

# Wrangle Report¶

## Introduction

Nowadays social media is a strong driving force in our day to day communication. Many information remain registered in their domain for years (at least until they broke or collapse their server). We can mine very useful information from Facebook, Twitter and Instagram using common protocols such as API in order to increase our understanding in some specific matter, person, page or target group.

## Objectives

The aim of this work is to go through all steps of data wrangling, that is, gather, assess, clean, store and analyze a whole dataframe. Important notice that this dataframe is full of flaws and inconsistent data so we have a chance to practice the skills we've learned during this course.

## Gathering¶

In order to complete this project I've had to establish a Twitter API using Tweepy and accessing a developer account. Once logged in Twitter I got my consumer key and private token and started downloading WeRateDogs data. I also downloaded data from an image prediction algorithm's output in which image from dogs were put to test.

## Assessing¶

All the data were assessed with regard to its shape, quality, tidiness and elusiveness. The main issues were pointed out after careful consideration and divided in 8 quality issues and 2 tidiness issues.

## Cleaning¶

The main issues were fixed properly using the available tools in Pandas and other useful packages. Some points required a more sophisticated strategy to get fixed.

## Storing and Visualization¶

All the cleaned data were stored in .csv files for further analysis. Later all data were analyzed and some graphs plotted to emphasize 4-5 conclusions. Important notice that some conclusion could be achieved during the previous process: assessing and cleaning.

## Conclusion¶

After going through all steps of data wrangling I could see that many data were discarded because they lack consistency, some were filled in the wrong fields, some were intentionally left blank and some were corrupted.