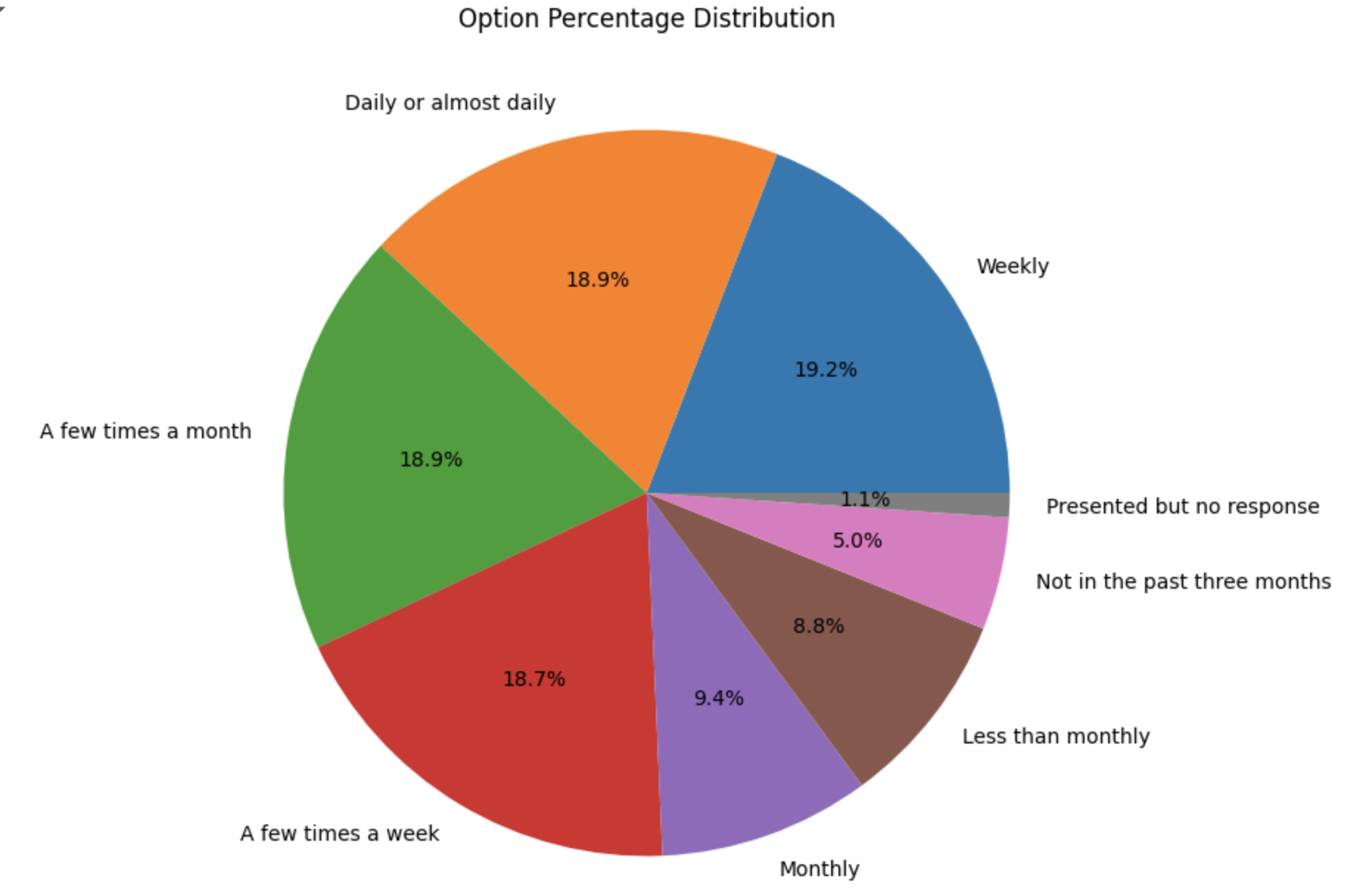
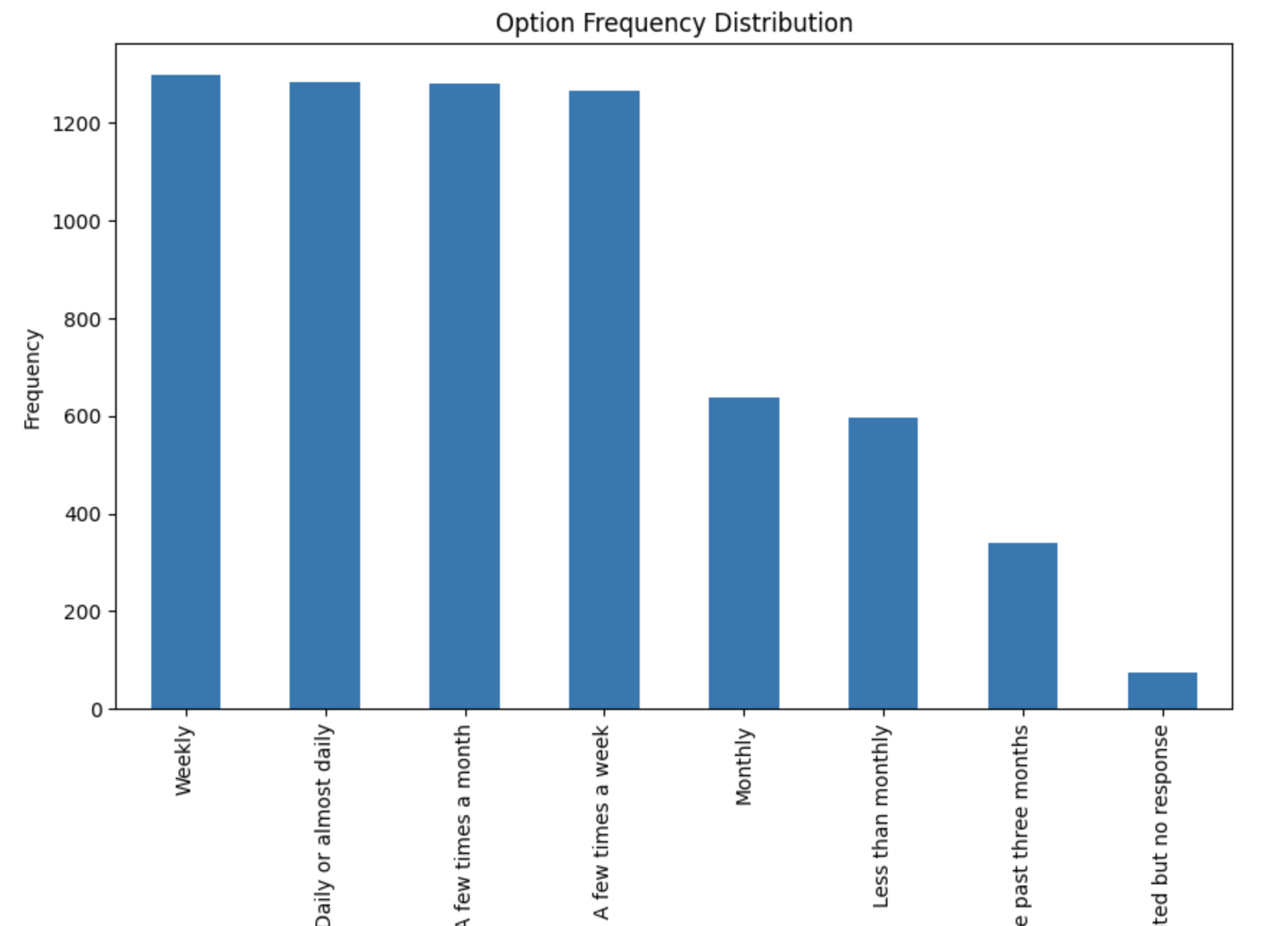
STA130 course project individual proposal

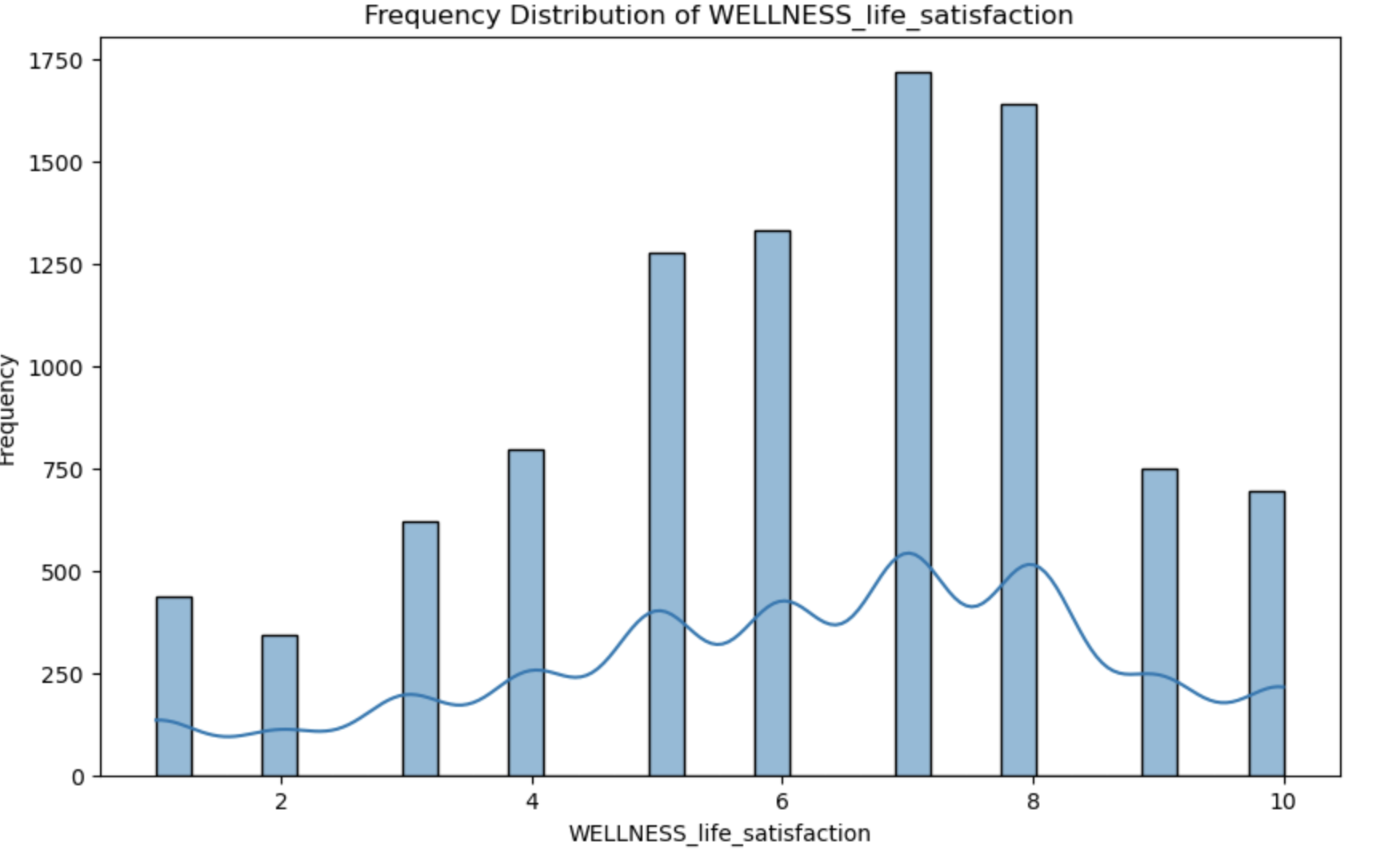
Analysis 1:

Variable1: In the PAST THREE MONTH, how often have you... - had a phone conversation with a friend or family member?-Q117\_7

Variable2: On a scale of 1 to 10, How do you feel about your life as a whole right now?-Q16

Analysis plan:

I plan to determine how people rate their lives based on how often they talk to friends or family, which fits the purpose of describing how social connections affect people’s well-beings. Since these two data are of different data types, I used descriptive statistics to show the distribution of the variables. For variable 1, since the data type is object, I used python to analyze the frequency and percentage distribution for the occurrence of each option. To better understand the data, I also analyzed mode to identify which option appeared the most times. For visualization, I plotted histograms and pie charts, as displayed below.

By looking at the charts, I found that most people call their friends or family weekly (19.2%), daily (18.9%), several times a week (18.7%), or several times a month (18.9%). This proves that most people have relaxed conversations frequently, which may also explain why most people rated their lives between 5-8 out of 10. Frequent conversations with friends and family allow people to express their emotions and feelings and have someone to talk to when negative feelings arise, thus enabling them to adjust their moods in a timely manner and ensure that they are positive and optimistic most of the time. For Variable 2, I calculate the mean, median, variance and standard deviation for the data, since the data type is integer. I also draw a frequency table and box plot for variable 2, which is shown below and also can be seen in my notebook.

Variable 2 has a mode of 7.0 and a mean of 6.28, which proves that most people rate their lives highly. I believe that high frequency of communication with family and friends may be the reason why the overall ratings are all high.

Analysis 2:

Variable 1: age(How old are you (in years)?)-Q4

Variable 2: number of closed friends(How many close friends do you have?)-Q19

Analysis plan:

I want to determine the relationship between people’s age and number of closed friends they have. I will do linear regression analysis in order to reach my goal. First step is to plot a scatterplot graph to see if two variables have obvious linear relationship. If the data points are distributed roughly along straight line, these two variables may be correlated. Next step is to calculate the correlation coefficient. If the correlation coefficient is close to ±1, there may be a strong linear relationship between age and number of closed friends. The further step is to fit a linear regression model to the data and examine the slope to see if it is statistically significant. By looking at the regression coefficient and p-value, I can see whether there is a significant linear relationship.Then, I need to check the R-squared value of the model, which indicates the proportion of variance explained by the model. The closer it is to 1, the better the linear fit.

I would argue that the results of my data analysis may provide evidence of a negative correlation between age and the number of close friends.Since as age increases, people feel more tired to keep contact with large number of friends, therefore will keep only small amounts of closed friends. If my prediction for result is true, we should increase awareness for mental healthiness of the elderly, since less number of closed friends may cause negative emotions since they don’t have strong social connections. This helps to reach the goal of project, which is to raise awareness of social connections and well-beings.

Analysis 3:

Variable 1: How many times per day do you visit social networking websites (e.g., Instagram, Facebook, Twitter), on Average?-Q54

Variable 2: How often do you have meaningful social interactions?-Q32

Analysis plan:

Since data type for both variable 1 and variable 2 are categorical, I can’t do mathematical data analysis for it, which is calculate its mean, median, variance and standard deviation. I can only do frequency count. For data visualization, box plot and histogram are drawn to show understanding for the distribution of the frequency of people visiting social network platform and distribution of having meaningful conversations. In order to see how these two variables affect each other, I would focus on observe whether the two sets of data have similar or significantly different distributional trends by comparing the distributional shapes, areas of concentration, and tail shapes of the two sets of data. If the regions where datas are highly concentrated(mode) in the two variable datasets are in opposite direction, then it is likely to indicate that if people spend a lot of time on social media, they may be neglecting to communicate meaningfully with others as a result. I believe my result will prove more time spent on social media platform reduces the frequency of having meaningful conversations. By analyzing the relationship between time that people spent on social media and the frequency for people to have insightful discussion, we can see how social media affect people’s well being. Exploring and focusing on the connection between time spent on social media and people’s well-being in the analysis coordinates with the theme of our project, which is to analyze factors in modern society that impact mental health.