IMDB Movie Analysis

Project Overview:

Problem Statement: The dataset provided is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

Objective

Analyze the relationship between movie budgets and their financial success.
Examine how different factors (genre, director, language, budget, duration) impact IMDb scores.

Dataset for the project

Dataset

XLS IMDB_Movies

Approach

Data Cleaning:
- Removed missing or irrelevant data.
- Standardized numerical formats (example - converted figures into millions).

Analysis Techniques Used:
- **Correlation Analysis:** Measured the strength of relationships between budget and financial success.
- **Factor Impact Analysis:** Assessed how genre, director, language, budget, and duration influence IMDb scores.
- **Profit/Loss Evaluation:** Identified the most profitable movies based on budget vs. gross earnings.

Visualization:
- Created scatter plots, bar charts, and tables for a clear presentation of findings.
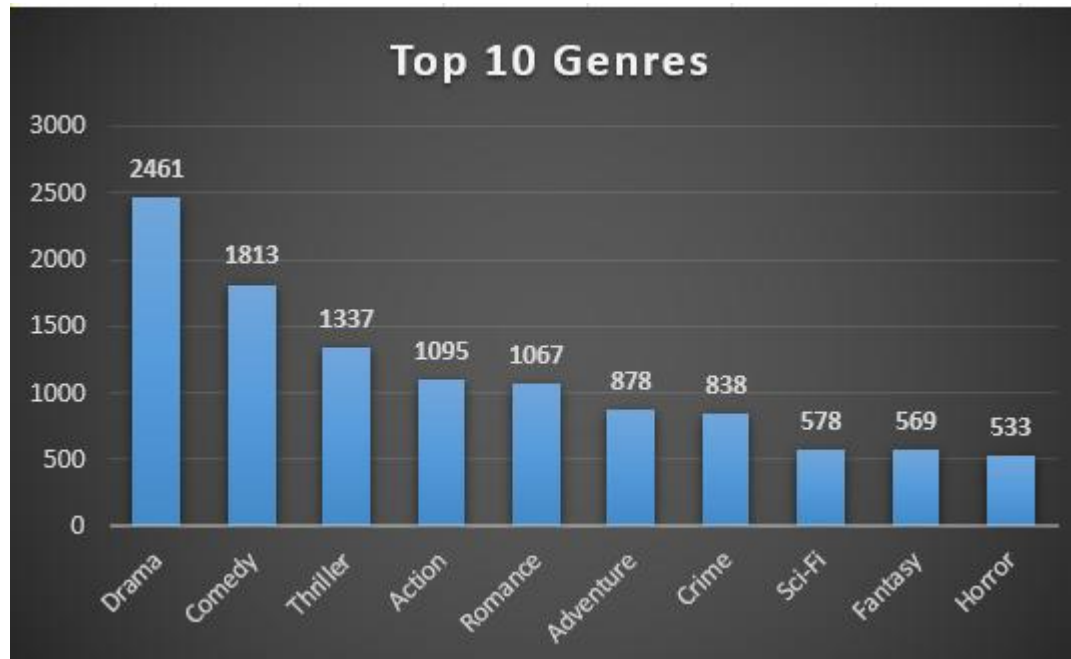
Tech-Stack Used

Microsoft Excel 2021
  Data Cleaning & Processing
  Correlation Analysis
  Chart Visualization
  Financial Insights

Project by – Alokk Joshi

After Cleaning the data,
we have
4817 rows
to apply analytics on

Task 1

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.

Top 10 Genres

Unique Genres and their occurance in the data		Descriptive Statistics of all the genres							
Genres	Count	Mean	Median	Mode	Max	Min	Range	VAR	SD
Drama	2461	6.7	6.8	7.2	8.8	3.3	5.5	0.9	0.9
Comedy	1813	6.2	6.3	6.3	9.5	1.9	7.6	1.2	1.1
Thriller	1337	6.3	6.4	6.4	7.5	3.4	4.1	1.1	1.0
Action	1095	6.2	6.3	6.6	9.1	3.7	5.4	1.2	1.1
Romance	1067	6.4	6.5	6.5	6.6	5.1	1.5	1.0	1.0
Adventure	878	6.4	6.6	6.6	7.8	3.3	4.5	1.3	1.1
Crime	838	6.5	6.6	6.6	0.0	0.0	0.0	1.0	1.0
Sci-Fi	578	6.2	6.3	6.7	6.3	2.8	3.5	1.5	1.2
Fantasy	569	6.3	6.4	6.7	6.8	6.4	0.4	1.3	1.2
Horror	533	5.8	5.9	6.2	8.0	2.2	5.8	1.2	1.1
Family	523	6.2	6.3	6.7	8.6	3.2	5.4	1.4	1.20
Mystery	461	6.4	6.5	6.4	0	0	0	1.1	1.06
Music	322	6.4	6.7	7.1	7.2	7.2	0	1.4	1.20
Biography	289	7.1	7.2	7.0	5.7	5.7	0	0.5	0.72
Animation	233	6.5	6.7	6.7	4.8	3.7	1.1	1.3	1.14
War	206	7.1	7.1	7.1	0	0	0	0.8	0.87
History	199	7.1	7.2	7.5	7.5	7.5	0	0.8	0.88
Sport	176	6.6	6.8	7.2	0	0	0	1.2	1.10
Musical	131	6.5	6.7	7.0	3.4	3.4	0	1.5	1.22
Documentary	120	7.2	7.4	7.5	8.7	2.7	6	1.1	1.06
Western	94	6.7	6.8	6.5	8.9	3.8	5.1	1.1	1.05
Film-Noir	6	7.6	7.7	0.0	0	0	0	0.2	0.39
Short	5	6.4	6.5	0.0	0	0	0	0.4	0.67
News	3	7.5	7.4	0.0	0	0	0	0.2	0.42

Project by – Alokk Joshi

Insights from top 10 genres and their imdb scores

Unique Genres and their occurance in the data		Descriptive Statistics of all the genres							
Genres	Count	Mean	Median	Mode	Max	Min	Range	VAR	SD
Drama	2461	6.7	6.8	7.2	8.8	3.3	5.5	0.9	0.9
Comedy	1813	6.2	6.3	6.3	9.5	1.9	7.6	1.2	1.1
Thriller	1337	6.3	6.4	6.4	7.5	3.4	4.1	1.1	1.0
Action	1095	6.2	6.3	6.6	9.1	3.7	5.4	1.2	1.1
Romance	1067	6.4	6.5	6.5	6.6	5.1	1.5	1.0	1.0
Adventure	878	6.4	6.6	6.6	7.8	3.3	4.5	1.3	1.1
Crime	838	6.5	6.6	6.6	0.0	0.0	0.0	1.0	1.0
Sci-Fi	578	6.2	6.3	6.7	6.3	2.8	3.5	1.5	1.2
Fantasy	569	6.3	6.4	6.7	6.8	6.4	0.4	1.3	1.2
Horror	533	5.8	5.9	6.2	8.0	2.2	5.8	1.2	1.1

Here are key insights based on the **top 10 genres**:

1. Drama is the most frequent genre (2,461 movies) and has the lowest variance (0.9), indicating stable IMDb ratings.
2. Comedy has the widest rating range (7.6), showing strong audience polarization.
3. Crime and Romance share the second-lowest variance (1.0), implying consistent audience reception.
4. Comedy movies have the highest maximum rating (9.5), suggesting some top-rated blockbuster hits.
5. Fantasy and Sci-Fi have the highest standard deviations (1.2), showing varied audience opinions.
6. Thriller, Action, and Horror have the lowest median ratings (6.3, 6.3, 5.9), indicating they are more critically mixed.
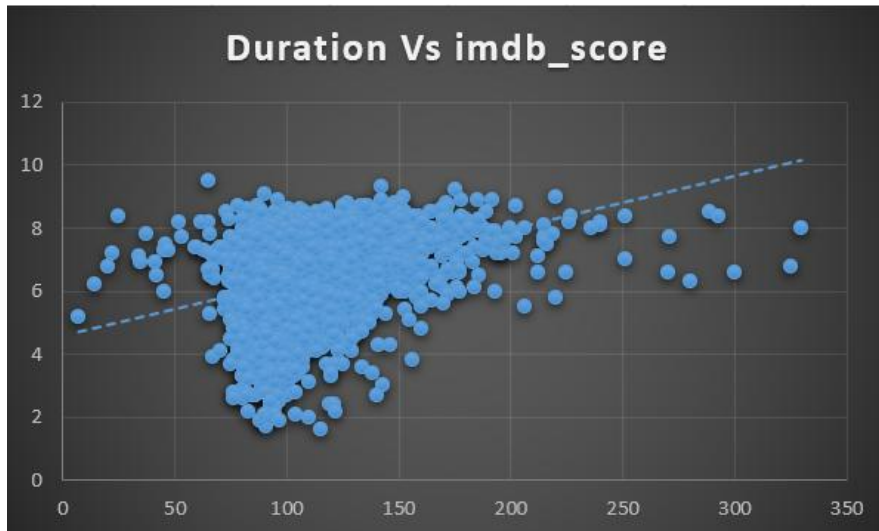
Task 2

Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.
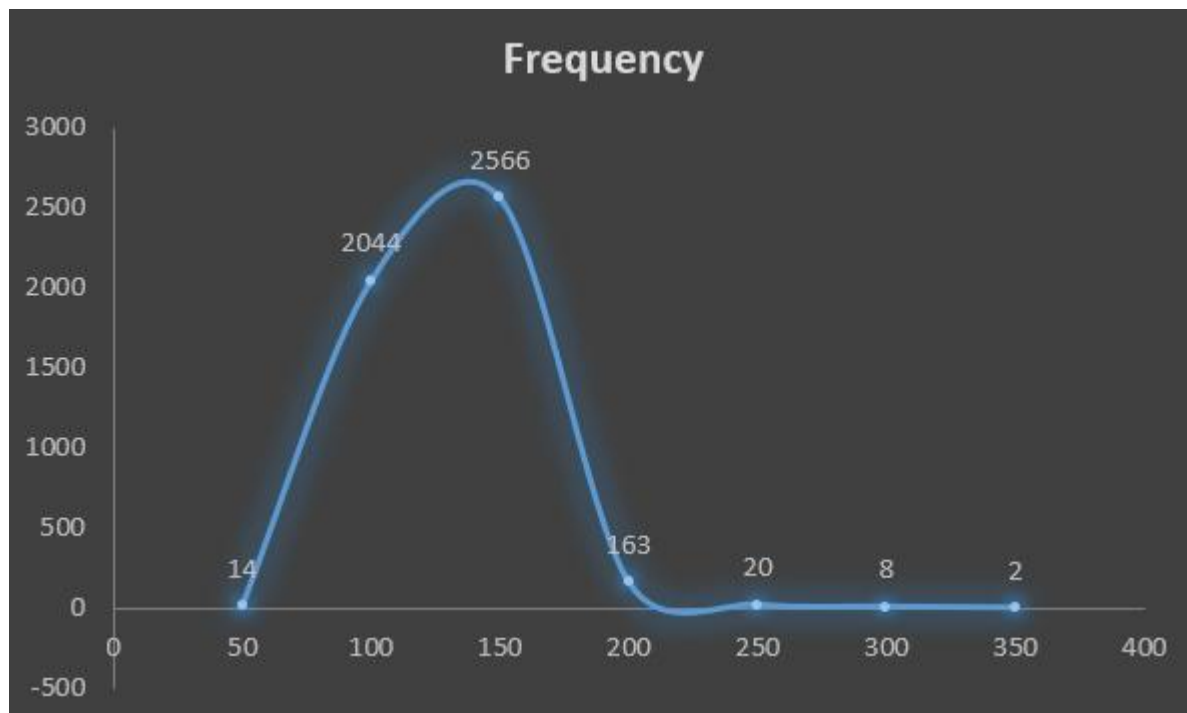
Class	Frequency
1-50	14
50-100	2044
100-150	2566
150-200	163
200-250	20
250-300	8
300-350	2

Maximum movies fall between 50 and 150 minutes of duration

Duration Vs imdb_score

- There is a weak positive correlation between movie duration and IMDb score, **meaning longer movies tend to have slightly higher ratings.**
- **Most movies fall between 80-150 minutes, forming a dense cluster**, indicating a standard duration range for films.
- **Movies shorter than 60 minutes mixed ratings between 5.5 and 8.2, whereas many movies with duration between 80 and 130 also receive lower ratings.**
- A few outliers above 250 minutes exist, but their ratings between 6 and 8, vary widely, showing no strong trend among extremely long films.
- The trendline suggests a gradual increase in IMDb score with duration, but the spread of data points implies other factors also influence ratings.

Task 3

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

Below are the descriptive stats for top 10 languages from the max movie count

Language	Movie Count	Mean	Median	Mode	Max	Min	Range	VAR	SD
English	4500	6.37	6.50	6.70	9.50	1.60	7.90	1.24	1.12
French	72	7.02	7.20	7.20	8.40	4.90	3.50	0.51	0.71
Spanish	40	6.94	7.15	7.20	8.20	4.40	3.80	0.71	0.84
Hindi	28	6.63	6.95	7.80	8.50	2.80	5.70	1.89	1.37
Mandarin	24	6.79	7.05	7.60	7.90	3.20	4.70	1.03	1.02
German	19	7.34	7.60	7.40	8.50	4.90	3.60	0.86	0.93
Japanese	16	7.37	7.60	8.20	8.70	5.60	3.10	0.99	1.00
Cantonese	11	6.95	7.20	6.50	7.80	5.30	2.50	0.45	0.67
Russian	11	6.36	6.50	5.30	8.10	4.10	4.00	1.74	1.32
Italian	10	7.08	7.15	0.00	8.90	5.10	3.80	1.31	1.14

Movie count-wise top 10 languages

Language	No of Movies	Average of imdb_score
English	4500	6.4
French	72	7.0
Spanish	40	6.9
Hindi	28	6.6
Mandarin	24	6.8
German	19	7.3
Japanese	16	7.4
Russian	11	6.4
Cantonese	11	7.0
Italian	10	7.1

Highest imdb rating wise top 10 languages

Language	No of Movies	Average of imdb_score
Telugu	1	8.4
None	2	8.0
Indonesian	2	7.9
Maya	1	7.8
Hebrew	5	7.6
Persian	4	7.6
Danish	5	7.5
Dari	2	7.5
Dzongkha	1	7.5
Portuguese	8	7.5

Here are major insights

English Dominates Quantity but Not Quality – English is the most common language with 4500 movies, but its average IMDB score (6.4) is lower than several other languages.

Japanese and German Have High Ratings – Among the top 10 most common languages, Japanese (7.37) and German (7.34) have the highest average IMDB scores.

French and Italian Perform Well – French (7.02) and Italian (7.08) maintain strong average IMDB scores, suggesting that movies in these languages are generally well-received.

Higher Scores in Less Common Languages – Less frequently used languages like Telugu (8.4), Indonesian (7.9), and Hebrew (7.6) tend to have higher IMDB scores, indicating that movies in these languages may be more selective or higher quality. **Due to a smaller number of movies, they cannot be considered in drawing big conclusions.**

Wide Score Range in English and Hindi – English movies have the widest IMDB score range (7.9), followed by Hindi (5.7), showing that quality varies significantly within these languages.

Lowest Variance in Cantonese Movies – Cantonese has the lowest variance (0.45) and standard deviation (0.67), meaning its IMDB scores are more consistent compared to other languages.

Task 4

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

Here are the analytics

Total Unique Directors 2398

Movie Count-wise top directors			
Directors	Movies	Avg IMDB Score	Percentile Rank
Steven Spielberg	26	7.5	89%
Woody Allen	22	7.0	75%
Martin Scorsese	20	7.7	93%
Clint Eastwood	20	7.2	83%
Spike Lee	16	6.6	56%
Ridley Scott	16	7.1	79%
Steven Soderbergh	15	6.7	61%
Renny Harlin	15	5.7	28%
Tim Burton	14	7.1	76%
Oliver Stone	14	7.0	72%
Ron Howard	13	6.9	72%
Robert Zemeckis	13	7.3	86%
Robert Rodriguez	13	5.7	26%
Joel Schumacher	13	6.4	51%
Barry Levinson	13	6.6	57%

IMDB Score-wise Directors			
Directors	Movies	Avg IMDB Score	Percentile Rank
John Blanchard	1	9.5	100%
Sadyk Sher-Niyaz	1	8.7	100%
Mitchell Altieri	1	8.7	100%
Cary Bell	1	8.7	100%
Mike Mayhall	1	8.6	100%
Charles Chaplin	1	8.6	100%
Ron Fricke	1	8.5	100%
Raja Menon	1	8.5	100%
Majid Majidi	1	8.5	100%
Damien Chazelle	1	8.5	100%
Sergio Leone	4	8.5	100%

Project by – Alokk Joshi

Here are major insights

Martin Scorsese has the best score among top directors – He has directed 20 movies with an average IMDB score of 7.7, placing him in the 93rd percentile, meaning his movies are highly rated.

Steven Spielberg is the most consistent – With 26 movies and an average IMDB score of 7.5 (89th percentile), he has maintained strong ratings across many films.

Clint Eastwood and Woody Allen have good ratings – Both have directed over 20 movies. Clint Eastwood has an average IMDB score of 7.2 (83rd percentile), while Woody Allen has 7.0 (75th percentile).

Some directors have made many movies but have lower ratings – Renny Harlin (5.7 IMDB score) and Steven Soderbergh (6.7 IMDB score) have directed several movies, but their ratings are not as high.

Directors with the highest IMDB scores have made only one movie – John Blanchard (9.5 IMDB score) and Sadyk Sher-Niyaz (9.0 IMDB score) have the best scores, but they have only directed a single film.

Tim Burton and Robert Zemeckis are consistently good – Tim Burton (14 movies, 7.1 score) and Robert Zemeckis (13 movies, 7.3 score) have made multiple movies that are well-rated by audiences.

Task 5

Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Some Cleaning First

Correlation = 0.0993131

It means there is very faith relationship between the budget and the gross earning

Even after filling the budget and gross base director the correlation is very far from 1. Removing all the 0 budget and gross.

Now we have 4131 row of data to analyze the budge and gross relationship

Here are top 10 movies with max profit

Top 10 most profit making movies using Large() function					
Max Profit	Budget	Director	Movie_title	% Profit over Budget	IMDB Score
523505847	237000000	James Cameron	Avatar	221%	7.9
502177271	150000000	Colin Trevorrow	Jurassic World	335%	7
458672302	200000000	James Cameron	Titanic	229%	7.7
449935665	11000000	George Lucas	Star Wars: Episode IV - A New Hope	4090%	8.7
424449459	10500000	Steven Spielberg	E.T. the Extra-Terrestrial	4042%	7.9
403279547	220000000	Joss Whedon	The Avengers	183%	8.1
377783777	45000000	Roger Allers	The Lion King	840%	8.5
359544677	115000000	George Lucas	Star Wars: Episode I - The Phantom Menace	313%	6.5
348316061	185000000	Christopher Nolan	The Dark Knight	188%	9
329999255	78000000	Gary Ross	The Hunger Games	423%	7.3

Project by – Alokk Joshi

Here are top 5 insights from profit study

A high-profit movie doesn't always have the best rating – The Dark Knight has the highest IMDb rating (9.0) but made a lower profit ($348M, 188% profit) than Avatar ($523M, 221%) and Jurassic World ($502M, 335%), which have lower ratings (7.9 and 7.0).

Small-budget movies can still make huge profits – Star Wars: Episode IV - A New Hope had a budget of only $11M but made a massive profit of $449M, achieving an outstanding 4090% profit margin with a strong IMDb rating of 8.7. Similarly, E.T. the Extra-Terrestrial turned a $10.5M budget into $424M profit (4042%). This shows that a good story and a strong fan base matter more than just big budgets.

Superhero and animated movies are both profitable and well-rated – The Avengers (IMDb 8.1, $403M profit, 183%) and The Lion King (IMDb 8.5, $377M profit, 840%) show that popular franchises can make big money and also get high ratings.

A low rating doesn't always mean low profit – Star Wars: Episode I - The Phantom Menace has the lowest IMDb score (6.5) but still made $359M profit (313%), likely because Star Wars has a huge fan following.

Christopher Nolan's The Dark Knight is the best-rated movie on the list – It has the highest IMDb rating (9.0) and still made a strong profit of $348M (188%), proving that a movie can be both critically loved and financially successful.

Project by – Alokk Joshi
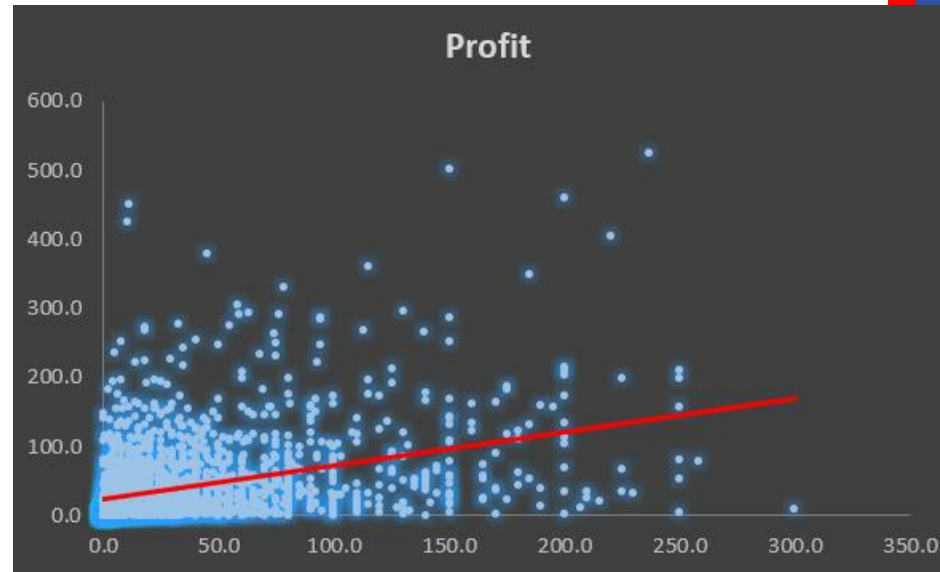
Scatter Plot (Budget vs. Profit)

- **Positive Correlation:** The red trend line indicates a positive relationship between budget and profit—higher budgets generally lead to higher profits.
- **High Variability:** While some high-budget movies earn significant profits, many lower-budget movies also perform well, showing that budget alone doesn't guarantee success.
- **Outliers Exist:** A few movies with extremely high profits stand out, suggesting that exceptional films can break trends and outperform expectations.
- **Cluster at Lower Budgets:** Most movies have lower budgets and profits, implying that smaller films dominate the industry, with only a few high-budget blockbusters.
- **Budget is a Factor, Not a Guarantee:** While there is a general upward trend, the wide spread of points shows that other factors play a crucial role in profitability.

3

5 Whys 1

Why do some low-budget movies still achieve high profits?
They have strong storytelling and audience engagement.
Why does strong storytelling lead to high profits?
Engaging plots attract word-of-mouth marketing, leading to more viewers.
Why does word-of-mouth marketing increase viewership?
People trust recommendations from friends and family more than advertisements.
Why do personal recommendations matter more than advertisements?
They feel more authentic and unbiased.
Why does authenticity matter in movie success?
It builds credibility and long-term fan loyalty.

5 Whys 2

Why do some movies with lower IMDb ratings still make high profits?

They belong to well-established franchises with loyal fan bases.

Why do established franchises attract large audiences despite lower ratings?

Fans are emotionally invested in the characters and storylines.

Why does emotional investment lead to box office success?

Fans watch the movies regardless of critical reception.

Why do fans continue watching despite poor reviews?

They want to stay updated with the franchise's story.

Why does staying updated matter?

It keeps them connected to the larger fan community and cultural trends.

5 Whys 2

Why do movies directed by well-known directors tend to perform better?

Famous directors have built trust and a loyal audience.

Why does a director's reputation impact movie success?

Audiences associate their name with quality filmmaking.

Why do people trust well-known directors?

Their past successful movies have created a strong track record.

Why does a strong track record influence audience behavior?

It reduces the perceived risk of watching a new movie.

Why does reducing risk make people more likely to watch a movie?

Viewers feel confident they will get a good experience, leading to strong box office performance.

My Key Learnings from IMDb Movie Analysis

➢ **Clean data is crucial** – Before analyzing, it's important to remove errors, standardize formats, and organize data properly.

➢ **Numbers need context** – Looking at percentages (like profit % over budget) alongside absolute values gives a clearer picture.

➢ **Data reveals patterns** – Genre, director, and budget influence a movie's success, just like key factors drive outcomes in any industry.

➢ **Statistics support decisions** – Using metrics like correlation, outliers, and averages helps in making informed business choices.

➢ **Presentation makes insights powerful** – Presenting data in a clear and structured way is as important as analyzing it.