

## Projeto II – WeRateDogs Tweeter

Por Leandro Almeida

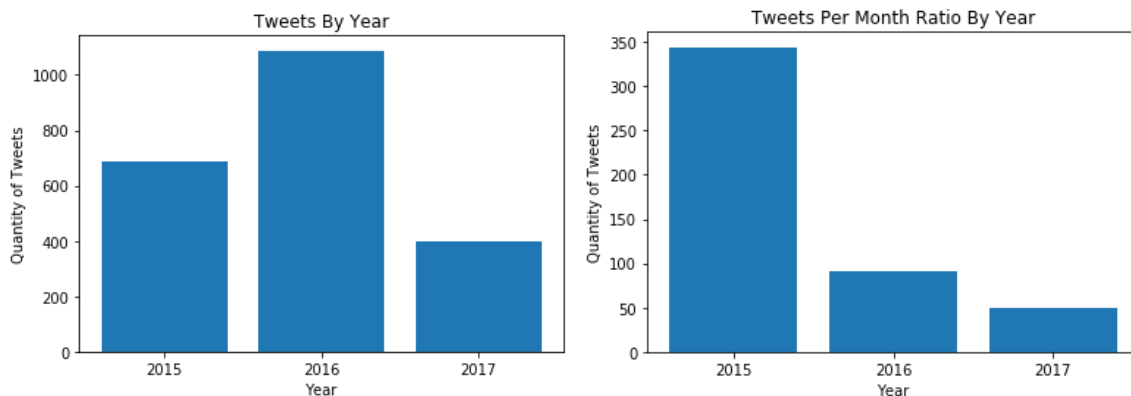
O objetivo deste relatório é comunicar as descobertas em relação aos tweets do usuário WeRateDogs. Os dados da conta passaram pelo processo de data wrangling, portanto estamos considerando os dados disponíveis já limpos e as etapas de coleta-avaliação-limpeza não serão documentadas nesse relatório. Com posse desses dados, serão reportadas algumas avaliações e visualizações para analisarmos algumas características dos tweets da conta.

O intuito neste momento não é chegar a conclusões com base nos dados, mas apenas gerar uma análise mais descritiva acerca dos dados que temos em mãos. Algumas perguntas podem ser levantadas nesse momento, mas ficaram sem resposta pois teríamos que trabalhar em modelos ou análises mais profundas que seriam da fase de exploração e não de data wrangling, que é a etapa do processo que esse relatório foi elaborado.

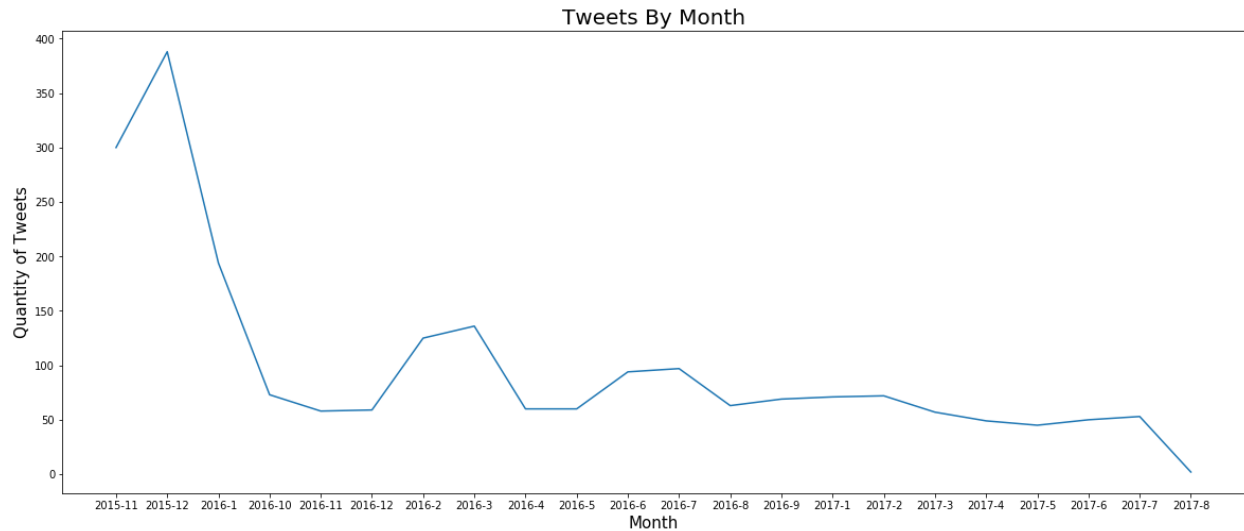
A conta WeRateDogs iniciou em 2015 com o objetivo de classificar e comentar, com muito humor sobre os cães que seus seguidores enviavam para avaliação. Atualmente, em 2019, a conta do twitter conta com mais de 7.69 milhões de seguidores. Os dados analisados foram as postagens da conta, sem considerar os retweets, com o status de 1 de agosto de 2017. No total, são 2175 mensagens analisadas.

- Freqüência de postagens

Como podemos ver no primeiro gráfico abaixo, em 2016 foi o ano em que houveram mais postagens. No entanto, essa não é exatamente uma análise justa já que apenas o ano de 2016 está sendo considerado os 12 meses. Para 2015 e 2017 temos menos meses nos dados sendo esperado que tenhamos menos postagens nesses anos. Se considerarmos algo como uma relação postagem/mês – em que o número de postagens é dividido pelo número de meses em que a conta estava ativa, vemos que o resultado é bem diferente, tendo 2015 como o ano com maior volume de postagem.



Esse comportamento, que 2015 foi o que teve o maior volume de postagem apesar de ter o menor período, pode ser comprovado se plotarmos o número de tweets por mês. Como podemos perceber no gráfico abaixo, os dois meses de 2015 tiveram, em média, 350 tweets por mês.



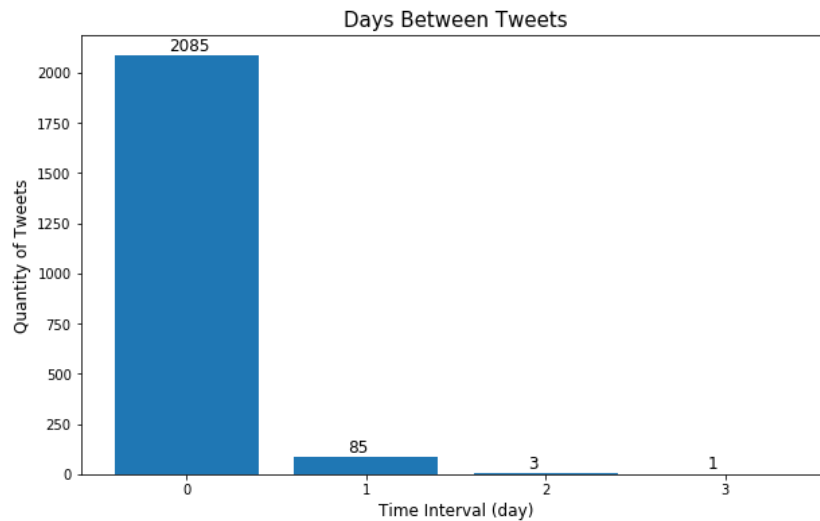
Uma análise interessante que pode ser feita com base na frequência é analisar se o número de postagem gera mais seguidores para a conta. Considerando que a postagem seria como a publicidade da conta, seria natural imaginar que quanto mais postagem houver, mais seguidores a conta terá.

- Fonte das postagens

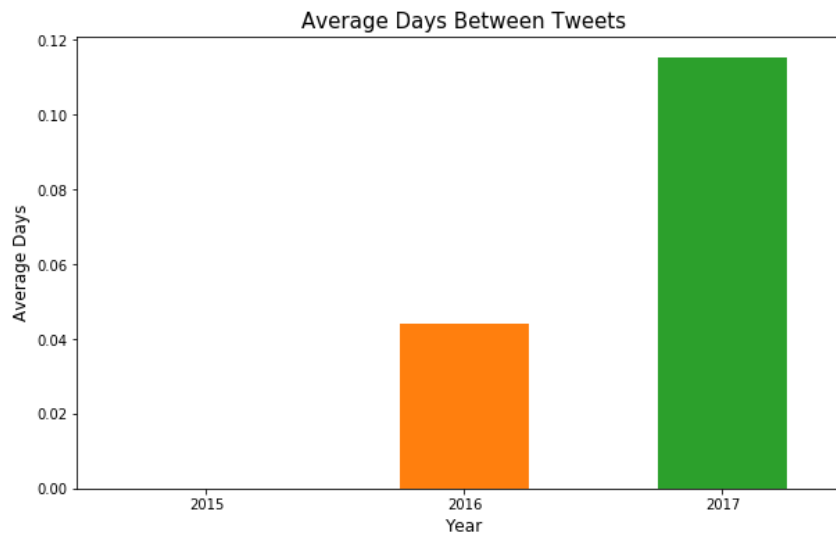
Outra característica sobre as postagens é que 94% delas é feita através do telefone móvel. É interessante constatar que uma conta no twitter com alcance de 7.7 milhões de pessoas, com destaque internacional, produtos relacionados a marca como jogo, livro e loja virtual seja mantida praticamente através do telefone celular e, possivelmente, em qualquer lugar que tenha conexão a internet. Algo bem diferente dos antigos modelos de negócio que sempre acreditavam que precisavam de um local fixo para trabalho, alto investimento em máquinas e publicidade para alcançar esse número de “clientes”.

- Periodicidade das postagens

Ainda em linha com a análise da fonte da postagem, a divulgação de conteúdo na conta tem uma latência muito baixa. Como podemos ver no gráfico abaixo, os dias entre uma postagem é praticamente zero para a maioria das postagens disponíveis para análise.



Podemos aprofundar um pouco mais e entender se as postagens com maior intervalo de dias têm algum viés com o passar dos anos. Desde modo, se plotarmos a média de tempo entre as postagens para cada ano – gráfico abaixo –, podemos perceber que a tendência é de crescimento, apesar da média sempre ser bem menor que um dia para todos os anos analisados.

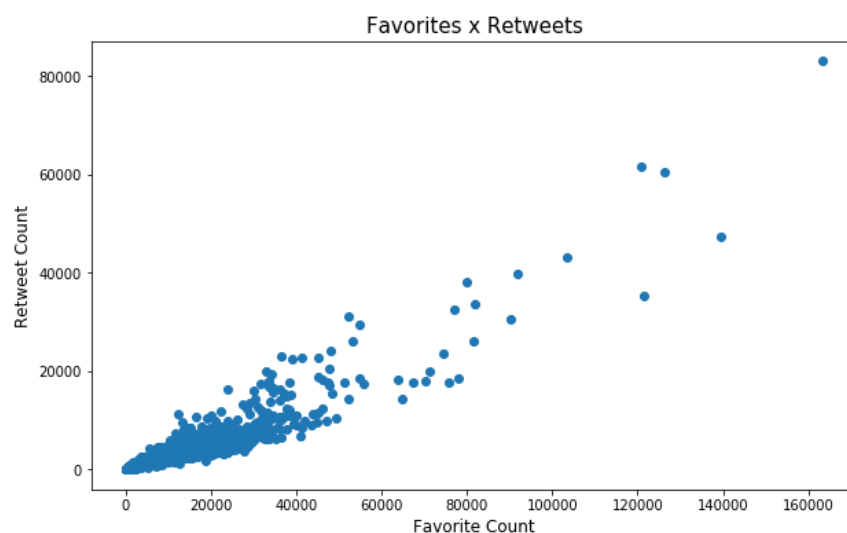


Uma outra análise que podemos explorar mais para entender é se o fato de utilizar o celular como principal fonte de postagem diminui a latência das postagens. Ou seja, sendo possível gerar conteúdo

para conta de uma fonte altamente móvel, podemos chegar ao equipamento perfeito para atualizar o conteúdo da conta. Extrapolando para a primeira pergunta, se chegarmos na primeira análise que postar mais vezes gera mais seguidores, seria o celular a melhor fonte de postagem ao invés de um escritório com pessoas dedicadas a isso?

- Qual a raça que gera mais interações entre os seguidores

Em relação ao conteúdo das postagens, vamos avaliar qual a raça de cachorro mais amada pelos seguidores do perfil We rate dogs. Temos duas formas de analisar: pela contagem de retweets do tweet ou pela contagem de marcações de favoritos do tweet. Portanto, primeiramente, vamos entender se existe uma grande diferença entre essas “métricas”. O objetivo é avaliar se precisamos aprofundar mais a análise delas para entender qual seria mais relevante para avaliar a raça de cachorro mais adorada.



Pelo gráfico acima, percebemos que existe uma boa correlação positiva – calculado sendo 0.92 – entre o número de retweets e contagem de marcações favoritas. Isso indica tendência de aumentar os retweets com o aumento da postagem de favoritos, embora os valores dessas duas métricas sejam bem diferentes. Desta forma, como guardam boa correlação, podemos considerar que as duas seriam boas métricas para avaliar a raça mais adorada pelos seguidores.

Nessa análise, vamos escolher a contagem de favoritos como métrica, entretanto, para ponderar que as raças não têm todas a mesma frequência nas postagens, o ranking da raça mais adorada será feito pela média do número de marcações como favoritas.

Na realidade, essa também não é exatamente a melhor forma de avaliação, pois as postagens com menor frequência têm mais chances de ter a média de marcações favoritas maiores. A forma mais correta seria ponderar a frequência e média para gerar o ranking – de forma semelhante ao que o site Rotten Tomatoes faz na sessão “Top Movies”. Entretanto, essa seria uma análise mais profunda e que seria mais adequada para a fase de exploração.

Como primeiro insight para a fase de Data Wrangling será apresentado um ranking com as 20 raças de cachorro com a maior média de marcações como favoritas. Contudo, como é uma informação relevante, será apresentado o ranking dessas raças com base na frequência de postagem.

breed	rank_by_avg_favorites	rank_by_post
Bedlington_terrier	1.0	69.0
Saluki	2.0	78.0
French_bulldog	3.0	11.0
Bouvier_des_Flandres	4.0	106.0
Afghan_hound	5.0	90.0
black-and-tan_coonhound	6.0	100.0
flat-coated_retriever	7.0	56.0
Irish_water_spaniel	8.0	90.0
Leonberg	9.0	90.0
whippet	10.0	42.0
English_springer	11.0	51.0
standard_poodle	12.0	42.0
cocker_spaniel	13.0	13.0
Samoyed	14.0	8.0
Mexican_hairless	15.0	61.0
Border_terrier	16.0	61.0
Eskimo_dog	17.0	15.0
Cardigan	18.0	16.0
Lakeland_terrier	19.0	25.0
golden_retriever	20.0	1.0

É interessante notar a raça Golden Retriever na posição 20 com base na média de marcações como favoritas, porém essa é a raça com o maior número de postagem. É possível que, ao considerar essas duas informações nessa análise, o Golden Retriever seja a raça mais amada entre os seguidores.