1. Project Overview

This project explores Netflix's movie catalog to uncover patterns in genre distribution, box office performance, popularity trends, and viewer ratings. By analyzing metadata from 499 movies, the goal is to identify dominant genres, high-performing titles, and release year trends that can inform content strategy, audience targeting, and recommendation systems.

2. Dataset Summary

- **Rows**: 499

- **Columns**: 9

- **Key Features**:

    Release Information: Release_Date (year of release)

    Movie Metadata: Title, Overview, Genre, Original_Language

    Performance Metrics: Popularity, Vote_Count, Vote_Average, Box_Office

    Visual Reference: Poster_Url (movie poster link)

3. Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

● **Data Loading:** Imported the dataset using pandas.

● **Initial Exploration:** Used df.info() to check structure and .describe() for summary statistics.

	Release_Date	Title	Overview	Popularity	Vote_Count	Vote_Average	Original_Language	Genre	Poster_Url
0	2021-12-15	Spider-Man: No Way Home	Peter Parker is unmasked and no longer able to...	5083.954	8940	8.3	en	Action, Adventure, Science Fiction	https://image.tmdb.org/t/p/original/1g0dhYtq4i...
1	2022-03-01	The Batman	In his second year of fighting crime, Batman u...	3827.658	1151	8.1	en	Crime, Mystery, Thriller	https://image.tmdb.org/t/p/original/74xTEgt7R3...
2	2022-02-25	No Exit	Stranded at a rest stop in the mountains durin...	2618.087	122	6.3	en	Thriller	https://image.tmdb.org/t/p/original/vDHsLnOWKl...
3	2021-11-24	Encanto	The tale of an extraordinary family, the Madri...	2402.201	5076	7.7	en	Animation, Comedy, Family, Fantasy	https://image.tmdb.org/t/p/original/4j0PNHkMr5...
4	2021-12-22	The King's Man	As a collection of history's worst tyrants and...	1895.511	1793	7.0	en	Action, Adventure, Thriller, War	https://image.tmdb.org/t/p/original/aq4Pwv5Xeu...

● **Missing/Duplicate Data Handling:** Checked for null values and duplicate value. There is no such data.
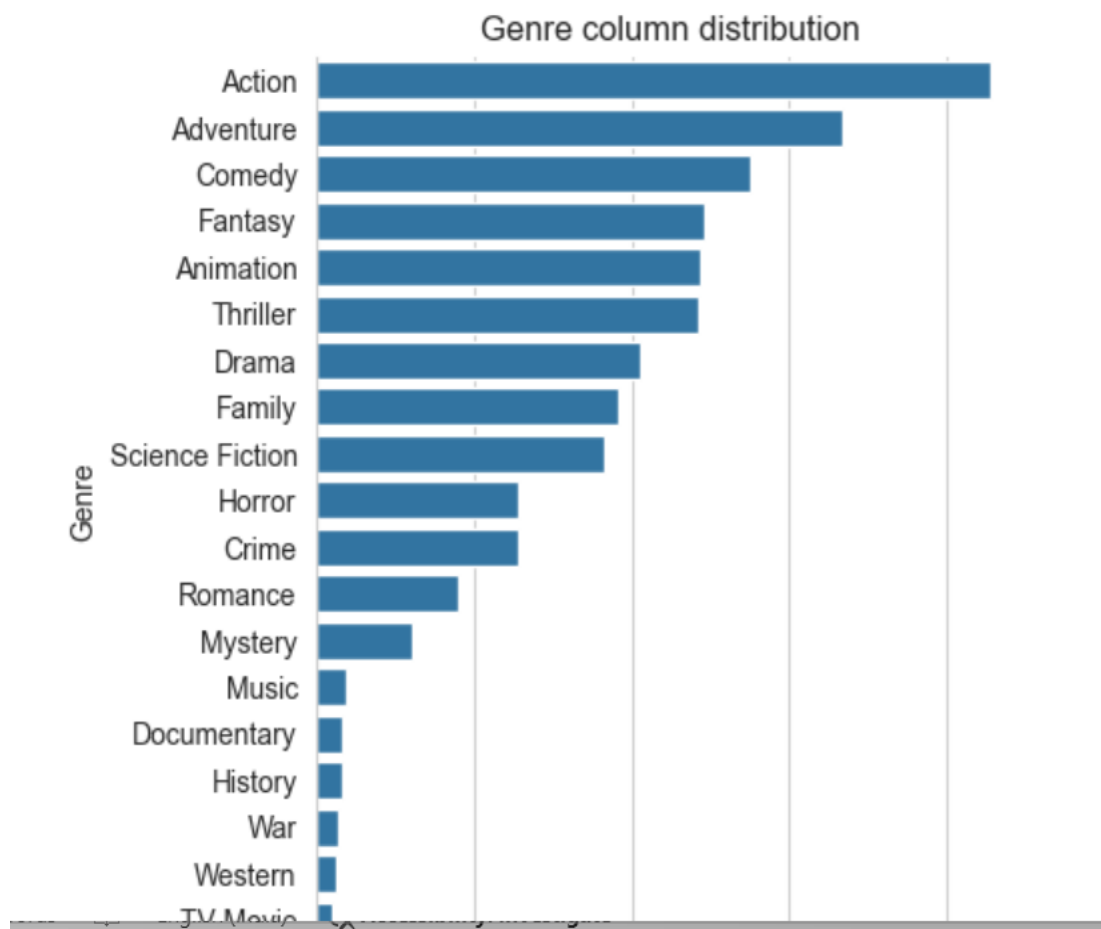
● **Feature Engineering:**

o Release_Date column needs to extract only the **year** value.

o **Overview, Original_Language, and Poster-URLs** would not be so useful during analysis, so we will drop these columns.

o Categorizing **Vote_Average** column into (super hit, hit, average, flop).

o Created new column **Box_Office** for categorized for Vote_Average.

o Splitting the **Genre** into a list and then explode our DataFrame to have only one genre per row for each movie.

o Casting **'Genre'** column into category for better memory utilization so that dataset could be more analysable and ensures insights.

4. Data Analysis using Seaborn (Business Transactions)

To answer the business questions, we applied structured analysis techniques and created visualizations with Seaborn, enabling clear identification of trends and patterns in the dataset.
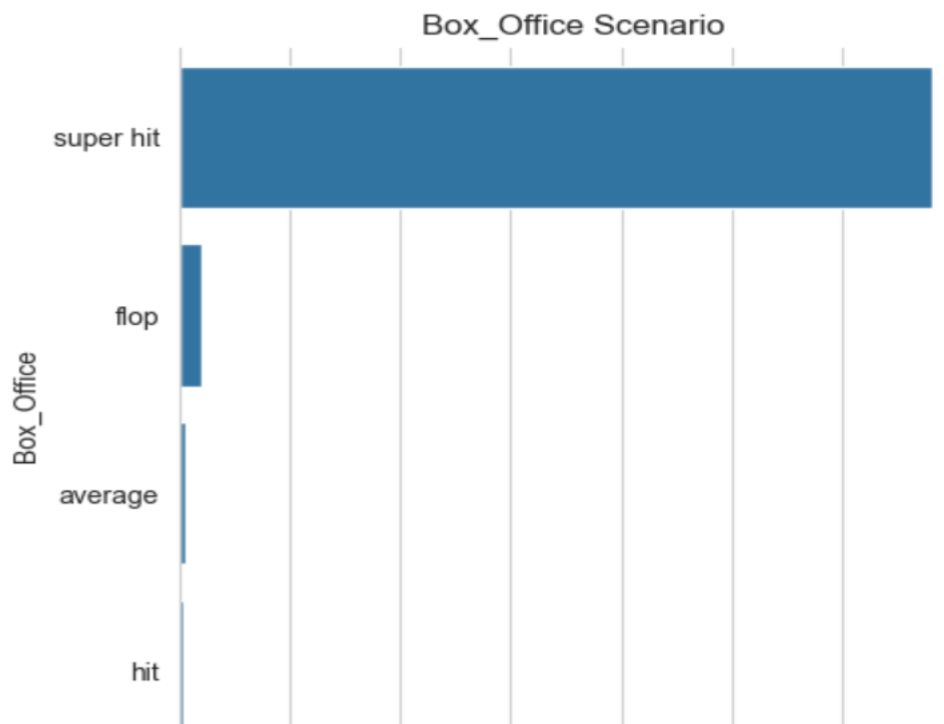
1. What is the most frequent genre in the dataset?

➢ Action genre is the most frequent genre in our dataset.

Genre column distribution

2. What is the overall Box Office scenario?

➢ Around 96% of the total movies are Superhit.

➢ Around 3% of the total movies are Flop.

Box_Office Scenario

3. Which movie has the highest popularity rate, and what is its genre?

➢ **Spider-Man: No Way Home** has the highest popularity rate in our dataset and it has genres of Action, Adventure, and Science Fiction.

```
df[df['Popularity']== df['Popularity'].max()]
```

	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Box_Office	Genre
0	2021	Spider-Man: No Way Home	5083.954	8940	8.3	super hit	Action
1	2021	Spider-Man: No Way Home	5083.954	8940	8.3	super hit	Adventure
2	2021	Spider-Man: No Way Home	5083.954	8940	8.3	super hit	Science Fiction

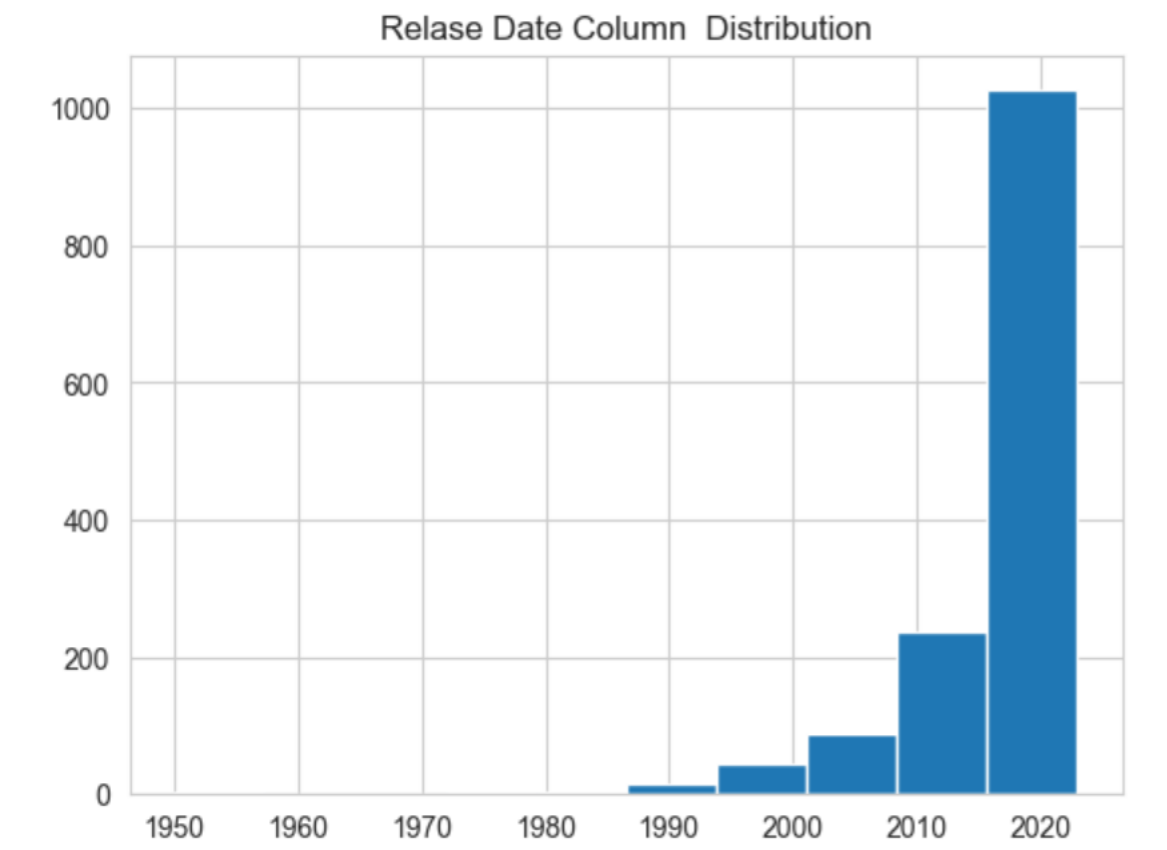## 4. Which movie has the lowest popularity rate, and what is its genre?

➢ **Furious 7** has the lowest popularity rate in our dataset and it has genres of Action, Thriller, crime, and adventure.

```
df[df['Popularity']== df['Popularity'].min()]
```

	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Box_Office	Genre
1417	2015	Furious 7	105.485	8973	7.3	super hit	Action
1418	2015	Furious 7	105.485	8973	7.3	super hit	Thriller
1419	2015	Furious 7	105.485	8973	7.3	super hit	Crime
1420	2015	Furious 7	105.485	8973	7.3	super hit	Adventure

5. Which year saw the highest number of movie releases?

➢ The year **2021** stands out as the peak release year in the dataset, with 537 movies added.

Relase Date Column Distribution

6. Business Recommendations

❖ **Content Strategy –** Focus on Action Genre

Since *Action* is the most frequent genre, Netflix should continue investing in action-oriented titles .

❖ **Box Office Success –** Leverage Superhit Titles

With more than 95% of movies classified as *Superhit*, Netflix can highlight these titles in marketing campaigns and recommendation algorithms to strengthen customer trust in its content quality.

❖ **Popularity Insights –** Promote High-Performing Movies

Spider-Man: No Way Home (Action, Adventure, Science Fiction) demonstrates strong audience appeal. Similar blockbuster franchises and superhero content should be prioritized for promotion.

❖ **Low Popularity Titles –** Revise Content Strategy

Furious 7 shows the lowest popularity despite being in multiple genres. Netflix should analyze why certain titles underperform (e.g., outdated franchise, poor reviews) and consider repositioning or limiting promotion of such content.

❖ **Release Year Trends –** Capitalize on Peak Years

With 2021 showing the highest number of releases, Netflix should investigate what drove this surge (e.g., pandemic-driven demand, production cycles) and replicate successful strategies in future years.