Movie Recommendation System — Workflow

1. Data Collection

• Load tmdb_5000_movies.csv and tmdb_5000_credits.csv.
• Merge both datasets using movie ID.

2. Data Cleaning

• Handle missing overview values.
• Parse JSON-like columns (genres, cast, crew).
• Extract director and top 3 actors.
• Normalize text and remove special characters.

3. Feature Engineering

• Combine overview, genres, cast, keywords, and director into a single 'tags' column.
• Prepare features for vectorization.

4. Text Vectorization

• Use CountVectorizer with max_features=5000.
• Remove English stopwords.

5. Similarity Matrix

• Generate cosine similarity matrix using the vectorized tags.
• Create a mapping between movie index and title.

6. Recommendation Function

• Input: Movie name.
• Output: Top 5 most similar movies.

7. Export & Deployment

- Save model files if needed.
- Finalize Jupyter Notebook.
- Create GitHub repository with notebook + scripts.
- Add documentation + sample outputs.