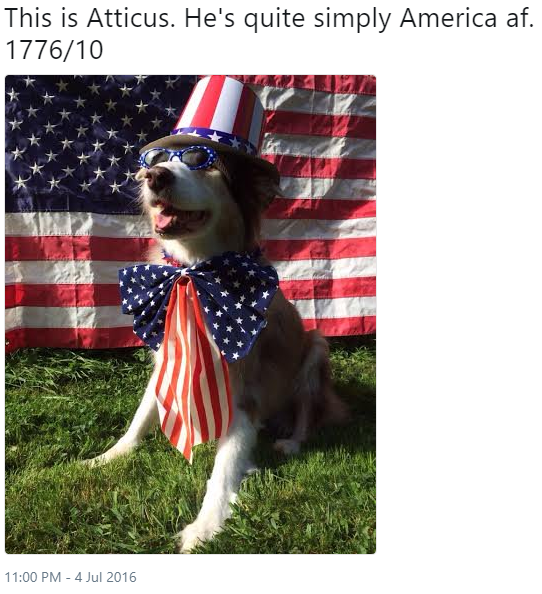
**WeRateDogs推特数据分析报告**

WeRateDogs 是一个推特主，他以诙谐幽默的方式对人们的宠物狗评分。这些评分通常以 10 作为分母。但是分子则一般大于 10：11/10、12/10、13/10 等等。为什么会有这样的评分？因为 "They're good dogs Brent." 也有一部分的原因是一张图像里有多个狗：165/170。

有时候也有图1中的极端情况。

图1. WeRateDogs的推特图片



通过分析WeRateDogs（@dog\_rates）的1994条推特数据，我们得到一些有趣的结论。

**1评分**

根据推文提取到1577个有效评分，标准评分（评分分子\*10/评分分母）的均值为10.68，大部分评分都在满分10分以上，最多次数的评分为12分，其次为11分。

那分数特别低的是些什么狗呢？Emm，多半是比较可爱的其他动物啦（见图3-6）。

图2. 标准评分直方图

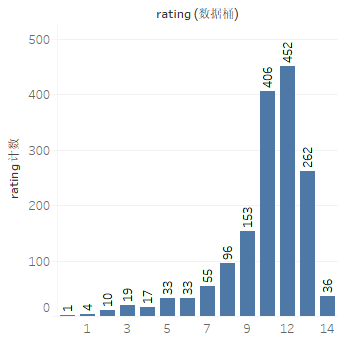
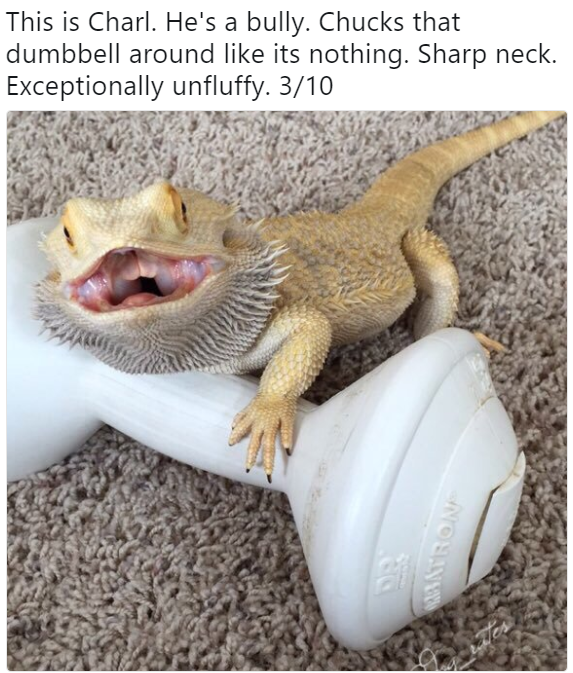
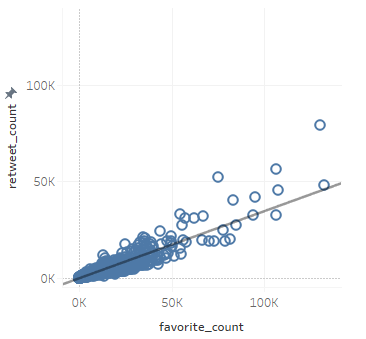


图3-6. WeRateDogs的推特图片

**2喜爱数和转发数**

图7. 喜爱数和转发数的关系图



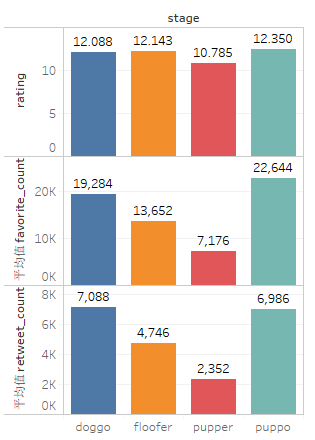
平均喜爱数为8923，平均转发数为2770，两者呈线性相关。

**3狗狗地位对评分、喜爱数和转发数的影响**

Puppo、doggo和floofer的评分都在12分以上，而且差距并不大。

puppo平均收到22644次喜爱，是平均收到最多次喜爱的狗狗地位。但平均被转发次数最多的狗狗地位却是doggo，dogggo平均被7088次转发。puppo平均被转发次数紧随其后，为6986次。

图8.狗狗地位对评分、喜爱数、转发数的影响



**4神经网络对狗狗种类的预测**

通过一个可以对狗狗种类进行分类的神经网络，对WeRateDogs推特档案中的所有图像进行预测，发现判断为狗的平均预测可信度为64.44%，判断为非狗的平均预测可信度为53.56%，两者差距并不算大，看来神经网络还要再多多加强机器学习呀。

而Golden retriever和labrador retriever分别是被神经网络识别出次数最多和第二多的狗狗种类。可能retriever是最经常被养的宠物狗，或者是最容易被神经网络识别出来的狗狗品种。

图9.神经网对狗狗种类预测的可信度

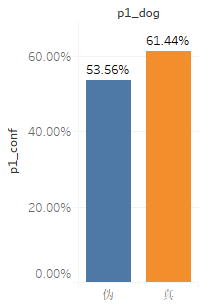


图10.被识别出最多次数的狗

