

WeRateDogs Act Report

Por: Anderson Carvalho,
em 21-05-2018.



O projeto foi baseado em dados fornecidos pela Udacity e cedidos pelo usuário Matt Nelson criador da conta do Twitter 'WeRateDogs'.

Os dados são postagens do Twitter (tweets), onde o 'WeRateDogs' pontua os cães classificações de 1 a 10/10, em que quase sempre atribui valores maiores que 10 ao numerador, 12, 13 ou 14/10. Isso acontece porque Matt Nelson os considera bons cães.

Alguns dados foram recebidos em arquivos, outros precisaram ser baixados. Dados adicionais foram coletados diretamente da base de dados do Twitter via uma API.

Com base nos dados disponíveis foram elaborados os seguintes questionamentos:

1. Quais os 5 cães que tiveram mais tweets compartilhados e suas classificações do WeRateDogs?
2. Qual a distribuição dos cães segundo os estágios propostos pelo WeRateDogs?
3. A raça dos cães parece ter alguma relação com o compartilhamento dos tweets?
4. Há muitos cães com o mesmo nome? Se sim, quais os 10 nomes mais comuns?
5. O estágio do cão sugere alguma relação com a classificação dada pelo WeRateDogs?

Foi feita uma avaliação dos dados para verificar as condições dos mesmos e perceber possíveis problemas de qualidade e arrumação.

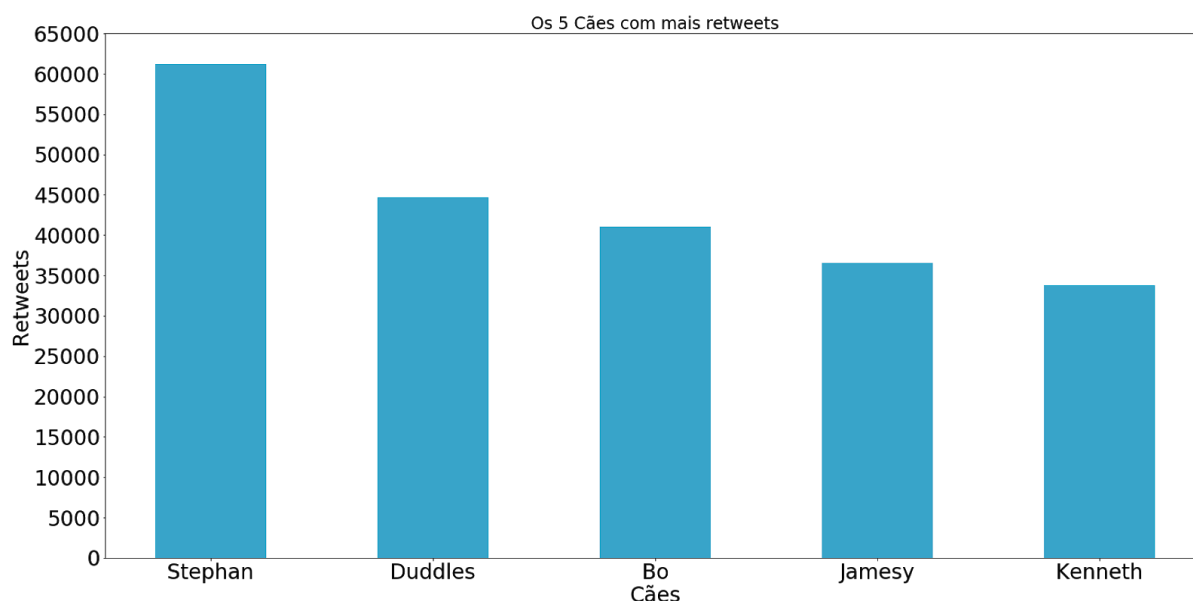
Após a avaliação dos dados e assinaladas as necessidades de limpeza, o processo foi realizado obtendo-se um conjunto de dados suficientemente limpo para análise e resposta aos questionamentos.

Questão 1

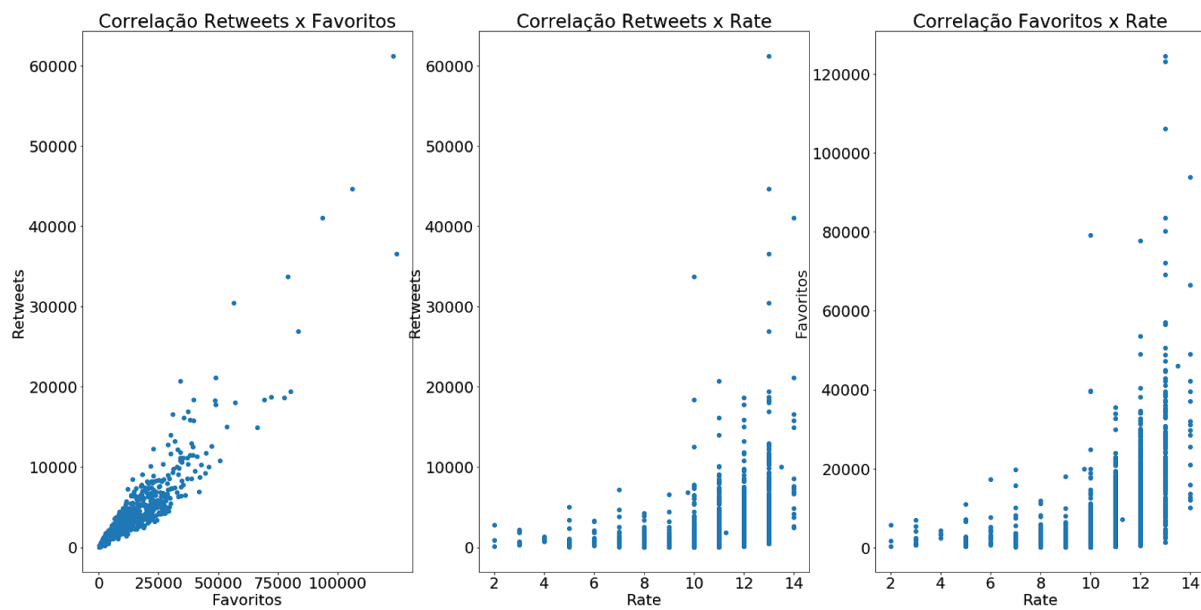
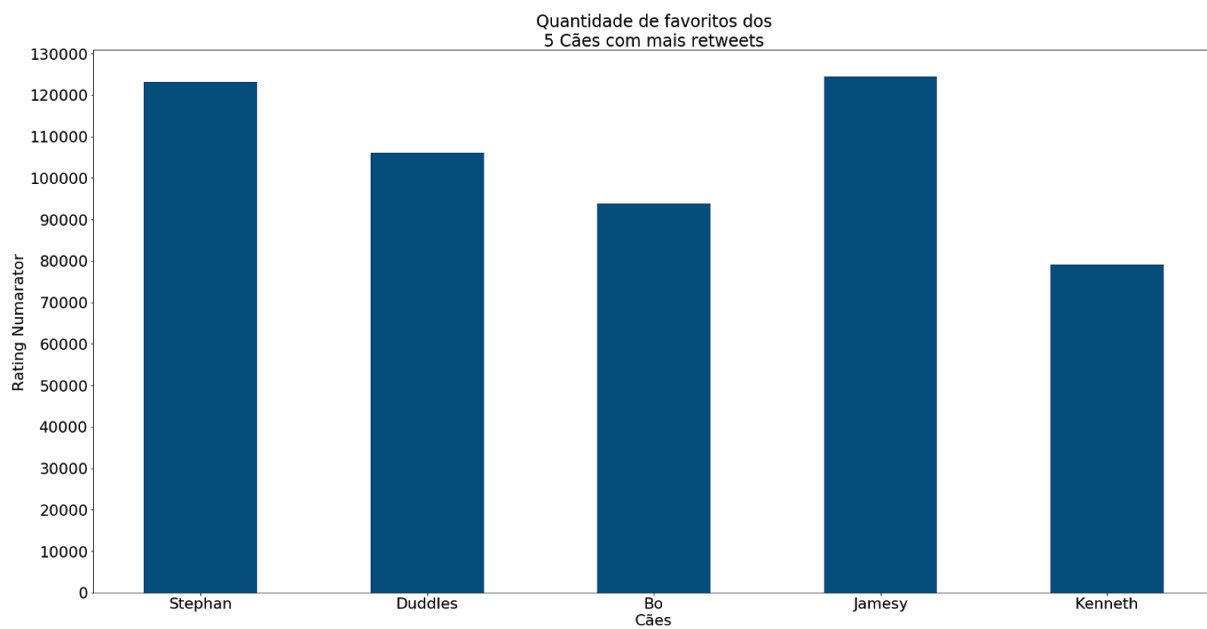
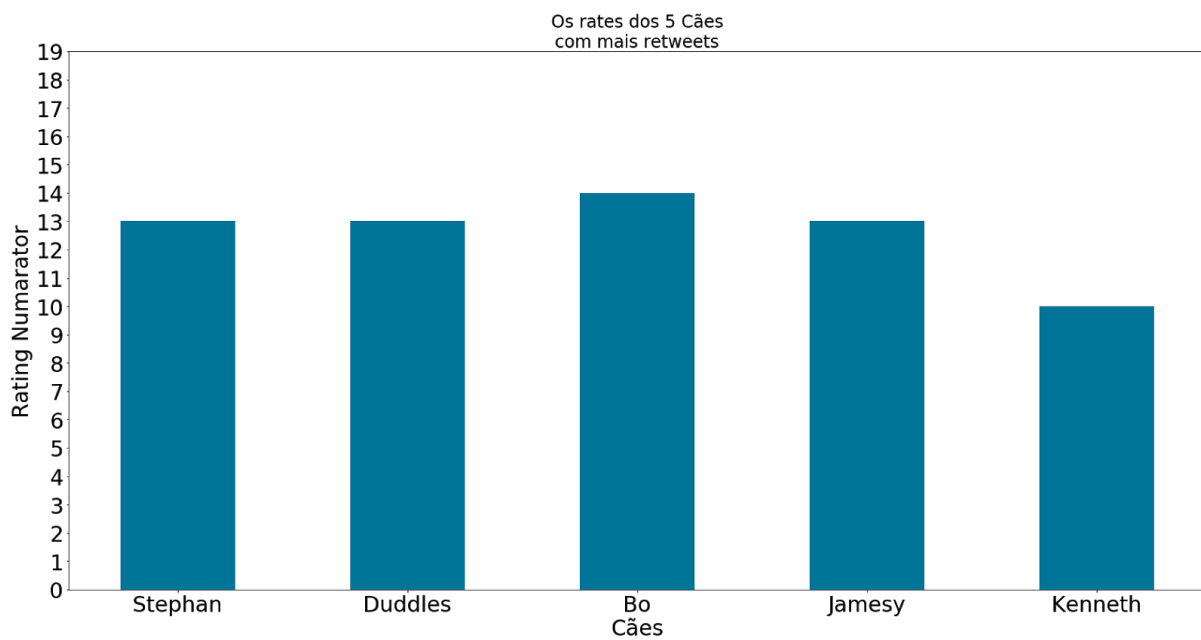
Quais os 5 cães que tiveram mais retweets, suas classificações do WeRateDogs e quantidade de marcações como favoritos?

Os dados parecem indicar que os 5 cães que tiveram mais retweets em suas postagens foram Stephan, Duddies, Bo Jamesy e Kenneth. É interessante que, embora Stephan apareça com mais retweets, ele aparece com o rate de WeRateDogs e marcações de favorito menores que outros cães. No entanto, Jamesy que aparece na quarta posição em retweets, com o mesmo rate de Stephan e mas com um total de marcações de favorito maior.

Ao comparar retweets, marcação como favorito e os rates do WeRateDogs de todos os cães observou-se uma correlação positiva entre retweets e favoritos.



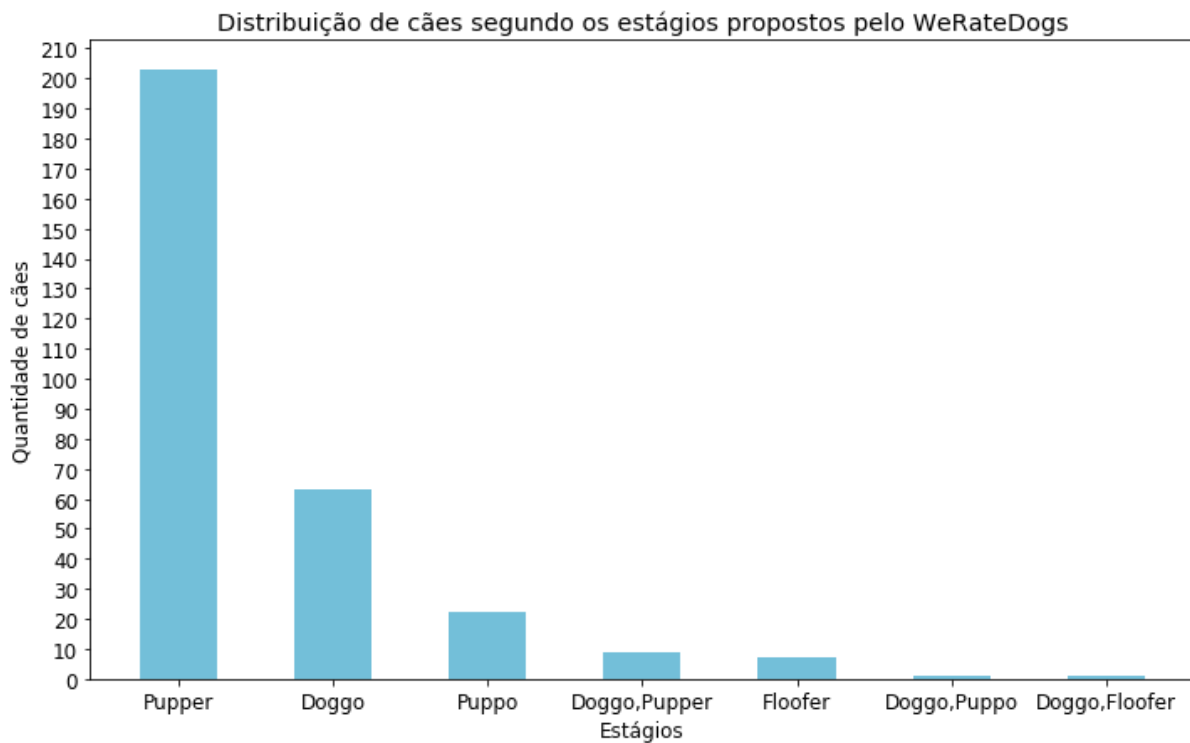
Este é o Stephan.



Questão 2

Qual a distribuição dos cães segundo os estágios propostos pelo WeRateDogs?

A distribuição dos cães de acordo com os estágios propostos pelo WeRateDogs sugere que a maioria dos cães são "Pupper".

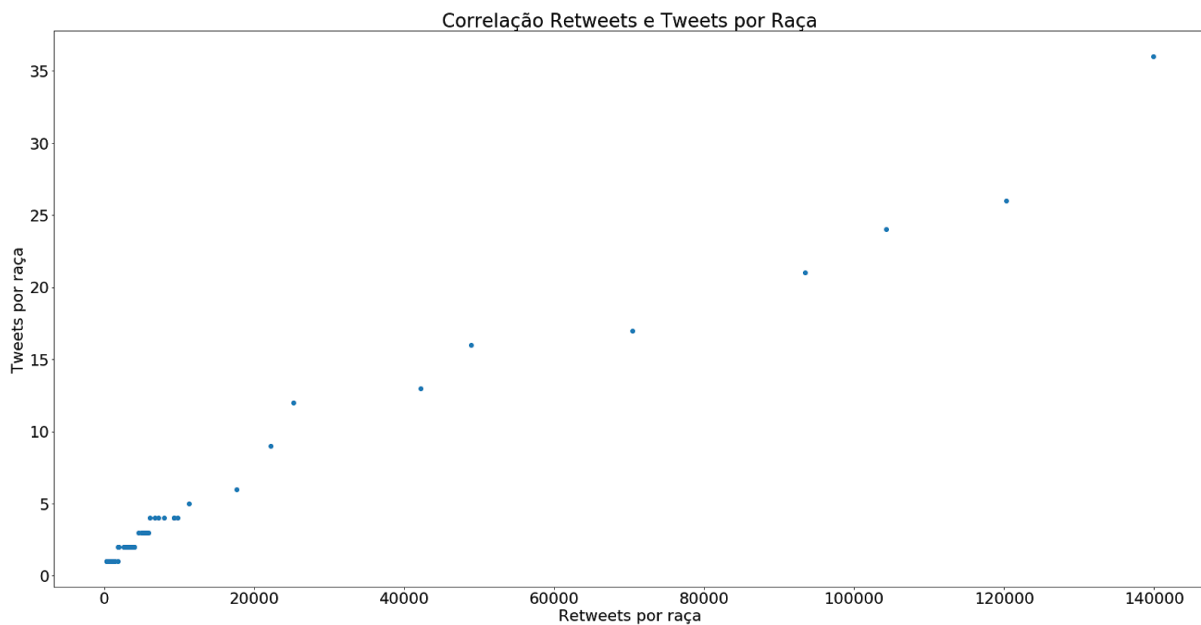


Um 'Pupper' chamado Olie

Questão 3

A raça dos cães parece ter alguma relação com o compartilhamento dos tweets?

Os dados apontam um maior número de retweets para raça Golden Retriever. Porém os dados também apontam que é a raça que mais tem postagem no WeRateDogs. Através do gráfico de dispersão percebe-se uma correlação positiva entre a quantidade de tweets e retweets. Ou seja, sem indicação de muitos retweets para uma raça que tenha tido poucos tweets.



Este é um Golden Retriever chamado Sunny

Questão 4

Há muitos cães com o mesmo nome? Se sim, quais 10 os nomes mais comuns?

Há muitos cães com o mesmo nome. Os 10 nomes mais comuns parecem ser: Charlie, Lucy, Oliver, Cooper, Tucker, Penny, Sadie, Winston, Lola e Toby.

Nomes	Charlie	Lucy	Oliver	Cooper	Tucker	Penny	Sadie	Winston	Lola	Toby
Qtde.	11	10	10	10	9	9	8	8	7	7

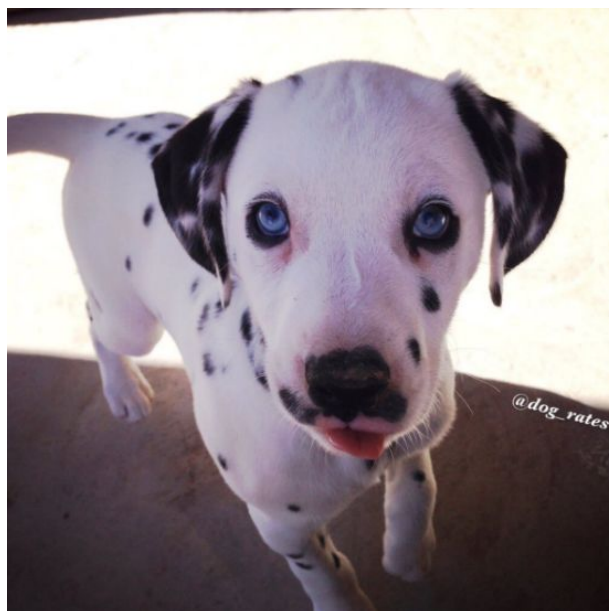
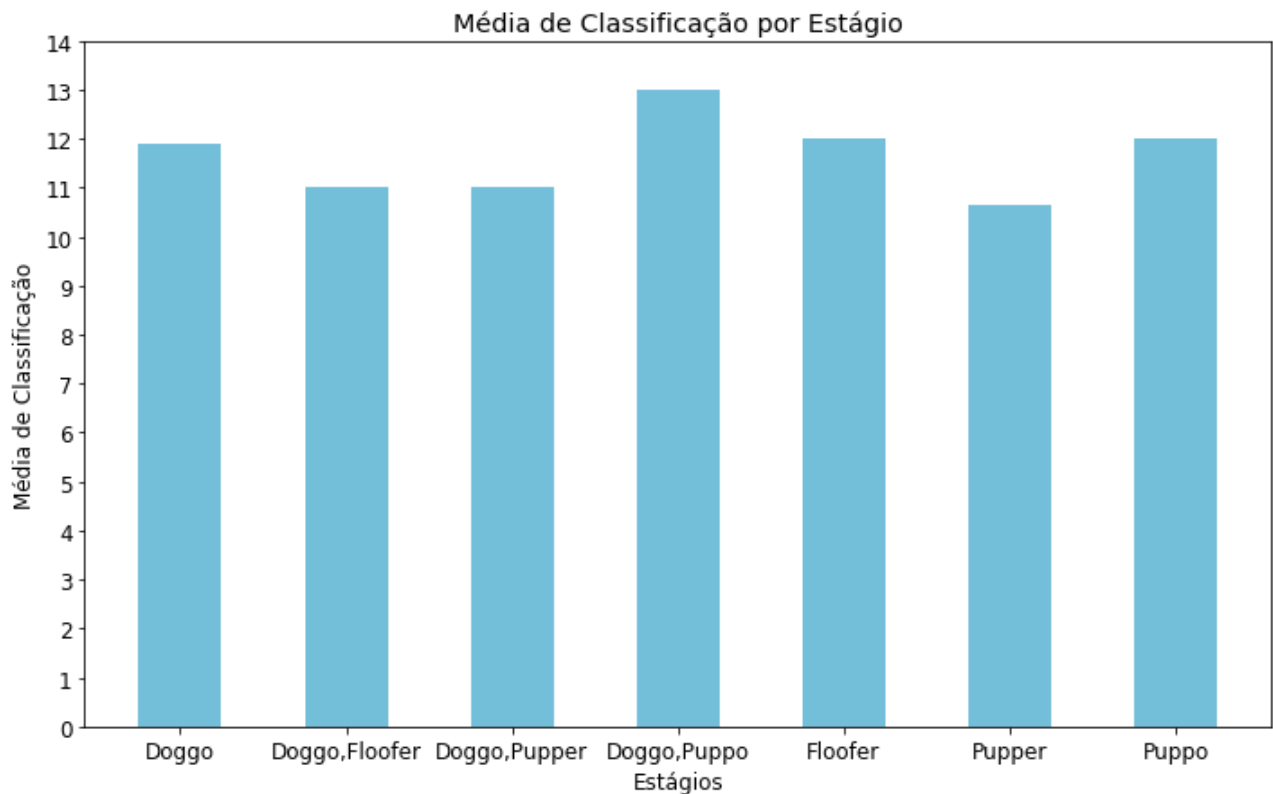


Este é um dos Charlies

Questão 5

O estágio do cão sugere alguma relação com a classificação dada pelo WeRateDogs?

Considerando que na Questão 2 os dados apontam que a maioria dos cães estão indicados como 'Pupper'. Aparentemente cães neste estágio recebem uma classificação menor que os demais.



Outro 'Pupper', um Dálmata chamado Cooper

Conclusão

De acordo com os dados os 5 cães que tiveram mais retweets em suas postagens foram Stephan, Duddies, Bo Jamesy e Kenneth, observou-se que cães com maior rate parecem receber mais retweets.

A maioria dos cães parece ser "Pupper" e a raça que mais teve retweets foi a Golden Retriever, que é também a raça que aparece com o maior de postagens. De uma maneira geral os gráficos não parecem indicar nenhuma raça que tenha poucas postagens, mas com número de tweets por postagem desproporcional.

Os nomes de cães mais comuns parecem ser: Charlie, Lucy, Oliver, Cooper, Tucker, Penny, Sadie, Winston, Lola e Toby. E aparentemente cães no estágio "Pupper" recebem uma classificação menor que os demais.

Um filhote de cão Golden Retriever, com nome 'Charlie' provavelmente fará muito sucesso, terá um ótimo rate, muitos retweets e marcações como favorito.



Limitações

Algumas das respostas tiveram limitações devido a dados indisponíveis por diversos fatores. Desta forma esta análise remete-se apenas os conteúdo parcialmente apurado através das postagens do WeRateDogs.

Fontes de consulta

- Udacity - <https://br.udacity.com>
- WeRateDogs - https://twitter.com/dog_rates
- Tweepy - <http://www.tweepy.org/>
- Twitter API - <https://developer.twitter.com>
- Slick Remix - <https://www.slickremix.com>
- Stack Overflow - <https://stackoverflow.com/>
- American Kennel Club - <https://www.akc.org/>
- Kaggle - <https://www.kaggle.com/>