

Twitter Dataset Analysis Summary

After investigating the dataset and going through the data wrangling process we've come to some conclusions and insights.

Questions about the data :

- *What's the highest rated tweet?*
- *Does having hashtags affect tweets popularity*
- *What's the most viral tweet?*
- *Which dogs are the most viral*
- *Which dogs are the least viral*

1. *The highest rated tweet was a tweet with a labrador dog called atticus*

```

[130]: tweetraterank = newdf2.sort_values(by='rating_numerator',ascending=False)
       tweetraterank.iloc[:10,[0,1,2,3,4,5,6]]

[130]:
```

	tweet_id	source_x	rating_numerator	rating_denominator	name	Nickname	retweet_count
1376	749981277374128128	<a href="https://about.twitter.com/products/tw...	1776	10	Atticus	None	2772
2041	835246439529840640	<a href="http://twitter.com/download/iphone" r...	960	0	None	None	83
2165	855860136149123072	<a href="http://twitter.com/download/iphone" r...	666	10	None	None	1058
2166	855862651834028034	<a href="http://twitter.com/download/iphone" r...	420	10	None	None	27
281	670842764863651840	<a href="http://twitter.com/download/iphone" r...	420	10	None	None	4324



WeRateDogs ✓
@dog_rates · [Follow](#)



This is Atticus. He's quite simply America af. 1776/10



5:00 PM · Jul 4, 2016



4.6 K



Reply



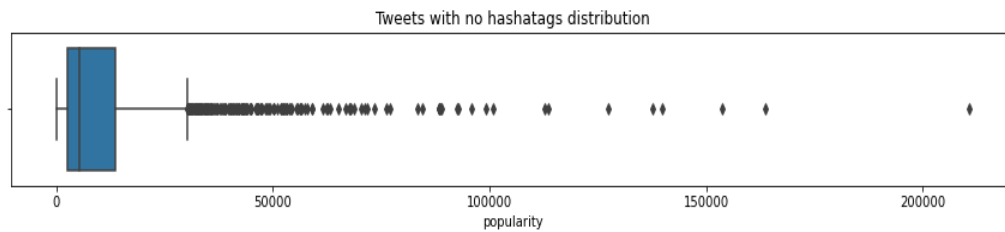
Share

We also concluded that the irregular denominators aren't wrong as some tweets do have those numbers



We can find that in the data gathering process some irregular values were added

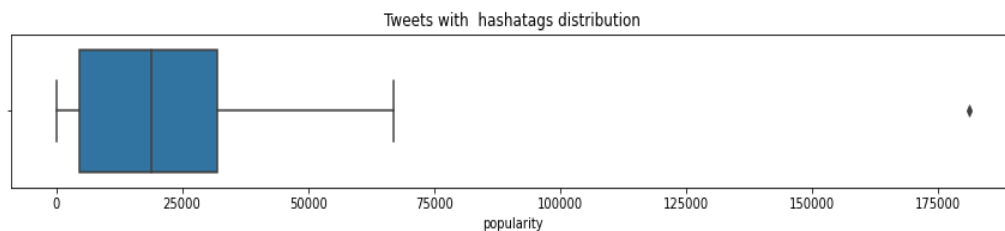
- 2. To find if having hashtags affect the popularity
First we look at the distribution of both the categories*



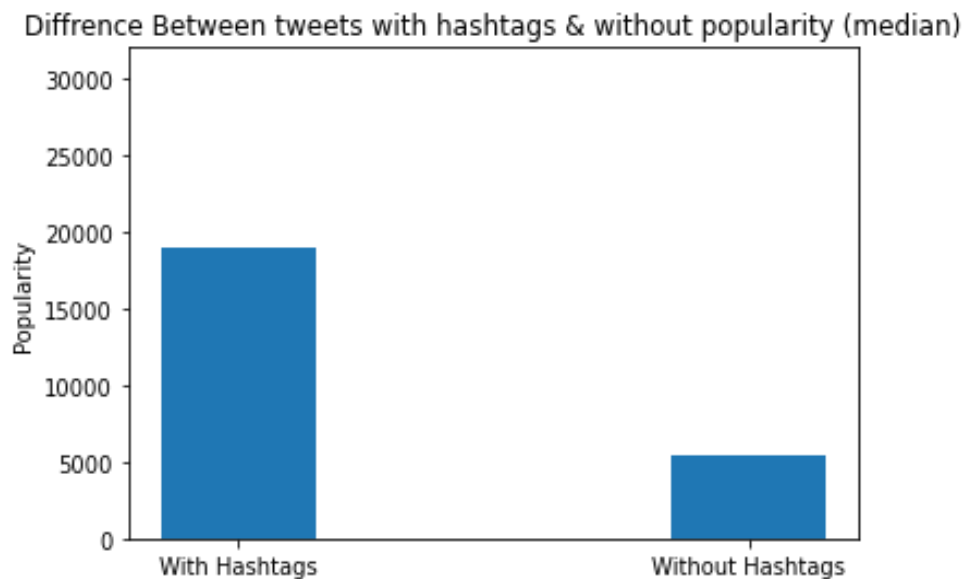
while Tweets with Hashtags has way less Data Points And outliers affect the mean alot

```
[174]: plt.figure(figsize=(15, 2))
sns.boxplot(x=hashtagwt['popularity'])
plt.title('Tweets with hashtags distribution')
```

```
[174]: Text(0.5, 1.0, 'Tweets with hashtags distribution')
```



We can see tweets with hashtags are a lot less but and have less outliers so we can't compare using the mean so we used median



We can see tweets with hashtags has a higher average of popularity (Note:popularity is measured by the sum of retweets , favourites)

3. The most viral tweet by popularity was found to be a tweet that has a golden retriever

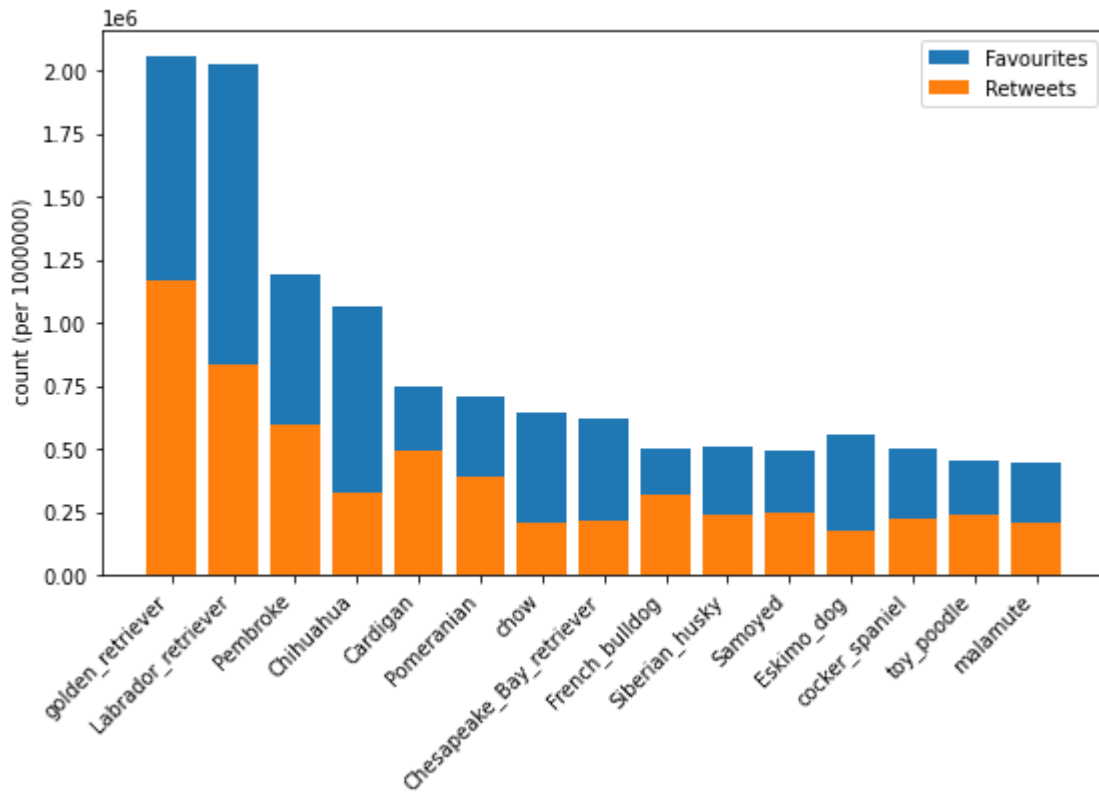
Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad)

```
[187]: viraldf = newdf2.copy()
viraldf['popularity'] = viraldf[['retweet_count', 'favorite_count']].sum(1)
viraldf.sort_values('popularity', ascending=False, inplace=True)
viraldf.iloc[:, [0, -1, 3, 4, 6, 7, 9]]
```

	tweet_id	popularity	rating_denominator	name	retweet_count	favorite_count	p2
1316	744234799360020481	210590	10	None	79515	131075	golden_retriever
1941	822872901745569793	181075	10	None	48265	132810	Pembroke
1820	807106840509214720	163640	10	Stephan	56625	107015	whippet
2287	879415818425184262	153805	10	Duddles	45849	107956	NaN
2218	866450705531457537	139710	10	Jamesy	32883	106827	NaN

4. From our analysis we can conclude that the golden retriever is the most favourable dog (the dogs were sorted on popularity first)

Most viral Dogs (retweets | favorties)



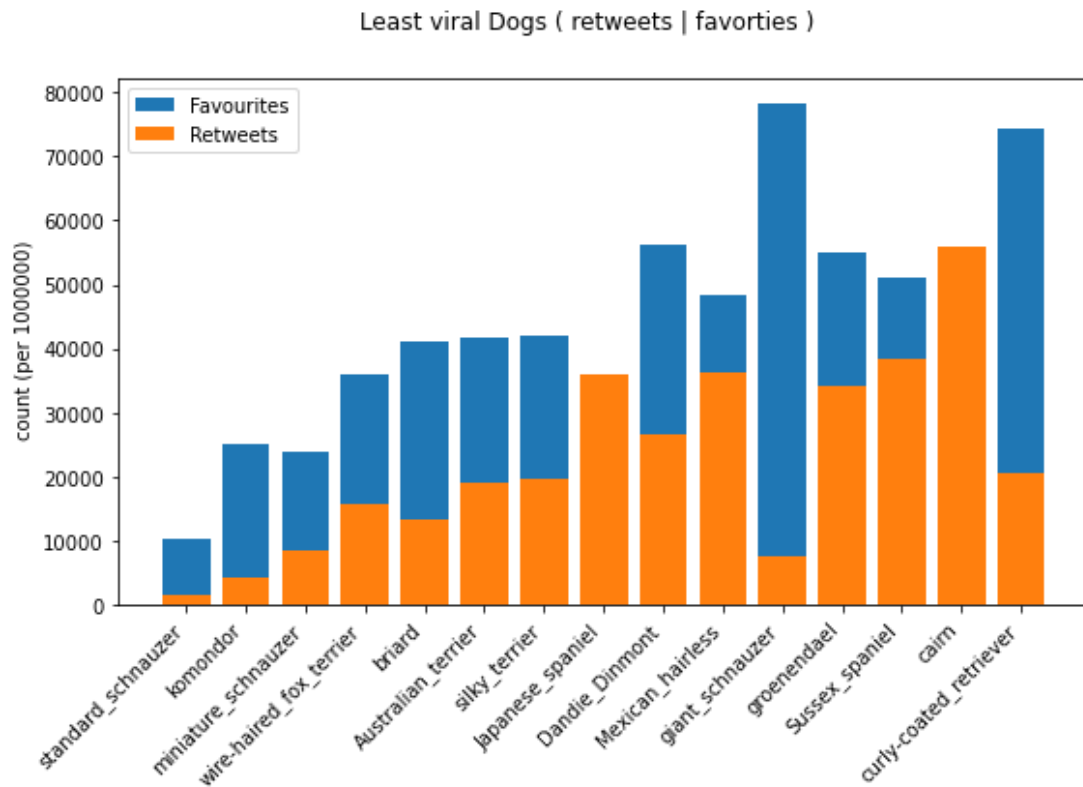
[161]:

hd

[161]:

	name	retweets_count	favorites_count	popularity
81	golden_retriever	1167577	2052552	3220129
32	Labrador_retriever	836109	2022311	2858420
44	Pembroke	594198	1189560	1783758
14	Chihuahua	326463	1063025	1389488
12	Cardigan	491668	744097	1235765
45	Pomeranian	392407	709674	1102081
74	chow	210842	641579	852421
13	Chesapeake_Bay_retriever	215829	619367	835196
21	French_bulldog	320133	502013	822146
53	Siberian_husky	241053	509428	750481
50	Samoyed	245457	496265	741722
20	Eskimo_dog	172559	553603	726162
75	cocker_spaniel	222739	499081	721820
100	toy_poodle	238448	453865	692313
87	malamute	207013	446082	653095

5. *Least Viral Dogs from our analysis the standard schnauzer is the least favourite dog by a big difference*



We also used popularity to sort the values and we sorted them from lowest to highest so the least is shown first

[162]:

ld

[162]:

	name	retweets_count	favorites_count	popularity
99	standard_schnauzer	1615	10265	11880
85	komondor	4356	25232	29588
91	miniature_schnauzer	8437	23792	32229
104	wire-haired_fox_terrier	15725	36126	51851
71	briard	13309	41083	54392
4	Australian_terrier	19236	41736	60972
96	silky_terrier	19777	41873	61650
31	Japanese_spaniel	35951	34604	70555
15	Dandie_Dinmont	26650	56237	82887
37	Mexican_hairless	36147	48436	84583
80	giant_schnauzer	7496	78080	85576
82	groenendael	34073	54912	88985
55	Sussex_spaniel	38408	51028	89436
73	cairn	55964	37367	93331
77	curly-coated_retriever	20704	74361	95065