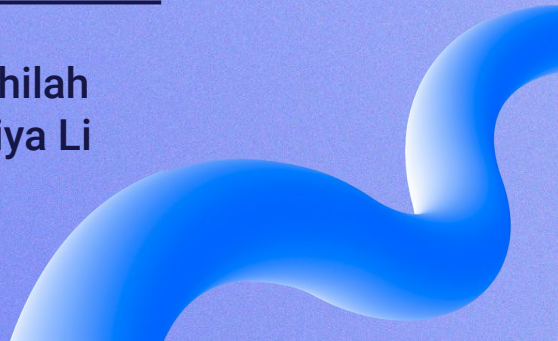





Analysing Movie Genre Popularity

Group 52 - Rafaëlla van Nee, Hari Joshithaa Aghilah
Senthilprathiban, Mara-Iuliana Dragomir, Viktoriya Li



Our Team

Rafaëlla van Nee



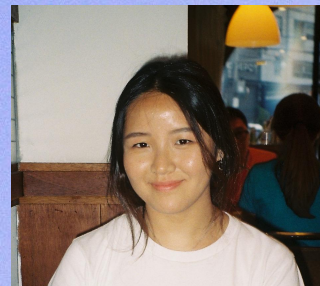
Mara-Iuliana Dragomir



Hari Joshithaa Aghilah
Senthilprathiban



Viktoriya Li



Contents



01

Research Question

02

Data Sources

03

**Data Wrangling
Methods**

04

Analysis

05

Conclusion





Research Question

How have the prevalence and popularity of different movie genres evolved over time in the film industry?

Sub Research Questions:



- What does the distribution of movie genres look like in general, and how did it change over time? Which genres were more prevalent in the movie industry in each decade?
- How have audience and critic ratings for specific movie genres (such as action, drama, comedy) evolved over the decades? Are certain genres experiencing an increase or decrease in popularity?
- Is there a correlation between movie genres and the revenue generated, and if so, which genres tend to exhibit a stronger correlation with higher revenues?
- Is there a correlation between a movie's genre and its likelihood of being part of a franchise?


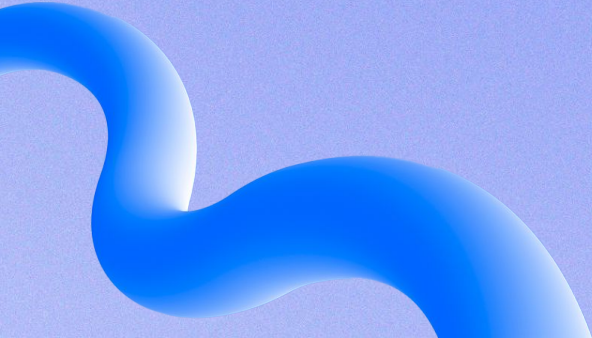




Data Sources

Kaggle



We use the “Movies Dataset” from Kaggle. The entire dataset consists of several CSV files such as `movies_metadata.csv`, `ratings.csv`, `links.csv`, `credits.csv`, and `keywords.csv`.



Introduction to the Dataset



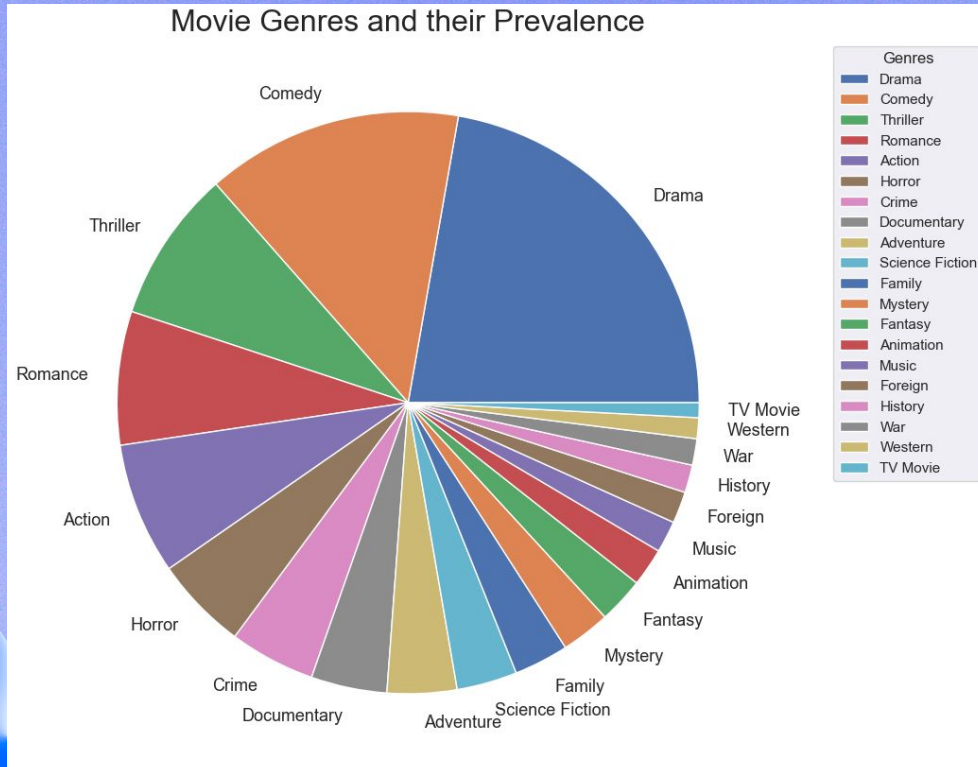
Overview of the "movies_metadata.csv" and "ratings.csv" Datasets

- Contains 45,466 rows and 24 columns.
 - Diverse data types: text, numbers, boolean, JSON-format, URLs.
 - Primarily a clean dataset with some data quality considerations.
- Contains 26,024,289 rows and 4 columns.
 - Numeric values
 - Clean dataset: no missing values or duplicate rows identified.
 - Aligned movieid column from ratings.csv with id column in movies_metadata.csv for accurate merging.
- 
- 

Data Wrangling Methods

Data Cleaning and Preparation	Extracting and Preparing Relevant Data
<ul style="list-style-type: none">• Addressed missing values and duplicates.• Removed 1,418 rows due to significant missing data.• Identified and removed 30 duplicate entries based on the movie's ID.	<ul style="list-style-type: none">• Extracted genre names from JSON-formatted entries.• Filtered to include only released movies.• Created separate dataset copies for different analyses as per subquestions.

What does the distribution of movie genres look like in general, and how did it change over time? Which genres were more prevalent in the movie industry in each decade?



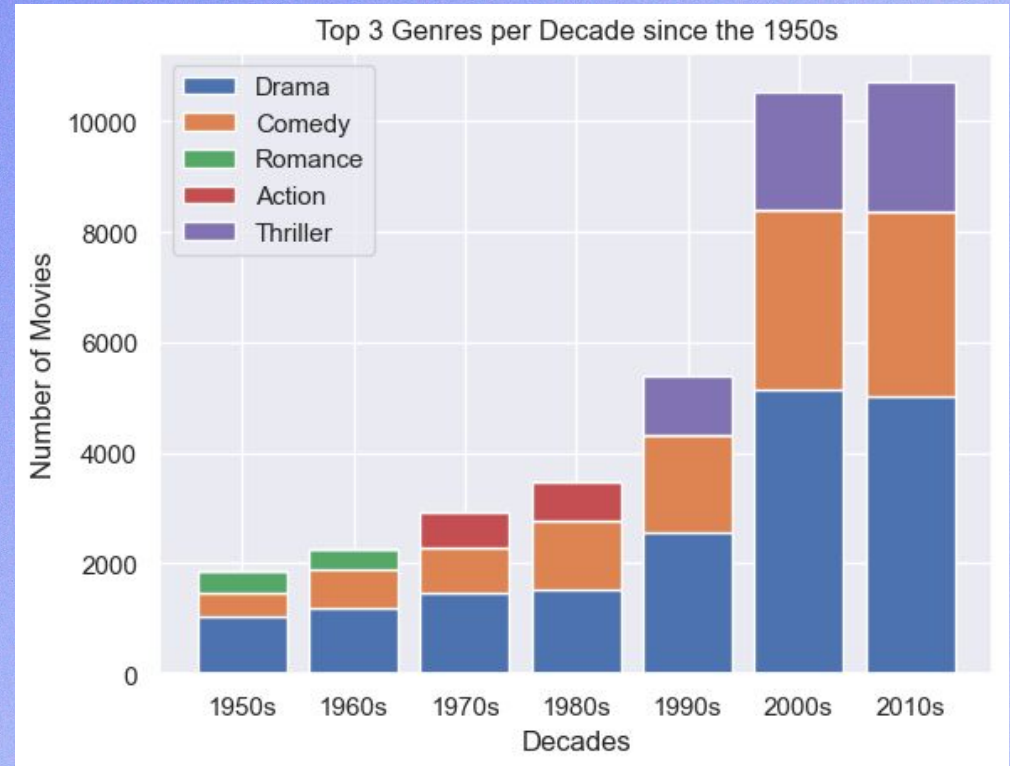
General distribution of genres:

- Most Common: Drama
- Least Common: TV Movie

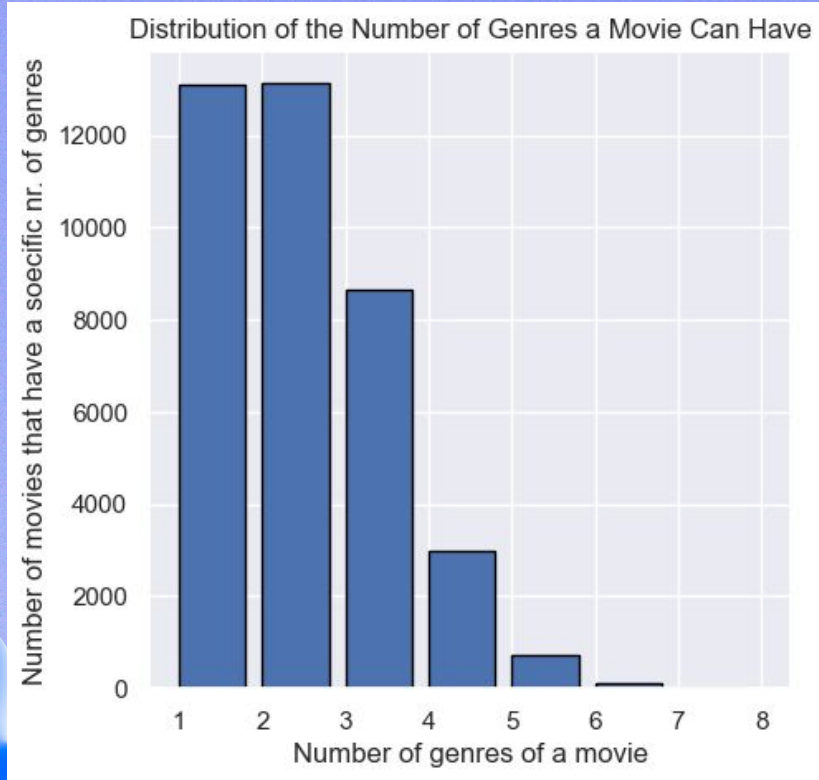
What does the distribution of movie genres look like in general, and how did it change over time? Which genres were more prevalent in the movie industry in each decade?

Top 3 Most common genre / decade

- Top 2 every decade since 1950: Drama and Comedy
- 3rd most common changes every 20 years: Romance, Action, Thriller



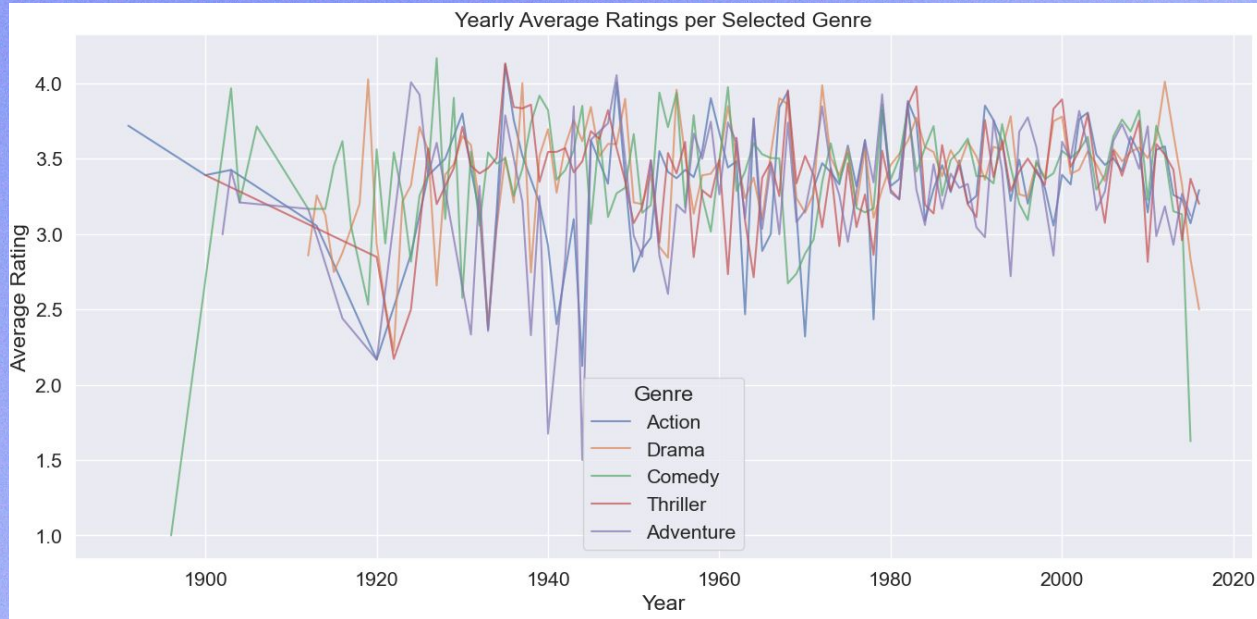
What does the distribution of movie genres look like in general, and how did it change over time? Which genres were more prevalent in the movie industry in each decade?



Number of genres a movie can have:

- Most movies have either 1 or 2 different genres

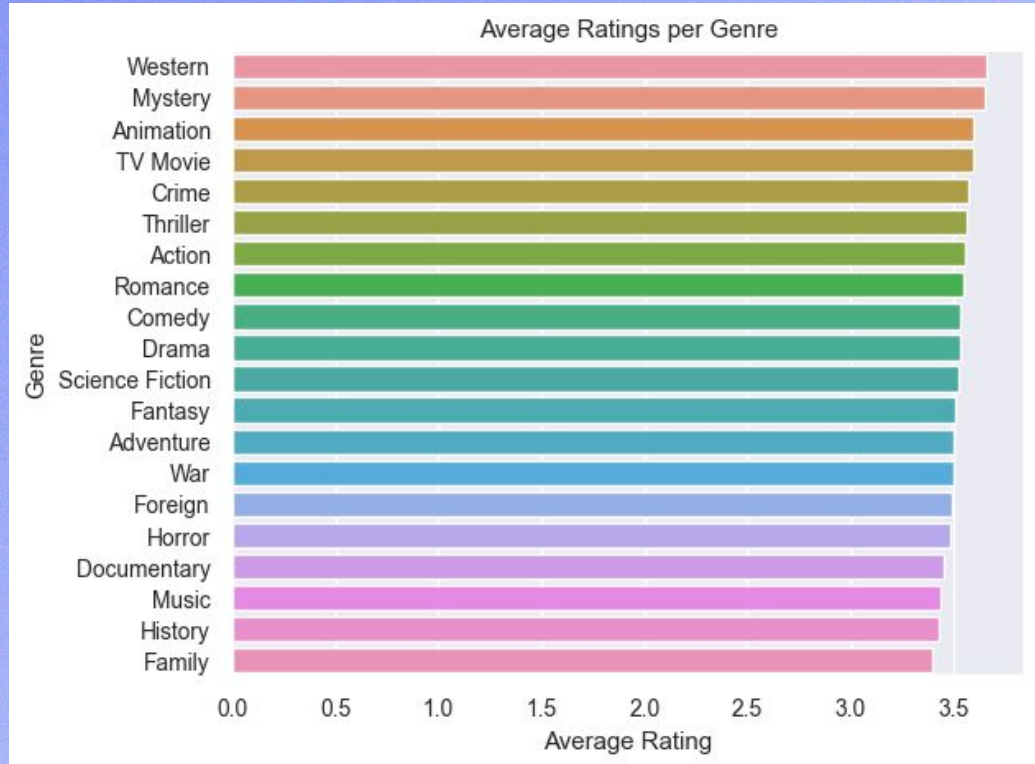
How have audience and critic ratings for specific movie genres (such as action, drama, comedy) evolved over the decades? Are certain genres experiencing an increase or decrease in popularity?



- Average ratings over time
- No specific trend for genres

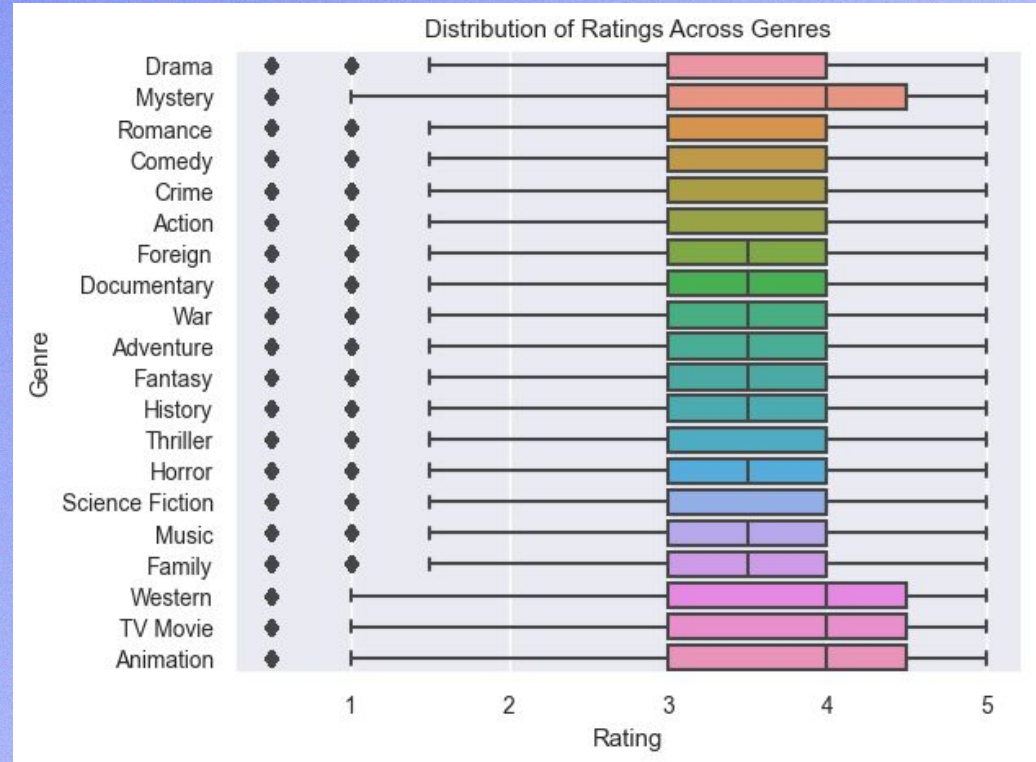
How have audience and critic ratings for specific movie genres (such as action, drama, comedy) evolved over the decades? Are certain genres experiencing an increase or decrease in popularity?

- Genres Average Rating
- Highest: Western
- Individual tastes

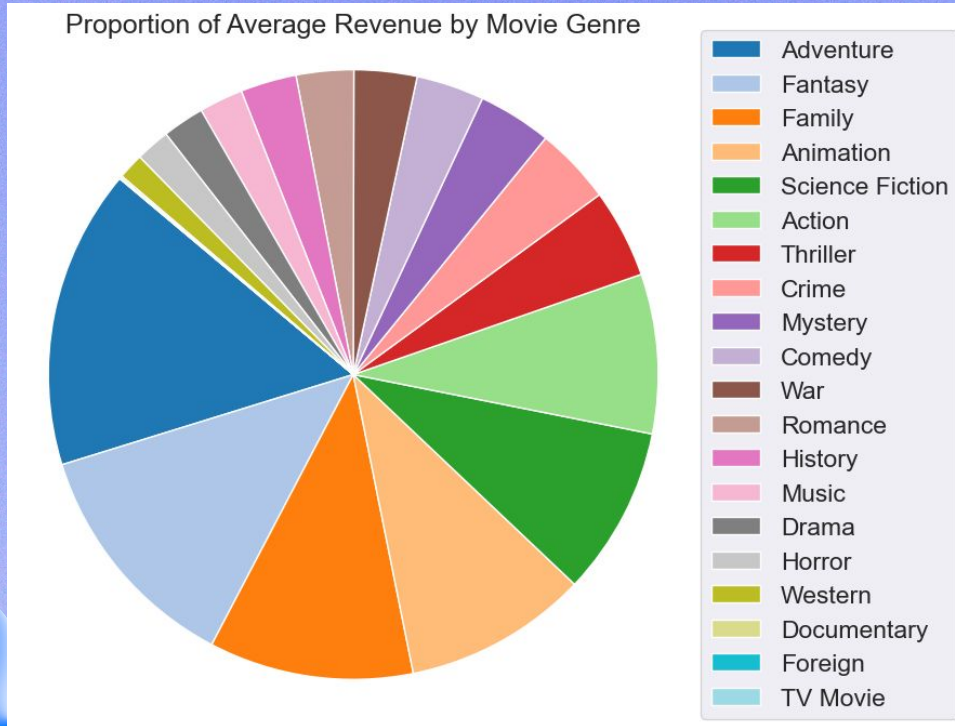


How have audience and critic ratings for specific movie genres (such as action, drama, comedy) evolved over the decades? Are certain genres experiencing an increase or decrease in popularity?

- Distribution for Ratings
- Some genres more dense than others



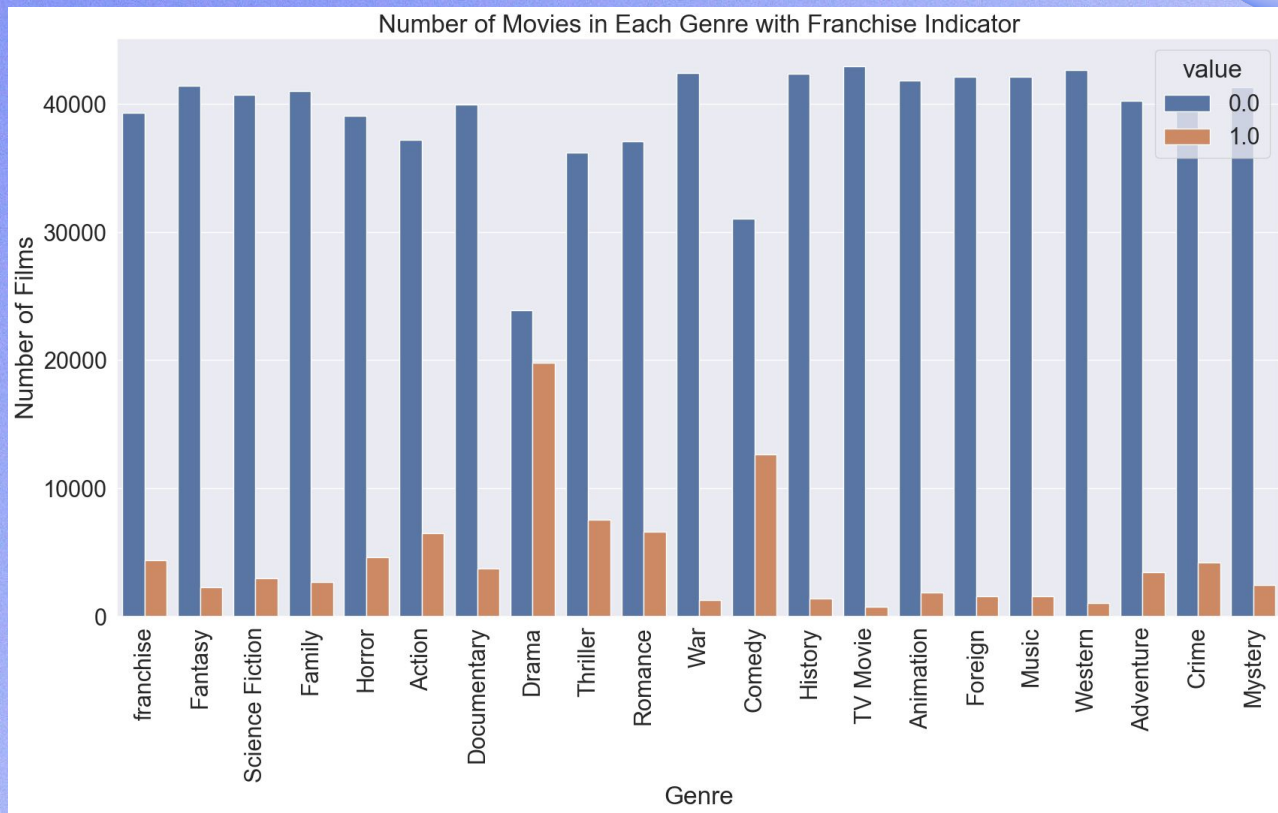
Is there a correlation between movie genres and the revenue generated, and if so, which genres tend to exhibit a stronger correlation with higher revenues?



- Adventure, Fantasy, Family: high revenues
- Documentary, Foreign, TV Movie: low revenues

Is there a correlation between a movie's genre and its likelihood of being part of a franchise?

- Class imbalance between Franchise and Non-franchise movies
- Drama: genre with the least imbalance



Is there a correlation between a movie's genre and its likelihood of being part of a franchise?

- **Franchise:**
 - **Drama, Comedy:** positive correlation
 - **All other genres:** no correlation
- **Non-franchise:**
 - **Drama, Comedy:** positive correlation
 - **All other genres:** no correlation



Conclusion: results from sub-questions

RQ1

Drama and Comedy - most common

RQ2

Western, Animation, Mystery - highest average ratings

RQ3

Adventure, Fantasy - highest total revenue

RQ4

Drama and Comedy - common in franchises

Limitations and Future Research

- Main focus on Western Cinema
 - Streaming platforms