Insights And Visualization Report

Introduction

Real-world data rarely comes clean. Using Python and its libraries, we will gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling. we will document our wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python (and its libraries) and/or SQL.

The dataset that we will analyze and visualize is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs.

MATERIALS

1- The WeRateDogs Twitter archive (twitter-archive-enhanced.csv)
2- The tweet image predictions (image_predictions.tsv)
3-Twitter API (tweet_json.txt)

PROCEDURE

- Storing, analyzing, and visualizing your wrangled data.

- Reporting on 1) your data wrangling efforts and 2) your data analyses and visualizations.
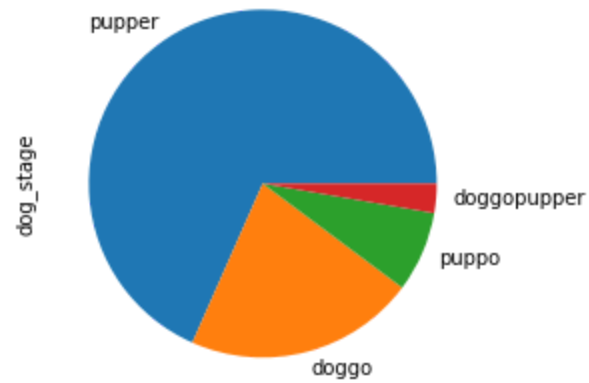
DATA

- df1 represents (twitter-archive-enhanced.csv)

- df2 represents (image_predictions.tsv)

- df3 represents (tweet_json.txt)

Storing data

We stored the clean DataFrame in a CSV file with the main one named twitter_archive_master.csv.
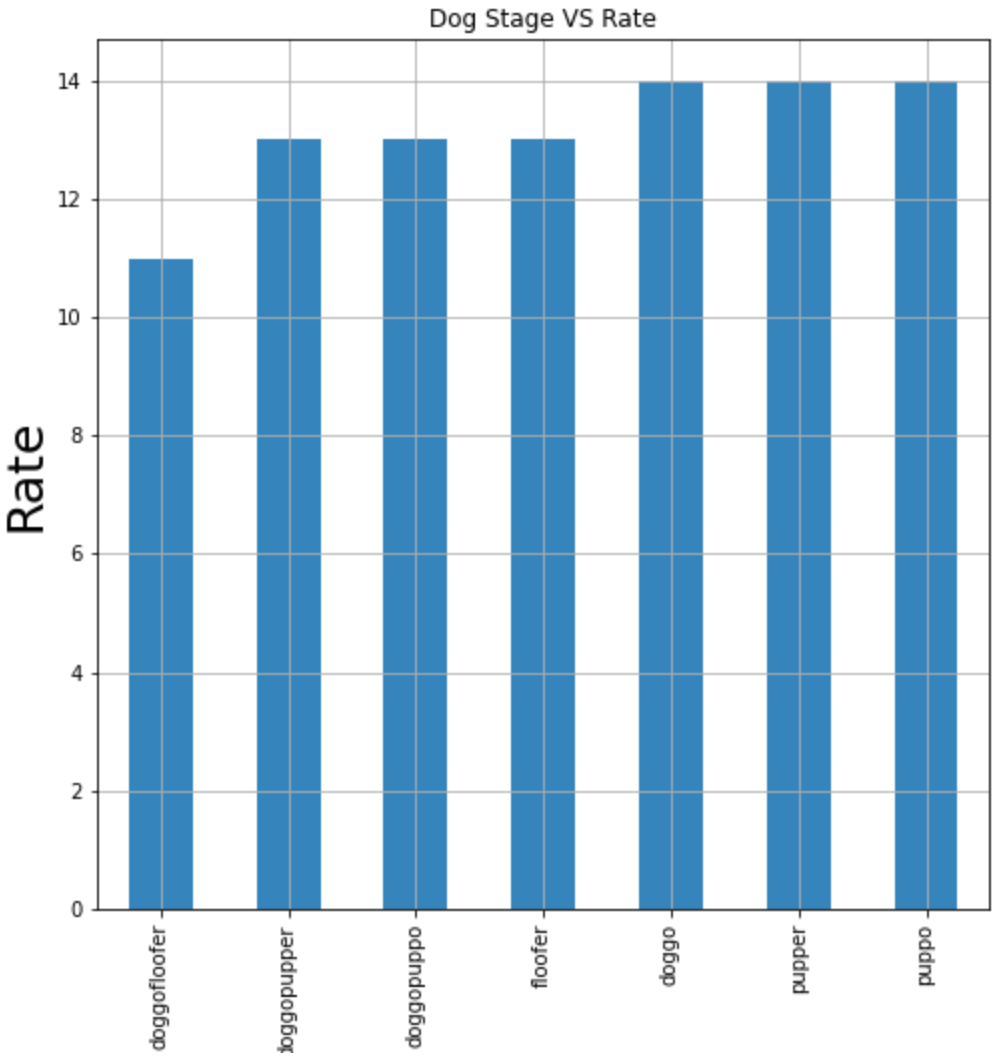
Analyzing, and Visualizing Data

1-Dog stages

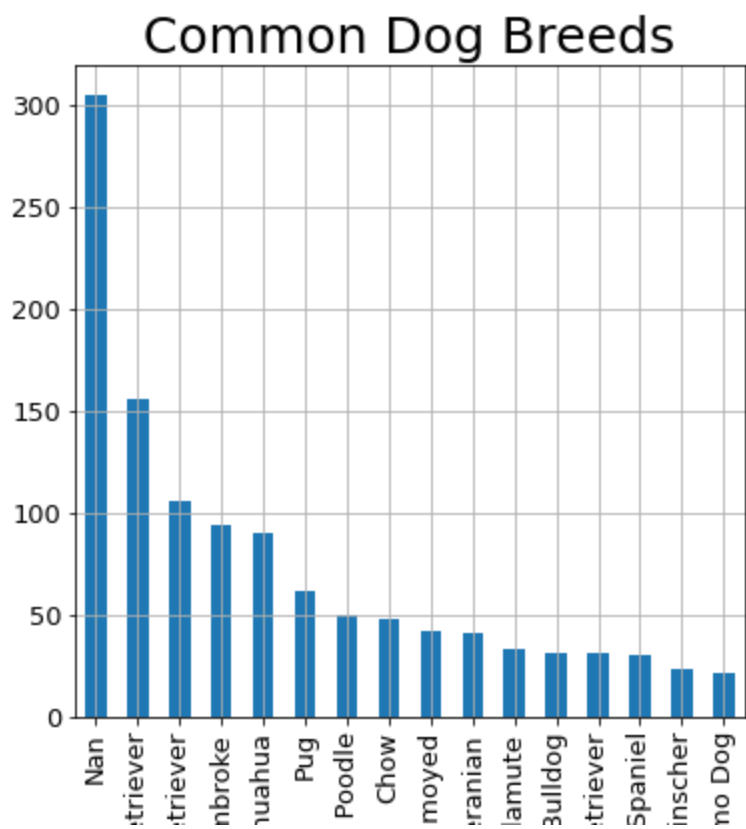## Insights (1)

- Pupper dog stage is the most common stage

2-Dog Stage VS Rate

Dog Stage VS Rate

Insights (2)

- Puppo, Pupper and doggo dog stages are the most rated stages
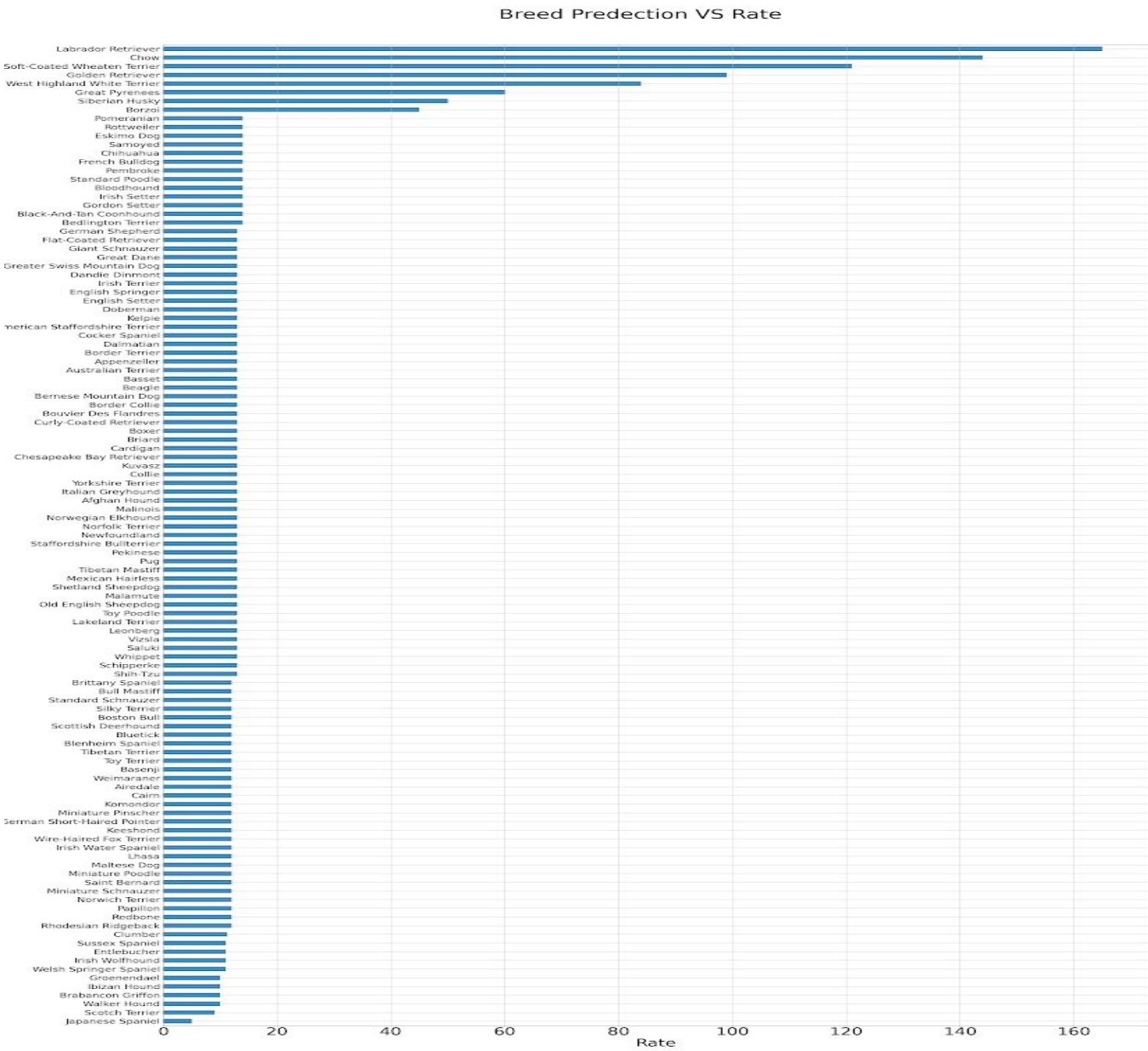
3-Common Dog Breeds

Common Dog Breeds

 Insights (3)

- Golden Retriever, Labrador Retriever and Pembroke breeds are the most common breeds
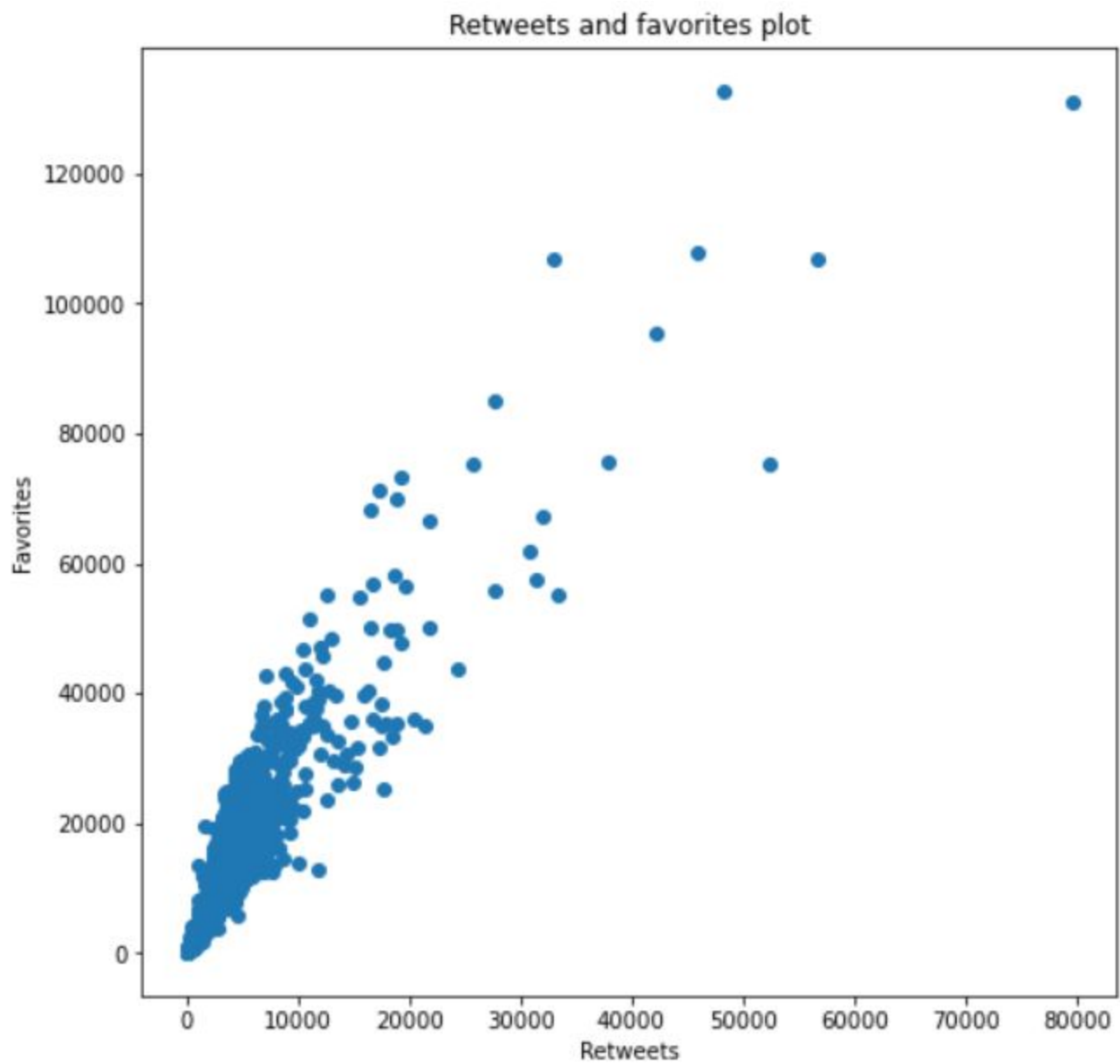
4-Breed Prediction VS Rate

Breed Predection VS Rate

Insights (4)

- Labrador Retriever breed is the most rated breed.

5-Retweets and favorites plot

Retweets and favorites plot

Insights (5)

- Retweet count is directly proportional with favorite count.

REFERENCES

1- matplotlib documentation

2- Data Wrangling course pages on Udacity

3- Stack overflow