

Wrangle and Analyze Data

Wrangle Report

Introduction

We will document our wrangling efforts, plus showcase them through analyses and visualizations using Python and its libraries and/or SQL. The purpose and goal of this project is to create a trustworthy and interesting analyses and Visualization based on the wrangled data

Questions

- Where the majority?
- Most Popular Dog Breeds?
- Retweet Count, Favorite Count and Ratings?

Project Details

This part of the project is divided into three steps, which are as follows:

- Data Gathering.
- Data Assessment.
- Data Cleaning

Gathering Data:

Gathering data from the twitter account “WeRateDogs”, download the file `twitter_archive_data` from the project page, then I downloaded the file “image-predictions” programmatically, and last thing the twitter API which I downloaded from the project page because I didn’t have twitter account.

Assessing Data:

For assessing data I first looked at the data that I have, then typed the problems (2 tidiness and 8 quality) which is:

After gathering the required pieces of data, we assessed them visually and programmatically for quality and tidiness issues. Visually, by printing the datasets in the Jupyter Notebook and by looking at it via Excel. And programmatically, by using Pandas functions and so on. Finally, we filtered the problems into quality problems (which are issues in validity, completeness, accuracy and consistency i.e. issues in the content), and tidiness problems (which are issues in the structure), and pointed them out.

Cleaning Data:

Finally after gathering and Assessing I started Cleaning, I started with Issues first because fixing it will make the rest easier, I mostly used pandas for cleaning.