

Data Wrangling

Introduction

This report is to briefly describe my wrangling efforts, the steps taken, and the decisions made. This will be divided into how Did I gather the data then how I assessed the data then how did I clean it.

Gathering

To gather the data, I used three resources. The first resource was from twitter archive containing all information about the tweets but missing some. The second resource was the image predictions which is the dog's prediction provided on Udacity website which is downloaded using the request library. Finally, the Data from twitter API which contains more information than the one in the archive. It is gathered using the twitter API.

Assessing

My way in doing this is after each resource I gather I look into the data information and I get familiar into how the data looks like then I start digging more and matching data from the different resources to validate the integrity of the data. The Final step of assessing the data is to write each note and mistake (Quality or tidiness problems) in the data to prepare it for the next step.

Cleaning

In the cleaning, I used the Define, Code then test method. I started with each problem by defining it and classify if it is quality or tidiness problem. After that I start coding the solution then testing it if there is a mistake, I return a step, or two steps back as needed.

Conclusion

In conclusion, Assessing the data is the most difficult part for me since it takes more time to see large data and understanding it. While the gathering is the most straight forward in this project.