数据可视化

- 推文主要来源
- 狗狗类型占比
- 狗狗类型 与 点赞、转发、评分 均值的关系

聚合、整理数据

按狗类型分组的 评分、转发、点赞均值数据

使用 groupby 根据 dog_type 聚合 (dog_type 就是 image-predictions.tsv 的 p1 值)

```
dog_type_gb = twitter_archive_clean.groupby(['dog_type'])
dog_type_counts = dog_type_gb.size()[dog_type_gb.size()>10].to_frame(name='counts')
```

使用 `dataframe.join(other_dataframe)`的方式聚合 均值数据

```
dog_type_counts = dog_type_counts \
.join(dog_type_gb.agg({'retweet_count':  'median'}).rename(columns={'retweet_count':
'retweet_count_median'})) \
.join(dog_type_gb.agg({'favorite_count':
'median'}).rename(columns={'favorite_count': 'favorite_count_median'})) \
.join(dog_type_gb.agg({'rating_numerator':
'median'}).rename(columns={'rating_numerator': 'rating_numerator_median'})) \
.join(dog_type_gb.agg({'retweet_count':    'mean'}).rename(columns={'retweet_count':
'retweet_count_mean'})) \
.join(dog_type_gb.agg({'favorite_count':  'mean'}).rename(columns={'favorite_count':
'favorite_count_mean'})) \
.join(dog_type_gb.agg({'rating_numerator':
'mean'}).rename(columns={'rating_numerator': 'rating_numerator_mean'}))
```
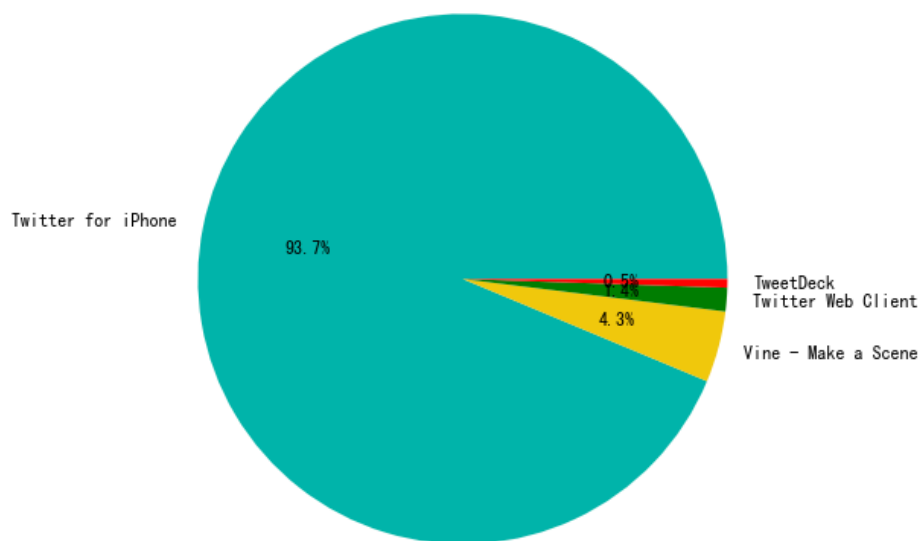
最后数据形式

	dog_type	counts	retweet_count_median	favorite_count_median	rating_numerator_median	retweet_count_mean	favorite_count_mean	rating_numerator_mean
0	golden_retriever	137	2283.0	8230.0	12.0	3680.919708	12439.430657	12.266423
1	Labrador_retriever	94	1736.0	5560.5	11.0	3862.382979	11835.574468	13.936170
2	Pembroke	88	1964.0	6521.0	12.0	3170.170455	11270.022727	11.443182
3	Chihuahua	78	934.5	2935.0	11.0	2800.500000	8603.653846	10.692308
4	pug	54	985.0	2624.0	10.0	1827.314815	5909.240741	10.240741
5	chow	41	2213.0	6837.0	12.0	2900.634146	10102.195122	11.609756
6	Samoyed	40	2679.0	7438.0	12.0	4597.425000	13533.625000	11.700000
7	Pomeranian	38	1138.0	3053.0	11.0	2863.105263	8103.210526	11.151316
8	toy_poodle	37	1146.0	4128.0	11.0	2499.351351	6746.540541	11.054054
9	malamute	29	2177.0	6984.0	11.0	2530.241379	8559.206897	10.896552
10	cocker_spaniel	27	2035.0	8444.0	12.0	3704.592593	11655.370370	11.296296
11	French_bulldog	26	2745.5	10379.5	12.0	4656.384615	17854.769231	11.269231
12	Chesapeake_Bay_retriever	23	1819.0	5729.0	10.0	2832.434783	10550.826087	10.782609
13	miniature_pinscher	21	1526.0	3864.0	11.0	2469.476190	8913.476190	10.666667
14	seat_belt	21	2024.0	4994.0	10.0	2753.333333	7936.333333	9.047619
15	German_shepherd	20	1873.5	7119.5	11.5	2643.050000	10116.400000	11.000000
16	Siberian_husky	20	1235.0	3620.0	11.5	1714.300000	6985.750000	13.250000
17	Staffordshire_bullterrier	19	1549.0	4340.0	11.0	2453.000000	9807.526316	10.684211
18	Maltese_dog	18	825.5	2464.5	10.0	1249.166667	2970.111111	9.277778
19	Eskimo_dog	18	1403.0	5386.5	12.0	4742.444444	11954.944444	11.777778
20	Shetland_sheepdog	18	1354.5	4248.0	10.5	2398.666667	8555.333333	10.388889
21	teddy	17	1101.0	3515.0	11.0	1646.000000	5817.000000	18.588235
22	Shih-Tzu	17	578.0	2756.0	11.0	1508.294118	3816.588235	10.529412
23	beagle	17	872.0	3911.0	10.0	1406.235294	4708.294118	10.352941
24	Cardigan	17	2796.0	10556.0	12.0	4513.647059	15251.705882	11.529412
25	Italian_greyhound	16	1028.0	2987.0	10.5	2310.250000	8698.000000	9.937500
26	Rottweiler	16	1082.5	3344.5	12.0	2436.062500	7950.937500	11.187500
27	Lakeland_terrier	15	1077.0	3086.0	11.0	4673.066667	13163.800000	10.466667
28	kuvasz	15	1691.0	4959.0	11.0	1543.266667	5593.733333	10.350667
29	web_site	14	4927.5	15149.5	13.0	5765.428571	17750.000000	12.214286
30	West_Highland_white_terrier	14	914.5	2925.0	10.0	1547.142857	6482.642857	15.642857
31	Great_Pyrenees	14	2032.5	6927.5	12.0	4704.000000	12878.357143	14.928571
32	basset	13	2654.0	8820.0	11.0	3245.846154	13830.076923	11.153846
33	American_Staffordshire_terrier	13	1697.0	5903.0	11.0	2193.461538	7528.230769	11.000000
34	Pekinese	13	1776.0	5092.0	12.0	2090.769231	7137.615385	11.153846
35	vizsla	13	1656.0	5735.0	11.0	1982.384615	7462.692308	10.846154
36	dalmatian	12	1380.0	5015.0	9.5	1731.000000	5164.583333	9.333333
37	Border_collie	12	1242.5	4108.0	11.5	2197.083333	8899.583333	10.583333
38	Airedale	12	987.0	3021.5	10.5	1387.000000	5259.000000	9.833333
39	kelpie	11	2433.0	9489.0	11.0	2461.181818	9944.181818	11.454545
40	Old_English_sheepdog	11	1489.0	4992.0	11.0	2578.818182	8241.727273	10.909091

推文来源比例图

```python
labels = twitter_archive_clean.source.value_counts().index
sizes = twitter_archive_clean.source.value_counts().values
fig = plt.figure(1, figsize=(8,8))
ax = fig.add_subplot(111)
colors = ('#01B8AA', '#F2C80F', 'g', 'r', 'c', 'm', 'y', 'burlywood', 'w')
ax.pie(sizes,colors=colors ,labels=labels, autopct='%1.1f%%')
#ax.title('Twitter sending platform',fontsize=15)
```

按 dog_type 分类统计 点赞数均值、转发数均值、评分数均值的关系

● 结论

1. 最常见的狗狗：golden_retriever、Labrador_retriever、Pembroke、Chihuahua 人们的评分、点赞、转发都偏高

2. 人们最喜欢的狗狗：French_bulldog、Gardigan，评分相对也高

3. 评分最高的是 teddy，但人们不喜欢点赞、转发和 teddy 相关的推文