

Análise dos Dados e Visualização

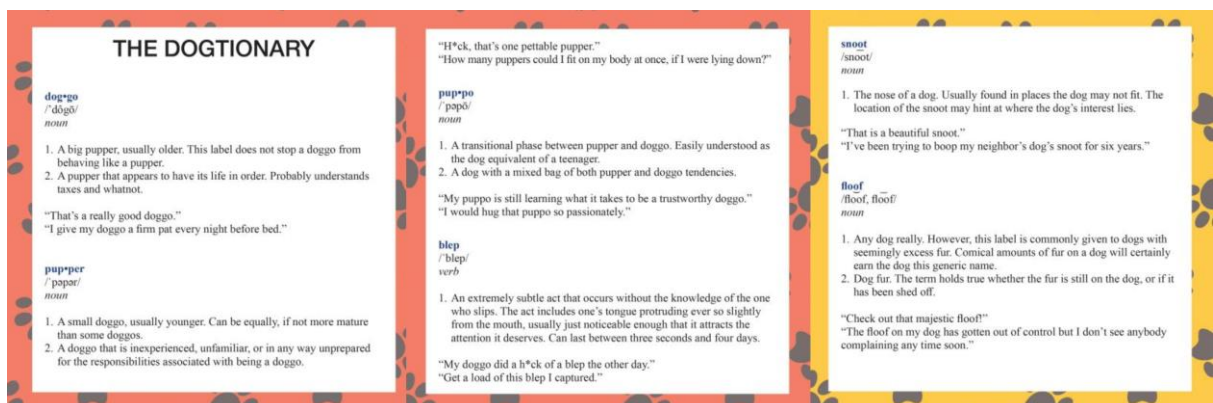
Introdução

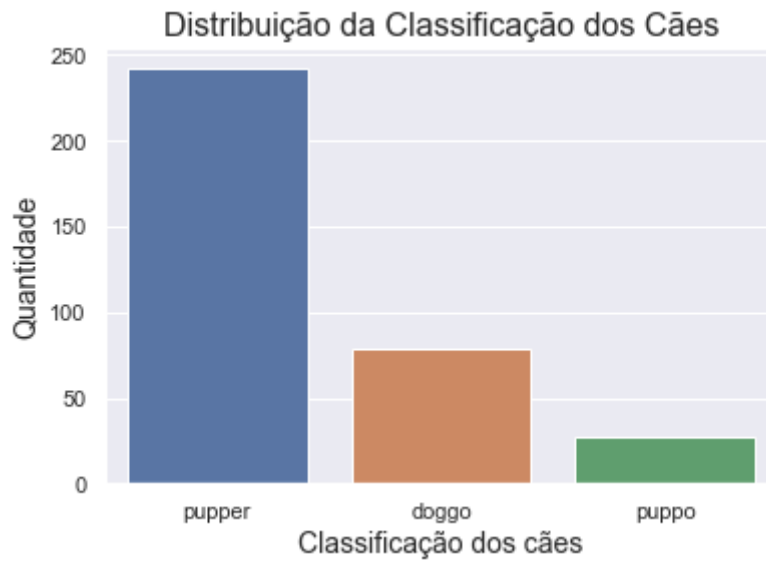
Este relatório tem como objetivo mostrar as conclusões, insights e resultados visuais das análises realizadas dos datasets *twitter_aarchive* e *image_predictions*.

Análise de dados e Visualizações

1. Distribuição da Classificação dos cães

O gráfico abaixo mostra qual é a classificação mais comum dada pelo WeRateDogs, entre Puppo, Doggo e Puppo. Na imagem abaixo o *Dogtitionary* explica o que significa estes estágios do cães:

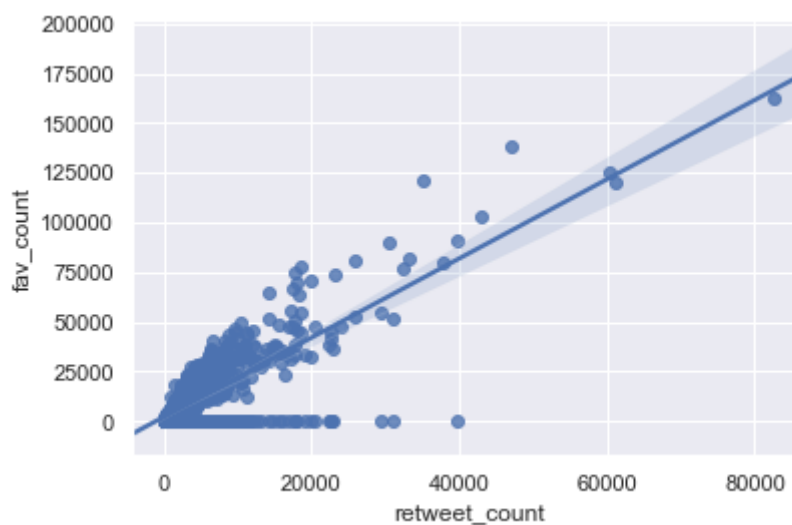




No gráfico fica clara a grande diferença do estágio Puppo para as outras duas, sendo Puppo a de menor frequência.

2. Correlação entre retweets e favoritos

Podemos observar no gráfico abaixo que há uma forte correlação positiva entre retweets e favoritos. Na tabela de correlação vemos um valor próximo de 1 (0.796), onde 1 significa correlação máxima.



	tweet_id	rating_numerator	rating_denominator	fav_count	retweet_count
tweet_id	1.000000	0.022882	-0.026000	0.518647	0.388633
rating_numerator	0.022882	1.000000	0.197295	0.016498	0.017111
rating_denominator	-0.026000	0.197295	1.000000	-0.022300	-0.021869
fav_count	0.518647	0.016498	-0.022300	1.000000	0.796528
retweet_count	0.388633	0.017111	-0.021869	0.796528	1.000000

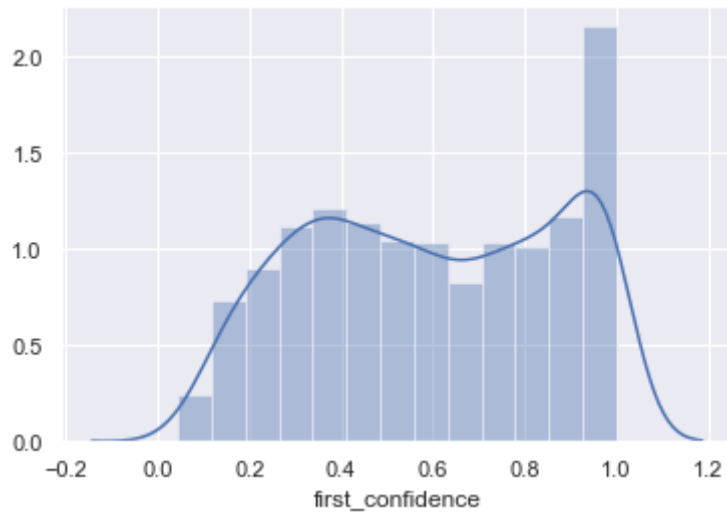
Outro ponto interessante é que a maior parte da contagem se concentra abaixo dos 20 mil retweets e 20 mil favoritos.

3. Raça do cachorro mais comum prevista pelo algoritmo

O arquivo *image_predictions* armazena os dados das previsões sobre a raça dos cães que o algoritmo realizou à partir das imagens dos tuítes.

```
Golden_retriever    150
Labrador_retriever  100
Pembroke            89
Chihuahua           83
Pug                 57
Chow                44
Samoyed             43
Toy_poodle          39
Pomeranian          38
Cocker_spaniel      30
```

De acordo com a tabela acima, Golden e Labrador são a raças de cachorros mais comuns, liderando as duas primeiras posições do top 10 da previsão realizada pelo algoritmo. Vejamos esse rank no histograma abaixo:



O Gráfico acima mostra a taxa de confiança (de 0 a 1, sendo 0 nada confiante e 1 totalmente confiante) na primeira previsão realizada. Podemos observar que a maior parte da confiança do nosso algoritmo tem sua maior quantidade próxima de 1 (entre 0.8 e 1)