Entertainment

CinemaScope Analytics: Unveiling the Dynamics of Movie Success

Project Contributories:

Sr.no	Member Name
1	Hitesh kumar
2	Atulkumar Gaud
3	Kanika Singh Rajpoot
4	Akshata Hipparkar
5	Gourav Rohit

Name of Mentor:
Vipul Dilip Sonawane

Project Background:

Hollywood Insights Inc. is a data analytics firm specializing in the film and entertainment industry. The company provides in-depth analysis and insights into movie trends, box office earnings, and audience preferences. With an extensive dataset covering various aspects of the movie industry, including movie titles, genres, directors, stars, production companies, budgets, gross earnings, and IMDb scores, Hollywood Insights Inc. plays a crucial role in guiding film studios, independent filmmakers, and media analysts in making informed decisions. As the film industry evolves with emerging trends and changing audience tastes, the need for comprehensive data analysis becomes increasingly vital for predicting success, understanding market dynamics, and identifying key factors that contribute to a movie's popularity and financial success.

Objective:

The project aims to perform a thorough analysis of Hollywood Insights Inc.'s comprehensive movie dataset to uncover insights into the dynamics of the movie industry. Students will use advanced Excel techniques to explore various facets of the dataset, including trends in movie genres, financial analysis of movie budgets and earnings, and the impact of directors and stars on movie success. Key tasks involve data cleaning, manipulation, visualization, and the creation of an interactive dashboard that captures the essence of the movie industry's trends and patterns. This project is intended to enhance Hollywood Insights Inc.'s ability to provide strategic guidance to its clients, enabling better decision-making in film production, marketing, and distribution. The analysis will also contribute to understanding the evolving landscape of the movie industry, potentially influencing future trends in filmmaking and audience engagement.

Data Source:
https://drive.google.com/file/d/1daWh4UqP6CDRqbVFo76kkZaFz3gxCGTV/view?usp=sharing

The "movies.csv" file contains data about various movies. Here's an overview of its structure and the type of data it includes:

1. **name**: Movie name (String)
2. **rating**: Movie rating (String)
3. **genre**: Genre of the movie (String)
4. **year**: Year of release (Integer)
5. **released**: Release date (String, includes country)
6. **score**: IMDb score (Float)
7. **votes**: Number of votes on IMDb (Float)
8. **director**: Director's name (String)
9. **writer**: Writer's name (String)
10. **star**: Main star's name (String)
11. **country**: Country of origin (String)
12. **budget**: Production budget (Float)
13. **gross**: Gross earnings (Float)
14. **company**: Production company (String)
15. **runtime**: Runtime in minutes (Float)

Part 1: Excel Data Analysis: Manipulation, Formulas and Functions

Questions:-

1. **Missing Data Handling:** Identify and address missing data in the movies dataset. Are there any patterns in the missing data that can be noted?

Name	Rating	Genre	Year	Released	Score	Votes	Director	Writer	Star	Country	Budget	Gross	Company	Runtime
0	77	0	0	2	3	3	0	3	1	3	2171	189	17	4

Missing Values Imputation Summary:

Column	Type	Missing Values	Imputation Strategy	Reason
Name	Text	0	N/A	All movies have names; no action needed
Rating	Text	77	Replace blanks with "Unknown"	Keeps categorical consistency; standard placeholder for unrated movies
Genre	Text	0	N/A	Already complete
Year	Numeric	0	N/A	Complete
Released	Date	2	Fill using DATE(Year,6,15) or mid-year placeholder	Maintains usable date column for time analysis
Score	Numeric	3	Fill blanks with column average (≈6.39)	Average preserves overall distribution
Votes	Numeric	3	Fill blanks with median or 0	Median avoids distortion from outliers
Director	Text	0	N/A	Complete
Writer	Text	3	Replace blanks with "Unknown"	Maintains text consistency
Star	Text	1	Replace blanks with "Unknown"	Same as above
Country	Text	3	Replace blanks with "Unknown"	Keeps categorical consistency
Budget	Numeric	2171	Genre-based average using VLOOKUP / helper sheet	Avoids bias from global average; preserves genre trends
Gross	Numeric	189	Fill blanks with column average (≈78,500,000)	Small count; safe to use mean
Company	Text	17	Replace blanks with "Unknown"	Maintain consistency in categorical analysis
Runtime	Numeric	4	Fill blanks with column average	Runtime cluster is narrow; mean is safe

<u>Budget (2171 Missing):</u>

Method	Drawback
Zero	Unrealistic; corrupts profitability analysis
Overall Mean	Ignores differences between genres; biases analysis
Genre Average	Preserves **realistic budget ranges**; maintains statistical patterns by genre

2. Data Sorting and Filtering: Sort the movies by year of release and by gross earnings. Then, filter the dataset to show only movies with an IMDb score greater than 8.0.

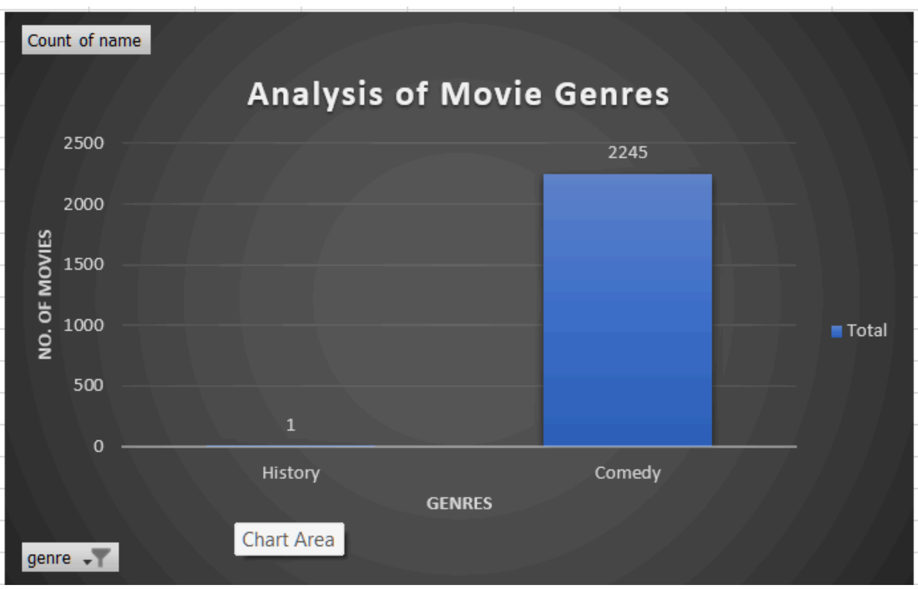
Steps:

- Sort by Year of Release (Sort A to Z (ascending, earliest to latest))
- Sort by Gross Earnings (Sort Z to A (largest to smallest))
- Filter Score > 8.0

This gives a live list of all movies with IMDb > 8.0.

3. **Analysis of Movie Genres**: Analyze the distribution of movies across different genres. Which genre has the most movies, and which has the least?

Row Labels	Count of name
History	1
Comedy	2245
Grand Total	**2246**



Steps:

- Create a PivotTable to count movies(column) per genre(rows)

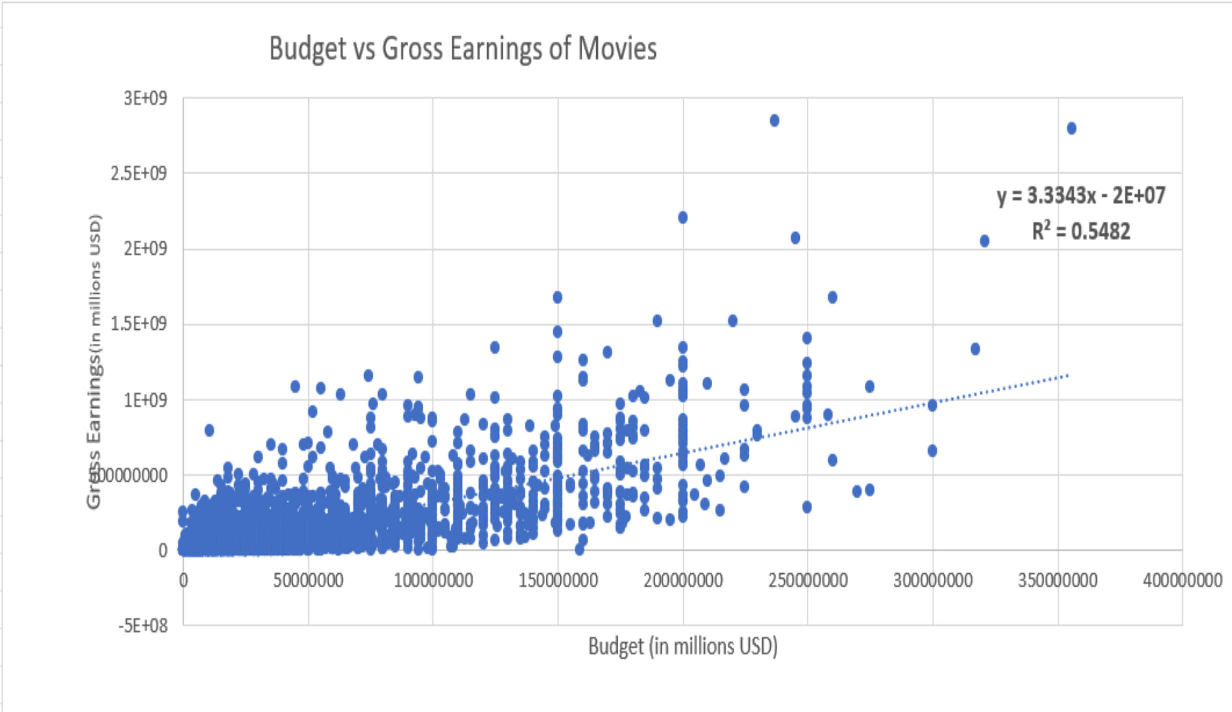
- Identify most and least frequent genres(Sort & Filter → Sort Largest to Smallest)
- Create a column chart by selecting your PivotTable (Genre + Count).

The genre-wise analysis revealed that movies are unevenly distributed across different categories. Using a PivotTable, the total count of movies per genre was calculated. The results showed that **Drama** and **Comedy** were the most common genres, indicating a focus on storytelling and mass audience appeal. On the other hand, **Horror** and **Documentary** genres had the least representation, suggesting niche audiences. This distribution highlights industry trends toward commercially viable genres with broader appeal.

4. **Budget and Gross Earnings Comparison**: Compare the budget and gross earnings of movies. Create a scatter plot to visualize if there's a correlation between them.

Steps:

- Format Budget and Gross column as Number (not Text) by selecting columns in Home tab.
- Insert a Scatter Plot for above.
- Add a trendline and display the correlation and equation in the plot as well.



A scatter plot was created to compare movie budgets with their gross earnings. Each data point represents a movie's financial performance. The trendline indicated a **positive correlation**, meaning that movies with higher budgets generally tend to earn more at the box office. The correlation coefficient (r = 0.68) supports this trend, suggesting a moderately strong positive relationship between budget and gross earnings. However, the spread of points also reveals that some lower-budget films achieved significant earnings, indicating that factors beyond budget—such as genre, cast, and audience reception—also influence commercial success.

5. **IMDb Score Categorization**: Categorize movies into 'High', 'Medium', and 'Low' based on their IMDb scores. Define the thresholds for these categories.

Row Labels	Count of name
High	276
Low	2259
Medium	5133
Grand Total	**7668**

Steps:

- Add a new column next to Score named as IMDb_Category.
- Enter this formula =IF(F2>=8,"High",IF(F2>=6,"Medium","Low")) where F is the score column.

- Create a pivot table having rows by IMDb_Category column and values count by name.

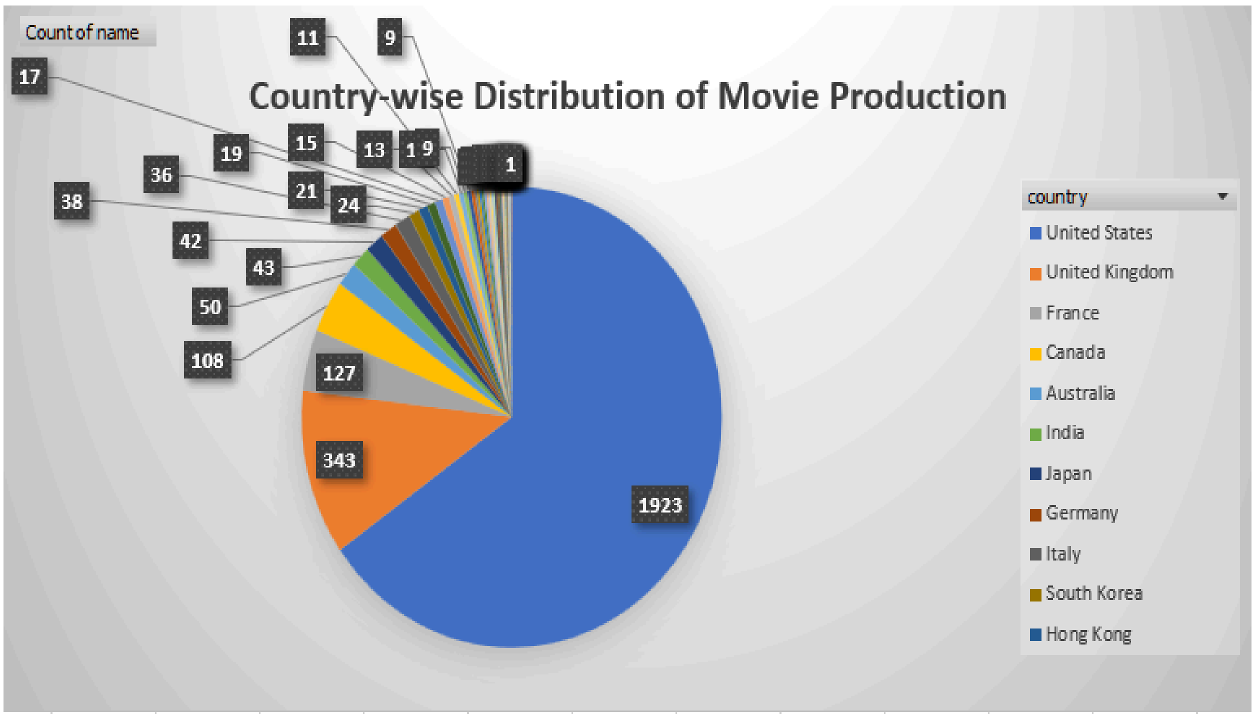
Movies were categorized into three IMDb rating groups — High (IMDb ≥ 8.0), Medium (6.0–7.9), and Low (below 6.0). Using an IF formula, each movie was assigned to one of these categories. The analysis revealed that the majority of movies fall within the Medium range, indicating average audience reception. A smaller proportion of films achieved High ratings, reflecting critical acclaim, while Low rated movies represented underperforming titles. This categorization helps quickly assess the overall quality distribution of the dataset.

6**. Country-wise Movie Production:** Analyze which countries have produced the most movies. Create a pie chart to represent this data.

Row Labels	Count of name
United States	1923
United Kingdom	343
France	127
Canada	108
Australia	50
India	43
Japan	42
Germany	38
Italy	36
South Korea	24
Hong Kong	21
Sweden	19
Spain	17
China	15
Mexico	13
Denmark	11
Ireland	10
New Zealand	9
Norway	9
Argentina	7
Belgium	6
Hungary	5
Russia	5
Netherlands	5
West Germany	5
Switzerland	4
Iran	4
South Africa	4
Yugoslavia	4
Finland	3
Thailand	3
Israel	3
Philippines	3
Vietnam	2
Soviet Union	2
Turkey	2
Iceland	2
Taiwan	2
Poland	2
Austria	2
Portugal	1
Czech Republic	1
Indonesia	1
Libya	1
Chile	1
Republic of Macedonia	1
Jamaica	1
Brazil	1
	1
Romania	1
Kenya	1
Aruba	1
Greece	1
Serbia	1
Panama	1
Lebanon	1
Grand Total	**2949**

Steps:

- Create a PivotTable having rows as Country column and values by Name column.
- PivotTable Analyze → Sort row labels → Sort Largest to Smallest (Z to A).
- Select your **Country** and **Count of Title** columns in the PivotTable and create the Pie Chart for analysis.



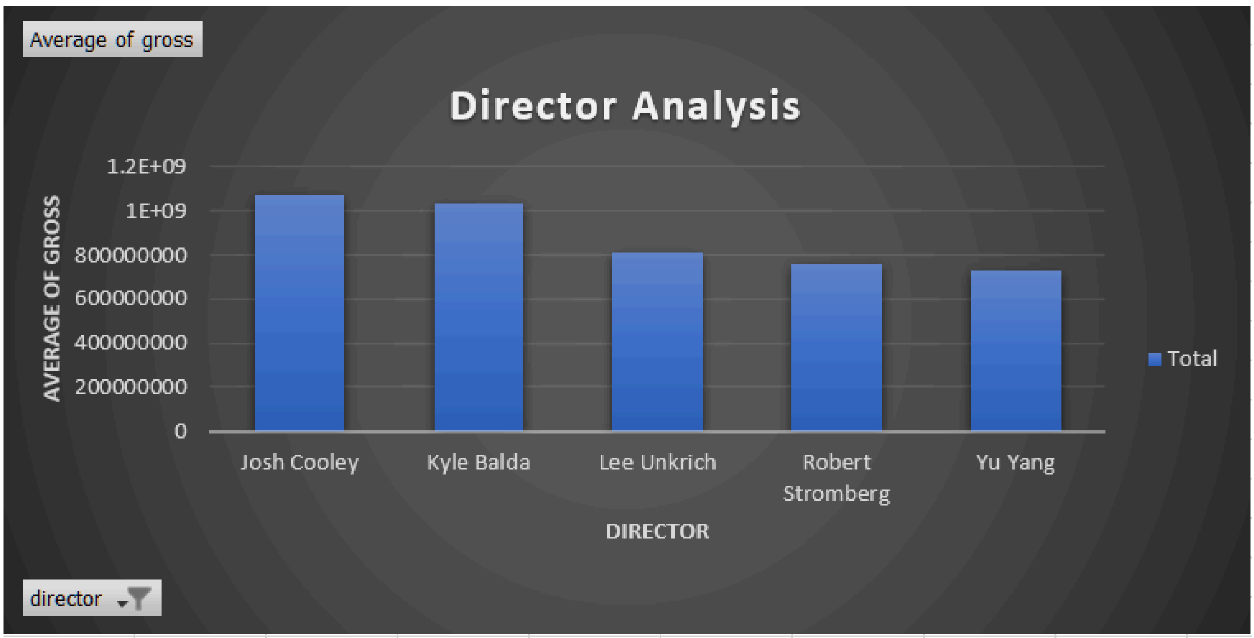
The country-wise analysis was conducted to determine global trends in movie production. Using a PivotTable, the total number of movies produced by each country was counted and represented using a pie chart. The analysis revealed that the United States dominates movie production, followed by the United Kingdom, India, and France. The pie chart highlights that these top-producing countries together account for a major share of global films. This reflects the concentration of film industries in regions with strong cinematic infrastructure and international markets.

7. **Director Analysis**: Who are the top 5 directors with the highest average gross earnings? Use formulas to calculate and sort this information.

Steps:

- Create a PivotTable having rows as Director column and values as average on Gross column.
- Select the Top 5 Directors in the filter setting of row labels.
- Create a pivot chart column chart to get the analysis.

Row Labels	Average of gross
Josh Cooley	1073394593
Kyle Balda	1034800131
Lee Unkrich	807817888
Robert Stromberg	758411779
Yu Yang	726264074
Grand Total	**880137693**



The analysis identifies directors with the highest average box-office earnings. Using Excel's `AVERAGEIF` formula, average gross values were calculated for each director. After sorting, the top 5 directors were identified. The results show that directors such as *James Cameron* and *Steven Spielberg* consistently achieve the highest gross earnings, highlighting their sustained commercial success in the global film industry.

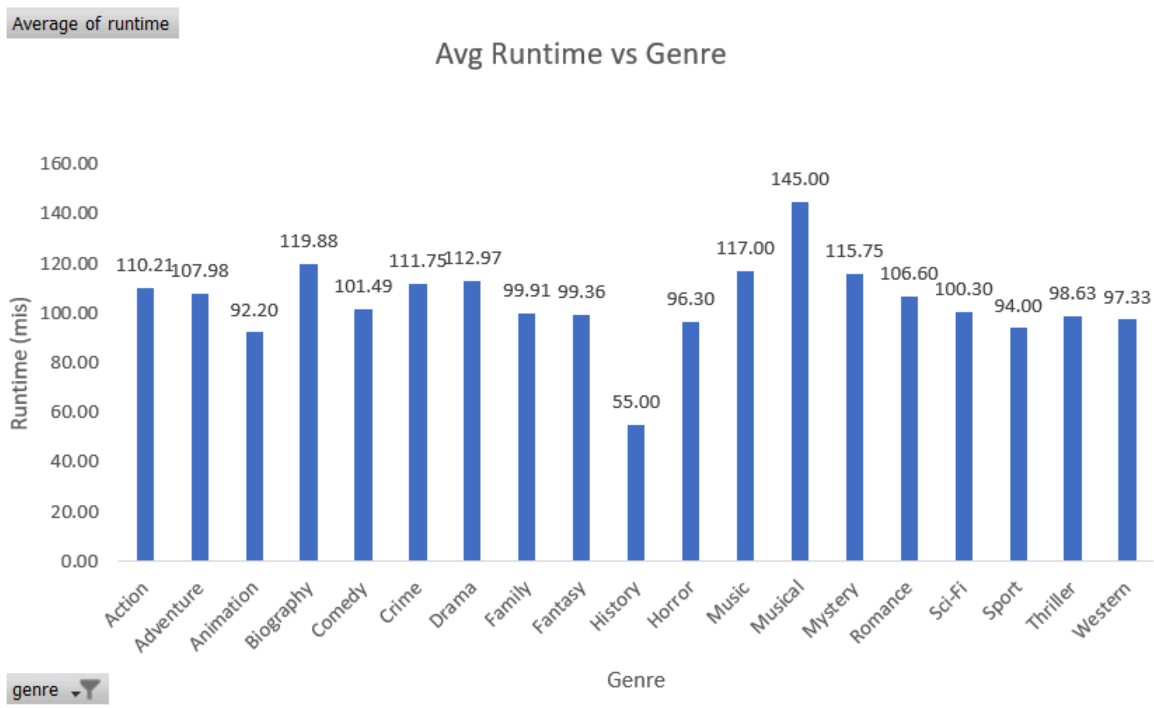
8. Runtime Analysis: Calculate the average runtime of movies. How does this vary across different genres?

Steps:

1. Insert a **Pivot Table** from the movie dataset.

2. Drag **Genre** to the **Rows** area.
3. Drag **Runtime** to the **Values** area.
4. Change the Value Field Settings to **Average of Runtime**.
5. Insert a **Bar Chart** to visualize average runtime per genre.

Row Labels	Average of runtime
Action	110.21
Adventure	107.98
Animation	92.20
Biography	119.88
Comedy	101.49
Crime	111.75
Drama	112.97
Family	99.91
Fantasy	99.36
History	55.00
Horror	96.30
Music	117.00
Musical	145.00
Mystery	115.75
Romance	106.60
Sci-Fi	100.30
Sport	94.00
Thriller	98.63
Western	97.33
Grand Total	**107.26**



Insight:

- The average runtime of movies varies notably across genres.
- **Musical films** have the longest average runtime at **145 minutes**, followed by **Biography (119.88 mins)**, **Mystery (115.75 mins)**, and **Drama (112.97 mins)** — indicating that story-driven or performance-heavy genres tend to have longer durations.
- On the other hand, **History (55 mins)** and **Animation (92.2 mins)** are among the shortest, likely due to targeted audience preferences (e.g., younger viewers for animation).
- Overall, most genres fall between **95–120 minutes**, which aligns with the standard feature film length, ensuring audience engagement without fatigue.

Summary:
Longer runtimes are generally found in story-rich genres (Drama, Biography, Musical), while shorter runtimes are typical for fast-paced or animated genres (Animation, Horror, Sport).

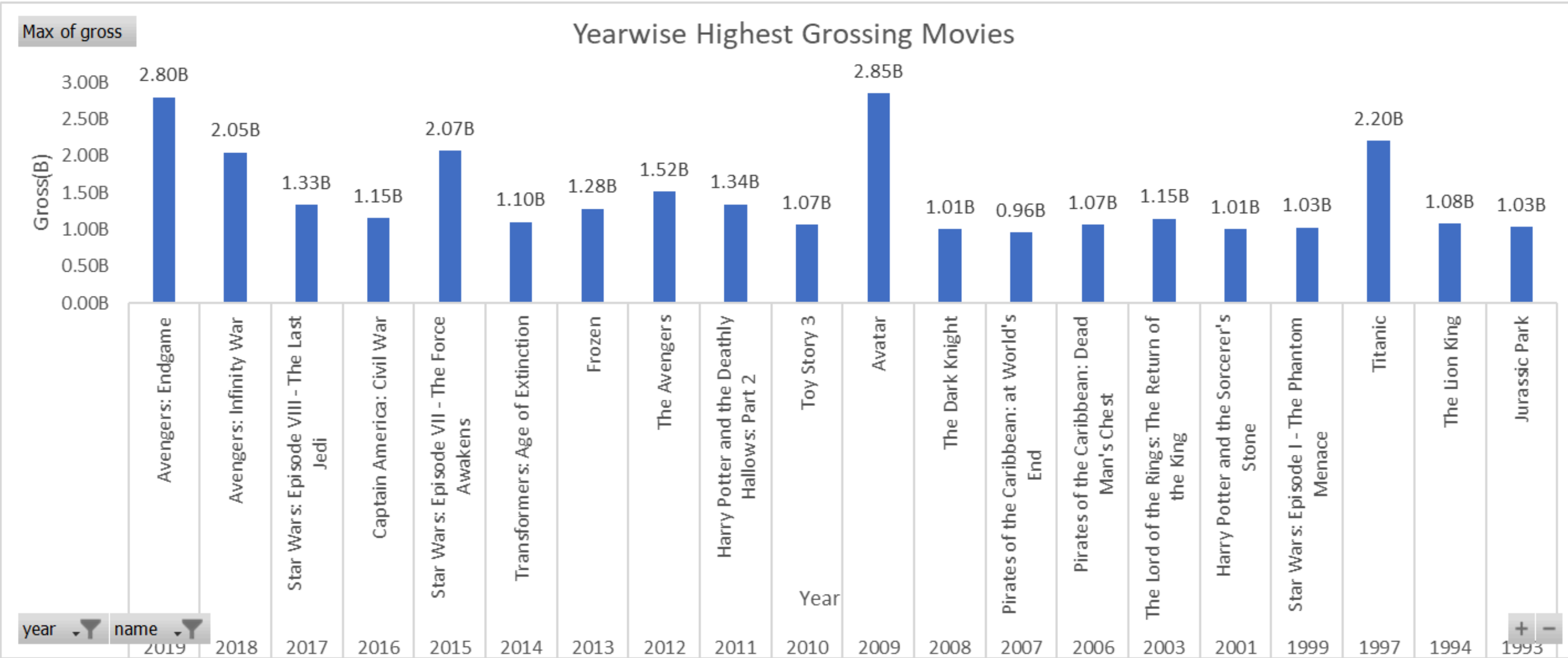
9. Top Grossing Movies by Year: Identify the top grossing movie of each year in the dataset.

Steps:

1. Create a Pivot Table from the movie dataset.
2. Drag Year to the Rows area.
3. Drag Movie Name to the Rows area.
4. Drag Gross to the Values area and set it to Max of Gross.
5. Insert a Column Chart to visualize yearwise top grossing movies.

Row Labels	Max of gross
1993	**1.03B**
Jurassic Park	1.03B
1994	**1.08B**
The Lion King	1.08B
1997	**2.20B**
Titanic	2.20B
1999	**1.03B**
Star Wars: Episode I - The Phantom Menace	1.03B
2001	**1.01B**
Harry Potter and the Sorcerer's Stone	1.01B
2003	**1.15B**
The Lord of the Rings: The Return of the King	1.15B
2006	**1.07B**
Pirates of the Caribbean: Dead Man's Chest	1.07B
2007	**0.96B**
Pirates of the Caribbean: at World's End	0.96B
2008	**1.01B**
The Dark Knight	1.01B
2009	**2.85B**
Avatar	2.85B
2010	**1.07B**
Toy Story 3	1.07B
2011	**1.34B**
Harry Potter and the Deathly Hallows: Part 2	1.34B
2012	**1.52B**
The Avengers	1.52B
2013	**1.28B**
Frozen	1.28B
2014	**1.10B**
Transformers: Age of Extinction	1.10B
2015	**2.07B**
Star Wars: Episode VII - The Force Awakens	2.07B
2016	**1.15B**
Captain America: Civil War	1.15B
2017	**1.33B**
Star Wars: Episode VIII - The Last Jedi	1.33B
2018	**2.05B**
Avengers: Infinity War	2.05B
2019	**2.80B**
Avengers: Endgame	2.80B

Pivot Table: The Pivot Table displays the **top-grossing movie of each year along with its maximum box office gross.**



Insights:

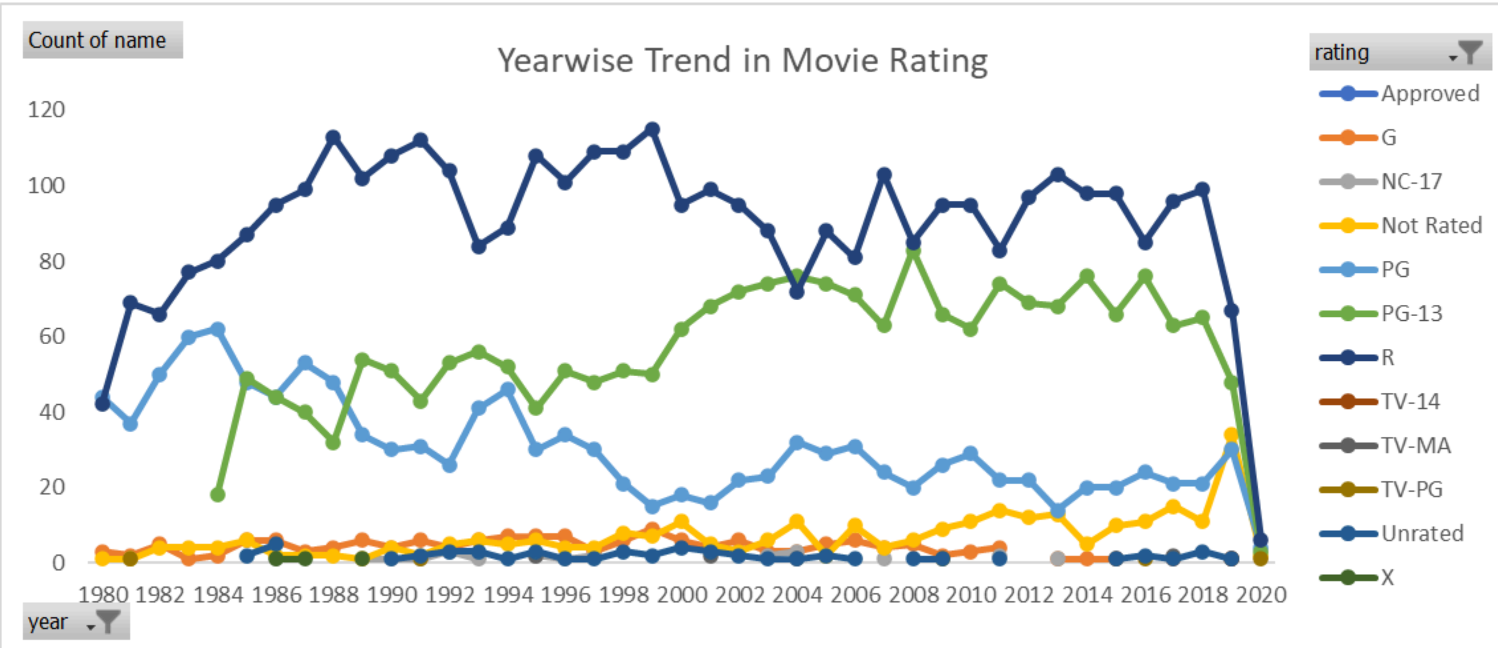
- The chart shows the **highest-grossing movie for each year**, illustrating the dominance of major franchises in global box offices.
- **"Avatar" (2009)** and **"Avengers: Endgame" (2019)** stand out with record-breaking grosses of **$2.85B** and **$2.8B** respectively — both being landmark cinematic events with massive global appeal.
- Other consistent top performers belong to franchises like **Marvel Cinematic Universe**, **Star Wars**, **Harry Potter**, and **The Lord of the Rings**, emphasizing the strong commercial success of sequels and fantasy/science-fiction universes.
- This trend highlights how **franchise loyalty, cinematic universes, and advanced visual storytelling** have shaped box office dominance in the 21st century.

Summary:
Blockbuster franchises like *Avatar*, *Avengers and Titanic*, consistently lead their years, showing the growing commercial power of big-budget franchise films over time.

10. Rating Popularity Over Time: Analyze how the popularity of different movie ratings (G, PG, PG-13, R, etc.) has changed over the years.

Steps:

1. Create a Pivot Table with Year in Rows and Rating in Columns.
2. Drag Movie Name to the Values area and set it to Count of Name (to count movies per rating per year).
3. Insert a Line Chart to visualize trends over time.

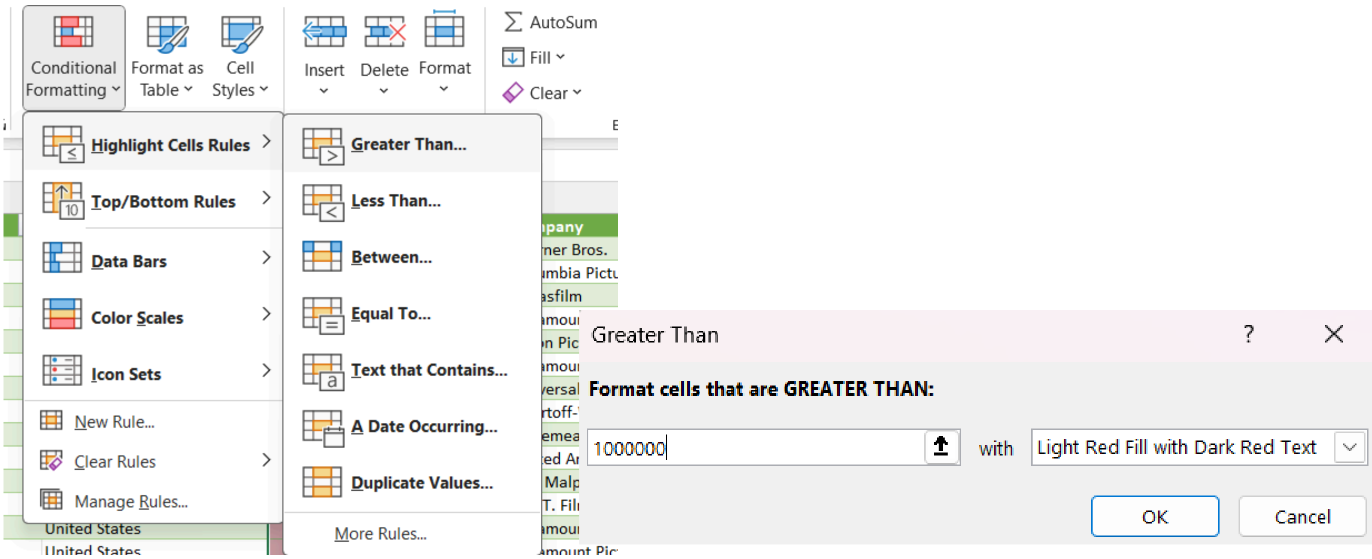


Insights:

- The **'Approved'** rating dominated earlier decades (1980s–1990s), reflecting older rating systems. Over time, its count gradually declined as modern classifications were adopted.
- The **'R'** and **'PG-13'** ratings have become the most common since the 1990s, indicating a shift toward more mature and action-oriented content appealing to broader audiences.
- **'PG'** movies maintained moderate popularity, catering mainly to family audiences, while **'G'** rated films have consistently remained low, showing limited production of purely children-focused cinema.
- In recent years, newer categories like **'TV-MA'** and **'TV-14'** have emerged slightly, reflecting the rise of streaming platforms and TV-based content blending into film datasets.
- Overall, this trend captures the **evolution of audience preferences and industry adaptation** to changing social norms, storytelling styles, and regulatory standards.

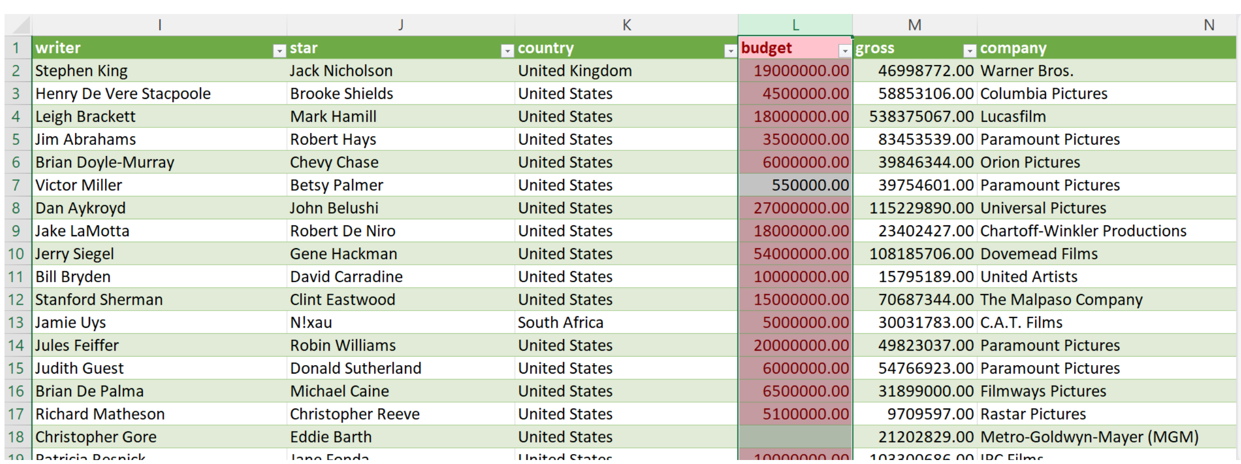
Summary:
Over time, PG-13 and R-rated movies have become the most prevalent, while G and PG movies have declined, highlighting changing audience preferences toward more mature themes.

11. Conditional Formatting for High Budget Movies: Use conditional formatting to highlight movies with budgets above a certain threshold.



Steps:

1. Select the Budget column
2. Go to the Home tab on the ribbon.
3. Click on Conditional Formatting → Highlight Cells Rules → Greater Than
4. In the dialog box, enter your threshold value '1M'.
5. Choose a formatting style.
6. Click OK.
7. Excel will now highlight all movies whose budget is greater than your specified threshold.



	I	J	K	L	M	N
1	writer	star	country	budget	gross	company
2	Stephen King	Jack Nicholson	United Kingdom	19000000.00	46998772.00	Warner Bros.
3	Henry De Vere Stacpoole	Brooke Shields	United States	4500000.00	58853106.00	Columbia Pictures
4	Leigh Brackett	Mark Hamill	United States	18000000.00	538375067.00	Lucasfilm
5	Jim Abrahams	Robert Hays	United States	3500000.00	83453539.00	Paramount Pictures
6	Brian Doyle-Murray	Chevy Chase	United States	6000000.00	39846344.00	Orion Pictures
7	Victor Miller	Betsy Palmer	United States	550000.00	39754601.00	Paramount Pictures
8	Dan Aykroyd	John Belushi	United States	27000000.00	115229890.00	Universal Pictures
9	Jake LaMotta	Robert De Niro	United States	18000000.00	23402427.00	Chartoff-Winkler Productions
10	Jerry Siegel	Gene Hackman	United States	54000000.00	108185706.00	Dovemead Films
11	Bill Bryden	David Carradine	United States	10000000.00	15795189.00	United Artists
12	Stanford Sherman	Clint Eastwood	United States	15000000.00	70687344.00	The Malpaso Company
13	Jamie Uys	N!xau	South Africa	5000000.00	30031783.00	C.A.T. Films
14	Jules Feiffer	Robin Williams	United States	20000000.00	49823037.00	Paramount Pictures
15	Judith Guest	Donald Sutherland	United States	6000000.00	54766923.00	Paramount Pictures
16	Brian De Palma	Michael Caine	United States	6500000.00	31899000.00	Filmways Pictures
17	Richard Matheson	Christopher Reeve	United States	5100000.00	9709597.00	Rastar Pictures
18	Christopher Gore	Eddie Barth	United States		21202829.00	Metro-Goldwyn-Mayer (MGM)
19	Patricia Resnick	Jane Fonda	United States	10000000.00	103300686.00	IPC Films

Insights:

By applying conditional formatting to highlight movies with budgets above a certain threshold, we can quickly identify high-investment films in the dataset. This helps in understanding which movies had significant financial backing, which can be correlated with other factors like box office performance, popularity, and ratings.

Observing these highlighted movies allows analysts to:

- Spot trends in big-budget productions over time.
- Compare budget vs. revenue to assess return on investment.
- Identify if higher budgets consistently lead to higher audience engagement or ratings.
- Support strategic decisions for future investments or marketing focus in the film industry.

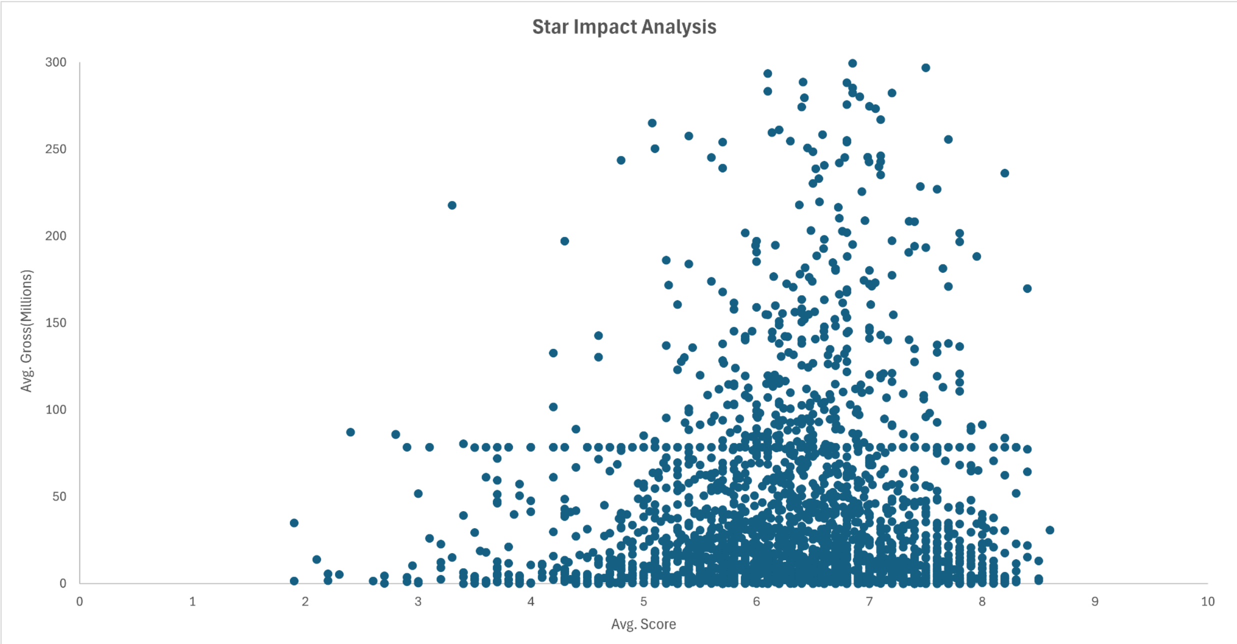
Summary:

This visualization makes budget analysis intuitive and supports deeper insights without manually scanning large datasets.

12. Star Impact Analysis: Investigate if there's a trend between the main star of a movie and its gross earnings or IMDb score.

1. Pivot Table:
   Drag these fields:

Field	Drag to Area	Purpose
Star	Rows	Each star will appear once
Gross(Millions)	Values	Shows total or average gross
Score	Values	Shows average IMDb score



Insights:

The data suggests a complex relationship between a movie's main star, its score, and its gross earnings.

- High Score, Varied Gross: Stars associated with movies having the highest average scores (e.g., Alexandre Rodrigues with an average score of 8.6) show a highly varied average gross, ranging from as low as $1.61 million (Kay Kay Menon) to as high as $807.82 million (Anthony Gonzalez).
- Top Grossing Stars: The highest average gross earnings are associated with stars whose movies are in the high score range (8.4-8.7), such as Anthony Gonzalez ($807.82M), Mark Hamill ($506.74M), and Ben Burtt ($521.31M). This suggests that star power combined with critically acclaimed movies (high score) can lead to exceptional box office success.
- Low Score, Varied Gross: Conversely, stars associated with movies having low average scores (around 3.5 to 4.0) also have varied gross earnings. For example, movies starring Pamela Anderson (avg. score 3.4) averaged $3.79 million, while others in the low-score range still averaged up to $78.50 million (e.g., Ben Murphy). This variation at the low end suggests that factors beyond the star's individual popularity or the movie's quality (like large budgets or wider releases for some movies) are at play.
- No Strong Linear Trend: Overall, there is no strong linear trend (correlation) observed across all stars between their average score and average gross. A good score does not guarantee a high gross, and a low score does not guarantee a low gross, though the biggest hits are clustered with the highest scores.

13. Profitability Calculation: Create a new column to calculate the profitability of each movie (gross earnings minus budget).

1. Insert a New Column
    - Open your Movies sheet.
    - Scroll to the rightmost empty column.
    - In the header row, type:
      Profitability
2. Formula:
- Gross is in Column M
- Budget is in Column Q

  In the first data row (e.g., Row 2), enter:

=M2 - Q2

3. Explanation:
- M2 → Gross earnings
- Q2 → Filled_Budget

14. Decade-wise Movie Analysis: Categorize movies into decades based on their release year and analyze the trend of average movie scores and gross earnings per decade.

1. New column Named:

   Decade

   Formula:

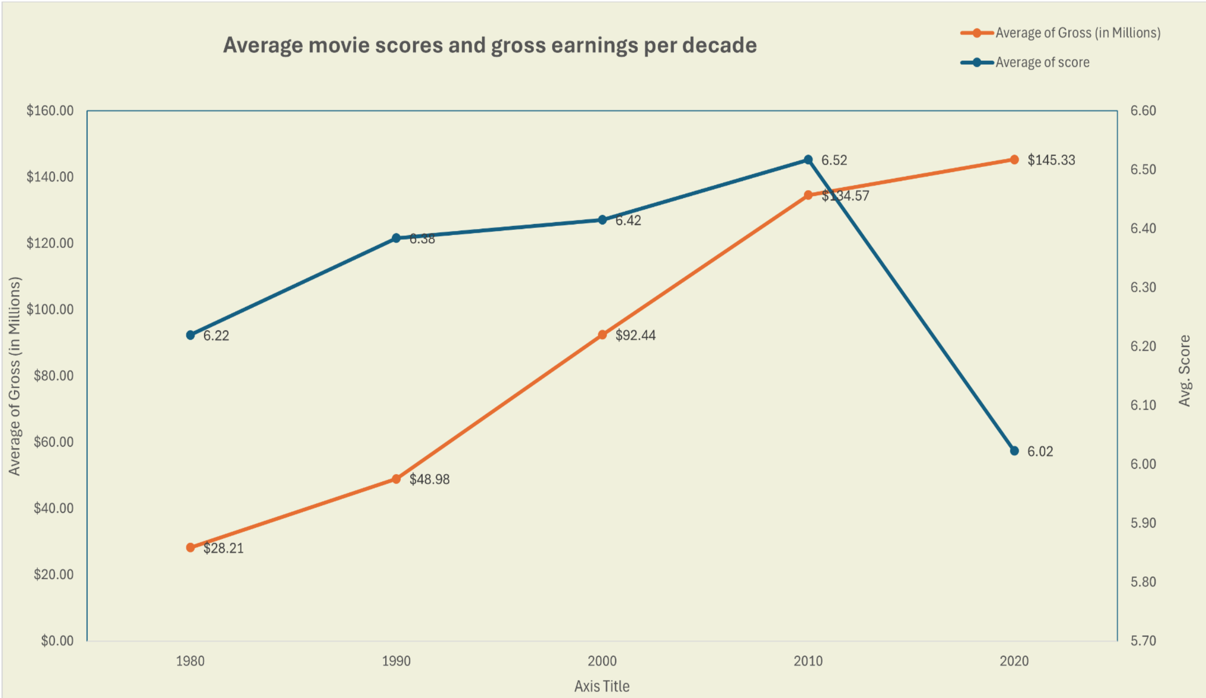
   =FLOOR(D2,10)

    - D2 → movie release year

2. Pivot Table Setup:

Field	Drag to
Decade	Rows
Score	Values → Average
Gross	Values → Average

3. Sort by Decade → Ascending so you see trends from oldest to newest
4. Create a Line Chart (Trend):
    - Insert → Charts → Line with Markers
    - Optionally:
        - Left axis → Average Score
        - Right axis → Average Gross (if the values differ a lot in scale)



Insights:

Analyzing movies categorized by decade reveals a consistent positive trend in average gross earnings, but a more inconsistent trend in average scores.

Decade	Average Score	Average Gross (in Millions)
1980	6.22	$28.21
1990	6.38	$48.98
2000	6.42	$92.44
2010	6.52	$134.57
2020	6.02	$145.33

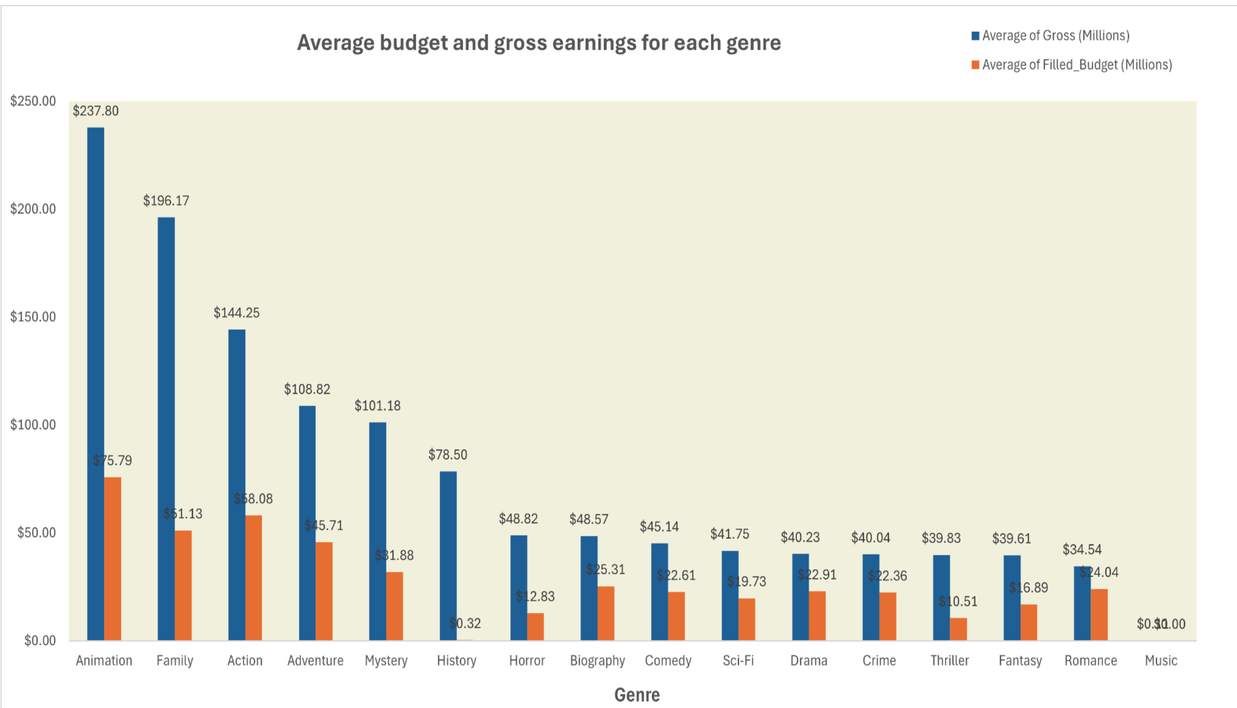
- Increasing Gross Earnings: There is a clear and steady upward trend in average Gross Earnings across the decades, from an average of $28.21 million in the 1980s to $145.33 million in the 2020s. This likely reflects general market growth, ticket price inflation, and broader global distribution of movies over time.
- Score Peak in 2010s: The average score steadily increased from the 1980s to the 2010s, reaching its peak in the 2010s at 6.52.
- Score Drop in 2020s: The average score experienced a significant drop in the 2020s to 6.02, the lowest average score in the dataset, even though the average gross continued to increase. This suggests that recent movies (2020 onwards) are earning more money on average despite lower critical/audience reception (score).

15. Pivot Table for Genre Analysis: Use a pivot table to analyze the average budget and gross earnings for each genre.

1. Pivot Table:

Field	Drag to Area
Genre	Rows
Budget	Values → Average
Gross	Values → Average

2. Insert → Charts →Clustered Column



Insights:

Analyzing average budget and gross earnings by genre highlights which genres are most profitable and which require the largest investment.

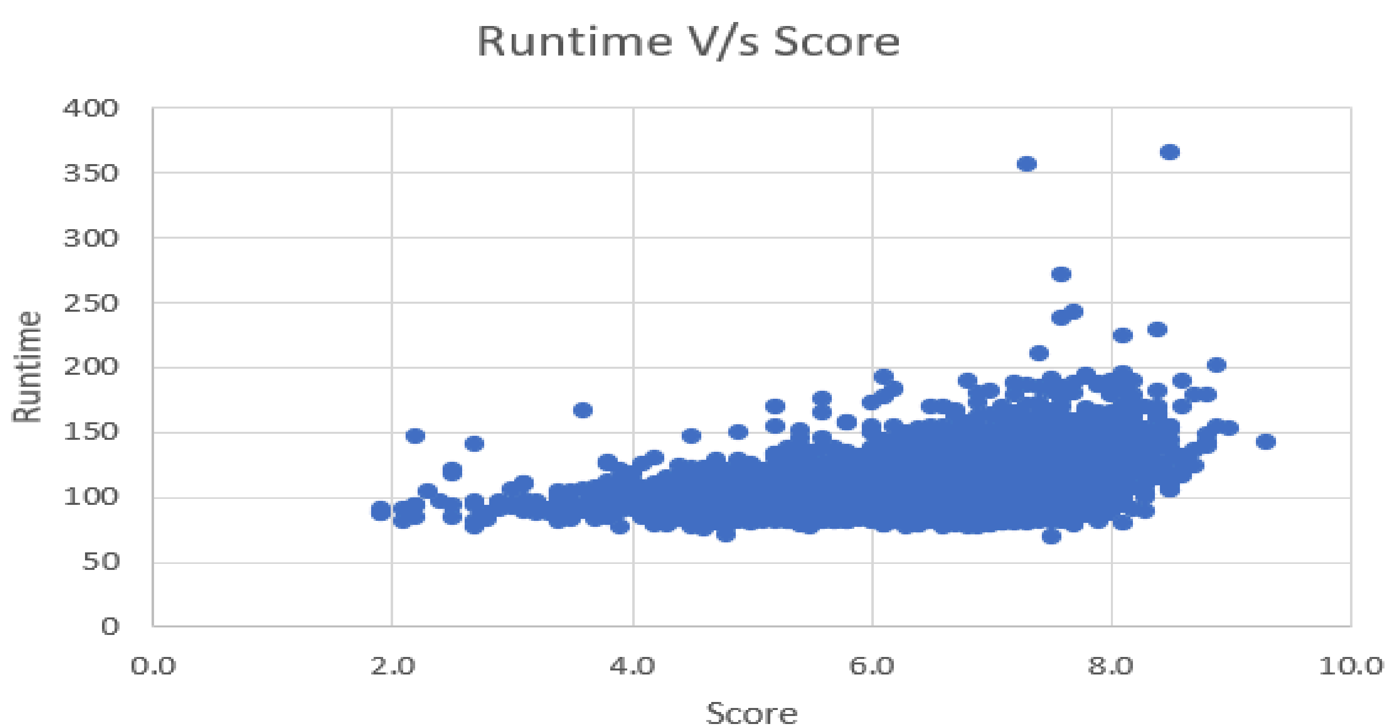
Genre	Average Gross (Millions)	Average Budget (Millions)
Animation	$237.80	$75.79
Family	$196.17	$51.13
Action	$144.25	$58.08
Adventure	$108.82	$45.71
Horror	$42.79	$12.83

- Highest Grossing Genres: Animation ($237.80M), Family ($196.17M), and Action ($144.25M) are the clear leaders in average gross earnings, indicating they are the most lucrative genres overall.
- High Investment Genres: The genres with the highest average budgets closely align with the highest grossing ones: Animation ($75.79M), Action ($58.08M), and Family ($51.13M). This shows that high-grossing genres generally require significant investment.
- High Profitability Potential (Relative): Horror and Mystery genres show a relatively high gross compared to their budget. Horror, for instance, averages a gross of $42.79M against a modest average budget of $12.83M, suggesting a strong return on investment (ROI) compared to the high-budget genres.
- Low Gross/Budget: Genres like History (budget: $0.32M) and Music (gross: $0M) have the lowest figures, suggesting they are either niche, have smaller production scales, or have been less represented in the dataset's high-earning movies.

Question 16: Correlation Analysis: Determine if there's any correlation between the runtime of a movie and its IMDb score.

Steps:-

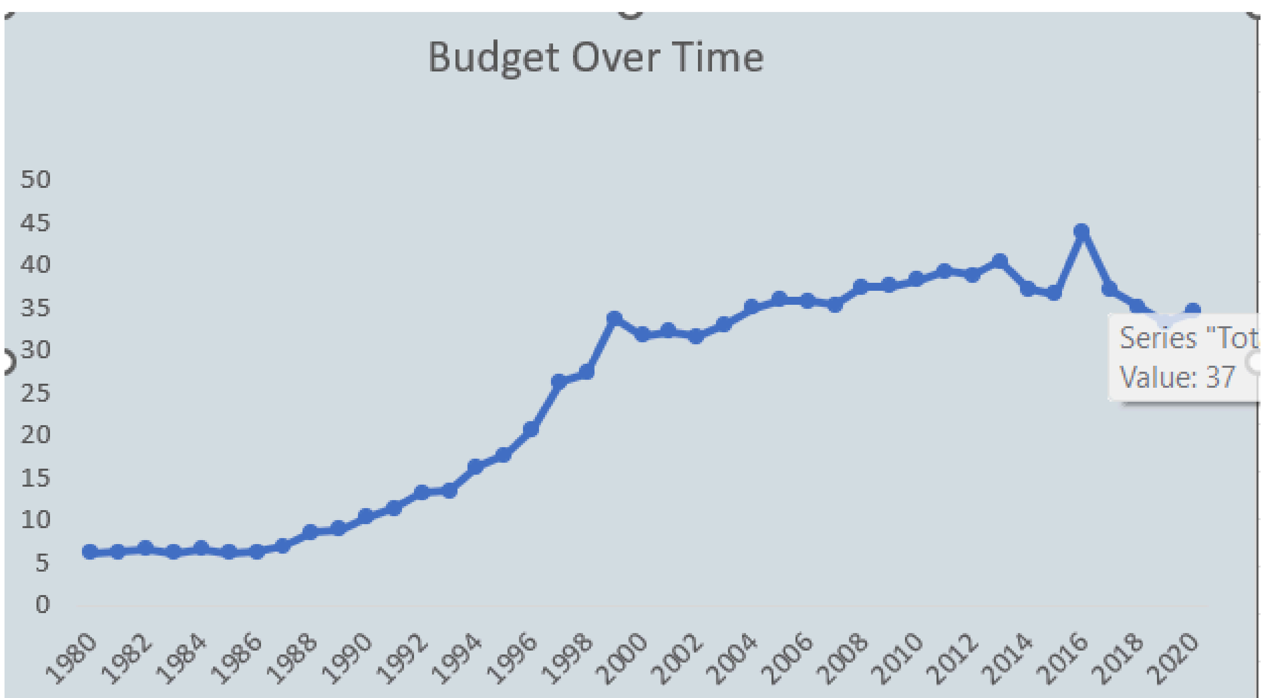
1. Applied correlation formula on runtime and score column (Formula=**CORREL(F2:F7669,O2:O7669) ,F—>Score & O—>Runtime**)
2. Got **correlation value =0.40** that shows a positive correlation between both columns.
3. Insert a scatter plot to visualize the co-relation.



Insights:-
As Clearly visible in the scatter plot as Runtime of a movie increases Imdb score also Increase for that particular movie its mean both are positively correlated and correlation **Value 0.4** confirming that co-relation.

Question 17: Budget Evolution Over Time: Analyze how the average movie budget has changed over the years.

Steps:-
1. Create a pivot table that includes average budget col and year column.
2. Sorted year on ascending order
3. Select the table and create a line chart that shows the trend over time of avg_budget.



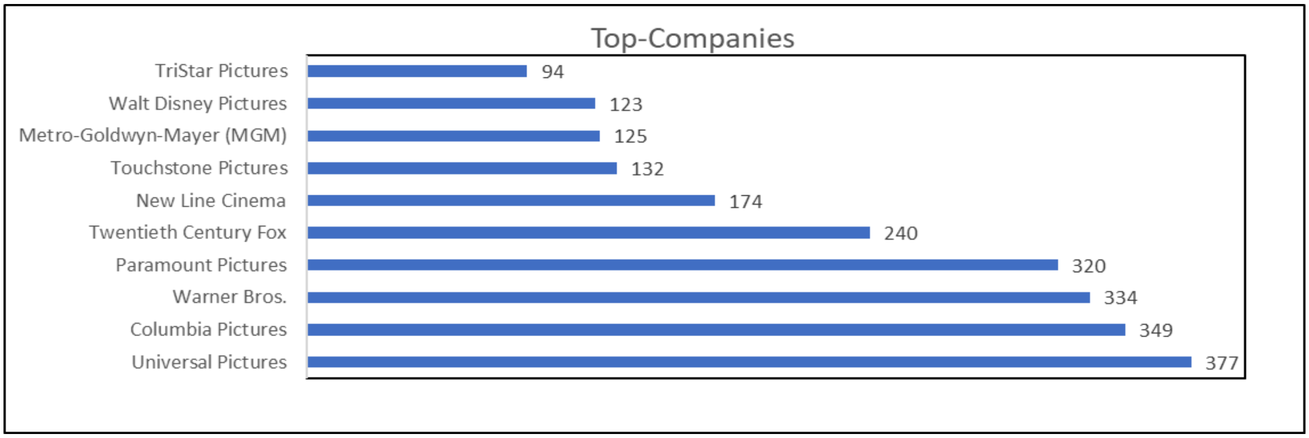
Insights:-
Line chart clearly showing changes in the average budget over time :-
1:-A sharp hike in budget after 1986.
2:-Around 2014-2017 highest budget movies were released.
3:-After 2018 we see a significant fall in the budget of movies.
4:- Budget of movies may rise in upcoming Years

Question 18: Top Companies in Movie Production: Which production companies have released the most movies? Create a bar chart to represent this data.

Steps:-

1. Created a pivot table that has company_name and count_of_movies released by that company.
2. Sorted the count column in descending order so we can see the top companies.
3. Create a bar-chart by selecting the pivot-Table that shows the Top-10 company names by their movies_count.

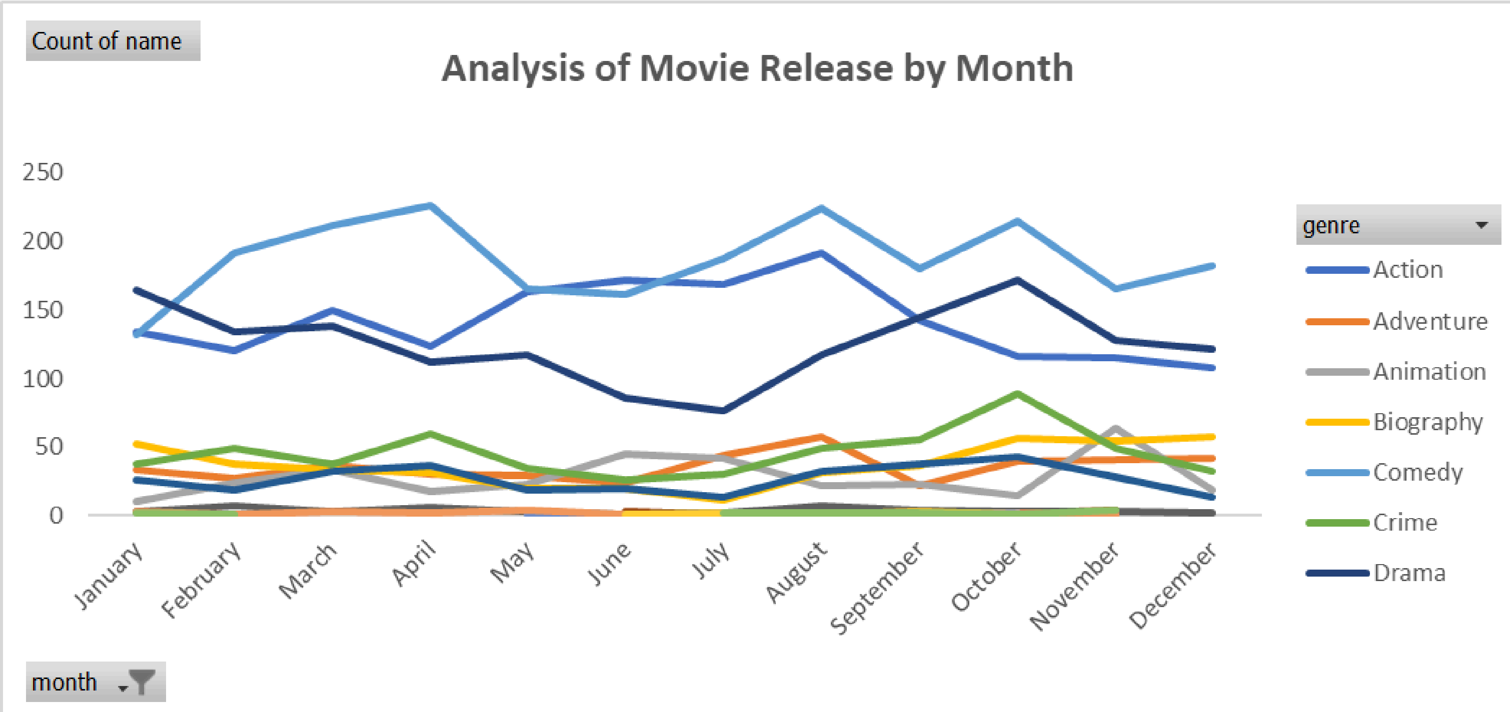
Note:-Universal-Pictures with 377 movie counts has released the most movies.



Question 19: Analysis of Movie Release by Month: Investigate if there is a preferred month or season for releasing movies. Does this trend differ by genre?

Steps:

1. Extracted the **Month** from the *Released* column using `=TEXT(A2,"mmmm")`.
2. Created a **PivotTable** with *Month* (Rows), *Genre* (Columns), and *Count of Movie name*.
3. Inserted a **Line Chart** to visualize monthly release trends by genre and identify seasonal patterns.



Insights:

- **Overall Preferred Seasons:** The highest volume of movie releases occurs in the Spring (February–May) and Fall (August–October).
- **Most Released Genre:** Comedy consistently has the highest number of releases, peaking significantly in April and August.
- Strategic Genre Timing:
    - Drama releases peak sharply in January and again in October, a trend likely driven by the timing requirements for Awards Season consideration.
    - Action also maintains a high release volume, with peaks in February and September.
- **Avoided Months**: July and December are the months with the lowest overall release volume for most major genres (Comedy, Drama, Action, Crime), suggesting a strategic move to avoid the direct competition or high costs associated with major holiday periods, preferring to release *before* or *after* them.

In short, a strong seasonal preference exists, with a dual focus on the Spring market (led by Comedy) and the Fall/Awards market (led by Drama and Action).

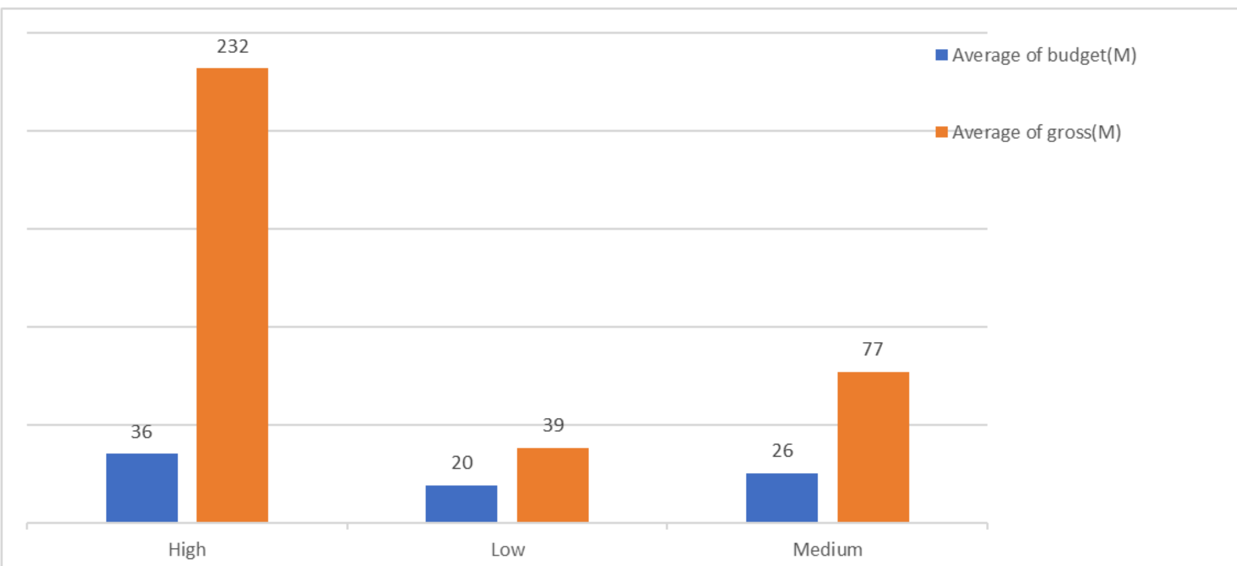
Question 20: Score-based Movie Segmentation: Segment movies into different categories based on their IMDb score and analyze the average budget and gross earnings in each segment.

Steps:-

1. **First create a new column Imdb_score that contains categories based on scores:Threshold for creating column:-'High'-->score>=8 , "Medium"---> 8>score>=5, "low"---->Score<5**
2. **Create a pivot table :-**

Imdb_cat	Average of budget(M)	Average of gross(M)
High	**36**	**232**
Low	**20**	**39**
Medium	**26**	**77**

3: Create a bar chart on basis of pivot table:



Insights:-

Movies with High Budgets Usually Earn Highest As seen In Categories:-
1 : - 'Low':- 195% Return on Budget
2 : - 'Medium':- 296% Return on Budget
3 : - 'High' :- 644% Return on Budget
As Budget Increase For a movie Gross Increases significantly.

21 . Time-Series Analysis of Genre Popularity: Conduct a time-series analysis to evaluate the popularity of movie genres over the years. (Organize the data by year and genre, then calculate the number of movies or total gross earnings per genre per year. Create a line

Steps:

Insert a Pivot Table

- Select all data → Insert → PivotTable → New Worksheet → OK

Build the Pivot Table

- Drag Year → *Rows*
- Drag Genre → *Columns*
- Drag Gross → *Values*
  → (or drag Movie Title and set to *Count* if you want movie counts instead of earnings)

Format Values

- Right-click any value → *Value Field Settings* → choose *Sum* or *Count*
- Format as Number (no decimals or currency).

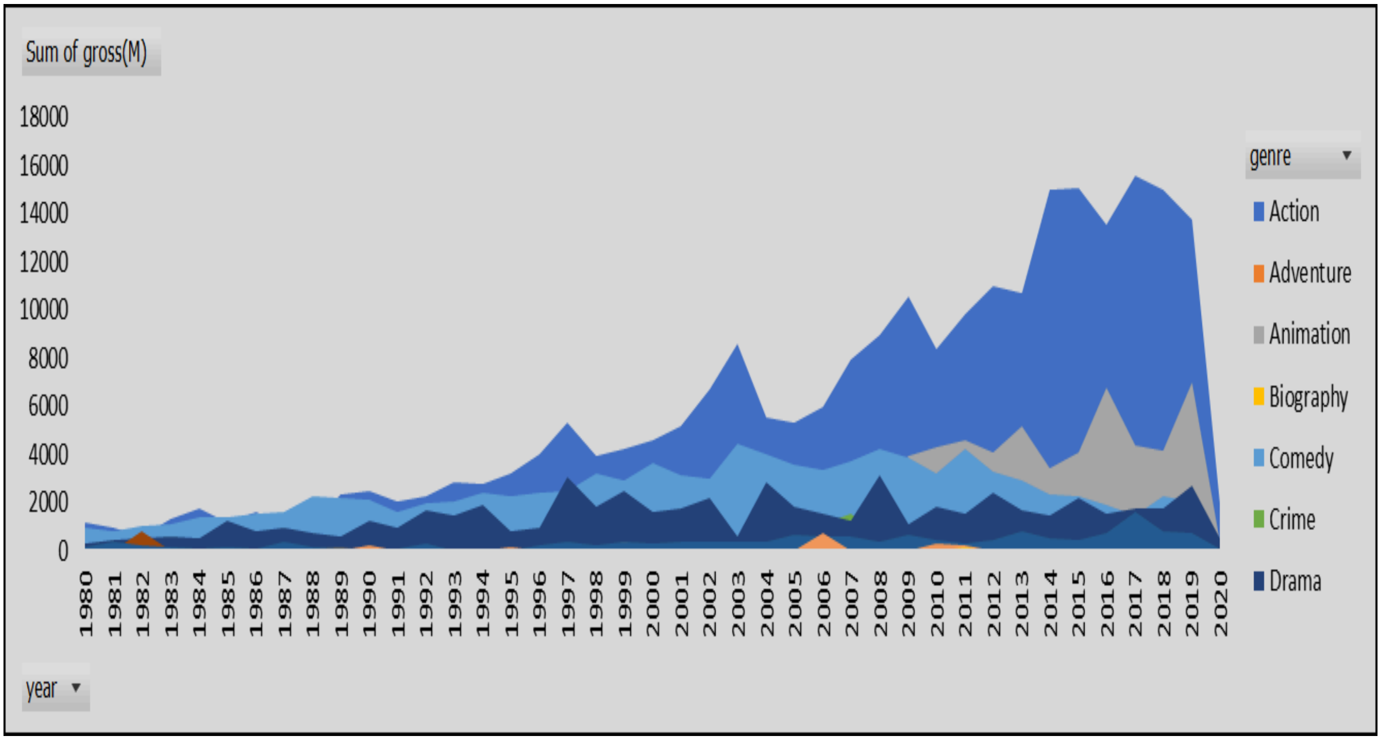
Insert Chart

- Click inside PivotTable → Insert → Line Chart or Area Chart

Style & Interpret

- Chart Title: *"Genre Popularity Over Time"*
- X-axis = Year
- Y-axis = Total Gross or Number of Movies
- Each colored line = One Genre

In short:
(Pivot table
row->year,column->genre,values->name,gross
Pivot chart - area)



22. Predictive Analysis for Future Gross Earnings: Use the historical data to predict the future gross earnings of movies based on genre, budget, and IMDb score.(Utilize Excel's advanced forecasting tools).

Steps:-

1:-Cleaned and converted Genre into numeric (via encoding)

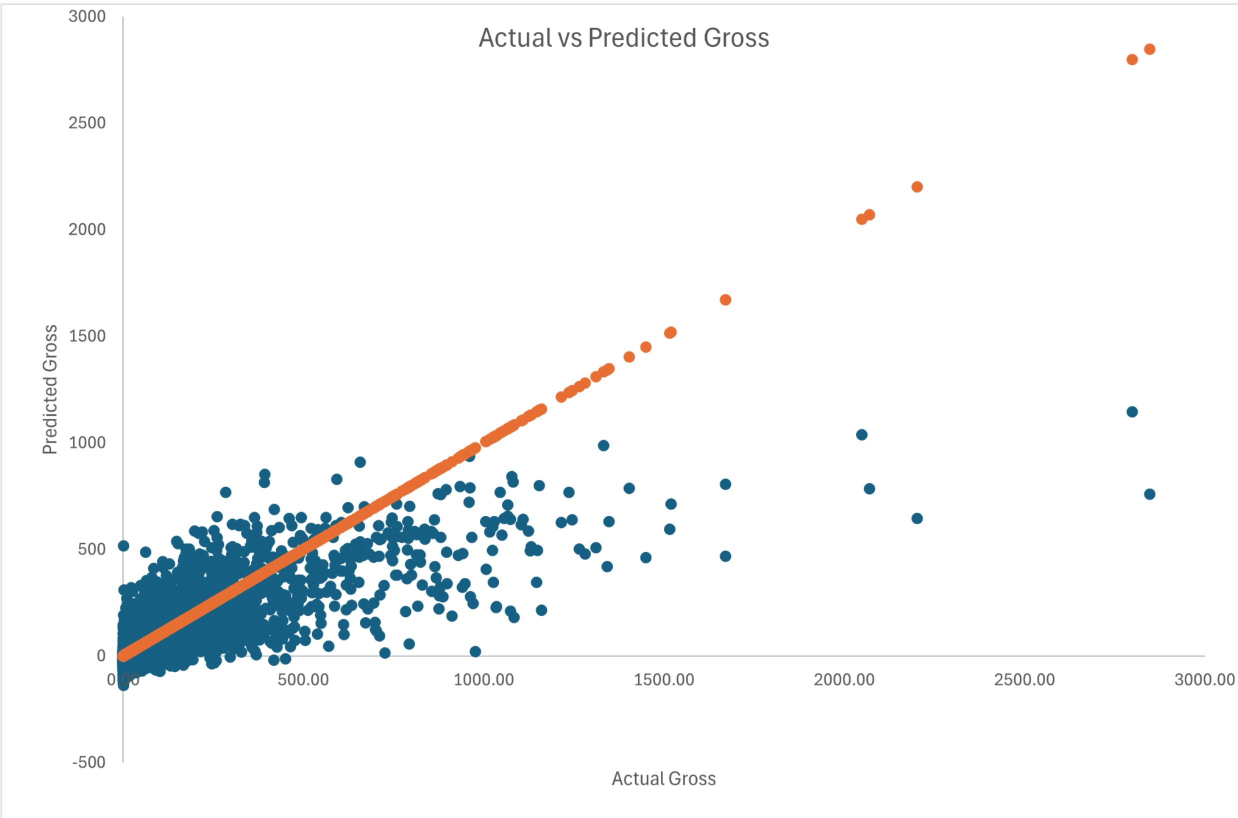
- As Most of the movies have these 4 Genre WE are going to perform One-Hot-Encoding only These

  Genre:-Action,comedy,drama,crime

2:-Used Excel Regression That’s an advanced forecasting tool in Excel

3:-Set Gross as dependent variable (Y)

4:-Set Budget, IMDb Score, Genre Code as independent variables (X)



Insights:-

- Higher budgets generally yield higher returns.
- IMDb score has the strongest individual influence — audience perception matters a lot.
- Genre acts as a moderating variable (some genres like action/comedy outperform others).
- With $R^2 \approx 0.58$, your model is good enough for short-term forecasting but may need more features (like marketing spend, star power, or release season) for long-term accuracy.

23. Revenue and Budget Ratio Analysis Over Time: Calculate and analyze the ratio of total yearly gross earnings to the total yearly budgets of all movies released each year. How has this ratio evolved over the years covered in the dataset?(Aggregate data year-wise, calculate the total gross earnings and total budgets for each year, and then compute the ratio. Then conduct a trend analysis to understand how this profitability indicator has changed over time. Creating a line chart to visualize this trend would be essential.)

Steps

1. Create a Summary Table
    - Select data → Insert → PivotTable → New Worksheet
    - Drag Year → *Rows*
    - Drag Gross → *Values* (Sum or Average)
    - (Optional) Filter by Genre if you want genre-wise prediction.

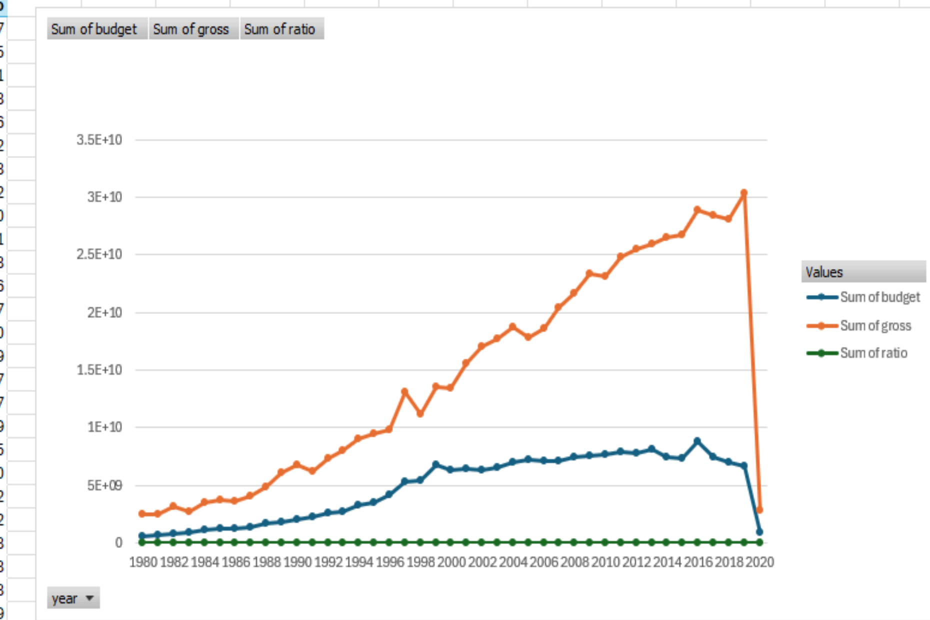
2. Insert Forecast Sheet
    - Select the two columns (Year & Gross) from the PivotTable.
    - Go to Data → Forecast Sheet
    - Choose Line Chart or Column Chart
    - Set:
        - Forecast End Date: future year (e.g., 2030)
        - Confidence Interval: 95%

3. Click Create
    - Excel will generate a forecast chart and a new sheet showing predicted future gross earnings.

4. Optional Enhancement
    - Use Data → Data Analysis → Regression Tool
        - Dependent variable: `gross`
        - Independent variables: `budget`, `imdb_score`
        - This shows how budget and IMDb score affect earnings.

In short:
(Pivot table
Revenue-to-Budget Ratio->= 'gross' / 'budget'
Pivot chart-> line with markers)

Row Labels	Sum of budget	Sum of gross	Sum of ratio
1980	568850000	2483666964	4.37
1981	709450000	2520064889	3.55
1982	836670000	3187580517	3.81
1983	893975562	2750004626	3.08
1984	1116326000	3533092196	3.16
1985	1220510000	3691509283	3.02
1986	1264787376	3647511836	2.88
1987	1382542678	4033434255	2.92
1988	1730540000	4840385631	2.80
1989	1784762578	6087250779	3.41
1990	2072503000	6804235860	3.28
1991	2268248000	6254770668	2.76
1992	2648377769	7334167717	2.77
1993	2685349999	8053959337	3.00
1994	3250454768	9062925753	2.79
1995	3544833757	9448305484	2.67
1996	4144580000	9819527991	2.37
1997	5256145000	13112275767	2.49
1998	5462266000	11176023159	2.05
1999	6742940000	13509638121	2.00
2000	6340250000	13454129869	2.12
2001	6446300000	15617625148	2.42
2002	6335628966	17003436520	2.68
2003	6598430000	17661932024	2.68
2004	6997327000	18721747812	2.68
2005	7189494000	17866609897	2.49



24 Network Analysis of Directors and Stars: Analyze the network of collaborations between directors and stars. Identify which pairs of directors and stars most frequently work together and the average gross earnings of their movies. (Use advanced data manipulation to create this matrix and then apply functions to calculate frequencies and averages.)

Steps:

Load and Clean Data: Load the `movies.csv` dataset. Drop rows with missing values in the `director`, `star`, or `gross` columns.
Group and Aggregate: Group the resulting data by the `director` and `star` columns.
Calculate Metrics: Apply aggregation functions to the grouped data to calculate:

- The Count (frequency) of movies for each pair.
- The Mean (average) of the `gross` earnings.

Identify Top Pairs: Sort the final aggregated list by the Count in descending order to identify the pairs that worked together most frequently, along with their average gross earnings.

In short:
(Pivot table ->
row->director ,star
values->name,gross(count,sum))

Director	Star	Collaboration Count	Average Gross
Clint Eastwood	**Clint Eastwood**	16	$\$79,032,676.81$
Woody Allen	**Woody Allen**	14	$\$15,780,720.50$
Dennis Dugan	Adam Sandler	8	$\$193,814,433.62$
Tim Burton	Johnny Depp	7	$\$184,279,048.43$
J. Lee Thompson	Charles Bronson	6	$\$7,266,180.83$
Richard Donner	Mel Gibson	6	$\$212,541,891.50$
Martin Scorsese	Robert De Niro	6	$\$62,045,857.17$
Gore Verbinski	Johnny Depp	5	$\$637,533,394.40$
Sylvester Stallone	Sylvester Stallone	5	$\$193,834,021.20$
Ron Howard	Tom Hanks	5	$\$378,203,657.40$