Report

7-8-2022

On Analysis and Insights into Final Data

Introduction:

Real-world data is rarely pristine. The dataset you will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog rates, commonly known as WeRateDogs. WeRateDogs is a Twitter account that scores people's pets and includes a hilarious comment about the dog

This project works through the data wrangling process, concentrating on data collection, analysis, and cleansing. There is also visualization and observation from the analysis offered.

Gather:

Data was acquired for this project from the following sources

The Twitter archive of WeRateDogs. Udacity students were given the 'Twitter-archive-enhanced.csv' file

This collection includes basic twitter data (tweet ID, date, content, and so on) for all 5000 tweets as of August 1, 2017

Tweet image prediction, i.e., a neural network predicts what breed of dog (or other objects, animal, etc.) is present in each tweet. This file was made available to Udacity students (Like me)

Twitter API using Python's Tweepy module to capture at a minimum each tweet's retweet and favorite ("like") count and any other relevant data.

Data Assessment:

After gathering the data, I proceeded to evaluate it in terms of both quality and tidiness

In terms of quality dimensions, there are four major issues to consider

1.Completeness: Incomplete data

2.Validity: Is the data logical

3.Accuracy: Incorrect data

4.Uniformity: Standardization

Furthermore, there are three primary needs for tidiness:

1.Each variable is represented by a column

2.Each observation constitutes a row

3.A table is formed by each sort of observation unit .

Clean:

Cleaning data is tedious and often iterative. Just when data analyst believe they found all quality and tidiness issue, they often found additional issue arises.

The cleaning process involves three steps:

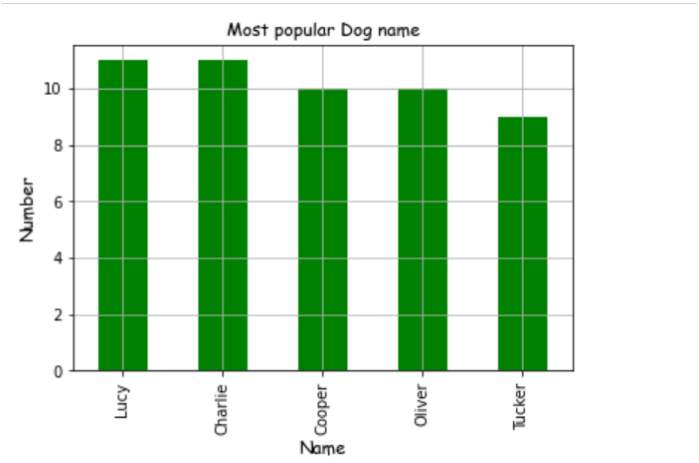
1. Define: Determine exactly what needs to be clean and how.

2. Code: Programmatically clean the code

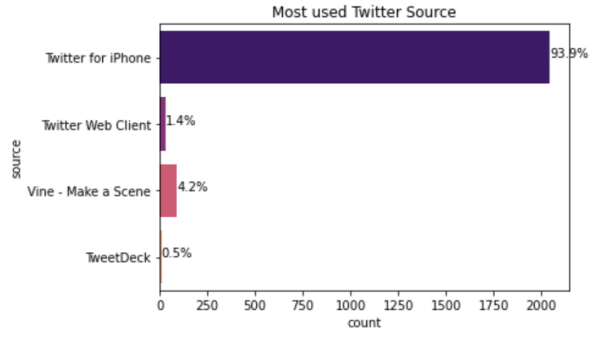
3. Test: Evaluate the code to ensure the data set was cleaned properly.

Analysis and Visualization:

There are serval analysis, which I have done and those are in following:

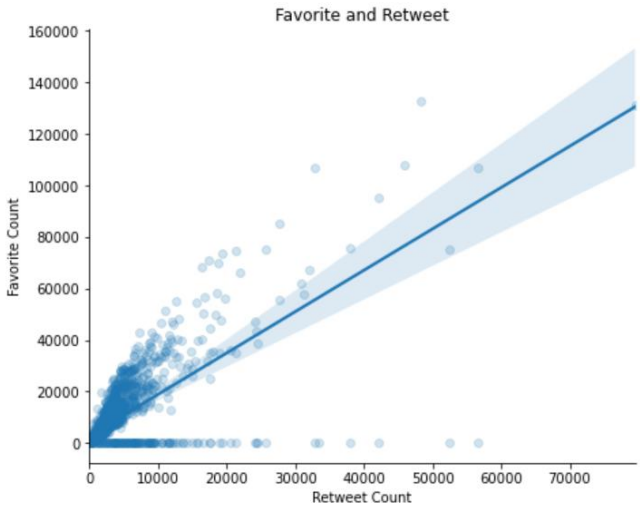


Lucy and Charlie the most popular dog names with(10) ,with the name coper, Oliver close with (9) and Tucker (8).



The users make use of Twitter for iphone ,twitter make a scene and twitter web client and the last tweetdeck.

And form our visualization ,twitter for iphone is the most used by users with people on the platform tweets numbering over 1800 tweets while other clients did'n even hit 250 tweets.



Favourite vs. Retweet Counts:

There is a positive relationship between favourite ("like") counts and how many times a post was retweeted. This correlation is critical for the owner of the WeRateDogs Twitter account to understand when deciding how to improve user traffic on the page. A data analysis team may propose prior articles with a high retweet or like count so that the page owner could model future posts after previously popular pieces.

Conclusion:

The article provides a straightforward look at the data wrangling process. There is so much more that can be done with this data collection.