# YouTube-Analytics

September 11, 2023

```python
[1]: import pandas as pd
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     import geopandas as gpd

     # Load the Dataset from Kaggle's directory structure with ISO-8859-1 encoding
     youtube_data_kaggle = pd.read_csv('/kaggle/input/global-youtube-statistics-2023/
      ↪Global YouTube Statistics.csv', encoding='ISO-8859-1')
```

```python
[2]: # Top 10 channels based on subscribers and video views
     top_10_subscribers = youtube_data_kaggle.nlargest(10, 'subscribers')
     top_10_video_views = youtube_data_kaggle.nlargest(10, 'video views')
     # Visualization for Top 10 Channels based on Subscribers
     plt.figure(figsize=(12, 8))
     sns.barplot(data=top_10_subscribers, y='Youtuber', x='subscribers',␣
      ↪palette='viridis')
     plt.title('Top 10 YouTube Channels by Subscribers')
     plt.xlabel('Number of Subscribers')
     plt.ylabel('YouTube Channel')
     plt.tight_layout()
     plt.show()

     # Visualization for Top 10 Channels based on Video Views
     plt.figure(figsize=(12, 8))
     sns.barplot(data=top_10_video_views, y='Youtuber', x='video views',␣
      ↪palette='viridis')
     plt.title('Top 10 YouTube Channels by Video Views')
     plt.xlabel('Number of Video Views')
     plt.ylabel('YouTube Channel')
     plt.tight_layout()
     plt.show()
```

Top 10 YouTube Channels by Subscribers

Top 10 YouTube Channels by Video Views
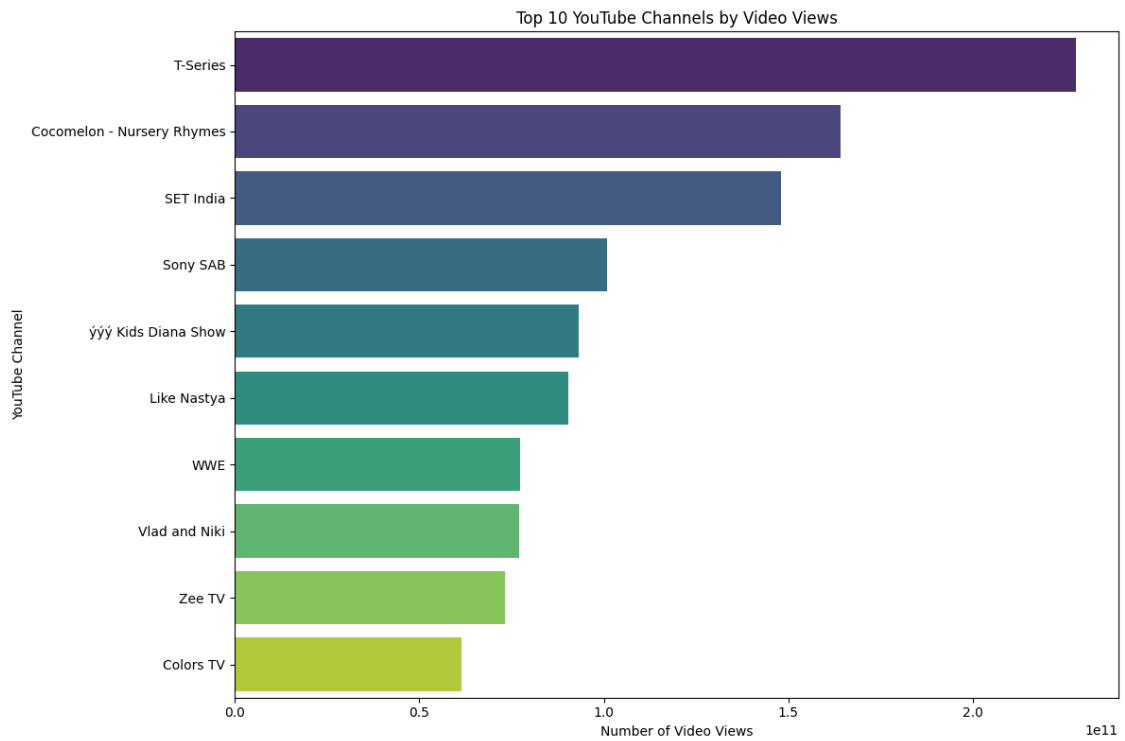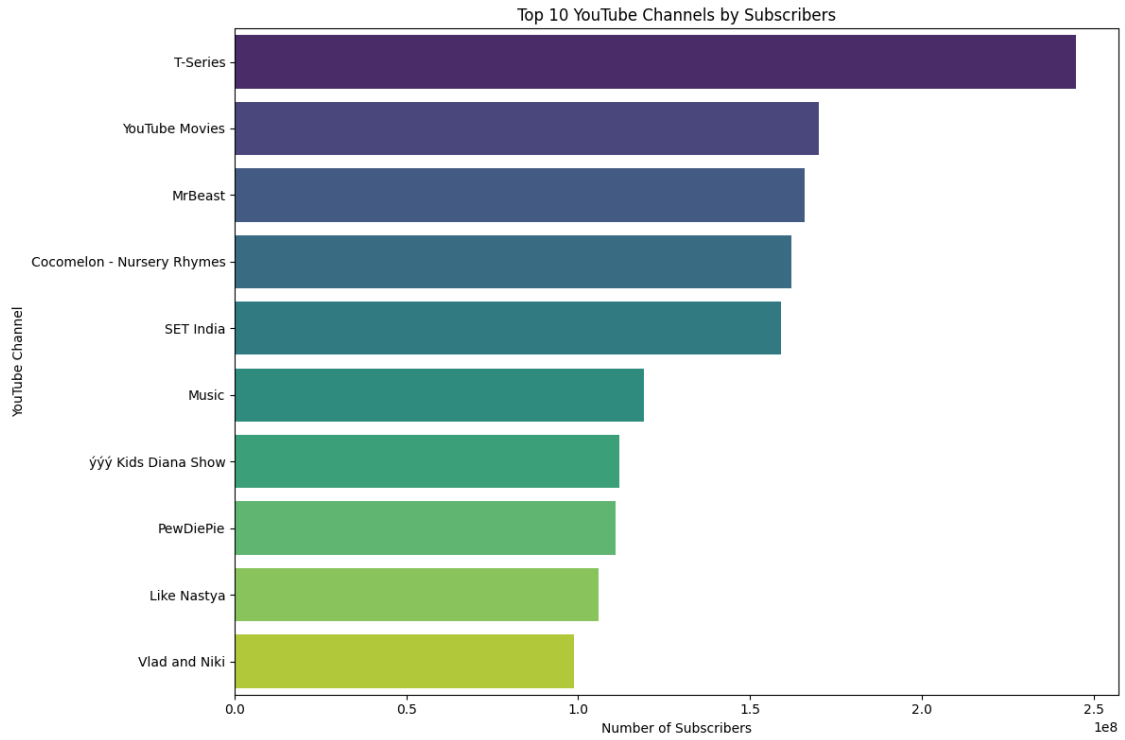
Subscribers:

T-Series from India leads the pack with 245 million subscribers. YouTube Movies and MrBeast follow with 170 million and 166 million subscribers respectively. Categories dominating the top 10 based on subscribers include Music, Film & Animation, Entertainment, and Education.

Video Views:

T-Series again tops the list with a whopping 228 billion video views. Cocomelon - Nursery Rhymes and SET India are next with 164 billion and 148 billion views respectively. Interestingly, YouTube Movies which ranks second in subscribers, doesn't have any video views recorded in this dataset. This might be because they are a specialized channel and the views might not be publicly disclosed.

From this initial analysis, we can infer that:

Indian channels, especially those in the music and entertainment categories, have a significant presence in the top ranks both in terms of subscribers and video views. US channels are also dominant, particularly in the entertainment and education categories.

```
[3]: # Most Popular Categories
popular_categories = youtube_data_kaggle['category'].value_counts().head(10)

# Plotting the Most Popular Categories
plt.figure(figsize=(12, 8))
popular_categories.plot(kind='barh', color='skyblue')
plt.title("Top 10 Most Popular Categories")
plt.xlabel("Number of Channels")
plt.ylabel("Category")
plt.gca().invert_yaxis()
plt.show()

# Distribution of Upload Frequencies
plt.figure(figsize=(12, 8))
sns.histplot(youtube_data_kaggle['uploads'], bins=50, color='salmon')
plt.title("Distribution of Upload Frequencies")
plt.xlabel("Number of Uploads")
plt.ylabel("Number of Channels")
plt.show()

# Average Uploads
average_uploads = youtube_data_kaggle['uploads'].mean()
print(f"Average number of uploads: {average_uploads:.2f}")
```

Top 10 Most Popular Categories

Distribution of Upload Frequencies
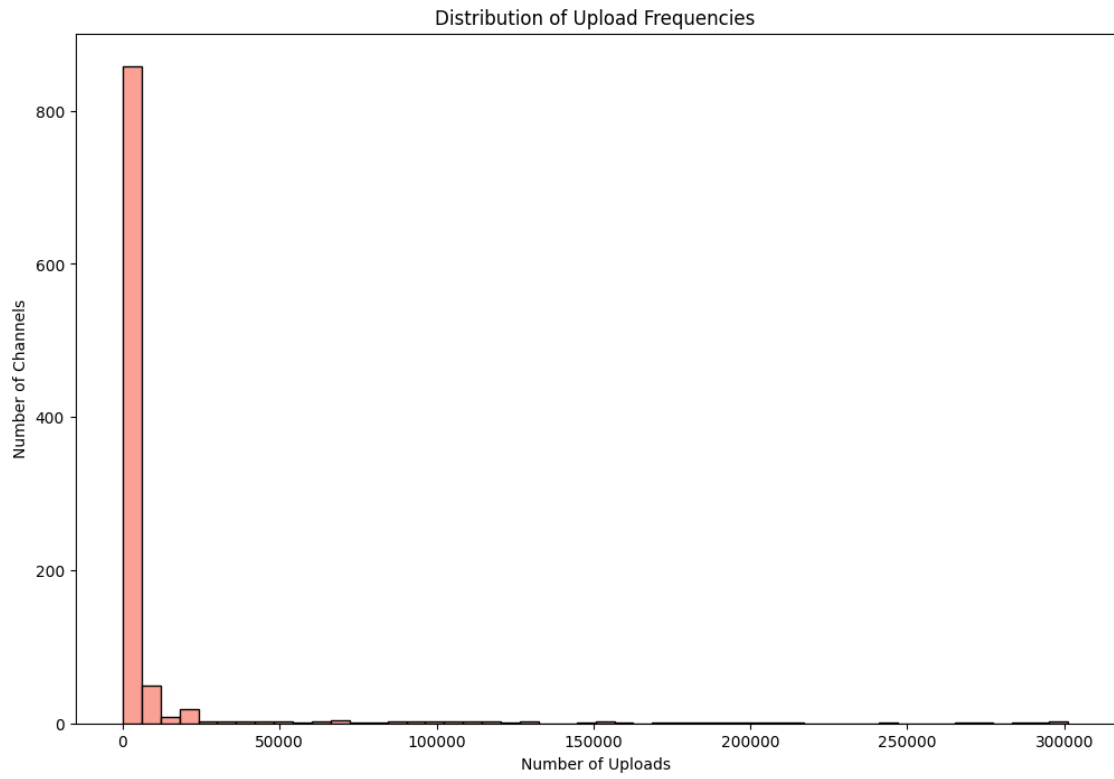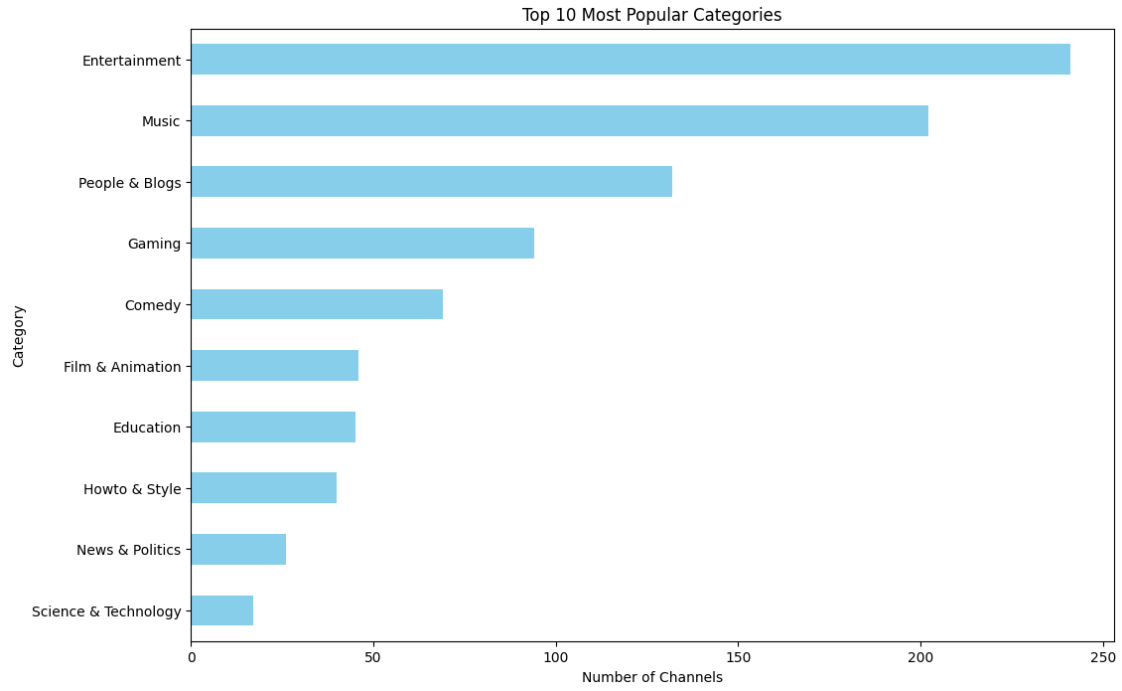
```
Average number of uploads: 9187.13
```

**Content Strategy:**

Most Popular Categories:

Entertainment is the most prevalent category with 241 channels. Music and People & Blogs follow next with 202 and 132 channels respectively. Other notable categories include Gaming, Comedy, and Film & Animation.

Upload Frequency:

On Average channels in this dataset have uploaded around 9,187 videos. From this data, we can infer that channels focused on entertainment, music, and vlogging (people & blogs) resonate most with audiences. Additionally, consistent and frequent content production seems to be a common trend among successful channels.

```python
[4]: # Top YouTube creators from different countries based on subscribers
     top_creators_by_country = youtube_data_kaggle.groupby('Country').apply(lambda x:
      ↪ x.nlargest(1, 'subscribers')).reset_index(drop=True)
     top_creators_by_country = top_creators_by_country.sort_values(by='subscribers',␣
      ↪ascending=False)

     # Visualization
     plt.figure(figsize=(15, 10))
     sns.barplot(data=top_creators_by_country.head(15), x='subscribers',␣
      ↪y='Country', palette='viridis')
     plt.title('Top 15 YouTube Creators by Country Based on Subscribers')
     plt.xlabel('Subscribers (in hundreds of millions)')
     plt.ylabel('Country')
     plt.tight_layout()
     plt.show()
```

Top 15 YouTube Creators by Country Based on Subscribers
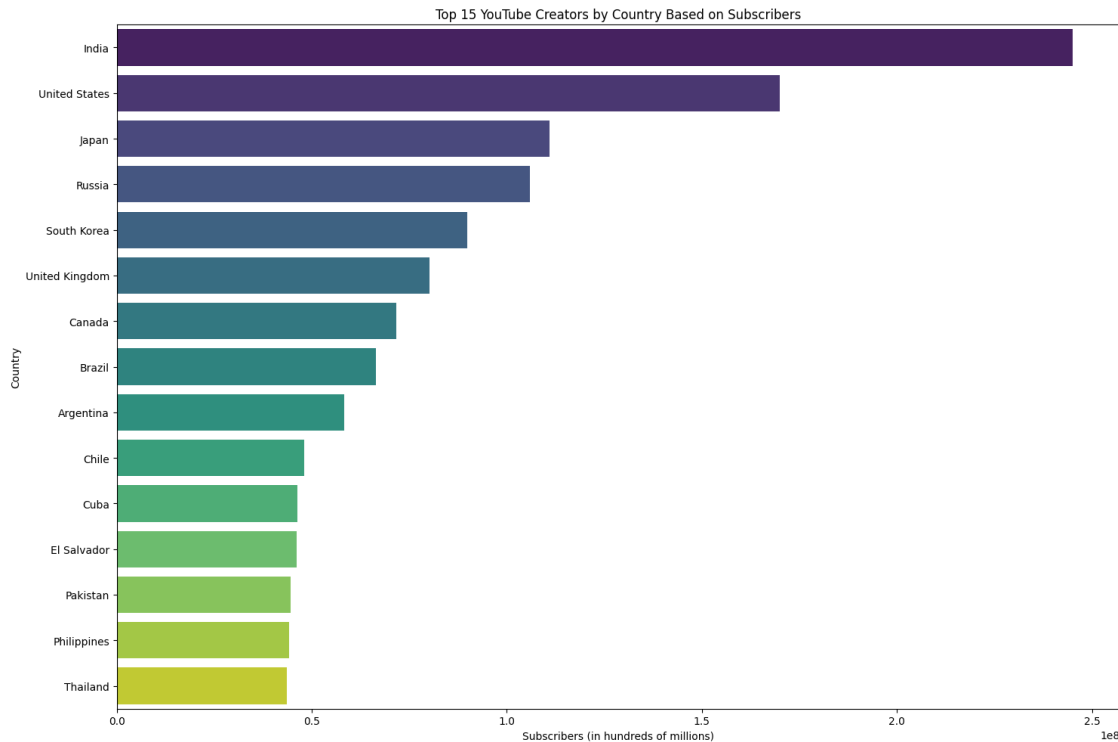
Regional Influencers:

T-Series from **India** is the top YouTube channel with 245 million subscribers. YouTube Movies from the **United States** follows next with 170 million subscribers. **PewDiePie** from Japan (though originally from Sweden and his content is primarily in English) is the top creator for Japan with 111 million subscribers. Like Nastya from **Russia**, BLACKPINK from **South Korea**, and 5-Minute Crafts from the **United Kingdom** are other noteworthy top creators from their respective countries.

```python
[5]:  # Correlation between Channel Performance and Earnings
      correlations = youtube_data_kaggle[['subscribers', 'video views',
       'lowest_monthly_earnings', 'highest_monthly_earnings',
       'lowest_yearly_earnings', 'highest_yearly_earnings']].corr()

      # Visualization of Correlations
      sns.set_style("whitegrid")
      plt.figure(figsize=(12, 8))
      sns.heatmap(correlations, annot=True, cmap='coolwarm', vmin=-1, vmax=1)
      plt.title("Correlation between Channel Performance and Earnings")
      plt.show()
```

Correlation between Channel Performance and Earnings
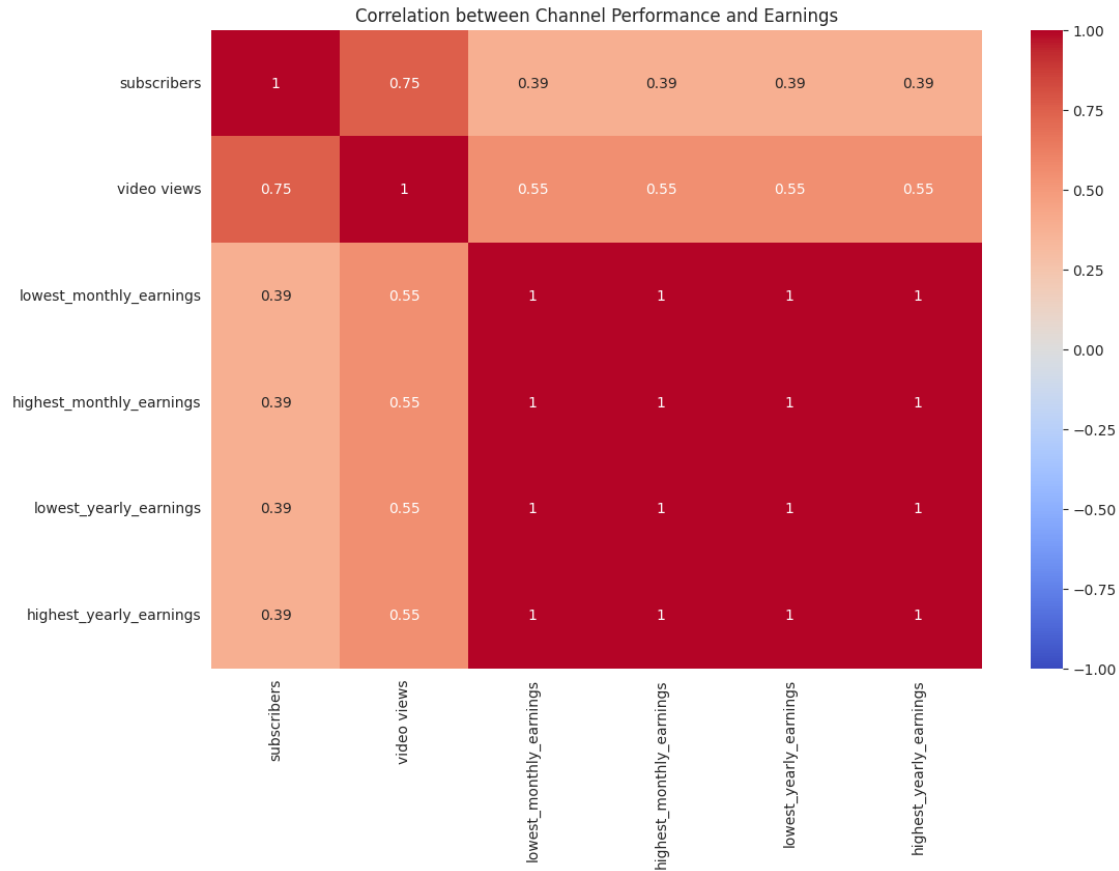
Earnings Analysis:

The heatmap displays the correlations between channel performance metrics (subscribers and video views) and estimated earnings (both monthly and yearly).

Observations:

Subscribers have a strong positive correlation with both monthly and yearly earnings. This indicates that channels with more subscribers generally tend to have higher earnings.

Video views also show a strong positive correlation with earnings. As the number of views increases, the estimated earnings for a channel also rise.

The correlation between subscribers and video views is high, implying that channels with more subscribers tend to get more views.

These correlations suggest that both subscribers and video views play a significant role in a channel's potential earnings. Channels that manage to grow their subscriber base and consistently achieve high video views are likely to have higher earnings.

```python
[6]:  # Geospatial Visualization
      world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))
      fig, ax = plt.subplots(figsize=(15, 10))
```

```
world.boundary.plot(ax=ax, linewidth=1, color='black')
ax.scatter(top_creators_by_country['Longitude'],␣
  ↪top_creators_by_country['Latitude'],␣
  ↪s=top_creators_by_country['subscribers'] / 1000000, color='blue', alpha=0.6)
ax.set_title("Distribution of Top YouTube Channels by Country (based on␣
  ↪Subscribers)")
plt.show()
```

/tmp/ipykernel_21/1551938021.py:2: FutureWarning: The geopandas.dataset module
is deprecated and will be removed in GeoPandas 1.0. You can get the original
'naturalearth_lowres' data from
https://www.naturalearthdata.com/downloads/110m-cultural-vectors/.
  world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))

Distribution of Top YouTube Channels by Country (based on Subscribers)
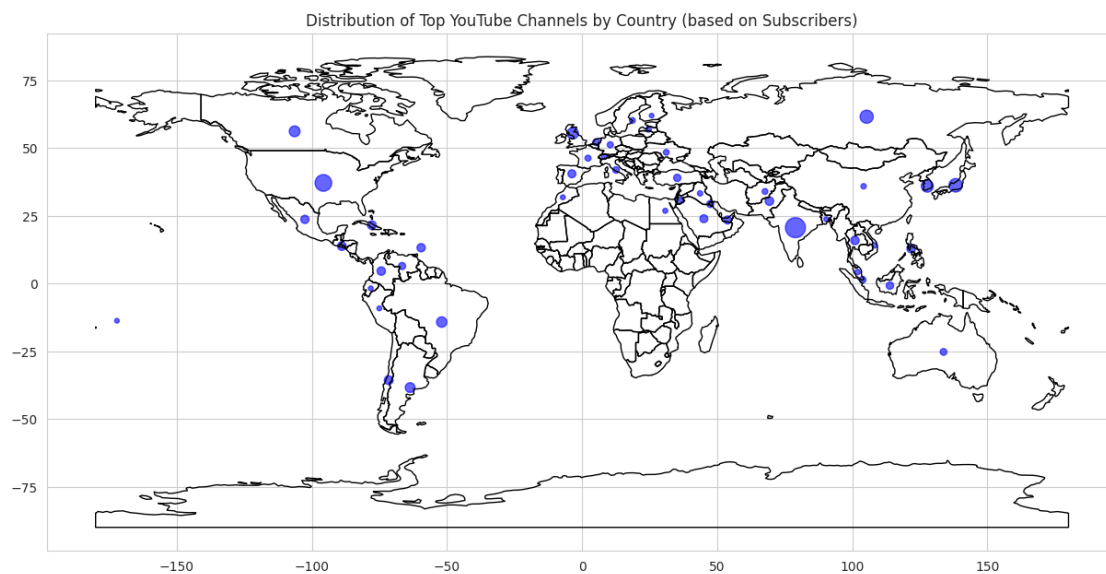
Geospatial visualization displaying the distribution of top YouTube channels based on subscribers:

The size of the blue dots represents the number of subscribers for the top YouTube channel in each country. As evident, countries like India and the United States have large dots, indicating a high number of subscribers for their top channels.

```
[7]: # Evolution of YouTube Channel Categories Over Time
category_evolution = youtube_data_kaggle.groupby(['created_year', 'category']).
  ↪size().unstack().fillna(0)

# Selecting the top 5 categories based on the latest year's data
top_categories = category_evolution.iloc[-1].nlargest(5).index

# Plotting the Evolution using a Line Chart for the top 5 categories
```

```
plt.figure(figsize=(15, 10))
for category in top_categories:
    plt.plot(category_evolution.index, category_evolution[category],␣
 ↪label=category, marker='o')
plt.title("Evolution of Top 5 YouTube Channel Categories Over Time")
plt.ylabel("Number of Channels Created")
plt.xlabel("Year")
plt.legend(title='Categories')
plt.grid(True, which='both', linestyle='--', linewidth=0.5)
plt.tight_layout()
plt.show()
```
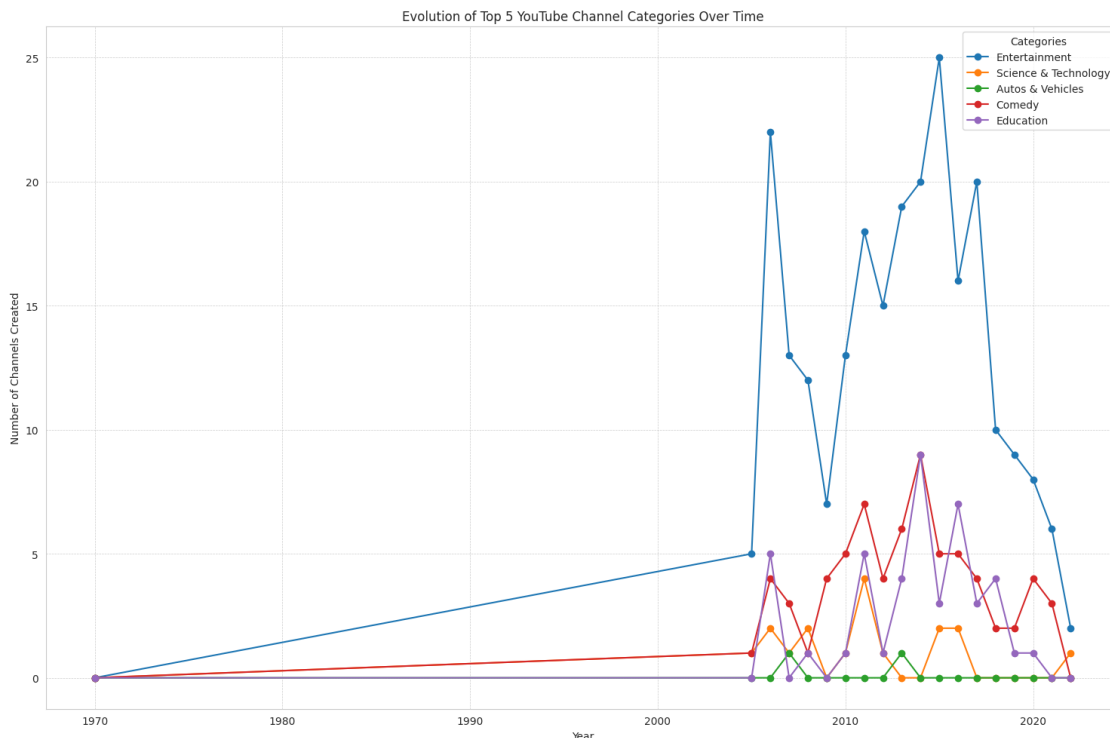

Evolution of Top 5 YouTube Channel Categories Over Time

The line chart provides insights into the trajectory of the top 5 YouTube channel categories, based on their year of creation:

Initial Years (around 2005-2007):

During the platform's early stages, there's a visible exploration across categories. Some categories might show a modest beginning, while others might have a more pronounced presence. Dominance of Key Categories:

As the years progress, certain categories, like Music and Entertainment, exhibit a consistent growth trend, highlighting their enduring popularity. Consistent Players:

The Music and Entertainment categories, among the top 5, showcase consistent growth, reaffirming their dominant position on the platform over the years. Emerging Trends:

Categories like Gaming, if showing a steep upward trajectory in recent years, signify the growing influence and popularity of gaming content on YouTube. Dynamics of Preference:

The ebbs and flows in the line chart, especially for the top 5 categories, give us a nuanced understanding of audience preferences and how they've shifted over the years