Act Report

分析步骤：

1. 提取合并表中有分析意义的列：
'tweet_id','timestamp','text','rating_numerator','rating_denominator','name',
'stage','source_name', 'favorite_count', 'retweet_count'加入新的 dataframe：dfa
2. 生成新的指标，
(1) 统计评分 rating_ratio=rating_numerator/rating_denominator
(2) text 的长度
3. 把 dfa 存入 dfa.csv。

分析过程：

为了使趋势更明显，除去 favorite_count 大于 50000 的值，然后使用 pd.plotting.scatter_matrix(dfaf,figsize=(15,15))和 dfa.corr()得到所有量之间的关系。

一、转发/喜爱数随时间的变化趋势

以时间为横轴，favorite_count 为纵轴作图。可见总体的转发/喜爱数随时间上升。推测是由粉丝量增加造成的。

二、rating_ratio 与喜欢/转发数的关系

相关系数为 0.3，0.4，中度正相关。

三、不同 stage 的喜爱数的中位数

在 tableau 中分组数据。可见 2016 季 3 开始稳定出现 4 种类型。为了排除关注人数变化带来的偏差，应仅使用 2016 年 7 月开始的数据。

	timestamp 季度							
stage	2015 季4	2016 季1	2016 季2	2016 季3	2016 季4	2017 季1	2017 季2	2017 季3
doggo			11	20	14	24	4	2
floofer		1	2	2	1			1
pupper	73	82	27	17	10	11	7	4
puppo			3	4	6	5	9	2

另外，floofer 类型数据量过小没有代表性。

以 stage 分类对转发的中位数作图，可见 puppo 是最受欢迎的成长阶段。

四、推文长度与受欢迎程度间的关系

推文长度与受欢迎程度有轻微正相关，可能是由于有趣的图片会受到更多描写。