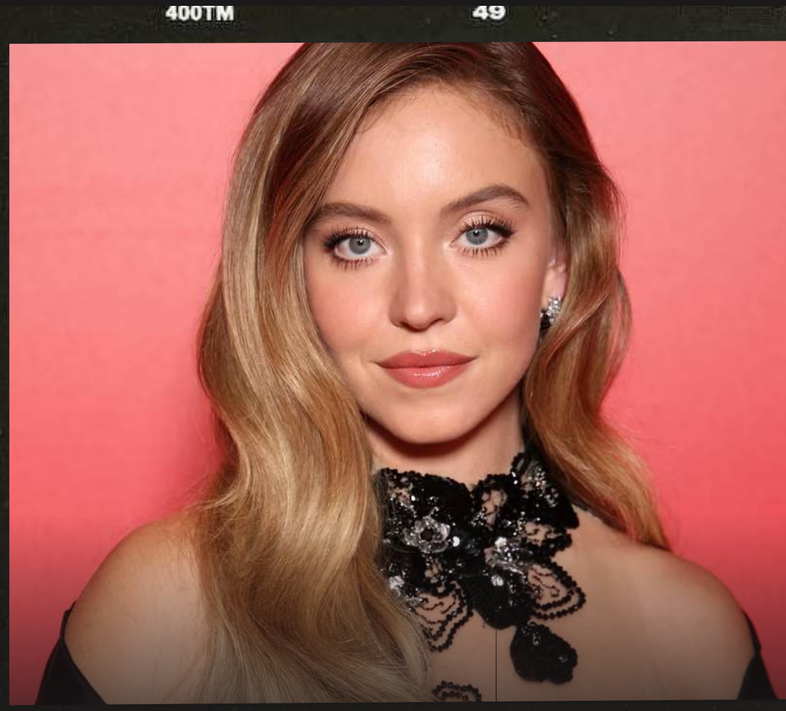
AAYUSH BASNET
PRODUCTIONS

UNDER INVESTIGATION

IMDb (an initialism for Internet Movie Database) is an online database of information related to films, television series, podcasts, home videos, video games, and streaming content online – including cast, production crew and personal biographies, plot summaries, trivia, ratings, and fan and critical reviews.



ABOUT THE ANALYSIS

Imagine it’s a quiet evening. You curl up on your couch, remote in hand, ready to escape into a world of stories. Whether it’s the spine-tingling suspense of a thriller, the heartwarming moments of a rom-com or the adrenaline rush of an action blockbuster, movies and shows have a unique way of capturing our imagination. They make us laugh, cry, and think, they inspire us, transport us to distant lands, and sometimes, even help us understand ourselves a little better.


The
King's Man


POLICE
POLICE
JOKER

TABLE OF CONTENTS

1. Objective
2. Tools and Libraries
3. Data Overview and Basic Exploration
4. Data Cleaning
5. Univariate and Bivariate Analysis
6. Genre-Specific Analysis
7. Year and Trend Analysis
8. Summary

OBJECTIVE

- Understand Evolution: Track how the movie industry has evolved in terms of genres, popularity, and production trends over time.

- Genre Exploration: Investigate the dominance of certain genres across different eras and how storytelling has adapted to changing times.

- Provide Insights: Present actionable insights for filmmakers, marketers, and cinephiles based on historical data and trends.

.

TOOLS & LIBRARIES

PYTHON

The director of this analysis—a versatile and powerful programming language that orchestrated every step of the journey.

PANDAS

The skilled editor, slicing, dicing, and organizing massive datasets into meaningful frames that tell a story.

NUMPY

The mathematical genius, crunching numbers and handling complex calculations with ease.

MATPLOTLIB

The artist, bringing raw data to life through clear and beautiful visualizations.

SEABORN

The stylist, adding flair and polish to our plots, making trends and patterns more captivating.

DATA OVERVIEW & BASIC EXPLORATION

```
# Display the descriptive statistic

df.describe()
```

	score	budget_x	revenue
count	10178.000000	1.017800e+04	1.017800e+04
mean	63.497052	6.488238e+07	2.531401e+08
std	13.537012	5.707565e+07	2.777880e+08
min	0.000000	1.000000e+00	0.000000e+00
25%	59.000000	1.500000e+07	2.858898e+07
50%	65.000000	5.000000e+07	1.529349e+08
75%	71.000000	1.050000e+08	4.178021e+08
max	100.000000	4.600000e+08	2.923706e+09

```
df.info()  # Display the summary
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10178 entries, 0 to 10177
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   names       10178 non-null  object 
 1   date_x      10178 non-null  object 
 2   score       10178 non-null  float64
 3   genre       10093 non-null  object 
 4   overview    10178 non-null  object 
 5   crew        10122 non-null  object 
 6   orig_title  10178 non-null  object 
 7   status      10178 non-null  object 
 8   orig_lang   10178 non-null  object 
 9   budget_x    10178 non-null  float64
 10  revenue     10178 non-null  float64
 11  country     10178 non-null  object 
dtypes: float64(3), object(9)
memory usage: 954.3+ KB
```

DATA CLEANING

```
# Fill the null values in 'genre' and 'crew' column with 'N/A'

df['genre'] = df['genre'].fillna('N/A')
df['crew'] = df['crew'].fillna('N/A')

df.isnull().sum()
```

```
names         0
date_x        0
score         0
genre         0
overview      0
crew          0
orig_title    0
status        0
orig_lang     0
budget_x      0
revenue       0
country       0
dtype: int64
```

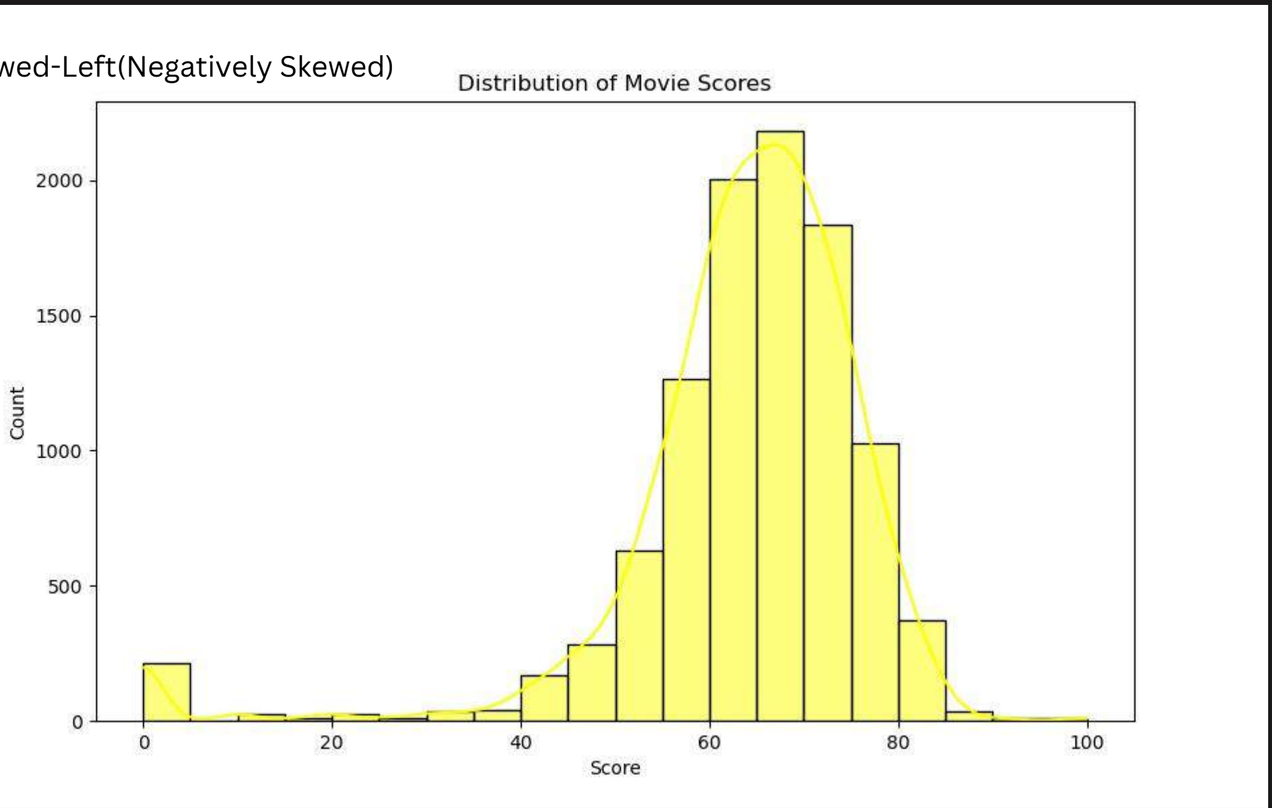
```
# Convert date_x column from object to datetime

df['date_x'] = pd.to_datetime(df['date_x'])

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10178 entries, 0 to 10177
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   names       10178 non-null  object
 1   date_x      10178 non-null  datetime64[ns]
 2   score       10178 non-null  float64
 3   genre       10093 non-null  object
 4   overview    10178 non-null  object
 5   crew        10122 non-null  object
 6   orig_title  10178 non-null  object
 7   status      10178 non-null  object
 8   orig_lang   10178 non-null  object
 9   budget_x    10178 non-null  float64
 10  revenue     10178 non-null  float64
 11  country     10178 non-null  object
dtypes: datetime64[ns](1), float64(3), object(8)
memory usage: 954.3+ KB
```

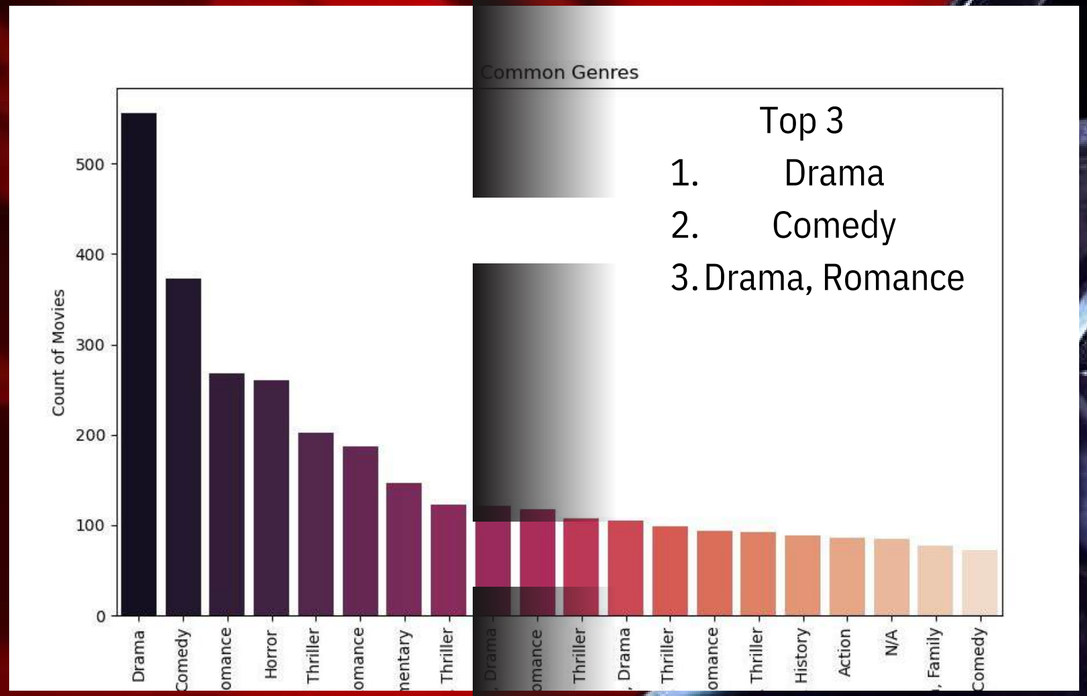
UNIVARIATE AND BIVARIATE ANALYSIS

What is the distribution of movie scores? Plot a histogram and describe its shape.

Skewed-Left(Negatively Skewed)



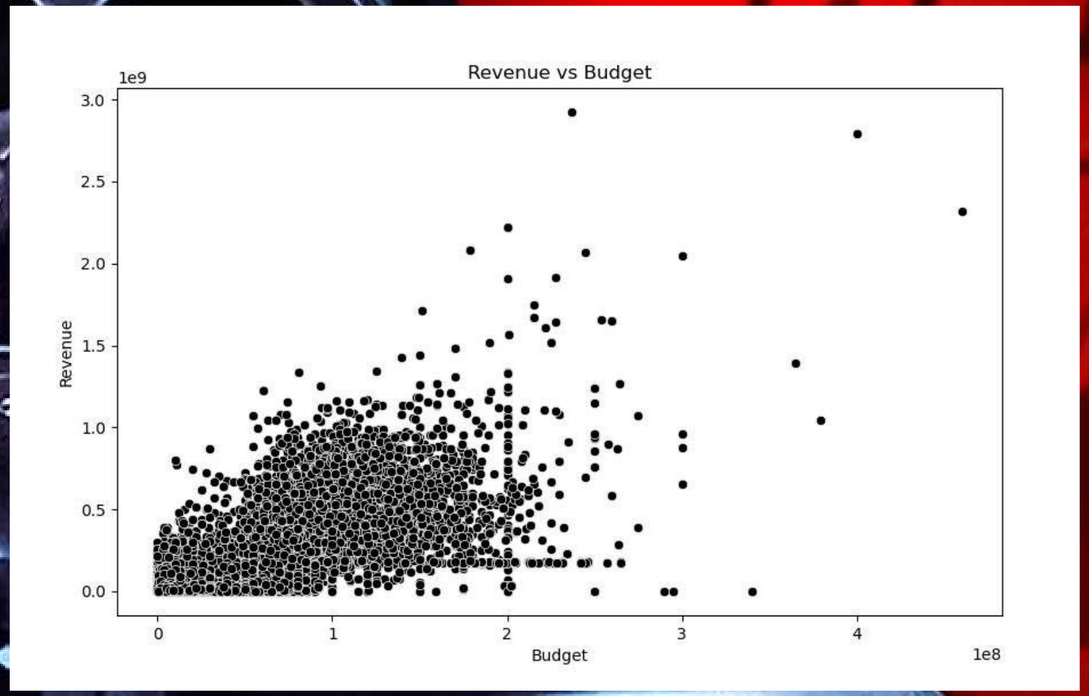
What is the most common genres in the IMDB? Use a bar chart to show their distribution.



Top 3

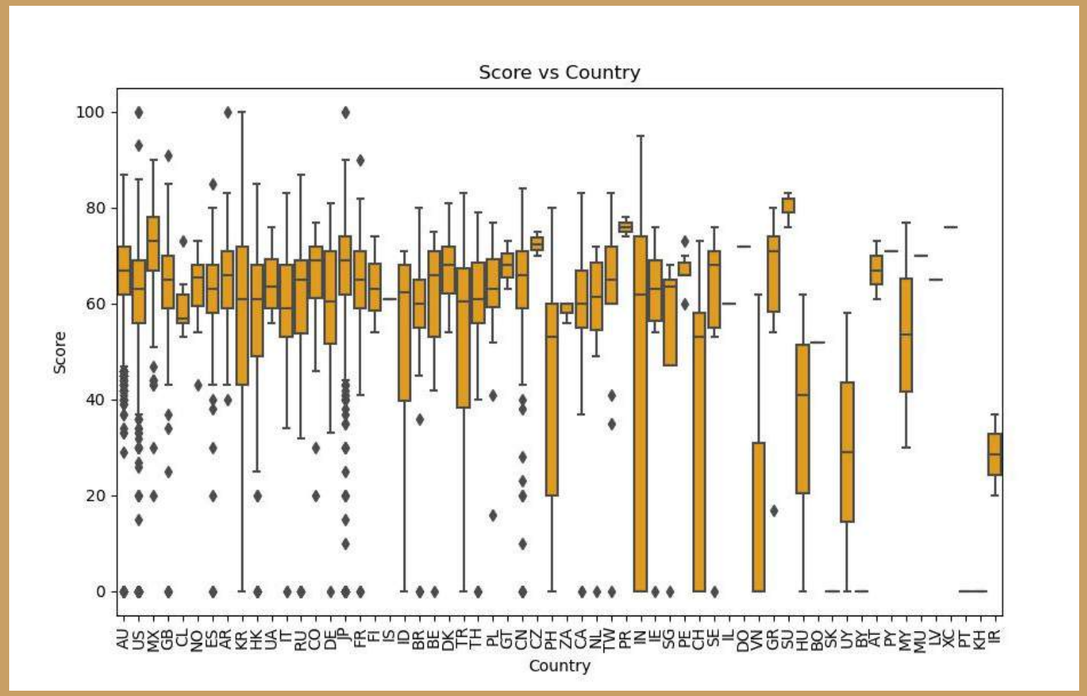
1. Drama
2. Comedy
3. Drama, Romance

Is there any relationship between a movie's budget and its revenue? Plot a scatter plot and describe any observed trend.

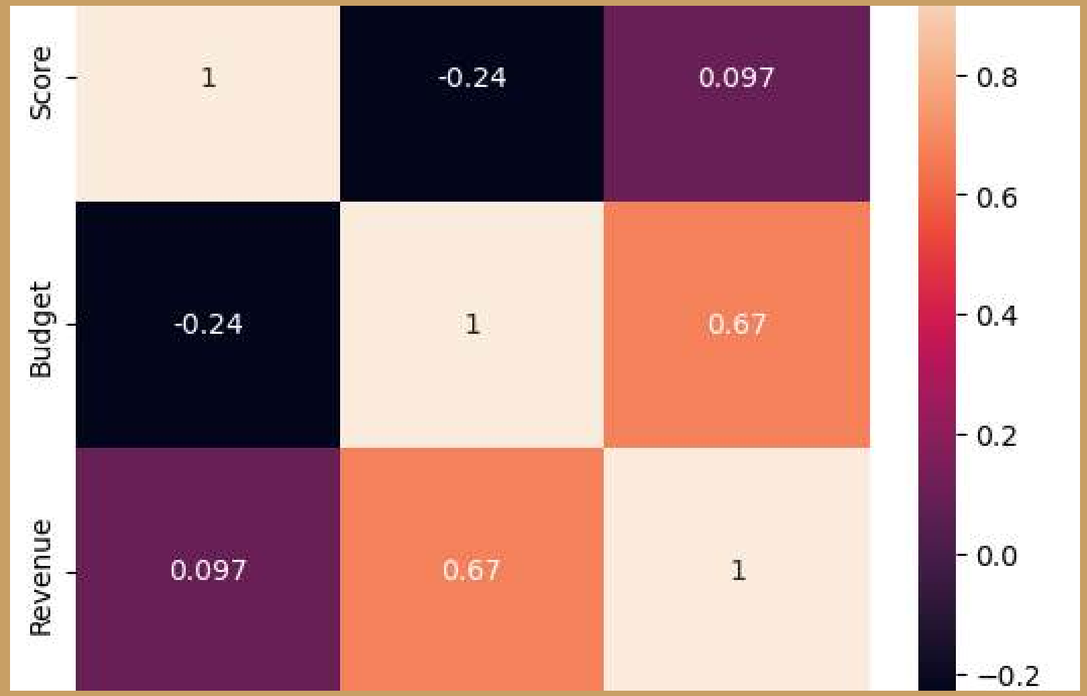


There appears to be a general postivite trend, where movies with higher budgets tend to generate higher revenue. However, the correlation is not perfect, as there is significant spred.

How do scores vary by country? Use a boxplot to visualize the difference in scores across country.

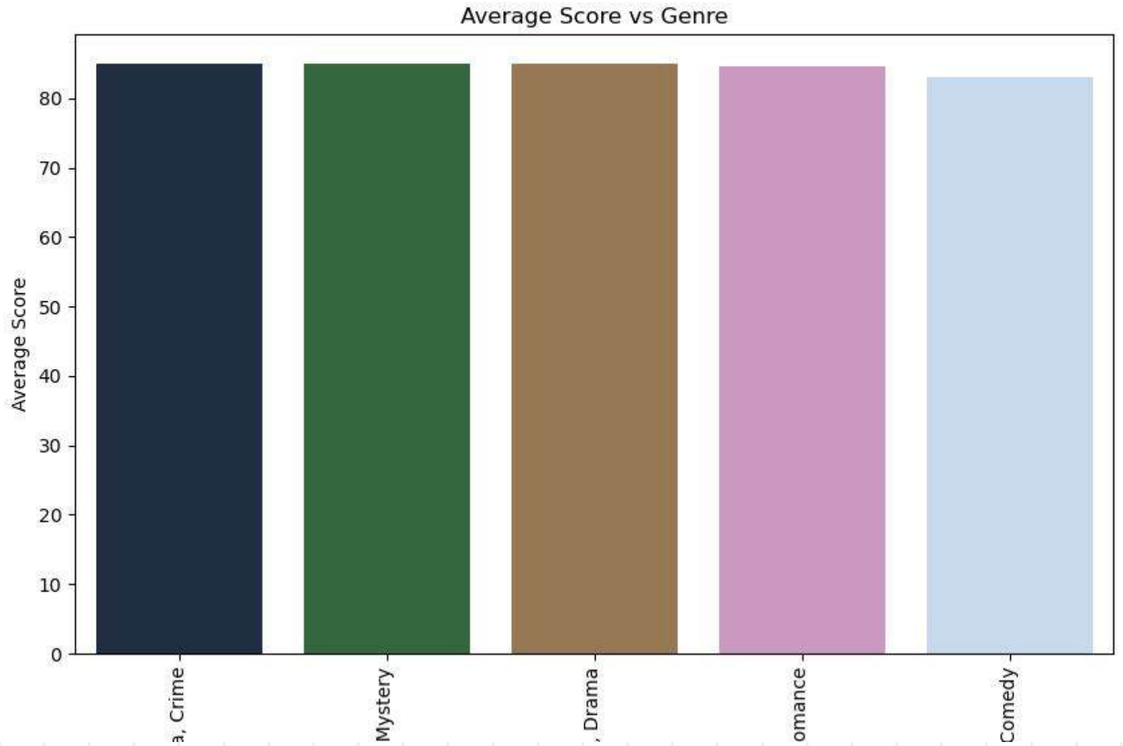


Is there a correlation between the score a movie received against its budget and revenue? Create a heatmap and calculate the correlation coefficient.



GENRE SPECIFIC ANALYSIS

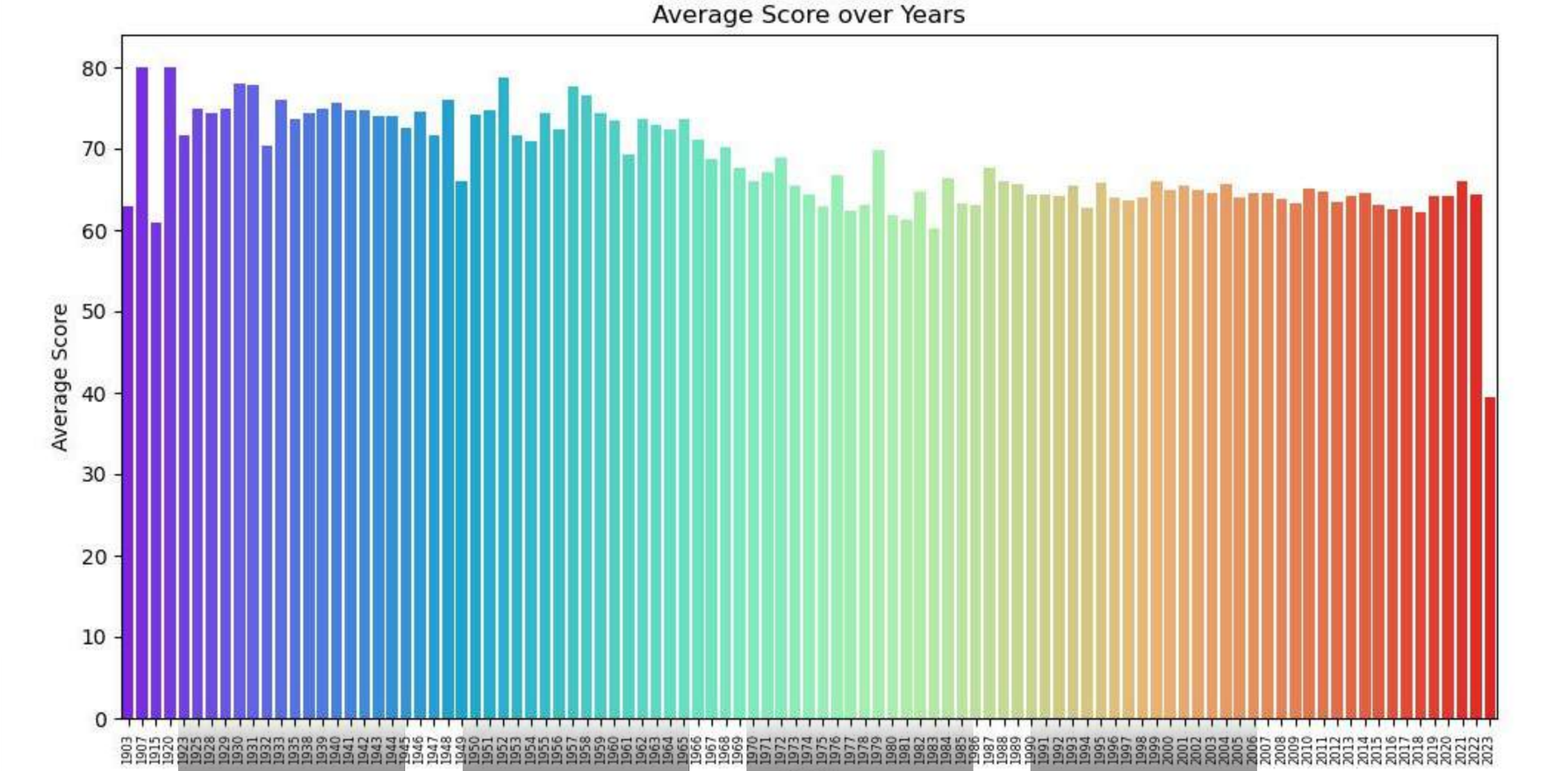
Which genre has the highest average score? Calculate the average score for each genre and plot the results.



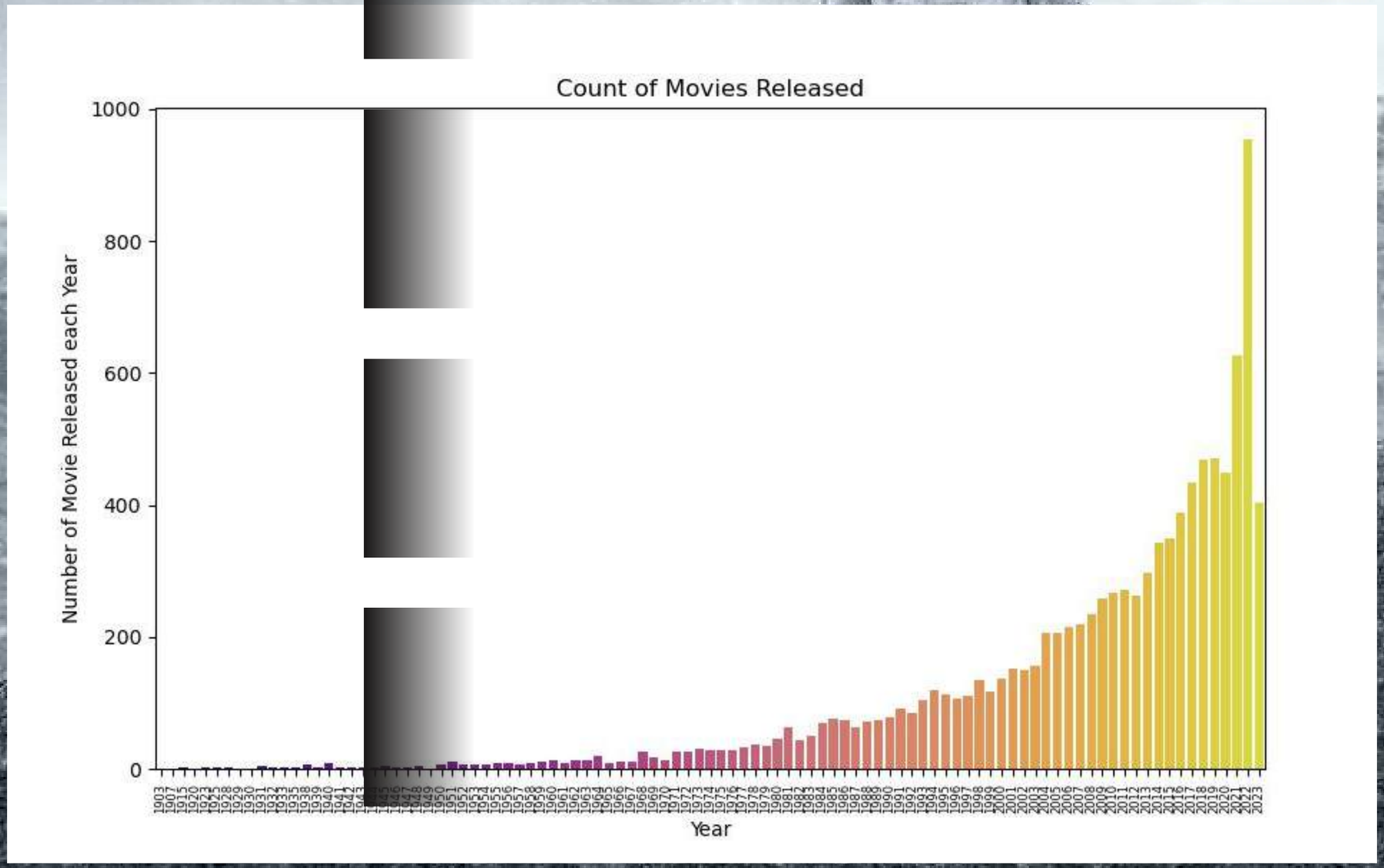
Fantasy, Drama and Crime is the genre with highest average score.

YEAR & TREND ANALYSIS

How has the average score changed over the years? Plot the average score for each year.



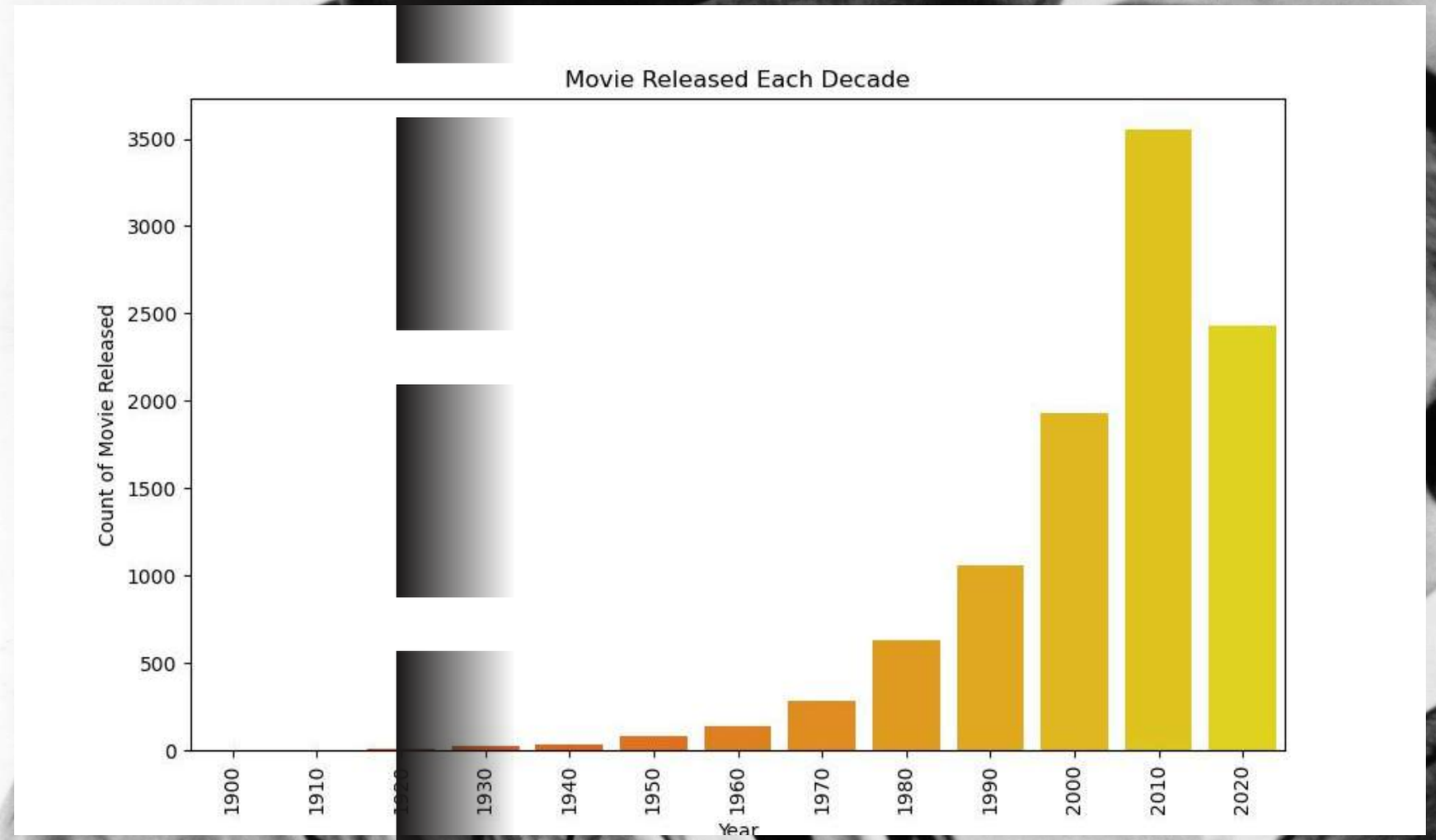
Which years had the highest and lowest number of movie releases? Plot the number of movies released each year



Year with the highest number of Moive Released is **2022**

Year with the lowest number of Moive Released is **1903**

Plot the number of movies released each decade.



2010 decade saw the most numbe of movie release(3553)

SUMMARY

- Genre Popularity over Time: Certain genres, such as action and adventure, have seen a consistent rise in popularity over the years, like driven by advancements in special effects and global box office appeal. In contrast, genre like westerns or musicals has experienced a decline, possibly due to change in audience preferences and cultural trends.
- Impact of Budget on Movie Scores: High budget moves often perform better in terms of audience and critic rating, as they can invest in better visual effects, removed directors and top-tier actors. However, there are exceptions with some low-budget films (e.g.: independent dramas or thrillers) achieving critical acclaim due to strong storytelling and innovative filmmaking.
- Seasonal Release Trends and Scores: Movies released during summer or holiday season tend to have higher box office earnings and audience rating. These periods are strategically chosen for blockbuster films that cater to family and mass audiences.

THANK YOU