

wrangle_report

July 12, 2019

1 WeRateDogs Project Report

Weratedogs is a Twitter profile that rates the cuteness level of various dogs that are sent to them. The goal of this project was to analyze the data that was obtained to ultimately see which dog breeds are the cutest. The project has worked with three separate data files that contain various relevant data. For analytical purposes, I used python's pandas to explore, clean, and create visualizations for the data set.

The first steps with this project were storing the data files into appropriate datasets, including scrubbing the JSON file to make it manageable. After making copies of the files and storing them into their respective datasets, I was able to merge the data into a single CSV file for analysis. I then was able to visually and programmatically assess areas in the data that required corrections or updates. I noted these issues down underneath the exploratory phase of my project. After exploring and assessing the data, I then proceeded to clean and correct the issues that I found within the dataset. This including correcting data types, extracting values from data, removing retweets, etc. I broke down each problem into three sections: Define, which outlined the issue and how I was going to address it. Code, the code used to solve the problem, and Test, the code used to test the data to verify the code worked. Following the cleaning, I then proceeded to create programmatic visualizations for the data which included graphs about highest rated dog breed as well as which dog stages were the most popular.

From this project, I was able to put my programming skills to the test. I was able to learn how to fully assess and clean dirty data and how to think about the problem ahead of time before leaping into trying to analyze the data.