Report: act_report

The dataset that I worked on is the tweet archive of Twitter user **@dog_rates**, also known as **WeRateDogs**. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. After wrangling the data,and did my analysis, I got some insights from the wrangled data. I discovered the following:

1. after isolating the texts from the tweets ,i removed stop words and also removed some unnecessary words like:

"HTTPS", "https", "h t t p s","t","co".

then I imported an image mask

```
#Importing image mask
image = Image.open("dog.jpg")
plt.axis('off')
plt.imshow(image);
```

which I used in creating the wordcloud below

Using the code below

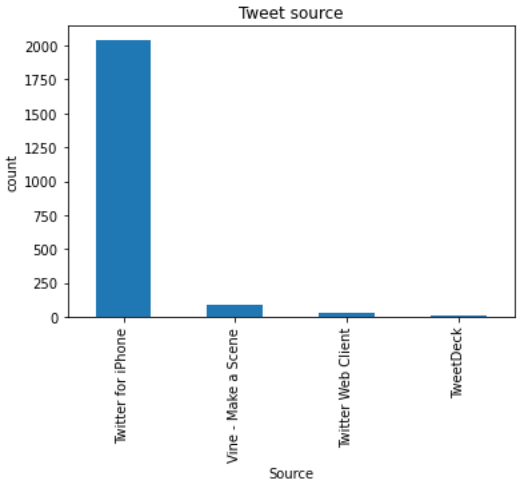
```
dog = np.array(image)
wordcloud = WordCloud(stopwords=STOPWORDS,mask=dog, background_color="white").generate(text)

# Display the generated image:
# the matplotlib way:
image_colors = ImageColorGenerator(dog)
plt.figure(figsize=[15,15])
plt.imshow(wordcloud.recolor(color_func=image_colors), interpolation="bilinear")
plt.axis("off")
plt.show()
```

After creating the wordcloud , I could see that the most used words in the tweets were pupper ,meet, dog, pop, and dog_rates

2.I also discovered that most of the tweets were gotten from twitter for i-phone while other sources were from Vine,Twitter web client,and TweetDeck respectively

```
twit_main.source.value_counts().plot(kind="bar")
plt.title('Tweet source')
plt.xlabel('Source')
plt.ylabel('count');
```



3. I also discovered that id with number 744234799360020481 has the most retweets with 79515 retweets with a favourite count of 131075 with a rating numerator of 13

4.Also it was notice that the ID with number 822872901745569793 had the most favorite count with a count of 132810

5.After plotting ratings with retweet count, it was discovered that tweets with higher ratings had more retweets as opposed to tweets with lower ratings which had lower retweets.

```
#creation of Bar charts
locations = [1, 2]
retweet_count = [low_ratings,high_ratings]
labels = ['Low', 'High']
plt.bar(locations, retweet_count, tick_label=labels,color=['black','grey'])
plt.title('Ratings by retweet count')
plt.xlabel('Ratings')
plt.ylabel('retweet count');
```