# Movies Dataset Analysis Project Report

## 1. Introduction

This project analyzes a dataset of movies released between 2006 and 2016. The main objective is to understand how factors such as Genre, Rating, Revenue, Votes, and Year influence movie performance. The analysis explores patterns and relationships among these variables to identify what drives audience engagement and box-office success.

## 2. Dataset Overview

• Dataset Source: IMDb Movie Dataset  
• Dataset Size: 1000 rows × 12 columns  
  
Columns Overview:  
- Title: Movie name  
- Genre: Movie category (e.g., Action, Drama, Comedy)  
- Director: Director of the movie  
- Actors: Main cast members  
- Year: Year of release  
- Runtime (Minutes): Duration of the movie  
- Rating: IMDb user rating (1–10)  
- Votes: Number of audience votes  
- Revenue (Millions): Box office revenue in millions  
- Metascore: Critics’ score (0–100)

## 3. Data Cleaning and Preprocessing

The dataset contained missing values in the 'Revenue (Millions)' and 'Metascore' columns. These missing values were handled using imputation with mean values. Data types were corrected where necessary — particularly for numerical columns like Revenue and Votes. Outliers in the 'Rating' column were treated using the Interquartile Range (IQR) method to remove extreme values. A new categorical column 'Rating\_Group' was created using rating bins (4★–8★) to simplify analysis.

## 4. Data Analysis and Visualization Summary

The exploratory analysis focused on genre distribution, revenue trends, ratings, metascore influence, and director performance. Key patterns and insights were visualized using histograms, scatter plots, and bar charts.

## 5. Key Insights and Observations

1. Action, Adventure, and Sci-Fi genres dominate movie production and revenue, showing that audiences prefer visually engaging films.  
2. Movies with higher ratings (7★–8★) generate higher revenue, proving that quality influences box-office success.  
3. Between 2006 and 2016, movie production increased nearly sixfold, indicating a strong industry expansion.  
4. Action–Adventure–Fantasy movies earned the highest revenue (~$936M), while low-rated films earned significantly less.  
5. Votes strongly correlate with revenue — movies watched and rated by more audiences tend to earn more.  
6. Metascore has a weak relationship with revenue, meaning critic reviews don’t necessarily determine financial success.  
7. Directors like J.J. Abrams, James Cameron, and Christopher Nolan achieved the highest box-office revenues with large-scale franchises.

## 6. Conclusion

The analysis shows that commercial movie success is driven by a mix of engaging genres, strong audience ratings, and wide audience reach. While critical reviews (Metascore) have limited impact, audience popularity and franchise value significantly influence box-office results. Overall, the data highlights that balancing entertainment value, storytelling, and large-scale appeal is key to achieving both popularity and profitability in the movie industry.