

Welcome to "Project Box Office: Sentiment Secrets"

The year is 2025. You've just been hired as a **Data Analyst** by a major movie studio to solve a high-stakes challenge:

The studio needs to **predict** whether their upcoming movies will have *smash-hit openings* — or *disappointing flops*.

You've been given **early access** to **critic reviews** that are pouring in before each movie's release.

Now it's up to you:

- Can you analyze the *sentiment* hidden inside these reviews?
- Can you use it to **predict** each movie's **opening weekend box office revenue**?
- Can you advise the studio to **double down** on marketing — or **cut their losses** early?

Your mission:

👉 Build a machine learning model that takes in critic reviews (plus some bonus movie information like release month and distributor) and predicts the movie's opening weekend revenue — as accurately as possible.

If you succeed, you'll save millions in marketing costs. If you fail... well, let's not talk about what happens to data analysts who back a box office bomb.


Deliverable:

You will complete a **Colab notebook** and a **short written report** explaining your results. Details (and grading criteria) can be found in the rubric document.

Resources Provided:

- A cleaned dataset of movie reviews and metadata.
 - A starter Colab notebook with guidance and tasks laid out.
 - Links to helpful resources on sentiment analysis and regression modeling.
-

Start your journey here:

 <https://github.com/SurvivrrHayde/CS3DS4002>

Good luck, analyst. Hollywood is counting on you.