Data Analysis Report

Project Summary

- **Purpose of the Project** -- Firstly , find the meaningful insights from the data. Secondly , understand the Trend or pattern between different types of variable in a dataset. My main purpose is to visualize the data to uncover trends and relationships which would have been difficult to identify from raw data alone.

- **Tools & Techniques –** Advance Excel , Pivot table .

- **Content**

- Overview of Data - Understand the theoretical part of data for strong foundation.
- Limitation of Project - Understand some limitation of the project.
- Investigate About Data - Find Basic insights about data to getting comfortable with dataset.
- Find Complex Answers - Showcase the power of pivot table to solve complex answers.
- Torture the Data – Find Trends or pattern with the help of advance charts like Scatter plot.
- Conclusion – Discussion about outcomes which we get through this analysis for future – decision making related to movies.

https://github.com/ShubhamPadiya2002/Excel-Project

Overview of Data

IMDb (Internet Movie Database) is a popular online database of information related to films, television programs, and video games, including cast, production crew, plot summaries, trivia, and user ratings.

How IMDb Platform Works ?

1.User Ratings: IMDb ratings are primarily based on the ratings submitted by registered users of the IMDb website or app. Users can rate movies on a scale from 1 to 10, with 10 being the highest rating. These ratings are aggregated to calculate the overall rating for each movie.

2.Weighted Average: IMDb uses a weighted average formula to calculate the overall rating for a movie. The formula takes into account factors such as the number of ratings received and the distribution of those ratings.

3.Ratings Distribution: IMDb considers the distribution of ratings when calculating the overall rating for a movie. Movies with a more uniform distribution of ratings may receive a higher overall rating compared to movies with a more polarized distribution.

4.User Reviews: In addition to numerical ratings, IMDb also allows users to submit written reviews and comments for movies. These user reviews can provide additional context and insight into the quality of a movie.

5.Top Rated Lists: IMDb also provides curated lists of top-rated movies based on user ratings. These lists highlight the highest-rated movies across various genres, decades, and other categories.

https://github.com/ShubhamPadiya2002/Excel-Project

Limitation of the Project

- **Data Scope**: The analysis might be limited to the data available in the IMDb database, which may not cover all movies or have complete information for each title.
- **Limitations Depth**: The project uses tools like Excel and pivot tables, which are powerful but may not offer the same depth of analysis as more advanced statistical software.
- **Timeframe**: The project seems to cover data up to a certain point, which means recent movies or trends might not be included.
- **Data Accuracy**: The accuracy of the data could be a limitation. Since the data is user-generated, there may be errors or inaccuracies that could affect the results of the analysis. Users may submit incorrect information or make subjective judgments about movie ratings, which could introduce biases into the analysis.

https://github.com/ShubhamPadiya2002/Excel-Project

Description of Data

- **Title** - Official name of the Movie.
- **Release Date** - Date which Movie Release.
- **Color/B&W** - Print of the Movie.
- **Genre** - Classification of Movie Based on style, content, and thematic element.
- **Language** - Official language of Movie.
- **Country** - Origin of a Movie.
- **Rating** - Classification of Movie Based on Quality.
- **Lead Actor** - Name of the Movie Actor.
- **Director Name** - Name of person who direct the movie
- **Lead Actor FB Likes** - Actor FaceBook likes.
- **Cast FB Likes** - Cast FaceBook likes.
- **Director FB Likes** - Director FaceBook likes.
- **Movie FB Likes** - Movie FaceBook likes.
- **IMDb Score (1-10)** - Assigning numerical value Based on Performance.
- **Total Reviews** - Sum of Movie reviews.
- **Duration (min)** - Time Length of a movie.
- **Gross Revenue** - Revenue Generate by Movie. (Excluding Expenses)
- **Budget** - Overall Expenses incurred for making a movie.

https://github.com/ShubhamPadiya2002/Excel-Project

Difference Between Genre & Rating

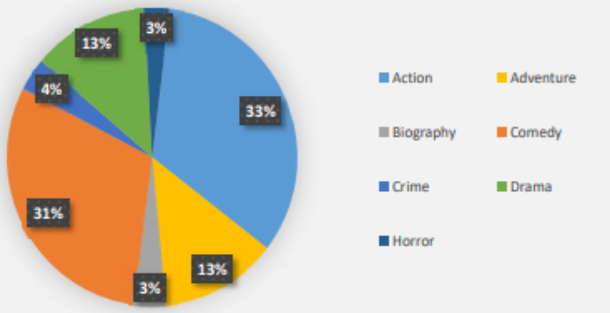
Aspect	Genre	Rating
Definition	Genre refers to a category or classification that describes the style, content, and thematic elements of a film.	Rating refers to an evaluation of the quality or content of a movie. It provides viewers with feedback or an opinion on various aspects of a film, such as its entertainment value, artistic merit, suitability for different audiences, and overall impact.
Purpose	The primary purpose of genre is to categorize movies and provide viewers with a framework for understanding and selecting films that align with their interests and preferences.	The primary purpose of rating is to inform viewers about the content and quality of a movie, helping them make informed decisions about which films to watch.
Classification	Genres describe the content, style, and themes of a film. For example, action movies typically feature fast-paced sequences involving physical feats, while romance movies focus on romantic relationships and emotional connections.	Ratings assess various elements of a movie, including plot, acting, direction, cinematography, special effects, music, and overall entertainment value.
Examples	Common movie genres include action, comedy, drama, horror, romance, science fiction, fantasy, thriller, animation, and documentary.	Common rating systems include G (General Audiences), PG (Parental Guidance), PG-13 (Parental Guidance-13 and up), R (Restricted), and NC-17 (No One 17 and Under Admitted), among others.
Content Analysis	Genre analysis involves examining narrative structure, themes, characters, visual style, and other elements to identify common patterns and characteristics within a particular genre.	Rating analysis involves evaluating a movie's strengths and weaknesses across different criteria to determine its overall quality and appeal.
Usage in Industry	In the film industry, genres play a crucial role in marketing and distribution strategies. They help filmmakers target specific audiences, position movies within the market, and identify trends in audience preferences.	Ratings also play a role in marketing campaigns, promotional materials, and awards considerations within the film industry. Movie ratings influence audience decisions, critical acclaim, box office success, and industry recognition.

https://github.com/ShubhamPadiya2002/Excel-Project

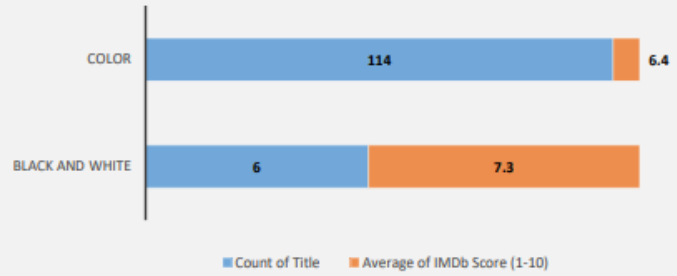
IMDB DASHBOARD

By Shubham Padiya

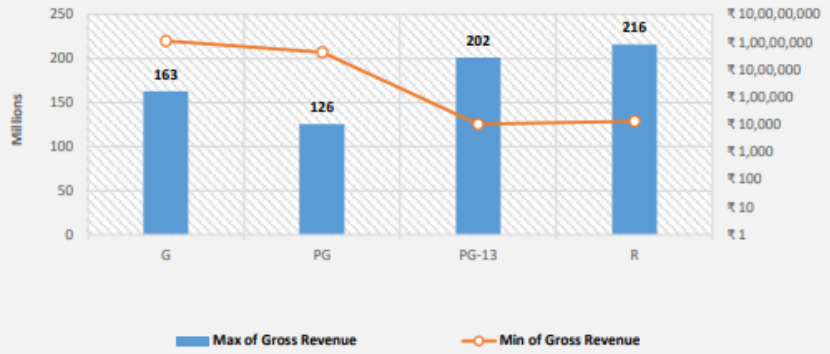
Revenue (%) (Based on Genue)



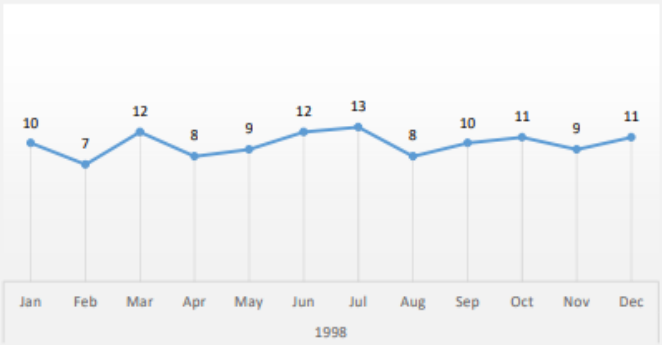
No. of Movies & Avg. Rating (Based On Colour)



Highest Revenue Vs Lowest Revenue (Based on Rating)

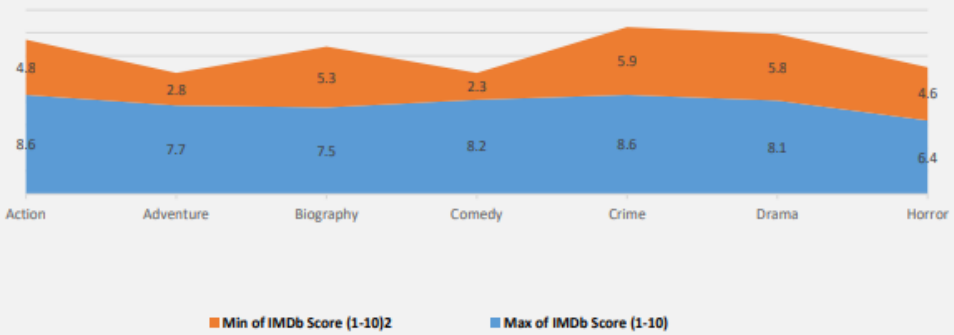


No. of Movies (Based On Month)

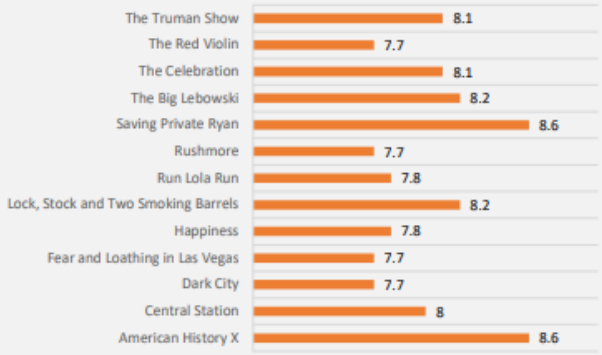


Top 10 Contries (Based On Revenue)

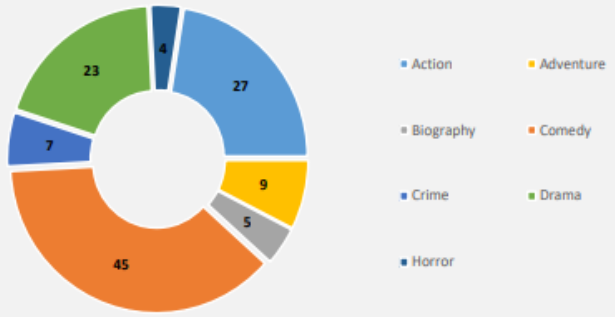
Highest VS Lowest IMDB (Based on Genre)

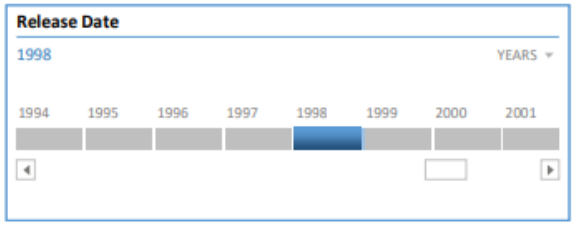


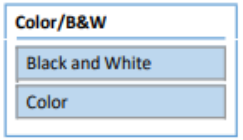
Top 10 Movies (Based On IMDB Rating)



No. of Movies (Based On Genre)

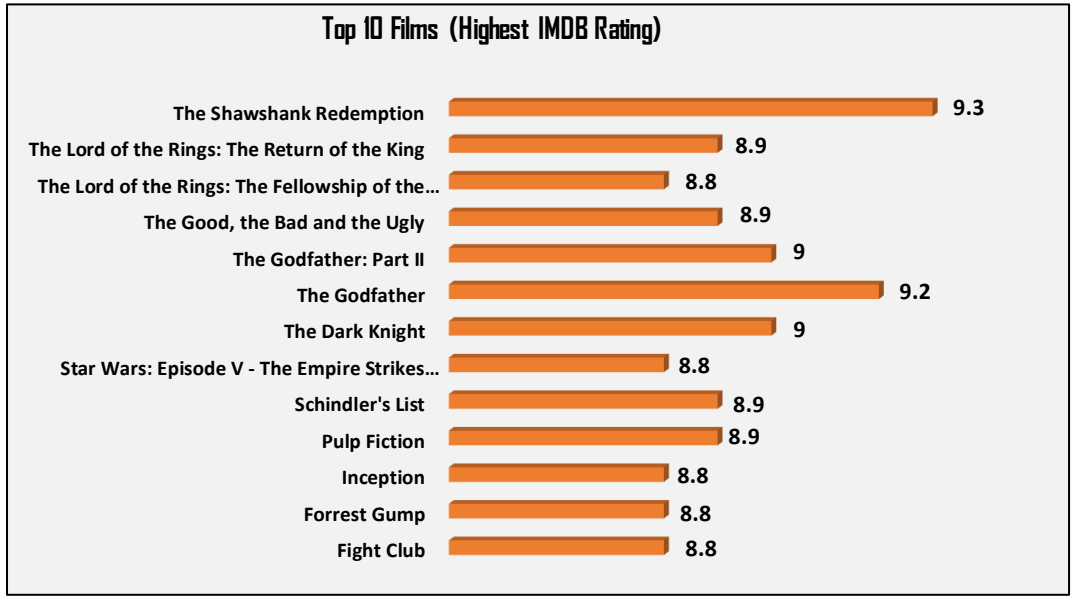


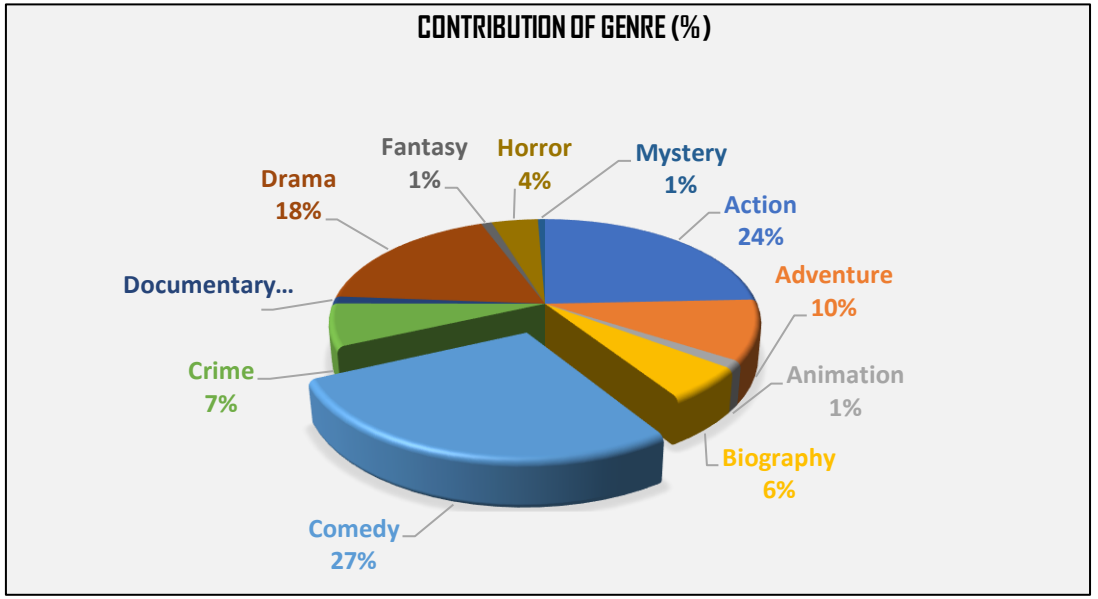




https://github.com/ShubhamPadiya2002/Excel-Project

Let's Investigate about Data

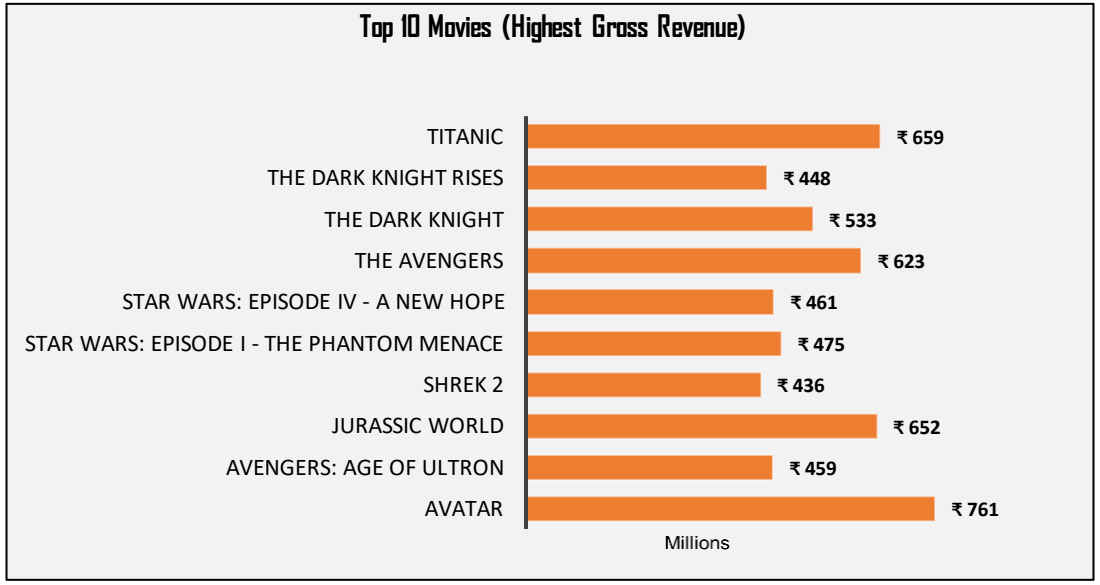


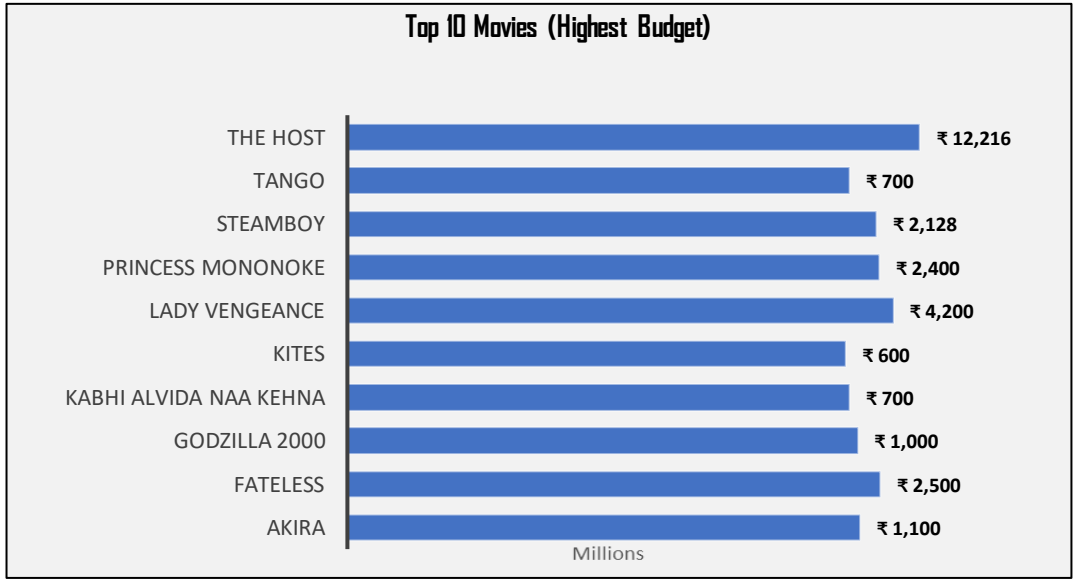


The Shawshank Redemption has the Highest IMDb Rating Movie of all time with a 9.3 rating & The Godfather has achieved the second highest IMDb rating of 9.2.

Comedy has the most popular or demand genre & after comedy Action has the 2nd most popular genre among all type of genres.

https://github.com/ShubhamPadiya2002/Excel-Project

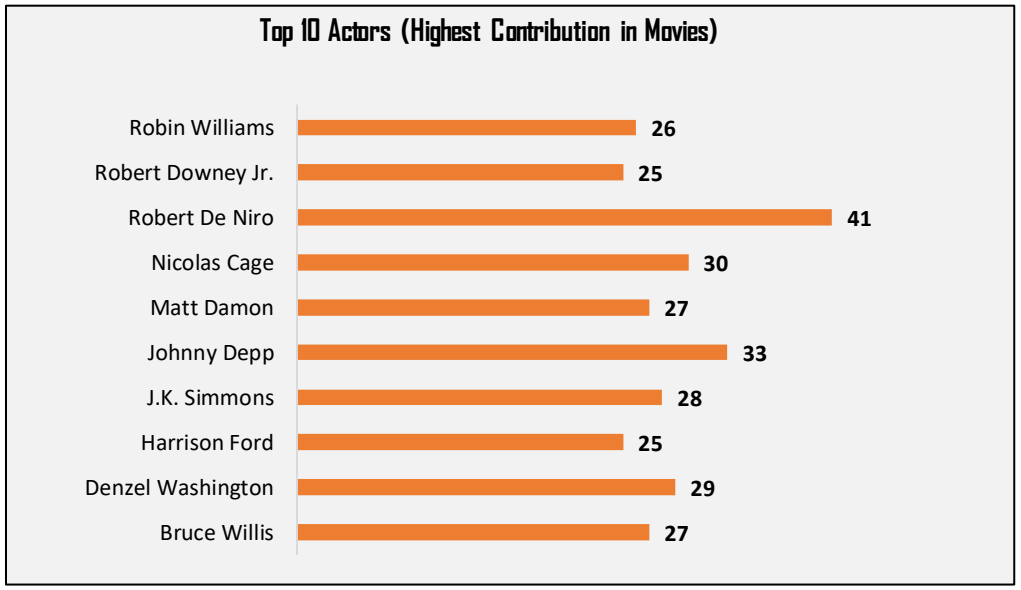


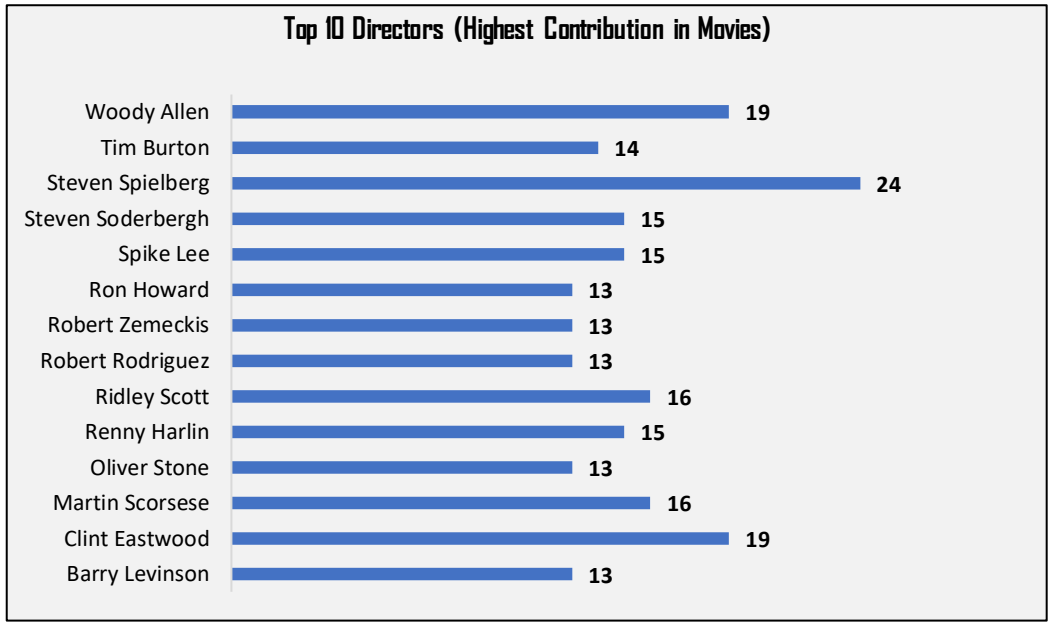


Avatar has the highest gross revenue with a 761 million & Titanic has the 2nd highest gross revenue with a 659 million.

The Host has the highest Budget movie with a 12k million (approx.) & Lady Vengeance has the 2nd highest Budget with a 4k million (approx.)

https://github.com/ShubhamPadiya2002/Excel-Project

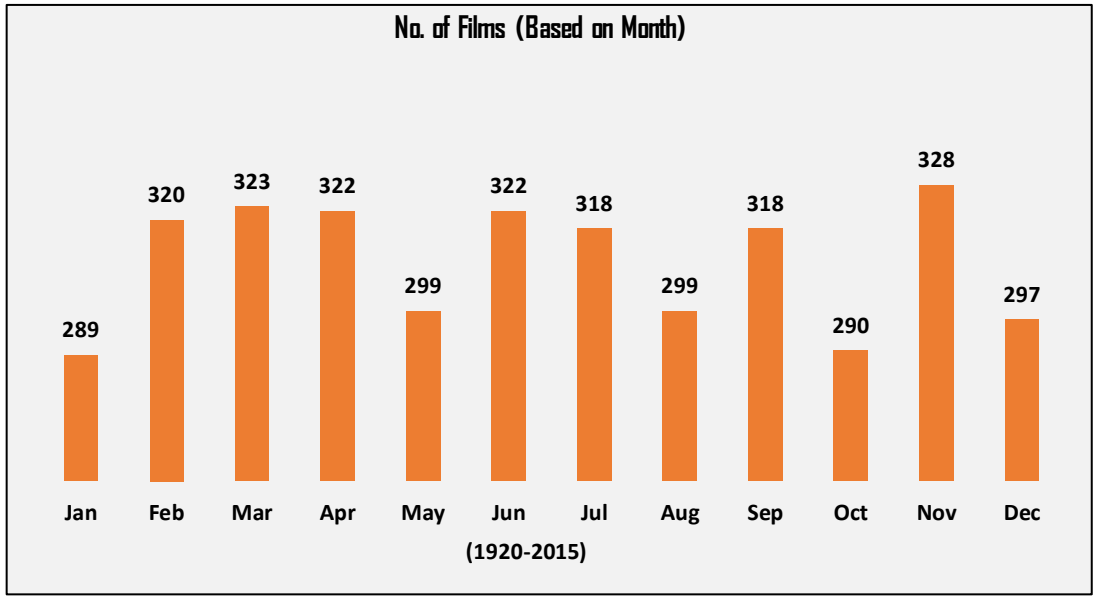


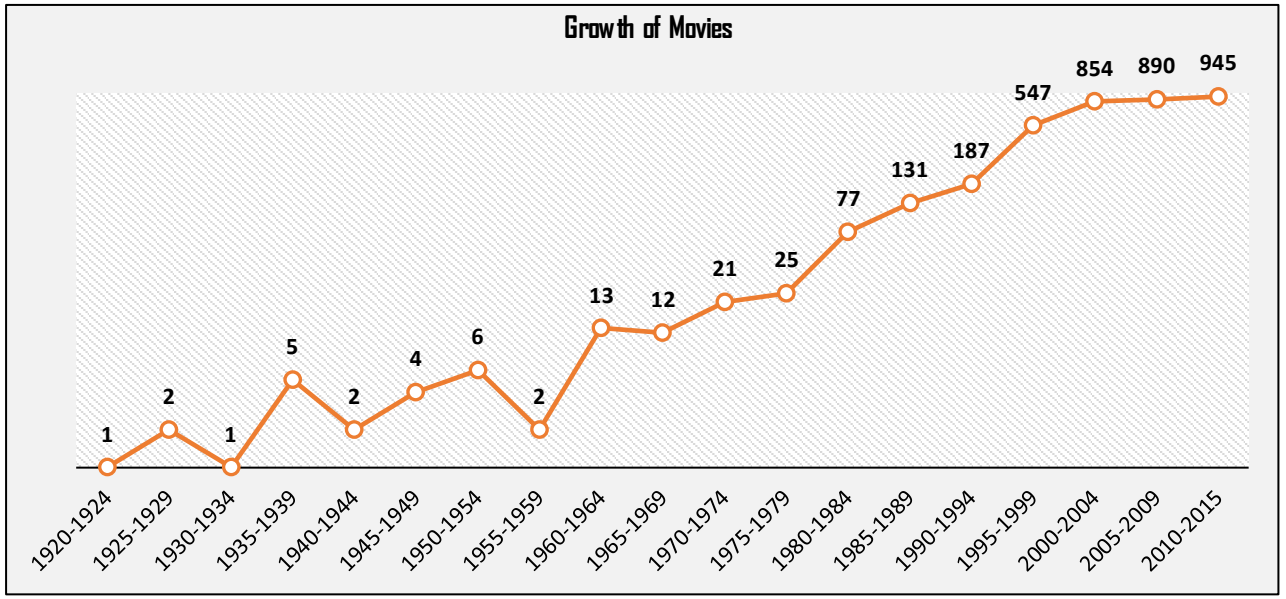


Robert De Niro has done highest no. of movies (41) & Johnny Depp has the 2nd actor which done highest no. of movies after robert De niro.

Steven Spielberg has direct highest no. of movies (24) & Woody Alleen , Clint Eastwood has the 2nd director which done highest no. of movies after steven.

https://github.com/ShubhamPadiya2002/Excel-Project





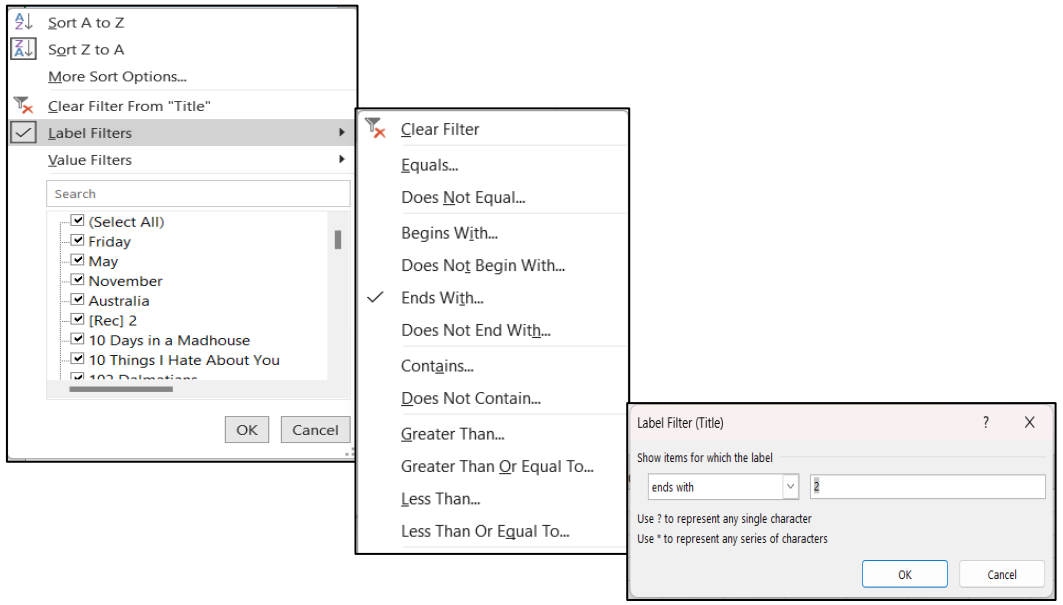
As per the above fig, we can see that in 1920 to 2015, Highest no. of Movies are release in November month & lowest no. of Movies are release in Jan.

As per the above fig, we can observe that growth of movies release continuously increase in every 5 years gap in period of 1920 – 2015.

https://github.com/ShubhamPadiya2002/Excel-Project

How to find Complex Answers with Pivot Table -

Row Labels	Sum of Gross Revenue
Shrek 2	₹ 43,64,71,036
Spider-Man 2	₹ 37,33,77,893
Despicable Me 2	₹ 36,80,49,635
Iron Man 2	₹ 31,20,57,433
The Twilight Saga: Breaking Dawn - Part 2	₹ 29,22,98,923
The Hunger Games: Mockingjay - Part 2	₹ 28,16,66,058
Toy Story 2	₹ 24,58,23,397
Rush Hour 2	₹ 22,61,38,454
X-Men 2	₹ 21,49,48,780
The Amazing Spider-Man 2	₹ 20,28,53,933
Cars 2	₹ 19,14,50,875
Pitch Perfect 2	₹ 18,34,36,380
How to Train Your Dragon 2	₹ 17,69,97,107
Hotel Transylvania 2	₹ 16,96,92,572
Kung Fu Panda 2	₹ 16,52,30,261
American Pie 2	₹ 14,50,96,820
Taken 2	₹ 13,98,52,971
The Santa Clause 2	₹ 13,92,25,854
Grown Ups 2	₹ 13,36,68,525
Rio 2	₹ 13,15,36,019
Cloudy with a Chance of Meatballs 2	₹ 11,97,93,567
Die Hard 2	₹ 11,75,41,000
Dr. Dolittle 2	₹ 11,29,50,721

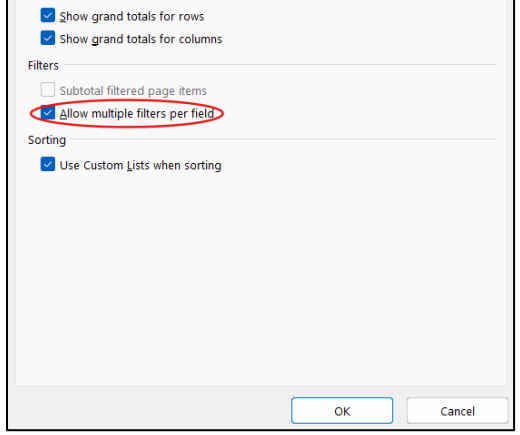


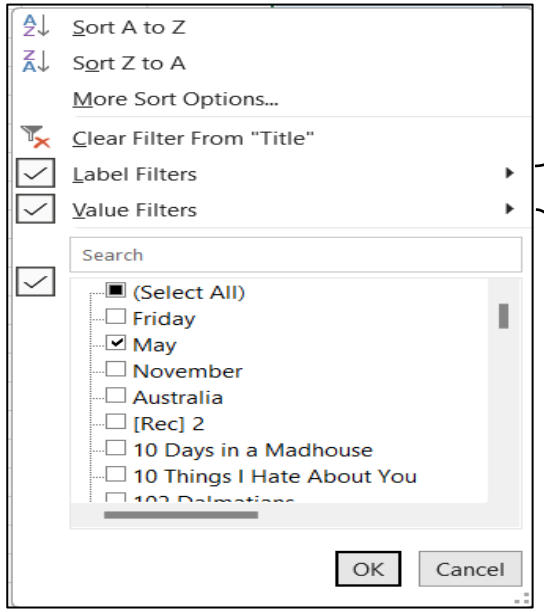
a) Which Sequel movie have the highest gross revenue among all sequel movies ?

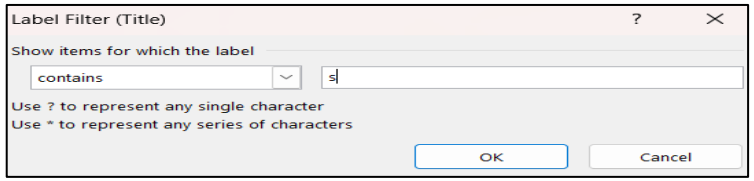
➢ **Shrek 2** have the 1st highest gross revenue sequel movie and Spider Man 2 has the 2nd highest gross revenue.

https://github.com/ShubhamPadiya2002/Excel-Project

Row Labels	Sum of Gross Revenue
Mallrats	₹ 21,22,561
Me and Orson Welles	₹ 11,86,957
Micmacs	₹ 12,60,917
Miss Potter	₹ 29,75,649
Grand Total	**₹ 75,46,084**





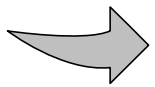


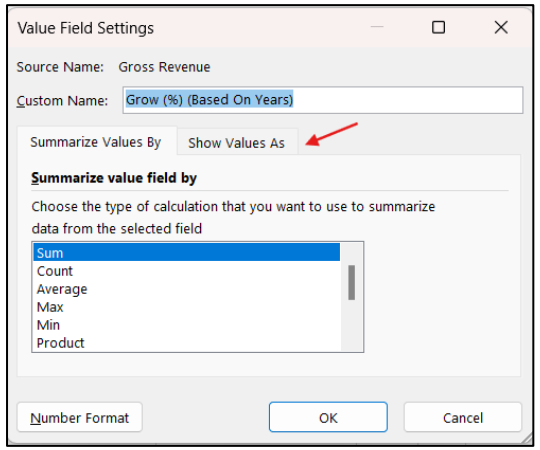
b) How many movies which title starts with ‘M’ & have gross revenue between 10 lacs to 30 lacs ?

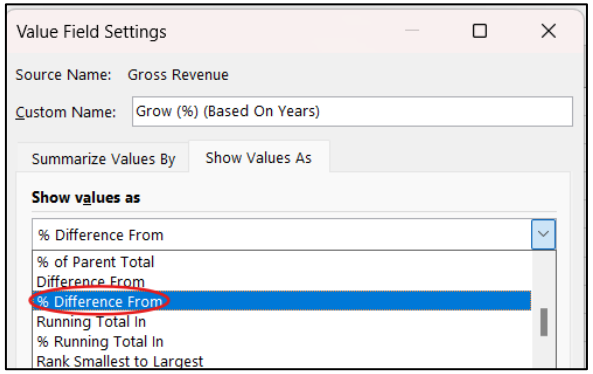
- There are **4** movies which starts with “M” & have gross revenue between 10 to 30 lacs.

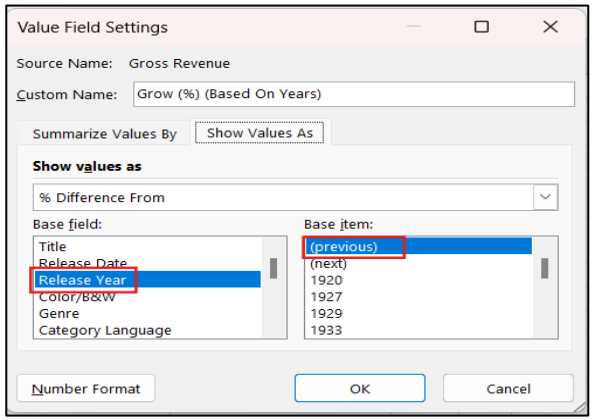
https://github.com/ShubhamPadiya2002/Excel-Project

Row Labels	Sum of Gross Revenue	Grow (%) (Based On Years)
Action	11,75,36,28,956.00	
2012	3,69,54,06,129.00	
2013	3,20,15,01,155.00	-13.37%
2014	4,85,67,21,672.00	51.70%







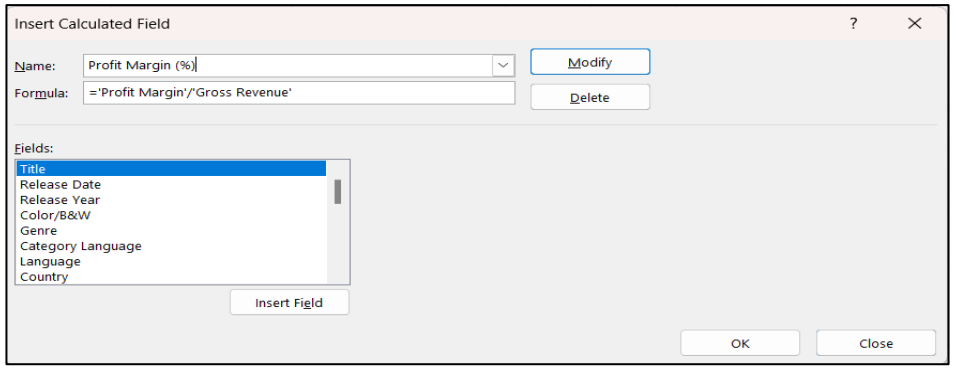


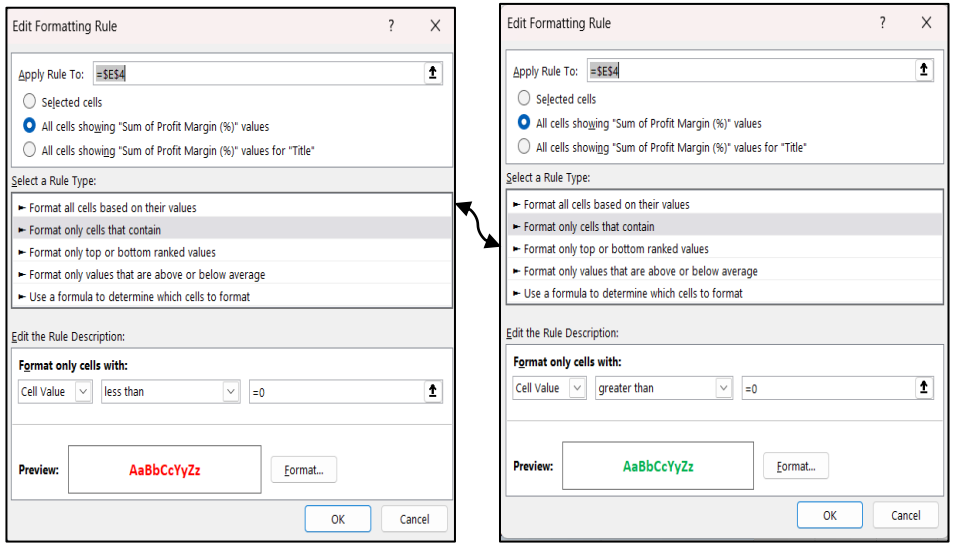
c) What % did action movies revenue grow in 2013 to 2014 ?

- Action movies revenue grow in 2013 to 2014 is **51.7 %** means action movies perform incredible in 2014 in comparison of 2013.

https://github.com/ShubhamPadiya2002/Excel-Project

Row Labels	Sum of Budget	Sum of Gross Revenue	Sum of Profit Margin	Sum of Profit Margin (%)
Ayurveda: Art of Being	3,00,000.00	16,892.00	-2,83,108.00	-1676.0%
Baahubali: The Beginning	1,80,26,148.00	64,98,000.00	-1,15,28,148.00	-177.4%
Chocolate: Deep Dark Secrets	15,00,000.00	49,000.00	-14,51,000.00	-2961.2%
Dil Jo Bhi Kahey...	7,00,00,000.00	1,29,319.00	-6,98,70,681.00	-54029.7%
Fiza	10,00,000.00	6,10,991.00	-3,89,009.00	-63.7%
Jab Tak Hai Jaan	72,17,600.00	30,47,539.00	-41,70,061.00	-136.8%
Kabhi Alvida Naa Kehna	70,00,00,000.00	32,75,443.00	-69,67,24,557.00	-21271.2%
Kites	60,00,00,000.00	16,02,466.00	-59,83,97,534.00	-37342.3%
Monsoon Wedding	70,00,000.00	1,38,76,974.00	68,76,974.00	49.6%
My Name Is Khan	1,20,00,000.00	40,18,695.00	-79,81,305.00	-198.6%
Neal 'N' Nikki	15,00,000.00	1,00,358.00	-13,99,642.00	-1394.6%
Paa	15,00,00,000.00	1,99,228.00	-14,98,00,772.00	-75190.6%
The Lunchbox	10,00,000.00	42,31,500.00	32,31,500.00	76.4%
Veer-Zaara	70,00,000.00	29,21,738.00	-40,78,262.00	-139.6%



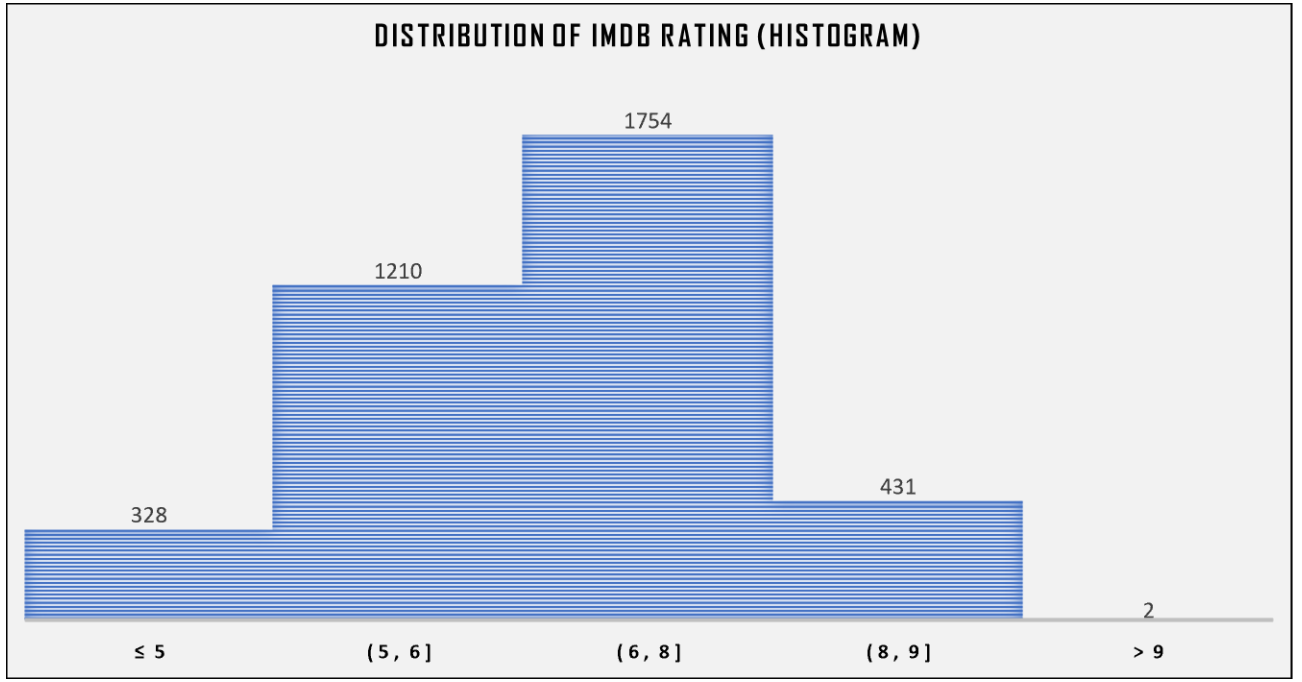


d) What is the profit margin (%) of Indian movie ‘Monsoon wedding’ ?

- The profit margin of Monsoon wedding is **49.6%** which means movie costs is approx 50%.

https://github.com/ShubhamPadiya2002/Excel-Project

Let's Torture the Data...

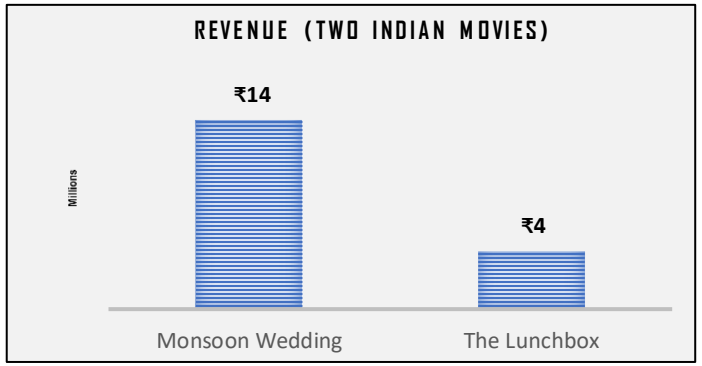


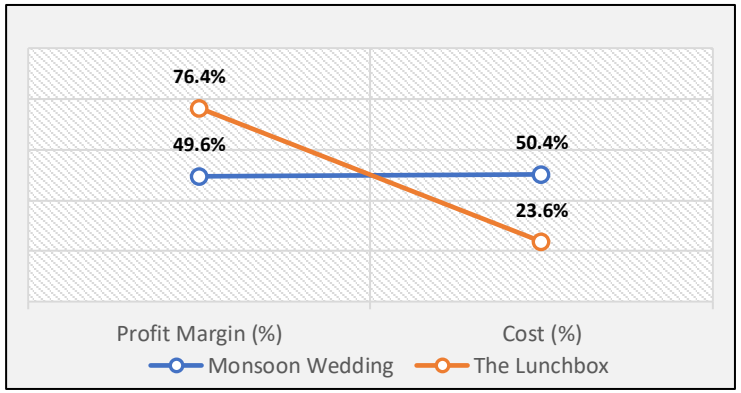
As per the above figure, We can observe that maximum no. of movies are falls under 6 – 8 IMDb Rating . There are 328 movies which has not even achieved IMDb rating greater then 5 indicating the prevalence of lower-rated films within the dataset & only 2 movies are reached beyond 9 IMDb rating showcasing the rarity of exceptionally high-rated films within the collection.

This distribution underscores the diversity and range of movie ratings present in the dataset, revealing insights into the overall quality distribution and the rarity of top-rated cinematic experiences.

https://github.com/ShubhamPadiya2002/Excel-Project

Cost Analysis in Movie Profitability



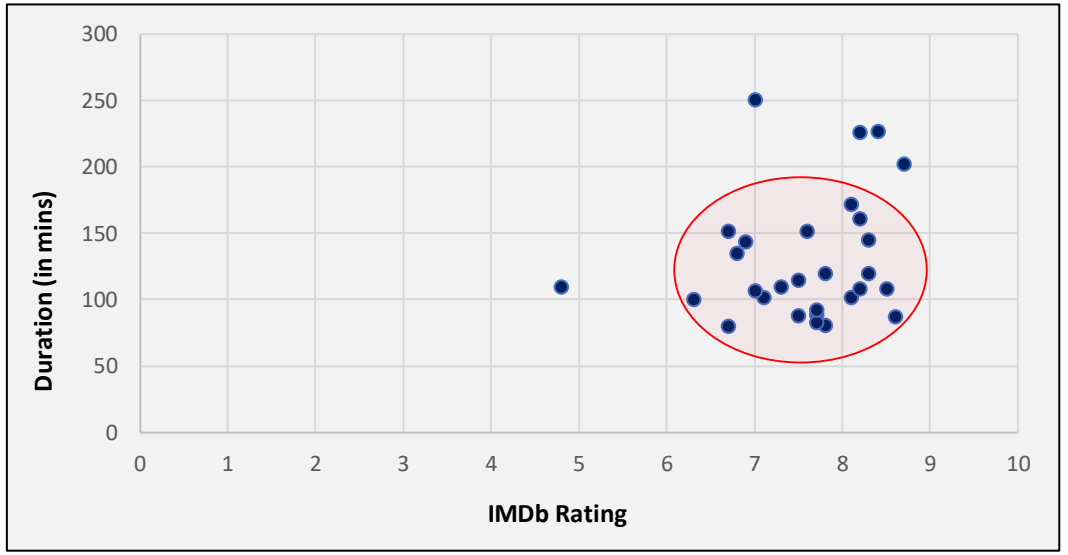


As per the above figure , we can observe that gross revenue of the movie “The LunchBox“ is very low in comparison of “Monsoon Wedding” , which does not mean that ‘Monsoon Wedding has high Profit Margin in comparison of ‘LunchBox’ because **Profit Margin is inversely proportional to Cost**. It means cost increases profit margin decreases , cost decreases profit margin automatically increases.

So, We can clearly see that “The LunchBox” has higher profit margin **(76.4%)** then “Monsoon Wedding” because “Lunchbox” movie has lower cost **(23.6%)** comparison to “Monsoon Wedding”.

https://github.com/ShubhamPadiya2002/Excel-Project

Correlation of IMDB Rating with Movie Duration



As per the above figure , We clearly see a pattern that mostly movies which has a range of 7 – 8 rating are duration of 100 – 150 mins which means audience mostly likes movies which time duration of between 1hr 50 mins to 2hr 10 mins (approximately). Some exceptional cases that movies have duration of 100 – 150 mins are very less IMDb rating of 4-5. & some movies have time duration of more than 200 mins but even this long duration doesn't affect movie rating that much which means movie have good story plots.

When we deeply observe , we can see another pattern that some of the movies has duration range of 200 – 250 mins which means over 3.5 hr to 4hr but still achieve a rating of 8-9 which is unbelievable

https://github.com/ShubhamPadiya2002/Excel-Project

Why some of the movies has duration of 200 – 250 mins , but still achieve a IMDb rating of 8 – 9 ?

Here are some reason behind it !

Quality of Content: The most crucial factor influencing IMDb ratings is the quality of the content itself. Movies with compelling storytelling, engaging characters, thought-provoking themes, and exceptional production values can captivate audiences regardless of their length.

Directorial Skill: Skilled directors can effectively manage longer runtimes by pacing the narrative, maintaining audience interest, and delivering a cohesive and immersive viewing experience. They understand how to leverage the extended runtime to develop characters, explore complex plotlines, and create memorable cinematic moments.

Fan Base: Certain directors or franchises have dedicated fan bases that eagerly anticipate longer films. These fans may actively contribute to higher IMDb ratings due to their loyalty and enthusiasm for the content.

Audience Expectations: Some genres, such as epic dramas, historical epics, and certain art-house films, are known for their longer runtimes. Audiences watching these genres may have different expectations and be more accepting of extended durations if they feel the storytelling justifies it.

Conclusion

- Through this analysis, We have build good theoretical knowledge about movies and their terms like **what is the difference between genre & rating, etc.**
- With the help of this analysis , We find insights related to some important movies terms like contribution of genre(%) ,gross revenue ,budget ,director, actors**.**
- In this analysis, we also discover some time patterns like **which month movies release the most**. This analysis help director to decide release month.
- Through this analysis, We understand the **power of pivot table** that how pivot table solve complex questions easily.
- Through cost analysis, We understand the very important concept of profit margin & cost. As a producer , we must should focus on cost or budget of a movie because **profit margin is inversely proportional to Cost.**
- With the help of analysis of time duration, the director decide **how many mins of movie should he make** for get best IMDb range.

https://github.com/ShubhamPadiya2002/Excel-Project

Thank You !!

For see till the end

Shubham Padiya

Data Enthusiast

-- Click Here

-- Click Here