

Trinity Assignment Five

IMDB Movie Analysis

Aayush Damani

Table of Contents

1. PROJECT OVERVIEW
2. APPROACH
3. TECH STACK USED
4. RESULTS AND INSIGHTS
5. GOOGLE DRIVE LINK

Project Description

- The goal of this project is to leverage Excel skills in order to analyze IMDB Movie Data in order to answer the question “What factors influence the success of a movie on IMBD?”
- The impact of factors such as genre, duration, directors etc. on IMDB score will be investigated in this project.
- Here, success is defined by high IMDB ratings. This analysis is crucial for movie producers, directors, and investors who aim to understand the determinants of a successful movie, enabling them to make informed decisions for future projects.
- Dataset Used: IMDB Movie Data

Project Overview

Movie Genre Analysis: Analyze the distribution of movie genres and their impact on the IMDB score.

Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

Language Analysis: Examine the distribution of movies based on their language.

Director Analysis: Influence of directors on movie ratings.

Budget Analysis: Explore the relationship between movie budgets and their financial success.

Approach

Data Preparation:

- Set up the database using provided commands and files.
- Ensure data integrity for accurate analysis.

Analysis Execution:

- Complete the tasks identified in the previous slide (Project Overview) in order to derive actionable insights using Excel Pivot Tables and the relevant functions.

Function/Chart Execution:

- Write and validate the correct Excel functions and make the relevant charts.

Reporting:

- Create a PowerPoint presentation with video explanation of all findings and insights.
- Include screenshots and summarize findings for leadership presentation.

Tech Stack Used

Microsoft Excel



Google Slides



Movie Genre Analysis

Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

Approach:

- First, the 'genre' column was split up into several different columns as one movie can have multiple genres (Using Text to Columns feature)
- Then, using Power Query, the 7 'genre' columns were unpivoted so that we can find the count of movies and the IMDB score for each genre.
- Lastly, using a combination of Pivot Tables and aggregate functions such as MODE or MEDIAN, the descriptive statistics for each genre were obtained.

Movie Genre Analysis

Genre	Count of movie_title	Average of imdb_score	Max of imdb_score	Min of imdb_score	StdDev of imdb_score	Var of imdb_score	Median of imdb_score	Mode of imdb_score
Action	1153	6.239895924	9.1	1.7	1.118835265	1.25179235	6.3	6.1
Adventure	923	6.441170098	8.9	1.9	1.131196138	1.279604703	6.6	6.7
Animation	242	6.576033058	8.6	1.7	1.139594803	1.298676314	6.7	6.7
Biography	293	7.150170648	8.9	4.5	0.722515799	0.52202908	7.2	7
Comedy	1872	6.195245726	9.5	1.7	1.090713849	1.189656701	6.3	6.7
Crime	889	6.564791901	9.3	2.4	1.02645633	1.053612597	6.6	6.6
Documentary	121	7.180165289	8.7	1.6	1.056536782	1.116269972	7.4	7.5
Drama	2594	6.763762529	9.3	2	0.957353999	0.916526679	6.9	7.2
Family	546	6.245054945	8.7	1.7	1.201598055	1.443837887	6.4	6.7
Fantasy	610	6.30704918	8.9	1.7	1.160685835	1.347191607	6.4	6.7
Film-Noir	6	7.633333333	8.2	7.1	0.43204938	0.186666667	7.65	#N/A
Game-Show	1	2.9	2.9	2.9	#DIV/0!	#DIV/0!	2.9	#N/A
History	207	7.083574879	8.9	2	0.887901843	0.788369682	7.2	7.5
Horror	565	5.843539823	8.7	2.2	1.130468522	1.277959079	5.9	6.2
Music	214	6.410280374	8.5	1.6	1.178838019	1.389659076	6.6	6.5
Musical	132	6.507575758	8.5	2.1	1.225718123	1.502384918	6.7	7
Mystery	500	6.4864	8.6	2.2	1.090758703	1.189754549	6.6	6.6
News	3	7.533333333	8.1	7.1	0.513160144	0.263333333	7.4	#N/A
Reality-TV	2	4.75	6.6	2.9	2.61629509	6.845	4.75	#N/A
Romance	1106	6.45	8.6	2.1	0.996294038	0.99260181	6.5	6.5
Sci-Fi	616	6.281818182	8.8	1.9	1.210816001	1.466075388	6.4	6.7
Short	5	6.38	7.1	5.2	0.746324326	0.557	6.5	#N/A
Sport	182	6.606043956	8.7	2	1.101940407	1.214272661	6.8	7.2
Thriller	1408	6.315767045	9	2.2	1.054799156	1.112601259	6.4	6.1
War	213	7.070422535	8.6	2.7	0.874706587	0.765111613	7.1	7.1
Western	97	6.689690722	8.9	3.8	1.042481468	1.086767612	6.8	6.5

RESULTS & INSIGHTS

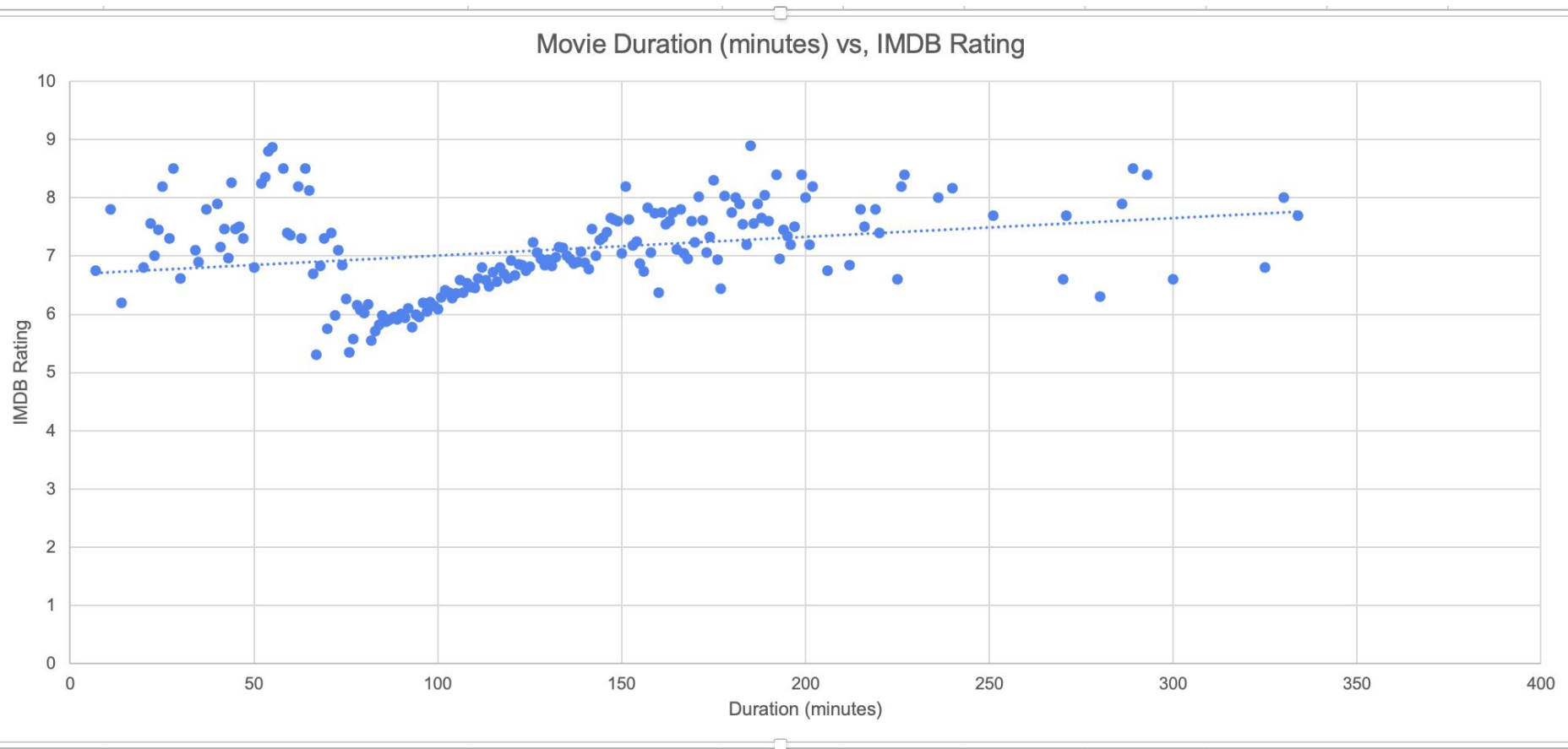
METRIC	RESULT	INSIGHT
Highest Rated Genre	Documentary has the highest average IMDb score (7.18).	Documentaries tend to be highly rated by audiences, potentially due to their educational and informative content.
Most Common Genres	Drama (2594 movies) and Comedy (1872 movies) are the most common genres	These genres dominate the movie industry, suggesting they have broad appeal and production.
Consistency in Ratings	Biography has a relatively low standard deviation (0.722), indicating consistent ratings.	Biographies are generally well-received and maintain a high standard of quality in the eyes of viewers.
Highest Maximum IMDB Rating	Comedy has the highest maximum IMDb score (9.5).	Despite varying quality, some comedies reach high levels of acclaim, reflecting the genre's potential for standout hits.

Movie Duration Analysis

Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score. [Using functions AVERAGE(), MEDIAN(), STDEV() etc.]

	Duration (in minutes)
Average of Movie Durations:	132.7329843
Median of Movie Durations:	127
Standard Deviation of Movie Durations:	73.46969953
Maximum of Movie Durations:	511
Minimum of Movie Durations:	7

Movie Duration Analysis



RESULTS & INSIGHTS

METRIC	RESULT	INSIGHT
Relationship between Duration and IMDB Rating	Movies with a duration around 120 minutes tend to have a wide range of ratings from 6.0 to 8.2.	Movie length does not necessarily correlate strongly with rating.
Relationship between Duration and IMDB Rating	The graph shows a very weak positive relationship between Movie Duration and Rating.	The same insight as above can be derived from the graph, as well as the aggregate statistics.

Language Analysis

Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

Language	Count of movie_title	Average of imdb_score	StdDev of imdb_score	Max of imdb_score	Min of imdb_score	Median of imdb_score
Aboriginal	2	6.95	0.777817459	7.5	6.4	6.95
Arabic	5	7.38	0.884307639	8.2	6	7.4
Aramaic	1	7.1	#DIV/0!	7.1	7.1	7.1
Bosnian	1	4.3	#DIV/0!	4.3	4.3	4.3
Cantonese	11	6.954545455	0.704788814	7.8	5.3	7.2
Chinese	3	5.666666667	0.550757055	6.2	5.1	5.7
Czech	1	7.4	#DIV/0!	7.4	7.4	7.4
Danish	5	7.5	1.077032961	8.3	5.7	8.1
Dari	2	7.5	0.141421356	7.6	7.4	7.5
Dutch	4	7.425	0.434932945	7.8	7	7.45
Dzongkha	1	7.5	#DIV/0!	7.5	7.5	7.5
English	4704	6.398426871	1.122067928	9.5	1.6	6.5
Filipino	1	6.7	#DIV/0!	6.7	6.7	6.7
French	73	7.038356164	0.726985812	8.4	4.9	7.2
German	19	7.342105263	0.954123093	8.5	4.9	7.6
Greek	1	7.3	#DIV/0!	7.3	7.3	7.3
Hebrew	5	7.58	0.334664011	8	7.2	7.6
Hindi	28	6.632142857	1.398955582	8.5	2.8	6.95
Hungarian	1	7.1	#DIV/0!	7.1	7.1	7.1
Icelandic	2	7.55	0.919238816	8.2	6.9	7.55
Indonesian	2	7.9	0.424264069	8.2	7.6	7.9
Italian	11	7.227272727	1.244259546	8.9	5.1	7.3
Japanese	18	7.394444444	0.990823913	8.7	5.6	7.6
Kannada	1	7.1	#DIV/0!	7.1	7.1	7.1
Kazakh	1	6	#DIV/0!	6	6	6
Korean	8	7.3875	0.825378701	8.4	5.7	7.5
Mandarin	26	6.788461538	1.042046802	7.9	3.2	7.05
Maya	1	7.8	#DIV/0!	7.8	7.8	7.8
Mongolian	1	7.3	#DIV/0!	7.3	7.3	7.3
None	2	7.95	0.777817459	8.5	7.4	7.95
Norwegian	4	7.15	0.574456265	7.6	6.4	7.3
Panjabi	1	6.6	#DIV/0!	6.6	6.6	6.6
Persian	4	7.575	1.203813385	8.5	5.9	7.95
Polish	4	8.25	0.981495458	9.1	7.4	8.25
Portuguese	8	7.4875	0.883883476	8.7	6.1	7.7
Romanian	2	7.2	0.989949494	7.9	6.5	7.2
Russian	11	6.363636364	1.383671007	8.1	4.1	6.5
Slovenian	1	6.4	#DIV/0!	6.4	6.4	6.4
Spanish	40	6.9375	0.855056603	8.2	4.4	7.15
Swahili	1	7.4	#DIV/0!	7.4	7.4	7.4
Swedish	5	7.44	0.756967635	8.2	6.6	7.6
Tamil	1	5.1	#DIV/0!	5.1	5.1	5.1
Telugu	1	8.4	#DIV/0!	8.4	8.4	8.4
Thai	3	6.633333333	0.450924975	7.1	6.2	6.6
Urdu	1	7	#DIV/0!	7	7	7
Vietnamese	1	7.4	#DIV/0!	7.4	7.4	7.4
Zulu	2	7.1	0.282842712	7.3	6.9	7.1

RESULTS & INSIGHTS

METRIC	RESULT	INSIGHT
English Movies	English movies have an average IMDb score of 6.40, a standard deviation of 1.12, with scores ranging from 1.6 to 9.5.	This high variability indicates that while English movies dominate in quantity (4704 films in this database), their quality varies widely, catering to diverse tastes and preferences.
Bollywood Industry	Hindi movies have an average IMDb score of 6.63, a standard deviation of 1.40, and scores ranging from 2.8 to 8.5	The large standard deviation suggests a significant disparity in quality, reflecting the diverse range of genres and production values in the Bollywood industry.
Spanish Movies	Spanish movies have an average IMDb score of 6.94, a standard deviation of 0.86, with scores ranging from 4.4 to 8.2.	Spanish films are generally well-received, maintaining a good average score with moderate variability, suggesting a steady quality in Spanish cinema.

Director Analysis

Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

Directors	Average of imdb_score	Count of movie_title	Percentile
John Blanchard	9.5	1	99.8
Cary Bell	8.7	1	99.2
Mitchell Altieri	8.7	1	99.2
Sadyk Sher-Niyaz	8.7	1	99.2
Charles Chaplin	8.6	1	98.8
Mike Mayhall	8.6	1	98.8
Raja Menon	8.5	1	98
Ron Fricke	8.5	1	98
Damien Chazelle	8.5	1	98
Majid Majidi	8.5	1	98
Sergio Leone	8.475	4	97.8
Christopher Nolan	8.425	8	97.6
Asghar Farhadi	8.4	1	95.6
Richard Marquand	8.4	1	95.6
Catherine Owens	8.4	1	95.6
Bill Melendez	8.4	1	95.6
Robert Mulligan	8.4	1	95.6
Jay Oliva	8.4	1	95.6
S.S. Rajamouli	8.4	1	95.6
Marius A. Markevicius	8.4	1	95.6
Rakeysh Omprakash Mehra	8.4	1	95.6
Moustapha Akkad	8.4	1	95.6
Fritz Lang	8.3	1	94.2
Stanley Donen	8.3	1	94.2
Lenny Abrahamson	8.3	1	94.2
Justin Paul Miller	8.3	1	94.2
Sut Jhally	8.3	1	94.2
Lee Unkrich	8.3	1	94.2
John Sturges	8.3	1	94.2
Pete Docter	8.233333333	3	94
Hayao Miyazaki	8.225	4	93.8
Quentin Tarantino	8.2	8	93.6

=PERCENTRANK.EXC(\$B\$4:\$B\$503, B4) * 100

RESULTS & INSIGHTS

METRIC	RESULT	INSIGHT
Highest Rated Director on IMDB	The data analysis indicates that John Blanchard has the highest average IMDb score of 9.5, despite having only 1 film.	Directors with a single highly rated film, like John Blanchard, can achieve very high average IMDb scores, suggesting that even a limited filmography can significantly impact their perceived excellence.

Therefore, we should apply a filter to the data from the previous slide, in order to remove the directors that have only 1 movie from this list.

Director Analysis (Filtered)

Directors	Average of imdb_score	Count of movie_title	Percentile
Sergio Leone	8.475	4	97.8
Christopher Nolan	8.425	8	97.6
Pete Docter	8.233333333	3	94
Hayao Miyazaki	8.225	4	93.8
Quentin Tarantino	8.2	8	93.6
Milos Forman	8.133333333	3	91.8
Frank Capra	8.06	5	89.2
Tony Kaye	8.033333333	3	89
Stanley Kubrick	8	7	86.4
David Lean	8	4	86.4
Frank Darabont	7.975	4	86
Billy Wilder	7.975	4	86
Denis Villeneuve	7.966666667	3	85.8
Joss Whedon	7.925	4	85.2
James Cameron	7.914285714	7	85
Elia Kazan	7.866666667	3	81.8
Charles Ferguson	7.866666667	3	81.8
Richard Brooks	7.8	4	76.2
Alfonso Cuarón	7.8	4	76.2
Alejandro G. Iñárritu	7.783333333	6	76
Dean DeBlois	7.766666667	3	75.8
Victor Fleming	7.766666667	3	75.4
Mel Gibson	7.766666667	3	75.4
Fred Zinnemann	7.76	5	75.2
David Fincher	7.75	10	74.8
Tom Hooper	7.733333333	3	74.4
Andrew Stanton	7.733333333	3	74.4
Peter Weir	7.725	4	74.2
Peter Jackson	7.675	12	68.2
Thomas Vinterberg	7.666666667	3	68
Martin Scorsese	7.66	20	67.8
Matthew Vaughn	7.65	4	67
David Lynch	7.64	5	66.8

=PERCENTRANK.EXC(\$B\$4:\$B\$503, B4) * 100

RESULTS & INSIGHTS

METRIC	RESULT	INSIGHT
Highest Rated Director on IMDB	The data analysis indicates that Sergio Leone has the highest average IMDb score of 8.475, having 4 films. The next director on the list is Christopher Nolan, with a rating of 8.425 while having 8 films.	Filtering the data helped us to lower the maximum IMDB rating, as it was over-inflated due to a low number of films for certain directors.

Budget Analysis

Correlation between Budget and Average Gross Revenue

-0.007752813

Sorted by Profit Margin

movie_title	budget	gross	profit	profit margin
Paranormal Activity	\$15,000.00	\$107,917,283.00	\$107,902,283.00	99.98610047
Tamaton	\$218.00	\$592,014.00	\$591,796.00	99.96317655
The Blair Witch Project	\$60,000.00	\$140,530,114.00	\$140,470,114.00	99.95730452
The Brothers McMullen	\$25,000.00	\$10,246,600.00	\$10,221,600.00	99.75601663
The Texas Chain Saw Massacre	\$83,532.00	\$30,859,000.00	\$30,775,468.00	99.72931074
The Texas Chain Saw Massacre	\$83,532.00	\$30,859,000.00	\$30,775,468.00	99.72931074
El Mariachi	\$7,000.00	\$2,040,920.00	\$2,033,920.00	99.65701742
The Gallows	\$100,000.00	\$22,757,819.00	\$22,657,819.00	99.56059058
Super Size Me	\$65,000.00	\$11,529,368.00	\$11,464,368.00	99.43622235
Halloween	\$300,000.00	\$47,000,000.00	\$46,700,000.00	99.36170213
Halloween	\$300,000.00	\$47,000,000.00	\$46,700,000.00	99.36170213
Halloween	\$300,000.00	\$47,000,000.00	\$46,700,000.00	99.36170213
American Graffiti	\$777,000.00	\$115,000,000.00	\$114,223,000.00	99.32434783
Rocky	\$960,000.00	\$117,235,247.00	\$116,275,247.00	99.18113364
In the Company of Men	\$25,000.00	\$2,856,622.00	\$2,831,622.00	99.12484046
Napoleon Dynamite	\$400,000.00	\$44,540,956.00	\$44,140,956.00	99.10195012
Facing the Giants	\$100,000.00	\$10,174,663.00	\$10,074,663.00	99.01716647
Snow White and the Seven Dwarfs	\$2,000,000.00	\$184,925,485.00	\$182,925,485.00	98.9184833
Benji	\$500,000.00	\$39,552,600.00	\$39,052,600.00	98.7358606
My Date with Drew	\$1,100.00	\$85,222.00	\$84,122.00	98.70925348
The Circle	\$10,000.00	\$673,780.00	\$663,780.00	98.51583603
Fireproof	\$500,000.00	\$33,451,479.00	\$32,951,479.00	98.50529778
Open Water	\$500,000.00	\$30,500,882.00	\$30,000,882.00	98.36070314
Primer	\$7,000.00	\$424,760.00	\$417,760.00	98.35201055
Pi	\$60,000.00	\$3,216,970.00	\$3,156,970.00	98.13489091
Slacker	\$23,000.00	\$1,227,508.00	\$1,204,508.00	98.12628512
The Devil Inside	\$1,000,000.00	\$53,245,055.00	\$52,245,055.00	98.12189132
Once	\$180,000.00	\$9,437,933.00	\$9,257,933.00	98.09280274
Hollywood Shuffle	\$100,000.00	\$5,228,617.00	\$5,128,617.00	98.08744836
Gone with the Wind	\$3,977,000.00	\$198,655,278.00	\$194,678,278.00	97.9980396
My Big Fat Greek Wedding	\$5,000,000.00	\$241,437,427.00	\$236,437,427.00	97.92907004

Budget Analysis

Correlation between Budget and Average Gross Revenue

-0.007752813

Sorted by Total Profit

movie_title	budget	gross	profit	profit margin
Avatar	\$237,000,000.00	\$760,505,847.00	\$523,505,847.00	68.83653151
Jurassic World	\$150,000,000.00	\$652,177,271.00	\$502,177,271.00	77.00011842
Titanic	\$200,000,000.00	\$658,672,302.00	\$458,672,302.00	69.63588732
Star Wars: Episode IV - A New Hope	\$11,000,000.00	\$460,935,665.00	\$449,935,665.00	97.61354982
E.T. the Extra-Terrestrial	\$10,500,000.00	\$434,949,459.00	\$424,449,459.00	97.58592641
The Avengers	\$220,000,000.00	\$623,279,547.00	\$403,279,547.00	64.70283662
The Avengers	\$220,000,000.00	\$623,279,547.00	\$403,279,547.00	64.70283662
The Lion King	\$45,000,000.00	\$422,783,777.00	\$377,783,777.00	89.35626142
Star Wars: Episode I - The Phantom Menace	\$115,000,000.00	\$474,544,677.00	\$359,544,677.00	75.76624382
The Dark Knight	\$185,000,000.00	\$533,316,061.00	\$348,316,061.00	65.31137659
The Hunger Games	\$78,000,000.00	\$407,999,255.00	\$329,999,255.00	80.88231803
Deadpool	\$58,000,000.00	\$363,024,263.00	\$305,024,263.00	84.02310647
The Hunger Games: Catching Fire	\$130,000,000.00	\$424,645,577.00	\$294,645,577.00	69.38623477
Jurassic Park	\$63,000,000.00	\$356,784,000.00	\$293,784,000.00	82.3422575
Despicable Me 2	\$76,000,000.00	\$368,049,635.00	\$292,049,635.00	79.35061123
American Sniper	\$58,800,000.00	\$350,123,553.00	\$291,323,553.00	83.20592845
Finding Nemo	\$94,000,000.00	\$380,838,870.00	\$286,838,870.00	75.31764549
Shrek 2	\$150,000,000.00	\$436,471,036.00	\$286,471,036.00	65.63345844
The Lord of the Rings: The Return of the King	\$94,000,000.00	\$377,019,252.00	\$283,019,252.00	75.0675862
Star Wars: Episode VI - Return of the Jedi	\$32,500,000.00	\$309,125,409.00	\$276,625,409.00	89.48646761
Forrest Gump	\$55,000,000.00	\$329,691,196.00	\$274,691,196.00	83.31772256
Star Wars: Episode V - The Empire Strikes Back	\$18,000,000.00	\$290,158,751.00	\$272,158,751.00	93.79649935
Home Alone	\$18,000,000.00	\$285,761,243.00	\$267,761,243.00	93.70103524
Star Wars: Episode III - Revenge of the Sith	\$113,000,000.00	\$380,262,555.00	\$267,262,555.00	70.28368991
Spider-Man	\$139,000,000.00	\$403,706,375.00	\$264,706,375.00	65.56903517
Minions	\$74,000,000.00	\$336,029,560.00	\$262,029,560.00	77.97812788
The Sixth Sense	\$40,000,000.00	\$293,501,675.00	\$253,501,675.00	86.37145768
Jaws	\$8,000,000.00	\$260,000,000.00	\$252,000,000.00	96.92307692
Frozen	\$150,000,000.00	\$400,736,600.00	\$250,736,600.00	62.56892932
The Secret Life of Pets	\$75,000,000.00	\$323,505,540.00	\$248,505,540.00	76.81647121

RESULTS & INSIGHTS

METRIC	RESULT	INSIGHT
Profit Margin	The movie 'Paranormal Activity' had a budget of only \$15,000, while getting a gross revenue of \$107,917,283 and a profit margin of 99.99%.	There are many such movies with a very low budget but a high gross revenue - these would be the 'ideal' film to make as it involves a low barrier to entry.
Highest Profit	The movie 'Avatar' had the highest gross profit, with \$523,505,847, while having a profit margin of 66.84%.	This suggests that profit margin should not be the only indicator considered when making a movie, as in Avatar's case, a higher budget led to a higher revenue.
Correlation between Budget and Gross Revenue	The correlation between budget and gross revenue is -0.001, which suggests that there is no correlation between the two variables.	As proven by movies like Paranormal Activity, with a low budget and high revenue, there is no direct correlation between budget and gross revenue, so no insights can be drawn from the result.

DRIVE LINK

<https://docs.google.com/presentation/d/1hkCe7on5KaAFSofkP4y9pkEbskwYvXgLPsYqG7f4R6g/edit?usp=sharing>



THANK YOU FOR YOUR TIME.

Please reach out to me at aayushd97@gmail.com for any feedback or suggestions.

