Let's first talk about Data Gathering

I have to collect complement datas (like and retweet count) for tweets that have been already gathered. I have found some difficulties for upgrading my tweeter developer account for having acess to more datas. Indeed Firstly i used `client` object instead of `api` and datas i gathered were not correct. Finally i have upgraded my account and collect the datas correctly.

The gathering of other datas was more easy.

Now, the assessing process

It is important to say that some issues are quickly detectable but others needs more attention.

- The assessing visualisation helps me a lot in terms of comprehension and detecting quality issues.
- The programmatic evaluation comes to confirm what i have observed and helps me to discover the remaining issues either it is quality issues or tidiness issues.

Many others issues could be raised like

- the fact that there are tweets where two dogs are noted at the same time,
- two notes were given to a dog : the first one for joking and second one was the real note Others quality issues, the most about missing values(for example, dog_stage or name columns) could also be raised, but it will be difficult to fix it since we will be obliged to lose almost all the records if we want to delete the records with missing values.
- Some dog's breed have weird name like laptop

The cleaning

I have used pandas methods to clean datas. Sometimes , i have writen my own functions for avoiding repeating code and make the cleaning more easy.

Some cleaning operations like change datatype or drop columns were very easy compared to cleaning operation which requires more reflexions and good ideas like search name in text column or solve tidiness issues.

The project was very interesting and i have take a real pleasure by doing it