wrangle_report

June 15, 2020

0.1 Report on Wrangling Efforts:

I divided my wrangling journey in three steps: - Gathering - Assessing - Cleaning

1. Gathering: One of the dataset is provided to us in the form of text file i just downloaded it and load it my notebook.The next dataset is on server so i use requests module to access the data from the web.For the last dataset we have an option to get the data from twitter api or just download the given file so due to lack of time i downloaded the file but i'm making sure that i will get the data using twitter api again in sapre time.

2. Assessing: After gathering the three datasets i assess all the datasets first visually and then programmatically and found many issues which are needed to resolve in order to clean the dataset.Whenever i found an issue i differentiate the issue as qulaity or tidiness and wrote the issue in the respective areas.To know whether an issue is of tidiness or not I used three below rules: - Each variable forms a column. - Each observstion forms a row. - Each type of observational unit forms a table. I found many quality issues and some tidiness issues using info(), describe(), duplicate(), value_counts(), etc methods.

3. Cleaning: In this step out of the found issues i cleaned 8 quality issues and 2 tidiness issues.First i make the copies of all the three dataset so that our original data should be safe in case of any failure.For each issue first i defined the issue then wrote the code for it and at last test it as was taught in the lessons.I use some methods like drop(), merge(), rename(), drop_duplicates(), etc.After cleaning i converted the copies to .csv files.

```
In [ ]:
```