

ANALYSIS AND VISUALIZATION

We Rate Dogs Data.

Introduction

This project was about wrangling the data of a twitter account user @dog_rates. This account Primarily rates people's dogs with humor attached to it. This report involves the visual representation of the insights made from the data and conclusions drawn from it.

Data gathering

The project gathered data from the following sources:

- Tweet image predictions
- Twitter archive of @WeRateDogs
- Twitter API

Assessing Data

I assessed the data on two factors; quality and tidiness issues.

Under the data [Quality](#) I looked out for the following;

1. Completeness
2. Validity
3. Accuracy
4. Consistency

Three requirements were met under the [Tidiness](#) issues

1. Each variable forms a column
2. Each observation forms a row
3. Each type of unit forms a table

Data Cleaning

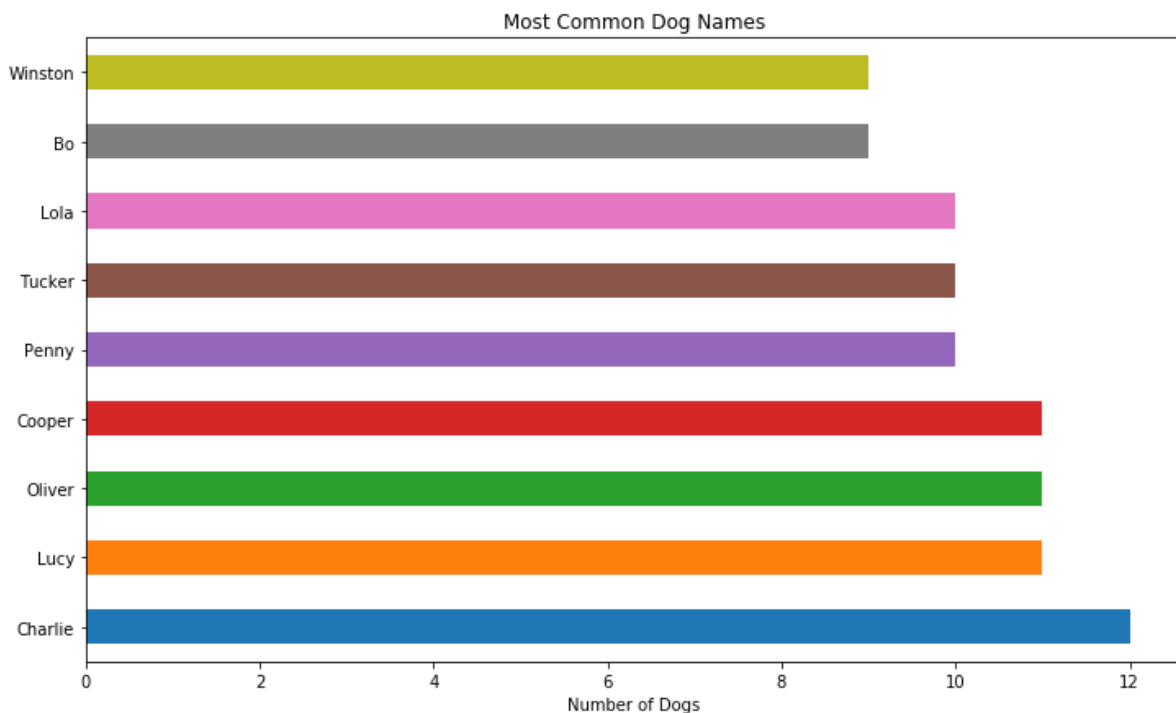
This iterative process involved three steps.

1. Define
2. Code
3. Test

Insights

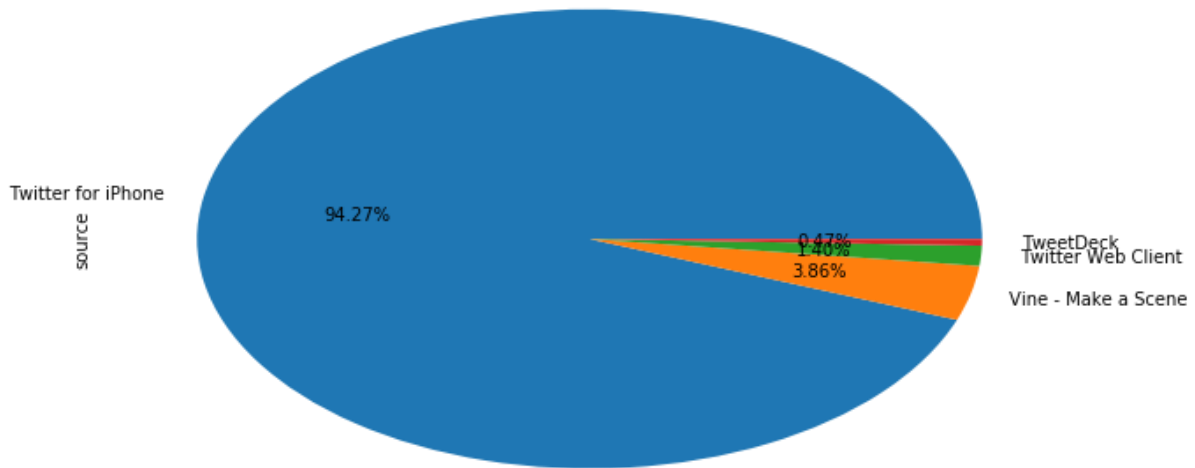
Three (3) insights were made in my project.

1. Most popular dog name



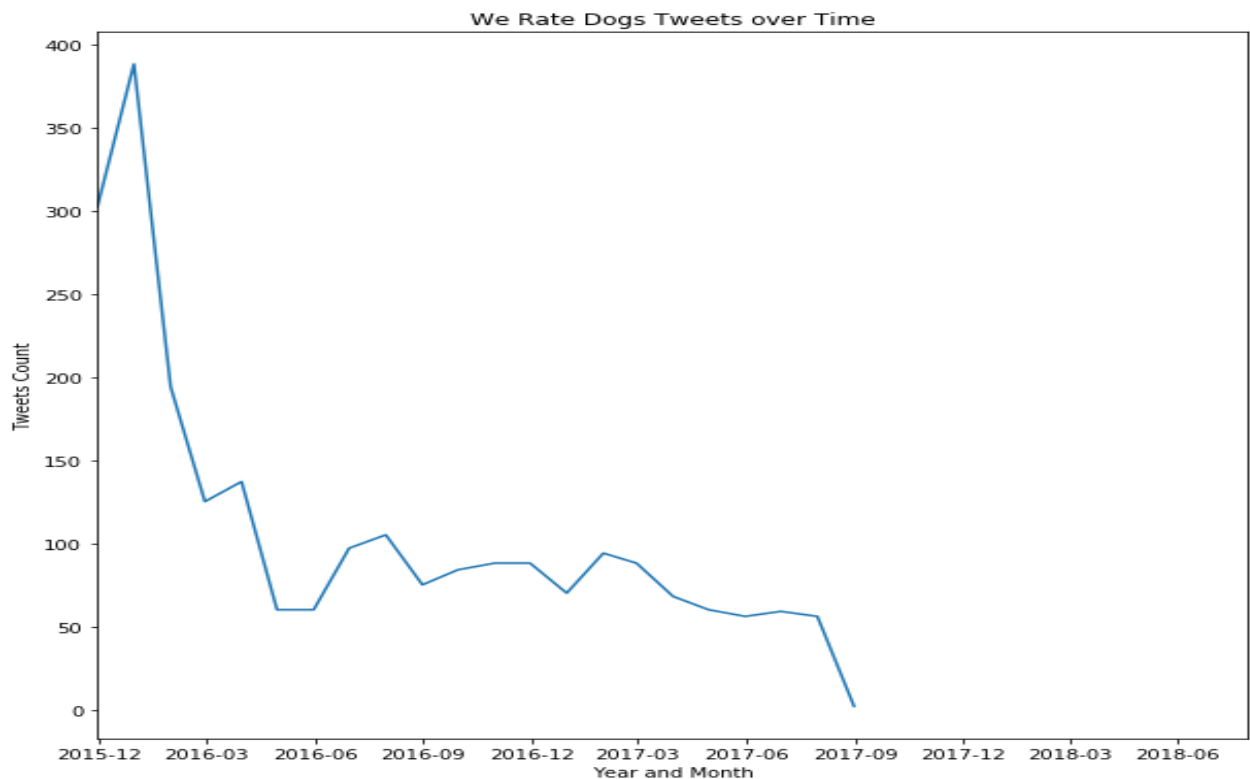
Twelve (12) dogs were named as Charlie which makes it the most popularly named dog. Lucy, Oliver and Cooper were tied at being the second most popular dogs' names with eleven (11) named dogs each.

2.Source with the most tweets



Four mediums were used by the users to share their tweets namely iPhone, Vine, Tweet Deck and Tweet Web Client. The Pie chart visualized above shows that amongst these four, iPhone was the most used medium for tweeting contributing for 94.27% of the tweets. Vine made up for 3.86% with Tweet Deck and Twitter Web Client making up for less than 2% of the tweets combined.

3. Tweet interactions overtime



The graph above interpretes the tweets by the the user and its corresponding interactions over a period of time(2015 – 2018). The tweet interactions saw a rise in numbers in the first quarter of 2016 but seemed to experience a negative regression afterwards until all interactions siezed at the beginning of the quarter of 2017.