Wrangle Report

1. Gather

-The data had to be gather in 3 different ways. One file was directly given, the other had to be programmatically downloaded from a page and the other had to be created by querying Twitter's API.

-Create the 2 first DataFrames was relatively simple. I was also able to construct the loop which extracts the data from Twitter's API but struggled with saving the file correctly (otherwise, I would have to wait about 30 min each time I opened my notebook.

2. Assess/Clean

-In this stage, the data had to be assessed and then organised to make it easier to understand.

-I first did a visual assessment and a programmatic assessment. The main quality and tidiness issue for each table were then noted in a markdown cell below.

-The main issues with the data lied in the twitter_archive table. Certain issues were complicated resolve, like the fact that certain names were invalid. There are several reasons for this: no name was given to the pet, or the pet wasn't a dog. Since the dataset is large, I chose to adopt a simplistic approach and drop these rows.

-The key cleaning steps involved creating a master DataFrame which compiled the 3.