**My Approach to Movie Rating Classification Project**

**Project Goal**
I wanted to use data science to figure out what makes some movies get high ratings (≥ 7.0) on IMDb. I used the IMDb Most Popular Movies (2006–2016) dataset from Kaggle to explore factors like votes, revenue, runtime, metascore, and genres, and then built a simple model to predict high vs low rated movies.

**Step 1: Loading & Exploring the Data**
I loaded the CSV file and examined the number of rows and columns, the meaning of each column, data types, and missing values. Revenue and Metascore had many missing values, which I noted early on.

**Step 2: Cleaning the Data**
I kept only useful columns such as Genre, Runtime, Votes, Revenue, Metascore, and Rating. Rows with missing Rating or Votes were dropped, and missing Revenue and Metascore values were filled using the median.

**Step 3: Creating the Target**
I created a binary label: 1 = High Rated (Rating ≥ 7.0)
0 = Low Rated (Rating < 7.0)

**Step 4: Preparing Features**
Genres were one-hot encoded, and numeric features were kept as they are. These became the inputs (X), while the label became the output (y).

**Step 5: Exploratory Data Analysis**
I visualized relationships using scatter plots, box plots, histograms, and bar plots to understand how features like votes, revenue, runtime, and genre relate to ratings.

**Step 6: Splitting the Data**
The dataset was split into 80% training and 20% testing data with stratification. Numeric features were scaled to avoid dominance by large values.

**Step 7: Model Selection & Training**
I used Logistic Regression because it is simple, effective for binary classification, and interpretable. The model was trained on the training data.

**Step 8: Evaluation**
I evaluated the model using accuracy, precision, recall, F1-score, and a confusion matrix to understand its performance on unseen data.

**Step 9: Key Insights**
Votes and Metascore were the strongest predictors of high ratings. Revenue had some influence, but quality mattered more than earnings. Genres like Drama and Sci-Fi appeared more frequently among high-rated movies.

**Overall Learning**
This project followed a complete data science workflow and helped me understand how to go from raw data to insights and prediction in a beginner-friendly way.