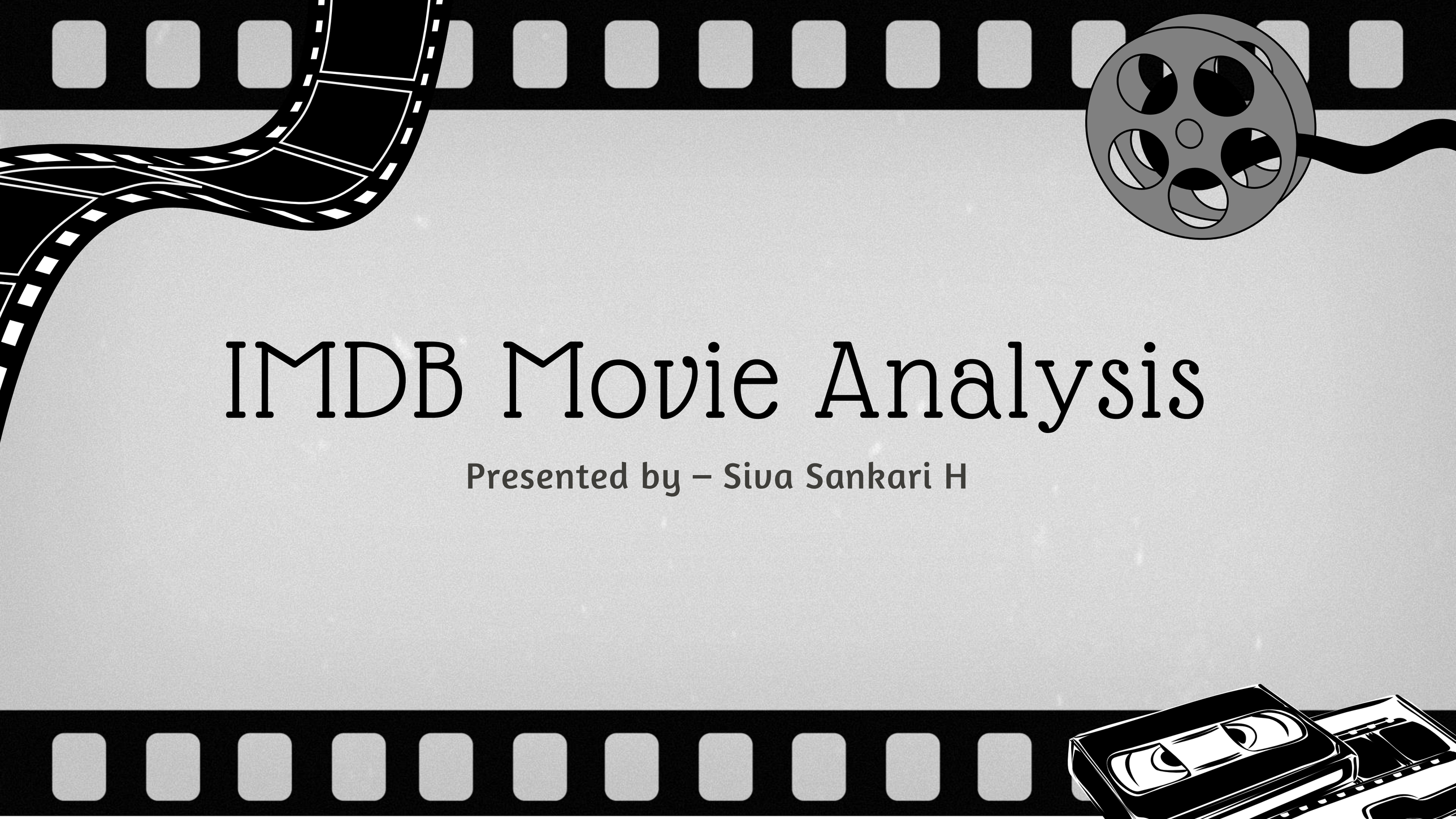


IMDB Movie Analysis

Presented by – Siva Sankari H

Project Description

- IMDb analysis refers to the process of examining and interpreting data from the Internet Movie Database (IMDb).
- 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?"
- 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.

PROJECT GOALS

Main aim of the project is to perform the following analysis.

- Movie Genre Analysis
- Duration Analysis
- Language Analysis
- Director Analysis
- Budget Analysis

Pre-Processing

Data Cleaning – I

- Removed All duplicate entries in the provided dataset.
- We are now left with 9 columns and 3786 rows.

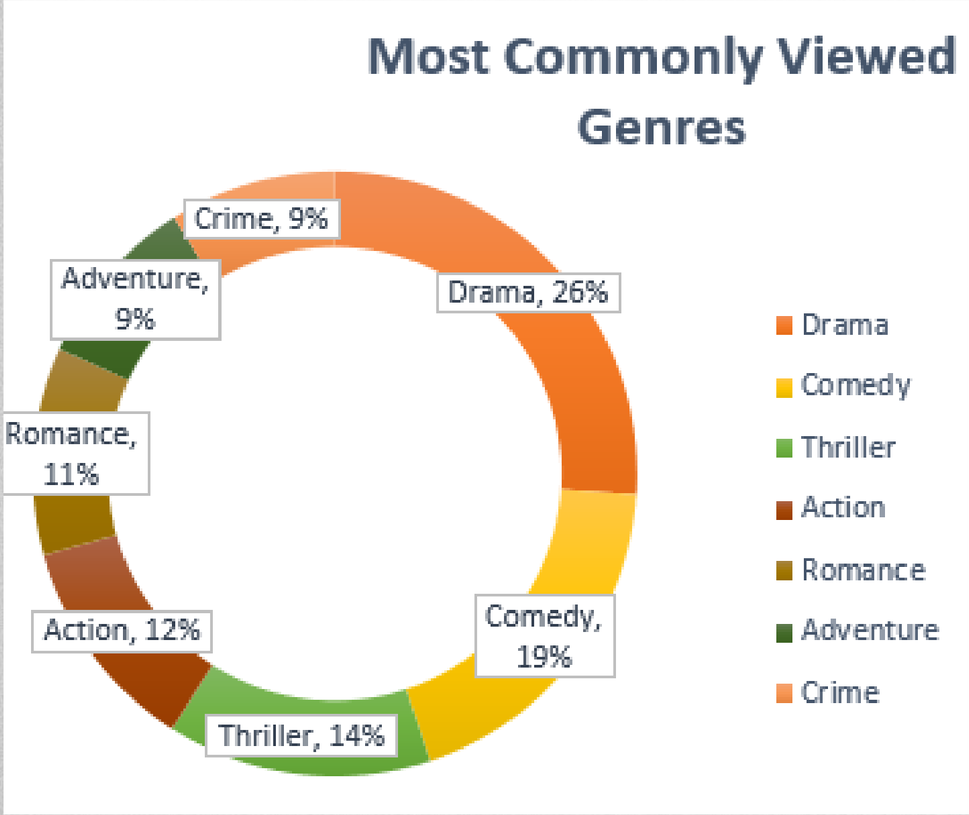
Data Cleaning – II

- Removed unwanted columns from the dataset for the ease of analysis

Data Handling

- There were certain blank rows in some columns – removed all of them.
- For certain analysis, I have normalized IMDB scores, and averages.

1. Movie Genre Analysis



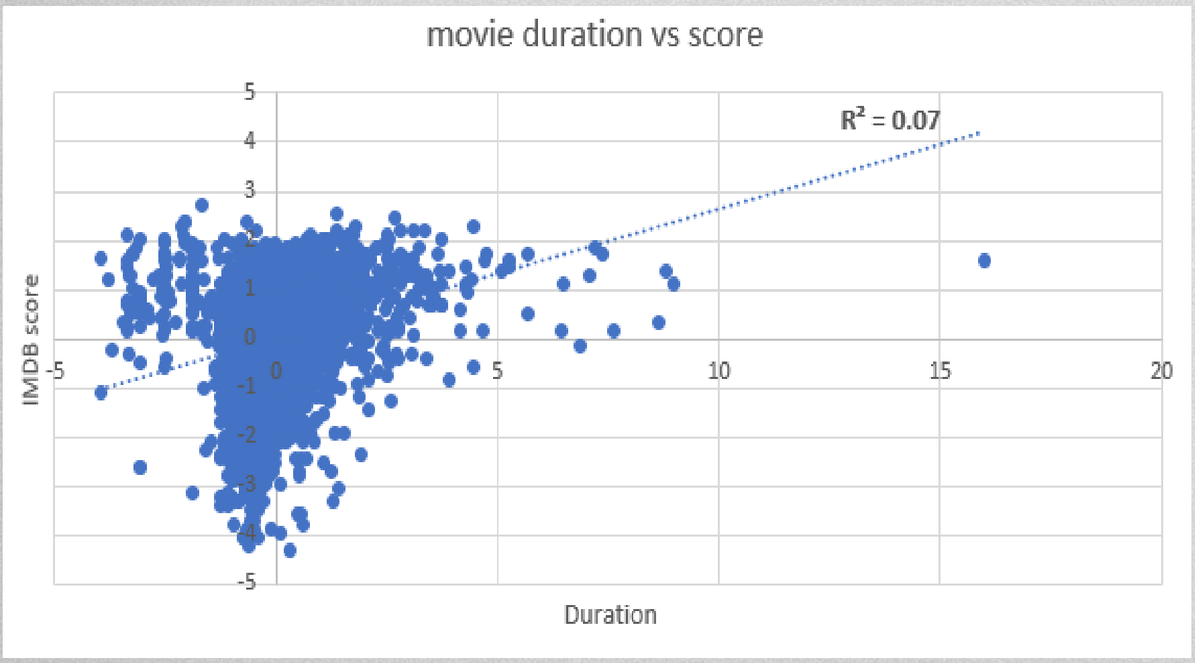
Task: Determine the most common genres of movies in the dataset.

Findings and Analysis –

1. Main stream genres are Drama, Comedy, Action, Romance, and thriller.
2. Movies outside the mainstream genres often struggle to find a large audience, leading to lower box office success.

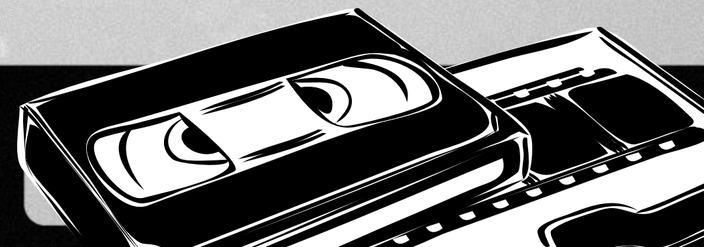
2. Movie Duration Analysis

Task: Analyse the distribution of movie durations and identify the relationship between movie duration and IMDB score.



Findings and Analysis –

1. Poor correlation between movie duration and IMDB score with $R^2 = 0.07$ – meaning only 7% of the score can be explained by the IMDB Rating.
2. The low R^2 value suggests that movie duration is not a good predictor of score.
3. Other factors, such as genre, plot, acting, and directing, likely play a much bigger role in determining a movie's score than its duration.



3. Language Analysis

Language	Count	Sum	Mean	Variance	SD	Maximum
English	4585	29315.30	6.39	1.27	1.13	9.50
French	73	513.80	7.04	0.52	0.72	8.40
Spanish	40	277.50	6.94	0.71	0.84	8.20
Hindi	28	185.70	6.63	1.89	1.37	8.50
Mandarin	24	162.90	6.79	1.03	1.02	7.90
German	19	139.50	7.34	0.86	0.93	8.50
Japanese	17	124.90	7.35	0.94	0.97	8.70
Italian	11	79.50	7.23	1.41	1.19	8.90
Russian	11	70.00	6.36	1.74	1.32	8.10
Cantonese	11	76.50	6.95	0.45	0.67	7.80

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

Findings and Analysis –

1. English is the most used language with an average rating of 6.9.
2. The large global audience for English-language films leads to more ratings, potentially boosting scores.

4. Director Analysis

Director name	Movies Made	Average imdb score	percentile score
Tony Kaye	1	8.6	0.988
Charles Chaplin	1	8.6	0.988
Alfred Hitchcock	1	8.5	0.984
Ron Fricke	1	8.5	0.984
Damien Chazelle	1	8.5	0.984
Majid Majidi	1	8.5	0.984
Sergio Leone	3	8.433	0.981
Christopher Nolan	8	8.425	0.981
S.S. Rajamouli	1	8.4	0.98
Richard Marquand	1	8.4	0.98

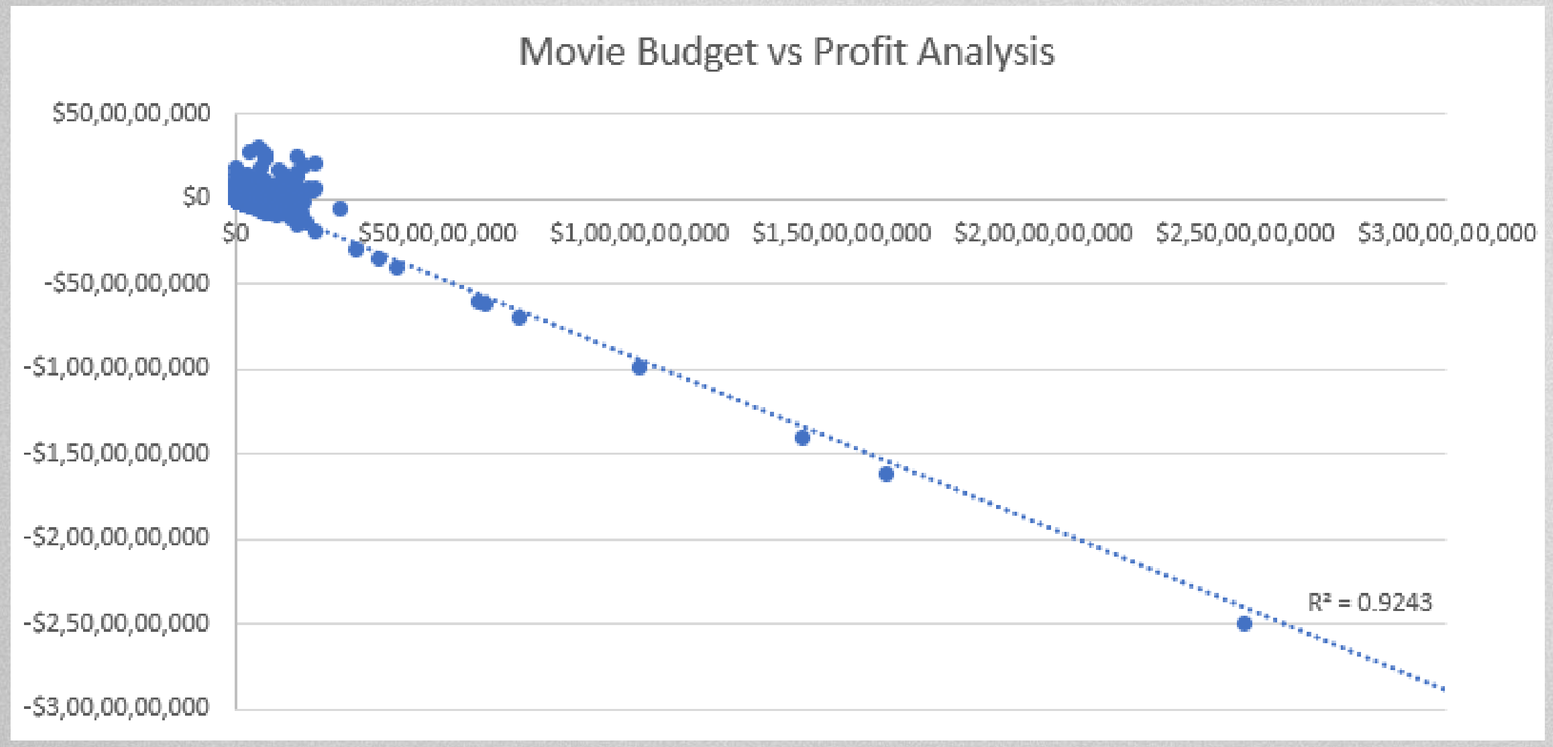
Task: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations

Findings and Analysis –

1. It can be seen that all of these directors have a success percent of above 98%.
2. All these directors belong to 98th percentile.

5. Budget Analysis

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



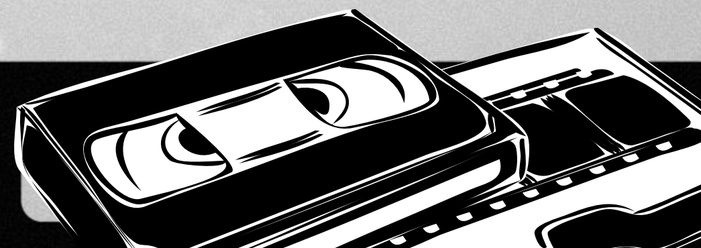
Findings and Analysis –

- Negative correlation – as movie budget increase it may not necessarily lead to profit. Profit generally decreases.
- R^2 = 0.9243 indicates a strong correlation between the variables. This means 92.4% of the profit can be explained by the budget.

Drive Links

- Excel Link –
- ExcelSheet_Modified_Dataset

- Report Link –
- Report_link



THANK YOU

Presented by: Siva Sankari H