Data Analyst Program

Wrangle and Analyze Project

By

Aisha M AL-mutairi

introduction

My goal in this project: wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The Twitter archive is great, but it only contains very basic tweet information. Additional gathering, then assessing and cleaning is required for "Wow!"-worthy analyses and visualizations.

In this project I was worked on the three data set

- **Enhanced Twitter Archive**

  The WeRateDogs Twitter archive contains basic tweet data for all 5000+ of their tweets, but not everything. One column the archive does contain though: each tweet's text, which I used to extract rating, dog name, and dog "stage" (i.e. doggo, floofer, pupper, and puppo) to make this Twitter archive "enhanced." Of the 5000+ tweets, I have filtered for tweets with ratings only (there are 2356).

- **Image Predictions File**

  a table full of image predictions (the top three only) alongside each tweet ID, image URL, and the image number that corresponded to the most confident prediction (numbered ١ to ٤ since tweets can have up to four images).

- `tweet_json.txt`

  I gatherd each tweet's count and favorite ("like") count and save it in one table

  I will talk about the steps of this project

- **Step one : Gathering Data**
  I gathered three pieces of data
  - "twitter-archive-enhanced.csv"
  - "image_predictions.tsv " - (downloaded programmatically using the Requests library)
  - "tweet-json.txt" (Gather each tweet's retweet count and favorite ("like"))

- **Step two : Assessing Data**
  After I gathered all three pieces of data, I was assess them visually and programmatically for quality and tidiness issues. And then I was document to clean them in the next step

- **Step three : Cleaning Data**
  After the assessing and document I have made a copy of the original data before cleaning and then I define ,code , testevery issue.
  I have successfully cleaned all issues .

- **Step four: Storing Data**
  After the cleaning I have saved the gathered , assessed , and cleening to a CSV file named ("tweeter_archive_master.csv")

- **Step five: Analyzing and Visualizing Data**
  I have produce three insights and three visualization

References

- https://www.w3schools.com/
- Udacity lessons