**1. 目标（第一阶段）**

* 每人从不同角度独立对同一份**已清洗数据**做分析。
* 每个人输出一份**组件得分表**（component scores），格式统一。
* 第二阶段由合并脚本把所有组件得分**统一归一化并按权重合成最终推荐分数**，输出 Top-N 推荐列表。
* 原始数据只读：data/raw/movies.xlsx（**不能改或重命名**）。若做预处理把结果放 data/processed/。

**2. 四人分工（每人输出一个组件分数 CSV）**

（每人只需负责输出 movie\_id,title,component\_score 三列）

* **01 — Content（内容与相似度）**  
  负责列：genres, keywords, overview, title  
  目标输出：reports/tables/01\_content\_scores.csv（列：movie\_id,title,component\_score）  
  思路举例：TF-IDF 或关键词权重、genre 相关性、overview 信息量等。
* **02 — Ratings & Popularity（评分与受欢迎度）**  
  负责列：vote\_average, vote\_count, popularity  
  目标输出：reports/tables/02\_rating\_scores.csv  
  思路举例：贝叶斯平均/加权投票数、popularity 合成得分。
* **03 — Business（票房/出品/制作方）**  
  负责列：budget, revenue, production\_companies  
  目标输出：reports/tables/03\_business\_scores.csv  
  思路举例：ROI、票房规模、知名制片公司加权等。
* **04 — Time & Locale（时间/语言/地域/时长）**  
  负责列：release\_date, runtime, spoken\_languages, original\_language  
  目标输出：reports/tables/04\_time\_scores.csv  
  思路举例：新片优先（时间衰减）、语言是否匹配目标用户、时长适配度。

每人内部实现细节可以自由发挥（可以生成一些图表），但**输出格式必须完全一致**（见下面第 3 条）。

**3. 输出与评分规范（必须严格遵守）**

**文件与位置**

* 每人输出文件命名（必须）：
* reports/tables/01\_content\_scores.csv
* reports/tables/02\_rating\_scores.csv
* reports/tables/03\_business\_scores.csv
* reports/tables/04\_time\_scores.csv
* **每个文件必须包含且仅需这三列（顺序最好相同）：**
* movie\_id,title,component\_score
  + movie\_id：与原数据中唯一标识一致（字符串或数字均可）
  + title：电影名（便于人工核对）
  + component\_score：分数（数值）

**统一分数尺度与校验（关键）**

**输出分数可先按你自己算法产生原始分数（最终结果建议最高分100，最低分0）**