



IMDb
Movies India

Home

Film

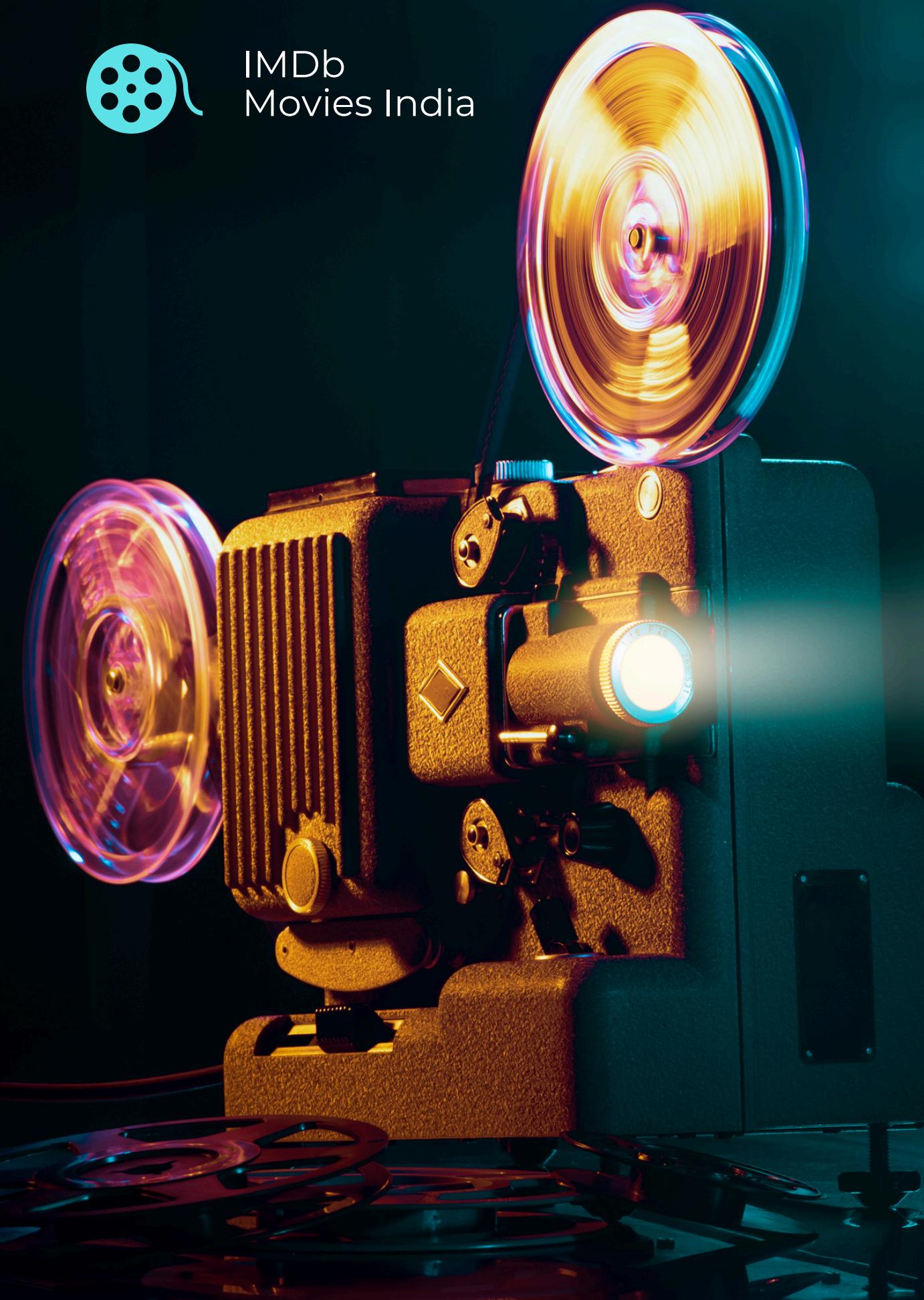
Theater

HW7

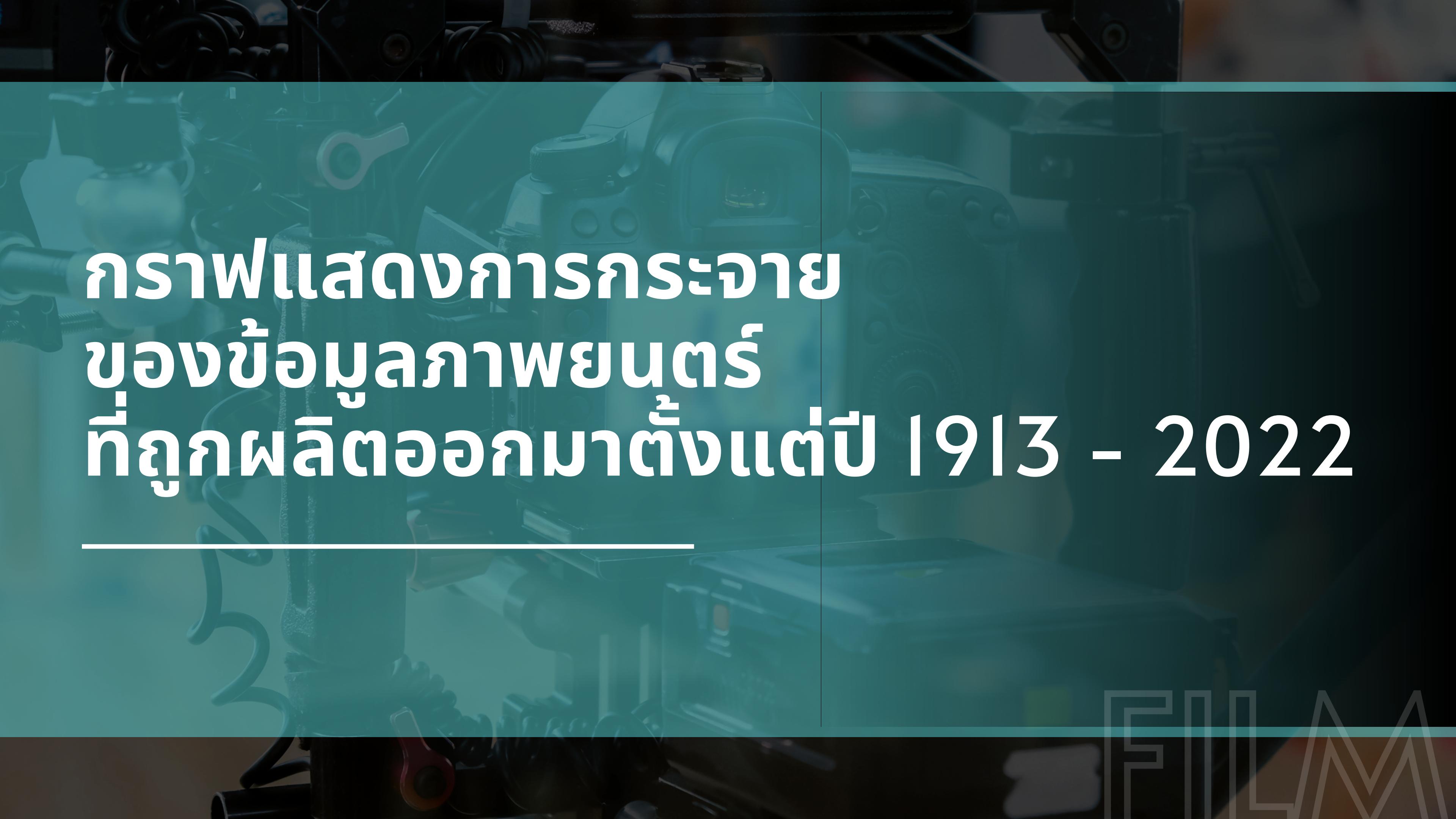
PLOT GRAPH

แสดงการกระจายของข้อมูลใน

“ INDIA IMDB ”



FILM



กราฟแสดงการกระจาย ข้อมูลสภาพน้ำต์ ที่ถูกผลิตออกมาตั้งแต่ปี 1913 - 2022





IMPORT DATA

```
[ ] import pandas as pd  
  
▶ 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).
```

```
[ ] file_path = '/content/drive/MyDrive/Dataviz_2024_DATA/IMDb Movies India.csv'  
India_Imdb = pd.read_csv(file_path, encoding='latin-1') # ລົບອັກຂະໜາດສີເຫຼືອກ  
India_Imdb.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





TRANSFORM DATA

```
[ ] # Data Cleansing  
# Drop ค่า Nan ออกจากตาราง Year และMovie_India_Imdb = India_Imdb.dropna(subset=['Year'])Movie_India_Imdb.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
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
5	...Aur Pyaar Ho Gaya	(1997)	147 min	Comedy, Drama, Musical	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan	Shammi Kapoor

FILM



```
[ ] # แปลง column Year ให้เป็นตัวเลข
# การแปลงค่าภายในคอลัมน์ Year ให้เป็นชนิดข้อมูลแบบสตริง (หรือข้อความ) ก่อนทำการประมวลผลอีกครั้ง
Movie_India_Imdb['Year'] = Movie_India_Imdb['Year'].astype(str).str.replace(r'^[^\d]', '', regex=True)
Movie_India_Imdb['Year'] = Movie_India_Imdb['Year'].astype(int)
Movie_India_Imdb.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
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
5	...Aur Pyaar Ho Gaya	1997	147 min	Comedy, Drama, Musical	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan	Shammi Kapoor

FILM



```
[ ] # สร้าง DataFrame ใหม่ที่จัดกลุ่มจำนวนหนังที่ออกฉายในแต่ละปี
Movies_per_year = Movie_India_Imdb.groupby('Year').size().reset_index(name='Movie_count')
Movies_per_year
```

	Year	Movie_count
0	1913	1
1	1914	1
2	1915	1
3	1917	1
4	1922	1
...
97	2018	381
98	2019	410
99	2020	320
100	2021	392
101	2022	38

102 rows × 2 columns

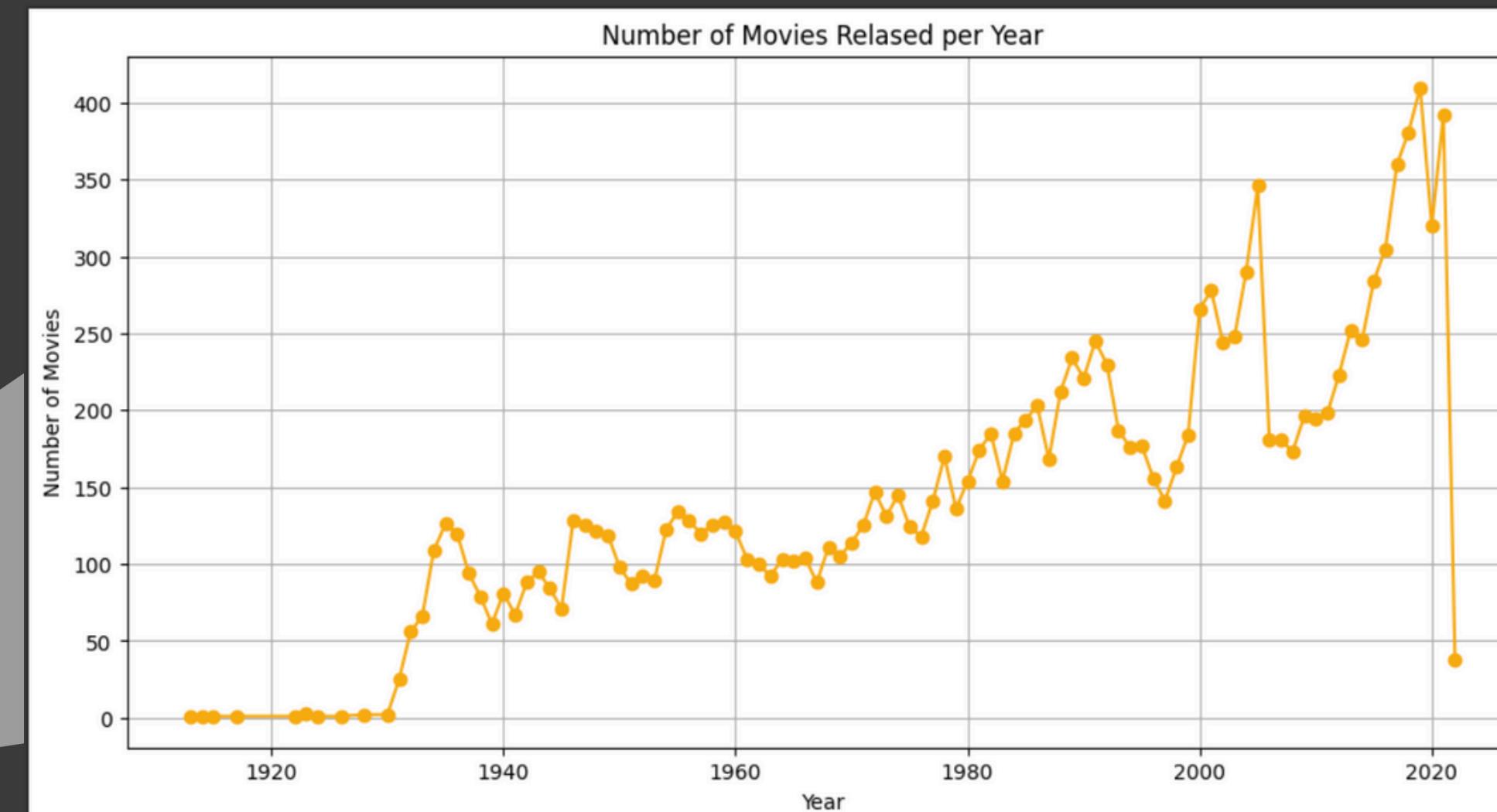
FILM

PLOT GRAPH LINE CHART

Film



```
# สร้างกราฟ Line chart
plt.figure(figsize=(12, 6)) # กำหนดขนาดของกราฟ
plt.plot(Movies_per_year[ 'Year' ], Movies_per_year[ 'Movie_count' ], marker='o', linestyle='-', color = '#f8ab11')
plt.title('Number of Movies Relased per Year')
plt.xlabel('Year')
plt.ylabel('Number of Movies')
plt.grid(True)
plt.show()
```





กราฟแสดงการกระจาย ของข้อมูลสภาพนั้นที่กำกับโดย ผู้กำกับแต่ละคน





IMPORT DATA

```
[ ] import pandas  
import pandas as pd  
import matplotlib.pyplot as plt
```

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

→ Mounted at /content/drive

```
[ ] df = pd.read_csv('/content/drive/MyDrive/IMDb Movies India.csv', encoding='latin')  
df
```



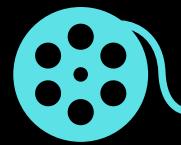


TRANSFORM DATA

```
[ ] # ลบข้อมูลที่มี NAN ออกจากตาราง  
df1 = df.dropna(subset='Director')  
df1
```

	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
...
15503	Zulm Ki Zanjeer	(1989)	125 min	Action, Crime, Drama	5.8	44	S.P. Muthuraman	Chiranjeevi	Jayamalini	Rajinikanth
15504	Zulm Ko Jala Doonga	(1988)	NaN	Action	4.6	11	Mahendra Shah	Naseeruddin Shah	Sumeet Saigal	Suparna Anand
15505	Zulmi	(1999)	129 min	Action, Drama	4.5	655	Kuku Kohli	Akshay Kumar	Twinkle Khanna	Aruna Irani
15506	Zulmi Raj	(2005)	NaN	Action	NaN	NaN	Kiran Thej	Sangeeta Tiwari	NaN	NaN
15508	Zulm-O-Sitam	(1998)	130 min	Action, Drama	6.2	20	K.C. Bokadia	Dharmendra	Jaya Prada	Arjun Sarja

14984 rows × 10 columns



```
# นับจำนวนผู้กำกับในคอลัมน์ Director  
num_directors = df1['Director'].nunique()  
# แสดงจำนวนผู้กำกับ  
print(f'จำนวนผู้กำกับในคอลัมน์ Director: {num_directors}')
```



จำนวนผู้กำกับในคอลัมน์ Director: 5938

FILM



```
[ ] #หาว่าผู้กำกับแต่ละคนทั้ง 5938 คน สร้างหนังคนละกี่เรื่อง  
director_counts = df1['Director'].value_counts()  
director_counts
```

Director	count
Jayant Desai	58
Kanti Shah	57
Babubhai Mistry	50
Mahesh Bhatt	48
Master Bhagwan	47
...	...
Naeem Siddiqui	1
Shadaab Khan	1
Mystelle Brabbee	1
Kunal Shivdasani	1
Kiran Thej	1

5938 rows × 1 columns
dtype: int64



จัดกลุ่มผู้กำกับที่สร้างหนังจำนวนเท่ากันให้อยู่กลุ่มเดียวกัน

```
director_groups = director_counts.groupby(director_counts).count()  
director_groups
```

count	
count	
1	3896
2	808
3	379
4	179
5	137
6	103
7	76
8	51
9	33
10	37
11	21
12	20
13	20
14	23
15	20
16	15
17	19
18	10
19	7
20	8
21	4
22	11
23	3
24	6
25	3
26	5
27	2
28	1
29	5
30	5
31	2
32	8
33	2
34	2
35	1
36	3
39	2
42	1
44	3
46	2
47	1
48	1
50	1
57	1
58	1

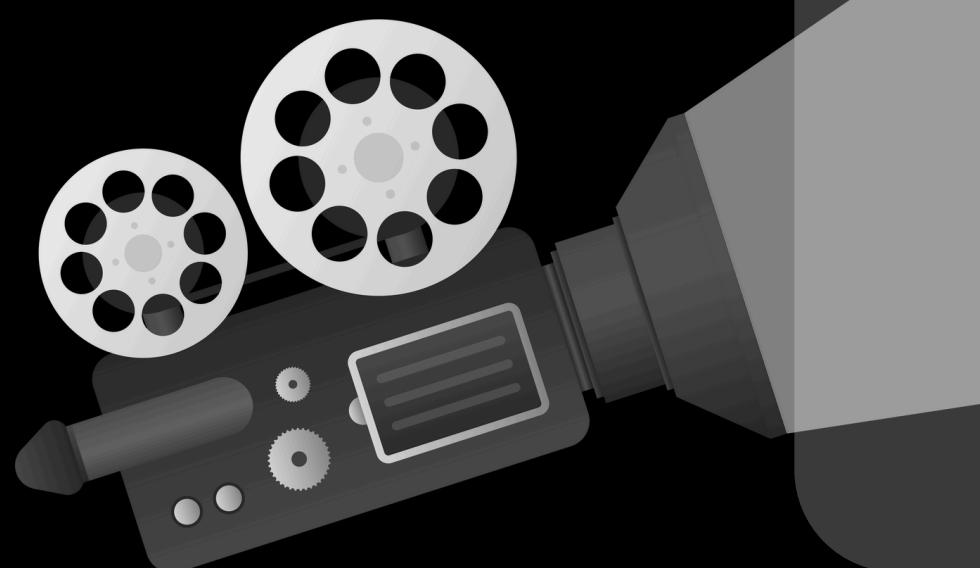
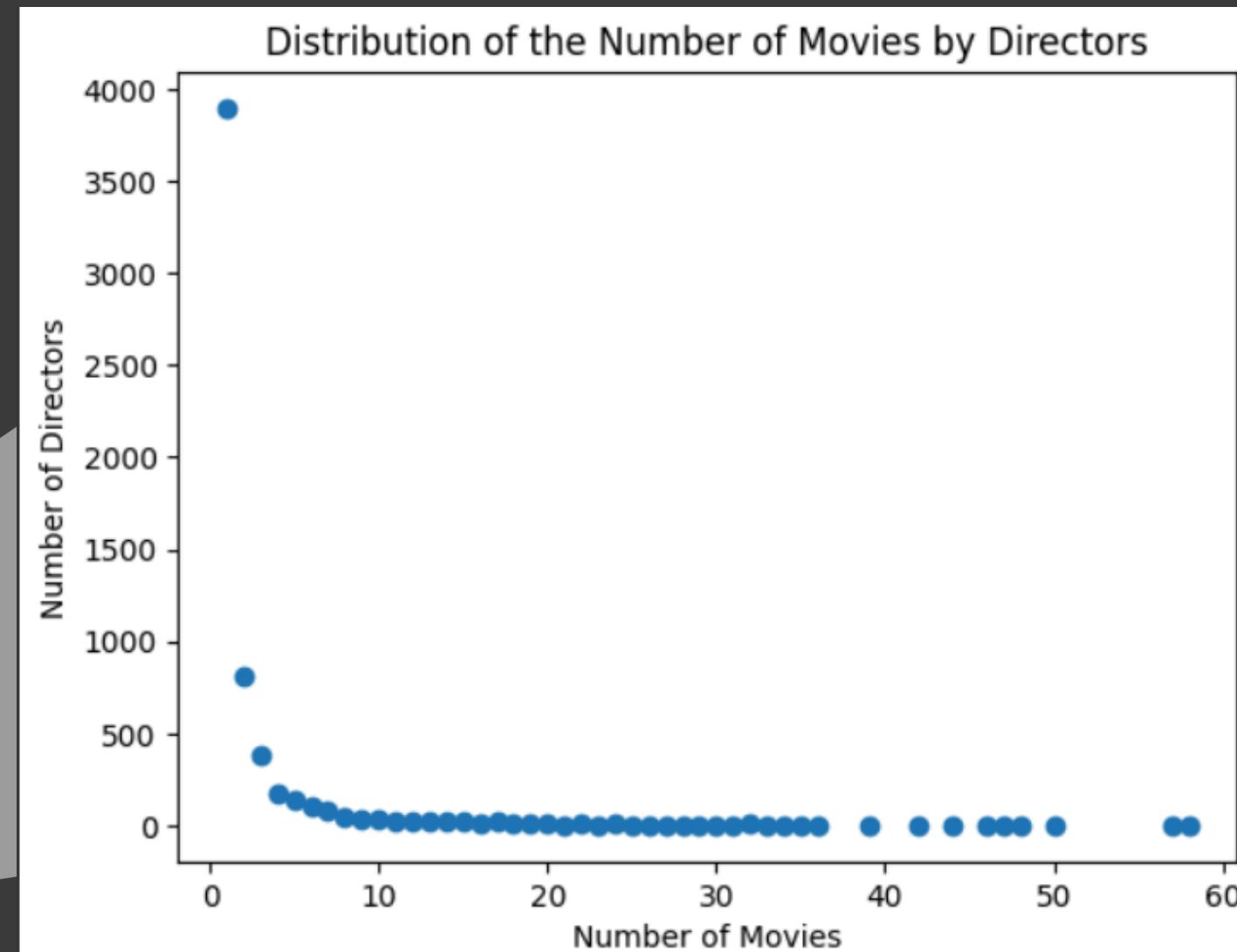
FILM

PLOT GRAPH SCATTER PLOT

FILM



```
[ ] # สร้างกราฟการกระจายข้อมูล
plt.scatter(director_groups.index, director_groups.values)
plt.xlabel('Number of Movies')
plt.ylabel('Number of Directors')
plt.title('Distribution of the Number of Movies by Directors')
plt.show()
```



FILM

สมาชิก

นางสาว กัญญา ดาโรฤทธิ์

นาย จิรภัทร แสงสุกใส

นางสาว พรวลัย พอกซ์ออล

นางสาว มนัสนันท์ มีชำนาญ

รหัสนักศึกษา 653020200-1

รหัสนักศึกษา 653020203-5

รหัสนักศึกษา 653020573-2

รหัสนักศึกษา 643020090-1



THANK YOU

FILM



IMDb
Movies India

Home

Film

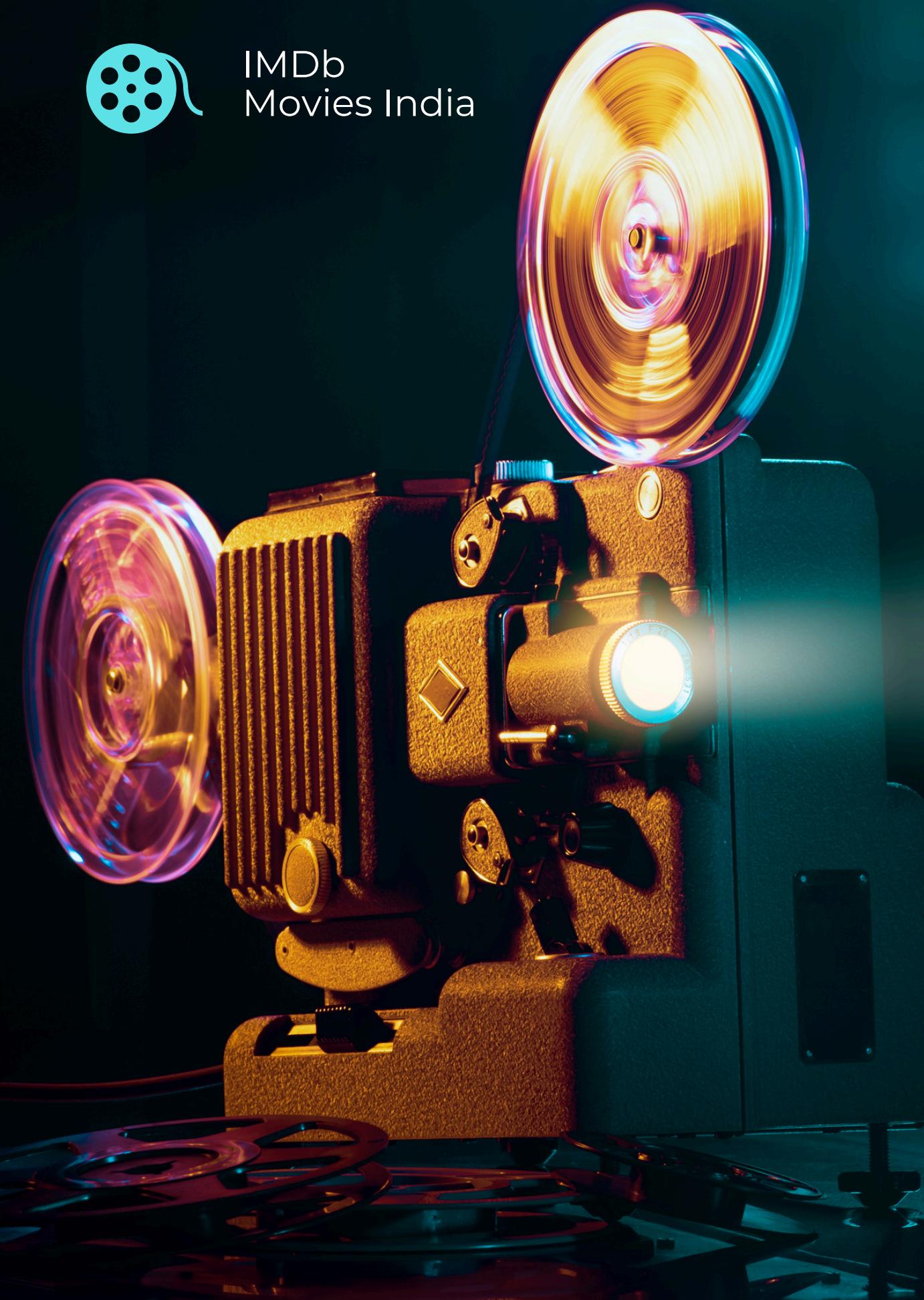
Theater

HW7

PLOT GRAPH

แสดงการกระจายของข้อมูลใน

“ INDIA IMDB ”



FILM