Data Analysis and Visualization

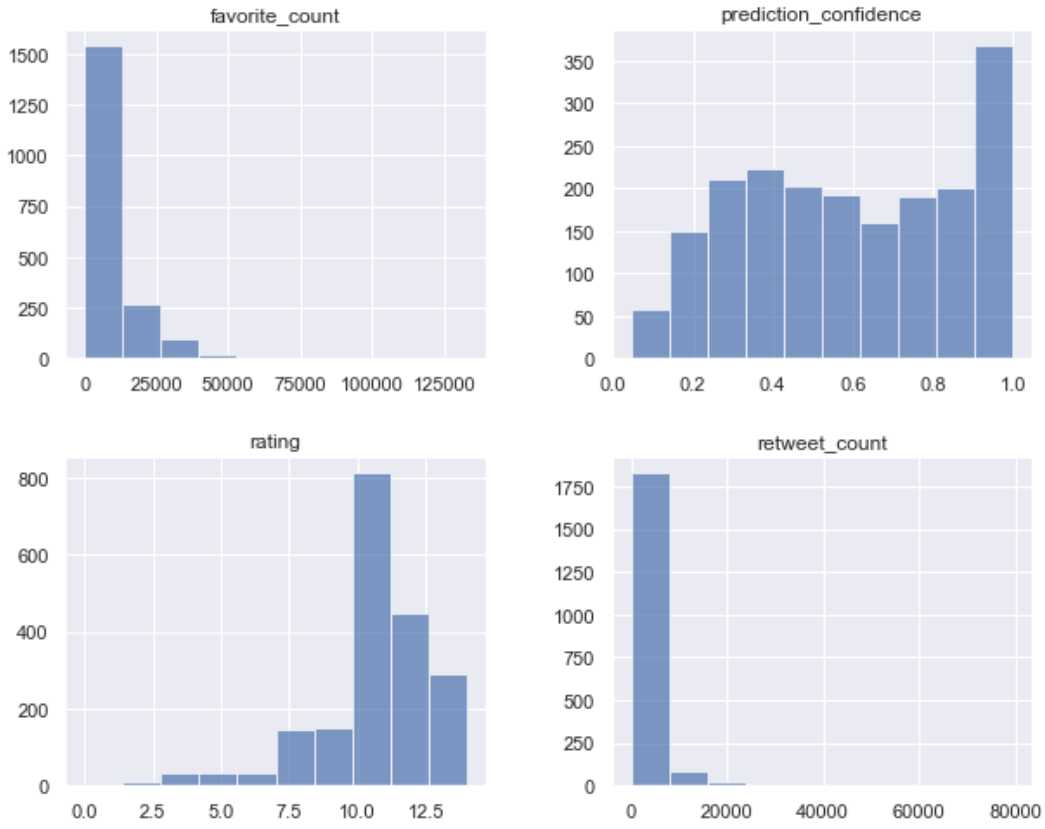
Summary

After finalizing the data wrangling step, here it comes the step to investigate the dataset in a different way to track the patterns and come up with the most representative insights.

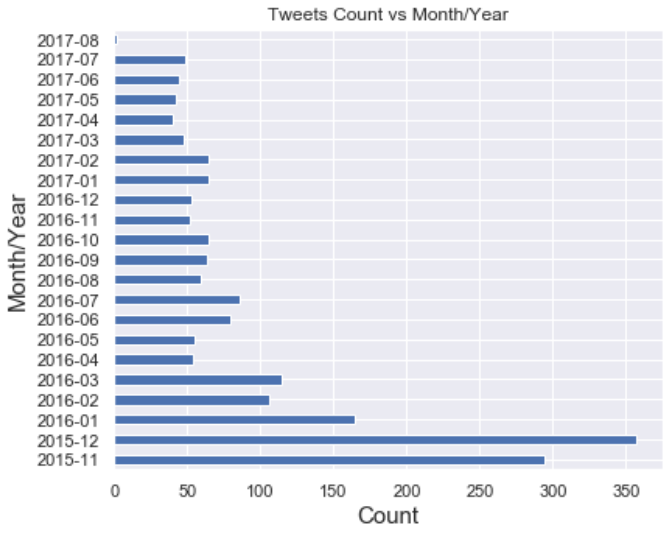
Data Assessment and Visualization

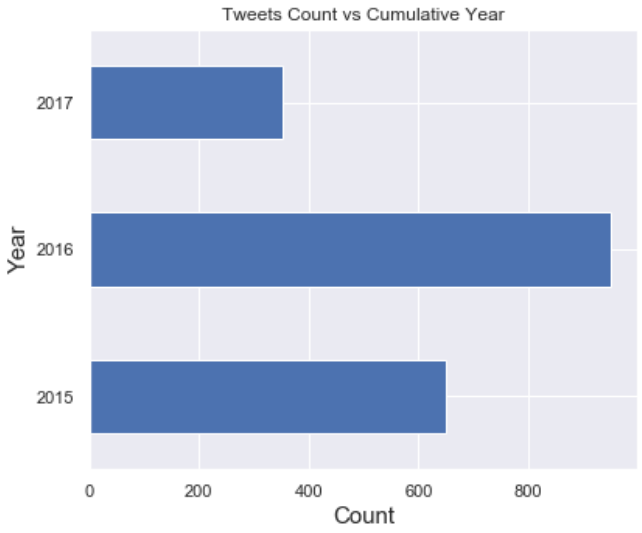
Here's a summary of the dataset statistics and histogram distribution:

	rating	retweet_count	favorite_count	prediction_confidence
count	1,956.00	1,956.00	1,956.00	1,956.00
mean	10.54	2,792.57	8,981.93	0.59
std	2.17	4,713.76	12,306.54	0.27
min	0.00	16.00	81.00	0.04
25%	10.00	623.75	1,967.50	0.36
50%	11.00	1,366.50	4,148.00	0.59
75%	12.00	3,245.25	11,455.75	0.85
max	14.00	79,515.00	132,810.00	1.00

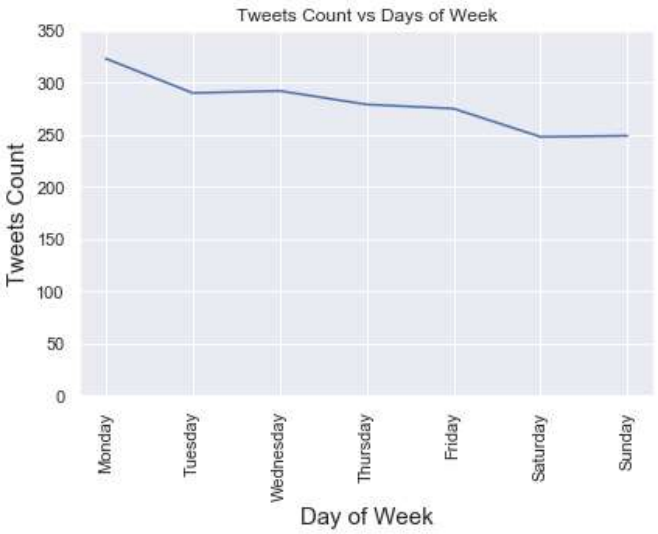


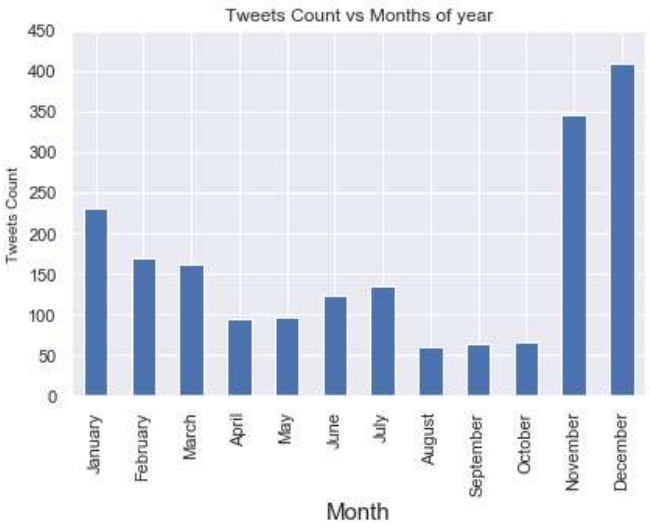
Then, having a look at the dataset on a chronological wider scale, it seems that tweets witnessed the largest number at the end of 2015, plunged by the beginning of 2016, and kept fluctuating after that In addition to total tweets count of 2016 was higher compared to 2016 and 2017 with taking into consideration the very low and low representation of 2015 and 2017.



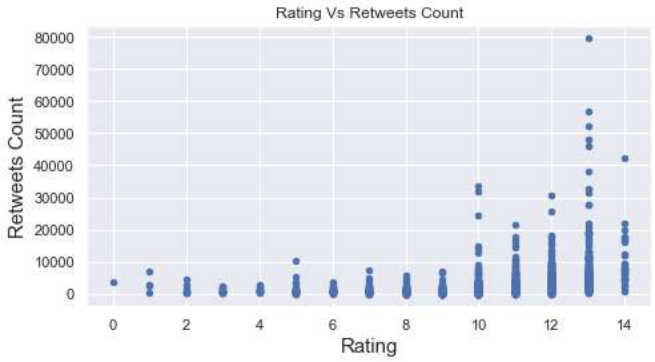


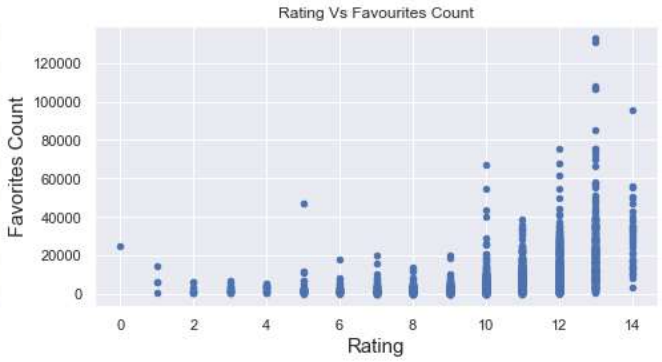
Moving to a narrower scale and how tweet posts increase on Monday and decrease gradually over the week. While months like December and November showcase the largest number of tweets.



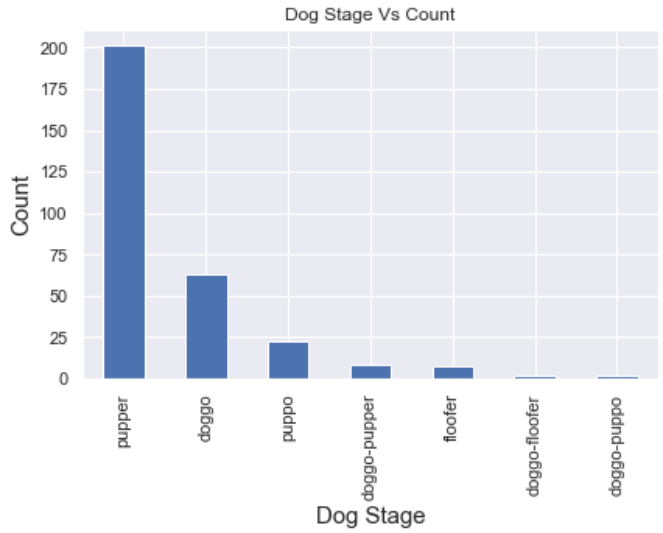


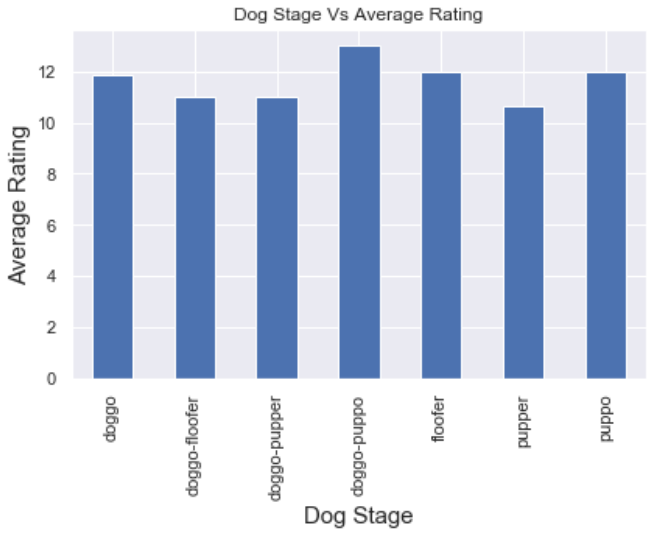
Then, diving deeper into the dataset to look for more patterns where the higher ratings take a higher number of retweets and favorites.



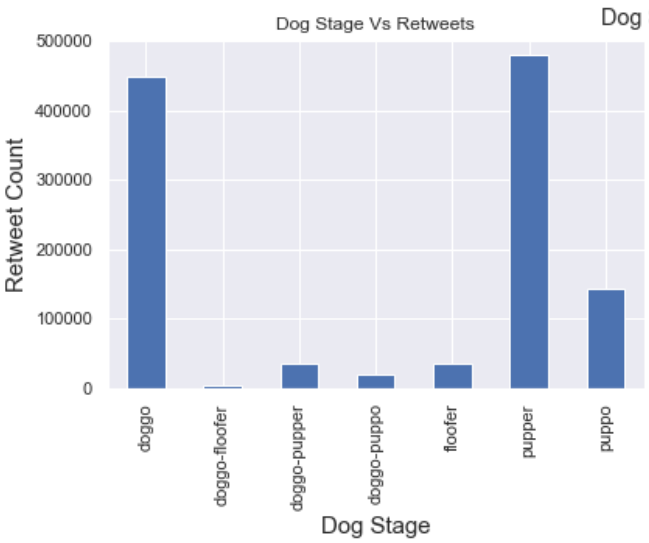


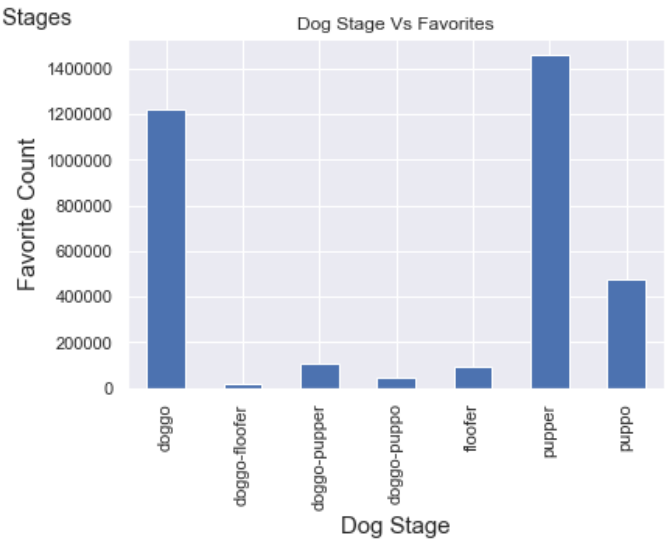
Furthermore, the Pupper dog stage has the highest presence by far in the dataset, then Doggo dog stage. However, their average rating is nearly similar.



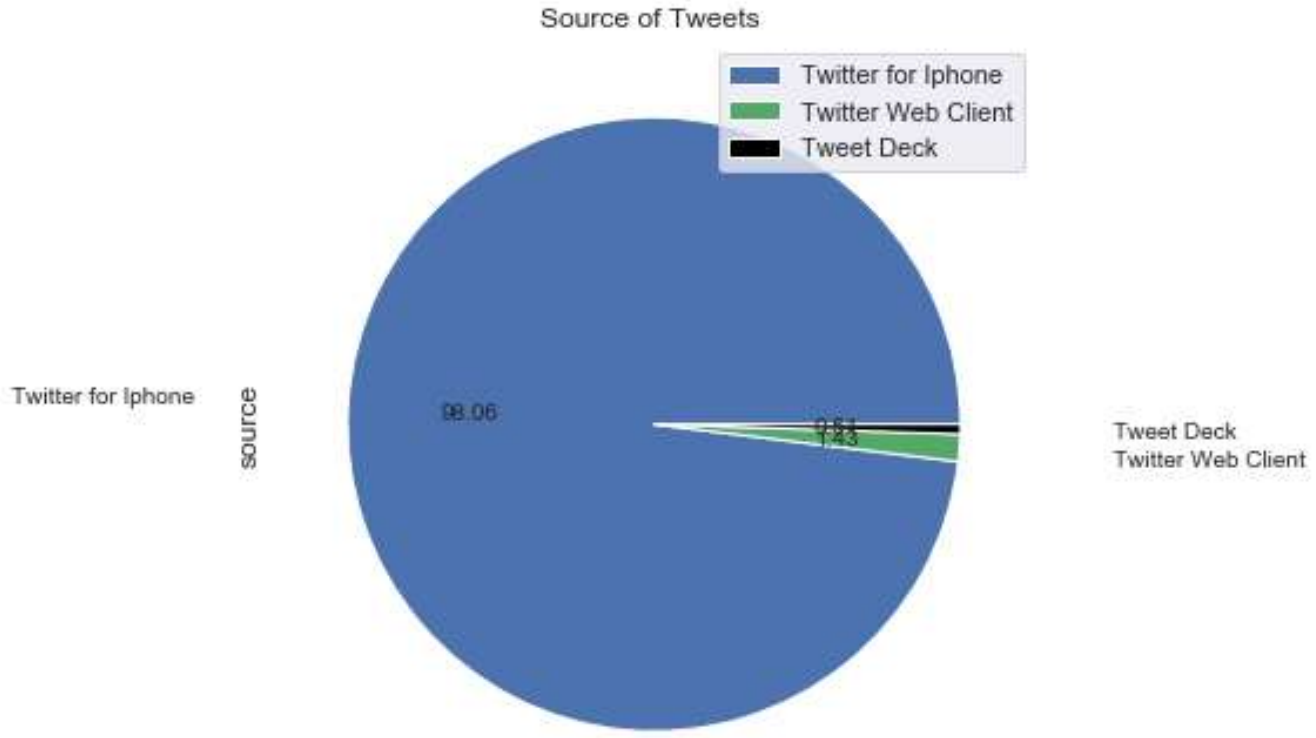


And the Pupper dog stage also has the highest number of retweets and favorites and it could be due to its high presence. Although, doggo dog stage has moderate tweets 'presence, it has a remarkable number of retweets and favorites.





Having a look at another part of data “source of tweets” and through which they are posted. The pie chart demonstrates that 98% were posted through iPhone.



Over and above, the bar charts below show the top 10 predicted dog breeds like golden retriever and how their confidence level of prediction is noteworthy which exceeds 60%.