

本数据集收集的是一个推特博主叫WeRateDog的每条推特。（超级有意思的博主）这个博主就是对狗狗打分 传送门->https://twitter.com/dog_rates

 Image Name

** 我在做的事情就是 **

- 1.数据进行收集
- 2.评估数据质量问题和清洁度（结构）问题，
- 3.清洗数据集 ☺

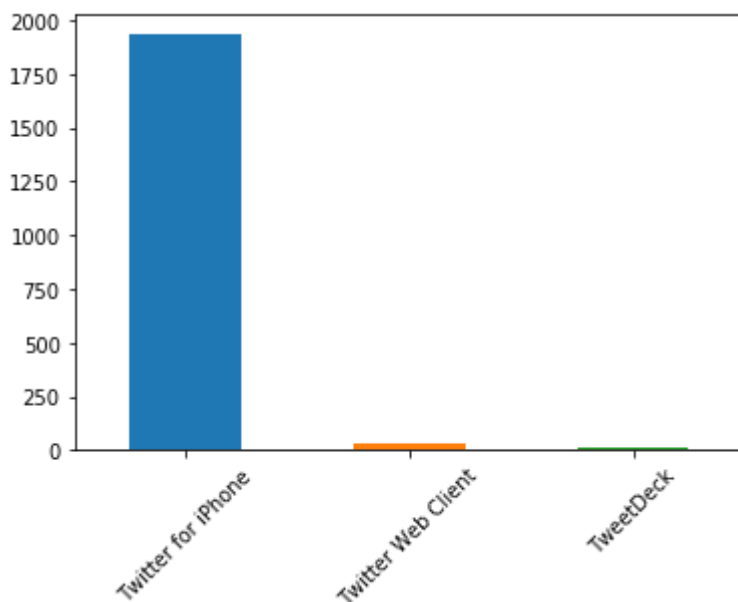
这个推特的作者时不时的发一些狗狗的照片，然后会给狗狗打分，这个分数一般是n/10,有时候n会大于10，就是狗狗的颜值爆表了。我这里的主要展示了：

- 高分的狗狗是否得到大家的认同
- 点赞数量和转发数量否成正比
- 作者是否对某些狗狗有偏见

分析和可视化

```
# 分析或可视化代码
%matplotlib inline
twitter_data = pd.read_csv("twitter_data_clean.csv")
twitter_data['source'].value_counts().plot.bar(rot = 45)
```

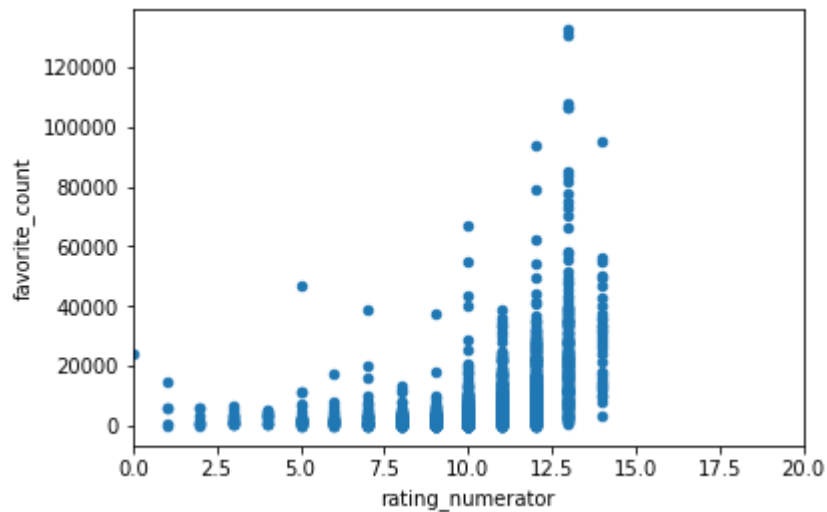
<matplotlib.axes._subplots.AxesSubplot at 0x25c8ae5cb00>



从该图中可以看出博主平时使用的是手机发送推特。我推测他可能不是全职去做这件事情的。

```
# 分析或可视化代码
twitter_data.plot.scatter(x="rating_numerator", y="favorite_count", xlim=(0,20))
```

<matplotlib.axes._subplots.AxesSubplot at 0x25c8ae941d0>

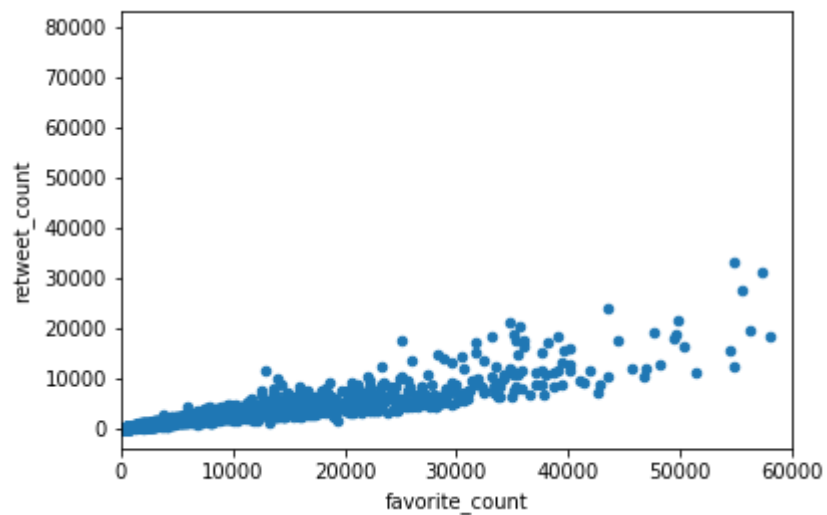


分数和点赞数的关系图

从图中可以看出有时候在高分的时候，会获得高点赞。也就是说大家可能十分认同这个狗狗得到高分。

```
twitter_data[['favorite_count', 'retweet_count']].plot.scatter(x="favorite_count", y =  
"retweet_count", xlim=(0, 60000))
```

<matplotlib.axes._subplots.AxesSubplot at 0x25c8af08ba8>

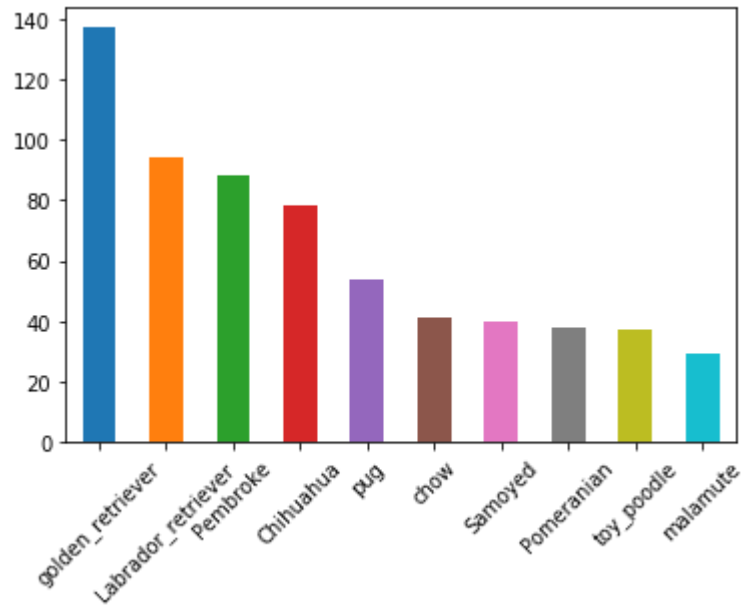


转发数和点赞数的关系图

点赞数和转发数成正相关性。

```
t = twitter_data['p1'].value_counts()[0:10]  
t.plot.bar(rot=45)
```

<matplotlib.axes._subplots.AxesSubplot at 0x25c8af7b320>



狗狗品种出现的次数统计图

可以看出作者转发更多的是“金毛犬”和“拉布拉多”。所以说作者可能存在“偏爱”的情况，没有“雨露均沾”。