

## Projeto WeRateDogs

### Act Report

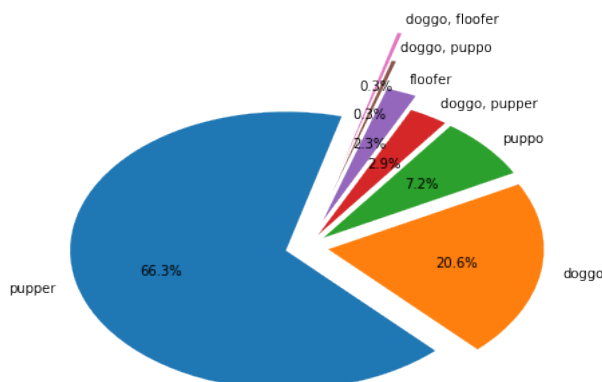
A análise dos dados do projeto WeRateDogs foi concluída com sucesso. Três fontes de dados foram utilizadas: (1) O arquivo inicial disponibilizado, identificado como `twitter_archive_enhanced.csv`, com dados do Tweeter; o arquivo (2) `image_predictions.tsv`, originário do servidor da Udacity via download por url previamente identificada; (3) dados do Tweeter obtidos via API pelo aplicativo Tweepy.

Os três conjuntos de dados puderam ser acessados e visualizados em sua plenitude. O arquivo `twitter_archive` possuía inicialmente 2.356 elementos, o `image_predictions` possuía 2.075, e os dados oriundos via API, armazenados no Data Frame `tweet_df`, totalizaram 2.343. Observa-se, assim, uma discrepância entre as fontes de dados.

A atividade de data wrangling foi composta, além da obtenção dos dados para uso no Notebook Jupyter, teve o processo de limpeza caracterizado por 3 etapas (definição, codificação e teste) para cada uma das atividades elencadas.

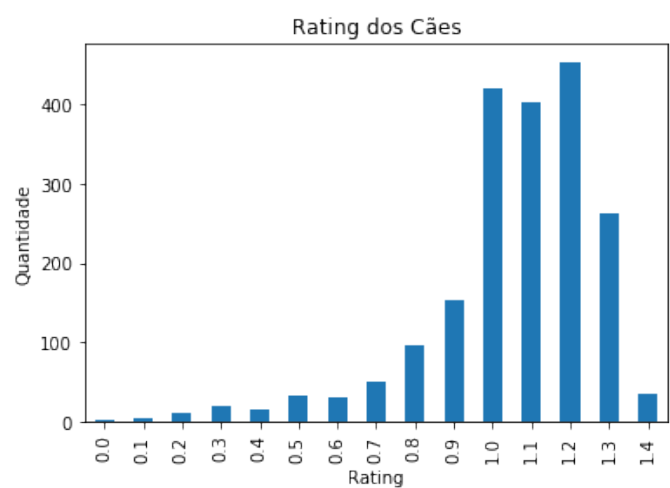
Foram encontradas 7 categorias quanto ao estágio dos cães. Dos 1992 casos analisados, apenas 306 possuíam identificação do estágio, sendo:

- 1) Pupper: 203
- 2) Doggo: 63
- 3) Puppo: 22
- 4) Doggo, pupper: 9
- 5) Floofer: 7
- 6) Doggo, puppo: 1
- 7) Doggo, floofer: 1



O gráfico acima apresenta a divisão das categorias dos cães por percentual. A maior categoria encontrada foi a dos "pupper" com 66,4%, seguida pela categoria 'doggo' com 20,6%.

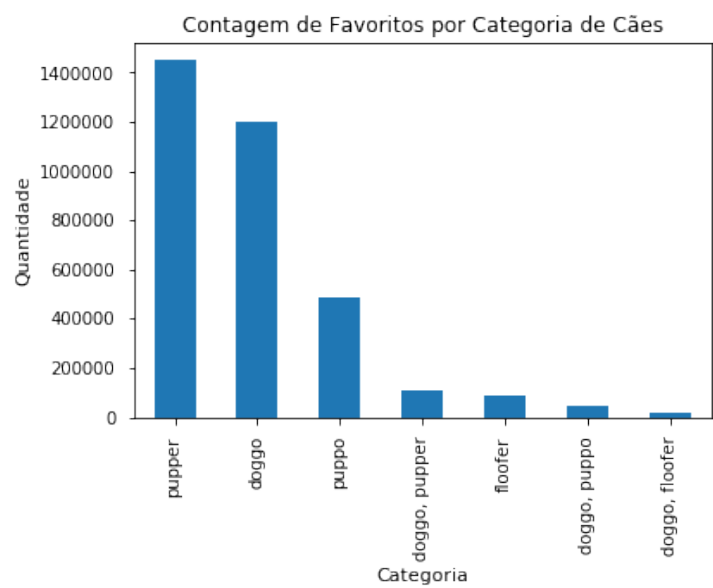
Quanto ao rating, a maior parte dos resultados são com valores de 1,0, 1,1 e 1,2. Com 1989 casos analisados, a média do rating foi de 1,05. O gráfico a seguir apresenta a distribuição dos valores de rating, variando de zero a 1,4.



Foi possível identificar e calcular a quantidade de ‘favoritos’ que os tweets receberam. Os dois primeiros cães com maiores pontuações como "favoritos", em 'favorite\_count', não possuem seus nomes identificados no dataframe. A checagem no campo 'text' confirma que em ambos os casos seus nomes não foram digitados nos tweets. Mas há o inidicativo da categoria (estágio) a que pertecem, sendo puppo e doggo, respectivamente. Os três seguintes, com maiores pontuações, foram: Jamesy, Stephan e Duddles. A maior pontuação foi de 143.260. A menor foi de 80. A média foi de 8.842

Ordem	Nome do Cão	Estágio	Quantidade de Favoritos
1	(Sem Nome)	puppo	143.260
2	(Sem Nome)	doggo	128.159
3	Jamesy	pupper	124.306
4	Stephan	--	122.909
5	Duddles	--	105.872

Na relação entre a classificação dos cães e a quantidade de favoritos que receberam, é possível observar no próximo gráfico que as duas categorias com maior quantidade de 'favoritos' são 'pupper' e 'doggo', com 1.448.836 e 1.198.384 'favoritos', respectivamente.



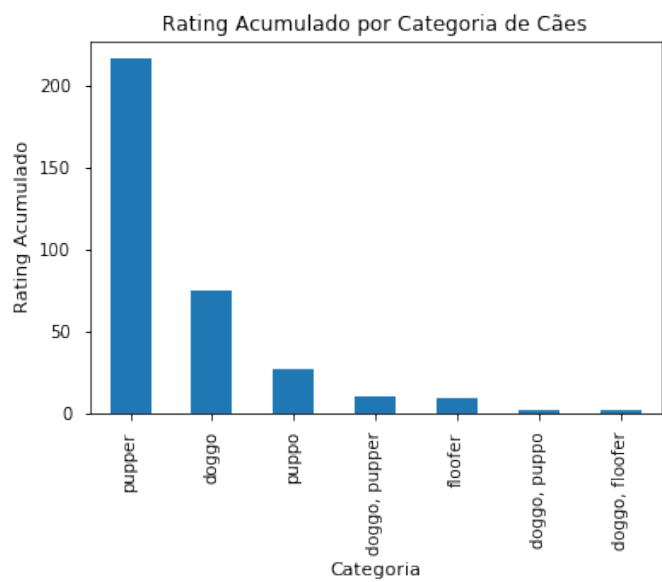
**Total da contagem de favoritos por classificação dos cães**

pupper	1.448.836
doggo	1.198.384
puppo	488.886
doggo, pupper	112.049
floofer	90.447
doggo, puppo	46.926
doggo, floofer	16.817

A categoria de cães com o maior acumulado do rating foi o ‘pupper’ com 215,9. O acumulado por categoria foi:

pupper	215.9
doggo	74.9
puppo	26.4
doggo, pupper	9.9
floofer	8.4
doggo, puppo	1.3
doggo, floofer	1.1

O gráfico a seguir apresenta o rating acumulado por categoria.



O rating médio por categoria pode ser visualizado no gráfico a seguir. A categoria ‘doggo, pupper’ apresentou a maior média e só possuiu um único elemento registrado no dataframe.

