The process of going through this project went through the data wrangling stages. That is, gathering, assessing, and finally, cleaning. For this project, the gathering had to be done from three different sources.

The first data source was a csv file, provide on the Udacity servers and had

to be downloaded, and loaded into the jupyter notebook.

The second data had to be obtained programmatically using the Python's request library, which was a really cool experience.

The third and most difficult in my experience, was requesting for data from Tweety, Twitter's API to obtain information from a json file.

My request to the API failed, so I used the provided data of what I would have had should the request gone through.

To gain useful data, I had to merge the different datasets.

I assessed the data visually, and programmatically to have a feel of where to

correct, general information like number of columns, the shape, the datatypes of the various columns and other relevant information.

Since the requirement was to look for about 8 different quality issues and 2 different tidiness issues, I came up with these ten issues to correct them:

Uder quality these 8

1. Duplicates From Retweets
2. Errors in Datatypes in timestamp column
3. Invalid dog names
4. Rating denominator not equal 10
5. Erroneous extraction of numerators
6. Tweet_id(s) from df table do not have corresponding images in the imgage_pred_tsv data
7. Rows with NaNs for expanded_urls column.
8. Source phone used in tweeting scattered and need categorisation

and Under tidiness, the two issues tackled were;

1. Separate datasets which need to be merged i.e. df and image_pred_tsv datasets
2. The different dog stages should be merged to one column

Each of these issues then went through the standard for addressing issues in data analysis wrangling stages.

I defined each issue, wrote code to correct them, and finally tested them to see if they have indeed been solved.

Finally, I came up with visualizations. These various visualizations made it easy to come up with insights from the data.

Among the insights include the most popular dog stage, which was pupper, the most source device tweets came from, the trend of count of tweets over the period from 2015 to 2017.

stage.