\*Meeting Report\*\*

---

### Agenda

- Discussion on Project Preprocessing and Data Selection

- Dataset Selection for Model Input

- User-Based Movie Selection and Filtering for Genre and Ratin

---

### Summary of Discussion

1. \*\*Dataset Selection\*\*

The team discussed and finalized the selection of datasets required for the project. Out of six available datasets, only two were chosen: `movies.csv` and `ratings.csv`. These two datasets contain the essential columns required for the analysis and modeling:

- \*\*User ID\*\*: Identification of individual users

- \*\*Movie ID\*\*: Unique identifier for each movie

- \*\*Genre\*\*: Movie genre(s)

- \*\*Rating\*\*: User ratings for each movie

- \*\*Movie Title\*\*: The title of each movie

2. \*\*User-Specific Movie Selection Process\*\*

- A specific user (User ID: 5249) was selected for preliminary analysis. This user has watched a limited selection of 20 movies, making him the only user in the dataset with a comparatively small watch history, which provides a simplified data subset for testing our model's functionality.

- Out of these 20 movies, five were selected as a starting point for analysis:

1. \*Maveric\*

2. \*Con...\*

3. \*Piano\*

4. \*Eraser\*

5. \*Toy Story\*

3. \*\*Genre Analysis Using Association Rules\*\*

- The genres of these selected movies were extracted and analyzed using association rules to determine prominent genre patterns. Based on this analysis, three key genres emerged as preferences for User ID 5249:

- Adventure

- Drama

- Sci-Fi

4. \*\*Filtering Process with Genre and Threshold\*\*

- A filter was applied to the remaining movies in the dataset to retain only those that matched the user’s identified preferred genres (Adventure, Drama, and Sci-Fi). Additionally, a threshold of `>3` was set for ratings to ensure that only highly-rated movies remained for further analysis.

- After applying these criteria, nine movies were shortlisted, as follows:

| Movie Title | Rating |

|---------------------------|--------|

| Legends of the Fall | 4 |

| The Client | 4 |

| PP | 4 |

| Austin Powers | 4 |

| Apocalypse | 4.5 |

| Hunt for Red October | 4.5 |

| Armageddon | 4 |

| Lord of the Rings | 5 |

| Pirates of the Caribbean | 4 |

5. \*\*Prediction Process Based on Rating and Genre\*\*

- To predict movies that align best with User ID 5249's interests, the model utilized both genre preferences and high ratings. The top five movies selected as recommendations for the user are as follows:

1. \*Lord of the Rings\* - Rating: 5

2. \*Apocalypse\* - Rating: 4.5

3. \*Hunt for Red October\* - Rating: 4.5

4. \*Legends of the Fall\* - Rating: 4

5. \*The Client\* - Rating: 4

---

### Conclusion

The meeting successfully outlined and implemented a structured approach to filtering and predicting movie recommendations for a specific user within the dataset. By focusing on relevant genres and ratings, the team identified key insights into User ID 5249's preferences and generated a concise recommendation list. Moving forward, similar preprocessing and filtering methods will be applied to refine the recommendation model further.

### Action Items

- \*\*Data Validation\*\*: Ensure the accuracy of genre and rating data across the datasets.

- \*\*Model Optimization\*\*: Explore potential improvements in association rule-based genre identification.

- \*\*Testing\*\*: Apply the developed filtering process on additional users to validate the robustness of the model.

\_