

WeRateDogs

Twitter Dataset

Data wrangling project report

Gather:

1. Given csv files
2. Website
3. API of Twitter.

I imported different packages to perform the wrangling process.

Access:

Accessed using different packages like tweepy, os, pandas, numpy, etc.

Clean:

1. Checked the data-types.
2. Removed the NaN values.
3. Data for 5 tweets was unrecovered.
4. Converted Tweet IDs, status IDs, and user IDs from integers to strings.
5. Timestamps to datetime objects.
6. Retweets and favourites from floating-point numbers to integers.