act_report

January 10, 2023

0.1 Reporting: `act_report` by Abdulraqib Omotosho

This document communicates all the insights and displays the visualizations I produced from the data I wrangled.

0.1.1 Insights and Visualizations

- Most popular dog names: Tucker, Lucy and Cooper
- Most popuar dog type: Pupper
- Dog type with the highest confidence interval: Doggo
- Most popular dog breeds: golden_retriever
- Most rated dog breeds: golden_retriever
- Highest Mean retweet counts of dog types: doggo at 4914.918033
- Highest Mean retweet counts of dog breeds: Bedlington_terrier at 6562.333333
- The average count of the most liked dog types: doggo at 15632.688525
- The average count of the most liked dog breeds: Bedlington_terrier at 20995.666667
- The most popular source of tweets: Twitter for iPhone (1584)
- The highest average retweet counts by on dog names: Stephan at 50687.00

Importing the libraries used in the process

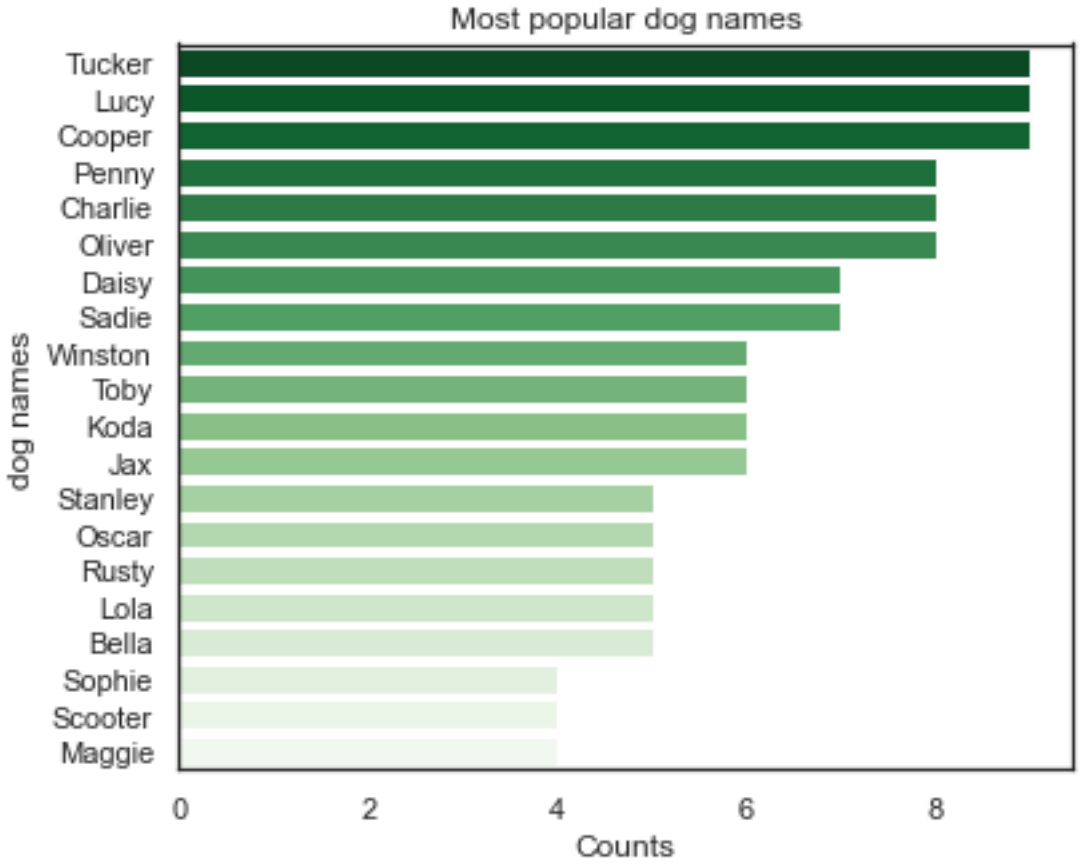
[28]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_theme(style='white')
from PIL import Image
import warnings
warnings.simplefilter(action='ignore')

df = pd.read_csv('twitter_archive_master.csv')
```

- Most popular dog names

[3]:
```
plt.figure(figsize=(6, 5))
x = df['names'].value_counts().head(20)
sns.barplot(x.values, x.index, palette='Greens_r').set(title='Most popular dog␣
 ↪names', xlabel='Counts', ylabel='dog names')
```

[3]:
```
[Text(0.5, 1.0, 'Most popular dog names'),
 Text(0.5, 0, 'Counts'),
```

```
Text(0, 0.5, 'dog names')]
```



[33]:
```
img = Image.open('tucker.jpg')
img
```

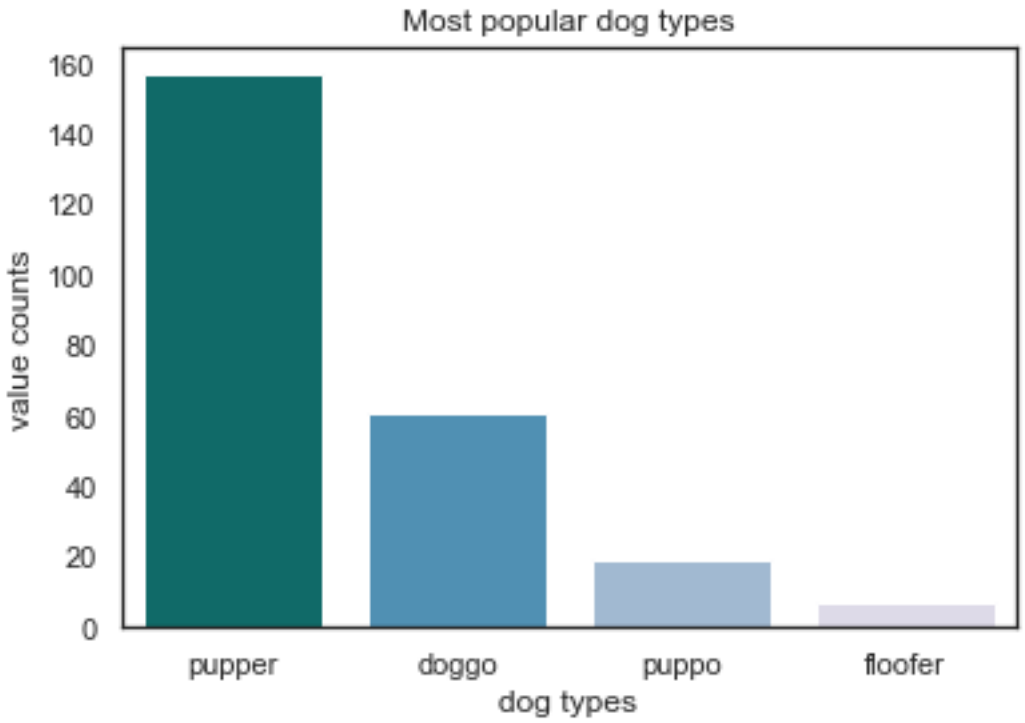
[33]:

The Tucker, Lucy and Cooper names are the most popular

- What dog types are the most popular

[5]:
```
x = df['types'].value_counts()
sns.barplot(x.index, x.values, palette='PuBuGn_r').set(title='Most popular dog␣
↪types', xlabel='dog types', ylabel='value counts')
```

[5]:
```
[Text(0.5, 1.0, 'Most popular dog types'),
 Text(0.5, 0, 'dog types'),
 Text(0, 0.5, 'value counts')]
```

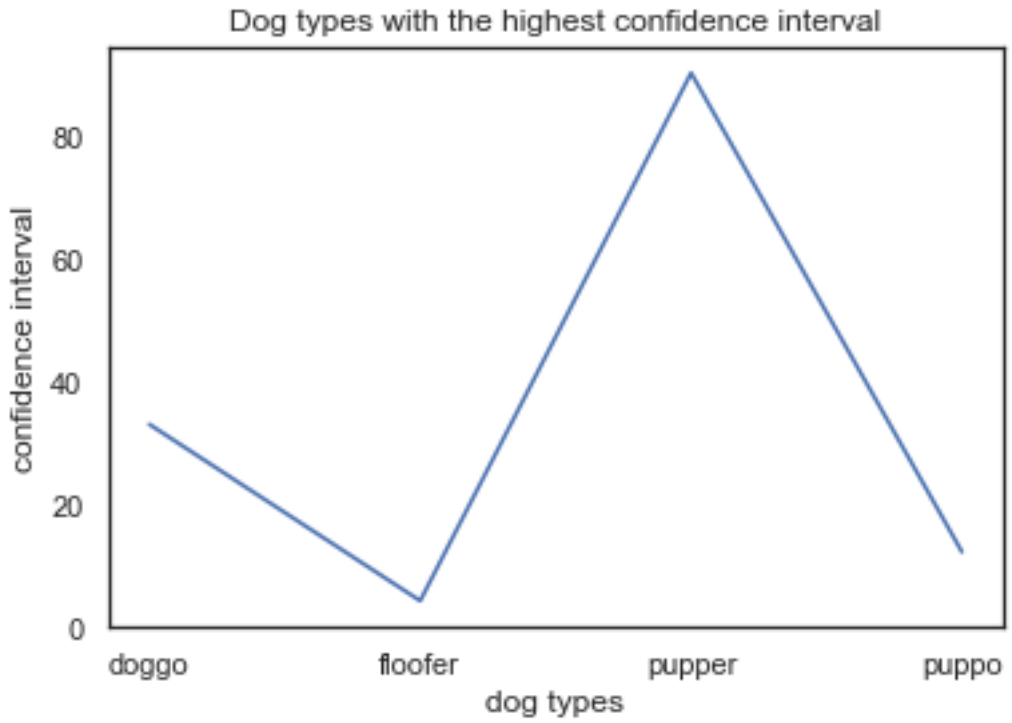


Pupper is the most popular dog type

- Which dog has the highest confidence interval

[7]:
```
x = df.groupby('types')['confidence'].sum()
sns.lineplot(x.index, x.values, ).set(title='Dog types with the highest␣
 ↪confidence interval', xlabel='dog types', ylabel='confidence interval')
```

[7]:
```
[Text(0.5, 1.0, 'Dog types with the highest confidence interval'),
 Text(0.5, 0, 'dog types'),
 Text(0, 0.5, 'confidence interval')]
```

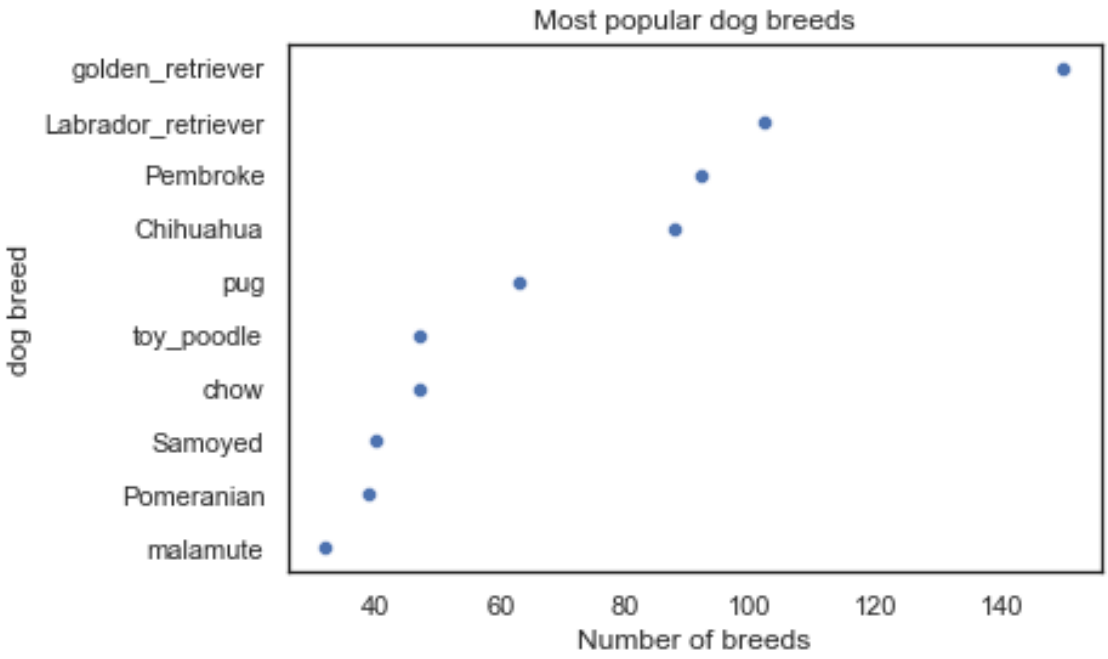


Pupper has the highest confidence interval

- Most popular dog breeds

[9]:
```
x = df['dog_breed'].value_counts().head(10)
sns.scatterplot(x.values, x.index).set(title='Most popular dog breeds',␣
 ↪xlabel='Number of breeds', ylabel='dog breed')
```

[9]:
```
[Text(0.5, 1.0, 'Most popular dog breeds'),
 Text(0.5, 0, 'Number of breeds'),
 Text(0, 0.5, 'dog breed')]
```



[35]:
```
Image.open('gr.jpg')
```

[35]:

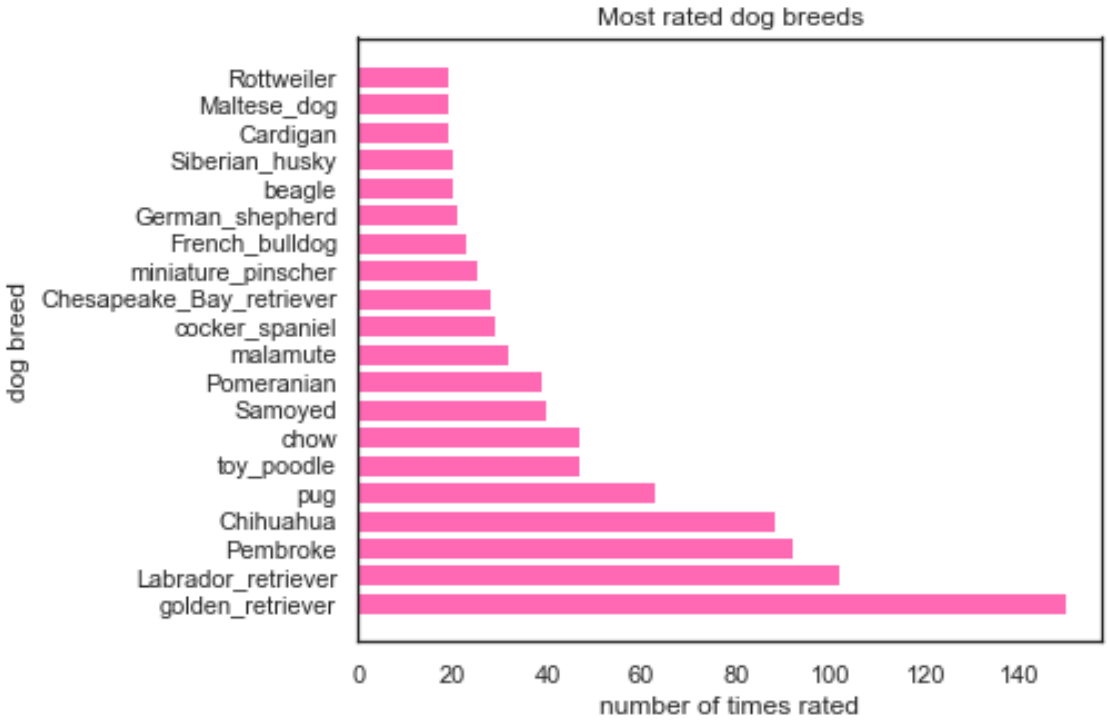
`golden_retrievers` are the most popular dog breeds

- Most rated dog breeds

[11]:
```
x = df.groupby('dog_breed')['rating'].size().sort_values(ascending=False).
↪head(20)
plt.figure(figsize=(6, 5))
```

```
plt.barh(x.index, x.values, color='hotpink')
plt.xlabel('number of times rated')
plt.ylabel('dog breed')
plt.title('Most rated dog breeds')
```

[11]:
```
Text(0.5, 1.0, 'Most rated dog breeds')
```

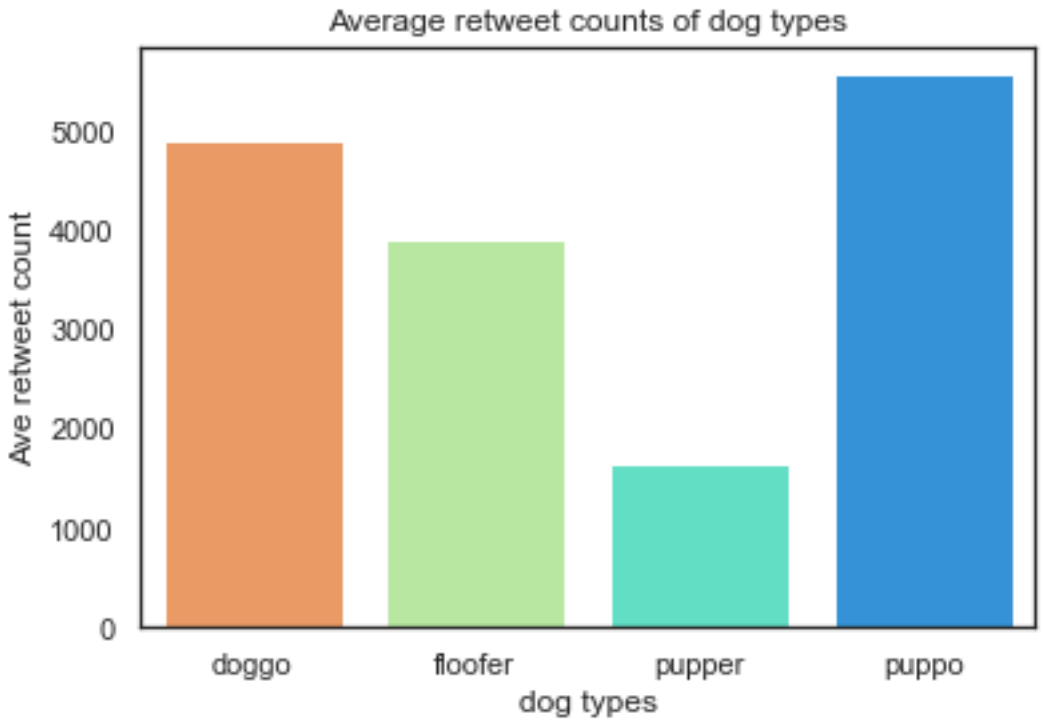


Rottweilers are the dogs rated the most times

- Mean retweet counts of dog types

[13]:
```
x = df.groupby('types')['retweet_count'].mean()
sns.barplot(x.index, x.values, palette='rainbow_r').set(title='Average retweet␣
↪counts of dog types', xlabel='dog types', ylabel='Ave retweet count')
```

[13]:
```
[Text(0.5, 1.0, 'Average retweet counts of dog types'),
 Text(0.5, 0, 'dog types'),
 Text(0, 0.5, 'Ave retweet count')]
```

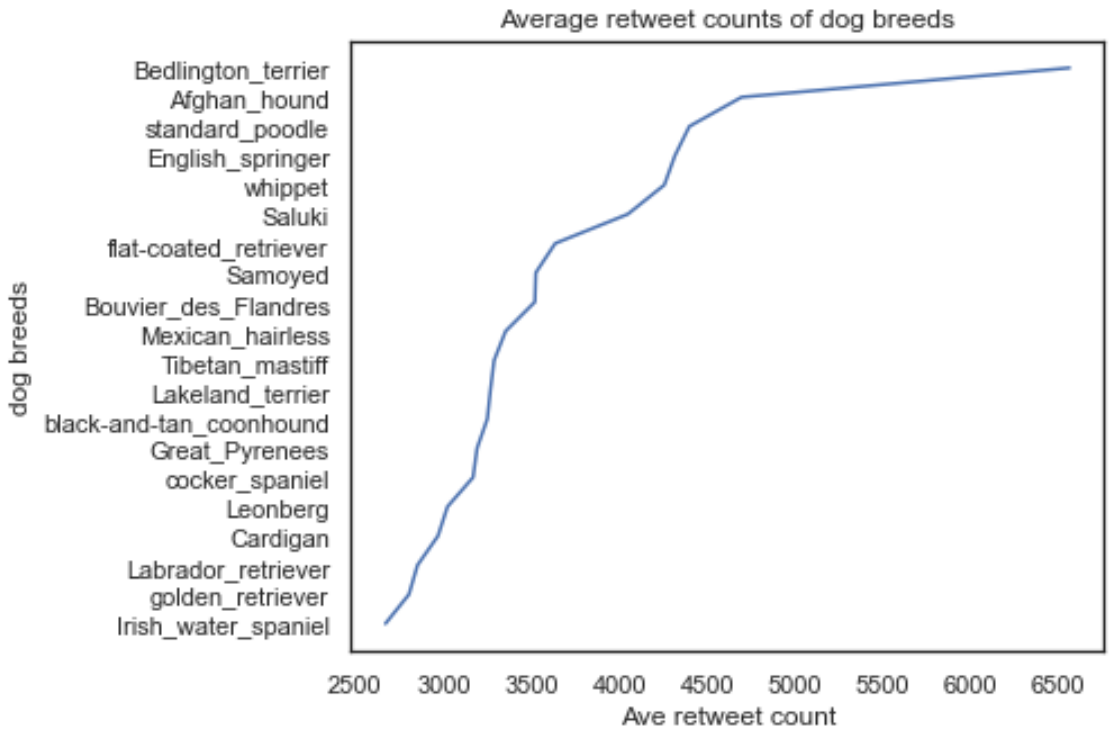


Puppo has the highest average retweet counts of any dog types

- Mean retweet counts of dog breeds

[15]:
```
plt.figure(figsize=(6, 5))
x = df.groupby('dog_breed')['retweet_count'].mean().
  ↪sort_values(ascending=False).head(20)
sns.lineplot(x.values, x.index, palette='rainbow_r').set(title='Average retweet␣
  ↪counts of dog breeds', ylabel='dog breeds', xlabel='Ave retweet count')
```

[15]:
```
[Text(0.5, 1.0, 'Average retweet counts of dog breeds'),
 Text(0, 0.5, 'dog breeds'),
 Text(0.5, 0, 'Ave retweet count')]
```

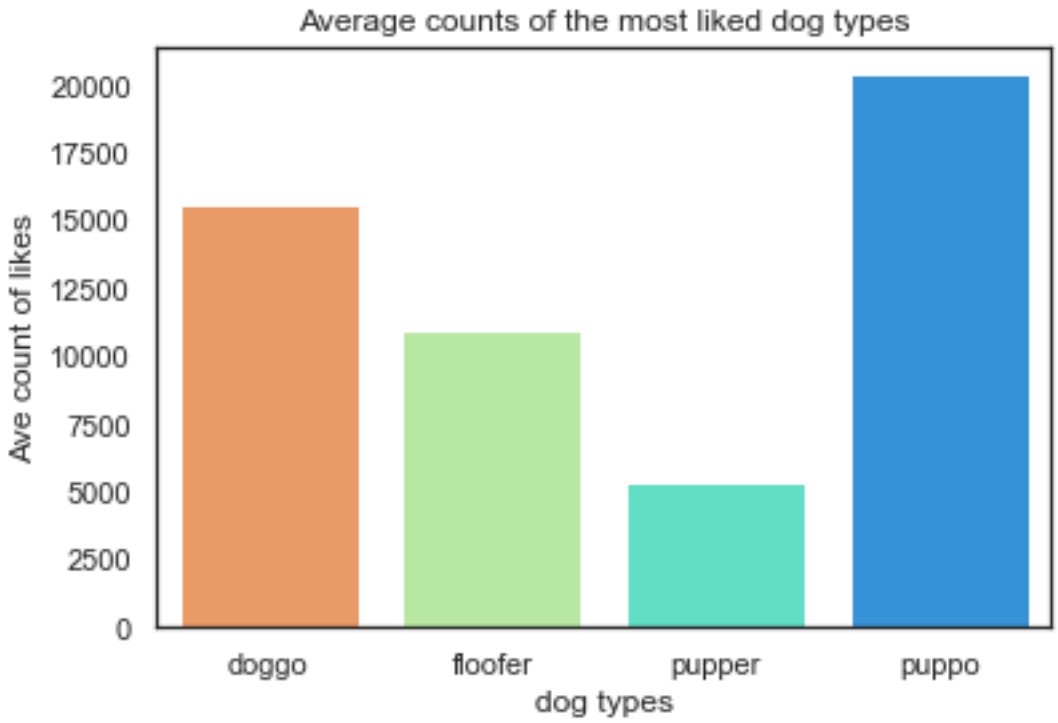


Bedlington_terrier has the highest average retweet counts of any dog breeds

- The average count of the most liked dog types

[17]:
```
x = df.groupby('types')['favorite_count'].mean()
sns.barplot(x.index, x.values, palette='rainbow_r').set(title='Average counts␣
 ↪of the most liked dog types', xlabel='dog types', ylabel='Ave count of␣
 ↪likes')
```

[17]:
```
[Text(0.5, 1.0, 'Average counts of the most liked dog types'),
 Text(0.5, 0, 'dog types'),
 Text(0, 0.5, 'Ave count of likes')]
```

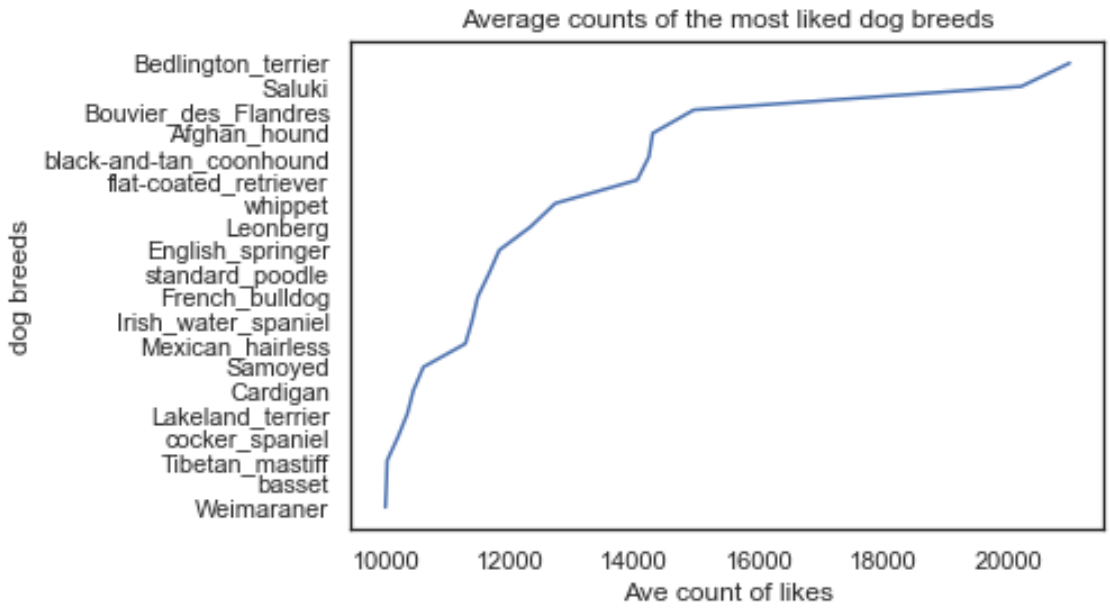


Puppo are the dogs with the highest favorite counts of any dog types

- The average count of the most liked dog breeds

[19]:
```
x = df.groupby('dog_breed')['favorite_count'].mean().
 ↪sort_values(ascending=False).head(20)
sns.lineplot(x.values, x.index, palette='rainbow_r').set(title='Average counts␣
 ↪of the most liked dog breeds', ylabel='dog breeds', xlabel='Ave count of␣
 ↪likes')
```

[19]:
```
[Text(0.5, 1.0, 'Average counts of the most liked dog breeds'),
 Text(0, 0.5, 'dog breeds'),
 Text(0.5, 0, 'Ave count of likes')]
```

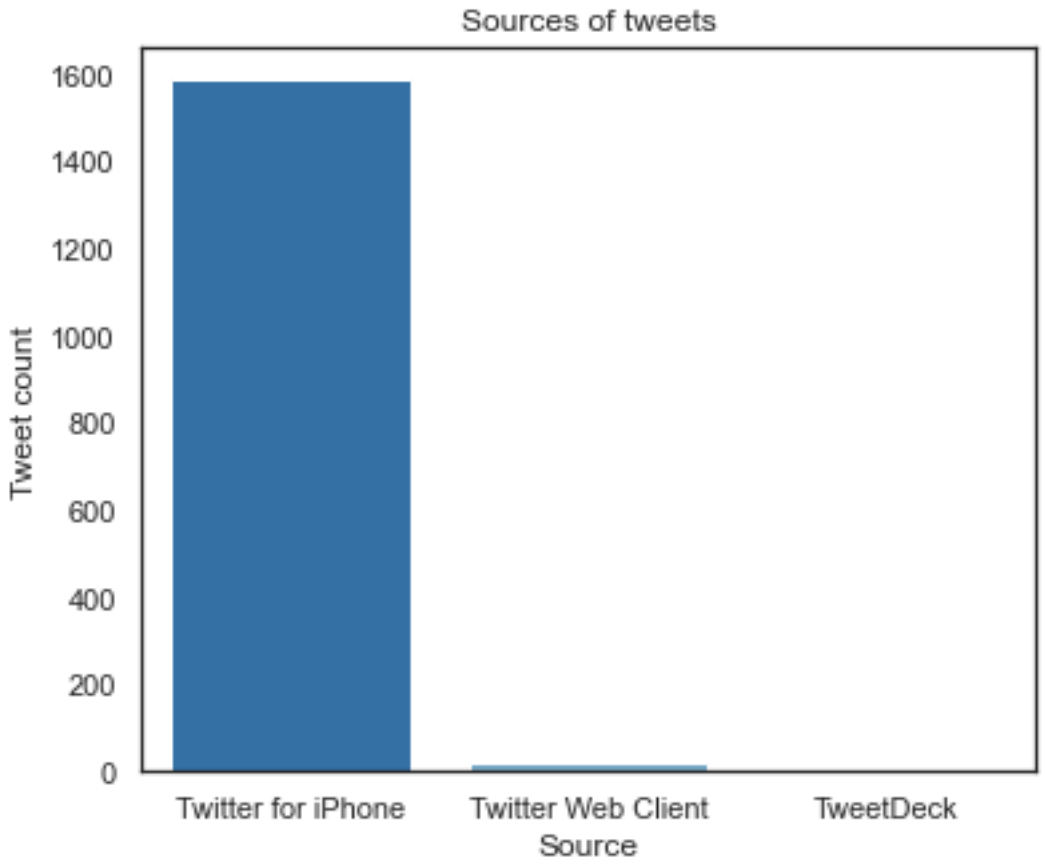


Bedlington_terrier are the dogs with the highest favorite counts of any dog breeds

- The most popular source of tweets

[21]:
```
plt.figure(figsize=(6, 5))
x = df.source.value_counts()
sns.barplot(x.index, x.values, palette='Blues_r').set(title='Sources of␣
↪tweets', xlabel='Source', ylabel='Tweet count')
```

[21]:
```
[Text(0.5, 1.0, 'Sources of tweets'),
 Text(0.5, 0, 'Source'),
 Text(0, 0.5, 'Tweet count')]
```

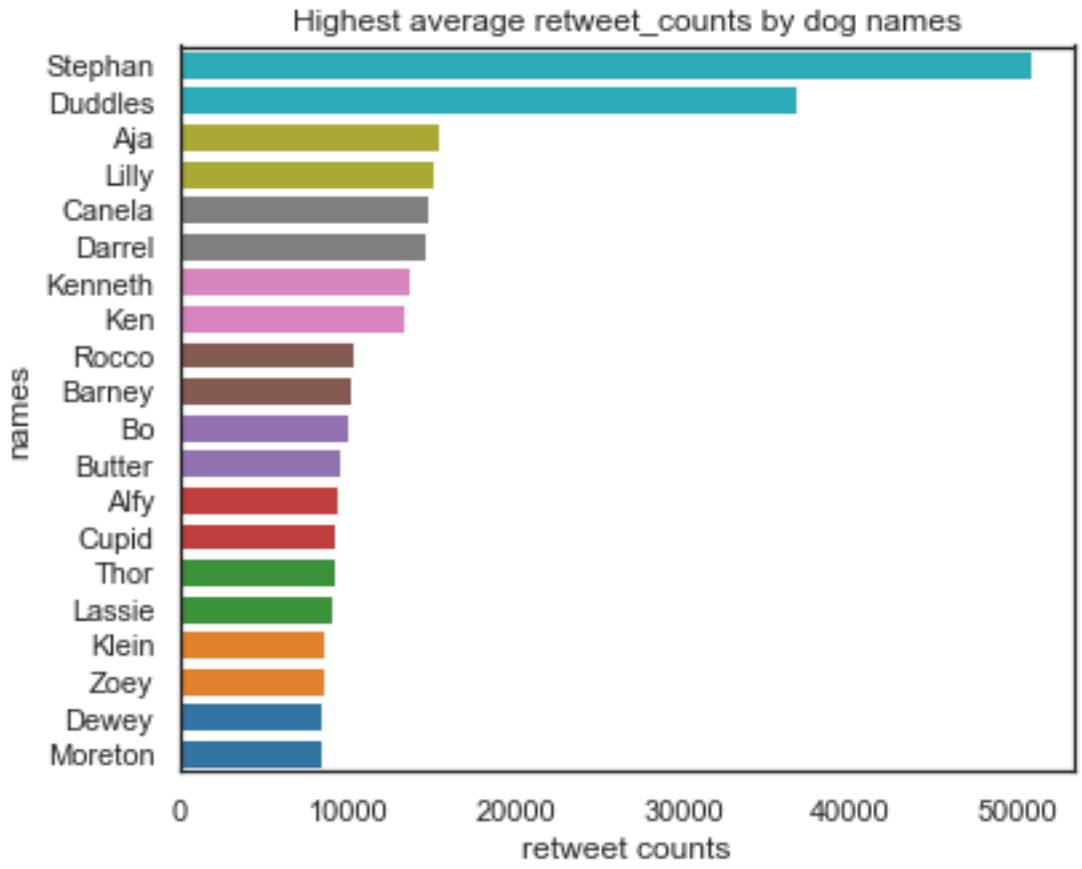


Most tweets where from an iPhone

- The highest average retweet counts by on dog names

[23]:
```
plt.figure(figsize=(6,5))
x = df.groupby('names')['retweet_count'].mean().sort_values(ascending=False).
 ↪head(20)
sns.barplot(x.values, x.index, palette='tab10_r').set(title='Highest average␣
 ↪retweet_counts by dog names', xlabel='retweet counts', ylabel='names')
```

[23]:
```
[Text(0.5, 1.0, 'Highest average retweet_counts by dog names'),
 Text(0.5, 0, 'retweet counts'),
 Text(0, 0.5, 'names')]
```



Dogs named `Stephan` had the highest average retweet counts

- Word Cloud of the tweets

[24]:
```
plt.figure(figsize=(10,10))
from wordcloud import WordCloud
text = " ".join(cat for cat in df.text)
wordcloud = WordCloud(collocations=False, background_color='white').
 ↪generate(text)

plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Word cloud of the tweets')
plt.show()
```

- Word Cloud of dog breeds

[25]:
```
plt.figure(figsize=(10,10))
from wordcloud import WordCloud
text = " ".join(cat for cat in df.dog_breed)
wordcloud = WordCloud(collocations=False, background_color='black', width=3000,␣
↪height=2000).generate(text)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Word Cloud of dog breeds')
plt.show()
```

Word Cloud of dog breeds