Report on

Wrangle and Analyze Data Project for FWD
By Amr Menshawy

- Introduction

This project to analyze tweeter account **@dog_rates** also known as **WeRateDogs**.
WeRateDogs is a Twitter account that rates people's dogs with a humorous comment
about the dog. These ratings almost always have a denominator of 10.
also part of the project to provide some insides,
A lot of process required like (Data Gathering , Data Assessing , Data Clean) will be
covered in this project last stages are sort and export the data also visualize this data to
show this insides.

- Data Gathering

Part from this project as we spoke are data garthering from different sources almost these data
was provided on the Udacity classroom like (twitter_archive_enhanced , image_predictions ,
tweet_json) the type of the source (csv,tsv,and json) ,we will deep inside these sources to
merge it to one file and start our data assessment.

- Data Assessment

  After we finished the data gathering and merge the data to one file we need to visual
  the data by excel also Coding by using pandas some useful library tail() , head() ,
  value_counts(), dtypes(),nunique() we found a lot of issues and tidiness which will you
  start clean it in the next part

- Data Clean

  After we made our assessment we found data quality issue and the tidiness we will use
  the pandas and numpy libraries to fix these issue and clean the data also testing the
  data to make sure that data is clean and ready to be used for sort, export and store.

  1. Change timesatmp columns
  2. Clean source columns
  3. Dog stage
  4. Drop retweets
  5. favorite_count & retweet_count change to int
  6. Remove unneeded columns
  7. Rename columns
  8. Fix rating
  9. tweet_id: is int, should be type object as no calculation is needed
  10. correct missing_names
  11. Merging the three dataframes, using the common tweet_id
  12. The puppo/doggo/floofer/pupper terms should be in one column

13.