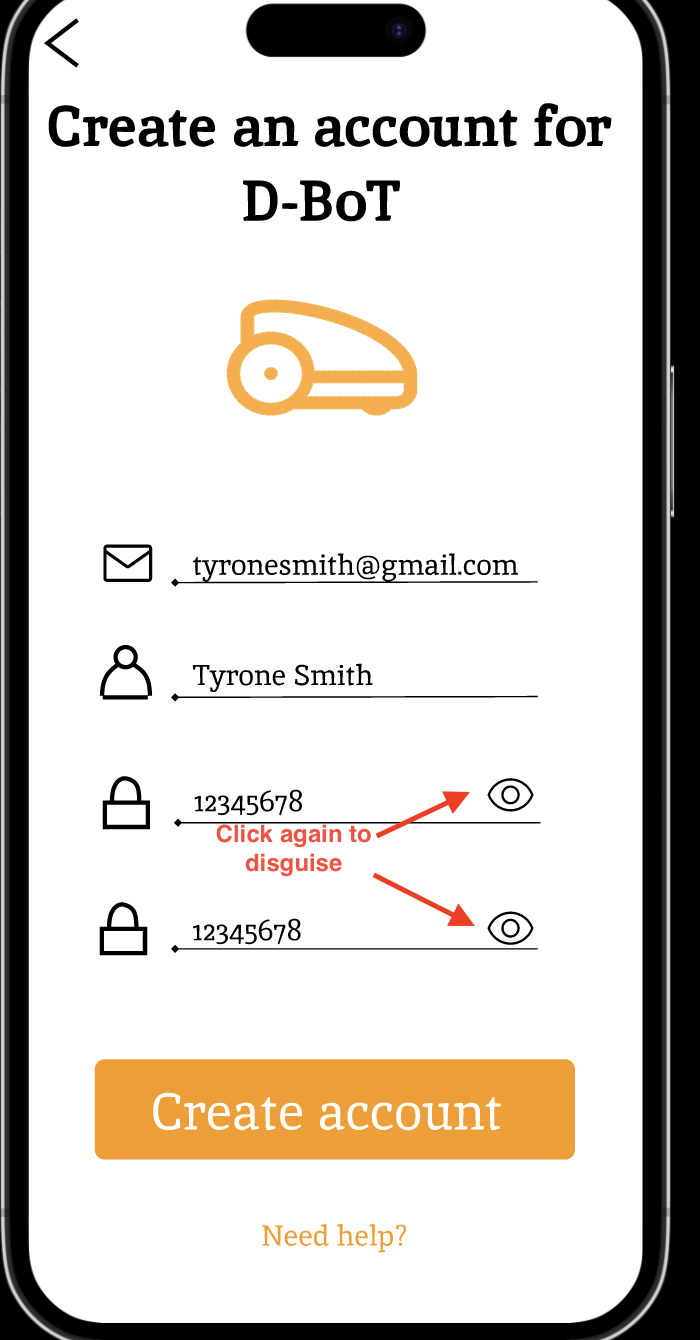
1. <Account Creation Page>
   1. layout order

Quote:”*I personally think my name should be asked first instead of my email address. It is more intuitive that way.*”



* 1. disguise option for password and confirmation password.

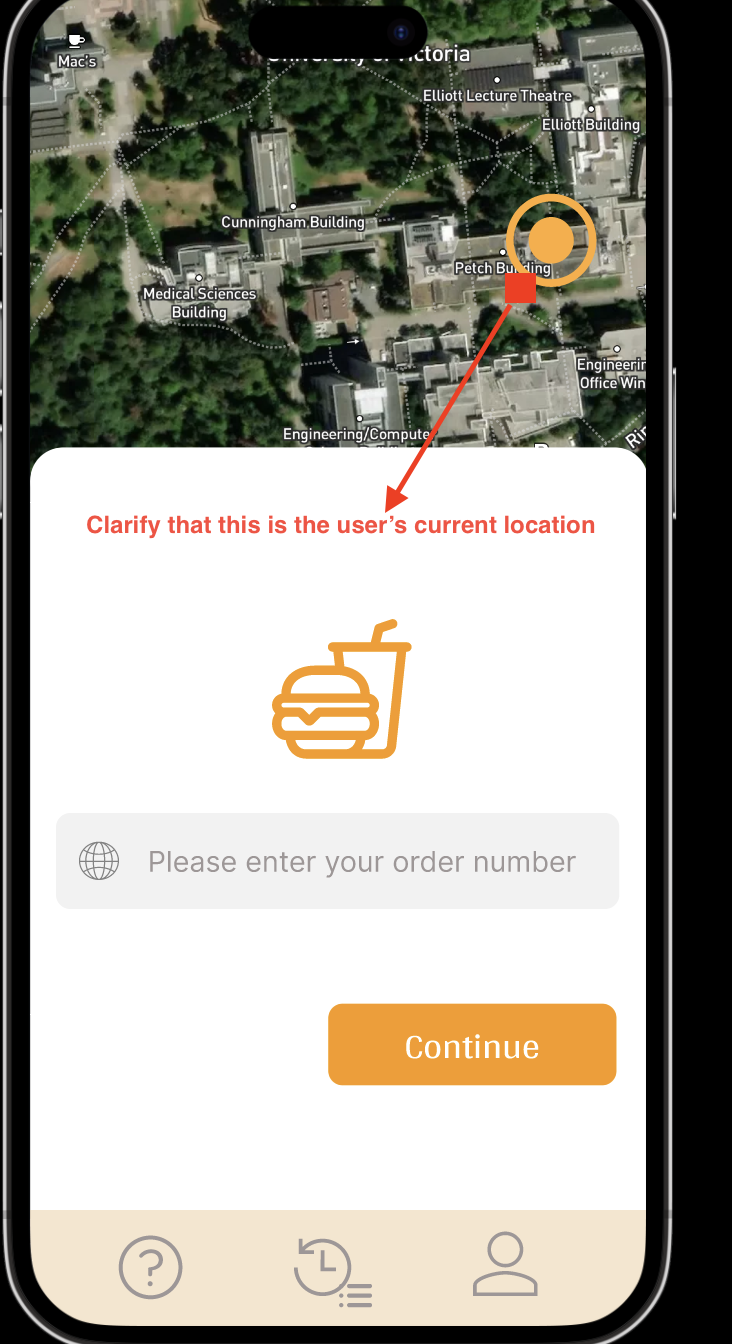
Quote:”*It would make sense for me to be able to both cover and uncover the password*.”



2.<Order Number and Destination Page>

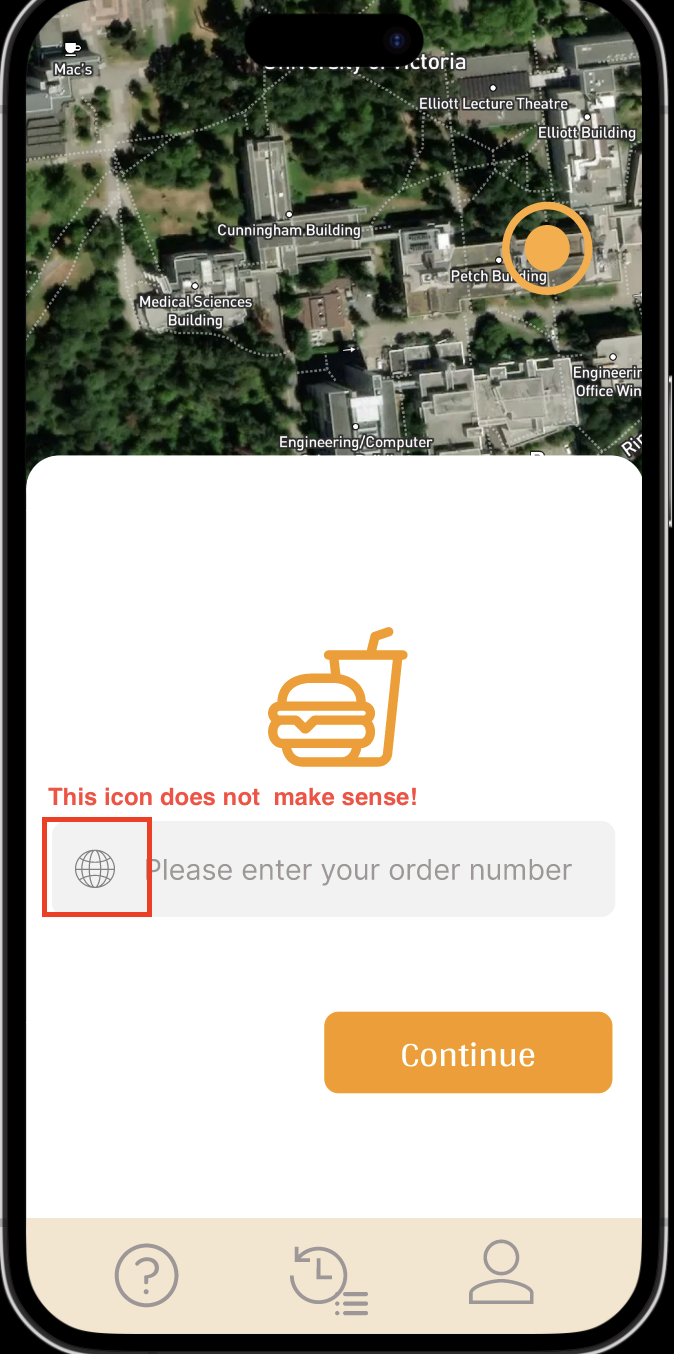
a. location indicator on map

Quote:”*I’m not sure what this location is supposed to be*.”

The user points out that this indicator does not clarify whether this is their location or the D-BoT location. Therefore we thought about clarifying by simply adding a text tag by the indicator.

b. enter order number icon does not map to its functionality

Quote: “*This icon does not make me think of anything that relates to order number*”



c. indicator of destination being selected

Quote: “*The color of the different kinds of location is confusing*.”

The user expressed that the indicators all have the same color,and had to come to us for clarification during the testing.

Namely, if the selected destination is not the same as the current user location, there should be two indicators with different colors.

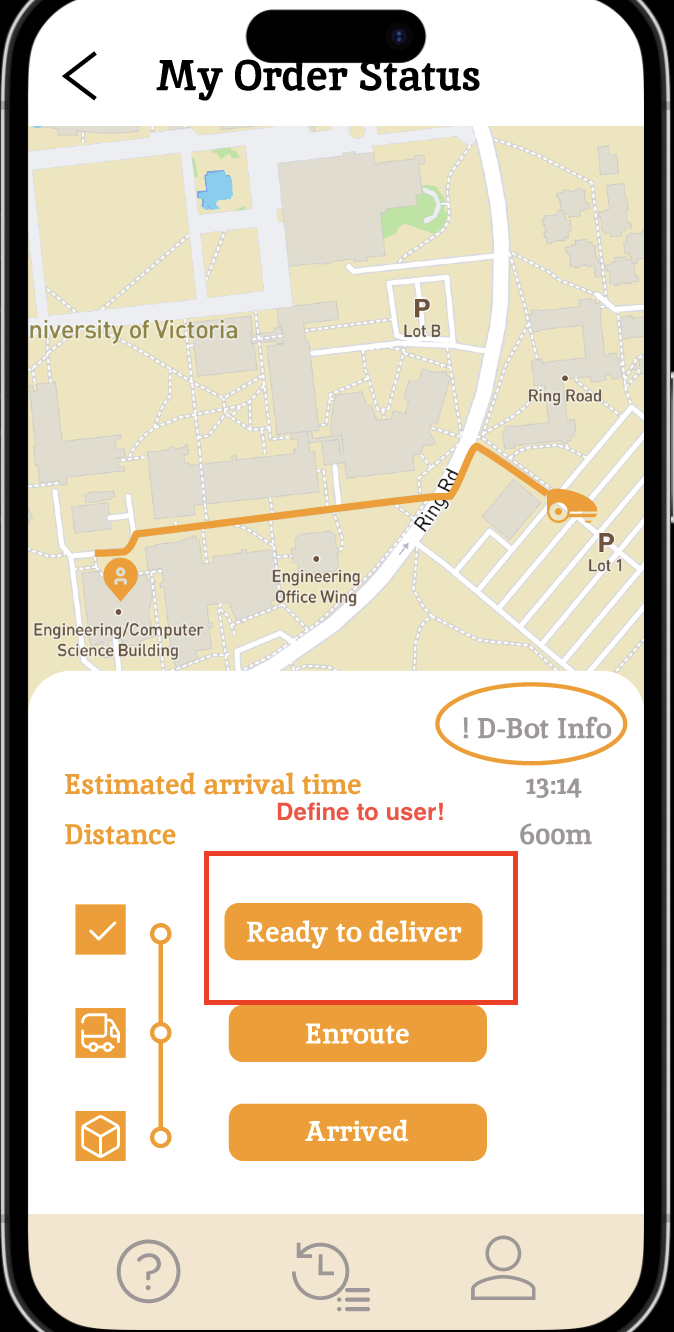
Therefore, in response to this concern, we proposed the solutions as follows: **Case 1(separate)**, if user location != destination, then two indicators should be present, color of user location indicator = yellow, color of destination = blue. **Case 2(merged)**, if user location = destination, then one indicator present with the color green.

3. <Oder Status Page>

a. “Ready to deliver” state

Quote: “*What makes an order ready and when would it not be ready?*”

The participant questions what exactly does ready to deliver indicate, we have explained that an order is ready to be delivered when it is place inside the D-BoT by the driver and there are no other orders waiting to be put into the same D-BoT(a D-BoT can carry up to 3 orders per trip) . Therefore to make this known to the user we would indicate this information beside the state.

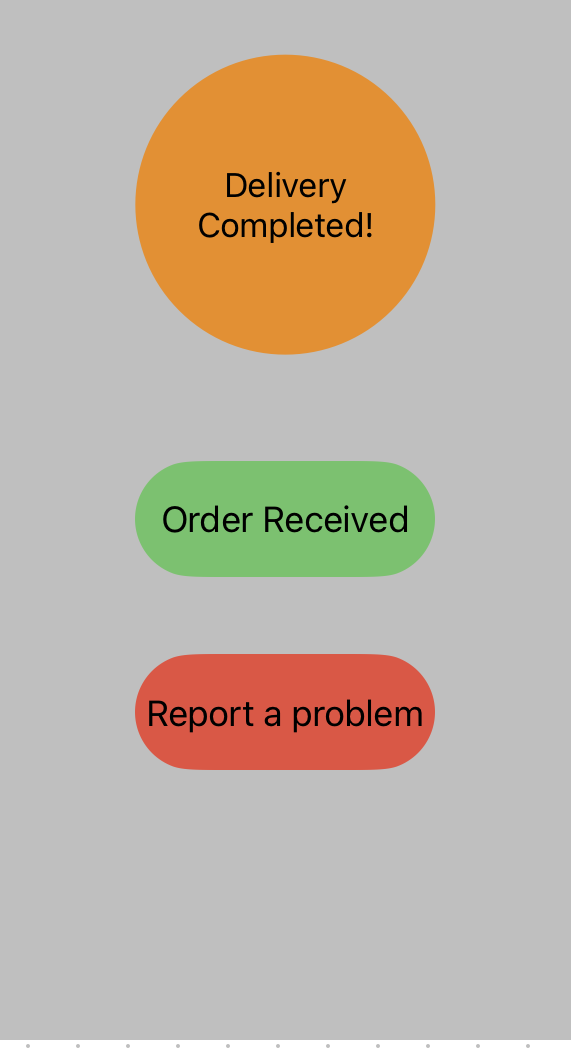


4. <A new page to be added for user to report if they have received or not received their orders!>

Quote: “*What if the order is not inside the D-BoT??*”

To address this issue we would add a new page after the “order status page” that allows the user to confirm they have received the order or to report to the platform that the order is not present.

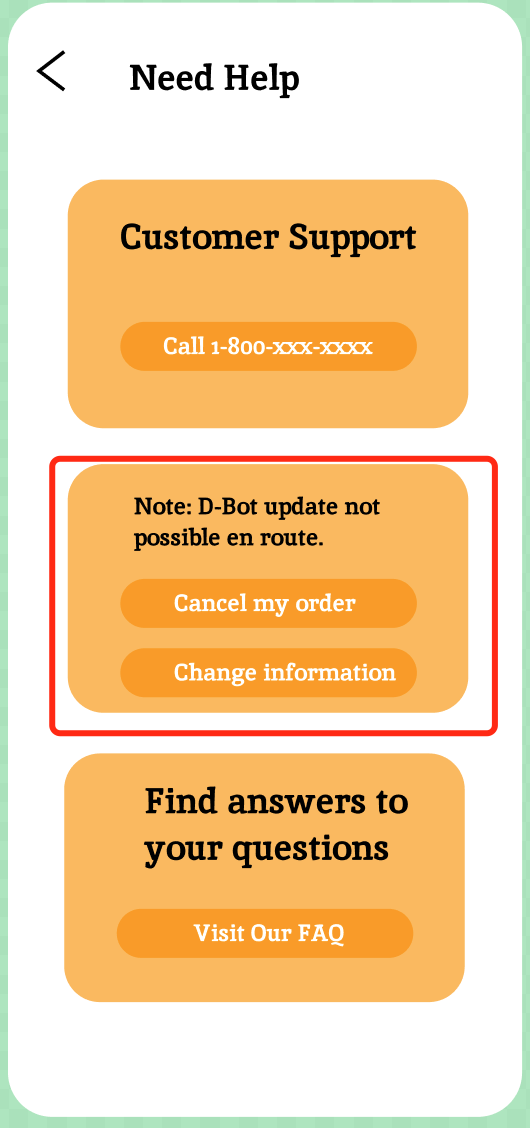
Below is a rough sketch of the proposed new page.



5. < If the users want to cancel/change destination of an order they just replace>

Quote: “*Where can I cancel an order?*”

We have included “Cancel item” and “Change order information” icons in our help page. Users can make modifications before the D-Bot departs from its station.

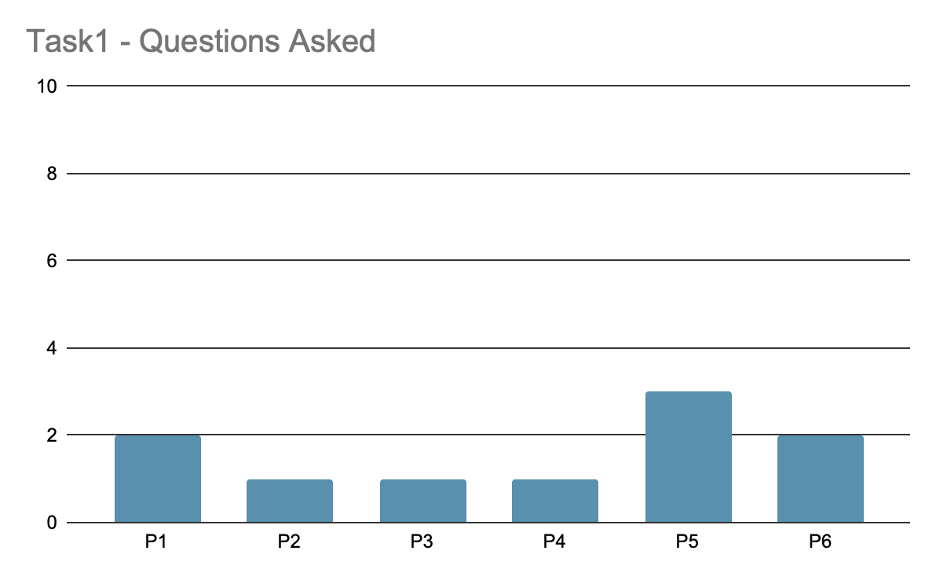
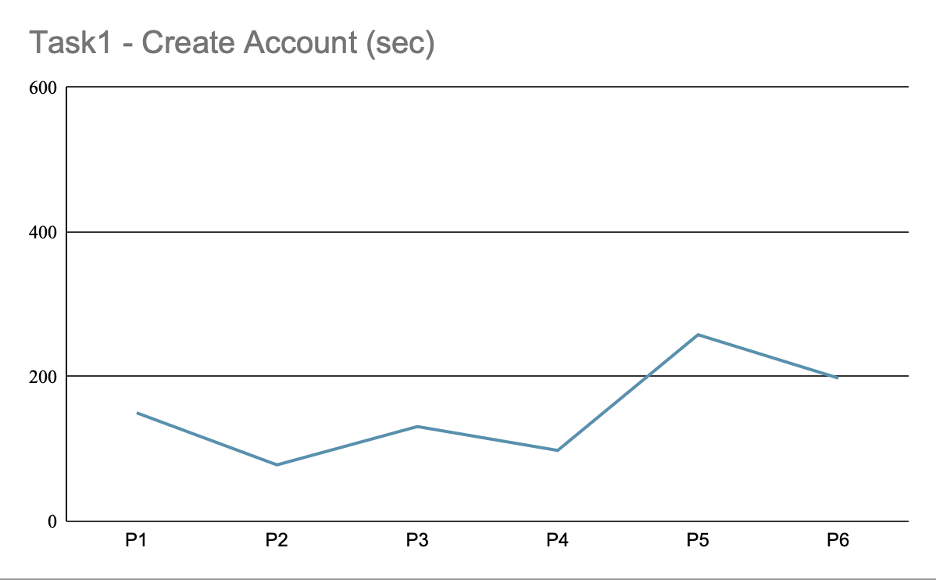


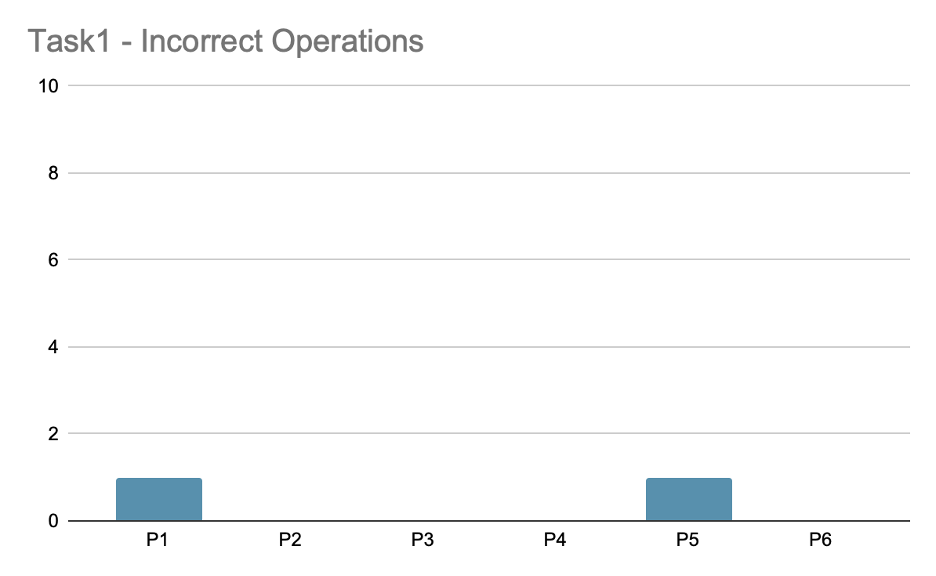
# Discussion

Data Analysis:

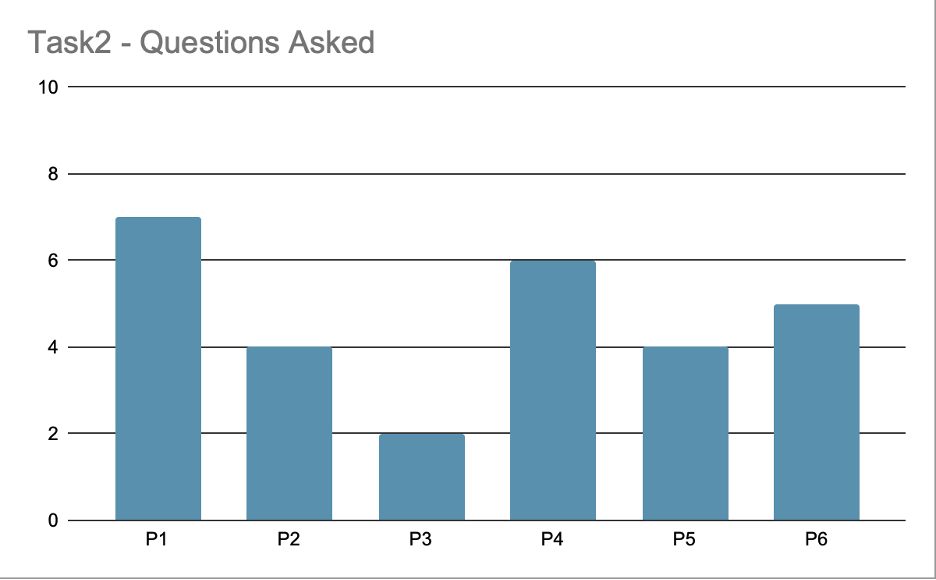
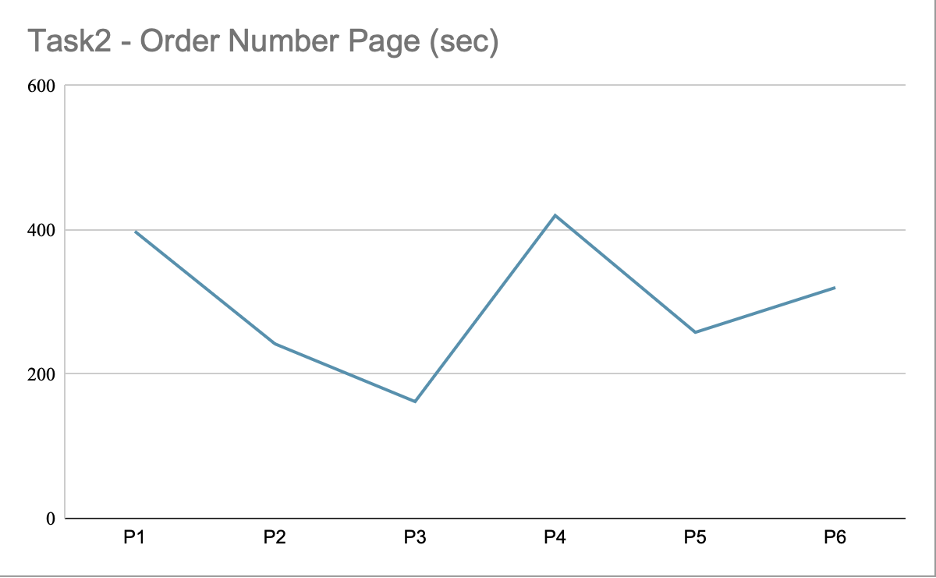
Based on the study, our group created the following tables for each task:

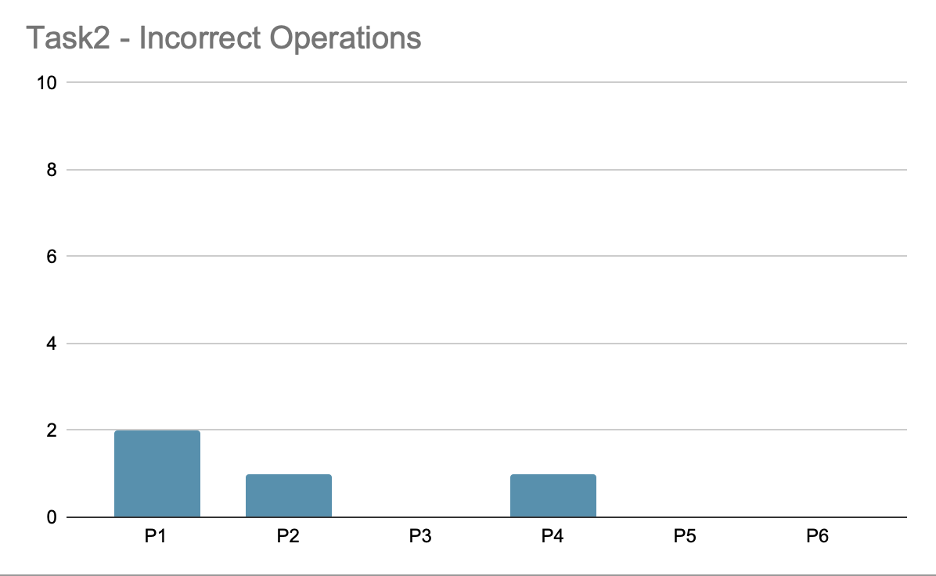
TASK1



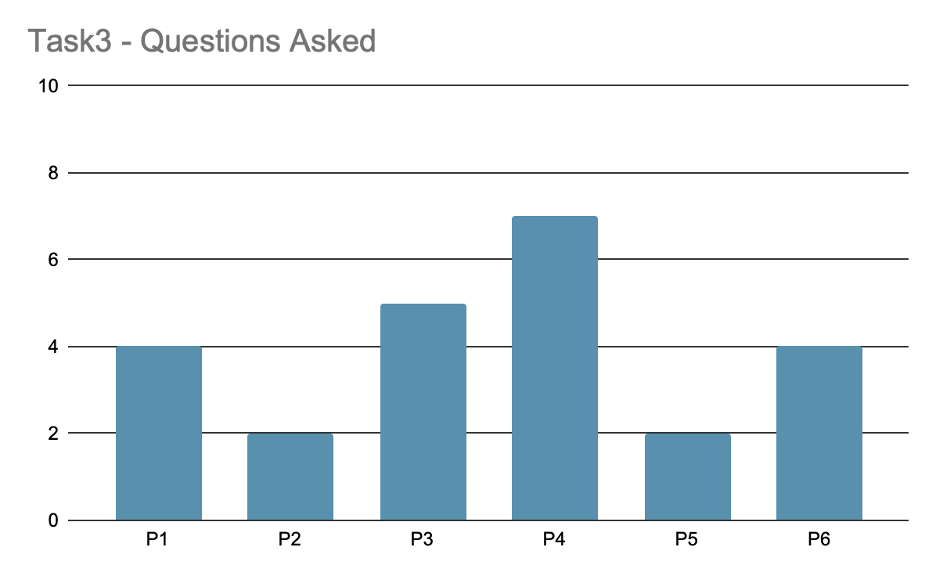
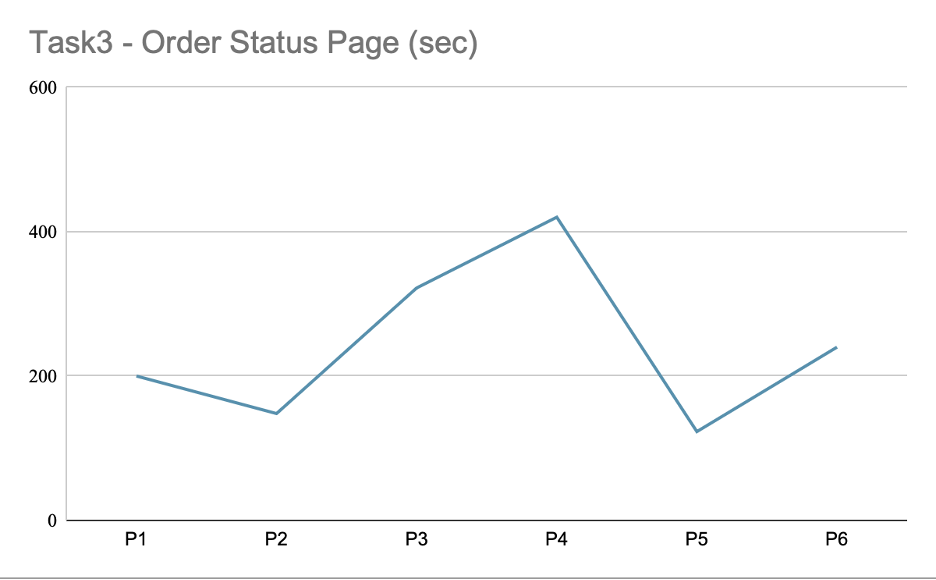


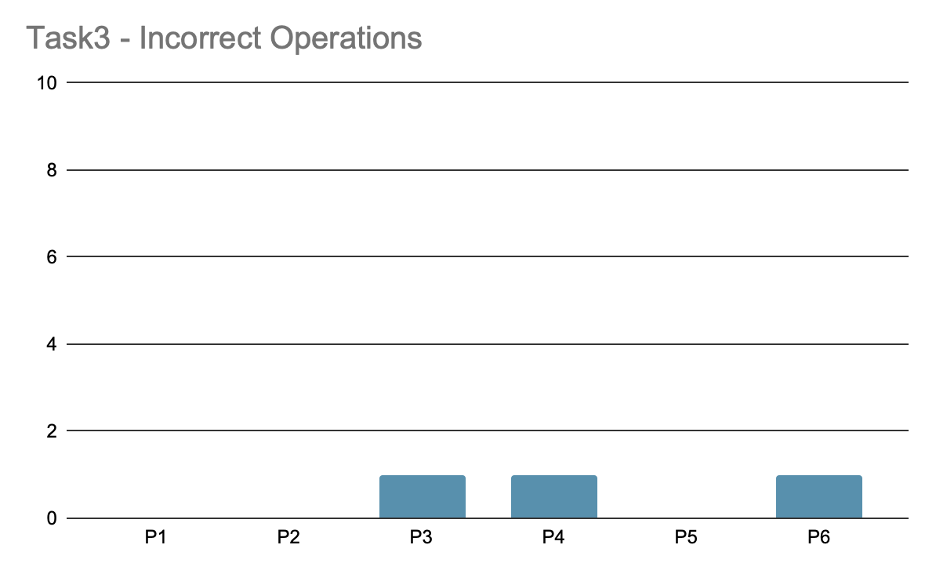
TASK2



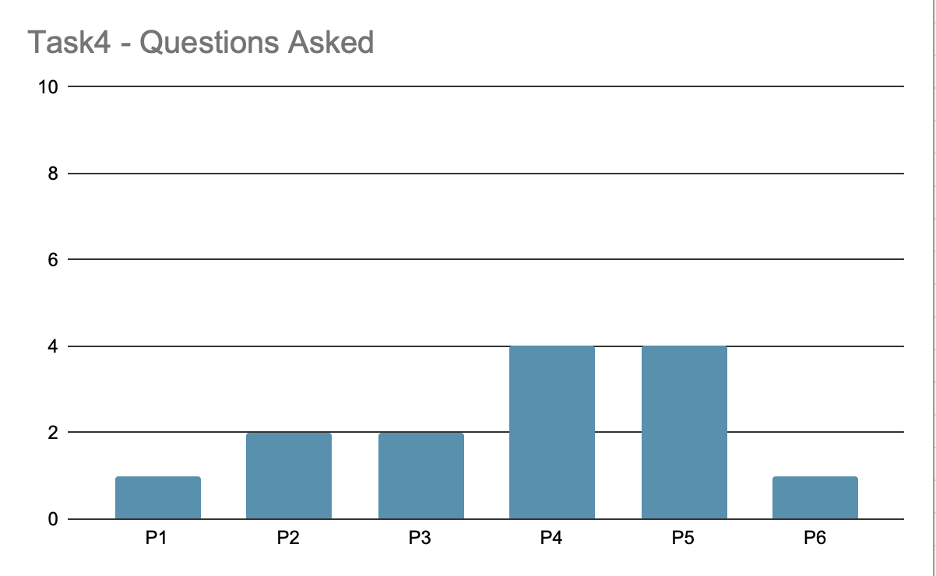
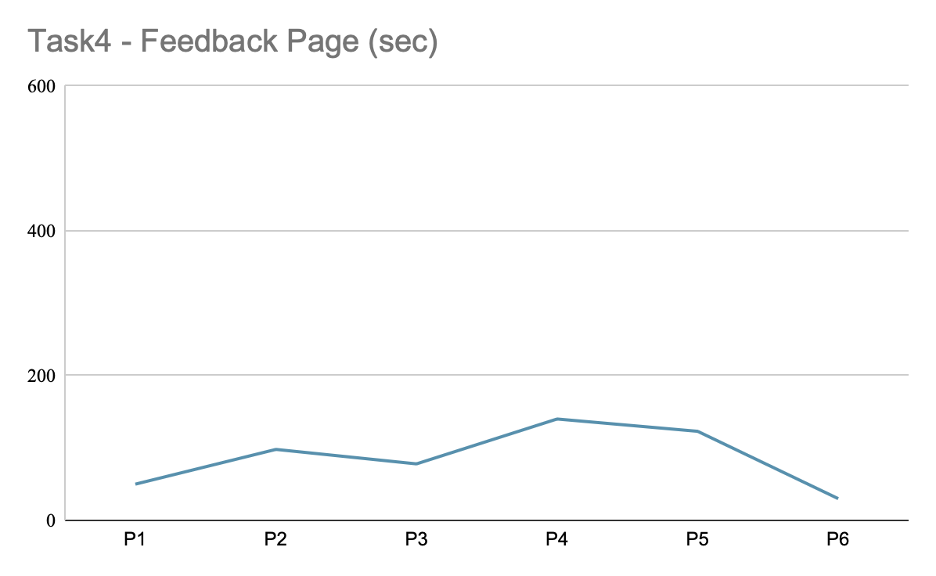


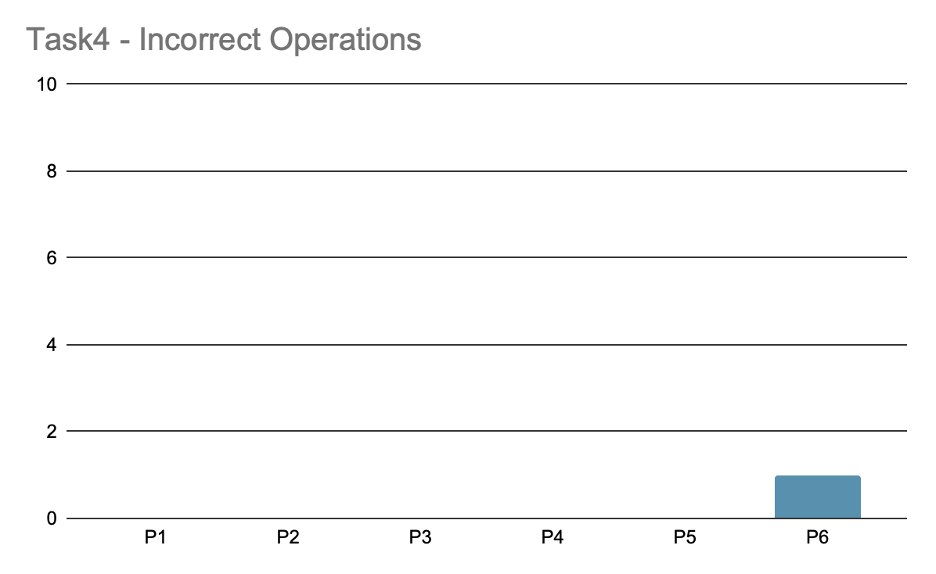
TASK3



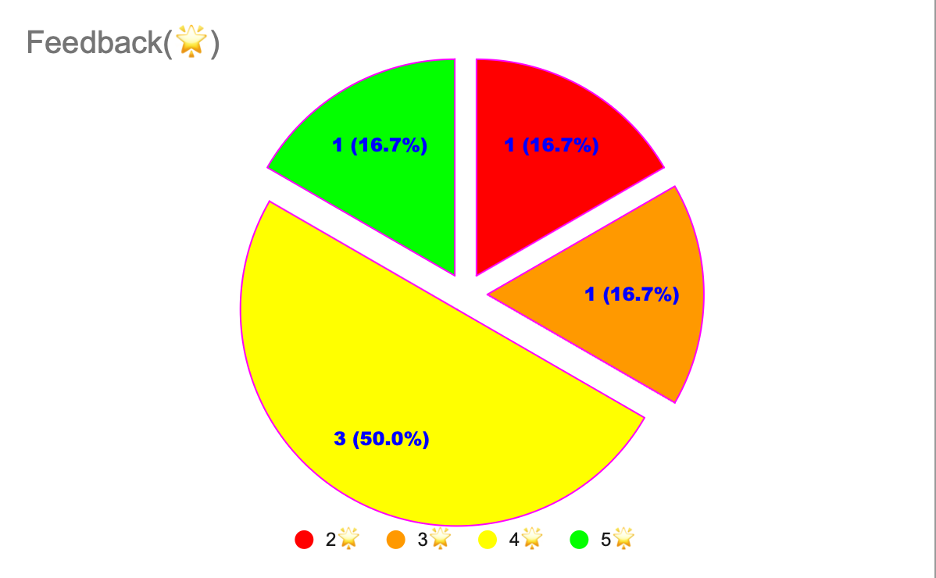


TASK4





FEEDBACK



The average amount of time spent on Task1 by users was 152 seconds with a minimum of 78 seconds and a maximum of 258 seconds; average time spent on Task 2 was 300 seconds with a minimum of 162 seconds and maximum of 420 seconds; average time spent on Task3 was 242 seconds with a minimum of 123 seconds and maximum of 420 seconds; average time spent on Task4 was 86 seconds with a minimum of 30 seconds and a maximum of 140 seconds. The **average total time** spent by users was **780** seconds with a minimum of 566 seconds and a maximum of 1078 seconds. These findings suggest that users may have different levels of proficiency or interest in each task, which can be further explored and investigated to improve user engagement and task completion rates.

The average number of questions asked on Task1 was 1.6 with a minimum of 1 maximum of 3; average number of questions asked on Task2 was 4.6 with a minimum of 2 maximum of 7; average number of questions asked on Task3 was 4 with a minimum of 2 maximum of 7; average number of questions asked on Task4 was 2.5 with a minimum of 1 maximum of 4. The **average total number of questions** asked by users was **12.7**. Task 2 had the highest average number of questions asked while Task4 had the lowest. This suggests that users may have encountered different levels of difficulty or uncertainty with each task, leading to varying amounts of questions asked. Identifying and addressing the most commonly asked questions can help to improve user experience and reduce barriers to task completion.

The average number of errors made on Task1 was 0.33 with a minimum of 0 and maximum of 1; average number of errors made on Task2 was 0.66 with a minimum of 0 and maximum of 2; average number of errors made on Task3 was 0.5 with a minimum of 0 and maximum of 1; average number of errors made on Task4 was 0.16 with a minimum of 0 and maximum of 1. The **average total incorrect operations** was **1.65**. This indicates that overall incorrect operations made by users are very limited yet still need improvement to minimize errors.

3(50%) of users rated our app **4 stars** and the other three(16.7% \* 3) gave us **2**, **3**, **5** stars respectively. This suggests that the app received mixed reviews from users, with some finding satisfactory and others having more varied opinions. Further analysis and feedback from users may be needed to identify areas for improvement and enhance overall user satisfaction.

Data analysis showed that most participants found the mobile app's mid-fidelity prototype easy to use and appreciated the convenience of being able to place orders and track their delivery status in real time. Participants also provided valuable feedback and suggestions for improvement, such as adding more functionality regarding need help functions and suggestions for ensuring anonymity and information security.

Overall, the positive feedback and suggestions for improvement suggest that the app has the potential to improve the customer experience of using the campus robo-delivery system. The insights gained from the study help to ensure that the app is designed to meet the needs and preferences of customers. The study also highlights the importance of conducting user testing and gathering feedback throughout the development process to ensure the app meets user needs. It is worth noting that there are limitations to our evaluation due to objective environmental constraints, and specific issues will be noted later.

# Study limitations and reflections

One of the limitations of this study is the absence of a functional mobile app, as the mid-fidelity prototype may not accurately reflect the full functionality and user experience of the actual app. For example, the study also pre-set the email address and password in the register page for ease of use, which prevented users from entering incorrect details during the registration process, thereby limiting the extent to which a comprehensive test could be performed.

In addition, the sample size of participants was relatively small. The study would have benefited from a larger and more diverse sample to ensure that the app met the needs and preferences of a wider range of customers. Moreover, the study was conducted over a relatively short period of time and it would be beneficial to conduct a longer-term study to assess the long-term usability and functionality of the application.

Overall, the pilot testing of the medium fidelity prototype was successful in identifying areas for improvement and gathering feedback from participants. This feedback and suggestions for improvement can inform future iterations of the prototype, and the insights gained from the study can help ensure that the prototype meets the needs and preferences of the customer.