Data Insights and Visualization

On conclusion of the data wrangling process, a describe method was used to summarize the database with the summarized outcome below.

	tweet_id	rating numerator	rating denominator	favourite count	retweet count	img_num	p1_conf	p2_conf	p3_conf
count	1.99E+03	1994	1994	1994	1994	1994	1994	1.99E+03	1.99E+03
mean	7.36E+17	12.215035	10.51003	8895.7257	2766.753	1.203109	0.593941	1.34E-01	6.02E-02
std	6.75E+16	41.463367	7.261522	12213.193	4674.698	0.560777	0.271954	1.01E-01	5.09E-02
min	6.66E+17	0	7	81	16	1	0.044333	1.01E-08	1.74E-10
25%	6.76E+17	10	10	1982	624.75	1	0.362857	5.39E-02	1.62E-02
50%	7.08E+17	11	10	4136	1359.5	1	0.587635	1.17E-01	4.95E-02
75%	7.88E+17	12	10	11308	3220	1	0.846285	1.95E-01	9.16E-02
max	8.92E+17	1776	170	132810	79515	4	1	4.88E-01	2.73E-01

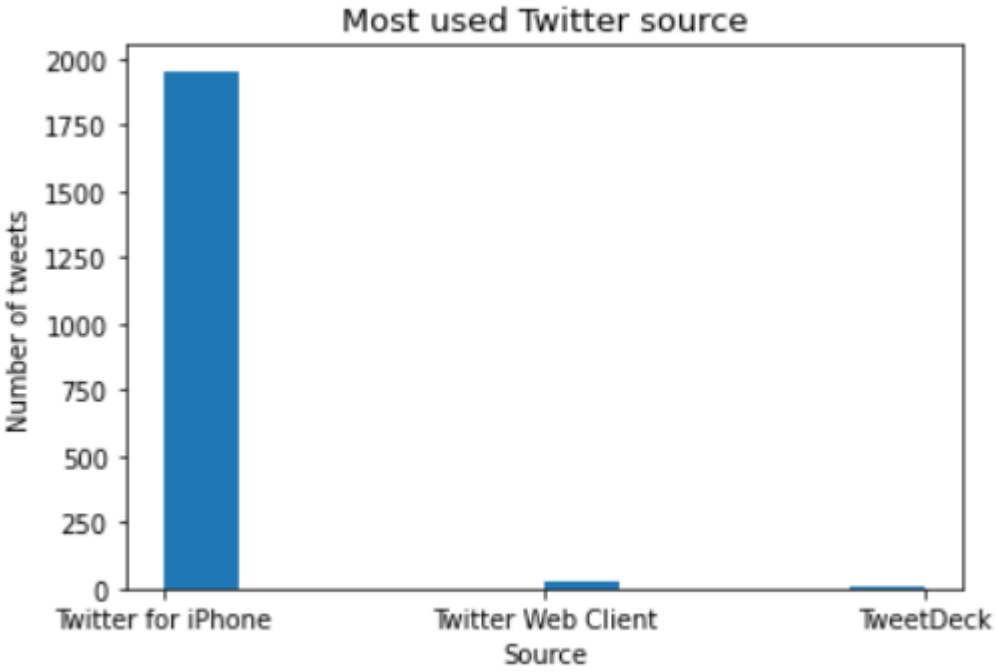
The following conclusions was deduced:

1. The most liked tweet has 132810 likes
2. The most retweeted tweet has 79515 retweets

The ratings numerator was divided by the denominator to obtain the following results. Dogs ratings equivalent to 1.2 is the most common with 454 dogs rated 1.2 can be deduced from the result

S/N	rating	counts
1	1.2	454
2	1	413
3	1.1	403
4	1.3	263
5	0.9	151
6	0.8	98
7	0.7	53
8	1.4	36
9	0.6	33
10	0.5	31
11	0.3	19
12	0.4	16
13	0.2	10
14	0.1	5
15	0	2
16	177.6	1
17	1.126	1
18	1.127	1
19	0.975	1
20	3.428571	1
21	1.35	1
22	42	1

Finally, a visualization to show which source was used for the most tweet in the dataset was made with the summary below. Twitter for iPhone was the predominant tweet source


Most used Twitter source
2000
1750
1500
1250
1000
750
500
250
0
Number of tweets
Twitter for iPhone
Twitter Web Client
TweetDeck
Source