

WeDogRate report

This project entailed data wrangling, which consists of gathering ,assessing data and cleaning data, Based on the project detail given about @dog_rates also known as WeRateDogs . I was instructed to gather each of the three pieces of data as described below in a Jupyter Notebook titled wrangle_act.ipynb

- a. The WeRateDogs Twitter archive,with link: twitter_archive_enhanced.csv.
- b. 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 hosted on Udacity's servers and downloaded programmatically using the Requests library and the following URL:https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv
- c. [tweet json.txt](#)

I gathered each of the data,firstly by importing the necessary packages such as pandas, numpy, requests,json matplotlib.pyplot e.t.c. then the gathered dataframe was name into different dataframe from (a),(b)(c) into :

- I. twitter_enhanced
- II. image_prediction
- III. twitter_df. Respectively.

After that have been done, I went to the next stage of data wrangling which is Assess. I did Assessing by visual assessment and programmatic assessment where issues of quality and tidiness arose starting from the columns I didn't need for that moment ,erroneous datatypes,erroneous columns ie four columns instead of one column in the twitter_enhanced table, different casing font in columns and unknown names of dogs e.t.c which I spot, noted and rested for the last stage of my data wrangling

Before cleaning the data need to be copy why the original keep for future preference while the copy one is used in the cleaning processes . the copy from I, ii, and iii named : df_enhanced, df_image and twitter_clean.finally cleaning started step by steps each data with structural issues of each dataframe been handled first resulting from irrelevant columns was merge to form a single columns,doggo,floofer,pupper and puppo column stages merged to dog_stages. After that was out of way I merged the three dataframe to form a dataframe steps by steps called df_table_clean . finally was the content issue from unknown dog names was erased from the specific dataframe column,wrong datatype was corrected to its appropriate datatype, and some datafame column was dropped. unknown dog names was erased from the dataframe column. at last my data frame was ready to be analyzed, visualized and stored.