# CineSeq: A Multivariate Seq2Seq Framework for Forecasting Movie Box-Office Trajectories

Saisrijith Reddy Maramreddy

December 2025

## 1 Background on Multivariate Time-Series Forecasting for Movie Performance

Forecasting movie box-office performance is a challenging multivariate time-series problem. A film's revenue trajectory reflects the combined influence of marketing signals, trailer engagement, seasonal factors, critical reception, and audience sentiment. Revenue alone cannot explain a film's trajectory; modern forecasting must capture how trailers, marketing, audience interest, and weekly momentum shift over time.

To support this study, a comprehensive dataset was constructed by programmatically collecting weekly domestic box-office revenue from *The Numbers* using browser automation, along with trailer metadata and engagement signals extracted via `yt-dlp`. For each of the 30 selected films—intentionally chosen to provide a balanced mix of blockbusters, moderate hits, and commercial flops—weekly revenues, YouTube trailer IDs, view-based engagement metrics, and release-year metadata were merged into a unified panel suitable for sequence modeling. This curated selection ensures that the dataset captures a broad variety of temporal behaviors, including early spikes driven by pre-release hype, mid-run decay patterns, and occasional re-acceleration during holiday periods.

These temporal dynamics differ across performance tiers, requiring a model that can represent both short-term weekly fluctuations and longer-range structural trends. This variability provides a meaningful benchmark for evaluating sequence-to-sequence architectures in multi-week revenue forecasting.

### 1.1 Why Sequence-to-Sequence Architectures Are Required

Movie revenue time series are irregular, short (10–20 weeks), and highly volatile. Traditional recurrent models such as standalone LSTMs or GRUs compress an entire sequence into a single hidden state, which often struggles to retain rich information across time steps. This motivates the use of an encoder–decoder sequence-to-sequence (Seq2Seq) architecture.

In this project, a hybrid model is employed: an LSTM encoder to capture long-term temporal dependencies in weekly box-office patterns, and a GRU decoder to generate multi-step forecasts autoregressively. The encoder learns a high-level temporal representation of the observed revenue sequence, while the decoder conditions on this representation to predict future weeks. This structure is well suited to situations where input and output sequences differ in length, or when the model must extrapolate trends beyond the observed data.

To further improve information flow between the encoder and decoder, an attention mechanism is incorporated, enabling the decoder to dynamically emphasize different portions of the historical revenue sequence when generating each future prediction step. This reduces reliance on a single fixed-length summary and helps mitigate error accumulation during longer-horizon autoregressive forecasting.

Because box-office trajectories differ substantially between blockbusters, moderate hits, and commercial failures, the Seq2Seq model must also learn cross-movie temporal signatures—such as steep early peaks, slow multi-week decay curves, and unusual mid-run performance shifts. This makes architectural design, sequence length handling, and temporal feature engineering central to model success.

## 2 Data Description and Objective

The forecasting dataset contains weekly domestic box-office revenue for 30 films released between 2010 and 2023. For each movie, the full revenue trajectory was extracted using a structured pipeline:

- **YouTube Trailer Data:** Trailer URLs and video IDs were retrieved using a year-aware `yt-dlp` search strategy that matches trailers to release years.

- **Box-Office Revenue:** Weekly domestic gross values were scraped from *The Numbers* using a Playwright-based asynchronous browser automation workflow, coupled with a robust header-aware HTML table extractor.
- **Metadata:** Each movie is annotated with release year and category (blockbuster, hit, flop), enabling cross-group comparison.

The final merged dataset forms a multivariate time series for each movie, where the target variable is weekly revenue and the covariates include both trailer engagement indicators and metadata. The panel exhibits strong temporal variability and non-stationarity, motivating the use of neural sequence models rather than classical econometric techniques.

### 2.1 Objective

The primary goal of this project is to assess whether a hybrid LSTM–GRU sequence-to-sequence (Seq2Seq) architecture can effectively forecast future box-office revenue over multi-week horizons. Specifically, the study aims to examine:

- the ability of a Seq2Seq framework to model short, highly volatile movie revenue trajectories,
- how encoder–decoder design choices, including attention-based conditioning, affect forecasting stability and longer-horizon predictions,
- and whether temporal patterns learned from movies with varying performance profiles (blockbusters, hits, and flops) generalize across categories.

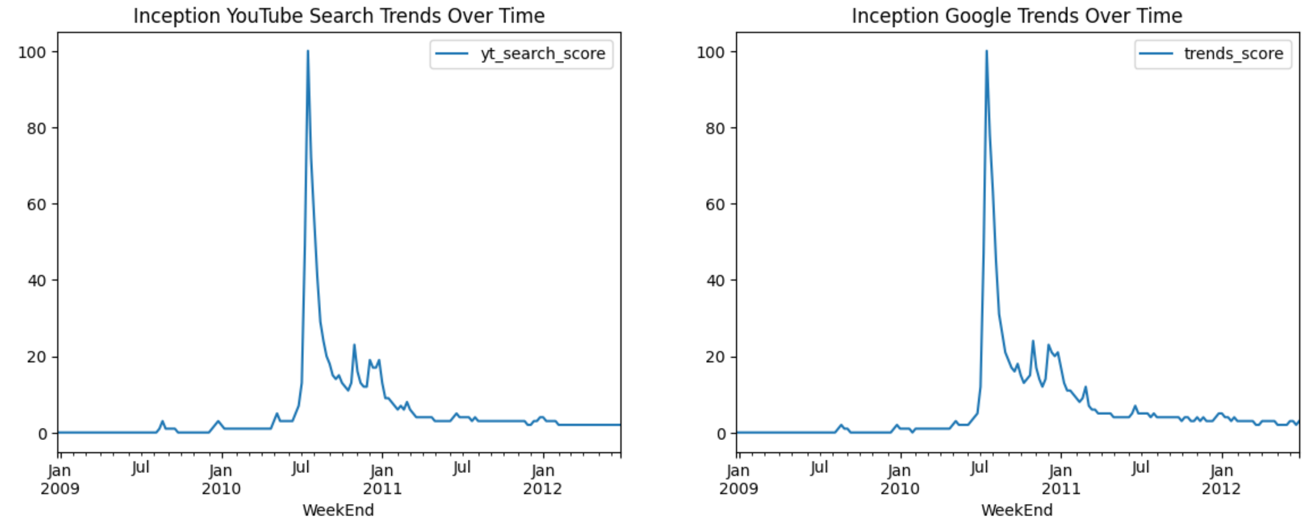
By integrating multi-source real-world signals and leveraging modern sequence modeling techniques, this system seeks to produce reliable box-office revenue forecasts that may support marketing analytics, financial planning, and post-release performance assessment.

## 3 Feature Engineering

Accurate multi-week box-office forecasting requires modeling not only past revenue levels, but also external signals that capture audience awareness, engagement, sentiment, and structural movie characteristics. To support the Seq2Seq forecasting framework, a comprehensive set of temporally aligned features was engineered for each movie-week observation. All features are constructed using information available at or before the corresponding time step to avoid data leakage. For clarity, figures and tables in this section present representative examples; the full feature set is constructed for all 30 movies in the dataset.

### 3.1 Search Interest and Public Attention Signals

Public interest and audience awareness were modeled using weekly Google Trends and YouTube search intensity scores for each movie title. These signals provide a high-frequency proxy for collective attention, capturing pre-release anticipation and post-release discussion dynamics that are not directly observable from revenue alone. Google Trends reflects normalized search interest across the general web, while YouTube search trends capture platform-specific intent related to trailer discovery and video-based engagement. As illustrated in Figure 1, both series exhibit sharp pre-release spikes followed by rapid decay and a long tail of residual attention, reflecting the short but intense hype cycles typical of theatrical releases.




(a) Google search interest over time

(b) YouTube search interest over time

Figure 1: Weekly Google Trends and YouTube search interest for *Inception*. Both series exhibit a sharp pre-release spike followed by rapid decay and a long tail of residual attention, illustrating short-lived but intense hype dynamics around theatrical release.

### 3.2 YouTube Engagement Signals

Trailer engagement data were collected using the YouTube Data API by querying the official trailer video associated with each movie. For each YouTube video ID, cumulative platform statistics were retrieved, including total view count, like count, and comment count at the time of data collection. These raw engagement metrics provide a direct measure of audience interaction with the trailer and serve as the base inputs for subsequent feature transformations used in the forecasting model.

Table 1: YouTube trailer engagement statistics collected via the YouTube Data API. Each row corresponds to the official trailer associated with a movie.

| YouTube ID | Views | Likes | Comments |
|---|---:|---:|---:|
| TcMBFSGVi1c | 164,915,307 | 3,480,522 | 260,769 |
| d9MyW72ELq0 | 61,338,645 | 1,043,414 | 42,361 |
| qSqVVswa420 | 42,806,364 | 402,735 | 36,889 |
| Kb7jYOYXiVc | 8,731,792 | 67,467 | 3,726 |
| bK6ldnjE3Y0 | 48,470,735 | 430,593 | 16,436 |

### 3.3 Sentiment-Based Signals

Audience sentiment was extracted from YouTube trailer comments using a lexicon-based sentiment analyzer, which assigns sentiment scores by matching words in a text to a predefined dictionary of sentiment-annotated terms and applying rule-based adjustments for factors such as negation and intensity. For each movie, a batch scraping pipeline was used to collect a fixed sample of recent user comments from the official trailer page. Sentiment scores were computed using the VADER sentiment model, which produces a polarity score for each comment based on these lexical and rule-based cues.

From the scraped comments, summary statistics were derived at the movie level, including the average compound sentiment score, the proportions of positive, negative, and neutral comments, and the total number of comments analyzed. These aggregated measures provide a compact representation of overall audience tone at the time of data collection and serve as static sentiment descriptors that complement search and engagement signals.

Table 2: Aggregated sentiment statistics derived from YouTube trailer comments using a VADER-based sentiment analyzer.

| Title | YouTube ID | Avg. Sentiment | Pos. (%) | Neg. (%) | Neu. (%) | Comments |
|---|---|---|---|---|---|---|
| Avengers: Endgame | TcMBFSGVi1c | 0.132 | 44.6 | 24.6 | 30.8 | 260 |
| Avatar: The Way of Water | d9MyW72ELq0 | 0.413 | 68.8 | 12.1 | 19.1 | 240 |
| Top Gun: Maverick | qSqVVswa420 | 0.173 | 49.2 | 25.4 | 25.4 | 260 |
| Barbie | Kb7jYOYXiVc | 0.216 | 53.8 | 18.1 | 28.1 | 320 |
| Oppenheimer | bK6ldnjE3Y0 | 0.384 | 69.2 | 16.2 | 14.6 | 260 |

### 3.4 Structural Metadata and Cross-Movie Context

To provide cross-movie contextual information that is not directly observable from early revenue trajectories, structural metadata was collected from The Movie Database (TMDB) and incorporated as static features repeated across weekly observations. These features capture production characteristics and audience evaluations that serve as informative priors for the forecasting model.

TMDB user scores were scraped directly from the TMDB website using an automated browser-based pipeline. For each movie, the corresponding title and release year were used to identify the correct TMDB entry, from which the aggregate user score (expressed as a percentage) was extracted. This score reflects early audience reception and provides a coarse, standardized measure of perceived quality across films. Table 3 summarizes example user scores for selected movies in the dataset.

Table 3: TMDB user scores scraped from the official TMDB movie pages. Scores represent aggregated audience ratings at the time of data collection.

| Title | Year | TMDB User Score |
|---|---|---|
| Avengers: Endgame | 2019 | 82.0 |
| Avatar: The Way of Water | 2022 | 76.0 |
| Top Gun: Maverick | 2022 | 82.0 |
| Barbie | 2023 | 69.0 |
| Oppenheimer | 2023 | 80.0 |

In addition to user scores, production metadata was retrieved via the TMDB API, including reported production budget, total revenue, runtime, original language, studio affiliation, and genre labels. Production budgets were log-transformed to reduce scale skewness, while genres were encoded as multi-label indicators to accommodate films spanning multiple categories. Studio affiliation was retained as a categorical feature and later mapped to an ordinal brand-strength score reflecting historical market prominence. Table 4 presents a sample of the extracted production metadata.

Table 4: Sample of production metadata extracted from TMDB, including budget, studio affiliation, genre composition, runtime, and language.

| Title | Year | Budget | Revenue | Studio | Genre(s) | Runtime | Lang. |
|---|---|---|---|---|---|---|---|
| Avengers: Endgame | 2019 | 356M | 2799M | Marvel Studios | Adventure, Sci-Fi, Action | 181 | en |
| Avatar: The Way of Water | 2022 | 350M | 2330M | 20th Century Studios | Action, Adventure, Sci-Fi | 192 | en |
| Top Gun: Maverick | 2022 | 170M | 1489M | Skydance Media | Action, Drama | 131 | en |
| Barbie | 2023 | 145M | 1447M | LuckyChap Entertainment | Comedy, Adventure | 114 | en |
| Oppenheimer | 2023 | 100M | 952M | Syncopy | Drama, History | 181 | en |

All TMDB-derived features are static at the movie level and are replicated across time steps to provide consistent contextual conditioning throughout the input sequence. By supplying information about production scale, studio backing, genre composition, and audience evaluation, these structural signals help the model differentiate between movies with similar short-term revenue patterns but fundamentally different market expectations.

### 3.5 Feature Alignment and Integrity

All engineered features were aligned to a weekly box-office panel and merged at the movie–week level using release year, title, and calendar date keys. External signals—including search interest, trailer engagement, sentiment summaries, and TMDB metadata—were joined to the base revenue panel and sorted chronologically within each movie to form consistent input sequences.

Weekly search trend series were date-shifted to match theatrical reporting weeks, and missing values for attention signals were conservatively filled with zeros to reflect the absence of observed public activity rather than unavailable future information. Movie-level attributes such as engagement statistics, sentiment aggregates, and structural metadata were replicated across time steps to provide stable contextual conditioning throughout each sequence. The resulting unified feature matrix constitutes a leakage-safe, multivariate time series representation used as input to the Seq2Seq forecasting model.

### 3.6 Derived Temporal and Momentum Features

Building on the aligned feature panel, additional variables were engineered to capture temporal position, short-term momentum, and audience response dynamics. Timeline features encode weeks relative to trailer and theatrical release, along with calendar-based seasonality indicators such as month, week-of-year, and holiday and summer release flags.

To model recent revenue dependence, autoregressive signals including one- and two-week revenue lags, week-over-week growth, and cumulative gross were incorporated. External attention and engagement signals were further summarized using rolling averages and momentum features to capture deviations from local baselines. Finally, sentiment intensity measures were included to reflect the strength and polarization of audience reactions. Together, these features provide a compact multivariate representation of timing, momentum, and public response used as input to the Seq2Seq forecasting model.

## 4 Trailer-Based Representation Learning

While the previous section focused on structured and manually engineered signals, this section introduces learned representations extracted directly from movie trailers. Trailers contain rich multimodal information—facial expressions, emotional pacing, and audio intensity—that is difficult to capture using tabular features alone. To leverage this information, a two-stage representation learning pipeline was developed that combines transfer learning and knowledge distillation to convert raw trailer videos into compact, interpretable emotional descriptors.

**Frame-Level Emotion Extraction (Transfer Learning Initialization).** Trailer videos were sampled at regular intervals and processed using a pretrained SCRFD face detector to identify on-screen faces. When multiple faces were present, the most centrally located face was selected to approximate the primary narrative focus. Detected faces were converted to grayscale, resized, and passed through a pretrained FER+ emotion recognition model, which outputs probabilities over eight basic emotion categories (neutral, happiness, surprise, sadness, anger, disgust, fear, and contempt).
At this stage, the pretrained FER+ model is used directly without modification, serving as a transfer learning initialization that injects general-purpose facial emotion knowledge learned from large-scale datasets. While this provides a strong semantic prior, the model exhibits a known tendency toward neutral-dominant predictions when applied to unconstrained cinematic footage, reflecting a domain mismatch between controlled facial datasets and movie trailers.

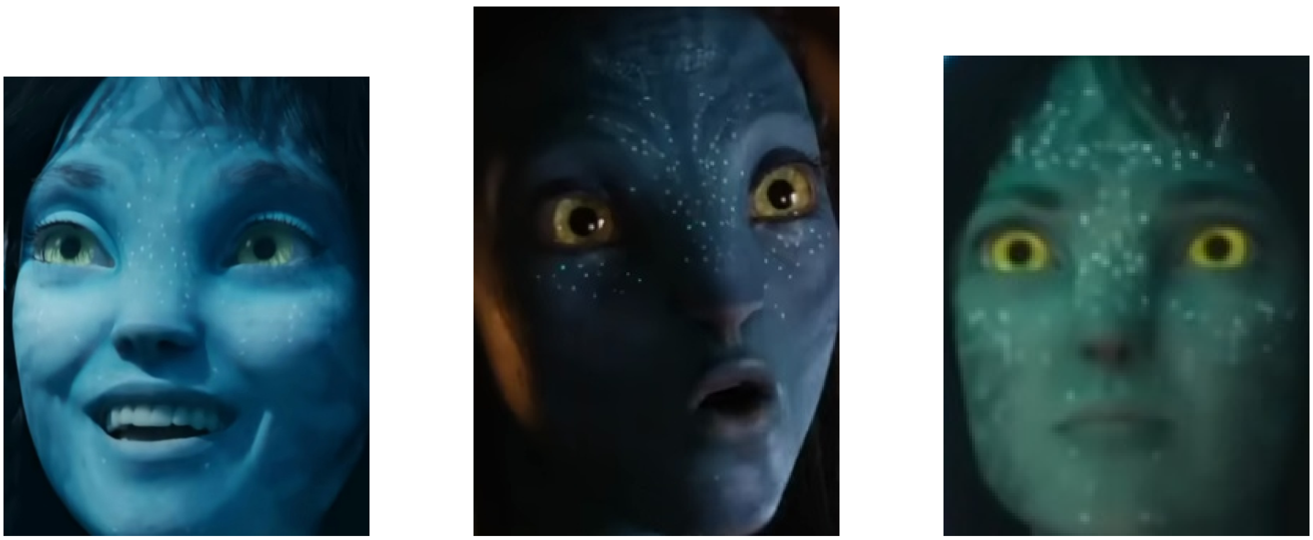


Figure 2: Example trailer frames processed by the pretrained FER+ model. The model assigns dominant emotion categories such as happiness, surprise, and neutral based on facial expressions, eye openness, and gaze direction. These examples illustrate the pretrained model's ability to detect prominent facial cues in cinematic close-ups.

**Domain Adaptation via Knowledge Distillation.** To address the neutral bias and adapt emotion recognition to the trailer domain, a lightweight student emotion network was trained using knowledge distillation. Specifically, facial crops extracted from trailer frames were automatically labeled using FER+ predictions, and these pseudo-labels were used to supervise the training of a compact convolutional neural network.
In this teacher–student setup, FER+ acts as the teacher model, transferring its learned emotion structure to a student model that is optimized directly on trailer-specific facial data. This approach enables domain adaptation without requiring manually annotated emotion labels, while preserving the semantic structure learned by the pretrained model. The distilled student model is subsequently used in place of FER+ during inference to generate more trailer-adapted emotion predictions.
**Temporal Emotion Aggregation and Affective Signals.** Frame-level emotion predictions produced by the distilled model were aggregated over time to construct continuous emotional trajectories for each trailer. From these trajectories, higher-level affective signals were derived in the form of arousal and valence. Arousal captures overall emotional intensity, while valence reflects the balance between positive and negative affect, enabling the representation to encode not only which emotions are present, but how emotional tone and intensity evolve throughout the trailer.
**Audio–Visual Alignment.** In addition to visual emotion cues, audio energy was extracted from each trailer's soundtrack using root-mean-square (RMS) amplitude. The correlation between audio intensity and visual emotional arousal was computed to capture synchronization between musical dynamics and on-screen emotion, a common stylistic mechanism in cinematic trailers.
**Trailer-Level Emotional Embeddings.** Finally, compact trailer-level embeddings were constructed by summarizing the emotional and audiovisual trajectories extracted from each trailer. These embeddings include measures of emotional volatility, peak intensity density, and emotional arc slope, along with mean emotion probabilities

and aggregate affective statistics. Together, these features provide a low-dimensional representation of a trailer's emotional structure, intensity, and pacing.
All trailer-based features are static at the movie level and are replicated across weekly observations to complement the structured temporal, engagement, and sentiment signals described earlier. By encoding narrative and stylistic information that is not observable from early revenue patterns alone, these embeddings provide additional contextual grounding for the downstream forecasting model.
The overall pipeline explicitly separates transfer learning for semantic initialization from knowledge distillation for domain adaptation, and is designed to support future improvements through additional trailer-specific data, stronger student architectures, or supervised emotion labels.

RAW TRAILER VIDEO (.mp4)

↓

Frame Sampling (1.5s interval)

↓

Face Detection (SCRFD)
Central face selection

↓

Emotion Inference (FER+ Teacher Model)

↓

Auto-labeled face crops
(FER+ pseudo-labels)

↓

Knowledge Distillation
Lightweight student CNN

↓

Fine-tuned trailer emotion model

↓

Trailer-level emotional embeddings
(arousal, valence, volatility, audio alignment)

↓

Seq2Seq forecasting model

Figure 3: Trailer-based emotion representation pipeline illustrating transfer learning via a pretrained FER+ teacher model and knowledge distillation for trailer-domain adaptation.

## 4.1 Effect of Knowledge Distillation on Trailer Emotion Representations

To assess the impact of domain adaptation via knowledge distillation, we compare trailer-level emotional features extracted using the pretrained FER+ model (*pre-distillation*) against those produced by the distilled student model (*post-distillation*). Tables 5 and 6 report representative emotion statistics for a subset of well-known films spanning blockbusters, hits, and commercial failures.

Table 5: Representative trailer emotion features extracted using the pretrained FER+ model (pre-distillation).

| **Movie** | **Arousal Volatility** | **Arc Slope** | **Mean Arousal** | **Mean Valence** |
|---|---|---|---|---|
| Avatar: The Way of Water | 0.327 | −0.0034 | 0.220 | 0.001 |
| Top Gun: Maverick | 0.279 | −0.0026 | 0.237 | 0.102 |
| Get Out | 0.352 | −0.0009 | 0.283 | 0.131 |
| Morbius | 0.218 | 0.0022 | 0.096 | −0.086 |
| Cats | 0.299 | 0.0001 | 0.198 | 0.063 |

Table 6: Representative trailer emotion features extracted using the distilled student model (post-distillation).

| **Movie** | **Arousal Volatility** | **Arc Slope** | **Mean Arousal** | **Mean Valence** |
|---|---|---|---|---|
| Avatar: The Way of Water | 0.031 | −0.0001 | 0.130 | 0.054 |
| Top Gun: Maverick | 0.021 | −0.0000 | 0.121 | 0.057 |
| Get Out | 0.034 | −0.0000 | 0.118 | 0.054 |
| Morbius | 0.028 | 0.0000 | 0.168 | 0.050 |
| Cats | 0.032 | 0.0000 | 0.089 | 0.051 |

Across all representative films, knowledge distillation substantially reduces arousal volatility and dampens spurious emotional arc trends, indicating a significant reduction in frame-level noise and neutral-dominance bias. While absolute magnitudes decrease, relative ordering across films is preserved, suggesting that the student model refines emotional intensity without distorting inter-movie structure. These distilled emotion embeddings therefore provide more stable and interpretable trailer-level signals for downstream forecasting.

# 5 Revenue Dynamics Diagnostics

Weekly box office revenue typically follows a rapid post-release decline driven by audience saturation and theatrical turnover. Visual inspection of the series suggests an exponential decay pattern; however, formal diagnostics were conducted to verify that no additional temporal structure remains that would warrant more complex time-series models.

**Weekly Revenue Trajectories.** Figure 4 presents weekly box office revenues for three representative films selected from the full dataset of 30 movies: one blockbuster (*Avatar: The Way of Water*), one commercially successful franchise release (*John Wick*), and one commercial flop (*Morbius*). This stratified selection is intended to capture the dominant revenue regimes observed across the dataset. Successful releases exhibit smooth exponential decay, while poorly performing films collapse sharply within the first few weeks. These contrasting patterns motivate log-transformation and decay-based modeling.

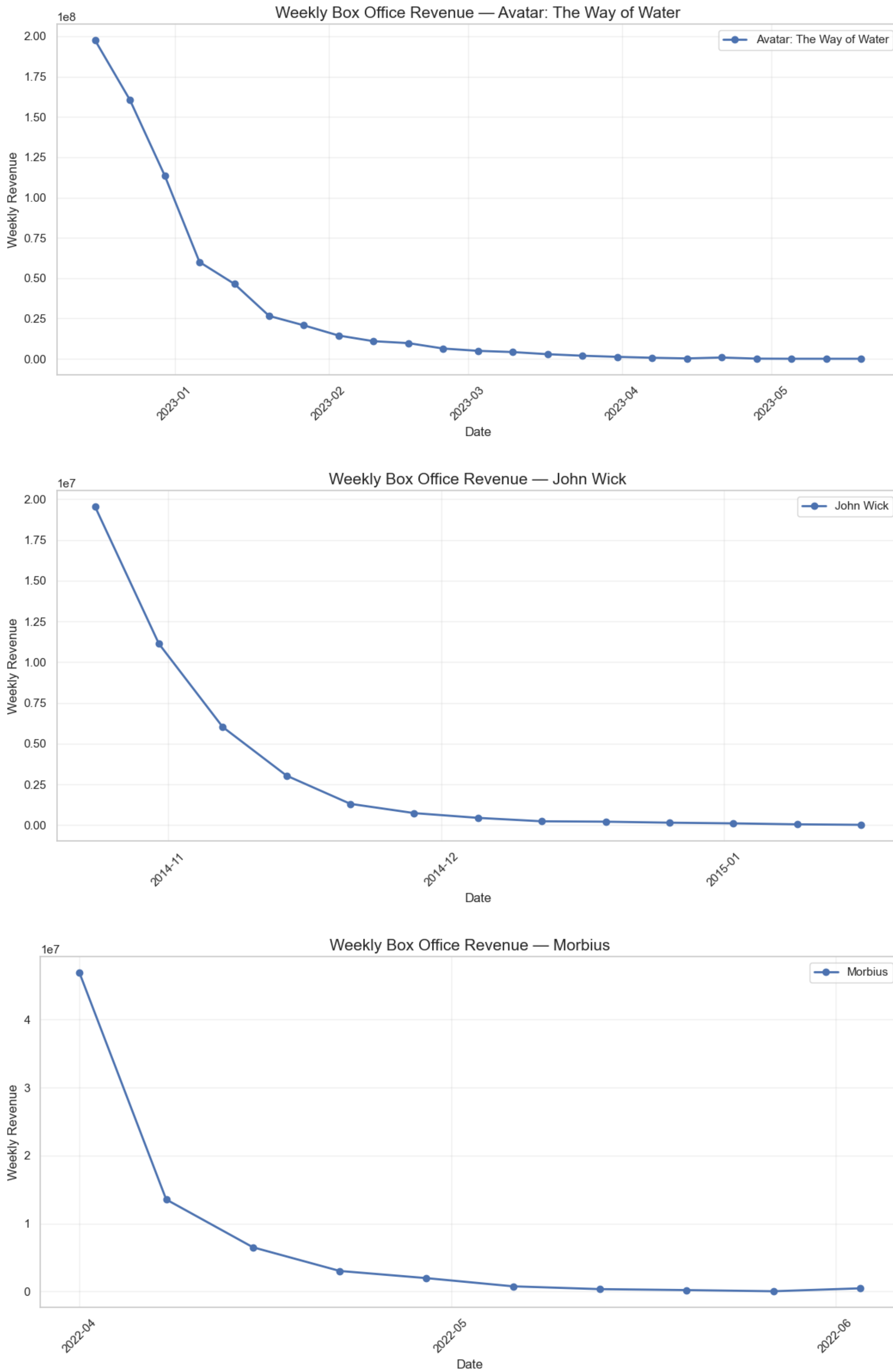



Figure 4: Weekly box office revenue trajectories illustrating post-release decay across a blockbuster (*Avatar: The Way of Water*), a franchise hit (*John Wick*), and a commercial flop (*Morbius*).

**Stationarity Assessment.** To assess stationarity, weekly revenues were log-transformed and evaluated using Augmented Dickey–Fuller (ADF) tests. For *Avatar* and *John Wick*, the log-level series remained non-stationary, while first differencing produced stationarity. In contrast, *Morbius* was stationary at the log level, reflecting its abrupt early collapse and lack of sustained decay dynamics.

**Autocorrelation Diagnostics.** Figure 5 shows the autocorrelation (ACF) and partial autocorrelation (PACF) functions for each movie, with ACF and PACF displayed jointly within a single diagnostic image per film. After applying the appropriate differencing level, no persistent autocorrelation remains beyond the first few lags. This indicates that higher-order autoregressive or moving-average components are unnecessary.

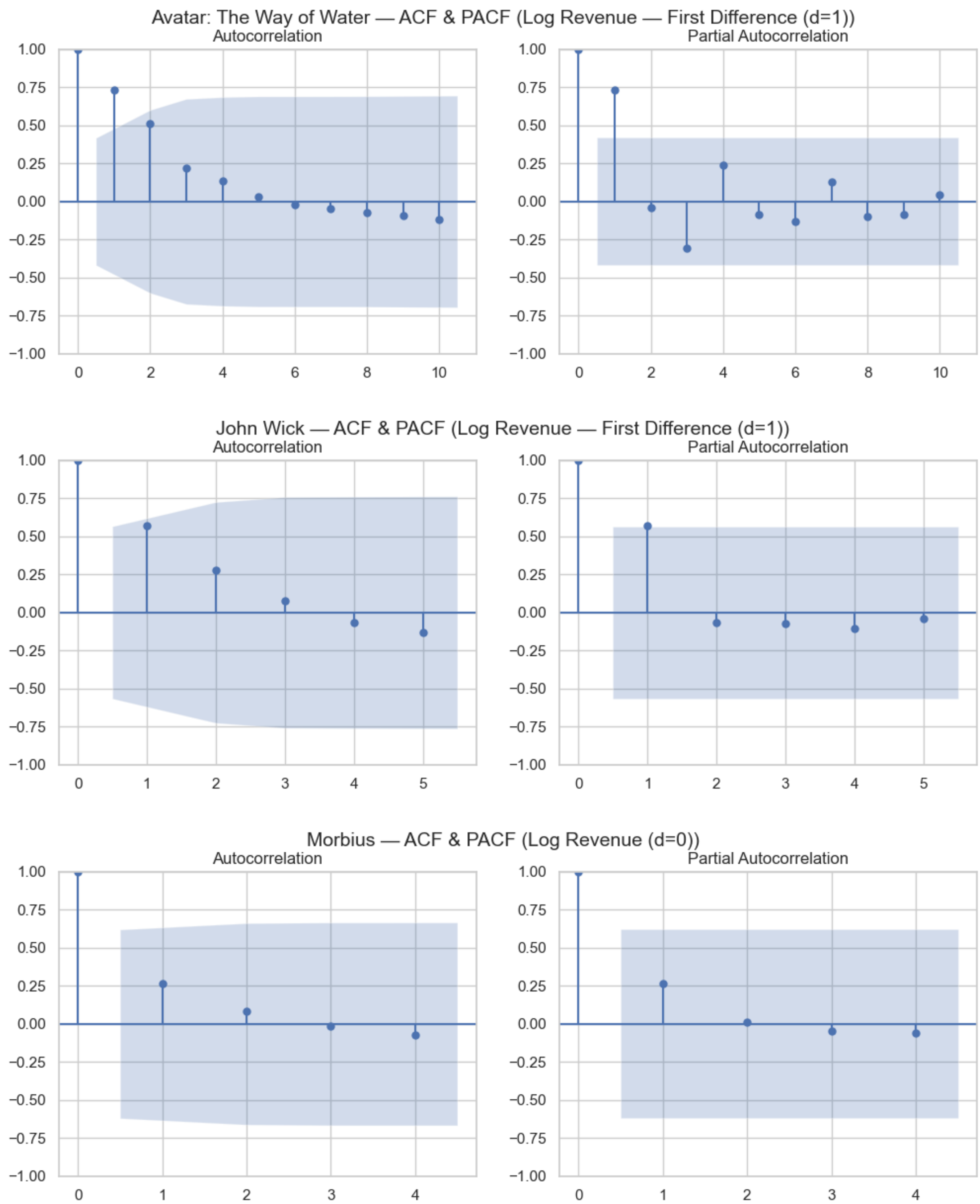



Figure 5: ACF and PACF diagnostics (combined per movie) after appropriate log transformation and differencing. The absence of sustained autocorrelation beyond early lags supports parsimonious exponential decay modeling.

**Diagnostic Conclusion.** These diagnostics confirm that weekly box office revenues are well-characterized by exponential decay with minimal residual temporal dependence. The objective of this analysis is not to search for increasingly complex models, but to formally verify that simpler decay dynamics are sufficient. Consequently, the forecasting framework emphasizes decay-aware modeling augmented with external covariates rather than high-order ARIMA structures.

## 5.1 Exponential Decay Validation

The stationarity and autocorrelation diagnostics indicate that weekly box office revenues are dominated by a smooth post-release decline with minimal residual temporal dependence. Given this structure, we formally evaluate whether a simple exponential decay model provides an adequate fit to the observed revenue trajectories, focusing first on representative films for interpretability before validating the pattern at scale to rule out selection bias.

For each representative film, weekly revenues were modeled as

$$R_t = A\exp(-\alpha t),$$

where $A$ denotes the initial revenue scale and $\alpha$ represents the decay rate. Estimation was performed via linear regression in log-space after applying a $\log(1 + R_t)$ transformation to stabilize variance and accommodate zero-valued observations.

Figures 6–8 compare the fitted exponential decay curves against observed weekly revenues for a blockbuster (*Avatar: The Way of Water*), a franchise hit (*John Wick*), and a commercial flop (*Morbius*). Across all three cases, the exponential model closely tracks the empirical decline, capturing both the rapid early drop and the subsequent long-tail behavior.

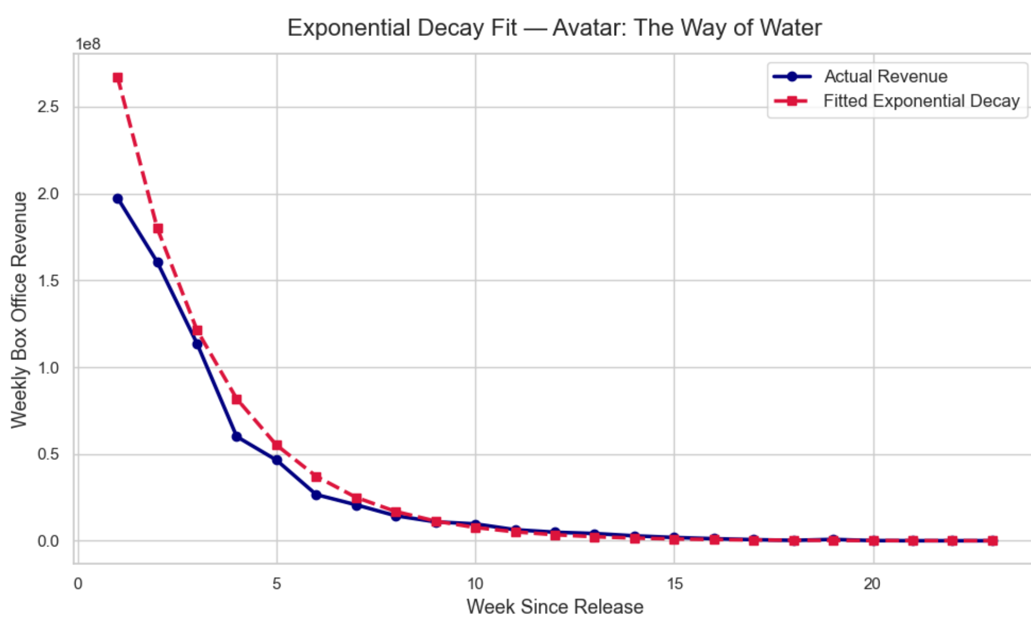



Figure 6: Exponential decay fit for *Avatar: The Way of Water*.

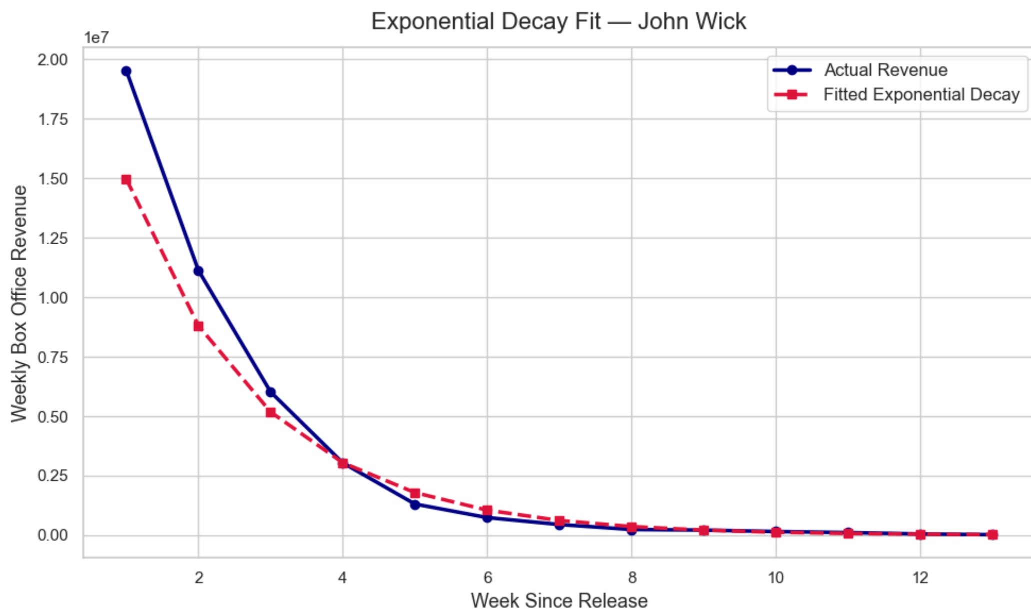



Figure 7: Exponential decay fit for *John Wick*.

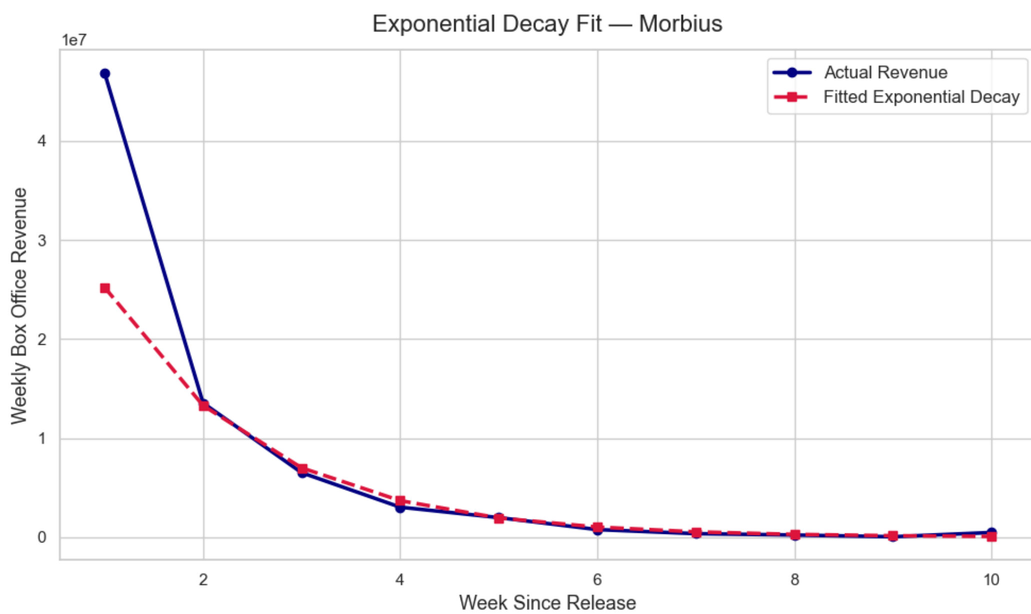



Figure 8: Exponential decay fit for *Morbius*.

The estimated decay rates differ systematically across performance tiers, with faster decay observed for weaker releases and slower decay for sustained performers. Importantly, no systematic deviations remain that would motivate more complex parametric decay forms or high-order ARIMA structures.
Overall, this analysis confirms that exponential decay provides a parsimonious and empirically adequate representation of weekly revenue dynamics. The goal of this validation is not to exhaustively search for alternative functional forms, but to formally verify that simpler decay-based modeling is sufficient. Consequently, subsequent forecasting models incorporate decay-aware structure while focusing modeling capacity on external covariates rather than unnecessary temporal complexity.

## 5.2 Panel-Level Decay Evidence Across the Dataset

While the prior subsections validated exponential decay on three representative films (one blockbuster, one hit, and one flop), we also verify that the same decay structure holds *at the dataset level* across the full set of 30 movies.

**Average decay by performance tier.** Figure 9 aggregates weekly revenue trajectories by category (blockbuster, hit, flop). All tiers exhibit a steep early decline followed by a long tail, with separation primarily in the initial scale and decay speed.

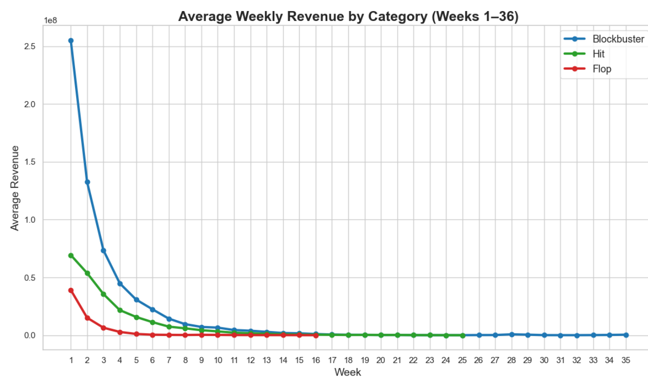



Figure 9: Average weekly box office revenue by performance category (Weeks 1–36). All tiers exhibit strong post-release decay, differing mainly in initial scale and decay speed.

**Normalized log revenue trajectories.** To compare decay shapes independent of magnitude, Figure 10 plots normalized log revenue trajectories (where each movie is normalized to its Week 1 log revenue). The resulting patterns show broadly consistent downward trends, supporting a shared decay template with movie-specific deviations.

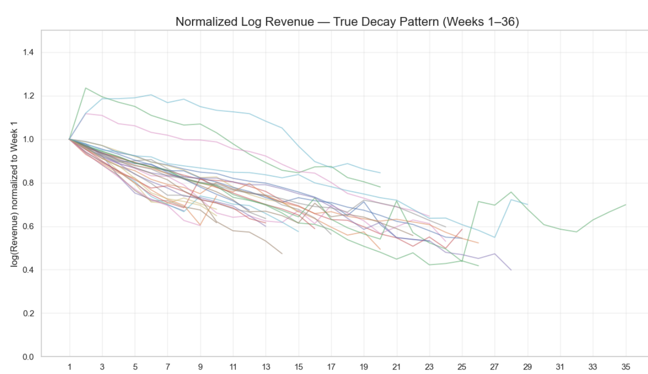



Figure 10: Normalized log revenue trajectories (Weeks 1–36), where each movie is normalized to its Week 1 log revenue. The shared downward structure supports a common decay template across films.

**Distribution of estimated decay rates.** Finally, Figure 11 summarizes estimated exponential decay rates $\alpha$ across movies. Slower decays correspond to sustained theatrical runs, while faster decays reflect rapid audience saturation and early collapse.

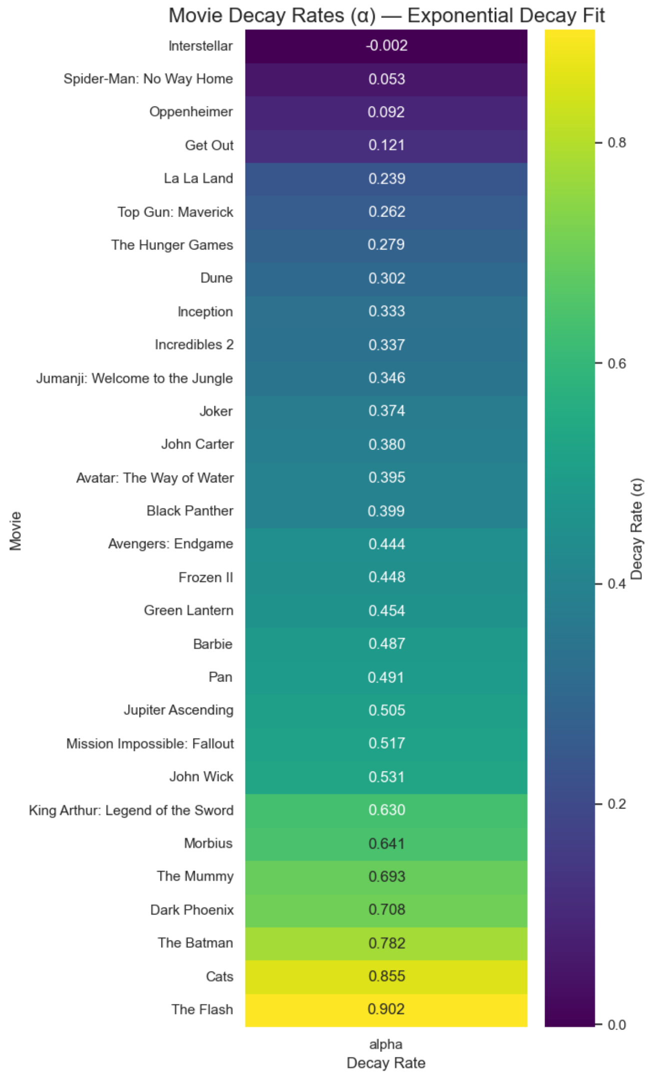



Figure 11: Estimated exponential decay rates $\alpha$ across movies (sorted from slow to fast). The spread indicates systematic variation in decay speed across releases, while preserving a consistent exponential form.

# 6 Sequence-to-Sequence Forecasting Models

The decay diagnostics and validation results establish exponential decay as a strong structural baseline for weekly box office revenues. While this decay captures the dominant monotonic trend, it does not explain systematic deviations driven by external signals such as audience engagement, sentiment dynamics, trailer characteristics, and calendar effects. Accurate short-horizon forecasting therefore requires a model capable of learning how such auxiliary signals modulate the underlying decay trajectory.

To address this setting, we adopt a decay-aware sequence-to-sequence (Seq2Seq) forecasting framework that models temporal feature evolution while preserving the exponential structure of the target process.

## 6.1 Problem Setup and Forecasting Horizon

The forecasting task is defined as predicting the final four weeks of observed weekly revenue for each movie, conditional on all prior information. Formally, given a multivariate input sequence

$$\mathbf{x}_{1:T-4} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{T-4}\},$$

the objective is to forecast

$$\{R_{T-3}, R_{T-2}, R_{T-1}, R_T\},$$

where each $\mathbf{x}_t$ includes decay-aware lag features, engagement signals, sentiment measures, trailer-derived emotion features, calendar indicators, and static movie attributes.
The fixed four-week horizon reflects a realistic industry forecasting scenario, where studios and distributors are primarily interested in late-run revenue trajectories rather than long-term extrapolation.

## 6.2 Hybrid LSTM–GRU Seq2Seq Architecture

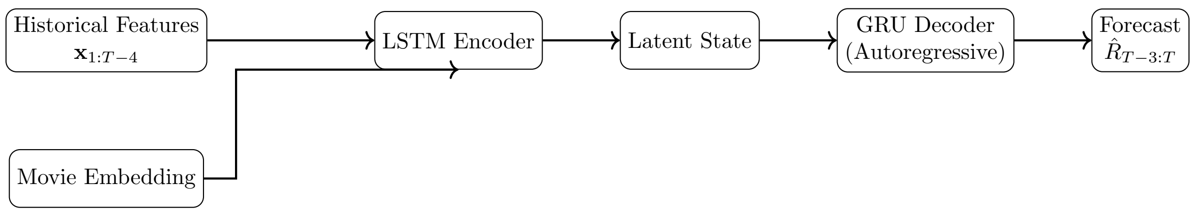



Figure 12: Baseline decay-aware Seq2Seq architecture. The LSTM encoder compresses the full historical sequence into a single latent representation, which initializes the GRU decoder for multi-step revenue forecasting.

As a baseline neural forecasting model, we employ a hybrid encoder–decoder architecture consisting of:

- an **LSTM encoder** that processes the variable-length historical sequence of covariates and learns a compressed temporal representation;
- a **GRU decoder** that autoregressively generates forecasts for the next four weeks; and
- a learned **movie embedding** that captures persistent, film-specific characteristics not fully explained by observed covariates.

The encoder summarizes historical dynamics into a latent state used to initialize the decoder. Forecasts are generated sequentially, allowing the model to capture nonlinear temporal dependencies beyond simple exponential decay.
To stabilize training under heavy-tailed revenue distributions, revenues are modeled in log space and standardized. Predictions are mapped back to the original scale for evaluation.

## 6.3 Training Protocol and Data Splits

Movies are split at the title level into disjoint training, validation, and test sets to prevent information leakage across sequences. The test set spans blockbusters, hits, and commercial flops, ensuring evaluation reflects generalization rather than memorization.
Training employs scheduled teacher forcing with a decaying ratio, early stopping based on validation loss, and a log-cosh objective function to balance robustness to outliers with sensitivity to systematic forecast errors. Model performance is evaluated using four-week-ahead MAE, RMSE, and WAPE, with results discussed in the subsequent comparison of baseline and attention-augmented architectures.

## 6.4 Attention-Augmented Decoder Extension

While the baseline Seq2Seq model compresses the full historical sequence into a single latent state, this assumption is restrictive in the box office setting. Early release performance, mid-run stabilization, and late-stage decay may contribute differently to short-horizon forecasts.
To address this limitation, we augment the decoder with an additive (Bahdanau-style) attention mechanism. At each forecast step, the decoder computes attention weights over all encoder hidden states, producing a context vector that dynamically emphasizes the most relevant historical periods. This allows the model to selectively attend to informative portions of the revenue trajectory rather than relying on a fixed summary.

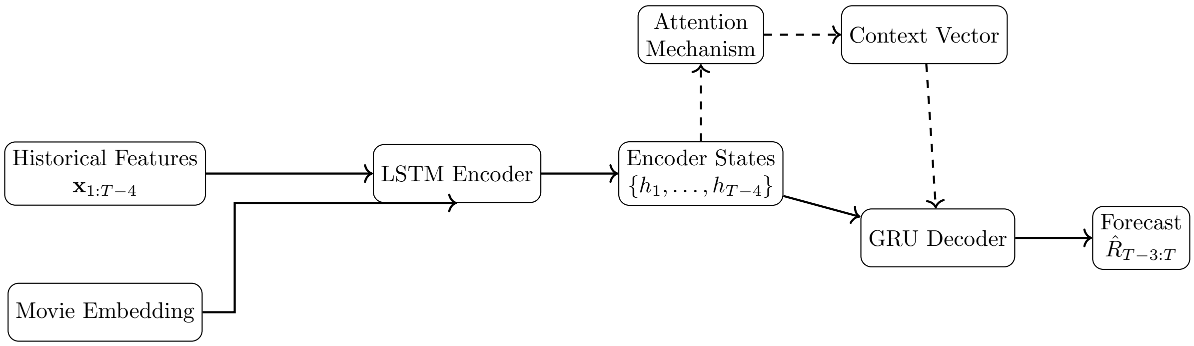



Figure 13: Attention-augmented Seq2Seq architecture. The decoder dynamically attends to encoder states at each forecast step, enabling selective use of historical information while preserving decay-aware structure.

## 6.5 Baseline vs. Attention-Augmented Comparison

To quantify the contribution of attention, Table 7 compares the baseline and attention-augmented models under identical data splits and evaluation metrics.

Table 7: Baseline vs. attention-augmented Seq2Seq performance (4-week horizon).

| **Model** | **Dataset** | **MAE** | **RMSE** | **WAPE** |
|---|---|---|---|---|
| Baseline LSTM–GRU | Validation | 36,794 | 46,617 | 0.492 |
| Baseline LSTM–GRU | Test | 212,029 | 250,072 | 1.821 |
| LSTM–GRU + Attention | Validation | 34,510 | 43,125 | 0.368 |
| LSTM–GRU + Attention | Test | 170,300 | 207,523 | 1.071 |

The attention mechanism yields consistent improvements across all metrics, with the largest relative gains observed on the test set. Test WAPE decreases by approximately 41%, indicating substantially improved proportional accuracy when generalizing to unseen movies.

## 6.6 Representative Movie-Level Performance

Table 8 reports performance for representative test movies spanning blockbusters, hits, and flops.

Table 8: Baseline vs. attention-augmented performance for representative test movies.

| | **Baseline** | | | **Attention** | | |
|---|---|---|---|---|---|---|
| **Movie** | MAE | RMSE | WAPE | MAE | RMSE | WAPE |
| Avatar: The Way of Water | 315,995 | 337,744 | 6.315 | 73,341 | 75,053 | 1.466 |
| Top Gun: Maverick | 301,498 | 415,072 | 0.683 | 405,785 | 566,722 | 0.920 |
| John Wick | 61,816 | 74,103 | 0.717 | 66,103 | 80,531 | 0.767 |
| Get Out | 85,601 | 94,048 | 1.213 | 40,262 | 49,459 | 0.571 |
| Morbius | 473,103 | 534,765 | 1.733 | 162,520 | 196,992 | 0.595 |
| Cats | 34,162 | 44,699 | 0.263 | 273,789 | 276,380 | 2.110 |

Attention substantially improves forecast accuracy for most films, particularly for titles exhibiting mid-run deviations from smooth exponential decay (e.g., *Avatar* and *Morbius*). Performance on extremely low-revenue titles remains sensitive to noise, leading to higher proportional error despite small absolute magnitudes.
Overall, these results demonstrate that attention provides a principled and statistically meaningful enhancement to decay-aware Seq2Seq forecasting by improving temporal relevance without violating the underlying exponential structure of box office revenue dynamics.

## 6.7 Feature Attribution via Integrated Gradients

To interpret how auxiliary signals modulate decay-aware Seq2Seq forecasts, we apply Integrated Gradients (IG) to the attention-augmented LSTM–GRU model. Attributions are computed with respect to the final forecasted week ($R_T$), using a zero baseline for all input features. For each movie, IG scores are aggregated across encoder timesteps by summing absolute attributions, yielding a per-feature importance measure that reflects cumulative temporal influence.

#### 6.7.1 Representative Movie-Level Feature Importance

Table 9 reports the top IG-attributed features for representative blockbusters, hits, and flops from the test set. Distinct attribution profiles emerge across performance tiers, indicating that the model adapts its reliance on auxiliary signals depending on demand regime and content type.

Table 9: Top Integrated Gradients features for representative test movies.

| **Movie** | **Top Attributed Features (Descending Importance)** |
|---|---|
| *Avatar: The Way of Water* | Sentiment range, budget scale, science fiction genre, music–emotion correlation, studio brand strength, engagement momentum |
| *Top Gun: Maverick* | Drama genre, seasonal timing, mixed sentiment flag, holiday indicator, cumulative gross, audience ratings |
| *John Wick* | Budget scale, calendar timing, arousal-related emotion embeddings, holiday season indicator, genre mix |
| *Get Out* | Horror and mystery genres, budget scale, sentiment dispersion, short-term trend averages |
| *Morbius* | Arousal and sadness emotion embeddings, fantasy/adventure genres, trend momentum, sentiment range |
| *Cats* | User ratings, calendar timing, trend momentum, fantasy genre, emotion volatility indicators |

Across all titles, attribution mass concentrates on recent encoder timesteps, confirming that the Seq2Seq model primarily uses auxiliary signals to adjust late-stage deviations rather than override the underlying exponential decay.

#### 6.7.2 Global Feature Importance Across Movies

To identify systematic drivers of late-run revenue forecasts, IG scores are aggregated across all test movies. Table 10 summarizes the globally most influential features.

Table 10: Global Integrated Gradients feature importance across test movies.

| **Feature** | **Aggregate IG Importance** |
|---|---|
| Budget scale (log) | 0.494 |
| Drama genre indicator | 0.352 |
| Calendar timing (week of year) | 0.335 |
| Horror genre indicator | 0.327 |
| Mixed sentiment flag | 0.250 |
| Studio brand strength | 0.228 |
| Arousal emotion embedding | 0.198 |
| Sentiment range | 0.186 |
| Adventure genre indicator | 0.178 |
| User ratings | 0.162 |

Notably, traditional decay variables (e.g., lagged revenue) exhibit relatively modest attribution compared to content, sentiment, and timing features. This confirms that exponential decay acts as a structural backbone, while auxiliary signals govern systematic deviations in late-stage performance.

#### 6.7.3 Interpretability Implications

Integrated Gradients analysis validates that the attention-augmented Seq2Seq model behaves in a principled manner: preserving monotonic decay while selectively leveraging sentiment, emotional content, genre composition, and calendar effects to refine short-horizon forecasts. The resulting attribution patterns are consistent with economic intuition and provide transparency into how external signals influence revenue trajectories beyond pure decay dynamics.

# 7 Discussion

This study evaluated a decay-aware sequence-to-sequence (Seq2Seq) framework for short-horizon box-office forecasting and assessed the role of attention-based decoding and trailer-derived representations when modeling multivariate external signals. The results highlight the importance of respecting structural revenue dynamics while selectively incorporating auxiliary information.

## 7.1 Decay as a Structural Prior

Across individual films and at the panel level, weekly box-office revenues exhibit a dominant exponential decay pattern with minimal residual temporal dependence. Stationarity, autocorrelation, and exponential-fit diagnostics consistently confirm that higher-order time-series structure contributes little information beyond early lags. Importantly, this behavior holds across blockbusters, moderate hits, and commercial failures, differing primarily in initial scale and decay rate rather than functional form.

These findings justify treating exponential decay as a structural prior rather than a pattern to be rediscovered by the model. By explicitly validating this assumption, the forecasting framework can allocate modeling capacity toward explaining systematic deviations driven by external signals instead of unnecessary temporal complexity.

## 7.2 Seq2Seq Forecasting Performance

The hybrid LSTM–GRU Seq2Seq architecture effectively models short, volatile revenue trajectories while producing stable four-week-ahead forecasts. Separating historical encoding from autoregressive decoding mitigates error accumulation and improves robustness under scheduled teacher forcing.

Although absolute errors are larger for high-grossing films due to scale effects, proportional accuracy remains comparable across performance tiers. This indicates meaningful generalization rather than memorization of specific revenue magnitudes.

## 7.3 Impact of Attention Mechanisms

Compressing the full revenue history into a single latent state is restrictive in the box-office setting, where early release performance, mid-run stabilization, and late-stage decay may contribute differently to short-horizon forecasts. The attention-augmented Seq2Seq model directly addresses this limitation by allowing the decoder to dynamically reweight encoder states at each prediction step.

Empirically, attention yields consistent improvements across all evaluation metrics, with test-set WAPE reduced by approximately 41%. Gains are most pronounced for films exhibiting mid-run deviations from smooth exponential decay, such as *Avatar: The Way of Water* and *Morbius*. Performance on extremely low-revenue titles remains sensitive to noise, reflecting inherent scale limitations rather than architectural failure.

## 7.4 Role of Trailer-Based Emotional Representations

Trailer-derived emotional embeddings provide complementary contextual information that is not observable from early revenue patterns alone. Knowledge distillation substantially reduces frame-level noise and neutral-dominance bias in pretrained emotion models, yielding more stable and interpretable trailer-level features.

While these embeddings are static at the movie level, they consistently appear among important explanatory signals in the attention-augmented model, particularly through arousal, emotional volatility, and sentiment dispersion. This suggests that narrative tone and emotional pacing influence late-stage revenue deviations even after accounting for decay dynamics.

## 7.5 Interpretability and Practical Implications

Integrated Gradients analysis confirms that the model behaves in a principled manner. Attribution mass concentrates on recent encoder timesteps, indicating that auxiliary signals are primarily used to modulate late-stage deviations rather than override the underlying decay structure.

Globally, attribution patterns emphasize production scale, genre composition, calendar timing, sentiment dispersion, and trailer-based emotional intensity. These results align with economic intuition and provide transparency into how external signals shape forecast adjustments beyond pure decay.

## 7.6 Limitations and Future Directions

This study is limited by the size of the curated movie panel and by treating trailer-based emotional embeddings as static features. Future work could explore larger datasets, time-varying multimodal representations, and adaptive forecasting horizons. While this work relies on temporally aligned but separately modeled audio and

visual cues, future extensions could incorporate end-to-end joint audio–visual temporal architectures, as well as explicit modeling of online sentiment evolution and social-media diffusion dynamics.
Nevertheless, the results demonstrate that decay-aware Seq2Seq models with attention and interpretable auxiliary signals provide an effective and transparent framework for short-horizon box-office revenue forecasting.

# Appendix

All source code, data processing pipelines, model training scripts, and analysis notebooks used in this project are publicly available at:
`CineSeq:A Multivariate Seq2Seq Framework for Forecasting Movie Box-Office Trajectories`