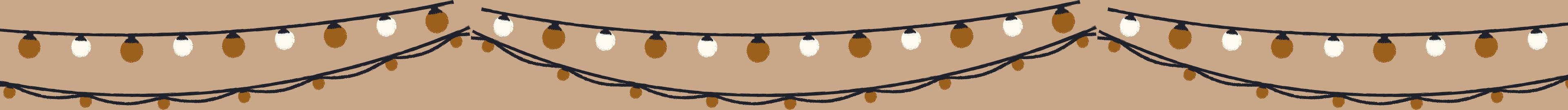




THE TREND IN FEATURE FILMS LENGTH



Hoàng Nam Phong
Group 1 BI21



OUTLINE



Introduction



Data Set



Trend of Length



Trend of Correlation
between Length, Runtime
and Watchers Rating



Conclusion

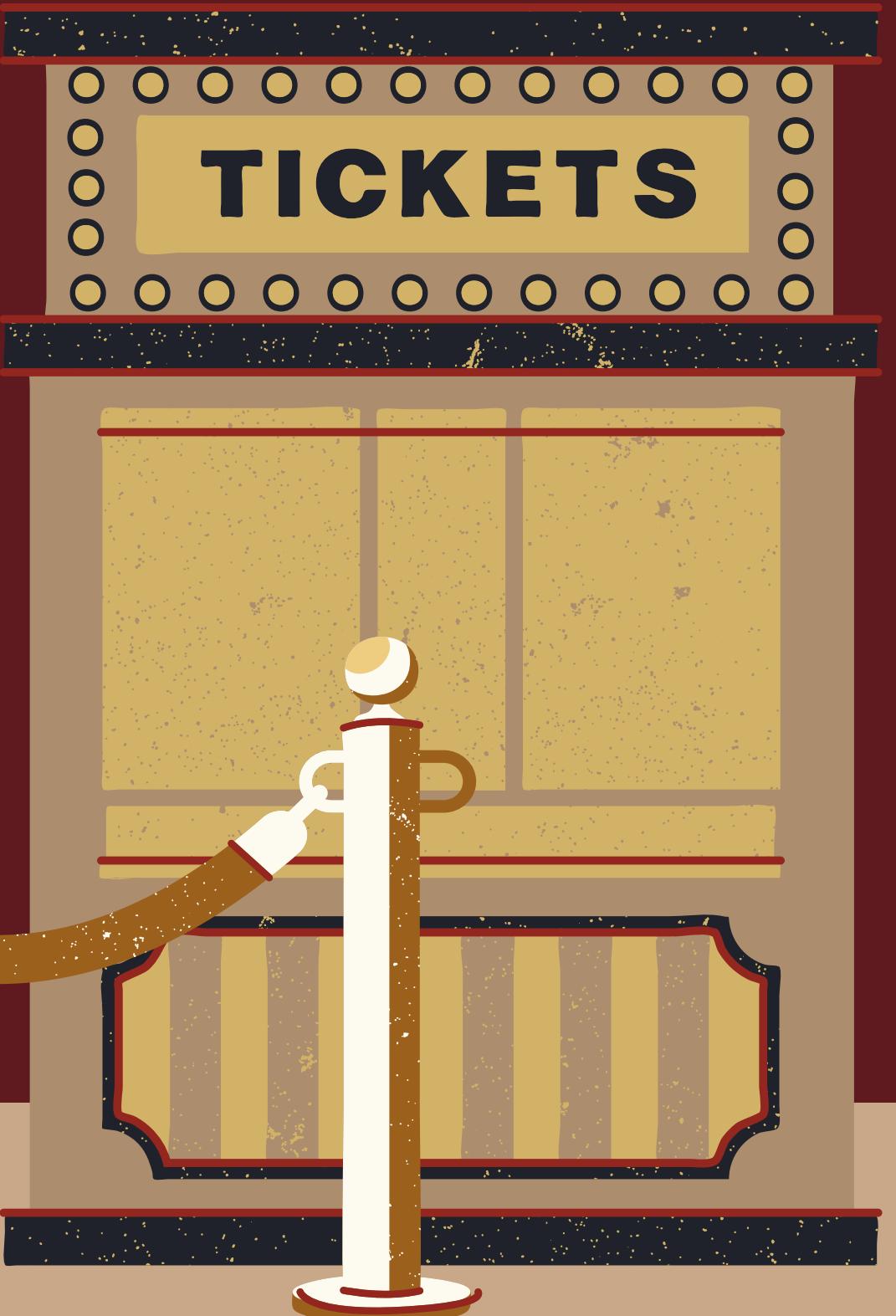
INTRODUCTION

H

- Are Movies Becoming Longer Over Time?
- Has The Proportion of Different Length Type Changed?
- Is There Any Correlation Between Movie Length and Revenue?
- Is There Any Correlation Between Movie Length and Revenue?
- Are the Trends still the Same between different Genres?



THE DATASET



#1.Boxofficemojo

Crawled using Web Scraper
addon. Name, Year, Genres,
Length

#2.Title.Rating

From IMDB database. Movies
codenames, user ratings

#3.Title.Basics

From IMDB database. Titles for
Movies with codenames in
Title.Rating

#4.Inflation Rate

Self-made. Including
inflationary multiplier of each
year.

BOXOFFICEMOJO

- Crawled using Web scraper Addon
- From Boxofficemojo.com
- 40 Tables (1982-2022)



Concat 40 Tables together (1982-2022)

```
#Concate tables

path_dir = "C:/Users/admin/OneDrive/Tài liệu/Basic economy/Python BI/Python BI/Project/csv/"

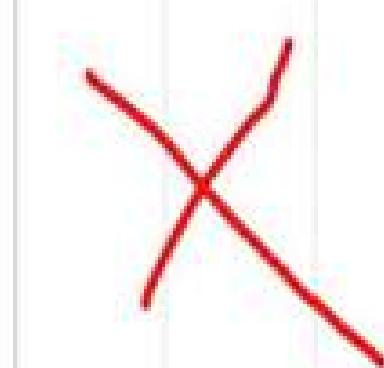
for files in os.listdir(path_dir):
    print(files)

df = pd.DataFrame()
for files in os.listdir(path_dir):
    Dataframe = pd.read_csv(path_dir+files)
    df = pd.concat([df, Dataframe], ignore_index= True)
```

Remove unwanted columns

```
#Drop unwanted column  
  
df = df.drop(['web-scraper-order', 'web-scraper-start-url', 'movie-select', 'movie-select-href', 'title-sum', 'title-sum-href'], axis=1)
```

Extract and reorganized data with wrong positions

name	year	boxoffice	release	mpaa	running-time	genres	pro	
Amelia (2009)	2009	\$19,643,086	Earliest Release Date October 23, 2009	(Domestic) MPAAPG	Running Time 1 hr 51 min	Genres Adventure, Biography, Drama, Romance	IMDbPro See more details at IMDbPro	
Mao's Last Dancer (2009)	2009	\$23,914,731	dàu MPAA phải ở cột mpaa	MPAAPG	Running Time 1 hr 57 min	Genres Biography, Drama, Music, Romance	IMDbPro See more details at IMDbPro null	
I Give My First Love to Yo	-2009	\$24,051,386	IMDbPro	null	null	null	null	

```
#Extracting duration information

time_ok = []

search_time = "Running Time"
for i in range(0,len(df)):
    if str(df['release'][i]).find(search_time) >-1:
        time_ok.append(df['release'][i])
    elif str(df['mpaa'][i]).find(search_time) >-1:
        time_ok.append(df['mpaa'][i])
    elif str(df['running-time'][i]).find(search_time) > -1:
        time_ok.append(df['running-time'][i])
    elif str(df['genres'][i]).find(search_time) > -1:
        time_ok.append(df['genres'][i])
    elif str(df['pro'][i]).find(search_time) > -1:
        time_ok.append(df['pro'][i])
    else:
        time_ok.append(np.NaN)

df["Duration"] = time_ok
df["Duration"] = df["Duration"].str.replace("Running Time", "").str.replace(" ", "")
df
```

Removing Row with NA in important columns

```
#drop na value from Duration  
df = df.dropna(subset=["Duration", "name"])
```

Removing unwanted symbols

```
#Remove symbol from boxoffice  
df["boxoffice"] = df["boxoffice"].str.replace("$","")
```

Convert hour-minute to minute

```
#changing Duration from hour minute form to minute  
df_new['Duration']= df_new['Duration'].str.replace('h', '**60').str.replace('r','+') .str.replace('min','*1')  
  
df_new["Duration"] = df_new["Duration"].apply(eval)
```

Duration	Duration
1hr14min	74
1hr34min	94
1hr36min	96
1hr47min	107
1hr45min	105
...	...
1hr27min	87
2hr6min	126
2hr27min	147
2hr10min	130
3hr12min	192

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 6218 entries, 0 to 6217
```

```
Data columns (total 5 columns):
```

#	Column	Non-Null Count	Dtype
0	name	6218 non-null	object
1	year	6218 non-null	int64
2	genres	6218 non-null	object
3	boxoffice	6218 non-null	float64
4	duration	6218 non-null	int64

```
dtypes: float64(1), int64(2), object(2)
```

```
memory usage: 243.0+ KB
```

		name	year	genres	boxoffice	duration
0		Avatar: The Way of Water	2022	Action, , Adventure, , ...	2.217480e+09	192
1		Top Gun: Maverick	2022	Action, , Drama	1.488730e+09	130
2		Jurassic World: Dominion	2022	Action, , Adventure, , ...	1.001980e+09	147
3		Doctor Strange in the Multiverse of Madness	2022	Action, , Adventure, , ...	9.557760e+08	126
4		Minions: The Rise of Gru	2022	Adventure, , Animation, , ...	9.396280e+08	87
...	
6213		Lookin' to Get Out	1982	Comedy	9.464610e+05	105
6214		The Ballad of Gregorio Cortez	1982	Comedy, , Drama, , Rom...	8.049630e+05	104
6215		Five Days One Summer	1982	Drama	1.990780e+05	96
6216		The Escape Artist	1982	Comedy, , Crime, , Drama	1.433690e+05	94
6217		Bugs Bunny's 3rd Movie: 1001 Rabbit Tales	1982	Adventure, , Animation, , ...	7.835000e+04	74

```
6218 rows × 5 columns
```

TITLE.BASICS & TITLE.RATING

- Official Datasets from IMDB
- Very heavy (over 8 millions lines)
including surplus information



	tconst	titleType	primaryTitle	startYear	genres
0	tt0000001	short	Carmencita	1894	Documentary,Short
1	tt0000002	short	Le clown et ses chiens	1892	Animation,Short
2	tt0000003	short	Pauvre Pierrot	1892	Animation,Comedy,Romance
3	tt0000004	short	Un bon bock	1892	Animation,Short
4	tt0000005	short	Blacksmith Scene	1893	Comedy,Short
...
9629585	tt9916848	tvEpisode	Episode #3.17	2010	Action,Drama,Family
9629586	tt9916850	tvEpisode	Episode #3.19	2010	Action,Drama,Family
9629587	tt9916852	tvEpisode	Episode #3.20	2010	Action,Drama,Family
9629588	tt9916856	short	The Wind	2015	Short
9629589	tt9916880	tvEpisode	Horrid Henry Knows It All	2014	Adventure,Animation,Comedy

8327117 rows × 5 columns



	primaryTitle
0	Vojtech, receny sirotek
1	The Ghost Dance
2	Le jardinier
3	Quiero soñar
4	Richard III
...	...
379174	An Actress' Dream
379175	One Love 2 Story
379176	Gaja
379177	The Battle Within
379178	Female Teacher: Chain and Bondage

Join 3 tables using SQL

name	year	genres	boxoffice	duration	averageRating
Avatar: The Way of Water	2022	Action, Adventure, Fantasy, Sci-Fi	2,217,479,936	192	7.8
Top Gun: Maverick	2022	Action, Drama	1,488,729,984	130	8.3
Jurassic World: Dominion	2022	Action, Adventure, Sci-Fi, Thriller	1,001,980,032	147	5.6
Doctor Strange in the Multiverse of Madness	2022	Action, Adventure, Fantasy, Horror, Sci-Fi	955,776,000	126	6.9
Minions: The Rise of Gru	2022	Adventure, Animation, Comedy, Crime, Family, Sci-Fi	939,628,032	87	6.6
Black Panther: Wakanda Forever	2022	Action, Adventure, Drama, Fantasy, Sci-Fi, Thriller	855,107,968	161	6.8
The Batman	2022	Action, Crime, Drama, Thriller	770,945,984	176	7.8
Thor: Love and Thunder	2022	Action, Adventure, Comedy, Fantasy, Romance, Sci-Fi	760,928,000	118	6.3
Water Gate Bridge	2022	Action, Drama, History	626,572,032	153	5.5
Fantastic Beasts: The Secrets of Dumbledore	2022	Adventure, Family, Fantasy	406,951,008	142	6.2
Sonic the Hedgehog 2	2022	Action, Adventure, Comedy, Family, Fantasy, Sci-Fi	402,656,992	122	6.5
Uncharted	2022	Action, Adventure	401,748,992	116	2.1
Puss in Boots: The Last Wish	2022	Adventure, Animation, Comedy, Family, Fantasy, Mystery	401,488,000	102	7.9
Black Adam	2022	Action, Adventure, Fantasy, Sci-Fi	392,952,000	125	6.4

Inflationary Multiplier Table

year	inflation_multiplier
1982	3.03
1983	2.94
1984	2.82
1985	2.72
1986	2.67
1987	2.58
1988	2.47
1989	2.36
1990	2.24
1991	2.15
1992	2.09
1993	2.03
1994	1.97

Creating Length Sorting Column

```
#select name, year, genres, boxoffice, inflation_multiplier, AdjustedRev, duration, CASE
#WHEN duration < 90 THEN 'Short'
#WHEN duration > 120 then 'Long'
#ELSE 'Medium'
#end as length_rate
```

Complete Dataset

	name	year	genres	boxoffice	inflation	AdjustedRev	averageRating	duration	length_rate
0	Avatar: The Way of Water	2022	Action, ,	2,217,479,936	1	2,217,479,936	7.8	192	Long
1	Top Gun: Maverick	2022	Action, ,	1,488,729,984	1	1,488,729,984	8.3	130	Long
2	Jurassic World: Dominion	2022	Action, ,	1,001,980,032	1	1,001,980,032	5.6	147	Long
3	Doctor Strange in the Multi	2022	Action, ,	955,776,000	1	955,776,000	6.9	126	Long
4	Minions: The Rise of Gru	2022	Adventure,	939,628,032	1	939,628,032	6.6	87	Short
5	Black Panther: Wakanda Fo	2022	Action, ,	855,107,968	1	855,107,968	6.8	161	Long
6	The Batman	2022	Action, ,	770,945,984	1	770,945,984	7.8	176	Long
7	Thor: Love and Thunder	2022	Action, ,	760,928,000	1	760,928,000	6.3	118	Medium
8	Water Gate Bridge	2022	Action, ,	626,572,032	1	626,572,032	5.5	153	Long
9	Fantastic Beasts: The Secret	2022	Adventure,	406,951,008	1	406,951,008	6.2	142	Long
10	Sonic the Hedgehog 2	2022	Action, ,	402,656,992	1	402,656,992	6.5	122	Long
11	Uncharted	2022	Action, ,	401,748,992	1	401,748,992	2.1	116	Medium
12	Puss in Boots: The Last Wish	2022	Adventure,	401,488,000	1	401,488,000	7.9	102	Medium
13	Black Adam	2022	Action, ,	392,952,000	1	392,952,000	6.4	125	Long
14	Elvis	2022	Biography,	287,340,000	1	287,340,000	7.4	159	Long

DESCRIPTIVE STATISTICS

- Average runtime 108 minutes(medium length), average rating 6.28/10, average adjusted revenue \$138mil.
- All numeric data are right-skewed except for rating.

pj.describe()					
✓ 0.4s	Column1	year	AdjustedRev	averageRating	duration
count	6218.000000	6218.000000	6.218000e+03	5845.000000	6218.000000
mean	3114.494693	2000.494693	1.382015e+08	6.285133	108.020103
std	1796.618671	11.016623	2.424857e+08	1.065222	19.358377
min	0.000000	1982.000000	6.411000e+03	1.600000	38.000000
25%	1561.250000	1991.000000	2.064220e+07	5.600000	95.000000
50%	3115.500000	2000.000000	5.263415e+07	6.400000	105.000000
75%	4669.750000	2009.000000	1.447073e+08	7.000000	117.000000
max	6224.000000	2022.000000	4.049540e+09	9.400000	288.000000

CORRELATION

- Positive, but not tight correlation between Duration, Revenue and Rating.

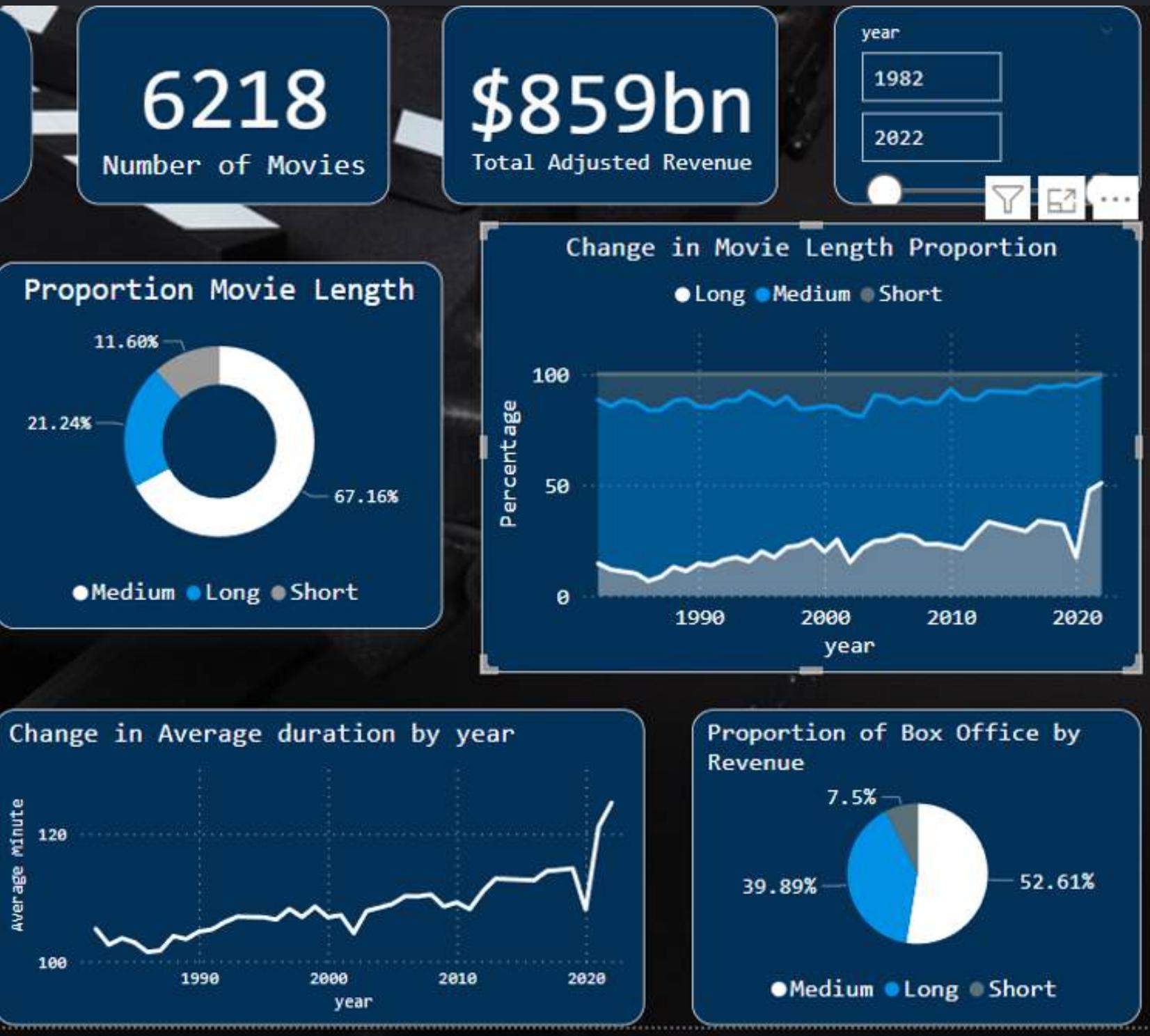


THE TEND IN MOVIES LENGTH



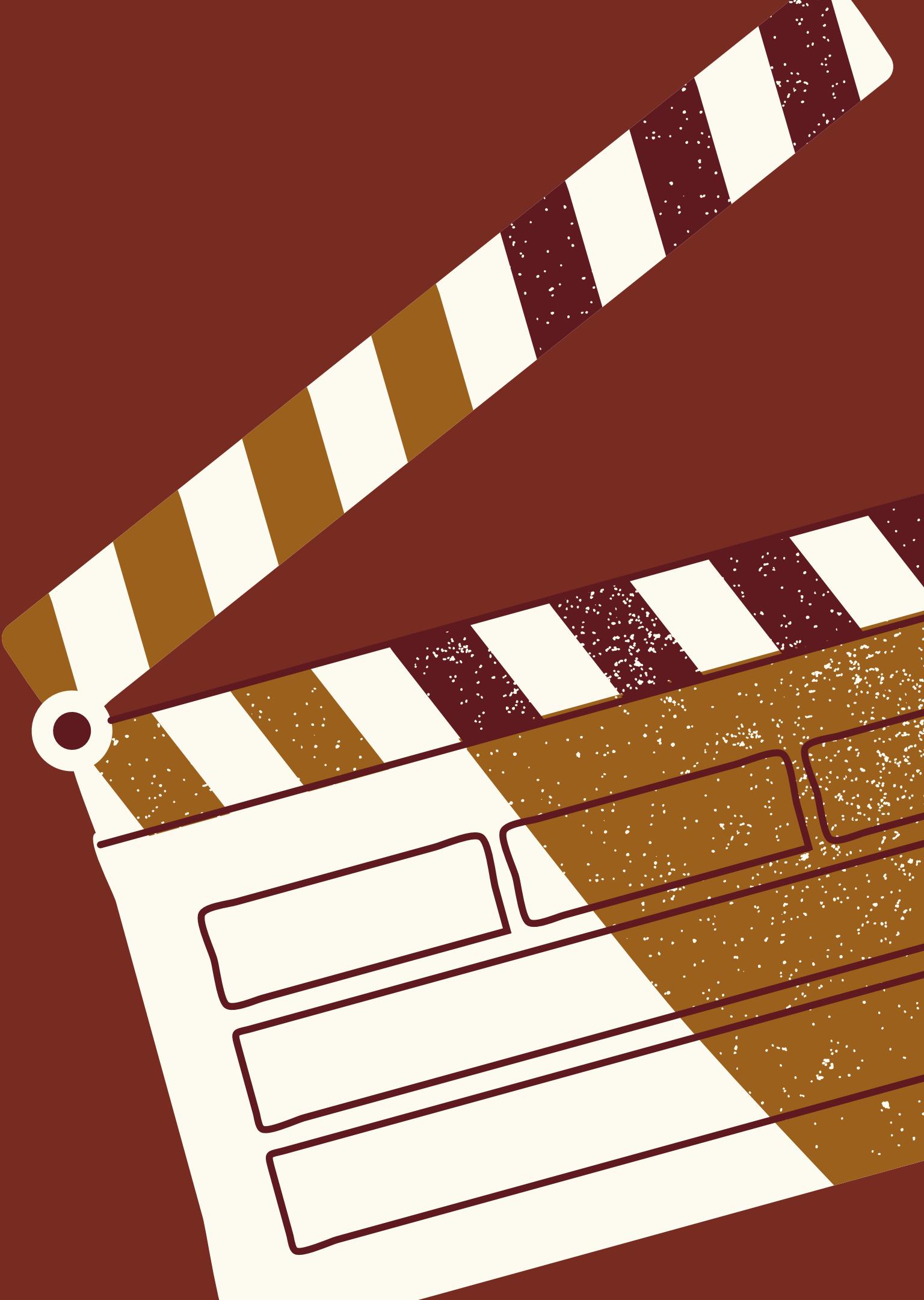
THE TREND IN MOVIES LENGTH

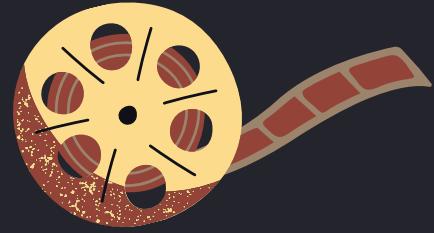
- Clear upward trend in average runtime of movies over 40 years
- Medium length dominant, upward trend in proportion of > 2 hours movie, short movies almost disappear
- Steep drop in 2020 followed by booms in 2021-2022 (due to Covid-19)



EXPLANATION

- Technical Advancement of the Industry:
 - Digital Camera
 - Disappearance of VHS tapes
 - Growth of Movie Theater industry
- Moviemakers are allowed to make longer movies that suit their ambitions and visions.

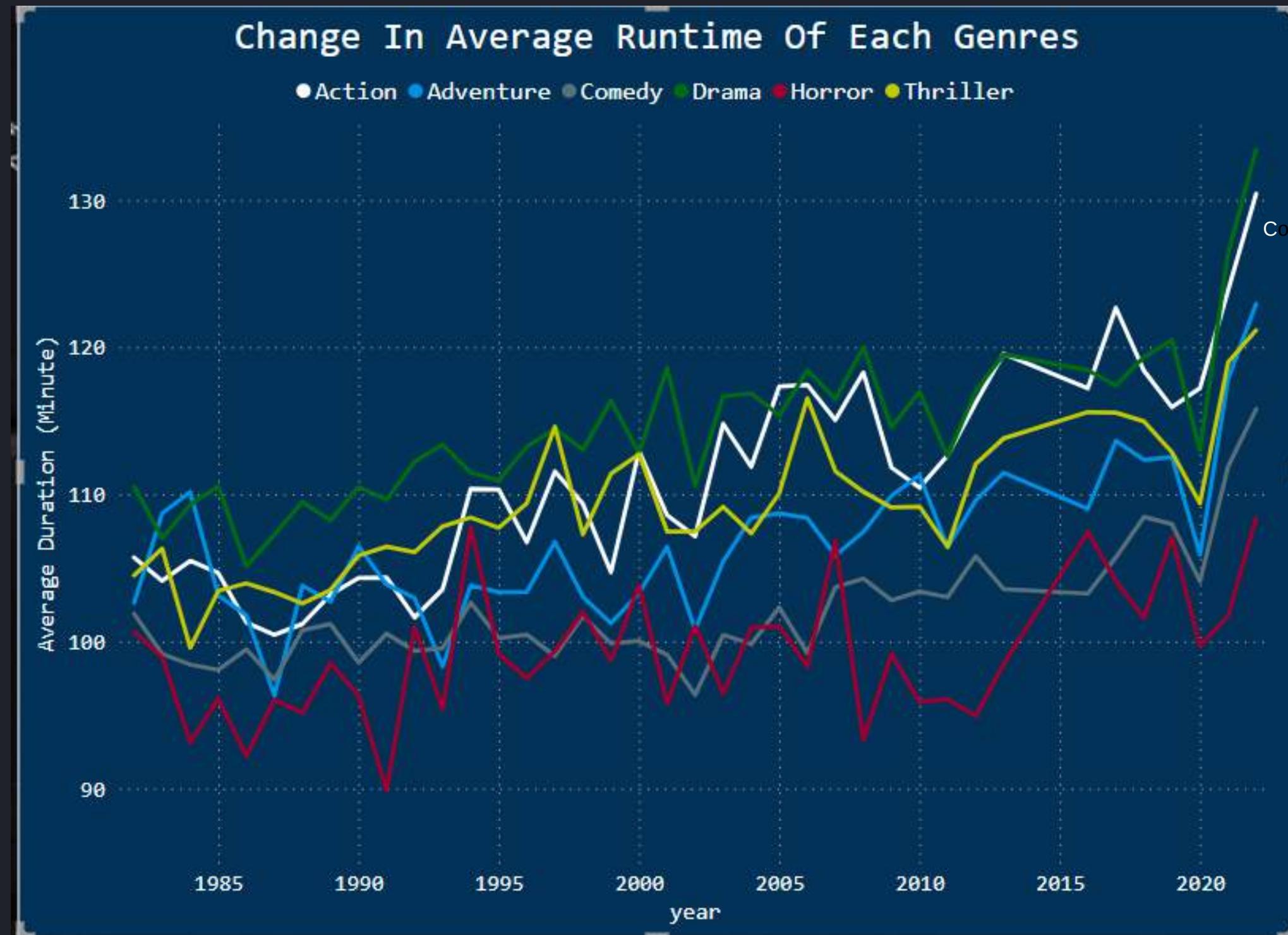




“The two-hour film has had a great run for about 100 years but it’s become a very difficult format to work in it”

-----Joe Russo-----

COMPARISON



- Most genres have the same trends
- Horror and Comedy having the least change. Horror only fluctuate within 90-110min range

COMPARISON



- Most genres have the same trends
- Horror has not changed much.
Reason:
 - Lower budget, often chosen by independent or small moviemakers
 - Basic and often cliché plot & characters
 - Hard to maintain tension effectively with longer movies

CORRELATION BETWEEN RUNTIME, REVENUE AND RATING

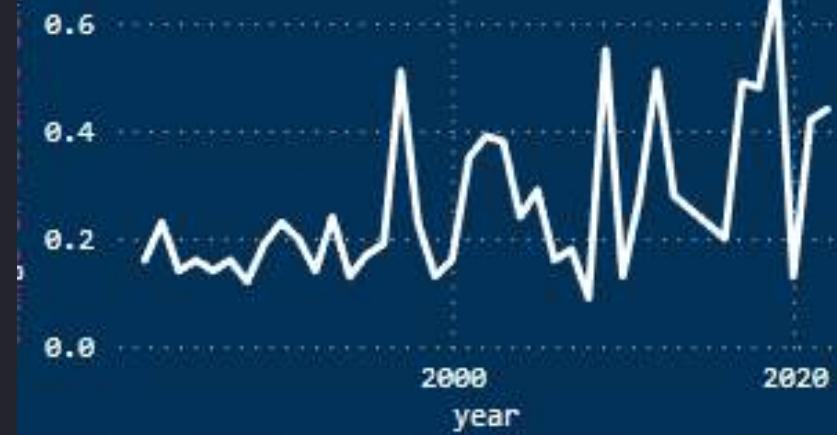


CORRELATION MOVIE LENGTH

0.26

Revenue-Length Correlation

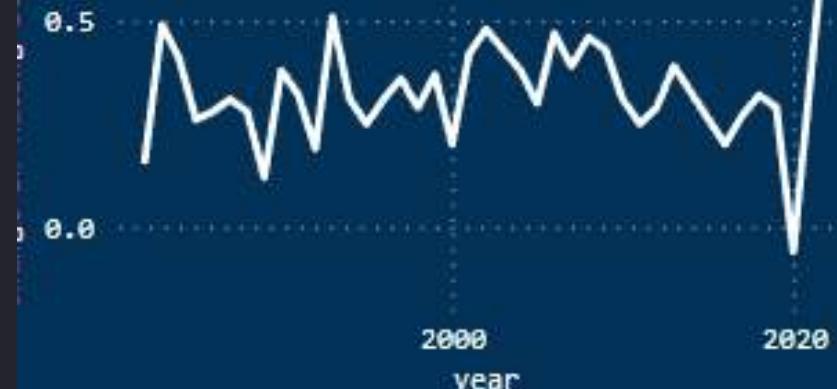
Change in Revenue-Length correlation by year



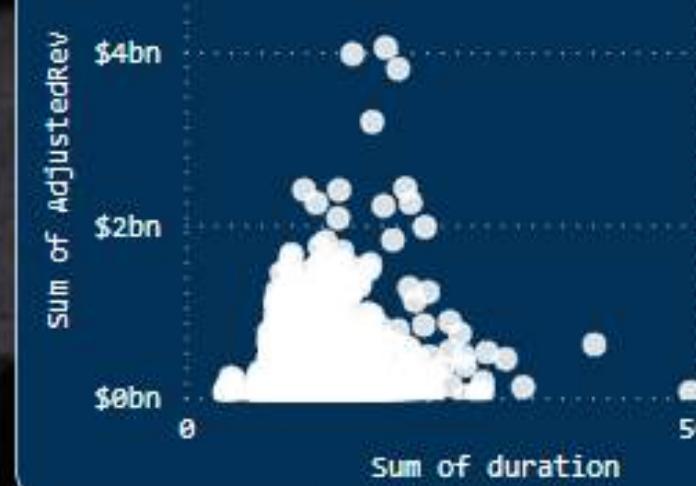
0.33

Rating-Length Correlation

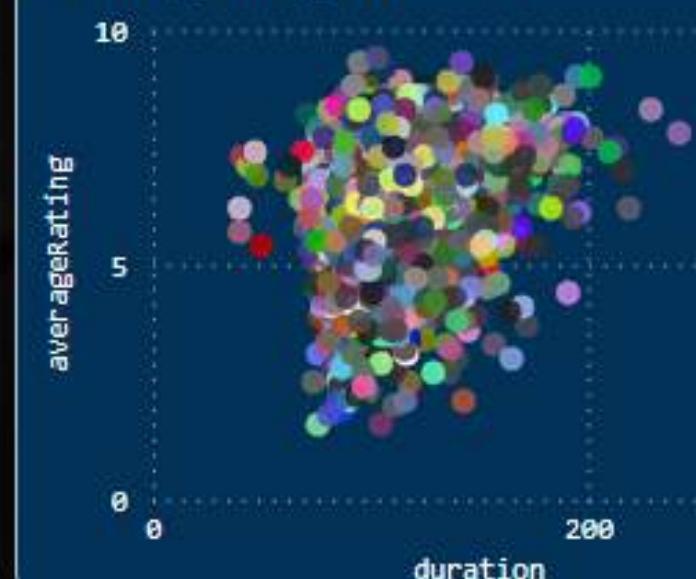
Change in Rating-Length correlation by year



Revenue-Length Correl



Rating-Length Corel

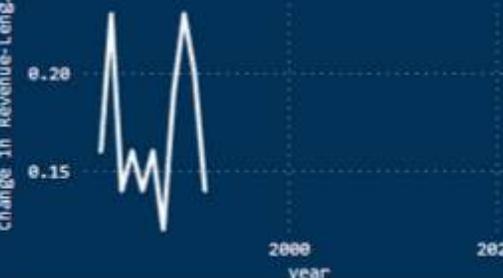


- Duration has positive, but not tight link to both
- All films, considered, Length has higher correlation with user rating
- Revenue-Lenght has grown while Rating-Length stable over 40 years

0.17

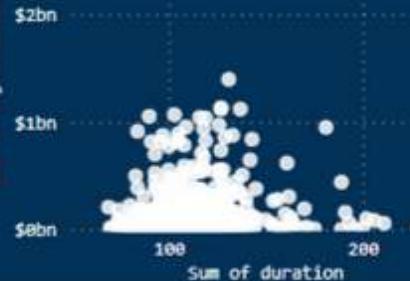
Revenue-Length Correlation

Change in Revenue-Length Correlation by year

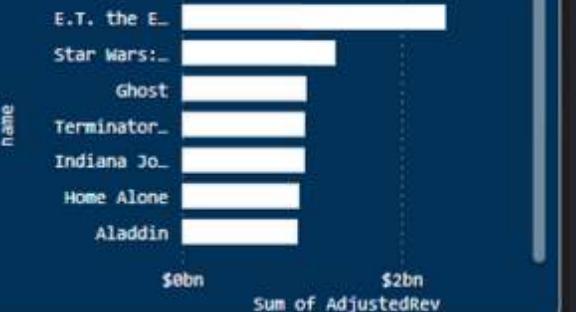
**0.29**

Rating-Length Correlation

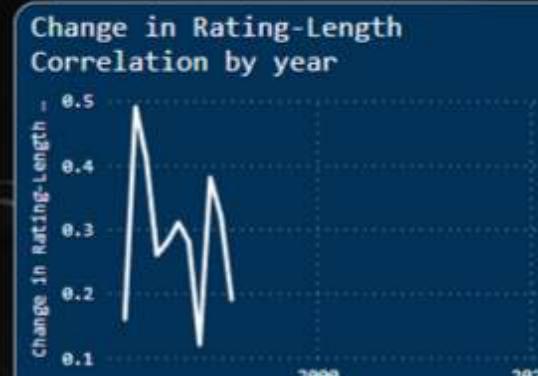
Revenue-Length Correl



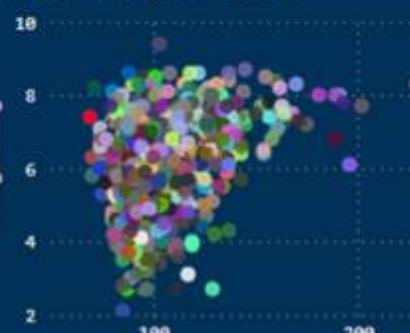
Sum of AdjustedRev by name



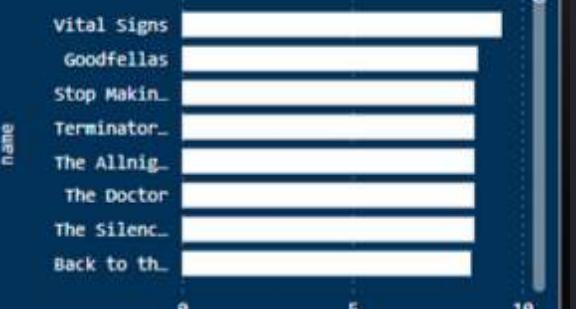
Change in Rating-Length Correlation by year



Rating-Length Corel

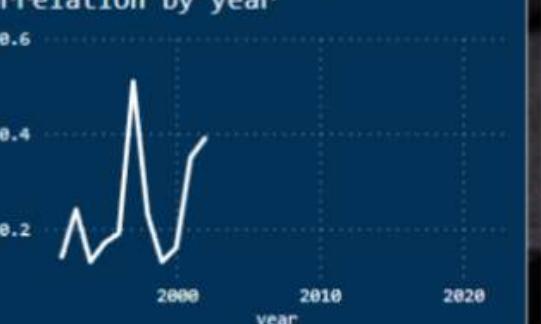


Average of averageRating by name

**0.24**

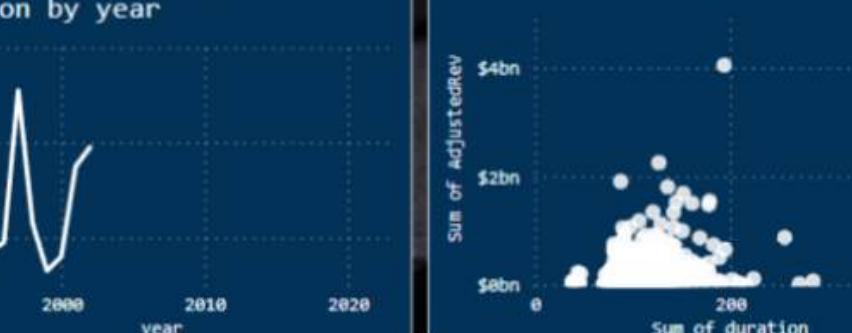
Revenue-Length Correlation

Change in Revenue-Length Correlation by year

**0.34**

Rating-Length Correlation

Revenue-Length Correl

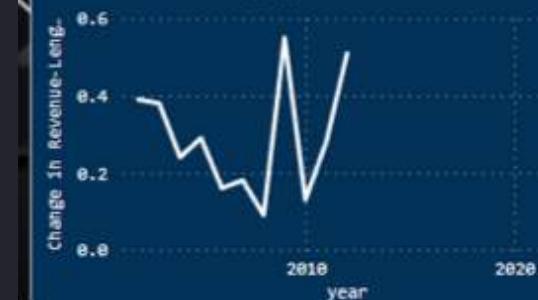


Sum of AdjustedRev by name

**0.29**

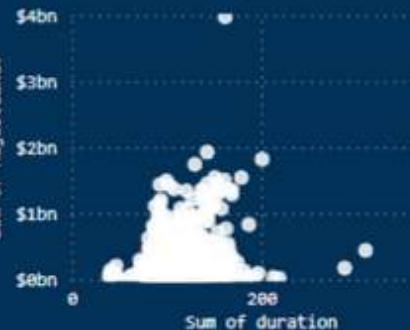
Revenue-Length Correlation

Change in Revenue-Length Correlation by year

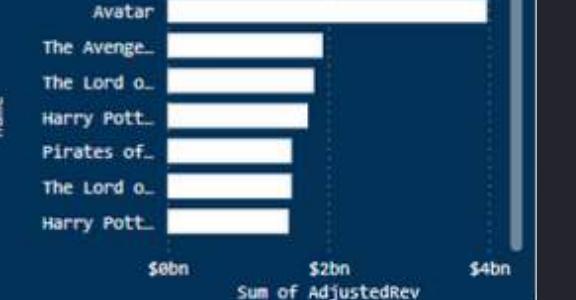
**0.38**

Rating-Length Correlation

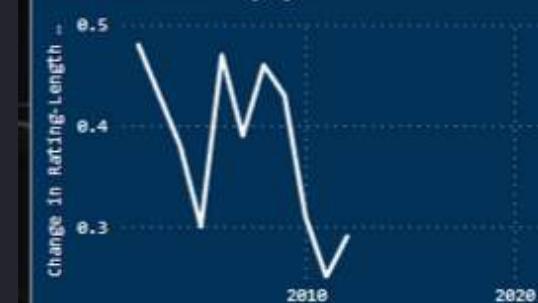
Revenue-Length Correl



Sum of AdjustedRev by name



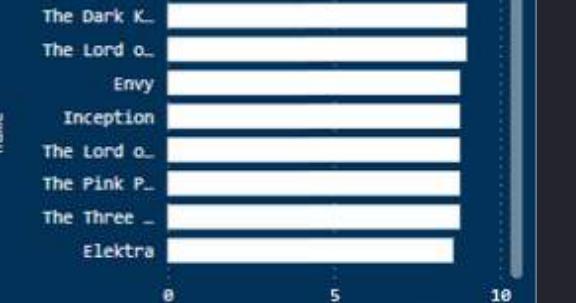
Change in Rating-Length Correlation by year



Rating-Length Corel

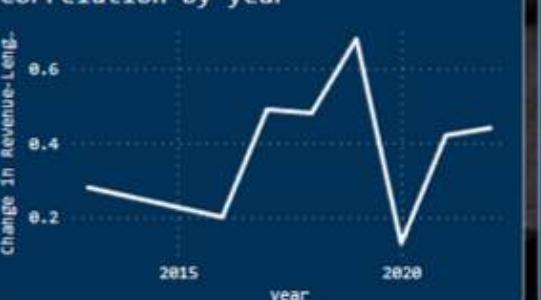


Average of averageRating by name

**0.39**

Revenue-Length Correlation

Change in Revenue-Length Correlation by year

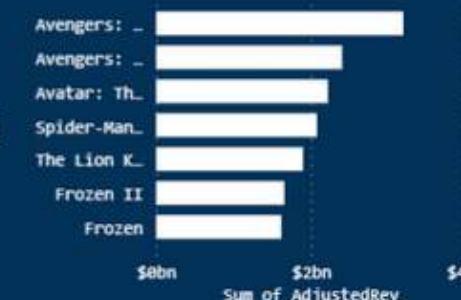
**0.32**

Rating-Length Correlation

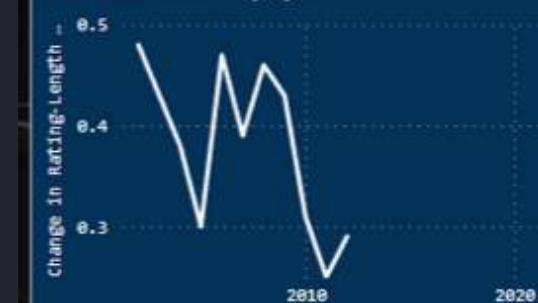
Revenue-Length Correl



Sum of AdjustedRev by name



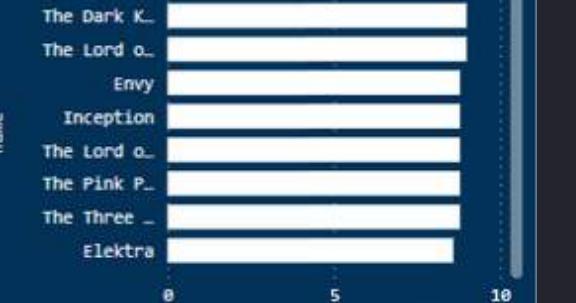
Change in Rating-Length Correlation by year



Rating-Length Corel



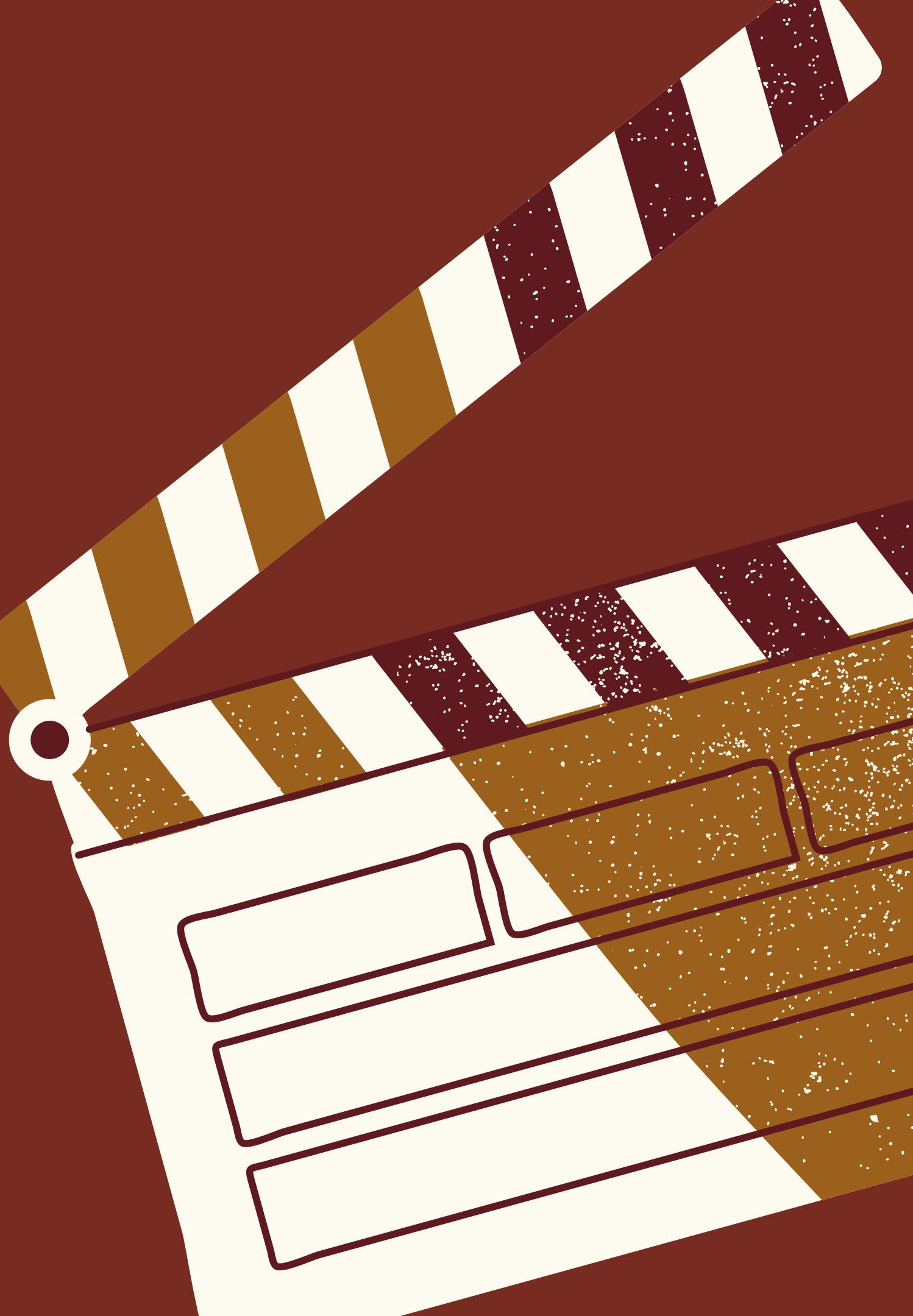
Average of averageRating by name



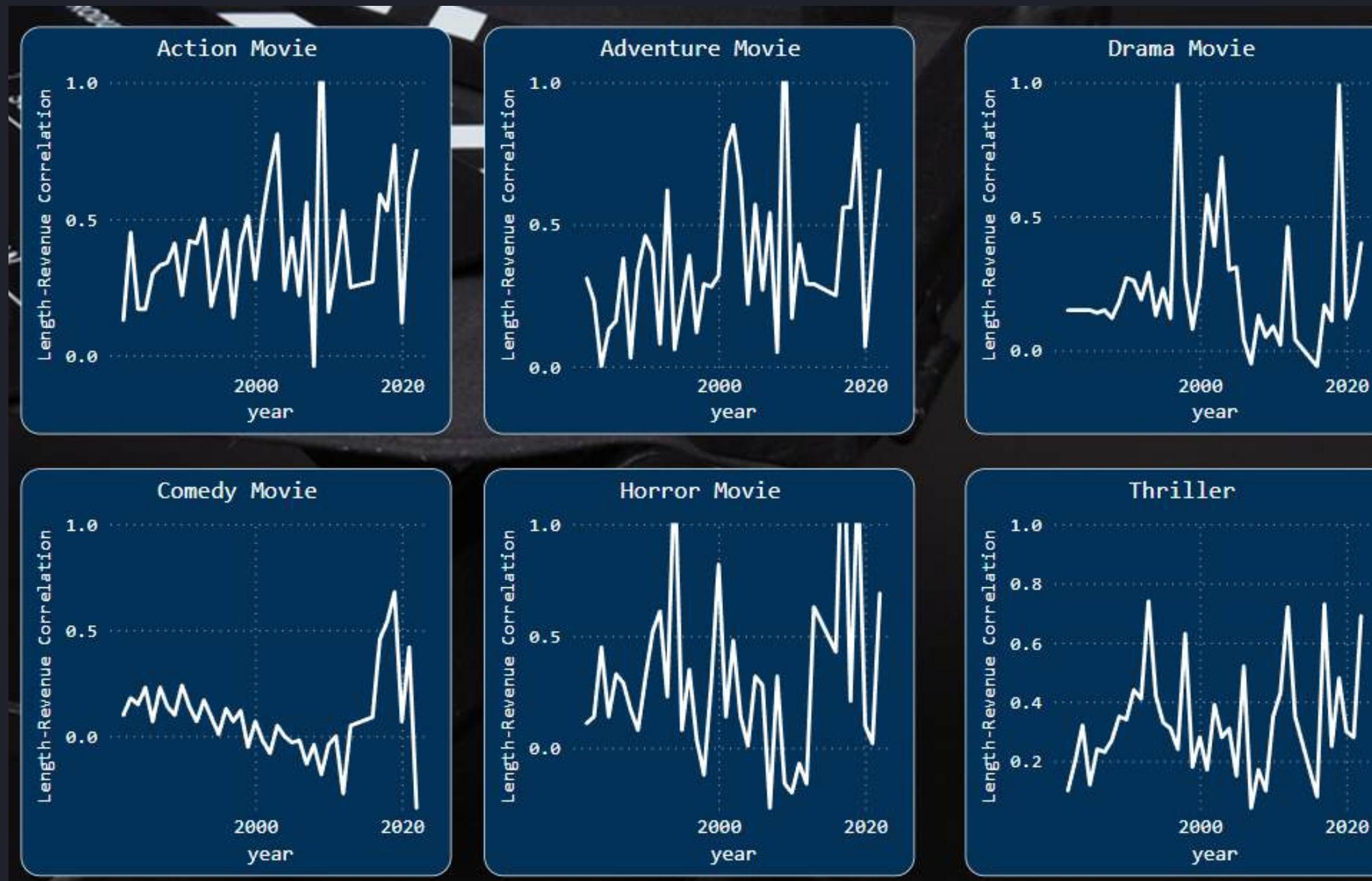
Rank	Title	Worldwide Lifetime Gross	Domestic Lifetime Gross	Domestic %	Foreign Lifetime Gross	Foreign %	Year
1	Avatar	\$2,923,706,026	\$785,221,649	26.9%	\$2,138,484,377	73.1%	2009
2	Avengers: Endgame	\$2,799,439,100	\$858,373,000	30.7%	\$1,941,066,100	69.3%	2019
3	Avatar: The Way of Water	\$2,267,606,306	\$665,544,119	29.4%	\$1,602,062,187	70.6%	2022
4	Titanic	\$2,254,128,488	\$673,853,032	29.9%	\$1,580,275,456	70.1%	1997
5	Star Wars: Episode VII - The Force Awakens	\$2,071,310,218	\$936,662,225	45.2%	\$1,134,647,993	54.8%	2015
6	Avengers: Infinity War	\$2,052,415,039	\$678,815,482	33.1%	\$1,373,599,557	66.9%	2018
7	Spider-Man: No Way Home	\$1,921,847,111	\$814,115,070	42.4%	\$1,107,732,041	57.6%	2021
8	Jurassic World	\$1,671,537,444	\$653,406,625	39.1%	\$1,018,130,819	60.9%	2015
9	The Lion King	\$1,663,075,401	\$543,638,043	32.7%	\$1,119,437,358	67.3%	2019
10	The Avengers	\$1,520,538,536	\$623,357,910	41%	\$897,180,626	59%	2012

EXPLANATION

- Longer Runtime allows movies makers to tell stories up to their ambitions and visions
- Better scripts or prestigious directors more likely to be greenlighted to make longer movies



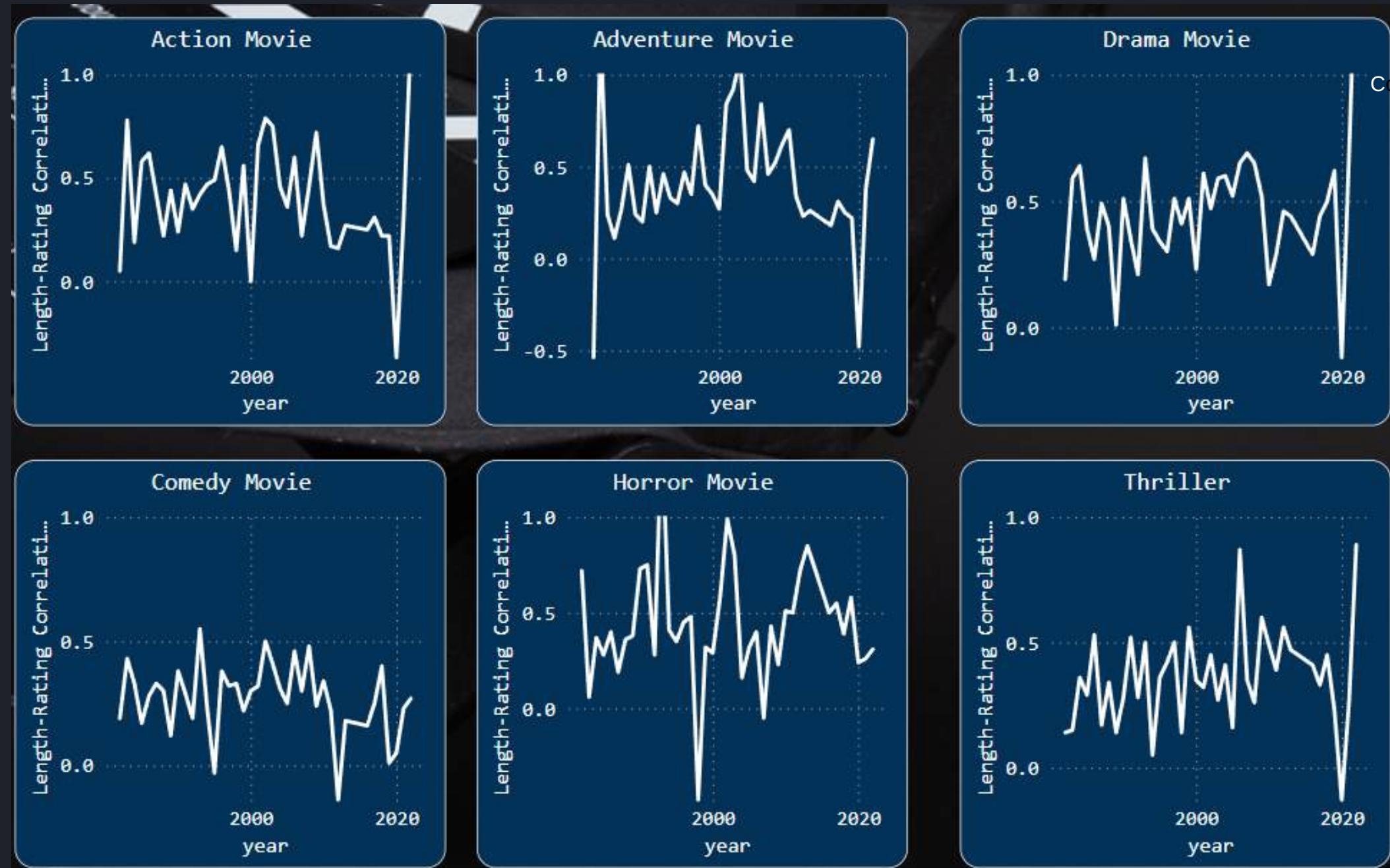
COMPARISON(REVENUE)



Come

- Most genres have the same trends. Fluctuated but overall upward
- Comedy has downward trend. Reason:
 - (similar to horror) harder to maintain intended emotion for audience for longer films -->lower rating-->lower revenue

COMPARISON(RATING)



- Both comedy and horror have a more downward, horizontal-looking trends.

CONCLUSION

- There Is An Upward Trend In The Average Length of Movies over 40 Years
- Upward Trend in Proportion of Longer Movies
- Positive, But Not Tight Link Between Duration and Revenue and Rating
- Trends are Overall similar Across different Genres, except Horror and Comedy



The Effect of Movies Length To Their Performance Over Time



6218

Number of Movies

\$859bn

Total Adjusted Revenue

Navigation

Length

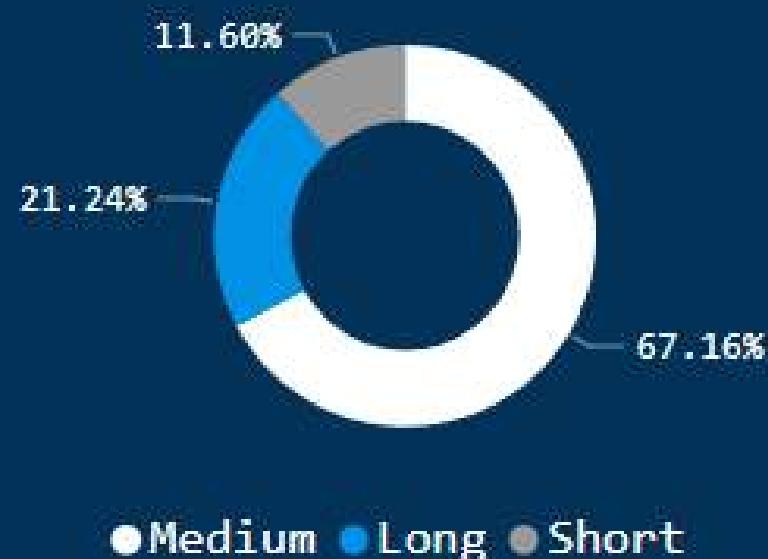
Correlation

Comparison 1

6.29

Average IMDB Rating

Proportion Movie Length



Change in Movie Length Proportion

● Long ● Medium ● Short

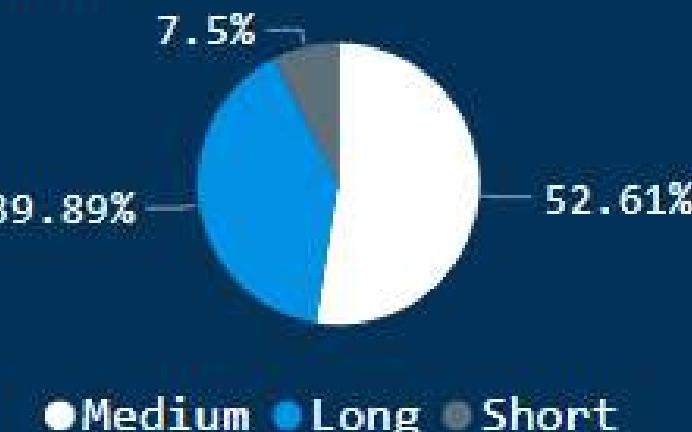


name	Duration	Adjusted Revenue
*batteries not included	106	\$167,928,991
10 Cloverfield Lane	103	\$134,464,991
10 Days with Dad	104	\$10,703,400
10 Things I Hate	97	\$94,124,000
Total	6718	\$359,337,140,582

Change in Average duration by year



Proportion of Box Office by Revenue



0.26

Revenue-Length Correlation

0.33

Rating-Length Correlation

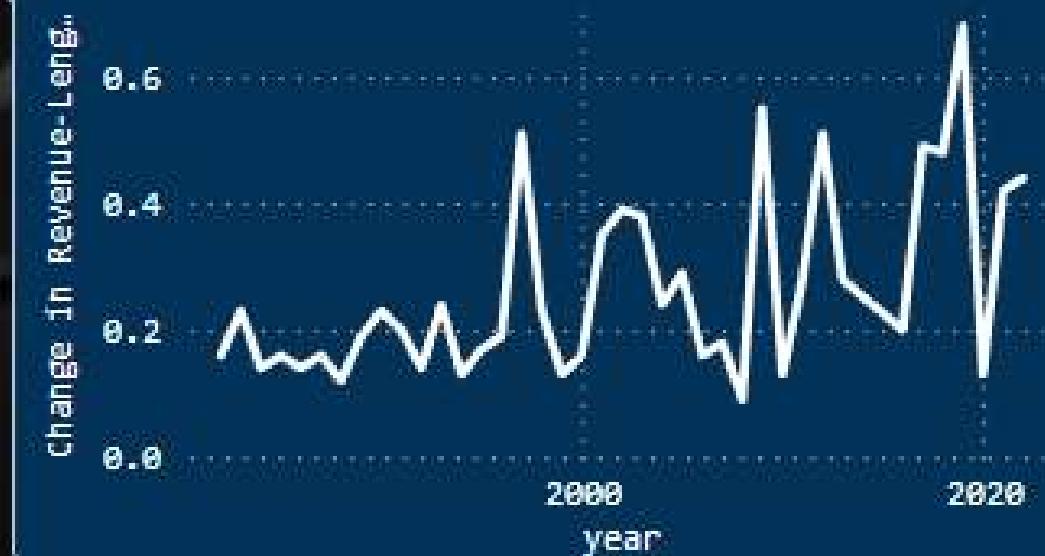
Navigation

Length

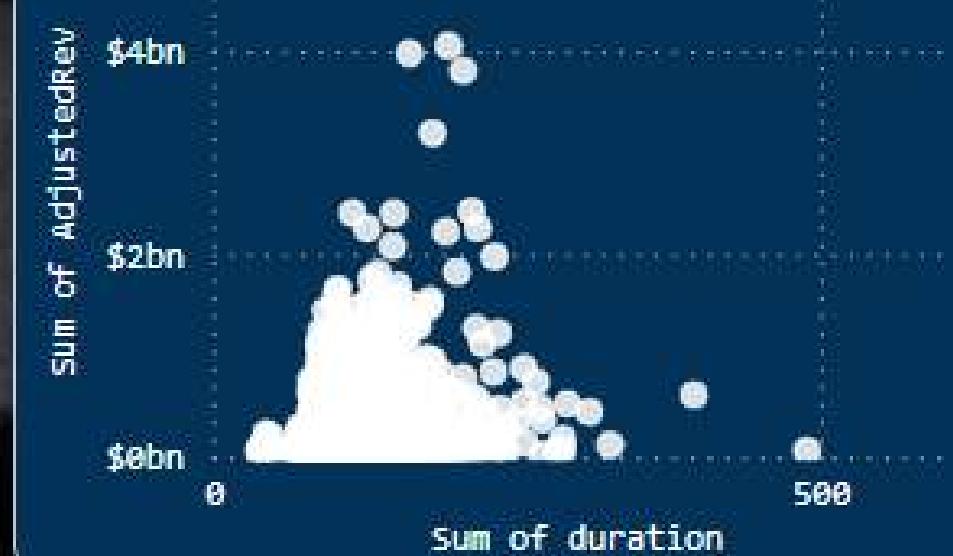
Correlation

Comparison 1

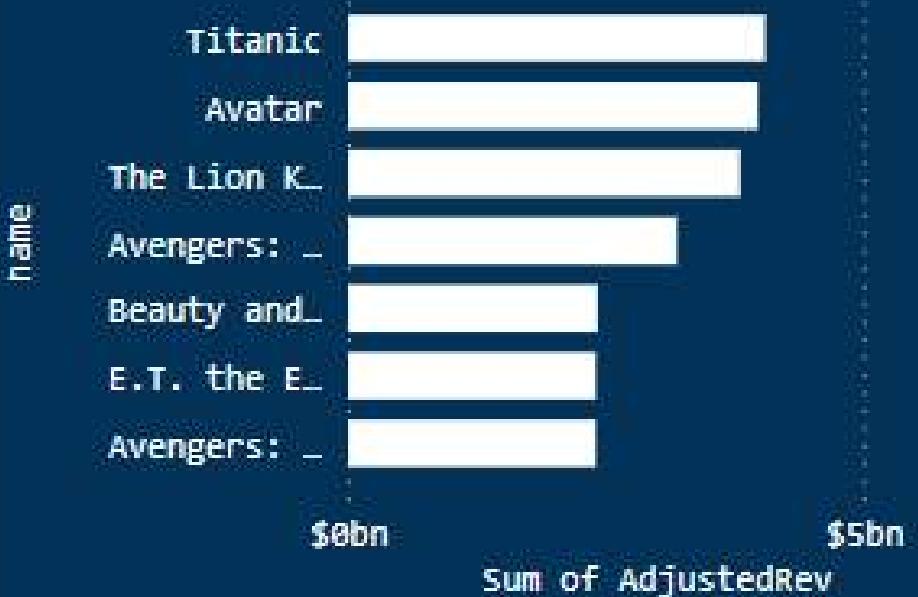
Change in Revenue-Length
Correlation by year



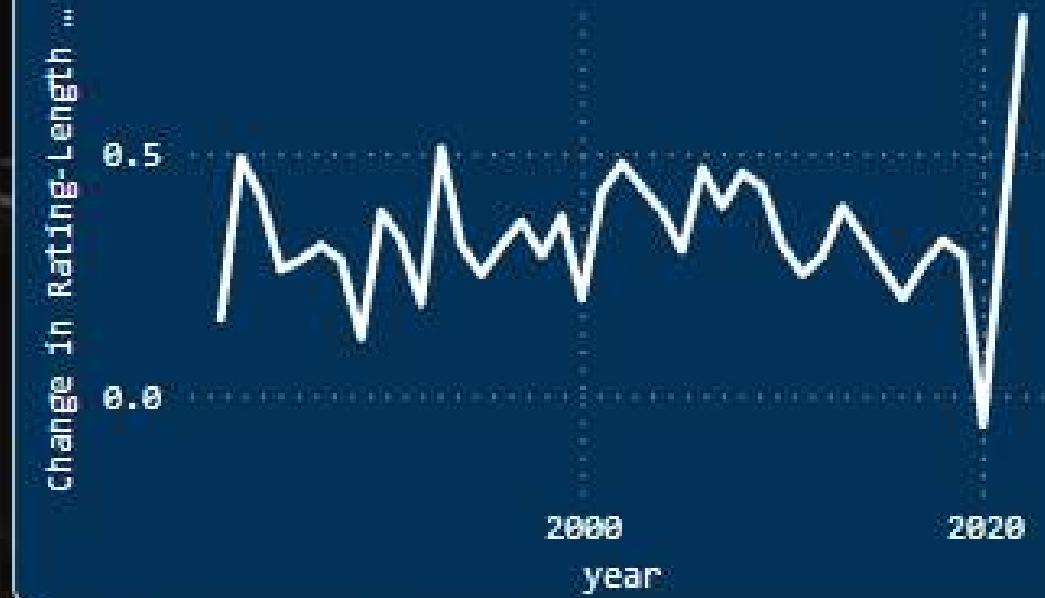
Revenue-Length Correl



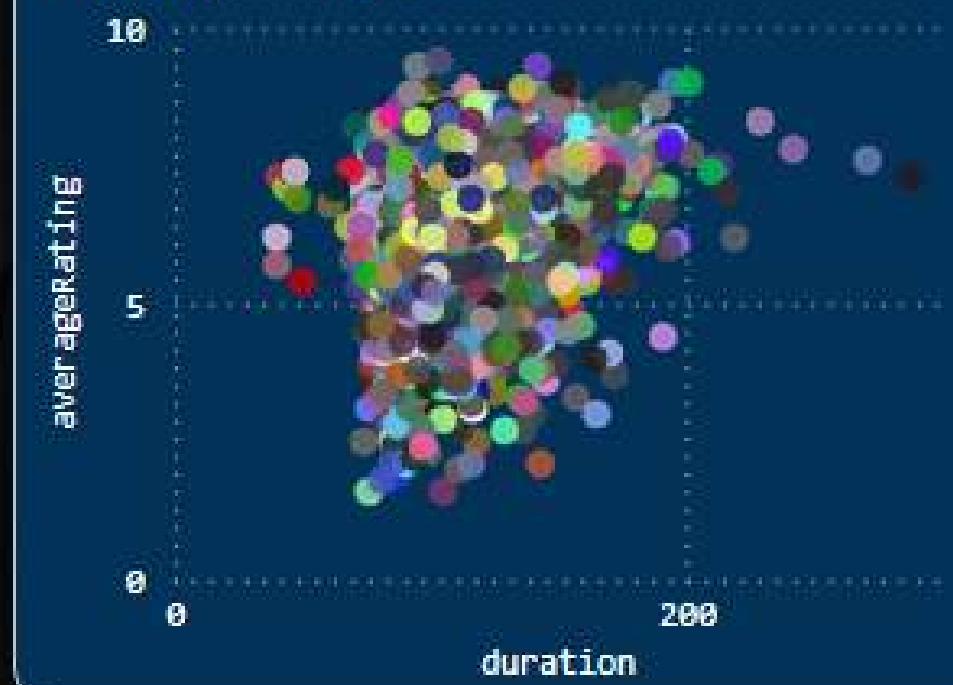
Sum of AdjustedRev by name



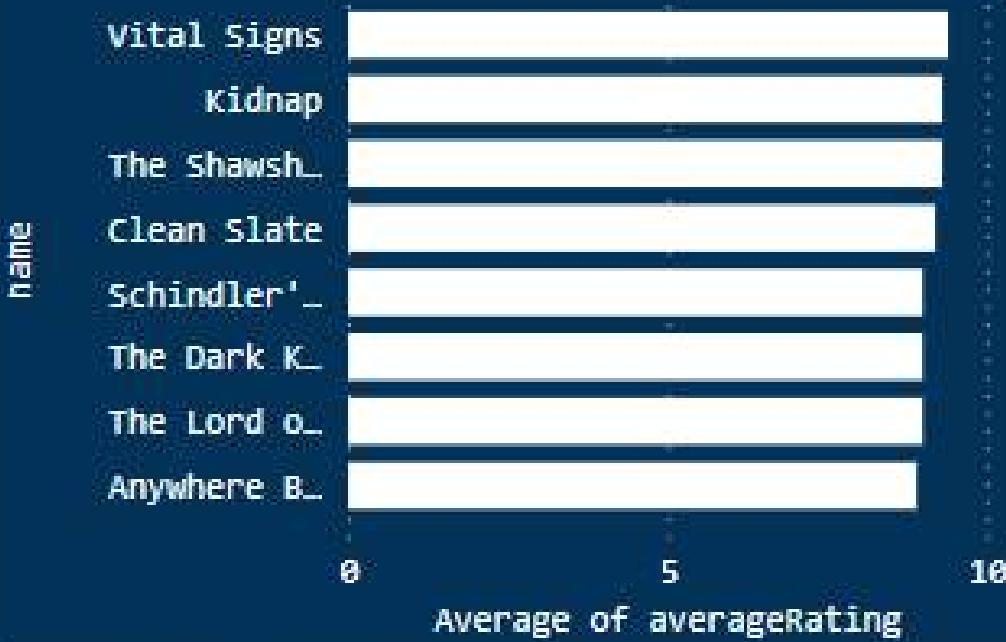
Change in Rating-Length
Correlation by year



Rating-Length Corel

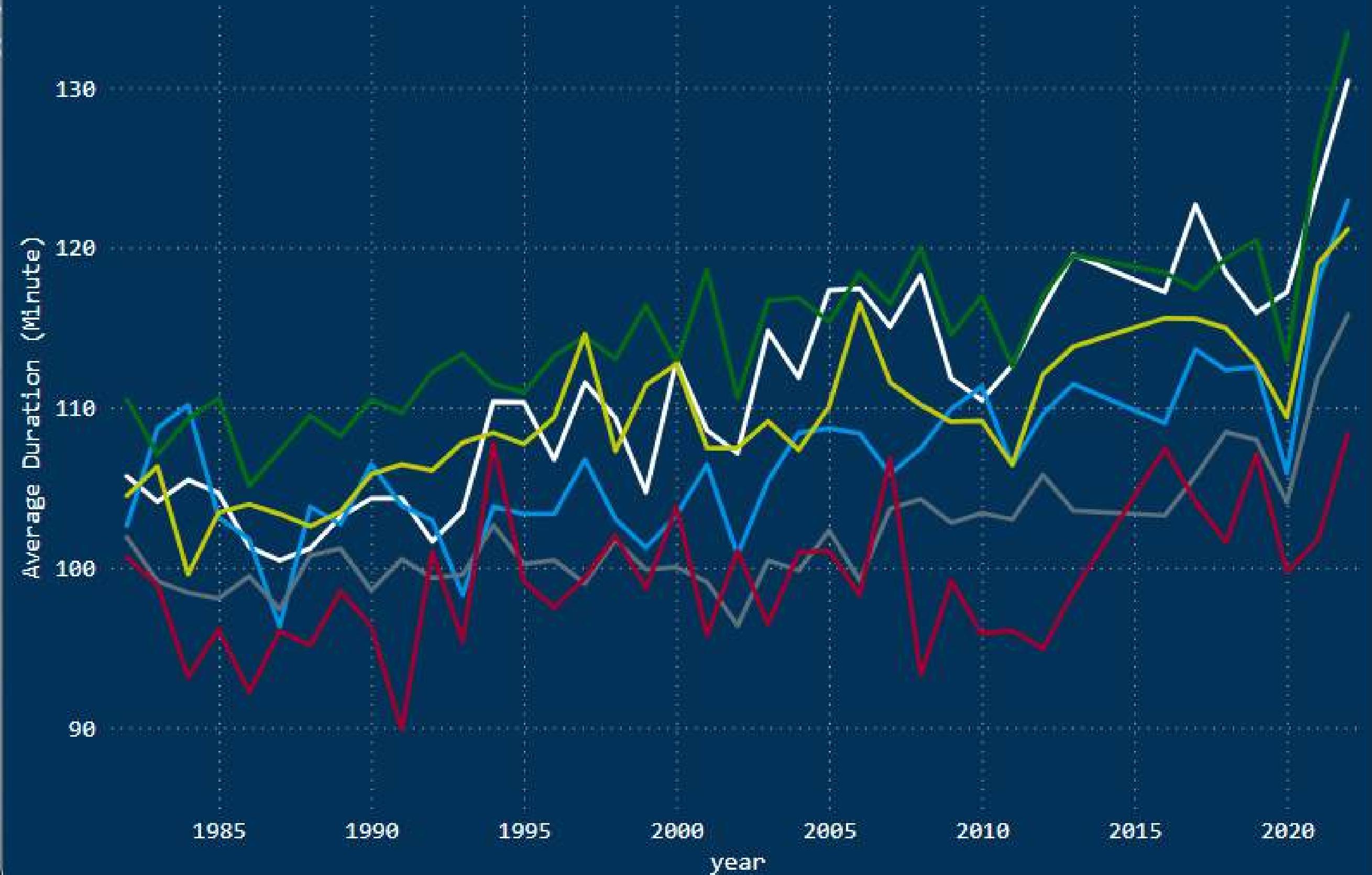


Average of averageRating by name



Change In Average Runtime Of Each Genres

● Action ● Adventure ● Comedy ● Drama ● Horror ● Thriller



Navigation

Comparison 1

Proportion

Length-Revenue

Length-Rating

Navigation

Comparison 1

Proportion

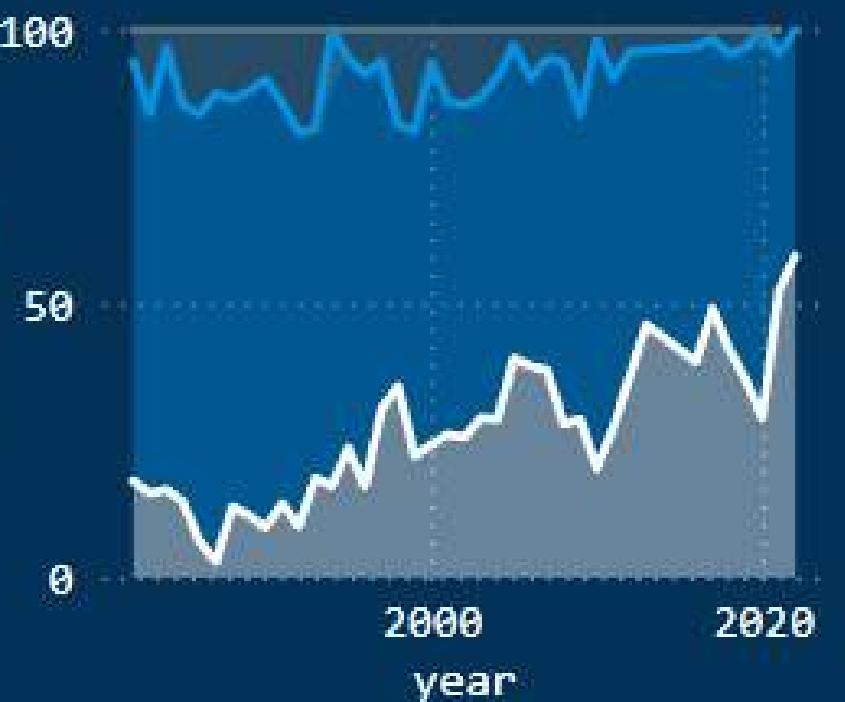
Length-Revenue

Length-Rating

Action Movie

● Long ● Medium ● Short

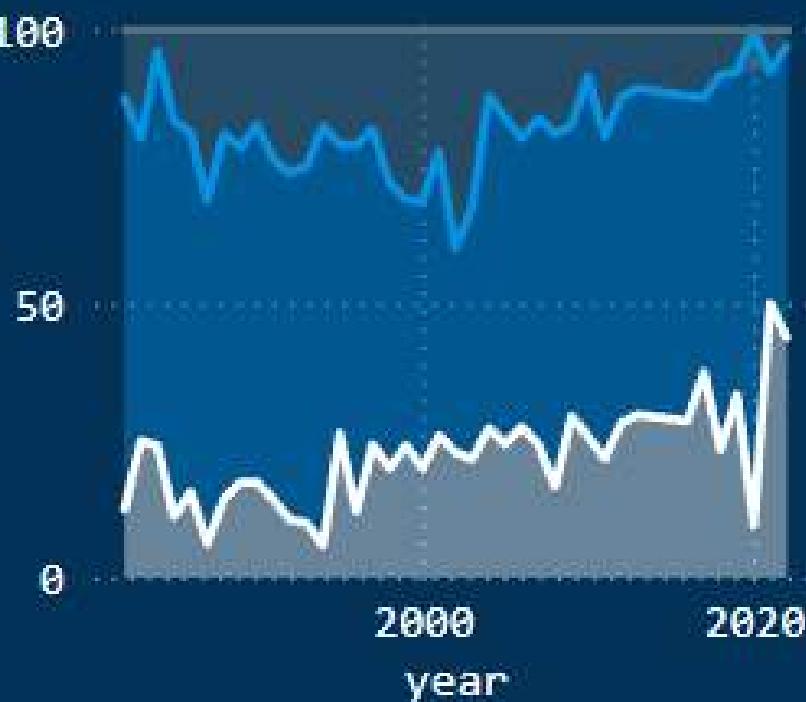
Percentage



Adventure Movie

● Long ● Medium ● Short

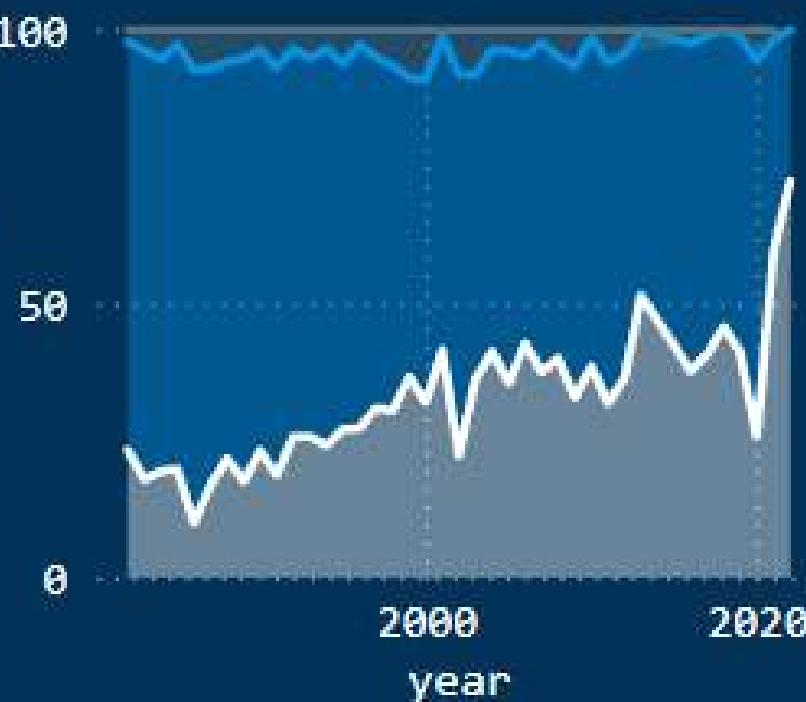
Percentage



Drama Movie

● Long ● Medium ● Short

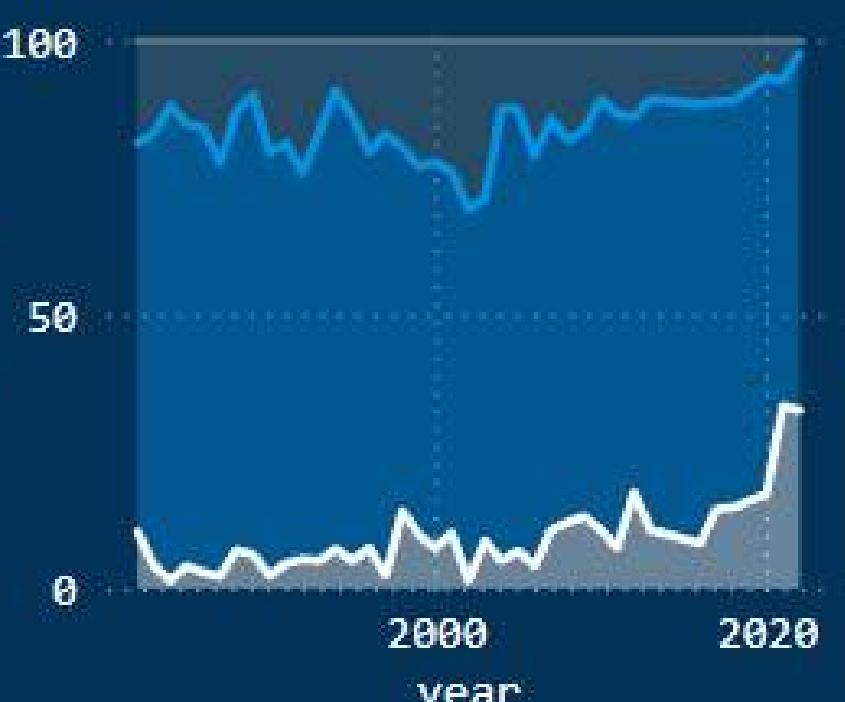
Percentage



Comedy Movie

● Long ● Medium ● Short

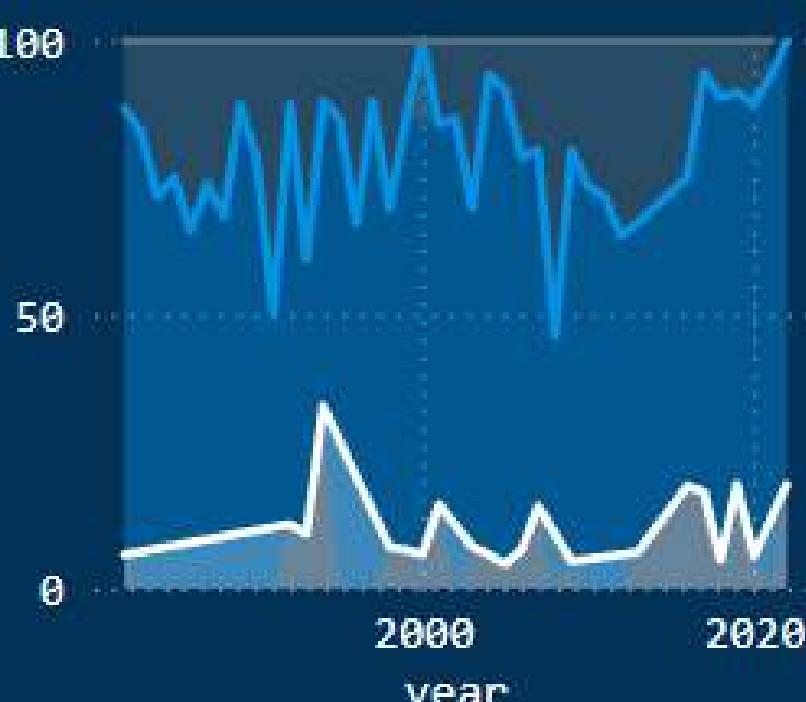
Percentage



Horror Movie

● Long ● Medium ● Short

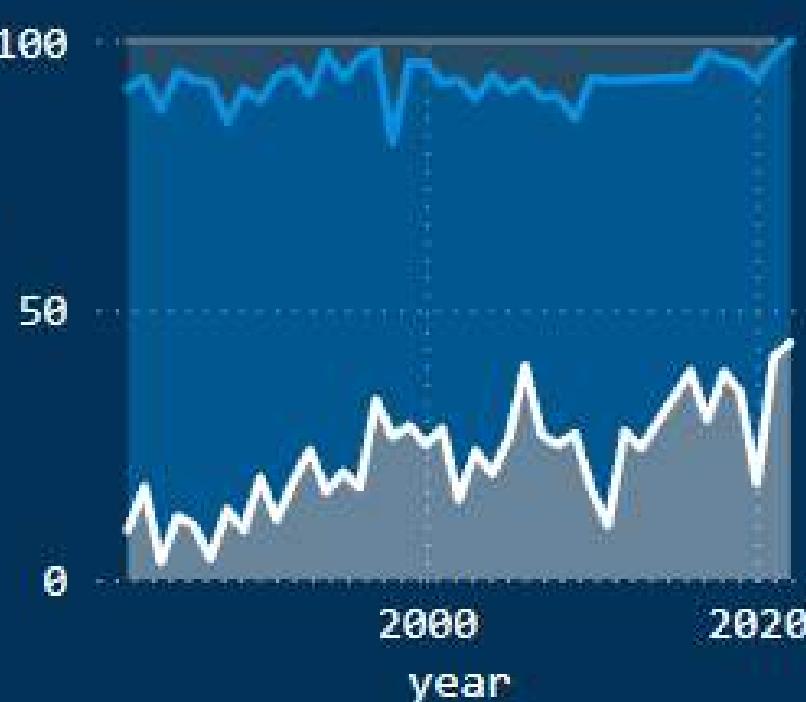
Percentage



Thriller Movie

● Long ● Medium ● Short

Percentage



Navigation

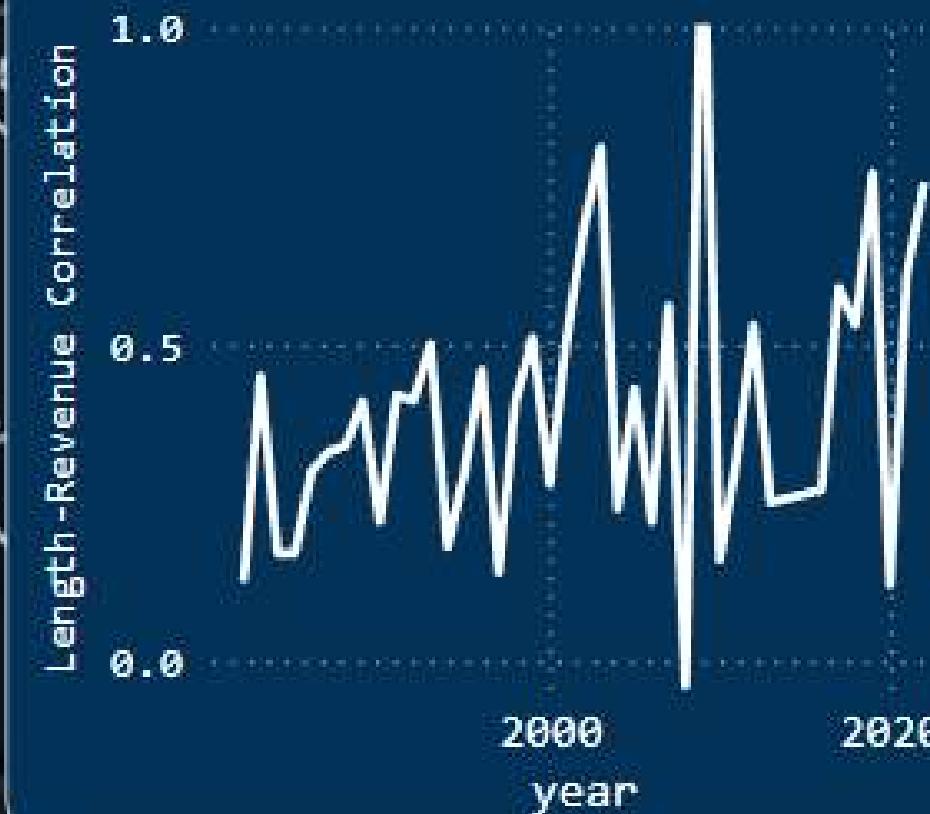
Comparison 1

Proportion

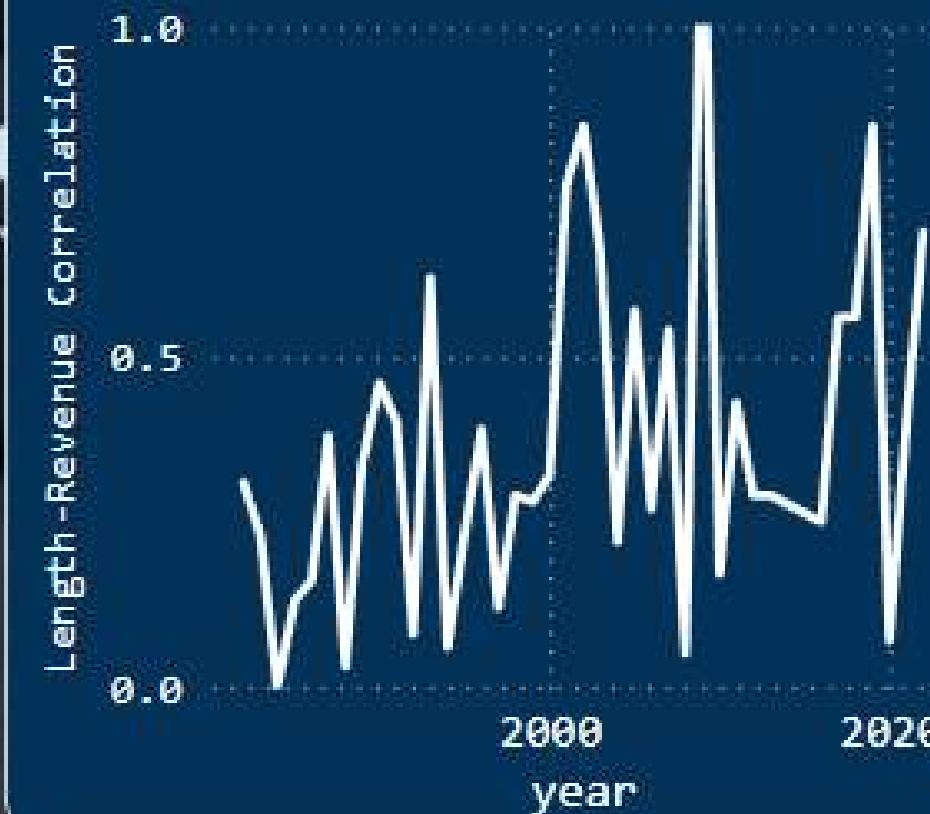
Length-Revenue

Length-Rating

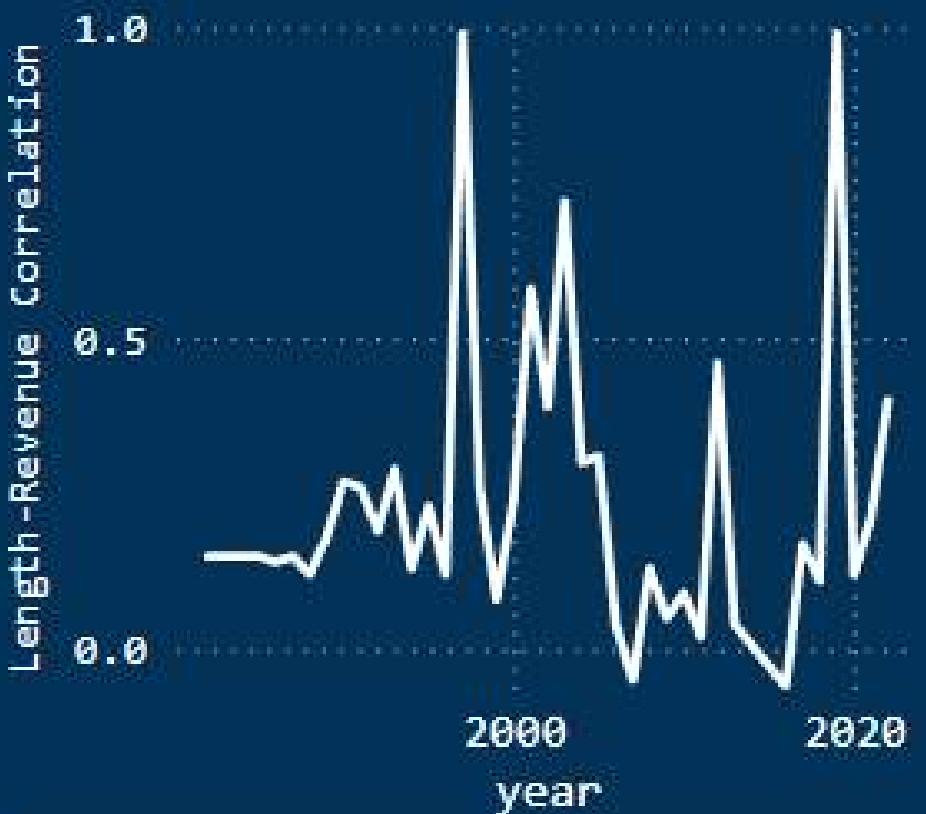
Action Movie



Adventure Movie



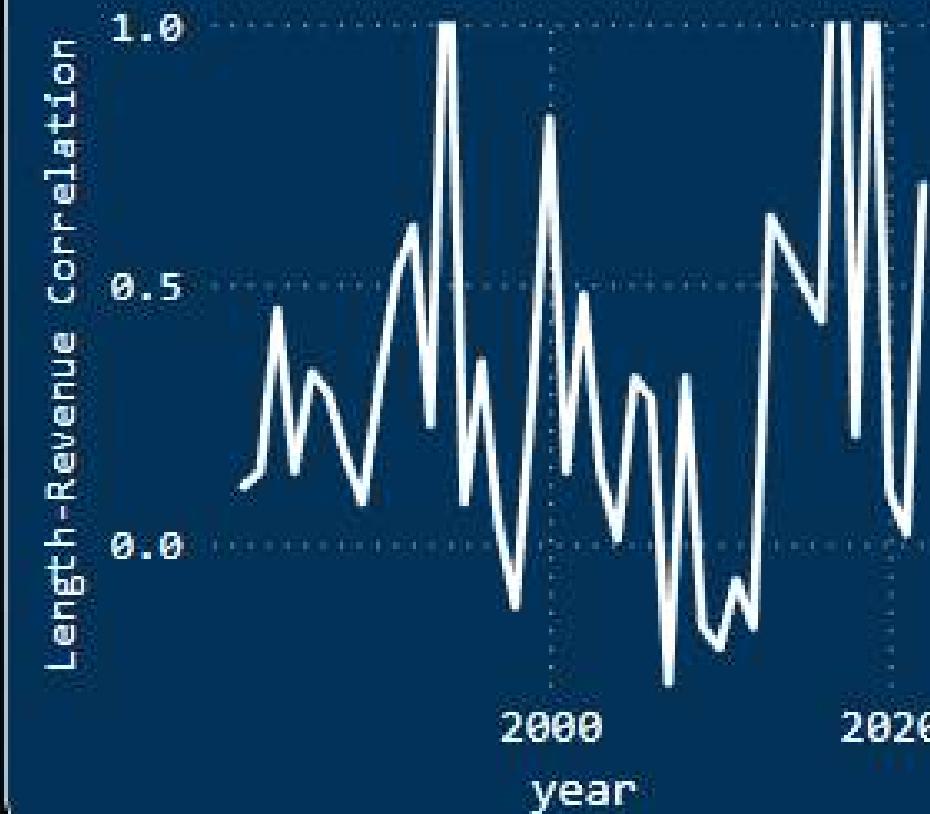
Drama Movie



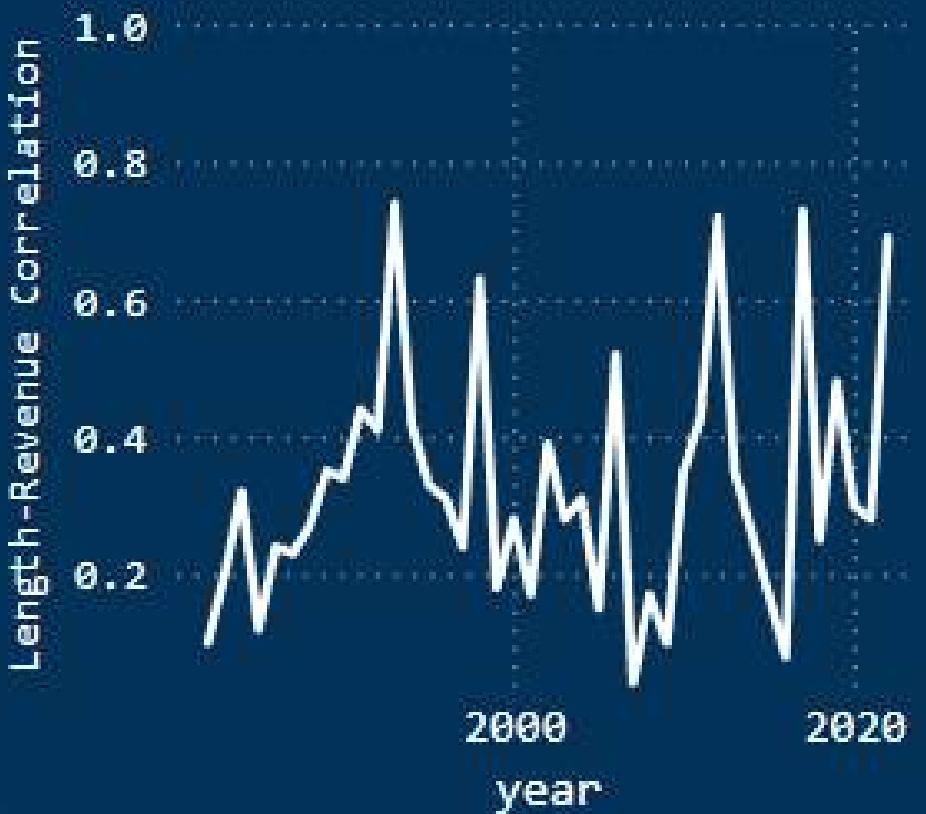
Comedy Movie



Horror Movie



Thriller



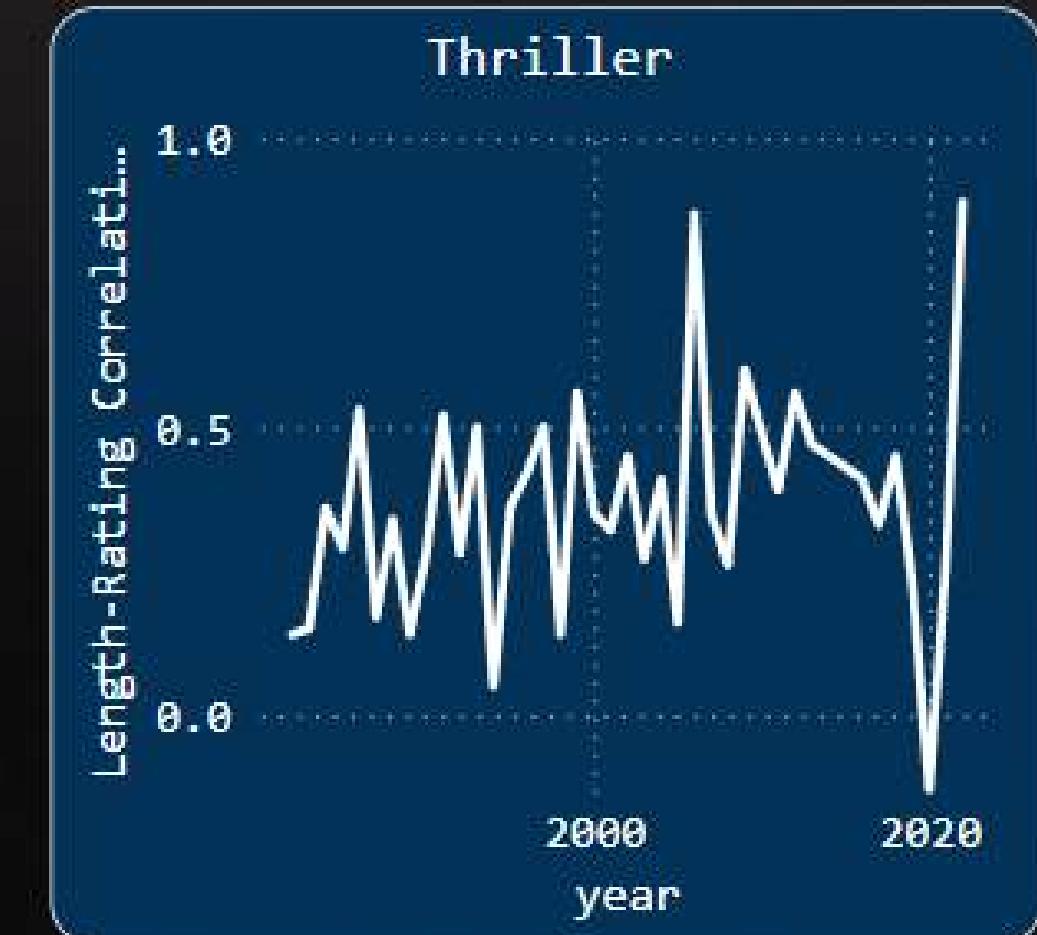
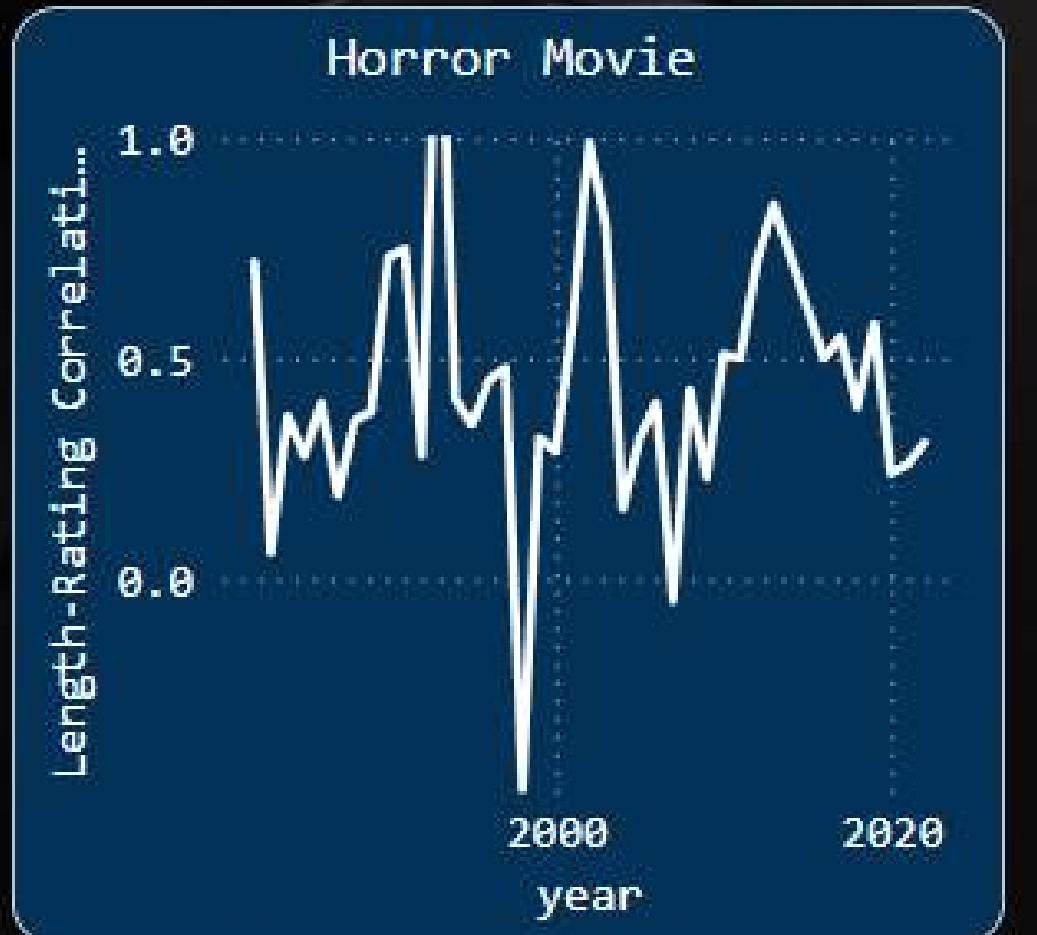
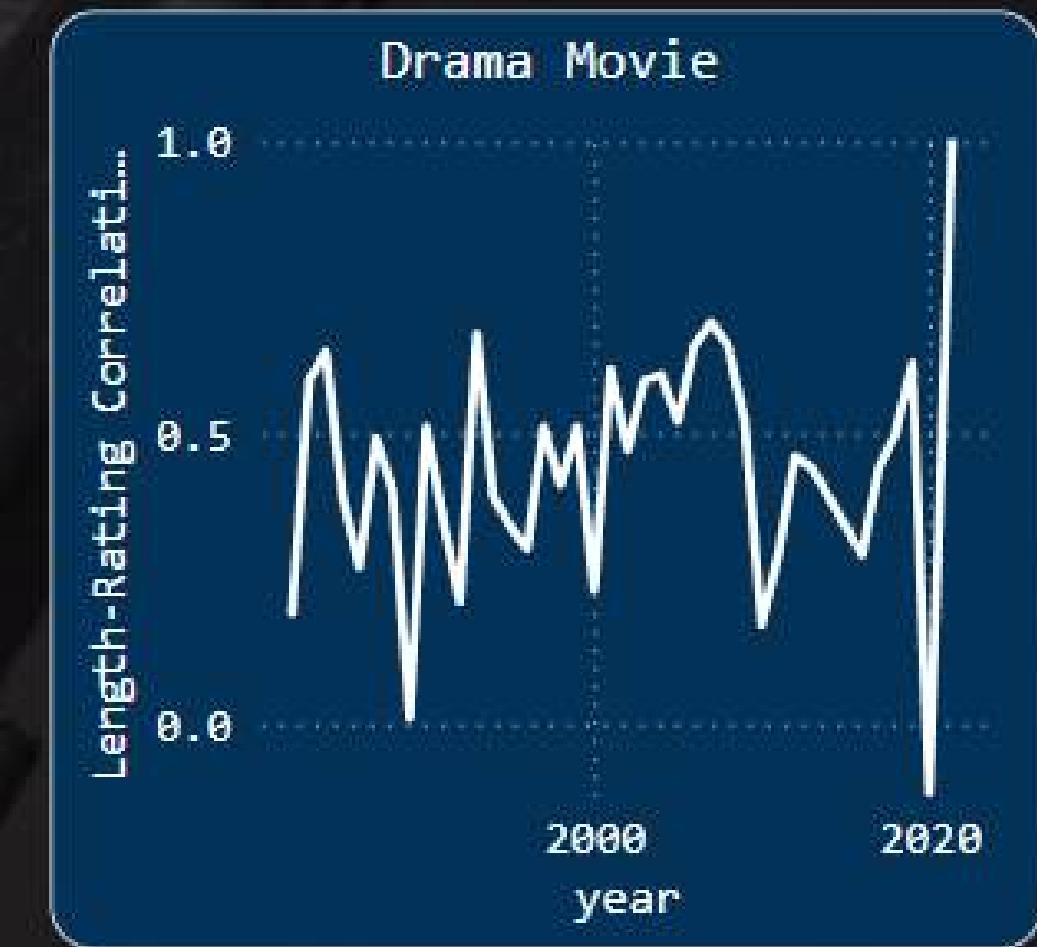
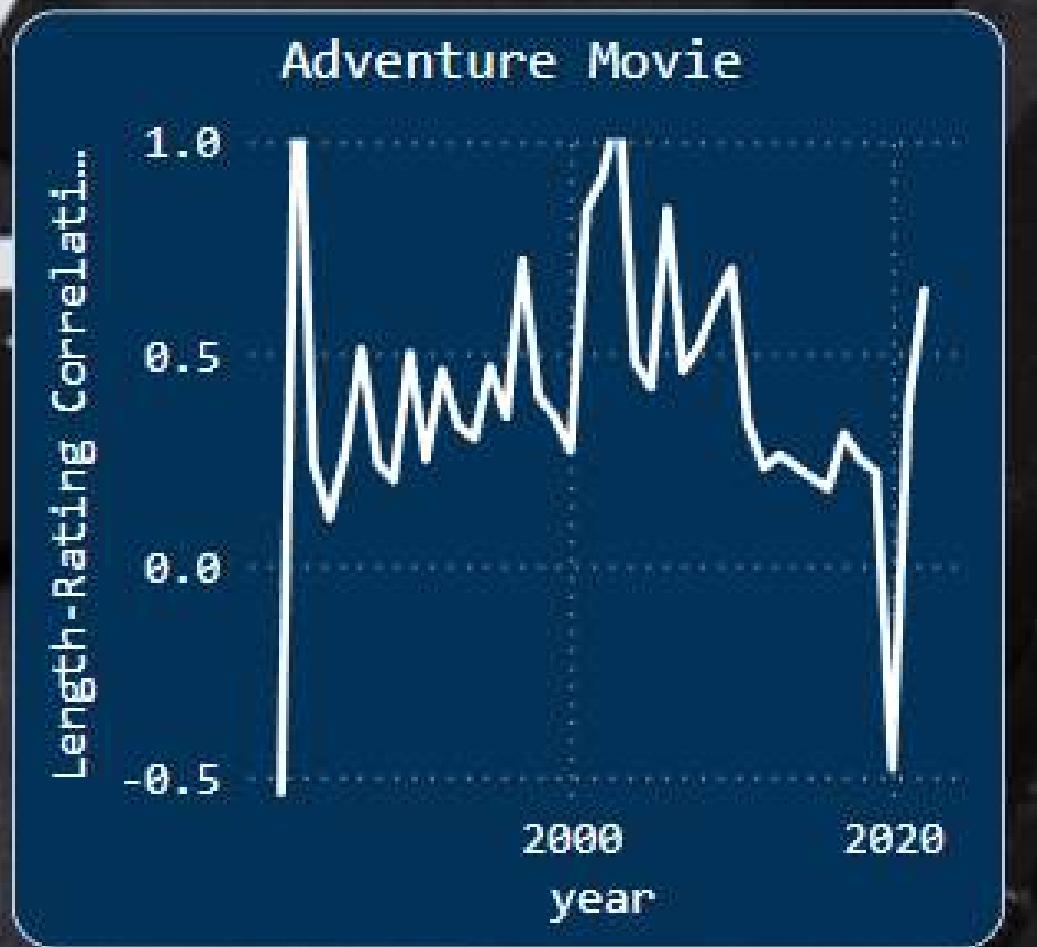
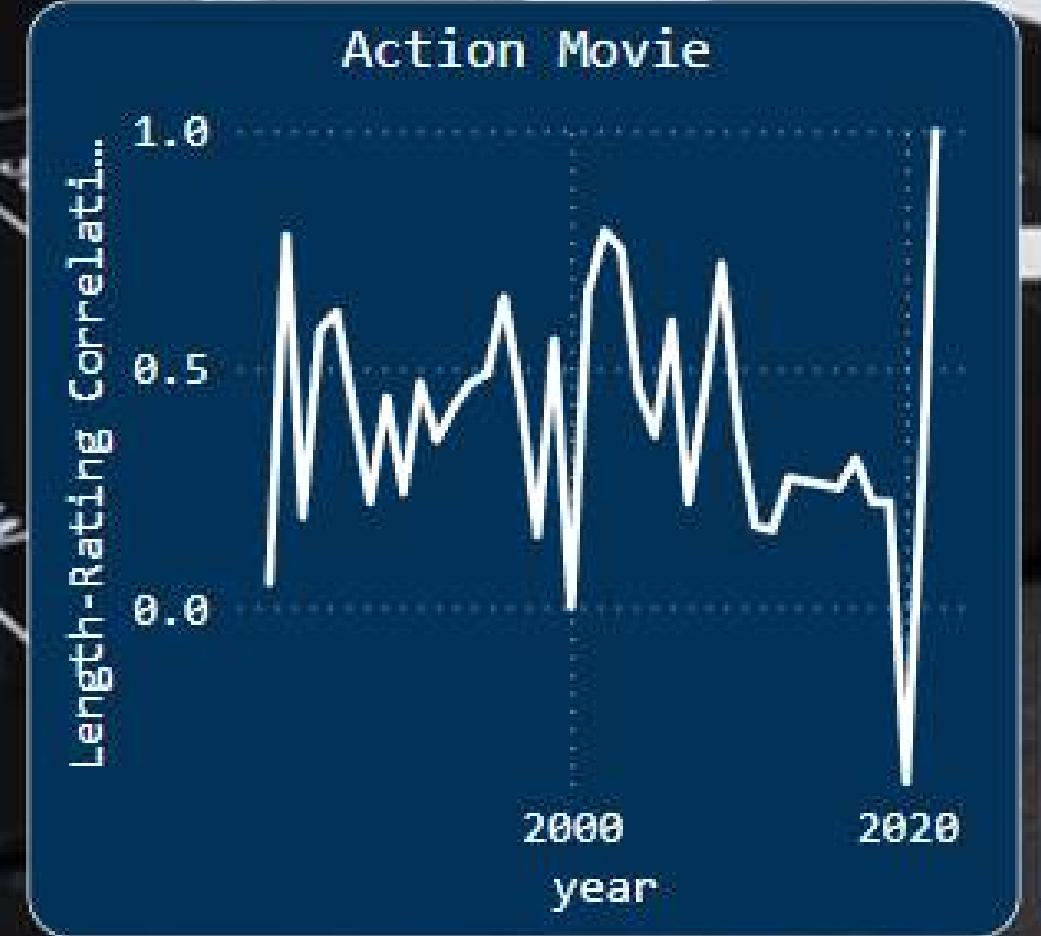
Navigation

Comparison 1

Proportion

Length-Revenue

Length-Rating



THANK YOU FOR LISTENING

Don't hesitate to ask any questions!

