Explanation of Code

The code is split into three Google Colaboratories: Data Analysis, ML Regression, and ML Classification.

To run the code in Data Analysis, it is necessary to conduct several of the data cleaning steps in excel (we conducted these steps and saved the changes to a copy of the dataset called "IMDB_movies_cleaner.xlsx"). These steps are: dropping all of the rows without domestic earnings or meta_scores, dropping the certificate, overview, stars, and poster columns, splitting the genre column into three columns called Genre, Genre2, and Genre3, and creating a new column for an inflation index for each year from 1925 through 2022 based on data from the Federal Reserve Bank of Minneapolis.

To run the code in ML Regression, it is necessary to encode the cleaned, output dataset with columns for indicator variables for the famous director and genre variables. It is also necessary to make sure when creating indicator variables for the genres to group the following genres: Action & Adventure, Sci-Fi & Fantasy, Thriller & Horror, Musical & Music, Crime & Film Noir & Mystery, and Biography & History (we conducted these steps and saved the changes to a copy of the dataset called "ml_movies.csv").

To run the code in ML Classification, it is necessary to create a column for the indicator variable for the Blockbuster classification (we conducted this step and saved the changes to a copy of the dataset called "ml_workshop.csv").