**2.2 Project 1: Project Check-In/Milestone 2**

1. *Any surprises from your domain from the data?*
   1. No surprises here. I have actually found it fun to work on a project with text. In the past I have dreaded the process, but Spark has made things easier. I have fallen in love with the Pipeline function. It takes a lot of confusion out of the text preprocessing process.
2. *The dataset is what you thought it was?*
   1. The dataset is just what I thought it was. It contains the variables I need to create a model very easily without tacking on superfluous information I would need to drop anyway. There was a misconception about how many classes were in the dataset that will be outlined in the next section.
3. *Have you had to adjust your approach or research questions?*
   1. I have had to adjust my approach a couple of times. When I started my first attempt at modeling I expected to use logistic regression. There are three classes in the data (good, neutral, and bad). After my first model I only had predictions for two of the variables (good and bad). This sent me back to the drawing board to figure out why there was no predictions for the third class. I’m at the point where I think I might need to create the model using another method. After trying a decision tree, I came to the same results, only two classes. Needless to say, I was very confused. Went back to the beginning and looked at the data and discovered my mistake. In the description of the dataset on Kaggle it mentions three classes, but the dataset only comes with two of them. All of that time wasted because I didn’t look at my data more carefully.
4. *Is your method working?*
   1. Besides the modeling issues above, the methods I have been using have been working well. Spark is not giving me any major challenges, the few that I have discovered will be outlined in the next section. One method that I am worried about working is creating a live dashboard in Python with Django. I am not as familiar with creating a server with Django (I am more well versed in NodeJS). If I encounter any issues I might change my methods and create a dashboard with NodeJS and Express.
5. *What challenges are you having?*
   1. The challenges I have been encountering have been with Spark running on my computer. I have a Mac with good RAM, memory, and processor. When I tried to model I encountered lots of Java errors about the connection being refused. I decided to move my code over to Databricks. At first, I found it a little bit clunky, but in the end, I was able to transfer my code over and have a notebook that runs without any error. Another minor challenge I might encounter is trying to export the model from Spark. I haven’t yet researched how to do this. This could lead to some minor challenges down the road.