

## **WeRateDogs Data Wrangling**

The analysis found in the wrangle\_act.ipynb notebook investigates user engagement of the WeRateDogs twitter feed, digging specifically into possible 1) commonalities of the top 10 favorited tweets 2) changes in consumer engagement with changes in the WeRateDogs tweeting behavior and 3) difference in consumer engagement based off the day of the week.

The data wrangling process of gather, assess, clean was followed. Firstly, three different datasets were gathered that provided data on historical tweets, image predictions and additional attributes. These datasets were collected through use of a provided .csv file, programmatically downloading a .tsv from a website and utilizing the twitter API to source a JSON file. These sources of data were then each reach into a pandas dataframe.

The data was assessed both programmatically through the use of python and visually in Excel. During the assessment step a list of issues was documented and separated into quality and tidiness. This list was then referenced during the cleaning process.

Lastly, the data was cleaned using the define, clean, test approach. During the cleaning a few additional quality issues emerged, which isn't uncommon as data wrangling is an iterative process. These additional issues were documented and cleaned during the cleaning process. Lastly, the databases were merged, resulting in one tidy database.

With the data wrangling completed, the Exploratory Data Analysis could commence. The cleaned data was analyzed through the use of pandas and a variety of methods such as query(), sort\_values(), and groupby(). Visualizations such as the scatter plot, histogram, bar charts and line graph were integrated to assist in bringing the consumer engagement insights to light and were of high importance when seeking trends at the engagement versus internal tweeting behavior and when perusing the day of the week data.

The data was then saved in a .csv file named "twitter\_archvie\_master.csv) and the notebook clearly commented for others staff members to use.