Work on this deliverable was distributed evenly between all members.

The client already gave us pre-processed data, so no scraping or other methods are required. Even though data collection has already been done, we were only given one of the five hypotheses to work on.

**Preliminary Analysis of Data:**
2 month data:
We found that there are 1016 data points for the daily AM data, 1347 data points for the daily PM data, and 830 data points for the daily end data, all of which starts at week 19 (the week of May 5, 2022) and ends at week 29 (the week of July 18, 2022). Based on the end-of-the-day data, which is also collected from one survey, there are a total of 50 unique participants involved in the study and 7 unique locations.

3 month data:
There are 2240 data points for the daily AM data, 2220 data points for the daily PM data, and 1830 data points for the daily end data. The 3 month data includes data points from the previous 2-month data and ends at week 33 (the week of Aug 15, 2022). Based on the end-of-the-day data (daily_end) there are a total of 69 unique participants involved in the study and 7 unique locations (no new location added).

4 month data:
There are a total of 2988 data points for the daily AM data, 2954 data points for the daily PM data, and 2384 data points for the daily end data. The 4 month data includes data points from both the previous 2-month and 3-month data and ends at week 40 (the week of Oct 3, 2022), which means that it includes the most recent data. The most updated data also shows that we have the same number of unique participants and locations, which means no new participants or locations were added.

**Hypothesis:** Participants will work from an average of three different locations per week (including different rooms of the same house).

Based on data collected from end-of-the-day reports, participants work from an average of 2.57 different locations per week. Because of limited understanding of the data as of now, we decided to omit the data from daily_am and daily_pm files from each months' data. As a result, there may be loss of data if participants start or end their days outside of the general time range.

**Project Scope Reassessment:**
We are currently only working on the hypothesis whether participants work an average of three locations per week. We are currently only using the data from the csv files of daily_end for the 2,

3, and 4 month data, because we are unsure of whether to include data from any other files like daily_am, daily_pm, and the Friday ones. It is very possible that participants worked in multiple locations per day, but grouping daily_end, daily_am, and daily_end altogether could result in incorrect data, so we are just using daily_end for now.

There aren't any large risks of achieving the project goal as it is essentially just calculating the mean from grouped data. However, perhaps answering this specific question and making assumptions from this dataset alone, as there are many biases involved with data collection in general, could be underwhelming. Whether or not the hypothesis is proven, it may just apply to the given group of WFH participants, while using another group would result in a vastly different average or spread of data.