

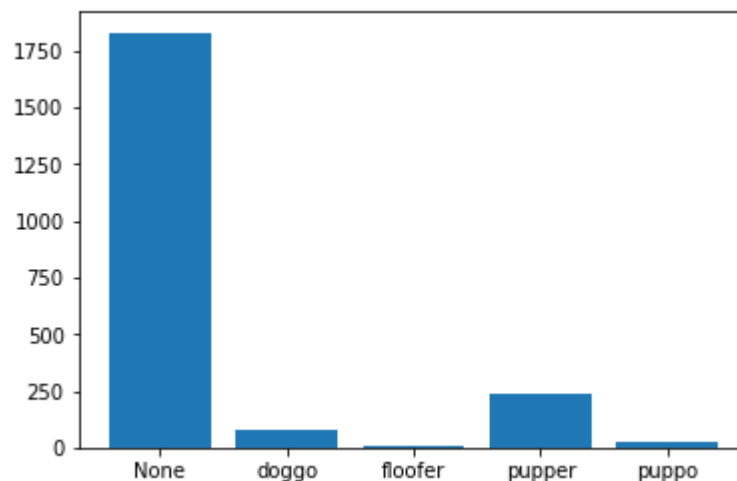
Wrangling Report (Insights e Visualizações):

1. Primeiro Insight:

Após terminar de limpar as informações para a criação do Dataset final (df_final) comecei a perceber algumas diferenças e insights dentro da base de dados criada.

Uma das informações finais que me preocupou foi a diferença, ou falta de informação, gigantesca entre as informações que estão vazias (“None”) com relação a coluna de “tipo_cachorro”, mas logo percebi que se devia que no próprio tweet não era relatado o tipo do cachorro de acordo com as variáveis presentes na base de dados.

Conseguimos reparar esta diferença com a visualização abaixo:

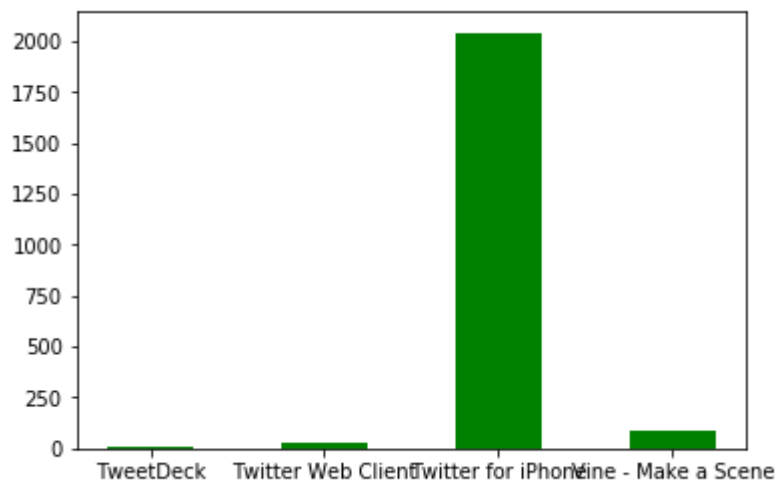


2. Segundo Insight:

Ainda focando na comparação de valores, ao realizar uma busca mais profunda na coluna de “Source” onde temos a fonte de onde o tweet foi publicado, consegui notar que a maioria dos tweets foram enviados pela plataforma “Iphone”.

Ainda estou aprendendo sobre visualizações em python mas já considero um avanço conseguir realizar esta análise com a base de dados que eu já possuo.

Segue visualização de comparação:

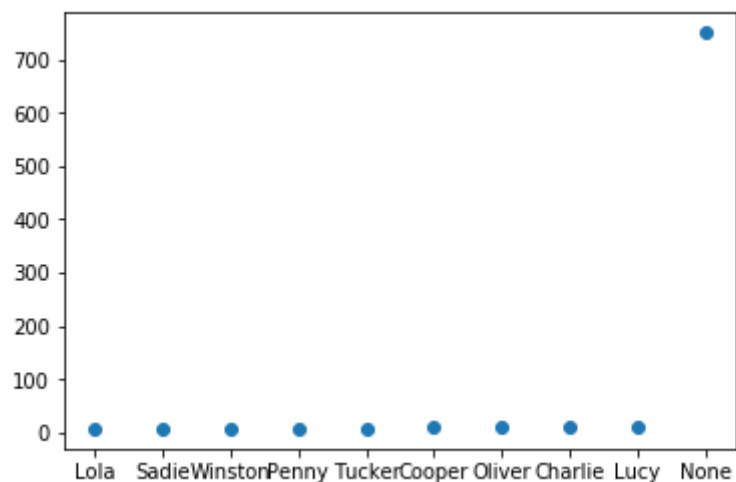


3. Terceiro Insight:

Realizando uma análise mais direta na base de dados, utilizando o Excel após a exportação, comecei a focar na coluna de “name” onde é a variável que contém o nome dos cachorros, e novamente tive a diferença de nomes que não estão preenchidos, nos deparamos com o mesmo problema.

A resposta para este problema é basicamente a do Insight de cima, onde o nome do cachorro muitas vezes não é descrito no tweet, gerando assim o valor “None” na coluna de nome.

Segue visualização:



Obs: Utilizei apenas os cachorros mais mencionados (Top 10)