**Analyzing the Effect of Global Inflation on Movie Revenues and Genre Trends Using Data Science Methods**

**Introduction**

This project analyzes whether there is a meaningful relationship between global inflation and trends in the movie industry, by looking at box office revenues and popular genres. The idea was to explore how economic conditions, like inflation, might influence financial outcomes in entertainment, and whether these conditions also affect creative patterns such as the genres that gain popularity.

To do this, I used two datasets. The first one was the TMDB Movies Dataset, which contains information such as revenue, budget, release date, vote counts, and genres. The second was the Global Inflation Dataset, which includes yearly inflation rates for countries worldwide. Since not all years had complete data across both sources, I limited the analysis to the years 2000 to 2010, where data was more consistent. I matched each movie with the global average inflation rate of its release year and merged the two datasets accordingly.

After preparing the data, I followed a full data science workflow: exploratory data analysis (EDA), hypothesis testing, and machine learning modeling. Each stage helped me better understand the data and answer the main questions of the project.

**Exploratory Data Analysis (EDA)**

I started with exploratory data analysis to get a better understanding of the patterns and relationships in the dataset before doing any formal modeling. I first cleaned the data by removing movies that had missing or zero revenue, and I fixed inconsistencies in genre formatting so I could analyze genre trends clearly. Then I calculated the global average inflation rate for each year from 2000 to 2010 and merged that with the movies data based on release year.

Using visualizations like bar charts, boxplots, and scatter plots, I looked at how revenue and genre distribution changed over time and across inflation levels. Revenue had a highly skewed distribution, a few movies earned extremely high amounts while most earned relatively little. I observed that average movie revenues seemed lower in years with higher inflation, especially around 2008, which aligns with the global financial crisis. However, to make this conclusion, further analysis is needed.

In contrast, genre popularity stayed mostly consistent throughout the years. Genres like Action, Drama, and Comedy were dominant across the entire period, and there was no clear change in genre frequencies during high or low inflation periods. These observations from EDA formed the basis for further statistical testing.

**Hypothesis Testing**

To confirm whether the patterns I observed in the EDA were statistically significant, I ran two hypothesis tests.

First, I used a two-sample t-test to compare average movie revenues between high-inflation years and low-inflation years. I categorized the years based on whether their inflation rate was above or below the average inflation during 2000–2010. The result of the t-test showed a p-value of 0.48, which is not statistically significant at the 0.05 level. This means that, although there appeared to be a difference in average revenues in the EDA visuals, this difference cannot be confirmed statistically. In other words, there is no strong evidence that inflation level affects movie revenue in this sample.

Next, I used a chi-square test to check if genre distribution was related to inflation levels. I grouped inflation into categories (low, medium, high) and counted how often each genre appeared in each group. The result gave a p-value of 0.18, which is also not statistically significant. Therefore, we do not have strong evidence that inflation levels influence which genres are produced or preferred. This is consistent with what was seen in the EDA, where genre frequencies remained fairly stable across the years.

**Machine Learning Methods**

In the final phase of the project, I used machine learning to see if inflation and other movie-related features could help predict revenues or genres. This part of the project helped me understand how inflation compares to other features in terms of predictive power.

**1. Predicting Revenue (Regression)**

First, I trained a linear regression model to predict a movie’s revenue using features like inflation, budget, vote count, rating, runtime, release year, and genre (as one-hot encoded variables). The model achieved an R² score of 0.73, which means it could explain about 73% of the variation in movie revenue. This is a good result, but when I looked at the model’s coefficients, inflation had a very small weight, especially compared to budget and vote count. So, it can be said that inflation contributes to revenue prediction, but not very strongly.

**2. High vs. Low Revenue Classification**

Then I created a binary classification problem by labeling movies as “high revenue” if they were above the median revenue, and “low revenue” otherwise. I tested several models: Logistic Regression, Decision Tree, K-Nearest Neighbors, Random Forest, and XGBoost. The best performance came from the Random Forest classifier, which achieved an accuracy of 80.3%. When I added genre as a feature, the accuracy increased to 81.8%, showing that genre information helps improve prediction performance, even if it’s not directly influenced by inflation.

**3. Predicting Genre**

Finally, I tried to predict a movie’s main genre using features like inflation, revenue, vote count, rating, and release year. This was a multi-class classification task. The best model only achieved 24.9% accuracy, which is not much better than random guessing. This result confirms that genre is not something that can be predicted based on inflation or other numeric features, which supports the earlier chi-square test result.

**Conclusion**

In this project, I explored how global inflation relates to movie revenues and genres using data science tools. I combined two datasets, cleaned and aligned them, explored patterns visually, tested hypotheses statistically, and applied machine learning models to predict outcomes.

The main results of the project were:

* Movies released during high-inflation years had lower average revenues, but the difference was not statistically significant (t-test p = 0.30)
* Genre distributions did not change significantly based on inflation levels (chi-square p = 0.18)
* Revenue could be predicted using regression (R² = 0.73), but inflation was not a strong factor
* Genre prediction from numeric features was unsuccessful (best accuracy = 24.9%)

Overall, inflation appears to have some possible influence on the financial side of the movie industry, but it does not significantly affect the creative aspects, such as which genres are produced or consumed. While economic conditions might impact how much people are willing to spend on movies, they don’t seem to change the types of stories being told. This project showed that combining economic and entertainment data can lead to useful insights, and that even when patterns are weak, they can still be explored and quantified using the right analytical tools.