For this project I had 4 main functions. I performed much of the cleaning of data in Pandas, I created the summary stats along with graphs, I managed the repo and helped with our git issues, and I helped solve issues with the JavaScript for the site we created. For the cleaning of the data, we had one key issue with the features and activities columns. We wanted to separate each of the items in the columns, but this presented a problem in that even though the items in the two columns looked like lists they were strings. After lot of googling for a long while I was able to find a method for converting the items to a list, move each item into their own column, and converted each of them into Boolean values. I then converted them to ones and zeros so they could be used in our machine learning right away. I also used Pandas to create some new data frames and created some graphs from those items. Adam got a lot of the working code for the JavaScript. My main function was adding ideas and solving some of the issues we had. One issue that I was able to solve was getting our map on the site to zoom in on the markers that showed up based off our filters. I ended up creating a group each time a new set of filters and placed that group inside of bounds. I placed this inside our function for creating our map. My greatest challenge was mentioned above, and that was getting the features and activities columns separated. Even though I was able to find a function to help me, it did not work right away. I figured out that I also needed get the unique values in the two columns and add them to a dictionary as a key so they could be used as columns in the new data frames I was creating.

Our team worked very well together. We set up meetings for Friday, Saturday, and Sunday for each week and met on zoom for a few hours as needed. We also worked together on slack and mentioned ideas we had as they came up. One thing l liked that we did was that we created our own slack channel to communicate. This really helped us streamline our ideas. Our key team strength was that we worked together to solve our problems and we kept open communication on everything we were doing.

The key topic for our assignment was to answer the question: can a trails popularity be predicted by its features, and if so, can we find some hidden gems, or trails that haven’t really been discovered by many, but would be enjoyed by the user. For the machine learning portion of the project, we used a few different modules. The first was the random forest module. We used this to get a ranking of the most import features and activities in the data set. We also used Vader to get a sentiment analysis on the reviews of these trails that were pulled from scraping trip advisor. The sentiment analysis was used to get a score based on the sentiment of the reviews and adding that back as a column to our original data set. We had issues with adding the sentiment analysis back to the data set, and tried to use FuzzyWuzzy, but we were unable to get an acceptable matching score when attempting to match the items from the two data frames we created. We were able to add a map to a JavaScript site that we created along with filters including a filter for trails we predicted to be hidden gems. We ended up using a confusion matrix to predict the trails popularity based on its features, and we had an F1 score of 72. For an app like this I feel that this was a success, as this is suggesting that a person would like a specific trail.