Data Nutrition

About:

The data set includes the status of each YouTube video's status from 2017 October 13th to 2018 December 31th. The data sets tallied the videos' likes, dislikes, comment count, like-view ratio, and video views as the dependent variables and their category as the independent variable.

———

Created by: Mitchell J

Content: An article with csv file.

Source: https://www.kaggle.com/datasets/datasnaek/youtube-new/

———

Use Cases

Potential real-world applications of the dataset

- What category of videos should a new YouTuber create to be popular?
- What categories of video are becoming more popular?
- What categories of video did people more likely to notice before COVID?

———

Description:

TELL US ABOUT THIS DATASET
This data set includes the status of each YouTubers' status constructed by Mitchell J and clean by

Yukang Zhao, who is our team member.

IS THERE AN INTENDED PURPOSE FOR THE DATASET? WHAT DOMAIN WAS IT DESIGNED FOR?
The purpose of this dataset is to help us figuring out what category is the most popular in
ARE THERE ANY ADDITIONAL RESTRICTIONS UNDER WHICH THE DATASET IS MADE AVAILABLE? ARE THERE ANY LEGISLATION, CODES OF PRACTICE OR GUIDANCE FROM THE JURISDICTION OR DOMAIN IN WHICH THE DATA WAS COLLECTED ABOUT THE USE OF THIS DATA? ARE THERE TASKS FOR WHICH THE DATASET CANNOT BE USED? PLEASE PROVIDE A DESCRIPTION OR GUIDANCE.
No

ARE THERE TASKS FOR WHICH THE DATASET WOULD BE CAUTIONED AGAINST BEING USED? IF SO, PLEASE PROVIDE A DESCRIPTION.
We do not believe so.

HAS THE DATASET BEEN USED FOR ANY TASKS ALREADY? IF SO, PLEASE PROVIDE A DESCRIPTION AND LINKS TO PAPERS OR SYSTEMS USING THE DATASET
We don't find any.

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?
It represents the US videos published in 2017 and 2018.

How many instances are there in total (of each type, if appropriate)?
23778

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?

This data set doesn't contain all possible instances, it is a sample from larger set.

Are relationships between individual instances made explicit?
Every individual instances are independent of each other.