Wrangling Report:

1.0 – **Gathering:**

Comecei mapeando todas as fontes de dados e como seria o processo para importação para o ambiente de desenvolvimento do Jupyter Notebook.

Tive alguns problemas com a importação das informações extraídas direto da API do twitter, mas após uma longe pesquisa e voltas em capítulos passados do módulo de Data-Wrangling, consegui desenvolver o código necessário para a importação.

Com todas as bases de dados importadas e mapeadas com uma nomenclatura única de cada base, passei para o próximo passo que é a exploração ou Assess.

2.0 **Assess:**

Com todas as bases importadas, comecei a explorar e acessar os dados disponíveis, utilizando todas as funções ali presentes. Foi desta exploração que surgiram os problemas a serem corrigidos.

Percebi vários valores nulos e nomes incorretos, o que me deixou intrigado para ajustar e passar para a próxima fase do Wrangling, a de Limpar os dados.

3.0 **Clean:**

Após a fase de exploração vem à fase onde limpamos e ajustamos todos os dados e problemas apontados anteriormente, quando acessamos os nossos dados.

Pude notar que uma grande parte das colunas estava com o tipo de dado do jeito incorreto para trabalharmos, como por exemplo, as colunas de “timestamp” e “tweet_id” que estavam incorretas e assim sendo necessária a alteração por meio da função astype().

Após outra análise, notei também que as últimas colunas do estágio do cachorro (“puppo”, “floofer”, etc) estavam como variáveis, e que na verdade deveriam ser valores e precisei assim tratar este problema com a função melt e pivotar as quatro colunas para linhas.

Existiam várias colunas que não eram necessárias para a análise e acabei dropando-as, deixando apenas o que achei relevante para a entrega do projeto.

Utilizei de técnicas de Regex para extrair informações presentes em strings e criar uma nova coluna de “Source”, deixando apenas a ultima informação extraída.