

**TASK**

**Exploratory Data Analysis on the Movies Data Set**

[](https://www.hyperiondev.com/)

Movies have always been an essential form of entertainment for masses around the world. Through this EDA, I aim to uncover insights about the movie industry, understand trends, and analyse the factors contributing to the success or failure of a movie.

**SUMMARY OF THE DATA SET**

The dataset comprises information on various movies, capturing attributes like budget, revenue, genres, popularity, and ratings. The dataset also offers insights into movie production details, such as release dates and production companies.

**DATA CLEANING**

Columns Removed: 'homepage', 'keywords', 'original\_language', 'original\_title', 'overview', 'production\_companies', 'status', 'tagline'

Reason:

To make the analysis concise and focused, columns that were deemed redundant or unlikely to contribute to the main analyses were removed.

**MISSING DATA**

Handling Zero Values:

Movies with a budget or revenue of zero were considered to have missing or unrecorded data and were discarded.

Reason:

Keeping such entries could skew the analysis, especially when comparing profitability or assessing return on investment.

**DATA STORIES AND VISUALISATIONS**

1.Most Expensive Movies

Visualized the top 5 movies with the highest budgets:

* Pirates of the Caribbean: On Stranger Tides 380 million USD
* Pirates of the Caribbean: At World's End 300 million USD
* Avengers: Age of Ultron 280 million USD
* Superman Returns 270 million USD
* John Carter 260 million USD

Insight: High budget doesn't necessarily correlate with high ratings or popularity.

2. Most Profitable Movies

Plotted the difference between revenue and budget for each movie:

• Avatar 2.6 billion USD

• Titanic 1.6 billion USD

• Jurassic World 1.4 billion USD

• Furious 7 1.3 billion USD

• The Avengers 1.3 billion USD

Maximum Profit: $2,550,965,087.00

Minimum Profit: $-165,710,090.00

Insight: Some of the most profitable movies weren't the most expensive ones.

3. Popularity Analysis

Sorted movies based on popularity:

1. Minions

2. Interstellar

3. Deadpool

4. Guardians of the Galaxy

5. Mad Max: Fury Road

6. Jurassic World

7. Pirates of the Caribbean: The Curse of the Black Pearl

8. Dawn of the Planet of the Apes

9. The Hunger Games: Mockingjay - Part 1

10. Big Hero 6

Insight: Popularity doesn't always correlate with budget or revenue. Some low budget movies can be extremely popular.

4. Highly Rated Movies:

Showcased movies with ratings above 7. Here are 10 examples:

1. Avatar

2. The Dark Knight Rises

3. Tangled

4. Avengers: Age of Ultron

5. Harry Potter and the Half-Blood Prince

6. Roger & Me

7. Eraserhead

8. Pi

9. Clerks

10. Cure

Insight: A significant number of movies have received good ratings, highlighting quality content production.

5. Successful Genres:

Analysed the distribution of movie genres.

Insight: Certain genres are more prevalent, reflecting viewer preferences and possibly indicating trends in profitable movie types. The bar graph told us that the most frequent genre of movie is drama, followed by comedy, thriller and then action. The least frequent genre is foreign.

**CONCLUSIONS**

The movie industry is multifaceted, with multiple variables influencing a movie's success. While factors like budget and genre play a role, they aren't sole determinants of a movie's success, as evidenced by our analysis. This EDA offers a foundation for further in-depth analyses and predictive modelling.

**THIS REPORT WAS WRITTEN BY: Rameez Bhayed**

