

A minha maior dificuldade encontrada no processo de data wrangling foi entender o funcionamento do tweepy, levei um tempo para entender o funcionamento do mecanismo de timer da biblioteca e que ele deveria ficar dentro do laço e que a propria biblioteca ia lidar com o timer.

Outro ponto a se destacar no processo de data wrangling é a impossibilidade de limpar alguns dados, por exemplo, tweets que não possuem nomes, pois, o usuario não colocou, ou então, alguns casos onde a nota do cachorro não está no padrao, pois existe mais de um cachorro na foto, por exemplo, tem um caso onde a nota do cachorro foi 4/20, a foto é o músico Snoob Dog, uma clara brincadeira com o musico, devido ao fato de que 4:20 é considerada a hora mundial do fumo da maconha (<https://www.significados.com.br/4-20/>), esses dados só foram possíveis de serem encontrados devido a Analise através da programação, pois, visualmente era praticamente improvavel de serem encontrados, e para essa analise através da programação a biblioteca pandas foi fundamental.

Voltando as dificuldades que eu encontrei com o Tweepy, utilizei outra referencia para buscar os dados, inclusive coloquei o link que eu me ajudou na resolução no problema no Jupyter Notebook.

Sobre o processo de data wrangling, seguir o passo a passo ensinado, Gather → Assess → Clean, facilita muito em criar o mapa mental e entender os problemas dos dados, ajuda também em focar em cada etapa do processo.