Self-Assessment

Entering this project with only the knowledge gained from this bootcamp created a nervous feeling. I was questioning myself on whether I could be of any use as I don’t have a background in this field of data analytics. Using the information from the instructor and TA that I had the privilege of working under, I was able to begin expressing my ideas and input when attempting to figure out what we (as a group) wanted to accomplish with this project. It became clear that each one of us had become proficient in certain areas of the course modules more than others and with that, we were able to divide the project into equal segments. I took on the responsibility of taking our data set and converting the results from a text-based answer into an integer. This would mean that each answer given within our data set would have a corresponding number. For example, if the answers given were “yes” or “no” then the code written would change the answer to either 1 or 0. This code was also used to convert answers such as “never”, “Somewhat”, “Rarely”, and “Very Frequently” into numbers. Once this was completed and the data set has been saved as a new csv file, it was then passed off to one of my partners who then used it within our machine learning model.

With that completed, my next task was to create the visualizations using Tableau. We came to an agreement that age played a large factor in our data set therefore I used age for each of my visualizations. I started making heatmaps for each type of music which would allow for easy representation of the effects of music per age using different colors to show the amount of people within that specific grouping. Next, I used a pie chart to show the amount of use per streaming service. I then create a bar graph the show the amount of mental illness per age group and how much music affected the results. With my part done, I was in communication with my team members and picked up any extra pieces. We each had an understanding that we all may have specific parts to play but if someone needed help, then we were there to figure out the issue together. During the model building process, we each helped to discuss how to make our data work to get the answers we needed. With that only being one person’s responsibility, we still gave our input to help alleviate any confusion or issues. I think the biggest takeaway from this experience was whether you know the answer, just speak up and no question is a stupid question. You never know who might also be thinking the same and can help you figure it out.

Project Summary

To summarize, this project was a challenge of everything we learned during this bootcamp. We had to create a way to communicate with each other that didn’t require the use of a laptop or computer. We chose to utilize our Slack platform and create a “Final Project” group to allow us to communicate with each other at any time. This proved very useful as we all have full-time jobs and don’t work from home. Even during a pre-planned trip, I was able to communicate with my team members over Slack and help answer any questions or issues found.

We each had our own strengths and weaknesses but showed that asking questions and conversing with one another ultimately brought us close and led us down the correct path for the answer we were looking to achieve.

The topic we chose was music’s effect on mental health. We chose this topic specifically because as we discussed with each other what we want to do, we all had music as a common suggestion since it has personally had its own effect on each of our individual lives.

We used several models in order to work correctly with our data set. The first being a “SVM” or Support Vector Model. The second model is a “SMOTE” model as we found the need to use a model which would allow for oversampling and handling a large amount of input variables.

The results proved to have lower accuracy than the first test but made better use of the data given due to the change from SVM to a SMOTE model. The precision decreased but the no effect and worsen classes would now show a true value as opposed to before when using the SVM those values would show either 1 or 0. The biggest issue we found with the final results of this model was that the data set simply was not large enough to have a more accurate result.