

Project 4: @WeRateDogs (Twitter)

Wrangle report

by Patricia Chandía

This project was a very challenging one. The data wrangling effort I put into it to get the final (master) data set was a lot.

In first place you need to get the data from 3 different sources with different techniques and apply all the contents of the nanodegree so far.

Anyway, it was very interesting to explore how to get data from Twitter because I think I could use it for my research in social sciences.

During the data wrangling process, it was difficult to find the right order to modify the dataframes and it was an essay/error process all the way. I will remember always do a copy of the dataframe to explore ideas, just in case I ruin it (it happened to me, haha).

You need to use the comments to understand what are you trying to get, because sometimes you will continue the work days before and don't want to enface that you don't remember what you were doing.

Also is important to identify concisely de data issues to solve, it's the only way you can go on.

Those data frames are not clean at all, there's a lot of changes to do in the way the data is stored, for example the NaN values are sometimes "None" so a check with `.info()` will told you wrong information.

Finally, I want to thanks to my session leader Luis Malavé for all the support and Udacity reviewers for put such a personalized report available to fix my code.