**Youtube trending videa**

**Tým:**

Erik Prchlík

Tereza Lukschová

Klaudia Balážová

**Téma:**

Analýza Youtube kanálov a trend-videí

**Zdroj dát:**

Plánované využitie troch datasetov z platformy Kaggle a z oficiálnych štatistických úradov vybraných štátov.

**Zaujímavé charakteristiky:**

Údaje o trending videách, zmenách v čase a top kanáloch môže byť dôležité pre nových alebo súčasných tvorcov video obsahu na platforme Youtube. Ponúka im to možnosť sa inšpirovať a naďalej zlepšovať.

**Štruktúra datasetov:**  
<https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset?select=GB_youtube_trending_data.csv>

CSV (dáta o videách) a JSON (definované kategórie) pre každú krajinu

Ide o dataset, ktorý denne zaznamenáva top trend videá na platforme Youtube. Jeden záznam predstavuje video nachádzajúce sa v top trending v konkrétny deň, tým pádom jedno video sa môže vyskytovať v datasete niekoľko krát, či už v neprerušenom intervale času alebo aj prerušenom (znovu sa dostal do trending po nejakom čase). K dnešnému dňu 26.10.2022 obsahuje 29 720 záznamov.

Minimum 12Aug20

Maximum 26Oct22 – aktualizovaný denne

video\_id: String (J78aPJ3VyNs)

title: String (I left youtube for a month and THIS is what happened.)

publishedAt: datetime (2020-08-11T16:34:06Z)

channelId: String (UCYzPXprvl5Y-Sf0g4vX-m6g)

channelTitle: String (jacksepticeye)

categoryId: Int (24)

trending\_date: datetime (2020-08-12T00:00:00Z)

tags: String (jacksepticeye|funny|funny meme|memes|jacksepticeye memes|reddit|subreddit|community|community memes|...)

view\_count: Int (2038853)

likes: Int (353790)

dislikes: Int (2628)

comment\_count: Int (40228)

thumbnail\_link: String (<https://i.ytimg.com/vi/J78aPJ3VyNs/default.jpg>)

comments\_disabled: Boolean (False)

ratings\_disabled: Boolean (False)

description: String (I left youtube for a month and this is what happenedMY COFFEE COMPANY: https://twitter.com/TOTMCoffe...)

<https://www.kaggle.com/datasets/harshithgupta/youtubes-channels-dataset>

JSON

Nie je dobre popísaný, no ide o dataset o kanáloch na základe trending videí platformy Youtube. Jeden záznam by mal predstavovať jedno video s údajmi o kanále, ktorý ho zverejnil. Z toho vyplýva, že sa v ňom nachádza duplikované popisy o kanáloch, avšak menené v čase, čo môže byť tiež použité v ďalšej analýze podľa toho za aké obdobie je trending video sledované. Cieľom využitia tohto datasetu je získať dodatočné dáta o kanáloch a dodatočné jednoduché štatistiky, ktoré by sme prepojili s trending videami. Obsahuje približne 11 500 záznamov, v súčasnej dobu bez znalosti za aký časový interval (zistíme z atributu videoPublished).

2020

"totalviews/channelelapsedtime":string"0.16519925596076884"

"channelId":string"UCdzU3DSGzyWzN2118yd9X9g"

"videoCategoryId":string"22"

"channelViewCount":string"14654"

"likes/subscriber":string"0.5555555555555556"

"views/subscribers":string"95.11111111111111"

"videoCount":string"30"

"subscriberCount":string"18"

"videoId":string"--DwgB78t-c"

"dislikes/views":string"0.0005841121495327102"

"channelelapsedtime":string"88705"

"comments/subscriber":string"0.0"

"likes/views":string"0.005841121495327103"

"channelCommentCount":string"1"

"videoViewCount":string"1712"

"likes/dislikes":string"-2"

"comments/views":string"0.0"

"totvideos/videocount":string"488.46666666666664"

"elapsedtime":string"50040"

"videoLikeCount":string"10"

"videoDislikeCount":string"1"

"dislikes/subscriber":string"0.05555555555555555"

"totviews/totsubs":string"814.1111111111111"

"views/elapsedtime":string"0.03421262989608313"

"videoPublished":string"2012-01-19T18:38:28Z"

"VideoCommentCount":string"0"

<https://www.kaggle.com/datasets/syedjaferk/top-1000-youtubers-cleaned>

CSV

Dataset o top 1 000 influenseroch (kanáloch) po celom svete na platforme Youtube. Jeden záznam predstavuje jeden kanál. Rank kanálu je pravdepodobne určený len na základe počtu odberateľov, no obsahuje informácie napríklad o kategórii videí, ktorými si kanál zaoberá, ale aj údaj o prevažujúcej krajine publika. Každopádne, dataset bol vytvorený jednorazovo v nejakom určitom čase, ale v akom je možné určiť iba na základe dátumu zverejnenia datasetu.

2022 Júl

Rank: Int - Overall rank of the account (1)

Username: String - username/handle of the account (tseries)

Youtube Url: String - URL like for the account (http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy4\_brA)

Name: String - Name of the accout holder. (T-Series)

Category: String - Category of the account (Music, Entertainment, etc…)

Subscribers: String - Total number of subscribers. (220.1M)

Audience Country: String - Country of the Majority of the audience. (India)

Avg. Views: String - Average views of the content. (52.7K)

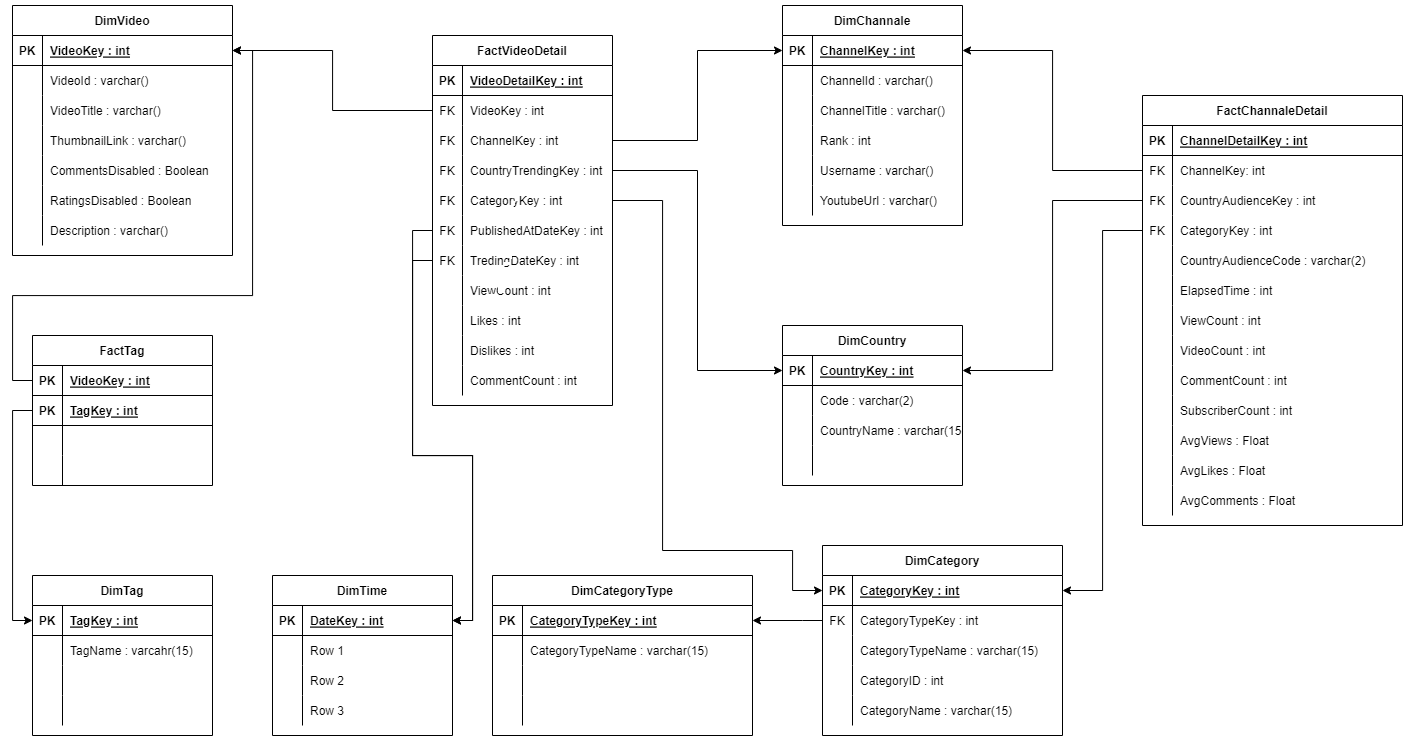
Avg. Likes: String - Average likes of the videos/content. (1.7K)

Avg. Comments: String - Average Comments of the video/content (108)

**KPI:**

1. Obľúbenosť kategórií v závislosti na krajine.
2. Obľúbenosť kategórií v závislosti na vekových kategóriách vybraných zemí\*
3. Počet trendových videí v rámci jedného kanálu pre každý mesiac.
4. Počet kanálov vyskytovaných sa v trend-videách, ktoré sa nenachádzajú v datasete TOP 1000 kanáloch.
5. Počet trend-videí kanálov na základe počtu odberateľov.
6. Zhoda publika TOP kanálov s výskytom v trendings.
7. Priemerná dĺžka zotrvania videa v trendings.

\*je to zaujimavé?



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **DimVideo** |  |  |  |  |  | | |
| **Key** | Name | Data Type | Null | Attributes | Description | | |
| **1.PK** | VideoKey | int |  | identity | VideoKey. Primary key for table DimVideo | | |
| **2.** | VideoId | varchar() |  |  | The Id of the video. (' | J78aPJ3VyNs | ') |
| **3.** | VideoTitle | varchar() |  |  | Title of video. (' | I left youtube for a month and THIS is what happened. | ') |
| **4.** | ThumnailLink | varchar() |  |  | Youtube link. (' | <https://i.ytimg.com/vi/J78aPJ3VyNs/default.jpg> | ') |
| **5.** | CommentsDisabled | Boolean |  |  | Video have disabled comments. (' | FALSE | ') |
| **6** | RatingsDisabled | Boolean | ok |  | Video have disabled ratings. (' | FALSE | ') |
| **7** | Description | varchar() | ok |  | Description of video. (' | I left youtube for a month and this is what happenedMY | ') |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **DimTag** |  |  |  |  |  | | |
| **Key** | Name | Data Type | Null | Attributes | Description | | |
| **1.PK** | TagKey | int |  | identity | TagKey. Primary key for table DimTag | | |
| **2.** | TagName | varchar(15) |  |  | Name of tag. (‘ | funny meme | ') |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **DimDate** |  |  |  |  |  | | |
| **Key** | Name | Data Type | Null | Attributes | Description | | |
| **1.PK** | DateKey | int |  | identity | DateKey. Primary key for table DimDate | | |
| **2.** |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **DimCategoryType** |  |  |  |  |  | | |
| **Key** | Name | Data Type | Null | Attributes | Description | | |
| **1.PK** | CategoryTypeKey | int |  | identity | CategoryTypeKey. Primary key for table DimCategoryType | | |
| **2.** | CategoryTypeName | varchar(15) |  |  | Category type of dimension. (‘ | Video | ') |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **DimCategory** |  |  |  |  |  | | |
| **Key** | Name | Data Type | Null | Attributes | Description | | |
| **1.PK** | CategoryKey | int |  | identity | CategoryKey. Primary key for table DimCategory | | |
| **2. FK** | CategoryTypeKey | int |  |  | Key to CategoryType. (' | 1 | ') |
| **3.** | CategoryId | int |  |  | Id of category from source dataset. (' | 1 | ') |
| **4.** | CategoryName | varcgar(15) |  |  | Category name of video or channel. (' | Channel | ') |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **DimChannel** |  |  |  |  |  | | |
| **Key** | Name | Data Type | Null | Attributes | Description | | |
| **1.PK** | ChannelKey | int |  | identity | ChanelKey. Primary key for table DimChannel | | |
| **2.** | ChannelId | float |  |  | The Id of the channel. (‘ | UCdzU3DSGzyWzN2118yd9X9g | ') |
| **3.** | ChannelTitle | varchar(30) |  |  | Title of channel. (' | jacksepticeye. | ') |
| **4.** | Rank | int |  |  | Overall rank of the account. (' | 1 | ') |
| **5.** | Username | varchar(15) |  |  | Username/handle of the account. (' | tseries | ') |
| **6.** | YoutuberUrl | varchar() |  |  | URL like for the account (' | <http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy4_brA> | ') |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DimCountry** |  | |  | | |  | |  | |  | | |
| **Key** | Name | | Data Type | | | Null | | Attributes | | Description | | |
| **1.PK** | CountryKey | | int | | |  | | identity | | CountryKey. Primary key for table DimCountry | | |
| **2.** | Code | | varchar(2) | | |  | |  | | Code of country. (' | GB | ') |
| **3.** | CountryName | | varchar(15) | | |  | |  | | Whole name of country. (' | Great Britain | ') |
| **FactTag** | |  | |  |  | |  | |  | | | |
| **Key** | | Name | | Data Type | Null | | Attributes | | Description | | | |
| **1.PK** | | VideoKey | | int |  | |  | | Key to connect tag with video. | | | |
| **2. PK** | | TagKey | | int |  | |  | | Key to connect video with tag. | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **FactVideoDetail** |  |  |  |  |  | | |
| **Key** | Name | Data Type | Null | Attributes | Description | | |
| **1.PK** | VideoDetailKey | int |  | identity | VideoDetailKey. Primary key for table FactVideoDetail | | |
| **2. FK** | VideoKey | int |  |  | The key of the video. (' | J78aPJ3VyNs | ') |
| **3. FK** | ChannelKey | int |  |  | The key of the channel. (' | 1 | ') |
| **4. FK** | CountryTrendingKey | int |  |  | Key of country where video was in trendings. (' | 1 | ') |
| **5. FK** | CategoryKey | int |  |  | The key of the category. (' | 1 | ') |
| **6 FK** | PublishedAtDateKey | int |  |  | Key for video published date. (' | 1 | ') |
| **7 FK** | TendingDateKey | int |  |  | Key for date in which video was trending. (' | 1 | ') |
| **8.** | ViewCount | int | ok |  | Number of video views. (' | 100 | ') |
| **9.** | Likes | int | ok |  | Number of likes. (' | 100 | ') |
| **10.** | Dislikes | int | ok |  | Number of dislikes. (' | 578 | ') |
| **11.** | CommentCount | int | ok |  | Numer of comments. (' | 23456 | ') |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **FactChannelDetail** |  |  |  |  |  | | |
| **Key** | Name | Data Type | Null | Attributes | Description | | |
| **1.PK** | ChannelDetailKey | int |  | identity | VideoDetailKey. Primary key for table FactChannelDetail | | |
| **2. FK** | ChannelKey | int |  |  | The key of the channel. (' | 1 | ') |
| **3. FK** | ChannelAudienceKey | int |  |  | Key of channel main audience country. (' | 1 | ') |
| **4. FK** | CategoryKey | int |  |  | The key of the category. (' | 1 | ') |
| **5.** | CountryAudienceCode | varchar(2) |  |  | Code of channel main audience country. (' | GB | ') |
| **6.** | ElapsedTime | int | ok |  | Chanel elapsed time. (' | 356654 | ') |
| **7.** | ViewCount | int | ok |  | Number of likes. (' | 100 | ') |
| **8.** | VideoCount | int | ok |  | Number of dislikes. (' | 578 | ') |
| **9.** | CommentCount | int | ok |  | Numer of comments. (' | 23456 | ') |
| **10.** | SubscriberCount | int | ok |  | Total number of subscribers. (' | 220.1M | ') |
| **11.** | AvgViews | float | ok |  | Average views of the content. (' | 57K | ') |
| **12.** | AvgLikes | float | ok |  | Average likes of the videos/content. (' | 1.7K | ') |
| **13.** | AvgComments | float | ok |  | Average Comments of the video/content. (' | 108 | ') |