wrangle_report

September 16, 2022

1 wragle report

1.0.1 Introduction

On this Project the main aim was to practise and put together all the intersting concepts acquired through learning the Data Wrangling section of the Udacity Data Analyst Nanodegree. Here I wrangled the Data from the twitter archive of the Twitter user @dog_rates.To ensure a little bit of Quality of the analysis,the Data was gathered from three different sources namely,`downloading the file from the internet`, `Programmatically downloading the file from Udacity server` and also `Quering data from a Twitter API`. So on this report I will be outlining my data wrangling efforts.

1.0.2 Project sections

The project consists of three main sections namely:
* Gathering * Assessing * Cleaning

1.0.3 Gathering

On this section I had three different sources to collect data from, there is a CSV file that I downloaded from the internet, TSV file programmaticaly downloaded from the Udacity server and finally the text file queried from the Twitter API using keys obtained from Twitter by opening the Developers account.

1.0.4 Assesing

Various operations were done to display data in different dimension, these operations are called programmatic assesment and then this was topped up by using visual assesment to detect issues that can be seen by just browsing through the tables. Issues of different nature were identified and cartegorised by either Quality issues(from the content/attributes of the data) or Tideness issues(arised from the issues of the structure).

1.0.5 Cleaning

Cleaning is the most interesting stage of the data wrangling process.Here all the issues identified while assesing the data are adressed using different python functions, methods and libraries. I followed the `Define`,`Code` and `Test` approach for a smooth and readable wrangling.For good and

quality analysis, this stage needs to be mastered, and the interesting thing about data cleaning is that we can always iterate and go back to the previous stages until data is in our desired format.

1.0.6   Conclusion

With the help of the Python tools, the desired format of our data was reached. identified issues were cleaned, but there is still a room for improvement on our wrangling efforts as some issues were visible but not adressed due to the requirement of the project rubric.