**Report: Top-Grossing Film Analysis and Relevancy**

Date: August 12, 2025

Members: Jess Yo, Linda Li, Octave

Group: 4

This project analyzes the popularity of 50 films by comparing data from two different platforms. The process began by scraping the HTML and web data from Wikipedia to identify the highest-grossing films, which is parsed and cleaned to populate a dataframe. Then, using each movie title as a query value, we called the TMDb API to retrieve related information: original title, release date, synopsis or overview, popularity, vote count and vote average. This data was stored in a separate dataframe, with the final df.describe() call used to gather statistical insights on comparisons of the films’ overall popularity against each other. Our analysis can be helpful in determining market trends and cultural shifts.

Market trends are becoming more and more indispensable in today’s digital age. Aggregating related datasets, as performed in our project, demonstrates how patterns can be identified to forecast demand for future projects. This is important for stakeholders in how they make decisions on film investments, as well as marketing and release strategies. For streaming platforms, dynamically updated data can provide better recommendations, leading to increased engagement and retention.

Moreover, these analytics go beyond commercial values, and can also be important cultural indicators, identify values or themes that adhere to targeted audiences, and provide markers for technological integrations. For example, the utilization of visual effects and inclusion of AI software can help researchers study the progression of techniques used in film making. Conclusions can provide insight into sociological changes and the popular sentiments including, but not limited to, political views, cultural symbols, and modern technologies.

Essentially, this project illustrates data collection through a combination of web scraping and API services. In its literal purpose, it is helpful in analyzing films based on public perception, and determining future predictions of media in the context of cultural shifts. Its application, however, demonstrates its convenience for future projects, creating action from data.