Wrangle Report

本次项目通过可视化和编程的方式对搜集的三个数据进行相应的处理。具体步骤包括：

1. 数据搜集和读取：

- 使用requests下载所需的图片预测数据image-predictions.tsv，直接下载使用tweet_json.txt和twitter-archive-enhanced.csv。
- 通过excel分别观察数据结构，找到对应的方法转为DataFrame格式方便在Jupyter清洗。

2. 数据清洗

数据的清洗通过【定义-清洗-测试的步骤】进行。

1. 通过观察，直观看到每个数据集存在的相应问题，并找到哪些数据是可用数据；
2. 再通过编程的方法检查每个数据集的基本问题；
3. 合并image-predictions.tsv和tweet_json.txt对应的DataFrame，因为只需要含有图片的原始评级 (不包括转发) image 数据集进行 merge 时选择 inner 方式；
4. 着重处理twitter-archive-enhanced.csv，分别处理狗狗类型，名字和评分的中脏乱数据，并使用正则的方法从text中提取相应的合理数据，最后与上面的数据进行合并和保存。

3. 可视化

针对可视化过程中发现一些数据问题，进行进一步清洗，以达到相应的可视化要求。