wrangle_report

December 7, 2018

This project is more challenging and was very excited to start this project because this impressed me alot of way.when i started project i got stuck alot of place where showing so much of error while accessing data from twitter.i enjoyed doing the precess of assessing and cleaning the data mannually by looking the dataset as well as programatically.Identifying issue was very challenging and it confused about to identify the significance of each columns and fixing the issue.I didnt went through the complete cleaning process and some issue i got by just checking the columns randomnly which have some suspecious value.

Gathering data ,Udacity provide twitter archives for the tweets of @WeRateDogs in csv format twitter_archive_enhanced.csv Download this file manually This file (image_predictions.tsv) is hosted on Udacity's servers and should be downloaded programmatically using the Requests library and the following URL: https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv querying the Twitter API using a Python library to download each tweet's JSON data which includes the favorites and retweets counts.