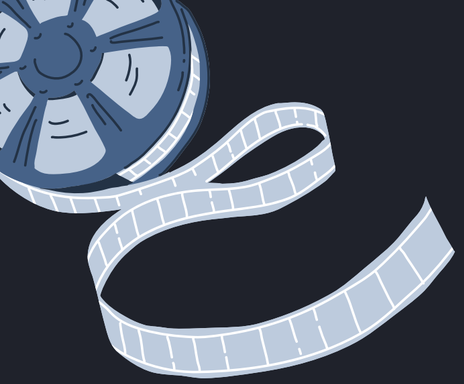
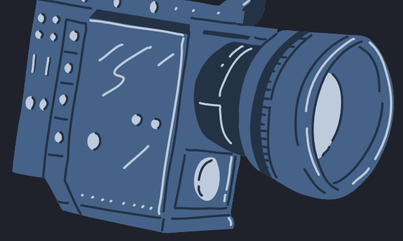
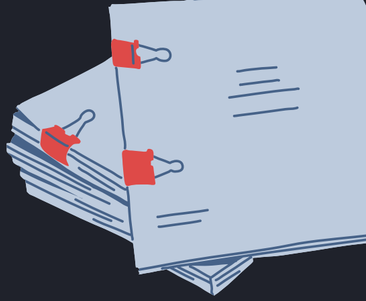
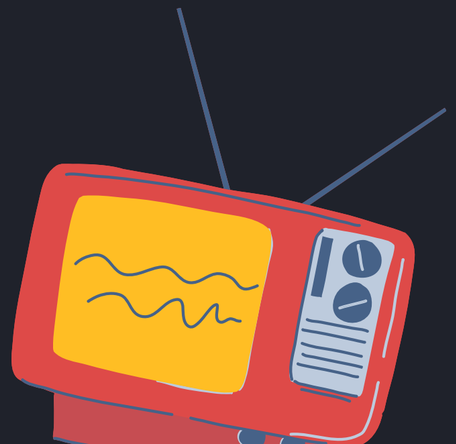
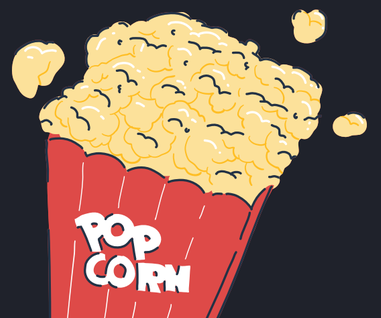


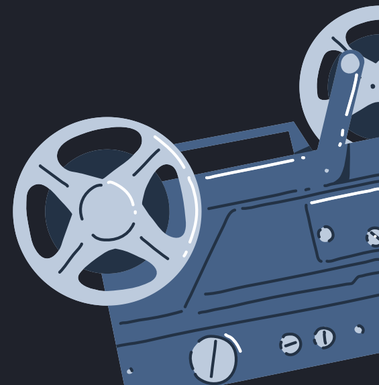





# IMDB INDIAN MOVIE





# นำเข้าข้อมูล

```
import numpy as np
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).
```

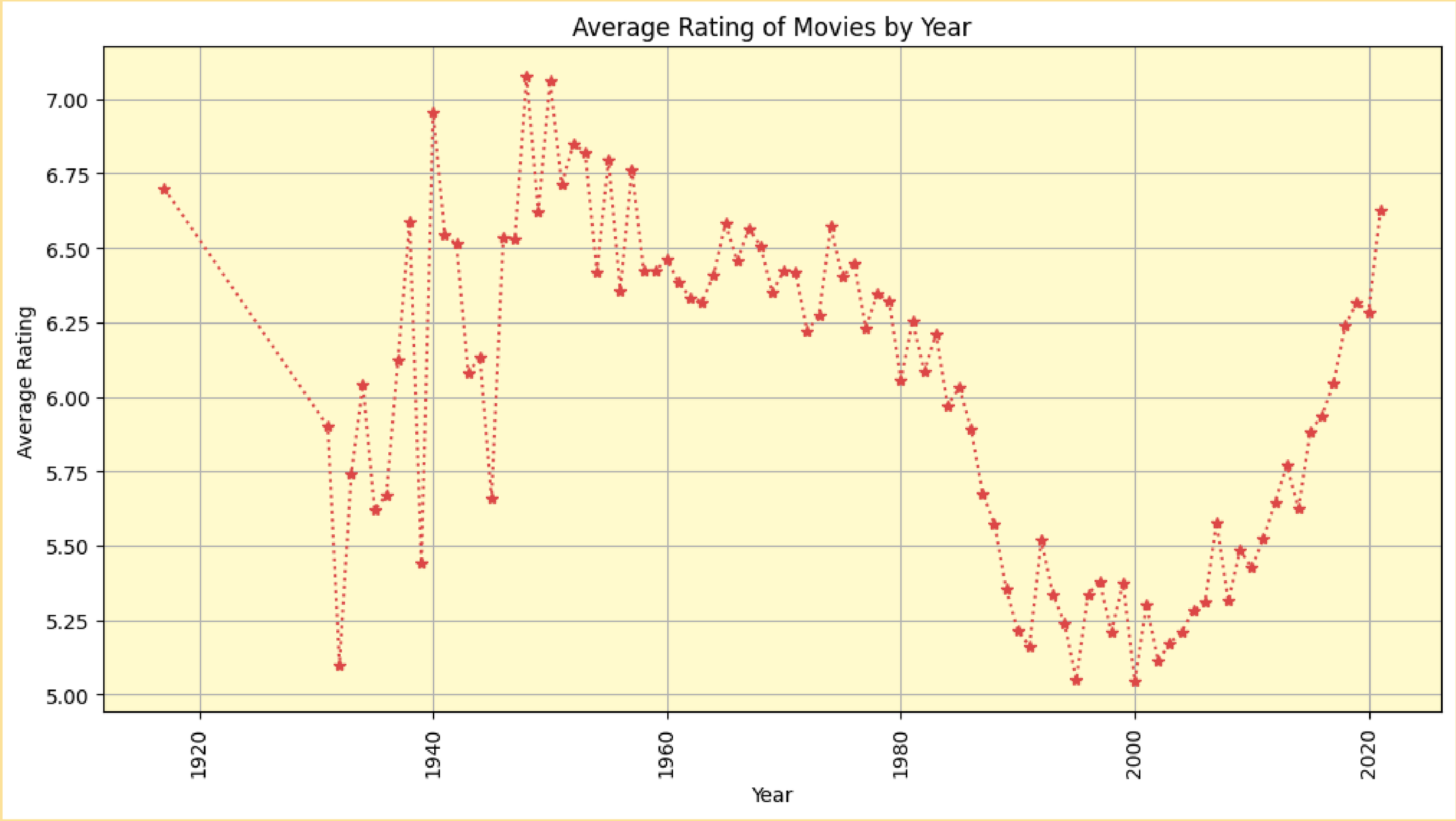
```
from google.colab import drive
drive.mount('/content/drive')
```

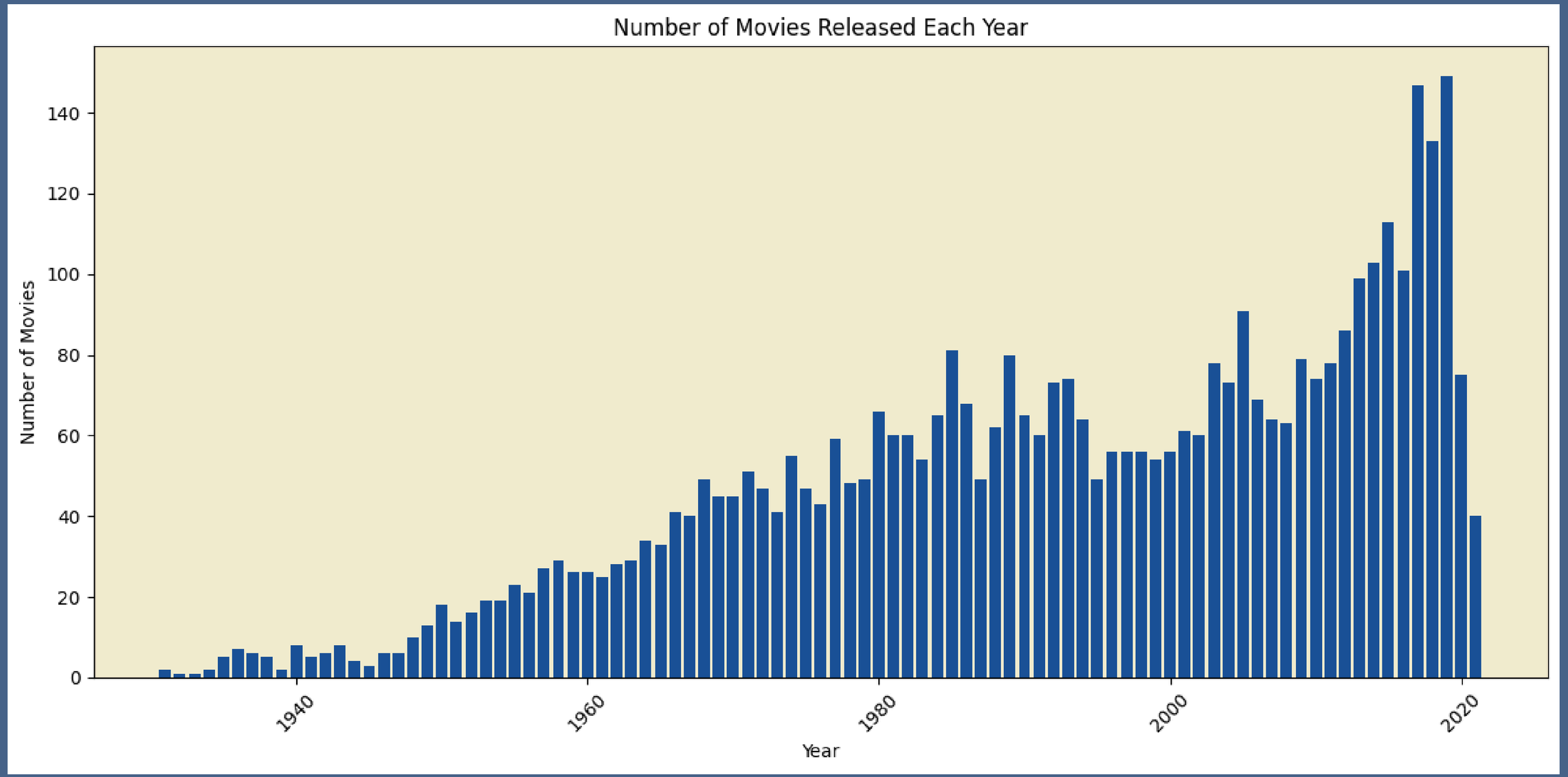
```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
df_movies = pd.read_csv('/content/drive/MyDrive/Data_Viz_2024_Data570-8/IMDb Movies India.csv', encoding='latin-1')
```

```
df_movies.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 |


Average Rating of Movies by Year
Average Rating
7.00
6.75
6.50
6.25
6.00
5.75
5.50
5.25
5.00
1920
1940
1960
1980
2000
2020
Year


Number of Movies Released Each Year
Number of Movies
0
20
40
60
80
100
120
140
1940
1960
1980
2000
2020
Year

# ทำความสะอาดข้อมูล

```
df_datamovies.shape
```

```
(15509, 10)
```

```
df_movies = df_datamovies.dropna(subset=['Year','Rating'])
```

```
df_movies.shape
```

```
(7919, 10)
```

```
remaining_percentage = (df_movies.shape[0] / df_datamovies.shape[0]) * 100
print(remaining_percentage)
```

```
51.06067444709523
```

# ทำความสะอาดข้อมูล (ต่อ)

df_movies

| | Name | Year | Duration | Genre | Rating | Votes | Director | Actor 1 | Actor 2 | Actor 3 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | #Gadhvi (He thought he was Gandhi) | (2019) | 109 min | Drama | 7.0 | 8 | Gaurav Bakshi | Rasika Dugal | Vivek Ghamande | Arvind Jangid |
| **3** | #Yaaram | (2019) | 110 min | Comedy, Romance | 4.4 | 35 | Ovais Khan | Prateik | Ishita Raj | Siddhant Kapoor |
| **5** | ...Aur Pyaar Ho Gaya | (1997) | 147 min | Comedy, Drama, Musical | 4.7 | 827 | Rahul Rawail | Bobby Deol | Aishwarya Rai Bachchan | Shammi Kapoor |
| **6** | ...Yahaan | (2005) | 142 min | Drama, Romance, War | 7.4 | 1,086 | Shoojit Sircar | Jimmy Sheirgill | Minissha Lamba | Yashpal Sharma |
| **8** | ?: A Question Mark | (2012) | 82 min | Horror, Mystery, Thriller | 5.6 | 326 | Allyson Patel | Yash Dave | Muntazir Ahmad | Kiran Bhatia |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

# ทำความสะอาดข้อมูล (ต่อ)

```
# Convert 'Year' column to string type before applying string operations
df_movies['Year'] = df_movies['Year'].astype(str).str.replace('(', '').str.replace(')', '').astype(int)
```

```
<ipython-input-30-1ead4a9e4f89>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_movies['Year'] = df_movies['Year'].astype(str).str.replace('(', '').str.replace(')', '').astype(int)
```

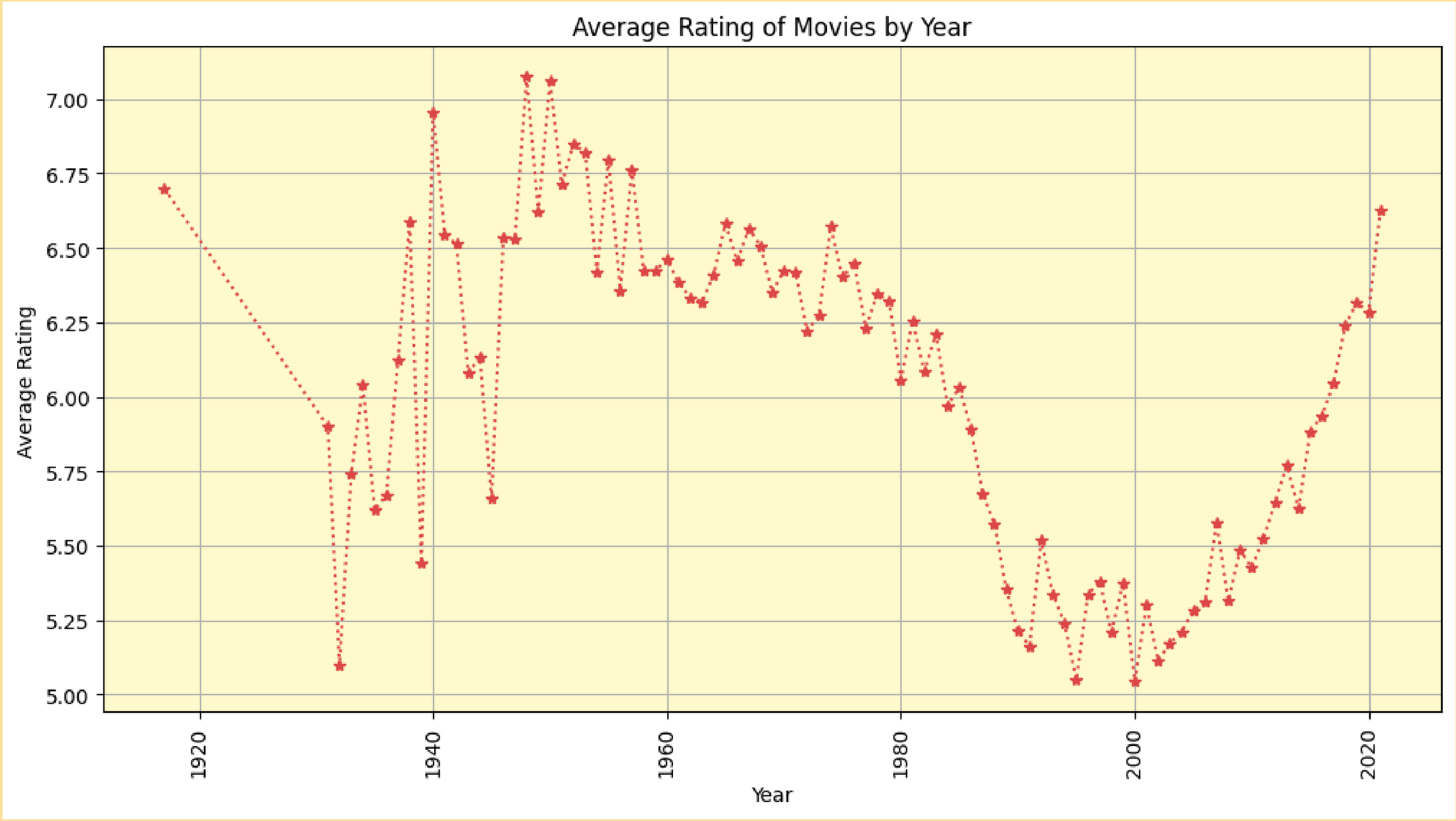
```
df_movies.head()
```

| | Name | Year | Duration | Genre | Rating | Votes | Director | Actor 1 | Actor 2 | Actor 3 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | #Gadhvi (He thought he was Gandhi) | 2019 | 109 min | Drama | 7.0 | 8 | Gaurav Bakshi | Rasika Dugal | Vivek Ghamande | Arvind Jangid |
| **3** | #Yaaram | 2019 | 110 min | Comedy, Romance | 4.4 | 35 | Ovais Khan | Prateik | Ishita Raj | Siddhant Kapoor |
| **5** | ...Aur Pyaar Ho Gaya | 1997 | 147 min | Comedy, Drama, Musical | 4.7 | 827 | Rahul Rawail | Bobby Deol | Aishwarya Rai Bachchan | Shammi Kapoor |
| **6** | ...Yahaan | 2005 | 142 min | Drama, Romance, War | 7.4 | 1,086 | Shoojit Sircar | Jimmy Sheirgill | Minissha Lamba | Yashpal Sharma |
| **8** | ?: A Question Mark | 2012 | 82 min | Horror, Mystery, Thriller | 5.6 | 326 | Allyson Patel | Yash Dave | Muntazir Ahmad | Kiran Bhatia |

# วิเคราะห์ข้อมูล-สร้างกราฟ

```
rating_per_year = df_movies.groupby('Year')['Rating'].mean().reset_index(name='Average Rating')
```

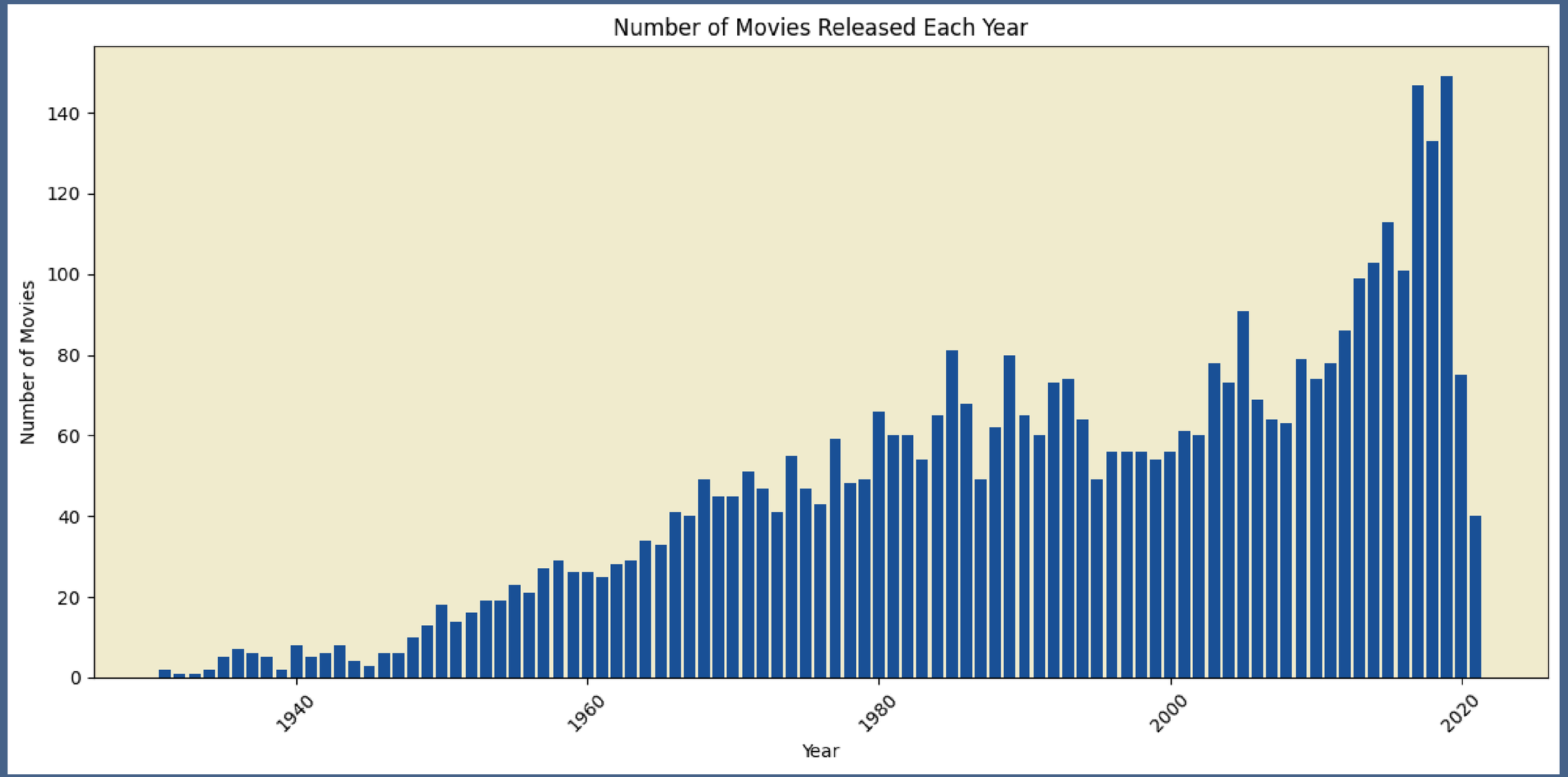
```
import matplotlib.pyplot as plt

plt.figure(figsize=(12, 6))
plt.plot(rating_per_year['Year'], rating_per_year['Average Rating'],'*:', color='#DE4A48')
plt.xlabel('Year')
plt.ylabel('Average Rating')
plt.title('Average Rating of Movies by Year')
plt.grid(True)

# เพิ่มรายละเอียดกราฟ
plt.title('Average Rating of Movies by Year')
plt.xlabel('Year')
plt.ylabel('Average Rating')
plt.xticks(rotation=90)  # หมุนป้ายปีเพื่อให้ดูง่ายขึ้น
plt.grid(True)

ax = plt.gca()  # Get the current axes object
ax.set_facecolor('#FFFACD')  # Set the background color

# แสดงกราฟ
plt.show()
```


Average Rating of Movies by Year
Average Rating
7.00
6.75
6.50
6.25
6.00
5.75
5.50
5.25
5.00
1920
1940
1960
1980
2000
2020
Year

# วิเคราะห์ข้อมูล-สร้างกราฟ

```
import matplotlib.pyplot as plt
import numpy as np

# นับจำนวนหนังในแต่ละปี
movie_counts = df_movies['Year'].value_counts().sort_index()

# สร้าง subplots
fig, ax = plt.subplots(figsize=(12, 6))

# สร้าง bar chart และกำหนดสี
ax.bar(movie_counts.index, movie_counts.values, color='#195498')

# กำหนดชื่อแกนและชื่อกราฟ
ax.set_xlabel('Year')
ax.set_ylabel('Number of Movies')
ax.set_title('Number of Movies Released Each Year')

# หมุนป้ายปีเพื่อให้อ่านง่าย
plt.xticks(rotation=45)
```

```
# ปรับพื้นหลังของพื้นที่ภายในแกน
ax.set_facecolor('#f4edd1')  # เปลี่ยนสีพื้นหลังของพื้นที่ภายในแกน

# แสดงกราฟ
plt.tight_layout()
plt.show()

# ปรับสีกริดไลน์
ax.grid(color='white')  # เปลี่ยนสีกริดไลน์เป็นสีขาว

# ปรับความโปร่งใสของกริดไลน์
ax.grid(alpha=0.5)  # เปลี่ยนความโปร่งใสของกริดไลน์
```


Number of Movies Released Each Year
Number of Movies
0
20
40
60
80
100
120
140
1940
1960
1980
2000
2020
Year

# สมาชิก

นายคณิศร กุดกลาง 653020202-7

นางสาวปีย์รดา ภู่ถนนนอก 653020211-6

นายณัฐภัทร ทักษิณ 653020568-5

นางสาวธนภรณ์ ดาษถนิม 653020570-8