

Como foi meu segundo projeto feito em python, foi bastante desafiador. A princípio quando li os detalhes do projeto, fiquei bastante perdido, achando até que não daria conta do projeto, após ler todo o roteiro do trabalho vi que não é tão difícil quanto achei.

Eu iniciei o projeto escrevendo um código para baixar da base da Udacity o arquivo "image-predictions". Em seguida, foi a vez de baixar os dados de tweets da conta *WeRateDogs*, como foi minha primeira vez utilizando a biblioteca "*tweepy*" fiquei bastante confuso no início, gastei bastante tempo em alguns sites pesquisando sobre o funcionamento da biblioteca e suas funções. Após entender o necessário sobre o funcionamento da biblioteca, chegou a hora de colocar em prática e desenvolver o código para baixar os dados dos tweets, como foi meu primeiro código que levasse tempo para executar, fiquei bastante ansioso.

Após a fase de coleta de dados, chegou a parte de avaliar os dados. Avaliação de dados que em minha opinião é a fase mais fácil no processo de *data wrangling*, porém necessária igual as outras. Durante a fase de avaliar os dataframe, utilizei bastante funções do pandas para colher algumas informações sobre os dataframes, informações importantes como por exemplo média, valor mínimo, valor máximo, algumas porcentagens e etc. Ainda avaliando os dados descobri que tinha bastante problemas de qualidade no arquivo *twitter-archive-enhanced* que foi fornecido pela Udacity, nomes inválidos, twittes sem imagens, muitas colunas que não tem grande utilidade no dataframe.

Depois de fazer toda a avaliação que achei necessária, chegou a vez de limpar os dados. O processo de limpeza de dados, não foi difícil como imaginei que seria, não houve a necessidade de um código complexo durante a limpeza, a maioria da limpeza realizada foi tipo dos dados, excluindo dados desnecessários ou que afetariam na análise e visualização.

Em resumo este projeto foi bastante desafiador, o processo mais complicado para mim foi o de coleta de dados, que foi no momento de coletar dados do twitter utilizando a biblioteca *tweepy*.