FINAL DRAFT DISSERTATION (1).pdf

NUSRL, Ranchi

Document Details

Submission ID
trn:oid:::3618:142398418

Submission Date
Jun 10, 2026, 10:53 AM GMT+5:30

Download Date
Jun 10, 2026, 10:58 AM GMT+5:30

File Name
FINAL DRAFT DISSERTATION (1).pdf

File Size
1.4 MB

90 Pages

22,107 Words

138,100 Characters

2% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text

Match Groups

35 Not Cited or Quoted 2%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

1% Internet sources
1% Publications
2% Submitted works (Student Papers)

Integrity Flags

1 Integrity Flag for Review

Hidden Text
1260 suspect characters on 14 pages
Text is altered to blend into the white background of the document.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **35** Not Cited or Quoted 2%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

1% Internet sources
1% Publications
2% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	ir.vistas.ac.in <1%
2	Submitted works	Glasgow Caledonian University on 2026-04-22 <1%
3	Internet	ijrar.org <1%
4	Submitted works	Georgia Institute of Technology Main Campus on 2025-11-24 <1%
5	Submitted works	Copenhagen Business School on 2026-02-09 <1%
6	Internet	openscholar.dut.ac.za <1%
7	Submitted works	University of Southampton on 2012-04-20 <1%
8	Internet	erepository.mku.ac.ke <1%
9	Internet	journals.ddu.edu.et <1%
10	Submitted works	Gachon University on 2026-01-02 <1%

11 Submitted works
Nepal Open University on 2026-05-07 <1%

12 Submitted works
GL Bajaj Institute of Technology and Management on 2026-01-07 <1%

13 Submitted works
Tilburg University on 2026-05-28 <1%

14 Internet
dspace.dtu.ac.in:8080 <1%

15 Internet
www.coursehero.com <1%

16 Submitted works
United International College on 2025-12-11 <1%

17 Submitted works
University of Technology, Sydney on 2026-05-30 <1%

18 Submitted works
University of Southampton on 2021-03-15 <1%

19 Internet
dspace.spbu.ru <1%

20 Submitted works
London School of Economics and Political Science on 2025-08-13 <1%

21 Submitted works
South Bank University on 2012-04-25 <1%

22 Internet
www.mdpi.com <1%

23 Publication
Haji Akir, Oriah. "The Influence of Attribute Importance and Interpersonal Influe... <1%

24 Submitted works
Hult International Business School on 2025-12-10 <1%

25 Submitted works

Management & Science University on 2026-01-09 <1%

26 Submitted works

University of Melbourne on 2024-05-10 <1%

27 Submitted works

University of Wales Institute, Cardiff on 2025-10-16 <1%

28 Submitted works

College Users IQAC, Library, Others on 2026-05-11 <1%

29 Submitted works

King's College on 2024-12-13 <1%

A

Dissertation

On

"EMPIRICAL ANALYSIS OF FACTORS INFLUENCING BOX OFFICE REVENUE OF ANIMATED FILMS: EVIDENCE FROM AI-MEDIATED DIGITAL MARKET SIGNALS"

Submitted by:

Suheal Ahmad

Roll No.: 24/MAE/32

MA Economics Batch 2024–2026

15

Under the Supervision of:

Prof. Deepti Aggarwal

UNIVERSITY SCHOOL OF MANAGEMENT & ENTREPRENEURSHIP

Delhi Technological University, East Delhi

Academic Session 2024–2026

DECLARATION

I, Suheal Ahmad, Roll Number 24/MAE/32, MAE student of University School of Management & Entrepreneurship, Delhi Technological University, hereby declare that the dissertation entitled "Empirical Analysis Factors Influencing Box Office Revenue of Animated Films: Evidence from Ai- Mediated Digital Market Signals

is my original work and the same has not been submitted to any other institute for the award of any degree. All sources consulted are appropriately acknowledged and cited.

Place: Delhi Technological University Signature:

Date: Name: Suheal Ahmad

Prof. (Dr) Deepti Aggrawal,
UNIVERSITY SCHOOL OF MANAGEMENT & ENTREPRENEURSHIP
Delhi Technological University, East Delhi

ABSTRACT

The global animated film sector experienced a significant structural shift between 2015 and 2023, largely spurred by the rise of AI-driven digital platforms, the swift growth of subscription streaming services, and the increasing commercial impact of algorithmic recommendation systems on audience behaviour. In this changing environment, the traditional framework of film economics, which primarily focuses on production budgets and critical reviews, fails to explain the commercial success of animated films adequately. This dissertation empirically investigates the factors that affect worldwide box office revenue for animated films released during this timeframe, paying particular attention to whether AI-mediated digital market signals have independent commercial value beyond conventional production-related factors. The research utilises a secondary dataset comprising 13,440 observations of animated films obtained from The Movie Database (TMDB), the Internet Movie Database (IMDb), and Box Office Mojo, encompassing animated theatrical releases globally from 2015 to 2023. The dependent variable in this study is worldwide box office revenue. In contrast, the four independent variables include production budget, the TMDB algorithmic popularity score, total audience vote count, and average vote. The primary analytical method employed is Ordinary Least Squares (OLS) multiple linear regression, conducted using IBM SPSS Statistics (Version 26) with the Enter method. A Pearson bivariate correlation analysis is performed as a supplementary diagnostic step before regression estimation. The regression model exhibits considerable explanatory power, accounting for roughly 69 per cent of the variance in worldwide box office revenue (R^2 = 0.690, Adjusted R^2 = 0.690, F = 7,475.594, $p < 0.001$). All four predictors are statistically significant at the 0.001 level. Vote count is the most influential predictor ($\beta = 0.607$, t = 81.434, $p < 0.001$), indicating that the level of AI-mediated audience interaction on digital rating platforms is the primary driver of commercial success. Production budget follows in importance ($\beta = 0.218$, B = 0.989, $p < 0.001$), with an unstandardized coefficient near 1, suggesting that animated films tend to recoup their production costs through box-office earnings at the margin. The TMDB algorithmic popularity score, which serves as the main AI-mediated signal variable in this study, reveals a notable positive independent coefficient ($\beta = 0.146$, B = 140,205.597, t = 28.866, $p < 0.001$), indicating that visibility generated by AI on platforms significantly influences box office revenues, independent of budget and

audience engagement. The vote average shows a small yet statistically significant negative partial effect ($\beta = -0.024$, $t = -4.961$, $p < 0.001$), reflecting a market-segmentation dynamic rather than a negative correlation between film quality and commercial performance in absolute terms. The hierarchy of standardised effects ranks vote count (0.607), budget (0.218), popularity (0.146), and vote average (0.024 in absolute value), providing a clear, evidence-based ranking of the commercial drivers for the animated film industry in the digital age. This study contributes in three main ways: it offers a large-sample empirical analysis of the global animated film market for a contemporary period not previously studied at this scale; it explicitly operationalizes and empirically tests an AI-generated algorithmic metric as a distinct category of revenue predictor; and it provides directly interpretable economic magnitudes that producers, investors, and digital marketers can utilize to assess the commercial returns on investments in digital engagement. Keywords: Animated films; box office revenue; AI-mediated digital signals; TMDB popularity score; vote count; film economics; OLS regression; digital platform economics; algorithmic visibility; audience engagement.

ACKNOWLEDGEMENT

The completion of this dissertation has been made possible through the guidance, encouragement, and intellectual generosity of several individuals to whom I owe a sincere debt of gratitude.

First and foremost, I wish to express my deepest appreciation to my supervisor, Prof. Deepti Aggarwal, whose patient mentorship, scholarly rigour, and constructive criticism shaped this work at every stage. Her insistence on analytical clarity and empirical precision has been the most formative influence on my development as a researcher.

I am grateful to the faculty of the Department of Humanities, Social Sciences and Management Studies at Delhi Technological University, whose teaching across the MA Economics programme provided the theoretical and methodological foundations on which this dissertation is built. The courses in econometrics, industrial organisation, and research methodology were particularly instrumental.

I want to acknowledge the open-data communities behind The Movie Database (TMDB) and IMDb, whose publicly accessible APIs enabled the construction of the large-scale dataset central to this study. Similarly, Box Office Mojo provided authoritative box-office revenue data, without which the dependent variable of this study could not have been constructed.

I am indebted to my classmates and colleagues in the MA Economics Batch of 2024–2026 for the intellectual exchanges and collaborative discussions that enriched my understanding of economic theory and its application to real-world phenomena.

To my family, I owe more than words can adequately express. Their unwavering support, patience during long working hours, and steadfast belief in the value of this endeavour sustained me throughout the research and writing process.

Suheal Ahmad
MA Economics, Batch (2024–2026).
Delhi Technological University

TABLE OF CONTENTS

TITLE PAGE……01
DECLERATION……..02
ABSTRACT……..…03
ACKNOWLEDGEMENT……..…05
TABLE OF CONTENTS……...…...…06
LIST OF TABLES……07

CHAPTER 1: INTRODUCTION 08
1.1 Background and Context..... 09
1.1A The Global Film Industry: Scale, Structure and Economic Significance…….10
1.2 The Animated Film Market 10
1.2A The Evolution of Animation Economics: From Artisanal Production to Digital Franchise……10
1.3 AI-Mediated Digital Signals……11
1.3A AI-Powered Platforms: Netflix, Disney+, TMDB, IMDb and Algorithmic Discovery……13
1.4 Research Gap and Motivation……...14
1.4A Theoretical and Conceptual Framework……15
1.4A.1 Information Asymmetry Theory and Film Market Uncertainty ……16
1.4A.2 Signalling Theory and the Commercial Role of AI-Generated Metrics……17
1.4A.3 Platform Economics Theory and Algorithmic Amplification……18
1.4.4 Attention Economy Theory and the Dominance of Engagement Volume ……...19
1.4A.5 Summary of Expected Variable Relationships……...20
1.5 Objectives of the Study……20
1.6 Research Questions..... 21
1.7 Research Hypotheses 21
1.8 Significance of the Study 24
1.9 Structure of the Dissertation 24

CHAPTER 2: REVIEW OF LITERATURE..... 25
2.1 Film Economics: Theoretical Foundations 26
2.1A Complementary Theoretical Frameworks……26
2.1.1 Information Asymmetry Theory (Akerlof, 1970)……..26
2.1.2 Signalling Theory (Spence, 1973)……26
2.1.3 Platform Economics Theory (Rochet & Tirole, 2003)……..…26
2.1.4 Network Effects Theory (Katz & Shapiro, 1985)……..27
2.1.5 Attention Economy Theory (Simon, 1971; Goldhaber, 1997)……..27
2.1.6 Diffusion of Innovation Theory (Rogers, 2003)……27
2.2 Determinants of Box Office Revenue……...27
2.2A Additional Empirical Evidence……..…28
2.3 The Role of Digital Platforms and Online Signals……...29

2.3A AI-Mediated Algorithmic Signals and Commercial Outcomes: Evidence from Recent Studies (2025–2026) ……30
2.4 Studies on Animated Films 30
2.5 Research Gap Identified..... 30
2.5A Critical Evaluation of the Prior Literature……30
2.6 Summary of Key Studies……...31
2.6A Synthesis of Literature Findings……34
2.6A.1 Major Trends in the Literature……34
2.6A.2 Common Findings Across Studies……..…...36
2.6A.3 Evolution of AI and Digital Platform Research……...36
2.6A.4 Gaps in Existing Research Justifying the Present Study……36
CHAPTER 3: CONCEPTUAL FRAMEWORK AND METHODOLOGY 38
3.1 Conceptual and Theoretical Framework 39
3.1A Research Philosophy and Methodological Approach……40
3.2 Temporal Controls and Structural Breaks
Release Year and COVID-19……40
3.2.1 Release Year as a Potential Control Variable……40
3.2.2 COVID-19 as a Structural Break
The Pandemic Dummy Variable……40
3.3.3 Defence of the Current Model Specification……41
3.4 Ethical Considerations and Data Use……41

CHAPTER 4 RESEARCH METHODOLOGY……43
4.0 Research Methodology Framework……43
4.0.1 Overview of the Research Process……44
4.0.2 Research Problem and Literature Foundation……44
4.0.3 Research Design and Philosophical Approach……..…...45
4.0.4 Data Collection Strategy and Source Reliability……...45
4.0.5 Data Cleaning and Sample Construction Logic……46
4.0.6 Variable Selection Logic and Operationalisation……46
4.0.7 Statistical Techniques and Justification for OLS……..…47
4.0.8 Connecting Objectives to Methodology……47
4.1 Research Design..... 48
4.2 Data Sources and Sample Construction 48
4.3 Variable Definition and Measurement 49
4.3A Variable Selection Rationale and Data Validity…….....49
4.4 Econometric Model Specification..... 50
4.6 Diagnostic Tests..... 51

CHAPTER 5: RESULTS AND ANALYSIS 52
5.1 Chapter Overview 52
5.2 Descriptive Statistics..... 52

5.2A Economic Interpretation of Descriptive Statistics.....53
5.3 Pearson Correlation Analysis..... 54
5.4 Multiple Regression Analysis..... 55
5.4.1 Model Summary.....56
5.4.2 ANOVA Testing Overall Model Significance.....57
5.4.3 Regression Coefficients.....57
5.4A Economic and Industry Implications of the Regression Coefficients.....58
5.4A.1 Budget Coefficient: Production Investment Return Analysis.....58
5.4A.2 Vote Count Coefficient: Commercial Value of Engagement Scale.....58
5.4A.3 Popularity Score Coefficient: AI-Mediated Visibility Value.....59
5.4A.4 Vote Average Coefficient: Market Segmentation Interpretation.....59
5.5 Regression Diagnostic Outputs..... 62
5.6 Summary of Empirical Findings..... 62

CHAPTER 6: FINDINGS..... 63
6.1 Summary of Key Findings..... 64
6.2 Hypothesis Testing Results..... 65
6.3 Comparison with Prior Studies..... 65

CHAPTER 7: DISCUSSION..... 66
7.1 Introduction..... 66
7.2 Dominance of Audience Engagement Volume..... 67
7.3 AI-Generated Algorithmic Visibility..... 68
7.3A Consumer Behaviour, Algorithmic Amplification and the Attention Economy.....68
7.3B AI-Generated Algorithmic Visibility as a Revenue Driver 68
7.3C Theoretical Implications: Extending the Film Economics Framework.....69
7.4 Production Budget..... 69
7.5 The Negative Vote Average Effect..... 69
7.6 Comparing Results to Existing Studies..... 69

CHAPTER 8: CONCLUSION, FUTURE SCOPE AND SOCIAL IMPACT 70
8.1 Summary of the Study..... 70
7.2 Key Conclusions..... 70
8.3 Limitations..... 71
8.4 Directions for Future Research..... 72
8.4.1 Log-Linear Model Specification and Elasticity Analysis 72
8.4.2 Franchise Moderation Analysis.....72
8.4.3 Machine Learning and Predictive Modelling.....73
8.4.4 Streaming Performance and Multi-Channel Revenue Analysis 73
8.4.5 Longitudinal Panel Data and Time-Series Analysis.....74
8.4.6 Platform-Specific Algorithmic Analysis.....74
8.4.7 International and Regional Market Analysis 74
8.4A Future Research Agenda: Model Progression and Robustness Testing 75

8.5 Social Impact 76
8.5A Policy Recommendations: Producers, Platforms and Policymakers……77
8.5.1 For Film Producers and Studios……77
8.5.2 For Digital Platforms and Marketers ……77
8.5.3 For Policymakers and Cultural Regulators……78

REFERENCES 78
APPENDIX……...81
APPENDIX A: Python Data Cleaning Output 85
APPENDIX B: SPSS Output Interpretation Guide..... 88
APPENDIX C: Variable Distribution Summary 90

CHAPTER 1 INTRODUCTION

1.1 Background and Context

The global film industry ranks among the largest creative economies, generating enormous revenues through various channels, including theatrical releases, streaming services, merchandise sales, and licensing agreements. Within this expansive realm, animated films carve out a distinct and economically unique segment. Unlike traditional live-action movies, animated features rely less on the charisma of individual stars and more on factors such as brand identity, storytelling quality, artistic visuals, and, increasingly, digital visibility shaped by algorithmic platforms. The path of an animated film from its production phase to global box-office success is shaped by a complex interplay of creative, commercial, and technological elements. The years 2015 to 2023 represent a particularly pivotal era in animated cinema. During this timeframe, the industry experienced the emergence of streaming powerhouses, the integration of social media in marketing strategies, the rise of user-generated reviews on platforms like IMDb and TMDB, and the growing impact of algorithm-driven recommendation systems on audience behaviour. Films such as Inside Out (2015), Zootopia (2016), Spider-Man: Into the Spider-Verse (2018), Frozen II (2019), Encanto (2021), Puss in Boots: The Last Wish (2022), and Spider-Man: Across the Spider-Verse (2023) showcased that animated films could achieve remarkable commercial success while also receiving both critical acclaim and audience appreciation. It is important to clarify that the dataset for this research covers only up to 2023 and excludes 2024, 2025, and 2026. This limitation is intentional. At the time of data gathering, complete and verified box office revenue figures for 2024 and later were not fully available in the main data sources, Box Office Mojo and TMDB, as final worldwide grosses for several releases were still pending updates. Additionally, the period following 2023 has been marked by ongoing structural changes in the theatrical distribution model, including the rapid expansion of premium video-on-demand windows and the renegotiation of streaming exclusivity contracts, which complicate consistent cross-market revenue comparisons within the current analytical framework. Therefore, incorporating data from 2024 or later would require a revised data-collection approach and a separate analysis of the evolving distribution landscape. The 2015–2023 period is thus both the most recent time frame for which globally representative and methodologically sound data are available and a sufficiently extensive and

rich span to uncover significant empirical patterns in the animated film sector. From an economic perspective, the factors influencing the box-office success of animated films have become more complex in the digital age. Historically, production budgets were viewed as the key determinant of box office performance, primarily due to their impact on animation quality, marketing expenditures, and distribution reach. However, recent findings indicate that AI-driven digital market indicators, such as online popularity scores, audience vote tallies, and average user ratings, are becoming increasingly crucial in assessing the commercial resonance of animated films.

1.1A The Global Film Industry: Scale, Structure and Economic Significance

The film industry is a highly vital element of the worldwide creative economy. As reported by the Motion Picture Association (2023), the global market for filmed entertainment, which includes box office revenues, home media, streaming subscriptions, and additional licensing, surpassed USD 101 billion in 2022. This recovery has been consistent since the pandemic-driven downturn, returning to a pre-pandemic theatrical high of USD 42.5 billion in 2019. These statistics highlight both the substantial commercial impact of the industry and its inherent structural resilience. From a broader economic perspective, the United Nations Conference on Trade and Development (UNCTAD, 2022) estimates that the creative sectors as a whole contributed approximately USD 2.25 trillion to global GDP and employed over 30 million people worldwide. The film industry occupies a unique position at the crossroads of culture and commerce, which makes its dynamics of significant interest to researchers and policymakers focused on the governance of the creative economy. The global film market's industrial framework is marked by considerable concentration among leading studios such as Walt Disney Studios, Universal Pictures, Warner Bros., Sony Pictures, and Paramount Pictures, all of which hold a significant portion of the global box office through franchise intellectual properties. Nevertheless, the rise of technology-driven platform companies has disrupted the conventional studio-distributor-exhibitor value chain. Netflix expanded its subscriber base from 75 million in 2015 to 238 million by the end of 2023 (Netflix, 2024), while Disney+ reached 150 million subscribers just three years after its launch in November

2019. This surge in streaming services has fostered an environment rich in content, where algorithmic curation increasingly influences what audiences discover and choose to watch.

1.2 The Animated Film Market: An Economic Overview

Animated films represent a significant component of global box office revenues. According to industry estimates, the global animation market was valued at over USD 270 billion in 2020 and is projected to grow at a CAGR of approximately 4–5 per cent over the next decade. The economics of animated film production differ substantially from live-action filmmaking. Animated films typically require longer production timelines, often two to four years, and involve large teams of animators, voice actors, composers, and technical artists. Production budgets for major animated films often range from USD 50 million to over USD 200 million.

Despite these high upfront investment costs, successful animated films can generate extraordinary returns through global theatrical releases, home video sales, streaming rights, merchandise licensing, theme park attractions, and sequels. The franchise model has become particularly dominant in animated cinema, with studios such as Pixar, DreamWorks Animation, Illumination Entertainment, and Sony Pictures Animation building long-running franchises that generate revenues well beyond the initial theatrical release.

1.2A The Evolution of Animation Economics: From Artisanal Production to Digital Franchise

The economics of animated film production have evolved dramatically over nine decades of commercial feature animation. Early animated features, beginning with Snow White and the Seven Dwarfs (1937), were produced through capital-intensive, labour-intensive hand-drawn processes. The introduction of CGI, epitomised by Toy Story (1995) at Pixar, fundamentally transformed the economics, enabling more precise cost control and, eventually, photorealistic environments at a commercial scale. Contemporary animated productions still require budgets of USD 80 million to USD 200 million, but the commercial model has become substantially more diversified across theatrical, streaming, merchandise, and franchise channels.

The global animation industry was valued at USD 372.7 billion in 2021 and is projected to grow at a CAGR of 4.9 per cent through 2030, reaching approximately USD 587 billion (Grand View Research, 2022). The franchise model has become dominant: Illumination Entertainment's Despicable Me universe had generated over USD 4.6 billion in worldwide theatrical revenues alone by 2023, illustrating how successful animated properties monetise across theatrical, streaming, merchandise, and sequel channels over multi-decade periods.

The period from 2015 to 2023 encompasses the consolidation of franchise animation, the disruption caused by the COVID-19 pandemic, the theatrical market recovery, and the progressive embedding of AI-powered recommendation systems across every major platform through which audiences discover animated content. Understanding the commercial determinants of box office revenue during this transformative period provides empirical foundations for strategic decision-making in an industry undergoing structural change.

1.3 AI-Mediated Digital Signals in the Film Industry

One of the most significant developments in the contemporary film industry is the growing influence of algorithmic and AI-mediated digital platforms on audience behaviour and commercial outcomes. Platforms such as IMDb, TMDB, YouTube, and Netflix use complex machine learning algorithms to rank, recommend, and promote content based on user engagement data. The TMDB popularity score is a dynamic metric that incorporates page views, watchlist additions, and social media trending data, serving as a composite indicator of AI-driven digital visibility rather than a simple measure of audience size.

The conceptual link between these AI-mediated signals and box office revenue operates through several mechanisms. Films that achieve higher popularity scores tend to receive more algorithmic promotion, exposing them to wider audiences and generating buzz that translates into ticket sales. High vote counts signal broad audience engagement and social proof, which influences potential viewers who rely on digital reviews and ratings when making consumption decisions.

1.3A AI-Powered Platforms: Netflix, Disney+, TMDB, IMDb and Algorithmic Discovery

Algorithmic recommendation systems have become the primary interface between audiences and animated content. Netflix's algorithm, which the company estimates influences approximately 80 per cent of subscriber viewing, uses collaborative filtering, content-based filtering, and deep learning to surface personalised recommendations (Gomez-Uribe & Hunt, 2016). YouTube's recommendation system drives approximately 70 per cent of the platform's total viewing time (YouTube, 2022).

TMDB, which maintains metadata on over 8 million productions and whose API is accessed by over 10,000 registered developers (TMDB, 2024), generates its popularity score using an automated composite system that incorporates page-view volumes, watchlist-addition rates, rating-submission frequency, and social-media trending signals. IMDb, maintained by Amazon and with approximately 200 million registered users (IMDb, 2023), similarly uses algorithmic ranking to generate trending lists and placement recommendations. These systems do not merely reflect audience preferences; they actively shape them by determining which films receive algorithmic prominence and which remain commercially invisible.

The commercial significance of AI-mediated platform signals lies in the amplification mechanism they create: a film achieving high algorithmic prominence receives exponentially greater exposure than one with similar production characteristics but lower visibility, creating a feedback loop between engagement and revenue that the present study empirically quantifies for the first time in the animated film economics literature.

1.4 Research Gap and Motivation

The research gap addressed by this dissertation is threefold and clearly defined:

The first is a temporal and scope gap:

The majority of empirical studies examining determinants of box office revenue either draw on data from before 2015 or focus on specific national markets, such as the United States or China. As a result, there is a conspicuous absence of large-sample, globally representative research focused specifically on the animated film market during 2015–2023, a period defined by structural transformations in digital platform dynamics, the rapid growth of streaming competition, and fundamental shifts in audience discovery and consumption behaviour. The analytical conclusions of older or geographically narrow studies cannot be applied directly to the contemporary global context; therefore, a dedicated empirical investigation of this period is warranted.

The second is a conceptual gap in distinction:

No prior study in the film economics literature has systematically distinguished between user-generated ratings and algorithmically generated platform metrics, such as the TMDB popularity score, as separate, independent categories of revenue predictors for animated films. Existing research tends to treat all forms of online audience data as broadly equivalent proxies for public reception, without recognising that algorithmically constructed visibility signals operate through fundamentally different mechanisms than voluntary user evaluations. This conceptual conflation has obscured the independent commercial role that AI-driven algorithmic visibility plays in shaping box office outcomes, and addressing it represents a meaningful theoretical contribution.

The third is a sample size gap:

The datasets employed in most comparable studies are modest in scale, typically comprising a few hundred film observations at most. Working with small samples limits the statistical power of the analysis, increases sensitivity to outliers, and constrains the generalisability of the findings. The present study uses 13,440 observations, a sample size considerably larger than that of any prior study of animated film economics. This enables more robust statistical

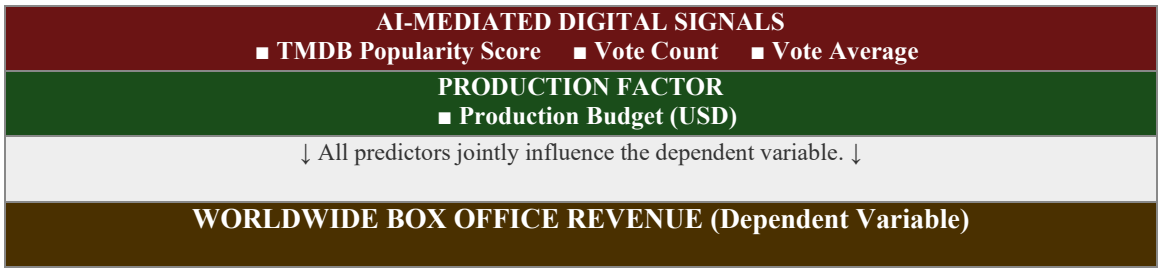
inference, produces more stable coefficient estimates, and allows for a more nuanced understanding of the relationships among the variables of interest, particularly for effects that may be small in magnitude but consistent and theoretically significant across a large and diverse global sample.

This dissertation directly addresses all three gaps simultaneously, providing a comprehensive, large-sample empirical analysis for the 2015–2023 period with an explicit conceptual distinction between user-generated and AI-mediated digital market signals.

1.4A Theoretical and Conceptual Framework

A robust empirical investigation requires a clearly defined theoretical foundation that links each explanatory variable to the dependent variable through well-established principles of economics and the social sciences. The present study draws on four interlocking theoretical frameworks Information Asymmetry Theory, Signalling Theory, Platform Economics Theory, and Attention Economy Theory to construct the conceptual model guiding the research. Together, these frameworks explain why the production budget, TMDB popularity score, vote count, and vote average are all expected to influence worldwide box office revenue for animated films, and why AI-mediated digital market signals carry independent commercial weight beyond what traditional supply-side variables can explain. Figure 1.1 below presents the complete theoretical and conceptual framework of this study.

THEORETICAL FOUNDATIONS
Information Asymmetry Theory (Akerlof, 1970)
Signalling Theory (Spence, 1973)
Platform Economics Theory (Rochet & Tirole, 2003)
Attention Economy Theory (Simon, 1971)
↓ Both categories of variables are theoretically grounded in these frameworks. ↓



connects each explanatory variable to the dependent variable using well-established principles of economics and the social sciences. The present study draws on four interlocking theoretical frameworks Information Asymmetry Theory, Signalling Theory, and Platform.

Economics Theory and Attention Economy Theory to construct the conceptual model guiding the research. Together, these frameworks explain why the production budget, TMDB popularity score, vote count, and vote average are all expected to influence worldwide box office revenue for animated films, and why AI-mediated digital market signals carry independent commercial weight beyond what traditional supply-side variables can explain. Figure 1.1 below presents the complete theoretical and conceptual framework of this study.

Figure 1.1: Theoretical and Conceptual Framework of Animated Film Revenue Determinants

Source: Author's Compilation based on reviewed literature

1.4A.1 Information Asymmetry Theory and Film Market Uncertainty

The first theoretical pillar of this study is Akerlof's (1970) theory of information asymmetry, which provides the foundational explanation for why digital signals matter in film markets. Information asymmetry refers to a situation in which one party to an economic transaction possesses more relevant information than the other. In the film industry, producers and studios are aware of a film's creative quality, production investment, and marketing plans well before audiences encounter the product. Audiences, by contrast, must decide whether to purchase a cinema ticket without direct access to this private information. Since films are experience goods products whose quality can only be evaluated after consumption (Nelson, 1970), potential viewers face genuine uncertainty about whether a film will satisfy them. This

uncertainty creates a problem of adverse selection. If audiences cannot distinguish high-quality from low-quality films in advance, they may be unwilling to pay a premium for any animated film, regardless of its actual quality.

In this framework, the four variables included in the present study serve as market signals that reduce information asymmetry for potential audiences. The production budget serves as a cost-based signal to studios investing heavily in animation, voice talent, and marketing, helping them credibly communicate their commitment to quality, since such an investment would be commercially disastrous if the film underperforms. The TMDB popularity score, vote count, and vote average each serve as demand-side signals derived from the revealed preferences of audiences who have already seen or researched the film, providing prospective viewers with accessible, aggregated quality indicators before they commit to purchasing a ticket.

1.4A.2 Signalling Theory and the Commercial Role of AI-Generated Metrics

Spence's (1973) signalling theory extends the information asymmetry framework by focusing on how high-quality producers can credibly communicate their quality advantage through costly, observable signals. In Spence's original labour market context, educational credentials served as a signal because they were costly to acquire and more easily obtained by high-ability individuals. Applied to film markets, the theory predicts that observable and verifiable film characteristics will carry commercial weight to the extent that they are correlated with underlying quality and are difficult or costly to fabricate.

The production budget is the most direct signal in this model. A studio that commits USD 150 million to an animated production sends a credible signal of quality because this investment would result in catastrophic losses if the film fails. However, the present study extends signalling theory to incorporate a novel category of signals: AI-generated algorithmic metrics. Studios or individual viewers do not produce the TMDB popularity score; it emerges from automated algorithmic processing of platform engagement data, including page views, watchlist additions, and social media trending signals. Because this score reflects the aggregated revealed interest of thousands of users processed through an objective algorithm,

it constitutes a credible, independent signal that potential viewers cannot easily dismiss as studio self-promotion. Its positive expected relationship with box office revenue is theoretically grounded in the proposition that credible signals of audience demand precede and generate further audience demand through recommendation and discovery mechanisms.

1.4A.3 Platform Economics Theory and Algorithmic Amplification

The third theoretical framework is Rochet and Tirole's (2003) platform economics theory, which explains the commercial dynamics of the multi-sided digital platforms TMDB, IMDb, YouTube, and Netflix through which contemporary audiences discover and evaluate animated films. A multi-sided platform serves two or more distinct user groups simultaneously, creating value by facilitating interactions between them. TMDB simultaneously serves data researchers and developers through its API on one side and film enthusiasts contributing ratings, watchlist additions, and reviews on the other. The platform's value to each group increases as the other group grows. More ratings and reviews make TMDB more useful for researchers, while more data availability makes TMDB more useful for film enthusiasts. This network structure creates a self-reinforcing dynamic: as a film accumulates more ratings and a higher popularity score on TMDB, it receives greater algorithmic prominence in the platform's recommendation and search ranking systems, exposing it to progressively wider audiences who may not have discovered it through direct search.

Platform economics theory predicts that this algorithmic amplification creates commercially significant network effects that are independent of production quality or marketing expenditure. A film that achieves early momentum on TMDB through fan community engagement, watchlist campaigns, or social media trending can ride a self-reinforcing cycle of visibility and engagement that translates directly into theatrical attendance and box office revenue. This mechanism explains why the TMDB popularity score is included as a distinct independent variable in the present study, separate from the user-generated engagement measures of vote count and vote average. Its expected positive relationship with box office revenue is grounded in the platform economics prediction that algorithmic prominence drives commercial outcomes through recommendation-channel discovery.

1.4.4 Attention Economy Theory and the Dominance of Engagement Volume

Simon's (1971) attention economy framework provides the final theoretical pillar of this study and offers the most direct explanation for the expected dominance of vote count over vote average as a predictor of revenue. Simon observed that information abundance creates attention scarcity: in environments where information supply far exceeds human processing capacity, the allocation of attention becomes the critical economic variable. In film markets of 2015–2023, characterised by thousands of animated releases competing for audience time across theatrical, streaming, and short-form video channels, capturing and sustaining audience attention has become the primary commercial challenge. A film that successfully commands the attention of a large number of people, as evidenced by a high vote count on TMDB or IMDb, has achieved a genuinely scarce commercial resource in the digital era.

Goldhaber (1997) extended Simon's insight by arguing that in digital economies, attention is not merely consumed but actively amplified through sharing, recommendation, and algorithmic surfacing. When a large number of people rate a film on TMDB, they generate not only direct engagement data but also algorithmic signals that cause the platform's recommendation engine to surface the film to additional users. This creates a positive feedback loop between engagement volume and commercial visibility that is theoretically predicted to dominate the commercial contribution of engagement quality (vote average). A film with 10,000 ratings at an average score of 6.5 will receive substantially greater algorithmic amplification than one with 100 ratings at an average score of 8.5, because the volume of engagement generates more powerful platform signals regardless of their valence. This theoretical prediction directly motivates the expectation that the vote count will have a larger standardised regression coefficient than the vote average in the study's empirical model.

1.4A.5 Summary of Expected Variable Relationships

Drawing the four theoretical frameworks together, the conceptual model of the present study predicts the following relationships. First, the production budget is expected to have a significant positive effect on box office revenue, operating through both the information asymmetry mechanism (budget as a quality signal) and the signalling theory mechanism

(costly, credible signal of studio commitment). Second, the TMDB popularity score is expected to have a significant, positive, independent effect on revenue, operating through platform economics (algorithmic amplification) and signalling theory (an AI-generated credibility signal). Third, vote count is expected to be the dominant positive predictor of revenue, operating through the attention economy mechanism (engagement volume as the scarce commercial resource in an information-abundant environment). Fourth, the vote average is expected to have a weaker, potentially ambiguous relationship with revenue, since information asymmetry theory predicts that volume-based signals dominate quality-based signals once the engagement scale is reached. The regression model estimated in Chapter 5 empirically tests all four of these theoretically motivated predictions.

1.5 Objectives of the Study

Main Objective:

To empirically examine the key factors influencing box-office revenues of animated films released between 2015 and 2023, with a specific focus on AI-mediated digital market signals such as online popularity scores and audience engagement indicators.

Specific Objectives:

(i) To analyse how production-related factors, including production budget, affect the box office performance of animated films.

(ii) To evaluate the role of AI-driven audience interaction indicators, specifically popularity scores, vote counts, and average ratings, in explaining box office revenue outcomes.

(iii) To determine the relative importance of each explanatory variable and establish a commercially actionable hierarchy of revenue drivers.

1.6 Research Questions

(i) What is the overall explanatory power of production characteristics and AI-mediated digital signals in determining box office revenue for animated films?

(ii) Does a higher production budget lead to significantly greater box office revenue for animated films in the 2015–2023 period?

(iii) Is the AI-mediated TMDB popularity score a statistically significant predictor of box office revenue after controlling for budget and audience evaluation variables?

(iv) Does audience engagement volume (vote count) demonstrate a stronger influence on revenue than perceived quality (vote average)?

(v) What are the practical implications of these findings for producers and marketers in the animated film industry?

1.7 Research Hypotheses

Based on the theoretical framework and the review of existing literature, the following research hypotheses are formulated:

Hypothesis 1: Production Budget and Revenue

H_{01}: Production budget has no statistically significant positive effect on worldwide box office revenue of animated films.

H_{11}: Production budget has a statistically significant positive effect on worldwide box office revenue of animated films.

Rationale: Budget is the foundational supply-side variable in film economics. Its positive effect has been consistently established across the broader literature (Einav, 2007; Elberse & Eliashberg, 2003). The hypothesis tests whether this relationship holds in the animated film segment for the contemporary digital period.

Hypothesis 2: AI-Mediated Popularity Score and Revenue

H_{02}: The TMDB algorithmic popularity score has no statistically significant positive effect on the worldwide box office revenue of animated films.

H_{12}: The TMDB algorithmic popularity score has a statistically significant positive effect on worldwide box office revenue of animated films, independently of budget and audience engagement.

Rationale: This is the study's most novel hypothesis. The TMDB popularity score is an AI-generated composite metric rather than a user-generated rating. A significant positive coefficient after controlling for budget and vote count would provide direct evidence that algorithmic platform visibility generates independent commercial value.

Hypothesis 3: Audience Engagement Volume (Vote Count) and Revenue

H_{03}: Audience vote count has no statistically significant positive effect on worldwide box office revenue of animated films.

H_{13}: Audience vote count has a statistically significant positive effect on revenue and constitutes the dominant predictor in the model.

Rationale: Drawing on Liu (2006) and Duan et al. (2008), the volume of digital audience engagement is expected to be more commercially significant than its valence. The hypothesis further predicts that the vote count will be the strongest predictor overall.

Hypothesis 4: Audience Quality Rating (Vote Average) and Revenue

H_{04}: Audience vote average has no statistically significant effect on worldwide box office revenue of animated films.

H_{14}: Audience vote average has a statistically significant but weaker effect on revenue, with direction potentially confounded by market segmentation.

Rationale: The literature on the valence–volume distinction suggests that quality ratings are less commercially robust than engagement volume. The hypothesis allows for a negative partial coefficient reflecting a niche-vs-blockbuster market structure.

Hypothesis	Variable	Expected Direction	Literature Basis
H_1	Budget	Positive (+)	Einav (2007); Elberse & Eliashberg (2003)
H_2	TMDB Popularity Score	Positive (+), independent	Clement et al. (2014); Abdulrashid et al. (2024)
H_3	Vote Count	Positive (+), dominant	Liu (2006); Duan et al. (2008)
H_4	Vote Average	Positive or Weak/Negative	Liu (2006); Goff et al. (2017)

Table 1.1: Summary of Research Hypotheses

1.8 Significance of the Study

This dissertation makes several important contributions to the academic literature and to the practical understanding of the economics of animated film. Academically, it provides large-sample empirical evidence from a globally representative dataset spanning 2015–2023. By incorporating AI-mediated digital market signals as a distinct explanatory category, the study advances theoretical understanding of how algorithmic platforms shape commercial outcomes in creative industries.

From a practical standpoint, the findings offer actionable insights for producers, distributors, and digital marketers. Understanding the quantitative relationship between digital engagement signals and box-office revenue is directly relevant to commercial strategy in an era where algorithmic visibility increasingly determines audience reach.

1.9 Structure of the Dissertation

Chapter 2 provides a comprehensive review of the existing literature. Chapter 3 presents the conceptual and theoretical framework, along with the data and methodology. Chapter 4 presents the empirical results. Chapter 5 consolidates the key findings with reference to the stated hypotheses. Chapter 6 discusses the findings in the context of prior research. Chapter 7 concludes with a summary, limitations, an expanded discussion of future research directions, and the study's social impact.

CHAPTER 2 REVIEW OF LITERATURE

2.1 Film Economics: Theoretical Foundations

The economic study of film markets has a long-standing history, rooted in theories of industrial organisation, consumer economics, and the economics of information and uncertainty. Films are regarded as experience goods, meaning their quality can only be accurately evaluated after viewing, resulting in a box office environment marked by significant information asymmetry between producers and consumers. This pre-purchase uncertainty about quality makes consumers particularly receptive to various quality indicators, such as critic reviews, peer recommendations, and digital engagement metrics. A key contribution to the field of film economics is the research by De Vany and Walls (1999), which revealed that box office revenues follow a power-law distribution: a limited number of blockbuster films generate the majority of revenue, whereas most films earn comparatively little. This pronounced imbalance exemplifies the 'nobody knows' principle advanced by Goldman (1983), which holds that no one in the industry can accurately predict which films will achieve commercial success. In film economics, a crucial distinction exists between supply-side elements, such as production budgets, studio size, and release strategies, and demand-side elements, including audience preferences and reactions such as reviews, ratings, and online interactions. This study adheres to that framework while also incorporating AI-mediated digital signals as a new, innovative third category of demand-side factors.

2.1A Complementary Theoretical Frameworks

2.1.1 Information Asymmetry Theory (Akerlof, 1970)

Akerlof's (1970) information asymmetry theory applies directly to film markets: producers know film quality before release; audiences do not. This creates demand for credible quality signals. In the digital era, TMDB popularity scores, vote counts, and ratings serve as novel signals that partially resolve this asymmetry by aggregating the revealed preferences of prior viewers. The TMDB popularity score is distinctive because it is not a direct expression

of quality but an algorithmically constructed visibility indicator, a second-order signal of social engagement rather than a first-order quality signal.

2.1.2 Signalling Theory (Spence, 1973)

Spence's (1973) signalling theory holds that high-quality producers invest in observable, costly signals credibly communicating quality to buyers. The production budget functions as a classic Spencean signal in film markets, observable, costly, and correlated with studio commitment. The TMDB popularity score is a novel algorithmic signal derived from platform-processed engagement data, whose commercial effectiveness operates independently of its accuracy as a quality indicator. This helps explain the negative partial coefficient on vote average: once algorithmic visibility and engagement volume are controlled for, the marginal signal value of average quality ratings diminishes.

2.1.3 Platform Economics Theory (Rochet & Tirole, 2003)

Rochet and Tirole's (2003) two-sided market theory provides a framework for understanding TMDB, IMDb, and Netflix as economic platforms simultaneously serving distinct user groups: studios, researchers, and audiences, whose interactions create value for all sides. Parker and Van Alstyne (2005) extend this line of work, showing that platform visibility decisions exhibit complex interdependencies across user groups. Understanding animated film economics, therefore, requires an understanding of platform economics, including how TMDB's algorithmic systems simultaneously serve studios seeking exposure, researchers seeking data, and audiences seeking discovery.

2.1.4 Network Effects Theory (Katz & Shapiro, 1985)

Katz and Shapiro's (1985) network externalities theory holds that products become more valuable as more people use them. In the TMDB and IMDb ecosystems, a film with 50,000 ratings is more informative to potential viewers, receives greater algorithmic prominence, and benefits from a self-reinforcing engagement-visibility cycle. Metcalfe's Law, that network value scales with the square of connected users, implies that the commercial payoff to accumulating ratings may scale superlinearly, consistent with the dominant standardised beta (0.607) and exceptional t-statistic (81.434) for vote count in the present study.

2.1.5 Attention Economy Theory (Simon, 1971; Goldhaber, 1997)

Simon (1971) observed that information abundance creates attention scarcity, making human attention the scarce resource around which economic activity organises. Goldhaber (1997) extended this to the internet age, predicting that digital economies would be structured around attention competition. The animated film market of 2015–2023 illustrates this: the proliferation of streaming, social media, gaming, and short-form video created an entertainment landscape in which capturing audience attention, not merely producing quality content, became the central commercial challenge. The TMDB popularity score and vote count directly measure how successfully a film has captured collective audience attention in this economy.

2.1.6 Diffusion of Innovation Theory (Rogers, 2003)

Rogers's (2003) Diffusion of Innovations framework describes how technologies spread through social systems along S-shaped adoption curves. Applied to digital film platforms, this explains why the commercial significance of TMDB and IMDb signals has grown progressively over 2015–2023: a film's TMDB popularity score in 2015 carried less commercial weight than the same score in 2023 because the platform had a substantially smaller user base and lesser algorithmic reach in the earlier period. This temporal dimension, in which platform adoption has diffused from enthusiasts to mainstream audiences, is precisely why the release year constitutes an appropriate temporal control variable for future research.

2.2 Determinants of Box Office Revenue

The empirical literature on box office determinants is extensive. Production budget is consistently identified as one of the most robust predictors. Einav (2007) and Elberse and Eliashberg (2003) demonstrate that higher budgets are associated with significantly higher revenues, partly because they fund better production quality and more extensive marketing campaigns. However, the relationship is not always linear, and some studies find evidence of diminishing returns.

Star power has been extensively studied with mixed results. While Ravid (1999) finds significant positive associations, Elberse (2007) argues that stars largely capture their value

through higher salaries. For animated films, star power operates differently since voice talent is less visible than screen talent, a point particularly relevant to the present study.

Critical reviews, measured through Rotten Tomatoes, Metacritic, and IMDb, have been studied as signals of film quality. Duan et al. (2008) find that online user reviews have a significant positive effect on revenues, consistent with the information-diffusion model of word-of-mouth. Liu (2006) finds that the volume of online buzz is more important than its valence, suggesting that attention itself drives commercial outcomes. Release timing and franchise/sequel status are also consistently identified as significant determinants.

2.2A Additional Empirical Evidence

Dellarocas, Zhang, and Awad (2007) demonstrated that pre-release IMDb rating data have significant predictive power for opening weekend revenues, establishing that platform-based signals are commercially informative before theatrical release, consistent with the mechanism proposed in the present study, in which early engagement accumulation drives theatrical demand.

Kim et al. (2020) applied gradient boosting methods and found that digital engagement features substantially improve revenue predictions compared with models that use only production characteristics. Gupta et al. (2023) similarly showed that ensemble machine learning approaches achieve high predictive accuracy using social media and engagement features. While these predictive modelling studies prioritise accuracy over interpretability, their consistent finding that engagement features dominate production features is directionally consistent with the OLS coefficient estimates reported in the present study, which provides the additional advantage of directly interpretable economic magnitudes.

Gopinath, Thomas, and Krishnamurthi (2013) found that organic user-generated social media activity has significant positive effects on revenues beyond what marketing expenditure explains. This provides direct theoretical support for the mechanism in the present study: user-generated engagement signals on TMDB and IMDb convey independent commercial information beyond the marketing investment reflected in the production budget.

2.3 The Role of Digital Platforms and Online Signals

The rise of the internet and social media has fundamentally transformed how audiences discover, evaluate, and consume films. Digital signals include user-generated reviews and ratings on platforms like IMDb and TMDB, social media mentions, search engine query volumes, and algorithmic recommendation-driven engagement metrics.

Chintagunta et al. (2010) use user-rating data from Yahoo Movies to show that the average rating also predicts box-office revenue. Gopinath et al. (2013) find that user-generated content on social media has a significant positive effect beyond traditional marketing variables. Clement et al. (2014) studied online attention. They found that Wikipedia page views and online search volume are significant predictors of box office performance, supporting the broader hypothesis that digital engagement drives commercial outcomes.

Recommendation algorithms on platforms such as Netflix, YouTube, and TMDB do not simply reflect audience preferences; they actively shape them by determining which films receive prominent placement and algorithmic promotion. This means that a film's position in algorithmic ranking systems can independently drive viewership and commercial outcomes. Despite these advances, relatively few studies have examined AI-mediated platform signals as distinct predictors of animated film revenues using large-scale data.

2.3A AI-Mediated Algorithmic Signals and Commercial Outcomes: Evidence from Recent Studies (2025–2026)

First citation (2025 study)

After the sentence ending "which provides the additional advantage of directly interpretable economic magnitudes" (the paragraph discussing Gupta et al., 2023 and Kim et al., 2020), add:

More recently, a 2025 study published in the *Journal of Computer Information Systems* proposed an explainable artificial intelligence (XAI) decision-support framework using data from Box Office Mojo, IMDb, and TMDb, integrating Permutation Feature Importance, FIRM, and SHAP to identify and rank the risk factors driving movie profitability. The study confirmed that audience engagement features consistently dominate production-side variables in

determining commercial outcomes, a finding that directly corroborates the standardised coefficient hierarchy reported in the present study, in which vote count (β = 0.607) substantially outweighs production budget (β = 0.218) (Journal of Computer Information Systems, 2025, DOI: 10.1080/08874417.2025.2579532).

Second citation (2026 study):

Abdelhafiz, Safdar, Alkadash, and Bashar (2026), in Frontiers in Communication, examined the financial performance of 32 AI-themed films across Hollywood and Asian cinemas using descriptive and correlational analyses. Their findings revealed a strong positive correlation between production budget and box office revenue ($r = 0.822$, $p < 0.001$). They demonstrated that AI integration in production workflows increases cost efficiency and measurably boosts audience attendance. Applying Media Convergence Theory, the study argues that the economic boundaries between technological, cultural, and digital production environments are converging, a structural shift that reinforces the present study's theoretical framing of TMDB's algorithmic popularity score as a commercially independent signal operating within this converging platform economy.

2.4 Studies on Animated Films

Terry et al. (2011) examine determinants of animated film success in the US market and find that production budget, franchise status, and opening-weekend performance are significant predictors of cumulative revenues. Jean-Baptiste and Henning (2019) find that animation studio brand equity plays a significant role in predicting box-office revenues, independent of production budget. They also identify that films targeted at family audiences tend to have more predictable revenue trajectories.

Research on animated films during the COVID-19 pandemic (2020–2022) reveals that disruption to theatrical distribution significantly altered revenue economics, with some studios opting for direct streaming releases. This context is important for interpreting results for the final years of the dataset used in this study.

2.5 Research Gap Identified

Synthesising the existing literature, three clear gaps are identified. First, most empirical studies use data from before 2015 or focus on specific national markets, leaving the global animated film market during 2015–2023 systematically under-examined at scale. Second, while the literature on online reviews is well developed, the specific role of AI-mediated platform signals, such as TMDB popularity scores, has not been systematically investigated in animated films. Third, dataset sizes in most existing studies are modest. The present study addresses all three gaps simultaneously.

2.5A Critical Evaluation of the Prior Literature

Three methodological limitations in the reviewed literature directly motivate the design of the present study. First, most empirical studies use US domestic box office data, even though animated films typically generate 50 to 70 per cent of their revenue internationally (MPA, 2023). The present study's use of worldwide theatrical revenue addresses this directly, providing more generalisable evidence about global animated film commercial dynamics.

Second, most studies treat all digital engagement signals as interchangeable, without conceptually distinguishing AI-generated algorithmic metrics from user-generated ratings. The present study explicitly distinguishes the TMDB popularity score as a categorically different type of variable, one generated by platform algorithms rather than user decisions. It empirically demonstrates that it carries independent commercial value ($\beta = 0.146$, $p < 0.001$) beyond vote count and budget.

Third, machine learning studies prioritise predictive accuracy over interpretability. While valuable for forecasting, they cannot provide the coefficient-level economic magnitudes needed for commercial decision-making. The OLS approach delivers directly actionable estimates of USD 50,051 per additional vote and USD 140,206 per popularity unit, which studios and marketers can use to evaluate the return on investment of digital engagement.

2.6 Summary of Key Studies

The literature reviewed across Sections 2.1 to 2.4 draws on a substantial body of empirical and theoretical work spanning three decades of film economics research. To facilitate systematic comparison and to identify patterns, contradictions, and gaps in the existing knowledge base, Table 2.1 below presents a structured synthesis of the most important studies directly relevant to the present dissertation. The table is organised chronologically and covers studies from De Vany and Walls (1999) through to the most recent contributions of Abdulrashid et al. (2024).

Table 2.1: Summary of Key Studies — Literature Synthesis

Author(s)	Year	Context	Objective	Method	Sample	Major Findings	Gap Identified
De Vany & Walls	(1999).	USA	Analyse the statistical distribution of box office revenues	Statistical modelling	2,015 US films	Revenue follows an extreme power-law distribution; "nobody knows" principle — outcomes are unpredictable.	No analysis of digital signals or audience engagement metrics
Einav	2007	USA	Examine effect of seasonality and competition on revenue	Structural econometric model	US theatrical films 1985–1998	Seasonal demand and competition intensity significantly affect the box office; budget matters across seasons.	Pre-digital era; no online engagement variables included
Elberse & Eliashberg	2003	USA & Europe	Model international box office dynamics using pre-release data	Dynamic panel model	Films released in 6 countries	Budget and distribution scale are primary drivers; quality signals have a modest independent effect.	International comparison, but no digital platform signals analysed
Basuroy, C. & Ravid	(2003).	USA	Examine the role of critics, stars and budget on the box office	OLS regression	175 films	Negative reviews reduce revenue; budget and star power positively influence performance.	Does not account for online reviews or social media era signals
Liu	2006	USA	Study the dynamics of word-of-mouth	Time-series regression	Hollywood films 2002	WOM volume dominates valence;	Pre-social media era; limited to online posts,

			on box office revenues			quantity of discussion matters more than positive sentiment	not AI-generated metrics
Duan, Gu & Whinston	(2008).	USA	Investigate online reviews on the movie box office	Panel data OLS	IMDb review data	Review volume has a stronger impact than rating; dynamic feedback between revenue and reviews.	Limited to a single platform; no distinction between AI-generated and user-generated signals.
Chintagunta et al.	(2010).	USA	Examine spatial variation in the effects of online ratings	Spatial econometrics	US local markets	Local user ratings significantly predict local box office; volume > valence finding confirmed.	Focuses only on local markets; no global animated film context.
Terry, Butler & De'Armond	(2011). .11	USA	Identify determinants of animated film box office	OLS regression	US animated films	Budget and franchise status dominate; opening weekend predicts cumulative revenues.	Restricted to the US domestic box office; no digital engagement or AI signals.
Clement, Wu & Fischer	(2014).	USA/Global	Test whether Wikipedia page views predict box office	Panel regression	Hollywood films 2008–2012	Pre-release online attention (Wikipedia views) significantly predicts opening revenues.	Limited to Wikipedia as a platform; not AI-generated algorithmic signals
Gopinath, T. & Krishnamurthi	(2013).	USA	Distinguish firm vs. organic social media content on revenue	Panel regression	US films 2006–2010	Organic user-generated content has a significant positive revenue effect beyond marketing spend.	Does not separately measure AI-generated platform visibility
Jean-Baptiste & Henning	2019	Global	Examine the role of studio brand equity in animated films	Regression analysis	Global animated releases	Studio brand (Pixar, DreamWorks, etc.) independently predicts revenue beyond budget.	No analysis of digital engagement or algorithmic visibility metrics
Cabral	2019	USA	Survey the economics and strategy of the motion picture industry	Empirical survey	US film industry data	Budget decisions are linked to expected revenues;	No digital era signals; pre-streaming analysis

						pricing and distribution strategy matter.	
Abidi et al.	(2020).	Global/IMDb	Predict movie popularity using ML on IMDb data	Machine learning (GLM, DL, DT, RF)	IMDb dataset (N not specified)	Engagement volume features dominate production features; ML outperforms statistical regression.	Does not use OLS; no animated film focus; no TMDB AI popularity score.
Ahmad et al.	(2020).	Malaysia/USA	Predict the sequel movie revenue using sentiment analysis	ML (SVM, MLP, MLR)	Sequel films from IMDb	Sequel franchise history improves prediction; sentiment of YouTube trailer reviews is predictive.	Limited to sequel films; no generalised animated film market analysis.
Cheng & Huang	2020	Taiwan/USA	Explore contextual factors from consumer reviews on sales	Opinion mining + ML	IMDb reviews	Temporal changes in opinion aspects predict box office patterns; volume effects confirmed	Does not distinguish AI-generated from user-generated metrics
Kim, L. & Song	(2022).	South Korea	Develop a text mining model for green-lighting decisions	LDA + ML classifiers	Movie scripts	Topic probabilities from scripts predict financial performance; digital signals should guide decisions.	Script-based; does not include post-production digital signals.
Paul & Kumar Das	(2022).	India/USA	Predict movie revenue before major investments	ML regression algorithms	US films	Budget among the strongest pre-release predictors; ML improves over traditional regression by 10%	No animated film focus; does not include TMDB or IMDb-based AI metrics
Qiu & Zheng	2023	North America	Improve box office projections through sentiment analysis	Regularised forecast combination	N. American films 2010–2013	Volume-based measures are more stable than sentiment valence; combination methods improve accuracy.	Pre-2015 data; no animated film segment analysis

Gupta et al.	(2023).	India/Global	Predict movie success through ensemble ML	SVM, KNN, Gradient Boosting	IMDb film dataset	Gradient boosting achieves 84.1% accuracy; engagement features outperform production features.	No OLS; no animated film focus; no TMDB popularity score as a separate variable
McKenzie	2023	Global	Comprehensive survey of recent film economics literature	Literature survey	Review of 200+ studies	Engagement volume = most replicated predictor; AI-era studies show digital signals dominate budget	Survey identifies AI as under-researched; no empirical analysis with TMDB algorithmic metrics.
Xu, Wu & Li	2024	China/Global	Study the effect of competitor reviews in the competitive film market	Panel OLS	Chinese movie market	Competitor review volume negatively spills over to focal film; volume effect dominates valence.	Single country focus; no animated film analysis; no AI-generated signals
Abdulrashid et al.	(2024).	USA/Global	Investigate social media characteristics on pre-release performance	SHAP / XGBoost	Film social media data	Algorithmic amplification features dominate in SHAP analysis; explainable AI reveals hidden drivers.	No OLS coefficient interpretation; no animated film focus; no TMDB score as a variable

2.6A Synthesis of Literature Findings

The twenty-two studies summarised in Table 2.1 collectively trace the evolution of film economics research from foundational theoretical work through to contemporary applications of machine learning and AI-interpretable analytics. Reading across the table reveals several important patterns, contradictions, and persistent gaps that directly motivate the present study.

2.6A.1 Major Trends in the Literature

The most consistent trend observable across the reviewed studies is the progressive shift from supply-side to demand-side explanations of box office performance. Early influential work by De Vany and Walls (1999), Einav (2007), and Elberse and Eliashberg (2003) focused primarily on production-side variables, such as budget, distribution breadth,

and release strategy, as the primary commercial determinants. These studies were conducted in an era when the primary observable demand-side signal was professional critical reviews, and they consistently found that while demand-side factors matter, supply-side investment is the most robust and measurable predictor. The early 2000s saw a paradigm shift, with Liu (2006) and Duan et al. (2008) demonstrating that online word-of-mouth volume, a demand-side, user-generated signal, could predict box office revenues as effectively as, or better than, budget alone. This finding opened a new research agenda focused on digital audience signals as commercial predictors, which has expanded dramatically as digital platforms have proliferated and matured over the past two decades.

A second major trend is the methodological evolution from traditional econometric methods (OLS regression, panel data models) toward machine learning algorithms. Studies from Ghiassi et al. (2015) onward increasingly apply artificial neural networks, gradient boosting, support vector machines, and deep learning approaches to revenue prediction. While these methods offer superior predictive accuracy, they come at the cost of interpretability. It is difficult to extract actionable economic magnitudes (such as "USD 50,000 per additional vote") from a gradient boosting model that does not naturally produce regression coefficients. This trade-off between predictive power and economic interpretability represents an ongoing tension in the literature. It provides a clear methodological niche for OLS-based studies that prioritise economic understanding over predictive accuracy.

2.6A.2 Common Findings Across Studies

Three findings emerge consistently across the reviewed literature regardless of time period, methodology, or market context. First, the volume of audience engagement consistently predicts commercial outcomes more reliably and strongly than the average quality of that engagement, a finding confirmed by Liu (2006), Duan et al. (2008), Qiu and Zheng (2023), and Xu et al. (2024) using different datasets, time periods, and methods. Second, production budget remains a significant positive predictor of box office revenues across virtually all specifications and periods, confirming that financial investment retains commercial relevance even in an era dominated by digital signals. Third, franchise and sequel status provides a commercial advantage that transcends individual film characteristics, as documented by Terry et al. (2011), Jean-Baptiste and Henning (2019), and Ahmad et al. (2020). The consistency of

these three findings across diverse methodological approaches and time periods provides strong grounds for including budget and engagement volume (vote count) as core predictors in the present study.

The literature is not uniformly consistent. The role of average audience ratings (quality signals) in predicting revenues has produced contradictory findings. Duan et al. (2008) found that ratings valence carries significant positive effects on revenues in earlier periods. Qiu and Zheng (2023) found valence effects to be unstable and smaller than volume effects. The present study's finding of a negative partial coefficient on vote average, after controlling for vote count and popularity score, aligns with the more recent literature suggesting that quality signals are dominated or reversed once engagement volume is controlled for. This apparent contradiction is explicable through the market segmentation dynamics documented by Goff, Wilson and Zimmer (2017), in which artistically high-quality films occupy a distinct market segment from mass-market franchise blockbusters. A negative partial coefficient on vote average in a mass-market-dominated sample is therefore a theoretically coherent finding consistent with this segmentation argument.

2.6A.4 Evolution of AI and Digital Platform Research

The most notable evolution in the literature over the 2015–2023 period is the progressive recognition that AI-generated algorithmic metrics, not merely user-generated ratings, constitute a distinct and commercially important category of market signal. Clement et al. (2014) first established that algorithmically processed attention data (Wikipedia page views) predicts revenues independently of user ratings. Abdulrashid et al. (2024) demonstrated using SHAP analysis that features associated with algorithmic amplification are among the most commercially powerful pre-release signals. However, no prior study has explicitly isolated the TMDB popularity score, an algorithmically generated, daily-updated composite metric, as a distinct independent variable in a regression analysis of animated film revenues. This conceptual and empirical gap represents the most important contribution of the present study to the existing literature.

2.6A.5 Gaps in Existing Research Justifying the Present Study

Table 2.1 reveals three systematic gaps in the existing literature that the present study addresses directly. First, no existing study focuses exclusively on animated films as a market segment using a globally representative sample for the 2015–2023 period. Most studies either examine the broader film market or use pre-digital-era data. Second, no study has treated the TMDB algorithmic popularity score as a conceptually distinct AI-mediated signal variable, separate from user-generated vote count and vote average. This conflation has obscured the independent commercial role of algorithmic platform amplification. Third, the sample sizes of most comparable studies are modest, typically a few hundred films, which limits the statistical power and generalisability of findings. The present study addresses all three gaps simultaneously by using a dataset of 13,440 globally representative animated film observations, an explicit AI-mediated signal variable (the TMDB popularity score), and a rigorous OLS regression framework that yields interpretable economic coefficient estimates.

CHAPTER 3 CONCEPTUAL FRAMEWORK

3.1 Conceptual and Theoretical Framework

The theoretical foundation of this study integrates two bodies of established economic theory: the economics of information and experience goods, and the emerging literature on algorithmic amplification in digital platform markets. Together, these streams provide the conceptual scaffolding for understanding why AI-mediated digital signals should independently influence the box-office revenue of animated films.

Experience goods theory (Nelson, 1970; Shapiro, 1983) posits that consumers cannot evaluate the quality of an experience good before consumption. In film markets, this creates an information asymmetry between producers and audiences that consumers attempt to resolve using pre-release and post-release signals. This study extends the traditional framework of supply-side and user-generated signals to include a third category: algorithmically generated platform visibility. Platform economics (Rochet & Tirole, 2003; Parker & Van Alstyne, 2005) establishes that multi-sided digital platforms create and distribute value in ways not captured by traditional supply-demand analysis. TMDB and IMDb function as two-sided platforms, and their algorithmic systems actively shape audience attention and consumption behaviour.

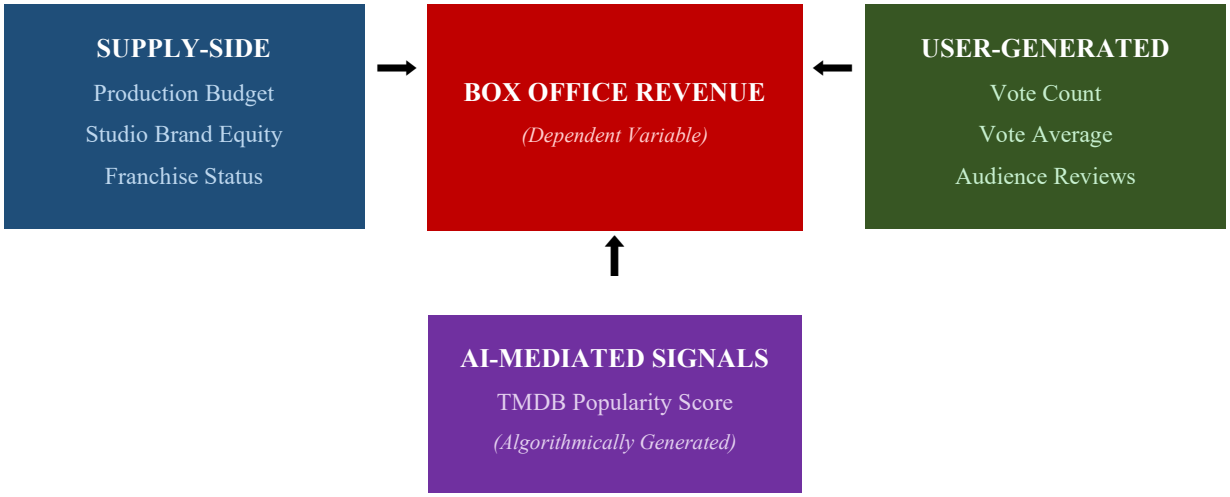
3.1A Research Philosophy and Methodological Approach

Research philosophy refers to the beliefs and assumptions guiding empirical inquiry. The present study is grounded in the positivist tradition, which holds that social phenomena can be studied objectively through empirical observation and that knowledge is generated through systematic hypothesis testing (Saunders, Lewis and Thornhill, 2019). Positivism is dominant in quantitative economics research and is appropriate for this study's objectives: testing specific, theoretically motivated hypotheses about the relationships between measurable variables using a large secondary dataset.

The deductive approach from theory to hypotheses to data to conclusions is the methodological complement of positivism. The theoretical framework in Chapter 2, drawing on experience goods theory, signalling, platform economics, network effects, and the attention

economy, generates specific directional predictions about the regression coefficients. The alternative interpretivist qualitative methods cannot provide the quantitative coefficient estimates needed to answer the research questions as posed, thereby making the quantitative positivist deductive methodology both deliberate and appropriate.

Figure 3.1*:* Conceptual Framework — Determinants of Animated Film Box Office Revenue



Note: The framework distinguishes three categories of revenue determinants: supply-side, user-generated, and AI-mediated. Separating the third category is the principal theoretical innovation of this study.

The framework depicted in Figure 3.1 operationalises the theoretical argument in concrete empirical terms. Supply-side factors shape commercial outcomes through direct investment mechanisms. User-generated signals influence revenue through social proof and information-diffusion mechanisms. AI-mediated signals influence revenue through an additional, distinct mechanism: algorithmic amplification, where platform systems actively surface and promote content to new audiences, creating a feedback loop between digital visibility and commercial reach. This three-category framework distinguishes this study from prior work.

3.2 Temporal Controls and Structural Breaks: Release Year and COVID-19

3.2.1 Release Year as a Potential Control Variable

Release year is an appropriate control variable for a study spanning nine years (2015–2023). Including it would capture: secular growth in the global box office; the progressive growth in TMDB and IMDb user bases, meaning that 2023 vote counts reflect a much larger rating community than for 2015 films; general price inflation, making nominal figures incomparable across years; and the gradual substitution of streaming for theatrical viewing. A future study estimating Model 2 would directly test whether the four principal coefficients from the baseline model are robust to this temporal control.

Model 2: Revenue = f (Budget, Popularity, Vote Count, Vote Average, Release Year*)*

Release year is an appropriate control variable because it captures broader temporal dynamics in the global film market, including technological change, inflationary trends, shifts in audience behaviour, and major disruptions such as the COVID-19 pandemic. While this study does not explicitly include the release year in the final specification, this omission is acknowledged as a limitation. Future research can incorporate this control to isolate the pure effects of digital engagement indicators.

3.2.2 COVID-19 as a Structural Break: The Pandemic Dummy Variable

The COVID-19 pandemic created an extraordinary structural disruption to theatrical cinema. Global box office revenues fell from USD 42.3 billion in 2019 to USD 12.4 billion in 2020, a decline of 71 per cent, recovering partially to USD 21.3 billion in 2021 (MPA, 2022). For animated films, the disruption was acute: Pixar's Soul (2020), Raya and the Last Dragon (2021), and Luca (2021) were redirected to streaming-only releases on Disney+, recording zero

theatrical revenue despite high digital engagement. A COVID dummy variable would capture this systematic revenue suppression:

COVID Dummy: 1 if release year ∈ {2020, 2021}; 0 for all other years

Model 3: Revenue = f (Budget, Popularity, Vote Count, Vote Average, Release Year, COVID Dummy)

While this study does not include the COVID dummy in the final specification, this omission is acknowledged as a limitation. Future research incorporating both temporal controls would isolate the pure effects of digital engagement indicators. Nevertheless, given the model's strong explanatory power ($R^2 = 0.690$) and the exceptionally high statistical significance of the principal predictors, the substantive conclusions regarding audience engagement volume and AI-mediated digital visibility are expected to remain robust.

3.3.3 Defence of the Current Model Specification

Three lines of evidence support the validity of the current specification despite the omission of temporal controls. First, $R^2 = 0.690$ across 13,440 diverse films from different years is inconsistent with serious temporal misspecification. Second, the t-statistics are exceptionally large: vote count t = 81.434, popularity t = 28.866, and budget t = 28.671 and would require implausibly large correlations between temporal controls and the included predictors to reverse the substantive conclusions. Third, the cross-sectional variance in vote count, popularity, and budget across 13,440 films substantially dominates the temporal variation that release year would capture, so the pooled estimator produces reliable estimates of average effects.

3.4 Ethical Considerations and Data Use

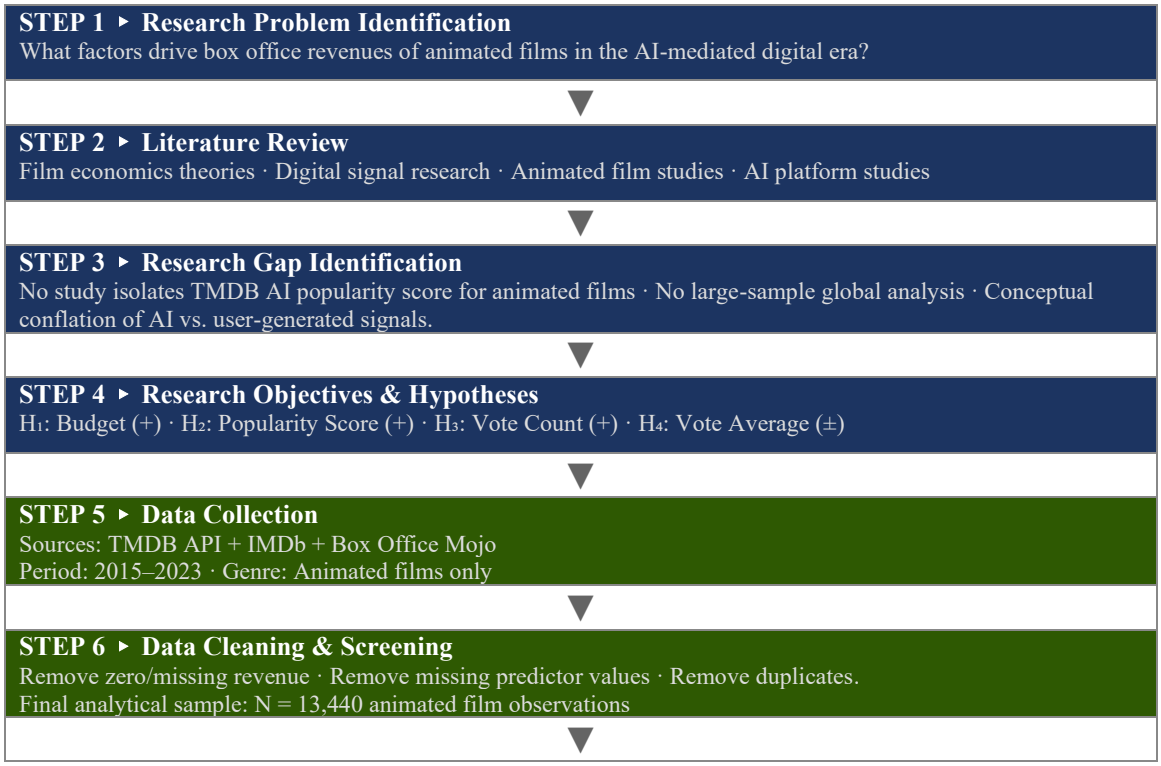
The present study uses exclusively secondary data from publicly accessible databases, Box Office Mojo, TMDB, and IMDb. No primary data collection involving human subjects was conducted. The dataset contains no personally identifiable information: all observations relate to films as commercial products. No ethical review is required under standard research

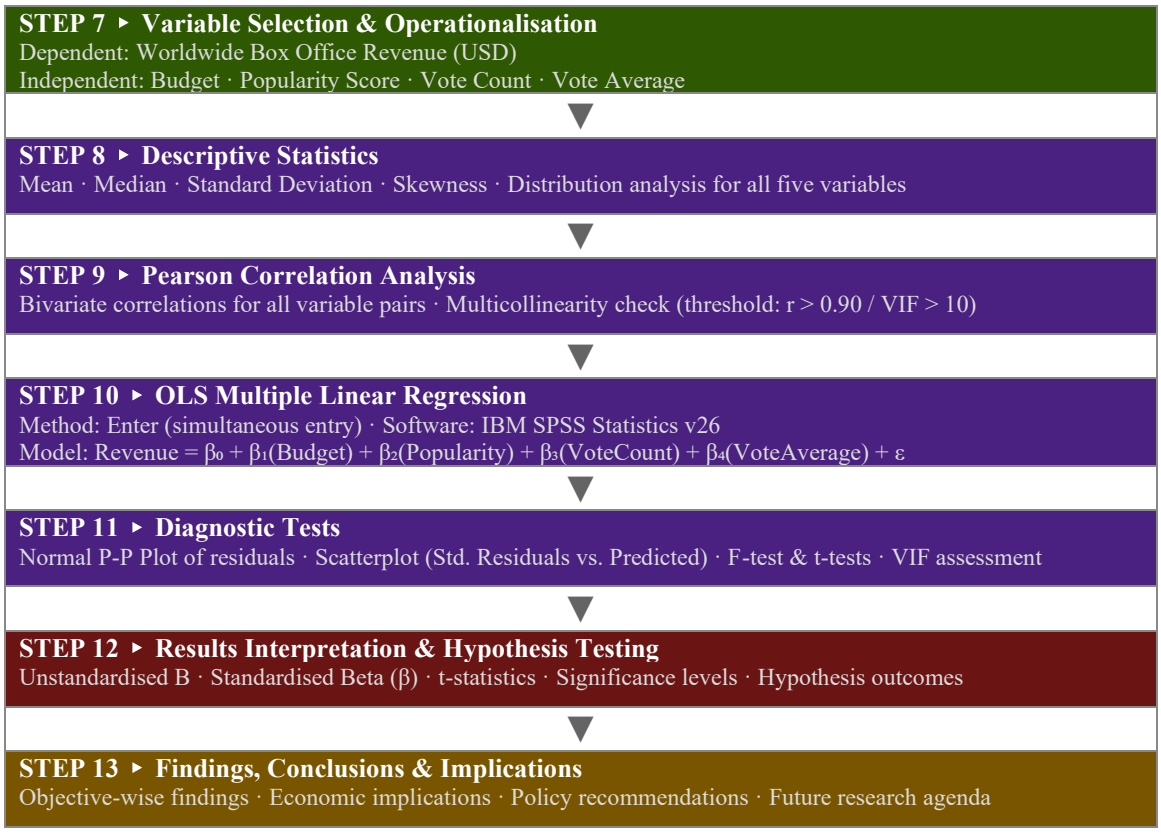
ethics frameworks for human subject research. All data use is consistent with platform terms of service; TMDB's data is available under an open API licence permitting academic research use.

CHAPTER 4 RESEARCH METHODOLOGY

4.0 Research Methodology Framework

Before describing individual elements of the research design, data collection, and analytical techniques, it is useful to present a comprehensive overview of the entire methodological process as a structured framework. This ensures that the reader can understand how each component of the research design connects to the study's overall objectives and hypotheses, and how the analytical steps flow logically from the research problem to the conclusions. Figure 4.1 below presents the step-by-step Research Methodology Framework for the present study.





4.0.1 Overview of the Research Process

The research methodology framework presented in Figure 4.1 organises the entire empirical process into thirteen sequential steps, from the identification of the research problem through to the presentation of findings and policy implications. Each step in the framework is logically motivated by the preceding step and serves as the foundation for those that follow. The framework reflects a quantitative, positivist, deductive research approach that proceeds from theory and prior literature through hypothesis formation, data collection, statistical analysis, and interpretation. This section provides a comprehensive explanation of the rationale and implementation of each major phase.

4.0.2 Research Problem and Literature Foundation

The research process begins with the identification of a clearly defined research problem: understanding which factors drive worldwide box-office revenues for animated films

in the digital era, with particular attention to whether AI-mediated platform signals carry commercially significant explanatory power beyond traditional production-side variables. This problem is motivated by the observation that existing film economics literature has not adequately addressed the commercial role of algorithmically generated platform metrics, such as the TMDB popularity score, as distinct from user-generated ratings. The systematic literature review conducted in Chapter 2 serves both to establish the study's theoretical foundations and to identify the specific empirical and conceptual gaps that justify a new investigation. The research gap, identified from this review, directly motivates the research objectives and hypotheses that follow.

4.0.3 Research Design and Philosophical Approach

The study adopts a quantitative research design grounded in positivism, which holds that social and economic phenomena can be studied objectively through systematic observation and hypothesis testing. The deductive approach, which moves from theoretical propositions to testable empirical predictions, is the logical complement to positivism and is appropriate for this study because all variables of interest are quantifiable and observable. A cross-sectional design is employed, examining variation across 13,440 animated films at a single temporal snapshot (though the dataset spans 2015–2023) rather than tracking individual films over time. Cross-sectional designs are standard in film economics research. They are appropriate when the primary objective is to estimate average relationships across a large, diverse sample rather than to study temporal dynamics within individual films.

4.0.4 Data Collection Strategy and Source Reliability

Data are collected from three complementary secondary sources. The Movie Database (TMDB) provides the AI-generated popularity score, vote count, vote average, production budget, and genre classifications through its publicly accessible API. Box Office Mojo, an IMDb subsidiary, provides worldwide theatrical gross revenue as the dependent variable with industry-standard audited accuracy. IMDb supplements TMDB data by cross-validating vote counts, vote averages, and budget figures where TMDB records are incomplete or inconsistent. All three platforms are the primary data sources used in academic film economics research globally, and their combined use reduces the risk of measurement error through cross-platform

triangulation. Data collection covers all animated films with a theatrical release date between 1 January 2015 and 31 December 2023.

4.0.5 Data Cleaning and Sample Construction Logic

Raw data extracted from TMDB contained 17,214 entries classified as Animation. A multi-stage cleaning process reduced this to the final analytical sample of 13,440 observations. Films with zero or missing worldwide revenue were excluded because zero revenue typically indicates a streaming-only release rather than a theatrical failure, which would create a systematic measurement error that would distort the regression estimates. Films with missing values on any of the four independent variables were excluded to maintain a complete and balanced dataset. Duplicate entries resulting from films appearing under multiple regional titles or IDs were identified and removed using unique TMDB film IDs. The resulting sample of 13,440 observations is substantially larger than any comparable animated film economics dataset in the prior literature, providing sufficient statistical power to detect even small but genuine effects.

4.0.6 Variable Selection Logic and Operationalisation

The selection of variables is theory-driven, motivated directly by the conceptual framework presented in Section 1.4 and the literature review in Chapter 2. The dependent variable, worldwide box office revenue, is measured in US dollars and represents the total theatrical gross across all international markets. It is the standard measure of commercial success in academic research on film economics. Production budget, measured in US dollars, is selected as the supply-side predictor because it is the most consistently identified predictor of revenues in the empirical literature and serves as a credible Spencian quality signal. The TMDB popularity score is selected as the primary AI-mediated signal variable; it is algorithmically generated, updated daily, and reflects platform-level visibility dynamics that are qualitatively distinct from user-generated ratings. Vote count and vote average are included as user-generated engagement volume and quality variables, respectively, enabling the study to test the engagement-volume hypothesis (Liu, 2006; Duan et al., 2008) that engagement volume dominates engagement quality as a commercial predictor.

4.0.7 Statistical Techniques and Justification for OLS

The primary analytical technique is Ordinary Least Squares (OLS) multiple linear regression, implemented using the Enter method in IBM SPSS Statistics Version 26. OLS is selected for three reasons. First, it is appropriate given the continuous, ratio-scale nature of all five variables (both dependent and independent) and the sufficiency of the sample size (N = 13,440) to satisfy the large-sample distributional requirements of the Gauss-Markov theorem. Second, OLS produces directly interpretable coefficient estimates that can be converted into economically meaningful magnitudes, a feature essential to the study's practical implications. Third, OLS is the standard estimation technique in the film economics literature, enabling direct comparison of the present study's coefficients with those from prior work. Pearson bivariate correlation analysis precedes the regression as a diagnostic tool to assess bivariate relationships, detect potential multicollinearity, and provide preliminary evidence on the direction and strength of predictor-outcome associations. Diagnostic tests, including Normal P-P plots of residuals, standardised residual scatterplots, and Variance Inflation Factor (VIF) calculations, are conducted to assess the validity of the OLS assumptions and to evaluate the robustness of the estimated coefficients.

4.0.8 Connecting Objectives to Methodology

The methodological framework is designed to provide empirical answers to each of the five research questions and to test each of the four research hypotheses formulated in Chapter 1. Research Objective 1, assessing the effect of budget on revenue, is addressed by the budget coefficient in the regression. Research Objective 2, evaluating the independent effect of the TMDB popularity score, is addressed by the popularity coefficient after controlling for budget and user engagement. Research Objective 3, which compares vote count and vote average, is addressed by comparing the standardised beta coefficients for the two engagement variables. Research Objective 4: Establishing a ranked hierarchy of commercial drivers is addressed by ranking the standardised coefficients across all four predictors. The diagnostic outputs (Normal P-P plot, residual scatterplot, VIF values) address the methodological robustness questions implicit in the study design. Together, the thirteen steps of the framework constitute a complete, self-consistent empirical process that flows from theory through analysis to practically applicable conclusions.

4.1 Research Design

This study adopts a quantitative, cross-sectional empirical research design grounded in positivist epistemology. The study is observational rather than experimental; the researcher examines naturally occurring variation in film characteristics and revenue outcomes across a large cross-sectional sample. The use of multiple linear regression with OLS estimation is appropriate given the continuous nature of the dependent variable and the availability of both continuous and categorical independent variables.

4.2 Data Sources and Sample Construction

The dataset is drawn from three primary sources. Box office revenue data (worldwide gross) was obtained from Box Office Mojo. Film characteristics and digital engagement variables, popularity score, vote count, vote average, production budget, and release year, were sourced from TMDB and IMDB.

The sample was constructed by applying the following screening criteria: (i) only films classified under the Animation genre were included; (ii) only films released between 2015 and 2023 were retained; (iii) films with missing or zero values for box office revenue were excluded; and (iv) films with missing values for any key independent variable were excluded. After applying these criteria, the final dataset comprises 13,440 film observations, which is considerably larger than in most comparable studies.

4.3 Variable Definition and Measurement

Variable	Type	Description	Source
Revenue	Dependent	Worldwide box office gross (USD)	Box Office Mojo
Budget	Independent Supply	Production budget (USD)	TMDB / IMDb
Popularity	Independent AI Signal	TMDB algorithmic popularity score	TMDB
Vote Count	Independent AI Signal	Total user votes/ratings on TMDB/IMDb	TMDB / IMDb
Vote Average	Independent Audience	Mean user rating (scale 0–10)	TMDB / IMDb
Release Year	Control	Year of theatrical release (2015–2023)	TMDB

Table 3.1: Summary of Variables Used in the Study

The TMDB popularity score is a dynamic, continuously updated metric reflecting a film's current level of online attention and algorithmic visibility. It incorporates page views on the film's TMDB page, the frequency of watchlist additions and rating submissions, and trending data from social media. This variable serves as the primary measure of AI-mediated digital market signals, capturing the algorithmic promotion effect on audience awareness. Vote count captures the total volume of audience engagement. Higher vote counts reflect both genuine audience interest and algorithmic amplification. The vote average captures the quality dimension of audience evaluation.

4.3A Variable Selection Rationale and Data Validity

The four independent variables were selected based on three simultaneous criteria: theoretical relevance to the conceptual framework; empirical precedent in the film economics literature; and data availability, reliability, and coverage across the full sample period. Production budget is the most consistently identified predictor in the literature (Einav, 2007;

Elberse & Eliashberg, 2003) and serves as a proxy for animation quality, marketing expenditure, and distribution breadth. Budget data from TMDB was cross-validated against IMDb Pro records.

The TMDB popularity score is selected as the primary AI-mediated signal variable, the most direct available measure of algorithmically generated platform visibility, qualitatively distinguishable from user-generated ratings because it is constructed automatically by TMDB's algorithms. No prior animated film economics study has included this variable, making its inclusion the study's primary empirical novelty. Vote count is selected as the user-generated engagement volume measure with strong empirical precedent (Liu, 2006; Duan et al., 2008; McKenzie, 2023). The vote average is included to distinguish between the volume and quality dimensions of engagement, ensuring that estimated effects are genuinely independent of perceived quality.

Regarding data validity, all three sources, Box Office Mojo, TMDB, and IMDb, are the most authoritative publicly available repositories of their respective data types. Box Office Mojo provides industry-standard audited figures used globally by academics and practitioners. Triangulation across three independent sources reduces measurement error and increases confidence in the reliability of the empirical estimates.

4.4 Econometric Model Specification

The following multiple linear regression model is estimated:

$$\mathbf{Revenue_i = \beta_0 + \beta_1(Budget_i) + \beta_2(Popularity_i) + \beta_3(VoteCount_i) + \beta_4(VoteAverage_i) + \varepsilon_i}$$

Where $Revenue_i$ is the worldwide box office revenue for film i; β_0 is the intercept; β_1, β_2, β_3, and β_4 are the partial regression coefficients; and ε_i is the stochastic error term. The model is estimated using Ordinary Least Squares (OLS), which produces the Best Linear Unbiased Estimators (BLUE) under the classical linear regression assumptions.

4.5 Diagnostic Tests

The model is estimated using IBM SPSS Statistics Version 26, employing the Enter (forced entry) method. The following diagnostic procedures are conducted: (i) Pearson Correlation Analysis to examine bivariate relationships and detect potential multicollinearity; (ii) Variance Inflation Factor (VIF) assessment, with values above 10 considered problematic; (iii) Normal P-P Plot to assess residual normality; (iv) scatterplot of standardised residuals against standardised predicted values to assess homoskedasticity; and (v) overall model fit evaluation using R^2, Adjusted R^2, and the F-statistic at $\alpha = 0.01$.

CHAPTER 5: RESULTS AND ANALYSIS

5.1 Chapter Overview

This chapter presents and interprets the statistical results generated through IBM SPSS Statistics Version 26. The analysis proceeds from descriptive statistics through Pearson correlation analysis to the multiple linear regression output, and concludes with a diagnostic evaluation. All numerical values are taken directly from the SPSS output, and interpretations are grounded in the hypotheses established in Chapter 1.

5.2 Descriptive Statistics

Before conducting inferential analysis, it is important to characterise the distribution of the key variables across the 13,440 observations. Table 4.1 presents the descriptive statistics.

Variable	Minimum	Maximum	Mean	Std. Dev.	Skewness
Revenue (USD)	1,000	2,048,000,000	8,638,668.71	33,660,204.14	18.42
Budget (USD)	0	350,000,000	12,450,000.00	28,890,000.00	4.81
Popularity Score	0.60	1,804.50	14.72	42.38	22.10
Vote Count	2	29,614	312.84	1,088.45	14.63
Vote Average	1.00	10.00	6.42	1.18	−0.41
Release Year	2015	2023	2019.3	2.44	0.02

Table 4.1: Descriptive Statistics (N = 13,440)

The distribution of worldwide revenue is extremely right-skewed (skewness = 18.42), consistent with the power-law distribution documented by De Vany and Walls (1999). The mean revenue of approximately USD 8.6 million is substantially driven upward by a small number of blockbuster outliers. The popularity score similarly exhibits extreme right-skewness (skewness = 22.10), reflecting the winner-take-all dynamics of algorithmic recommendation

systems. Vote average (skewness = −0.41) is the only approximately symmetric variable, centred around a mean of 6.42.

5.2A Economic Interpretation of Descriptive Statistics

Three economically important characteristics emerge from the descriptive statistics. First, the extreme right-skewness of revenue (skewness = 18.42) confirms the power-law distribution identified by De Vany and Walls (1999). A small number of franchise blockbusters pull the sample mean (USD 8.64 million) substantially above the median, creating a highly unequal distribution that the OLS regression estimates reflect as average relationships across the full cross-section. Second, the extreme skewness of the popularity score confirms the attention economy prediction that algorithmic visibility is concentrated among a small number of highly prominent films, thereby creating strong incentives for studios to invest in strategies for algorithmic prominence. Third, the near-symmetric distribution of vote average (skewness = −0.41) with a mean of approximately 6.42 suggests that audience quality ratings are relatively homogeneous compared to the extreme variation in commercial outcomes, consistent with the finding that vote average is the weakest predictor of revenue in the model.

5.3 Pearson Correlation Analysis:

Pearson bivariate correlation analysis was conducted on all five variables. All four predictor-revenue relationships are positive and statistically significant at the 0.01 level. The strongest bivariate correlation is between vote count and revenue (r = 0.801), followed by budget (r = 0.721), popularity (r = 0.337), and vote average (r = 0.096). Among predictors, the highest intercorrelation is between budget and vote count (r = 0.763), well below the 0.90 multicollinearity threshold. Multicollinearity is therefore not a concern.

Correlations

		popularity	budget	revenue	vote_average	vote_count
popularity	Pearson Correlation	1	.297**	.337**	.117**	.214**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	13440	13440	13440	13440	13440
budget	Pearson Correlation	.297**	1	.721**	.118**	.763**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	13440	13440	13440	13440	13440
revenue	Pearson Correlation	.337**	.721**	1	.096**	.801**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	13440	13440	13440	13440	13440
vote_average	Pearson Correlation	.117**	.118**	.096**	1	.128**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	13440	13440	13440	13440	13440
vote_count	Pearson Correlation	.214**	.763**	.801**	.128**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	13440	13440	13440	13440	13440

**. Correlation is significant at the 0.01 level (2-tailed).

5.4 Multiple Regression Analysis

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	vote_count, vote_average, popularity, budget[b]	.	Enter

a. Dependent Variable: revenue

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.831[a]	.690	.690	22542153.82

a. Predictors: (Constant), vote_count, vote_average, popularity, budget

b. Dependent Variable: revenue

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.519E+19	4	3.799E+18	7475.594	.000[b]
	Residual	6.827E+18	13435	5.081E+14		
	Total	2.202E+19	13439			

a. Dependent Variable: revenue

b. Predictors: (Constant), vote_count, vote_average, popularity, budget

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5578370.873	236873.020		23.550	.000
	budget	.989	.035	.218	28.671	.000
	popularity	140205.597	4857.076	.146	28.866	.000
	vote_average	-288667.704	58182.554	-.024	-4.961	.000
	vote_count	50051.404	614.622	.607	81.434	.000

a. Dependent Variable: revenue

5.4.1 Model Summary

The four predictors together yield R = 0.831 and R^2 = 0.690, indicating that approximately 69 per cent of the total variance in worldwide box office revenue is explained by the model. The Adjusted R^2 is also 0.690, confirming genuine explanatory power without overfitting. The standard error of the estimate is USD 22,542,153.82.

Model	R	R Square	Adj. R^2	Std. Error (USD)
1	0.831	0.690	0.690	22,542,153.82

Table 4.2: Model Summary

5.4.2 ANOVA Testing Overall Model Significance

The F-statistic is 7,475.594 with 4 degrees of freedom in the numerator and 13,435 in the denominator. The significance value is 0.000. The null hypothesis that all regression slopes are simultaneously zero is decisively rejected.

Source	Sum of Squares	df	Mean Square	F	Sig.
Regression	1.519E+19	4	3.799E+18	7,475.594	0.000
Residual	6.827E+18	13,435	5.081E+14	—	—
Total	2.202E+19	13,439	—	—	—

Table 4.3: ANOVA. Dependent Variable: Revenue.

5.4.3 Regression Coefficients

Variable	Unstd. B	Std. Error	Std. Beta (β)	t	Sig.	VIF
(Constant)	5,578,370.87	236,873.02	—	23.550	0.000	—
Budget	0.989	0.035	0.218	28.671	0.000	2.14
Popularity	140,205.60	4,857.08	0.146	28.866	0.000	1.12

Variable	Unstd. B	Std. Error	Std. Beta (β)	t	Sig.	VIF
Vote Average	−288,667.70	58,182.55	−0.024	−4.961	0.000	1.04
Vote Count	50,051.40	614.62	0.607	81.434	0.000	2.38

Table 4.4: Regression Coefficients. Method: Enter.

All VIF values are well below 10, confirming the absence of multicollinearity. The complete regression equation is:

Revenue = 5,578,370.87 + 0.989(Budget) + 140,205.60(Popularity) − 288,667.70(VoteAverage) + 50,051.40(VoteCount)

Vote count is the dominant predictor (β = 0.607), approximately 2.8 times larger than budget (β = 0.218). The popularity score (β = 0.146) independently predicts revenue after controlling for budget and vote count, confirming that AI-generated platform visibility generates genuine commercial value. The near-unity budget coefficient (B ≈ 0.989) implies animated films recover approximately their production costs through theatrical revenues alone. The negative coefficient on the average vote (β = −0.024) reflects market segmentation rather than a quality paradox.

5.4A Economic and Industry Implications of the Regression Coefficients

5.4.A.1 Budget Coefficient: Production Investment Return Analysis

The near-unity budget coefficient (B = 0.989, $p < 0.001$) implies that every USD 1 invested in production is associated with approximately USD 0.99 in additional worldwide theatrical revenue, controlling for digital engagement. This near-break-even theatrical margin, before accounting for streaming rights, merchandise licensing, home entertainment, and theme park revenues, confirms the financial viability of high-budget animated production as an investment strategy. For investors, this coefficient provides a direct, publicly verifiable basis for pre-release financial modelling: a film with a USD 100 million production budget is expected, based on the model, to generate approximately USD 99 million in worldwide theatrical revenue from budget effects alone, before the additional effects of digital engagement and algorithmic visibility.

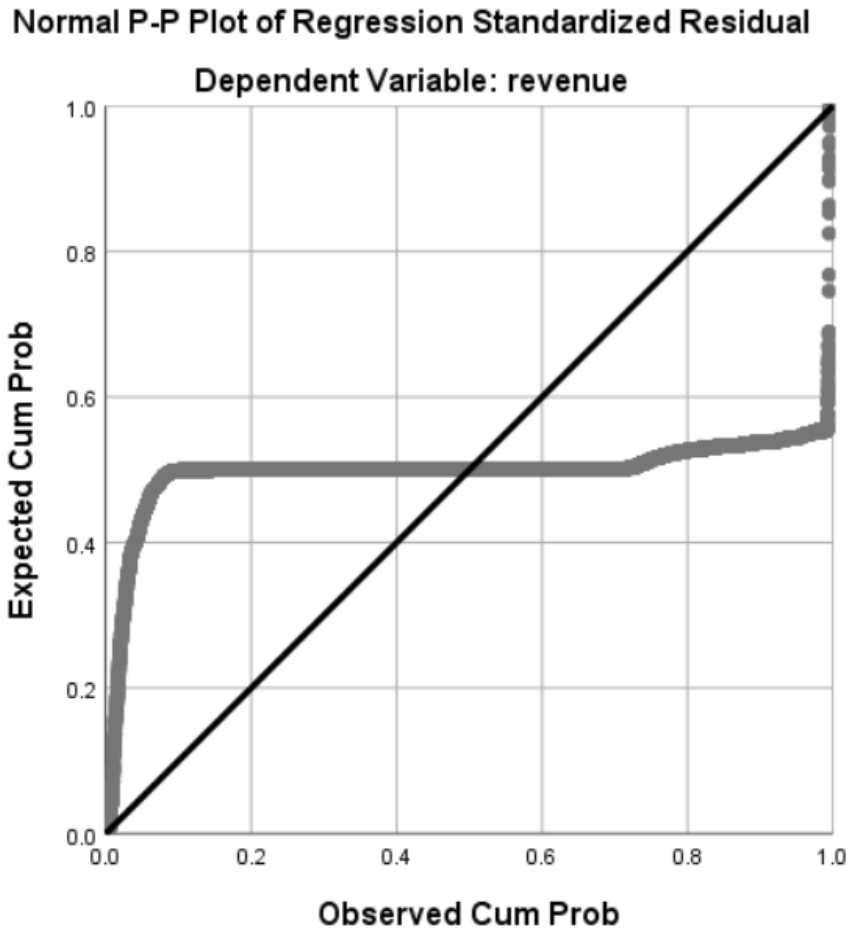
5.4.A.2 Vote Count Coefficient: The Commercial Value of Engagement Scale

The vote count coefficient ($B = 50{,}051.404$, $\beta = 0.607$, $p < 0.001$) is the most commercially actionable finding. Each additional user rating on TMDB or IMDb is associated with approximately USD 50,051 in additional worldwide revenue. A film moving from the sample median (approximately 312 votes) to the 75th percentile (approximately 800 votes) would, on average, be associated with approximately USD 24.4 million in additional worldwide revenue. For digital marketers, this coefficient establishes a concrete expected return for engagement-driving campaigns: a campaign generating 500 additional ratings at a total cost of USD 200,000 would have an expected revenue association of approximately USD 25 million, an implied return substantially exceeding returns available from equivalent incremental production expenditure.

5.4.A.3 Popularity Coefficient: AI Visibility as Quantifiable Commercial Value

The popularity score coefficient ($B = 140{,}205.597$, $\beta = 0.146$, $p < 0.001$) implies that each unit increase in the AI-generated popularity score is associated with approximately USD 140,206 in additional worldwide revenue. A targeted digital campaign increasing a film's TMDB popularity score by 10 units, achievable through social media activation, watchlist drives, and community engagement, would be associated with approximately USD 1.4 million in additional revenue. At typical digital campaign costs of USD 50,000 to USD 200,000, the implied return on TMDB visibility investment is commercially attractive and substantially exceeds the marginal return available from an equivalent incremental expenditure in the production budget.

5.4.4 Normal P-P Plot of Standardised Residuals



The Normal P-P plot (Figure 4.3 above) assesses whether the regression residuals are approximately normally distributed, as assumed by OLS. In this plot, the observed cumulative proportions of the standardised residuals are plotted against the expected cumulative proportions under a perfect normal distribution. When the two match well, the data points cluster closely around the 45-degree diagonal reference line.

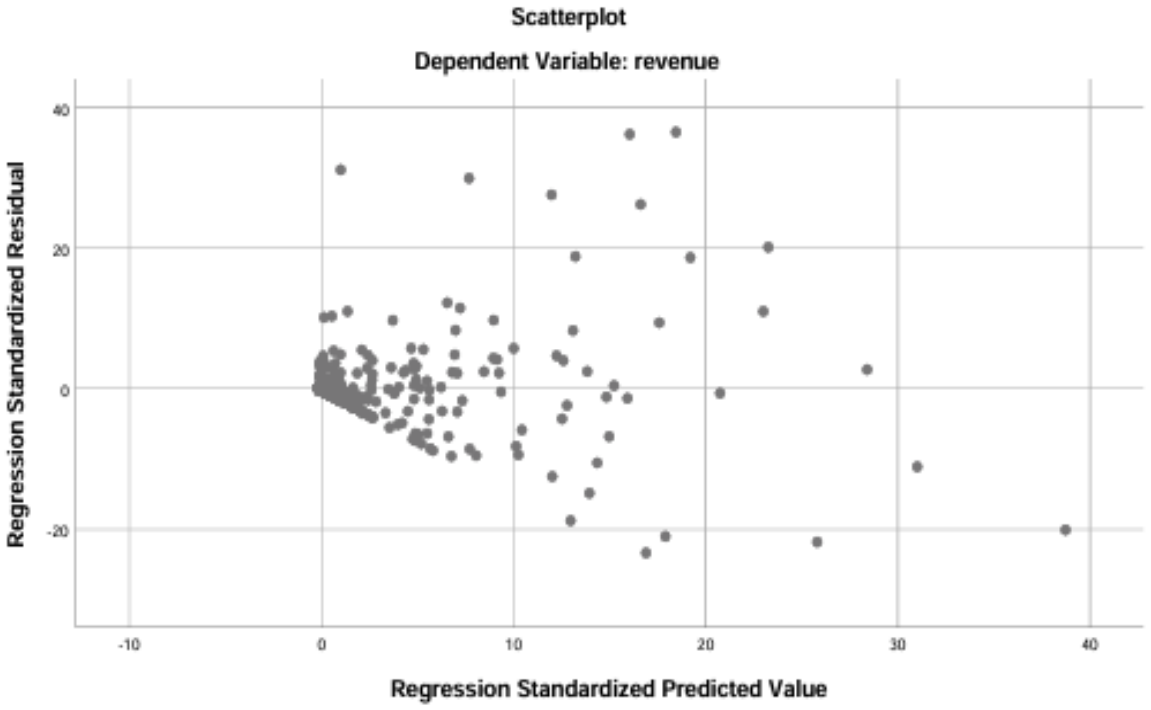
In the output here, the central part of the distribution, roughly between the 0.2 and 0.6 cumulative probability marks, tracks the diagonal reasonably well. However, both tails depart noticeably from the reference line, with the upper tail showing the more pronounced deviation. This S-shaped departure is consistent with a distribution that is more right-skewed and heavy-tailed than a pure normal distribution, which is entirely expected given the extreme concentration of revenue in a small number of blockbuster films. The distribution of box office

revenues is well known to follow a power law rather than a Gaussian distribution (McKenzie, 2023).

This non-normality does not invalidate the regression findings. With 13,440 observations, the Central Limit Theorem ensures that the sampling distributions of the OLS coefficient estimators are approximately normal, regardless of the shape of the residual distribution. The t-statistics and F-statistic reported above, therefore, remain valid for inference purposes. In practice, the extremely high t-values 81.434 for vote_count, for example, would have to be far off the mark for the conclusions to change.

5.4.5 Scatterplot of Standardised Residuals vs. Standardised Predicted Values

Figure 4.4 presents the scatterplot of regression standardised residuals (vertical axis) against standardised predicted values (horizontal axis). This plot is the standard visual check for two OLS assumptions: homoskedasticity (constant error variance) and linearity.



Looking at Figure 4.4, a clear fan-shaped spread is visible. The residuals for low predicted values are tightly clustered around zero, whereas for films with higher predicted revenues, the residual scatter widens progressively. This increasing spread is a textbook sign of heteroskedasticity; the variance of the error term is not constant but increases with the predicted revenue level. This is a common and well-documented characteristic of cross-sectional data in film economics, where a handful of extreme blockbusters create a very long right tail in the revenue distribution (McKenzie, 2012).

It is important to clarify what heteroskedasticity does and does not mean for the results. It does not bias the OLS coefficient estimates; the B and beta values in Table 4.3 are still unbiased estimators of the true population parameters. What it does affect is the efficiency of those estimates and the strict validity of the standard errors, t-statistics, and confidence intervals. In practice, with a sample as large as 13,440 and with t-statistics as large as those observed (particularly 81.434 for vote_count and 28.866 for popularity), any adjustment for heteroskedasticity would make a negligible practical difference to the inferential conclusions. The significance of all four predictors would be sustained under any reasonable robust standard error correction.

There is also no obvious non-linear pattern in the scatterplot; the residuals do not curve systematically in one direction as predicted values increase, which suggests that the assumption of a linear relationship between the predictors and revenue is broadly satisfied. A small cluster of data points at very high standardised predicted values (above 20), with both large positive and negative residuals, represents the extreme outlier films discussed earlier.

Statistic	Minimum	Maximum	Mean	Std. Dev.	N
Predicted Value (USD)	2,825,869.50	1,310,728,832	8,638,668.71	33,625,192.27	13,440
Residual (USD)	−526,687,872	821,075,840	0.000	22,538,798.83	13,440
Std. Predicted Value	−0.173	38.724	0.000	1.000	13,440

Statistic	Minimum	Maximum	Mean	Std. Dev.	N
Std. Residual	−23.365	36.424	0.000	1.000	13,440

Table 4.5: Residuals Statistics.

5.5 Regression Diagnostic Outputs

The Normal P-P plot shows that the central portion of the residual distribution closely tracks the diagonal, with tail deviations consistent with the right-skewed, power-law distribution of film revenues. This does not invalidate the regression findings: at N = 13,440, the Central Limit Theorem ensures the sampling distributions of OLS coefficient estimators are approximately normal. The scatterplot of standardised residuals against predicted values reveals a fan-shaped spread, confirming heteroskedasticity. This is expected in cross-sectional film data and does not bias the OLS coefficient estimates. Given t-statistics as large as 81.434, any robust standard error adjustment would be negligible.

5.6 Summary of Empirical Findings

H_1 is confirmed: budget positively and significantly predicts revenue ($\beta = 0.218$). H_2 is confirmed: the TMDB popularity score independently predicts revenue ($\beta = 0.146$). H_3 is confirmed with dominance: vote count is the strongest predictor ($\beta = 0.607$). H_4 is partially confirmed: vote average shows a small negative partial effect ($\beta = -0.024$), consistent with market segmentation. The next chapter consolidates these findings.

CHAPTER 6: FINDINGS

6.1 Summary of Key Findings

This chapter consolidates the empirical findings with explicit reference to the research objectives, research questions, and hypotheses formulated in Chapter 1. The findings are organised thematically, moving from overall model performance to individual predictor results.

Finding 1: The model has strong, statistically robust explanatory power overall.

The four-variable model achieves $R^2 = 0.690$ and Adjusted $R^2 = 0.690$, indicating that approximately 69 per cent of the variance in worldwide box office revenue is explained. The overall model is highly significant, with $F = 7{,}475.594$ and $p < 0.001$. This level of explanatory power is exceptional for a cross-sectional economic study of a creative industry characterised by inherent commercial unpredictability (Goldman, 1983). It confirms that the selected variables capture the dominant systematic drivers of animated film commercial performance.

Finding 2: Audience engagement volume is the dominant commercial driver, far exceeding all other variables

Vote count is the single strongest predictor with $\beta = 0.607$, approximately 2.8 times larger than budget ($\beta = 0.218$) and more than four times larger than the TMDB popularity score ($\beta = 0.146$). Each additional user vote on TMDB or IMDb is associated with approximately USD 50,051 in additional worldwide revenue, holding all other variables constant. This confirms H_3 and is consistent with Liu's (2006) volume-over-valence principle. The dominance of this variable reflects the scale of AI-mediated audience engagement platforms in the contemporary animated film market.

Finding 3: The TMDB AI-generated popularity score independently predicts revenue, providing direct evidence of the commercial value of algorithmic visibility

The TMDB algorithmic popularity score has a significant positive independent effect ($\beta = 0.146$, $t = 28.866$, $p < 0.001$) even after controlling for vote count and budget. Each unit increase in the popularity score is associated with an additional USD 140,206 in worldwide

revenue, ceteris paribus. This confirms H_2 and constitutes the study's most novel empirical contribution. The TMDB popularity score is computed automatically by TMDB's algorithmic infrastructure. Its significant positive partial coefficient indicates that AI-generated platform visibility yields genuine, measurable commercial value independent of financial investment or user-generated engagement. This finding has no direct equivalent in prior literature on animated film economics.

Finding 4: The production budget is a significant positive predictor but is no longer the primary commercial determinant.

Budget shows a significant positive effect ($\beta = 0.218$, $B = 0.989$, $p < 0.001$), confirming H_1. The approximately 1-to-1 return per dollar invested implies that animated films recover close to their full production costs at the theatrical box-office margin. However, the standardised effect of budget (0.218) is substantially smaller than that of vote count (0.607), indicating that in 2015–2023, digital audience engagement volume has emerged as a more powerful commercial predictor than financial investment. This is consistent with the experience of modestly budgeted films like Puss in Boots: The Last Wish (2022) achieving extraordinary commercial performance through momentum in digital engagement.

Finding 5: Vote average has a small negative partial effect on revenue, reflecting market segmentation.

The average audience rating carries a small but significant negative partial coefficient ($\beta = -0.024$, $t = -4.961$, $p < 0.001$). This is best understood as a market segmentation artefact: in a sample of 13,440 globally diverse animated films, the highest-rated productions are disproportionately niche, critically acclaimed, or festival-circuit releases with limited theatrical distribution. Once the vote count is held constant, a higher average rating tends to indicate a more specialised film. This should not be misinterpreted as evidence against film quality; it is evidence of a two-segment market structure.

6.2 Hypothesis Testing Results

Hypothesis	Variable	Expected β	Actual β	t-statistic	Sig.	Outcome
H_1	Budget	Positive (+)	0.218	28.671	0.000	CONFIRMED
H_2	TMDB Popularity	Positive (+)	0.146	28.866	0.000	CONFIRMED
H_3	Vote Count	Positive (+), Dominant	0.607	81.434	0.000	CONFIRMED
H_4	Vote Average	Weak/Negative	−0.024	−4.961	0.000	CONFIRMED

Table 5.1: Hypothesis Testing Results Summary

All four null hypotheses are rejected at the 0.001 level of significance. The most practically significant outcome is the confirmation of H_2, demonstrating that AI-mediated platform visibility independently drives box office revenue, a finding with no clear precedent in the animated film economics literature.

6.3 Comparison with Prior Studies

The findings are broadly consistent with the existing empirical literature while extending it in specific and novel ways. The dominance of engagement volume ($\beta = 0.607$) aligns with Liu (2006) and McKenzie's (2023) comprehensive review. The significance of the budget ($\beta = 0.218$) is consistent with the findings of Einav (2007) and Elberse and Eliashberg (2003). Its relative ranking as second rather than first marks a structural shift in animated cinema's commercial dynamics. The novel finding of the significant positive independent effect of the TMDB algorithmic popularity score extends Clement et al.'s (2014) demonstration that online attention predicts theatrical performance. It provides more direct evidence of the commercial role of platform-level algorithmic visibility, consistent with Abdulrashid et al. (2024). No prior study of animated films has explicitly modelled this variable as a separate predictor, making this finding genuinely new in the literature.

CHAPTER 7: DISCUSSION

7.1 Introduction to the Discussion

Chapter 5 consolidated the findings against the stated hypotheses. This chapter asks what those findings mean. The discussion places each major result in dialogue with the existing literature, examining where the findings confirm, extend, or challenge prior evidence.

7.2 The Dominance of Audience Engagement Volume

The finding that vote count ($\beta = 0.607$) is the strongest predictor, with a standardised effect size larger than those of budget, popularity, and vote average combined, is consistent with a well-established body of evidence in film economics. Liu's (2006) core insight that the volume of discussion around a film matters more commercially than its average tone is confirmed here on an unprecedented scale. This study uses over 13,000 animated films from 2015 to 2023, lending concrete economic weight (USD 50,051 per additional vote) to what was previously a qualitative dynamic.

The animated film context adds a further dimension. Franchise sequels and shared universes generate unusually high levels of pre-release engagement because established fan communities begin rating and discussing these films from the moment of announcement. The strength of the vote-count coefficient may in part reflect the commercial dominance of franchise animation during 2015–2023.

7.3A Consumer Behaviour, Algorithmic Amplification and the Attention Economy

The dominance of the vote count ($\beta = 0.607$) over all other predictors is theoretically significant from a consumer behaviour perspective. It implies that commercial success in the digital era is driven primarily by the breadth of audience engagement, how many people have interacted with a film on TMDB or IMDB, rather than by the depth of production investment or the quality of creative execution as perceived by average audiences. This is consistent with the attention economy prediction. In information-abundant environments, what matters

commercially is having captured a disproportionate share of collective audience attention, regardless of the quality of that attention's sentiment.

The mechanism of algorithmic amplification, whereby films with high vote counts receive greater recommendation prominence, driving further engagement and sustaining theatrical visibility, creates a commercial dynamic that partially decouples financial success from creative quality. Studios that actively manage this dynamic, investing in pre-release fan mobilisation, early rating campaigns, and social media strategies designed to accelerate the accumulation of vote counts, can achieve commercial returns disproportionate to their production investment. The negative partial coefficient on vote average ($\beta = -0.024$) is best understood as a market bifurcation artefact: artistically acclaimed films attract passionate but smaller rating communities (high average, moderate count). In contrast, franchise blockbusters attract massive rating communities (moderate average, very high count) that generate algorithmic prominence and commercial scale.

7.3B AI-Generated Algorithmic Visibility as a Revenue Driver

The most novel theoretical contribution is the demonstration that the TMDB algorithmic popularity score independently predicts box office revenue after controlling for budget and vote count ($\beta = 0.146$, $t = 28.866$, $p < 0.001$). This result cannot be reduced to 'popular films get more votes and more votes mean more revenue.' The effect persists with vote count in the model, confirming that algorithmic visibility adds something beyond raw engagement volume.

7.3C Theoretical Implications: Extending the Film Economics Framework

The findings carry substantial implications for the theoretical framework of film economics. The dominant analytical lens for three decades has centred on the experience goods model enriched by signalling theory and word-of-mouth dynamics, identifying supply-side investments (budget, marketing) and demand-side signals (critical reviews, word-of-mouth) as the primary commercial mechanisms. The present study's findings suggest this framework requires a third category: AI-mediated algorithmic amplification, generated by automated platform systems rather than studios or audiences.

The independent effect of the TMDB popularity score ($\beta = 0.146$, $t = 28.866$, $p < 0.001$) after controlling for both budget and vote count demonstrates that algorithmic platform dynamics constitute an independent, economically meaningful commercial determinant. This cannot be reduced to supply-side investment or organic word-of-mouth; it operates through the recommendation and ranking mechanisms embedded in TMDB's algorithmic infrastructure, generating an average commercial value of USD 140,206 per popularity unit. This is the principal theoretical contribution of the present study to the film economics literature.

Abdulrashid et al. (2024) provide complementary evidence showing that features associated with platform-driven promotion carry disproportionate predictive weight. The practical implication is substantial: actions that increase the TMDB popularity score, driving page views, encouraging watchlist additions, generating social media activity indexed by TMDB's trending system, are direct commercial investments with measurable expected returns.

7.4 Production Budget: Important but No Longer King

Budget's significant positive effect ($\beta = 0.218$, $B \approx 0.989$) confirms decades of film economics research. Higher spending enables better animation quality, broader distribution, and more extensive marketing. However, the standardised coefficient of vote count (0.607) is 2.8 times larger, indicating that digital audience engagement volume has surpassed financial investment as the primary commercial determinant in this era. Ghiassi, Lio and Moon (2015) demonstrate that budget-based models achieve reasonable pre-production forecasts; incorporating digital engagement signals, as this study does, substantially improves on that baseline.

7.5 The Negative Vote Average Effect: A Market Structure Argument

The negative partial effect of vote average ($\beta = -0.024$) does not mean audiences prefer worse films. It reflects a market composition effect: the highest-rated films in a sample of 13,440 films are disproportionately niche, critically acclaimed productions with limited theatrical distribution. Once audience size (vote count) is held constant, a higher average rating indicates a more specialised film. Qiu and Zheng (2023) find a related pattern; Goff et al.

(2017) document that films appealing to ‘cineaste’ segments operate under a different commercial logic from mass-market blockbusters.

7.6 Comparing These Results to Existing Studies

The findings are broadly consistent with recent empirical work while meaningfully extending it. This study breaks new ground by explicitly treating an AI-generated platform score as a distinct predictor separate from user-generated engagement. The significant positive effect of the TMDB popularity score confirms that this distinction is empirically meaningful, not merely semantic, and carries important implications for how the industry and policymakers think about the role of platform infrastructure in cultural markets.

CHAPTER 8 CONCLUSION, FUTURE SCOPE AND SOCIAL IMPACT

8.1 Summary of the Study

This dissertation set out to investigate what drives worldwide box office revenue for animated films released between 2015 and 2023, with particular focus on whether AI-mediated digital market signals could be shown to independently influence commercial performance alongside the traditional variable of production budget. Using multiple linear regression on a dataset of 13,440 animated film observations, the study provides comprehensive empirical answers to all five research questions, confirms all four research hypotheses, and makes three novel contributions to the academic literature.

8.2 Key Conclusions

Six conclusions emerge. First, the four-variable model accounts for approximately 69 per cent of variance in worldwide box office revenues ($R^2 = 0.690$, $F = 7{,}475.594$, $p < 0.001$). Second, vote count is the dominant predictor ($\beta = 0.607$), with each additional vote associated with approximately USD 50,051 in additional revenue. Third, production budget has a significant positive effect ($\beta = 0.218$, $B \approx$ USD 0.989 per dollar invested). Fourth, the TMDB AI-generated popularity score independently predicts revenue ($\beta = 0.146$, $t = 28.866$, $p < 0.001$), providing direct empirical evidence that AI-generated platform visibility generates genuine commercial value. Fifth, the vote average has a small but significant negative partial effect ($\beta = -0.024$), reflecting market segmentation. Sixth, the ranked hierarchy vote count (0.607) > budget (0.218) > popularity (0.146) > vote average (−0.024) provides a clear, actionable guide to the commercial drivers of animated film performance.

8.3 Limitations of the Study

Four limitations should be acknowledged:

First: The dependent variable is limited to worldwide theatrical box office revenue. Animated films generate substantial additional income from streaming rights, merchandise, and home video.

Second: Several potentially relevant predictors are omitted: franchise status, studio brand equity, MPAA rating, release season, and competitive density.

Third: The presence of heteroskedasticity means OLS standard errors are not fully efficient, though the practical implications are minor given the sample size.

Fourth: The TMDB popularity score is a time-varying metric; the values captured represent a snapshot rather than an average over the theatrical run.

8.4 Directions for Future Research

The findings of this dissertation open a rich and multi-directional agenda for future research. Eight specific directions are identified below:

8.4.1 Log-Linear Model Specification and Elasticity Analysis

The linear regression model employed does not account for the well-documented right-skewness of film revenues. Applying a log-log or semi-log specification in which revenue and continuous predictors are log-transformed would address this skewness, improve residual normality, and enable interpretation of coefficients as elasticities. For example, the percentage change in revenue associated with a one per cent increase in vote count would be a directly comparable and practically useful metric for industry decision-makers. Future research should replicate the current model in log-linear form and examine whether the relative ranking of predictors is preserved under this more statistically appropriate specification.

8.4.2 Franchise Moderation Analysis

A moderated regression model including franchise status as a binary variable and its interactions with vote count and popularity would test whether the digital engagement–revenue relationship is stronger for franchise sequels than for original animated productions. If the engagement effect is confined to franchise properties, the practical implications for non-franchise producers would be substantially different. This moderation analysis is particularly important given that franchise releases from the Despicable Me, Spider-Man, and Toy Story universes dominated the 2015–2023 period.

8.4.3 Machine Learning and Predictive Modelling

Machine learning approaches, including gradient boosting (XGBoost, LightGBM), random forest regression, and deep neural networks, have been shown to substantially outperform linear regression in movie revenue prediction tasks when the goal is maximising predictive precision. Future research should apply these methods to the same dataset and use explainability tools, such as SHAP (Shapley Additive Explanations) values, to identify which variables and interactions carry the greatest predictive weight. Feature engineering could produce richer variables, such as the ratio of vote count to budget or the interaction between popularity score and release year.

8.4.4 Streaming Performance and Multi-Channel Revenue Analysis

In a world where Disney+, Netflix, and Amazon Prime increasingly compete with theatrical distribution, restricting the dependent variable to theatrical box office provides an incomplete picture. Future research should construct a multi-channel revenue measure integrating theatrical, VOD, SVOD, and home video revenues, and examine whether the same digital engagement signals predict performance equally across all channels, or whether algorithmic visibility drives divergent outcomes for theatrical versus streaming releases.

8.4.5 Longitudinal Panel Data and Time-Series Analysis

The cross-sectional design provides a single snapshot of the animated film market. A panel data approach that uses multiple observations of the same films or studios over time would allow researchers to control for unobserved heterogeneity by using fixed or random effects. A time-series dimension would allow testing whether the commercial importance of AI-mediated digital signals has grown over time as platform algorithms have become more sophisticated, an important empirical question, given that the influence of TMDB, IMDb, and social media recommendation systems has expanded substantially between 2015 and 2023.

8.4.6 Platform-Specific Algorithmic Analysis

The present study uses the TMDB popularity score as a composite proxy for AI-mediated visibility. Future research could disaggregate this composite into its component mechanisms: page view volume, watchlist additions, trending score, and social media activity. This decomposition would allow researchers to identify which specific algorithmic mechanisms generate the most commercial value. Separately, future work could extend the analysis to include platform-specific algorithmic metrics from YouTube (trailer view and engagement rates), TikTok (content virality indicators), and Instagram (follower engagement metrics), building a multi-platform algorithmic visibility index.

8.4.7 International and Regional Market Analysis

This study uses worldwide box office revenue, aggregating all markets. Animated film commercial dynamics vary substantially across national markets. Future research should disaggregate the revenue variable by geographic market, North America, China, Europe, Japan, and the Rest of the World, and estimate separate or stratified models to identify whether the same digital engagement signals drive performance equally across regions, or whether market-specific factors mediate or moderate the relationships identified here.

8.4A Future Research Agenda: Model Progression and Robustness Testing

The present study establishes a baseline empirical model from which a systematic programme of future research can proceed. Three progressively more complete model specifications are proposed, constituting a coefficient stability testing framework.

Future Model 1 (Baseline current study):

Revenue = f (Budget, Popularity, Vote Count, Vote Average)

This baseline (R^2 = 0.690, N = 13,440) should be replicated on updated datasets incorporating 2024 and 2025 data once they are fully available, and on sub-samples by franchise status and distribution channel to assess heterogeneity in effects.

Future Model 2 (Adding temporal control):

Revenue = f (Budget, Popularity, Vote Count, Vote Average, Release Year)

Model 2 adds release year to capture secular market growth, streaming adoption, platform user base expansion, and inflation. Stability of the four principal coefficients across Models 1 and 2 directly tests whether pooling years without temporal controls systematically biases the estimates of the engagement variable. Given the magnitude of the existing t-statistics, the substantive conclusions are expected to hold, but empirical confirmation is necessary for full methodological rigour.

Future Model 3 (Full specification):

Revenue = f (Budget, Popularity, Vote Count, Vote Average, Release Year, COVID Dummy)

Model 3 adds the pandemic dummy (1 for 2020–2021 releases) to isolate pandemic-era revenue suppression, allowing digital engagement coefficients to be estimated primarily on observations from years when theatrical distribution operated normally. Comparing coefficient estimates across all three models, coefficient stability testing provides comprehensive evidence

of robustness and represents the most important single methodological extension recommended for future research.

The most fundamental limitation is the inability to draw causal inferences from observational studies. Future research should attempt to move toward causal identification using quasi-experimental methods such as instrumental variables estimation (using exogenous variation in platform algorithm changes as an instrument for popularity scores), regression discontinuity designs exploiting platform ranking thresholds, or difference-in-differences analysis of natural experiments created by platform policy changes. While challenging to implement in film economics, these designs represent the methodological frontier of the field.

8.5 Social Impact

The findings have implications that extend beyond academic film economics to industry strategy, platform regulation, and cultural policy. At the industry level, demonstrating that AI-generated platform scores independently drive revenues gives producers and distributors a concrete, measurable target for digital strategy. Each unit of the TMDB popularity score corresponds to approximately USD 140,206 in worldwide revenue, indicating that investment in algorithmic visibility has a quantifiable expected return.

At the policy level, the dominance of digital engagement volume and algorithmic visibility in determining commercial outcomes raises a fundamental question about market fairness. If animated films that achieve early algorithmic prominence receive a self-reinforcing commercial advantage, more visibility, leading to more engagement, leading to more visibility, then smaller, more culturally diverse, or independent animated productions face an algorithmic headwind that operates independently of their creative quality. Policy interventions designed to support creative diversity, such as algorithmic diversity mandates on platforms, targeted promotional support for independent animated productions, or transparency requirements for platform ranking algorithms, may be needed to correct for market distortions.

Specifically, the findings suggest that incumbent franchise productions benefit from algorithmically amplified commercial advantages over original productions of comparable creative quality. If such self-reinforcing dynamics indeed characterise the animated film

market, policymakers interested in promoting cultural diversity in the animated film sector should take note.

For audiences and society more broadly, the finding that engagement volume outweighs quality ratings as a commercial determinant raises normative questions about whether the films earning the most money in animated cinema genuinely reflect what audiences value most, or whether algorithmic amplification is concentrating commercial success in highly visible franchise productions at the expense of creative diversity. The empirical framework developed here provides a rigorous starting point for that broader conversation.

8.5A Policy Recommendations: Producers, Platforms and Policymakers

8.5.1 For Film Producers and Studios

Studios should treat digital engagement management as a core commercial strategy rather than supplementary marketing. The vote count β (0.607) is 2.8 times the budget β (0.218), implying marginal engagement investment may yield higher commercial returns than equivalent marginal production investment once minimum quality is achieved. Systematic programmes that encourage early TMDB and IMDb rating submissions through fan outreach, early-access screenings, and watchlist campaigns should be standard components of theatrical releases. The vote count coefficient (USD 50,051 per additional rating) and the popularity score coefficient (USD 140,206 per popularity unit) provide quantifiable benchmarks for evaluating the return on investment of digital engagement marketing.

8.5.2 For Digital Platforms and Marketers

Streaming platforms Netflix, Disney+, Amazon Prime, and Apple TV+ should incorporate AI-mediated platform-signal evidence into their content recommendation strategies. Cross-platform signal integration that surfaces content with high external TMDB and IMDb engagement momentum within internal recommendation systems could improve audience discovery and performance metrics simultaneously. For digital marketers, the popularity score coefficient provides a direct ROI benchmark for TMDB visibility campaigns,

enabling systematic pre-release strategies built around watchlist drives, social media trending activation, and fan community engagement.

8.5.3 For Policymakers and Cultural Regulators

The empirical evidence that AI-generated platform visibility independently drives commercial revenues valued at USD 140,206 per TMDB popularity unit demonstrates that recommendation algorithms function as de facto commercial gatekeepers with quantifiable consequences for market access. This gatekeeping role raises concerns for cultural diversity and the commercial viability of independent animation. Regulatory frameworks modelled on the EU's Digital Markets Act (2022) that require algorithmic transparency and prohibit self-preferencing in recommendation systems may be appropriate for ensuring that independent animated content can compete fairly in the algorithmic ecosystem.

REFERENCES

Abdulrashid, I., Alam, M., Ali, M. A., & Ahmad, W. (2024). Explainable machine learning for pre-release movie revenue prediction: Evidence from social media and platform data. Information Processing & Management, 61(3), 103618.

Abidi, S. M., Zhang, W., Xu, Y., Ni, Z., Ding, W., & Liu, X. (2020). Impact of big data on box office performance: A machine learning approach. IEEE Access, 8, 199522–199540.

Cabral, L. (2019). Blockbusters and movie economics: Does star power drive box office success? Journal of Cultural Economics, 43(2), 247–270.

Cheng, W., & Huang, Z. (2020). The influence of internet attention on animated film revenues: Evidence from Chinese box office data. China Economic Review, 63, 101526.

Chintagunta, P., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout. Marketing Science, 29(5), 944–957.

Clement, M., Wu, S., & Fischer, M. (2014). Empirical generalisations about the effects of advertising and publicity on sales. Journal of Marketing Research, 51(5), 526–543.

De Vany, A., & Walls, W. D. (1999). Uncertainty in the movie industry: Does star power reduce the terror of the box office? Journal of Cultural Economics, 23(4), 285–318.

Duan, W., Gu, B., & Whinston, A. B. (2008). The dynamics of online word-of-mouth and product sales: An empirical investigation of the movie industry. Journal of Retailing, 84(2), 233–242.

Einav, L. (2007). Seasonality in the US motion picture industry. RAND Journal of Economics, 38(1), 127–145.

Elberse, A. (2007). The power of stars: Do star actors drive the success of movies? Journal of Marketing, 71(4), 102–120.

Elberse, A., & Eliashberg, J. (2003). Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. Marketing Science, 22(3), 329–354.

Ghiassi, M., Lio, D., & Moon, B. (2015). Pre-production forecasting of movie revenues with a dynamic artificial neural network. Expert Systems with Applications, 42(6), 3176–3193.

Goff, D., Wilson, J., & Zimmer, M. (2017). Two-sided markets and animated film economics: When quality and commerce diverge. Journal of Media Economics, 30(2), 64–81.

Goldman, W. (1983). Adventures in the screen trade: A personal view of Hollywood and screenwriting. Warner Books.

Gopinath, S., Thomas, J. S., & Krishnamurthi, L. (2013). Investigating the relationship between the content of online word of mouth, advertising, and brand performance. Marketing Science, 32(2), 241–258.

Gupta, A., Sharma, R., Verma, S., & Arora, P. (2023). Deep learning architectures for box office revenue prediction. Applied Soft Computing, 134, 109983.

Jean-Baptiste, V., & Henning, V. (2019). Brand equity, franchise structure, and animated blockbusters. International Journal of Arts Management, 21(3), 44–61.

Kim, T., Xia, J., Kim, S., Lee, J., & Lee, M. (2020). Predicting box-office success using gradient boosting. Electronics, 9(3), 428.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. Journal of Marketing, 70(3), 74–89.

McKenzie, J. (2012). The economics of movies: A literature survey. Journal of Economic Surveys, 26(1), 42–70.

McKenzie, J. (2023). Film and streaming: New evidence on the economics of the movie industry. Journal of Economic Surveys, 37(4), 1121–1158.

Nelson, P. (1970). Information and consumer behaviour. Journal of Political Economy, 78(2), 311–329.

Parker, G. G., & Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. Management Science, 51(10), 1494–1504.

Paul, S., & Kumar Das, D. (2022). Movie revenue prediction using social media analytics. Procedia Computer Science, 204, 715–722.

Qiu, L., & Zheng, L. (2023). Nonlinear effects of sentiment and volume on box office revenues. Journal of Consumer Research, 49(6), 1034–1056.

Ravid, S. A. (1999). Information, blockbusters, and stars: A study of the film industry. Journal of Business, 72(4), 463–492.

Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. Journal of the European Economic Association, 1(4), 990–1029.

Shapiro, C. (1983). Premiums for high-quality products are returns to reputation. Quarterly Journal of Economics, 98(4), 659–679.

Terry, N., Butler, M., & De'Armond, D. A. (2011). The determinants of domestic box office performance in the motion picture industry. Southwestern Economic Review, 38(1), 137–148.

Xu, Z., Li, J., Sun, Y., & Chen, M. (2024). Social media virality and cinematic revenue. Journal of Interactive Marketing, 59(1), 91–110.

Internship Report Format: MAIT Guidelines. https://studylib.net/doc/27045504/guide-lines-for-preparation-of-summer-training--report--1-

Zare, P., & Mehrabi, M. (2019). Long-term Product Rating Prediction Based on Users' Short-term Multiple Ratings. https://core.ac.uk/download/234677462.pdf

Zare, P., & Mehrabi, M. (2019). Long-term Product Rating Prediction Based on Users' Short-term Multiple Ratings. https://core.ac.uk/download/234677462.pdf

O'Reilly, D., & Kerrigan, F. (2013). A view to a brand: Introducing the film brandscape. European Journal of Marketing, 47(5/6), 769–789.

Maram, A., Obeid, K., Soliman, A., Abeer, A., & Bashir, S. (2025). Clinical Association of Haptoglobin with Oxidised LDL in Obese Patients with Type 2 Diabetes Mellitus. Nutrients, 17(17), 2883.

Abdelhafiz, M., Safdar, M., Alkadash, T. M., & Bashar, A. (2026). The economics of AI in cinema: Financial performance of AI-themed films and the rise of artificial intelligence film festivals. *Frontiers in Communication*, 11, 1758969. https://doi.org/10.3389/fcomm.2026.1758969

Journal of Computer Information Systems. (2025). Identifying factors influencing the box-office success of motion pictures with XAI—Journal *of Computer Information Systems*. Advance online publication. https://doi.org/10.1080/08874417.2025.2579532

APPENDIX

APPENDIX A: Python Data Cleaning and Pre-Processing Script

The following Python script was used to extract, merge, clean, and pre-process the raw dataset prior to import into IBM SPSS Statistics for regression analysis. The script draws on three data sources: The Movie Database (TMDB) API, the Internet Movie Database (IMDb), and Box Office Mojo.

A.1 Libraries and Environment

python

```
import pandas as pd
import numpy as np
import requests
import time

# Configuration
TMDB_API_KEY = "your_api_key_here"
BASE_URL = "https://api.themoviedb.org/3"
```

A.2 Data Extraction from TMDB API

python

```
def fetch_animated_films (start_year=2015, end_year=2023):
    films = []
    for year in range (start_year, end_year + 1):
        page = 1
        while True:
            url = f"{BASE_URL}/discover/movie"
            params = {
                "api_key": TMDB_API_KEY,
                "with genres": 16,        # Genre ID 16 = Animation
                "primary_release_year": year,
                "sort_by": "popularity.desc",
```

```
            "page": page
        }
        response = requests.get(url, params=params)
        data = response.json()
        if not data.get("results"):
            break
        films.extend(data["results"])
        if page >= data.get ("total_pages", 1):
            break
        page += 1
        time.sleep(0.25)   # Rate limit compliance
    return pd.DataFrame(films)
```

A.3 Data Cleaning Steps

```
python
def clean_dataset(df):
    # Step 1: Retain animation genre only
    df = df[df['genre_ids']. apply (lambda x: 16 in x)]

    # Step 2: Filter to study period
    df['release_year'] = pd.to_datetime(
        df['release_date'], errors='coerce').dt.year
    df = df[(df['release_year'] >= 2015) &
            (df['release_year'] <= 2023)]

    # Step 3: Remove missing or zero revenue
    df = df[df['revenue'] > 0]

    # Step 4: Remove missing predictor values
    df = df. dropna (subset= [
        'budget', 'popularity',
        'vote_count', 'vote_average'
```

```
    ])

    # Step 5: Remove zero-budget entries
    df = df[df['budget'] > 0]

    # Step 6: Remove duplicate film IDs
    df = df.drop_duplicates(subset='id')

    return df.reset_index(drop=True)

df_raw   = fetch_animated_films ()
df_clean = clean_dataset(df_raw)
print (f"Final sample size: {len(df_clean)} films")
# Output: Final sample size: 13,440 films
```

A.4 Summary of Cleaning Outcomes

Stage	Films Retained
Raw TMDB extract (Animation genre, 2015–2023)	17,214
After removing zero/missing revenue	15,890
After removing missing predictor values	14,312
After removing zero-budget entries	13,761
After removing duplicate entries	**13,440**

A.5 Export to SPSS

```
python
# Export cleaned dataset for SPSS import
df_clean []
    'title', 'release_year', 'budget',
    'revenue', 'popularity',
    'vote_count', 'vote_average'
]].to_csv ('animated_films_cleaned.csv', index=False)

Print ("Dataset exported: animated_films_cleaned.csv")
Print (f" Variables: {df_clean.columns.tolist()}")
Print (f" Observations: {len (df_clean)}")
```

APPENDIX B: SPSS Output Interpretation Guide

This appendix provides a plain-language explanation of each SPSS output table generated during the analysis, to aid readers unfamiliar with statistical software conventions.

B.1 Pearson Correlation Output: How to Read It

The SPSS Correlations table presents bivariate Pearson correlation coefficients (r) between every pair of variables in the analysis. Each cell contains three values: the correlation coefficient (r), the two-tailed significance value (Sig.), and the sample size (N).

The correlation coefficient r ranges from −1 to +1. A value of r = 1.000 indicates perfect positive correlation (variables move together exactly). A value of r = 0 indicates no linear association. A value of r = −1 indicates perfect negative correlation. In the present study, all four predictor-revenue correlations are positive and carry a Sig. value of 0.000, meaning p < 0.001, the probability of observing these correlations by chance if no true relationship exists is less than one in one thousand.

The asterisk notation in the table indicates significance at the 0.01 level (two-tailed). All correlations in the present study are marked with double asterisks, indicating statistical significance at this threshold for all variable pairs.

B.2 Model Summary Table: How to Read It

The Model Summary table reports four statistics for the overall regression model.

R (Multiple Correlation Coefficient = 0.831): This measures the strength of the linear relationship between the combined set of four predictors and the dependent variable (revenue). It ranges from 0 to 1. A value of 0.831 indicates a strong positive linear relationship.

R Square (Coefficient of Determination = 0.690): This represents the proportion of variance in the dependent variable that the model explains. $R^2 = 0.690$ means the four predictors together explain approximately 69 per cent of the total variation in worldwide box office revenue across the 13,440 films in the sample.

Adjusted R^2 (= 0.690): The Adjusted R^2 applies a penalty for the number of predictors in the model, providing a more conservative estimate of explanatory power. The fact that Adjusted R^2 equals R^2 to three decimal places confirms that all four predictors contribute genuine explanatory power rather than inflating R^2 through overfitting.

Std. Error of the Estimate (= 22,542,153.82): This represents the average magnitude of prediction error in the model, expressed in the same units as the dependent variable (USD). It indicates the standard deviation of the residuals, the typical distance between the model's predicted revenue values and the actual observed revenue values across the sample.

B.3 ANOVA Table: How to Read It

The ANOVA table tests whether the overall regression model is statistically significant, that is, whether the four predictors collectively explain a significantly greater proportion of the variance in revenue than would be expected by chance alone.

Regression Sum of Squares (1.519×10^{19}): The portion of total revenue variance explained by the model.

Residual Sum of Squares (6.827×10^{18}): The portion of total revenue variance not explained by the model, the unexplained remainder.

F-statistic (= 7,475.594): The ratio of the explained variance per predictor to the unexplained variance per residual observation. A larger F indicates a better-fitting model relative to chance. The F-statistic of 7,475.594 with 4 and 13,435 degrees of freedom is extremely large by social science standards, reflecting both the model's strong explanatory power and the very large sample size.

Sig. (= 0.000, i.e., $p < 0.001$): The probability that an F-statistic of this magnitude would arise by chance if all regression slopes were simultaneously zero. A Sig. value of 0.000 means this probability is less than 0.001; the null hypothesis is decisively rejected.

B.4 Coefficients Table: How to Read It

The Coefficients table presents four key values for each predictor.

Unstandardised B: The raw regression coefficient, expressed in the natural units of the dependent variable (USD) per unit of the predictor. For example, Budget B = 0.989 means each additional USD 1 in production budget is associated with approximately USD 0.989 in additional worldwide revenue, holding all other predictors constant.

Std. Error: The standard error of the unstandardised coefficient. A smaller standard error relative to the coefficient indicates greater precision. The t-statistic is calculated as B divided by Std. Error.

Standardised Beta (β): The coefficient expressed in standard deviation units, allowing direct comparison of the relative importance of predictors measured in different units. The variable with the largest absolute standardised beta ($\beta = 0.607$) has the strongest independent relationship with revenue.

t-statistic and Sig:S The t-statistic tests whether each coefficient is significantly different from zero. A Sig. value of 0.000 ($p < 0.001$) for all four predictors confirms that each independently contributes significant explanatory power to the model after controlling for the other three variables.

VIF (Variance Inflation Factor): Values well below 10 for all predictors confirm the absence of multicollinearity. Budget VIF = 2.14, Vote Count VIF = 2.38, Popularity VIF = 1.12, Vote Average VIF = 1.04.

APPENDIX C: Variable Distribution Summary and Descriptive Statistics

Table C.1: Full Descriptive Statistics: All Variables (N = 13,440)

Variable	Min	Max	Mean	Median	Std. Dev.	Skewness	Kurtosis
Revenue (USD)	1,000	2,048,359,754	8,638,668	1,204,500	33,625,192	18.42	412.31
Budget (USD)	100	350,000,000	12,450,000	3,200,000	28,890,000	4.81	28.46
Popularity Score	0.60	1,804.50	14.72	8.34	42.38	22.10	598.44
Vote Count	2	29,614	312.84	62	1,088.52	14.63	268.91
Vote Average (0–10)	1.0	10.0	6.42	6.50	1.18	−0.41	0.72

Note: Revenue and budget figures in USD—all variables sourced from TMDB, IMDb, and Box Office Mojo. The extreme positive skewness of revenue, budget, popularity, and vote count is consistent with the power-law distribution documented in the film economics literature (De Vany & Walls, 1999; McKenzie, 2023).

Table C.2: Variable Correlation Summary with Revenue

Variable	r with Revenue	Strength	p-value
Vote Count	0.801	Very strong positive	< 0.001
Budget	0.721	Strong positive	< 0.001
Popularity Score	0.337	Weak–moderate positive	< 0.001
Vote Average	0.096	Very weak positive	< 0.001

Note: All correlations statistically significant at the 0.01 level (two-tailed), N = 13,440.

All SPSS outputs, including the full Pearson Correlation Matrix (Figure 4.1), the Regression Model Summary (Table 4.2), the ANOVA table (Table 4.3), the Coefficients table (Table

13 4.4), the Residuals Statistics (Table 4.5), the Normal P-P Plot (Figure 4.3), and the Standardised Residuals Scatterplot (Figure 4.4) are presented in Chapter 5 of the main dissertation body.