

分析和可视化

1，我们先来看看所有狗狗评分的最大、最小、中位数，分别是：0 分，20 分，11 分。

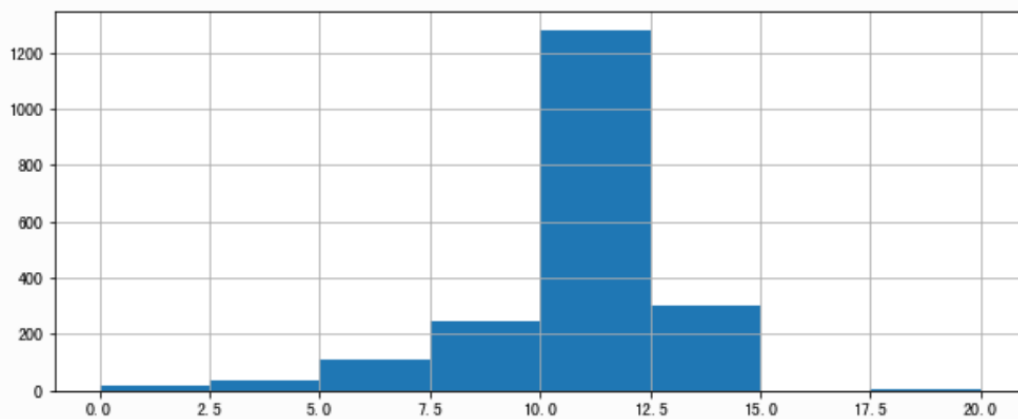
```
df_clean["评分"].describe()
```

```
count    1993.00
mean      10.58
std       2.18
min       0.00
25%      10.00
50%      11.00
75%      12.00
max       20.00
```

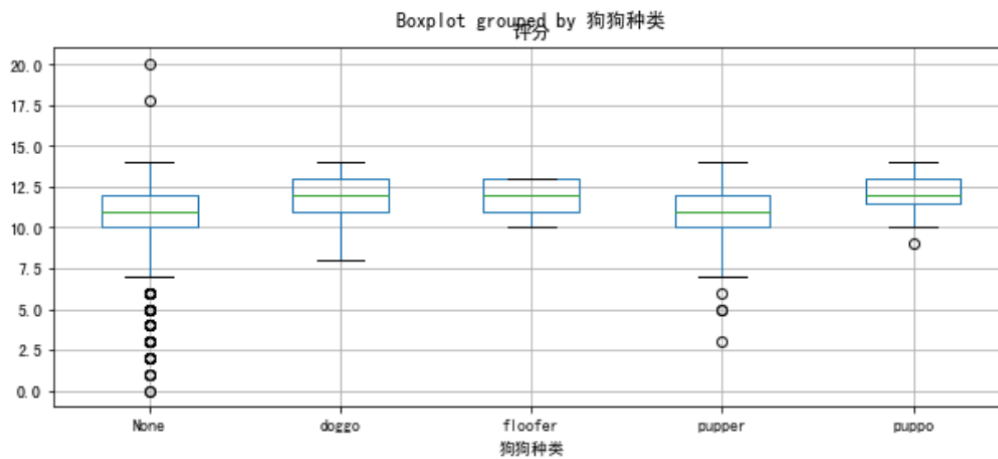
```
Name: 评分, dtype: float64
```

2，来看看分布的直方图，看下分布情况怎么样：

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb876ec3828>
```



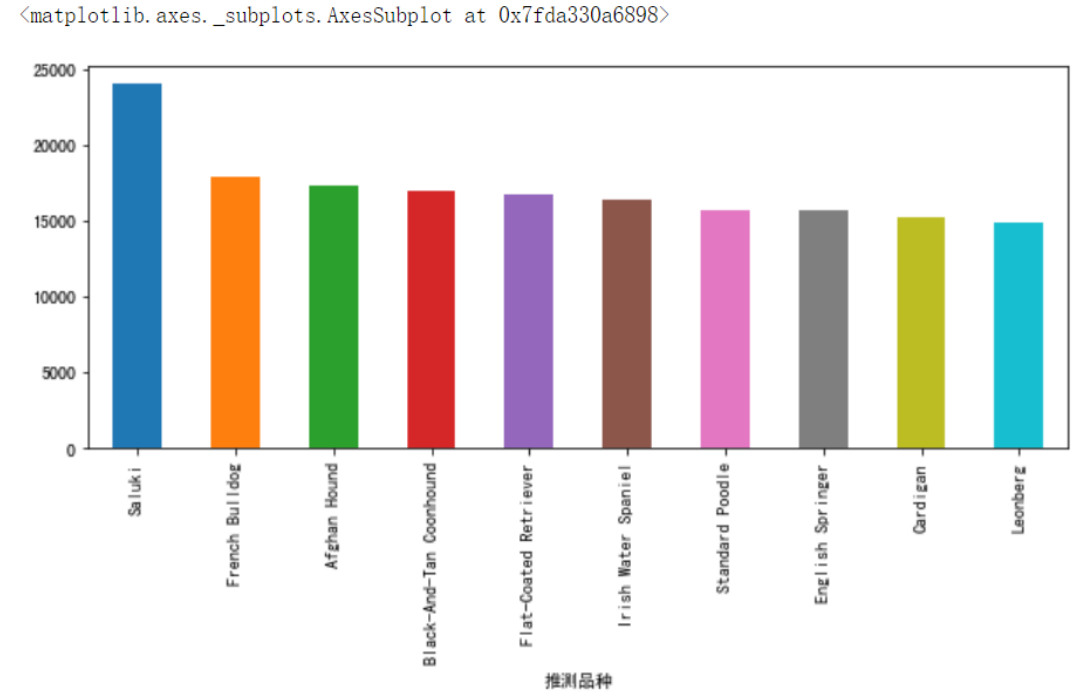
狗狗评分的平均值为 10.58 分，范围从 0 到 20 分，呈现中部集中态势。由图可见，大部分集中在 10-12.5 分。



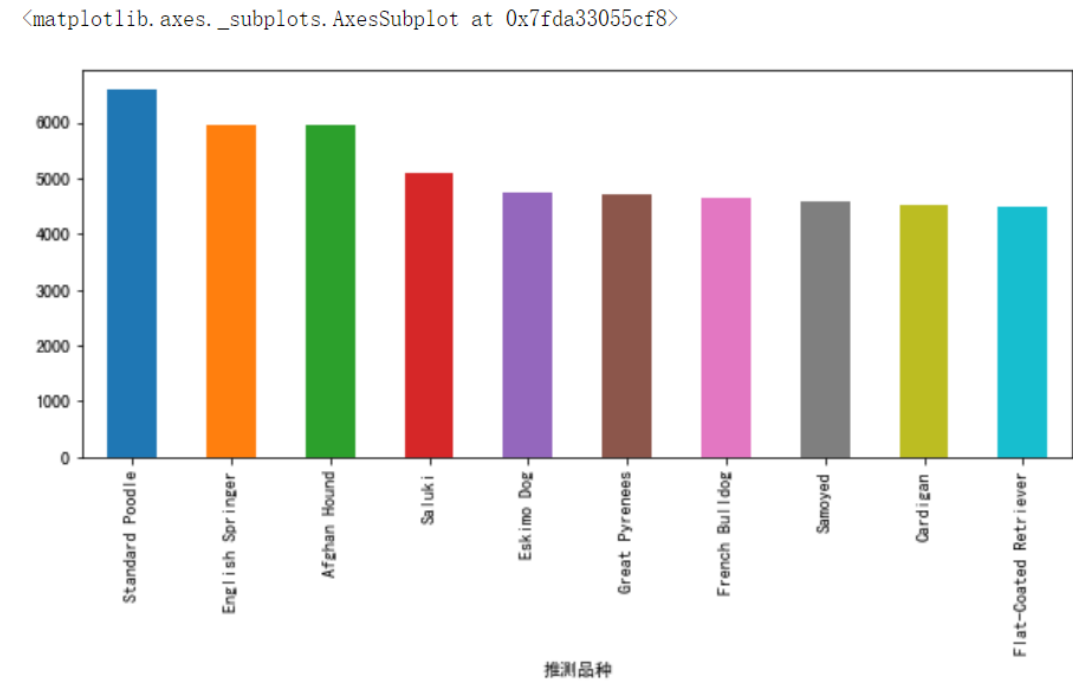
以上是按照狗狗分类的分数箱线图，似乎各品种狗狗的评分没有明显差异。

3，下面看看狗狗按品种的各项表现：

点赞数前十的狗狗品种

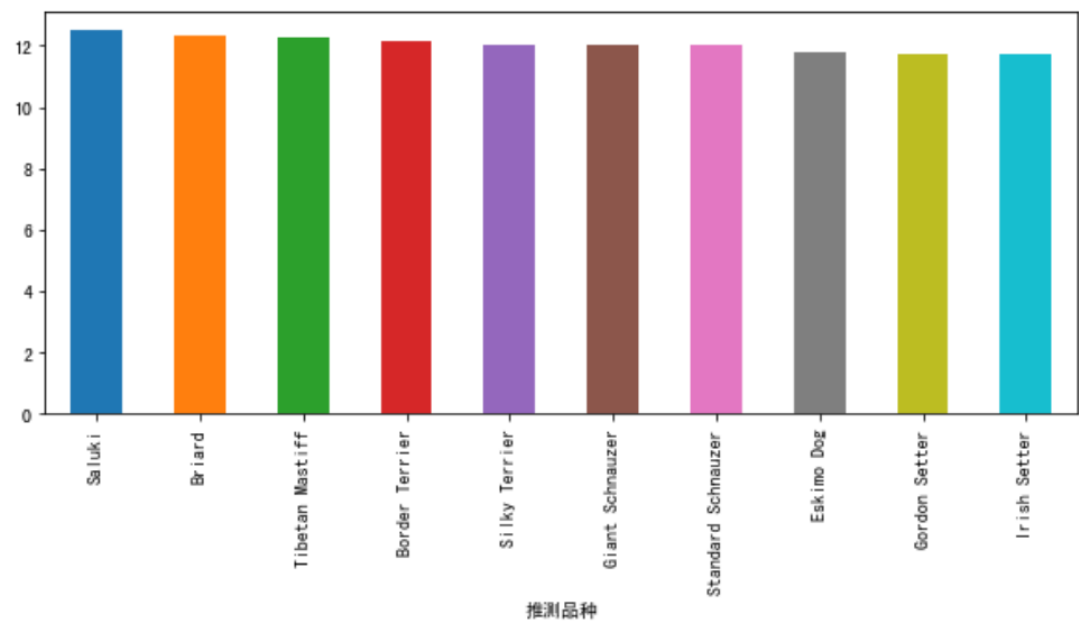


转发数最高的十大狗狗品种



平均评分最高的十大狗狗品种

<matplotlib.axes._subplots.AxesSubplot at 0x7fb874ceabe0>



点赞前十推测品种

推测品种

Saluki	23996.25
French Bulldog	17854.77
Afghan Hound	17250.67
Black-And-Tan Coonhound	16952.50
Flat-Coated Retriever	16731.62
Irish Water Spaniel	16335.33
Standard Poodle	15732.14
English Springer	15631.22
Cardigan	15251.71
Leonberg	14870.00

Name: 点赞数, dtype: float64

转发前十推测品种

推测品种

Standard Poodle	6602.14
English Springer	5948.67
Afghan Hound	5946.00
Saluki	5103.25
Eskimo Dog	4742.44
Great Pyrenees	4704.00
French Bulldog	4656.38
Samoyed	4597.43
Cardigan	4513.65
Flat-Coated Retriever	4500.12

Name: 转发数, dtype: float64

平均评分最高的十大推测品种

推测品种

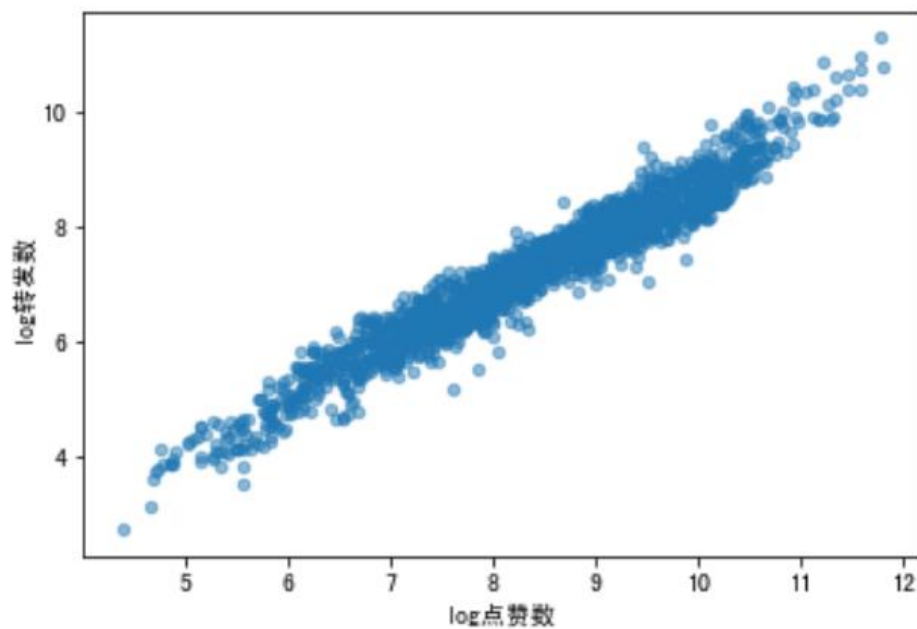
Saluki	12.50
Briard	12.33
Tibetan Mastiff	12.25
Border Terrier	12.14
Silky Terrier	12.00
Giant Schnauzer	12.00
Standard Schnauzer	12.00
Eskimo Dog	11.78
Gordon Setter	11.75
Irish Setter	11.75

Name: 评分, dtype: float64

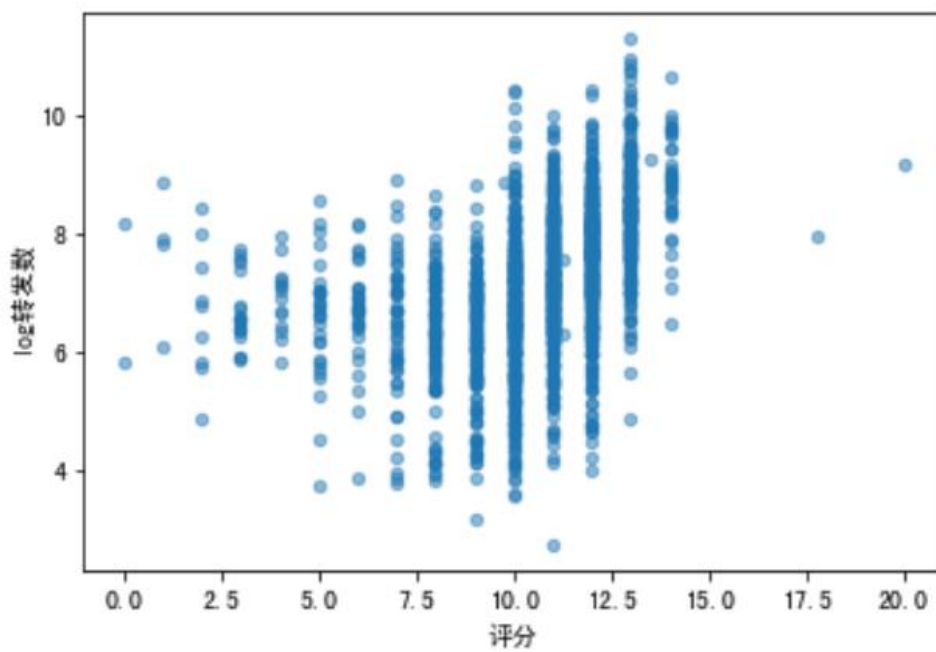
大部分点赞数高的狗狗品种转发数越高。Saluki 三项排名均靠前，算得上是最受欢迎的狗狗品种了。

4, 最后, 看下评分、点赞数、转发数之间有没有相关性。

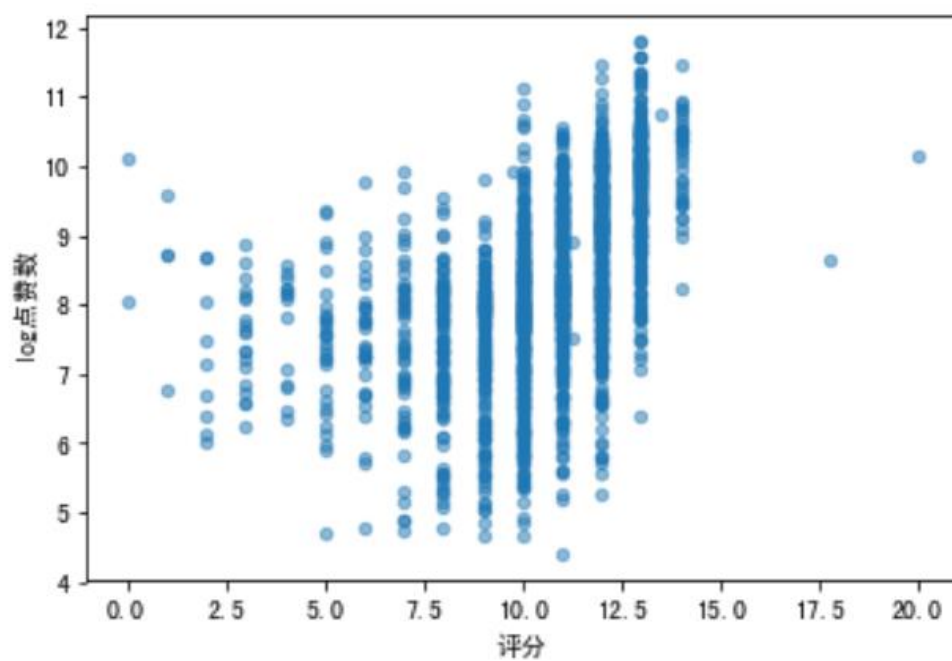
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb874c32080>
```



```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb874bc9eb8>
```



<matplotlib.axes._subplots.AxesSubplot at 0x7fb874af0d68>



评分较高的狗狗，转发数与点赞数也较高。转发数、点赞数有正向线性的关系。