# Movie Dataset Analysis Report

## 1. Introduction

This report presents an analysis of a movie dataset consisting of 9,827 rows and 9 columns. The dataset appears tidy with no missing (NaN) or duplicate values. The 'release\_date' column was converted into datetime format, and only the year value was extracted for analysis. Columns such as 'overview', 'original\_language', and 'poster\_url' were deemed unnecessary for the analysis and were dropped. Notable outliers were found in the 'popularity' column, and 'vote\_average' was categorized for better insights. The 'genre' column, containing comma-separated values and extra white spaces, was cleaned and transformed into a categorical format.

## 2. Data Cleaning Summary

- Removed unnecessary columns: 'overview', 'original\_language', 'poster\_url'.  
- Converted 'release\_date' to datetime and extracted the year.  
- Detected and noted outliers in 'popularity'.  
- Categorized 'vote\_average' for better interpretation.  
- Split and cleaned 'genre' column values.

## 3. Insights & Findings

### Q1: What is the most frequent genre in the dataset?

The Drama genre is the most frequent in the dataset, appearing in more than 14% of the entries among 19 different genres.

### Q2: What genre has the highest votes?

Approximately 25.5% of the dataset (6,520 movies) have high popularity votes. Drama once again leads in popularity, being the top choice among fans with more than 18.5% of popular movies.

### Q3: What movie got the highest popularity? What's its genre?

'Spider-Man: No Way Home' recorded the highest popularity rate in the dataset. Its genres include Action, Adventure, and Science Fiction.

### Q4: What movie got the lowest popularity? What's its genre?

'The United States, Thread' recorded the lowest popularity rate in the dataset. Its genres include Music, Drama, War, Sci-Fi, and History.

### Q5: Which year has the most filmed movies?

The year 2020 recorded the highest number of movie releases in the dataset.

## 4. Conclusion

From this analysis, Drama emerged as the most frequent and popular genre, both in terms of occurrences and audience votes. 2020 was the most active year in terms of movie releases. The findings suggest that genres like Drama, Action, and Adventure have a significant influence on popularity. These insights could help guide movie production, marketing strategies, and content curation.