**Báo Cáo Dự Án Cá Nhân**

Người trình bày: Huỳnh Bảo KHANG

Tên dự án: Phân tích top 1000 youtube channel

**Quá Trình:**

1. **Xác định mục tiêu:** Xác định mục tiêu, chọn chủ đề việc phân tích dữ liệu.
2. **Thu thập dữ liệu:** Thu thập dữ liệu cần thiết để giải quyết mục tiêu phân tích.
3. **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.
4. **Áp dụng phương pháp phân tích:** Tiến hành trực quan hóa và phân tích.
5. **Trình bày kết quả:** Tổng kết lại quá trình, kết quả đạt được.

**Bài Làm:**

1. **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.

1. **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.

Ảnh có chứa văn bản, ảnh chụp màn hình, hàng, Phông chữ

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, số, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, Sơ đồ, hàng

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động

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.

1. **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.

Ảnh có chứa văn bản, số, phần mềm, Phông chữ

Mô tả được tạo tự động

Dữ liệu làm sạch.

1. **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 Dashbroad bằng Power BI. Dashbroad 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.

Ảnh có chứa văn bản, phần mềm, số, Biểu tượng máy tính

Mô tả được tạo tự động

Ảnh có chứa văn bản, Phông chữ, hàng, số

Mô tả được tạo tự động Ảnh có chứa văn bản, ảnh chụp màn hình, hàng, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, bản đồ, ảnh chụp màn hình, tập bản đồ

Mô tả được tạo tự động

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.

1. **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.**