

Báo Cáo Dự Án Cá Nhân

NGƯỜI TRÌNH BÀY: HUỲNH BẢO KHANG

TÊN DỰ ÁN: PHÂN TÍCH TOP 1000 YOUTUBE CHANNEL

Quá Trình:

- I. **Xác định mục tiêu:** Xác định mục tiêu, chọn chủ đề việc phân tích dữ liệu.
- II. **Thu thập dữ liệu:** Thu thập dữ liệu cần thiết để giải quyết mục tiêu phân tích.
- III. **Khám phá và xử lý dữ liệu:** Gồm khám phá dữ liệu để hiểu rõ hơn về nội dung, cấu trúc và tính chất của dữ liệu. Xem dữ liệu mẫu, kiểm tra giá trị thiếu, xử lý dữ liệu không hợp lệ và tìm hiểu mối quan hệ giữa các biến.
- IV. **Áp dụng phương pháp phân tích:** Tiến hành trực quan hóa và phân tích.
- V. **Trình bày kết quả:** Tổng kết lại quá trình, kết quả đạt được.

Bài Làm:

I. Xác định mục tiêu:

Mục tiêu của bài báo cáo là nghiên cứu và phân tích các yếu tố ảnh hưởng đến sự phát triển và thành công của các kênh YouTube dựa trên bộ dữ liệu Top 1000 YouTube Channels. Dưới đây là mục tiêu cụ thể của bài báo cáo:

1. Nghiên cứu yếu tố ảnh hưởng đến sự phát triển của các kênh YouTube: Bài báo cáo sẽ tìm hiểu và phân tích các yếu tố quan trọng mà các kênh YouTube cần để phát triển và đạt thành công. Các yếu tố này có thể bao gồm số lượng lượt đăng ký, lượt xem, số lượng video, và các thuộc tính khác của kênh.
2. Hiểu rõ hơn về sự tương tác giữa số lượng lượt đăng ký, lượt xem và số lượng video: Bài báo cáo sẽ phân tích mối quan hệ giữa các yếu tố này để hiểu rõ hơn về tương quan và tác động của chúng lên nhau. Mục tiêu là xác định xem có mối quan hệ tương đồng, tương phản hoặc không có tương quan giữa các yếu tố này.
3. Phân tích xu hướng phát triển của các lĩnh vực chủ đạo trên YouTube: Bài báo cáo sẽ tìm hiểu xu hướng phát triển của các lĩnh vực chủ đạo trên YouTube như giải trí, âm nhạc, giáo dục, tin tức và chính trị. Mục tiêu là phân tích sự phổ biến, sự thay đổi về lượt đăng ký, lượt xem và số lượng video trong các lĩnh vực này để hiểu xu hướng và sự phát triển của chúng.
4. Đánh giá yếu tố thành công của các kênh YouTube nổi tiếng: Bài báo cáo sẽ xem xét và đánh giá các yếu tố thành công của các kênh YouTube nổi tiếng như T-series, Cocomelon Nursery Rhymes, ... và các kênh khác trong top. Mục tiêu là tìm hiểu những yếu tố đặc biệt mà các kênh này đã áp dụng để đạt được lượng lượt đăng ký và lượt xem cao.
5. Tạo ra một khung tham chiếu để giúp nhà sáng tạo nội dung YouTube định hướng và phát triển kênh của mình một cách hiệu quả.

Tổng quát, bài báo cáo sẽ cung cấp một cái nhìn tổng quan và chi tiết về các yếu tố ảnh hưởng đến sự phát triển của các kênh YouTube thông qua phân tích dữ liệu từ bộ dữ liệu Top 1000 YouTube Channels.

II. Thu thập dữ liệu:

Bước thứ hai trong quá trình phân tích dữ liệu là thu thập dữ liệu từ nguồn đáng tin cậy.

Trong bài phân tích này, nguồn dữ liệu được sử dụng là trang web <https://us.youtubers.me/global/all/top-1000-most-subscribed-youtube-channels>.

Trang web này cung cấp thông tin chi tiết về các kênh YouTube hàng đầu trên toàn cầu và được cập nhật thường xuyên, đảm bảo tính chính xác và tin cậy của dữ liệu.

Sau khi xác định được nguồn dữ liệu, tôi tiến hành quá trình lấy dữ liệu từ trang web này. Để thu thập dữ liệu, tôi sử dụng công cụ Excel để tổ chức và trích xuất thông tin cần thiết từ trang web. Bằng cách sử dụng các công cụ và chức năng của Excel, tôi đã tạo ra một bảng dữ liệu chứa thông tin về lượt đăng ký, lượt xem, số lượng video và các thuộc tính khác của top 1000 kênh YouTube.

Từ Web

☒ Cơ bản

☐ Nâng cao

URL

://us.youtubers.me/global/all/top-1000-most_subscribed-youtube-channels

OK

Hủy bỏ

Trình điều hướng

☐ Chọn nhiều mục

Hiện thị Tùy chọn

https://us.youtubers.me/global/all/top-1000-...

Document

Table 0

Table 1

Table 2

Dạng xem bảngDạng xem web

Table 0

Bản xem trước được tải xuống vào Thứ Sáu

rank	Youtuber	subscribers	videoc
1	T-Series	243,000,000	2
2	Cocomelon - Nursery Rhymes	160,000,000	1
3	SET India	157,000,000	1
4	Sony SAB	81,600,000	9
5	Kids Diana Show	111,000,000	9
6	Like Nastya	106,000,000	8
7	WWE	95,200,000	7
8	Vlad and Niki	97,700,000	7
9	Zee TV	69,400,000	7
10	Colors TV	63,200,000	6
11	Movieclips	59,400,000	5
12	Zee Music Company	95,500,000	5
13	El Reino Infantil	57,600,000	5
14	netd müzik	24,000,000	5
15	Ryan's World	34,900,000	5
16	ABS-CBN Entertainment	43,900,000	5
17	Toys and Colors	39,500,000	4
18	ChuChu TV Nursery Rhymes & Kids Songs	65,100,000	4
19	Super Simple Songs - Kids Songs	38,600,000	4

Tải

Chuyển đổi dữ liệu

Hủy bỏ

The top screenshot shows an Excel spreadsheet with a table named 'Table 0' being created. The table has columns for rank, Youtuber, subscribers, video views, video count, and category. The data is imported from a YouTube channel list.

rank	Youtuber	subscribers	video views	video count	category
1	T-Series	244,000,000	226,005,973,821	19,823	Music
2	Cocomelon - Nursery Rhymes	160,000,000	162,315,519,374	936	Education
3	SET India	157,000,000	146,803,659,362	114,448	Shows
4	Sony SAB	82,000,000	99,161,162,424	69,799	Shows
5	Kids Diana Show	112,000,000	92,619,392,007	1,099	People & Blogs
6	Like Nastya	106,000,000	90,098,693,755	803	People & Blogs
7	WWE	95,400,000	76,834,983,796	69,433	Sports
8	Vlad and Niki	98,400,000	76,670,329,838	564	Entertainment
9	Zee TV	69,700,000	72,011,528,995	126,504	Entertainment
10	Colors TV	63,700,000	60,611,184,375	111,143	Shows
11	Movieclips	59,400,000	59,274,715,045	38,94	Film & Animation
12	Zee Music Company	95,700,000	57,174,762,737	8,387	Music
13	El Reino Infantil	57,800,000	56,774,393,613	1,478	Music
14	netd müzik	24,000,000	55,715,693,427	23,146	Music
15	Ryan's World	34,900,000	55,066,267,708	2,422	Entertainment
16	ABS-CBN Entertainment	44,000,000	50,145,684,242	192,022	People & Blogs
17	Toys and Colors	39,800,000	46,557,552,060	978	Entertainment
18	ChuChu TV Nursery Rhymes & Kids Songs	65,400,000	45,372,725,200	612	Education
19	Super Simple Songs - Kids Songs	38,800,000	44,250,283,638	729	Education
20	Wave Music	57,600,000	40,411,613,168	19,418	Music
21	HAR PAL GEO	43,300,000	39,910,183,092	97,138	Entertainment
22	Little Baby Bum - Nursery Rhymes & Kids Songs	40,800,000	39,343,050,964	2,4	Education

The bottom screenshot shows the completed table with 22 rows of YouTube channel data. The table has columns for rank, Youtuber, subscribers, video views, video count, and category. The data is imported from a YouTube channel list.

Quá trình thu thập dữ liệu là một bước quan trọng và đảm bảo rằng chúng ta có nguồn dữ liệu chính xác và đầy đủ để tiến hành phân tích và khám phá các yếu tố quan trọng trong việc xây dựng và quản lý kênh YouTube.

III. Khám phá và xử lý dữ liệu:

Bước thứ ba trong quá trình phân tích dữ liệu là khám phá dữ liệu. Sau khi đã thu thập dữ liệu từ nguồn đáng tin cậy, tôi sử dụng các thư viện và công cụ xử lý dữ liệu của Python để tiến hành các bước xử lý và chuẩn bị dữ liệu cho việc trực quan hóa và phân tích.

Cụ thể, tôi sử dụng các thư viện như Pandas, NumPy và Matplotlib để thực hiện các bước xử lý dữ liệu. Từ bảng dữ liệu ban đầu, tôi thực hiện việc làm sạch và chuyển đổi dữ liệu để nó trở nên dễ hiểu và thích hợp cho việc trực quan hóa và phân tích sau này.

Qua các bước xử lý dữ liệu, tôi đã loại bỏ các giá trị thiếu, xử lý dữ liệu không hợp lệ và áp dụng các phép tính toán cần thiết để thu được thông tin quan trọng từ bộ dữ liệu. Kết quả là ta thu được bộ dữ liệu sạch và chuẩn bị sẵn sàng cho các bước trực quan hóa và phân tích tiếp theo.

Các chi tiết về các bước xử lý dữ liệu và phân tích sẽ được tôi đính kèm trong file notebook, giúp bạn có cái nhìn rõ ràng và cụ thể hơn về quá trình xử lý dữ liệu và các kết quả thu được từ bộ dữ liệu Top 1000 YouTube Channels.

	A	B	C	D	E	F	G	H	I	J
	Youtube Channel		Subscribers	Video Views	Video Count	Category	Started			
1	0 T-Series		243000000	224881706554	19757	Music	2006			
2	1 Cocomelon - Nursery Rhymes		160000000	161435301255	927	Education	2006			
3	2 SET India		157000000	146246275148	113723	Shows	2006			
4	3 Sony SAB		81600000	98603821529	69353	Shows	2007			
5	4 àà Kids Diana Show		111000000	92321006440	1097	People & Blogs	2015			
6	5 Like Nastya		106000000	89556905855	800	People & Blogs	2016			
7	6 WWE		95200000	76590809077	69232	Sports	2007			
8	7 Vlad and Niki		97700000	76255951806	561	Entertainment	2018			
9	8 Zee TV		69400000	71539733613	126034	Entertainment	2005			
10	9 Colors TV		63200000	60190628294	11059	Shows	2008			
11	10 Movieclips		59400000	59176881671	38871	Film & Animation	2006			
12	11 Zee Music Company		95500000	56961287918	8349	Music	2014			
13	12 El Reino Infantil		57600000	56551023544	1471	Music	2011			
14	13 netd mÀ/zik		24000000	55536940351	23036	Music	2014			
15	14 Ryan's World		34900000	54991159129	2414	Entertainment	2015			
16	15 ABS-CBN Entertainment		43900000	50090155660	191466	People & Blogs	2008			
17	16 Toys and Colors		39500000	46335699342	971	Entertainment	2016			
18	17 ChuChu TV Nursery Rhymes & Kids Songs		65100000	45127639835	605	Education	2013			
19	18 Super Simple Songs - Kids Songs		38600000	44034714118	724	Education	2006			
20	19 Wave Music		57500000	40343309988	19404	Music	2014			
21	20 HAR PAL GEO		42800000	39343220321	96133	Entertainment	2008			
22	21 Little Baby Bum - Nursery Rhymes & Kids Songs		40800000	39304951144	2393	Education	2011			

Dữ liệu làm sạch.

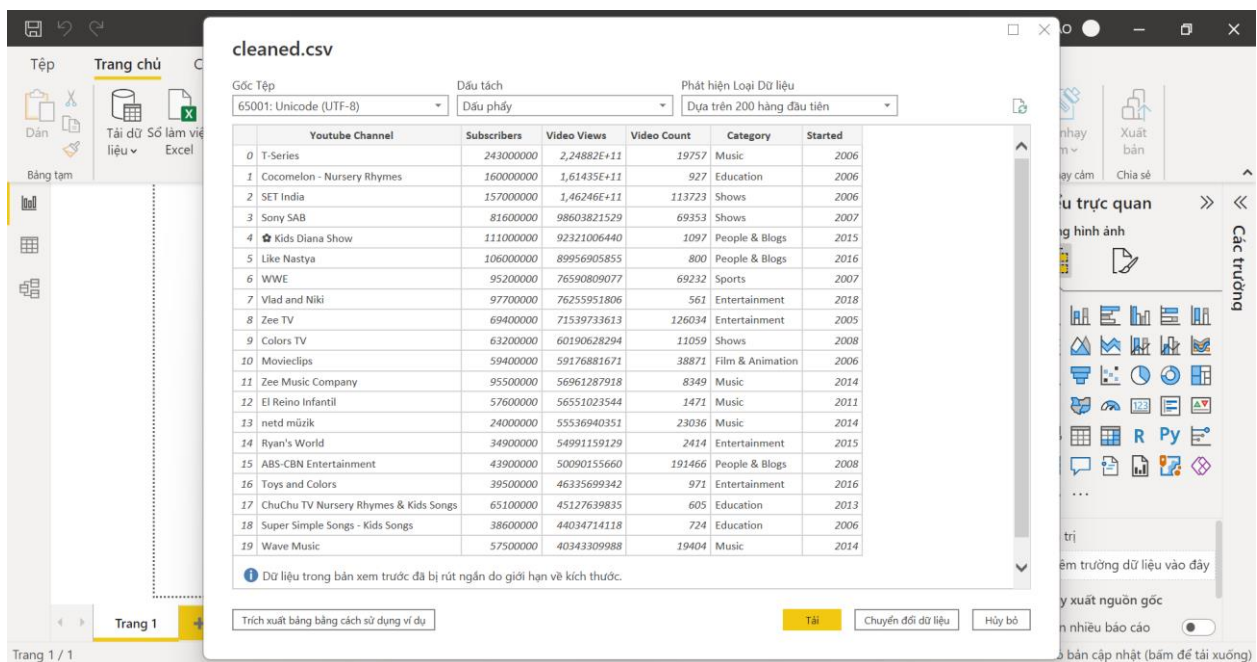
IV. Áp dụng phương pháp phân tích:

Bước tiếp theo trong quá trình phân tích dữ liệu là áp dụng phương pháp phân tích để hiểu rõ hơn về thông tin chứa đựng trong bộ dữ liệu Top 1000 YouTube Channels. Sau khi đã có dữ liệu sạch và đã xử lý, tôi sử dụng các thư viện phân tích và trực quan hóa dữ liệu trong Python để khám phá các mô hình, xu hướng và mối quan hệ trong dữ liệu.

Tôi sử dụng các thư viện như Matplotlib, Seaborn và Plotly để tạo các biểu đồ, đồ thị và biểu đồ tương tác. Các biểu đồ này giúp tôi trực quan hóa dữ liệu và hiểu rõ hơn về mối quan hệ giữa các biến số trong bộ dữ liệu.

Tôi cũng sử dụng các phép tính thống kê và phân tích để tìm hiểu các đặc điểm, xu hướng và mô hình trong dữ liệu.

Ngoài ra, để trình bày và thể hiện kết quả phân tích một cách chuyên sâu hơn, tôi đã tạo một Dashboard bằng Power BI. Dashboard này cung cấp một giao diện trực quan và tương tác để trình bày thông tin, đồ thị và kết quả phân tích một cách dễ hiểu và hấp dẫn hơn.



The screenshot shows a CSV file named 'cleaned.csv' with the following data:

	Youtube Channel	Subscribers	Video Views	Video Count	Category	Started
0	T-Series	243000000	2,24882E+11	19757	Music	2006
1	Cocomelon - Nursery Rhymes	160000000	1,61435E+11	927	Education	2006
2	SET India	157000000	1,46246E+11	113723	Shows	2006
3	Sony SAB	81600000	98603821529	69353	Shows	2007
4	Kids Diana Show	111000000	92321006440	1097	People & Blogs	2015
5	Like Nastya	106000000	89956905855	800	People & Blogs	2016
6	WWE	95200000	76590809077	69232	Sports	2007
7	Vlad and Niki	97700000	76255951806	561	Entertainment	2018
8	Zee TV	69400000	71539733613	126034	Entertainment	2005
9	Colors TV	63200000	60190628294	11059	Shows	2008
10	Movieclips	59400000	59176881671	38871	Film & Animation	2006
11	Zee Music Company	95500000	56961287918	8349	Music	2014
12	El Reino Infantil	57600000	56551023544	1471	Music	2011
13	netd müzik	24000000	55536940351	23036	Music	2014
14	Ryan's World	34900000	54991159129	2414	Entertainment	2015
15	ABS-CBN Entertainment	43900000	50090155660	191466	People & Blogs	2008
16	Toys and Colors	39500000	46335699342	971	Entertainment	2016
17	ChuChu TV Nursery Rhymes & Kids Songs	65100000	45127639835	605	Education	2013
18	Super Simple Songs - Kids Songs	38600000	44034714118	724	Education	2006
19	Wave Music	57500000	40343309988	19404	Music	2014

BÁO CÁO PHÂN TÍCH DỮ LIỆU TOP 1000 YOUTUBE CHANNELS THEO DANH MỤC

Thời Gian
Tất cả

Danh Mục
Tất cả

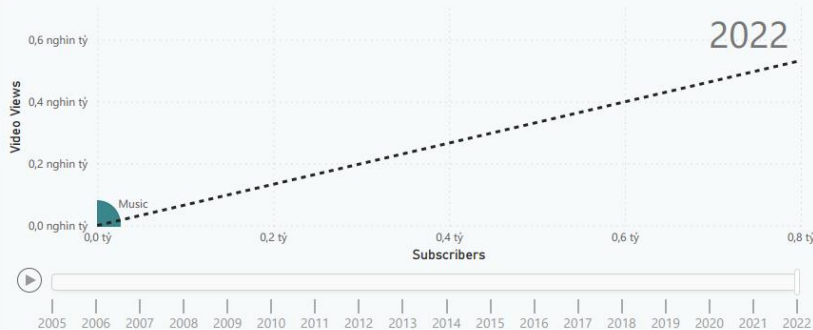
Các Danh Mục Kênh



Lượng Đăng Ký
19013,65 triệu

Số Lượt Xem
12661774,06 triệu

Số Lượng Video
16,15 triệu



Đặt câu hỏi về dữ liệu của bạn

Hãy thử một trong các bước sau để bắt đầu

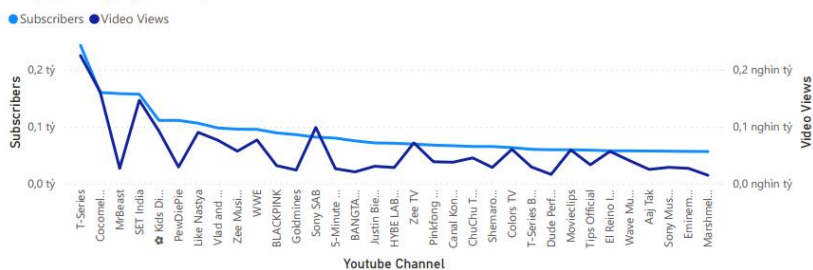
- average started
- count categories
- average column 1

BÁO CÁO PHÂN TÍCH DỮ LIỆU TOP 1000 YOUTUBE CHANNELS

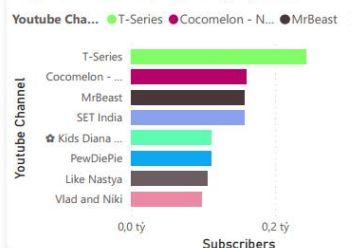
Thời Gian
Tất cả

Danh Mục
Tất cả

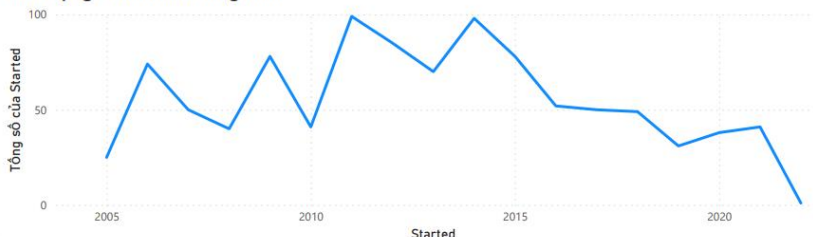
Lượng Đăng Ký Và Lượt Xem Theo Kênh



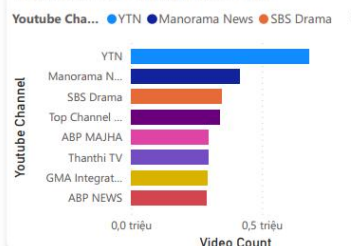
Top Kênh Có Lượng Đăng Ký Cao



Số Lượng Kênh Theo Từng Năm



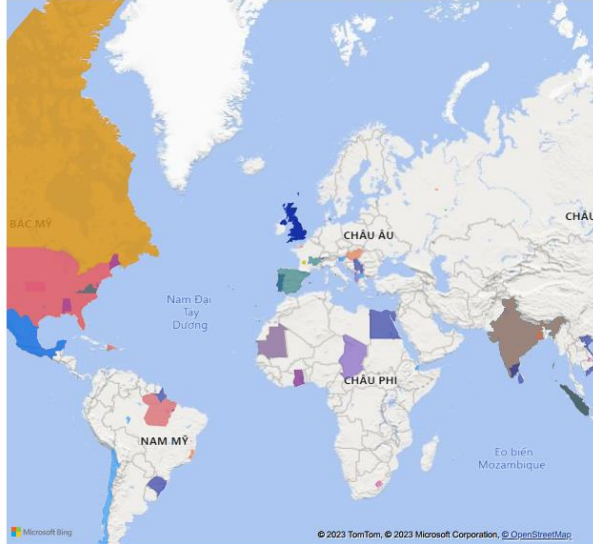
Top Kênh Có Lượng Video Cao



BÁO CÁO PHÂN TÍCH DỮ LIỆU TOP 1000 YOUTUBE CHANNELS

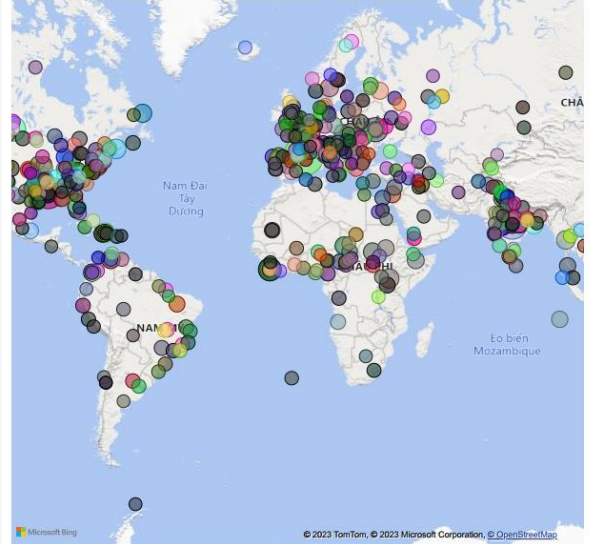
Phân Bố Danh Mục Trên Toàn Thế Giới

Category Autos & Ve... Comedy Education Entertain... Film & ... Gaming Howto ...



Phân Bố Kênh Trên Toàn Thế Giới

Youtube Channel T-Series Cocomelon - ... MrBeast SET India Kids Di... PewDiePie



Thông qua việc áp dụng phương pháp phân tích và sử dụng các công cụ phân tích và trực quan hóa, tôi có thể tìm ra các xu hướng, mô hình và thông tin quan trọng trong bộ dữ liệu Top 1000 YouTube Channels. Những phân tích này sẽ giúp ta hiểu rõ hơn về sự phát triển và thành công của các kênh YouTube, mối quan hệ giữa số lượng đăng ký, lượt xem và số lượng video, cũng như đánh giá và đưa ra gợi ý cho việc xây dựng và quản lý kênh YouTube trong tương lai.

Tất cả các kết quả phân tích, biểu đồ, và thông tin chi tiết về quá trình phân tích sẽ được trình bày chi tiết và minh bạch trong file notebook và Dashbroad, giúp bạn có cái nhìn tổng quan và chi tiết về bộ dữ liệu Top 1000

YouTube Channels và các kết quả phân tích mà tôi đã thực hiện.

V. Trình bày kết quả:

Dự án phân tích dữ liệu trên top 1000 YouTube Channels đã đem lại nhiều thông tin quan trọng về môi trường YouTube và yếu tố ảnh hưởng đến sự phát triển và thành công của các kênh. Dưới đây là những kết quả đạt được trong quá trình phân tích:

1. Phân loại các danh mục kênh: Dự án đã xác định và phân loại các danh mục kênh phổ biến nhất trên YouTube. Điều này giúp hiểu rõ hơn về sở thích và nhu cầu giải trí của người dùng YouTube, từ đó tạo ra các nội dung phù hợp và thu hút đối tượng khán giả mong muốn.
2. Tương quan giữa lượt đăng ký, lượt xem và số lượng video: Qua phân tích, đã được nhìn thấy mối quan hệ giữa các yếu tố này. Lượt đăng ký có mối tương quan mạnh với lượt xem, trong khi số lượng video không có mối tương quan đáng kể. Điều này cho thấy rằng sự quan tâm của khán giả chủ yếu tập trung vào lượt xem và chất lượng nội dung, chứ không phải chỉ số lượng video.
3. Các kênh nổi tiếng: Dự án đã xác định được các kênh có số lượng người đăng ký cao nhất trên YouTube. Việc nghiên cứu các kênh này có thể giúp hiểu được các yếu tố thành công và chiến lược xây dựng kênh hiệu quả.

4. Phân tích theo năm thành lập: Bằng cách phân tích số lượng kênh được tạo ra trong các năm khác nhau, chúng ta có thể nhận thấy xu hướng và sự phát triển của YouTube qua thời gian. Điều này có thể giúp các nhà sáng tạo nội dung lên kế hoạch và đưa ra chiến lược phù hợp.

5. Thay đổi và cạnh tranh trên YouTube: Dự án đã tìm hiểu về sự thay đổi và cạnh tranh trong các danh mục kênh và chất lượng nội dung trên YouTube. Điều này nhấn mạnh sự quan trọng của việc duy trì sự sáng tạo và đổi mới trong việc sản xuất nội dung để thu hút và giữ chân khán giả.

6. Tương quan giữa lượt xem và số lượng người đăng ký: Dự án đã phân tích tương quan giữa lượt xem và số lượng người đăng ký trên YouTube. Kết quả cho thấy rằng sự tăng trưởng số lượng người đăng ký có thể tác động đến lượt xem của một kênh. Điều này đặt ra tầm quan trọng của việc xây dựng và duy trì một cộng đồng người hâm mộ trung thành để tăng cường tương tác và lượt xem trên YouTube.

Tổng quát lại, dự án đã cung cấp cái nhìn tổng quan và thông tin chi tiết về môi trường YouTube và các yếu tố quan trọng đối với sự phát triển kênh. Những kết quả này có thể hỗ trợ các nhà sáng tạo nội dung và người dùng YouTube trong việc hiểu rõ hơn về cách tạo dựng và phát triển kênh thành công trên nền tảng này.

Lời kết:

Dự án này là một thử thách thú vị và cũng là cơ hội để tôi áp dụng kiến thức và kỹ năng trong lĩnh vực phân tích dữ liệu và trực quan hóa. Tuy nhiên, tôi nhận thức rằng dự án này chưa hoàn thiện và còn tiềm ẩn những thiếu sót.

Trong thời gian thực hiện dự án, tôi đã gặp phải một số hạn chế về thời gian và tài nguyên. Điều này đã ảnh hưởng đến khả năng của tôi trong việc tạo ra một bài trình bày hoàn chỉnh và chi tiết hơn. Tôi chưa thể đáp ứng tất cả các yêu cầu và mong muốn của dự án trong thời gian ngắn như mong đợi.

Tuy nhiên, tôi rất cảm kích sự thông cảm và đánh giá từ DataGenius. Tôi xin cam đoan rằng tôi sẽ tiếp tục nỗ lực hơn để hoàn thiện dự án này và đáp ứng mọi yêu cầu còn thiếu sót trong tương lai.

Cuối cùng, tôi xin gửi lời cảm ơn chân thành đến DataGenius và mọi người đã quan tâm và hỗ trợ trong quá trình dự án. Mọi thắc mắc và góp ý xin vui lòng liên hệ với tôi qua địa chỉ email để chúng ta có thể trao đổi và cải thiện dự án thêm.

Hết.
