**项目统一说明（请大家都遵守）**

**1️⃣ 运行位置**

* 所有代码都要在**项目根目录**下运行。  
  也就是你能看到这些文件夹的地方：
* data/
* reports/
* scripts/
* notebooks/
* 在运行脚本前，先进入项目根目录，比如：
* cd D:\HuaweiMoveData\Users\xsb\Desktop\python-project

**2️⃣ 数据说明**

* 原始数据路径：
* data/raw/movies.xlsx

**这份数据不能改动（包括文件名）！**

* 如果自己需要做数据清洗或格式处理，处理后的文件统一放在：
* data/processed/

比如：

data/processed/movies\_cleaned.csv

**3️⃣ 脚本命名规范**

* 每个人按照自己的编号命名，编号就是分工后号码，比如你是 1 号：
* scripts/01\_你的主题.py（python文件名称一定要用英文！！！）

例：

scripts/01\_popularity\_ratings.py

scripts/02\_financials\_production.py

scripts/03\_genres\_content.py

scripts/04\_release\_geo.py

* 代码中读取数据时统一写：
* import pandas as pd
* df = pd.read\_excel("data/raw/movies.xlsx")

不要用自己的电脑路径（比如 C:\Users\...）。

* 输出文件统一放：
  + 表格结果 → reports/tables/
  + 图像结果 → reports/figures/

**4️⃣ Git 操作流程（非常重要）**

**建议每个人先创建自己的分支进行开发，合并前再推送到 main。**

**推荐做法：**

首次从远程仓库拉取项目后，在根目录创建自己的分支（用编号或名字命名）在这个分支上编写和测试自己的代码：

git checkout -b feat/01\_content

或者

git checkout -b feat/你的名字

1. 提交更改：

git add .

git commit -m "feat: 完成内容分析组件"

1. 在推送（push）前，一定要先拉取（pull）最新版本，防止覆盖别人改动：

git pull origin main ← 先拉取最新版本

git push origin feat/01\_content ← 推送到自己的分支

1. 开发完成后，在 GitHub 上发起 **Pull Request (PR)** 到 main 分支，让组长或负责人合并。  
   这样可以避免多人直接改同一个分支造成冲突。

**5️⃣ 其他注意事项**

* 不要修改别人脚本。
* 不要上传临时文件（.venv、.ipynb\_checkpoints、.DS\_Store）。
* 不要直接改 data/raw 里的内容。
* 各自只维护自己的脚本和输出结果。