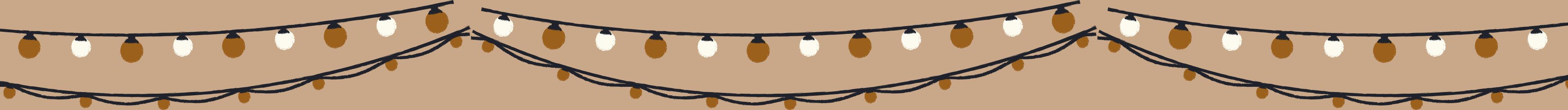




จำนวนการยับตัวแต่ละประเภท ที่ผลิตออกมานามาตามช่วงเวลา





IMPORT DATA

```
[ ] import pandas as pd
      import os

[ ] from google.colab import drive
drive.mount('/content/drive')

→ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

[ ] path = '/content/drive/MyDrive/Data_viz'

[ ] movie_india = pd.read_csv(os.path.join(path,'IMDb Movies India.csv'), encoding='latin-1')
movie_india.head()
```

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2	Actor 3
0		NaN	NaN	Drama	NaN	NaN	J.S. Randhawa	Manmauji	Birbal	Rajendra Bhatia
1	#Gadhvi (He thought he was Gandhi)	(2019)	109 min	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghamande	Arvind Jangid
2	#Homecoming	(2021)	90 min	Drama, Musical	NaN	NaN	Soumyajit Majumdar	Sayani Gupta	Plabita Borthakur	Roy Angana
3	#Yaaram	(2019)	110 min	Comedy, Romance	4.4	35	Ovais Khan	Prateik	Ishita Raj	Siddhant Kapoor
4	...And Once Again	(2010)	105 min	Drama	NaN	NaN	Amol Palekar	Rajat Kapoor	Rituparna Sengupta	Antara Mali

DATA CLEANSING

```
[ ] movie_india['Year'] = movie_india['Year'].str.replace('(', '').str.replace(')', '')
movie_india['Duration'] = movie_india['Duration'].str.replace('min', '')
movie_india.head()
```

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2	Actor 3
0		NaN	NaN	Drama	NaN	NaN	J.S. Randhawa	Manmauji	Birbal	Rajendra Bhatia
1	#Gadhvi (He thought he was Gandhi)	2019	109	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghambade	Arvind Jangid
2	#Homecoming	2021	90	Drama, Musical	NaN	NaN	Soumyajit Majumdar	Sayani Gupta	Plabita Borthakur	Roy Angana
3	#Yaaram	2019	110	Comedy, Romance	4.4	35	Ovais Khan	Prateik	Ishita Raj	Siddhant Kapoor
4	...And Once Again	2010	105	Drama	NaN	NaN	Amol Palekar	Rajat Kapoor	Rituparna Sengupta	Antara Mali

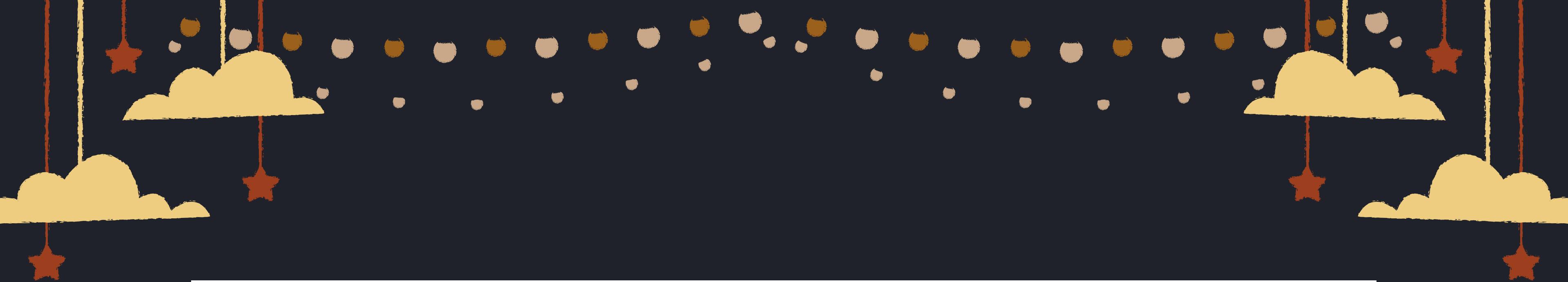
```
[ ] def split_genres(row):
    # Check if the genre value is a string before splitting
    if isinstance(row['Genre'], str):
        genres = row['Genre'].split(',')
        return [genre.strip() for genre in genres]
    else:
        # Handle cases where the genre is not a string (e.g., return an empty list or NaN)
        return []

movie_india['Genre'] = movie_india.apply(split_genres, axis=1)
```

```
[ ] import pandas as pd
from matplotlib import pyplot as plt
from google.colab import drive
import os

# Function to split genres and create new rows
def split_genres(row):
    if isinstance(row['Genre'], list):
        new_rows = []
        for i in range(len(row['Genre'])):
            new_row = row.copy()
            new_row['Genre'] = row['Genre'][i]
            new_rows.append(new_row)
        return new_rows
    else:
        return [row]

# Apply the function to split genres and create new rows
new_md = pd.concat([pd.DataFrame(split_genres(row)) for _, row in movie_india.iterrows()], ignore_index=True)
```



```
new_md.head()
```

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2	Actor 3
0		NaN	NaN	Drama	NaN	NaN	J.S. Randhawa	Manmauji	Birbal	Rajendra Bhatia
1	#Gadhvi (He thought he was Gandhi)	2019	109	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghamande	Arvind Jangid
2	#Homecoming	2021	90	Drama	NaN	NaN	Soumyajit Majumdar	Sayani Gupta	Plabita Borthakur	Roy Angana
3	#Homecoming	2021	90	Musical	NaN	NaN	Soumyajit Majumdar	Sayani Gupta	Plabita Borthakur	Roy Angana
4	#Yaaram	2019	110	Comedy	4.4	35	Ovais Khan	Prateik	Ishita Raj	Siddhant Kapoor

```
new_md = new_md.dropna(subset=['Rating', 'Year', 'Duration'])
```

```
new_md['Genre'].value_counts()
```

	count
Genre	
Drama	3847
Action	1699
Romance	1383
Comedy	1352
Crime	878
Thriller	684
Family	428
Musical	416
Mystery	308
Adventure	284

Mystery	308
Adventure	284
Horror	205
Fantasy	148
Documentary	130
Biography	126
History	103
Animation	56
Music	55

Sport	44
Sci-Fi	35
War	34
News	2
Western	2



```
# Filter the DataFrame to include only the specified genres
selected_genres = ['Drama', 'Action', 'Romance', 'Comedy', 'Crime']
filtered_data = new_md[new_md['Genre'].isin(selected_genres)]
```

filtered_data

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2	Actor 3
1	#Gadhvi (He thought he was Gandhi)	2019	109	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghamande	Arvind Jangid
4	#Yaaram	2019	110	Comedy	4.4	35	Ovais Khan	Prateik	Ishita Raj	Siddhant Kapoor
5	#Yaaram	2019	110	Romance	4.4	35	Ovais Khan	Prateik	Ishita Raj	Siddhant Kapoor
7	...Aur Pyaar Ho Gaya	1997	147	Comedy	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan	Shammi Kapoor
8	...Aur Pyaar Ho Gaya	1997	147	Drama	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan	Shammi Kapoor
...
23065	Zulm Ki Zanjeer	1989	125	Drama	5.8	44	S.P. Muthuraman	Chiranjeevi	Jayamalini	Rajinikanth
23067	Zulmi	1999	129	Action	4.5	655	Kuku Kohli	Akshay Kumar	Twinkle Khanna	Aruna Irani
23068	Zulmi	1999	129	Drama	4.5	655	Kuku Kohli	Akshay Kumar	Twinkle Khanna	Aruna Irani
23071	Zulm-O-Sitam	1998	130	Action	6.2	20	K.C. Bokadia	Dharmendra	Jaya Prada	Arjun Sarja
23072	Zulm-O-Sitam	1998	130	Drama	6.2	20	K.C. Bokadia	Dharmendra	Jaya Prada	Arjun Sarja

9159 rows × 10 columns

```

# Define a function to categorize years
def categorize_year(year):
    if 1917 <= year <= 1951:
        return '1917 - 1951'
    elif 1952 <= year <= 1986:
        return '1952 - 1986'
    else:
        return '1987 - 2021'

# Apply the function to create a new column 'Year Category'
filtered_data['Year Category'] = pd.to_numeric(filtered_data['Year']).apply(categorize_year)

```

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2	Actor 3	Year Category
1	#Gadhvi (He thought he was Gandhi)	2019	109	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghamande	Arvind Jangid	1987 - 2021
4	#Yaaram	2019	110	Comedy	4.4	35	Ovais Khan	Prateik	Ishita Raj	Siddhant Kapoor	1987 - 2021
5	#Yaaram	2019	110	Romance	4.4	35	Ovais Khan	Prateik	Ishita Raj	Siddhant Kapoor	1987 - 2021
7	...Aur Pyaar Ho Gaya	1997	147	Comedy	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan	Shammi Kapoor	1987 - 2021
8	...Aur Pyaar Ho Gaya	1997	147	Drama	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan	Shammi Kapoor	1987 - 2021
...
23065	Zulm Ki Zanjeer	1989	125	Drama	5.8	44	S.P. Muthuraman	Chiranjeevi	Jayamalini	Rajinikanth	1987 - 2021
23067	Zulmi	1999	129	Action	4.5	655	Kuku Kohli	Akshay Kumar	Twinkle Khanna	Aruna Irani	1987 - 2021
23068	Zulmi	1999	129	Drama	4.5	655	Kuku Kohli	Akshay Kumar	Twinkle Khanna	Aruna Irani	1987 - 2021
23071	Zulm-O-Sitam	1998	130	Action	6.2	20	K.C. Bokadia	Dharmendra	Jaya Prada	Arjun Sarja	1987 - 2021
23072	Zulm-O-Sitam	1998	130	Drama	6.2	20	K.C. Bokadia	Dharmendra	Jaya Prada	Arjun Sarja	1987 - 2021

9159 rows × 11 columns

```
group_data = filtered_data.groupby(['Genre', 'Year Category'])['Genre'].count().unstack()  
group_data
```

Year Category 1917 - 1951 1952 - 1986 1987 - 2021

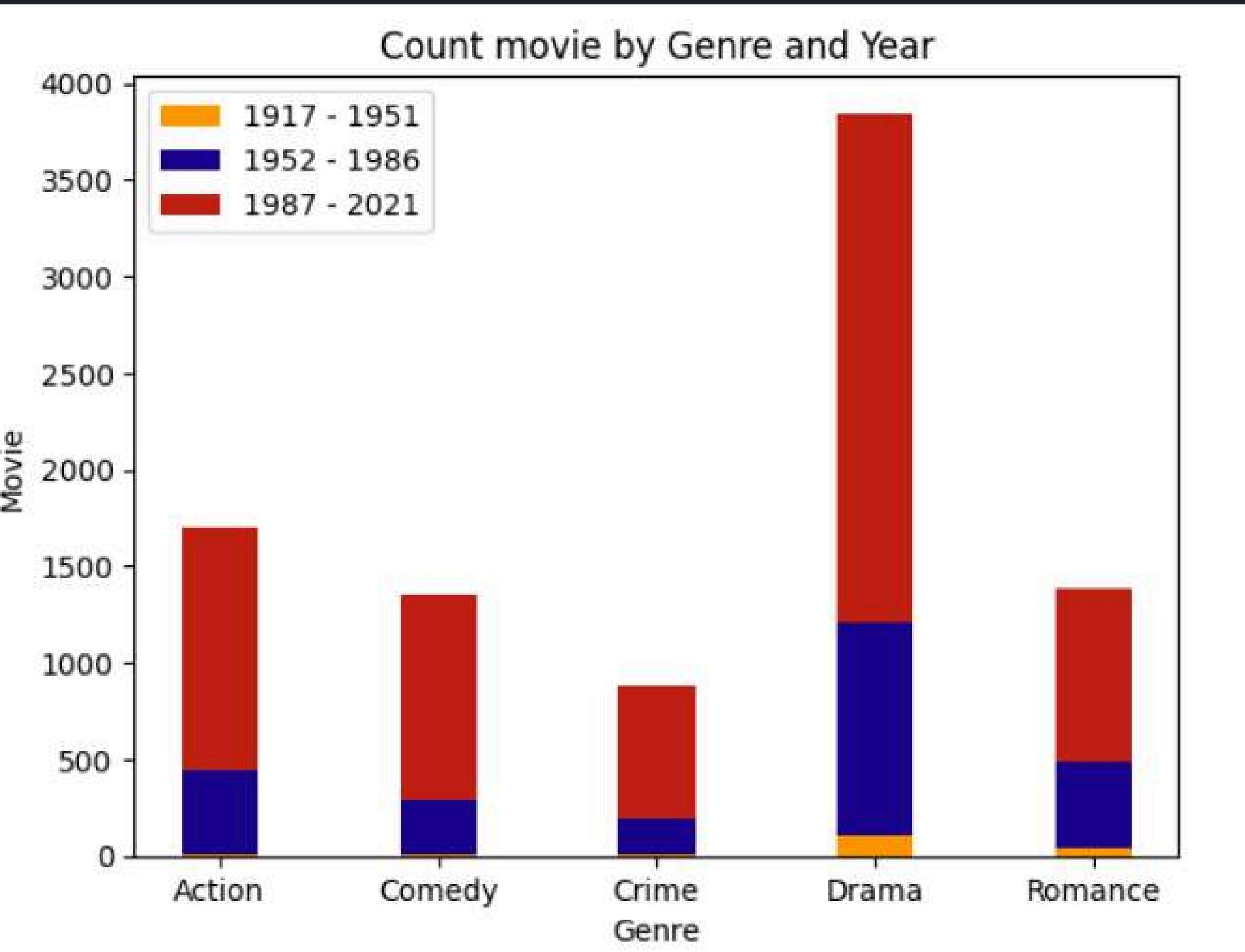
Genre

Action	5	438	1256
Comedy	7	290	1055
Crime	4	188	686
Drama	104	1102	2641
Romance	39	449	895

PLOT STACKED BAR CHART

```
1 import matplotlib.pyplot as plt
2
3 width = 0.35      # the width of the bars: can also be len(x) sequence
4
5 fig, ax = plt.subplots()
6
7 ax.bar(group_data.index, group_data['1917 - 1951'].values, width,
8        label='1917 - 1951',color = '#fc9700')
9 ax.bar(group_data.index, group_data['1952 - 1986'].values, width,
10       bottom=group_data['1917 - 1951'].values,
11       label = '1952 - 1986',color = '#19038a')
12 ax.bar(group_data.index, group_data['1987 - 2021'].values, width,
13        bottom=group_data['1917 - 1951'].values+group_data['1952 - 1986'].values,
14        label = '1987 - 2021',color = '#be2111')
15 ax.set_ylabel('Movie')
16 ax.set_xlabel('Genre')
17 ax.set_title('Count movie by Genre and Year')
18 ax.legend()
19
20 plt.show()
```







เปรียบเทียบจำนวนกาพยนตร์ที่ผลิตในแต่ละช่วงเวลา
จากความยาวของกาพยนตร์



GROUPED BAR CHART

```
# Define a function to categorize duration
def categorize_duration(duration):
    if duration < 90:
        return 'short'
    elif 90 <= duration <= 150:
        return 'medium'
    else:
        return 'long'

# Apply the function to create a new column 'Duration Category'
filtered_data['Duration Category'] = pd.to_numeric(
    filtered_data['Duration']).apply(categorize_duration)
```

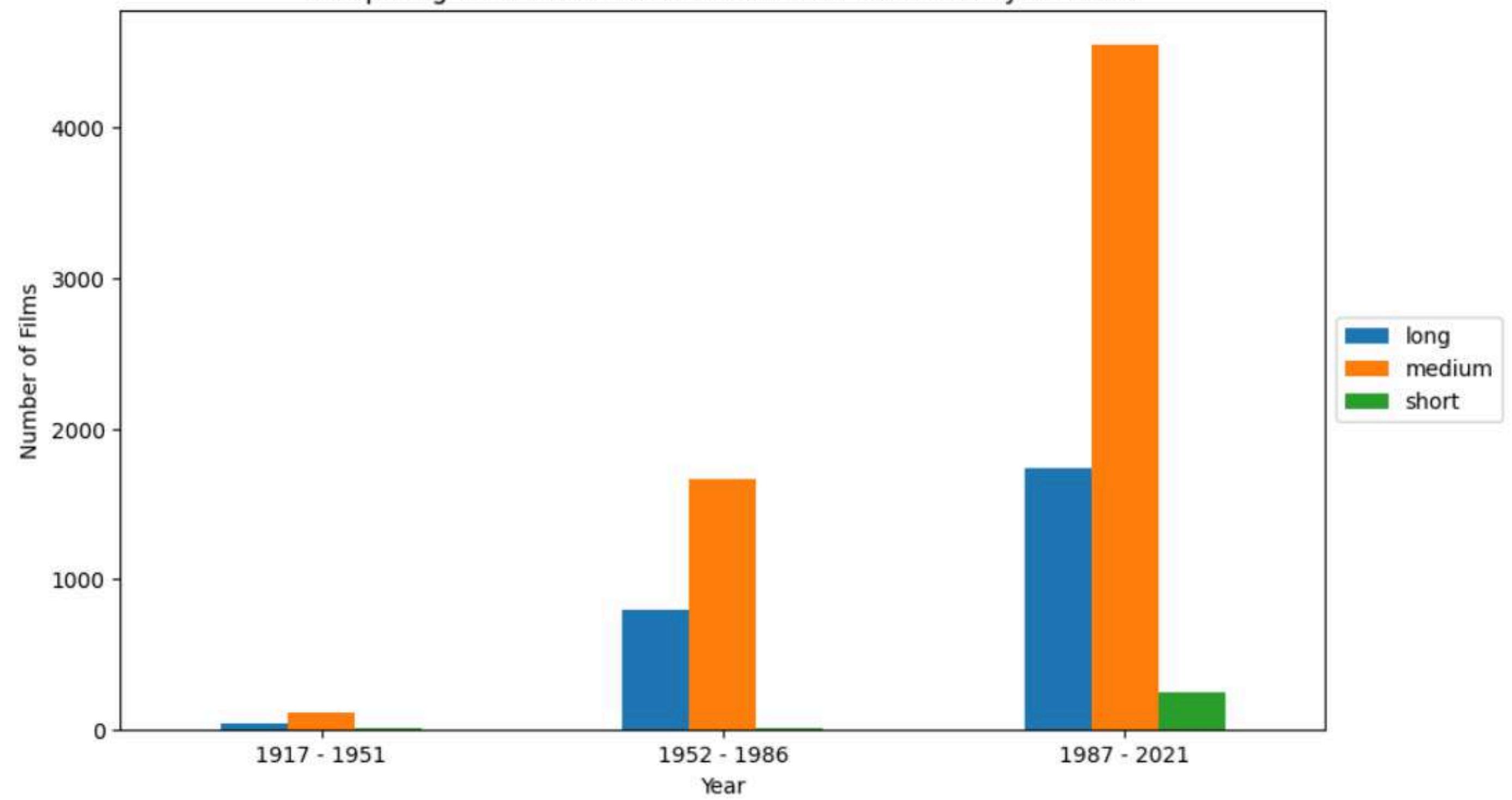


```
1 group = filtered_data.groupby(  
2     ['Year Category','Duration Category'])['Duration Category'].count().unstack()  
3 group
```

Year Category	Duration Category	long	medium	short
1917 - 1951	40	114	5	
1952 - 1986	793	1667	7	
1987 - 2021	1734	4555	244	

```
1 # Create the grouped bar chart
2 group.plot(kind='bar', figsize=(10, 6))
3
4 # Add labels and title
5 plt.ylabel('Number of Films')
6 plt.xlabel('Year')
7 plt.title('Comparing the Number of Films Produced Each Year by Duration')
8
9 plt.xticks(rotation=0)
10
11 # Move the legend outside the graph
12 plt.legend(loc='center left', bbox_to_anchor=(1.0, 0.5))
13
14 # Display the chart
15 plt.show()
```

Comparing the Number of Films Produced Each Year by Duration



THANK YOU

