



Giới thiệu

Mục tiêu của Notebook là phân tích Kênh YouTube được đăng ký nhiều nhất, với dữ liệu thời gian thực mới nhất được thu thập từ SUKHMANDEEP SINGH BRAR.

Nội dung Notebook gồm 5 phần sau:

1. Khám phá Dataset về số lượng các kênh Youtube
2. Số lượng kênh có thương hiệu và không có thương hiệu
3. Số lượng người đăng ký của các kênh YouTube
4. So sánh các kênh có thương hiệu và không có thương hiệu riêng biệt về số lượng người đăng kí
5. Các ngôn ngữ phổ biến
6. Các loại hình thức phổ biến của các kênh YouTube
7. Tổng số lượng người đăng kí của từng danh mục
8. So sánh độ phát triển của các kênh trên 1 lĩnh vực
9. Số lượng kênh YouTube theo từng quốc gia
10. Kết luận

✓ 1. Khám phá Dataset về cổ phiếu META

Bảng dữ liệu bao gồm 6 cột, tương ứng thuộc tính sau:

- Name: Tên kênh
- Brand channel: Kênh thương hiệu
- Subscribers (millions): Người đăng ký (triệu)
- Primary language: Ngôn ngữ
- Category: Thể
- Country: Quốc gia

Dữ liệu có 50 mẫu, tương ứng 50 kênh Youtube.

```
# load thư viện
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Cài đặt các thư viện cần thiết
!pip install pandas matplotlib seaborn
```

```
# Nhập các thư viện cần thiết
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (2.0.3)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (3.7.1)
Requirement already satisfied: seaborn in /usr/local/lib/python3.10/dist-packages (0.13.1)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2024.1)
Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/dist-packages (from pandas) (1.25.2)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (1.2.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (4.51.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (24.0)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (3.1.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
```

```
import pandas as pd
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)
data
```



	Name	Brand channel	Subscribers (millions)	Primary language	Category	Country
0	T-Series	Yes	249.0	Hindi[7][8]	Music	India
1	MrBeast	No	184.0	English	Entertainment	United States
2	Cocomelon	Yes	165.0	English	Education	United States
3	Sony Entertainment Television India	Yes	162.0	Hindi[9]	Entertainment	India
4	Kids Diana Show	Yes	113.0	English[10][11][12]	Entertainment	Ukraine-United States
5	PewDiePie	No	111.0	English	Entertainment	Sweden
6	Like Nastya	No	107.0	English	Entertainment	Russia-United States
7	Vlad and Niki	No	101.0	English	Entertainment	Russia-United States
8	Zee Music Company	Yes	99.5	Hindi[13][14]	Music	India
9	WWE	Yes	97.1	English	Sports	United States
10	Blackpink	No	91.2	Korean	Music	South Korea
11	Goldmines	Yes	89.5	Hindi	Film	India
12	Sony SAB	Yes	85.2	Hindi	Entertainment	India
13	5-Minute Crafts	Yes	80.2	English	How-to	Cyprus[a]
14	BangtanTV	No	76.4	Korean	Music	South Korea
15	Hybe Labels	Yes	72.6	Korean	Music	South Korea
16	Zee TV	Yes	72.4	Hindi	Entertainment	India
17	Justin Bieber	No	71.9	English	Music	Canada
18	Pinkfong	Yes	69.5	English	Education	South Korea
19	ChuChu TV Nursery Rhymes & Kids Songs	Yes	67.5	Hindi[16]	Education	India
20	Canal KondZilla	Yes	66.6	Portuguese	Music	Brazil
21	Colors TV	Yes	66.6	Hindi	Entertainment	India
22	Shemaroo Entertainment	Yes	66.3	Hindi	Music	India
23	T-Series Bhakti Sagar	Yes	62.4	Hindi	Music	India
24	Tips Official	Yes	60.8	Hindi	Entertainment	India
25	Movieclips	Yes	59.8	English	Film	United States
26	Dude Perfect	No	59.7	English	Sports	United States
27	El Reino Infantil	Yes	59.7	Spanish	Music	Argentina
28	Aaj Tak	Yes	58.8	Hindi	News	India
29	Wave Music	Yes	58.7	Bhojpuri	Music	India
30	Sony Music India	Yes	57.8	Hindi	Music	India
31	Eminem	No	57.4	English	Music	United States
32	Marshmello	No	56.5	English	Music	United States
33	Yash Raj Films	Yes	55.7	Hindi	Music	India
34	Infobells	Yes	55.6	Hindi	Education	India
35	LooLoo Kids	Yes	54.7	English	Music	Romania
36	Taylor Swift	No	54.1	English	Music	United States
37	Ed Sheeran	No	53.7	English	Music	United Kingdom
38	Billion Surprise Toys	Yes	53.0	English	Entertainment	United States
39	Ariana Grande	No	52.9	English	Music	United States

✓ 2. Số lượng kênh có thương hiệu và không có thương hiệu

```
import pandas as pd
import matplotlib.pyplot as plt

# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Đếm số lượng kênh có thương hiệu và không có thương hiệu
brand_counts = data['Brand channel'].value_counts()

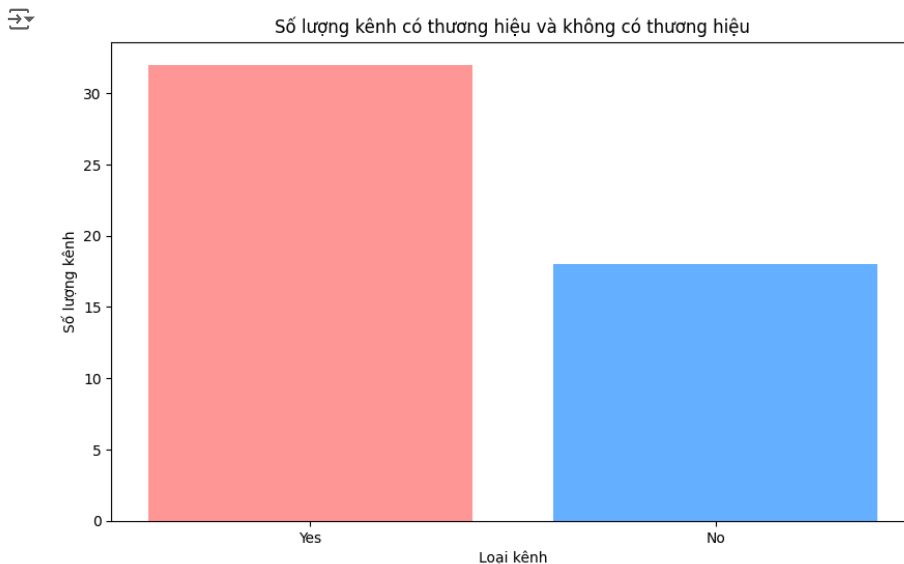
# Tạo biểu đồ cột để hiển thị số lượng kênh có thương hiệu và không có thương hiệu
plt.figure(figsize=(10, 6))

# Xác định vị trí các cột và chiều cao các cột
bar_positions = brand_counts.index
bar_heights = brand_counts.values

# Vẽ biểu đồ cột
plt.bar(bar_positions, bar_heights, color=['#ff9999', '#66b3ff'])

# Thêm tiêu đề và nhãn
plt.title('Số lượng kênh có thương hiệu và không có thương hiệu')
plt.xlabel('Loại kênh')
plt.ylabel('Số lượng kênh')

# Hiển thị biểu đồ
plt.show()
```



Qua biểu đồ trên ta thấy được:

- Kênh có thương hiệu có 32 kênh chiếm 64% tổng các kênh Youtube
- Kênh không có thương hiệu có 18 kênh chiếm % tổng các kênh Youtube

Kênh YouTube có thương hiệu (Brand Channel): Thường được gắn với một thương hiệu, công ty hoặc sản phẩm cụ thể. Sử dụng logo và hình ảnh thương hiệu để tạo sự nhận diện. Nội dung được sản xuất chuyên nghiệp, có tính nhất quán về phong cách và chất lượng. Có thể được quản lý bởi nhiều người (quản trị viên), giúp việc kiểm soát và phát triển nội dung hiệu quả hơn. Thường sử dụng cho mục đích tiếp thị, quảng cáo sản phẩm hoặc dịch vụ của thương hiệu. Thường có URL tùy chỉnh để dễ dàng nhận diện và chia sẻ.

Kênh YouTube không có thương hiệu (Personal Channel): Thường được gắn với tên cá nhân hoặc biệt danh của người dùng. Sử dụng hình ảnh cá nhân hoặc không có yêu cầu đặc biệt về hình ảnh thương hiệu. Nội dung có thể đa dạng về chủ đề và phong cách, không nhất thiết phải tuân theo một chuẩn mực nhất định. Thường chỉ do một người quản lý, người sở hữu kênh cũng chính là người sản xuất và đăng tải nội dung. Chủ yếu được sử dụng cho mục đích giải trí, chia sẻ sở thích cá nhân, hoặc làm vlog. Sử dụng URL tiêu chuẩn do YouTube cung cấp, không tùy chỉnh.

✓ 3. Số lượng người đăng ký của các kênh YouTube

```
import pandas as pd
import matplotlib.pyplot as plt

# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Sắp xếp dữ liệu theo số lượng người đăng ký
data_sorted = data.sort_values(by='Subscribers (millions)', ascending=False)

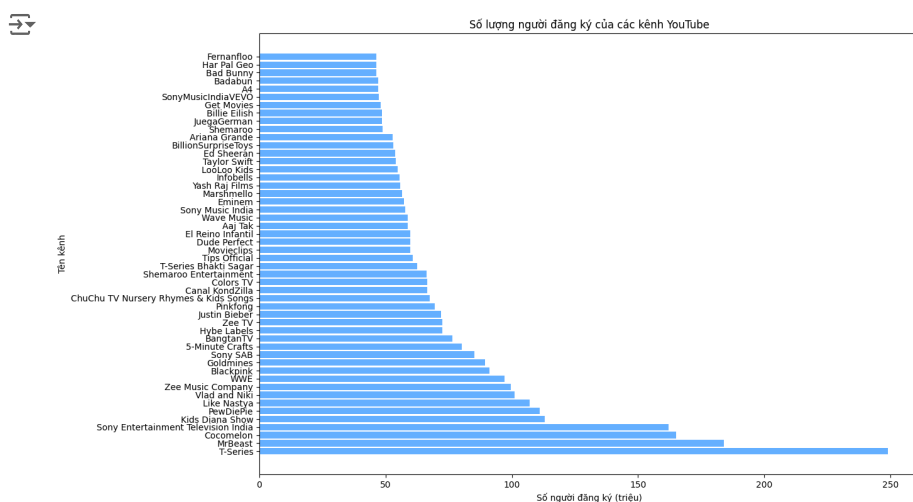
# Tạo biểu đồ cột ngang để hiển thị số lượng người đăng ký của mỗi kênh
plt.figure(figsize=(14, 8))

# Xác định vị trí các cột và chiều cao các cột
bar_positions = data_sorted['Name']
bar_heights = data_sorted['Subscribers (millions)']

# Vẽ biểu đồ cột ngang
plt.barh(bar_positions, bar_heights, color='#66b3ff')

# Thêm tiêu đề và nhãn
plt.title('Số lượng người đăng ký của các kênh YouTube')
plt.xlabel('Số người đăng ký (triệu)')
plt.ylabel('Tên kênh')

# Hiển thị biểu đồ
plt.tight_layout()
plt.show()
```



- Kênh có số lượng người đăng kí cao nhất là 249tr người.
- kênh có số lượng người đăng kí thấp nhất 46,3tr người.

✓ 4. So sánh các kênh có thương hiệu và ko thương hiệu riêng biệt về số lượng người đăng kí.

```
import pandas as pd
import matplotlib.pyplot as plt

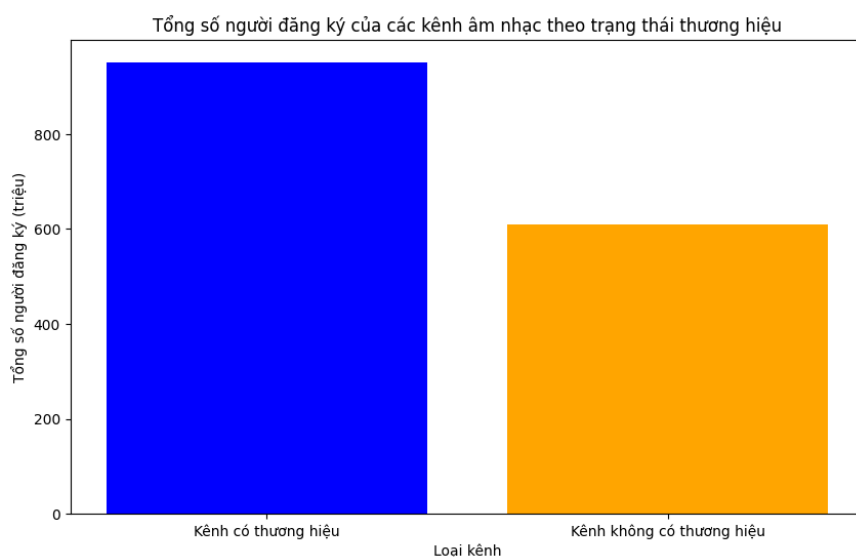
# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Lọc dữ liệu cho thể loại âm nhạc
music_channels = data[data['Category'] == 'Music']

# Tính tổng số người đăng ký cho các kênh có thương hiệu và không có thương hiệu
brand_subscribers = music_channels[music_channels['Brand channel'] == 'Yes']['Subscribers (millions)'].sum()
non_brand_subscribers = music_channels[music_channels['Brand channel'] == 'No']['Subscribers (millions)'].sum()

# Chuẩn bị dữ liệu cho biểu đồ cột
labels = ['Kênh có thương hiệu', 'Kênh không có thương hiệu']
subscribers = [brand_subscribers, non_brand_subscribers]

# Vẽ biểu đồ cột
plt.figure(figsize=(10, 6))
plt.bar(labels, subscribers, color=['blue', 'orange'])
plt.xlabel('Loại kênh')
plt.ylabel('Tổng số người đăng ký (triệu)')
plt.title('Tổng số người đăng ký của các kênh âm nhạc theo trạng thái thương hiệu')
plt.show()
```



```
import pandas as pd
import matplotlib.pyplot as plt

# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Lọc dữ liệu cho thể loại giải trí
entertainment_channels = data[data['Category'] == 'Entertainment']

# Tính tổng số người đăng ký cho các kênh có thương hiệu và không có thương hiệu
brand_subscribers = entertainment_channels[entertainment_channels['Brand channel'] == 'Yes']['Subscribers (millions)'].sum()
non_brand_subscribers = entertainment_channels[entertainment_channels['Brand channel'] == 'No']['Subscribers (millions)'].sum()

# Chuẩn bị dữ liệu cho biểu đồ cột
labels = ['Kênh có thương hiệu', 'Kênh không có thương hiệu']
subscribers = [brand_subscribers, non_brand_subscribers]

# Vẽ biểu đồ cột
plt.figure(figsize=(10, 6))
plt.bar(labels, subscribers, color=['blue', 'orange'])
plt.xlabel('Loại kênh')
plt.ylabel('Tổng số người đăng ký (triệu)')
plt.title('Tổng số người đăng ký của các kênh giải trí theo trạng thái thương hiệu')
plt.show()
```



```
import pandas as pd
import matplotlib.pyplot as plt

# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Lọc dữ liệu cho thể loại giáo dục
education_channels = data[data['Category'] == 'Education']

# Tính tổng số người đăng ký cho các kênh có thương hiệu và không có thương hiệu
brand_subscribers = education_channels[education_channels['Brand channel'] == 'Yes']['Subscribers (millions)'].sum()
non_brand_subscribers = education_channels[education_channels['Brand channel'] == 'No']['Subscribers (millions)'].sum()

# Chuẩn bị dữ liệu cho biểu đồ cột
labels = ['Kênh có thương hiệu', 'Kênh không có thương hiệu']
subscribers = [brand_subscribers, non_brand_subscribers]

# Vẽ biểu đồ cột
plt.figure(figsize=(10, 6))
plt.bar(labels, subscribers, color=['blue', 'orange'])
plt.xlabel('Loại kênh')
plt.ylabel('Tổng số người đăng ký (triệu)')
plt.title('Tổng số người đăng ký của các kênh giáo dục theo trạng thái thương hiệu')
plt.show()
```



```
import pandas as pd
import matplotlib.pyplot as plt

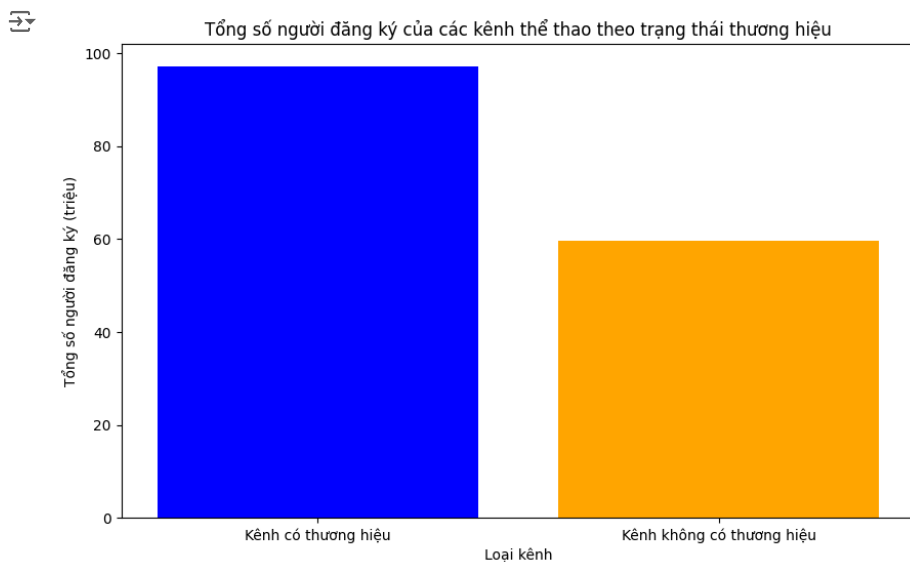
# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Lọc dữ liệu cho thể loại thể thao
sports_channels = data[data['Category'] == 'Sports']

# Tính tổng số người đăng ký cho các kênh có thương hiệu và không có thương hiệu
brand_subscribers = sports_channels[sports_channels['Brand channel'] == 'Yes']['Subscribers (millions)'].sum()
non_brand_subscribers = sports_channels[sports_channels['Brand channel'] == 'No']['Subscribers (millions)'].sum()

# Chuẩn bị dữ liệu cho biểu đồ cột
labels = ['Kênh có thương hiệu', 'Kênh không có thương hiệu']
subscribers = [brand_subscribers, non_brand_subscribers]

# Vẽ biểu đồ cột
plt.figure(figsize=(10, 6))
plt.bar(labels, subscribers, color=['blue', 'orange'])
plt.xlabel('Loại kênh')
plt.ylabel('Tổng số người đăng ký (triệu)')
plt.title('Tổng số người đăng ký của các kênh thể thao theo trạng thái thương hiệu')
plt.show()
```

```
import pandas as pd
import matplotlib.pyplot as plt

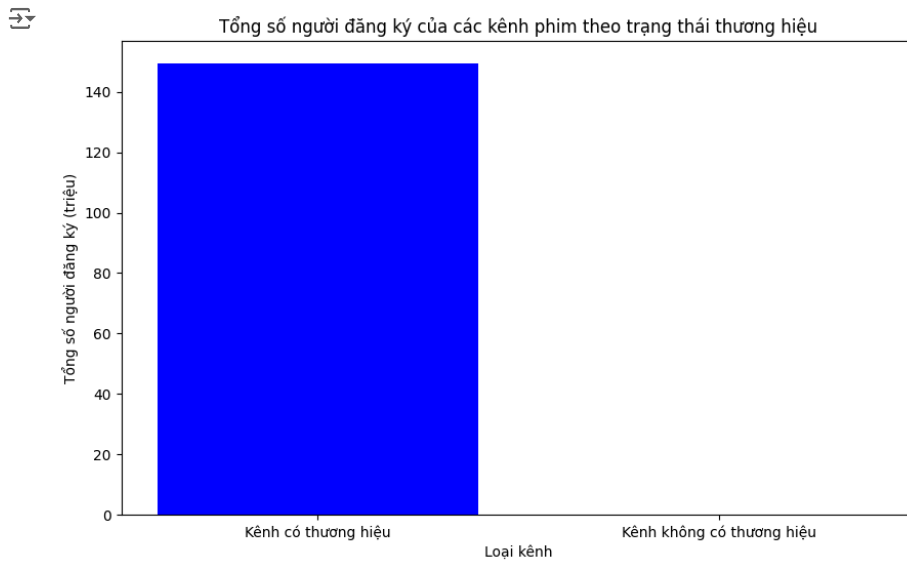
# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Lọc dữ liệu cho thể loại phim
film_channels = data[data['Category'] == 'Film']

# Tính tổng số người đăng ký cho các kênh có thương hiệu và không có thương hiệu
brand_subscribers = film_channels[film_channels['Brand channel'] == 'Yes']['Subscribers (millions)'].sum()
non_brand_subscribers = film_channels[film_channels['Brand channel'] == 'No']['Subscribers (millions)'].sum()

# Chuẩn bị dữ liệu cho biểu đồ cột
labels = ['Kênh có thương hiệu', 'Kênh không có thương hiệu']
subscribers = [brand_subscribers, non_brand_subscribers]

# Vẽ biểu đồ cột
plt.figure(figsize=(10, 6))
plt.bar(labels, subscribers, color=['blue', 'orange'])
plt.xlabel('Loại kênh')
plt.ylabel('Tổng số người đăng ký (triệu)')
plt.title('Tổng số người đăng ký của các kênh phim theo trạng thái thương hiệu')
plt.show()
```



```
import pandas as pd
import matplotlib.pyplot as plt

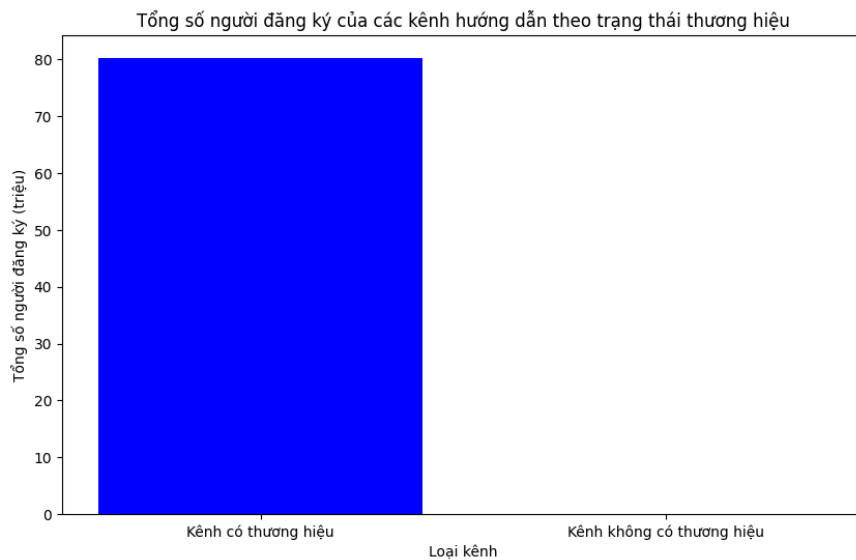
# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Lọc dữ liệu cho thể loại hướng dẫn
howto_channels = data[data['Category'] == 'How-to']

# Tính tổng số người đăng ký cho các kênh có thương hiệu và không có thương hiệu
brand_subscribers = howto_channels[howto_channels['Brand channel'] == 'Yes']['Subscribers (millions)'].sum()
non_brand_subscribers = howto_channels[howto_channels['Brand channel'] == 'No']['Subscribers (millions)'].sum()

# Chuẩn bị dữ liệu cho biểu đồ cột
labels = ['Kênh có thương hiệu', 'Kênh không có thương hiệu']
subscribers = [brand_subscribers, non_brand_subscribers]

# Vẽ biểu đồ cột
plt.figure(figsize=(10, 6))
plt.bar(labels, subscribers, color=['blue', 'orange'])
plt.xlabel('Loại kênh')
plt.ylabel('Tổng số người đăng ký (triệu)')
plt.title('Tổng số người đăng ký của các kênh hướng dẫn theo trạng thái thương hiệu')
plt.show()
```



```
import pandas as pd
import matplotlib.pyplot as plt

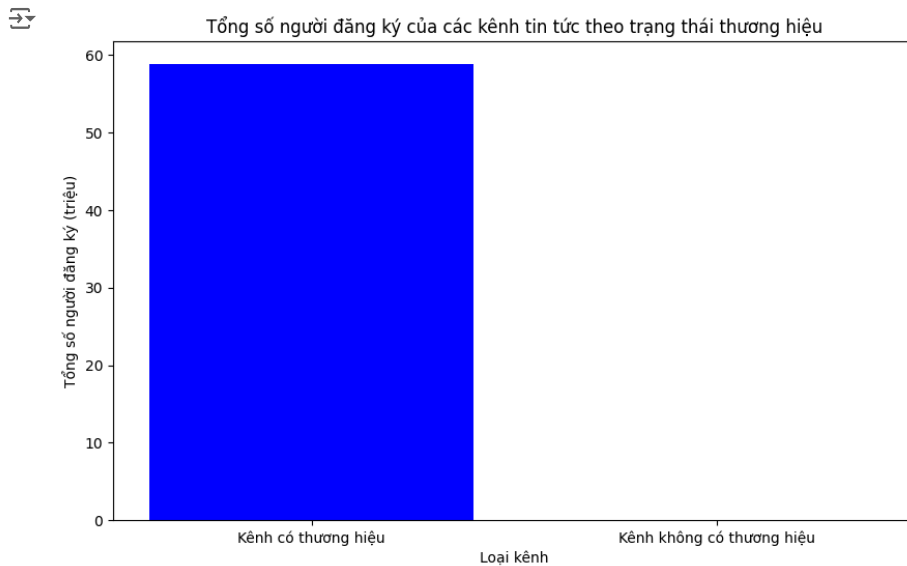
# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Lọc dữ liệu cho thể loại tin tức
news_channels = data[data['Category'] == 'News']

# Tính tổng số người đăng ký cho các kênh có thương hiệu và không có thương hiệu
brand_subscribers = news_channels[news_channels['Brand channel'] == 'Yes']['Subscribers (millions)'].sum()
non_brand_subscribers = news_channels[news_channels['Brand channel'] == 'No']['Subscribers (millions)'].sum()

# Chuẩn bị dữ liệu cho biểu đồ cột
labels = ['Kênh có thương hiệu', 'Kênh không có thương hiệu']
subscribers = [brand_subscribers, non_brand_subscribers]

# Vẽ biểu đồ cột
plt.figure(figsize=(10, 6))
plt.bar(labels, subscribers, color=['blue', 'orange'])
plt.xlabel('Loại kênh')
plt.ylabel('Tổng số người đăng ký (triệu)')
plt.title('Tổng số người đăng ký của các kênh tin tức theo trạng thái thương hiệu')
plt.show()
```



```
import pandas as pd
import matplotlib.pyplot as plt

# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Lọc dữ liệu cho thể loại trò chơi
games_channels = data[data['Category'] == 'Games']

# Tính tổng số người đăng ký cho các kênh có thương hiệu và không có thương hiệu
brand_subscribers = games_channels[games_channels['Brand channel'] == 'Yes']['Subscribers (millions)'].sum()
non_brand_subscribers = games_channels[games_channels['Brand channel'] == 'No']['Subscribers (millions)'].sum()

# Chuẩn bị dữ liệu cho biểu đồ cột
labels = ['Kênh có thương hiệu', 'Kênh không có thương hiệu']
subscribers = [brand_subscribers, non_brand_subscribers]

# Vẽ biểu đồ cột
plt.figure(figsize=(10, 6))
plt.bar(labels, subscribers, color=['blue', 'orange'])
plt.xlabel('Loại kênh')
plt.ylabel('Tổng số người đăng ký (triệu)')
plt.title('Tổng số người đăng ký của các kênh trò chơi theo trạng thái thương hiệu')
plt.show()
```



- Lợi thế thương hiệu: Trên tất cả các danh mục, kênh thương hiệu luôn có tổng số người đăng ký cao hơn so với các kênh không có thương hiệu. Lợi thế này có thể bắt nguồn từ sự nhận biết, nguồn lực và sức mạnh tiếp thị đã có từ trước mà thương hiệu sở hữu.
- Thu hút người đăng ký: Các kênh mang thương hiệu được công nhận có thể thu hút người đăng ký hiệu quả hơn, có thể do chất lượng được cảm nhận, độ tin cậy và quảng cáo rộng rãi.
- Xu hướng cụ thể theo danh mục: Mặc dù xu hướng thống trị của thương hiệu là rõ ràng nhưng mức độ khác biệt lại khác nhau. Các danh mục như âm nhạc và giải trí có sự chênh lệch rõ rệt hơn, trong khi các danh mục như thể thao và cách làm có khoảng cách tương đối nhỏ hơn.

✓ 5. Ngôn ngữ phổ biến

```
import pandas as pd
import matplotlib.pyplot as plt

# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Đếm số lượng kênh cho mỗi ngôn ngữ
language_counts = data['Primary language'].value_counts()

# Tạo biểu đồ cột để hiển thị số lượng kênh cho mỗi ngôn ngữ
plt.figure(figsize=(14, 8))

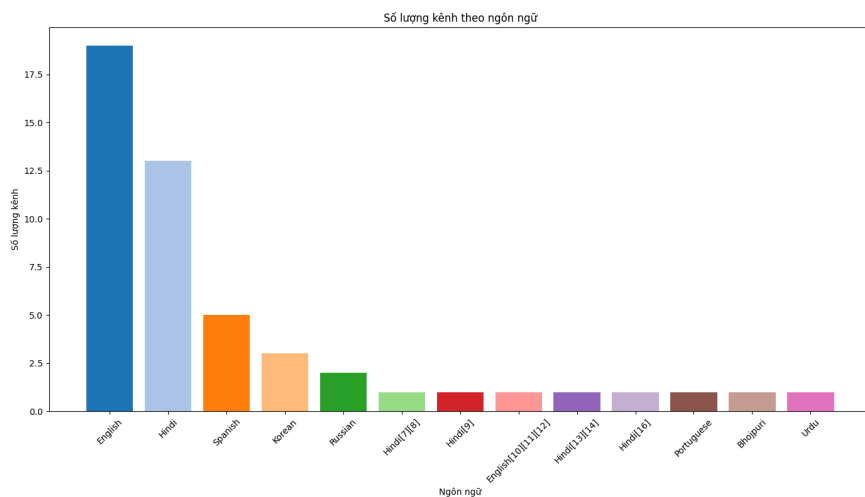
# Xác định vị trí và độ cao của cột
bar_positions = language_counts.index
bar_heights = language_counts.values

# Vẽ biểu đồ cột
plt.bar(bar_positions, bar_heights, color=plt.cm.tab20.colors)

# Thêm tiêu đề và nhãn
plt.title('Số lượng kênh theo ngôn ngữ')
plt.xlabel('Ngôn ngữ')
plt.ylabel('Số lượng kênh')

# Xoay nhãn trục x để dễ đọc hơn
plt.xticks(rotation=45)

# Hiển thị biểu đồ
plt.tight_layout()
plt.show()
```



- Tiếng Anh có 20 kênh sử dụng.
- Tiếng Hindi có 17 kênh sử dụng.
- tiếng Spanish có 5 kênh sử dụng.
- Tiếng korean có 3 kênh sử dụng.
- Tiếng Russian có 2 kênh sử dụng.
- Tiếng Bồ Đào Nha có 1 kênh sử dụng.
- Tiếng Bhojpuri có 1 kênh sử dụng.
- Tiếng Urdu có 1 kênh sử dụng.

=> Có thể nhận thấy Tiếng Anh là tiếng dùng phổ biến nhất

✓ 6. Các loại hình thức phổ biến của các kênh YouTube

```
import pandas as pd
import matplotlib.pyplot as plt

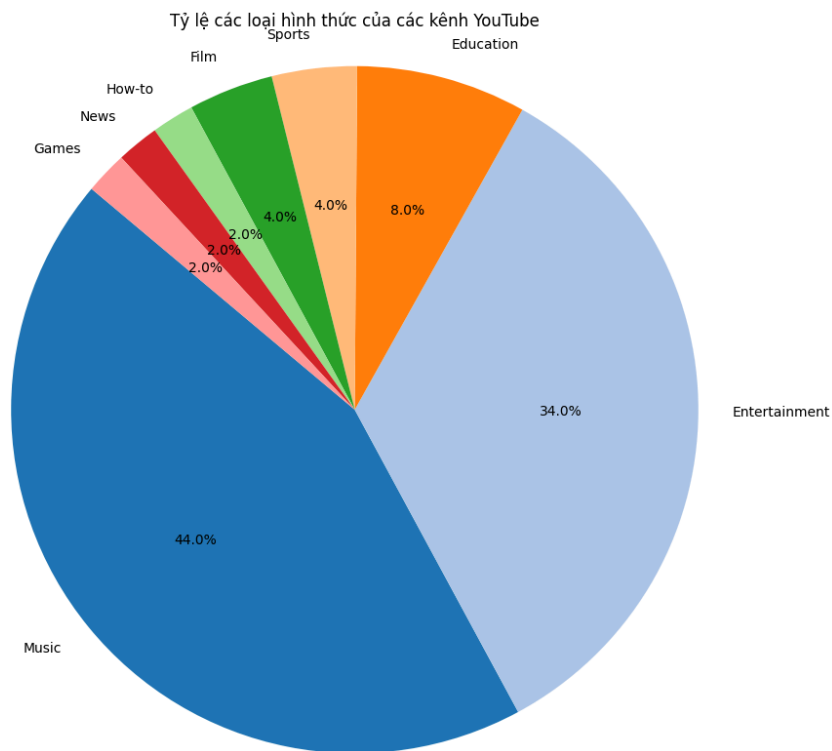
# Tải tệp CSV
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Đếm số lượng kênh cho mỗi loại hình thức
category_counts = data['Category'].value_counts()

# Tạo biểu đồ tròn
plt.figure(figsize=(10, 10))

# Xác định nhãn và kích thước cho biểu đồ tròn
labels = category_counts.index
sizes = category_counts.values

# Vẽ biểu đồ tròn
plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=140, colors=plt.cm.tab20.colors)
plt.axis('equal') # Đảm bảo tỉ lệ trục x và trục y bằng nhau để vẽ hình tròn.
plt.title('Tỷ lệ các loại hình thức của các kênh YouTube')
plt.show()
```



=> Music là hình thức phổ biến nhất

=> Games, News, How-to là hình thức chưa được phổ biến

✓ 7. Tổng số lượng người đăng kí của từng danh mục

```
import matplotlib.pyplot as plt
import seaborn as sns
# Tổng số người đăng ký theo danh mục
subscribers_by_category = data.groupby('Category')['Subscribers (millions)'].sum().sort_values(ascending=False)
print("\nTổng số người đăng ký theo danh mục:")
print(subscribers_by_category)

# Trực quan hóa
plt.figure(figsize=(14, 7))
sns.barplot(x=subscribers_by_category.values, y=subscribers_by_category.index, palette="cubehelix")
plt.xlabel('Subscribers (in millions)')
plt.title('Total Subscribers by Category')
plt.show()
```



Tổng số người đăng ký theo danh mục:

Category

Music 1559.4

Entertainment 1402.0

Education 357.6

Sports 156.8

Film 149.3

How-to 80.2

News 58.8

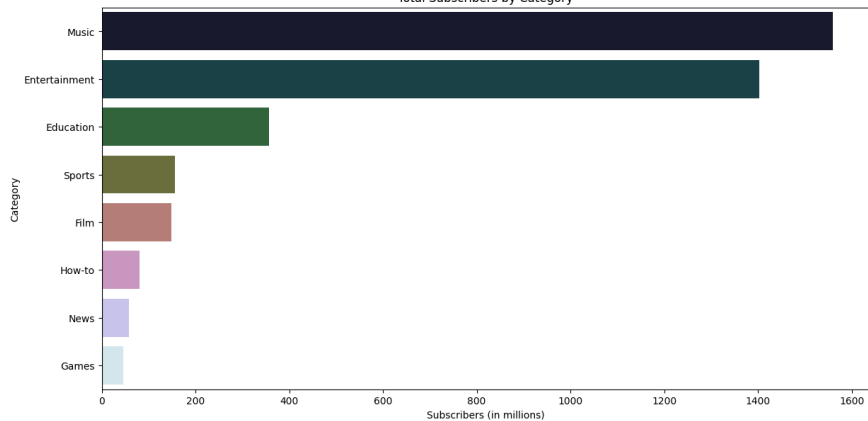
Games 46.3

Name: Subscribers (millions), dtype: float64

<ipython-input-7-87a5689cfdcf>:10: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. As

sns.barplot(x=subscribers_by_category.values, y=subscribers_by_category.index, palette='
Total Subscribers by Category



Âm nhạc (Music) và Giải trí (Entertainment) có tổng số lượng người đăng ký cao nhất, vượt trội so với các danh mục khác.

Giáo dục (Education) là danh mục tiếp theo với số lượng người đăng ký khá lớn, nhưng vẫn thấp hơn so với giải trí và âm nhạc.

Phim (Film) và Thể thao (Sports) là danh mục tiếp theo với số lượng người đăng ký khá thấp.

Tin tức (News) và Hướng dẫn (How-to) Trò chơi (Games) có số lượng người đăng ký thấp nhất trong tập dữ liệu.

8. So sánh độ phát triển của các kênh trên 1 lĩnh vực

▼ Music

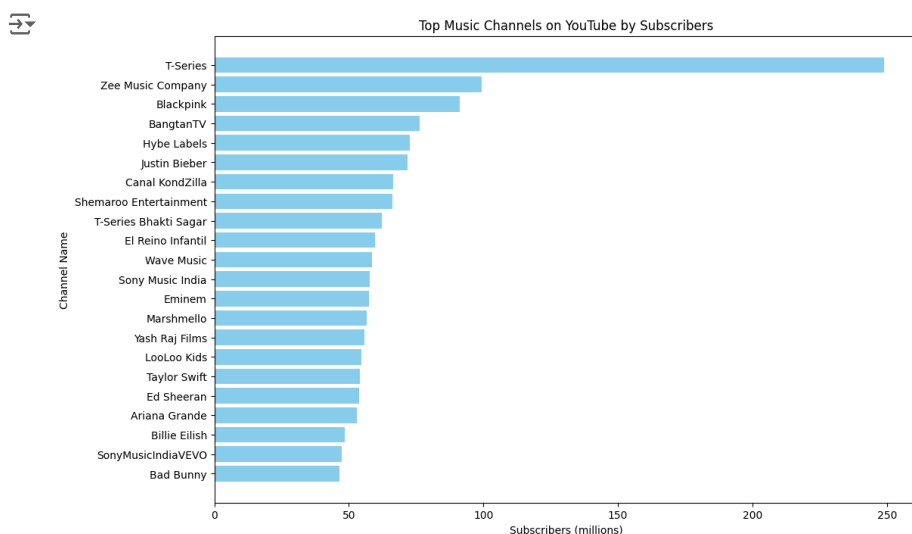

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Extract relevant data for music channels
music_channels = data[data['Category'] == 'Music']

# Sort channels by subscriber count
music_channels_sorted = music_channels.sort_values(by='Subscribers (millions)', ascending=False)

# Plotting the data
plt.figure(figsize=(12, 8))
plt.barh(music_channels_sorted['Name'], music_channels_sorted['Subscribers (millions)'], color='skyblue')
plt.xlabel('Subscribers (millions)')
plt.ylabel('Channel Name')
plt.title('Top Music Channels on YouTube by Subscribers')
plt.gca().invert_yaxis() # Highest values at the top
plt.show()
```



✓ Entertainment

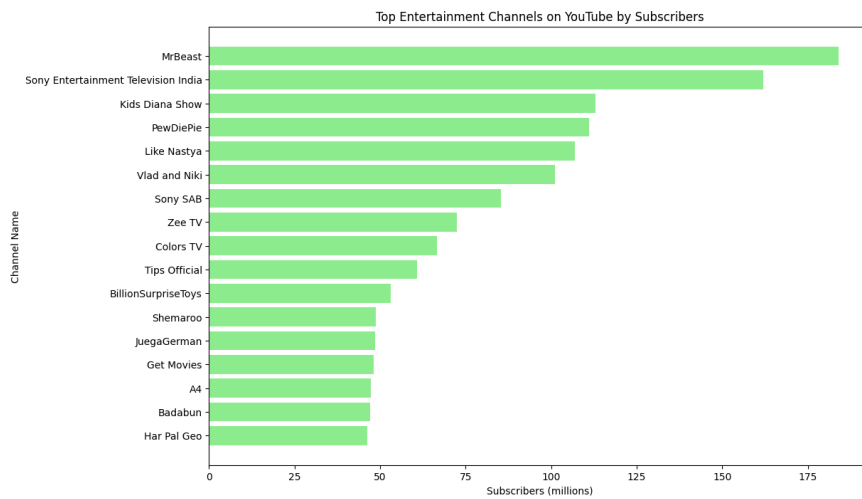
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Extract relevant data for entertainment channels
entertainment_channels = data[data['Category'] == 'Entertainment']

# Sort channels by subscriber count
entertainment_channels_sorted = entertainment_channels.sort_values(by='Subscribers (millions)', ascending=False)

# Plotting the data
plt.figure(figsize=(12, 8))
plt.barh(entertainment_channels_sorted['Name'], entertainment_channels_sorted['Subscribers (millions)'], color='lightgreen')
plt.xlabel('Subscribers (millions)')
plt.ylabel('Channel Name')
plt.title('Top Entertainment Channels on YouTube by Subscribers')
plt.gca().invert_yaxis() # Highest values at the top
plt.show()
```



Education

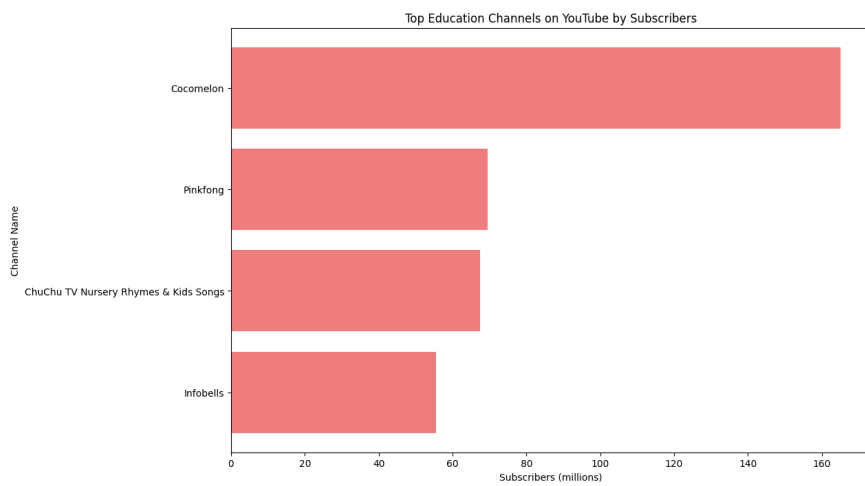
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file
file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Extract relevant data for education channels
education_channels = data[data['Category'] == 'Education']

# Sort channels by subscriber count
education_channels_sorted = education_channels.sort_values(by='Subscribers (millions)', ascending=False)

# Plotting the data
plt.figure(figsize=(12, 8))
plt.barh(education_channels_sorted['Name'], education_channels_sorted['Subscribers (millions)'], color='lightcoral')
plt.xlabel('Subscribers (millions)')
plt.ylabel('Channel Name')
plt.title('Top Education Channels on YouTube by Subscribers')
plt.gca().invert_yaxis() # Highest values at the top
plt.show()
```



▼ Sports

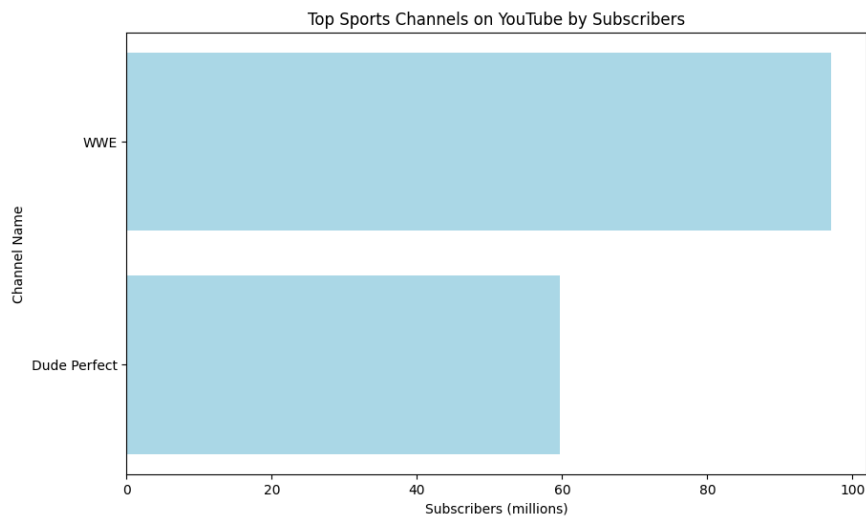
```
import pandas as pd
import matplotlib.pyplot as plt

file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Extract relevant data for education channels
sports_channels = data[data['Category'] == 'Sports']

# Sort channels by subscriber count
sports_channels_sorted = sports_channels.sort_values(by='Subscribers (millions)', ascending=False)

# Plotting the data
plt.figure(figsize=(10, 6))
plt.barh(sports_channels_sorted['Name'], sports_channels_sorted['Subscribers (millions)'], color='lightblue')
plt.xlabel('Subscribers (millions)')
plt.ylabel('Channel Name')
plt.title('Top Sports Channels on YouTube by Subscribers')
plt.gca().invert_yaxis() # Highest values at the top
plt.show()
```



▼ Film

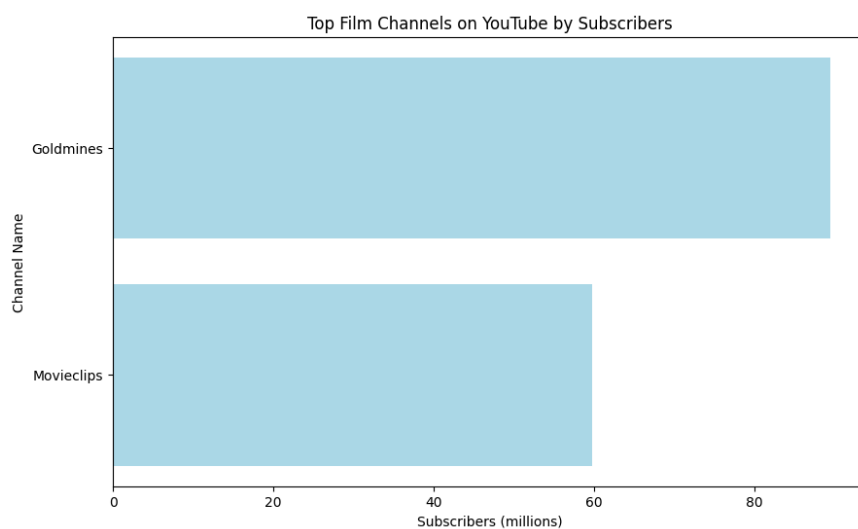
```
import pandas as pd
import matplotlib.pyplot as plt

file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Extract relevant data for education channels
film_channels = data[data['Category'] == 'Film']

# Sort channels by subscriber count
film_channels_sorted = film_channels.sort_values(by='Subscribers (millions)', ascending=False)

# Plotting the data
plt.figure(figsize=(10, 6))
plt.barh(film_channels_sorted['Name'], film_channels_sorted['Subscribers (millions)'], color='lightblue')
plt.xlabel('Subscribers (millions)')
plt.ylabel('Channel Name')
plt.title('Top Film Channels on YouTube by Subscribers')
plt.gca().invert_yaxis() # Highest values at the top
plt.show()
```



✓ How-to

```
import pandas as pd
import matplotlib.pyplot as plt

file_path = '/content/Most Subscribed YouTube Channels_exported.csv'
data = pd.read_csv(file_path)

# Extract relevant data for education channels
howto_channels = data[data['Category'] == 'How-to']

# Sort channels by subscriber count
howto_channels_sorted = howto_channels.sort_values(by='Subscribers (millions)', ascending=False)

# Plotting the data
plt.figure(figsize=(8, 6))
plt.barh(howto_channels['Name'], howto_channels['Subscribers (millions)'], color='orange')
plt.xlabel('Subscribers (millions)')
plt.ylabel('Channel Name')
plt.title('Top How-to Channels on YouTube by Subscribers')
plt.gca().invert_yaxis() # Highest values at the top
plt.show()
```

