

# Business Intelligence using Excel

*Prepared by Tam Nguyen*

## I. Performing ETL operation

### 1. Data Set Link

<https://www.kaggle.com/datasets/michaelmatta0/movies-ultimate-metrics-features-and-metadata?resource=download>

### 2. Data Explanation:

- The data set provides general information and financial and performance metrics of 6,500+ movies. However, after removing duplicated movies and movies that don't have enough financial information, the data set used for this assignment contains information on 4,460 unique observations.
- I broke the data set into four different tables
  - Cost\_Sales: Information on financial performance

	Label	Description
1	Movie ID	Unique identifier for each movie
2	Movie Name	Title of the movie
3	Production Budget (USD)	Estimated cost to produce the movie
4	Domestic Box Office (USD)	Total earnings in the US & Canada
5	International Box Office (USD)	Total earnings from international (excluded US and Canada)
6	Worldwide Box Office (USD)	Total worldwide earnings
7	Est. Domestic DVD Sales (USD)	Estimated DVD sales revenue in US and Canada
8	Est. Domestic Blu-ray Sales (USD)	Estimated Blu-ray sales revenue in US and Canada
9	Total Est. Domestic Video Sales (USD)	Combined video sales revenue
10	Opening Weekend (USD)	Revenue generated during the opening weekend
11	Legs	A measure of how well a movie performed after its opening weekend (domestic box office/biggest weekend))
12	Inf. Adj. Dom. BO (USD)	Domestic revenue adjusted for inflation

- Movie Infor: General movie information, including production and genre details

	Label	Description
1	Movie ID	Unique identifier for each movie
2	Movie Name	Title of the movie
3	MPAA Rating	Age classification (e.g., PG, R, etc.)
4	Running Time (minutes)	Duration of the movie
5	Franchise	The movie's franchise (if applicable)
6	Keywords	Keywords describing the movie's themes
7	Source	Whether the movie is an original screenplay, adaptation, etc.
8	Genre	Primary genre (e.g., Action, Comedy, Drama)
9	Production Method	Live-action or animated
10	Creative Type	Storytelling category
11	Production/Financing Companies	Studios involved in production
12	Production Countries	Countries where the movie was produced
13	Languages	Languages spoken in the movie

- Release\_Infor: Information about Theatrical & Digital Releases

	Label	Description
1	Movie ID	Unique identifier for each movie
2	Movie Name	Title of the movie
3	Release Date	Date when the movie was released in theaters
4	Video Release	Date when the movie was released for home viewing
5	Domestic Releases	Number of domestic releases
6	International Releases	Number of international releases
7	Movie URL	Direct link to the movie's page on The Numbers

- Theater\_performance: Information about the number of theaters that show the movie and the average run weeks in theaters of the movie.

	Label	Description
1	Movie ID	Unique identifier for each movie
2	Movie Name	Title of the movie
3	Theater counts	Number of theaters the movie was released in
4	Domestic Share Percentage	Domestic box office share of total revenue

### 3. ETL

- Cost\_Sales tables: Before loading these data sets to the worksheet, I promoted the first row as a label and changed the movie ID type from string to number.

The screenshot shows the Power Query Editor interface. The main area displays a table with columns: Movie ID, Movie Name, and Production Budget. The first row of data is highlighted in green. The 'Movie ID' column is selected, and the 'Changed Type' step is visible in the 'APPLIED STEPS' pane on the right. The status bar at the bottom indicates '12 COLUMNS, 999+ ROWS' and 'Column profiling based on top 1000 rows'.

- Movie\_Infor Table:
  - Promoted the first row as a label and changed the movie ID type from string to number.
  - Replaced empty cells from “Source”, “Production Method”, “Creative Type”, "Genre" columns to “Not Identified”

Movie\_Infor - Power Query Editor

File Home Transform Add Column View

Group By Use First Row as Headers Count Rows

Data Type: Text Detect Data Type Rename

Replace Values Fill Pivot Column Convert to List

Unpivot Columns Move Split Column Format Extract Merge Columns

Statistics Standard Rounding Scientific Trigonometry Rounding Information Date Time Duration Structured Column

Queries [4] Movie\_Infor Cost\_Sales Release Theater\_performance

fx = Table.ReplaceValue("#Replaced Value2","", "Not Identified",

Company	Production Countries	Languages
1	United States	English
2	United States	English
3	United States	English
4	United States	English
5	United States	English
6	United States	English
7	United States	English
8	United States	English
9	United States	English
10	United States	English
11	United States	English
12	United States	English
13	United States	English
14	United States	English
15	United States	English
16	United States	English, Apache
17	United States	English
18	United States	English
19	United States	English
20	United States	English
21	United States	English
22	United States	English
23	United States	English
24	United States	English

Query Settings

PROPERTIES Name Movie\_Infor

APPLIED STEPS Source Promoted Headers Changed Type Replaced Value Replaced Value1 Replaced Value2 Replaced Value3

TN

13 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 1:51 PM

• Release Table:

- Promoted the first row as a label and changed the movie ID type from string to number.
- Added year/quarter column. Based on the quarter column, I added a conditional column to determine which season that movies are released (Spring/Summer/Fall/Winter)
- Renamed "Year" to "Release Year" and "Quarter" to "Release Quarter" for more clarification.

Release - Power Query Editor

File Home Transform Add Column View

Close & Load Refresh Preview Manage

Advanced Editor

Choose Columns Remove Columns Keep Rows Remove Rows Sort

Data Type: Whole Number Use First Row as Headers Replace Values

Merge Queries Append Queries Combine Files

Manage Parameters Data source settings Parameters Data Sources New Source Recent Sources Enter Data New Query

Queries [3] Movie\_Infor Cost\_Sales Release

fx = Table.RenameColumns(#"Removed Columns1",{"Year", "Release Year"})

Movie Title	Release Year	Release Quarter	Release Season
1	2015	4	Winter
2	2022	4	Winter
3	2023	2	Summer
4	2019	2	Summer
5	2011	2	Summer
6	2015	2	Summer
7	2023	2	Summer
8	2018	2	Summer
9	2018	2	Summer
10	2007	2	Summer
11	2017	4	Winter
12	2023	3	Fall
13	2016	4	Winter
14	2019	4	Winter
15	2023	4	Winter
16	2012	1	Spring
17	2016	1	Spring
18	2017	4	Winter

Query Settings

PROPERTIES Name Release

APPLIED STEPS Source Promoted Headers Changed Type Inserted Year Inserted Quarter Removed Columns Renamed Columns Added Conditional Column Removed Columns1 Renamed Columns1

TN

10 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 11:08 PM

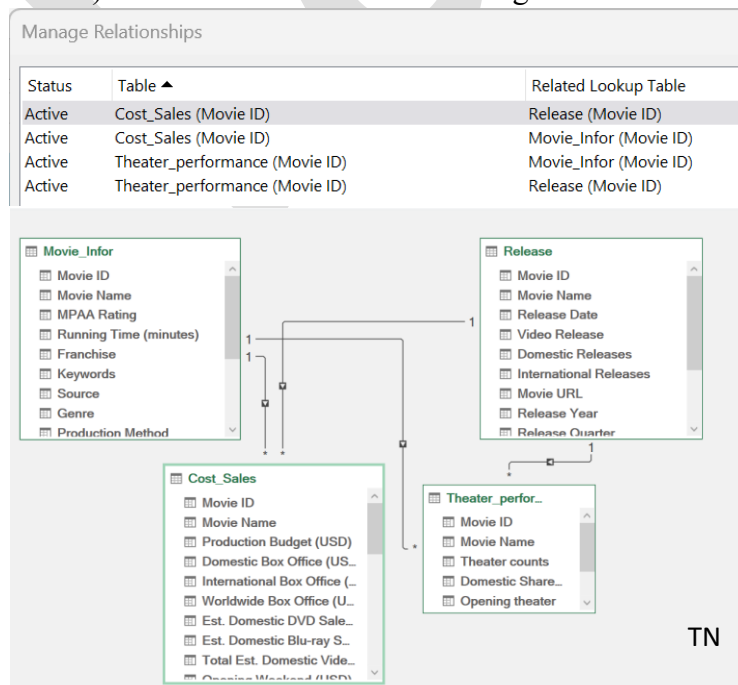
- Theater\_Performance Table:
  - Promoted the first row as a label and changed the movie ID type from string to number.
  - Extracted text from the “Theater Counts” column to get information about the number of opening theaters and average weeks run per theater.
  - Renamed newly added columns for more details.

7 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 11:12 PM

## II. Making table relationship with the Data Model

I have two data tables (Cost\_Sales and Theater\_Performance) and two lookup tables (Movie\_Infor and Release). I connected all four tables using Movie ID.



### III. Use the pivot table and DAX operation to analyze data and generate three fruitful conclusions.

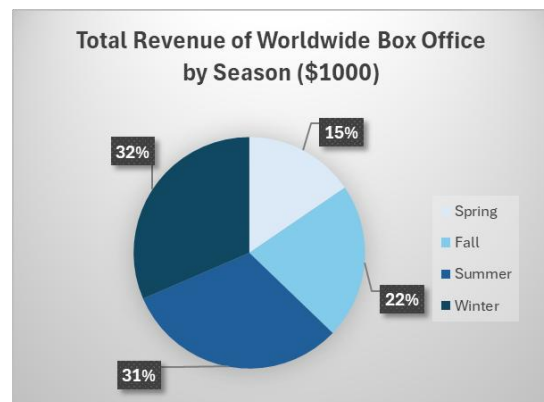
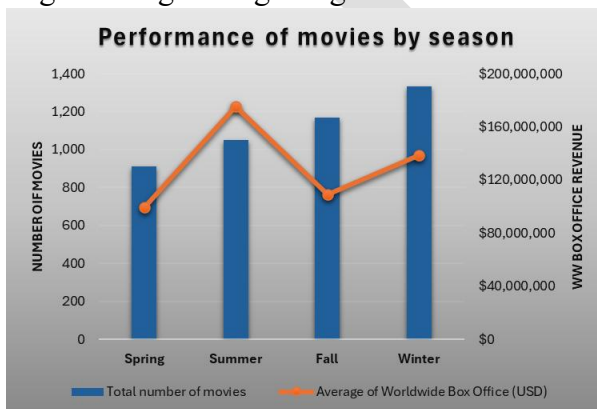
#### 1. DAX

Manage Measures		
<div> <div>New</div> <div>Edit</div> <div>Delete</div> </div>		
Measure	Formula	
Average of Ave. weeks run per theater	AVERAGE([Ave. weeks run per theater])	
Average of Domestic Box Office (USD)	AVERAGE([Domestic Box Office (USD)])	
Average of Est. Profit	AVERAGE([Est. Profit])	
Average of Est. Revenue	AVERAGE([Est. Revenue])	
Average of International Box Office (USD)	AVERAGE([International Box Office (USD)])	
Average of Production Budget (USD)	AVERAGE([Production Budget (USD)])	
Average of ROI	AVERAGE([ROI])	
Average of Running Time (minutes)	AVERAGE([Running Time (minutes)])	
Average of Worldwide Box Office (USD)	AVERAGE([Worldwide Box Office (USD)])	
Count of Movie Name	COUNTA([Movie Name])	
Sum of Domestic Box Office (USD)	SUM([Domestic Box Office (USD)])	
Sum of Est. Profit	SUM([Est. Profit])	
Sum of Est. Revenue	SUM([Est. Revenue])	
Sum of International Box Office (USD)	SUM([International Box Office (USD)])	
Sum of Opening theater	SUM([Opening theater])	
Sum of Production Budget (USD) 2	SUM([Production Budget (USD)])	TN
Sum of Worldwide Box Office (USD)	SUM([Worldwide Box Office (USD)])	

#### 2. Pivot Table and conclusions

- Conclusion 1: Financial Performance by Season**

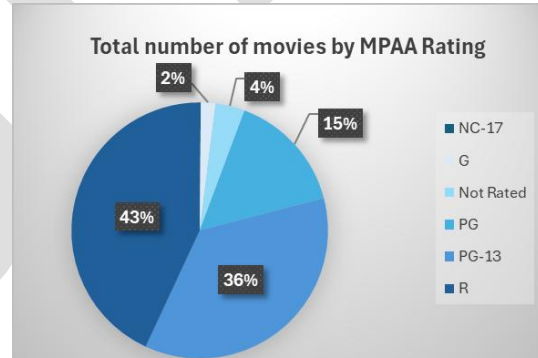
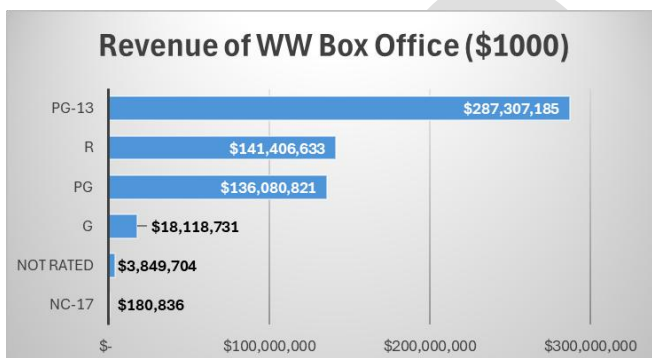
Spring has the fewest movie releases, leading to the lowest box office revenue. Winter sees the highest number of releases (1,333), resulting in the greatest worldwide box office earnings—around \$183 billion. Still, movies released in the summer tend to perform better, generating the highest global revenue with an average of about \$175 million per movie.



Row Labels	Name	Count of Movie	Sum of Worldwide Box Office (USD)	Sum of International Box Office (USD)	Sum of Domestic Box Office (USD)	Average of Worldwide Box Office (USD)	Average of International Box Office (USD)	Average of Domestic Box Office (USD)
Spring		911	\$ 90,772,043,915	\$ 48,863,706,554	\$ 41,908,337,361	\$ 99,640,004	\$ 53,637,439	\$ 46,002,566
Summer		1,049	\$183,932,987,784	\$106,086,643,237	\$ 77,846,344,547	\$175,341,266	\$101,131,214	\$ 74,210,052
Fall		1,167	\$127,472,259,852	\$ 70,145,991,424	\$ 57,326,268,428	\$109,230,728	\$ 60,107,962	\$ 49,122,766
Winter		1,333	\$184,766,618,345	\$106,641,213,949	\$ 78,125,404,396	\$138,609,616	\$ 80,000,911	\$ 58,608,705
<b>Grand Total</b>		<b>4,460</b>	<b>\$586,943,909,896</b>	<b>\$331,737,555,164</b>	<b>\$ 255,206,354,732</b>	<b>\$131,601,774</b>	<b>\$ 74,380,618</b>	<b>\$ 57,221,156</b>

### • Conclusion 2: Financial Performance by MPAA rating

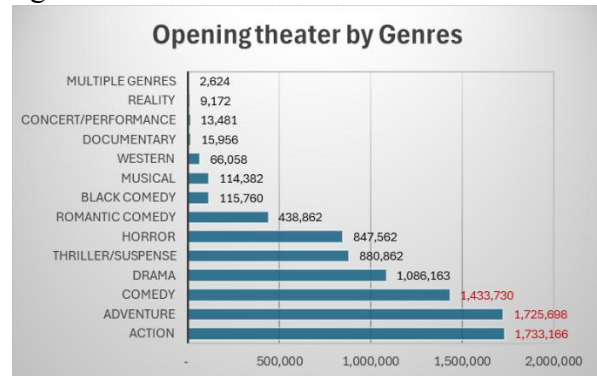
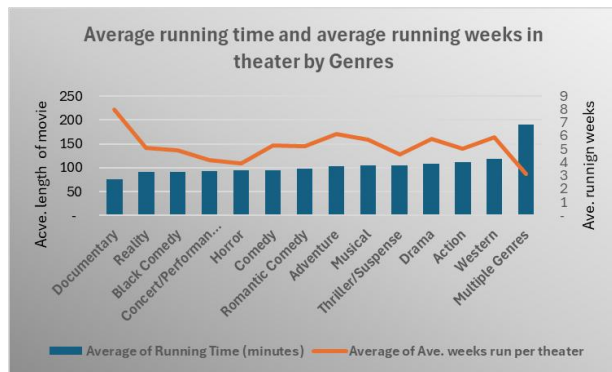
Movies with an R rating (under 17 requires a parent) make up the largest share of released films, accounting for 43% of the total. However, movies with a G rating (suitable for all ages) perform the best and have the highest average revenue both domestically and internationally (approximately \$95 million per G movie domestically, and around \$125 million internationally).



Row Labels	Name	Count of Movie	Sum of Worldwide Box Office (USD)	Average of Worldwide Box Office (USD)	Average of Domestic Box Office (USD)	Average of International Box Office (USD)
G		82	\$ 18,118,730,977	\$ 220,960,134	\$ 95,423,114	\$125,537,020
NC-17		6	\$ 180,836,394	\$ 30,139,399	\$ 7,414,965	\$ 22,724,435
Not Rated		165	\$ 3,849,703,624	\$ 23,331,537	\$ 1,864,536	\$ 21,467,001
PG		681	\$136,080,820,845	\$ 199,824,994	\$ 86,846,135	\$112,978,859
PG-13		1,605	\$287,307,185,135	\$ 179,007,592	\$ 75,674,756	\$103,332,836
R		1,921	\$141,406,632,921	\$ 73,610,949	\$ 34,580,593	\$ 39,030,356
<b>Grand Total</b>		<b>4,460</b>	<b>\$586,943,909,896</b>	<b>\$ 131,601,774</b>	<b>\$ 57,221,156</b>	<b>\$ 74,380,618</b>

### • Conclusion 3: Theater Performance by Genre

Documentaries tend to run the longest in theaters, averaging about 8 weeks, despite having the shortest runtime at approximately 76 minutes. However, there are few theaters that screen documentaries. Action, Adventure, and Comedy are the top three genres preferred by theaters, with the highest number of openings.



**Conclusion 3**

Row Labels	Average of Running Time (minutes)	Average of Ave. weeks run per theater	Sum of Opening theater
Action	112	5	1,733,166
Adventure	103	6	1,725,698
Black Comedy	91	5	115,760
Comedy	95	5	1,433,730
Concert/Performance	93	4	13,481
Documentary	76	8	15,956
Drama	108	6	1,086,163
Horror	95	4	847,562
Multiple Genres	191	3	2,624
Musical	105	6	114,382
Reality	91	5	9,172
Romantic Comedy	98	5	438,862
Thriller/Suspense	105	5	880,862
Western	118	6	66,058
<b>Grand Total</b>	<b>103</b>	<b>5</b>	<b>8,483,476</b>

**PivotTable Fields**

Active: All

Choose fields to add to report:

Search

☐ fx Average of Theater coverage (%)

☒ fx Average of Ave. weeks run per theater

Drag fields between areas below:

**Filters**

**Columns**

Σ Values

**Rows**

Genre

Σ Values

Average of Ave. weeks run per th...

Sum of Opening theater

#### IV. Use the Calculate function and generate at least two meaningful results with justifications.

##### 1. Calculated function (3)

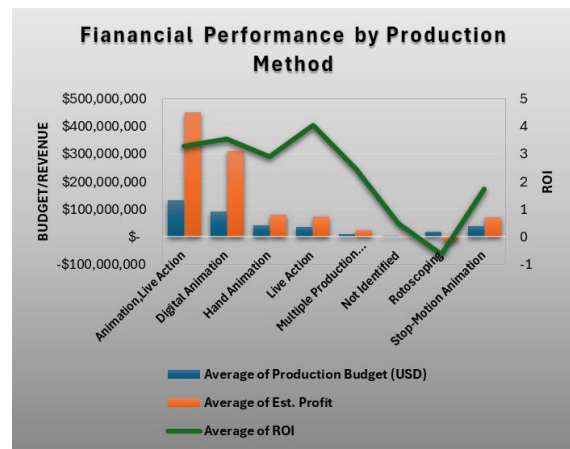
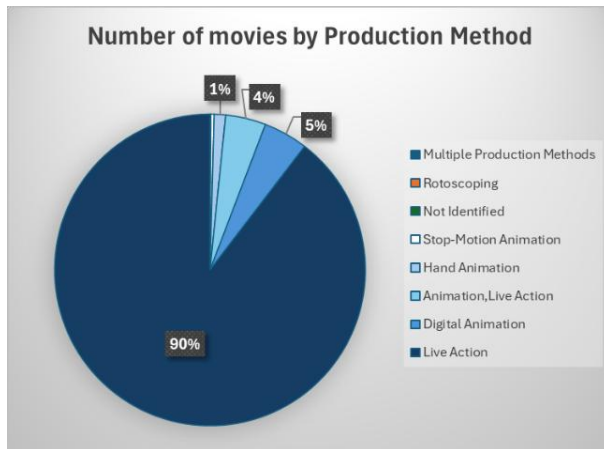
- **Est. Profit** = Cost\_Sales[Worldwide Box Office (USD)]+Cost\_Sales[Total Est. Domestic Video Sales (USD)]
- **Est. Revenue** = Cost\_Sales[Worldwide Box Office (USD)]+Cost\_Sales[Total Est. Domestic Video Sales (USD)]
- **ROI** = Cost\_Sales[Est. Profit]/Cost\_Sales[Production Budget (USD)]

##### 2. Conclusion

###### • **Conclusion 4: Financial Performance by Production Method**

In the movie industry, most films use live action as a production method (90% of movies are live-action), and live-action movies have the highest ROI in the industry (about 4%). However, based on the average budget and profit per film, animated movies have the highest production costs and the greatest profits.





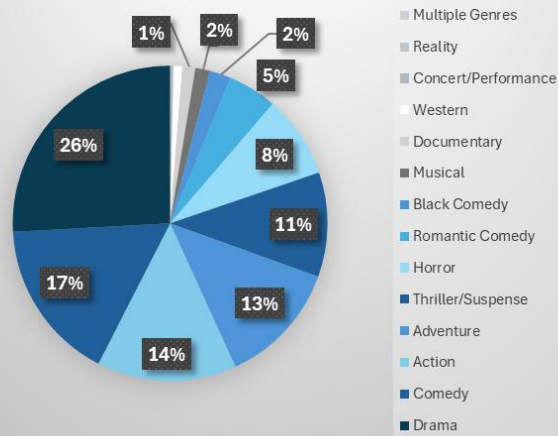
Drama	B	C	D	E	F	G	H	I
TN		1153		Adventure	\$ 249,203,681	3.17		
Row Labels	Count of Movie Name	Sum of Production Budget (USD) 2	Average of Production Budget (USD)	Sum of Est. Profit	Average of Est. Profit	Average of ROI		
Animation, Live Action	187	\$ 25,116,500,000	\$ 134,312,834	\$ 84,316,515,896	\$ 450,890,459	3.28		
Digital Animation	206	\$ 18,788,600,000	\$ 91,206,796	\$ 64,120,144,291	\$ 311,262,836	3.56		
Hand Animation	53	\$ 2,183,650,000	\$ 41,200,943	\$ 4,314,397,185	\$ 81,403,720	2.92		
Live Action	3,995	\$ 144,290,141,584	\$ 36,117,682	\$ 301,195,854,609	\$ 75,393,205	4.04		
Multiple Production Methods	1	\$ 10,000,000	\$ 10,000,000	\$ 24,708,080	\$ 24,708,080	2.47		
Not Identified	2	\$ 2,300,000	\$ 1,150,000	\$ 750,717	\$ 375,359	0.49		
Rotoscoping	1	\$ 20,000,000	\$ 20,000,000	\$ 12,594,916	\$ 12,594,916	0.63		
Stop-Motion Animation	15	\$ 578,000,000	\$ 38,533,333	\$ 1,088,166,028	\$ 72,544,402	1.75		
<b>Grand Total</b>	<b>\$ 4,460</b>	<b>\$ 190,989,191,584</b>	<b>\$ 42,822,689</b>	<b>\$ 455,047,941,890</b>	<b>\$ 102,028,687</b>	<b>3.96</b>		

• **Conclusion 5: Financial performance by Genre**

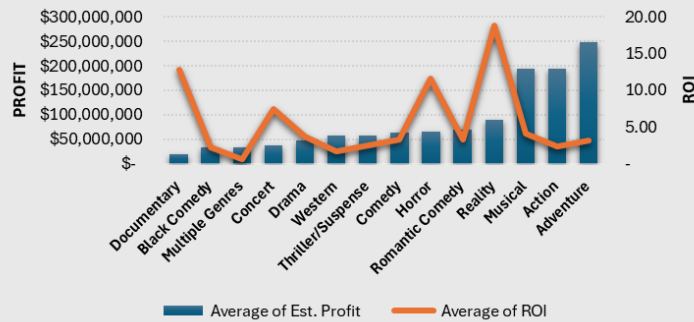
The drama genre has the highest number of released films (1,153 movies). However, in terms of profit, adventure is the most profitable genre, with an average of \$249 million per movie. In terms of return on investment, reality has the highest ROI, at 18.83%.



### Number of movies by Genre



### Profit and ROI



Row Labels	Count of Movie Name	Sum of Est. Profit	Average of Est. Profit	Average of ROI
Action	640	\$ 124,478,309,125	\$ 194,497,358	2.36
Adventure	566	\$ 141,049,283,715	\$ 249,203,681	3.17
Black Comedy	98	\$ 3,197,371,221	\$ 32,626,237	2.27
Comedy	741	\$ 47,347,941,292	\$ 63,897,357	3.26
Concert/Performance	11	\$ 399,459,078	\$ 36,314,462	7.50
Documentary	59	\$ 1,142,135,190	\$ 19,358,224	12.84
Drama	1,153	\$ 53,422,856,145	\$ 46,333,787	3.65
Horror	378	\$ 25,039,955,832	\$ 66,243,269	11.54
Multiple Genres	1	\$ 34,190,589	\$ 34,190,589	0.65
Musical	68	\$ 13,144,193,410	\$ 193,296,962	4.15
Reality	3	\$ 267,030,959	\$ 89,010,320	18.83
Romantic Comedy	223	\$ 15,695,012,134	\$ 70,381,220	3.36
Thriller/Suspense	478	\$ 27,502,916,111	\$ 57,537,481	2.43
Western	41	\$ 2,327,287,089	\$ 56,763,100	1.70
<b>Grand Total</b>	<b>4,460</b>	<b>\$ 455,047,941,890</b>	<b>\$ 102,028,687</b>	<b>3.962429361</b>

PivotTable Fields

Active

All

Choose fields to add to report:

prof

Profitable

Sum of Est. Profit

Drag fields between areas below:

Filters

Columns

Rows

Values