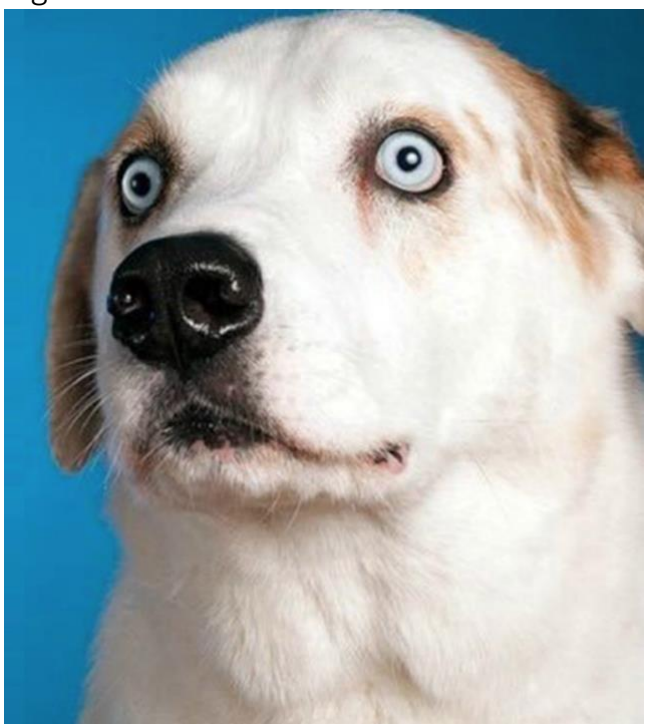
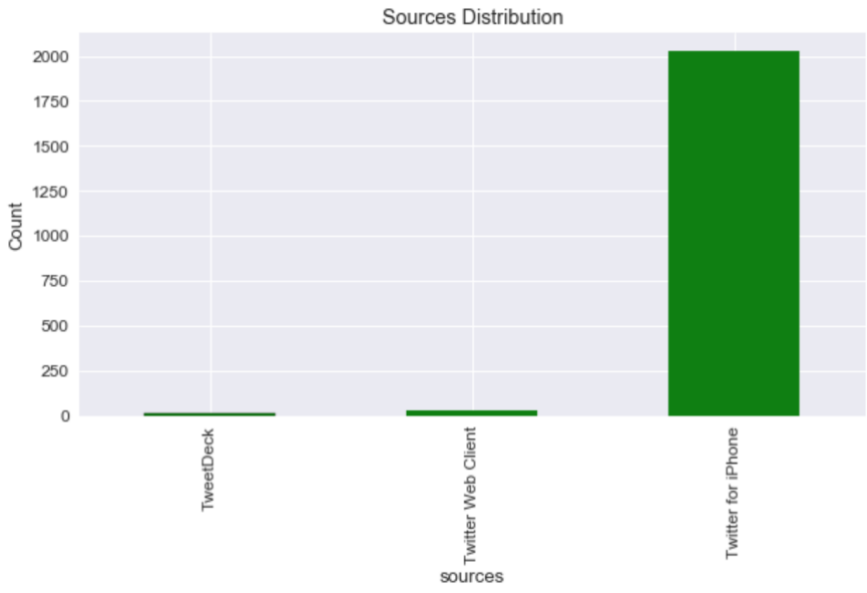
Wrangle and Analyze Data: WeRateDogs™(@dog_rates)

Chen Qingqing

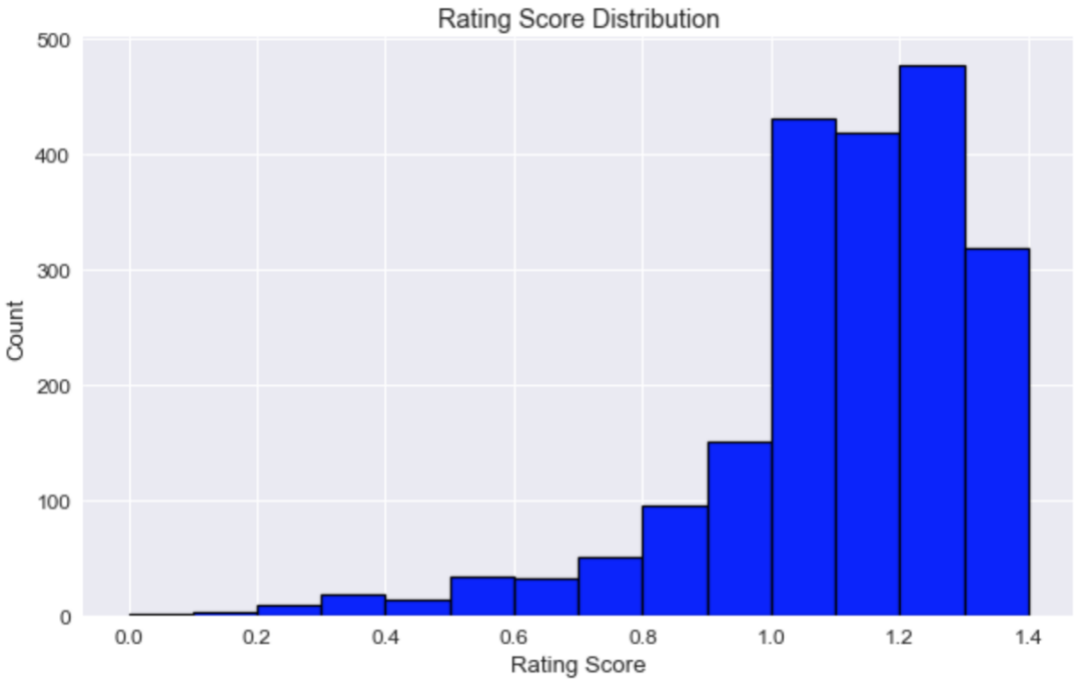
The dataset that I wrangled is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

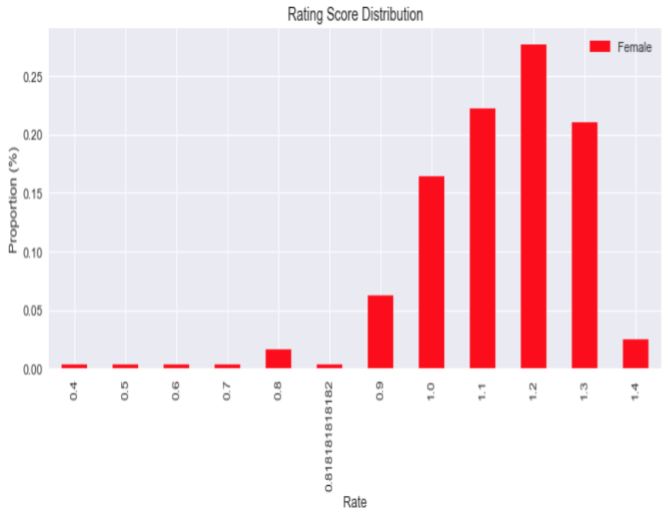


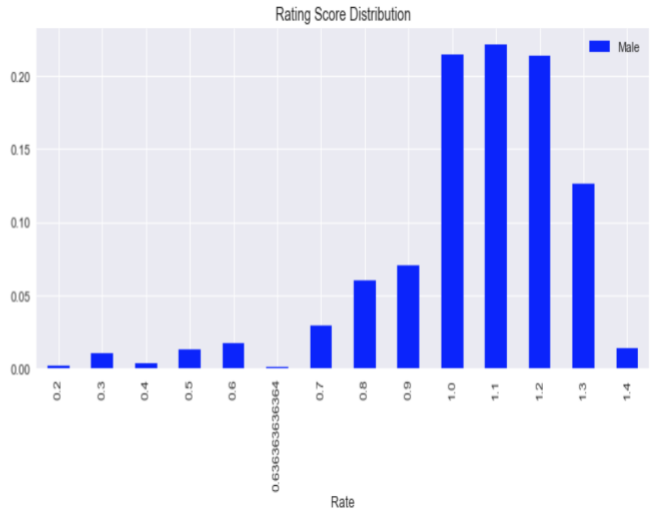
Firstly, I want to see the distribution of rating source and I founded that mostly people (around 98%) just use their Twitter through iPhone.



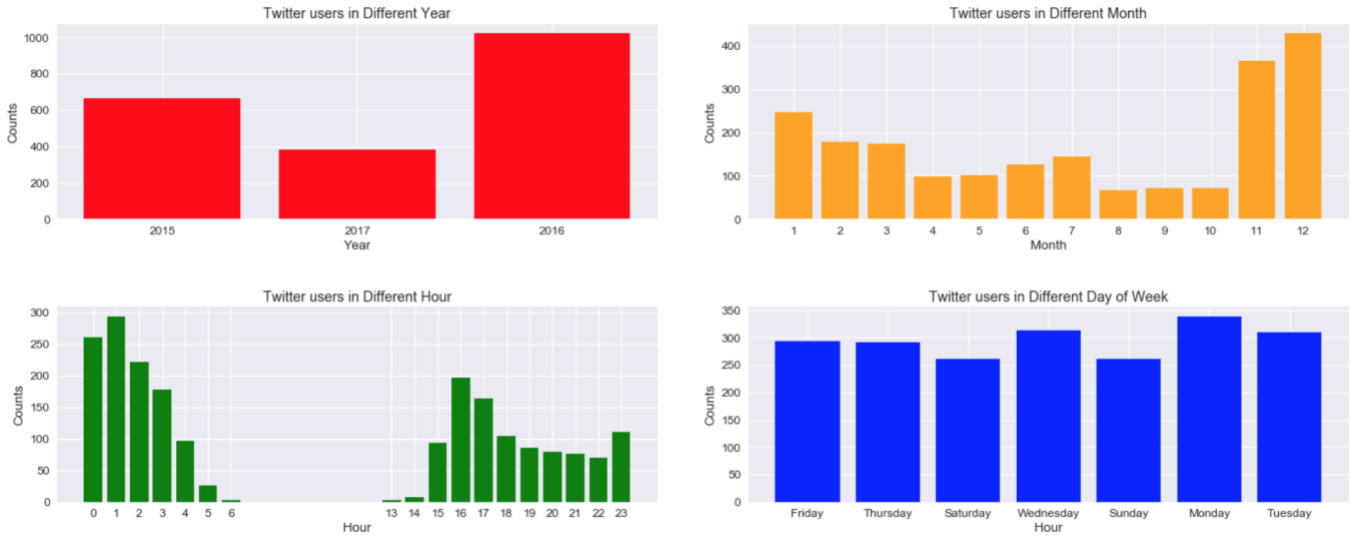
Then I explored the distribution of the rating score which is less than 2. And it shows a right skewed rating score distribution. In order to see whether the rating score distribution will changed by the dog's gender or not, I separated the dog rating score by gender, and the rating score still shows right skewed distribution in both gender, even though there is a small difference between female and male dogs.



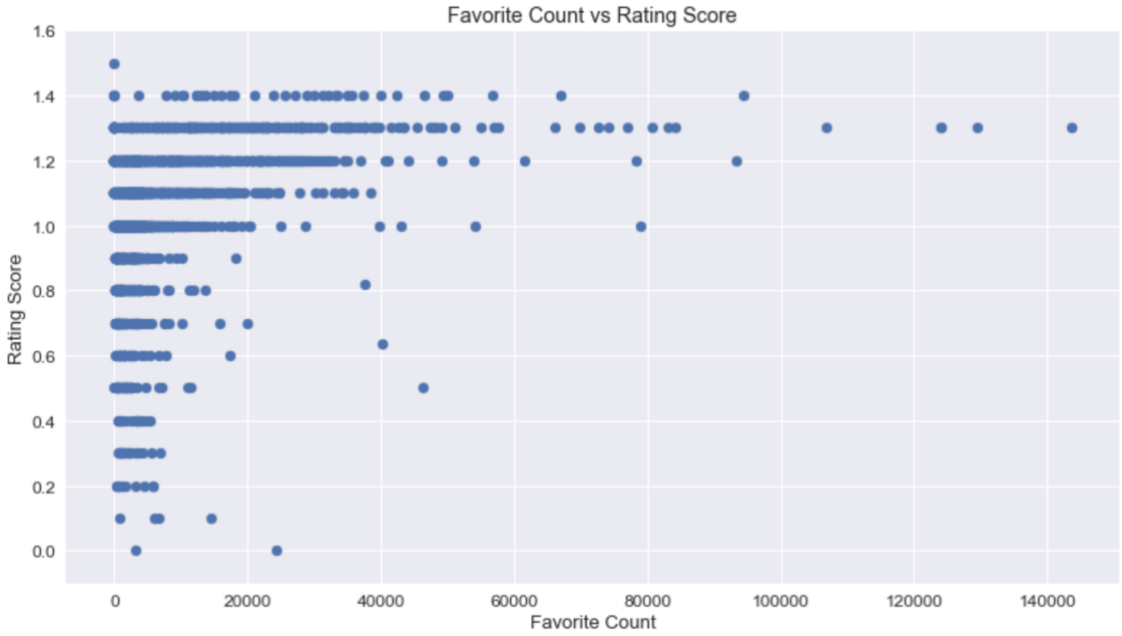




After this, I explored the difference among year, moth, hour and day of week for people using their twitter. The results show that 2016 has the highest users. And the highest number of twitter users appears in December while the lowest number of twitter users appears in August; The highest number of twitter users appears in 1 am and there are few users using twitter during 7 am to 12pm. There is no significant difference among different day of weeks. The number of twitter users on Monday is just slightly higher than the rest of other days of the week.



In addition, there is a relationship between rating score and favourite count where it shows that with the increasing of the rating score, the favourite count increases;



The last but not the least, the number of different dog stages, prediction algorithm is plotted, and the top three used algorithms is Golden retriever, Labrador retriever, and Pembroke.