Phase 1: Problem Definition and Design Thinking

Problem Definition: The problem is to develop a machine learning model that predicts IMDb scores of movies available on Films based on features like genre, premiere date, runtime, and language. The objective is to create a model that accurately estimates the popularity of movies, helping users discover highly rated films that match their preferences. This project involves data preprocessing, feature engineering, model selection, training, and evaluation.

Design Thinking:

1. Data Source: Utilize a dataset containing information about movies, including features like genre, premiere date, runtime, language, and IMDb scores.
2. Data Preprocessing: Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.
3. Feature Engineering: Extract relevant features from the available data that could contribute to predicting IMDb scores.
4. Model Selection: Choose appropriate regression algorithms (e.g., Linear Regression, Random Forest Regressor) for predicting IMDb scores.
5. Model Training: Train the selected model using the preprocessed data.
6. Evaluation: Evaluate the model's performance using regression metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.

Dataset Link: **https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores**

Conclusion:

Based on the given dataset and the strategy planned, a machine learning model to predict IMDb scores of available movies by a certain criteria will be built such that it provides a better user experience.