Data Analyst Nanodegree

We Rate Dogs @dog_rates

Wrangle and Analyze Data

Bayan Bin Manea - Project 5

UDACITY

أكاديمية مسك
MISK ACADEMY

Introduction

Real-world data rarely comes clean. In this project, I will be using Python and its libraries, to 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.

The dataset that I will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog.

These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

Insights

1. After evaluating models performances, the second model has the highest true predictions.
2. The most common word used in @dog_rates tweets is "This is"
3. Golden Retriever is the most common dog breed in @dog_rates retweets
4. The most common dog stage in @dog_rate tweet is "pupper" by 65.8% of overall tweets
5. There is a positive correlation between favorites count and retweet count

Visualizations

Most Common Dog breed in @dog_rates Tweets

Dog Stage @dog_rates tweets

Favorites vs. Retweets @dog_stage