**Group1:**

Our data came from a class survey and the survey had (11) questions. The data consist of 26 observations and (11) variables as follows: sushi, pet, early riser, age, major, height, Hug Length, School work hours, social media hours, Credit hours, Exam and we are interested with the major variable, social media hours, sushi, and pet ownership. We analyzed the data by creating a new data frame with a column containing both the CS majors and other-majors, and we checked for normality by visualizing the data with Histogram, F test, box plot, and QQ plot.

**In group 1 question:**

Is the average time spent on social media significantly higher for Computer Science?

We decided to use the one-tail Two Sample T-test method since we are comparing the means of two groups samples (computer science majors and other-majors) which indicated that we had a parametric test and, ideally this method was the most appropriate in this scenario.

This test makes four assumptions:

**Assumption 1:** Are the two samples' independents?

We determine that the two groups are mutually exclusive based on the prior knowledge we have about the data and how was collected, and we didn't see any relationships between each observation.

**Assumption 2:** Are the data from each of the two groups follow a normal distribution?

We began the analyses by checking if the data in each group follow a normal distribution using the "Shapiro-Wilk" test to check normality. We also used the QQ plot, and Histogram to check for normality for both groups in two separate graphs one for the CS majors and another for the other majors. Finally, we summarized both mean and standard deviation for the groups.

**Assumption 3:** Do the two populations have the same variances?

We used the F-test on both data sets to check our assumption regarding the variances.

**Assumption 4:** Were the samples obtained using random sampling?

Our approach was to check if there was any randomization done after data collection.

**Group2:**

In this group we used the same analysis as groups 1. The data was analyzed by creating two bar graphs to check the percentage of people who like sushi or pets, and we shift our interest to two different variables:

the sushi and pet ownership. In this group we went with the question to check the association between a person's preference for sushi and pet ownership. After looking at this question, we determined that there were two categorical variables present in the data, and for this type of scenario usually, the best fit will be a Chi-Square test of independence. This test was used to determine whether there is a significant association between two categorical variables.

**In group 2 question:**

Is there an association between a person's preference for sushi and pet ownership?

This test makes four assumptions:

**Assumption 1:** Both variables are categorical?

For testing if both variables are categorical, the two groups have nominal cells that identify that we had two data sets with categorical variable

Data representation:

sushi ("yes", "no")

Pet ("yes", "no")

**Assumption 2:** All observations are independent?

We established that none of the individuals were counted more than once. We also determined that each student had only one observations value related to them.

**Assumption 3:** Are the cells in the contingency table are mutually exclusive?

We used the contingency table to check if both groups are mutually exclusive.

**Assumption 4:** Expected value of cells should be 5 or greater in at least 80% of cells.

We visualized the contingency table results to check if 80% of the cells had a value grater then 5 and not less then 1.