**Project Goals**

* I started this project curious to see what films were the well received, currently popular, and did well in the box office.
* How the box office has changed over time
* If there were any contributing factors in a film’s popularity/ critical reception

**Criteria**

I had gathered the data of about 9,000 movies who met at least 1 of the 6 following criteria:

1. Films that were in the database for box office mojo. These were films that:
   1. Were released in 1977 or later. The database unfortunately did not have any financial data for any years earlier.
   2. Films that were among the top 200 grossing films of its year.
2. Films that were among the top 250 highest rated films on Letterboxd (with over 25,000 watches).
3. Films that were among the top 250 highest rated films on IMDb (with over 25,000 votes).
4. Films that had over a million Letterboxd users who had watched the film
5. The 250 films that were featured on a Letterboxd profile as one of their favorite films (each member gets 4).
6. I had seen the film and logged it on Letterboxd.

**Notable Films**

The following films were on all 4 letterboxd/imdb lists and made over a hundred million dollars in the box office. The Pianist was the only film to fit all the criteria that I had not seen. Grave of the Fireflies and the Before Sunrise were 2 films that I had not seen and they had met 4 of the 5 other criteria.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| movie\_name | year | Worldwide | Box\_Office\_Rank | Milestone | rating |
| The Lord of the Rings: The Fellowship of the Ring | 2001 | $ 880,839,846 | 2 | Half Billion | 5 |
| The Lord of the Rings: The Two Towers | 2002 | $ 936,689,735 | 1 | Half Billion | 5 |
| The Lord of the Rings: The Return of the King | 2003 | $ 1,140,682,011 | 1 | Billion | 5 |
| The Departed | 2006 | $ 291,465,034 | 14 | Hundred Million | 5 |
| No Country for Old Men | 2007 | $ 171,627,166 | 32 | Hundred Million | 5 |
| The Dark Knight | 2008 | $ 1,003,045,358 | 1 | Billion | 5 |
| Parasite | 2019 | $ 258,710,043 | 31 | Hundred Million | 5 |
| Spider-Man: Across the Spider-Verse | 2023 | $ 690,542,303 | 6 | Half Billion | 5 |
| Fight Club | 1999 | $ 100,853,753 | 39 | Hundred Million | 5 |
| Good Will Hunting | 1997 | $ 138,435,239 | 11 | Hundred Million | 5 |
| Pulp Fiction | 1994 | $ 108,143,702 | 10 | Hundred Million | 5 |
| The Silence of the Lambs | 1991 | $ 130,875,105 | 4 | Hundred Million | 5 |
| Star Wars: Episode V - The Empire Strikes Back | 1980 | $ 209,398,025 | 1 | Hundred Million | 5 |
| Spider-Man: Into the Spider-Verse | 2018 | $ 375,540,831 | 25 | Hundred Million | 4.5 |
| Terminator 2: Judgment Day | 1991 | $ 205,068,630 | 2 | Hundred Million | 4.5 |
| Dune: Part Two | 2024 | $ 714,844,358 | 7 | Half Billion | 4.5 |
| The Green Mile | 1999 | $ 286,801,374 | 10 | Hundred Million | 4.5 |
| Inglourious Basterds | 2009 | $ 321,455,689 | 20 | Hundred Million | 4.5 |
| Django Unchained | 2012 | $ 425,368,238 | 16 | Hundred Million | 4.5 |
| Prisoners | 2013 | $ 122,126,687 | 58 | Hundred Million | 4.5 |
| Interstellar | 2014 | $ 681,070,397 | 10 | Half Billion | 4.5 |
| The Prestige | 2006 | $ 109,676,311 | 54 | Hundred Million | 4.5 |
| Howl's Moving Castle | 2004 | $ 235,347,345 | 17 | Hundred Million | 4.5 |
| Spirited Away | 2001 | $ 274,925,095 | 15 | Hundred Million | 4 |
| The Grand Budapest Hotel | 2014 | $ 172,936,941 | 49 | Hundred Million | 4 |
| Dead Poets Society | 1989 | $ 175,480,949 | 3 | Hundred Million | 4 |
| The Pianist | 2002 | $ 120,072,577 | 38 | Hundred Million | NULL |

The following movies were released before 1977 (the first year in our data to show box office earnings) but met all other criteria.

|  |  |  |
| --- | --- | --- |
| movie\_name | year | rating |
| 2001: A Space Odyssey | 1968 | 3.5 |
| Psycho | 1960 | 4 |
| The Godfather | 1972 | 4.5 |
| One Flew Over the Cuckoo's Nest | 1975 | 5 |
| 12 Angry Men | 1957 | 4.5 |
| The Godfather Part II | 1974 | 4.5 |

These films did not initially make a hundred million but did so after adjusting for inflation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **movie\_name** | **year** | **Adjusted\_Worldwide** | **Box\_Office\_Rank** | **rating** |
| Alien | 1979 | $ 368,251,281.50 | 5 | 4.5 |
| Apocalypse Now | 1979 | $ 368,148,534.40 | 6 | 4.5 |
| Schindler's List | 1993 | $ 213,991,253.11 | 9 | 4.5 |
| The Shining | 1980 | $ 181,407,286.52 | 14 | 4.5 |
| Heat | 1995 | $ 142,508,797.36 | 27 | 4.5 |
| Eternal Sunshine of the Spotless Mind | 2004 | $ 126,950,944.94 | 73 | 4 |
| There Will Be Blood | 2007 | $ 119,314,670.81 | 73 | 4.5 |
| Goodfellas | 1990 | $ 117,482,420.72 | 26 | 4.5 |

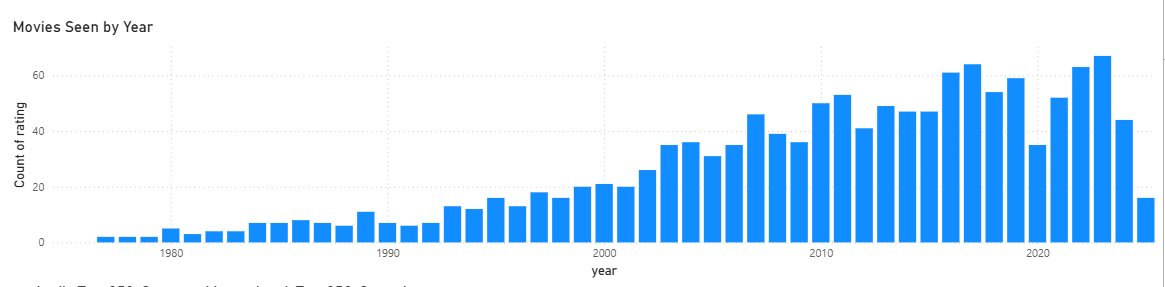
I wanted to see how many films I had seen in each of the lists and how I enjoyed them. I had not seen most of the films on the letterboxd top 250 (172 films or 68%). This being due to the average film of the letterboxd top 250.

|  |  |  |
| --- | --- | --- |
| List | Average Year | Percentage Seen |
| Letterboxd | 1981 | 31.20% |
| IMDb | 1988 | 70.40% |
| Top Favorited Letterboxd | 2001 | 78.00% |
| Million Watches | 2008 | 89.08% |

**Ratings**

A group of pie charts

AI-generated content may be incorrect.



**Yearly Grossing**

Yearly grossing was gradually increasing until covid hit in 2020 which caused it to suddenly plummet. However, it’s been gradually increasing again since then.

A screen shot of a graph

AI-generated content may be incorrect.

I wanted to see if a film making one of the 4 lists had anything to do with box office performance or release year. There was virtually no relationship between release year and being on the imdb or letterboxd top 250 list. There was no correlation between box office rank and the Letterboxd top 250 list. There was a mild correlation between box office rank and letterboxd top fans and being in the imdb top 250. The strongest relationship with box office rank was with the Letterboxd million watches list.

A screenshot of a graph

AI-generated content may be incorrect.

**Milestone**

Films with no box office data were on average the oldest. This makes sense a lot of these films had no data because they were before 1977. Films that made over a billion were on average much more recently. From that point on the greater the category, the more recent the average year became.

|  |  |  |
| --- | --- | --- |
| Milestone | avg\_year | movie\_count |
| No Gross Data | 1992 | 449 |
| Under Hundred Million | 2001 | 6806 |
| Hundred Million | 2009 | 1482 |
| Half Billion | 2013 | 192 |
| Billion | 2016 | 55 |

**Regression Analysis**

I did a regression analysis to see the impact year had on gross and another to see the impact having a million watches on letterboxd had on adjusted gross. For the year regression, saw that there was a decently strong correlation due to the small p-value of 9.2E-11 and the fairly high multiple R of .78. I found that this matched the data and its trends. With the exception of 2020(covid-19), each year was making ethe same amount of money or more than the last. The regression analysis on the films’ gross in relation to if they had a million watches on letterboxd did not yield the same results. The multiple R was much lower at a number of .49 and the p value was exactly 0. The p-value alone tells me that it was insufficient data to make a correlation. This probably being due to the dependent being a boolean value and there were not as many million watched films as there were in the database (568 out of 8984)

**Findings**

* Letterboxd users (including myself) watched more recent movies.
* Using the criteria mentioned, I should watch The Pianist, Grave of the Fireflies, and Before Sunrise.
* The box office was increasing in gross until covid (this is even adjusting for inflation).
* While release year had some relationship with the amount of money earned and the current popularity of a film, it did not have any impact on whether a film was to be highly rated.
* Films that were rated higher were much more on average much older than those that were the most seen/popular.

The following are additional data that is not currently in the dataset that I would like to add in the future:

* Average Letterboxd/IMDb ratings
* Genres
* Cast/Crew
* More films that didn’t fit the 6 criteria