WeRateDogs分析结果

1. 不同的dog\_status水平的狗狗的受欢迎水平

针对不同的dog\_status水平和喜爱数量进行交叉分析，试图分析出不同狗狗的不同受欢迎程度。

|  |  |  |  |
| --- | --- | --- | --- |
| **dog\_status** | **favorite\_counts** | **counts** | **mean\_favorite** |
| floofer | 68550 | 3 | 22850.00 |
| puppo | 634995 | 29 | 21896.38 |
| doggo | 1116454 | 68 | 16418.44 |
| none | 14316477 | 1734 | 8256.33 |
| pupper | 1650658 | 238 | 6935.54 |
| total | 17787134 | 2072 | 8584.52 |

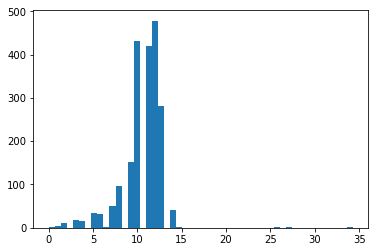
从上面的分析结果可以看出：

1. 一共2072ID，有高达1734个未进行“地位”评价，占比发到84%；
2. 在已有评价的狗狗中，从平均喜爱程度来看，floofer地位的狗狗更受人喜爱，但是评价数量偏低；
3. 在已有评价的狗狗中，从平均喜爱程度来看，pupper地位的狗狗最不受人喜爱，但是评价数量最高；
4. 从总体趋势来看，平均喜爱程度和评价数量成反比（除无dog\_status数据），可以假设本次的平均狗狗喜爱程度受极端值影响。从经验来看，喜欢狗狗的人更容易逛论坛也更容易评价。因此受此行为习惯的影响，评价数量可能对狗狗的平均喜爱程度产生影响
5. 本次分析不能直接得出结论说floofer从平均水平看更受人喜欢，而pupper更不受人喜欢。需要搜集更多的数据，评论等待足够长的时间，对以上假设进行验证，并排除一些其他因素的干扰。
6. 按照评分，对狗狗进行分组分析

首先对评级进行标准化处理，让不同的评级可以放在同一水平线上进行比较。

评级和喜爱程度进行交叉分析，确定评级是否与喜爱转发数一致。以及评估评级的准确性。

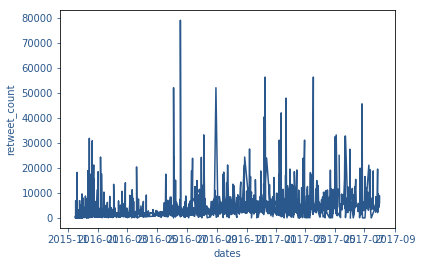
|  |  |  |  |
| --- | --- | --- | --- |
| **rating\_cut** | **favorite\_counts** | **counts** | **mean\_favorite** |
| 10级以下 | 2936256 | 847 | 3466.654 |
| 10到20级 | 14789920 | 1220 | 12122.89 |
| 20级以上 | 60958 | 5 | 12191.6 |
| total | 17787134 | 2072 | 8584.52 |



1. 从数据可以看出，评级和喜爱评分呈现一致趋势
2. 10级以下和10到20级的数据均几百甚至到千级的评级，而20级以上仅仅有5名评级，该级的评级准确性需要评估。
3. 而从评级的分布图可以看出，该评级呈现正偏态分布，该评级可能受到数据量的影响，同时也可能受到极端值的影响
4. 三种图片预测算法的准确率对比

执行结果可知，p1=11.29%,p2=p3=0%。

1. 从上面数据可以看出，p1的准确率是11.29%，p2和p3的准确率都是0%。也就是说p1的预测算法准确率要明显高于p2和p3.
2. P1的准确率也仅仅只有11.29%，准确率太低，需要提高准确率。但是可以优先优化p1的算法。
3. 不同时间转发量retweet\_count的趋势图



1. 从上图可以看出，转发量对着时间的推移有巨大的波动，且不断的有高峰值的出现，我们可以确定一下，这段时间是否有活动出现。
2. 先中间位置有一段转发量特别低的时间段，我们可以重点研究转发量低的原因，以及采取了什么措施提高的转发量。