WRANGLE REPORT

The dataset that was given was the tweet archive of Twitter user @dog_rates , also known as WeRateDogs, additional data via the Twitter API and the Image Predictions file.

The steps taken in wrangling the data are as follows;

- Gathering data
- Assessing data
- Cleaning data
- Storing data
- Analyzing and visualizing data

In the gathering stage the tweet archive was given, and was downloaded manually and uploaded into the workspace. Additional data was gotten through the Twitter API using the tweet IDs in the tweet archive. The additional data was required to get the tweet's retweet count and favorite count for the tweets in the tweet archive. The last dataset the Image Predictions file was downloaded programmatically using the Requests library through the given URL.

In the assessing stage the gathered data was assessed visually and programmatically for quality and tidiness issues. The datasets were assessed visually in the Jupyter Notebook (internally) and also through Microsoft Excel software (externally), the datasets were also assessed programmatically by using pandas' functions. After the assessment eight quality issues and two tidiness issues were discovered.

In the cleaning stage, all the issues documented while assessing were addressed. Before the cleaning, a copy of the original data was made, and the copy was used in the cleaning. During the cleaning the define-code-test framework was used to clearly document the steps taken

while cleaning the data. During the cleaning process the three datasets were merged into one master dataset.

In the storing stage the cleaned master dataframe was stored in a CSV file named "twitter_archive_master.csv".

In the analyzing and visualizing stage, the saved cleaned dataframe was read into the Jupyter Notebook and the data was analyzed. After analysis, four insights were produced and derived at, with two visualizations to aid the insights for better presentation.

After all the steps were taken the data was successfully wrangled and the Jupyter Notebook was saved as an html file.