Data Analysis Nano Degree

Data Wrangling Project

This project is part of Udacity's Data Analyst Nanodegree. It is based on the "WeRateDogs" Twitter page, a page which that rates pictures and videos of dogs out of ten. To show how kind dogs are, all of WeRateDogs' ratings are above ten. They also tag each dog with a different category : "doggo", "floofer", "pupper", or "puppo".

Udacity provided an archive of this Twitter data for WeRateDogs' tweets as a CSV file to facilitate this project. Two more sources of data were also gathered as part of this project:
predictions for which type of dog is present in each picture and additional tweet information acquired from Twitter.

The three steps of data wrangling: gather, assess, clean performed successfully and carefully throughout this project. In gathering phase, the image prediction data was downloaded using Python's Requests library. The additional Twitter information was downloaded using the Twitter API(after registering an account at twitter and getting necessary keys). Next was assess step, an inspection carried out on the generated data frames in order to find any quality or tidiness issues like type mismatch and having different columns to express one kind of data and so on. Finally cleaning step took place to fix the quality and tidiness issues that were previously identified (in previous steps).

As a result of data wrangling process an insights bullet indicates the final result after some analysis such as:
•The most popular dog, based on image predictions, is a Golden Retriever
•The most popular dog type is a "pupper"
•The median and mean ratings of numerator 10.86 and 11 respectively

Hazem Nuseir
Feb 2020