

# act\_report

March 29, 2022

## 1 Act Report: Wrangling and Analysing Project

After the gathering, assessing and cleaning of the data, I could finally start drawing conclusions from the data.

### 1.1 Insights

In the scope of this project I looked at three different metrics insights wise.

#### 1.1.1 Tweets per Year

The first question that came to my mind was: From which year were most tweets? To answer this I made use of the timestamp column in datetime format that I created before. Only three years are in the data. 2015, 2016, 2017 The channel was most active in the first year present in the data set, 2015, with 1054 tweets. That in an impressive 2.86 posts per day average. In the year 2017 they were much less productive with only 375 tweets.

#### 1.1.2 Most frequent dog names

The second metric I was interested in was the dogs' names. Unfortunately, a lot of the dogs' names are unknown. But some of the most popular names are Lucy, Charlie and Olivier, which is weird, because to me these are not even dog names. Nevertheless, all of these three names were given to three dogs.

#### 1.1.3 Dog breeds ratings

I was also curious which breeds were the highest rated. To find out I grouped the dogs by breed and looked at these ratings. It turns out that Gordon Setter and Pembroke were the highest rated ones with an average of 1.4. (14/10) I was surprised by these results because I have never even heard of these breeds before and expected to see Labradors and German Sheppards up there.



## 1.2 Visual Insights

I also created some visuals to get ahold of the data.

### 1.2.1 Dog breeds with the highest favorite counts

For that I created a distinct list of all breeds and made a list with all favorite counts per breed. When plotted as a bar chart we can see that most favorite counts are on unknown dogs. That aside, Labrador retrievers and Chihuahuas are very popular.

### 1.2.2 Dog Stages

Another plot I created is the dog stages. It shows, that most dogs in the data are in the pupper stage. Floofers are almost non-existent in that data set.

