Data Wrangling & Analysis Project

By Najwa Almulhem
February 3, 2019

In this project, I have gathered, assessed, cleaned, stored , analyzed & visualized twitter archive data for Waterdogs twitter account.

This was a challenging project and I have faced few challenges. In gathering the data, I had to use three different sources: csv file, Requests library over the internet and twitter API. Getting files from csv or from Requests was very easy. However, getting it from Twitter API was challenging, I have requested a developer account from twitter and my request was rejected, then I used the code that was provided by Udacity and it failed in every tweet. Luckily, the json file was also provided for us and I used it.

In Assessing the data, there are a lot of quality and tidiness issues, I was not sure which fields I should use for analysis .After thinking , I have choose the fields that I want or would like to analyze. For example, dog stage, breed , rate, at what time the tweets was created and how many time the tweet was retweeted or liked I had to asses data manually and programmatically.

In cleaning, making the right decision was not easy .For example, I have found that in 3 tweets either numerator or denominator was 0 and after investigation I decided to drop those records. In tweet number 835152434251116546 the images are stolen and the rate was 0 which is unusual .For Tweet 746906459439529985, there is no dog in the tweet. Tweet 835246439529840640 is a reply not a tweet.

I had to iterate the wrangling steps gather asses and clean many times, to make sure I am in the right track.

Overall, this was a very interesting project and it demonstrated the challenges of data wrangling and how much time and effort it takes. But getting the cleaned organized data at the end was very encouraging.