

Relatório de Revisão - Projeto de Data Wrangling

para #WeRateDogs tweets

1. Dados usados

O dataframe *tweet_img_clean* foi usado para esta análise e visualização. Mais precisamente, as seguintes colunas:

- **prediction:** Contém a predição das possíveis raças de cães, de acordo com suas imagens nos *tweets*.
- **favorite_count:** Contém a contagem de favoritos para cada *tweet*.
- **retweet_count:** Contém a contagem de *retweet* para cada *tweet*.
- **rating_total:** Contém a classificação dada pelo #WeRateDogs, para cada imagem.

2. Visualizações

Para a visualização, escolhi analisar as cinco raças de cães, da coluna *prediction*, com maior contagem de tweets, que são:

1. Golden Retriever - 157 *tweets*
2. Labrador Retriever - 108 *tweets*
3. Pembroke - 95 *tweets*
4. Chihuahua - 91 *tweets*
5. Pug - 62 *tweets*

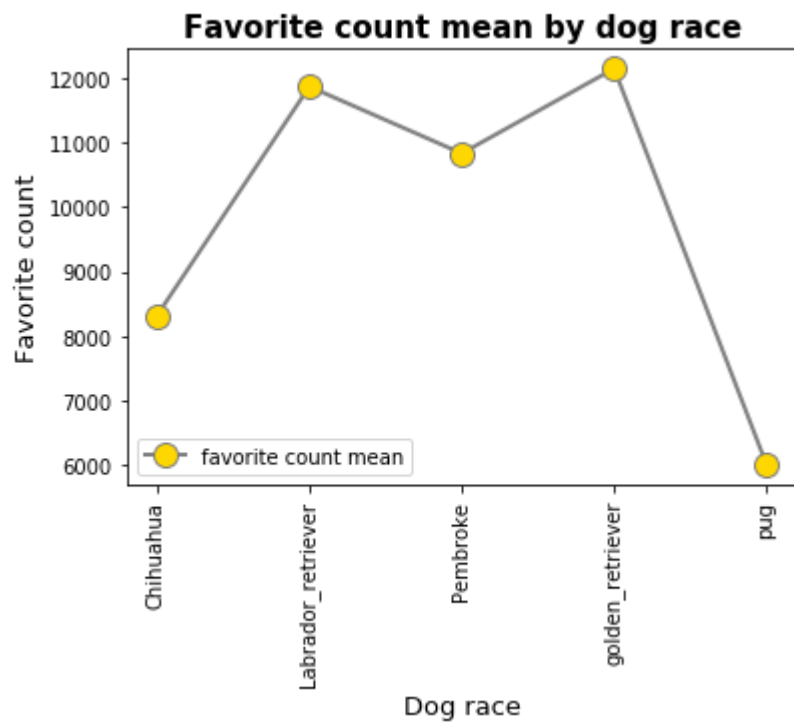
2.1. Questões

Qual a classificação das cinco raças por:

- Média de *favorite_count*?
- Média de *retweet_count*?
- Média de *rating_total*?

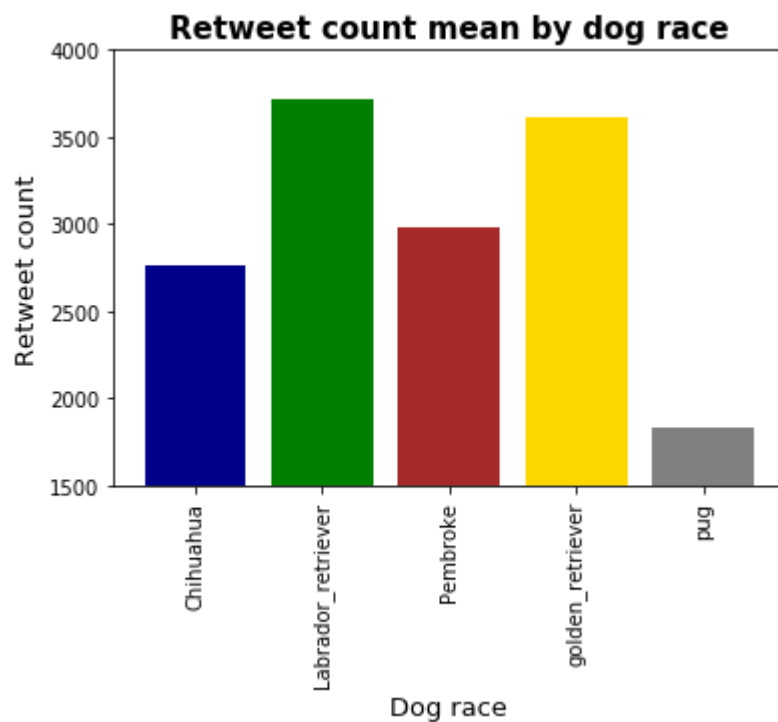
2.2. Gráficos

2.2.1. Média de *favorite_count* por raça de cachorro



Foi possível notar que o Golden Retriever ficou no topo, tendo a maior média de contagem de favoritos, seguido pelo Labrador Retriever, Pembroke, Chihuahua e Pug.

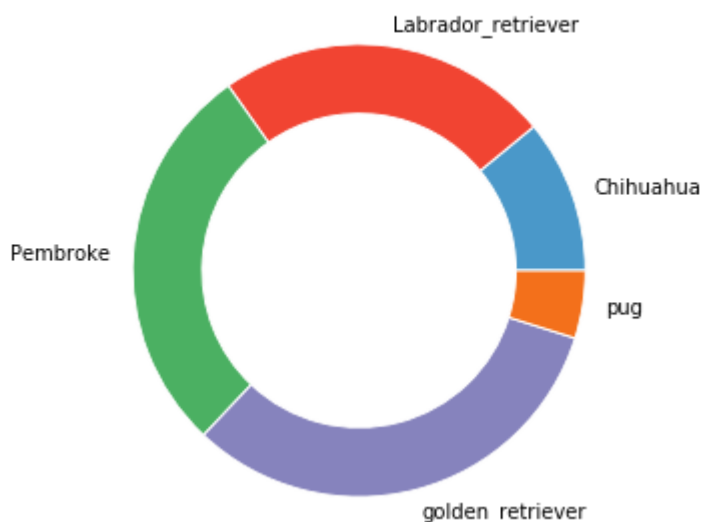
2.2.2. Média de *retweet_count* por raça de cachorro



Essa foi por pouco. O Labrador Retriever ficou no topo por uma pequena vantagem, seguido pelo Golden Retriever, Pembroke, Chihuahua e Pug.

2.2.3. Média de *rating_total* por raça de cachorro

Rating mean by dog race



Golden Retriever, de novo, fica no topo. Surpreendentemente, o Pembroke fica com o segundo lugar, seguido pelo Labrador Retriever, Chihuahua e Pug.

3. Percepções (*Insights*)

Apesar do Golden Retriever ter a maior contagem de *tweets*, ele perdeu para o Labrador Retriever em média de *retweet_count*. Outra coisa interessante aconteceu com o *rating_total*, onde o Pembroke venceu contra o Labrador Retriever, que possui a segunda maior contagem de *tweet*.

Então, enumerando, é possível perceber:

1. Golden Retriever tem, de longe, a maior quantidade de *tweets* e a maior média de *favorite_count*;
2. Mesmo com tanta quantidade de *tweets*, o Golden Retriever perdeu para o Labrador Retriever em *retweet_count*;
3. Pembroke, que fica em terceiro em quantidade de *tweets*, *favorite_count* e *retweet_count*, teve a segunda maior média de *rating_total*, com uma grande vantagem sobre o Labrador Retriever.

4. Conclusões

4.1. Limitações

As percepções aqui mostradas não são conclusivas, uma vez que toda a análise está baseada em predições. Os dados utilizados e análises realizadas foram feitas com propósito de estudo e criação de um projeto de Data Wrangling. Ainda assim, isso não diminui a seriedade e profissionalismo do trabalho apresentado.

4.2. Considerações

A #WeRateDogs apresenta uma grande variedade de *tweets* de várias raças de cães, mas as cinco que mais se destacam são: Golden Retriever, Labrador Retriever, Pembroke, Chihuahua e Pug. Com os três gráficos montados foi possível perceber algumas informações interessantes, como raças que possuíam menor contagem de *tweet* ficarem a frente de outras, que tinham maior quantidade de *tweet*, em alguns quesitos.

Isso mostra a variedade de reações do público para determinadas raças de cachorro, onde algumas tendem a ter maior quantidade de *retweet* do que outras (Labrador x Golden). Assim como a própria #WeRateDogs, que aparentam gostar mais de Pembrokes do que Labradores.

