

ANALYSIS - REPORT

1) Most common source to tweet for rating of dogs ?

A histogram was plotted between the count and the source from which they are tweeting.

Value_counts() for data in source column indicates the following insights:

```
iPhone          1259
Web Client       16
TweetDeck        9
Name: source, dtype: int64
```

These insights reveal's that about 98.05 percent of tweets are tweeted using iPhone as source and then follows 1.24 percent from Web Client and 0.7 percent tweets from TweetDeck.

We can say from these insights that they prefer iPhone as source for rating dogs.

2) Most common rating among all dogs?

A histogram was plotted between the count and rating of dogs. These insights reveal that nearly 297 dogs have most common rating of 1.0 and then follows nearly 286 dogs having rating of 1.2.

These insights also revealed the outliers in the rating data. The outliers present in these data are at 177.6 and 42.0

3) Most common dog?

A histogram was plotted between the count and the dog having value_counts greater than or equal to 20.

The frequency of dogs ranges from 1 to 94.

From the visualization, we came to know that most common dog is golden_retriever and then follows pembroke.

4) Favorite_count vs Retweet_count?

To get the clear picture about the relation between favorite_count and retweet_count, we plotted a scatter plot between favorite_count and retweet_count.

From the visualization, we identified that there is an positive linear relationship between these two columns of data in the dataset.

5) Most popular dog based on favorite_count?

To get the most popular dog based on favorite_count, we subset ten rows of data having highest favorite_count using pandas nlargest() method, then plotted a bar plot between these 10 dogs.

From the visualizations the most popular dog based on favorite_count was chihuahua, with favorite_count 127321.

6) Least popular dog based on favorite_count?

To get the least popular dog based on favorite_count, we subset ten rows of data having lowest favorite_count using pandas nsmallest() method, then plotted a bar plot between these 10 dogs.

From the visualizations the least popular dog based on favorite_count was english setter, with favorite_count 80.

7) Most popular dog based on retweet_count?

To get the most popular dog based on retweet_count, we subset ten rows of data having highest favorite_count using pandas nlargest() method, then plotted a bar plot between these 10 dogs.

From the visualizations the most popular dog based on retweet_count was chihuahua, with retweet_count 61115.

8) Least popular dog based on retweet_count?

To get the least popular dog based on retweet_count, we subset ten rows of data having lowest favorite_count using pandas nsmallest() method, then plotted a bar plot between these 10 dogs.

From the visualizations the least popular dog based on retweet_count was english setter, with retweet_count 12.

- As expected, the most popular dog based on favorite_count and the most popular dog based on retweet_count are same (chihuahua).
- And even the least popular dog based on favorite_count and the least popular dog based on retweet_count are same (english setter).
- Because there is a positive linear relationship between favorite_count and retweet_count.

9) Retweets and favorites over time?

We plotted a graph between retweet_count and favorite_count over timestamp to dig deeper about the relationship between these over time.

From the visualization, we came to know that the favorite_count is greater than retweet_count over timestamp.

