Wrangle Report

Introduction

This project was taxing. It began with downloading and installing files from three separate sources. Due to not being able to uses the Twitter API process. I used the download and install from the resources provided. This presented it own unique challenges as attempting to initially read the file and create a dataframe stalled my progress. I was able to overcome this hurdle by using pandas to read and extract the columns within the file.

Assessment

When assessing the data I began by looking through the sizes and shape of the three files. This was able to show when there was missing values and what kinds of column, and data, was available to be looked at. This helped in decided which columns and rows could be removed for better data and clarity purposes. The next step I took was to view sample rows of the data. This led to some interesting revelations, such as None replacing NaN values. It also showed that the Source column was unintuitive to read. It provided insight as to the p1 and its related columns were trying to convey. Due to the lack inability to confirm the breeds with certainty it was decided that the most probable breed selection would determine the breed and the breed and the column would be renamed for clarity. I also relied on manual visual assessment to learn that the name values had not been transferred completely.

Cleaning

The cleaning section reminded me of the third project in the nanodegree. Dropping columns and dropping rows were a familiar technique. I had to rename columns which was unfamiliar, but a solution was reached in stack overflow. The merging process was very reminiscent of programming with SQL and felt very comfortable to complete. Upon my second submission I received feedback that dropping columns fell more under the Quality assessment. This limited the Tidiness to one item, to balance I used a similar extract method from the numerators on the doggo and other slang descriptors into one column.

Revision

The feedback provided was very helpful when converting the ID to a string. There was lessons to be learned in the presentation. I did have to submit to addition research when researching how to update the numerators and learned how to search within strings and extract partial data within the strings.

Conclusion

This project was challenging. It definitely help expanded my knowledge on the process of converting data within python. This was well appreciated and the ability to transform data within frame and save the converted files as necessary as much appreciated and will continue to be studied and practiced moving forward.