

# Wrangle report

This project is about gathering, assessing, cleaning, analyzing, and visualizing data, data in the real world always comes with problems, sometimes complex this project helps us improve our skills in the steps mentioned above. This project is defined in several steps which we mention below.

## Gathering

This is the process to obtain our data with which we will play later, the steps were:

- The first file (twitter-archive-enhanced.csv) was provided to us manually, the result was saved in the variable (df\_enhanced)
- The second was downloaded using the **request** library for this [url](#), the result was saved in the variable (df\_predictions)
- The last file was downloaded via twitter api, for this we use the **tweepy** library, finally we parse the result to a dataframe and the result was saved in the variable (df\_tweets)

## Assessing

In this step we evaluate the data, visually and programmatically review the possible problems in the data and think about the possible solutions. Most of all, it is a step to get a general idea of the project.

## Cleaning

Here we are dedicated to eradicating unnecessary columns, we correct and enrich new data as necessary for the final product we want, when these steps are ready we have to merge the data into a single dataframe which will be analyzed.

## Analyzing and Visualizing data

We analyze the resulting dataframe and the different conclusions that we can draw, in addition to visualizing them, but since the project is not based on this aspect we can make basic conclusions of the dataframe, another reason not to draw too many conclusions since we are not conducting an EDA at the data.

## Store data

As a final step, we save the data in a master .csv, which we make available in our repository.

## References

- <https://stackoverflow.com/questions/1323364/in-python-how-to-check-if-a-string-only-contains-certain-characters>
- [https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.dt.day\\_name.html](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.dt.day_name.html)
- [https://pandas.pydata.org/pandas-docs/version/0.25.3/reference/api/pandas.Series.dt.day\\_name.html](https://pandas.pydata.org/pandas-docs/version/0.25.3/reference/api/pandas.Series.dt.day_name.html)
- <https://developer.twitter.com/en/docs/tweets/tweet-updates>
- [https://developer.twitter.com/en/docs/twitter-for-websites/twitter-for-websites-supported-languages/overview\(languages\)](https://developer.twitter.com/en/docs/twitter-for-websites/twitter-for-websites-supported-languages/overview(languages))
- <https://stackoverflow.com/questions/20250771/remap-values-in-pandas-column-with-a-dict>
- <https://stackoverflow.com/questions/23668427/pandas-three-way-joining-multiple-dataframes-on-columns>

