Report

Introduction:
The We Rate Dogs project is a data wrangling project that involves collecting, assessing, and cleaning data from various sources to create a final, cleaned dataset. The data for this project was gathered from three sources: the twitter_archive_enhanced dataset, which was manually downloaded, the image_predictions dataset, which was obtained using the requests library, and the twitter_json.txt dataset, which was also obtained using the requests library. The goal of this project was to clean the data and merge it into a single, cohesive dataset that could be used for further analysis.

Gathering:
The first step in the We Rate Dogs project was to gather the data. The twitter_archive_enhanced dataset was manually downloaded from a website, and the image_predictions and twitter_json.txt datasets were obtained using the requests library. However, due to the developer account not being verified, the Tweepy API could not be used to gather the data.

Assessing:
After gathering the data, the next step was to visually and programmatically assess the data for any issues. This included checking for null values, duplicates, and inconsistencies in the data. The twitter_archive_enhanced dataset was found to have several null values, and the image_predictions and twitter_json.txt datasets were found to be relatively clean.

Cleaning:
The next step was to clean the data. This involved dropping null values from the twitter_archive_enhanced dataset, changing data types for the tweet_id and timestamp columns, capitalizing dog names, merging all lifestages into one column, removing extra strings on the source of the tweet, and finally merging the three datasets into a single dataset. The cleaned dataset was then saved to a new csv file for further analysis.

Conclusion:
The We Rate Dogs project was a data wrangling project that involved collecting, assessing, and cleaning data from various sources to create a final, cleaned dataset. The data was gathered from three sources: the twitter_archive_enhanced dataset, the image_predictions dataset, and the twitter_json.txt dataset. The data was visually and programmatically assessed for any issues, and several cleaning steps were performed to create a cohesive, cleaned dataset that can be used for further analysis.