DATA 612 Final Project: Recommender Systems – One-Week Work Plan

Day 1: Project Kickoff & Data Prep

- Review division of labor and finalize roles
- Load and clean *MovieLens 1M* dataset
- Enrich movies using TMDB API
- Finalize enriched dataset (e.g., `movies_enriched_full.csv`)

Day 2: Content-Based Filtering (CBF)

- Build CBF models using genre, keywords, cast, and directors
- Compare *TF-IDF* vs *CountVectorizer*
- Test *Cosine* and *Jaccard* similarity metrics
- Save similarity matrices and results

Day 3: Collaborative Filtering (CF)

- Implement memory-based (User-User, Item-Item) and model-based (SVD, ALS)
- Tune SVD (Surprise) and ALS (PySpark) models
- Handle centering, normalization, and sparse data
- Save prediction outputs and evaluation scores

Day 4: Hybrid Recommender System

- Combine CF and CBF scores via weighted average
- Implement fallback logic (cold-start support)
- Add reranking modules for diversity, novelty, and serendipity
- Run initial comparison between hybrid vs standalone models

Day 5: Evaluation & Metrics

- Compute RMSE, MAE, Precision@K, Recall@K
- Evaluate beyond-accuracy: *coverage*, *diversity*, *novelty*, *serendipity*
- Plot evaluation metrics and document insights
- Propose A/B test design for real-world simulation

Day 6: Spark-Based Scaling

- Implement ALS on PySpark (Databricks or local)
- Compare runtime and memory usage vs in-memory ALS
- Write analysis on scalability and system limitations

Day 7: Finalization

- Prepare project report and presentation slides

- Finalize and clean notebooks/codebase
- Review everything as a team
- Submit deliverables