

## Summary of findings from WeRateDogs

### Key Findings

- Golden Retriever was the most common dog type on WeRateDogs (excluding the Unclassified dog breeds)
- The breed with the highest average WeRateDogs rating was the Afghan Hound with an average rating of 1.3. The lowest average rating was for the poor Scotch Terrier.
- On average the Bedlington Terrier breed receives the most Likes (Favorite)

We shall now review how I arrived at the above conclusion

```
In [3]: import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

## Data Gathering

First of I will import my data that was gathered, assessed and cleaned. This data came from a variety of sources:

1. Local Source - twitter-archive-enhanced.csv
2. Downloading .tsv file from [url](https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv)  
([https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad\\_image-predictions/image-predictions.tsv](https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv))
3. Retrieving data from Twitters API

```
In [4]: #import saved data
df = pd.read_csv('twitter_archive_master.csv')
df.head(2)
```

Out[4]:

|   | tweet_id           | timestamp                    | text  | expanded_u  |
|---|--------------------|------------------------------|---|---|
| 0 | 892420643555336193 | 2017-08-01<br>16:23:56+00:00 | This is<br>Phineas.<br>He's a<br>mystical<br>boy.<br>Only<br>eve... | <a href="https://twitter.com/dog_rates/status/89242064">https://twitter.com/dog_rates/status/89242064</a> |
| 1 | 892177421306343426 | 2017-08-01<br>00:17:27+00:00 | This is<br>Tilly.<br>She's<br>just<br>checking<br>pup on<br>you.... | <a href="https://twitter.com/dog_rates/status/89217742">https://twitter.com/dog_rates/status/89217742</a> |

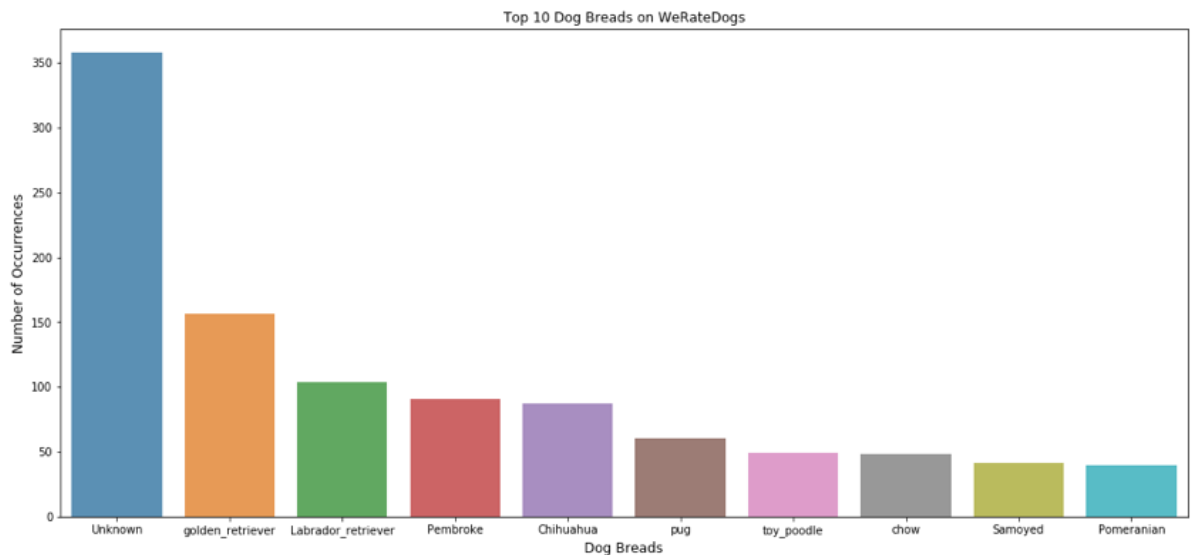
## On to the EDA

```
In [5]: #What is the most common dog type on WeRateDogs?
df.Dog_Prediction.value_counts()
```

```
Out[5]: Unknown                358
golden_retriever             156
Labrador_retriever          104
Pembroke                     91
Chihuahua                    87
...
EntleBucher                   1
Irish_wolfhound               1
silky_terrier                 1
Bouvier_des_Flandres         1
standard_schnauzer           1
Name: Dog_Prediction, Length: 113, dtype: int64
```

As we can see, if we ignore the Unkown breads of dog, the Golden Retriever bread is by far the most common. This is better visualied in the bar chart below

```
In [6]: # Top 10
dog_types = df.Dog_Prediction.value_counts()[:]
dog_types = dog_types[:10,]
plt.figure(figsize=(18,8))
sns.barplot(dog_types.index, dog_types.values, alpha=0.8)
plt.title('Top 10 Dog Breeds on WeRateDogs')
plt.ylabel('Number of Occurrences', fontsize=12)
plt.xlabel('Dog Breeds', fontsize=12)
plt.show()
```



```
In [7]: #Which breed of dog receives the highest/lowest average rating?
df['Rating'] = df.apply(lambda x: x['rating_numerator']/x['rating_denominator'],axis=1)
```

```
In [8]: df['Rating'].groupby(df['Dog_Prediction']).mean().sort_values(ascending=False)
```

```
Out[8]: Dog_Prediction
Unknown                1.696089
Afghan_hound           1.300000
Bouvier_des_Flandres   1.300000
Saluki                 1.250000
briard                 1.233333
...
redbone                0.966667
Welsh_springer_spaniel 0.950000
Tibetan_terrier        0.925000
Ibizan_hound           0.925000
Scotch_terrier         0.900000
Name: Rating, Length: 113, dtype: float64
```

Once again, the Unknown dog breed appears top of the rankings, however this is omitted from the assessment.

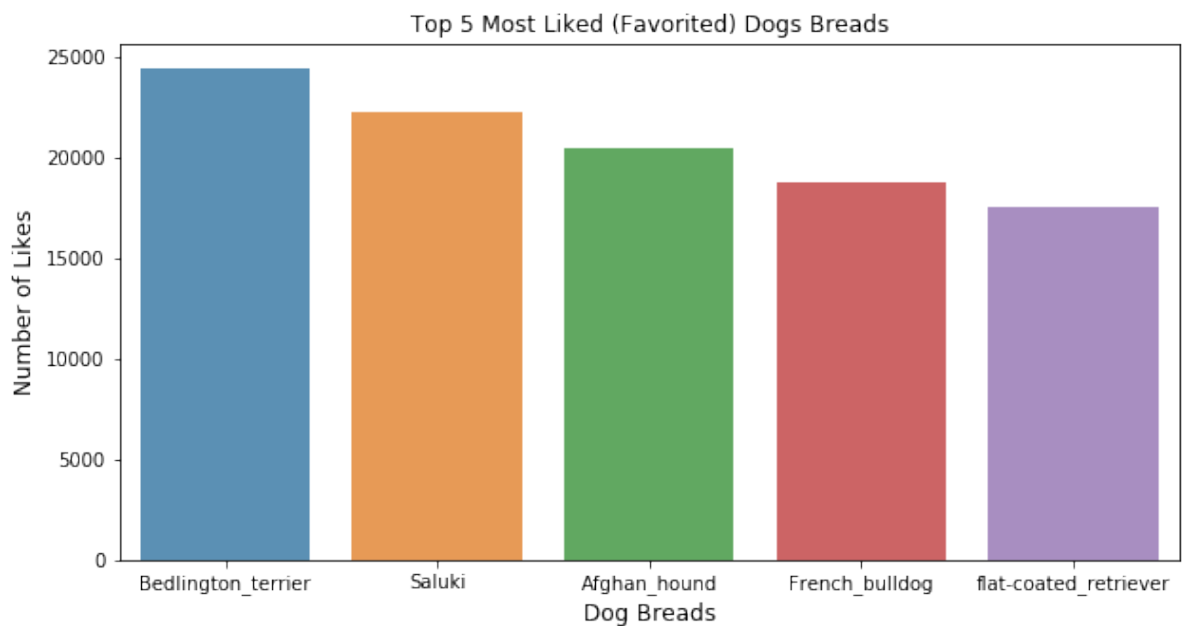
As we can see, if we ignore the Unknown breeds of dog, the Afghan Hound breed has the highest average WeRateDog rating with an excellent 1.3 (13/10)

Unfortunately for the Scotch Terrier, it has the lowest average WeRateDog rating of only 0.9 (9/10).

```
In [9]: #Which breed of dog on average receives the most likes?  
df['favorite'].groupby(df['Dog_Prediction']).mean().sort_values(ascending=False)
```

```
Out[9]: Dog_Prediction  
Bedlington_terrier      24450.600000  
Saluki                  22272.750000  
Afghan_hound            20521.500000  
French_bulldog          18787.866667  
flat-coated_retriever   17542.857143  
...  
standard_schnauzer      1855.000000  
Tibetan_terrier         1681.750000  
Irish_wolfhound         1392.000000  
groenendael             1055.500000  
Brabancon_griffon       807.000000  
Name: favorite, Length: 113, dtype: float64
```

```
In [20]: # Top 10
dog_types = df['favorite'].groupby(df['Dog_Prediction']).mean().sort_values(ascending=False)
dog_types = dog_types[:5,]
plt.figure(figsize=(10,5))
sns.barplot(dog_types.index, dog_types.values, alpha=0.8)
plt.title('Top 5 Most Liked (Favorited) Dogs Breads')
plt.ylabel('Number of Likes', fontsize=12)
plt.xlabel('Dog Breads', fontsize=12)
plt.show()
```



The above chart clearly outlines that the Bedlington Terrier breed does on average receive the most likes on the WeRateDogs Twitter feed.

This is very fortunate for all those Bedlington Terrier loving owners

```
In [30]: #The top rated on Dog of WeRateDogs
df[df['favorite'] == df['favorite'].max()]
```

Out[30]:

|     | tweet_id           | timestamp                 | text  | expanded_   |
|-----|--------------------|---------------------------|---|---|
| 808 | 744234799360020481 | 2016-06-18 18:26:18+00:00 | Here's a doggo realizing you can stand in a po... | <a href="https://twitter.com/dog_rates/status/744234799360020481">https://twitter.com/dog_rates/status/744234799360020481</a> |

In [31]: `df.describe()`

Out[31]:

|              | tweet_id     | rating_numerator | rating_denominator | img_num     | retweets     |     |
|--------------|--------------|------------------|--------------------|-------------|--------------|-----|
| <b>count</b> | 1.972000e+03 | 1972.000000      | 1972.000000        | 1850.000000 | 1972.000000  | 1   |
| <b>mean</b>  | 7.391910e+17 | 12.603844        | 10.476166          | 1.212973    | 2622.258621  | 8   |
| <b>std</b>   | 6.739823e+16 | 41.551655        | 6.845541           | 0.572532    | 4541.156738  | 12  |
| <b>min</b>   | 6.660209e+17 | 7.000000         | 7.000000           | 1.000000    | 11.000000    |     |
| <b>25%</b>   | 6.783704e+17 | 10.000000        | 10.000000          | 1.000000    | 572.500000   | 1   |
| <b>50%</b>   | 7.141778e+17 | 11.000000        | 10.000000          | 1.000000    | 1297.500000  | 4   |
| <b>75%</b>   | 7.916980e+17 | 12.000000        | 10.000000          | 1.000000    | 2984.750000  | 10  |
| <b>max</b>   | 8.924206e+17 | 1776.000000      | 170.000000         | 4.000000    | 78119.000000 | 156 |

In [ ]: