Analysis of Film Dataset with Google Sheets and SQL

Report by Riina Kikkas

Introduction

This report analyzes a dataset of 26 films using Google Sheets and SQL queries. The goal is to explore patterns in genre preferences, IMDb ratings, box office performance, release years, and changes in ranking after adjusting revenues for inflation The analysis highlights how different factors relate to a film's commercial and critical success.
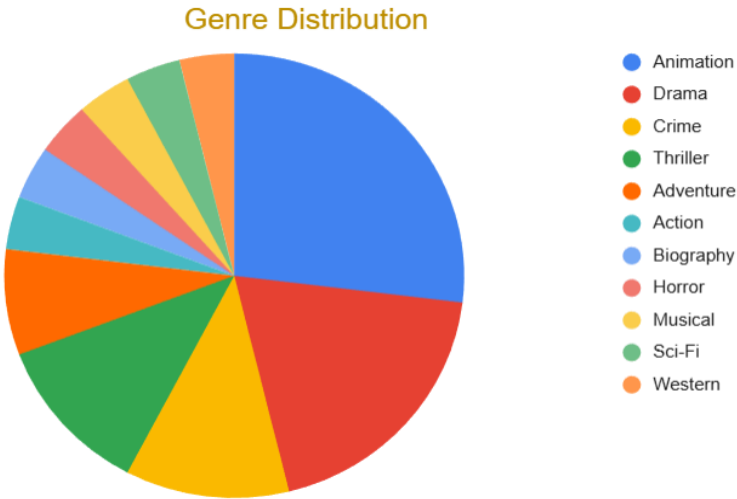
Dataset:

Title	Genre	Year	Rating	Box Office (adjusted)	Duration min	BO / Duration	Box Office (Orig)	Rank Cor	Rank Orig	Rank Change
Forrest Gump	Drama	1994	8.8	1,482,652,478	142	10441214.63	678,226,465	2	2	0
Life Is Beautiful	Drama	1997	8.6	464,464,535	116	4004004.61	230,099,013	12	8	-4
Léon: The Professional	Thriller	1994	8.5	98,937,312	110	899430.1126	45,258,012	19	18	-1
Alien	Sci-Fi	1979	8.5	468,114,083	116	4035466.23	104,900,000	11	17	6
Aliens	Action	1986	8.4	541,880,765	137	3955334.05	183,316,455	8	13	5
No Country for Old Men	Crime	2007	8.2	268,170,861	122	2198121.814	171,627,166	16	14	-2
The Handmaiden	Thriller	2016	8.1	51,110,677	144	354935.2597	37,863,670	20	20	0
Only Lovers Left Alive	Drama	2013	7.2	11,403,835	123	92714.10602	8,200,000	24	24	0
The Rocky Horror Picture Show	Musical	1975	7.4	698,535,613	100	6985356.134	116,000,000	7	16	9
My Neighbor Totoro	Animation	1988	8.1	112,492,541	86	1308052.797	41,076,708	18	19	1
Kiki's Delivery Service	Animation	1989	7.9	27,218,217	102	266845.2607	10,417,620	23	23	0
Ponyo	Animation	2008	7.6	307,229,238	101	3041873.646	204,173,694	15	10	-5
Whisper of the Heart	Animation	1995	7.9	9,756,901	111	87900.00685	4,589,697	25	25	0
Only Yesterday	Animation	1991	7.6	1,450,994	118	12296.55788	610,000	26	26	0
The Mummy	Adventure	1999	7.1	1,004,991,630	125	8039933.04	422,500,000	4	5	1
Panic Room	Thriller	2002	6.8	354,914,080	112	3168875.714	197,079,546	13	11	-2
A Bronx Tale	Crime	1993	7.8	38,787,438	121	320557.3393	17,300,000	22	21	-1
Dances with Wolves	Western	1990	8	1,051,496,704	181	5809374.055	424,200,000	3	4	1
Bram Stoker's Dracula	Horror	1992	7.4	498,462,805	128	3894240.662	215,862,692	9	9	0
The Departed	Crime	2006	8.5	469,381,939	151	3108489.664	292,081,509	10	7	-3
Nausicaä of the Valley of the Wind	Animation	1984	8	44,589,575	117	381107.4751	14,300,000	21	22	1
Spirited Away	Animation	2001	8.6	724,461,586	125	5795692.692	395,800,000	6	6	0
Walk the Line	Biography	2005	7.8	310,207,435	136	2280937.02	187,000,000	14	12	-2
Water for Elephants	Drama	2011	6.9	168,657,234	120	1405476.952	117,100,000	17	15	-2
Titanic	Drama	1997	7.9	4,571,479,607	195	23443485.17	2,264,743,305	1	1	0
King Kong	Adventure	2005	7.2	923,831,545	187	4940275.641	556,906,378	5	3	-2

Genre Distribution

This chart shows the distribution of genres among the 26 films in the dataset. Animation is the most frequent genre with 7 films, followed by Drama (5 films). Crime and Thriller are moderately represented, while several genres—such as Action, Biography, Horror, Musical, Sci-Fi, and Western—appear only once.
Overall, the selection is diverse, but it leans most strongly toward Animation and Drama.

Genre	count Genre
Animation	7
Drama	5
Crime	3
Thriller	3
Adventure	2
Action	1
Biography	1
Horror	1
Musical	1
Sci-Fi	1
Western	1

Genre Distribution
- Animation
- Drama
- Crime
- Thriller
- Adventure
- Action
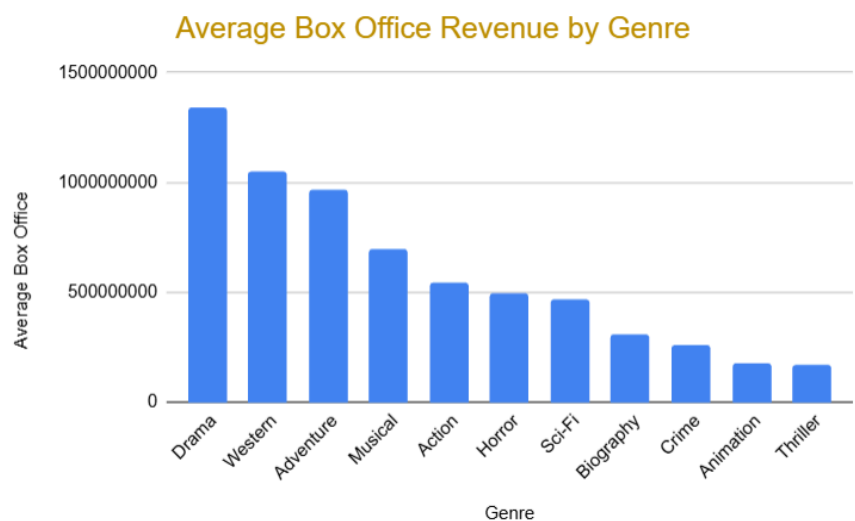- Biography
- Horror
- Musical
- Sci-Fi
- Western

Average Box Office Revenue by Genre

This chart shows the average inflation-adjusted box office revenue for each genre in the dataset. Drama has the highest average revenue, exceeding **$1.3 billion**, followed by Western and Adventure films. Musical and Action genres also perform strongly, while Horror, Sci-Fi, and Biography fall into the mid-range.
At the lower end, Crime, Animation, and Thriller have the lowest average box office performance among the films included.

Overall, the results suggest that large-scale genres such as Drama, Western, and Adventure tend to generate higher revenue, while smaller or niche genres typically earn less.

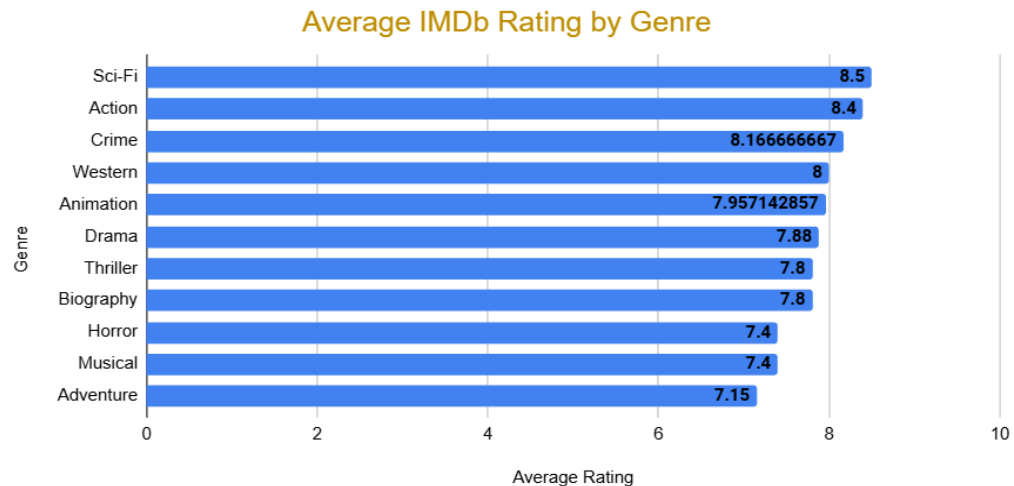| Genre	avg Box Office (adjusted)
Drama	1339731538
Western	1051496704
Adventure	964411587
Musical	698535613
Action	541880765
Horror	498462805
Sci-Fi	468114083
Biography	310207435
Crime	258780080
Animation	175314150
Thriller	168320690

Average IMDb Rating by Genre

This chart shows the average IMDb rating for each genre in the dataset. Sci-Fi has the highest average rating (8.5), followed closely by Action (8.4) and Crime (8.2). Western and Animation also perform strongly with averages around 8.0.
Genres such as Drama, Thriller, and Biography fall in the mid-range, with ratings between **7.8 and 7.9**.
At the lower end, Adventure, Musical, and Horror have the lowest average ratings, scoring between **7.1 and 7.4**.

Overall, the results suggest that genres with strong narrative or imaginative elements, such as Sci-Fi, Action, and Crime, tend to receive higher viewer ratings, while lighter or more niche genres typically receive slightly lower scores.

Genre	avg Rating
Sci-Fi	8.5
Action	8.4
Crime	8.2
Western	8.0
Animation	8.0
Drama	7.9
Thriller	7.8
Biography	7.8
Horror	7.4
Musical	7.4
Adventure	7.2

Average IMDb Rating by Genre

Genre	Average Rating
Sci-Fi	8.5
Action	8.4
Crime	8.166666667
Western	8
Animation	7.957142857
Drama	7.88
Thriller	7.8
Biography	7.8
Horror	7.4
Musical	7.4
Adventure	7.15

Distribution Analysis: Box Office and IMDb Ratings

Box Office Distribution (in millions USD)

The box plot shows a very wide spread in box office revenues after inflation adjustment. The minimum revenue is around **1.45 million**, while the maximum exceeds **4,571 million**, indicating an extremely large range. The interquartile range (IQR) is also wide, showing high variability between films. This suggests that box office performance is highly uneven, with a few very high-grossing films pulling up the upper quartile.
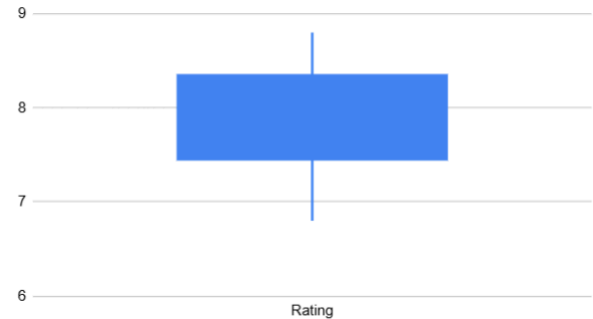
IMDb Rating Distribution

In contrast, the ratings distribution is much tighter.
IMDb ratings range from **6.8 to 8.8**, and most films fall between **7.0 and 8.35**, which creates a narrow IQR. This means that even though the box office revenue varies greatly, the films in the dataset are consistently well-rated.

Release Year Distribution

The box plot for release years shows how the films in the dataset are distributed over time. The earliest film was released in **1975**, while the most recent one came out in **2016**, giving the dataset a **41-year span**.
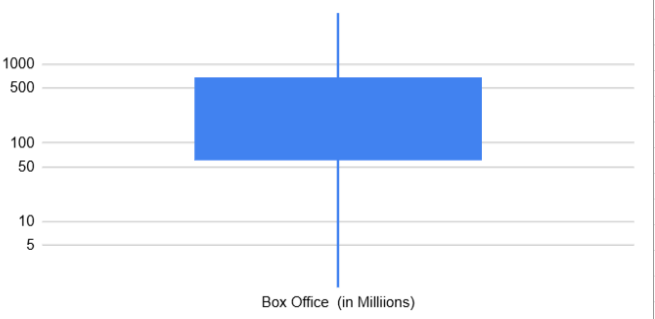
	minimum	lower quartile	upper quartile	maximum
Rating	6.8	7	8.35	8.8

	minimum	lower quartile	upper quartile	maximum
Box Office	1,450,994	63067336	659371901.2	4,571,479,607
Box Office (in Milliions)	1.45099383	63.06733615	659.3719012	4571.479607

Ratings by IMDb

Box Office Revenue Distribution (in millions USD)

	minimum	lower quartile	upper quartile	maximum
Year	1975	1990	2005	2016

Distribution of years
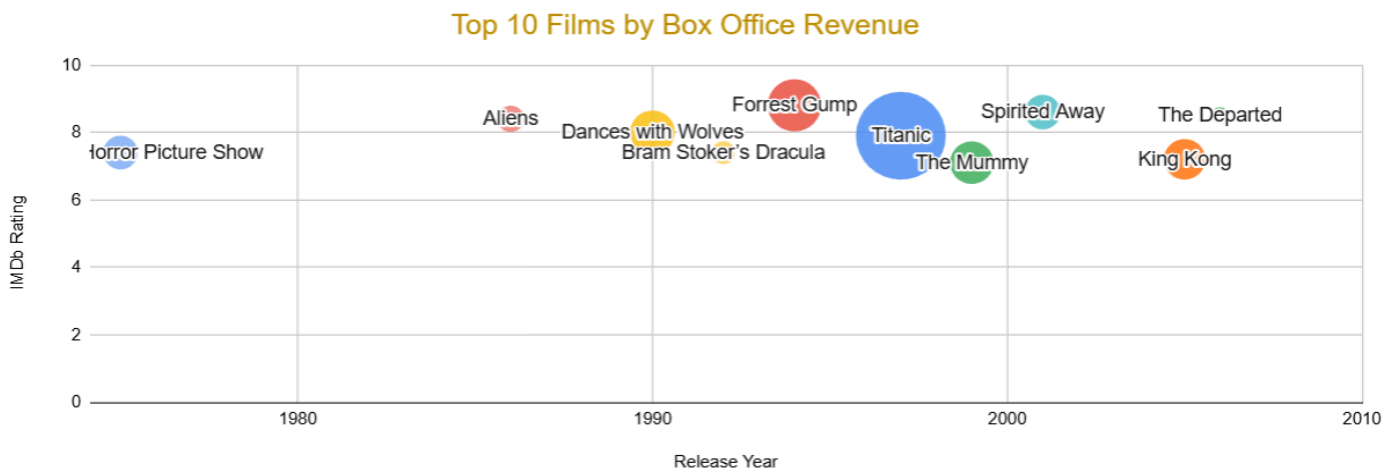

Top 10 Films by Box Office Revenue

This bubble chart shows the ten highest-grossing films in the dataset based on inflation-adjusted box office revenue. Each bubble represents one film, with its size indicating revenue, its horizontal position showing the release year, and its vertical position showing its IMDb rating.

Titanic (1997) stands out clearly as the top performer, with by far the largest bubble, indicating exceptional box office success. Other strong performers include **Forrest Gump (1994)**, **Dances with Wolves (1990)**, and **The Mummy (1999)**, all of which achieved high revenues along with solid IMDb ratings.

Most of the top box office films were released between **1990 and 2006**, showing that this period dominates the upper end of the revenue rankings. Older films like *The Rocky Horror Picture Show* (1975) and *Aliens* (1986) appear as smaller bubbles, showing that their adjusted revenue—while notable—is still much lower compared to major releases from the 1990s and early 2000s.

Overall, the chart suggests that films with strong audience appeal and high production values, especially from the 1990–2006 period, tend to achieve both high IMDb ratings and strong box office performance.

Title	Year	Rating	Box Office (adjusted)
Titanic	1997	7.9	4,571,479,607
Forrest Gump	1994	8.8	1,482,652,478
Dances with Wo	1990	8	1,051,496,704
The Mummy	1999	7.1	1,004,991,630
King Kong	2005	7.2	923,831,545
Spirited Away	2001	8.6	724,461,586
The Rocky Horr	1975	7.4	698,535,613
Aliens	1986	8.4	541,880,765
Bram Stoker's D	1992	7.4	498,462,805
The Departed	2006	8.5	469,381,939

Top 10 Films by Box Office Revenue
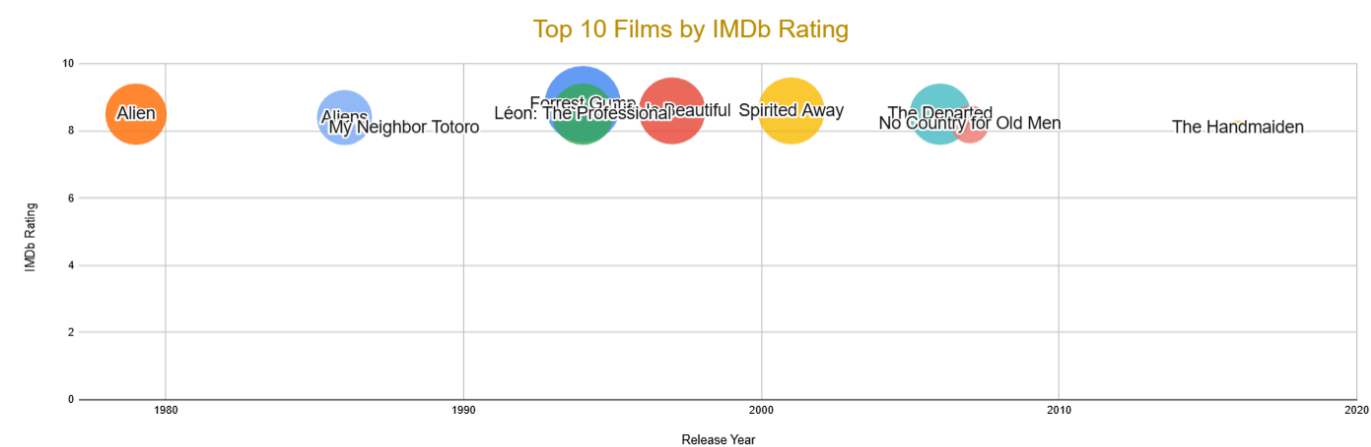

Top 10 Films by IMDb Rating

This chart presents the ten highest-rated films in the dataset based on IMDb scores. The ratings range from **8.1 to 8.8**, showing that all selected films are critically well-received.

Several films stand out with exceptionally high scores, including **Forrest Gump (1994)**, **Life is Beautiful (1997)**, and **Spirited Away (2001)**, all rated at **8.6 and above**. Classic films such as **Alien (1979)** and **Léon: The Professional (1994)** also rank highly.

Most of the top-rated films were released between **1988 and 2007**, indicating that this period produced many of the most acclaimed titles in the dataset. Despite spanning different genres and decades, the IMDb scores remain closely clustered, suggesting a consistently high level of critical appreciation.

Overall, the chart highlights how a variety of film styles and eras can achieve strong viewer ratings, with no single genre or decade dominating the top of the ranking.

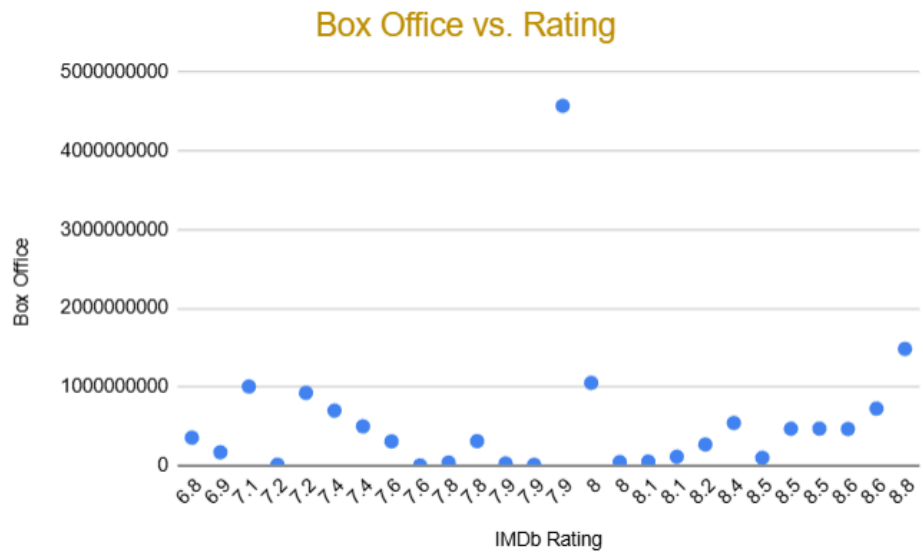| Title	Year	Rating
Forrest Gump	1994	8.8
Life Is Beautiful	1997	8.6
Spirited Away	2001	8.6
Léon: The Profe:	1994	8.5
Alien	1979	8.5
The Departed	2006	8.5
Aliens	1986	8.4
No Country for C	2007	8.2
The Handmaidei	2016	8.1
My Neighbor Tot	1988	8.1

Top 10 Films by IMDb Rating

Box Office vs. IMDb Rating

This scatter plot compares each film's IMDb rating with its inflation-adjusted box office revenue. The distribution shows **no strong correlation** between rating and revenue: films with similar ratings can have very different box office results.

While most films fall below the 200 million range regardless of rating, a few titles stand out as major outliers. For example, *Titanic* (7.9 rating) achieved exceptionally high box office revenue compared to other films with similar or even higher ratings. This demonstrates that **commercial success does not necessarily reflect IMDb rating**.

The chart indicates that audience ratings remain relatively stable, mostly between 7.0 and 8.8, while box office earnings vary dramatically. Overall, high ratings do not guarantee high box office revenue, and some moderately rated films can still become major box office hits.

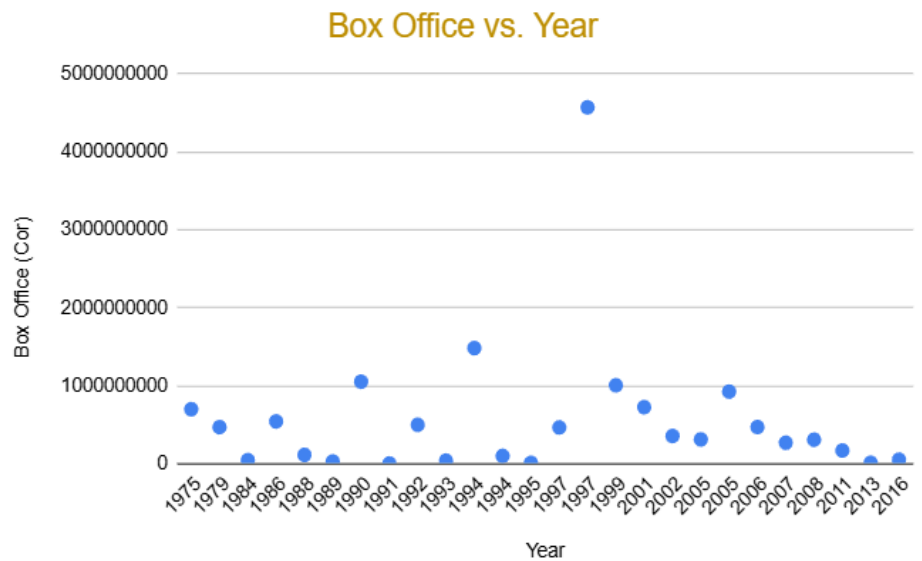| Rating	Box Office (adjusted)
6.8	354,914,080
6.9	168,657,234
7.1	1,004,991,630
7.2	11,403,835
7.2	923,831,545
7.4	698,535,613
7.4	498,462,805
7.6	307,229,238
7.6	1,450,994
7.8	38,787,438
7.8	310,207,435
7.9	27,218,217
7.9	9,756,901
7.9	4,571,479,607
8	1,051,496,704
8	44,589,575
8.1	51,110,677
8.1	112,492,541
8.2	268,170,861
8.4	541,880,765
8.5	98,937,312
8.5	468,114,083
8.5	469,381,939
8.6	464,464,535
8.6	724,461,586
8.8	1,482,652,478

Box Office vs. Rating

Box Office vs. Year

This scatter plot shows how box office revenue (adjusted for inflation) is distributed across release years. The data spans from **1975 to 2016**, and the plot reveals a very uneven pattern.

A few films stand out as extremely high-grossing outliers—for example, *Titanic* (1997) and *Forrest Gump* (1994)—far above the rest of the dataset. Most films, however, fall well below the 200 million range, regardless of release year.

There is **no clear trend of increasing or decreasing box office revenue over time**. Both older and newer films appear at very different revenue levels, suggesting that a film's release year alone does not predict its financial success. Instead, revenue seems to depend more on individual film popularity, marketing, and cultural impact.

Year	Box Office (adjusted)
1975	698,535,613
1979	468,114,083
1984	44,589,575
1986	541,880,765
1988	112,492,541
1989	27,218,217
1990	1,051,496,704
1991	1,450,994
1992	498,462,805
1993	38,787,438
1994	1,482,652,478
1994	98,937,312
1995	9,756,901
1997	464,464,535
1997	4,571,479,607
1999	1,004,991,630
2001	724,461,586
2002	354,914,080
2005	310,207,435
2005	923,831,545
2006	469,381,939
2007	268,170,861
2008	307,229,238
2011	168,657,234
2013	11,403,835
2016	51,110,677

Box Office vs. Year

Rank Changes after Inflation Adjustment

This chart shows how each film's box office ranking changed after adjusting all revenues for inflation. A positive value means the film moved *up* in the ranking, while a negative value means it moved *down* compared to its original position.
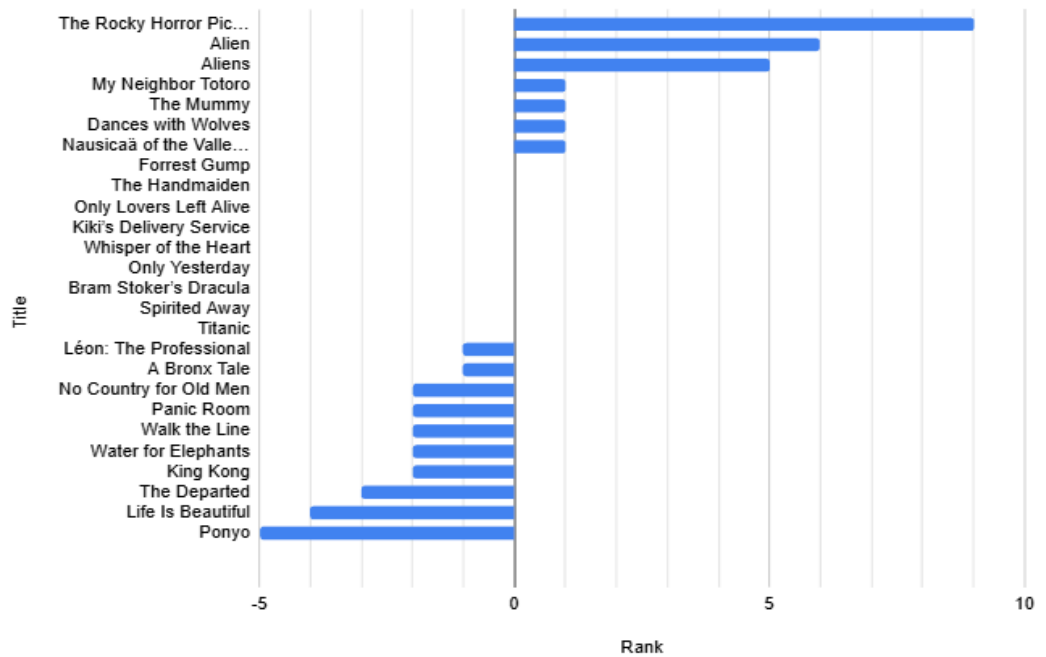
Several older films experienced the largest upward shifts. The strongest improvement is seen in **The Rocky Horror Picture Show**, which jumped **9 positions**, followed by **Alien** (+6) and **Aliens** (+5). This indicates that once inflation is taken into account, these older titles earned significantly more relative to the others than their non-adjusted numbers suggested.

More recent films generally declined in rank. Movies like **Ponyo**, **Life Is Beautiful**, **The Departed**, and **King Kong** dropped several positions, showing that their box office success was less impressive when compared on equal economic terms.

Overall, the analysis demonstrates that **inflation adjustment tends to benefit older films**, highlighting their true financial impact, while newer films lose some of their relative standing.

Title	Rank Change
The Rocky Horro	9
Alien	6
Aliens	5
My Neighbor Tot	1
The Mummy	1
Dances with Wol	1
Nausicaä of the	1
Forrest Gump	0
The Handmaider	0
Only Lovers Left	0
Kiki's Delivery Se	0
Whisper of the H	0
Only Yesterday	0
Bram Stoker's D	0
Spirited Away	0
Titanic	0
Léon: The Profes	-1
A Bronx Tale	-1
No Country for C	-2
Panic Room	-2
Walk the Line	-2
Water for Elepha	-2
King Kong	-2
The Departed	-3
Life Is Beautiful	-4
Ponyo	-5

Rank Changes after Inflation Adjustment to Box Office

Conclusion

This analysis shows that although the films in the dataset have consistently high IMDb ratings, their box office performance varies widely. Genres like Sci-Fi, Action, and Crime receive the highest ratings, while Drama, Western, and Adventure earn the strongest revenues. There is no clear correlation between rating and revenue—successful films can perform well critically, financially, or both. Inflation adjustment also reveals that several older films rise significantly in rank, showing their lasting impact. Overall, the dataset highlights how differently films can succeed depending on genre, audience appeal, and historical context.

Link to analysis