VISUALIZATION PROCESS

Visualization 1 - “Which dog stage "Doggo", "Pupper", "Puppo" and "Fluffer have the highest favorites and retweets on average?”.

I used the the groupby function to get the average of the favorite count and retweet count:

df_clean.groupby(['stage']).mean()[['favorite_count', 'retweet_count']]

Then I use the stacked bar plot for visualization:

fav_retweet.plot.bar(stacked=True)

The stage "puppo" ranks the highest average favorite count with 18225. In contrast, "pupper" receives the lowest favorite count with only 6750 on average. Tweets contain dog stage "doggo" that has the highest average retweet count with 7295.On contrary, tweets with no dog stage at all ranked lowest retweet count at 2948.

Visualization 2 - “What are the top 10 most popular dog breed make appearance on WeRateDog's Twitter Account?”

I created a new dataframe from df_clean called df_breed contains 'tweet_id', 'p1', 'p1_dog', 'p2', 'p2_dog', 'p3', 'p3_dog', and 'dog_breed' columns.

I then used DataFrame.loc( ) function to input value for the column ‘dog_breed’.

My logic was: If p1 == True -> p1 == dog_breed, elif p2 == True -> p2 == dog_breed, elif p3 == True -> p3 == dog_breed, else Null == dog_breed

After getting the all the dog breeds, I dropped all rows that couldn’t define a breed, and used the breed.head(11) to find the top ten most popular breeds got featured. I used head(11) instead of (10) because there were two breeds both came int the tenth place. Finally I used breed.plot.bar( ) to show the visualization.

My finding is the top ten most popular dog breeds that make appearance on the WeRateDog Twitter account are: Golden Retriever, Labrador Retriever, Pembroke, Chihuahua, Pug, Toy Poodle, Chow, Samoyed, Pomeranian, Malamute and Cocker Spaniel respectively.