act_report

June 25, 2022

0.1 Report: act_report

- Create a **250-word-minimum written report** called "act_report.pdf" or "act_report.html" that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

1 Data Wrangling Project

1.0.1 D.K. Oluwadare

1.1 Introduction

This project demonstrates the data wrangling process for the tweet archive of Twitter user @dog_rates, @dog_rates is a Twitter account that rates people's dogs with a humorous comment about the dog. In this analysis I demonstrate the data wrangling techniques that were used to gather, assess and clean the dog twitter archive.

1.2 Project Overview

1.2.1 Gather data

The following files were gathered for the analysis:

- The WeRateDogs (@dog_rates) Twitter archive - This file (archive.csv) was downloaded using Twitter's API and consists of basic tweet data for 2300+ tweets from WeRateDogs.
- The tweet image predictions - i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. This file (image_predictions.tsv) was downloaded programmatically from Udacity.
- Each tweet's retweet_count and favorite_count -This file (tweet_json) contains JSON data for each tweet indicating the retweet and favorite counts.

1.2.2 Assess data

The three files obtained in the gathering phase were loaded into individual Pandas data frames for assessment. Each of the data frames were evaluated visually and programmatically.

1.2.3 Clean data

The quality and tidiness issues were cleaned using programmatic techniques such as:

```
Dropping unnecessary columns from the tables
Removing rows that consisted of null retweets
Removal of rows with duplicate information
Deleted rows that did not have any dog predictions at all
Combining all three data frames into a single data frame
```

In [ ]: