**GLOBAL YOUTUBE STATISTICS DATASET**

* **Sharon Blessy J V**

1. **What are the top 10 YouTube channels based on the number of subscribers?**

**Ans.**

Code Snippet Used:

subs\_sorted = yt.sort\_values(by='subscribers', ascending=False)

top\_10 = subs\_sorted.head(10)

print(top\_10[['rank','Youtuber','subscribers','Title']])

Output:

****

1. **Which category has the highest average number of subscribers?**

**Ans.**

Shows have the highest average number of subscribers. Below given is the code snippet used to find the same.

yt\_1 = yt[yt['subscribers'] != 0]

avgsubs\_by\_category = yt\_1.groupby('category')['subscribers'].mean().sort\_values(ascending=False)

print(avgsubs\_by\_category)

print("Category with the highest average number of subscribers : " + avgsubs\_by\_category.index[0])

1. **How many videos, on average, are uploaded by YouTube channels in each category?**

**Ans.**

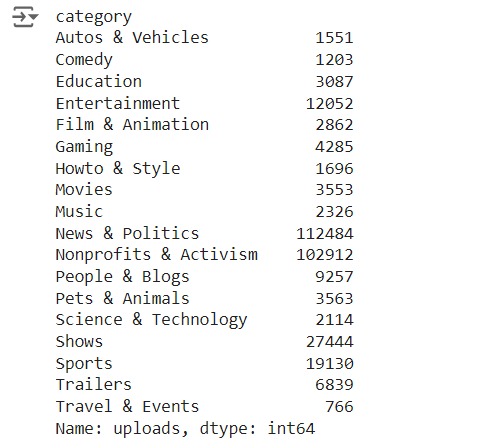
**Code Snippet Used:**

yt\_1 = yt[yt['subscribers'] != 0]

avgvid\_by\_category = yt\_1.groupby('category')['uploads'].mean().round().astype(int)

print(avgvid\_by\_category)

**Output:**

****

1. **What are the top 5 countries with the highest number of YouTube channels?**

**Ans.**

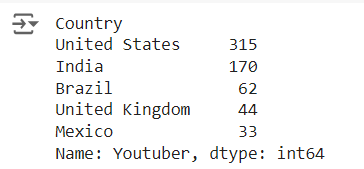
Top 5 countries include the United States, India, Brazil, United Kingdom and Mexico.

**Code Snippet:**

**top\_5 = yt.groupby('Country')['Youtuber'].count().sort\_values(ascending=False).head(5)**

**print(top\_5)**

**Output:**

****

1. **What is the distribution of channel types across different categories?**

**Ans.**

**Code Snippet:**

**distribution = yt.groupby(['category', 'channel\_type']).size().unstack(fill\_value=0)**

**distribution.plot(kind='bar', stacked=True)**

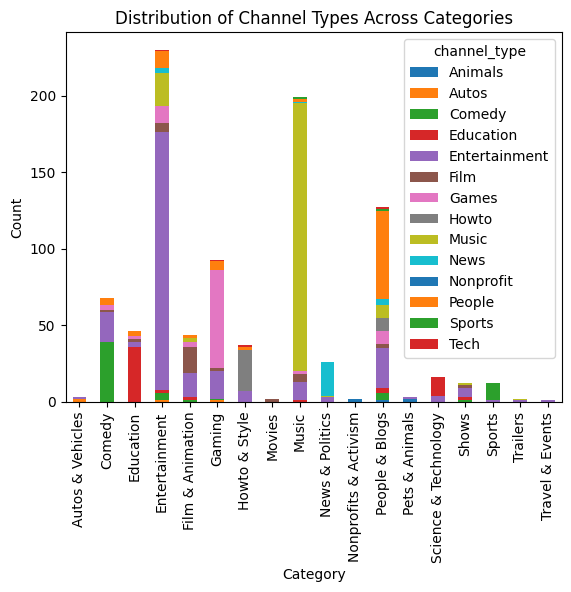
**plt.title('Distribution of Channel Types Across Categories')**

**plt.xlabel('Category')**

**plt.ylabel('Count')**

**plt.show()**

**Output:**

****

1. **Is there a correlation between the number of subscribers and total video views for YouTube channels?**

**Ans.**

**Code Snippet:**

**sub\_view\_corr = yt['subscribers'].corr(yt['video views'])**

**print(sub\_view\_corr)**

**Output:**

****

It indicates a strong positive relationship between no. of subscribers and total video views for the Youtube Channels.

1. **How do the monthly earnings vary throughout different categories?**

**Ans.**

**Code Snippet:**

**monthly\_earnings\_dist = yt.groupby('category')['highest\_monthly\_earnings'].mean().sort\_values(ascending=False)**

**sns.barplot(x=monthly\_earnings\_dist.index, y=monthly\_earnings\_dist.values, palette='viridis')**

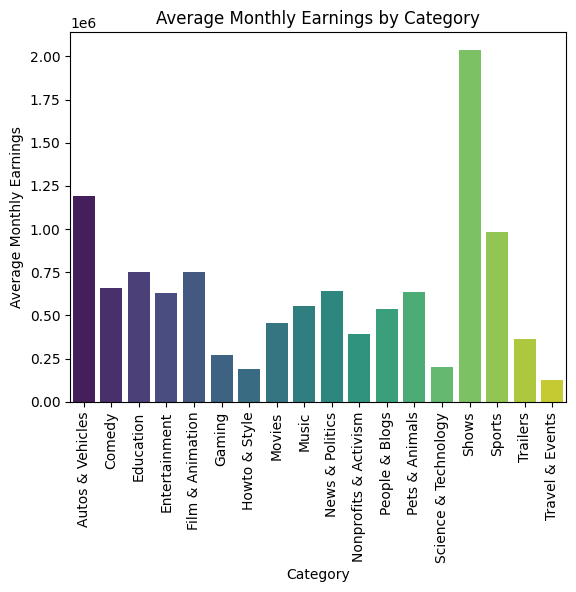
**plt.title('Average Monthly Earnings by Category')**

**plt.xlabel('Category')**

**plt.ylabel('Average Monthly Earnings')**

**plt.xticks(rotation=90)**

**plt.show()**

****

1. **What is the overall trend in subscribers gained in the last 30 days across all channels?**

**Ans.**

**Code Snippet Used:**

**mean\_sub\_gain = yt['subscribers\_for\_last\_30\_days'].mean()**

**highest = yt['subscribers\_for\_last\_30\_days'].max()**

**ch\_highest = yt[yt['subscribers\_for\_last\_30\_days'] == highest]['Title'].values**

**lowest = yt['subscribers\_for\_last\_30\_days'].min()**

**ch\_lowest = yt[yt['subscribers\_for\_last\_30\_days'] == lowest]['Title'].values**

**print(f"Average Subscribers Gained: {mean\_sub\_gain}")**

**print(f"Highest Subscribers Gained: {highest}")**

**print("Channels with highest subscribers Gained: ",\*ch\_highest)**

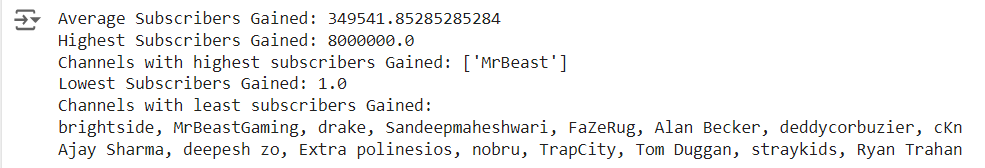
**print(f"Lowest Subscribers Gained: {lowest}")**

**print("Channels with least subscribers Gained:")**

**print(\*ch\_lowest[0:8], sep = ", ")**

**print(\*ch\_lowest[8:len(ch\_lowest)], sep = ", ")**

**Output:**

****

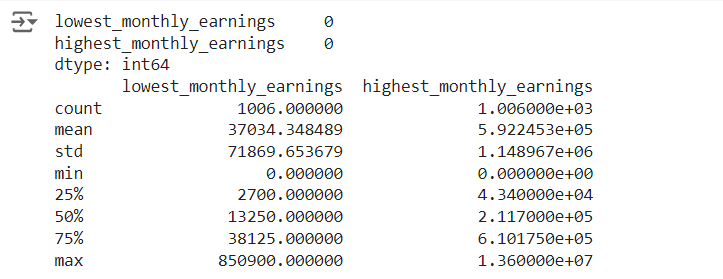
1. **Are there any outliers in terms of yearly earnings from YouTube channels?**

**Ans.**

**Code Snippet:**

**print(yt.isnull().sum()[['lowest\_monthly\_earnings','highest\_monthly\_earnings']])**

**print(yt.describe()[['lowest\_monthly\_earnings','highest\_monthly\_earnings']])**

**Output:**

Some channels have monthly earnings of 0 .

1. **What is the distribution of channel creation dates? Is there any trend over time?**

**Ans.**

**Code Snippet:**

**yt['created\_year'] = pd.to\_datetime(yt['created\_year'], format='%Y')**

**time\_trend = yt.groupby('created\_year')['Youtuber'].count()**

**plt.figure(figsize=(10, 6))**

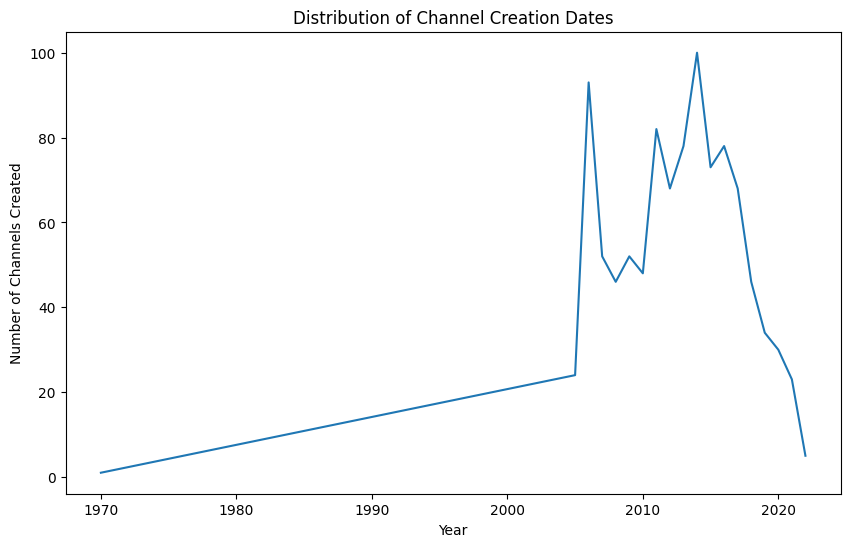
**plt.plot(time\_trend.index, time\_trend.values)**

**plt.title('Distribution of Channel Creation Dates')**

**plt.xlabel('Year')**

**plt.ylabel('Number of Channels Created')**

**plt.show()**

****

1. **Is there a relationship between gross tertiary education enrollment and the number of YouTube channels in a country?**

**Ans.**

**Code Snippet:**

**countries = yt['Country'].unique()**

**countries = [country for country in countries if pd.notna(country)]**

**countries.sort()**

**channel\_count\_by\_country = yt.groupby('Country')['Youtuber'].count()**

**gross\_edu\_by\_country = yt.groupby('Country')['Gross tertiary education enrollment (%)'].mean()**

**plt.bar(countries, gross\_edu\_by\_country)**

**plt.title('Gross Tertiary Education Enrollment by Country')**

**plt.xlabel('Country')**

**plt.ylabel('Gross Tertiary Education Enrollment (%)')**

**plt.xticks(rotation=90)**

**plt.show()**

**plt.bar(countries, channel\_count\_by\_country)**

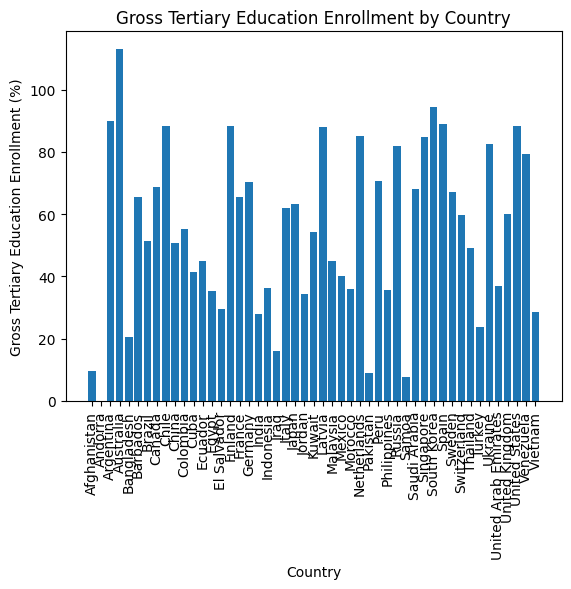
**plt.title('Number of YouTube Channels by Country')**

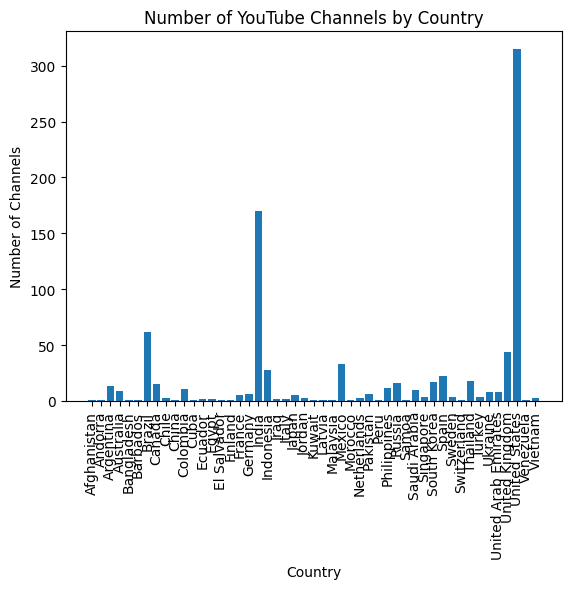
**plt.xlabel('Country')**

**plt.ylabel('Number of Channels')**

**plt.xticks(rotation=90)**

**plt.show()**

****

****

There is no visible relationship that can be observed.

1. **How does the unemployment rate vary among the top 10 countries with the highest number of YouTube channels?**

**Ans.**

**Code Snippet:**

**top\_10\_countries = yt.groupby('Country')['Youtuber'].count().sort\_values(ascending=False).head(10).index**

**top\_10\_data = yt[yt['Country'].isin(top\_10\_countries)]**

**unemployment\_rate\_by\_country = top\_10\_data.groupby('Country')['Unemployment rate'].mean()**

**plt.bar(unemployment\_rate\_by\_country.index, unemployment\_rate\_by\_country.values)**

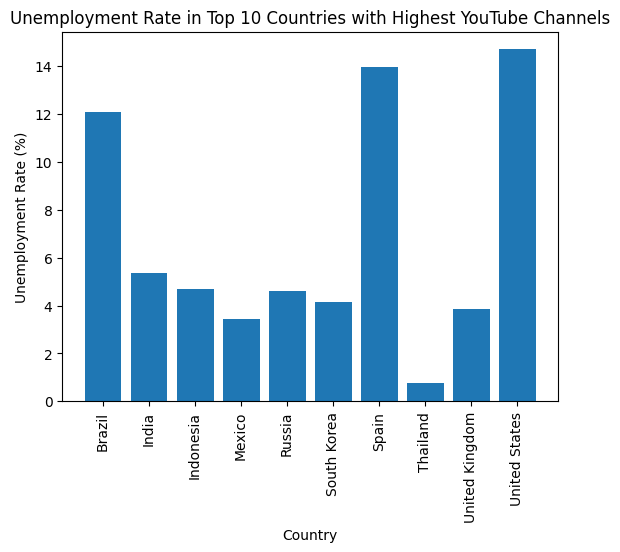
**plt.title('Unemployment Rate in Top 10 Countries with Highest YouTube Channels')**

**plt.xlabel('Country')**

**plt.ylabel('Unemployment Rate (%)')**

**plt.xticks(rotation=45)**

**plt.show()**

****

1. **What is the average urban population percentage in countries with YouTube channels?**

**Ans.**

**Code Snippet:**

**avg\_urban\_population\_by\_country = yt.groupby('Country')['Urban\_population'].mean().dropna()**

**print(avg\_urban\_population\_by\_country)**

****

1. **Are there any patterns in the distribution of YouTube channels based on latitude and longitude coordinates?**

**Ans.**

**Code Snippet:**

**from mpl\_toolkits.basemap import Basemap**

**latitudes = yt['Latitude'].dropna()**

**longitudes = yt['Longitude'].dropna()**

**plt.figure(figsize=(10, 7))**

**m = Basemap(projection='merc', llcrnrlat=-60, urcrnrlat=80, llcrnrlon=-180, urcrnrlon=180, resolution='c')**

**m.drawcoastlines()**

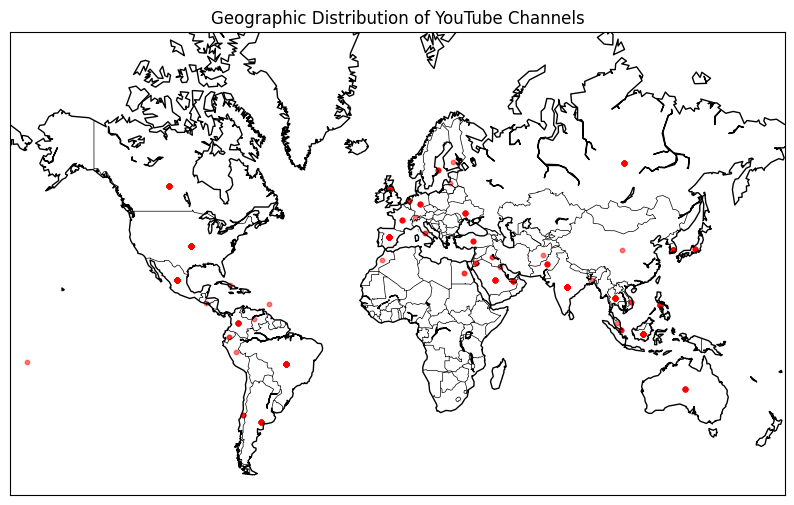
**m.drawcountries()**

**x, y = m(longitudes.values, latitudes.values)**

**m.scatter(x, y, marker='o', color='red', s=10, alpha=0.5)**

**plt.title('Geographic Distribution of YouTube Channels')**

**plt.show()**

****

1. **What is the correlation between the number of subscribers and the population of a country?**

**Ans.**

**Code Snippet:**

**correlation = yt['subscribers'].corr(yt['Population'])**

**print("Correlation between no. of subscribers and the population : ",correlation)**

****

1. **How do the top 10 countries with the highest number of YouTube channels compare in terms of their total population?**

**Ans.**

**Code Snippet:**

**top\_10\_countries = yt.groupby('Country')['Youtuber'].count().sort\_values(ascending=False).head(10).index**

**top\_10\_data = yt[yt['Country'].isin(top\_10\_countries)]**

**population\_by\_country = top\_10\_data.groupby('Country')['Population'].mean().sort\_values(ascending=False)**

**plt.figure(figsize=(10, 6))**

**plt.bar(population\_by\_country.index, population\_by\_country.values)**

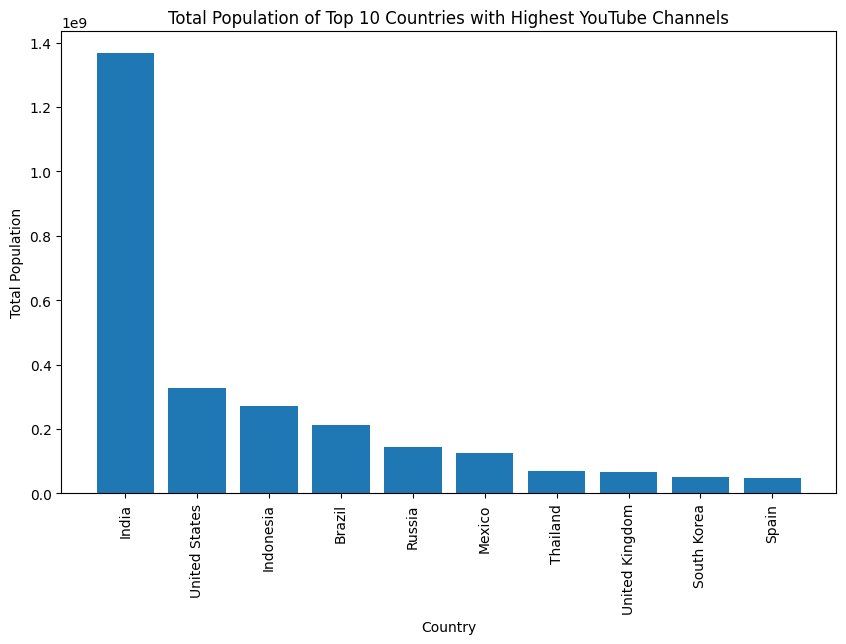
**plt.title('Total Population of Top 10 Countries with Highest YouTube Channels')**

**plt.xlabel('Country')**

**plt.ylabel('Total Population')**

**plt.xticks(rotation=90)**

**plt.show()**

****

1. **Is there a correlation between the number of subscribers gained in the last 30 days and the unemployment rate in a country?**

**Ans.**

**Code Snippet:**

**correlation = yt['subscribers\_for\_last\_30\_days'].corr(yt['Unemployment rate'])**

**print("Correlation between subscribers gained in the last 30 days and unemployment rate:", correlation)**

**Output:**

This indicates a very weak negative relationship.

1. **How does the distribution of video views for the last 30 days vary across different channel types?**

**Ans.**

**Code Snippet:**

**plt.figure(figsize=(10, 6))**

**sns.barplot(x='channel\_type', y='video\_views\_for\_the\_last\_30\_days', data=yt)**

**plt.xticks(rotation=90)**

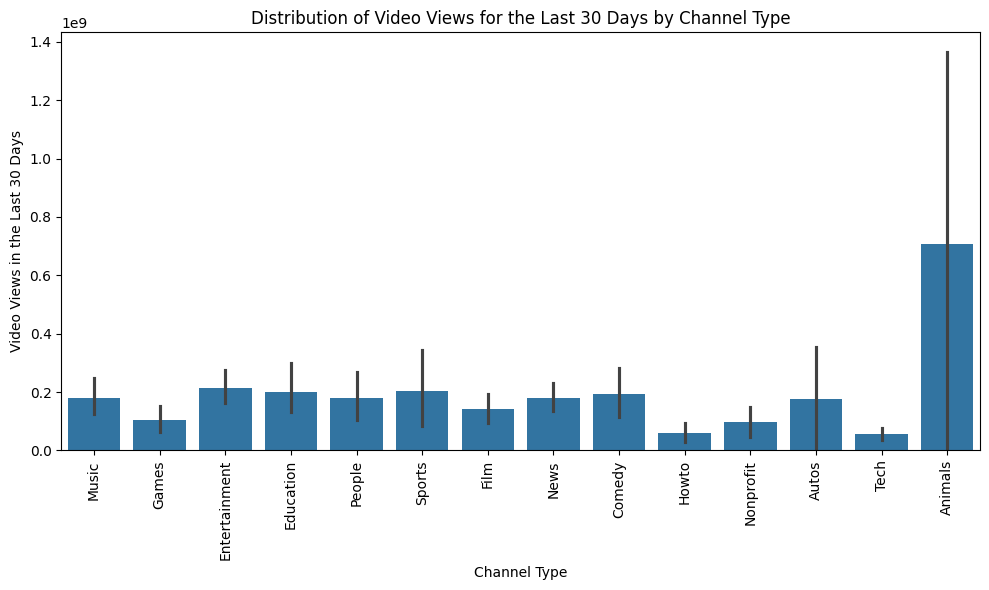
**plt.title('Distribution of Video Views for the Last 30 Days by Channel Type')**

**plt.xlabel('Channel Type')**

**plt.ylabel('Video Views in the Last 30 Days')**

**plt.tight\_layout()**

**plt.show()**

****

1. **Are there any seasonal trends in the number of videos uploaded by YouTube channels?**

**Ans.**

**Code Snippet:**

**upload\_trend = yt.groupby(yt['created\_month'])['uploads'].sum()**

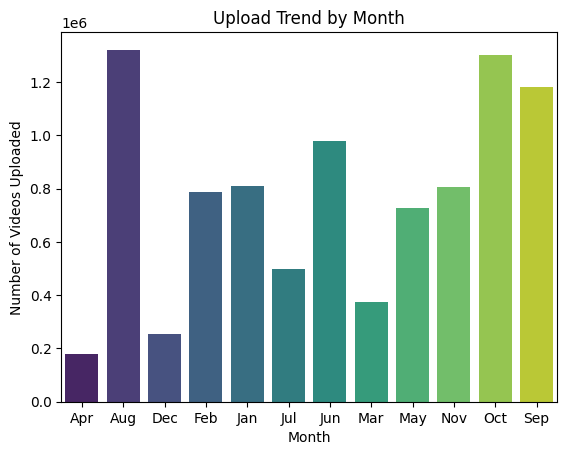
**sns.barplot(x=upload\_trend.index, y=upload\_trend.values, palette='viridis')**

**plt.title('Upload Trend by Month')**

**plt.xlabel('Month')**

**plt.ylabel('Number of Videos Uploaded')**

**plt.show()**

****

1. **What is the average number of subscribers gained per month since the creation of YouTube channels till now?**

**Ans.**

**Code Snippet:**

**avg\_subscribers\_gained\_per\_month = yt.groupby(yt['created\_month'])['subscribers\_for\_last\_30\_days'].mean()**

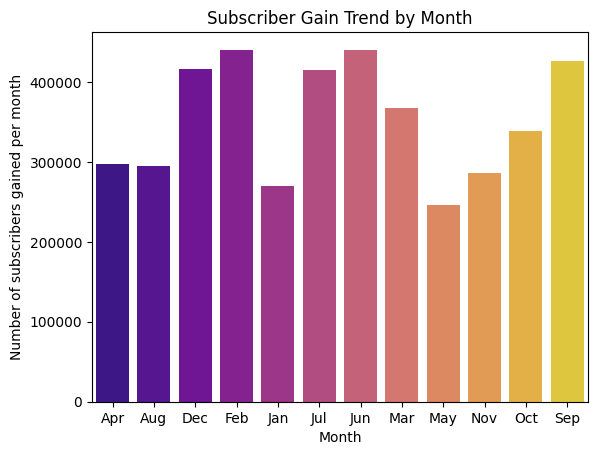
**sns.barplot(x=avg\_subscribers\_gained\_per\_month.index, y=avg\_subscribers\_gained\_per\_month.values, palette='plasma')**

**plt.title('Upload Trend by Month')**

**plt.xlabel('Month')**

**plt.ylabel('Number of subscribers gained per month')**

**plt.show()**

****