Project Report – Davis Hoffstatter

Movie Revenue Trends (2000-2024)

1. Introduction

Box office revenue is often viewed as the most recognizable measure of a film's success, yet many factors such as genre, budget, or international appeal shape how audiences respond. Revenue may lead to higher box office performance, but it is interesting to analyze if smaller films with a strong word of mouth can outperform big budget productions with massive marketing campaigns.

To explore this, I focused on three research questions:

> *1) How has the balance between domestic and foreign box office revenue shifted over time?*
>
> *2) Which years between 2000 and 2024 had the highest average worldwide box office revenue per film?*
>
> *3) How do inflation and ticket price growth affect box office trends when revenues are adjusted?*

Throughout the duration of this project, I plan to use a historical movie dataset from Kaggle: *The Movies Box Office Collection Data 2000- 2024*[1] alongside web scraping market summaries from *The Numbers*[2], a site that tracks box office trends. My objective is to combine individual film data with broader market insights to identify unique patterns that shed light on the status of the movie industry both domestically and internationally.

2. Data

This project uses two sources of data. The first is the Kaggle dataset which includes each film's title, release year, and worldwide, domestic, and foreign box office performance. The second source is *The Numbers*, from which I scraped market data such as average ticket prices and annual ticket sales. Together, these datasets provide a comprehensive foundation for evaluating trends in the movie market and analyzing the factors that influence box office success.

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024
2 https://www.the-numbers.com/market/

2.1 Kaggle Data Set

I located a data set on Kaggle that contains movie data from 2000-2024. The dataset is titled: The Movies Box Office Collection Data 2000-2024. The dataset consists of three separate excel files broken up into years 2000-2009, 2010-2024, and 2024. My first obstacle was combining the three separate excel files into a Jupyter notebook data frame to analyze the data concurrently. Upon completion, the data set concatenated 5,000 separate films and 9 columns of data: title, domestic views, domestic percent, foreign views, foreign percent, year, and rank. I created a data frame that easily interprets this information, so you can see what films have more success in their respected market and the trajectory of the industry concurrently.

2.2 The Numbers

To complement the Kaggle dataset, I scraped market level data from *The Numbers*, a site that tracks box office performance and industry trends. This data includes annual ticket sales, average ticket prices, and overall market revenue, which provides valuable context for understanding the overall landscape of the film industry over the years. The scraping process consisted of collecting information from multiple pages and then consolidating it into a data frame on Jupyter notebook. The data captures year by year changes in consumer and market behavior within the movie industry such as ticket sales, ticket prices, and attendance. After cleaning, I merged the files into a single dataset which offers a clear view of the movie market.

2.3 Combining Box Office and Market Data

My main goal was to merge the two datasets into a single data frame to make the data easier to read and analyze. The process was straightforward; I merged the datasets together with the "year" column as the common key for the merge. The only challenge I faced was a formatting mismatch: in the Kaggle dataset "Year" was capitalized and on *The Numbers*, I scrapped it was lower case "year", I resolved this by standardizing the column names before merging.

I merged the data frame, "df_all" which represents the Kaggle dataset numeric with the data frame "df" which represents the market information I scraped from *The Numbers*. The resulting data frame is called df_merged. I proceeded to save this out as a csv file and made sure to cross check the observations within the union, which remained 5,000.

Table 1 Data Dictionary

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024
2 https://www.the-numbers.com/market/

Column	Type	Source	Description
Unnamed	Integer	Kaggle	Not needed for analysis
Release Group	String	Kaggle	Title of movie (film name).
Worldwide	Numeric (currency)	Kaggle	Total Worldwide box office revenue for the film.
Domestic	Numeric (currency)	Kaggle	US box office revenue for the film
Domestic_percent	Float %	Kaggle	Percentage of worldwide revenue earned domestically.
Foreign	Numeric	Kaggle	International box office revenue
Foreign_percent	Float %	Kaggle	Percentage of worldwide revenue earned internationally.
year	Integer	Both	Release year of the film
Rank	Integer	Kaggle	Film's rank within its release year
tickets_sold	Integer	The Numbers	Total tickets sold across the U.S. market for that year
total_box_office	Numeric	The Numbers	Total box office revenue across all films for that year
inflation_adjusted_box_office	Numeric	The Numbers	Total box office revenue adjusted for inflation
average_ticket_price	Float	The Numbers	Average movie ticket price in the U.S. for that year

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024

2 https://www.the-numbers.com/market/

3. Analysis

3.1 Balance Between Domestic and Foreign Revenue

The first thing I wanted to analyze was the balance between domestic and foreign box office revenues between 2000-2024, and what this reveals about the United States role in box office performance globally. My hypothesis is that with the growing number of streaming platforms and accessibility to films away from the theatre, the domestic market is going to shift negatively. On the flip side, I think entertainment is more valued domestically, so there could not be a shift whatsoever. My first question attempts to address this. To do this, I calculated the average domestic and foreign revenue percentages for each year.

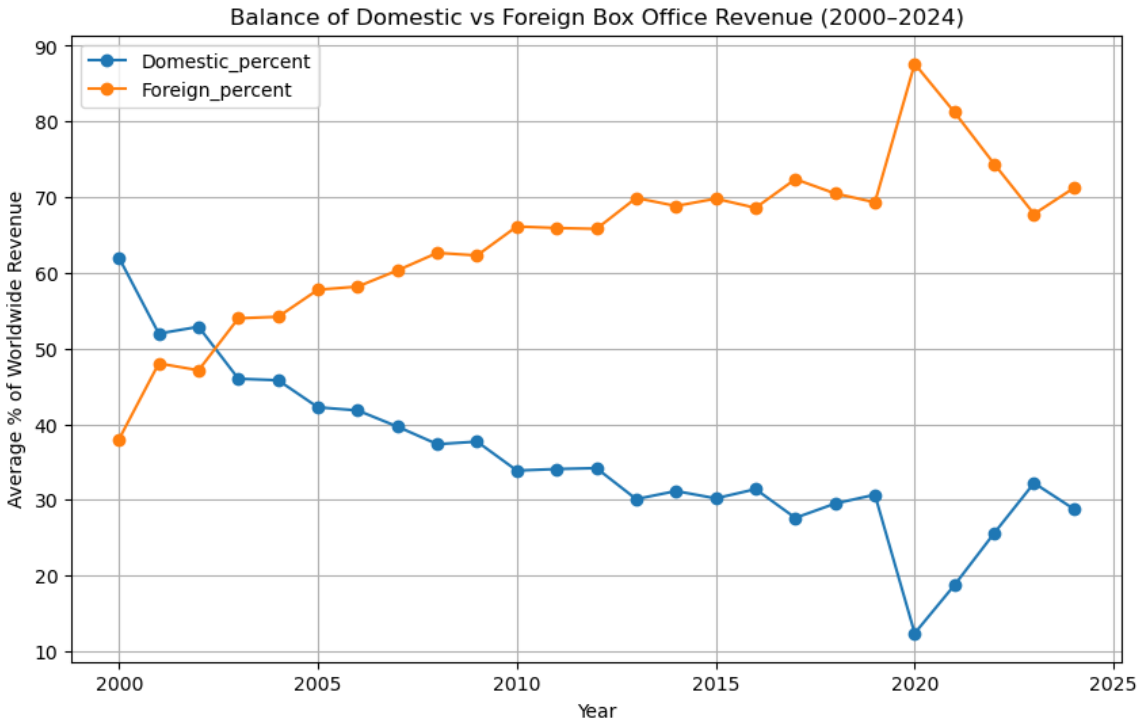

Figure 1 Domestic vs International Market Share

The results show a decline in domestic share of worldwide revenue, with foreign markets increasingly contributing to a larger portion. As my hypothesis states, in the early 2000's we see the domestic share accounting for nearly half of the market and briefly accounting for close to 70% of the share. By 2010 and especially into 2020, foreign percentages consistently exceed domestic shares. A sharp drop occurs around 2020 and is paired with a spike in foreign share, and this reflects the disruption of the COVID-19 pandemic and the

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024

2 https://www.the-numbers.com/market/

resilience of international markets. This trend highlights the growth of global audiences within box office success.

A line plot of domestic versus foreign percentages **(Figure 1)** illustrates this shift clearly. The foreign share rises steadily while domestic is steadily declining. This suggests that the global reach of the movie market has become more critical than its domestic performance alone.

	Release Group	year	Domestic	Worldwide	Domestic_percent
140	Drunken Master II	2000	11555430.0	11555430.0	100.0
4983	2024 Oscar Nominated Short Films: Documentary	2024	3162293.0	3162293.0	100.0
4199	The SpongeBob Movie: Sponge on the Run	2020	4810790.0	4810790.0	100.0
1191	The Gospel	2005	15778152.0	15778152.0	100.0
151	3 Strikes	2000	9832166.0	9832166.0	100.0

Figure 2: Top 5 Domestic Shares

During this analysis, I got curious regarding films with 100% domestic share, meaning they are only performing domestically. **Figure 2** displays the top five films with strict US audiences. These cases highlight films that either have no international distribution or are only targeted with domestic audiences. Although they are not the highest grossing films overall, this emphasizes how certain releases can rely on the domestic market for their success.

3.2 Peak Years for Box Office Revenue

For my second question, I wanted to identify which years between 2000 and 2024 had the highest average worldwide box office revenue per film and what this could suggest about the movie industry over this time. To do this, I calculated the average worldwide revenue per film for each of the years between 2000- 2024 and visualized the results in a bar chart **(Figure 3)**. My hypothesis is that pos pandemic individuals have been hesitant to attend movie theatres and prefer viewing movies at home.

The bar chart clearly shows a steady upward trend in average revenue from 2000 through 2019, and it peaks in 2019. The peak in 2019 could potentially occur due to significant blockbusters and franchise films during this period. As we saw in **Figure 1**, there is a sharp

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024

2 https://www.the-numbers.com/market/

drop in 2020 that represents the COVID-19 pandemic and theatres closing globally. Following this decline, the theatre revenues have not been the same, although there have been slight increases.

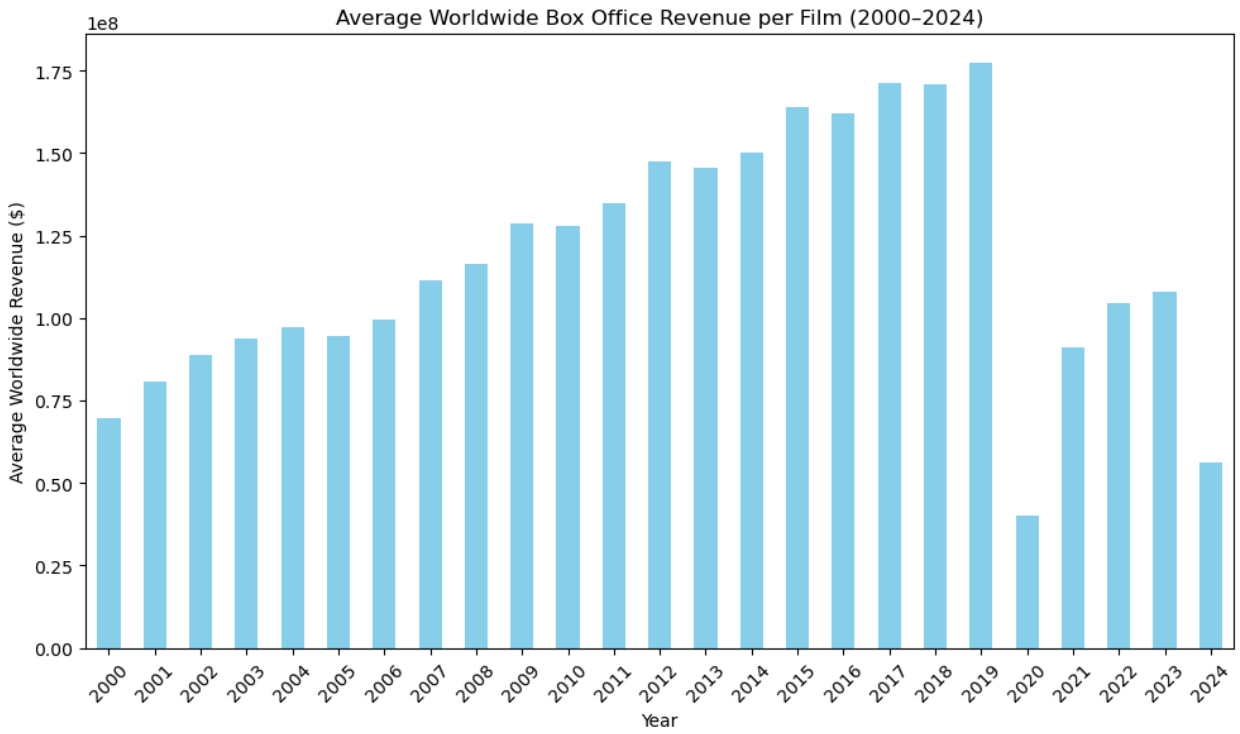

Figure 3: Peak Years for Box Office Revenue

While analyzing this question, my main area of interest was to see how the COVID-19 pandemic affected the industry as a whole. It seems like all the theatres in my hometown were affected by the pandemic and continue to be affected today. Seeing this visually represented by our bar chart in **Figure 3** speaks volume to what really occurred and the result of the pandemic. As you can see in **Figure 3**, the years are placed on the x axis and the average worldwide revenue in billions of dollars is represented on the y axis. It is unfortunate that the industry was affected as much as it was because worldwide revenue was progressively growing on a yearly basis up to this point.

3.3 Inflation Adjusted Revenue Trends

For my final analysis, I wanted to examine how average box office revenue per film on an individual level has evolved over time when adjusting for inflation and ticket price changes.

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024
2 https://www.the-numbers.com/market/

To do this, I calculated the mean adjusted box office revenue per film for each year and compared it to historical ticket price data going back multiple decades.

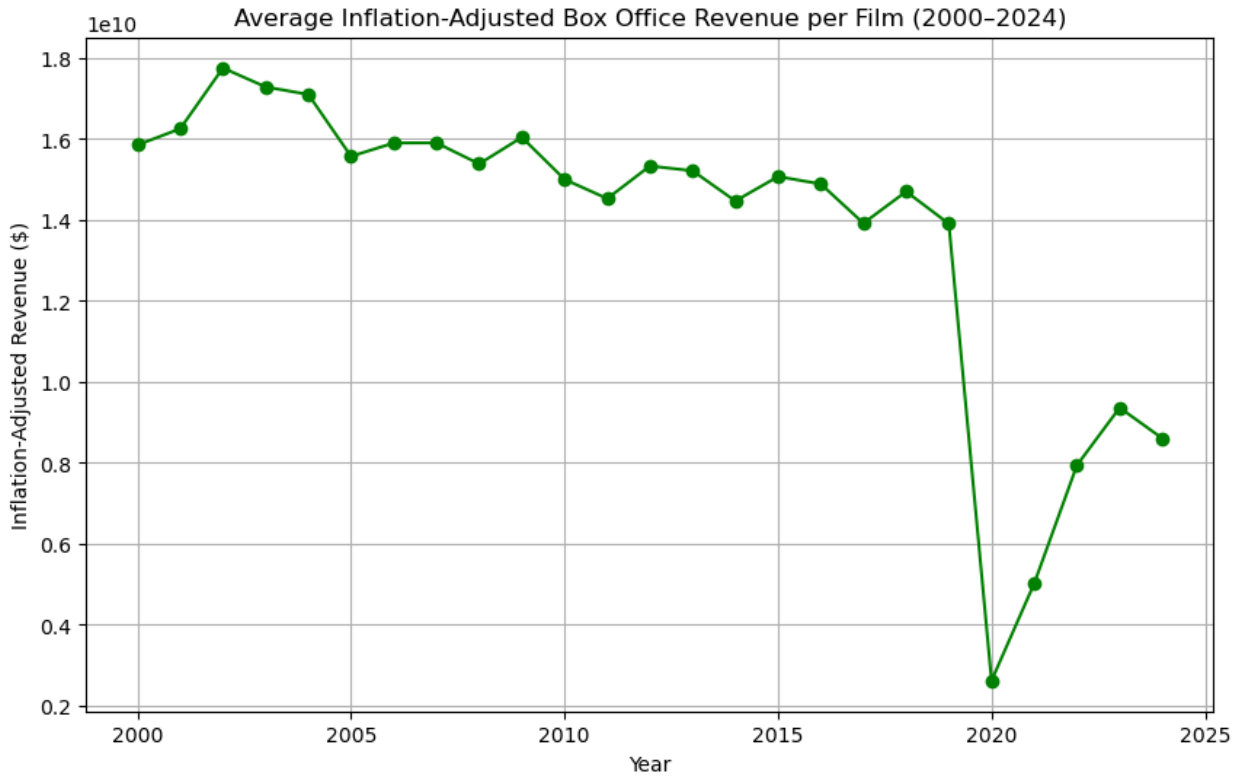

Figure 4: Average Inflation Adjusted Revenue ($ Billions)

Figure 4 shows the inflation adjusted box office revenue per film. This reveals that while overall revenue has risen on a yearly basis, adjusted averages flatten in the later years, especially post 2020. This suggests that rising ticket prices, rather than increased attendance, have been driving the majority of the industry's revenue growth. This leads us into the next portions of this analysis that specifically analyzes the changes in the ticket prices within the industry.

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024
2 https://www.the-numbers.com/market/

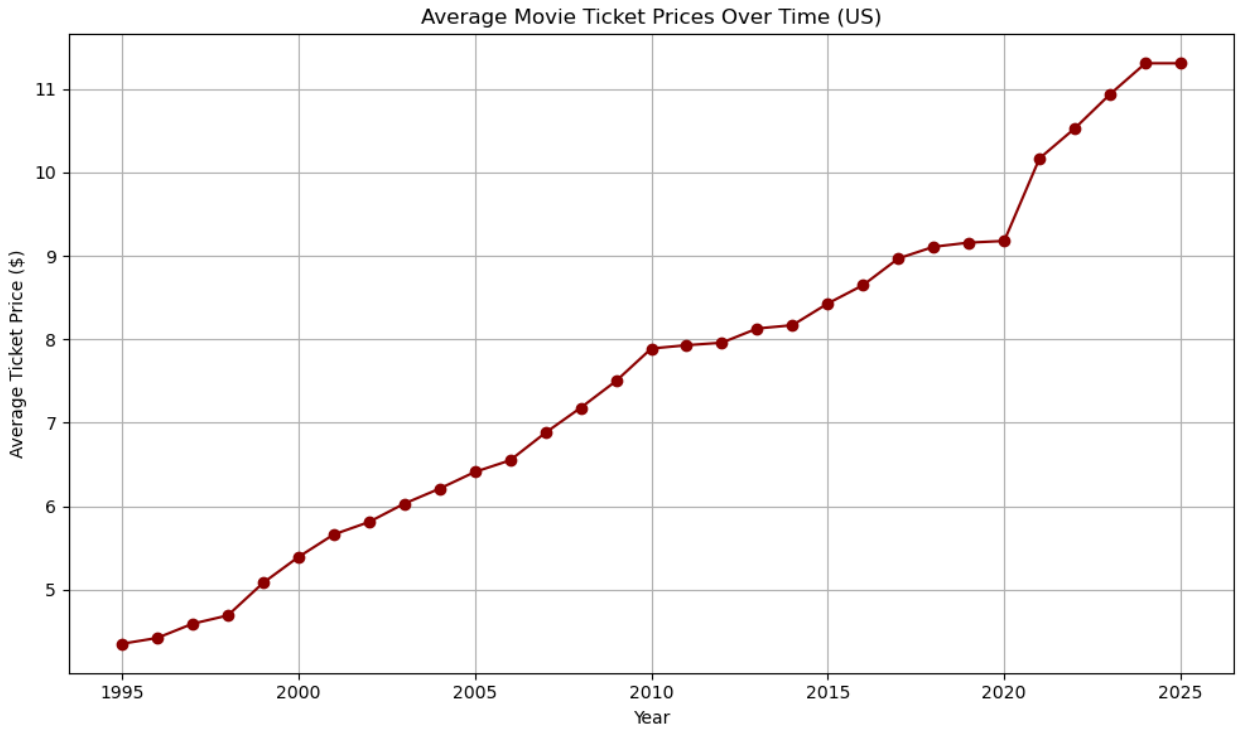

Figure 5: Average Movie Ticket Prices Over Time (USD)

To support the claims above, **Figure 5** displays the average movie ticket price in the US from 1995 to 2025. The graph shows a consistent increase and upward trend with sharper increases around 2020 and 2021. This is probably an effort to recover lost revenues due to the pandemic. This inflation reflects rising costs for consumers and explains why box office totals can appear strong even when the actual audience's turnout is very stagnant. Think about the last time you attended a movie that actually had a good audience turnout.

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024
2 https://www.the-numbers.com/market/

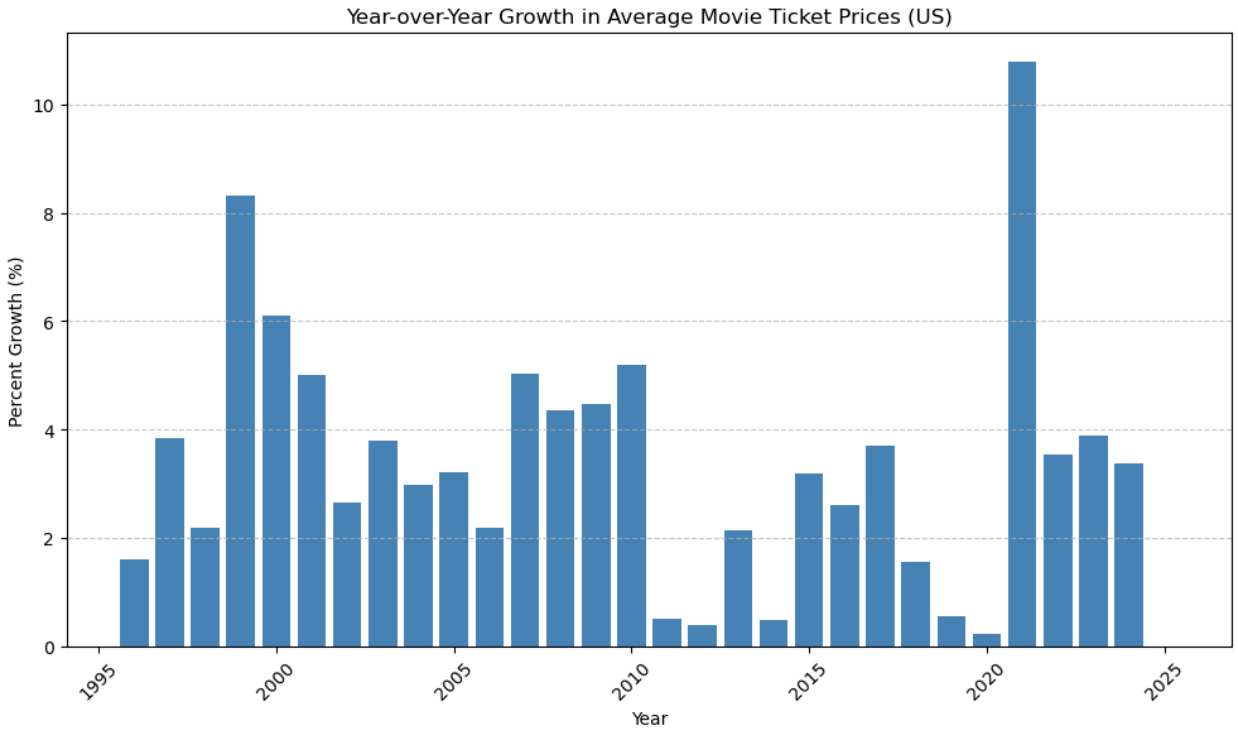

Figure 6: Year to Year Percentage Growth in Ticket Prices

Finally, **Figure 6** displays the year-to-year percent growth in ticket prices. This bar graph highlights key inflation spikes, which we see on display approximately every five years. Keep in mind that the graph is year-to-year, and every year represents its own growth. These surges back the claim that box office revenue growth in recent years is driven more by pricing than attendance volumes. Note how much of a price increase occurred post pandemic- we are facing increased ticket prices and lower attendance.

4. Conclusion

In this project, I analyzed three key aspects of movie box office performance.

1. The balance between domestic and foreign revenue
2. The peak years for worldwide box office revenue per film
3. Inflation adjusted revenue trends with ticket price growth

The main goals were to uncover how external factors such as release timing, geographic market, and ticket prices influence industry outcomes. Based on the analysis questions outlines in my proposal, I found the following results:

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024
2 https://www.the-numbers.com/market/

1. *Balance Between Domestic and Foreign Revenue*

Question: Has the balance between domestic and foreign box office revenue shifted over time?

Findings: Foreign markets are showing steady growth and have overtaken domestic markets in most years since around 2010. The looser regulations due to the pandemic caused a sharp shift in this dynamic. **Figure 1** shows the shift clearly with international audiences drawing much more box office success.

2. *Peak Years for Box Office Revenue*

Question: What years saw the greatest increase in box office revenue?

Findings: The average worldwide revenue per film rose consistently throughout the 2000s, peaking in 2019 before collapsing in 2020 due to the pandemic. **Figure 2** displays this distribution in bar chart form and shows some slight recovery post pandemic but is still below pre- pandemic levels.

3. *Inflation Adjusted Revenue Trends*

Question: How has the average box office revenue per film evolved over time when adjusted for inflation and ticket price changes, and what does this say about the industry's growth?

Findings: The Nominal box office totals suggest growth, but the inflation adjusted figures flatten after 2015 and dip post 2020. **Figures 3 through 5** reveal that rising ticket prices, not increased attendance, explain the majority of the industry's recent revenue bump.

This project faces several limitations, the first being inconsistent release date formatting, the next being incomplete data on streaming competition and international distribution. The domestic outlay is projected well, but the international trends are inconsistent. We do not necessarily have all the data to showcase international streaming trends.

As far as suggestions go, we clearly see a dip in revenue and box office attendance post pandemic. My suggestions would entail approaching the situation with a long-term mindset to get the numbers back to pre-pandemic levels. This can be done by offering incentives to get to the theatres such as promotions or marketing campaigns to promote

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024
2 https://www.the-numbers.com/market/

movies. I will also suggest lowering the average time per film to make films easier to watch for a busy person who does not have time to watch multiple hour-long films.

1 https://www.kaggle.com/datasets/parthdande/movies-box-office-collection-data-2000-2024
2 https://www.the-numbers.com/market/