

Projeto: Preparar e analisar dados Data Wrangling - WeRateDogs

Este projeto foi um grande desafio para mim do começo ao fim. Foi muito bom e importante aprender sobre o processo de Data Wrangling e ter uma noção básica sobre como consultar dados da API do Twitter por meio de uma conta autenticada de desenvolvedor e da biblioteca Tweepy.

A parte mais difícil do projeto para mim, foi como consultar os dados da API do Twitter. As minhas primeiras tentativas resultaram em erro e tive que obter ajuda de um mentor da Udacity para prosseguir. Depois de muitas tentativas, descobri que precisava atualizar minhas chaves de autenticação da conta do Twitter para que o processo de consulta funcionasse efetivamente.

Depois de reunir todos os dados, copieei os arquivos para os processos de avaliação e limpeza de dados. Procurei selecionar alguns maiores problemas de qualidade e arrumação nos três dataframes e em seguida comecei a corrigi-los. Em princípio, transformei a coluna de data e hora para o formato de dados adequados, assim como transformar as colunas de classificação em formato Float e a coluna Tweet_Id para String. Em seguida, corriji nomes de cachorro errados e removi colunas desnecessárias. Utilizando as funções `str.replace()` e `str.title()` da biblioteca pandas, eu removi o sublinhado entre as palavras e capitalizei a letra inicial em cada palavra para criar uma tabela mais coesa. A etapa final no processo de limpeza de dados foi juntar todos os três conjuntos de dados em um documento final contendo todas as informações relevantes. Para esta tarefa eu usei a biblioteca pandas usando a função `pd.merge()`.

Esse projeto foi muito importante para adquirir novas habilidades de análise de dados e poder ampliar a minha visão sobre como a organização e limpeza de dados contribui para ter melhores intuições a respeito de questões relevantes que não conseguiríamos responder com dados bagunçados e de qualidade ruim.