

Wrangle Report

This is a list of the quality issues and Tidiness issues that I fixed in the report:

Quality Issue

- 1) Remove non-null entries in `retweeted_status_id`, `retweeted_status_user_id`, `in_reply_to_status_id`, `in_reply_to_user_id` and `retweeted_status_timestamp`.
- 2) Drop columns: `retweeted_status_id`, `retweeted_status_user_id`, `in_reply_to_status_id`, `in_reply_to_user_id` and `retweeted_status_timestamp`.
- 3) There is inconsistency between lower and uppercase names for `p1`, `p2`, `p3`. Use lower to make all `p1`, `p2`, `p3` names lowercase.
- 4) Some dog types under `p1`, `p2`, `p3` are not dog types so remove these rows by seeing if `p1_dog`, `p2_dog` and `p3_dog` are all True.
- 5) Drop columns: `p1_dog`, `p2_dog` and `p3_dog`
- 6) Convert `Tweet_id` columns from int to str.
- 7) Convert timestamp column from object to datetime.
- 8) Denominators do not add up to 10 so multiply the numerator by whatever you multiply the denominator by to get 10.
- 9) Drop Rating Denominator column.
- 10) Replace None, an .etc with NaN under name that are not actual names.

Tidiness Issue

- 1) Merge `doggo`, `floofer`, `pupper`, `puppo` into one column.
- 2) Merge all tables together (done above)

The quality issues that were addressed were done through numerous commands such as `lower()`, `drop()`, `isnull()`, `query()`, `apply()` etc. and addressed the numerous flaws within the data to reach a point where the data could be analyzed without glaring anomalies that would damage the impact of analysis. By moving to merge all of the tables together before analysis allowed me to fully comprehend how I wanted to piece the entire puzzle of data together. Although there were some drawbacks to easily visually assessing the entire dataset when it was put together, I felt that having all of the data side by side was worth the hassle. Furthermore, dropping the columns that were not structurally necessary provided a much clearer picture of the dataset and relieved a lot of the hassle of trying to view all of the columns side by side.