VISUALIZATION PROCESS

Visualization 1 - “Does using a nickname like "Doggo", "Pupper", "Puppo" and "Fluffer" have an impact on favorites or retweets?”.

To do this, I first created a new data frame called df_nickname with 7 columns 'tweet_id', 'favorite_count', 'retweet_count', 'doggo', 'floofer', 'pupper', 'puppo' from df_clean. Second, I changed all the values contain nickname to “1” and None to “0”. Then, I added a ‘no_nickname’ column using df_nickname.loc( ) function so if there is a nickname in the row, the output value is ‘0’, and if there is no nickname, the output is ‘1’. This is where my df_nickname data frame is completed.

From the df_nickname, I then extract the info into 5 small data frames for each nickname and no nickname by dropping other columns and that rows where values are different than ‘0’. For example, for the excerpt dataframe ‘doggo’ for the nickname “doggo”, I only keep the “doggo”, “favorite_count”, “retweet_count” columns, and where rows ‘doggo[doggo.doggo != '0']’. I then use the DataFrame.describe( ) to see get the mean of “favaorite_count” and “retweet_count”.

After finishing excerpting the 5 small data frames, I combine these 5 into a new data frame called fav_reweet with the index of the nicknames ‘Doggo’, ‘Floofer’, ‘Pupper’, ‘Puppo’, and ‘No nickname’, and two columns ‘Favorite mean’ and ‘Retweet mean’. I then finish the visualization by using fav_retweet.plot.bar(…, stacked = True) to show the average favorites and retweets of tweets that contain one of these nickname or no nick name at all.

My finding is tweets contain dog nicknames "puppo" ranked the highest average favorite count with 18225. In contrast, tweets with dog nickanme "pupper" in them receivw the lowest favorite count with only 6750 on average. So go for "puppo" not "pupper". Tweets contain dog nicknames "doggo" that has the highest average retwwet count with 7295. On contraty, tweets with no dog nicknames at all ranked lowest retweet count at 2948. So use a dog slang for more retweet.

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.