Udacity - Data Analyst Nanodegree

Project: Wrangling and Analyze Data

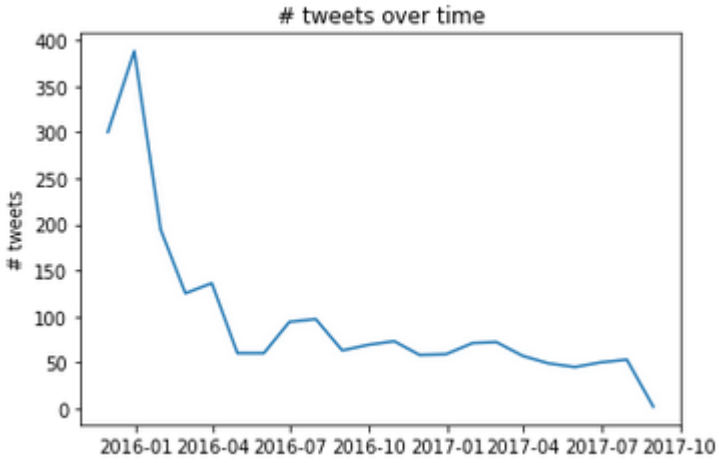
1. Summary

The goal of this project is to gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling.

2. Visualizing Data

- ✓ 2.1 Tweets over time

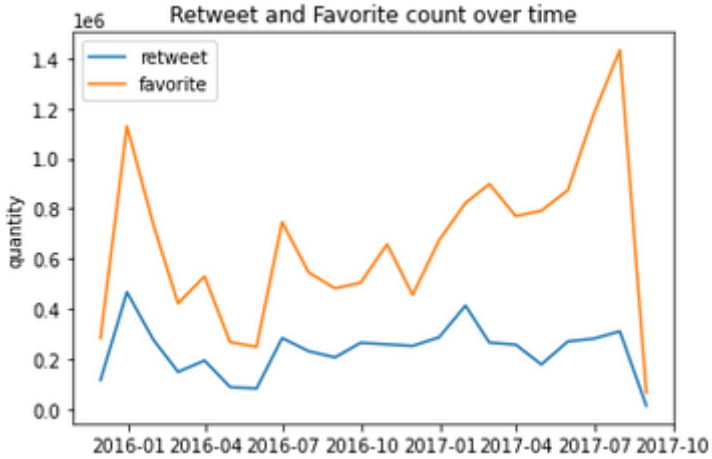
I have grouped the cleaned dataset by field “timestamp” to calculate the number of tweets over time and then I’ve plotted a line chart, as we can in the below image.



Insight: Through time, the number of tweets has decreased, in 2016 there were almost 400 tweets per month, however in 2017 this quantity decreased near to zero.

- ✓ 2.2 Retweet and favorite count over time

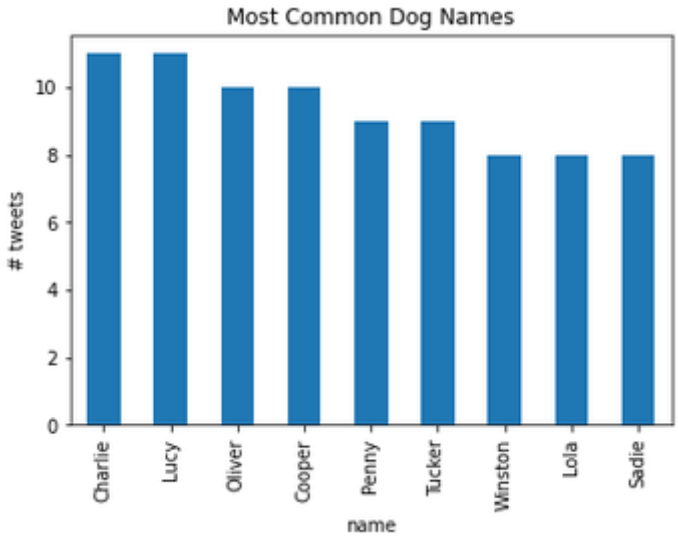
I have grouped the cleaned dataset by field “timestamp” to calculate the retweet and favorite count over time and then I’ve plotted a line chart, as we can in the below image.



Insight: Through time, the number of retweets and favorites have followed a similar trend, the only difference is that the number of favorites has been greater than retweets.

✓ 2.3 Most common dog names

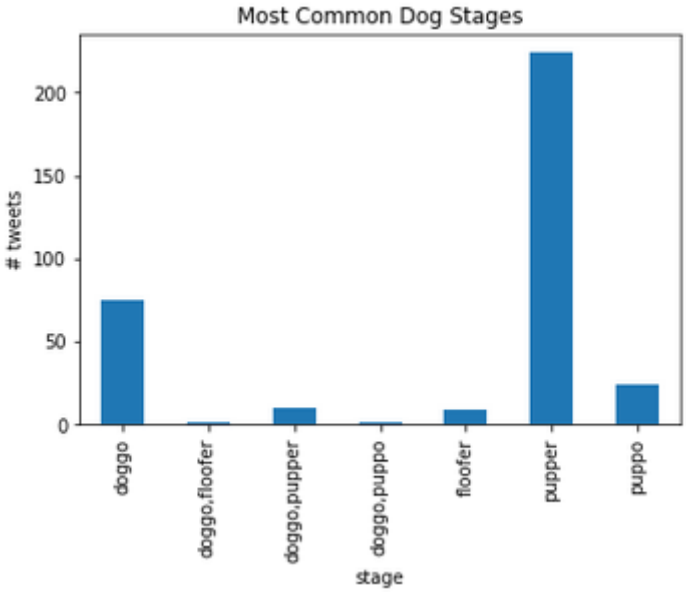
I have grouped the cleaned dataset by field “name” to calculate number of tweets of certain dog names and then I’ve plotted a bar chart, as we can in the below image.



Insight: Through time, Charlie, Lucy and Oliver are the three most common names.

✓ 2.4 Most common dog stage

I have grouped the cleaned dataset by field “stage” to calculate number of tweets of certain dog stages and then I’ve plotted a bar chart, as we can in the below image.



Insight: Through time, pupper is the most common dog stage.