IMDB-PROJECT PRESENTATION

PRESENTED BY MUHAMMED FATHY

IMDb TOP 10 HIGHEST-RATED

FILMS OF THE '00s

1	 The Dark Knight	★9.0
2	 The Lord of the Rings: The Return of the King	★9.0
3	 The Lord of the Rings: The Two Towers	★8.8
4	 The Lord of the Rings: The Fellowship of the Ring	★8.8
5	 City of God	★8.6
6	 Spirited Away	★8.6
7	 The Prestige	★8.5
8	 The Departed	★8.5
9	 The Pianist	★8.5
10	 Gladiator	★8.5

INTRODUCTION

Data Overview & Preparation

We worked with a dataset containing 10,866 movies and 21 features per entry. To begin our analysis, we explored the dataset's structure using functions like df.columns, df.shape, df.head(), and df.info() to understand the types of data we were dealing with. The features include details such as budget, revenue, cast, director, genres, release dates, and more. Initial exploration helped us get a sense of the dataset's scale and potential for insight extraction.

IMDB

INTRODUCTION

Data Cleaning

After identifying key issues during our initial exploration, we performed a thorough cleaning process. Using df.isnull().sum(), we detected several missing values — most notably, over 7,834 nulls in the 'homepage' column, and smaller but important gaps in fields like 'cast', 'genres', 'overview', and 'director'. For critical features like 'cast' and 'director', we dropped rows with missing values to maintain the quality of insights. For textual fields like 'homepage', 'tagline', and 'keywords', we replaced nulls with 'Unknown'.

We also checked for duplicated entries using df.duplicated().sum() and found only 1 duplicate, which we removed. This step ensured each movie record was unique. Additionally, we reviewed the 'director' column and discovered there were 5,014 unique directors in the dataset, highlighting its diversity.

INTRODUCTION

Data Cleaning

```
df.isnull().sum()
✓ 0.0s
```

id	0
imdb_id	0
popularity	0
budget	0
revenue	0
original_title	0
cast	0
homepage	7834
director	0
tagline	2725
keywords	1422
overview	0
runtime	0
genres	0
production_companies	954
release_date	0
vote_count	0
vote_average	0
release_year	0
budget_adj	0
revenue_adj	0

dtype: int64


```
df.isnull().sum()
✓ 0.0s
```


id	0
imdb_id	0
popularity	0
budget	0
revenue	0
original_title	0
cast	0
homepage	0
director	0
tagline	0
keywords	0
overview	0
runtime	0
genres	0
production_companies	0
release_date	0
vote_count	0
vote_average	0
release_year	0
budget_adj	0
revenue_adj	0

dtype: int64

```
#df.duplicated().any()
df.duplicated().sum()
```

Chat (CTRL + I) / Share (CTRL

```
[31] ✓ 0.0s
```

```
np.int64(1)
```

```
df = df.drop_duplicates()
#df = df.drop_duplicates(subset=['column1', 'column2'])
[32] ✓ 0.0s
```

```
df.duplicated().any()
[33] ✓ 0.0s
```

```
np.False_
```

DATA HAS BEEN CLEANED

IMDB

INTRODUCTION

Conclusion of Initial EDA Phase

By the end of the cleaning phase, we had a refined and reliable dataset ready for deeper exploration and visualization.

With all critical nulls handled and duplicate rows removed, the data is now structured to support meaningful insights and accurate visual storytelling. These preparations laid a strong foundation for the next phase of the analysis — exploratory data analysis and identifying key performance trends.

ANALYSIS INTRODUCTION

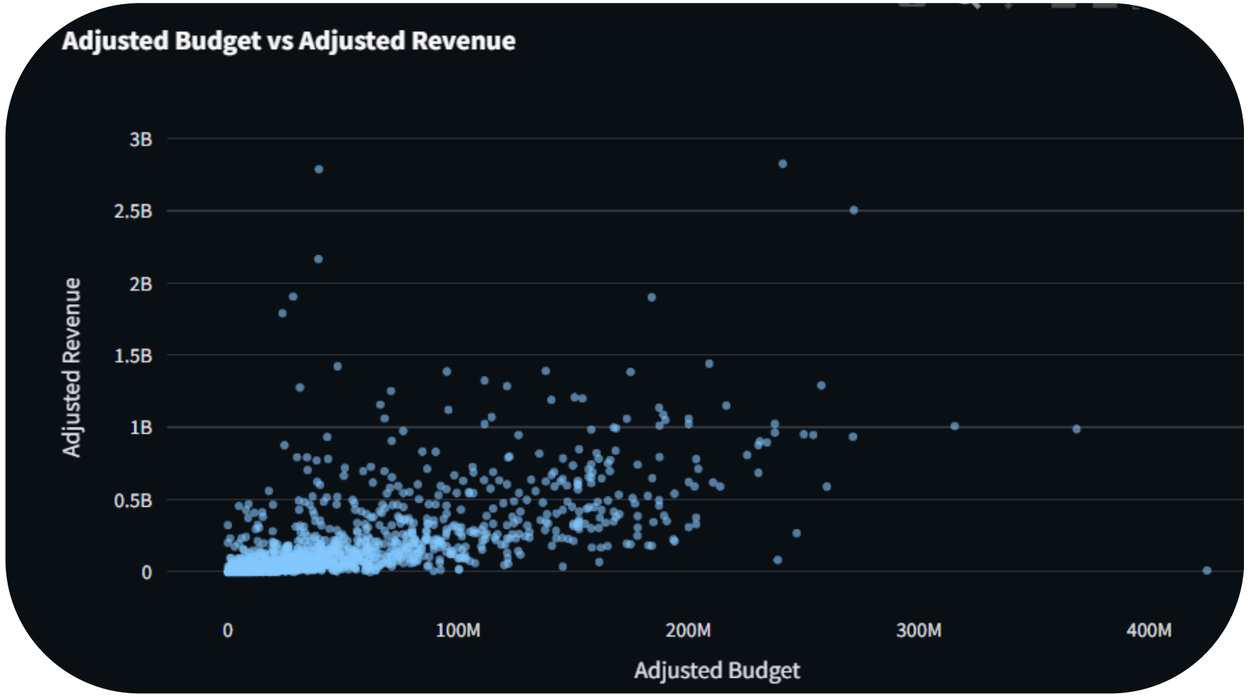
I began my analytical journey with a desire to understand what makes a movie successful: is it the budget? the genre? the runtime? the audience ratings?
To answer these questions, I used IMDB movie data and organized the analysis into 10 main axes, each supported by an interactive visual plot using plotly express and other libraries.

IMDB

ANALYSIS

1. **Adjusted Budget vs Adjusted Revenue:**

The first thing I did was draw a scatter plot showing the relationship between adjusted budget and adjusted revenue. I noticed a clear positive correlation: the higher the budget, the higher the revenue in most cases. But there were exceptions: some high-budget films earned surprisingly low revenue. This analysis showed that while budget is important, it is not the only factor.

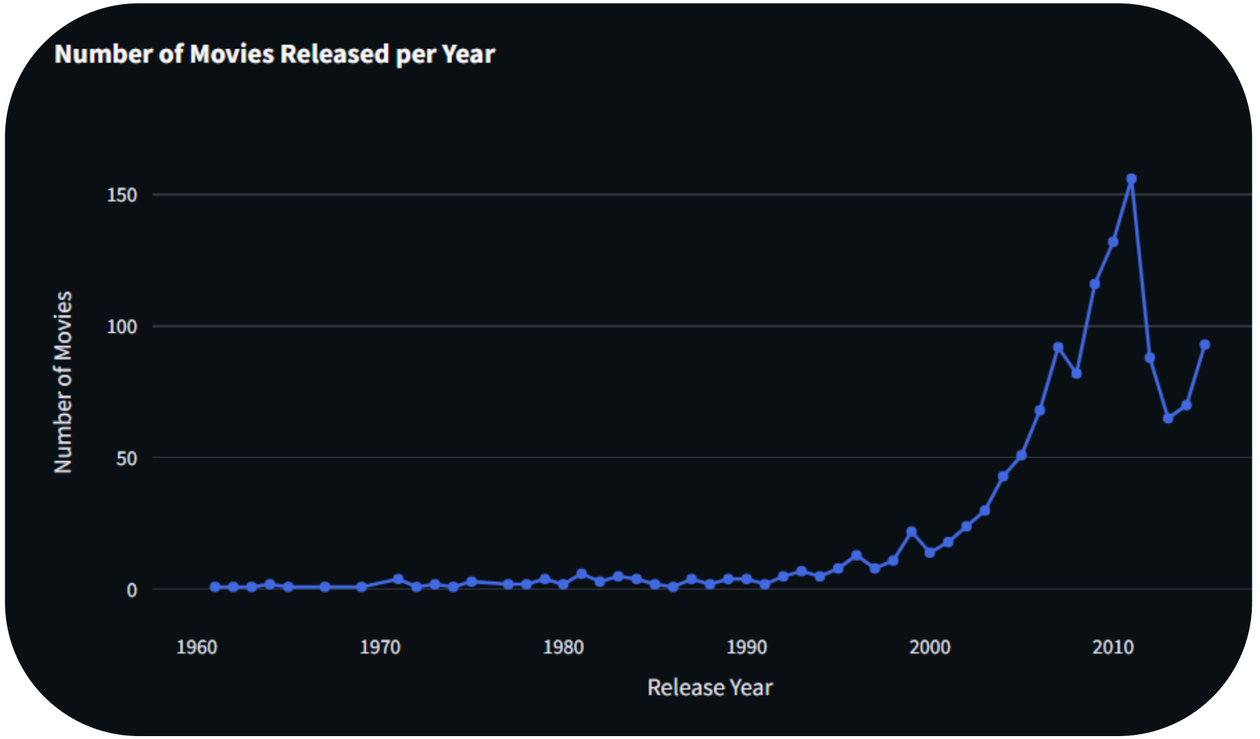


IMDB

ANALYSIS

2. **Movies Produced per Year:**

I created a bar chart showing the number of films produced each year. The result was impressive: there's a clear upward trend in production over time, especially after the 1990s. This indicates that the film industry is growing, and the opportunities for success are increasing with time.

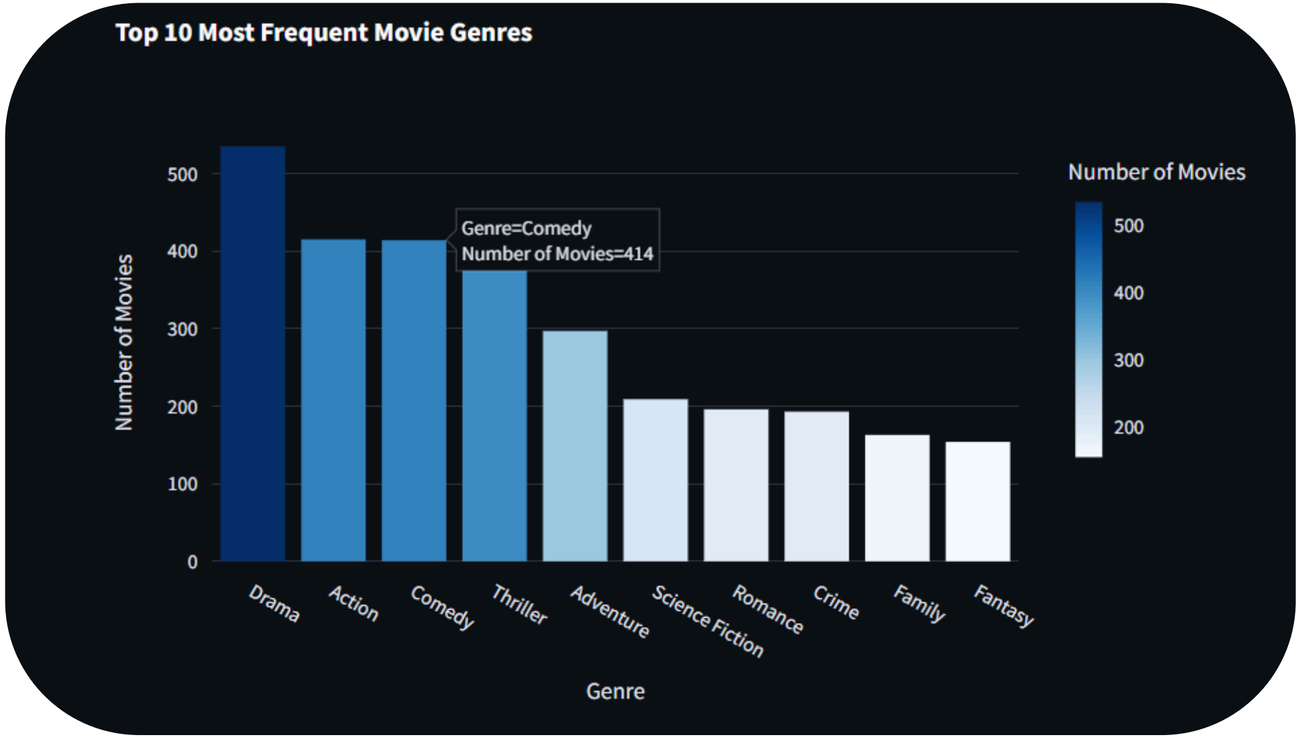


IMDB

ANALYSIS

3. **Most Common Genres:**

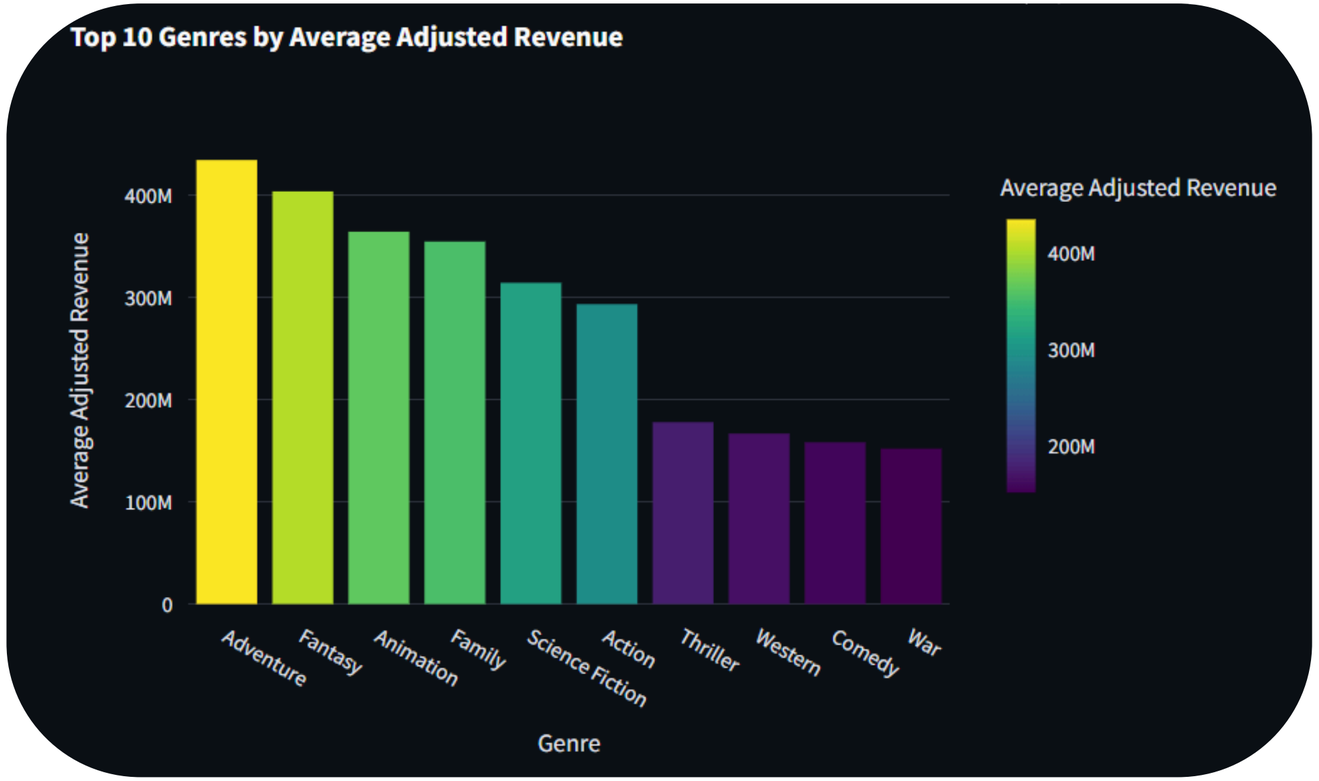
Next, I analyzed movie genres. Since each movie can belong to more than one genre, I exploded the genre column and counted how many films fall under each. I found that the most common genres are Drama, Action, and Comedy. But this doesn’t mean they are the most financially or critically successful.



ANALYSIS

4. Average Revenue by Genre:

Here, I dove into profitability. I calculated the average adjusted revenue for each genre. Adventure and Fantasy films had the highest average revenues. Even though these genres aren't the most common, they tend to earn the most. This proved that genre has a strong impact on a film's earnings.

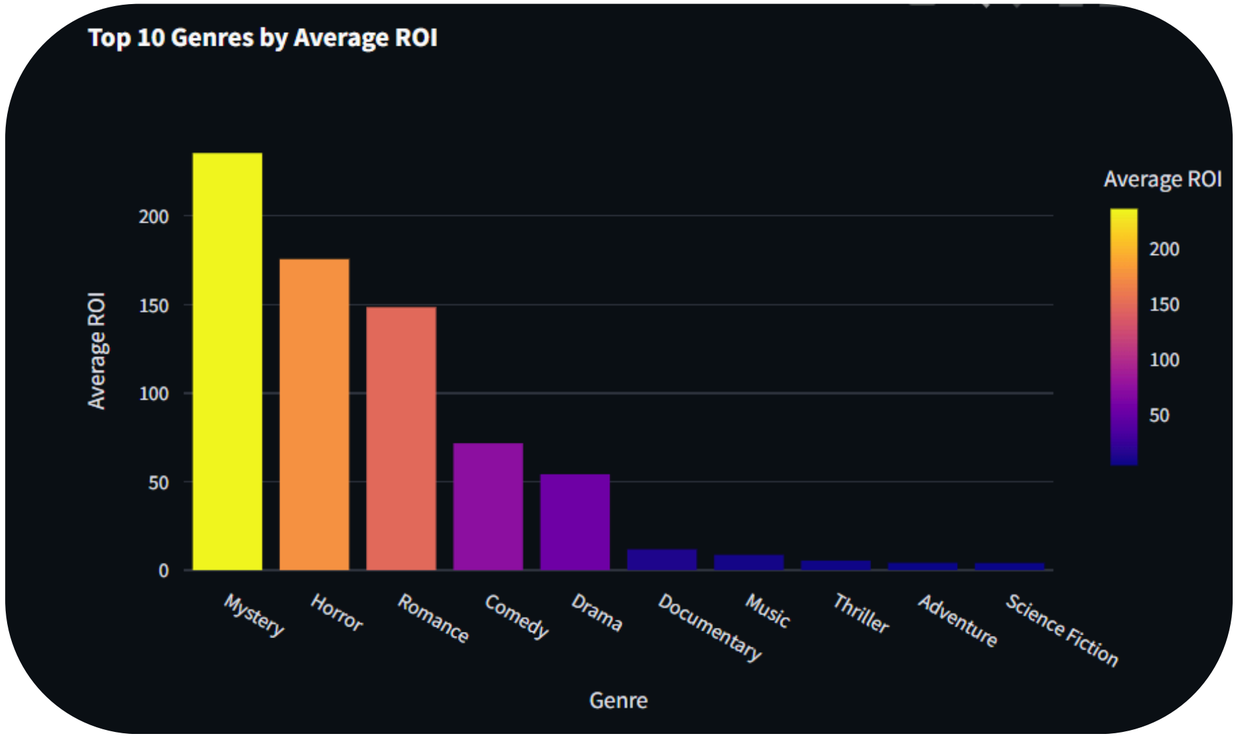


IMDB

ANALYSIS

5. Return on Investment (ROI) by Genre:

To complete the picture, I calculated the ROI metric for each movie, which links profit to budget. Then I aggregated the data by genre and found that genres like Horror and Documentary achieved very high ROI. That's because they require small budgets but can still earn solid revenues. So ROI gives a deeper insight than just absolute revenue.



IMDB

VISULAIZATION BY STREAMLIT

To make the analysis more interactive and engaging, Streamlit was used to display the visualizations. This allowed for a clean and responsive presentation of the data insights, making it easier to explore patterns, trends, and KPIs dynamically.
So you can find the rest of the plots there

The link is here Click here

MORE ANALYSIS

6. Average Rating by Genre:

Here I focused on audience opinion. I calculated the average IMDB rating for each genre. The result? Foreign and War movies received the highest ratings.
This is crucial for directors and producers who care about audience satisfaction, not just profit.

7. Runtime vs Revenue:

I wanted to know if movie length affects profit. So I drew another scatter plot, this time with runtime on the X-axis and revenue on the Y-axis. I observed that medium to long-duration films tend to earn higher revenues. However, the relationship isn't strongly linear—runtime alone isn't decisive.

MORE ANALYSIS

8. Top Movies (by Revenue or Rating):

I looked at the highest-grossing or highest-rated movies and analyzed their features: genre, runtime, and budget. The goal here was to learn from successful models.

9. Top Directors by Average Adjusted Revenue:

To add a human factor to the analysis, I focused on the most prolific directors and how successful their movies are. I found that certain directors consistently produce high-revenue films, suggesting the influence of directorial vision and experience. This helps identify the people behind the success—not just the movies themselves.

MORE ANALYSIS

10. Budget vs Audience Ratings:

Lastly, I examined whether higher budgets correlate with better audience ratings. Surprisingly, there was no strong pattern. Some low-budget films received high ratings, and some high-budget films received mediocre ones. This highlights that audience satisfaction doesn’t depend solely on budget size, but rather on storytelling, originality, and direction.

KEY INSIGHTS & PATTERN

Key Insights:

- High budgets usually lead to higher revenue, but not always.
- Some underproduced genres (like Horror) generate very high ROI.
- Highly rated genres aren't always the most profitable but often satisfy audiences.
- Film production has increased significantly over time.
- Medium to long-duration films tend to perform better.
- Certain directors consistently produce high-performing movies.
- Large budgets don't guarantee high audience ratings.

IMDB

KEY INSIGHTS & PATTERN

Notable Pattern:

Genres and directors that balance quality (high ratings), low cost (small budgets), and strategic decisions often achieve excellent results in both profit and popularity.

This opens the door for producers, investors, and creators to make smarter and more informed decisions.

IMDB

Thank you very much!

PRESENTED BY MUHAMMED FATHY