# ARIZONA STATE UNIVERSITY

Ira A. Fulton Schools of Engineering

TEM 505 Data Driven Decision Making

## Final Exam

Date of Submission: 8th Dec 2025

**Submitted by:**
Apurva Annasaheb Dange
ASU ID: 1233667918
Ankush Sanjay Harishchandre
ASU ID: 1233868378

Course Instructor: Dr. Steve T Cho

Fall 2025

## Table of Contents

## 1. Introduction

The given data analysis investigates whether Marvel continues to enjoy the same level of success that it had during the beginning of its operation or whether the franchise has already entered its decline phase. The analysis relies on data-driven decision analysis, relying on measurable evidence, statistical testing, and predictive modeling for each developed conclusion rather than personal opinion. Three major objectives of this analysis are as follows:

1. Critical Factors which tend to the success of marvel movies?
2. Is the "Marvel Era" coming to an end?
3. Can we predict which future Marvel films will do well?

By integrating financial data, audience response, critic ratings, and statistical tests, this report provides a series of insights that could help decision-makers such as film studios, investors, and production strategists plan for the future.

**Data Analysis Project Setup:**

**Code IDE:** Jupyter Notebook
**Language:** Python 3
**Libraries & Packages:** Pandas, Numpy, Matplotlib, Seaborn, Scikit Learn, SciPy
**Dataset Name:** Marvel_Movies_Dataset.csv

Code can be found in the canvas submission file name “data_analysis.ipnyb”

Several additional metrics were defined to translate raw data into decision-useful information:

- Profit: To calculate the profit, the worldwide gross minus production budget formula was used. This reflects the true outcome financially, beyond revenue figures.

- Return on Investment (ROI): Calculated as profit divided by budget, multiplied by 100. This shows the efficiency of the use of capital.

- Score: Cinema Score grades, like A+, A, A− were converted to numerical values to make them mathematically comparable.

These are essential variables for data-driven decision-making, which requires measurable and comparable indicators.

**Data Analysis Methods Used:**

In this report the trends are analyzed, changes are detected, and future outcomes are forecasted using the following techniques:

- Time-series visualization: This kind of visualization will allow for the observation of performance patterns over time.

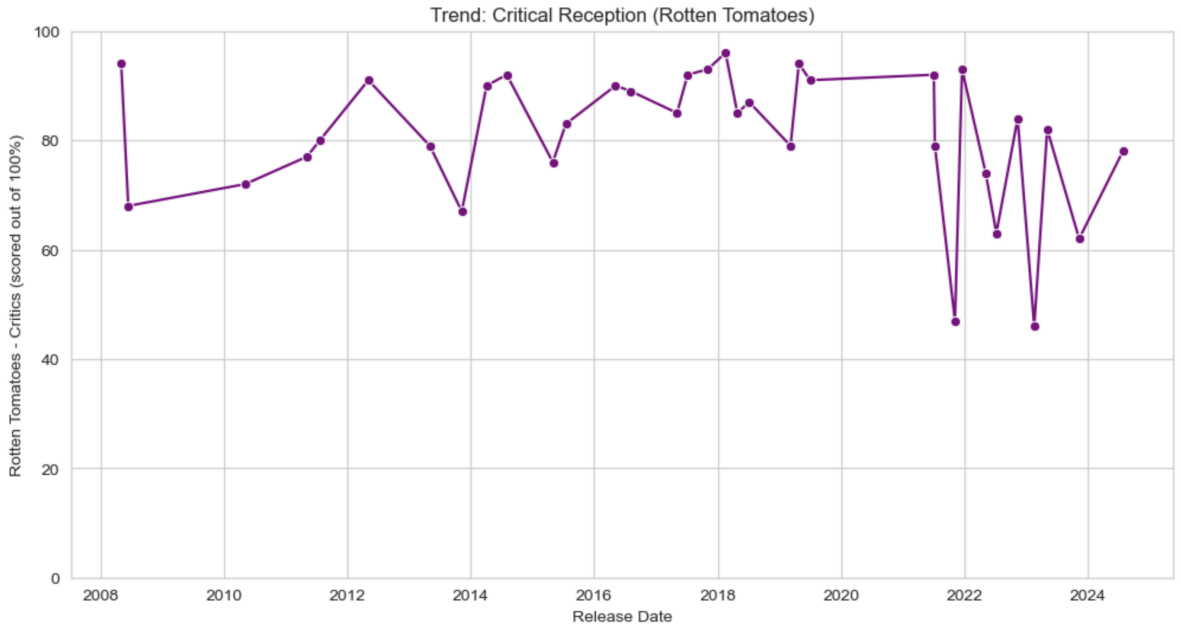
- Pearson correlation: This correlation shows the strength of the relationship between the variables with respect to profit.

- Welch's T-test: This statistical test is used to statistically compare pre-2021 and post-2021 film eras.

- Linear regression model: Is used to predict profit with respect to budget and audience/critic scores.

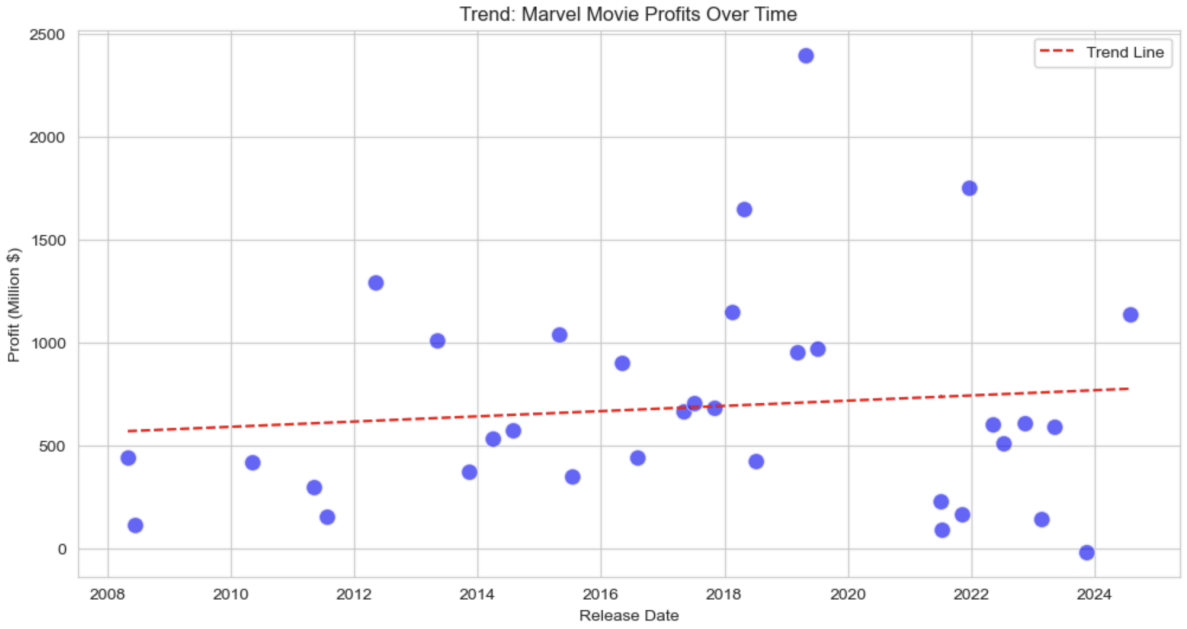
**Analysis of the trends:**

The following graph plots the trend line in Marvel's performance to see whether it has improved or declined.




Overall, the performance data from 2008 to 2019 shows a steady growth in profits, audience rating, and critic score, which shows:

- Strong narrative continuity
- Character development
- Heightened public interest impelled by crossovers across multiple films




During the period, Marvel movies had been looked at as relatively low-risk investments since historical data indicated favorable performance for nearly all releases.

The trend is considerably more variable post-2021. While some films are still performing well, the majority tend to underperform. Visually, the trend suggests:

- Increased fluctuation
- More mixed audience reviews
- More divided critic responses

From the graph, the trend line shows a slow moving of the slope, and the profit outcomes become more volatile moving forward into 2021, which shows some films performing really well while others below expectation.

These patterns suggest that Marvel has moved from predictable, reliable returns to uncertain performance and greater risk. In business terms we can say that, Marvel shifted from being a portfolio asset that was stable into one that was volatile, profits, therefore, cannot be predicted just from the reputation of the brand.

**Statistical Testing:**

To determine if the detected drop of the marvel movies is real or not an artifact of the small number of movies, Welch's T-test was applied to the two groups pre- and post-2021 releases. The results are as follows:

```
--- Profitability Analysis ---
Pre-2021 Mean: 764.93 | Post-2021 Mean: 530.36
P-value: 0.2354
Conclusion: Difference is NOT statistically significant (likely due to high variance).

--- Critical Quality Analysis ---
Pre-2021 Mean: 84.78 | Post-2021 Mean: 72.73
P-value: 0.0385
Conclusion: Statistically SIGNIFICANT decline.
```

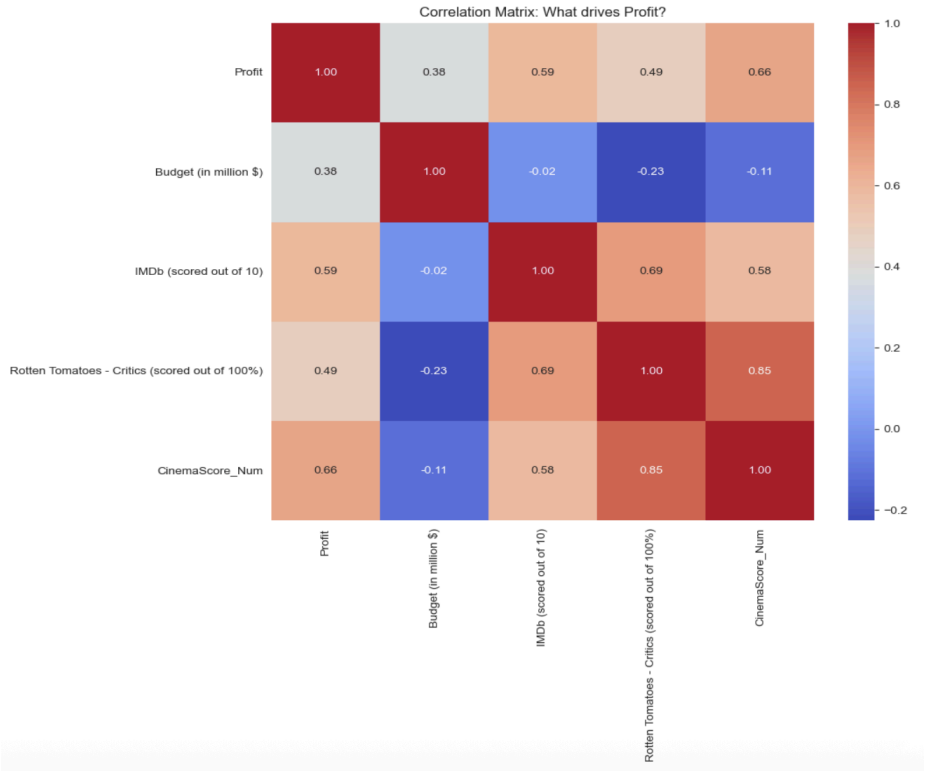
The p-value of 0.038 confirms that there is a decline in critic scores is statistically significant. The profit decline, however, is not statistically significant because of high variability, the presence of both blockbuster hits and major disappointments.
Critic scores have dramatically decreased since 2021, and at a level of statistical significance indicating that the drop reflects more than random fluctuation. There is no clarity on how the profit has gone down because of low data this generally says that the business is in risk of unpredictable financial situation. Overall, the Marvel has a significant decline in quality and a rise in financial variance, making strategic planning more challenging.

**Predictive Modeling**

A multiple linear regression model projects the profitability of a film, using as inputs those things that can be measured, like budget, critic score, and audience rating.This data-driven approach supports forecasting and reduces the financial uncertainty, informing investment decisions beyond the reliance on brand legacy.

**Key Takeaways From the Predictive Model**

After training the model on historical data, three factors has emerged as the core predictors of the profit:



As we can see in the correlation matrix, The CinemaScore have the strongest correlation with profit (0.66), followed by IMDb ratings by users, proving audience-driven value is more influential than budget.

1) Audience CinemaScore

The strongest influence on profit is by the opening-night audience score.

Characteristics:

- The audience score indicates authentic customer satisfaction.
- Word-of-mouth is a low-cost, yet powerful form of marketing.

It follows that positive audience response would mean more repeat viewing, stronger weekend growth, increased merchandise and streaming sales, and less negative impact of the press. In practical terms we can say that audience enthusiasm is positively linked to business outcomes; while fan dissatisfaction has immediate negative consequences. The finding also supports the customer-retention view, in that retaining existing fans is less costly than attracting new fans.

2) Critic Score
Critic scores are the second most robust predictor but have a lesser impact than audience perception. According to business interpretation, critical voices build primary public confidence and influence news headlines, social media sentiment, and perceived industry credibility. However, critic ratings cannot fully represent the broader audience, a title can be "critic-loved" and at the same time "audience-ignored," or vice versa. Thus, critic ratings serve as professional validation rather than the sole determinant of financial success.

```
Correlation with Profit:
Profit                                          1.000000
CinemaScore_Num                                 0.659428
IMDb (scored out of 10)                         0.593189
Rotten Tomatoes - Critics (scored out of 100%)  0.490728
Budget (in million $)                           0.377820
Name: Profit, dtype: float64
```

Budget shows a much weaker correlation, supporting the conclusion that expense does not guarantee revenue.

3) Production Budget
This model suggests that higher production budgets have the least predictive impact on profit. Implications such as increased spending does not directly translate to higher returns.
Instead of large-scale effects, such as elaborate CGI or star casts, do not determine automatically whether movies are successful. A great story is essential, and a big-budget production might flop with a weak storyline.

**Model Strength and Prediction Reliability**

The best regression model resulted in an $R^2$ value of 0.66, which shows that about 66% of the profit variance is explained by the identified predictors: CinemaScore, Critic Score, and Budget. This value represents a robust explanatory power in the entertainment and consumer-behavior domain, where emotional factors contribute substantially to the outcomes. This level of predictability, from a decision-making perspective, will add the following actionable value:

- Strategic budget allocation
- Early script review adjustments

Pre-release test screenings:

- Focus of marketing aligned with fan expectations.
- Decisions on theatrical release versus streaming.

In other words, data modeling acts as a financial safeguard by lessening investment uncertainties.

The model explains two-thirds of the difference in profit using measurable predictors with an $R^2$ of 0.65 , which makes it a  forecasting tool for movie studios.

```
Model R-Squared: 0.653 (The model explains 65.3% of the variance in profit)

Model Coefficients (Impact on Profit in Millions):
                                        Feature  Coefficient
0                         Budget (in million $)     1.999454
1                       IMDb (scored out of 10)   220.487211
2                                CinemaScore_Num   991.411923
3  Rotten Tomatoes - Critics (scored out of 100%)   -12.299619
```

**Business Interpretation:**

The results are clear with an perspective of budget does not count that much other than finance but the real success comes with customer satisfaction and feedback. This dynamic represents a remarkable shift in entertainment economics: the first Marvel era succeeded not simply because of how much was spent but as a function of value-laden storytelling. The true factor of realization of the metric of movie success, customers generally compare old movies with the new one's and try to give their critiques based on what was interesting in this part of the movie from the last one, which does not really related with the marketing efforts which the marvels will make unless and until the movie is properly connected with the storyline.

**Impact  of Predictive Modeling on the Industry:**

Predictive modeling offers several practical advantages to industry practice as such as it can guide on script approvals and further decisions, forecasting in financial terms and profit gains from the test data which they had for the previous movies. Most importantly this predictive model is used in making the budgets negotiations and allocations which reduce brand name assurances and preventing the over expenditure on low profit projects. In the conclusion for this predictive modelling we can say that it reduces the business risk from guessworking and guides in production planning.

**6. Conclusion**

The purpose of this project is to determine answers to the question which we mentioned in the introduction section which were Critical Factors which tend to the success of marvel movies? Is the "Marvel Era" coming to an end? Can we predict which future Marvel films will do well?. After data analysis, statistical testing, and predictive modeling, we came to an results as follows:

1) From the data analysis from graph 1 and graph 2, Marvel movies no longer are the stable production house because of its declining rates which it has between 2008-2019. Declines in critical and audience reception since 2021 are highly significant, and the volatility of the profits has increased, further evidence that success is no longer a mathematical certainty provided simply by association with the brand.

2) Audience satisfaction should be placed above production expenditure. The predictive model results illustrate that the views of actual movie goers have more weight than the size of the production budget. It is not the case that each lavishly-budgeted film will make a fair share of returns; an excess of money spent without checking the audience's expectations may cause a rise in losses. People going out happy is what mainly drives success, rather than the amount of CGI, stunts, or visual effects. This changes the funding financial rationale to writing, character development, and emotional engagement.

3) Predictable success is still achievable. The predictive model can show that success in the future is not random and can be reasonably forecast if studios monitor the key indicators that count. These indicators are measurable, trackable, and actionable; they include:

- Early audience reactions
- Test Screenings
- Consistency in character development Emotional responses to trailers and previews
- Consistency and Continuity in Franchising Decision-making.

Therefore, has to be proactive, not reactive. Marvel should also measure the perception of the audience in advance of a release rather than take a retroactive view.

**Strategic Implications for the Industry**

These findings extend well beyond Marvel into the greater film industry, as many high-budget franchises boast of huge fanbases. The results are indicative of a shift in entertainment economics:

- Audiences are better informed
- Expectations are higher
- Loyalty is conditional
- Value is determined beyond the visual spectacle

Studios should incorporate predictive modeling into standard workflows from the scripting phase through test screenings, and before promotional budget allocation. In tune with the priorities of the audience, studios can minimize risk, maximize financial return, and build durable engagement over temporary hype.