DATA WRANGLING PROJECT

This document briefly describes the final result of the data wrangling steps: gather, assess, and clean. All of them were executed during the completion of the Project: "Data Wrangling", from the Nano Degree Data Science II course, and here are presented.
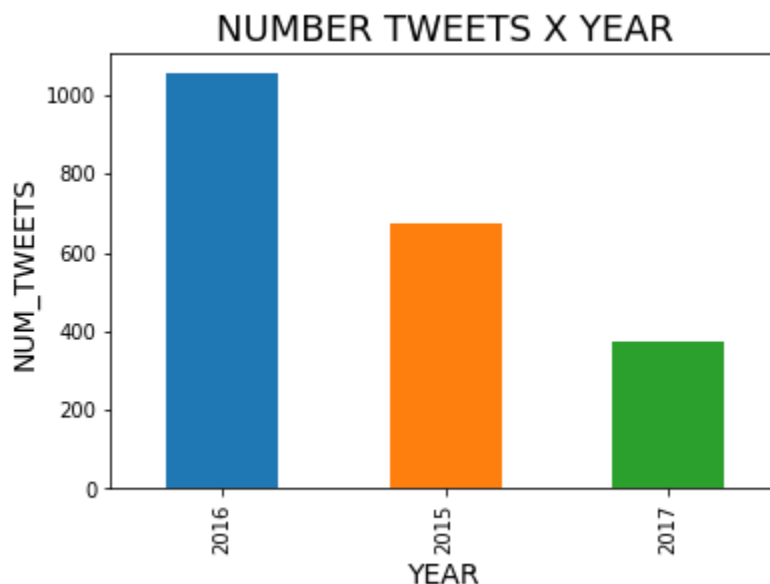
Analysis and Insights into final data

Insight 1

A. Looked for 'peaks' in tweet posting sorting by Year.
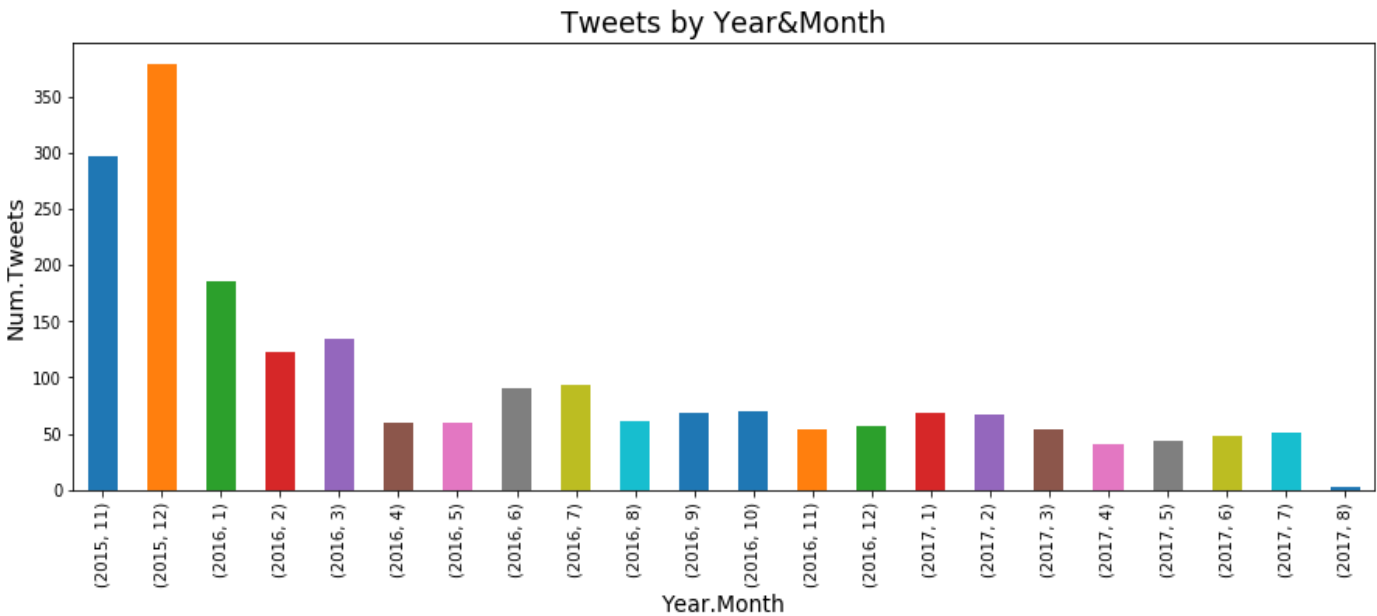
o   Found that **2016** had the greatest number of tweets.

YEAR	NUMBER OF TWEETS
2015	676
2016	**1054**
2017	374

NUMBER TWEETS X YEAR

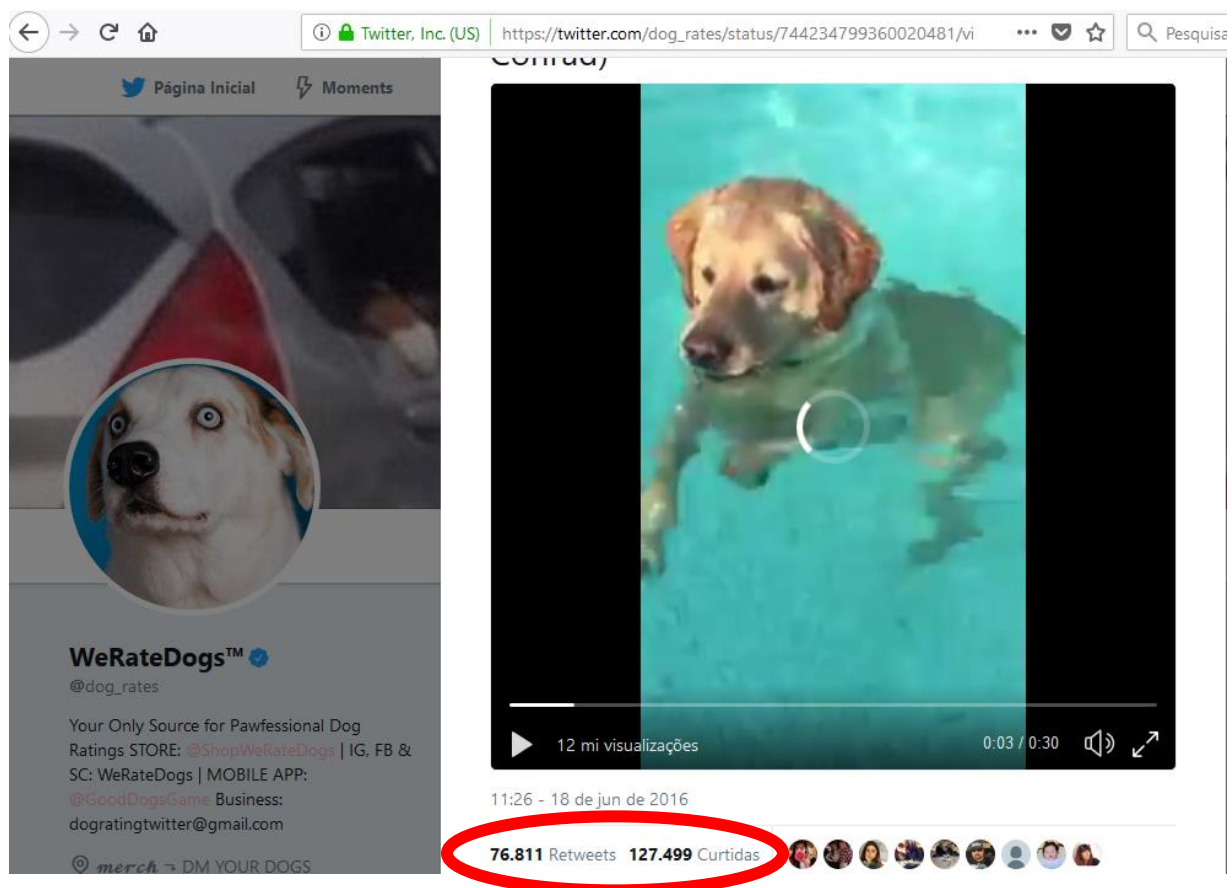**B. Looked for 'peaks' in tweet posting sorting by Year & Month**

- o Found that December 2015 had the tweet peak (379) overall.
- o Found that '2016' seems to have a greater number of tweets overall (1054) because of a biased sample distribution.

Tweets by Year&Month

Insight 2

A. Looked for the most retweeted tweet over the years…

- o Tweet 'ID' *744234799360020481* was the most retweeted, with **76811** retweets.
    - https://twitter.com/dog_rates/status/744234799360020481/video/1

Insight 3

A. Looked for tweet most marked "favorite" ...

- ○ Tweet ID **822872901745569793** is the most marked "Favorite" tweet, with **142504** "favorites".
- ○ https://twitter.com/dog_rates/status/822872901745569793/photo/1

Warm regards,

Rodrigo Neves de Barros
Data Scientist

Document Category: **External**