Camera, Lights, Action!

Foundations in Data Science - Group 4 Term Project

Virginia Massignan
Oleksandr (Alex) Rud
Joshua Au-Yeung
Nahid Bhoja

The problem

Using the TMDB (The Movie Database) Movie dataset, we ventured to understand historical success trends in the film industry and model how these trends can predict future successes

TMDB Movie Dataset v11:

- 1.18 million movies identified from 1860 - future non-released movies still in production
- Dataset includes budget, viewer ratings, popularity, revenue, genres, production locations, languages etc

Language-based Analysis of Movie Popularity

Virginia

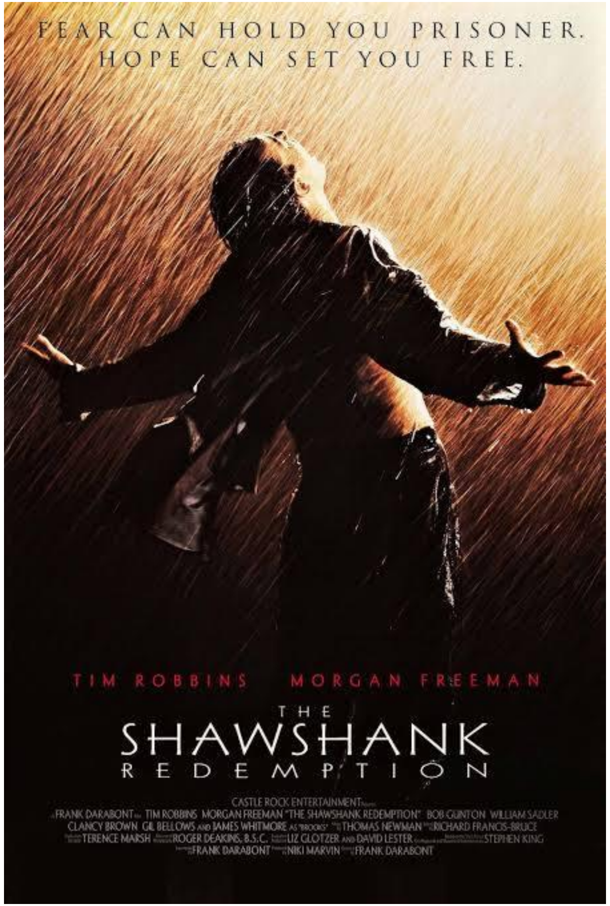
Categorizing Popularity

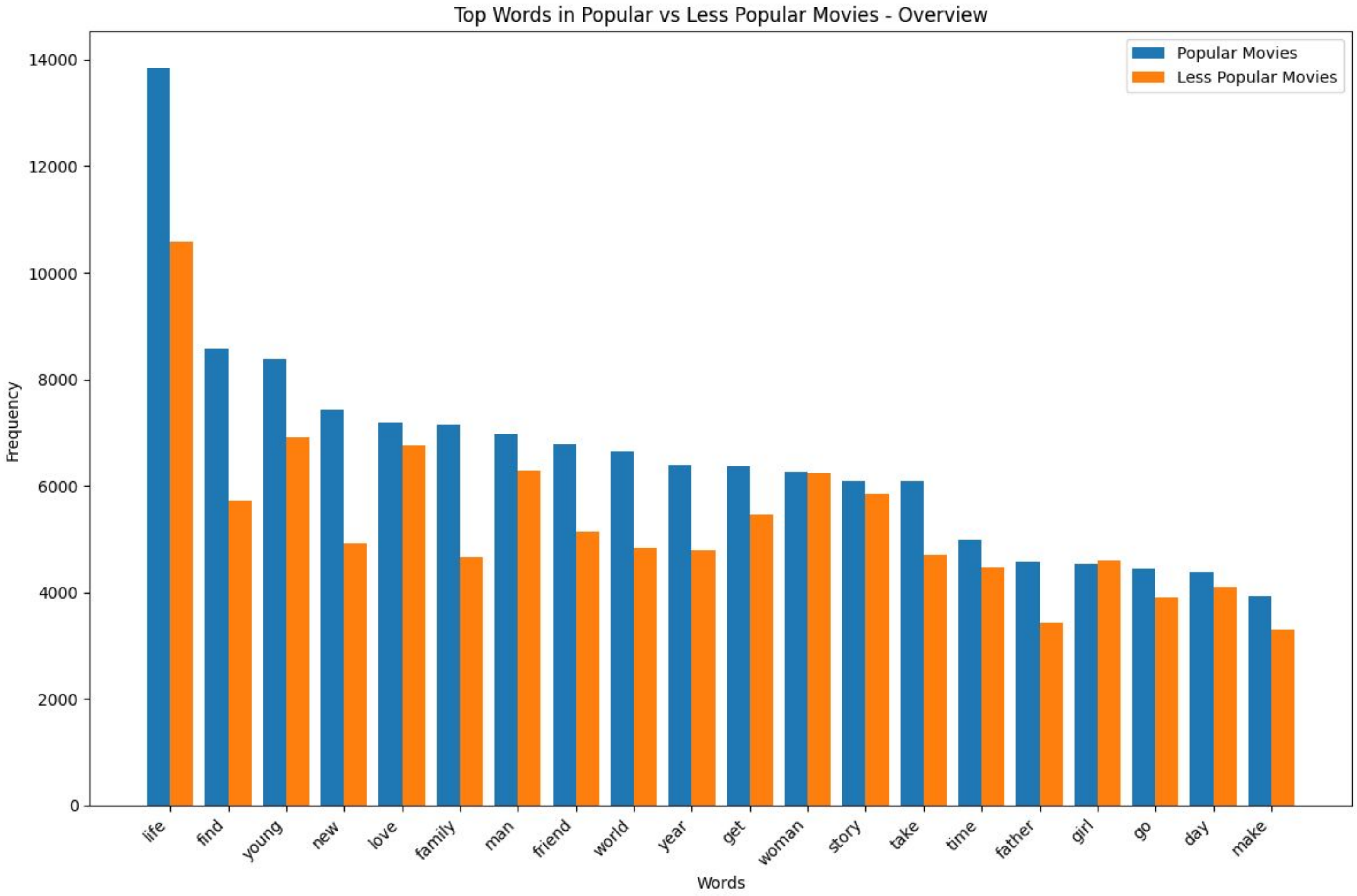
Dataset: Movies with both an overview and a tagline.

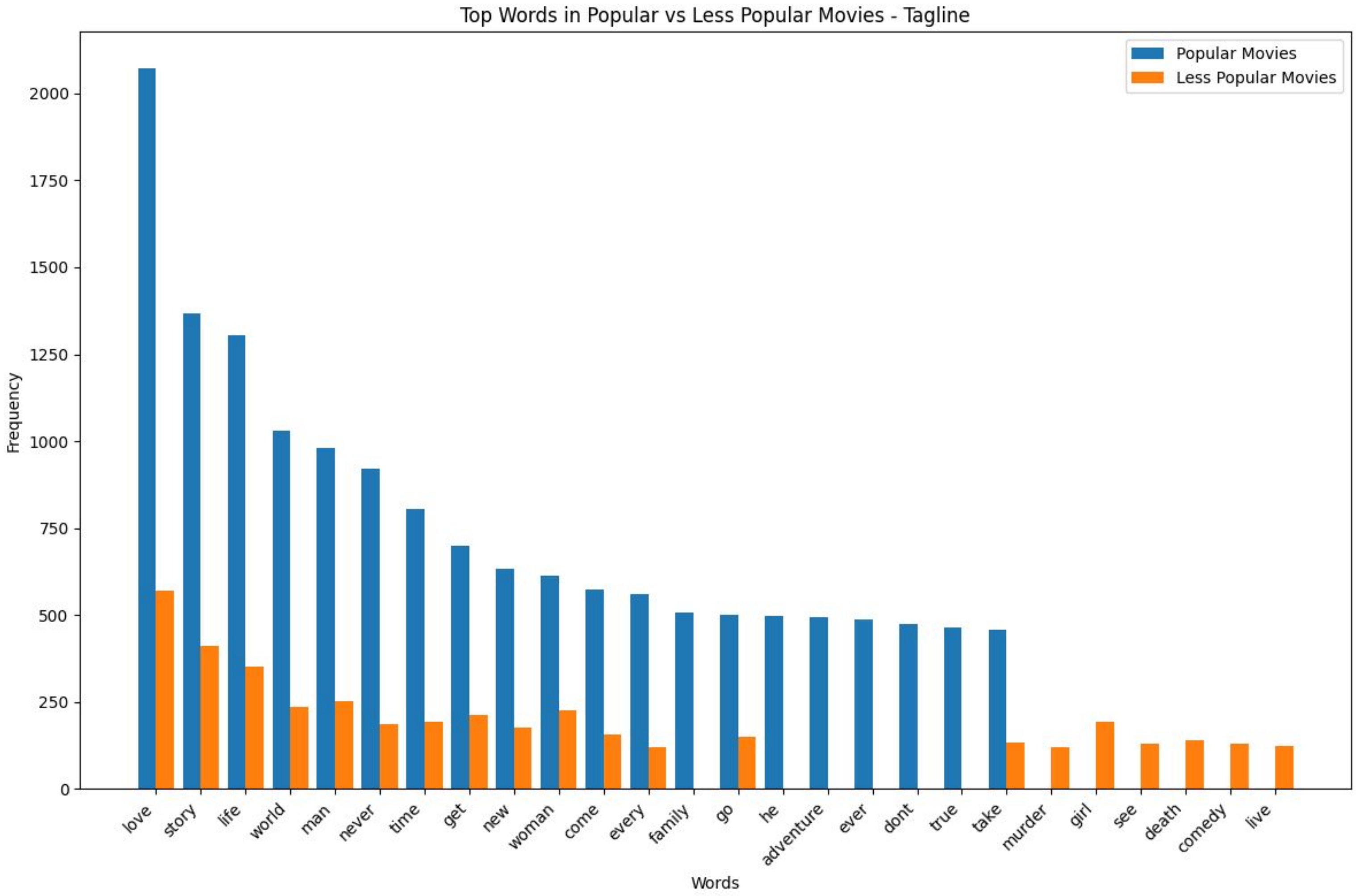
Computed a **Weighted Vote Score**: product of vote average and the logarithm of vote count.

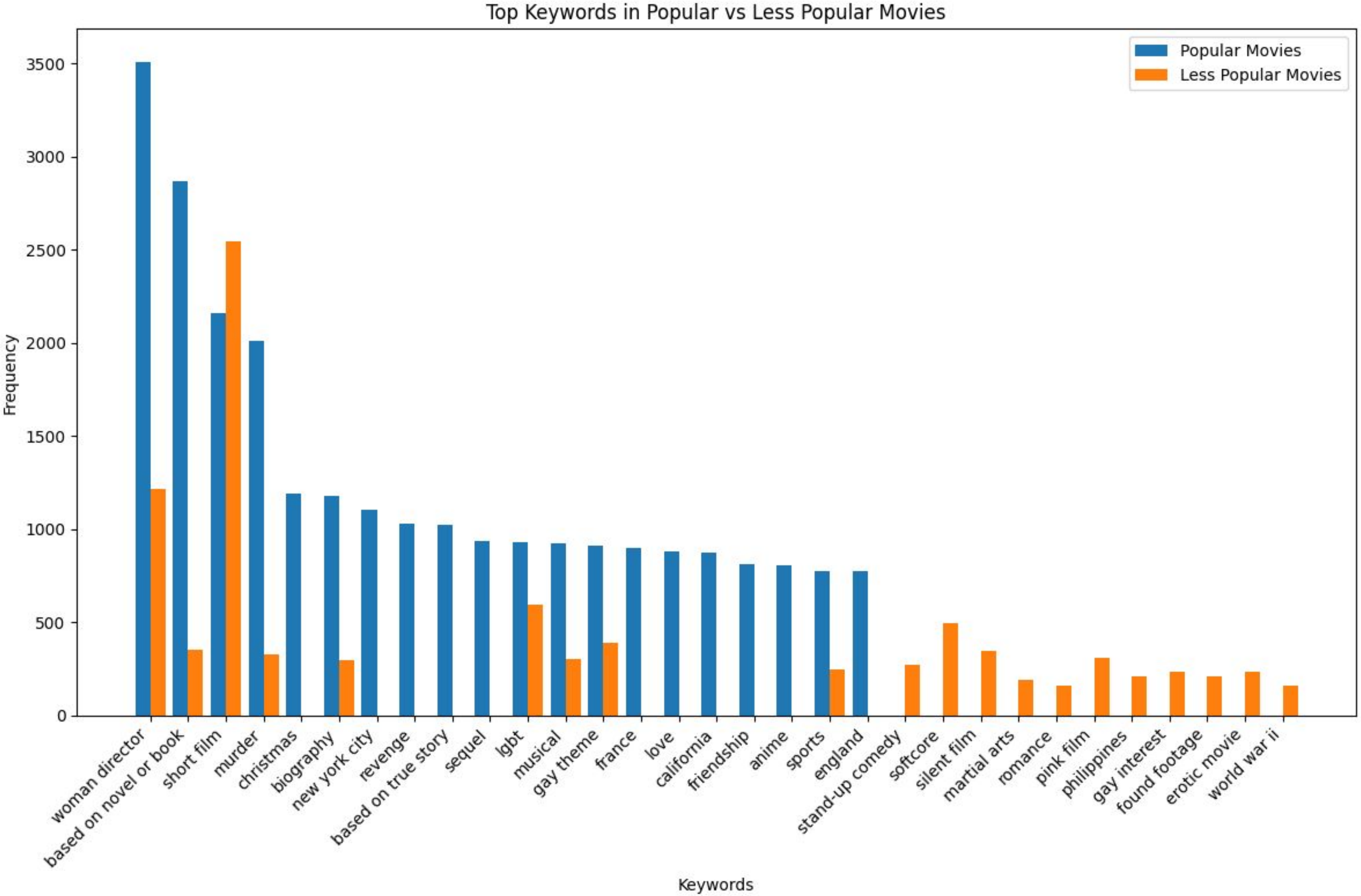
Categorized movies into:

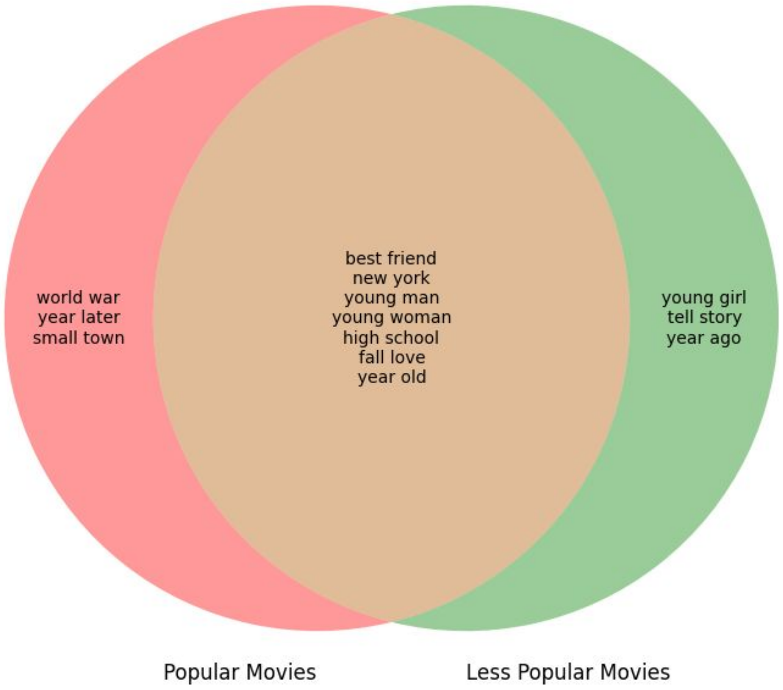
- Popular: Top 20%.
- Moderately Popular: Middle 60%.
- Less Popular: Bottom 20%.


FEAR CAN HOLD YOU PRISONER.
HOPE CAN SET YOU FREE.
TIM ROBBINS
MORGAN FREEMAN
THE
SHAWSHANK
REDEMPTION
CASTLE ROCK ENTERTAINMENT
FRANK DARABONT TIM ROBBINS MORGAN FREEMAN "THE SHAWSHANK REDEMPTION" BOB GUNTON WILLIAM SADLER
CLANCY BROWN GIL BELLOWS AND JAMES WHITMORE AS "BROOKS" THOMAS NEWMAN RICHARD FRANCIS-BRUCE
TERENCE MARSH ROGER DEAKINS, B.S.C. LIZ GLOTZER AND DAVID LESTER STEPHEN KING
FRANK DARABONT NIKI MARVIN FRANK DARABONT

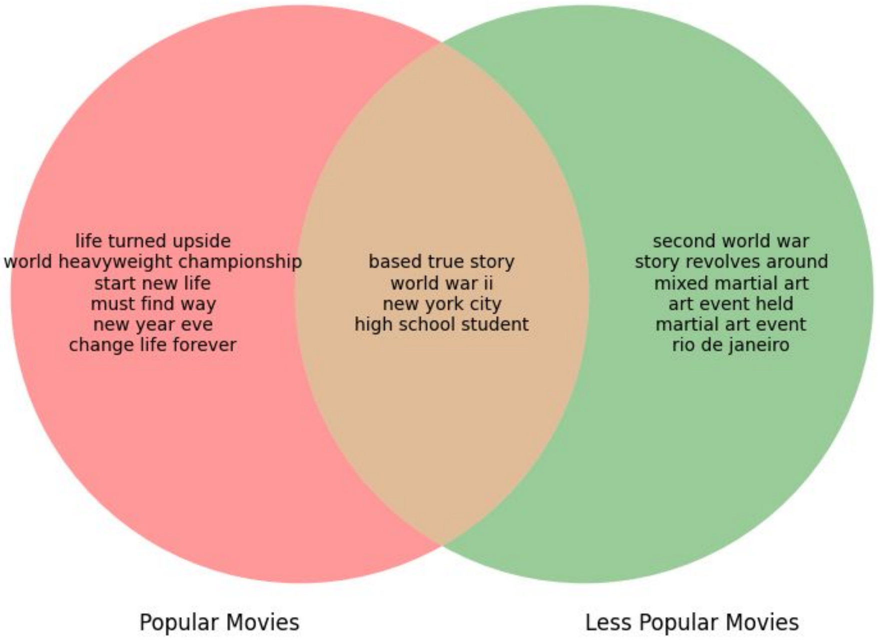

Top Words in Popular vs Less Popular Movies - Overview
Popular Movies
Less Popular Movies
Frequency
0
2000
4000
6000
8000
10000
12000
14000
life
find
young
new
love
family
man
friend
world
year
get
woman
story
take
time
father
girl
go
day
make
Words


Top Words in Popular vs Less Popular Movies - Tagline
Popular Movies
Less Popular Movies
Frequency
0
250
500
750
1000
1250
1500
1750
2000
love
story
life
world
man
never
time
get
new
woman
come
every
family
go
he
adventure
ever
dont
true
take
murder
girl
see
death
comedy
live
Words

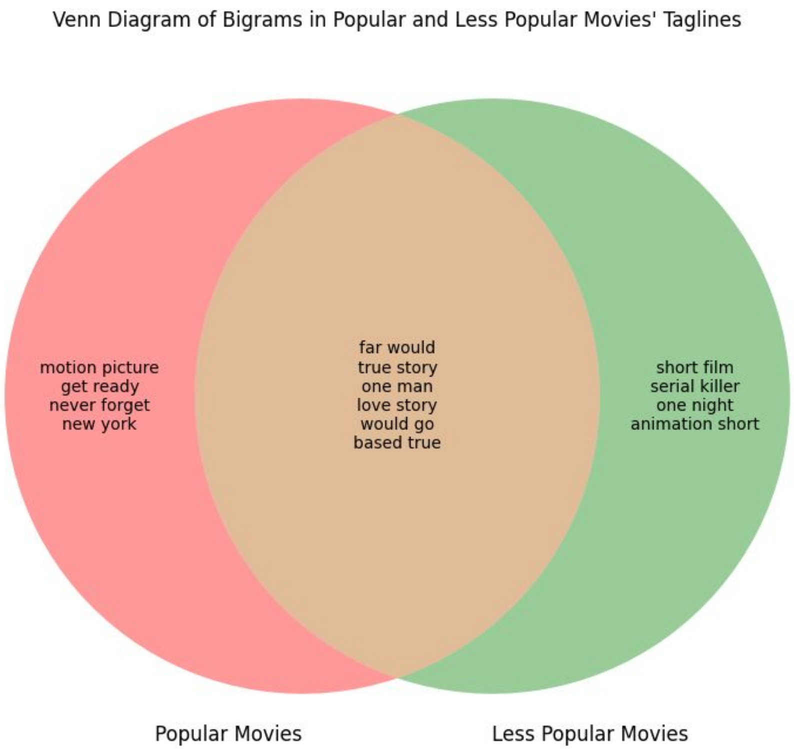

Top Keywords in Popular vs Less Popular Movies
Popular Movies
Less Popular Movies
Frequency
0
500
1000
1500
2000
2500
3000
3500
Keywords
woman director
based on novel or book
short film
murder
christmas
biography
new york city
revenge
based on true story
sequel
lgbt
musical
gay theme
france
love
california
friendship
anime
sports
england
stand-up comedy
softcore
silent film
martial arts
romance
pink film
philippines
gay interest
found footage
erotic movie
world war ii

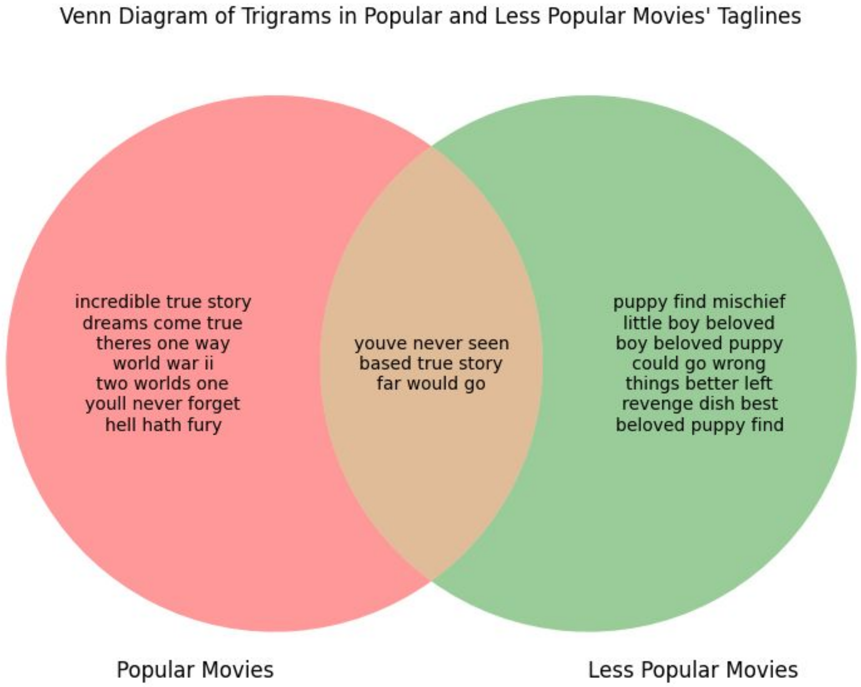


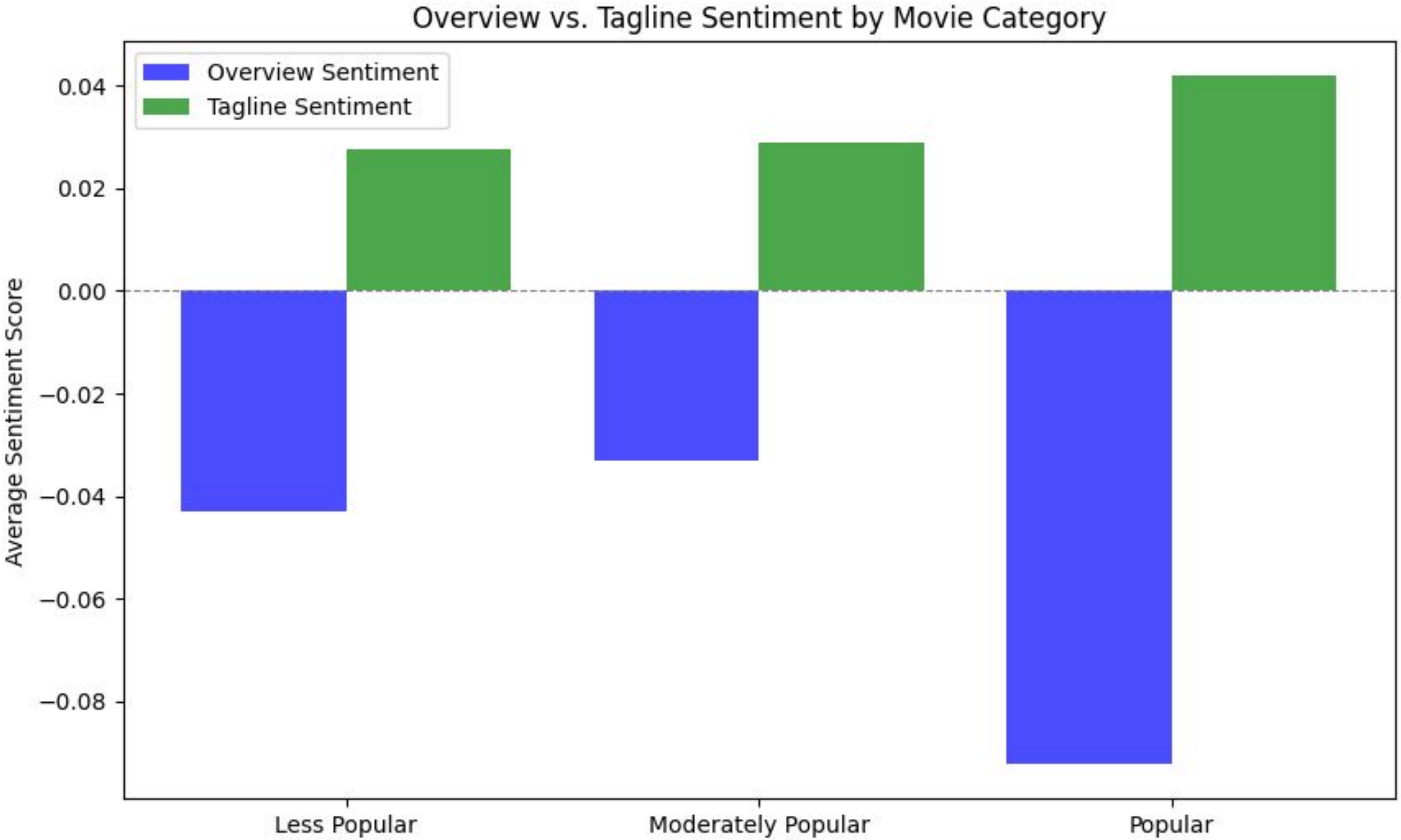
Venn Diagram of Bigrams in Popular and Less Popular Movies' Overview



Venn Diagram of Trigrams in Popular and Less Popular Movies' Overview


Venn Diagram of Bigrams in Popular and Less Popular Movies' Taglines
motion picture
get ready
never forget
new york
far would
true story
one man
love story
would go
based true
short film
serial killer
one night
animation short
Popular Movies
Less Popular Movies


Venn Diagram of Trigrams in Popular and Less Popular Movies' Taglines
incredible true story
dreams come true
theres one way
world war ii
two worlds one
youll never forget
hell hath fury
youve never seen
based true story
far would go
puppy find mischief
little boy beloved
boy beloved puppy
could go wrong
things better left
revenge dish best
beloved puppy find
Popular Movies
Less Popular Movies


Overview vs. Tagline Sentiment by Movie Category
Overview Sentiment
Tagline Sentiment
Average Sentiment Score
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.08
Less Popular
Moderately Popular
Popular

Case Study: Sentiment Patterns in Top 20 Movies

- Highlight the contrast between negative overviews and positive taglines in popular movies.
- Example: The Joker (-0.934 overview, 0.572 tagline).
- Tagline: “Put on a happy face.”
- Overview: “During the 1980s, a failed stand-up comedian is driven insane and turns to a life of crime and chaos in Gotham City while becoming an infamous psychopathic crime figure.”

Conclusion: The Power of Language

- Popular movies combine intense narratives with emotionally engaging marketing.
- Less popular movies have consistent but milder tones
- Text analysis complements quantitative metrics, revealing why audiences are drawn to movies

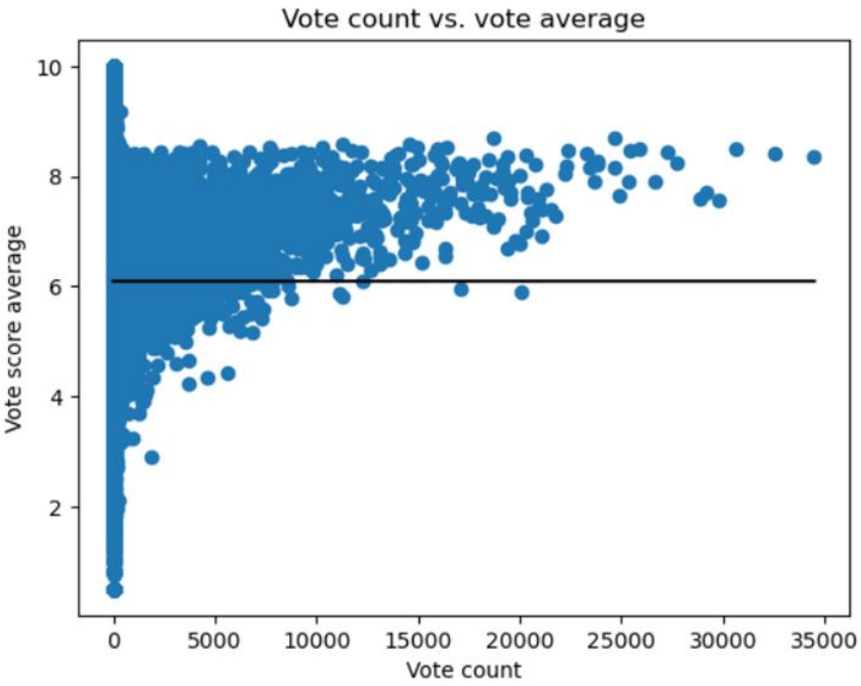
Analysis of the audience voting activity

Oleksandr

Voting activity overview

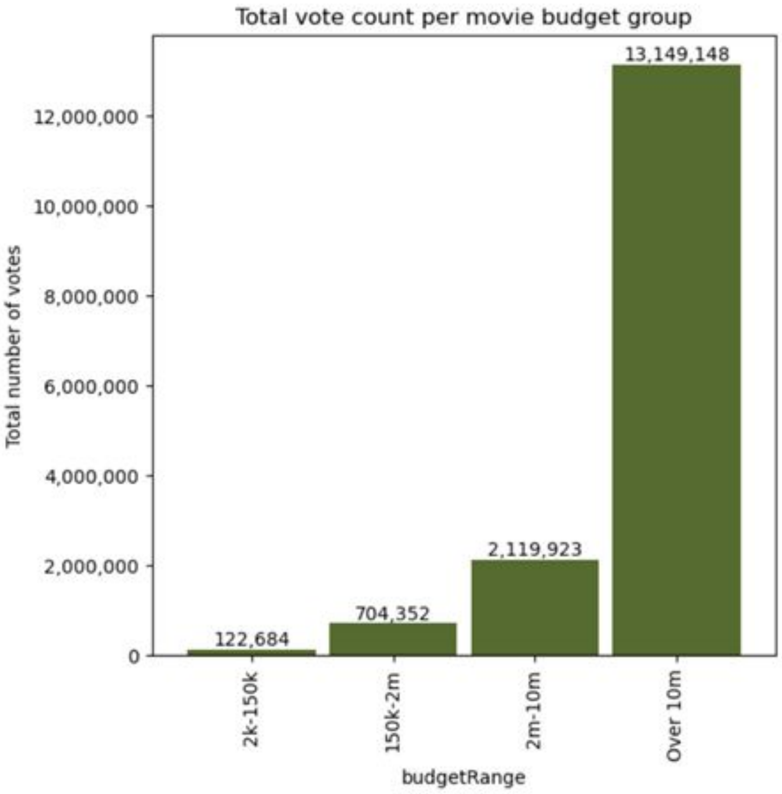
- What factors influence the number of votes and average voting scores?
- Source data - TMDB dataset with 1.1 million movies
- Columns used for calculations - *vote_count* and *vote_average*
- Data cleanup - remove records where *vote_count* and/or *vote_average* is null or 0
- Resulting dataset after cleanup contains 350k movies
- Questions to answer for the analysis:
    - Is there a correlation between vote counts and the average scores?
    - Does vote count and average score depend on the movie budget?
    - Does the movie runtime have an impact on its voting count and average score?

Vote Count vs. Average Vote Score



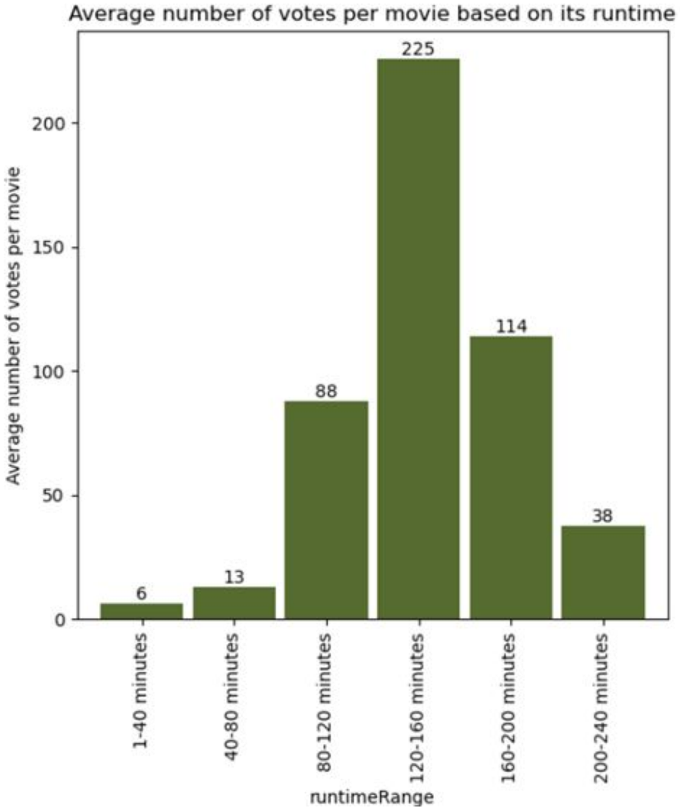
- Movies with higher vote count are unlikely to receive low average score
- Movies with vote count over 10k receive score of at least 5.6 or higher
- Low average scores are assigned only to movies with low or medium vote count (1k or less)
- Linear regression line is not influenced by the vote count
- Very high density of movies in the left sector defines almost 0 incline of the regression line

Audience activity and movie budget



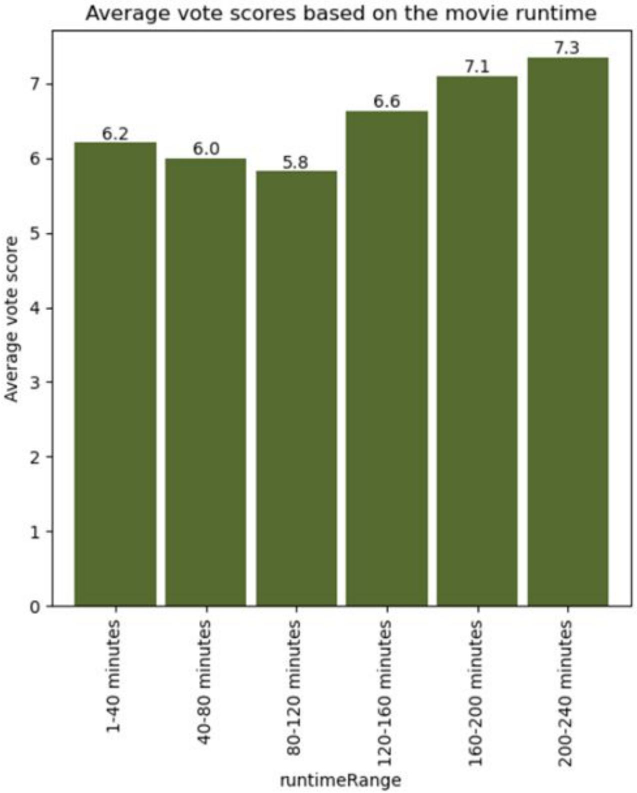
- Split movies into equal budget groups (qcut function)
- Around 5.5k movies in each budget group
- The group of movies with budget over 10m is an absolute leader in vote counts - 13.1m votes (2.2k votes per movie)
- Movies with budgets below 150k are least voted - 122.7k of votes per group (or 23 votes per movie).
- A higher budget movie receives 95x votes compared to a movie in low budget group

Movie runtime and vote counts



- Split movies into 6 groups based on their runtime
- Large difference in vote counts depending on movie runtime
- Most voted movies (225 votes) have runtime between 120 and 160 minutes
- Movies with runtimes between 1 and 40 minutes have lowest vote counts (avg. 6 votes per movie)

Movie runtime and average score



- Some differences in the average movie score depending on the runtime
- Long movies on average get higher scores.
- Movies in runtime range between 200 and 240 minutes on average receive score of 7.3
- The lowest average score (5.8) is received by movies with runtime between 80 and 120 minutes.

Relationship between budget, revenue and popularity to determine movie success

Joshua

Movie Success = Profit?

Profit of $1 million…success?

Profit margin = (revenue – budget) / budget x 100%

	Title	Revenue	Budget	Profit	Profit Margin
0	Inception	825,532,764	160,000,000	665,532,764	4.159580
1	Interstellar	701,729,206	165,000,000	536,729,206	3.252904
2	The Dark Knight	1,004,558,444	185,000,000	819,558,444	4.430046
3	Avatar	2,923,706,026	237,000,000	2,686,706,026	11.336312
4	The Avengers	1,518,815,515	220,000,000	1,298,815,515	5.903707

Column	count	unique	top	freq	mean	std	min	25%	50%	75%	max
title	1,101,128	947,634	Home	152	NaN	NaN	NaN	NaN	NaN	NaN	NaN
revenue	1,101,139	NaN	NaN	NaN	704,848.5	18,026,270	-12	0	0	0	3,000,000,000
budget	1,101,139	NaN	NaN	NaN	267,276.4	4,963,690	0	0	0	0	1,000,000,000
vote_average	1,101,139	NaN	NaN	NaN	1.938	3.048	0	0	0	5	10
popularity	1,101,139	NaN	NaN	NaN	1.248	7.703	0	0.6	0.6	0.873	2,994.357
profit	1,101,139	NaN	NaN	NaN	437,572.1	15,168,740	-1,000,000,000	0	0	0	2,780,000,000
profit_margin	61,778	NaN	NaN	NaN	∞	NaN	-2	-1	-1	0	∞

Cleaning Data

We don’t know exactly what films in our dataset are bad without looking individually at films – an unfeasible task, but what we can do is set a list of criteria based on domain knowledge to weed out films that are clearly wrong.

```
revenue_threshold = 3_000_000_000  # 3 billion. I know that Avatar is highest grossing of all time, so that's our max value

budget_threshold = 500_000_000      # 500 million. I know that Avatar 2 and Star Wars Force Awakens are the most expensive movies ever made, so we can use that as our cap

profit_threshold = 3_000_000_000     # 3 billion. Using Avatar 2 revenue upper bound
```

Column	count	unique	top	freq	mean	std	min	25%	50%	75%	max
title	11,075	10,762	Godzilla	3	NaN	NaN	NaN	NaN	NaN	NaN	NaN
revenue	11,075	NaN	NaN	NaN	62,075,530	156,289,300	1,015	1,000,000	10,000,000	50,697,340	3,000,000,000
budget	11,075	NaN	NaN	NaN	21,701,540	36,741,750	4	1,028,882	7,520,000	25,000,000	460,000,000
profit	11,075	NaN	NaN	NaN	40,373,990	131,825,600	-199,546,000	-801,546	1,300,000	28,000,000	2,780,000,000
profit_margin	11,075	NaN	NaN	NaN	7.3189	52.1582	-0.9999	-0.2676	0.8533	3.0000	999.9474
vote_average	11,075	NaN	NaN	NaN	5.8393	2.2232	0	5.5685	6.34	7	10
popularity	11,075	NaN	NaN	NaN	15.101	17.806	0	3.5	11.824	19.2705	241.285

Surprising results...but they make sense

Our mean profit margin for a film is 7.3. Very high

The profit margin standard deviation is 52 though, so there seem to be a lot of variation amongst results.

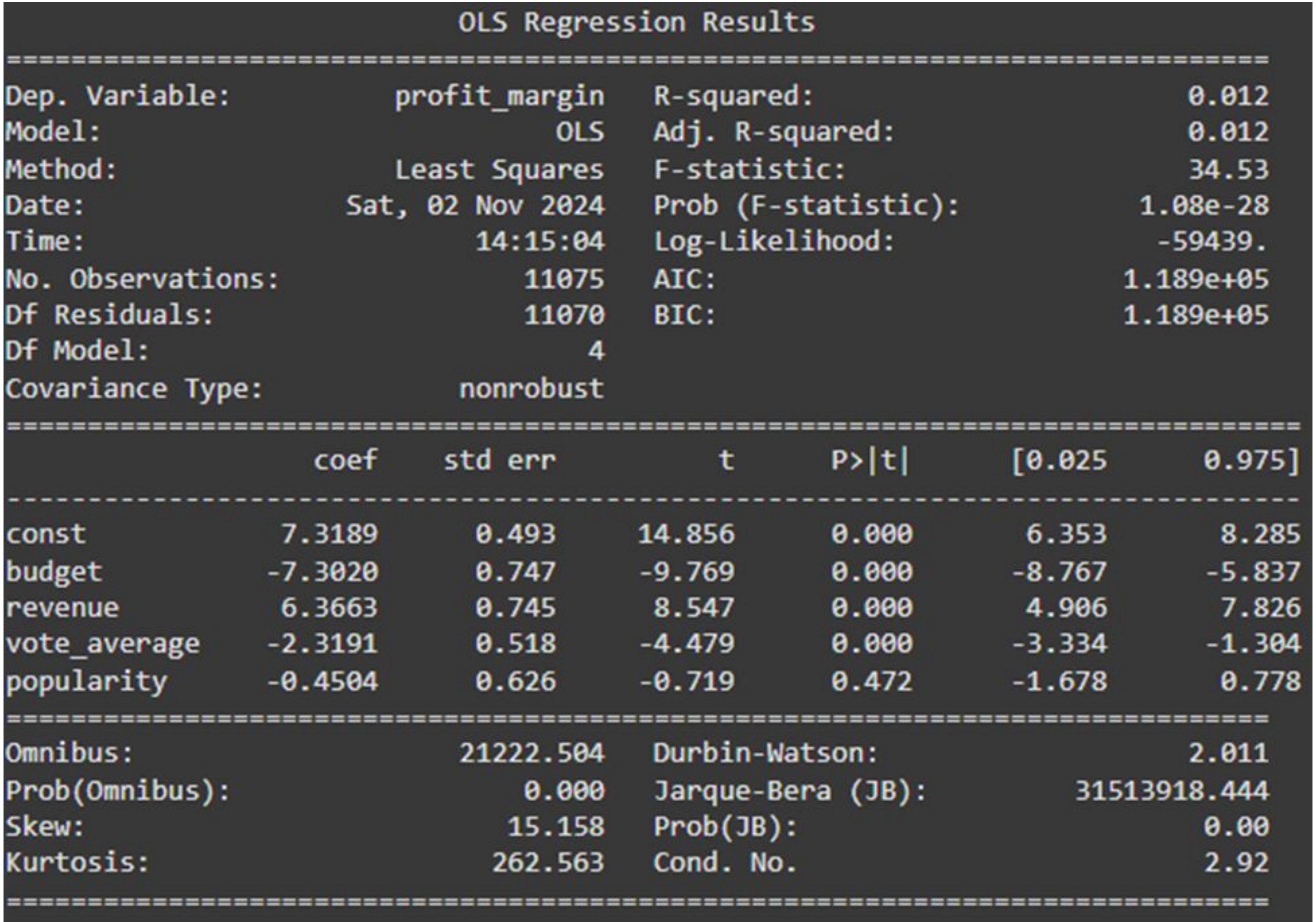
Another surprising result is that the mean budget of the films is $21 million. Cleaning our dataset likely skewed our results The dataset is supposed to be inclusive of films regardless their size. That means it should incorporate small budget indi films. Short and indi films are much more numerous than large budget films simply due to their ease of creation. Those rarely have a budget of even $1 million. Likely this means that small indi films and short films didn't list their budget in the dataset. Our resulting data possibly omitted almost all short and indi films.

Top 5 highest Revenue films

Title	Revenue	Budget	Profit	Profit Margin	Vote Average	Popularity
TikTok Rizz Party	3,000,000,000	250,000,000	2,750,000,000	11.00%	10.00	0.00
Bee Movie	2,930,000,000	150,000,000	2,780,000,000	18.53%	0.00	1.40
Avatar	2,923,706,026	237,000,000	2,686,706,026	11.34%	7.57	79.93
Avengers: Endgame	2,800,000,000	356,000,000	2,444,000,000	6.87%	8.26	91.76
Avatar: The Way of Water	2,320,250,281	460,000,000	1,860,250,281	4.04%	7.65	241.29
Titanic	2,264,162,353	200,000,000	2,064,162,353	10.32%	7.90	102.35

Top 5 highest Budget films

Title	Revenue	Budget	Profit	Profit Margin	Vote Average	Popularity
Avatar: The Way of Water	2,320,250,281	460,000,000	1,860,250,281	4.04%	7.65	241.29
Lost in the Stars	334,039,200	417,549,000	-83,509,800	-20.00%	6.33	23.73
Pirates of the Caribbean: On Stranger Tides	1,045,713,802	379,000,000	666,713,802	1.76%	6.54	79.19
Avengers: Age of Ultron	1,405,403,694	365,000,000	1,040,403,694	2.85%	7.28	96.57



```
                            OLS Regression Results
==============================================================================
Dep. Variable:          profit_margin   R-squared:                       0.012
Model:                            OLS   Adj. R-squared:                  0.012
Method:                 Least Squares   F-statistic:                     34.53
Date:                Sat, 02 Nov 2024   Prob (F-statistic):           1.08e-28
Time:                        14:15:04   Log-Likelihood:                -59439.
No. Observations:               11075   AIC:                         1.189e+05
Df Residuals:                   11070   BIC:                         1.189e+05
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const            7.3189      0.493     14.856      0.000       6.353       8.285
budget          -7.3020      0.747     -9.769      0.000      -8.767      -5.837
revenue          6.3663      0.745      8.547      0.000       4.906       7.826
vote_average    -2.3191      0.518     -4.479      0.000      -3.334      -1.304
popularity      -0.4504      0.626     -0.719      0.472      -1.678       0.778
==============================================================================
Omnibus:                    21222.504   Durbin-Watson:                   2.011
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         31513918.444
Skew:                          15.158   Prob(JB):                         0.00
Kurtosis:                     262.563   Cond. No.                         2.92
==============================================================================
```

OLS Multiple Regression Result Interpretation

Low of an R-squared. Model does a very poor job of explaining the variance in the profit margin. So there is weak explanatory power.

The Prob F statistic very low at a statistically significant level, indicating that at least one of the predictor variables is a significant factor in predicting the profit margin.

The budget, revenue, and vote_average all have p-values that are statistically significant, so they have an effect on the profit margin.

A one standard deviation increase in revenue is associated with an increase of approximately 6.3663 units in the profit margin. So higher revenue leads to an increase in profit margin. Profit is derived from revenue, so this is not particularly useful result.

For each 1 standard deviation increase in the scaled budget, the profit margin is expected to decrease by approximately -7.3020 units. So its a negative relationship, higher budgets lead to lower profit margins. This implies that if you wanted to increase your film's profit margin, its not as simple as just dumping more money into the budget.

OLS Multiple Regression Result Interpretation

A one unit increase in vote_average is associated with a decrease of approximately -2.3191 units in profit margin. This indicates that, counterintuitively, as a movie becomes more more highly rated, the profit margin might decrease. This is a very interesting finding. Following the logic to its extreme implications, the model predicts that you should make your film have a worse vote_average if you want your film to have a high profit margin.

It appears that popularity has a statistically insignificant effect upon the profit margin of the film. It appears that even if a film is popular on IMDB, it doesn’t make a difference to the resulting profit margin of the film.

What does revenue and budget coefficient mean in common sense English? After performing a few calculations, we can see that. For every increase in $1 million in revenue, profit margin is predicted by the model to increase by 4%. For every increase in $1 million in budget, profit margin is predicted by the model to decrease by 20%.

Analysis - Historical trends of budget, revenue and popularity of genres

Nahid

Analysis of movie trends and revenue over time

Looking at trends and profitability of the film industry from inception to its current state using the TMDB dataset v11

Study:

- Understand the growth of the film industry and how popularity of different genres have influenced investment and revenue

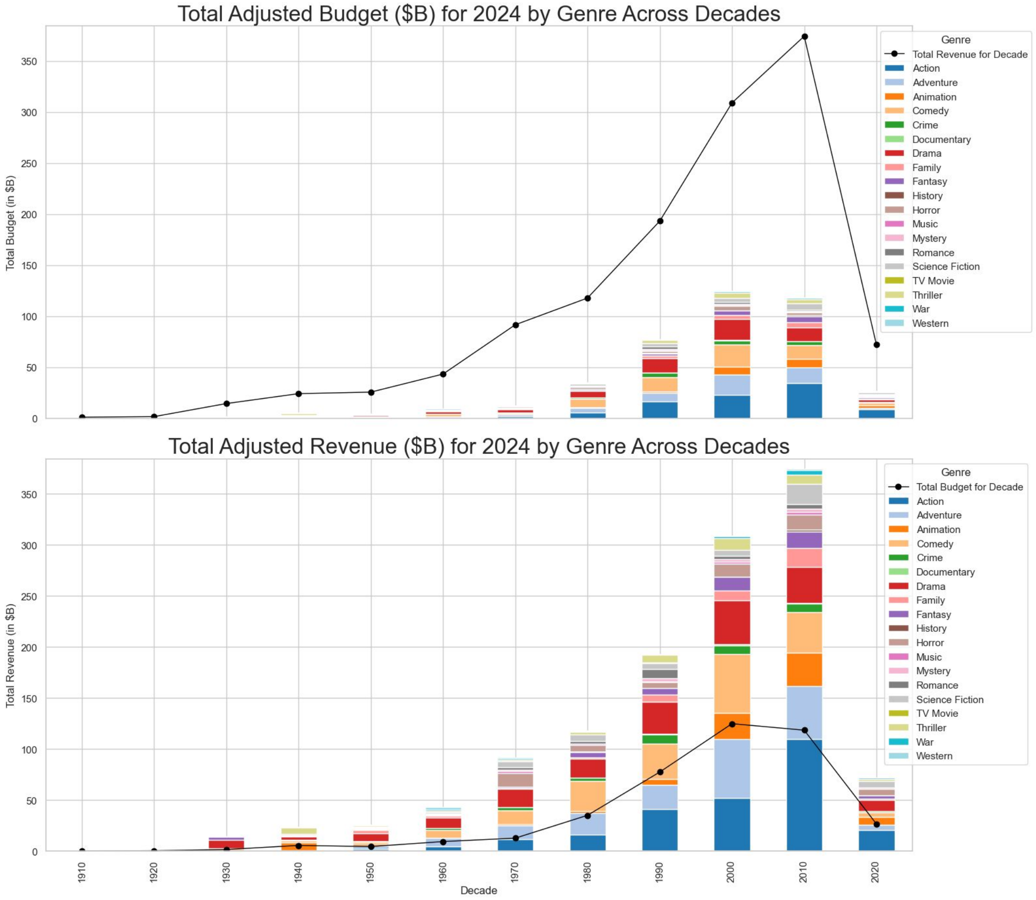
Data Cleansing: Initial set of 1.1M rows reduced to 9K useful movies!

- Removed movies with no titles or no release dates or outside 1915-2025
- Removed movies with $0 or NAN budgets and revenue
- Removed Adult movies and several cherry-picked off-skew movies

Data Transformation:

- Since this is a study over time, categorized the data into decades based on release date
- Also as this is a study over time looking at financial trends, corrected all historical $$ to 2024 CPI to account for inflation

Comparing Movie Genre Budgets vs Revenue



Total Adjusted Budget ($B) for 2024 by Genre Across Decades

Total Adjusted Revenue ($B) for 2024 by Genre Across Decades

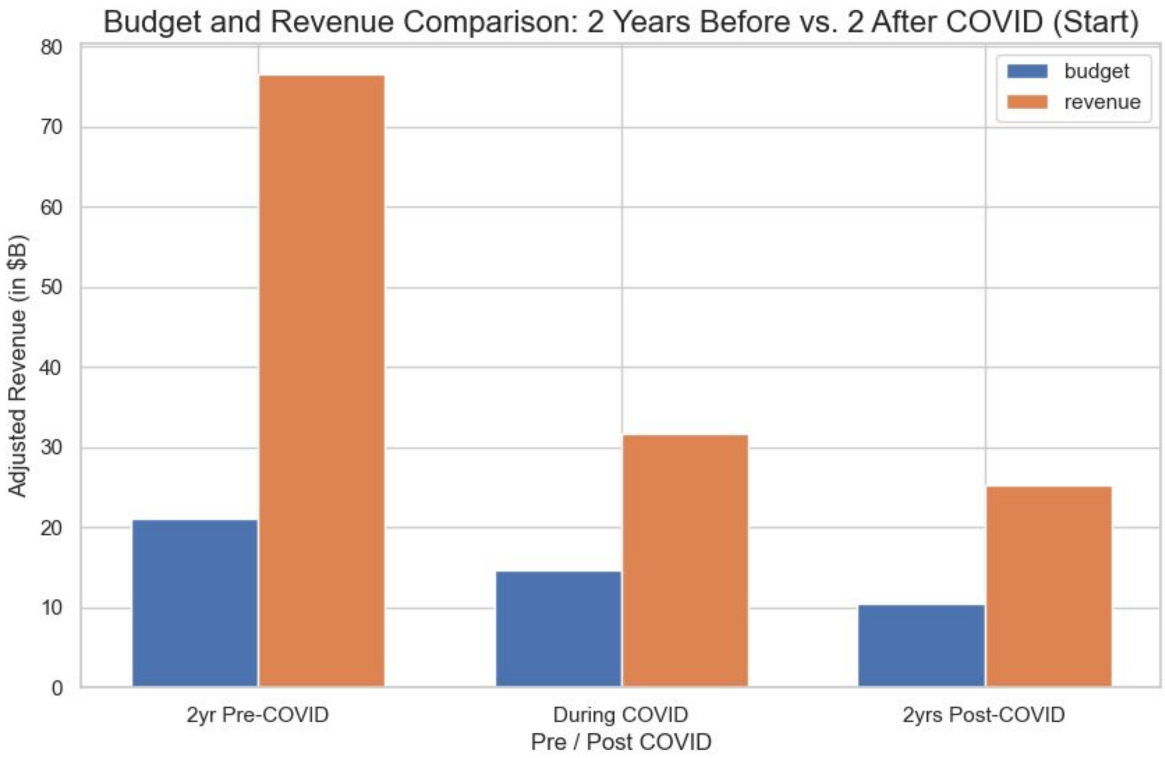
A simple comparison of movie budgets* and revenue* over time shows a consistent increment in both over the decades … until 2020!

At a macro level, we see budgets favouring genres like Action, Adventure, Comedy and Drama. These investments continuously reflect significant revenue returns even through large initial investments.

Let's talk 2020s!

*All budgets and revenue are correct for 2024 Consumer Price index (CPI) to level out inflation variances

How big was/is the COVID Impact??



Comparison of Revenue* before and after COVID is very telling!

Post COVID, the film industry as not caught up to revenue DURING COVID!**

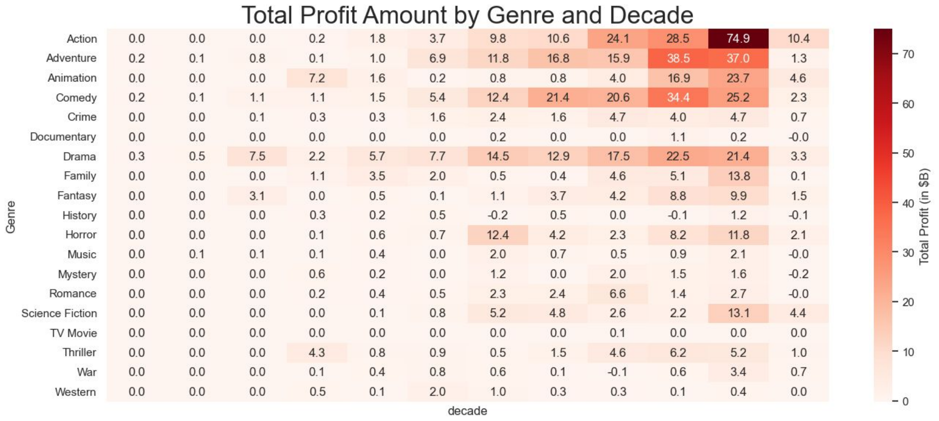
The Budget variance is not proportional to Revenue

*All budgets and revenue are correct for 2024 Consumer Price index (CPI) to level out inflation variances

** Using US Dates for COVID lockdown start/stop

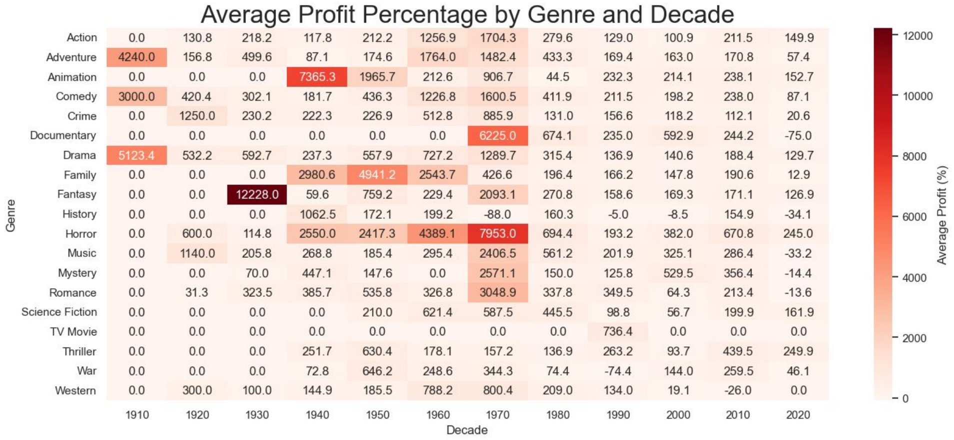
We stopped going to the movies during COVID … and we never went back!

A Deeper Look



Total Profit Amount by Genre and Decade

Action	0.0	0.0	0.0	0.2	1.8	3.7	9.8	10.6	24.1	28.5	74.9	10.4
Adventure	0.2	0.1	0.8	0.1	1.0	6.9	11.8	16.8	15.9	38.5	37.0	1.3
Animation	0.0	0.0	0.0	7.2	1.6	0.2	0.8	0.8	4.0	16.9	23.7	4.6
Comedy	0.2	0.1	1.1	1.1	1.5	5.4	12.4	21.4	20.6	34.4	25.2	2.3
Crime	0.0	0.0	0.1	0.3	0.3	1.6	2.4	1.6	4.7	4.0	4.7	0.7
Documentary	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	1.1	0.2	-0.0
Drama	0.3	0.5	7.5	2.2	5.7	7.7	14.5	12.9	17.5	22.5	21.4	3.3
Family	0.0	0.0	0.0	1.1	3.5	2.0	0.5	0.4	4.6	5.1	13.8	0.1
Fantasy	0.0	0.0	3.1	0.0	0.5	0.1	1.1	3.7	4.2	8.8	9.9	1.5
History	0.0	0.0	0.0	0.3	0.2	0.5	-0.2	0.5	0.0	-0.1	1.2	-0.1
Horror	0.0	0.0	0.0	0.1	0.6	0.7	12.4	4.2	2.3	8.2	11.8	2.1
Music	0.0	0.1	0.1	0.1	0.4	0.0	2.0	0.7	0.5	0.9	2.1	-0.0
Mystery	0.0	0.0	0.0	0.6	0.2	0.0	1.2	0.0	2.0	1.5	1.6	-0.2
Romance	0.0	0.0	0.0	0.2	0.4	0.5	2.3	2.4	6.6	1.4	2.7	-0.0
Science Fiction	0.0	0.0	0.0	0.0	0.1	0.8	5.2	4.8	2.6	2.2	13.1	4.4
TV Movie	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
Thriller	0.0	0.0	0.0	4.3	0.8	0.9	0.5	1.5	4.6	6.2	5.2	1.0
War	0.0	0.0	0.0	0.1	0.4	0.8	0.6	0.1	-0.1	0.6	3.4	0.7
Western	0.0	0.0	0.0	0.5	0.1	2.0	1.0	0.3	0.3	0.1	0.4	0.0



Average Profit Percentage by Genre and Decade

	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	2020
Action	0.0	130.8	218.2	117.8	212.2	1256.9	1704.3	279.6	129.0	100.9	211.5	149.9
Adventure	4240.0	156.8	499.6	87.1	174.6	1764.0	1482.4	433.3	169.4	163.0	170.8	57.4
Animation	0.0	0.0	0.0	7365.3	1965.7	212.6	906.7	44.5	232.3	214.1	238.1	152.7
Comedy	3000.0	420.4	302.1	181.7	436.3	1226.8	1600.5	411.9	211.5	198.2	238.0	87.1
Crime	0.0	1250.0	230.2	222.3	226.9	512.8	885.9	131.0	156.6	118.2	112.1	20.6
Documentary	0.0	0.0	0.0	0.0	0.0	0.0	6225.0	674.1	235.0	592.9	244.2	-75.0
Drama	5123.4	532.2	592.7	237.3	557.9	727.2	1289.7	315.4	136.9	140.6	188.4	129.7
Family	0.0	0.0	0.0	2980.6	4941.2	2543.7	426.6	196.4	166.2	147.8	190.6	12.9
Fantasy	0.0	0.0	12228.0	59.6	759.2	229.4	2093.1	270.8	158.6	169.3	171.1	126.9
History	0.0	0.0	0.0	1062.5	172.1	199.2	-88.0	160.3	-5.0	-8.5	154.9	-34.1
Horror	0.0	600.0	114.8	2550.0	2417.3	4389.1	7953.0	694.4	193.2	382.0	670.8	245.0
Music	0.0	1140.0	205.8	268.8	185.4	295.4	2406.5	561.2	201.9	325.1	286.4	-33.2
Mystery	0.0	0.0	70.0	447.1	147.6	0.0	2571.1	150.0	125.8	529.5	356.4	-14.4
Romance	0.0	31.3	323.5	385.7	535.8	326.8	3048.9	337.8	349.5	64.3	213.4	-13.6
Science Fiction	0.0	0.0	0.0	0.0	210.0	621.4	587.5	445.5	98.8	56.7	199.9	161.9
TV Movie	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	736.4	0.0	0.0	0.0
Thriller	0.0	0.0	0.0	251.7	630.4	178.1	157.2	136.9	263.2	93.7	439.5	249.9
War	0.0	0.0	0.0	72.8	646.2	248.6	344.3	74.4	-74.4	144.0	259.5	46.1
Western	0.0	300.0	100.0	144.9	185.5	788.2	800.4	209.0	134.0	19.1	-26.0	0.0

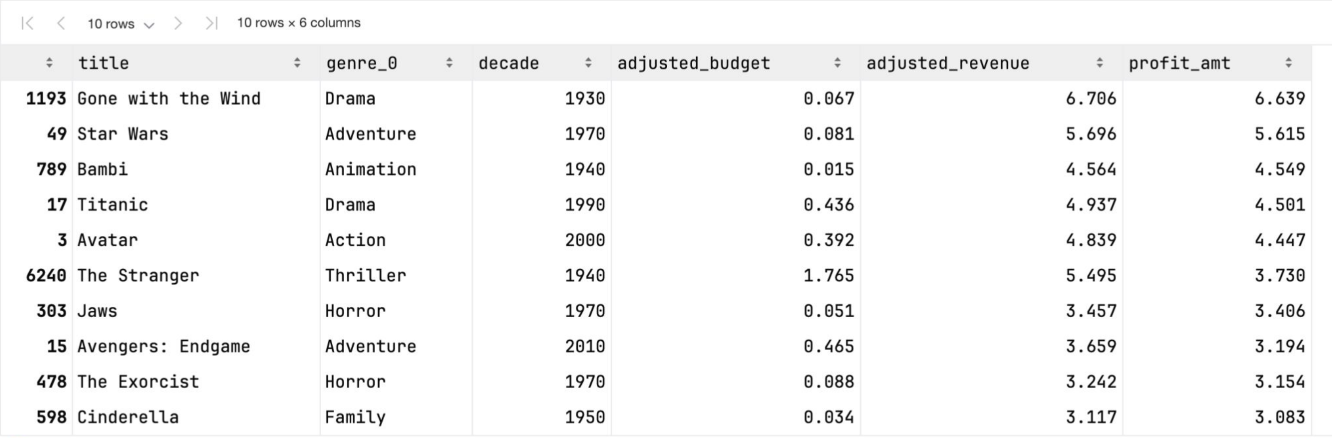
In evaluating profit based on value returned on initial budget *, we are sure to think that the big budget movies of the 21st century top the bill.

They do!

But if you look at profit from a different lens … the percentage returned based on initial budget *, we see a very different story.

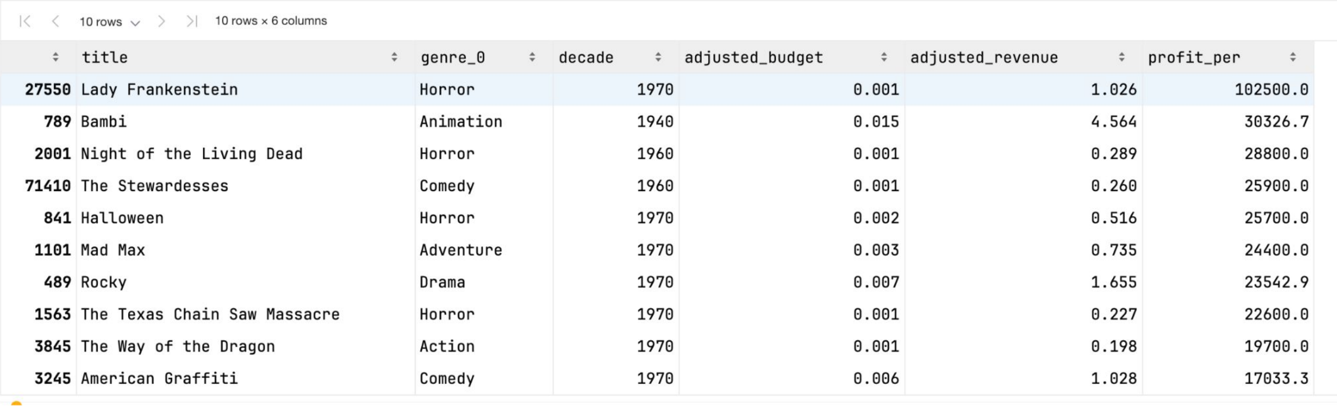
How can a movie made in the 1930s crush the latest Marvel miracle?

Top 10 Movies by Profit Amount (in $B):


10 rows
10 rows × 6 columns

	title	genre_0	decade	adjusted_budget	adjusted_revenue	profit_amt
1193	Gone with the Wind	Drama	1930	0.067	6.706	6.639
49	Star Wars	Adventure	1970	0.081	5.696	5.615
789	Bambi	Animation	1940	0.015	4.564	4.549
17	Titanic	Drama	1990	0.436	4.937	4.501
3	Avatar	Action	2000	0.392	4.839	4.447
6240	The Stranger	Thriller	1940	1.765	5.495	3.730
303	Jaws	Horror	1970	0.051	3.457	3.406
15	Avengers: Endgame	Adventure	2010	0.465	3.659	3.194
478	The Exorcist	Horror	1970	0.088	3.242	3.154
598	Cinderella	Family	1950	0.034	3.117	3.083

Top 10 Movies by Profit Percentage:

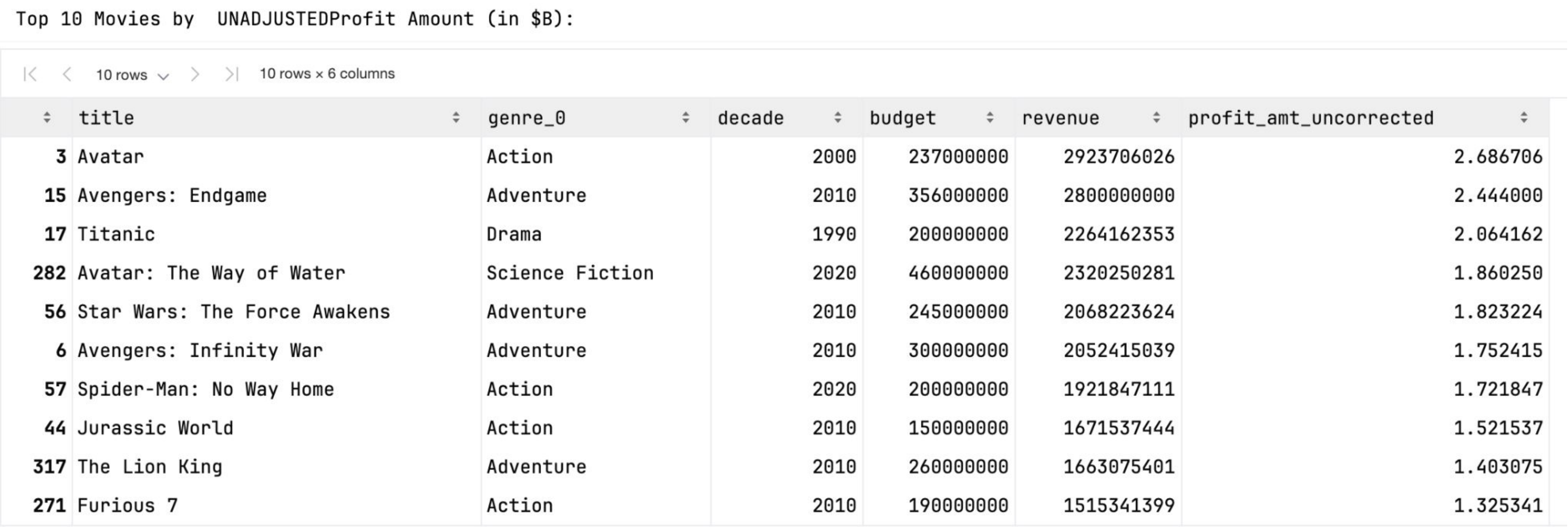

10 rows
10 rows × 6 columns

	title	genre_0	decade	adjusted_budget	adjusted_revenue	profit_per
27550	Lady Frankenstein	Horror	1970	0.001	1.026	102500.0
789	Bambi	Animation	1940	0.015	4.564	30326.7
2001	Night of the Living Dead	Horror	1960	0.001	0.289	28800.0
71410	The Stewardesses	Comedy	1960	0.001	0.260	25900.0
841	Halloween	Horror	1970	0.002	0.516	25700.0
1101	Mad Max	Adventure	1970	0.003	0.735	24400.0
489	Rocky	Drama	1970	0.007	1.655	23542.9
1563	The Texas Chain Saw Massacre	Horror	1970	0.001	0.227	22600.0
3845	The Way of the Dragon	Action	1970	0.001	0.198	19700.0
3245	American Graffiti	Comedy	1970	0.006	1.028	17033.3

Corrected for time, we really do see a wide spread of genres and eras that top the profit bill. Only two titles from the 21st century mega-movies make the top 10!

In addition, we see small budget cult-movies from the 1970s dominate the profit percentage!

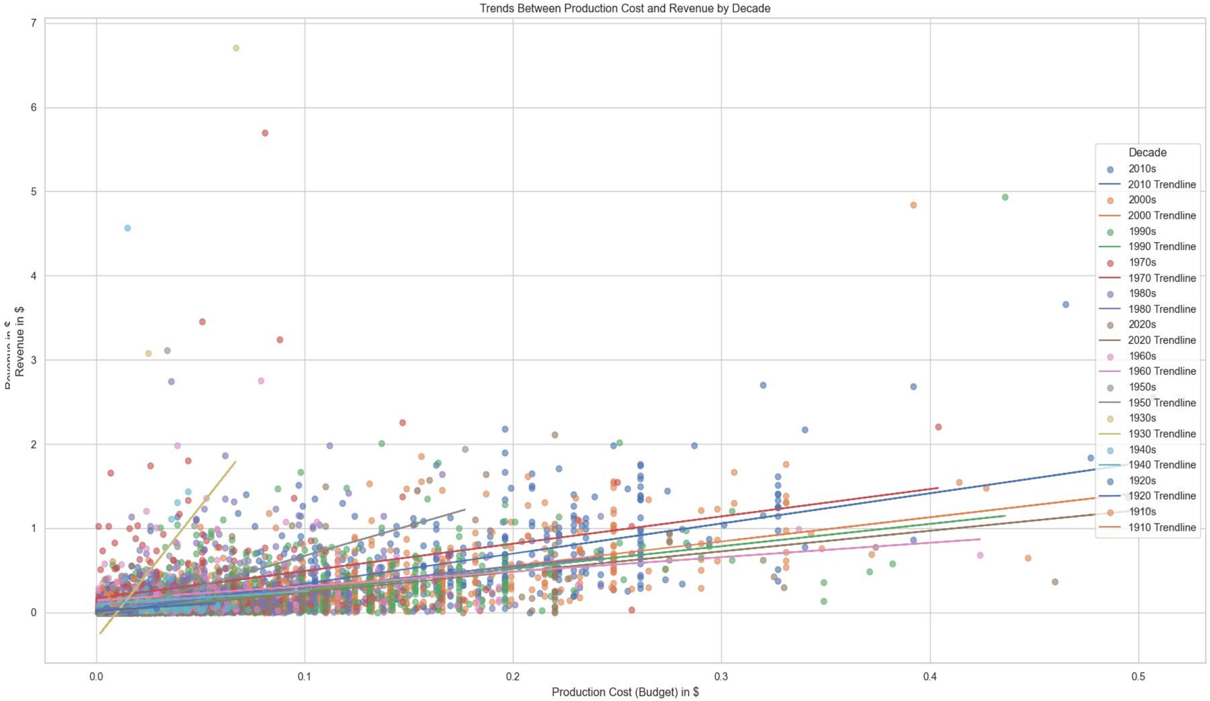
FYR … Unadjusted Top Grossing Movies

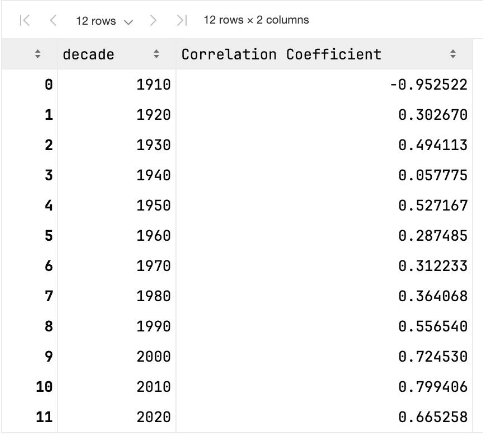


Top 10 Movies by UNADJUSTEDProfit Amount (in $B):

	title	genre_0	decade	budget	revenue	profit_amt_uncorrected
3	Avatar	Action	2000	237000000	2923706026	2.686706
15	Avengers: Endgame	Adventure	2010	356000000	2800000000	2.444000
17	Titanic	Drama	1990	200000000	2264162353	2.064162
282	Avatar: The Way of Water	Science Fiction	2020	460000000	2320250281	1.860250
56	Star Wars: The Force Awakens	Adventure	2010	245000000	2068223624	1.823224
6	Avengers: Infinity War	Adventure	2010	300000000	2052415039	1.752415
57	Spider-Man: No Way Home	Action	2020	200000000	1921847111	1.721847
44	Jurassic World	Action	2010	150000000	1671537444	1.521537
317	The Lion King	Adventure	2010	260000000	1663075401	1.403075
271	Furious 7	Action	2010	190000000	1515341399	1.325341

Is there a correlation with Budget??



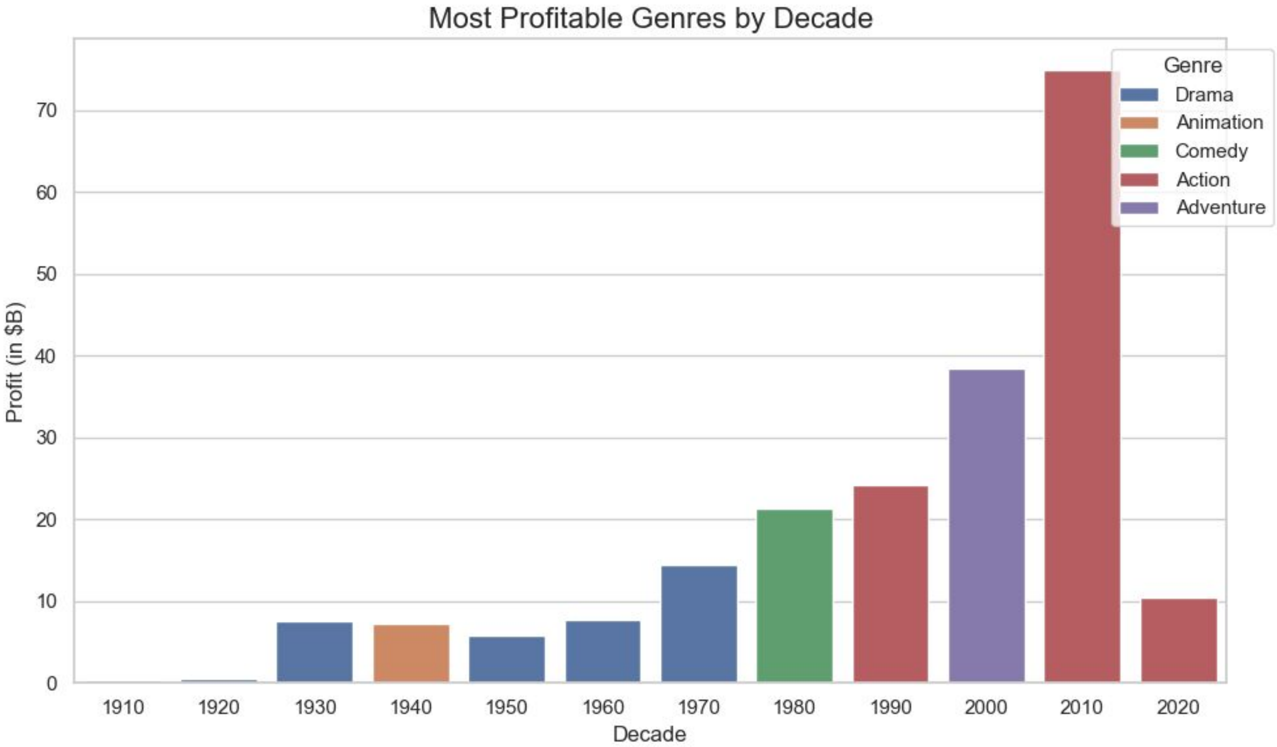


12 rows ∨ 12 rows × 2 columns

	decade	Correlation Coefficient
0	1910	-0.952522
1	1920	0.302670
2	1930	0.494113
3	1940	0.057775
4	1950	0.527167
5	1960	0.287485
6	1970	0.312233
7	1980	0.364068
8	1990	0.556540
9	2000	0.724530
10	2010	0.799406
11	2020	0.665258

History shows a weak to medium correlation between initial budget and revenue generated. However, the last 3 decades have shown a much stronger correlation … except 2020s!

Most profitable Genres over time



While there is a very weak correlation between genre and revenue, out of interest, these have been the most successful genres across the decades.

The earlier part of the last decade enjoyed Drama while the last 40yrs have gravitated towards Action.

But we don’t need data-science to know that! :-)