

# Analyzing Youtube Trends

Extracting Insights From Youtube Analytics Data

---

Satori

Data Scientist and Researcher

# Group Members

---

**20127214**

Nguyễn Trương Minh Khôi

**20127444**

Bùi Duy Bảo

**20127514**

Trần Thị Bảo Hương

**20127606**

Nguyễn Vinh Quang

**20127612**

Đỗ Khánh Sang

# Why YouTube?

As this is a large, influential social networking platform and can produce good income. Our project crawls data from YouTube to analyze and predict trending videos through API usage.



# Nội dung

**O1 Thu thập dữ liệu**

**O2 Tổ chức dữ liệu – Pipeline**

**O3 Phân cấp**

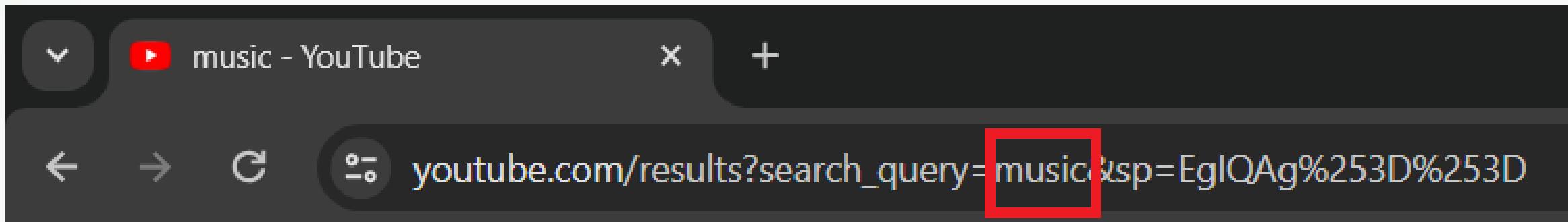
**O4 Phân tích**

**O5 Dashboard**

**O6 Đánh giá**

# 1. Thu thập dữ liệu

- Sử dụng đường link dưới đây để lấy list channel tương ứng với keyword tìm kiếm



# 1. Thu thập dữ liệu

List các channel tương ứng với music



ZayChik Sama

@MeiMusicRemix • 21.2K subscribers

Chào Mừng Mọi Người Đã Đến Kênh Của Minh

Subscribe



Wanji Music

@Wanjimusic • 150K subscribers

Đây là kênh bắt trend nhạc tiktok douyin siêu nhanh. Lạc vào kênh của tôi, bạn sẽ bị mê hoặc bởi những bản nhạc cực hay, siêu ...

Subscribe



Nam Việt Music ✅

@NamVietMusic • 3.71M subscribers

Nam Việt Kênh Youtube Giải trí - Âm nhạc - Hài kịch phong phú nhất Việt Nam hiện nay. ↗ Nam Việt Music phát hành nhiều sản ...

Subscribe



Music Attracts Wealth

@MusicAttractsWealth • 128K subscribers

Music Attracts Wealth I am Phupham and I compose and make music videos in with different frequencies, Natural Frequency, ...

Subscribe

# 1. Thu thập dữ liệu

List các channel tương ứng với podcast

	<p><b>Next Level Soul Podcast</b> ✓</p> <p>@NextLevelSoul • 805K subscribers</p> <p>Next Level Soul founder Alex Ferrari is a #1 best-selling author, podcaster, speaker, entrepreneur, &amp; filmmaker. Throughout his ...</p> <p><b>Subscribe</b></p>
	<p><b>Lore Podcast</b> ✓</p> <p>@lorepodcast • 19.7K subscribers</p> <p>The official YouTube channel for Lore, the award-winning podcast about true life scary stories. Lore exposes the darker side of ...</p> <p><b>Subscribe</b></p>
	<p><b>Bloomberg Podcasts</b> ✓</p> <p>@BloombergPodcasts • 37.9K subscribers</p> <p>Official home of Bloomberg Podcasts. Visit <a href="https://www.bloomberg.com/podcasts">https://www.bloomberg.com/podcasts</a>.</p> <p><b>Subscribe</b></p>
	<p><b>OfflineTV Podcast</b> ✓</p> <p>@OfflineTPodcast • 592K subscribers</p> <p>We're OfflineTV! We are a group of content creators who got our start in gaming. Now we live in a house together and make cool ...</p> <p><b>Subscribe</b></p>

# 1. Thu thập dữ liệu

31

Video  
Files

6

Channel  
Files

# 1. Thu thập dữ liệu

-Up to now we collected total 37 csv files that contain data about videos and channel on Youtube

-Each file approximately 9000 rows of data

-The meaning of each column:

Channel ID: ID of channel corresponding to video

Channel name: name of channel corresponding to video

Topic: the topic of video

Like: number of likes

Video ID: ID of video

Title: title of video

Duration: duration of video

Num Comments: number of comments

Published At: time video got published

View: number of viewers

Date: the date that data was collected

# 1. Thu thập dữ liệu

#	Column	Non-Null Count	Dtype
0	Channel ID	7447 non-null	object
1	Channel Name	7447 non-null	object
2	Video ID	7447 non-null	object
3	Title	7447 non-null	object
4	Topic	7447 non-null	object
5	Duration	7447 non-null	int64
6	Published At	7447 non-null	object
7	View	7447 non-null	int64
8	Like	7447 non-null	int64
9	Num Comments	7447 non-null	int64
10	Date	7447 non-null	object
11	Description	7009 non-null	object
12	Tag	7447 non-null	object

#	Column	Non-Null Count	Dtype
0	Channel ID	184 non-null	object
1	Channel Name	184 non-null	object
2	Description	168 non-null	object
3	View	184 non-null	int64
4	Num Subscriber	184 non-null	int64
5	Topic	184 non-null	object
6	Num Video	184 non-null	int64

# 1. Thu thập dữ liệu

Dataframe chứa thông tin của các channel

	Channel ID	Channel Name	Description	View	Num Subscriber	Topic	Num Video
0	UCBq5hp6DvwWam2XOXw6lYJA	Awakening Music	Awakening Music is a pioneering and dynamic re...	5820572933	13900000	[Music]	897
1	UC-56FOaYhovgtPBFgNPRXEg	TOSHI MUSIC	Artist. Performer. Storyteller.	8134	459	[Hip_hop_music, Music_of_Asia, Music, Pop_music]	10
2	UCeGJ6v6KQt0s88hGKMfybuw	Cozy Jazz Music	"Welcome to Cozy Jazz"\n\nWe hope our music ...	37490656	108000	[Music, Jazz, Music_of_Latin_America]	987
3	UCHJpwujMeHBaUJuRBxDx5Nw	ZinK Music ツ	Đây Là Kênh Chính Thức Của ZinK Music ツ Nha!\n...	12591633	35300	[Pop_music, Music, Electronic_music]	89
4	UCZh0PiKTUoxL_GnJs6KITsQ	MUSIC official		7337608	38100	[Electronic_music, Music]	8
...	...	...	...	...	...	...	...

# 1. Thu thập dữ liệu

Dataframe chứa thông tin của các video

Channel ID	Channel Name	Video ID	Title	Topic	Duration	Published At	View	Like	Num Comments	Date	Description	Tag
UCBq5hp6DvwWam2XOXw6lYJA	Awakening Music	HCRZbxqU	Maher Zain - Sanahya Keraman - "Khawater 7"  ...	[Music]	0	2011-08-01T12:38:23Z	7067710	44646	1361	03/03/2024	Visit www.exploringislam.com/en/n-Maher Zain ...	سنحيا كراما، خواطر، أغنية حواطر، أغاني خواطر،...
UCBq5hp6DvwWam2XOXw6lYJA	Awakening Music	m0MqT-S9fil	Raef - It's Jumuah [Friday]   (Rebecca Black C...	[Music, Pop_music]	0	2011-12-11T12:01:07Z	3878422	34103	3075	03/03/2024	Visit www.exploringislam.com/en/n-Out Now: Ra...	[Raef, Jummah, Jomaah, Jumah, Friday, Rebecca ...]
UCBq5hp6DvwWam2XOXw6lYJA	Awakening Music	nY-19pWQJZ8	Raef - Alhamdu Lillah (Official)	[Christian_music, Music, Pop_music]	0	2019-01-13T11:57:43Z	613597	13986	698	03/03/2024	Visit www.exploringislam.com/en/n-Official mu...	[Alhamdu Lillah, Raef, alhamdulillah, alhamdol...
UCBq5hp6DvwWam2XOXw6lYJA	Awakening Music	iRhy-ijrFw	Maher Zain & Humood - Tahayya   Behind The Sce...	[Music]	0	2023-01-06T09:00:23Z	680270	11145	355	03/03/2024	Behind the scenes of the filming of Maher Zain...	[world cup song, maher zain, humood, ماهر زين,...]
UCBq5hp6DvwWam2XOXw6lYJA	Awakening Music	KnKyMP-6y8o	#palestine tomorrow will be free	[Religion]	0	2022-08-10T08:00:06Z	63529	4718	136	03/03/2024	Visit exploringislam.com/en/n-	None
...	...	...	...	...	...	...	...	...	...	...	...	...
UCjTp-nBKswYLumqmVeBPwYw	The Jeff Nippard Podcast	yTzTjKxx098	Gender Differences in Training feat. Menno Hen...	[Health, Lifestyle_(sociology)]	0	2017-04-13T12:44:42Z	14561	442	33	03/03/2024	► SUBSCRIBE TO MY PODCAST CHANNEL → https://bi...	[podcast, icecream4prs, science, fitness, body...]

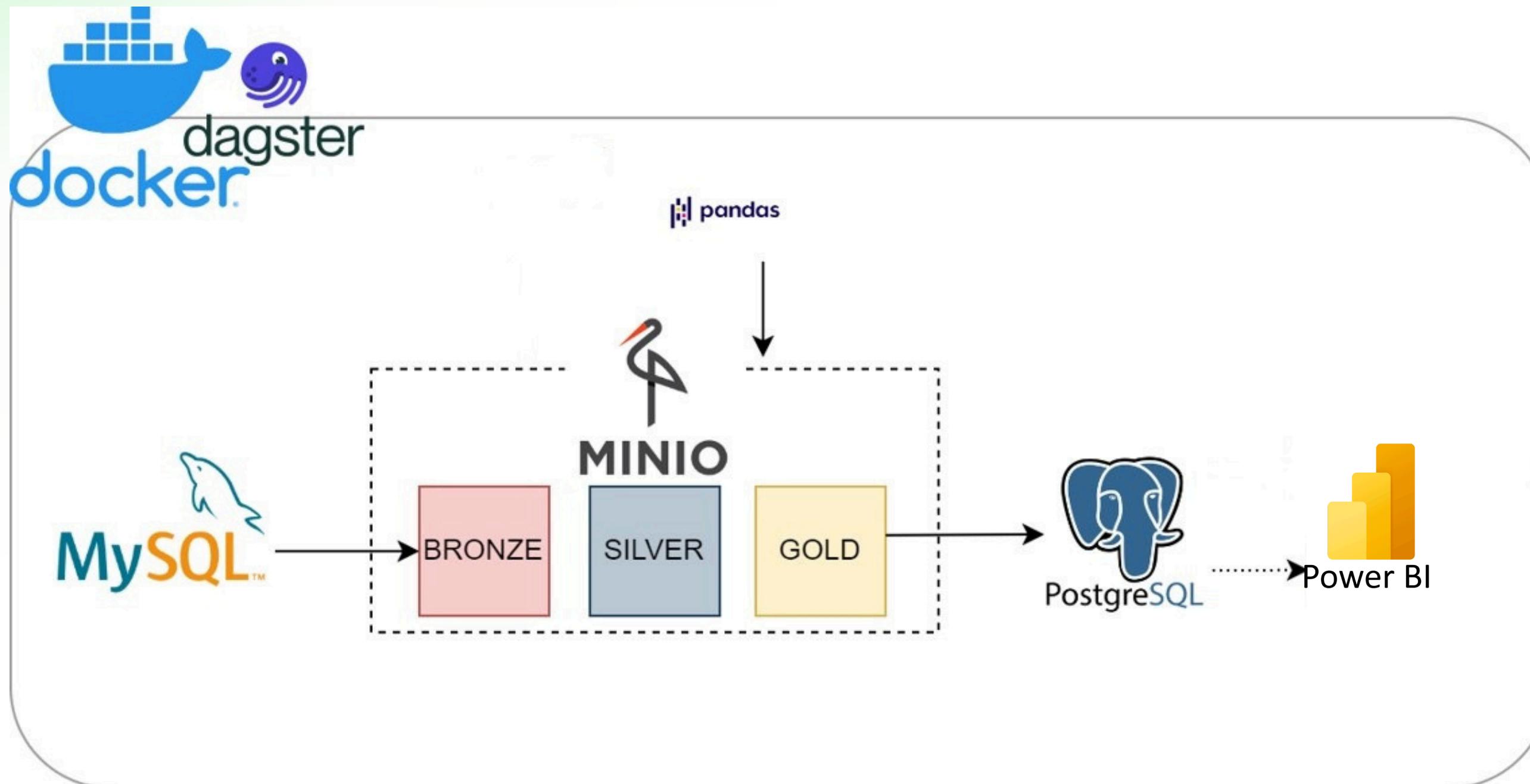
# 1.1 Các công cụ được sử dụng

Tool in use:

- Crawling: Python - BeautifulSoup, Selenium  
Sử dụng Docker để đóng gói ứng dụng và Dagster để orchestrate assets (theo định nghĩa của Dagster)
- MySQL: Dữ liệu ban đầu có định dạng .csv sẽ được nạp vào cơ sở dữ liệu MySQL
- MinIO: Xây dựng nền tảng dữ liệu (Data Lake), trong đó:
  - +Bronze: Dữ liệu thô được trích xuất từ MySQL
  - +Silver: Làm sạch dữ liệu
  - +Gold : Dữ liệu dùng để phân tích, sẽ được nạp vào PostgreSQL (Data Warehouse)
- PostgreSQL: Data Warehouse
- PowerBI: Dùng cho mục đích trực quan dữ liệu
- Pandas: Chuyển đổi dữ liệu ở lớp Bronze, Silver và Gold

# 2. Pipeline

Xây dựng pipeline lưu trữ dữ liệu



# 2. Pipeline

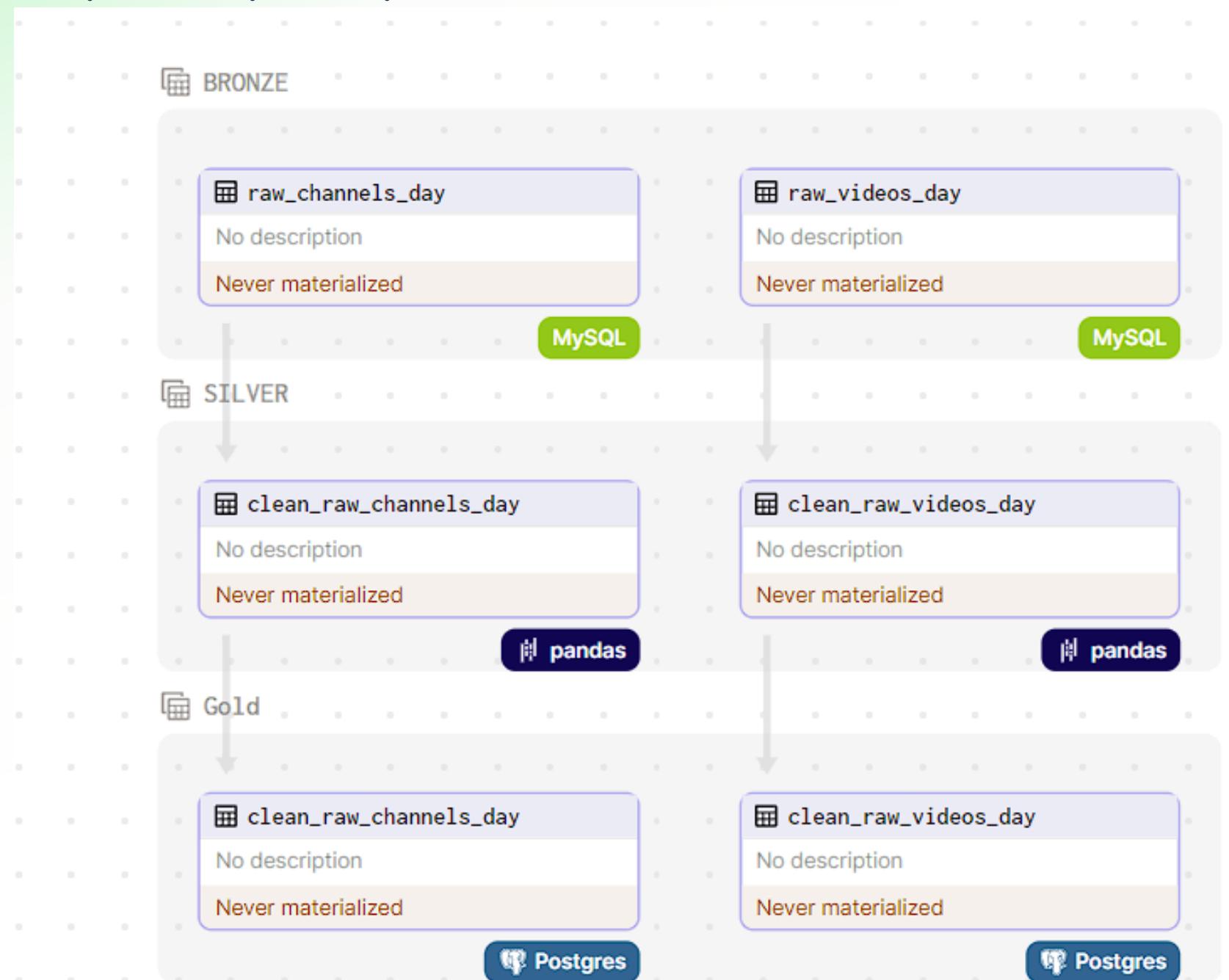
Dữ liệu crawl về được lưu trong MySQL

ChannelID	ChannelName	Description	View	NumSubscriber	Topic	NumVideo
UCIdW2PnzoVAfq2o-K9vPAYA	Deep Tune Music	default_value	9454057	13900	Christian_music, Electronic_music, Po...	76
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	Orinn Music LÃ Ä»i nqÅ© sáº£n xuá... - Dear Artists, Producers and Labels - ...	28878383	83600	Electronic_music, Music, Pop_music	461
UC7HwMQwUXwpw5RnmOS0Fmzw	Music Of Wibu	CÃ’ng ty truyá»n thÃ’ ng giáº£i trÃ- ...	755547	3360	Electronic_music, Music_of_Asia, Pop...	230
UCpHhPOfjPC1Z5MHTU_3mVg	HL Media Music	CÃ’ng ty truyá»n thÃ’ ng giáº£i trÃ- ...	13039629	21300	Music, Music_of_Asia	559
UCiq69t2ThqzU0NQ8yXcfAZA	Yen Music	CHÃ©O MÃ»NG CÃ¢C BÃº N ÃÄN VÃ»I ...	31940889	277000	Pop_music, Music, Electronic_music	631
UCQGxATbsTxI-rwWsDNYFVKq	Music Passionate	welcome to my channel where you c...	1774421	33800	Pop_music, Music, Jazz	111
UCETrW0knlc6rrvlu07SxEq	Fendi Music	Fendi Music LÃ KÃ¢nh Ãm Nháº£ic Cá... WM MEDIA lÃ cÃ’ng ty sáº£n xuáº¥...	22922707	97300	Electronic_music, Music	261
UChlcwJLWLQ26mDTUwjoDbjq	WM Music	WM MEDIA lÃ cÃ’ng ty sáº£n xuáº¥...	17209995	35100	Electronic_music, Music, Pop_music	64
UCtVt72Xq0qMvRSfyDDiqHhw	MEGA GS MUSIC	ChÆ°Æing trÃ¬nh âSÃ n Chiáº£n Gi...	982982628	1550000	Television_program, Entertainment, ...	603
UCOxqqCwgOqc2lMqC5PYz_Dq	Chillhop Music	Laidback beats for a more relaxed ev...	614244256	3280000	Lifestyle_(sociology), Music	769
UC_c26yCvit6WhOKfBp_84ew	Via»t Hiphop Co...	default_value	3125265	11800	Hip_hop_music, Music	2043
UC7anZcjCFXhot9dpN2DxBew	Retro Music India	Retro Music India, a distinguished mu...	5995509	1100000	Music_of_Asia, Music	480
UCRmTtUMRGr2uC-c6L3r90NQ	Miu Music	CÃ’ng ty TNHH BD Media lÃ cÃ’ng t...	97877302	300000	Pop_music, Music_of_Asia, Electronic...	1264
UCNGZ8S8G6cN0z17IoiqJPCw	Tennessee Tech...	This is the official Recording and Lives...	46338	794	Christian_music, Lifestyle_(sociology)...	187
UCe8UjG9ecgPmIDqRvFsEW4q	Thiáº£u Nhi MUS...	â Thiáº£u Nhi MUSIC LÃ kÃ¢nh Ã¢m ...	426943163	693000	Film, Music, Entertainment	37
UC5xi7kLHCZE6Osk5hPA2lHQ	RIN Music	RIN MUSIC thuá»c RIN Media lÃ Ä»á»i ...	217814511	685000	Pop_music, Music, Electronic_music	515
UCa0nKWXJ8ab_BVaPaxDxk_a	JW Music â¤	wanna listen to some tunes ? watch a...	27782062	66700	Independent_music. Music. Pop_music	28

ChannelID	ChannelName	VideoID	Title	Topic	Duration	PublishedAt	Views	Likes	NumComments	Date	^
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	5Fa-Ex2lXks	VIET DEEP 2023 - NHẠC HOUSE LAK ...	Electronic_music, Music, Pop_music	7893.0	2023-12-13T11:00:03Z	23949	184.0	7.0	02/01/20:	<
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	Jxx_co_m6M	VIET DEEP 2023 - NHẠC HOUSE LAK ...	Electronic_music, Music	8004.0	2023-12-03T11:00:22Z	25257	161.0	5.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	O25DqYbLXqY	VIET DEEP 2023 - NHẠC HOUSE LAK ...	Electronic_music, Music, Pop_music	8593.0	2023-10-30T11:00:30Z	220130	1386.0	44.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	CIU86jdO0_q	VIET DEEP 2024 - SIÊU PHẨM PHẢI C...	Electronic_music, Music	7290.0	2024-01-17T11:00:10Z	3645	31.0	2.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	S6Z390onmio	VIET DEEP 2023 - MIXTAPE KHOÂNG ...	Electronic_music, Music	7350.0	2023-12-11T11:00:30Z	11035	63.0	2.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	YDcjJtYqQLo	VIET DEEP 2023 - MIXTAPE NHẠC N...	Electronic_music, Music, Pop_music	7548.0	2023-11-26T11:00:24Z	49537	291.0	13.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	cW9c1T45sjU	CHILL BASS 2023 - NHẠC DEEP HOU...	Electronic_music, Music, Pop_music	7298.0	2023-09-18T11:00:28Z	71145	385.0	17.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	S1cQHmi8MmA	VIET DEEP 2023 - MIXTAPE VÀI LẦN ...	Electronic_music, Music, Pop_music	7296.0	2023-08-24T11:00:16Z	128777	630.0	24.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	R9ONeRxndU	VIET DEEP 2023 - MIXTAPE KHÓC R...	Electronic_music, Music	7499.0	2023-11-29T11:00:31Z	12580	66.0	8.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	7OyUhb0gfiQ	CHILL BASS 2024 - NHẠC DEEP HOU...	Electronic_music, Music	7714.0	2024-01-21T11:00:11Z	34133	251.0	14.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	Y47FcrlCaQk	VIET DEEP 2024 - MIXTAPE NHƯ VẬY...	Electronic_music, Music, Pop_music	7484.0	2024-01-20T11:00:26Z	8283	70.0	6.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	uYfV-0I9Nvk	VIET DEEP 2023 - NHẠC HOUSE LAK ...	Electronic_music, Music	7773.0	2023-11-07T11:00:48Z	63310	316.0	16.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	j7oA6OCdudA	CHILL BASS 2023 - NHẠC DEEP HOU...	Electronic_music, Music, Pop_music	3895.0	2023-06-24T11:00:28Z	144145	772.0	36.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	FpQpDL7EQIA	VIET DEEP 2023 - NHẠC HOUSE LAK ...	Electronic_music, Music	3861.0	2023-07-25T11:00:12Z	518285	3531.0	63.0	02/01/20:	
UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	tnfDnhy3x40	CHILL BASS 2023 - NHẠC DEEP HOU...	Electronic_music, Music, Pop_music	6862.0	2023-11-22T11:00:51Z	38078	225.0	9.0	02/01/20:	
UCKCzhWD9rOF74RhOnWXF19w	Deluxe Music	nDnS-l7r730	CHILL BASS 2023 - NHẠC DEEP HOU...	Electronic_music, Music, Pop_music	8621.0	2023-08-14T11:00:40Z	197658	1015.0	64.0	02/01/20:	>

# 2.Pipeline

- Các asset đại diện cho từng lớp
- Mỗi asset có một nhiệm vụ khác nhau



# 2. Pipeline

```
z usages
@asset(
    io_manager_key="minio_io_manager",
    required_resource_keys={"mysql_io_manager"},
    key_prefix=["bronze", "ecom"],
    group_name="BRONZE",
    compute_kind="MySQL"
)
def raw_videos_day(context) -> Output[pd.DataFrame]:
    sql_stm = "SELECT * FROM videos_day"
    pd_data = context.resources.mysql_io_manager.extract_data(sql_stm)
    return Output(
        pd_data,
        metadata={
            "table": "videos",
            "records count": len(pd_data),
        },
    )
```

```
z usages
@asset(
    io_manager_key="minio_io_manager",
    required_resource_keys={"mysql_io_manager"},
    key_prefix=["bronze", "ecom"],
    group_name="BRONZE",
    compute_kind="MySQL"
)
def raw_channels_day(context) -> Output[pd.DataFrame]:
    sql_stm = "SELECT * FROM channels_day"
    pd_data = context.resources.mysql_io_manager.extract_data(sql_stm)
    return Output(
        pd_data,
        metadata={
            "table": "channels",
            "records count": len(pd_data),
        },
    )
```

# 2. Pipeline

```
@asset(
    ins={
        "raw_videos_day": AssetIn(key_prefix=["bronze", "ecom"]),
    },
    key_prefix=["silver", "ecom"],
    io_manager_key="minio_io_manager",
    required_resource_keys={"minio_io_manager"},
    group_name="SILVER",
    compute_kind="Pandas"
)
def clean_raw_videos_day(raw_videos_day: pd.DataFrame) -> Output[pd.DataFrame]:
    pd_data = raw_videos_day
    pd_data['Description'].fillna('No Description', inplace=True)
    pd_data['Topic'].fillna('No Topic', inplace=True)
    pd_data['Topic'] = pd_data['Topic'].replace('None', 'No Topic')
    pd_data['Likes'].fillna('0', inplace=True)
    pd_data['Likes'] = pd_data['Likes'].replace('None', '0')
    pd_data['NumComments'].fillna('0', inplace=True)
    pd_data['NumComments'] = pd_data['NumComments'].replace('None', '0')
    pd_data['Tag'].fillna('No Tag', inplace=True)
    pd_data['Tag'] = pd_data['Tag'].replace('None', 'No Tag')

    return Output(
        pd_data,
        metadata={
            "table": "videos",
            "records counts": len(pd_data),
        },
    )
```

```
@asset(
    ins={
        "raw_channels_day": AssetIn(key_prefix=["bronze", "ecom"]),
    },
    key_prefix=["silver", "ecom"],
    io_manager_key="minio_io_manager",
    required_resource_keys={"minio_io_manager"},
    group_name="SILVER",
    compute_kind="Pandas"
)
def clean_raw_channels_day(raw_channels_day: pd.DataFrame) -> Output[pd.DataFrame]:
    pd_data = raw_channels_day
    pd_data['Description'].fillna('No Description', inplace=True)
    pd_data['Description'] = pd_data['Description'].replace('default_value', 'No Description')
    pd_data['Description'] = pd_data['Description'].replace("\n", "No Description")
    return Output(
        pd_data,
        metadata={
            "table": "channels",
            "records counts": len(pd_data),
        },
    )
```

# 2. Pipeline

```
def dwh_channels_day(clean_raw_channels_day) -> Output[pd.DataFrame]:  
    return Output(  
        clean_raw_channels_day,  
        metadata={  
            "schema": "public",  
            "table": "clean_raw_channels_day",  
            "records counts": len(clean_raw_channels_day),  
        },  
    )
```

```
def dwh_videos_day(clean_raw_videos_day) -> Output[pd.DataFrame]:  
    return Output(  
        clean_raw_videos_day,  
        metadata={  
            "schema": "public",  
            "table": "clean_raw_channels_day",  
            "records counts": len(clean_raw_videos_day),  
        },  
    )
```

# 2. Pipeline

- Dữ liệu lưu trên Minio

warehouse / ecom		
Name	Last Modified	Size
AssetKey(['bronze', 'ecom', 'raw_channels_day']).csv	Sat, Mar 30 2024 02:03 (GMT+7)	658.5 KiB
AssetKey(['bronze', 'ecom', 'raw_videos_day']).csv	Sat, Mar 30 2024 02:04 (GMT+7)	309.0 MiB
AssetKey(['silver', 'ecom', 'clean_raw_channels_day']).csv	Sat, Mar 30 2024 02:03 (GMT+7)	663.1 KiB
AssetKey(['silver', 'ecom', 'clean_raw_videos_day']).csv	Sat, Mar 30 2024 02:04 (GMT+7)	312.4 MiB

# 2. Pipeline

Dữ liệu thu thập được đưa vào PostgreSQL

select * from public.clean_raw_channels_day						
	ABC ChannelID	ABC ChannelName	ABC Description	123 View	123 NumSubscriber	ABC Topic
1	UCIdW2PnzoVAfq2o-K9vPAYA	Deep Tune Music	No Description	9,454,057	13,900	Christian
2	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	Orinn Music IÃ© Áá»i ngÃ© sÃºn xuáºt nhá»ng dÃ¡»±	28,878,383	83,600	Electronic_music
3	UC7HwMQwUXwpw5RnmOS0Fmzw	Music Of Wibu	- Dear Artists, Producers and Labels - Music Of Wibu is	755,547	3,360	Electronic_music
4	UCpHhPOfjPC1Z5iVHTU_3mVg	HL Media Music	CÃ©ng ty truyá»n thÃ© giÃ¡c trÃ©nh nhÃºc Viá»t N	13,039,629	21,300	Music
5	UCiq69t2ThqzU0NQ8yXcfAZA	Yen Music	CHÃO MÃ»NG CÃC BÃºN ÃÃN VÃ»I KÃNH YEN MUSIC	31,940,889	277,000	Pop_music
6	UCQGxATbsTxI-rwWsDNYFVKg	Music Passionate	welcome to my channel where you can find the best Mu	1,774,421	33,800	Pop_music
7	UCETrW0knlcgm6rrvlu07SxEg	Fendi Music	Fendi Music LÃnh NhÃºc Cá»ša Sea Team!----	22,922,707	97,300	Electronic_music
8	UCHlcwJLWLQ26mDTUwoDbjg	WM Music	WM MEDIA IÃ© cÃ©ng ty sÃºn xuáºt vÃ© truyá»n thÃ©	17,209,995	35,100	Electronic_music
9	UctVt72Xq0qMvRSfyDDiqHhw	MEGA GS MUSIC	ChÃÆng trÃ¬nh âSÃ n Chiáºn Giá»ng HÃ¡jtâ (Singer	982,982,628	1,550,000	Television
10	UCOxqgCwgOqC2IMqC5PYz_Dg	Chillhop Music	Laidback beats for a more relaxed everyday life. Õ¹¹¹õ	614,244,256	3,280,000	Lifestyle
11	UC_c26yCvit6WhOKfBp_84ew	Viá»t Hiphop Contest	No Description	3,125,265	11,800	Hiphop
12	UC7anZcjCFXhot9dpN2DxBew	Retro Music India	Retro Music India, a distinguished music label, has swift	5,995,509	1,100,000	Music
13	UCRmTtUMRGr2uC-c6L3r90NQ	Miu Music	CÃ¢n tv TNHH BD Media IÃ© cÃ¢n tv hoáºt Ãn ph	97,877,302	300,000	Pop_music
14	UCmi6OO6sTxuZfmKX1q12x9Q	The Minimalists Podcast (Old Episod				
15	UCNGZ8S8G6cN0z17loiqJPCw	Tennessee Tech School of Music				
16	UCe8UjG9ecgPmIDgRvFsEW4g	Thiáºu NhÃ MUSIC				
ABC ChannelID						
	ABC ChannelID	ABC ChannelName	ABC VideoID	ABC Title	ABC Topic	123 Duration
1	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	sa6Ci124qYQ	VIET DEEP 2022 - MIXTAPE NÃI ĐÃU XÓT XA - NHÃAC	Electronic_music, Music, Pop_music	
2	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	zwhbwtvMPWY	VIET DEEP 2023 - MIXTAPE SET NHÃAC CÃU 9X - DEEP H	Electronic_music, Music, Pop_music	
3	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	4hIZ-ADm8-k	CHILL BASS 2024 - NHÃAC DEEP HOUSE REMIX 2024 HA	Electronic_music, Music, Pop_music	
4	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	5Fa-Ex2IXKs	VIET DEEP 2023 - NHÃAC HOUSE LAK REMIX 2023 HAY	Electronic_music, Music, Pop_music	
5	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	Jxx_co_m6M	VIET DEEP 2023 - NHÃAC HOUSE LAK REMIX 2023 HAY	Electronic_music, Music	
6	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	O25DqYbLXqY	VIET DEEP 2023 - NHÃAC HOUSE LAK REMIX 2023 HAY	Electronic_music, Music, Pop_music	
7	UC2fu6CiFfNYz5UFORvFyc0w	Vie Channel - MUSIC	oQ3P6fcQnv4	CÙng nắm tay nhau và vào giai điệu này #shorts	Hip_hop_music, Music, Music_of_Asia	
8	UC2fu6CiFfNYz5UFORvFyc0w	Vie Channel - MUSIC	mjhjQbRmaZk	Chờ mãi nàng Mây cất giọng ca khúc này thôi#shorts	Music, Music_of_Asia, Pop_music	
9	UC9519glhmQLqbw5wxlypAPg	Atus Music	XXbD4mvR4xY	PARTY REMIX 2023 ↗ Mashups & Remixes of Popular	Electronic_music, Music, Pop_music	
10	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	CIU86Jd00_g	VIET DEEP 2024 - SIÊU PHÃM PHÃI CHIA TAY THÔI REM	Electronic_music, Music	
11	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	S6Z390onmio	VIET DEEP 2023 - MIXTAPE KHOÄNG CÁCH TÌNH YÊU F	Electronic_music, Music	
12	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	YDcjJtYgQLo	VIET DEEP 2023 - MIXTAPE NHÃAC NGHE TRÊN BAR CLI	Electronic_music, Music, Pop_music	
13	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	cW9c1T45sjU	CHILL BASS 2023 - NHÃAC DEEP HOUSE REMIX 2023 HA	Electronic_music, Music, Pop_music	
14	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	S1cQHmi8MmA	VIET DEEP 2023 - MIXTAPE VÀI LẦN ĐÓNA REMIX	Electronic_music, Music, Pop_music	
15	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	R9ONeRxndU	VIET DEEP 2023 - MIXTAPE KHÓC REMIX - LÃ CAY KHÓ	Electronic_music, Music	
16	UCW9o3ENwPXKd9gc3ffqRs9w	Wanji Music	sV3CTLxusYA	很任性 0.8X (DJ抖音版 2023) Thật Cố Chấp (Remix Tikto	Electronic_music, Music, Pop_music	

# 3. Phân cấp cho dữ liệu

Dữ liệu được chia thành 2 thể loại chính là Music và Podcast



Music



Podcast

### 3. Phân cấp cho dữ liệu

- Chia các channel thành 3 phân khúc dựa trên tỉ lệ giữa số lượng subscriber của channel và thời gian mà channel tồn tại
- Tạo thêm 1 cột Days tương ứng với thời gian tồn tại của channel
- Tạo thêm 1 cột Ratio tương ứng với tỉ lệ subscriber/days



# 3. Phân cấp cho dữ liệu

	Channel ID	Channel Name	Description	View	Num Subscriber	Topic	Num Video	Published At	Days	Ratio
0	UCIdW2PnzoVAfq2o-K9vPAYA	Deep Tune Music	NaN	9454057	13900	['Christian_music', 'Electronic_music', 'Pop_m...']	76	2023-10-16T21:16:28.149912Z	139	100.000000
1	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	Orinn Music là đội ngũ sản xuất những dự án âm...	28878383	83600	['Electronic_music', 'Music', 'Pop_music']	461	2022-10-06T07:06:51.635929Z	514	162.645914
2	UC7HwMQwUXwpw5RnmOS0Fmzw	Music Of Wibu	- Dear Artists, Producers and Labels -\nMusic ...	755547	3360	['Electronic_music', 'Music_of_Asia', 'Pop_mus...']	230	2020-12-11T15:18:41.084001Z	1178	2.852292
3	UCpHhPOfjPC1Z5iVHTU_3mVg	HL Media Music	Công ty truyền thông giải trí âm nhạc Việt Nam...	13039629	21300	['Music', 'Music_of_Asia']	559	2022-07-06T08:10:04.716296Z	606	35.148515
4	UCiq69t2ThqzU0NQ8yXCfAZA	Yen Music	CHÀO MỪNG CÁC BẠN ĐẾN VỜI KÊNH YEN MUSIC\n\nKê...	31940889	277000	['Pop_music', 'Music', 'Electronic_music']	631	2019-05-28T14:17:41Z	1741	159.103963
...	...	...	...	...	...	...	...	...	...	...
567	UCyoFGdNz4k7i8R0gtSoTZrA	Rooster Teeth Podcast	Join Armando, Griff, and Andrew for weekly epi...	253128	27600	['Television_program', 'Entertainment', 'Film']	88	2019-05-16T20:02:18Z	1753	15.744438
568	UCrboPj-9ZOZvr3IVyyhXz6Q	Happy Hour Podcast	NU Welcome to the official YouTube channel for...	227870135	764000	['Film', 'Entertainment']	688	2013-09-28T23:01:46Z	3809	200.577579
569	UCaqFv9hD-XJmPMzC5Yvtneg	The Dispatch Podcasts	Welcome to The Dispatch, where every podcast i...	686352	3780	['Society', 'Politics']	816	2023-01-21T01:29:22.026763Z	407	9.287469
570	UCu9c6ZVPLxaBXmSSps-antA	Palcast Podcast	The Palcast is hosted by Shaggy 2 Dope and Keg...	1084674	15100	['Lifestyle_(sociology)', 'Entertainment', 'Te...']	144	2021-04-21T23:39:59.01941Z	1047	14.422159
571	UCBw0bL8MtOll1UX7db2MT3A	The Always Sunny Podcast	Home of the podcast hosted by Charlie Day, Gle...	97984503	651000	['Film', 'Entertainment', 'Television_program']	161	2021-11-22T17:24:41.390167Z	832	782.451923

### 3. Phân cấp cho dữ liệu

- Với mỗi channel rank, chia các video tương ứng thành 3 phân khúc dựa trên tỉ lệ giữa lượt view của video và thời gian mà video tồn tại
- Tạo thêm 1 cột Days tương ứng với thời gian tồn tại của video
- Tạo thêm 1 cột Ratio tương ứng với tỉ lệ view/days



# 3. Phân cấp cho dữ liệu

Video ID	Video Rank	Title	Topic	Duration	Published At	View	Like	Num Comments	Date	Description	Tag	Genre	Days	Ratio
sa6Ci124qYQ	0	VIET DEEP 2022 - MIXTAPE NỖI ĐAU XÓT XA - NHÀ...	['Electronic_music', 'Music', 'Pop_music']	3682.0	2022-10- 28T11:00:46Z	323058	2651.0	97.0	02/01/2024	► Website: <a href="https://daorentertainment.com/">https://daorentertainment.com/</a> \n► Ng...	['DELUXE', 'DELUXE MUSIC', 'viet deep', 'viet ...	Music	492	656.621951
zwHbwtvMPWY	0	VIET DEEP 2023 - MIXTAPE SET NHẠC CŨ 9X - DEEP...	['Electronic_music', 'Music', 'Pop_music']	3535.0	2022-11- 15T11:00:44Z	623873	4543.0	72.0	02/01/2024	VIET DEEP 2023 - MIXTAPE SET NHẠC CŨ 9X - DEEP...	['DELUXE', 'DELUXE MUSIC', 'viet deep', 'viet ...	Music	474	1316.187764
4hiZ-ADm8-k	0	CHILL BASS 2024 - NHẠC DEEP HOUSE REMIX 2024 H...	['Electronic_music', 'Music', 'Pop_music']	7612.0	2024-01- 08T11:00:04Z	29929	159.0	6.0	02/01/2024	CHILL BASS 2024 - NHẠC DEEP HOUSE REMIX 2024 H...	['DELUXE', 'DELUXE MUSIC', 'viet deep', 'viet ...	Music	55	544.163636
5Fa-Ex2lXKs	0	VIET DEEP 2023 - NHẠC HOUSE LAK REMIX 2023 HAY...	['Electronic_music', 'Music', 'Pop_music']	7893.0	2023-12- 13T11:00:03Z	23949	184.0	7.0	02/01/2024	VIET DEEP 2023 - NHẠC HOUSE LAK REMIX 2023 HAY...	['DELUXE', 'DELUXE MUSIC', 'viet deep', 'viet ...	Music	81	295.666667
O25DqYbLXqY	0	VIET DEEP 2023 - NHẠC HOUSE LAK REMIX 2023 HAY...	['Electronic_music', 'Music', 'Pop_music']	8593.0	2023-10- 30T11:00:30Z	220130	1386.0	44.0	02/01/2024	VIET DEEP 2023 - NHẠC HOUSE LAK REMIX 2023 HAY...	['DELUXE', 'DELUXE MUSIC', 'viet deep', 'viet ...	Music	125	1761.040000

# 3. Phân cấp cho dữ liệu

Kết quả kiến trúc sau khi phân cấp dữ liệu

Music	Channel rank 1	Video rank 1	video 1, video 2,...
		Video rank 2	video 1, video 2,...
		Video rank 3	video 1, video 2,...
	Channel rank 2	Video rank 1	video 1, video 2,...
		Video rank 2	video 1, video 2,...
		Video rank 3	video 1, video 2,...
	Channel rank 3	Video rank 1	video 1, video 2,...
		Video rank 2	video 1, video 2,...
		Video rank 3	video 1, video 2,...
Podcast	Channel rank 1	Video rank 1	video 1, video 2,...
		Video rank 2	video 1, video 2,...
		Video rank 3	video 1, video 2,...
	Channel rank 2	Video rank 1	video 1, video 2,...
		Video rank 2	video 1, video 2,...
		Video rank 3	video 1, video 2,...
	Channel rank 3	Video rank 1	video 1, video 2,...
		Video rank 2	video 1, video 2,...
		Video rank 3	video 1, video 2,...

# 3. Phân cấp cho dữ liệu

Data frame của dữ liệu Channel

	ChannelID	ChannelName	Description	View	NumSubscriber	Topic	NumVideo	PublishedAt	Days	Ratio
0	UCIdW2PnzoVAfq2o-K9vPAYA	Deep Tune Music	No Description	9454057	13900	Christian_music, Electronic_music, Pop_music, ...	76	2023-10-16T21:16:28.149912Z	139	100.000000
1	UCKCzhWD9rOF74RbOnWXEj9w	Deluxe Music	Orinn Music IÃ Ä á» i ngÃ© sáº£n xuáº¥t nháº¬...	28878383	83600	Electronic_music, Music, Pop_music	461	2022-10-06T07:06:51.635929Z	514	162.645914
2	UC7HwMQwUXwpw5RnmOS0Fmzw	Music Of Wibu	- Dear Artists, Producers and Labels -\nMusic ...	755547	3360	Electronic_music, Music_of_Asia, Pop_music, Music	230	2020-12-11T15:18:41.084001Z	1178	2.852292
3	UCpHhPOfjPC1Z5iVHTU_3mVg	HL Media Music	CÃng ty truyá» n thÃng giáº£i trÃ Ä¢m nháº°;...	13039629	21300	Music, Music_of_Asia	559	2022-07-06T08:10:04.716296Z	606	35.148515
4	UCiq69t2ThqzU0NQ8yXcfAZA	Yen Music	CHÃ¢NG MÃ»NG CÃ C BÃ° NÃ Ä N VÃ» IKÃ NH YEN	31940889	277000	Pop_music, Music, Electronic_music	631	2019-05-28T14:17:41Z	1741	159.103963
...	...	...	...	...	...	...	...	...	...	...
567	UCyoFGdNz4k7i8R0gtSoTZrA	Rooster Teeth Podcast	Join Armando, Griff, and Andrew for weekly epi...	253128	27600	Television_program, Entertainment, Film	88	2019-05-16T20:02:18Z	1753	15.744438
568	UCrboPj-9ZOZvr3IVyyhXz6Q	Happy Hour Podcast	Welcome to the official YouTube chann...	227870135	764000	Film, Entertainment	688	2013-09-28T23:01:46Z	3809	200.577579
569	UCaqFv9hD-XJmPMzC5Yvtneg	The Dispatch Podcasts	Welcome to The Dispatch, where every podcast i...	686352	3780	Society, Politics	816	2023-01-21T01:29:22.026763Z	407	9.287469
570	UCu9c6ZVPLxaBXmSSps-antA	Palcast Podcast	The Palcast is hosted by Shaggy 2 Dope and Keg...	1084674	15100	Lifestyle_(sociology), Entertainment, Televisi...	144	2021-04-21T23:39:59.01941Z	1047	14.422159
571	UCBw0bL8MtOII1UX7db2MT3A	The Always Sunny Podcast	Home of the podcast hosted by Charlie Day, Gle...	97984503	651000	Film, Entertainment, Television_program	161	2021-11-22T17:24:41.390167Z	832	782.451923

572 rows x 10 columns

# 3. Phân cấp cho dữ liệu

Data frame của dữ liệu Video sau khi phân cấp

	ChannelID	ChannelRank	ChannelName	VideoID	VideoRank	Title	Topic	Duration	PublishedAt	Views	Likes	NumComments	Date	Description	Tag	Genre	Days	Ratio
0	UCTPioC8tmJMcpdB-XKMch4Q	1	music meets heaven	-4Nx2hEhVRE	1	Lauren Daigle ~ Look Up Child (Lyrics)	Christian_music, Music, Pop_music	184.0	2018-09-09T21:00:00Z	5053406	46866.0	1390.0	2024-02-03	Lauren Daigle ~ Look Up Child (Lyrics)\nall ...	christian music, music meets heaven, music mee...	Music	2002	2524.178821
1	UCTPioC8tmJMcpdB-XKMch4Q	1	music meets heaven	-4Nx2hEhVRE	1	Lauren Daigle ~ Look Up Child (Lyrics)	Christian_music, Music, Pop_music	184.0	2018-09-09T21:00:00Z	5061498	46926.0	1394.0	2024-02-18	Lauren Daigle ~ Look Up Child (Lyrics)\nall ...	christian music, music meets heaven, music mee...	Music	2002	2528.220779
2	UCTPioC8tmJMcpdB-XKMch4Q	1	music meets heaven	-4Nx2hEhVRE	1	Lauren Daigle ~ Look Up Child (Lyrics)	Christian_music, Music, Pop_music	184.0	2018-09-09T21:00:00Z	5066428	46966.0	1394.0	2024-02-27	Lauren Daigle ~ Look Up Child (Lyrics)\nall ...	christian music, music meets heaven, music mee...	Music	2002	2530.683317
3	UCTPioC8tmJMcpdB-XKMch4Q	1	music meets heaven	-4Nx2hEhVRE	1	Lauren Daigle ~ Look Up Child (Lyrics)	Christian_music, Music, Pop_music	184.0	2018-09-09T21:00:00Z	5067632	46977.0	1395.0	2024-02-29	Lauren Daigle ~ Look Up Child (Lyrics)\nall ...	christian music, music meets heaven, music mee...	Music	2002	2531.284715
4	UCTPioC8tmJMcpdB-XKMch4Q	1	music meets heaven	-4Nx2hEhVRE	1	Lauren Daigle ~ Look Up Child (Lyrics)	Christian_music, Music, Pop_music	184.0	2018-09-09T21:00:00Z	5068289	46982.0	1395.0	2024-03-01	Lauren Daigle ~ Look Up Child (Lyrics)\nall ...	christian music, music meets heaven, music mee...	Music	2002	2531.612887
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
256293	UCLfaW3fd5Oo-nFvKsL_1CfQ	3	TSL Movie Podcast	zz7ErPh8SRg	3	1985 Part 2 - Cool Cruel Summers - The Slaught...	Entertainment, Film, Television_program	4080.0	2021-07-11T23:00:11Z	949	61.0	22.0	2024-02-21	Continuing our discussion on the summer of 1985...	the slaughtered lamb movie podcast, cool cruel ...	Podcast	966	0.982402
256294	UCLfaW3fd5Oo-nFvKsL_1CfQ	3	TSL Movie Podcast	zz7ErPh8SRg	3	1985 Part 2 - Cool Cruel Summers - The Slaught...	Entertainment, Film, Television_program	4080.0	2021-07-11T23:00:11Z	952	61.0	22.0	2024-02-25	Continuing our discussion on the summer of 1985...	the slaughtered lamb movie podcast, cool cruel ...	Podcast	966	0.985507
256295	UCLfaW3fd5Oo-nFvKsL_1CfQ	3	TSL Movie Podcast	zz7ErPh8SRg	3	1985 Part 2 - Cool Cruel Summers - The Slaught...	Entertainment, Film, Television_program	4080.0	2021-07-11T23:00:11Z	952	61.0	22.0	2024-02-25	Continuing our discussion on the summer of 1985...	the slaughtered lamb movie podcast, cool cruel ...	Podcast	966	0.985507
256296	UCLfaW3fd5Oo-nFvKsL_1CfQ	3	TSL Movie Podcast	zz7ErPh8SRg	3	1985 Part 2 - Cool Cruel Summers - The Slaught...	Entertainment, Film, Television_program	4080.0	2021-07-11T23:00:11Z	952	61.0	22.0	2024-02-27	Continuing our discussion on the summer of 1985...	the slaughtered lamb movie podcast, cool cruel ...	Podcast	966	0.985507
256297	UCLfaW3fd5Oo-nFvKsL_1CfQ	3	TSL Movie Podcast	zz7ErPh8SRg	3	1985 Part 2 - Cool Cruel Summers - The Slaught...	Entertainment, Film, Television_program	4080.0	2021-07-11T23:00:11Z	953	61.0	22.0	2024-02-29	Continuing our discussion on the summer of 1985...	the slaughtered lamb movie podcast, cool cruel ...	Podcast	966	0.986542

256298 rows x 18 columns

# 3. Phân cấp cho dữ liệu

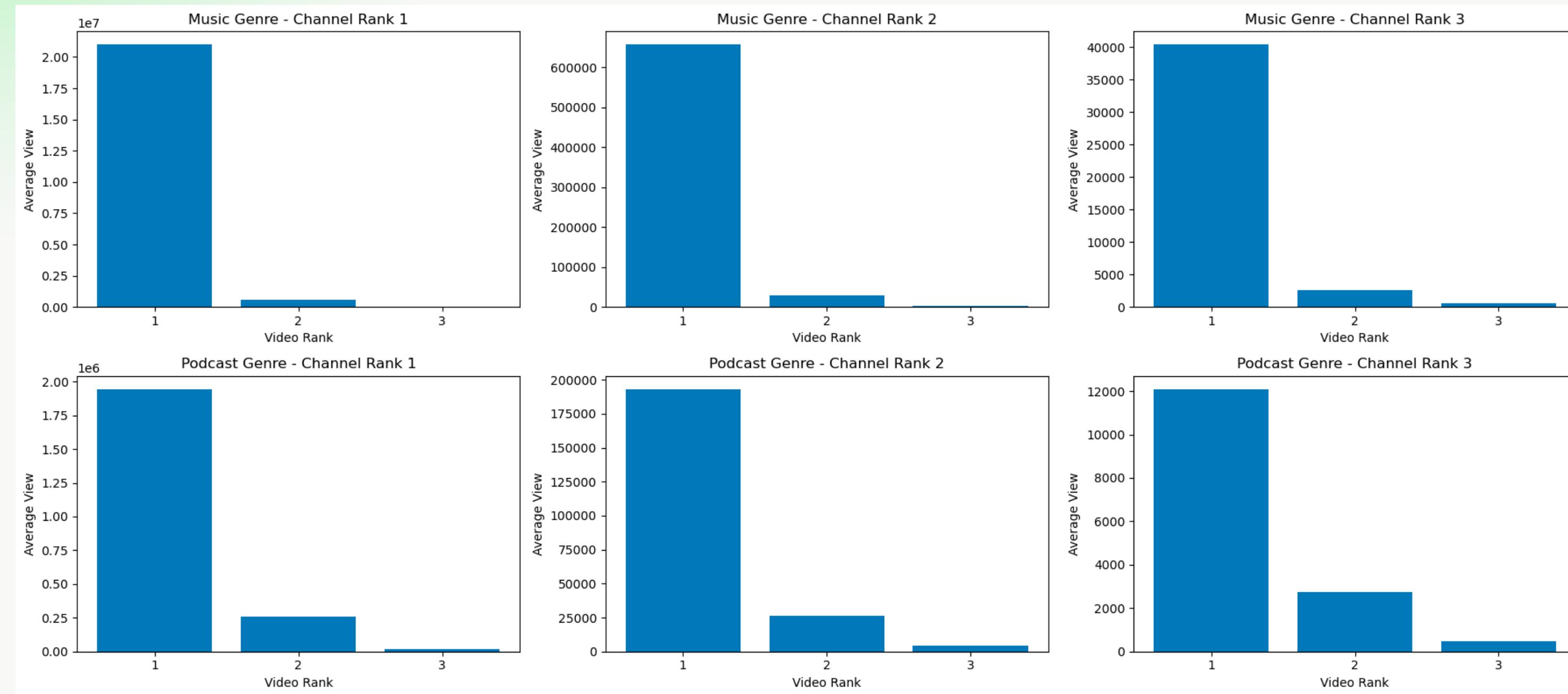
Thông tin ngẫu nhiên của 2 video

ChannelID	UC7qiJAwCaWY57G0TsO_M6Lw
ChannelRank	1
ChannelName	ACV Music Trending
VideoID	T9570V_tAww
VideoRank	1
Title	NHẠC TRẺ REMIX HAY NHẤT 2023 - LIÊN KHÚC NHẠC ...
Topic	Electronic_music, Music, Pop_music
Duration	4687.0
PublishedAt	2023-05-03T03:11:06Z
Views	744248
Likes	0.0
NumComments	52.0
Date	2024-02-27
Description	NHẠC TRẺ REMIX HAY NHẤT 2023 - LIÊN KHÚC NHẠC ...
Tag	acv trending, acv, nhac tre moi nhat, nhac tre...
Genre	Music
Days	305
Ratio	2440.157377

ChannelID	UCRpjHHu8ivVWs73uxHlwFA
ChannelRank	1
ChannelName	Eurovision Song Contest
VideoID	iMu47raqbcc
VideoRank	1
Title	Ronela Hajati - Sekret - Albania AL - Official Entry - Eurovision 2022
Topic	Music
Duration	215.0
PublishedAt	2022-03-04T15:00:12Z
Views	6065999
Likes	76370.0
NumComments	8126.0
Date	2024-02-02
Description	No Description
Tag	No Tag
Genre	Music
Days	730
Ratio	8309.587671

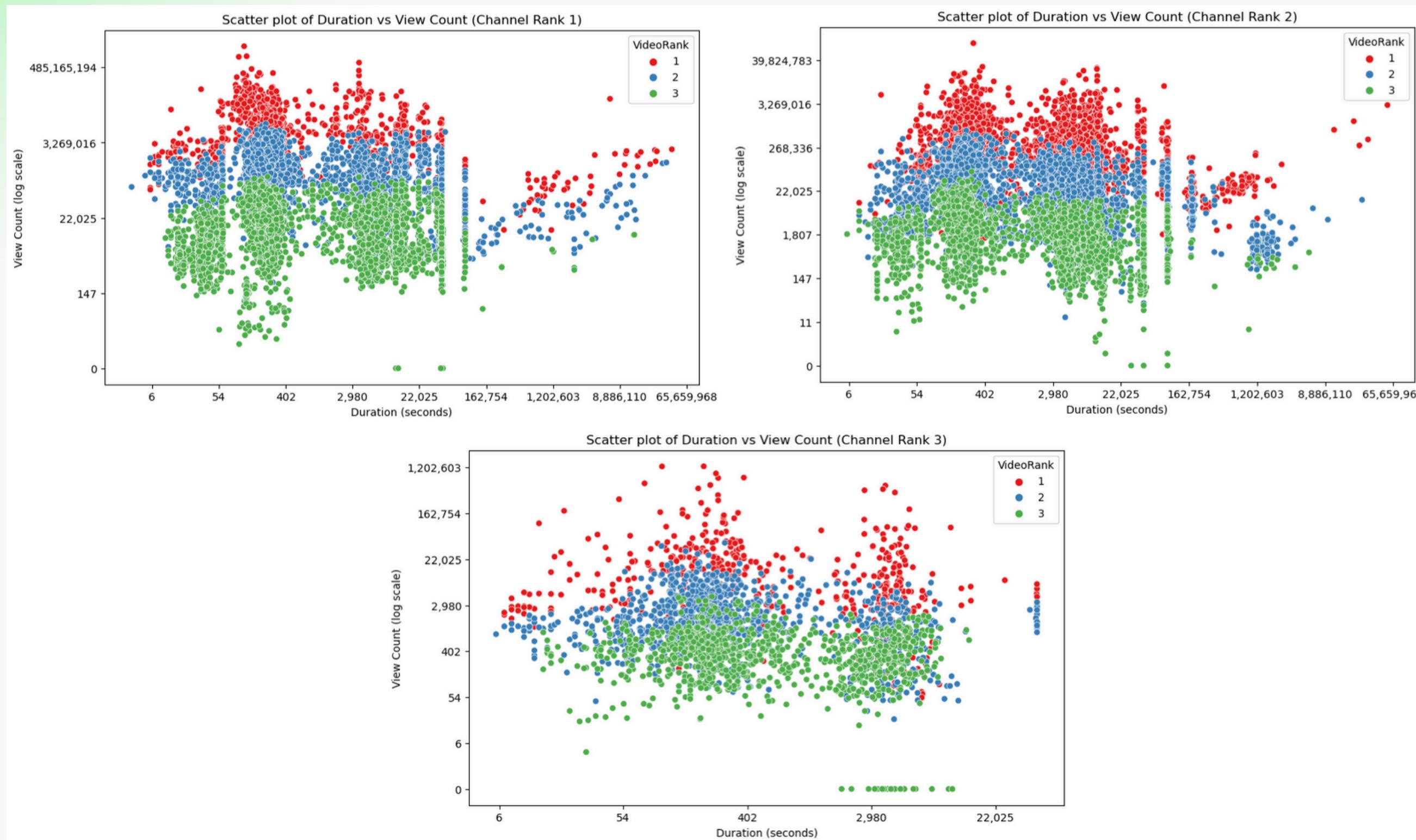
# 4. Phân tích dữ liệu

View trung bình của mỗi video rank trong từng channel rank của 2 thể loại  
Music và Podcast



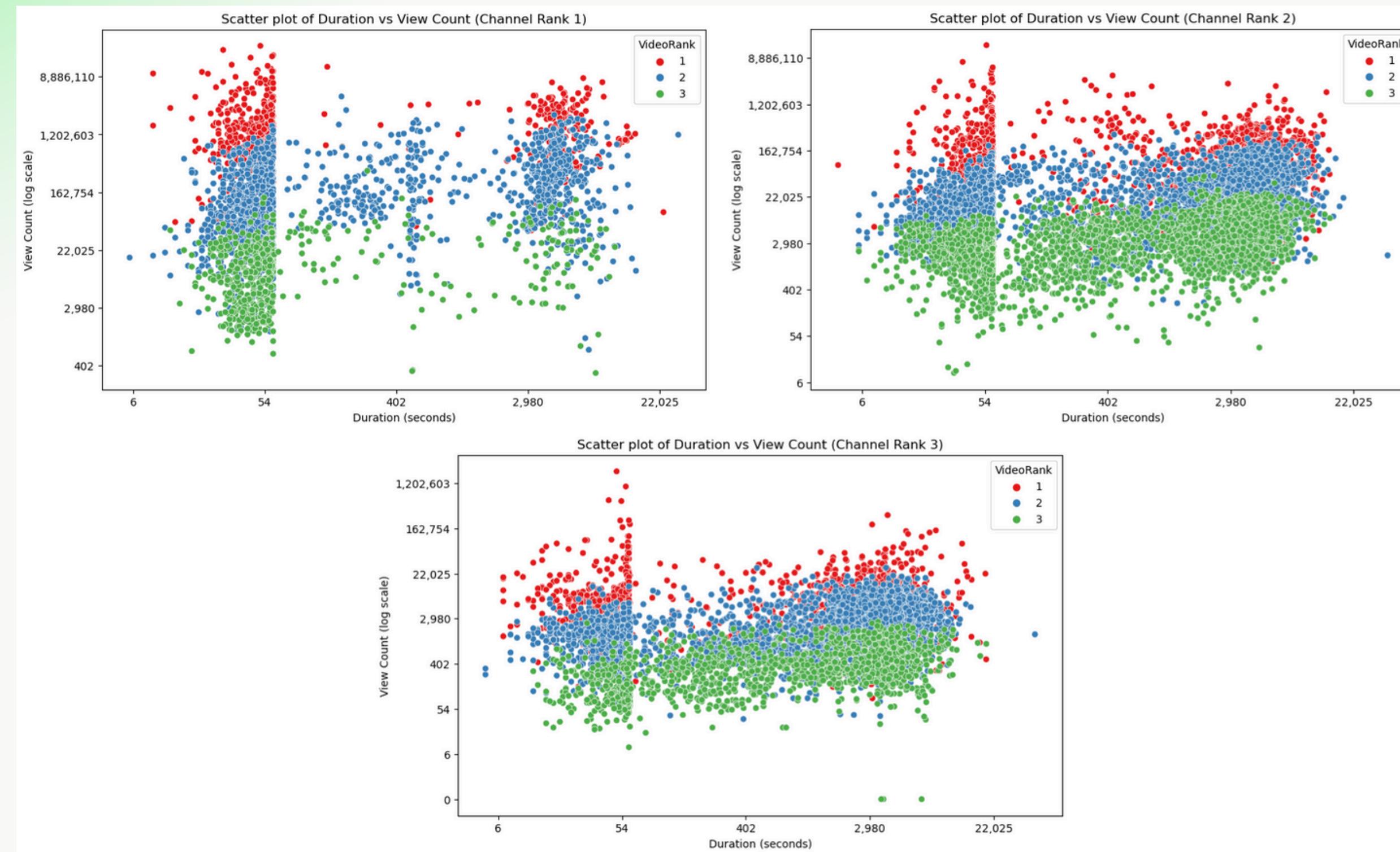
# 4. Phân tích dữ liệu

Tương quan giữa duration và view của thể loại Music



# 4. Phân tích dữ liệu

Tương quan giữa duration và view của thể loại Podcast



# 4. Phân tích dữ liệu

Tương quan pearson

Genre	ChannelRank	VideoRank
Music	1	1 -0.026742
		2 -0.025679
		3 -0.019915
	2	1 0.007682
		2 -0.041949
		3 -0.048958
	3	1 -0.082326
		2 -0.073721
		3 -0.064012
Podcast	1	1 -0.047725
		2 0.244415
		3 0.210704
	2	1 -0.061228
		2 0.303584
		3 0.345496
	3	1 -0.021942
		2 0.265563
		3 0.312083

# 4. Phân tích dữ liệu

Phân tích Title video của 2 thể loại Music và Podcast

Các khía cạnh phân tích:

- Tìm từ xuất hiện nhiều trong mỗi channel rank -> Sự khác biệt và xu hướng đặt title của các Channel
- Tỷ lệ nội dung tích cực và tiêu cực trong channel rank
- Độ dài title đối với mỗi thể loại -> Sự khác biệt giữa các phân khúc channel

# 4. Phân tích dữ liệu

Tìm từ xuất hiện nhiều trong mỗi channel rank

1

2

3

Lọc title của từng video:

- Dùng thư viện “`from langdetect import detect`” để kiểm tra ngôn ngữ tiếng Anh

Xử lý title tiếng Anh:

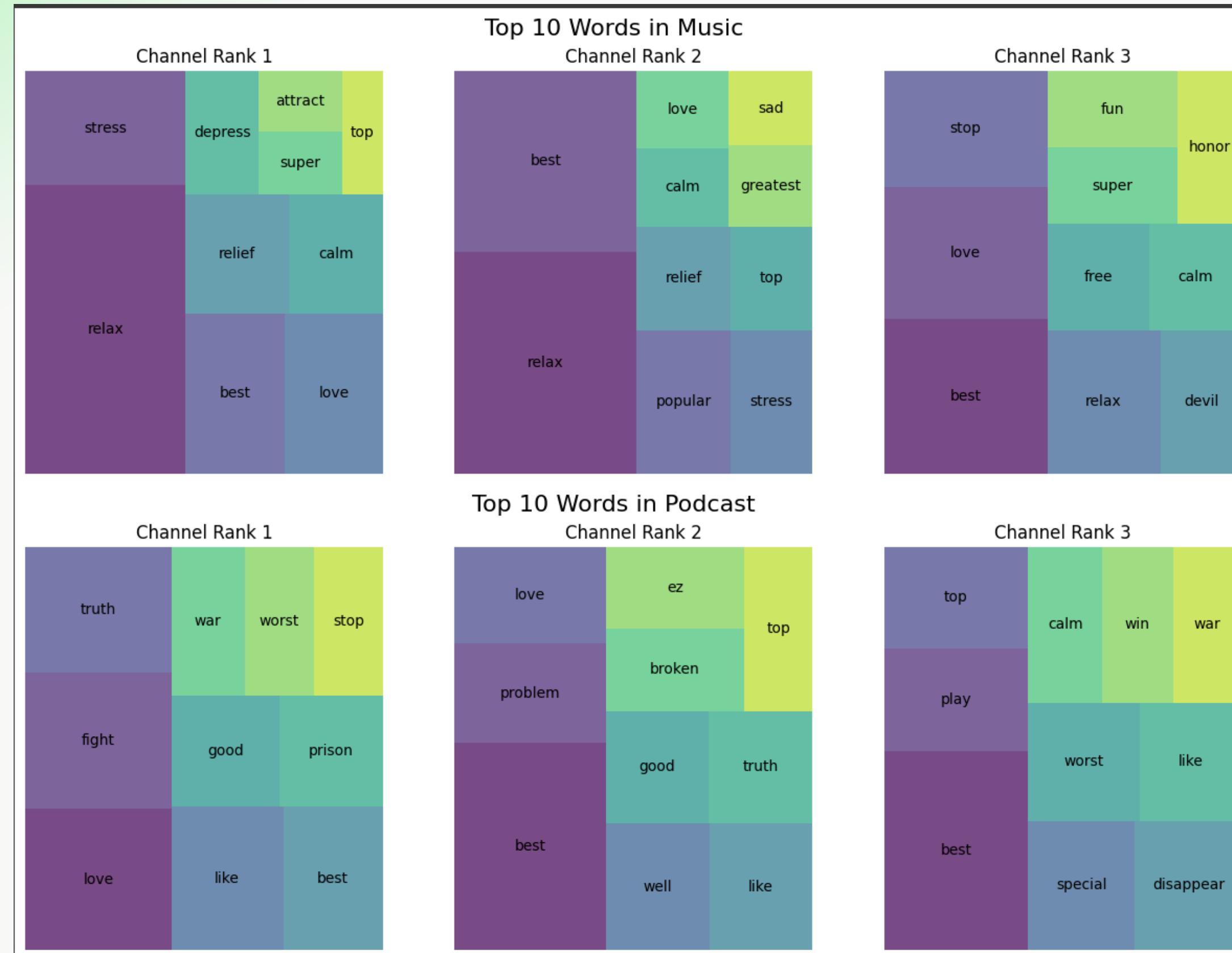
- Chuyển đổi về chữ thường
- Regex loại bỏ các ký tự không phải chữ cái và không phải số
- Tokenize tiêu đề bằng `word_tokenize`.
- Loại bỏ stop words bằng `stopwords`
- Đưa từ về dạng gốc của nó bằng `stem`

Đếm từ và lưu vào một dictionary `analyzed_data` có dạng như sau:

```
{'Music': {1: Counter({'monster': 12, 'remix': 148, 'tiktok': 18, 'viral': 4, 'see': 12, 'alex': 12, .....}}}
```

# 4. Phân tích dữ liệu

Top 10 từ xuất hiện nhiều trong mỗi channel rank



# 4. Phân tích dữ liệu

Tỉ lệ từ tích cực và tiêu cực mỗi channel rank

1

2

3

Với mỗi từ đã đếm :

- Dùng thư viện `SentimentIntensityAnalyzer` của Python để phân tích cảm xúc trong văn bản

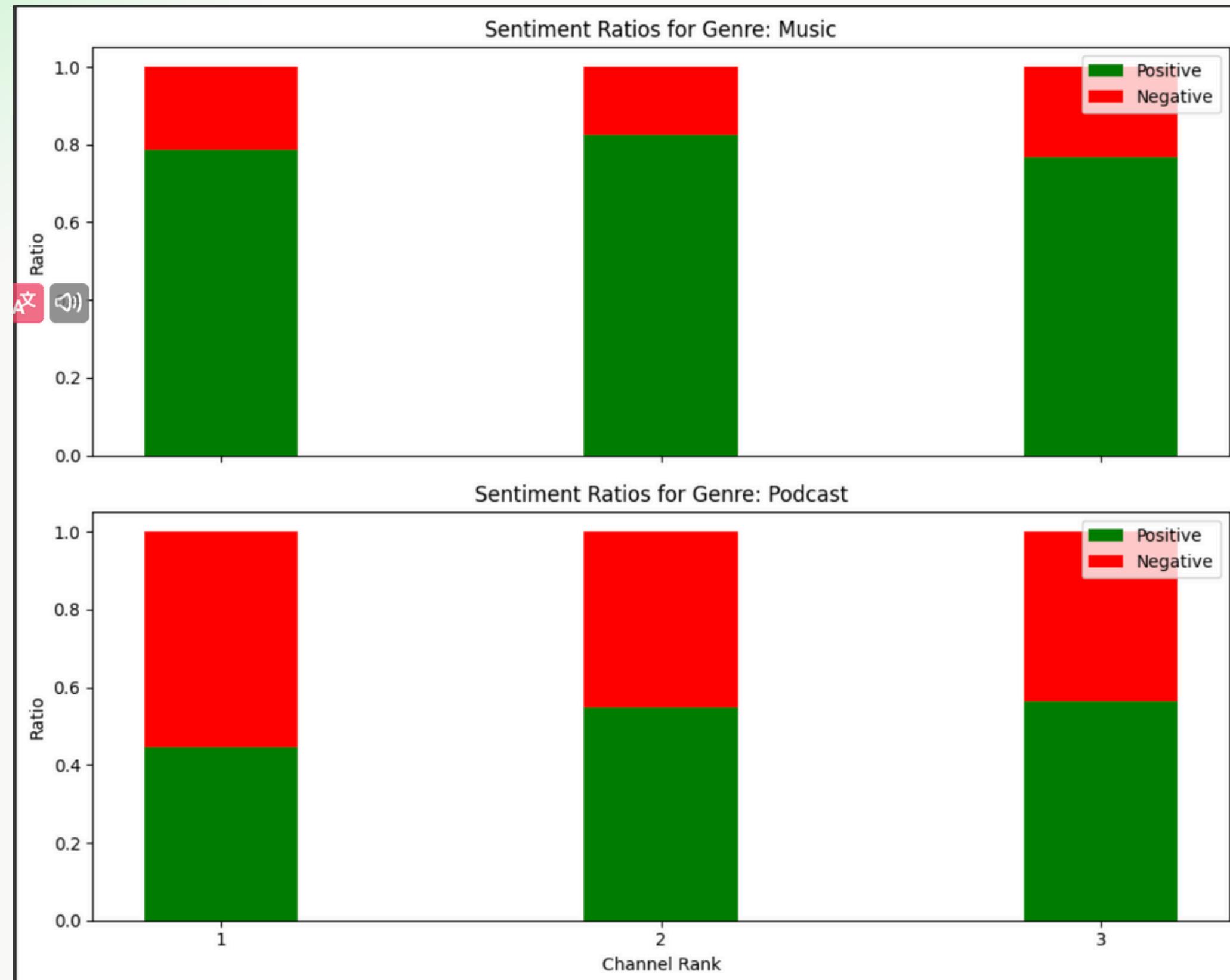
Chấm điểm cảm xúc:

- Dựa trên chỉ số ‘`compound`’ của thư viện, -1 thể hiện cảm xúc tiêu cực, 1 thể hiện cảm xúc tích cực.
- 0.2 được sử dụng làm ngưỡng. Nếu điểm số ‘`compound`’ của từ > 0.2 là tích cực. Nếu nhỏ hơn -0.2 là tiêu cực

Tính tổng số lượng từ của mỗi loại và tính tỷ lệ và lưu vào một dictionary.  
{'Music': {1: {'positive\_ratio': 0.78603, 'negative\_ratio': 0.21396, .....}}

# 4. Phân tích dữ liệu

Tỉ lệ từ tích cực và tiêu cực mỗi channel rank



# 4. Phân tích dữ liệu

Tìm độ dài Title phù hợp cho mỗi thể loại

1

2

3

Lấy độ dài của mỗi title video

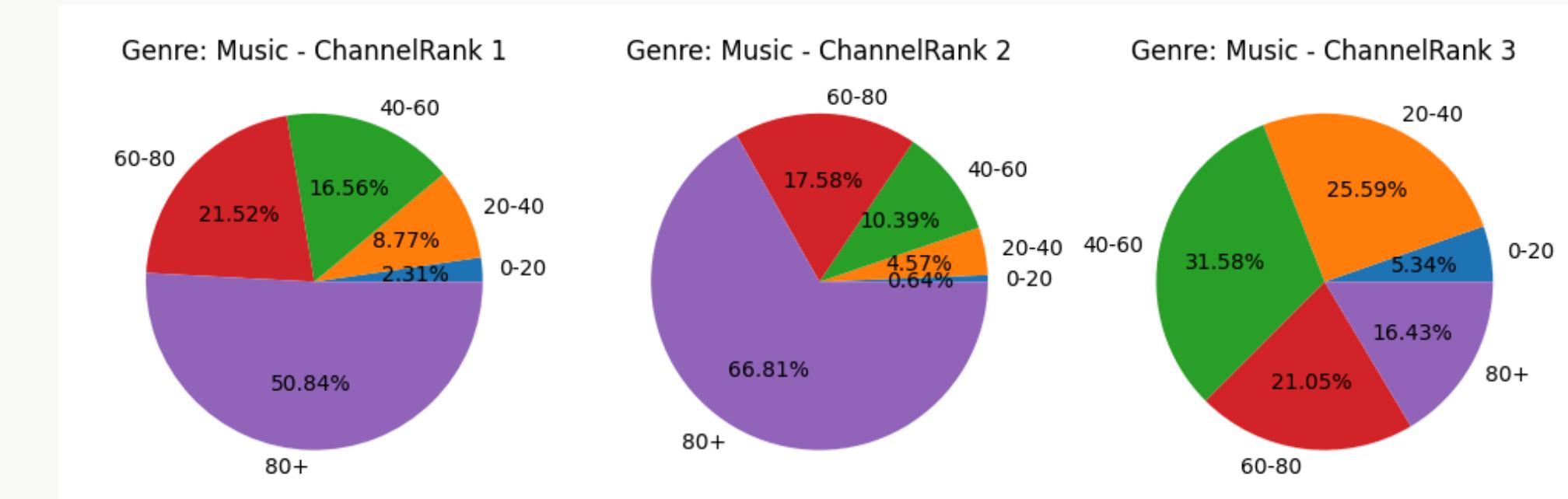
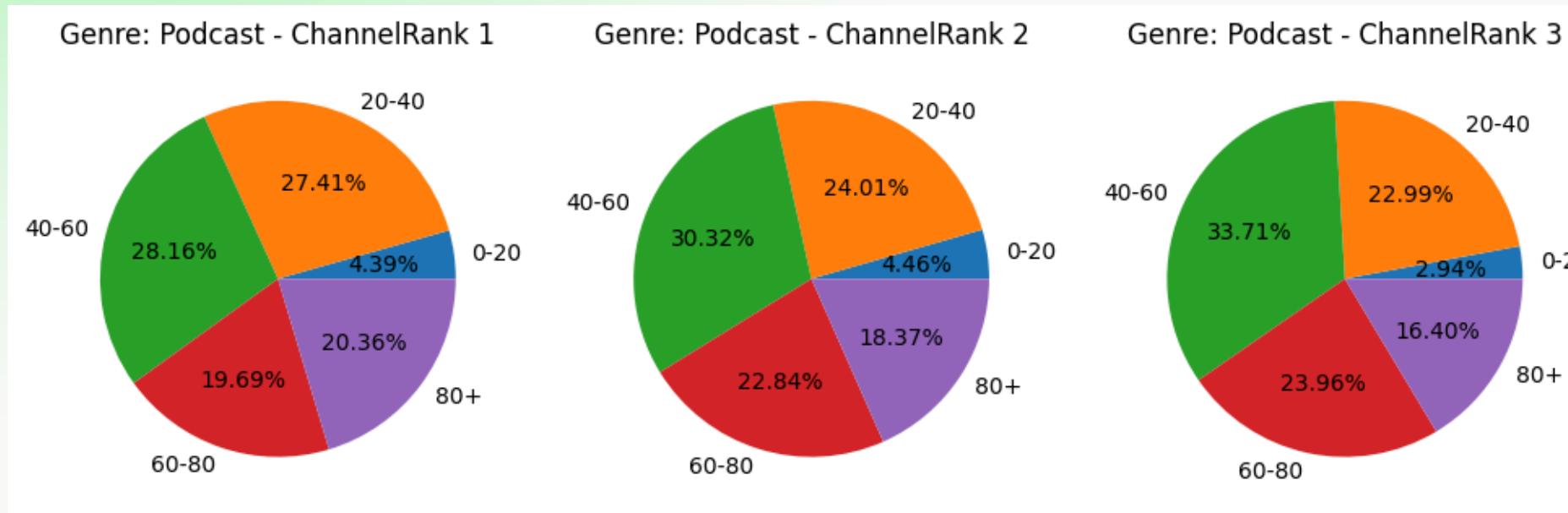
Chia thành các bin tương ứng:

- 0-20
- 20-40
- 40-60
- 60-80
- 80+

Tỉ lệ của mỗi bin ứng với thể loại và channel rank

# 3. Phân tích dữ liệu

## Phân tích Title Video



# 4. Phân tích dữ liệu

Tìm độ dài Title phù hợp cho mỗi thể loại

1

2

3

Kiểm tra format title:

Tên channel có trong Title video

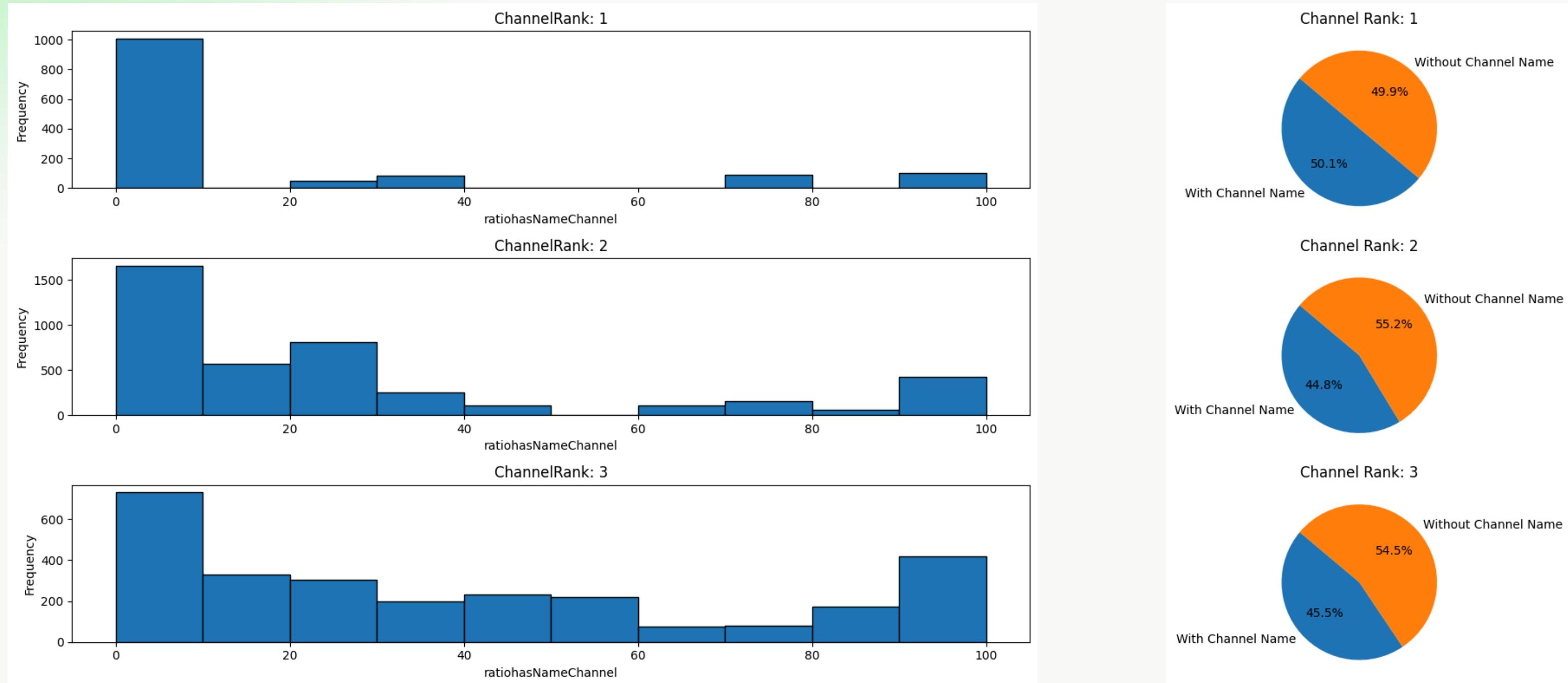
Title là một câu hỏi

Đánh giá:

- Số lượng Video title có tên Channel là bao nhiêu. Số lượng Video có title là câu hỏi

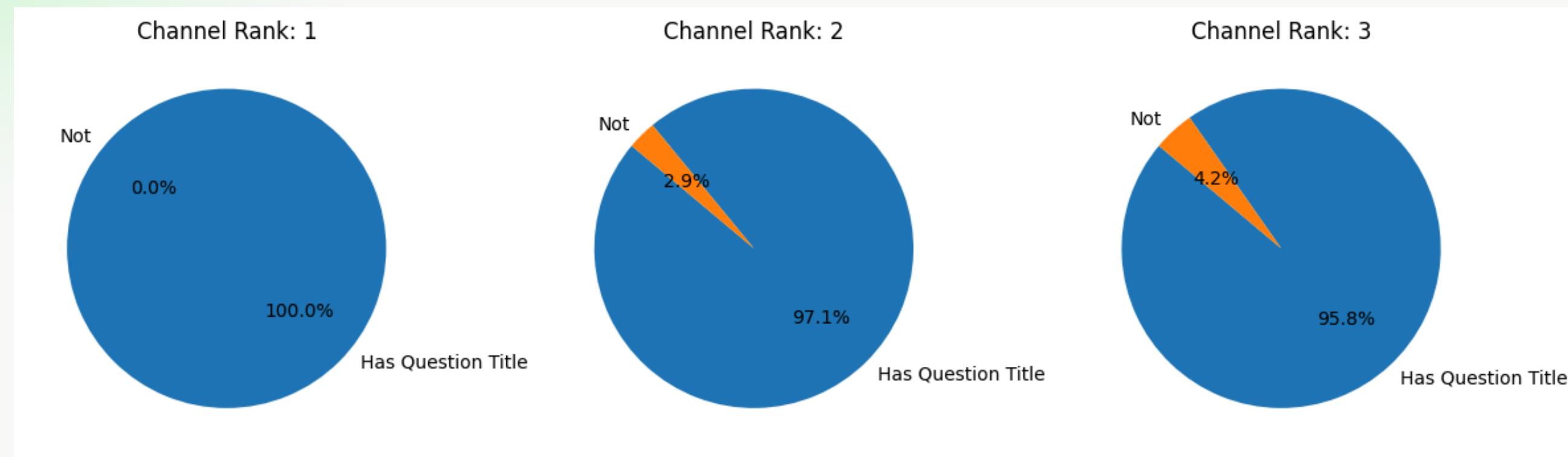
Tỉ lệ đối với mỗi phân khúc Channel

# 4. Phân tích dữ liệu



# 4. Phân tích dữ liệu

## Phân tích Title Podcast Video



# 4. Phân tích dữ liệu

Phân tích thời gian đăng video của mỗi phân khúc channel

Các khía cạnh phân tích:

- Thống kê các khoảng thời gian đăng video ở mỗi phân khúc video
- Tìm ra ‘khung giờ vàng’ để đăng video

# 4. Phân tích dữ liệu

Tìm khung giờ xuất hiện nhiều trong mỗi channel rank

1

2

3

Lọc giờ đăng của từng channel rank, của từng video rank:

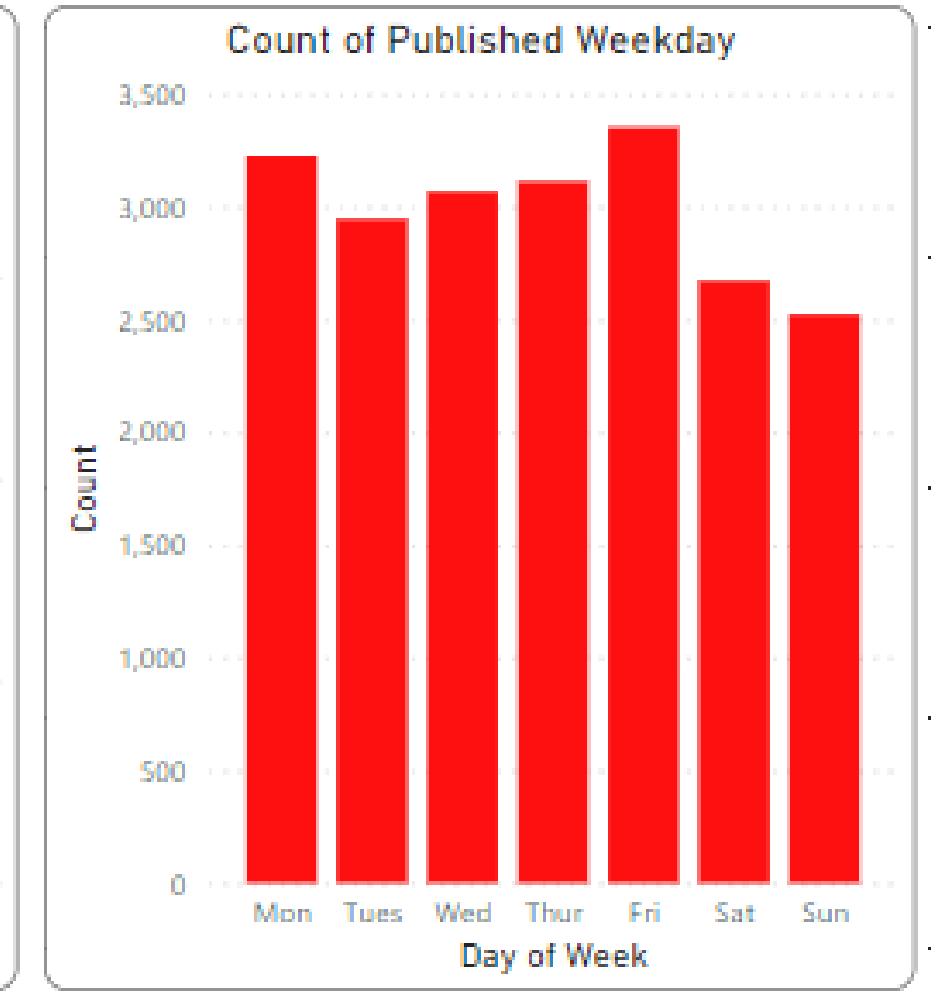
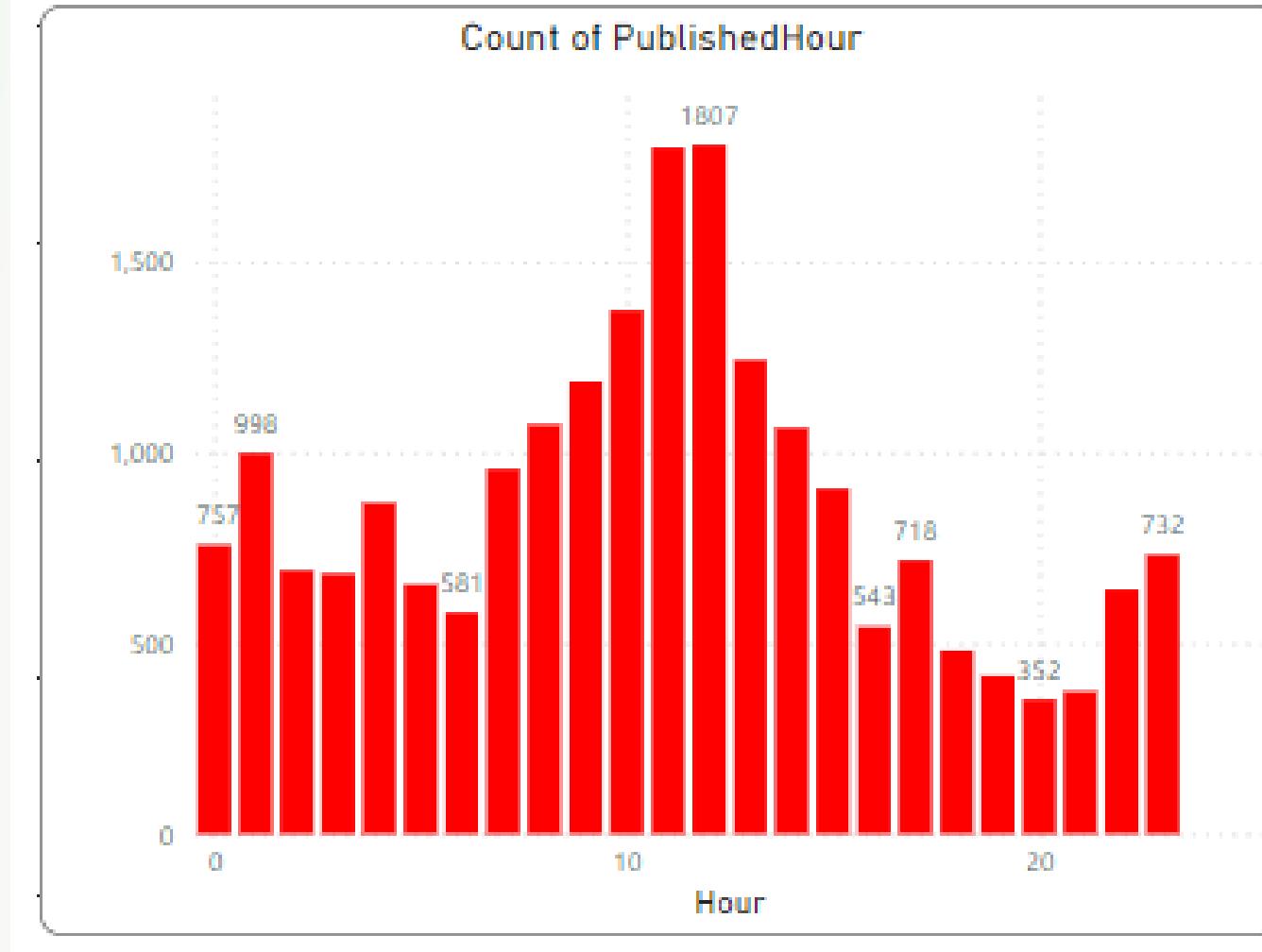
- Tách cột PublishedAt ra khỏi dataframe
- Sử dụng thư viện `datetime` để tách tiếp giờ ra khỏi thời gian đăng

Đếm các giờ và lưu vào dictionary có dạng {giờ\_1: số lần xuất hiện, giờ\_2: số lần xuất hiện, giờ\_3: số lần xuất hiện...}

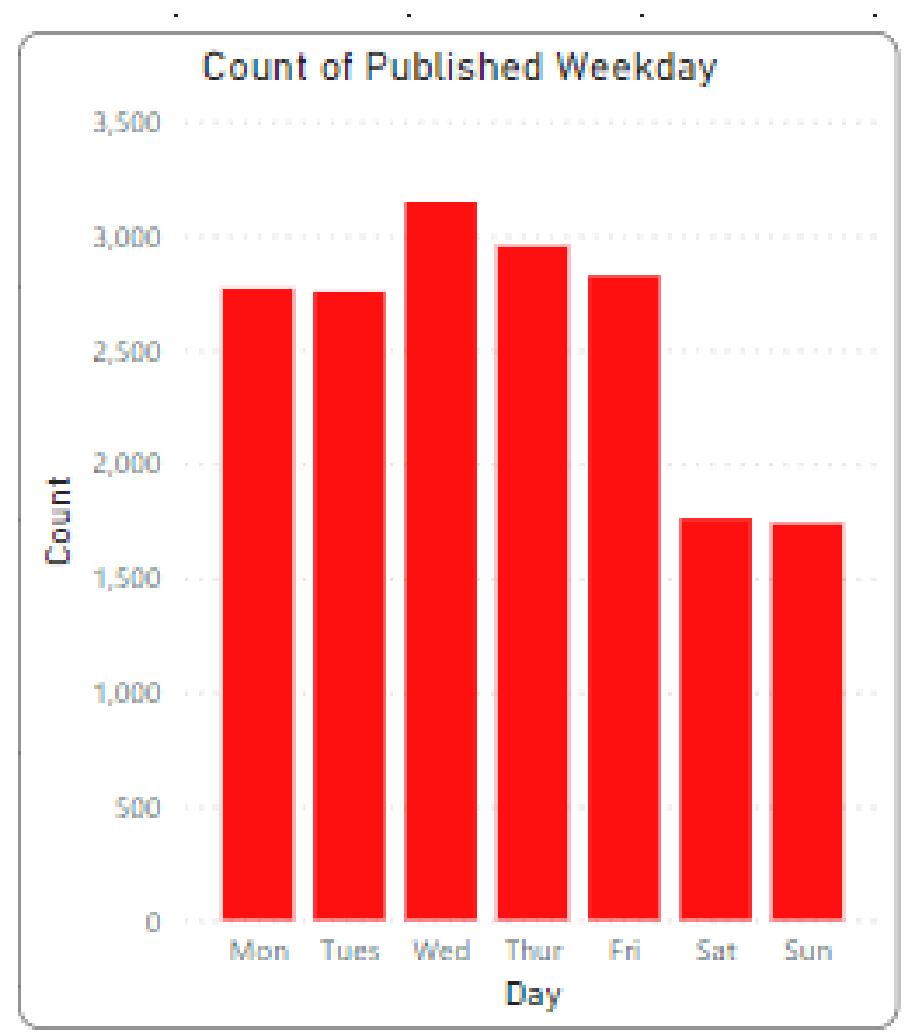
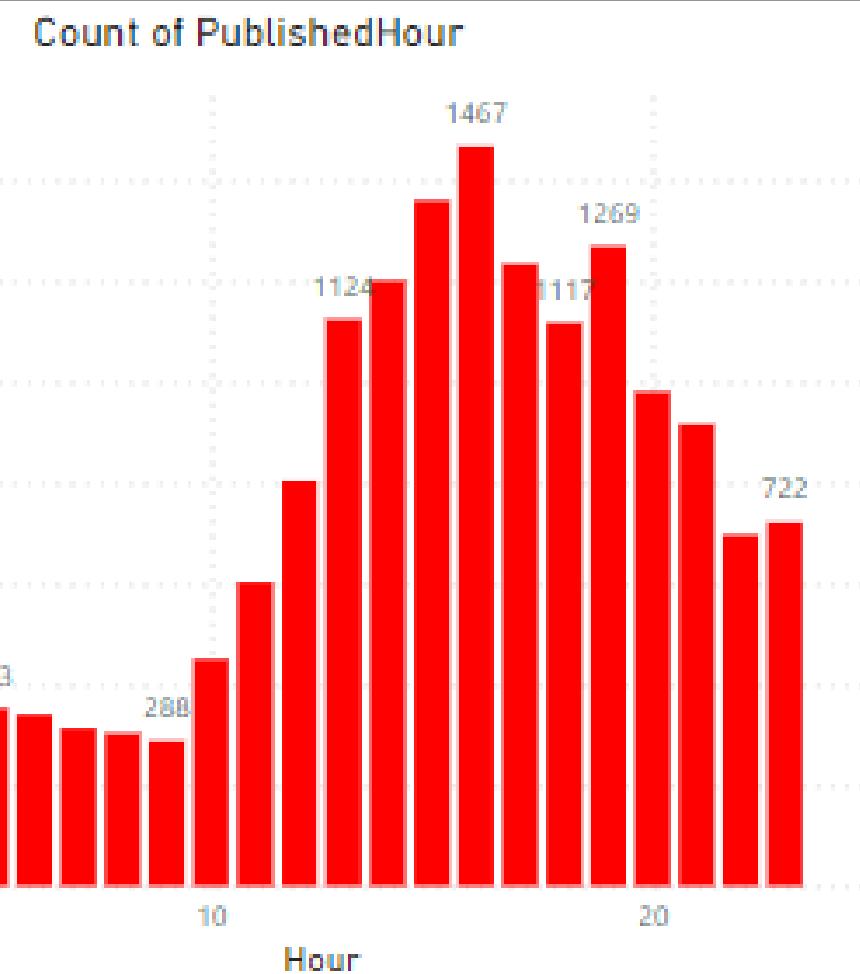
Lấy top các khoảng thời gian xuất hiện nhiều nhất tương ứng trong từng channel rank

# 4. Phân tích dữ liệu

Podcast



Music



# 5. Dashboard

# 6. Tự đánh giá

Tổng quan:

- Nắm quy trình khai thác dữ liệu
- Hoàn thành được việc thu thập dữ liệu
- Tiền xử lý và phân tích dữ liệu từ đó được ra được các insight phù hợp đối tượng hướng tới

Khó khăn:

- Dữ liệu thu thập bị thiếu (channel), thuộc tính không đồng nhất, một số lỗi khi thực thu thập
- Bài toán phân cấp dữ liệu
- Phân tích NLP gặp khó khăn trong vấn đề ngôn ngữ. (chỉ phân tích được tiếng Anh)

Thanks for watching