

Act Report

Introduction:

The Twitter user @dog_rates, also known as WeRateDogs is account rates people's dogs with a humorous comment about the dogs, his tweet was used through this project that focuses on gathering, assessing and cleaning data, and then analyzing and visualizing data.

Gather:

Three data was gathered from different sources:

- 1- **Enhanced Twitter archive:** contains WeRateDogs tweet for all 5000+ of their tweets.
- 2- **Image predictions:** a table full of image predictions provided from Udacity to download.
- 3- **Tweet data:** contains retweet and favorite count.

Assess:

Assess data began with assessing data visually and programmatically to detect and document issues. Asses data on both quality and tidiness issues.

The quality has Four main quality dimensions: completeness, validity, accuracy, and consistency. And there are three main requirements for tidiness:

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

Clean:

Clean is fixing the quality and tidiness issues that are identified in the assessment step. The data cleaning process has three steps: Define, code, test

Storing Data:

the cleaned data was saved in twitter_archive_master.csv.

Analyzing and visualizing data:

Insights:

1. The Labrador retriever breed has the highest retweet and favorite.
2. Puppy has the average highest rating, retweet and favorite.
3. Lucy and Charlie are the most popular dog name.

Visualization

