

Fundamentos de Data Science II

Projeto 2 – Data Wrangling

Evandro Silva



Act Report

Sumário

Introdução.....	3
Objetivo.....	3
Analises e Visualizações	4

Introdução

O Intuito deste projeto é colocar em prática as lições aprendidas no módulo de Data Wrangling do Nanodegree Data Science Foundations II da Udacity.
Que estão divididas em: Coletar, Avaliar e Limpar/Arrumar.

E tem como base 3 arquivos, dois fornecidos através de link disponibilizados e um que deve ser montado através de API que faz leitura de contas no Twitter (mais detalhes serão vistos nos passos carga dos de cada arquivo):

- `twitter_archive_enhanced.csv`
Este arquivo contém dados básicos de tweets (WeRateDogs) para os tweets em 1º de agosto de 2017. O arquivo WeRateDogs do Twitter contém dados de tweets básicos para mais de 5000 dos seus tweets, mas não tudo. Uma coluna o arquivo contém com certeza: cada texto de tweet, o que eu usei para extrair classificação, nome e "estágio" do cachorro (ou seja, doggo, floofer, pupper e puppo).. Foi fornecido pela própria WeRateDogs para usa na Udacity.
- `image_predictions.tsv`
As previsões de imagens em tweets, é uma previsão de qual raça de cachorro está presente em cada tweet. Este arquivo foi criado pela Udacity com base no arquivo WeRateDogs, utilizando Redes Neurais e Reconhecimento de Imagem.
- `tweet_json.txt`
A contagem de re-tweets e favoritos de cada tweet.
Arquivo criado com o auxílio da API de consulta no Twitter (usando as ID de tweets do arquivo WeRateDogs) para cada ID tweet usar a biblioteca Tweepy do Python.

Objetivo

Fazer o Wrangler dos dados do Twitter WeRateDogs para criar análises e visualizações interessantes e confiáveis.

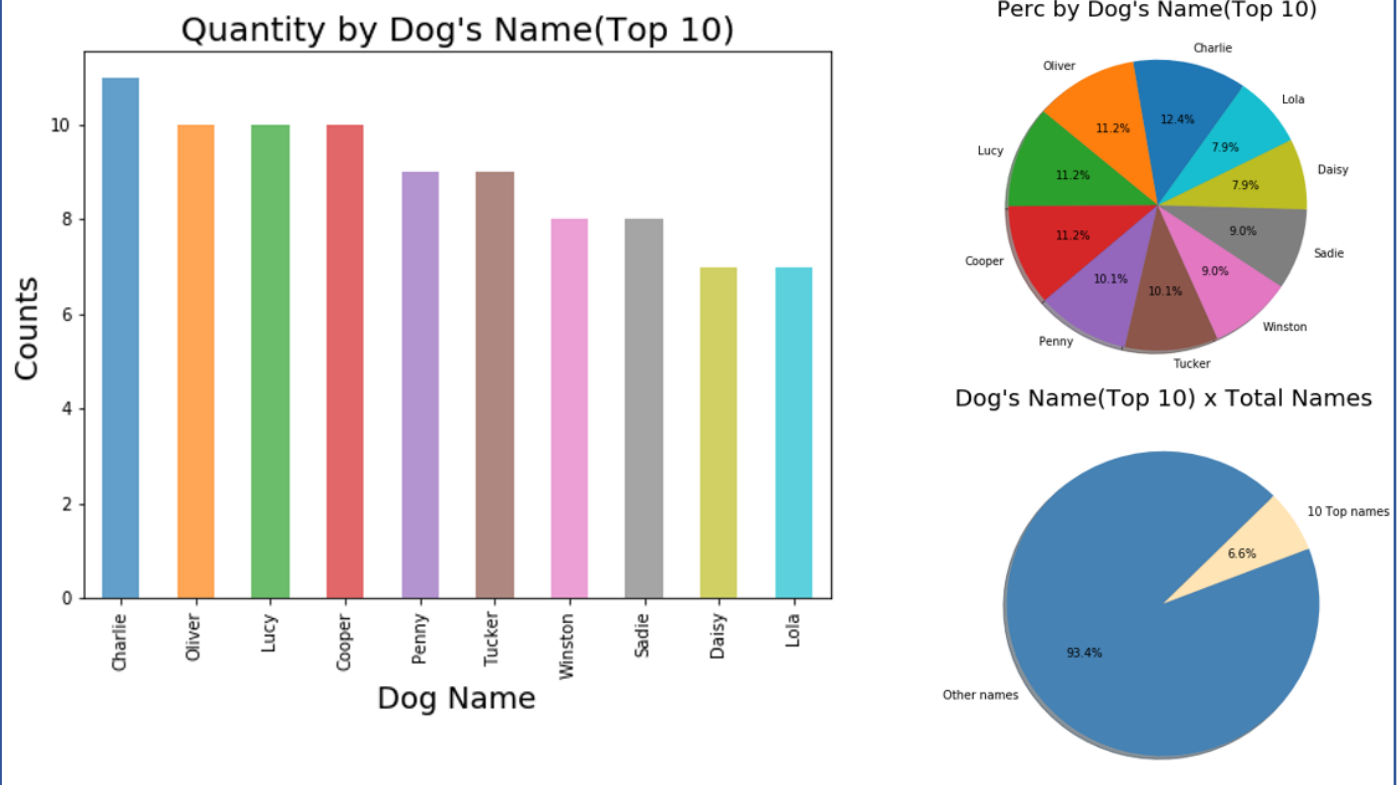
Complementando as informações do arquivo do Twitter e em seguida, fazer análises e visualizações.

Analises e Visualizações

Nos gráficos abaixo, temos uma visão de algumas informações baseada na análise dos nomes mais comuns no WeRateDogs.

- Gráfico da Lista de 10 nomes mais atribuídos aos cachorros
- Gráfico do Percentual de 10 nomes mais atribuídos aos cachorros (comparado com a soma dos 10)
- Gráfico do Percentual dos 10 nomes mais X Demais nomes

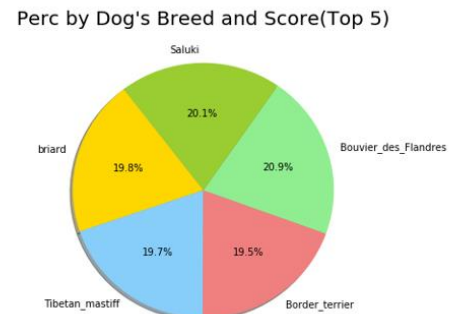
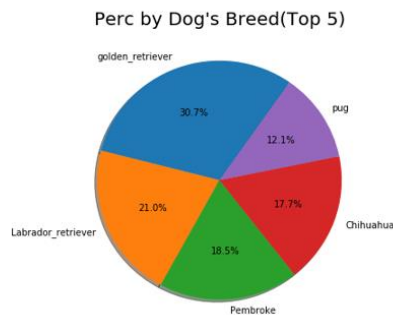
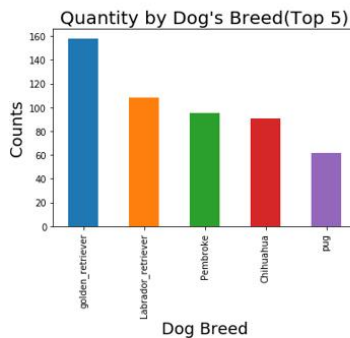
Nomes mais utilizados - Top 10 Dog Names



Charlie é o nome mais utilizado para cachorros no site WeRateDogs, com 11 ocorrências. Do total de cachorros com 10 nomes mais utilizados, Charlie foi atribuído para 12,4% dos cachorros. Quantidade de cachorro dos 10 nomes mais utilizados, corresponde a um pouco mais de 6% de todos os nomes de cachorros no site.

- Gráfico da Lista de 5 raças com mais cachorros
- Gráfico do Percentual 5 raças com mais cachorros (comparado com a soma dos 5)
- Gráfico do Percentual 5 raças com mais cachorros X Score

Raças com mais cachorros - Top 5 Dog Breed



Golden Retriever
First in Top five Breed



- Gráfico de Retweets por Score
- Gráfico do Retweets por Favoritos

Retweets – por Score e por Favotiros

