**Movie Recommendation System(Deliverable-2)**

My movies dataset consists of users and the movies they have watched. Each user has mentioned five favorite movies. Since the dataset is textual and not visual, I applied **text-specific preprocessing techniques** instead of image-based ones.

**Preprocessing Steps Applied:**

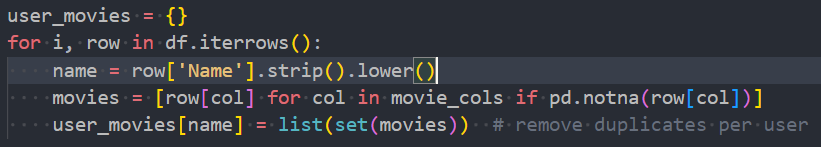
* **Lowercasing**  
  All movie titles were converted to lowercase to avoid mismatches due to capitalization (e.g., "Inception" vs "inception").
* **Whitespace Removal**  
  Leading and trailing whitespace was stripped from each movie title to ensure consistency.



* **Typo Fixing and Normalization**  
  A dictionary of common typos and alternate titles was created to standardize movie names.  
  Example fixes:
  + "muna bhi mbbs" → "munna bhai mbbs"
  + "top gun" → "top gun maverick"



* **Duplicate Removal**  
  If a user listed the same movie multiple times, duplicates were removed using Python’s set function.



These steps were essential for preparing the dataset for clustering and recommendation without semantic conflicts.

### **Cross-Correlation Analysis**

To mimic a **cross-correlation analysis**, I calculated the **pairwise similarity between users based on their movie choices**.

