BU Sargent Working from the Home Environment & Well-Being Study Data

Albert Ang, Hany Jasmine, Victoria Luo

Background

This study was conducted over the course of 6 months, where a total of 70 participants with full-time jobs worked remotely, using the computer at least 4 hours a day, and periodically recorded their progress. Participants were also given a Garmin watch to wear which "pinged" them three times a day to share information such as their current location, musculoskeletal discomfort, and the number of breaks they took. This project aims to study the effects of remote work on overall health, and any findings from the data given can potentially provide insight on the future of remote work.

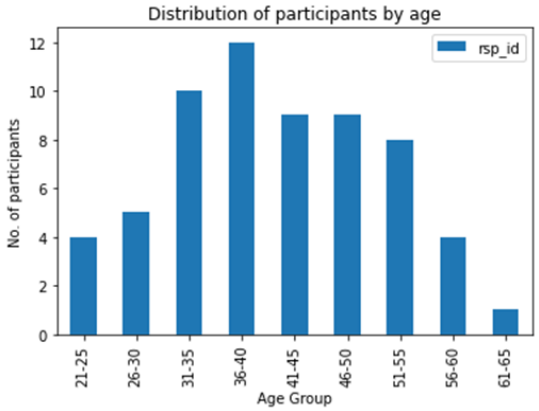
Preliminary Analysis of Data Collected

Throughout this project, we utilized different datasets including daily_AM, daily_PM, daily_END, friday_AM, friday_PM, friday_end, as well as data related to each participant's work stations. These data points were collected over a span of 6 months and were recorded either daily or weekly through the use of Garmin watches paired with the Ecological Momentary Assessment app (mEMA). In order to gain a better understanding of the participants that helped conduct this study, we have decided to also factor in their age and occupation into our data. Collecting the participants' demographic data can ultimately provide us with a more comprehensive understanding of how different variables affect a participant's mental health.

Prior to analyzing the datasets even further, we made sure to preprocess the existing data, which includes handling null values; for example, by replacing them with the numerical mean of a certain variable, and performing feature extraction out of existing data.

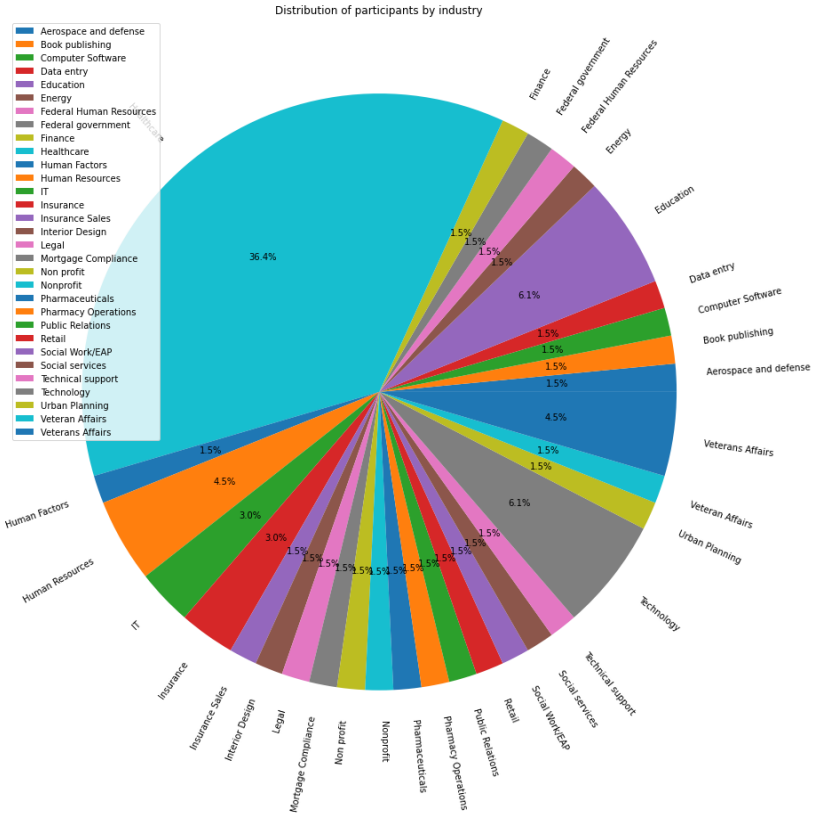
Age distribution of the participants involved in the study

The graph below conveys the age distribution of all participants. We figured this information would provide more context into any findings or results that we had.

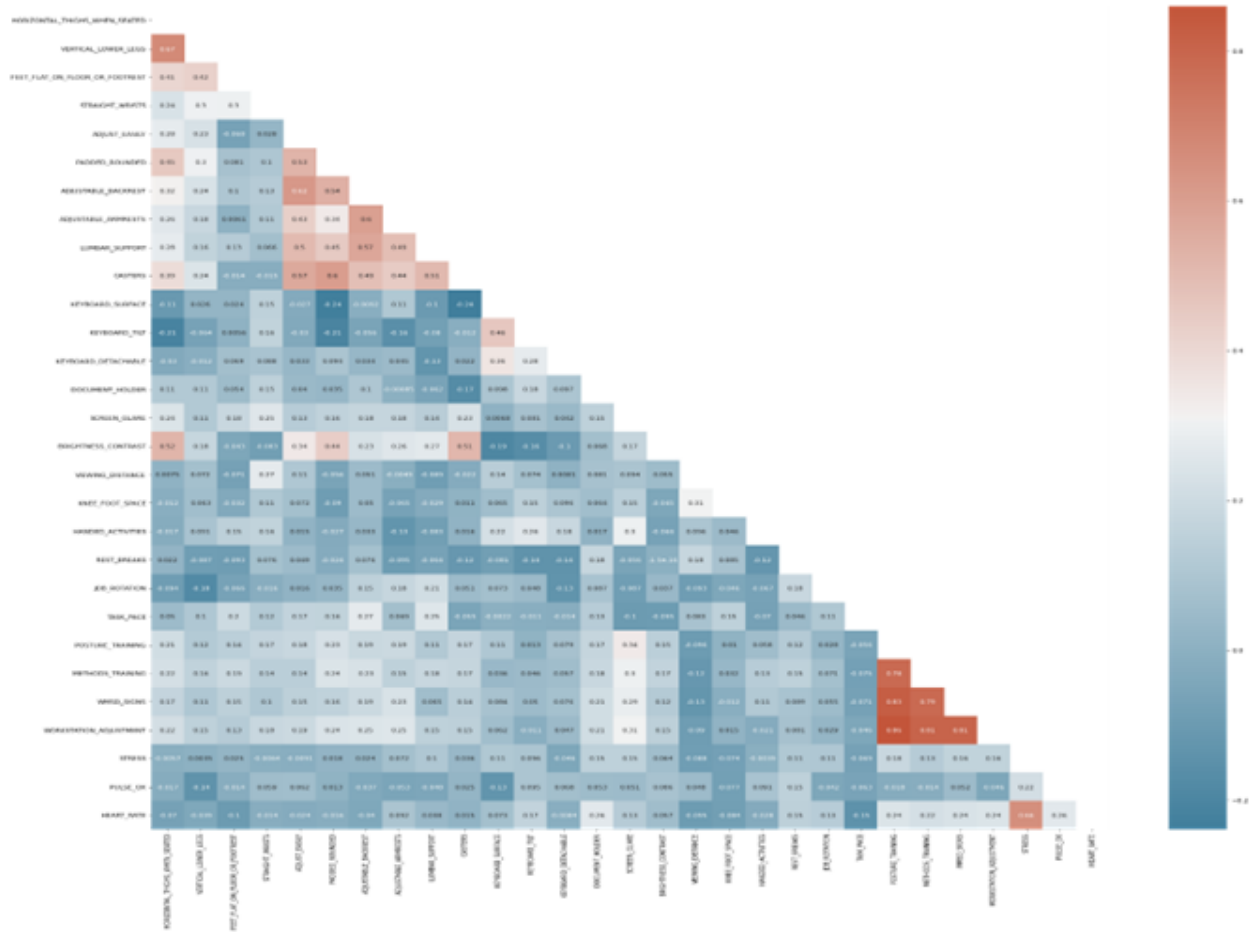


Industry/work distribution of the participants involved in the study

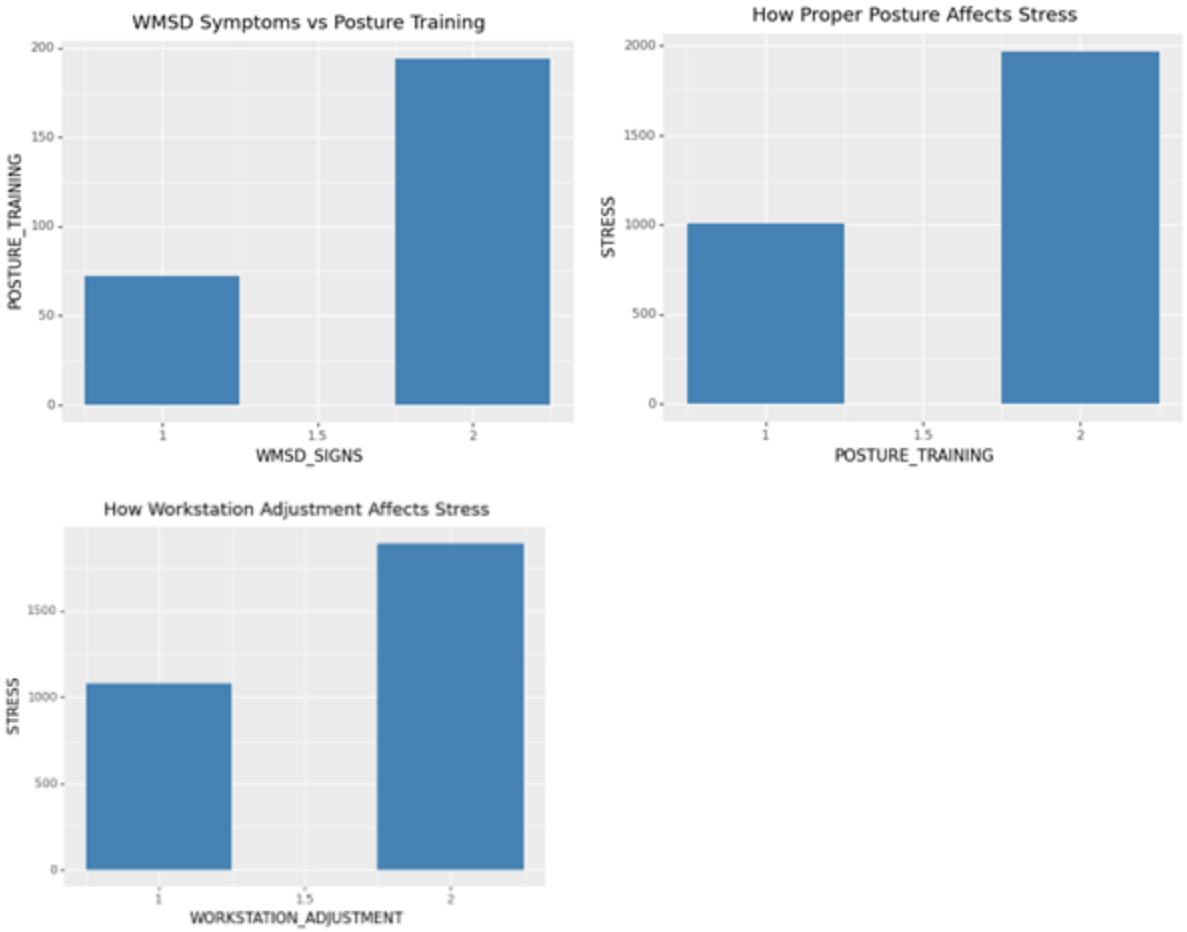
Similar to the age distribution, we analyzed the industry distribution among the participants in order to give us more context. Out of 69 total participants throughout the 6 month study, 24 participants, or 36.4%, work in healthcare.



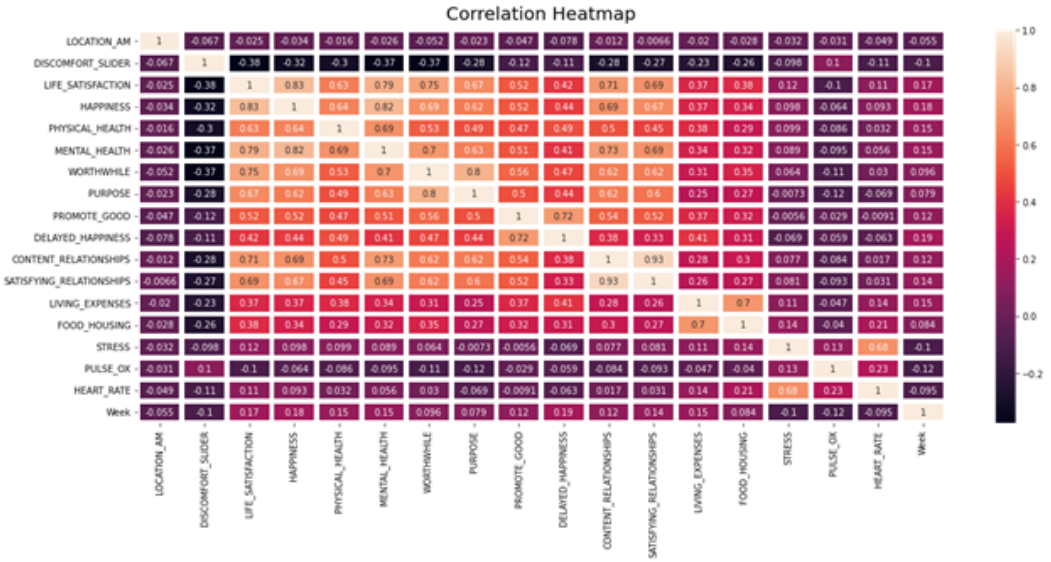
We realized that since quantitative reports such as heart rate and stress levels may not have been adequate to show meaningful findings, we should also examine more closely any qualitative or ergonomic variabilities in participants' environments. Based on information gleaned from correlation heatmaps of the Computer Workstation and Friday AM data, we included more factors in our research to produce a better understanding of how they affected overall mental and physical health, in addition to the number of locations each participant worked each week.



Shown above is the correlation heatmap used to analyze the Computer Workstation data. We found that posture_training, methods_training, wmsd_signs (Work-related Musculoskeletal Disorders), and workstation_adjustment (where workers were taught when and how to adjust their workstations to avoid musculoskeletal discomfort) were closely related and plotted it against information about stress to see how the presence of these types of practices and symptoms affected it.

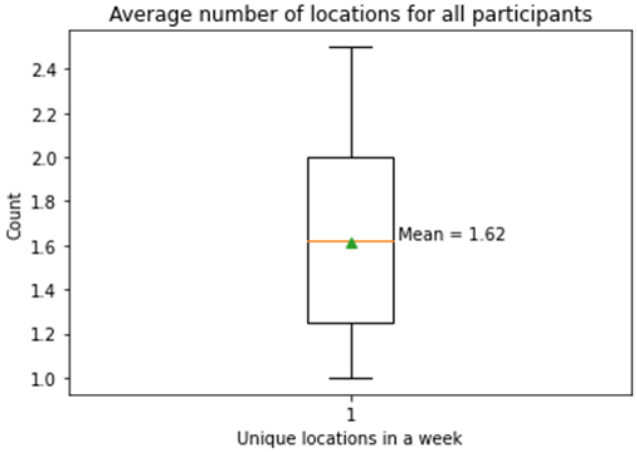


In all three barplots, where 1 is “True” and 2 is “False” for the presence of signs or prior training in proper posture and work methods, we can see that stress tends to be lower when the worker did receive proper training and higher when the worker did not receive proper training.

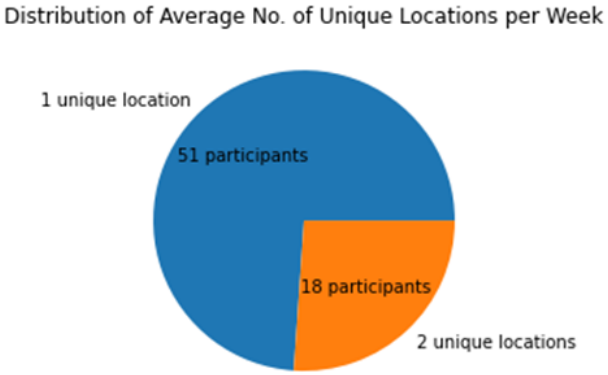


Data Analysis based on Hypothesis

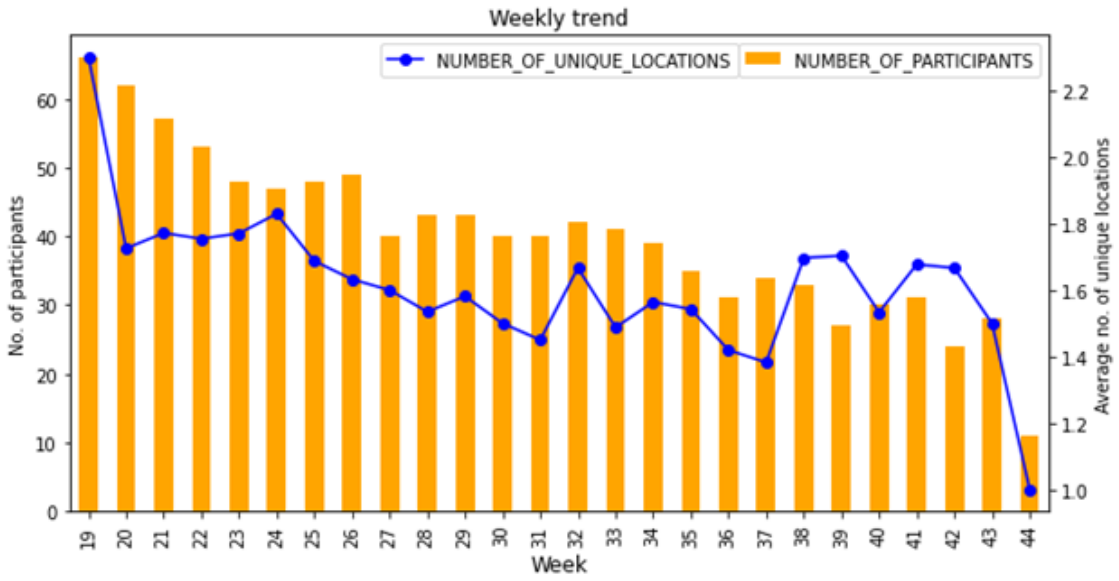
Hypothesis: Participants will work from an average of three different locations per week (including different rooms of the same house).



Based on our observations, we found that participants actually work from an average of **1.62** different locations per week. This is based on the collective data from the start to the end of the study at 6 months and is found by grouping daily_end data, which is recorded daily at the end of a participant's day, by each participant and week and finding the average number of unique locations that they worked in. Our findings also suggest that the highest average number of unique work locations did not even reach the value of 3, and overall the results disprove our initial hypothesis.

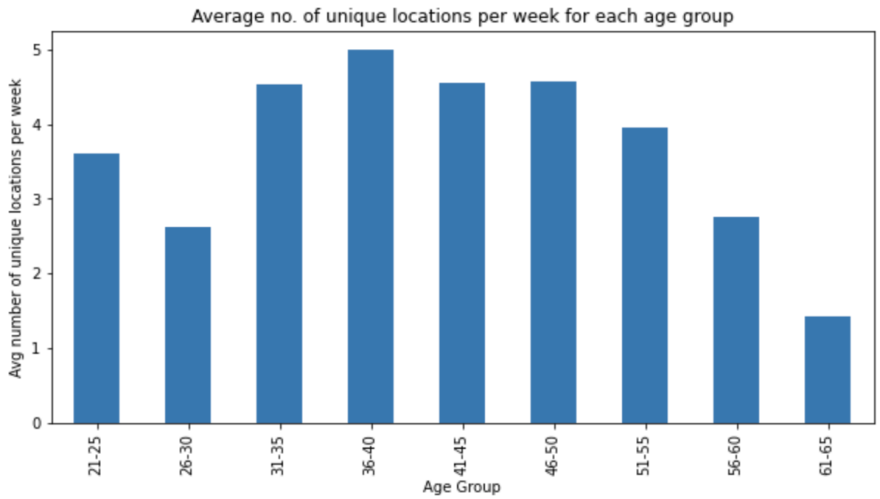


Rounding the averages down into whole values, we analyzed the distribution of the average number of unique locations per week for each participant and found that a total of 51 participants (73.9%) worked in only 1 unique location in a week and that the remaining 18 participants (26.1%) worked in 2 unique locations in a week.



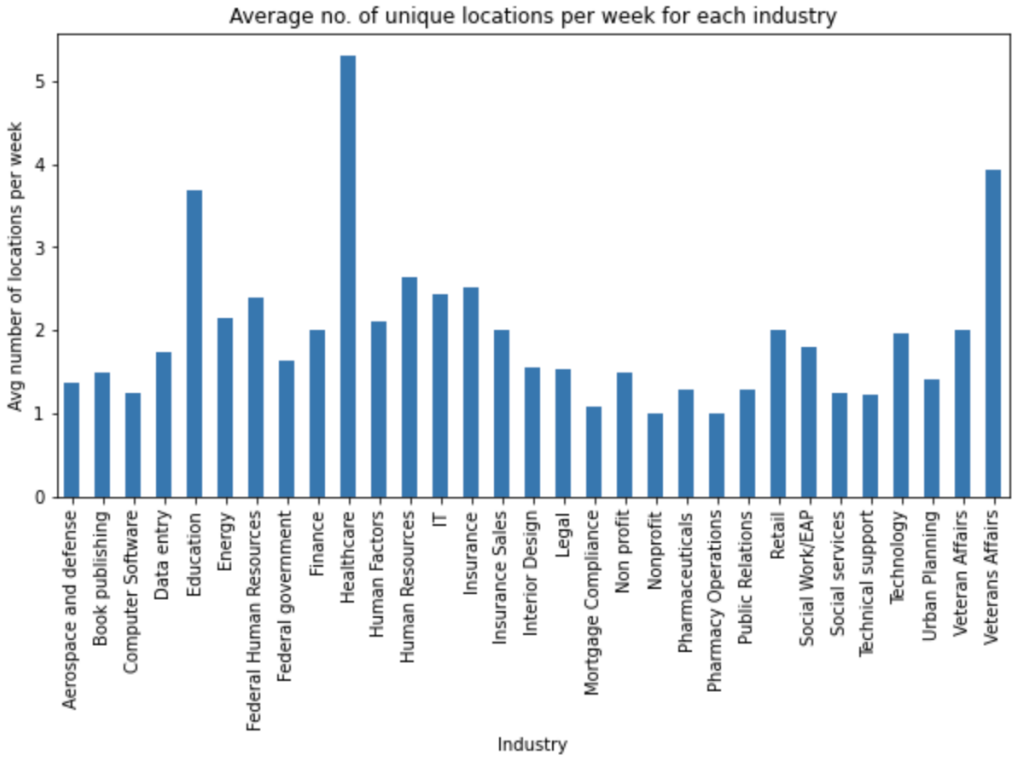
Throughout the entire 6 month duration, we noticed a decrease in number of participants involved in the study and believed that this trend and information are crucial in further analysis, as there had been a considerable decrease from 69 participants at week 19 to only 12 participants at week 44. Therefore, we analyzed how our insights related to the hypothesis had also changed over time. The figure shown above proves that the average number of unique locations in which participants worked at had decreased overall, which changes in accordance with the decrease in number of participants. Based on this analysis, we noticed that a large pool of participants would work in a high number of unique locations in one or a few number of weeks and then work only in one location for the remaining weeks of the study. This variation is the reason why we can find an increase followed by an immediate decrease between different weeks and a few drastic changes similar to one between week 19 and 20.

<u>Age distribution of participants in relation to the number of locations they work in a week</u>



We observed that participants between the ages of 36 and 40 tended to have higher variation in the locations they worked at and had the highest average number of unique work locations in a week. This defies our initial expectations that the 21-25 age group would be more likely to work in different locations. In contrast, we found that participants between the ages of 61 and 64 had the fewest variation or the least number of unique locations in a week. Although there is not enough data to support this, for the older demographic, less physical mobility due to age could be a factor worth noting.

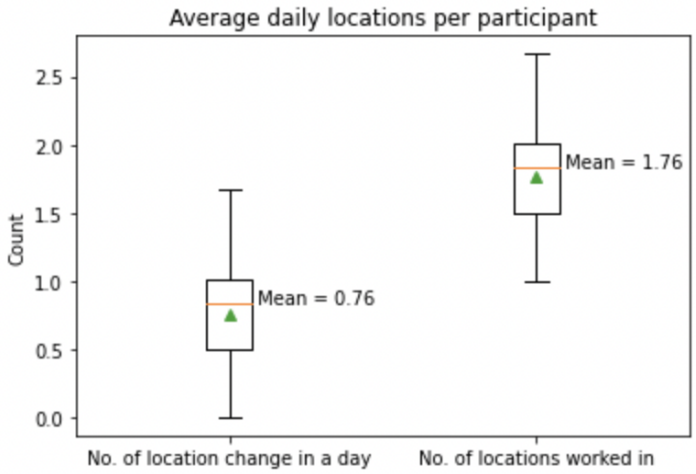
Industry/work distribution of participants in relation to the number of locations they work in a week



Based on the figure shown above, participants in the healthcare industry worked at an average of 5.31 different locations and had the largest variation in terms of average locations worked per week, followed by Veteran Affairs in second (average of 3.92) and Education in third (average of 3.68). It should be noted, however, that we had a disproportionate number of participants working in healthcare, which meant that it may not accurately depict the other industries as much.

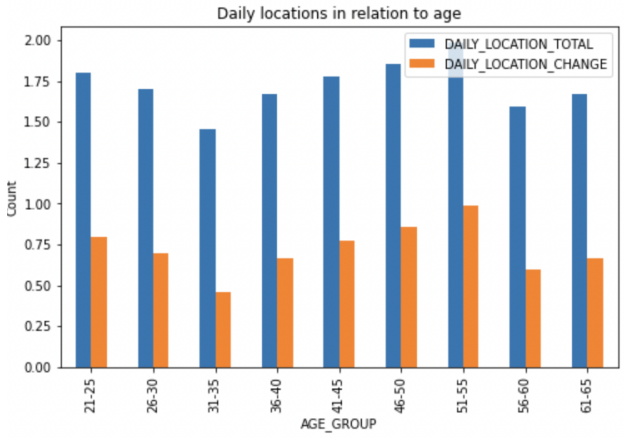
Hypothesis-based Data Analysis in the Scope of Daily Activities

Diving further into the main hypothesis that we worked on, we thought it would be valuable to analyze and get insights on the average number of locations a participant worked in a day and how other features may correlate to it.



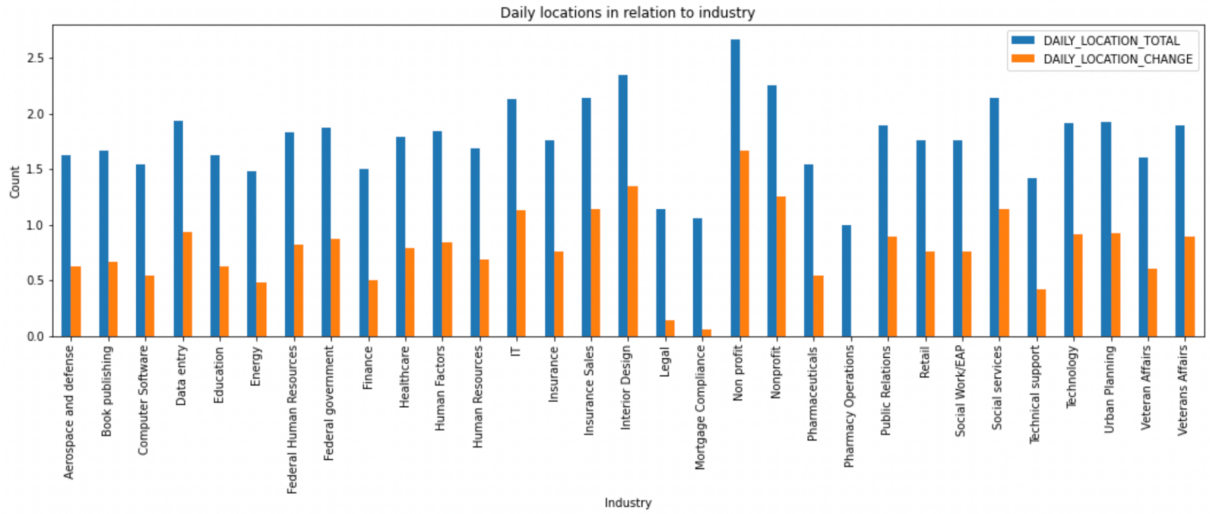
According to our analysis, we found that participants worked at an average of **1.76** unique locations in a day, which is also equivalent to an average of **0.76** number of location changes in a day. This result is derived from merging daily_AM, daily_PM, and daily_END data together and included data points for participants that had data recorded for at least 2 of the datasets. The result also factors in room changes in one location; for example, a participant that worked in location A in the morning, location B in the afternoon, and ended up in location A at the end of the day would indicate that the participant worked in 2 locations and had 1 location change in a day. Based on the weekly findings, it was no surprise that there was not much difference when it came to comparing the locations per participant in the scope of daily settings.

<u>Age distribution of participants in relation to the number of locations they work in a day</u>



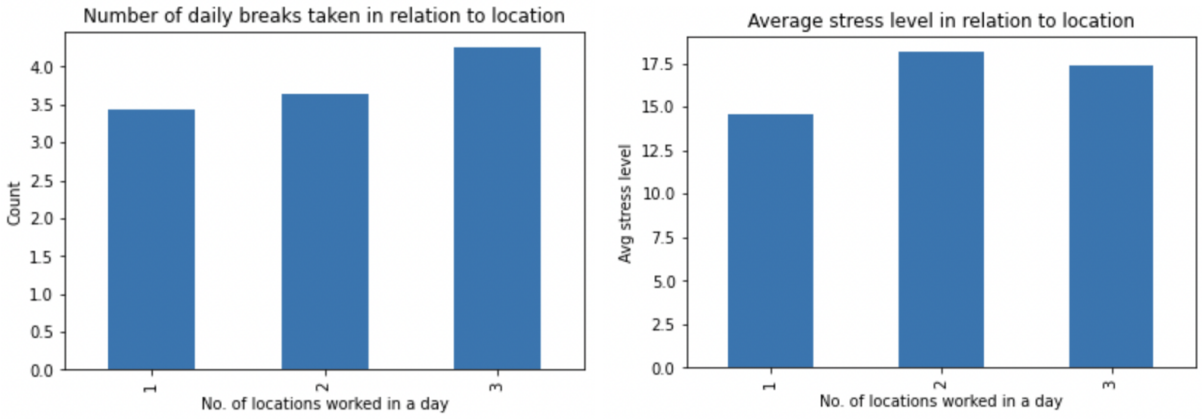
We observed that participants between the ages of 51 and 55 tended to have higher variation in the locations they worked at and had the highest average number of unique work locations, which was at 2 locations in a day. Surprisingly, we found that other age groups had roughly similar numbers in terms of how many unique locations they worked at in a day, and we found that using a compiled data of 6 months–without analyzing the changes in monthly intervals–participants had worked at a rough average of 1 unique location in a day.

Industry/work distribution of participants in relation to the number of locations they work in a day



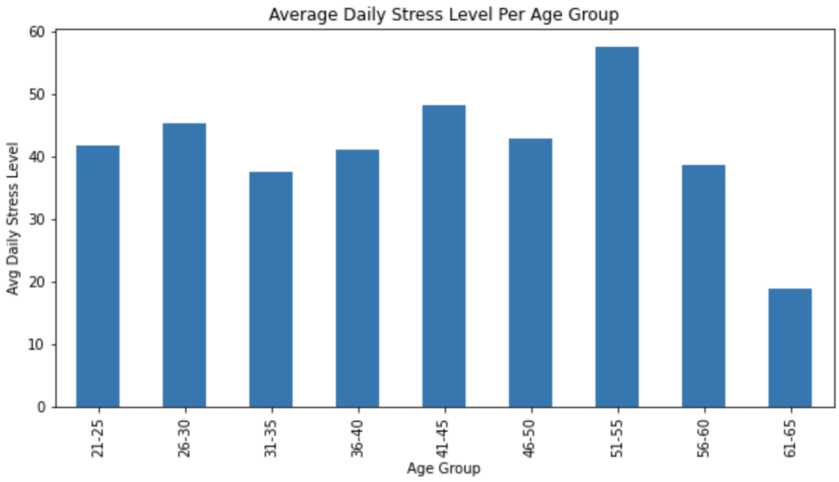
From the figure shown above, we found that participants working in nonprofit organizations worked at an average of >2.5 locations in a day and had the highest variation among participants working in other industries. We also found that participants working in pharmacy operations tended to work in only 1 location in a day and had the least variation in terms of unique locations.

Other insights related to the number of locations participants worked in a day

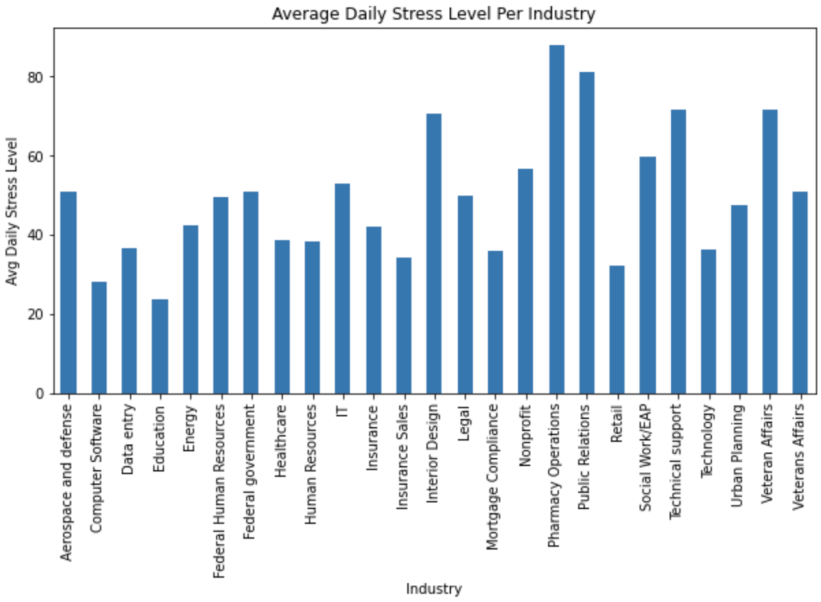


From the compiled 6 month data, we also viewed how other variables compare to the average number of locations worked in a day. As shown in the first figure, there is a positive correlation between the number of locations worked in a day and the number of daily breaks taken. Participants that worked at an average of 3 locations took an average of more than 4 breaks in a day. When looking at the stress levels, the average stress level per day surprisingly does not seem to have a strong correlation with the average number of locations worked in a day. People who only worked in one location did have lower stress levels but people who worked in two locations had higher stress levels than people who work in three locations.

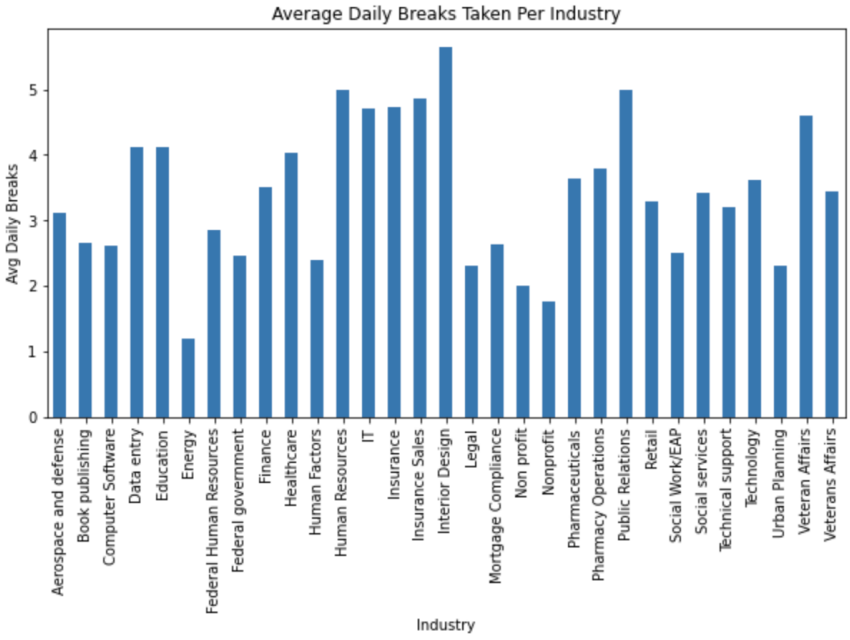
Further Insights based on Daily Data



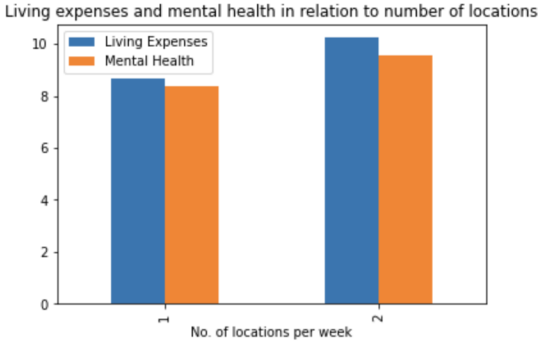
Utilizing the compiled daily data over the span of 6 months, we took the maximum stress level that an individual experienced in a day and compared the results across participants of different ages and occupations. We observed that stress levels are fairly similar amongst all the age groups except for 51-55 year olds whose stress levels are significantly higher and 61-65 year olds whose stress levels are lower. This stress level is measured based on an algorithm within the Garmin watch.

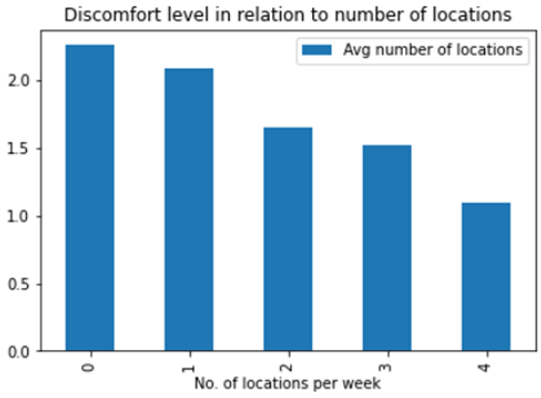


When it came to grouping participants by industry, we found that on a daily scale, participants working in pharmacy operations have the highest stress levels and that participants working in education have the lowest stress levels.



As we looked into other variables, we also noticed that participants in the interior design industry took the most amount of daily breaks while participants in the energy industry took the least amount.





We also analyzed how living expenses played a part in relation to the number of locations and the participants' mental health. Both participants' mental health and living costs levels increased as the average number of locations increased. However, mental health did not increase in proportion with living costs, implying that living costs do have an effect on mental health along with the number of average locations. On the other hand, we noticed that there was a decrease of discomfort levels as the number of locations in a week increased, which suggests that a change in location or setting would help alleviate discomfort. In general, it appears that working from home decreased the overall health of the participants as they have less mobility working from home.

Challenges Faced

We have noticed there is a disproportionate amount of participants working in healthcare in relation to other occupations. This can play a role in massively skewing the data and potentially creating bias in our results, perhaps for certain age ranges. There are an innumerable amount of factors that play a role in the overall health of people and it would be impossible to take all of them into consideration. Many of the factors that we did take into consideration were also difficult to quantify, in particular stress levels that were determined by a Garmin algorithm, which can lead to possible discrepancies. The Garmin watches were effectively the only authoritative source in the data, and thus it is difficult to say whether they adequately captured the more nuanced and subtle aspects of a person's overall health, such as the impact of social isolation or the demands of balancing work and household responsibilities.

The survey data used is also based on individual measures with no prior standards. As a result, factors such as discomfort or pain level can be misrepresented and purely based on prior experience of the participant, which can vary when compared to other participants. Additionally, people may not always produce accurate reports if they are not aware of certain external factors that could be affecting their health. If the measurements are inaccurate, they could also play a role in skewing the data incorrectly.