

# 项目报告

数据源于twitter上 WeRateDogs的推送消息，twitter用户 WeRateDogs 以诙谐幽默的方式对人们的宠物狗评级。数据采集所得。

## 报告思路：

### 1、现状：

- (1) 基础统计描述。
- (2) 通过时间序列展示推文的retweet\_count和favorite\_count。

### 2、目标（即要解决问题）：

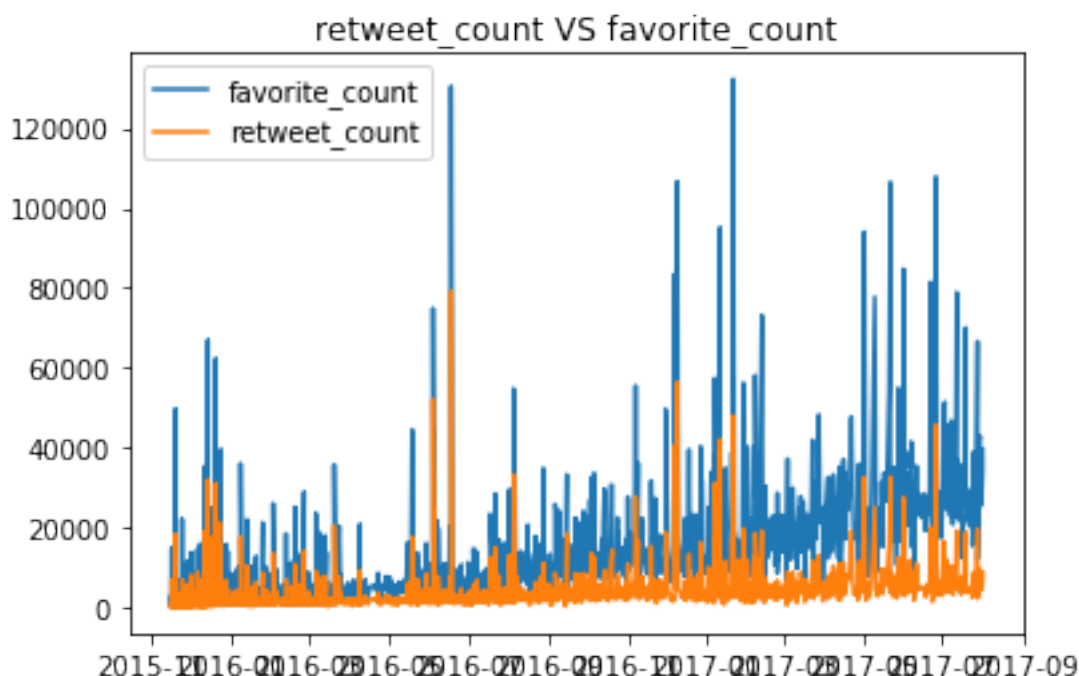
- (1) 想看图片识别准确率0.9以上，twitter用户喜欢的top10的狗的品种是哪些？
- (2) 所发twitter中哪个成长阶段到dog最多

## 1.1 基础统计描述

	tweet_id	rating_numerator	rating_denominator	img_num	p_conf	favorite_count	retweet_count
count	1.504000e+03	1504.000000	1504.0	1504.000000	1504.000000	1504.000000	1504.000000
mean	7.444885e+17	11.203457	10.0	1.231383	0.560212	9927.767952	3009.592420
std	6.858126e+16	1.240732	0.0	0.593533	0.297679	13222.557189	5056.234907
min	6.660334e+17	9.000000	10.0	1.000000	0.000010	81.000000	15.000000
25%	6.804916e+17	10.000000	10.0	1.000000	0.316669	2423.750000	703.000000
50%	7.248777e+17	11.000000	10.0	1.000000	0.568361	4953.000000	1554.000000
75%	8.005729e+17	12.000000	10.0	1.000000	0.828943	12579.000000	3488.500000
max	8.921774e+17	14.000000	10.0	4.000000	0.999956	132318.000000	79116.000000

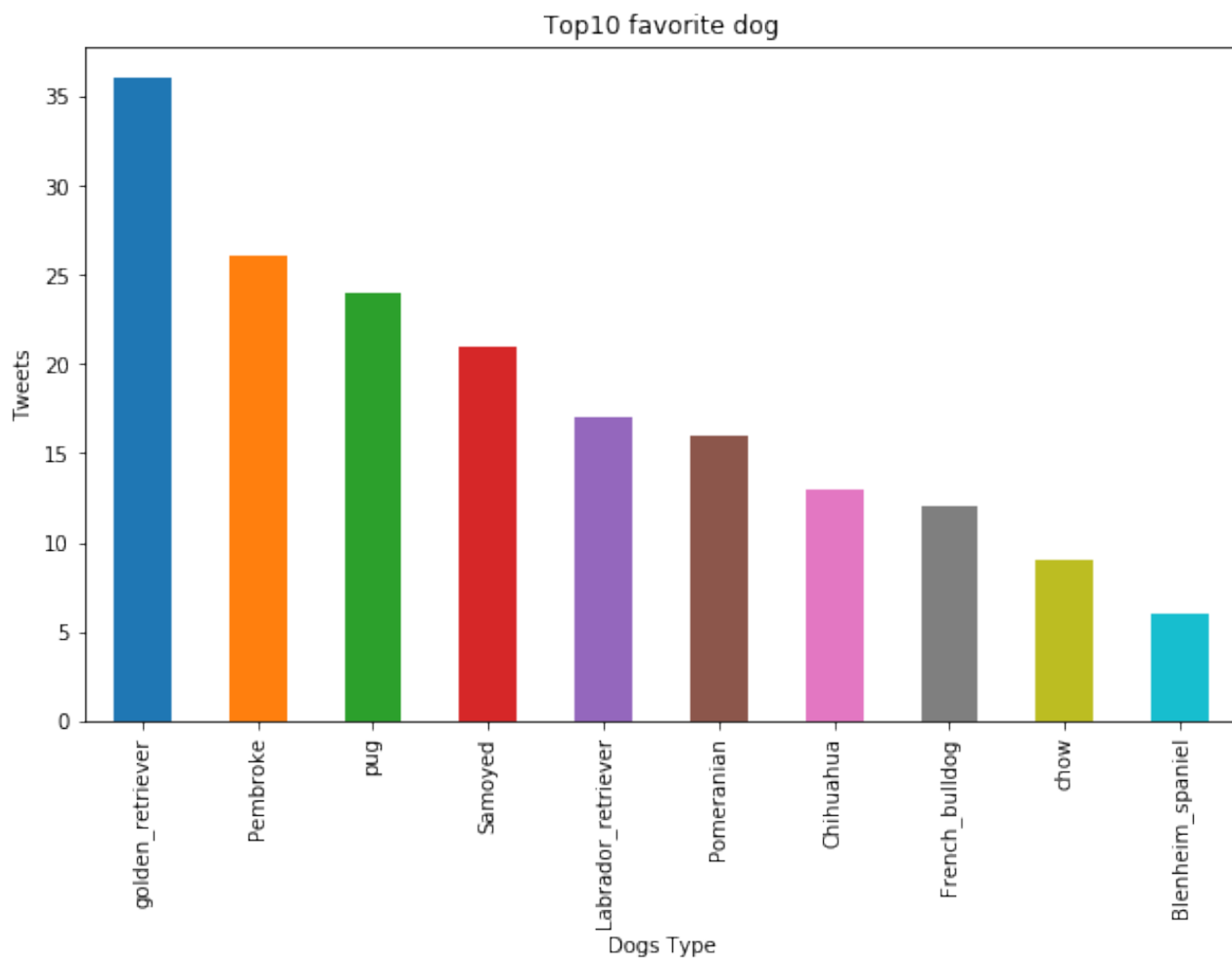
对清理后dataframe的基础描述性统计，可以对分析有一个基础性了解。比如，可知道有1504条数据，判定狗准确率均值在0.56等。

## 1.2 通过时间序列展示推文的retweet\_count和favorite\_count。



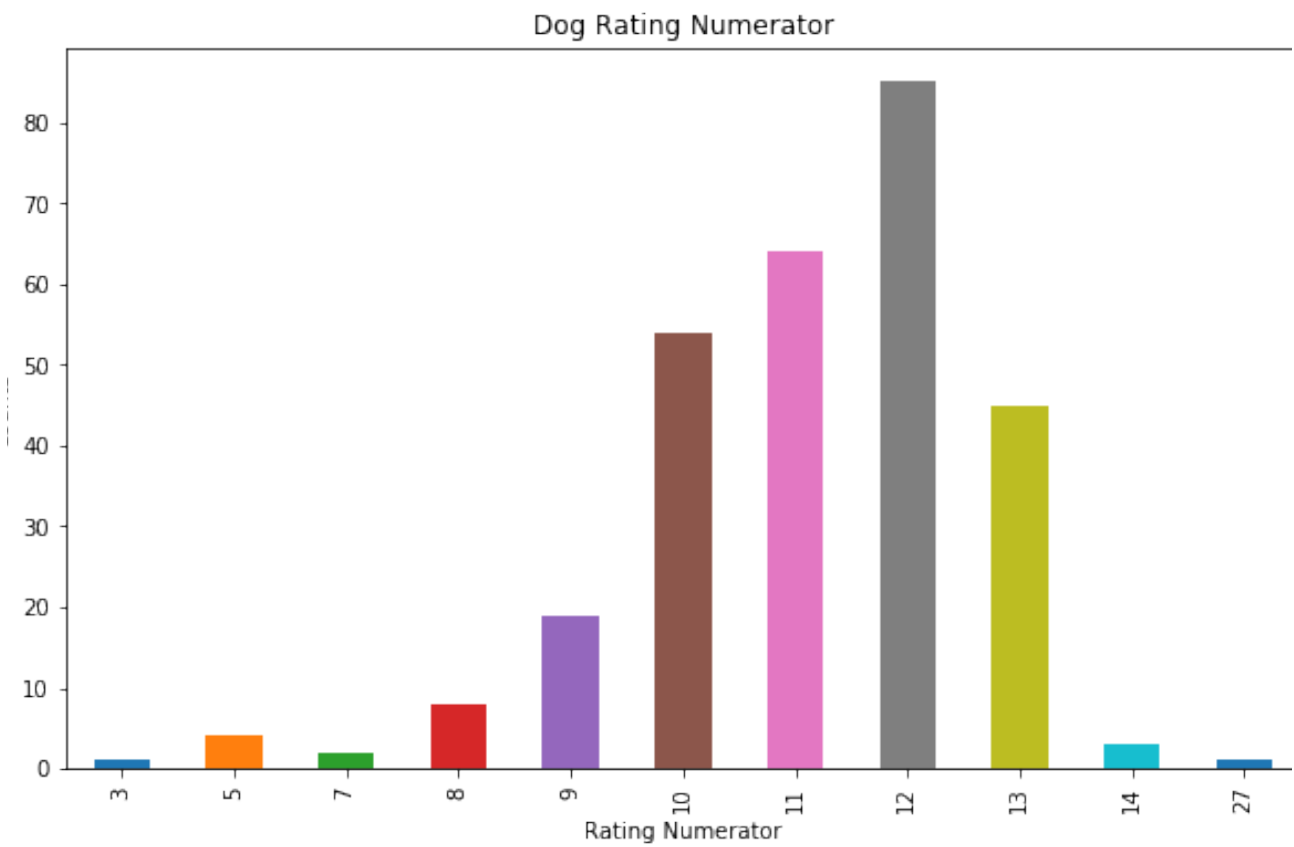
通过分析可以看出retweet\_count和favorite\_count均有增长趋势，favorite\_count比retweet\_count量高。

## 2.1 想看图片识别准确率0.9以上，twitter用户喜欢的top10的狗的品种是哪些



分析可得top10分别为：golden\_retriever、Pembroke、pug、Samoyed、Labrador\_retriever、Pomeranian、French\_bulldog、Chihuahua、chow、Blenheim\_spaniel。

## 2.2 所发twitter中哪个评分的dog最多



根据数据中反应情结果：rating\_numerator/rating\_denominator为12/10的狗狗数量最多。