Report: Studio Ghibli + BoxOfficeMojo Integration

Students: Serikbayeva Aruzhan & Saparbay Nursaya
Course: Data Collection & Preparation
Topic: Movie Performance and Ratings Analysis

Project Goal:

The main goal of this project was to combine **two real data sources**, one from an **API** and one from **web scraping**, and analyze how **Studio Ghibli movies** perform in terms of box office and ratings.

We wanted to see:

- Which directors and producers bring the **highest box office results**,
- Correlation Between Key Movie Variables,
- Top 10 Most Efficient Movies (Gross per Minute),
- Number of Studio Ghibli Movies per Decade,
- And if there is any connection between movie ratings and total earnings.

Data Sources:

1. API - Studio Ghibli API
We used the official open API of *Studio Ghibli* (https://ghibliapi.vercel.app/films), which provides structured information about all movies, including title, director, producer, release year, running time, and Rotten Tomatoes score.
This API is perfect for our topic because it contains verified and clean movie data in JSON format.

2. Web Scraping - BoxOfficeMojo
We collected financial data from the *BoxOfficeMojo* website, which lists movie earnings, number of theaters, and opening weekend gross.
Using requests and BeautifulSoup, we extracted the main table with movie titles and box office performance.
This site was chosen because it contains real financial performance of Studio Ghibli films, which allowed us to connect it with the API data.

Data Cleaning and Preparation:

We performed several cleaning steps:

- Standardized movie titles (lowercase, removed special symbols).
- Summed up duplicates in BoxOfficeMojo data(like re-released versions and etc.)
- Merged both datasets using the cleaned movie titles as a key.
- Dropped all missing values and kept only complete rows.

Final dataset columns:
title, director, producer, release_year, running_time (min), rt_score (/100), lifetime_gross ($), opening_gross ($), max_theaters (count)

Descriptive Statistics and Correlation:

We calculated key descriptive statistics (mean, max, std, count and etc.) and correlation values.
The **correlation between RT score and lifetime gross** was **0.11**, showing a **weak positive relationship**.
That means higher ratings slightly relate to better box office results, but the connection is not strong.

Visualizations and Insights:

1. **Box Office Efficiency- Gross per Minute**
   Ponyo ranked **#1** as the most efficient movie, earning the highest amount of money per minute.
   The Tale of the Princess Kaguya ranked **last**, showing that longer or artistic films may earn less per minute but rating is 100.

```
HIGHEST BOX OFFICE
   Movie: Ponyo
   Total Gross: $15,090,399

 LOWEST BOX OFFICE
   Movie: When Marnie Was There
   Total Gross: $561,085


============================================================

 HIGHEST RATED MOVIE
   Movie: The Tale of the Princess Kaguya
   Rating: 100 /100

 LOWEST RATED MOVIE
   Movie: From Up on Poppy Hill
   Rating: 83 /100


============================================================

 LONGEST MOVIE
   Movie: The Tale of the Princess Kaguya
   Duration: 137 min

 SHORTEST MOVIE
   Movie: The Red Turtle
   Duration: 80 min


============================================================

MOVIE SHOWN IN MOST THEATERS
   Movie: Ponyo
   Theaters: 927.0

MOVIE SHOWN IN FEWEST THEATERS
   Movie: The Tale of the Princess Kaguya
```

```
     Theaters: 29.0
```

2. Movies per Decade

- 2010s — 5 movies
- 2000s — 3 movies
- 1990s — 1 movie
  This shows that Studio Ghibli was most active in the 2010s decade.

2. **Top Directors by Average Rating**

Isao Takahata – 100

Michaël Dudok de Wit – 93

Hiromasa Yonebayashi – 92

Hayao Miyazaki – 91.4

Gorō Miyazaki - 83

Isao Takahata received the **highest critical ratings**, followed by *Michaël Dudok de Wit*.

3. **Top Producers by Average Rating**

Yoshiaki Nishimura - 96

Toshio Suzuki – 90

Toshio Suzuki, Isao Takahata, Vincent Maraval -93

Yoshiaki Nishimura is the top producer with the highest-rated films.

5. Rating Distribution
Most Studio Ghibli movies have ratings around **92.5 / 100**, showing that their films consistently receive strong critical acclaim.

Key Results Summary:

```
HIGHEST BOX OFFICE
   Movie: Ponyo
   Total Gross: $15,090,399

 LOWEST BOX OFFICE
   Movie: When Marnie Was There
   Total Gross: $561,085


========================================================

 HIGHEST RATED MOVIE
   Movie: The Tale of the Princess Kaguya
```

```
    Rating: 100 /100

 LOWEST RATED MOVIE
   Movie: From Up on Poppy Hill
   Rating: 83 /100


========================================================

 LONGEST MOVIE
   Movie: The Tale of the Princess Kaguya
   Duration: 137 min

 SHORTEST MOVIE
   Movie: The Red Turtle
   Duration: 80 min


========================================================

 MOVIE SHOWN IN MOST THEATERS
   Movie: Ponyo
   Theaters: 927.0

 MOVIE SHOWN IN FEWEST THEATERS
   Movie: The Tale of the Princess Kaguya
   Theaters: 29.0


========================================================
```

Conclusion:

This project demonstrates how two independent data sources, an **API** and **web scraping**, can be successfully combined for real-world analysis.
Studio Ghibli films show **very high and consistent ratings**, but **box office success** does not always depend on ratings.
Our analysis highlights the power of data collection and cleaning for discovering meaningful cultural and business insights.