

NETFLIX

By សុខាន់មិនការម៉ាប



MEMBER



นางสาวณัฐกฤดา ไชยโภว
นางสาวสิริญาพร วงศ์จักร์
นางสาวพรวลัย พีอกซ์ออล
นางสาวรุ่น ศิริพราหมณกุล

653020206-9
653020218-2
653020573-2
653020603-9

Popular on Netflix



Recently Watched



ABOUT DATA

SHOW_ID : รหัสที่ไม่ซ้ำกันของรายการทีวี/ภาพยนตร์
ทุกเรื่อง

TYPE : ตัวระบุประเภท TV SHOW หรือ MOVIE

TITLE : ชื่อภาพยนตร์ / รายการทีวี

DIRECTOR : ชื่อผู้กำกับ

CAST : นักแสดงที่ร่วมแสดงในภาพยนตร์/รายการนี้

COUNTRY : ประเทศที่ภาพยนตร์/การแสดงดังกล่าวถูกผลิตขึ้น

DATA_ADDED : วันที่ถูกเพิ่งลงใน NETFLIX

RELEASE_YEAR : ปีที่ออกจำหน่ายจริง

RATING : เ雷ตติ้งตามเรตอายุและประเภทของสื่อ

DURATION : ระยะเวลาความยาวของภาพยนตร์/รายการทีวีโดย TV SHOW เป็นชั่วโมง MOVIE เป็นนาที

LISTED_IN : ประเภทของภาพยนตร์หรือรายการ

DESCRIPTION : เรื่องย่อโดยสรุปของภาพยนตร์หรือรายการ

PREPARING DATA

สร้างคอลัมน์ CONTINENT 🔎

```
def get_continent(countries):
    if isinstance(countries, str):
        countries = [country.strip() for country in countries.split(',')]

    continents = set()
    for country in countries:
        found_continent = False
        for continent, country_list in continent_mapping.items():
            if country in country_list:
                continents.add(continent)
                found_continent = True
                break
        if not found_continent:
            # Handle cases where a country is not in the mapping
            pass

    if len(continents) == 1:
        return list(continents)[0]
    elif len(continents) > 1:
        return 'International'
    else:
        return 'Unknown'

df_Netflix['Continent'] = df_Netflix['country'].apply(get_continent)
df_Netflix.head()
```

Genre2	Genre3	Cast1	Cast2	month_added	Continent
Unknown	Unknown	Unknown	Unknown	September	North America
TV Dramas	TV Mysteries	Ama Qamata	Khosi Ngema	September	Africa
International TV Shows	TV Action & Adventure	Sami Bouajila	Tracy Gotoas	September	Unknown
Reality TV	Unknown	Unknown	Unknown	September	Unknown

PREPARING DATA

สร้างคอลัมน์ FESTIVAL



```
# สร้างฟังก์ชันเพื่อกำหนดเทศกาลตามเดือน
def get_festival(month):
    festivals = {
        'January': 'New Year',
        'February': 'Valentine's Day',
        'October': 'Halloween',
        'November': 'Thanksgiving',
        'December': 'Christmas'
    }
    return festivals.get(month, 'No Festival')
```

```
# สร้างคอลัมน์เทศกาลใน df_Netflix
df_Netflix['festival'] = df_Netflix['month_added'].apply(get_festival)

# ดูผลลัพธ์
print(df_Netflix[['month_added', 'festival']].head())
```

Genre2	Genre3	Cast1	Cast2	month_added	Continent	festival
Unknown	Unknown	Unknown	Unknown	September	North America	No Festival
V Dramas	TV Mysteries	Ama Qamata	Khosi Ngema	September	Africa	No Festival
International TV Shows	TV Action & Adventure	Sami Bouajila	Tracy Gotoas	September	Unknown	No Festival
Reality TV	Unknown	Unknown	Unknown	September	Unknown	No Festival
Mantic TV Shows	TV Comedies	Mayur More	Jitendra Kumar	September	Asia	No Festival

PREPARING DATA

สร้างตารางสำหรับประเทศไทยและทวีป 

```
new_rows = []
for index, row in df_Netflix.iterrows():
    continents = row['country'].split(',') if isinstance(row['country'], str) else [row['country']]
    for continent in continents:
        new_row = row.copy()
        new_row['country_all'] = continent.strip()
        new_rows.append(new_row)
```

```
Netflix_2 = pd.DataFrame(new_rows)
Netflix_2
```

```
def get_continent_2(country):
    if isinstance(country, str):
        country = country.strip()
    for continent, country_list in continent_mapping.items():
        if country in country_list:
            return continent
    return 'Unknown'
```

```
Netflix_2['continent_all'] = Netflix_2['country_all'].apply(get_continent_2)
Netflix_2.head()
```



Genre3	Cast1	Cast2	month_added	Continent	festival	Era	country_all	continent_all
Unknown	Unknown	Unknown	September	North America	No Festival	2020-Present	United States	North America
TV Mysteries	Ama Qamata	Khosi Ngema	September	Africa	No Festival	2020-Present	South Africa	Africa
TV Action & Adventure	Sami Bouajila	Tracy Gotoas	September	Unknown	No Festival	2020-Present	Unknown	Unknown
Unknown	Unknown	Unknown	September	Unknown	No Festival	2020-Present	Unknown	Unknown

PREPARING DATA

สร้างตารางสำหรับข้อมูล RADAR CHART 🔎

```
# Filter DataFrame for type = 'Movie'
movie_df = radar_df[radar_df['type'] == 'Movie']

# Convert 'duration' to numeric, coercing errors to NaN
movie_df['duration'] = pd.to_numeric(movie_df['duration'], errors='coerce')

# Create summary table
summary = movie_df.groupby('Genre1').agg(
    total_titles=('show_id', 'nunique'),
    unique_directors=('director', lambda x: x.nunique() if 'Unknown' not in x else x[x != 'Unknown'].nunique()),
    avg_duration=('duration', 'mean'),
    unique_ratings=('rating', 'nunique'),
    unique_countries=('country_all', 'nunique')
).reset_index()

summary
```

```
# Filter DataFrame for type = 'Movie'
movie_df = radar_df[radar_df['type'] == 'TV Show']

# Convert 'duration' to numeric, coercing errors to NaN
movie_df['duration'] = pd.to_numeric(movie_df['duration'], errors='coerce')

# Create summary table
summary2 = movie_df.groupby('Genre1').agg(
    total_titles=('show_id', 'nunique'),
    unique_directors=('director', lambda x: x.nunique() if 'Unknown' not in x else x[x != 'Unknown'].nunique()),
    avg_duration=('duration', 'mean'),
    unique_ratings=('rating', 'nunique'),
    unique_countries=('country_all', 'nunique')
).reset_index()

summary2
```

```
# เลือกคอลัมน์ที่ต้องการ normalization (ยกเว้น 'Genre1')
cols_to_normalize = ['total_titles', 'unique_directors', 'avg_duration', 'unique_ratings',
'unique_countries']

# สร้าง scaler
scaler = MinMaxScaler(feature_range=(1, 10))

# ทำ normalization ข้อมูลในคอลัมน์ที่เลือก
summary[cols_to_normalize] = scaler.fit_transform(summary[cols_to_normalize])

# เปลี่ยนชื่อคอลัมน์ Genre1 เป็น Genre
summary.rename(columns={'Genre1': 'Genre'}, inplace=True)

# แสดงผลลัพธ์หลังจากทำ normalization
summary
```

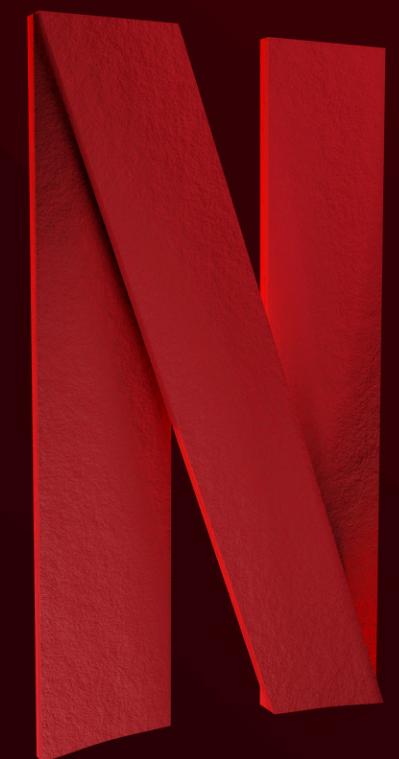
```
# เลือกคอลัมน์ที่ต้องการ normalization (ยกเว้น 'Genre1')
cols_to_normalize = ['total_titles', 'unique_directors', 'avg_duration', 'unique_ratings',
'unique_countries']

# สร้าง scaler
scaler = MinMaxScaler(feature_range=(1, 10))

# ทำ normalization ข้อมูลในคอลัมน์ที่เลือก
summary2[cols_to_normalize] = scaler.fit_transform(summary2[cols_to_normalize])

# เปลี่ยนชื่อคอลัมน์ Genre1 เป็น Genre
summary2.rename(columns={'Genre1': 'Genre'}, inplace=True)

# แสดงผลลัพธ์หลังจากทำ normalization
summary2
```



PREPARING DATA

สร้างตารางสำหรับข้อมูล RADAR CHART 

MOVIE

	Genre	total_titles	unique_directors	avg_duration	unique_ratings	unique_countries	type
0	Action & Adventure	5.829268	5.392697	9.238374	9.1	6.966292	Movie
1	Anime Features	1.112570	1.127422	6.576286	6.4	1.303371	Movie
2	Children & Family Movies	4.399625	3.903875	5.416579	9.1	5.853933	Movie
3	Classic Movies	1.444653	1.382265	10.000000	7.3	2.820225	Movie
4	Comedies	7.804878	7.498510	8.451606	8.2	7.067416	Movie
5	Cult Movies	1.061914	1.073770	7.701389	3.7	1.505618	Movie
6	Documentaries	5.660413	5.600596	5.721580	8.2	8.382022	Movie
7	Dramas	10.000000	10.000000	9.111068	10.0	10.000000	Movie
8	Horror Movies	2.542214	2.602832	7.423803	5.5	5.348315	Movie
9	Independent Movies	1.106942	1.127422	7.516349	3.7	1.606742	Movie
10	International Movies	1.714822	1.757824	8.606189	5.5	4.640449	Movie

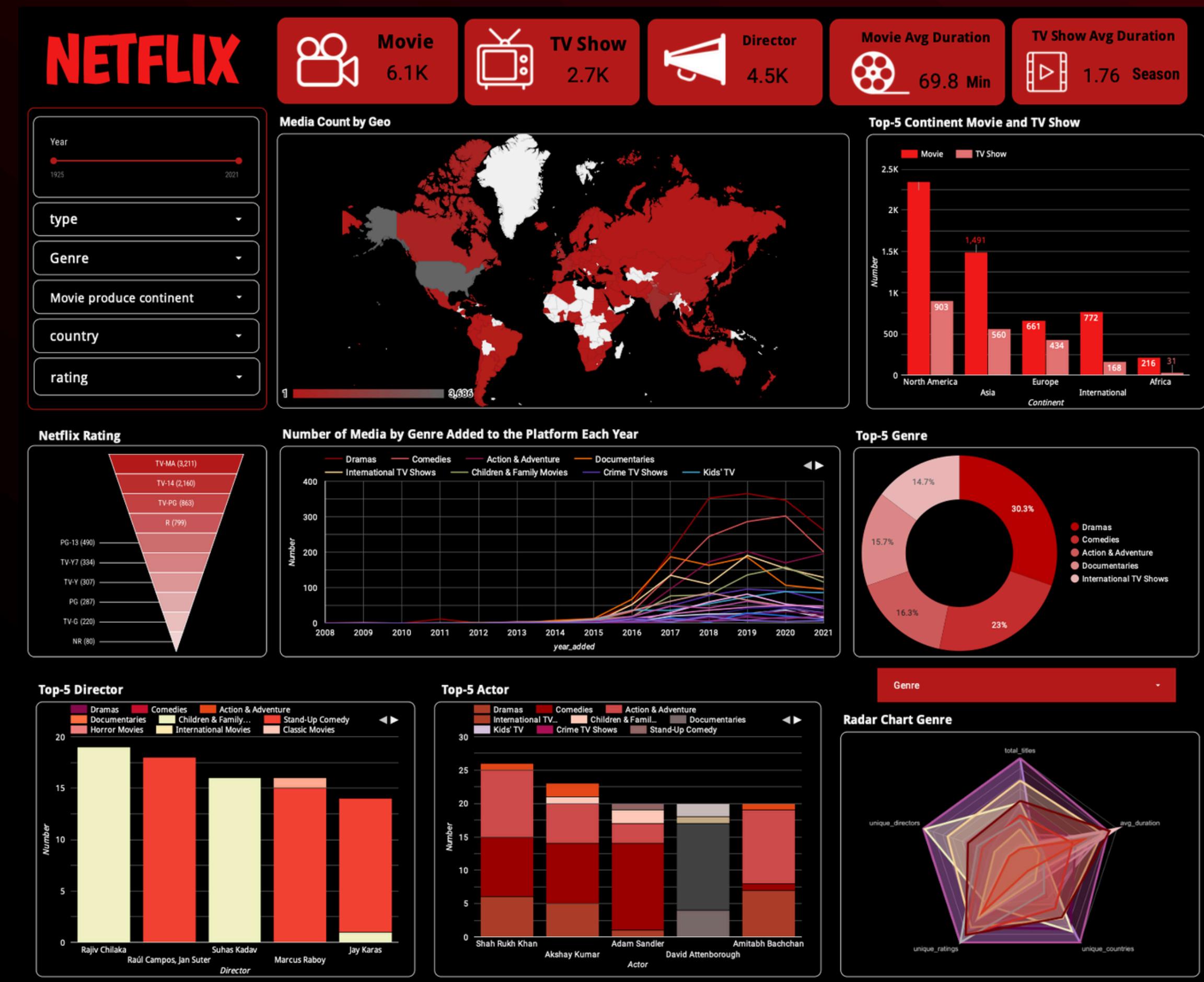
TV SHOW

	Genre	total_titles	unique_directors	avg_duration	unique_ratings	unique_countries	type
0	Anime Series	3.037516	2.65	1.938889	7.0	2.500000	TV Show
1	British TV Shows	3.934023	4.45	2.387800	10.0	4.000000	TV Show
2	Classic & Cult TV	1.244502	1.30	10.000000	7.0	1.833333	TV Show
3	Crime TV Shows	5.633894	9.85	2.335953	5.5	8.666667	TV Show
4	Docuseries	3.561449	4.90	1.614346	7.0	4.500000	TV Show
5	International TV Shows	10.000000	10.00	1.608894	10.0	10.000000	TV Show
6	Kids' TV	5.505821	3.10	2.788251	8.5	5.666667	TV Show
7	Reality TV	2.385511	1.15	1.988542	5.5	2.833333	TV Show
8	Romantic TV Shows	1.360931	1.15	4.081481	4.0	1.833333	TV Show
9	Spanish-Language TV Shows	1.011643	1.00	1.000000	2.5	1.166667	TV Show
10	Stand-Up Comedy & Talk Shows	1.384217	2.05	3.039216	5.5	1.166667	TV Show
11	TV Action & Adventure	1.454075	1.30	4.141667	5.5	2.000000	TV Show
12	TV Comedies	2.385511	1.45	4.550538	4.0	1.833333	TV Show



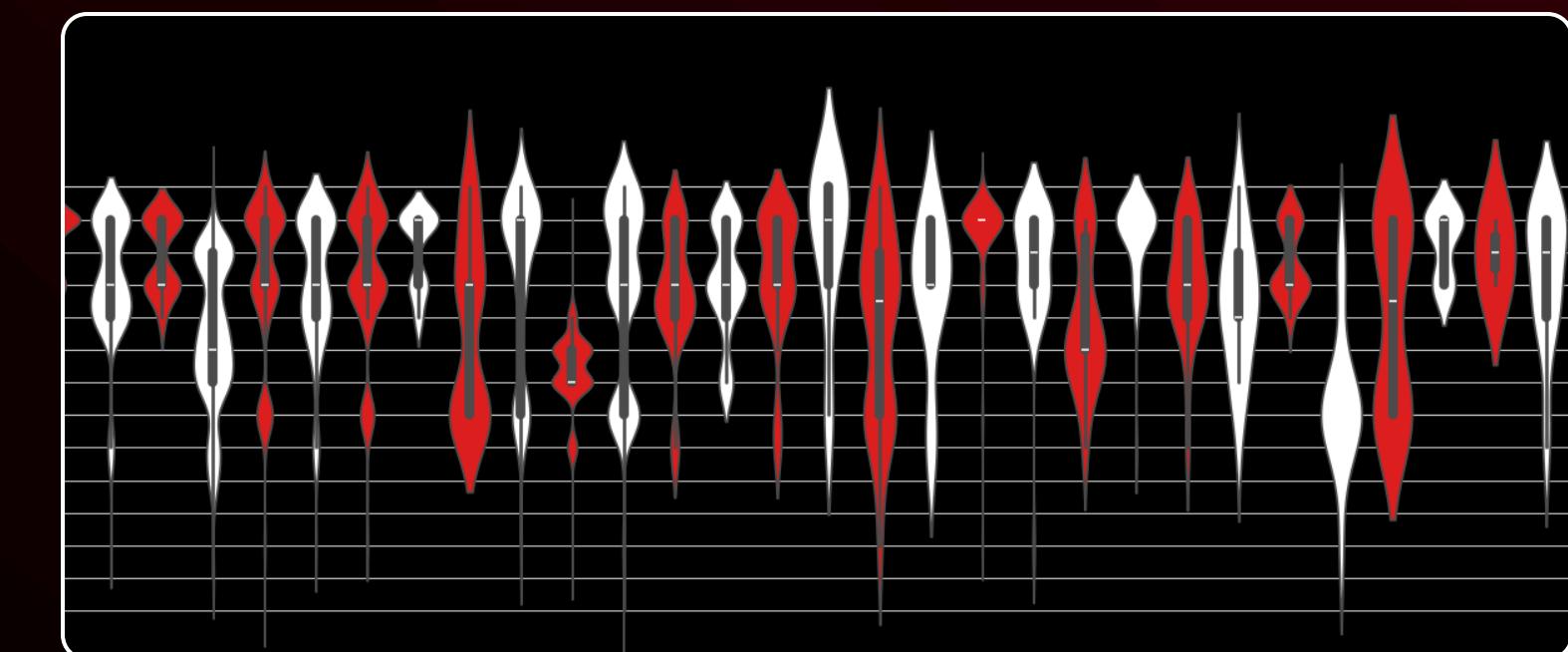
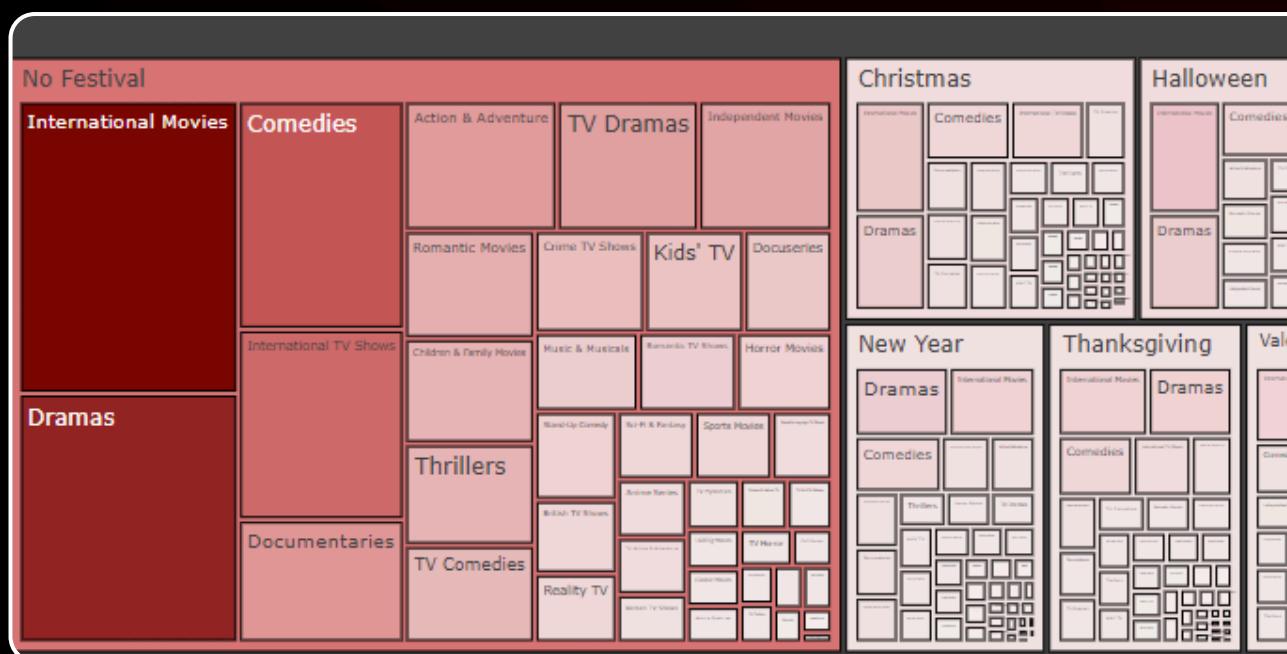
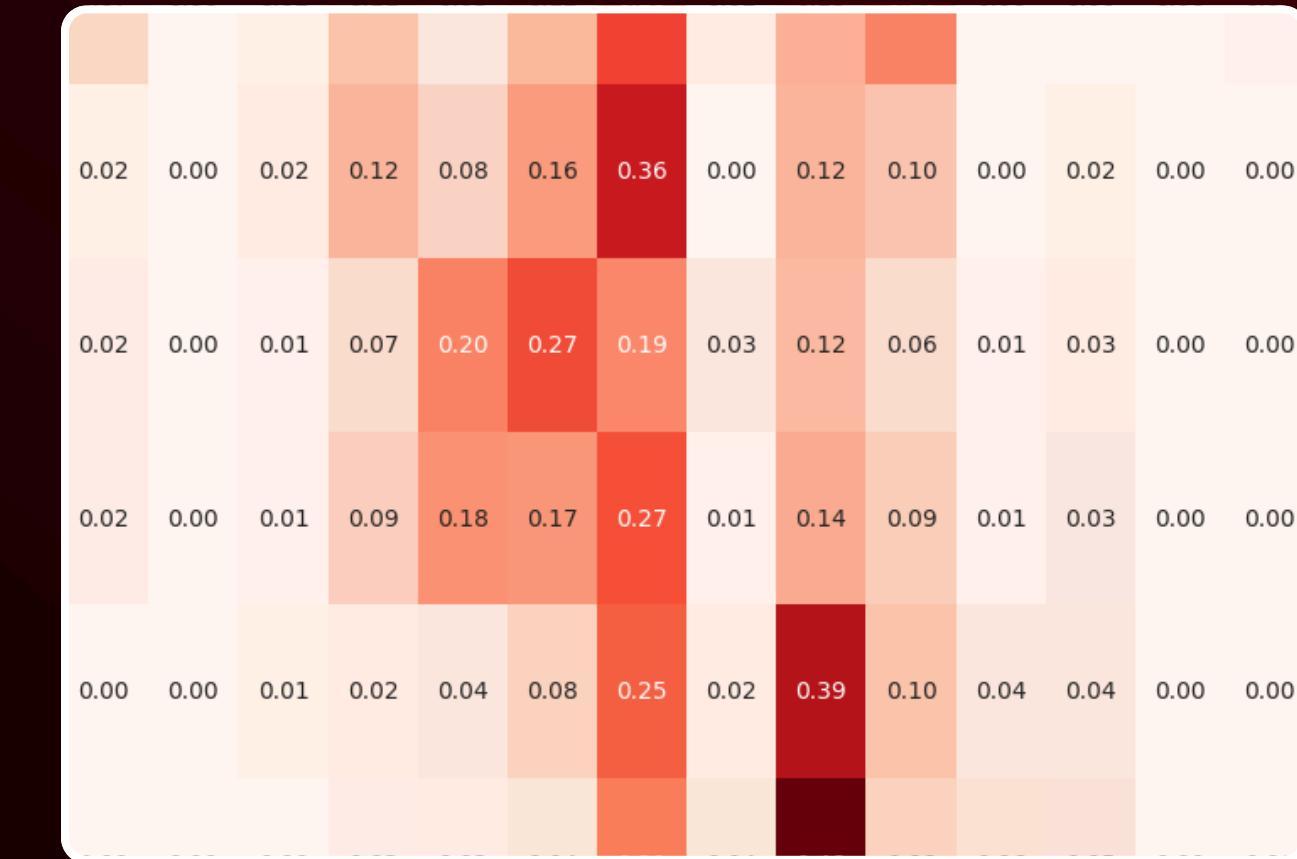
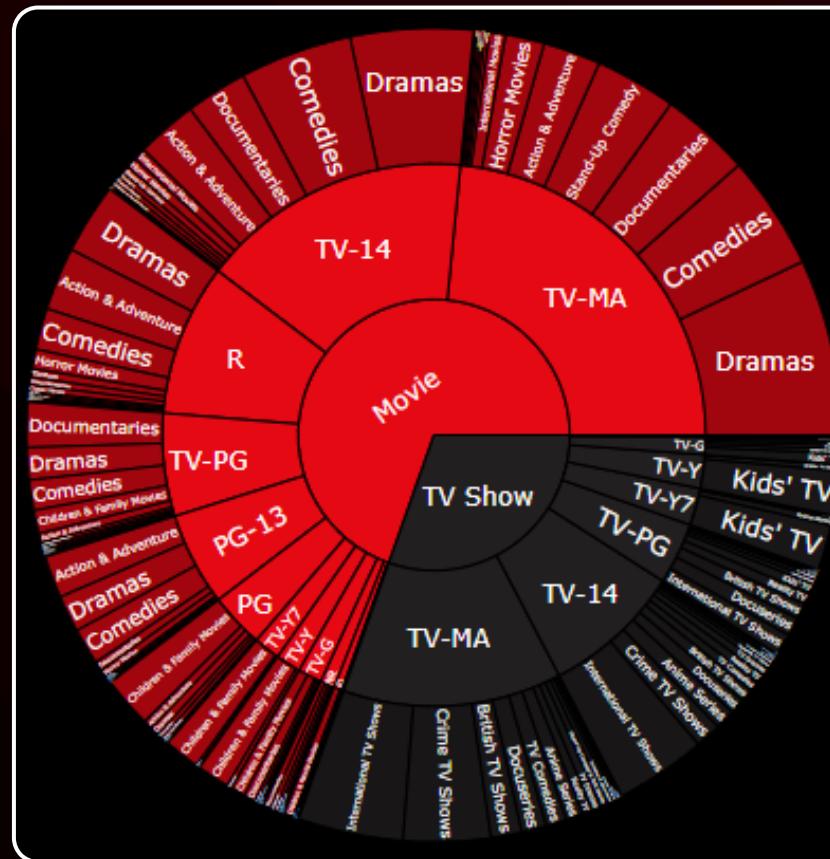
DASH BOARD

7



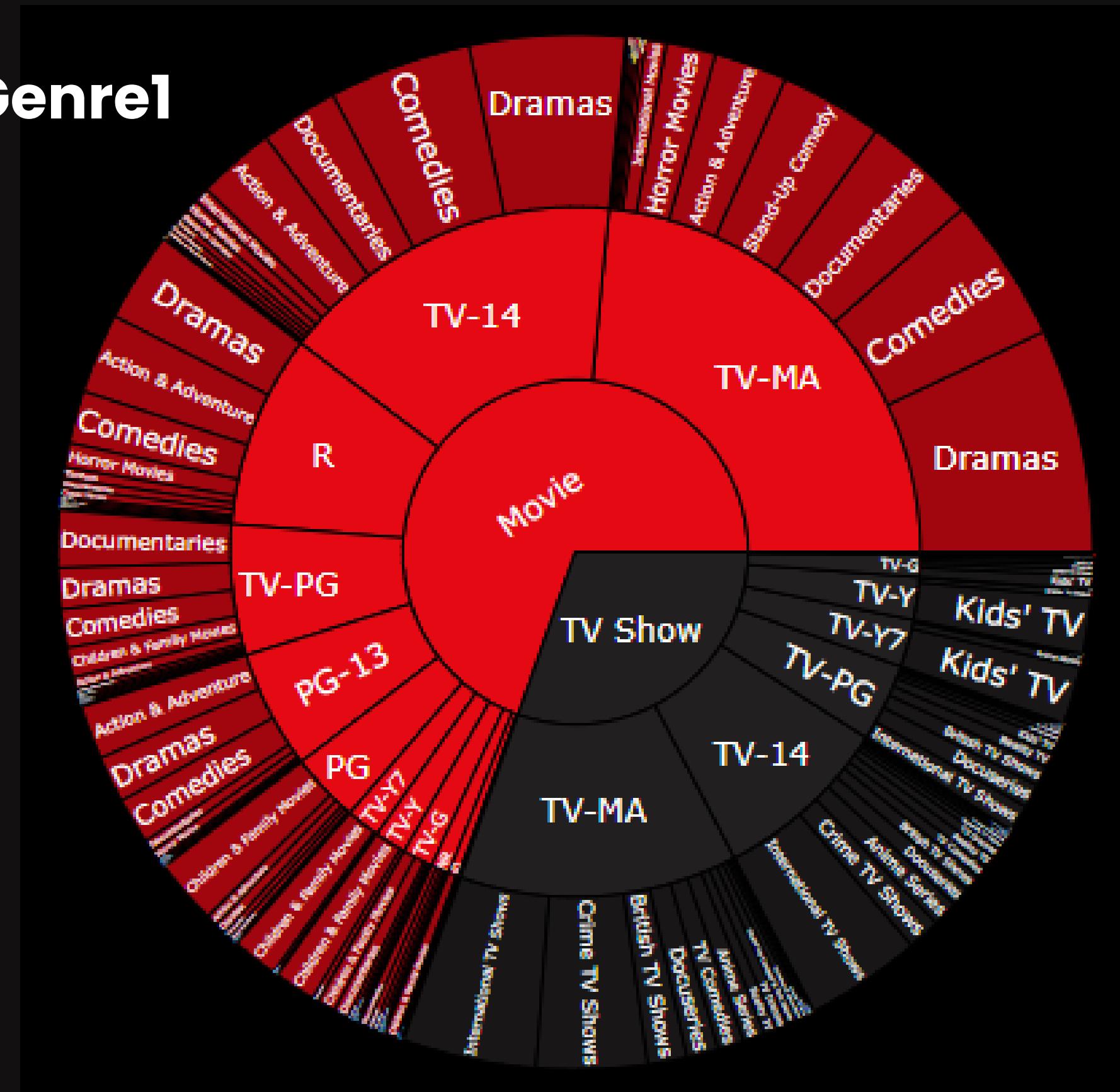
N

GRAPH



Sunburst Chart

Type → Rating → Genre

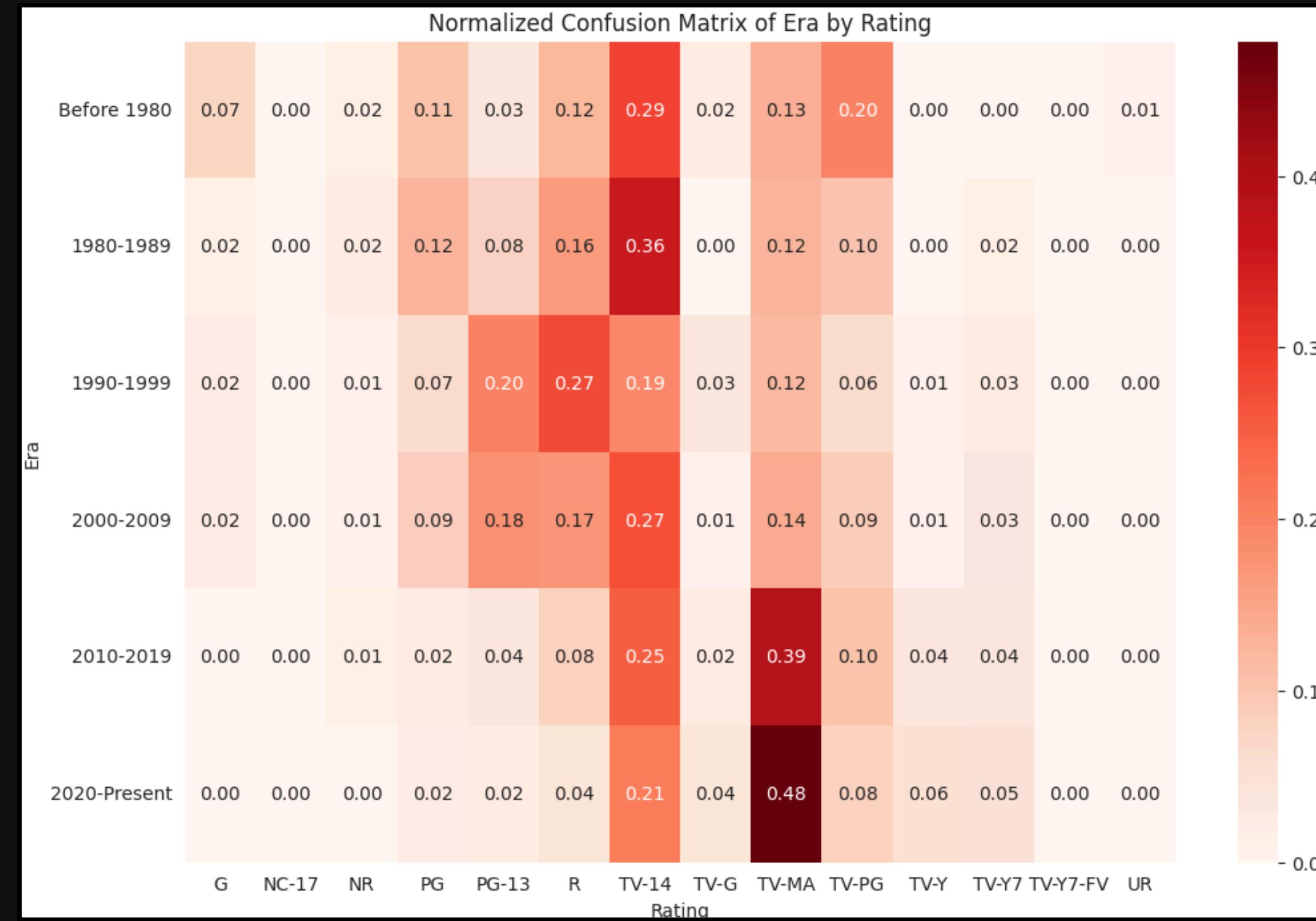


CLICK
HERE

CO

heatmap

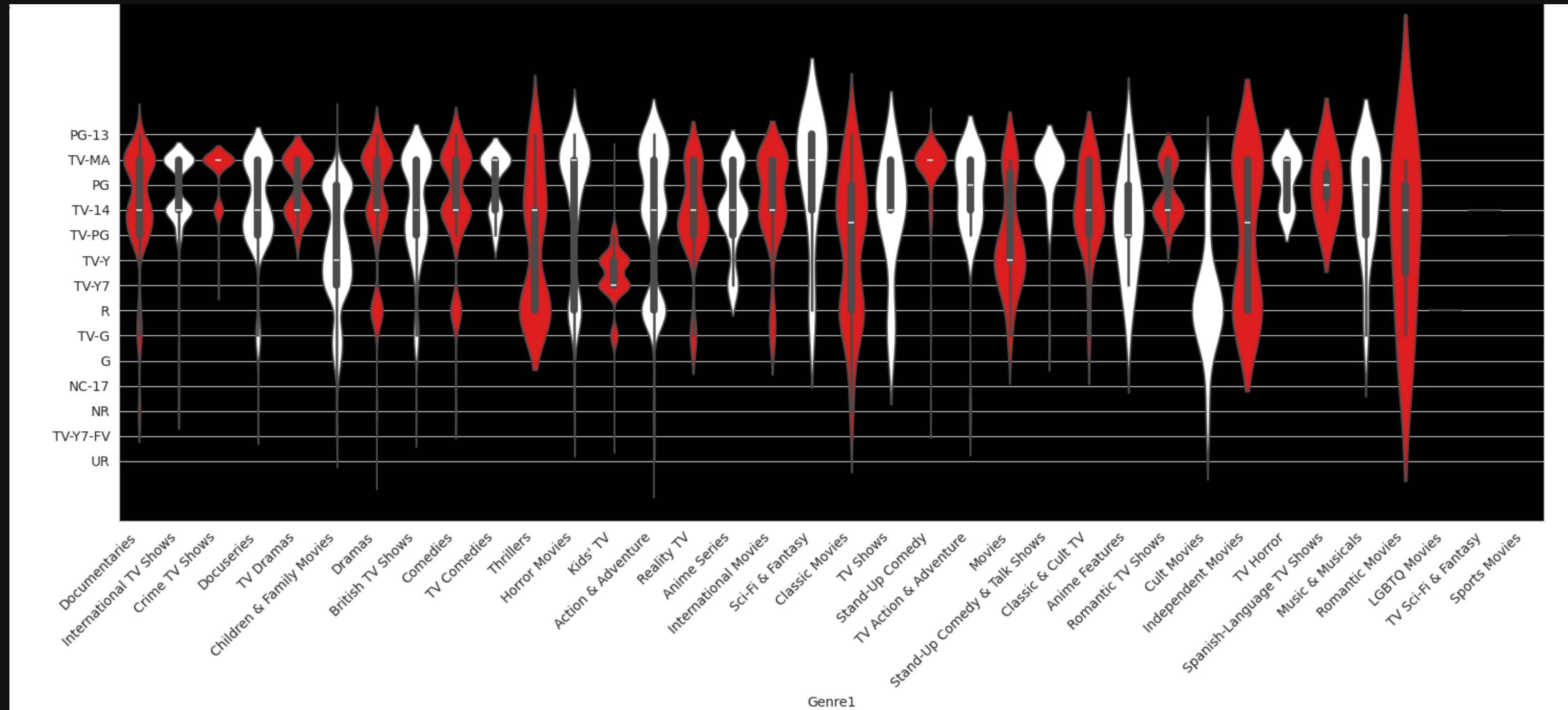
70



Violin Plot

11

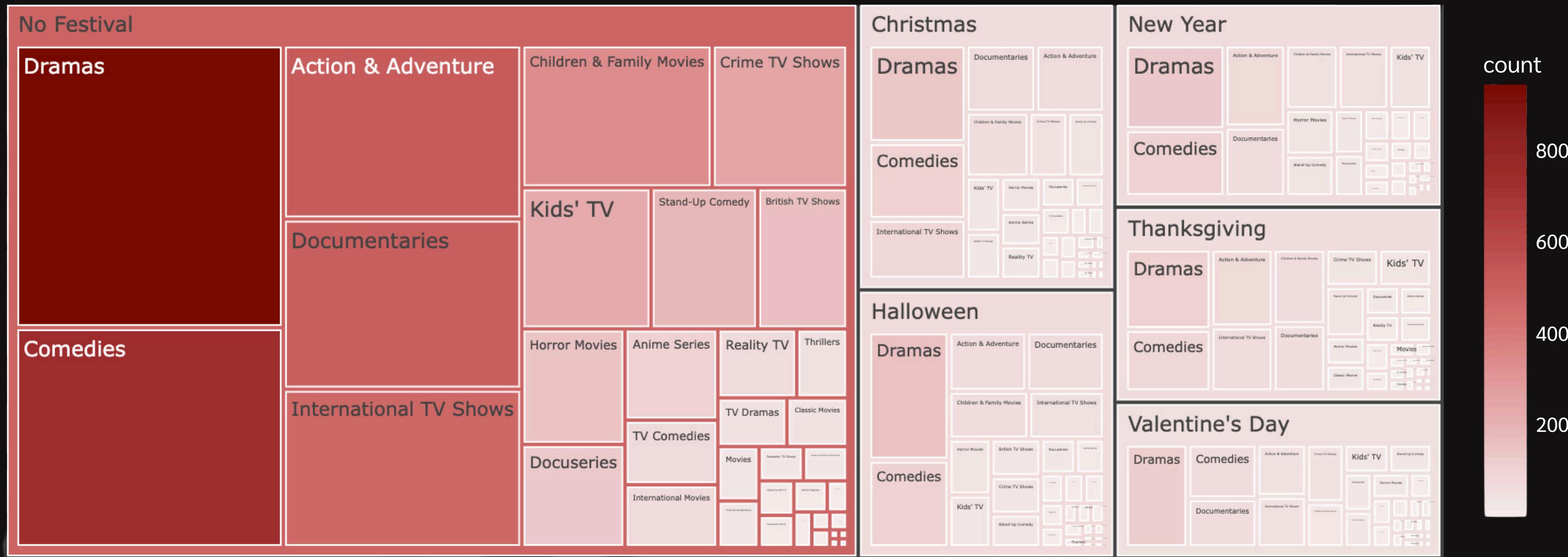
Violin Plot of Ratings by Genre



Treemap

12

Treemap of Genres by Festival



THANK YOU

For Your Attention