Data Report Analysis And Visualizations

Mohammed M. Huda

Version: 1.00
Date: May 24, 2020

ABSTRACT

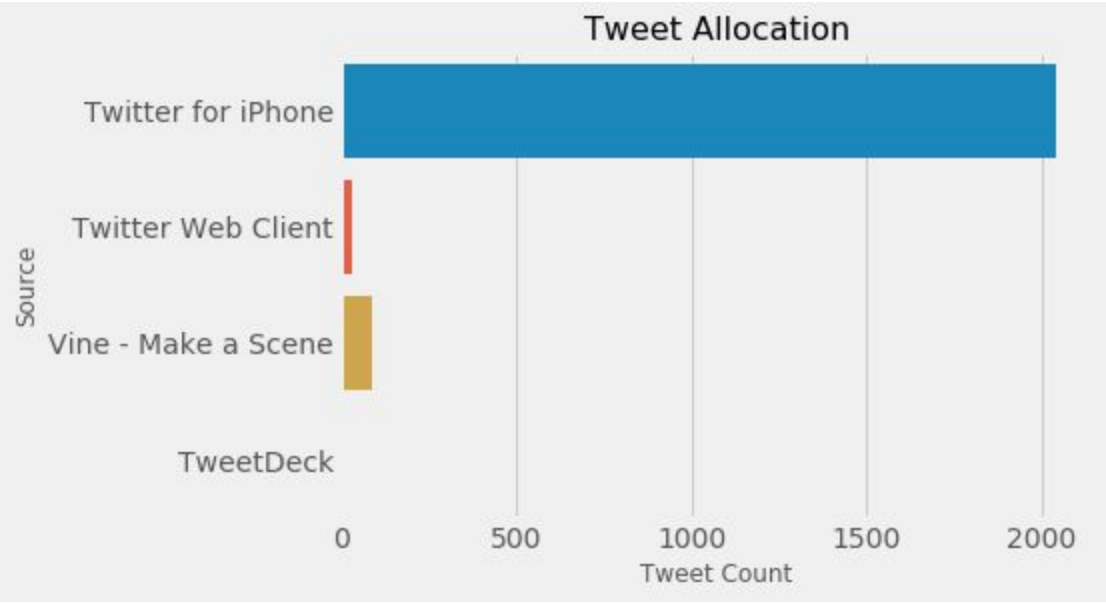
WeRateDogs is a Twitter account (@dog_rates) that rates dogs and promotes itself as "Your Only Source For Professional Dog Ratings." The account shares photos of people's dogs in a humorous manner and rates them on a scale of 0 to 10 (and sometimes higher than 10). WeRateDogs was launched in November 2015 and has over 8.7 million followers as of May 2020.

After obtaining the Twitter archive of WeRateDogs, we scraped the data, assessed it, cleaned it, and created some insights that helped us better understand the data. We came across some interesting observations as shown in the following visualizations.

VISUALIZATIONS

What is the most popular Application/Platform?

The most common application/platform used to post tweet messages to WeRateDogs is Twitter for iPhone, as shown in the figure below:



Code:

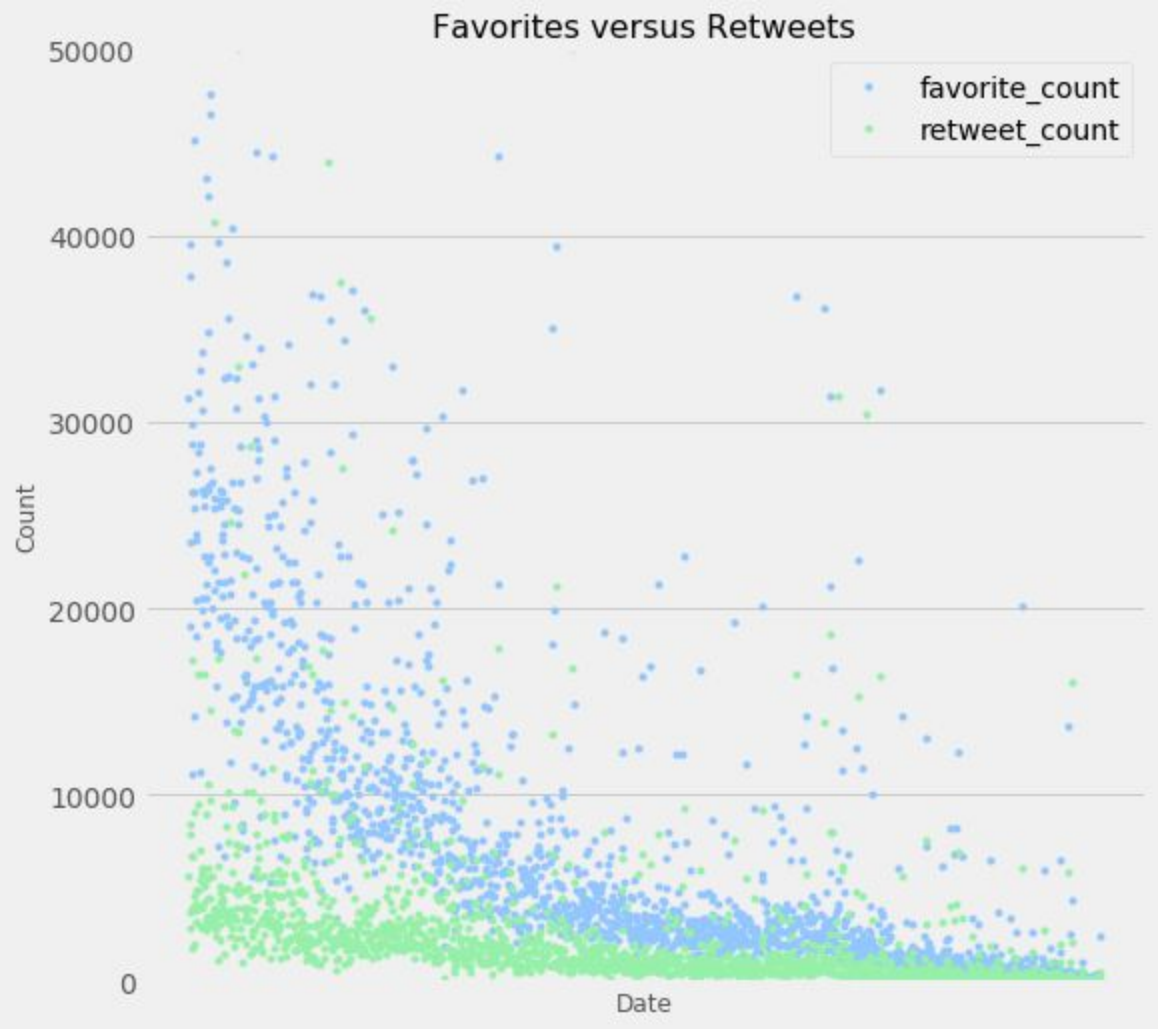
```
# Here we will display a graph showing which applications were used to
# post tweets to the WeRateDogs Twitter page.  We will use
# fivethirtyeight for this example.  The graph will display that
# Twitter for iPhone was the most commonly used application/platform
plt.style.use('fivethirtyeight')

sns.countplot(data = twitter_archive_clean, y = 'source')
plt.title('Tweet Allocation', fontsize=16)
plt.xlabel('Tweet Count', fontsize=12)
plt.ylabel('Source', fontsize=12)
plt.savefig('tweet-source.png')
```

Our data revealed that Twitter for iPhone accounted for 2042 tweet messages. Tweetdeck (which Twitter now owns) had the least number of tweet messages with just 11.

What is the difference between Favorites versus Retweets?

Favorites are more popular than retweets early on. Later, the number of favorites decreases and is in line with the number of retweets. Both the number of favorites and number of retweets are decreasing.



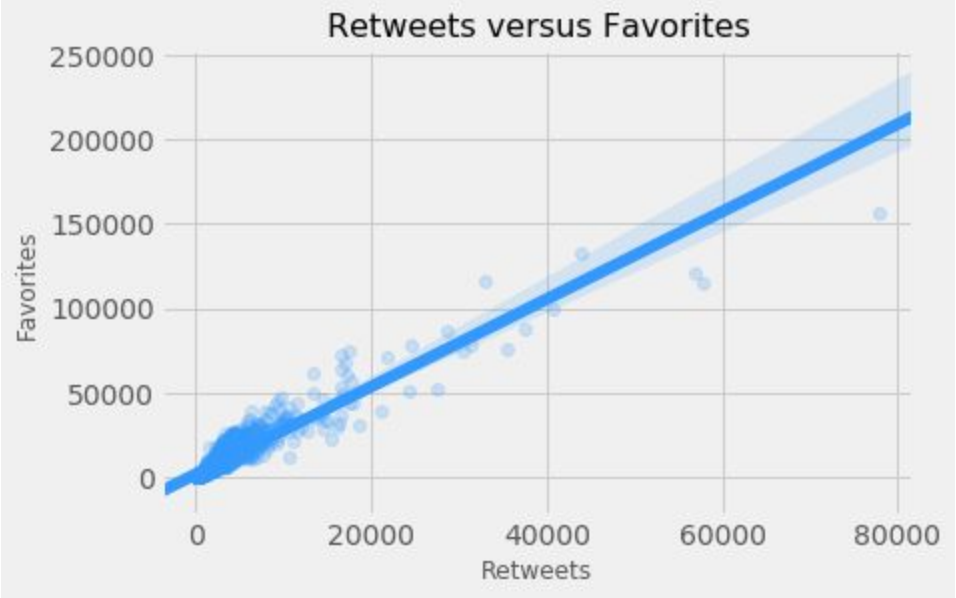
Code:

```
# Plotting favorites versus retweets.  This graph shows that favorites
# and retweets are sporadic in the beginning, but the data points
# converge over time and both decreasing

new_twitter_archive[['favorite_count', 'retweet_count']].plot(style =
'.', ylim=[0, 50000], figsize=(8,8))
plt.title('Favorites versus Retweets', size=16)
plt.xlabel('Date', size=12)
plt.xticks([], [])
plt.ylabel('Count', size=12)
plt.legend(ncol=1, loc='upper right')
plt.savefig('favorites-versus-retweets.png')
```

What is the difference between Retweets versus Favorites?

There is a strong positive correlation between the number of retweets and the number of favorites. The more a post is retweeted, the more the post is viewed and the more favorites the post receives.

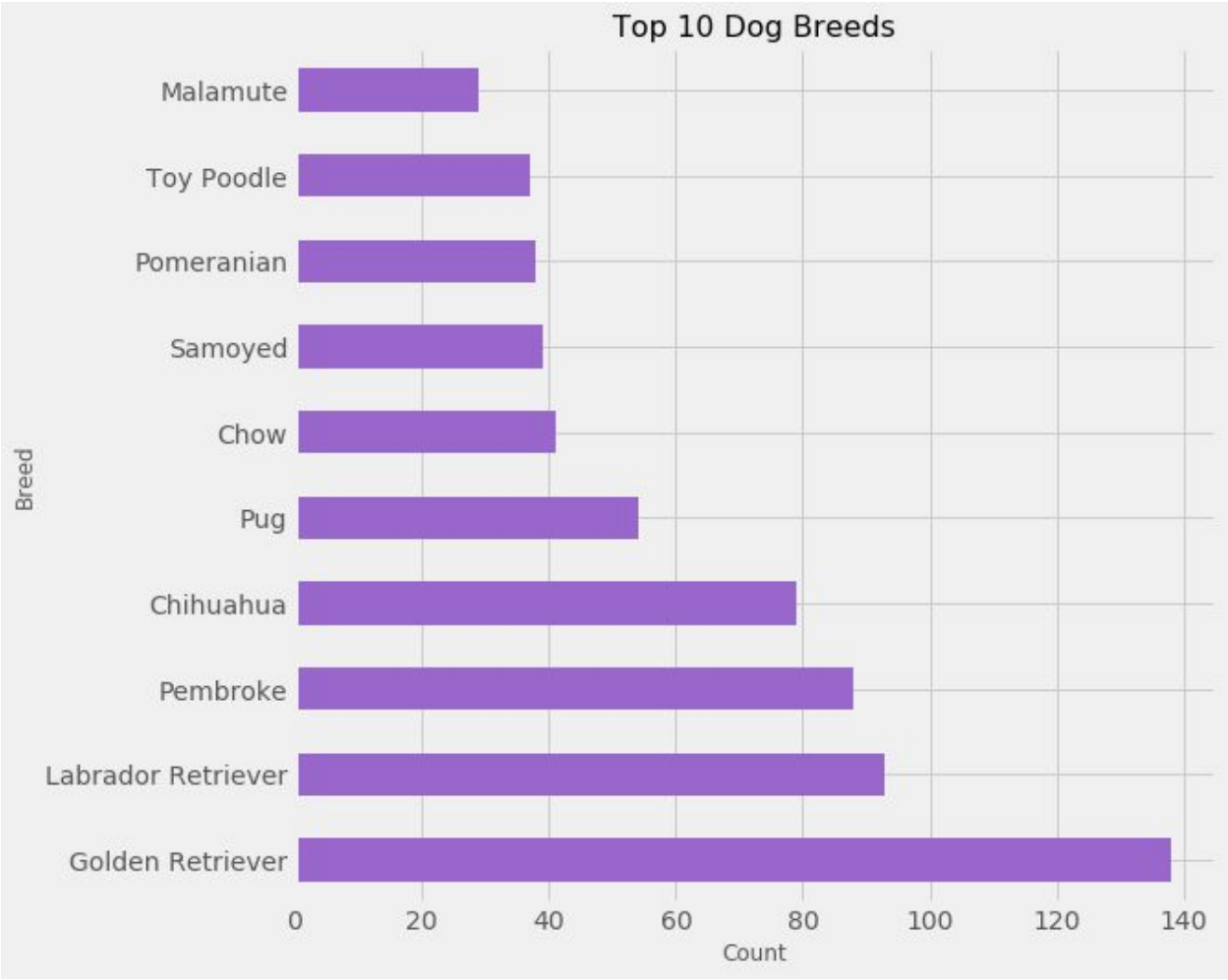


Code:

```
sns.regplot(x='retweet_count', y='favorite_count', color='#3399ff',
data=new_twitter_archive, scatter_kws={'alpha':0.2})
plt.title('Retweets versus Favorites', size=16)
plt.xlabel('Retweets', size=12)
plt.ylabel('Favorites', size=12)
plt.savefig('retweets-versus-favorites.png')
```

What are the most popular dog breeds?

Of the ten dog breeds, the most popular is the Golden Retriever. There are 138 instances of Golden Retriever, as shown in the figure:



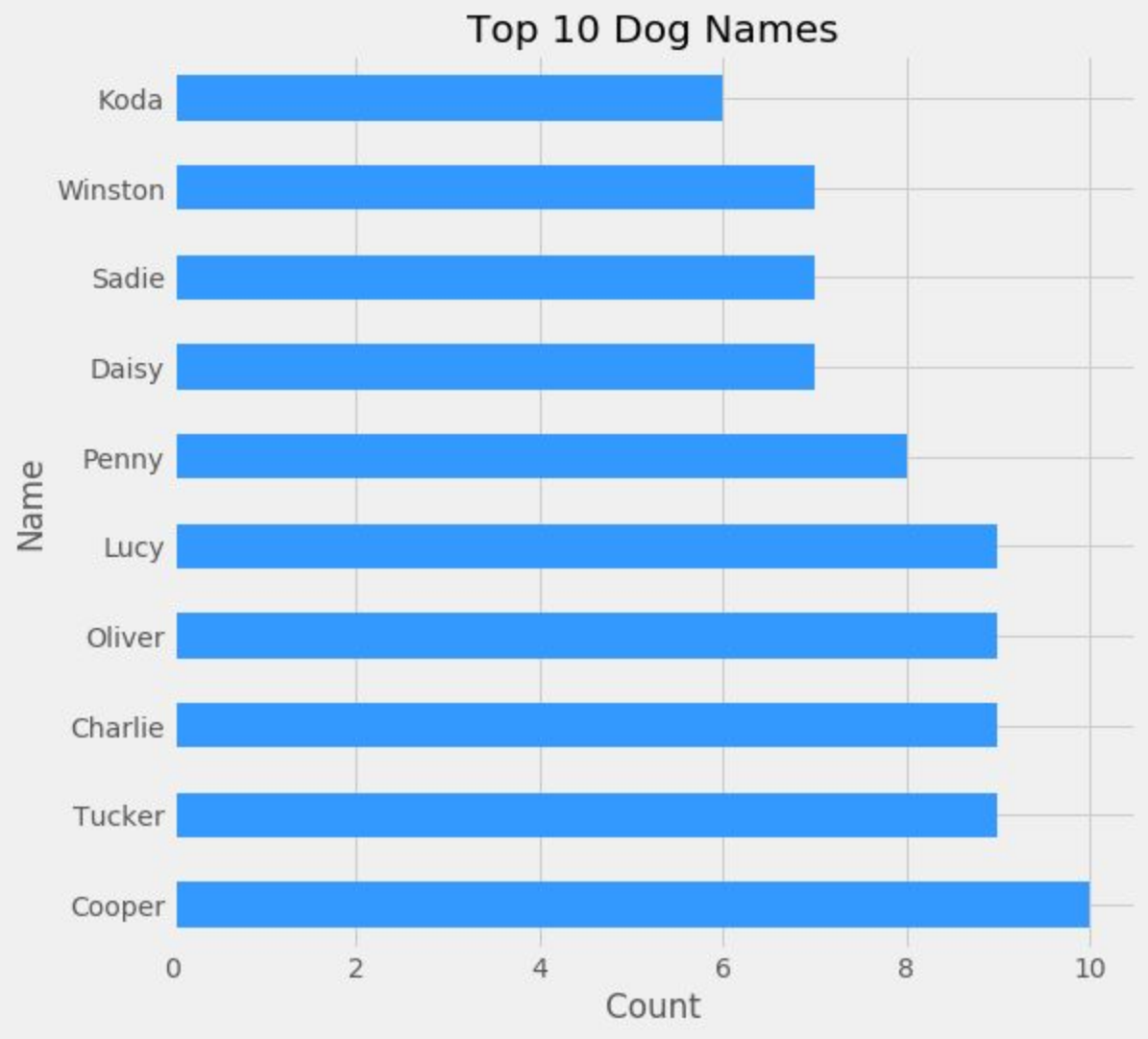
Code:

```
# Displaying the Top 10 Dog Breeds
top_breeds =
new_twitter_archive.dog_prediction_1.value_counts()[0:10].sort_values(
axis=0, ascending=False)
top_breeds.plot(kind = 'barh', color=['#9966CC'])
plt.title('Top 10 Dog Breeds', size=16)
plt.xlabel('Count', size=12)
plt.ylabel('Breed', size=12)
plt.savefig('top-breeds.png')

new_twitter_archive['dog_prediction_1'].value_counts()
```

What are the most popular dog names?

The most popular name according to the figure below is Cooper

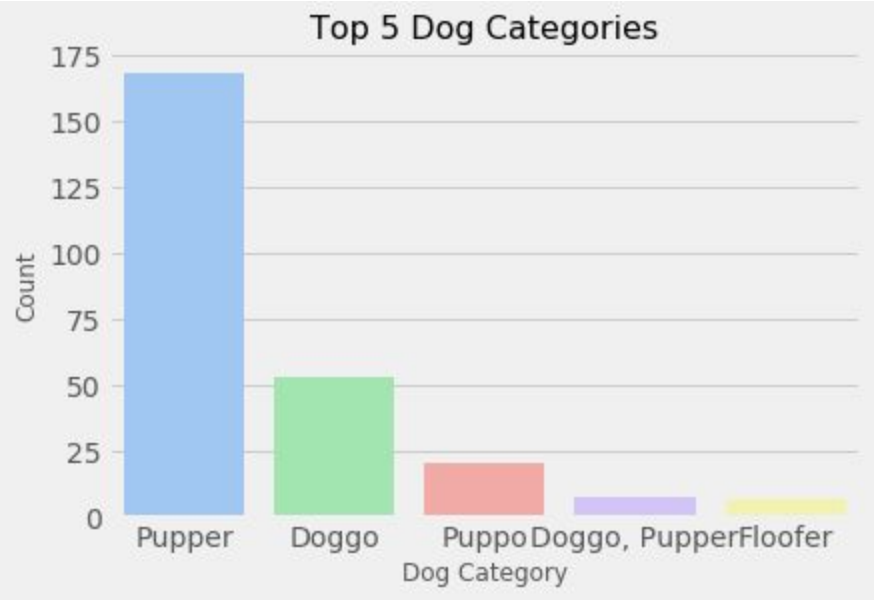


Code:

```
# Displaying the Top 10 Dog Names
top_names =
new_twitter_archive.name.value_counts()[1:11].sort_values(axis=0,
ascending=False)
top_names.plot(kind = 'barh', color='#3399ff')
plt.title('Top 10 Dog Names')
plt.xlabel('Count')
plt.ylabel('Name')
plt.savefig('top-names.png')
```

What are the most popular dog categories?

According to the figure, Pupper is the most popular dog category with 168 instances:



Code:

```
# Displaying the Top 5 Dog Categories
sorted_stage =
new_twitter_archive['dog_category'].value_counts()[1:6].index
sns.countplot(data = new_twitter_archive, x = 'dog_category', order =
sorted_stage, orient = 'h')
plt.xlabel('Dog Category', fontsize=12)
plt.ylabel('Count', fontsize=12)
plt.title('Top 5 Dog Categories',fontsize=16)
plt.savefig('top-categories.png')

new_twitter_archive['dog_category'].value_counts()
```