

Final Project Proposal

Team Members: Xavier Robbins and Paige Rosynek

Dataset: YouTube Trending Video Dataset - US (Kaggle)

https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset?select=US_youtube_trending_data.csv

The dataset we have chosen to explore is the YouTube Trending Video Dataset from Kaggle. The dataset contains information on the top trending YouTube videos (up to 200 videos) each day from August 3, 2020 to January 10, 2023, organized in a csv format. The dataset has a total of 177,390 entries with 16 columns. The columns and their descriptions are included in the table below:

Column	Type	Description
video_id	string	YouTube video ID that uniquely identifies each video on the platform.
title	string	Title of the video.
publishedAt	string	Date and time the video was uploaded.
channelId	string	YouTube channel ID that uniquely identifies each video on the platform.
channelTitle	string	Title of the YouTube channel the video was uploaded by.
categoryId	int	Integer value that represents the category of the video. The corresponding category value can be looked up in the JSON file described below.
trending_date	string	The date the video was on YouTube's trending list.
tags	string	The keywords listed by the video creator that relate to the content of the video. The keywords are given as a single string with each keyword separated by a ' ' character.
view_count	int	The number of total views the video had from 'publishedAt' up until 'trending_date'.
likes	int	The number of total likes the video had from 'publishedAt' up until 'trending_date'.
dislikes	int	The number of total dislikes the video had from

		'publishedAt' up until 'trending_date'.
comment_count	int	The number of total comments the video had from 'publishedAt' up until 'trending_date'.
thumbnail_link	string	Link to the thumbnail image of the video, which is the cover image that is displayed before the user clicks on the video.
comments_disabled	boolean	Whether or not the comments were disabled on the video, so users could not comment on the video.
ratings_disabled	boolean	Whether or not the video is age restricted. True - The video is age restricted and can only be viewed by users 18+. False - The video is not age restricted and can be viewed by any user on the platform.
description	string	The video description provided by the uploader.

The dataset is accompanied by a separate JSON file (US_category_id.json) that contains mappings for the categoryId column of the video dataset and has a total of 33 different categories. This dataset can be used to get the actual category string for a video based on its categoryId.

Objectives:

We both watch YouTube often for entertainment and learning, and we thought it would be interesting to explore trending videos and what characteristics of a video have an impact on whether or not it is trending. With this data set, we want to explore what features of a video make it trend, and we want to predict features of trending videos on YouTube, such as how many likes or comments a video will get. We want to utilize the trending date column and feature engineering to determine how many likes, dislikes, or comments a video may get in a day on the trending page. On top of this, it would be interesting to predict the category of a trending video on YouTube, and explore what categories of videos are most likely to end up on the trending page. Some more questions we want to explore and/or answer with our dataset are listed below:

- What features make a video popular?
- Predicting the number of likes or dislikes a trending video will get.
- Predicting the video category for a trending video.
- How many views will a trending video get?

- How many comments will a trending video get?
- Are certain video categories more likely to be trending?
- Are certain channels more likely to upload a video that ends up on the trending list?
- How long from the time of upload will it take for a video to become trending?
- How many days will a video be on the trending list?
- Does the thumbnail of a video play a role in whether or not it is trending?
- What features determine if a video will be trending?

First 10 Records:
US_youtube_trending_data.csv

	video_id	title	publishedAt	channelId	channelTitle	categoryId	trending_date	tags	view_count	likes	dislikes	comment_count	thumbnail_link
0	3C66w5Z0ixs	I ASKED HER TO BE MY GIRLFRIEND...	2020-08-11T19:20:14Z	UCvtRTOMP2TqYqu51xNrqAzg	Brawadis	22	2020-08-12T00:00:00Z	brawadis prank basketball skits ghost funny vi...	1514614	156908	5855	35313	https://i.ytimg.com/vi/3C66w5Z0ixs/default.jpg
1	M9Pmf9AB4Mo	Apex Legends Stories from the Outlands - "Th...	2020-08-11T17:00:10Z	UC0ZV6M2THA81QT9hrVWJG3A	Apex Legends	20	2020-08-12T00:00:00Z	Apex Legends Apex Legends characters new Apex ...	2381688	146739	2794	16549	https://i.ytimg.com/vi/M9Pmf9AB4Mo/default.jpg
2	J78aPj3VyNs	I left youtube for a month and THIS is what ha...	2020-08-11T16:34:06Z	UCYzjPXprvlSY-Sf0g4vX-m6g	jacksepticeye	24	2020-08-12T00:00:00Z	jacksepticeye funny funny meme memes jacksepti...	2038853	353787	2628	40221	https://i.ytimg.com/vi/J78aPj3VyNs/default.jpg
3	kXLn3HkppjaA	XXL 2020 Freshman Class Revealed - Official An...	2020-08-11T16:38:55Z	UCbg_UMjIHJg_19SZckaKajg	XXL	10	2020-08-12T00:00:00Z	xxl freshman od freshmen 2020 xsl freshman 20...	496771	23251	1856	7647	https://i.ytimg.com/vi/kXLn3HkppjaA/default.jpg
4	VIUo6yapDbc	Ultimate DIY Home Movie Theater for The LaBran...	2020-08-11T15:10:05Z	UCDVPcEbVLQgLZX0Rt6jo34A	Mr. Kate	26	2020-08-12T00:00:00Z	The LaBrant Family DIY Interior Design Makeove...	1123889	45802	964	2196	https://i.ytimg.com/vi/VIUo6yapDbc/default.jpg
5	w-aid8dvZo8	I Haven't Been Honest About My Injury. Here's...	2020-08-11T20:00:04Z	UC5zJwsfEs9WYe3A76p7xdA	Professor Live	24	2020-08-12T00:00:00Z	Professor injury professor achilles professor ...	949491	77487	746	7506	https://i.ytimg.com/vi/w-aid8dvZo8/default.jpg
6	uet14uf9NsE	OUR FIRST FAMILY INTRO!!	2020-08-12T00:17:41Z	UCDSjCBYqL7VQrBkfr1RtwA	Les Do Makeup	26	2020-08-12T00:00:00Z	[None]	470446	47990	440	4558	https://i.ytimg.com/vi/uet14uf9NsE/default.jpg
7	ua4QMFOATco	CGP Grey was WRONG	2020-08-11T17:15:11Z	UC2C_jShlt725hvb1ar5V9w	CGP Grey	27	2020-08-12T00:00:00Z	cpggrey education hello internet	1050143	89190	854	6455	https://i.ytimg.com/vi/ua4QMFOATco/default.jpg
8	SnsfZj91R7E	SURPRISING MY DAD WITH HIS DREAM TRUCK!! Lou...	2020-08-10T22:26:59Z	UCZDdF_p-L88NWVpzF0vjvMQ	Louie's Life	24	2020-08-12T00:00:00Z	surprising dad father papa with dream car truc...	1402687	95694	2158	6613	https://i.ytimg.com/vi/SnsfZj91R7E/default.jpg
9	SsWHMAhshPQ	Ovi x Natanael Cano x Aleman x Big Soto - Veng...	2020-08-11T23:00:10Z	UC648rgJOboZlgcDbW00vTSA	Rancho Humilde	10	2020-08-12T00:00:00Z	Vengo De Nadal Aleman Ovi Big Soto Trap Ovi Nat...	741028	113983	4373	5618	https://i.ytimg.com/vi/SsWHMAhshPQ/default.jpg

US_category_id.json

	kind	etag	id	snippet.title	snippet.assignable	snippet.channelId
0	youtube#videoCategory	lfWa37JGcqZs-jZeAyFGkbeh6bc	1	Film & Animation	True	UCBR8-60-B28hp2BmDPdntcQ
1	youtube#videoCategory	5XGylls7zkjHh5940dsT5862m1Y	2	Autos & Vehicles	True	UCBR8-60-B28hp2BmDPdntcQ
2	youtube#videoCategory	HCjFMARbBeWjpm6PDfReCOMOZGA	10	Music	True	UCBR8-60-B28hp2BmDPdntcQ
3	youtube#videoCategory	ra8H7xyAfmE2FewsDabE3TUSq10	15	Pets & Animals	True	UCBR8-60-B28hp2BmDPdntcQ
4	youtube#videoCategory	7mqChSJogdF3hSIL-88BfDE-W8M	17	Sports	True	UCBR8-60-B28hp2BmDPdntcQ
5	youtube#videoCategory	0Z6uGkj97NgjD-X3pkA-nL18Hqk	18	Short Movies	False	UCBR8-60-B28hp2BmDPdntcQ
6	youtube#videoCategory	K_-7stg0kIU7eUBOPUa6j5fqIMQ	19	Travel & Events	True	UCBR8-60-B28hp2BmDPdntcQ
7	youtube#videoCategory	I3IL9xGIM3MsULlqR4tvCsEkv98	20	Gaming	True	UCBR8-60-B28hp2BmDPdntcQ
8	youtube#videoCategory	D1W6tq5mMMCV0wtNxf9A6g9wWjU	21	Videoblogging	False	UCBR8-60-B28hp2BmDPdntcQ
9	youtube#videoCategory	QMEBz6mxVdklVaq8JwesPEw_4nl	22	People & Blogs	True	UCBR8-60-B28hp2BmDPdntcQ