**PROJECT REPORT**

## **1. Project Title and Objective**

**Title:** Movie Recommendation System Based on Genre and Ratings

**Objective:**The objective of this project is to build a basic movie recommendation system using Python. This system suggests movies to the user based on their selected genres and the average ratings of movies from a provided dataset. It is a simple rule-based system that uses filtering logic instead of machine learning algorithms.

## **2. Technologies Used**

* Programming Language: Python
* Libraries: pandas, re (regular expressions)
* Development Environment: Google Colab
* Dataset: MovieLens Small Dataset (from Kaggle)
* Files Used:
  + ratings.csv
  + movies\_metadata.csv
  + links\_small.csv

## **3. Project Description and Working**

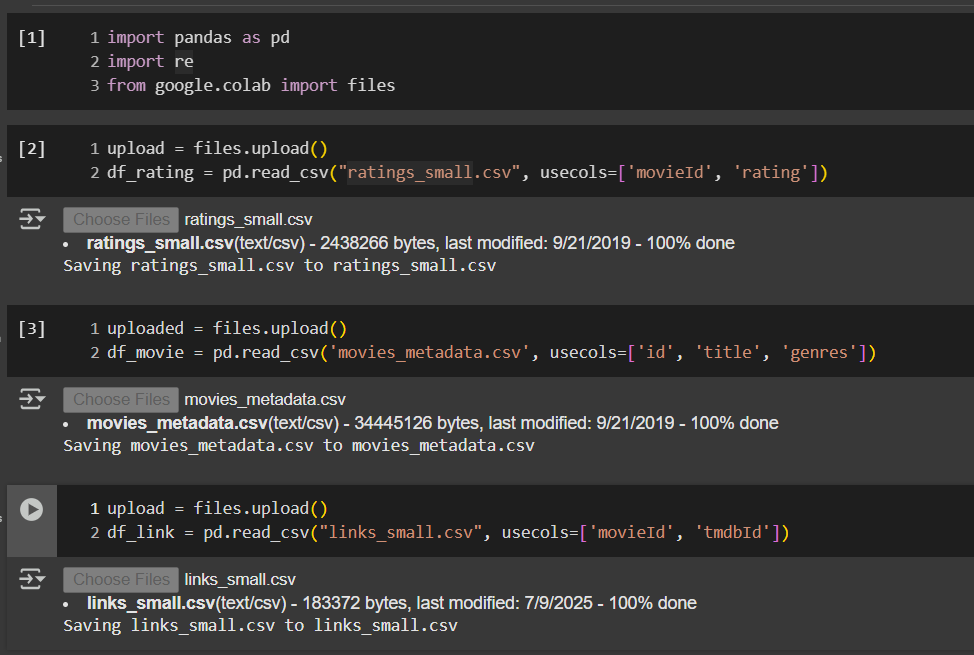
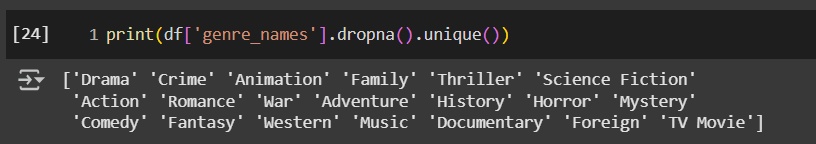
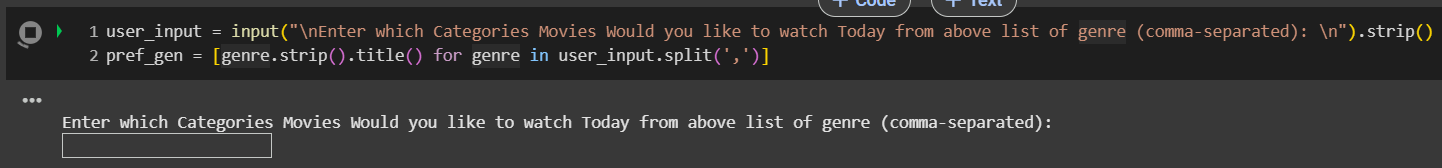
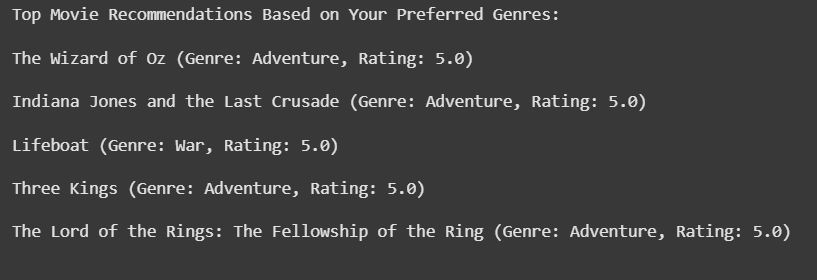
The movie recommendation system works in the following steps:

1. **Loading Data:** The system loads three datasets – ratings, movie metadata, and links – and reads only the necessary columns.
2. **Data Cleaning:** The data is cleaned by removing missing or invalid entries. Data types are also corrected for consistency.
3. **Genre Extraction:** The genres column contains JSON-like text. Using regular expressions, genre names are extracted and expanded so that each genre appears in a separate row.
4. **User Input:** The user is asked to enter one or more preferred genres (e.g., Action, Comedy, Drama). The input is processed and standardized.
5. **Filtering and Sorting:** The system filters the movies that match the selected genres and sorts them by their average rating in descending order.
6. **Recommendation Output:** Finally, the top 5 highest-rated movies from the filtered list are displayed to the user.

## **4. Key Features**

* Allows user to select one or more genres
* Recommends movies based on average ratings
* Displays top 5 results
* Works without machine learning (uses basic filtering logic)
* Uses real-world movie dataset

## **5. Screenshots**

* Dataset loading  
  
* Genre list displayed to the user  
  
* User input prompt  
  
* Final output with top 5 recommendations  
  

## 

## **6. Challenges Faced**

* Merging three datasets with different ID formats (movieId, tmdbId, id) was difficult.
* The genres were stored as JSON strings, so extracting them required regular expressions.
* Some entries in the dataset were missing or not usable and had to be removed.
* Ensuring that user input matched the actual genre names required careful string handling.

## **7. What I Learned**

* How to work with real-world datasets using pandas
* How to clean and preprocess data in Python
* How to merge datasets and handle missing values
* How to use regular expressions to extract information
* How to implement a simple recommendation system using rule-based logic

## **8. Sample Output**

Top Movie Recommendations Based on Your Preferred Genres:

1. The Dark Knight (Genre: Action, Rating: 4.5)

2. Inception (Genre: Action, Rating: 4.4)

3. Avengers: Endgame (Genre: Action, Rating: 4.3)

4. The Matrix (Genre: Action, Rating: 4.3)

5. Gladiator (Genre: Action, Rating: 4.2)

## **9. Dataset Source**

Kaggle - MovieLens Small Dataset  
URL: <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>