Wrangle Report

1.Quality

1. some of the names in p1, p2 and p3 are start with an uppercase letter and some with lowercase letters, we should make them all lowercase.
2. p1_conf, p2_conf and p3_conf are 6-digit decimal values we should round it up to 3 digits
3. 'tweet_id' type is integer while it should be string
4. 'in_reply_to_status_id' and 'in_reply_to_user_id' contain 2278 missing values, we won't use the data in these columns so we will drop it
5. retweeted_status_id', 'retweeted_status_user_id' and 'retweeted_status_timestamp' contain 2175 missing values, they should be dropped because we won't use retweet data.
6. 'name' has 'None' 745, 'a' 55, 'an' 8 and 'the' 7 as names, these are not names so we will replace it with nan value
7. 'rating_denominator' has the number 0 we should handle the 0 because it's invalid.
8. 'source' values are always one of these 4 categories [Twitter for iPhone, Vine - Make a Scene, Twitter Web Client, TweetDeck] we should extract the cateogry from 'source' column.
9. Source column dtype is integer we should change the data type to category.

2.Tidiness

1. these columns 'doggo', 'floofer', 'pupper', 'puppo' represents dog 'stage' so we should add a new column called 'dog_stage' and merge them toghether in it.
2. - columns names are not clear and doesn't make sense to a reader so I will change it to make it clearer and more understandable for readers.