**Project4 : Sunilduth Baichoo, Felipe Suarez , Nazim ben**

**Title : Box Office Performance vs. Sentiment Analysis**

**Goal:** Investigate whether **movie sentiment scores** (from IMDb, Rotten Tomatoes, or other review platforms) correlate with **box office earnings**.

Additionally, explore whether early sentiment trends can predict a movie’s financial success.

**RQ** : Does movie sentiment (from IMDb, Rotten Tomatoes) correlate with box office earnings, and can early sentiment trends predict a movie’s financial success?

**Datasets :** Below are some **popular movie review datasets**:

### IMDB Movie Reviews (Large Dataset)

### Rotten Tomatoes Movie Reviews

### Sentiment140 (Twitter-Based)

## **Deliverables**

The project will produce **insights, correlations, and predictive models**. The output will include:

### **A. Structured & Cleaned Data**

* A **CSV or database file** containing:
  + Movie title, release date, sentiment score, audience & critic ratings
  + Box office earnings (opening weekend, total domestic, total worldwide)
  + Other features like budget, genre, star cast, director, etc.

### **B. Correlation & Statistical Analysis**

* **Does sentiment correlate with box office revenue?**
* **Do critic scores or audience scores matter more?**
* **Does early sentiment (before release) predict success?**

**Example Outputs:**

* **Pearson Correlation Coefficients** between sentiment score & box office revenue
* **Statistical Significance Test (p-values)** to confirm relationships
* **Regression Coefficients:** Measuring the effect of sentiment on revenue