目录

[一、基本功能 1](#_Toc22591)

[二、适用系统 1](#_Toc13706)

[（一）查看自己的系统构架方式 2](#_Toc21363)

[1.使用系统信息工具 2](#_Toc775)

[2.使用命令提示符 3](#_Toc6222)

[三、使用说明 5](#_Toc27809)

[（一）生成签到表 7](#_Toc29341)

[（二）生成签到汇总表 12](#_Toc22908)

[（三）统计每学期青字班报名的人员 16](#_Toc30041)

[（四）设置文件 19](#_Toc16259)

[四、各个文件夹说明与规范 22](#_Toc6705)

[（一）合法的表头关键词 22](#_Toc26456)

[（二）input/all 22](#_Toc3647)

[（三）input/app 25](#_Toc15520)

[（四）input/att\_imgs 26](#_Toc7594)

[（五）input/sign\_for\_QingziClass/all 32](#_Toc30137)

[（六）output/app\_out 33](#_Toc31993)

[（七）output/att\_out 34](#_Toc30239)

[（八）output/sign\_for\_QingziClass\_out 35](#_Toc26215)

[1./报名.xlsx 35](#_Toc10513)

[2./cmt 35](#_Toc29309)

[3./unpdf 35](#_Toc15918)

[（九）output/unknown.xlsx 36](#_Toc3760)

[（十）output/temp.jpg 37](#_Toc16415)

[（十一）storage/storage.xlsx 37](#_Toc18914)

[五、搜索工具 38](#_Toc32150)

[六、手动透视矫正 40](#_Toc11764)

[七、Bug解决 42](#_Toc16158)

[八、参与开发 43](#_Toc22212)

[（一）仓库地址 43](#_Toc22674)

[（二）注意事项 43](#_Toc14932)

[九、开发人员 44](#_Toc23356)

SmartSheet使用说明

发行日期：250911

版本号：V0.9.2

一、基本功能

本程序包含三个功能：

（一）**[生成签到表](#生成签到表)**：本程序可以解析青字班人员花名册，生成标准的青字班签到表，可用于打印；

（二）**[生成签到汇总表](#生成签到汇总表)**：本程序可以解析线下签到的图片，统计生成线下签到汇总表；

（三）**[统计每学期青字班报名的人员](#统计每学期青字班报名的人员)**：本程序可以统计青字班报名的人员信息。

你可以从[这里](https://gitcode.com/lscatfish/SmartSheet/releases)下载最新版本。

二、适用系统

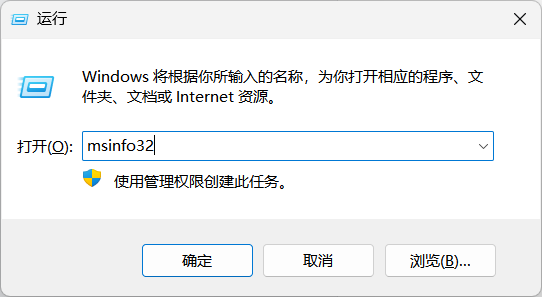
受到程序编写者水平限制，本程序运行的最适系统为Windows10，基于x86构架的64位系统（CISC）。ARM构架（RISC）的系统按理来说也可以兼容，但是尚未经过兼容性测试，可能会出现意外的错误。至于兼容到Linux或是macOS系统，希望后来者能够构建。

（一）查看自己的系统构架方式

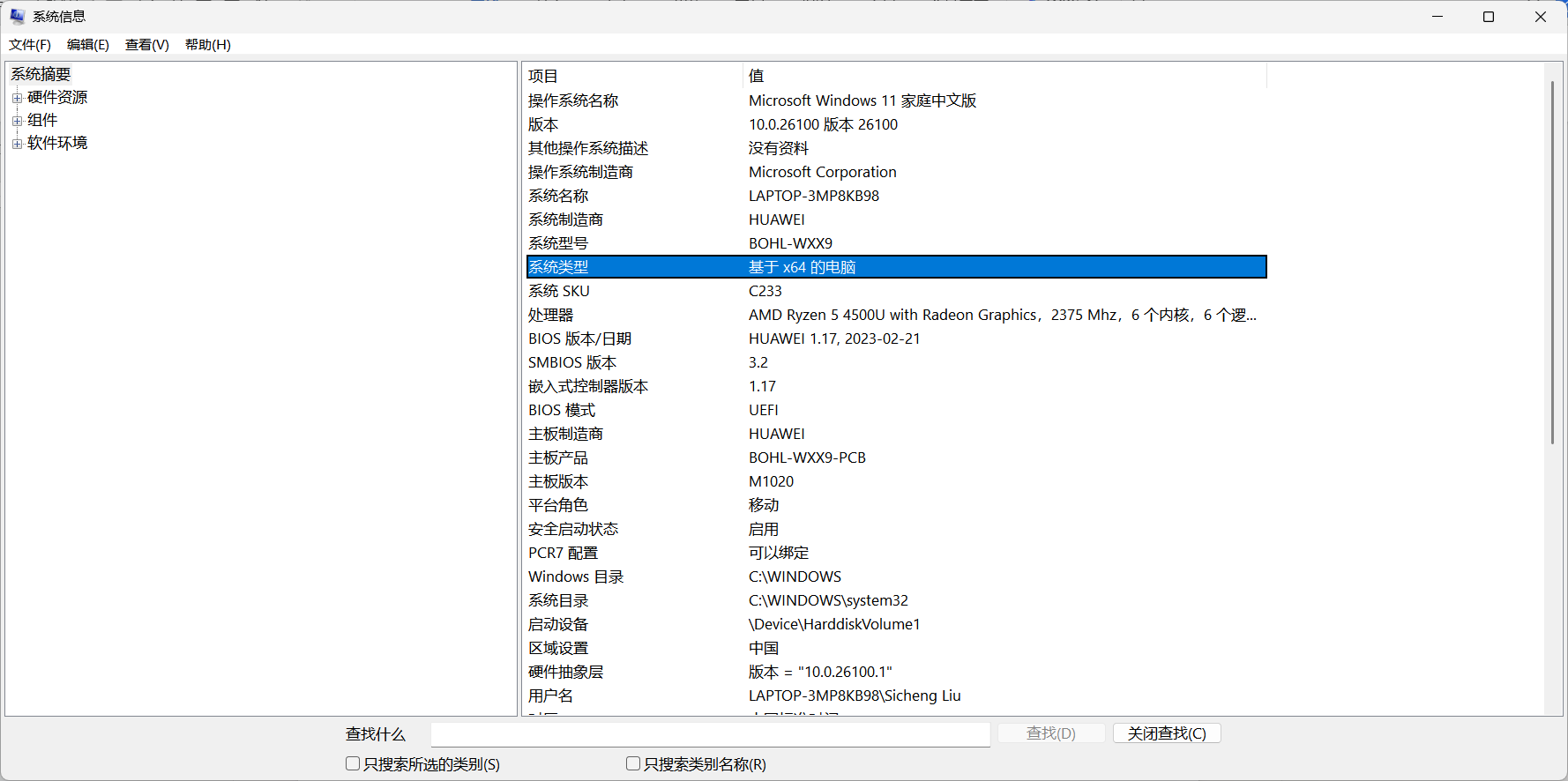
**1.使用系统信息工具**

同时按下Win+R，打开“运行”对话框，输入msinfo32并按回车键。在弹出的“系统信息”窗口中，查找“系统类型”字段，如果显示“x64基于PC”，则系统是x86架构的64位系统；如果显示“x86基于PC”，则系统是x86架构的32位系统；如果显示“ARM基于PC”，则系统是ARM架构。

打开“运行”对话框，输入msinfo32并按回车键

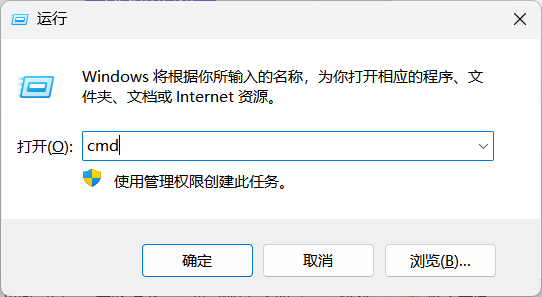


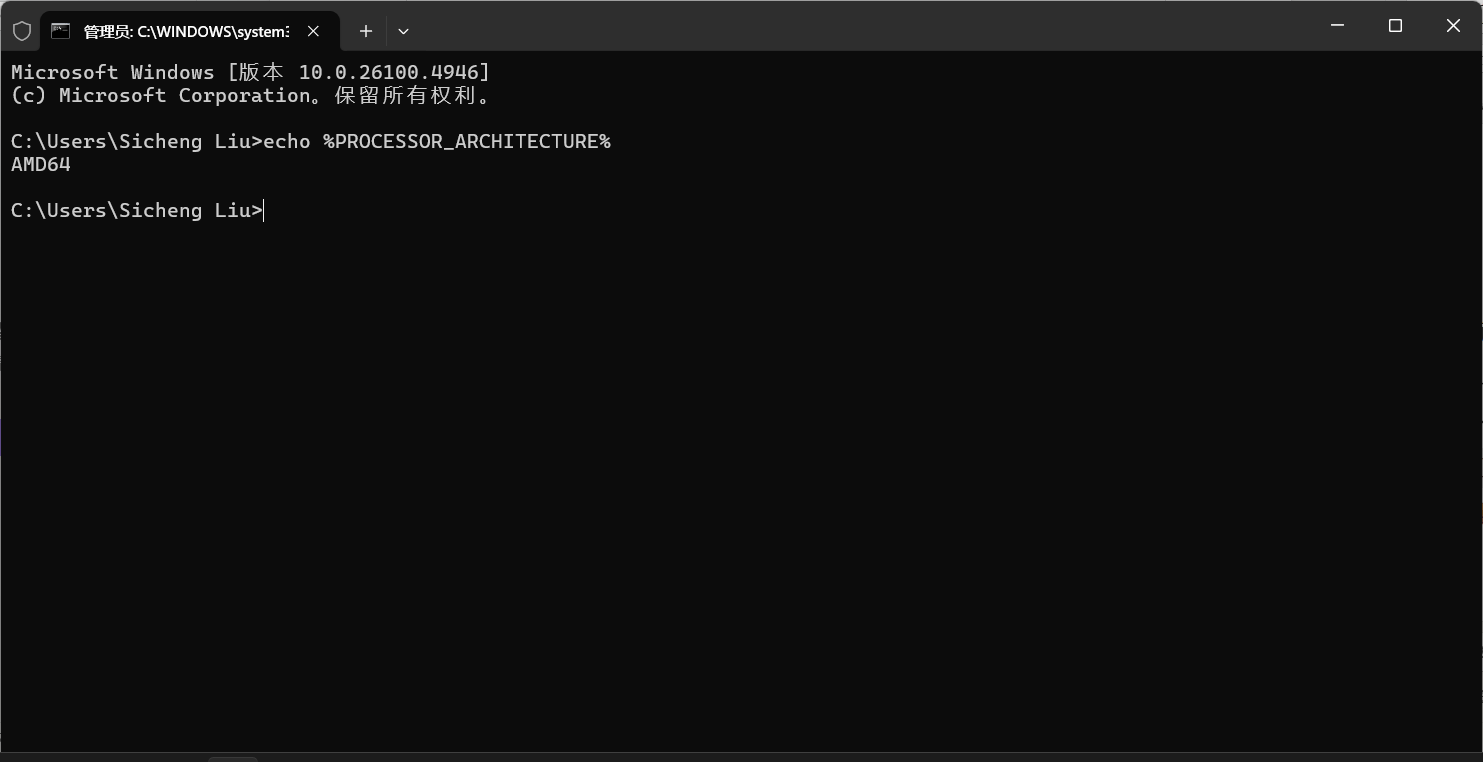
在“系统信息”窗口中，查找“系统类型”字段



**2.使用命令提示符**

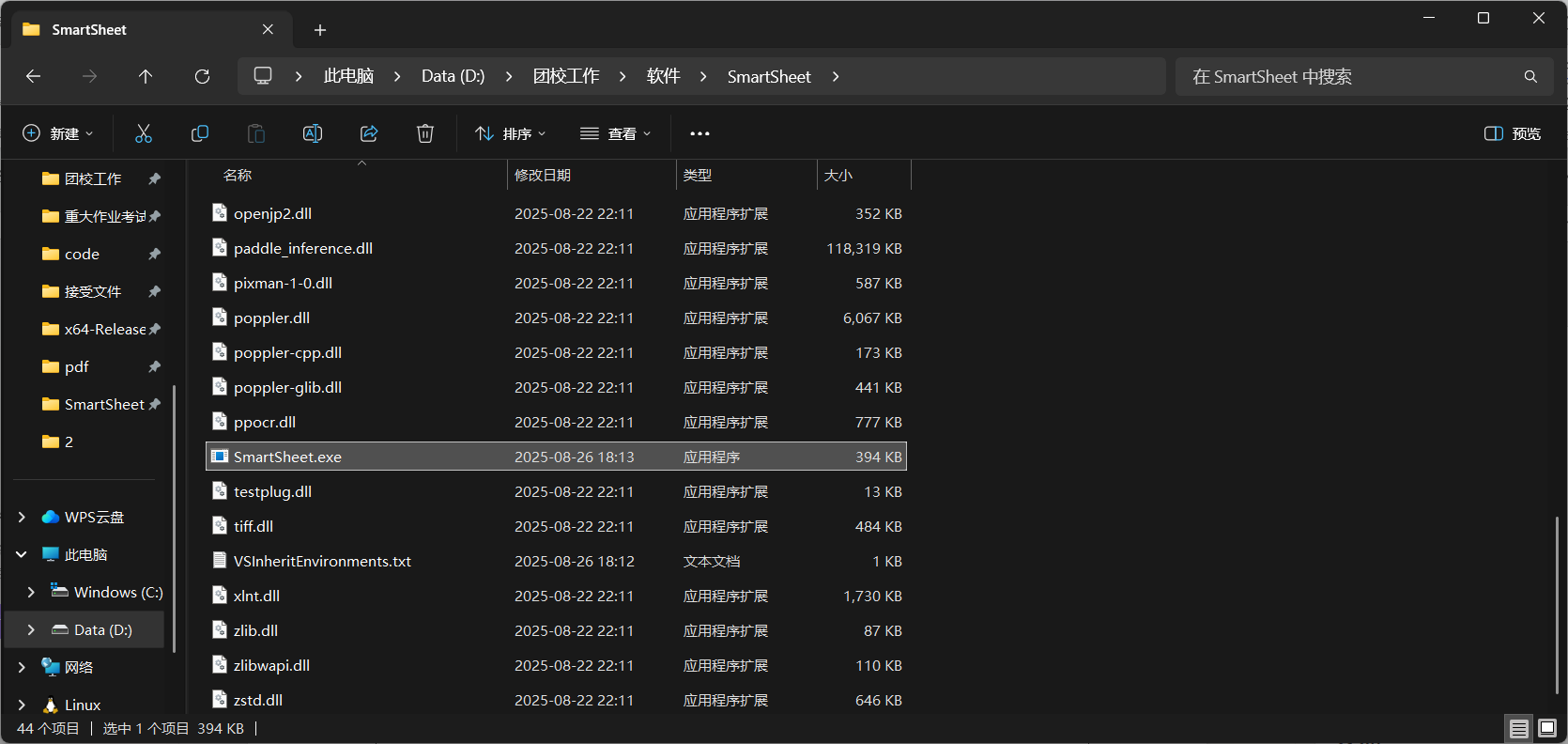
按下Win+R快捷键，输入cmd并按回车键，打开命令提示符。在命令行中输入echo %PROCESSOR\_ARCHITECTURE%并回车，若结果是“AMD64”，则表示系统是x86架构的64位系统；若结果是“x86”，则表示系统是x86架构的32位系统；若结果是“ARM”，则表示系统是ARM架构。

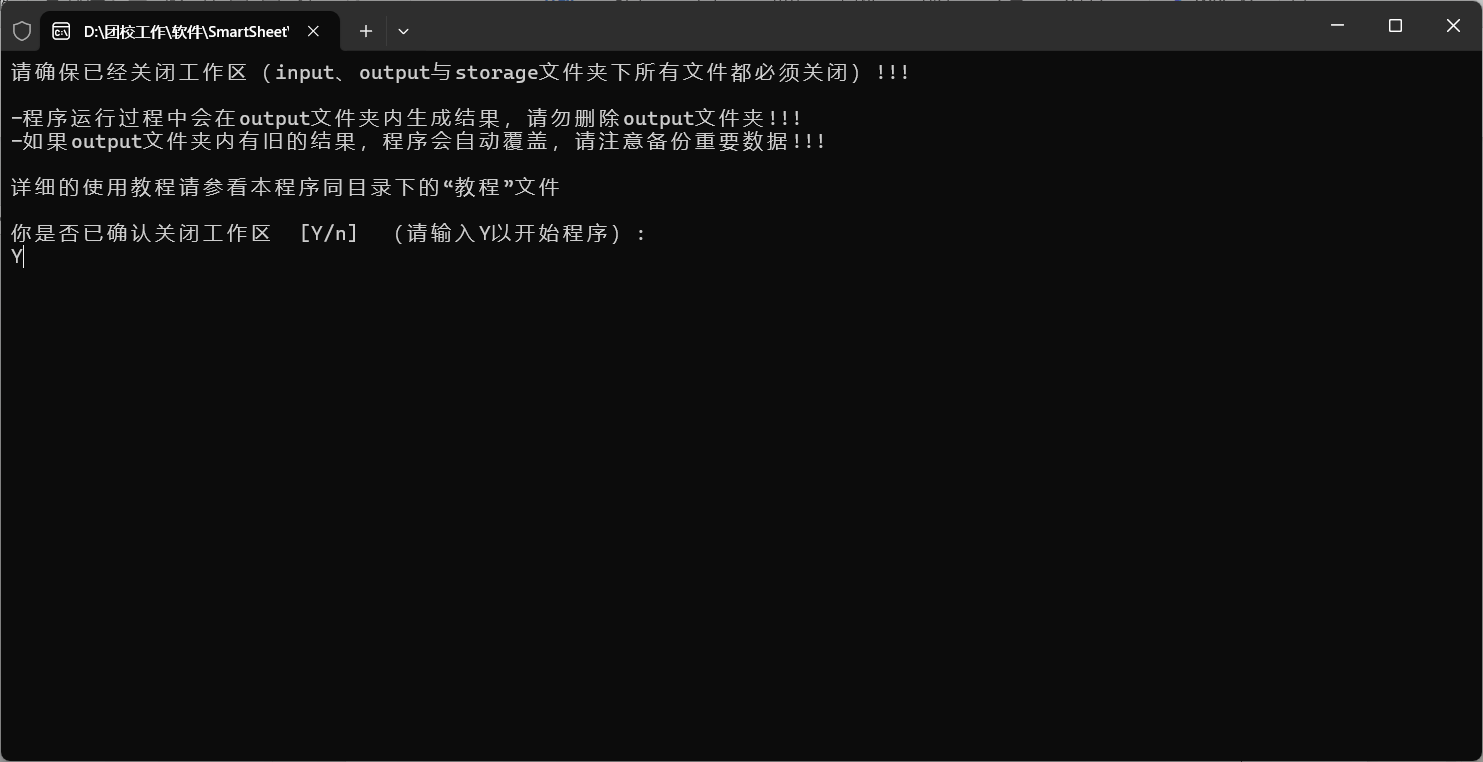


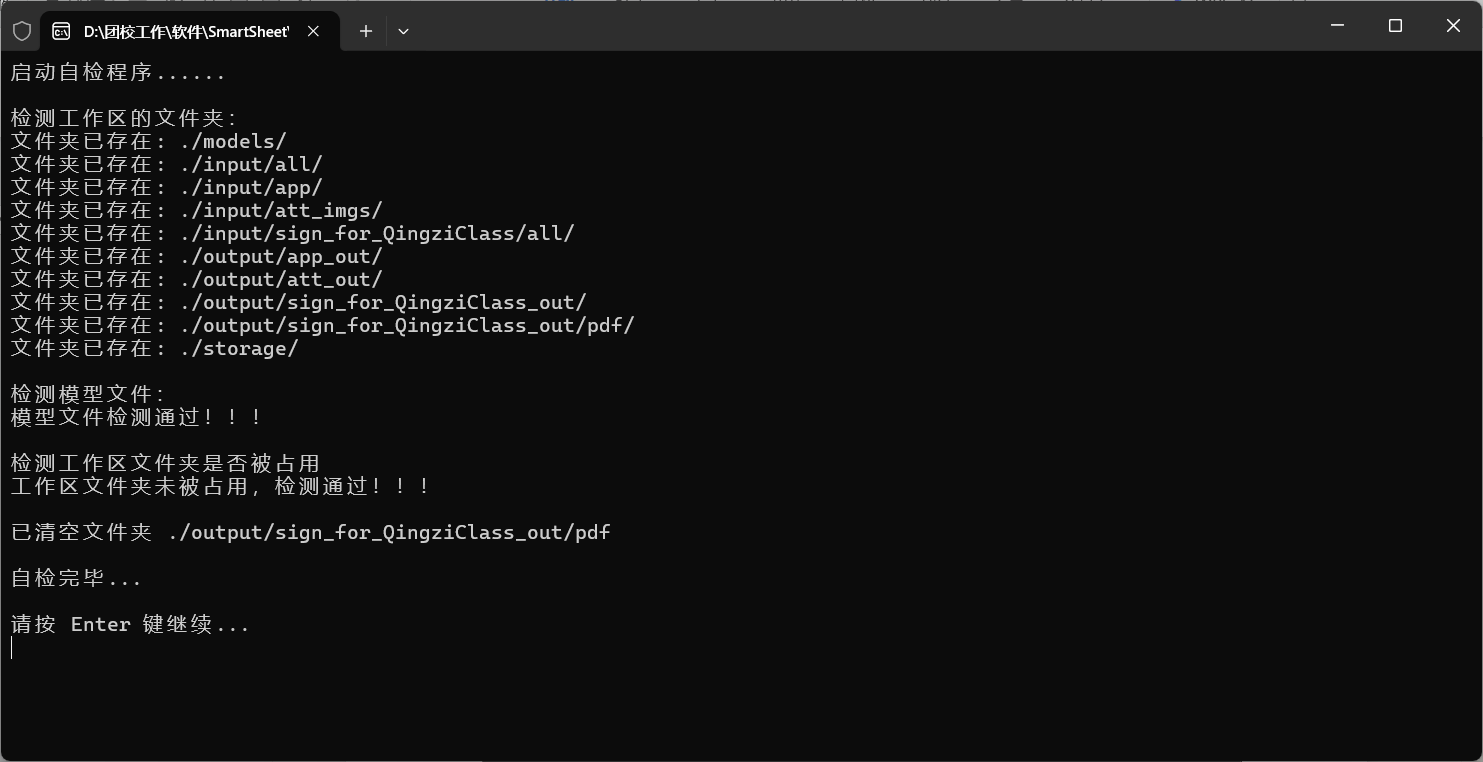


三、使用说明

双击SmartSheet.exe启动程序，确认工作区下所有的文件都关闭之后，输入“Y”并回车，进入程序自检页面。自检通过之后，按照提示按回车键，进入功能选择页面。（[请确认各文件夹文件的命名方式](#根文件夹说明与规范)）





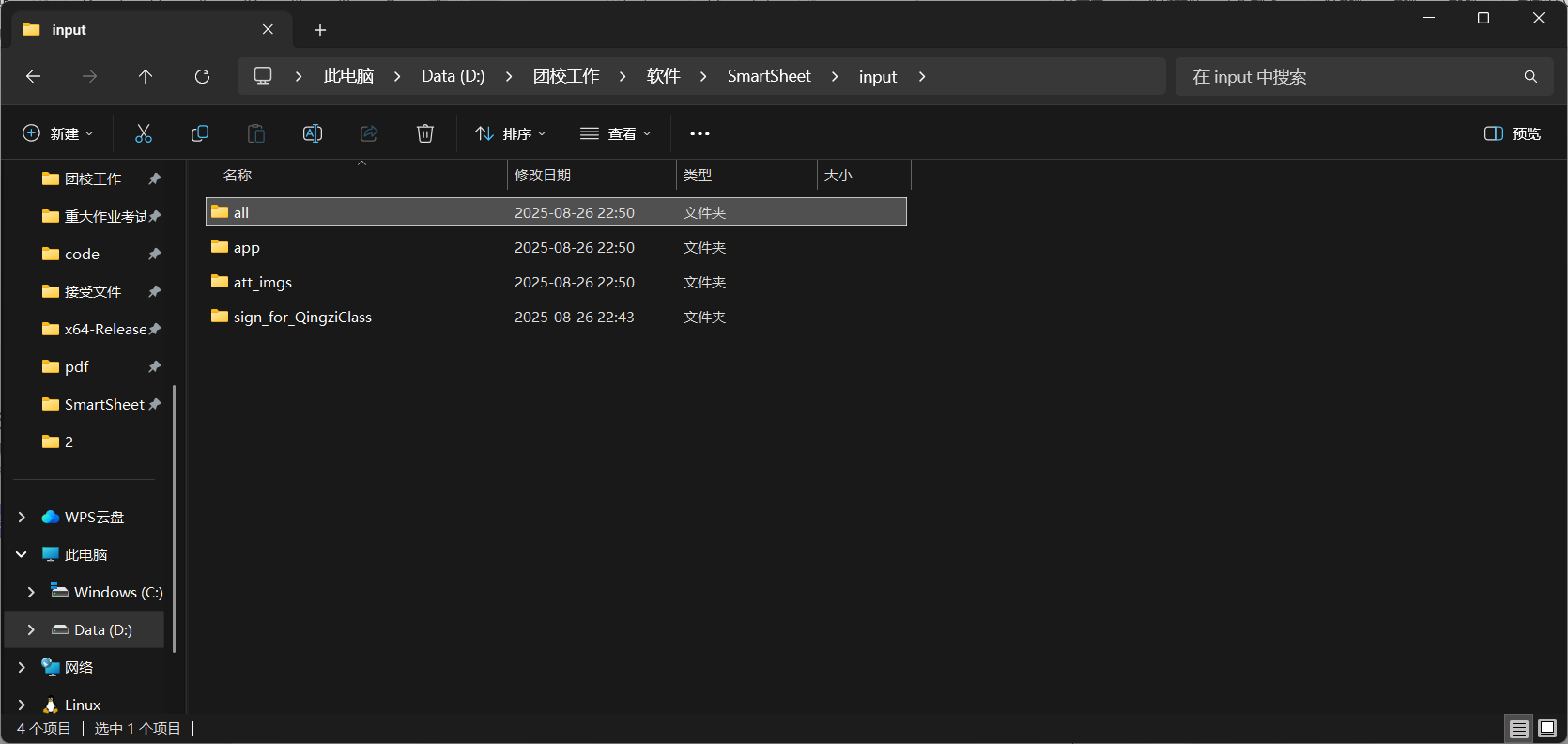


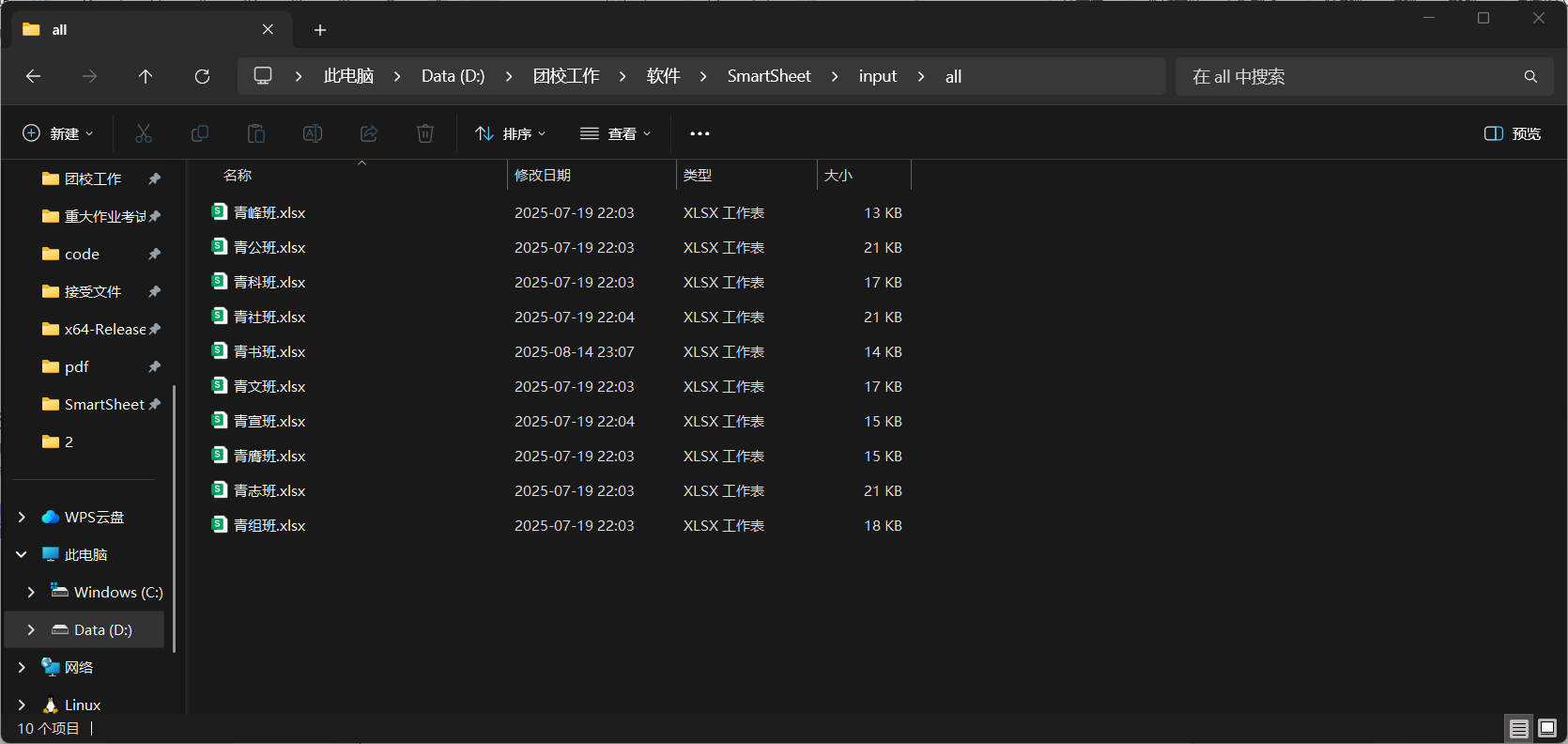
（[请确认各文件夹文件的命名方式以及内部格式](#根文件夹说明与规范)）

（一）生成签到表

1.确保文件夹“input/all/”中，青字班各班的花名册文件都已按照“班级名+.xlsx”命名，且再次确认没有被其他程序打开；

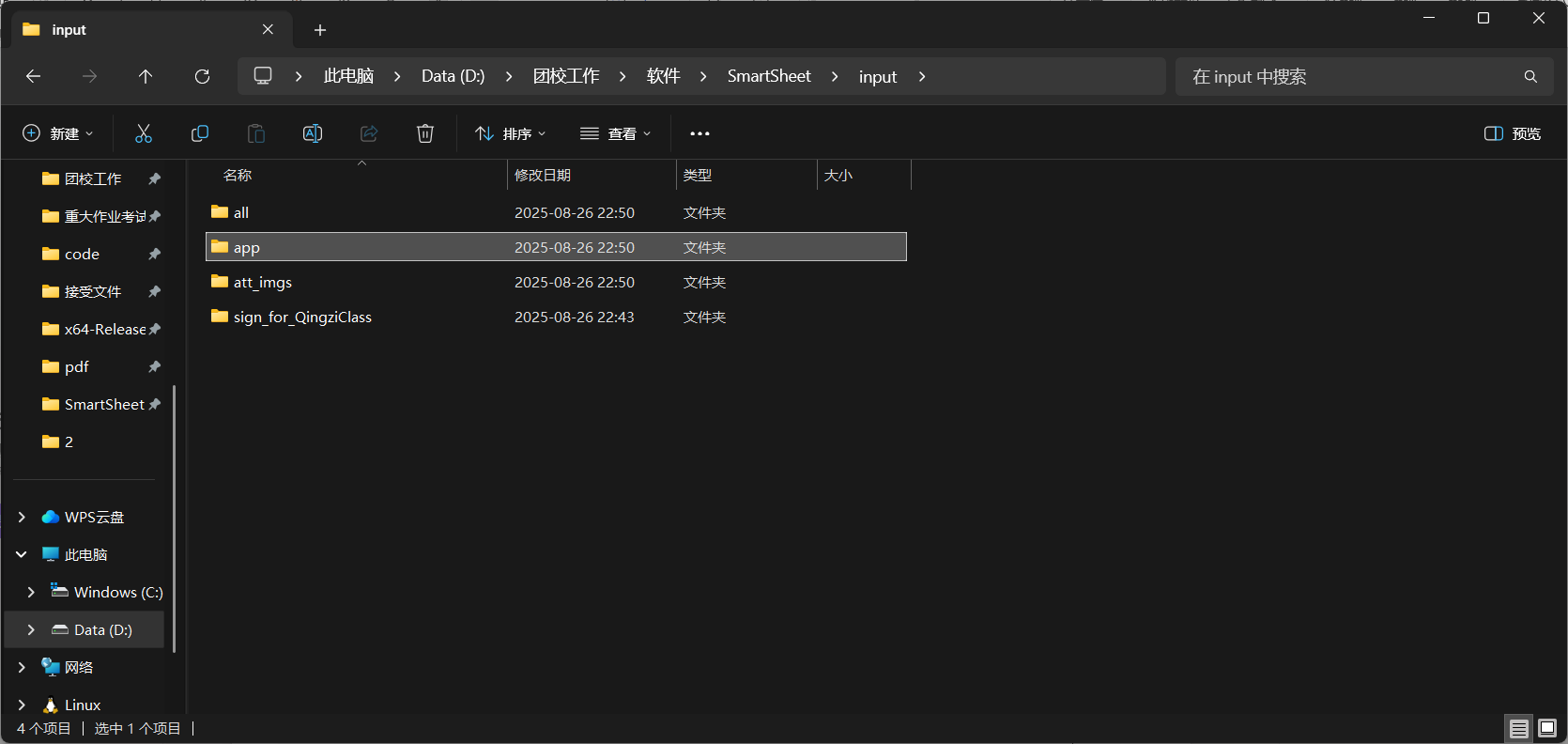


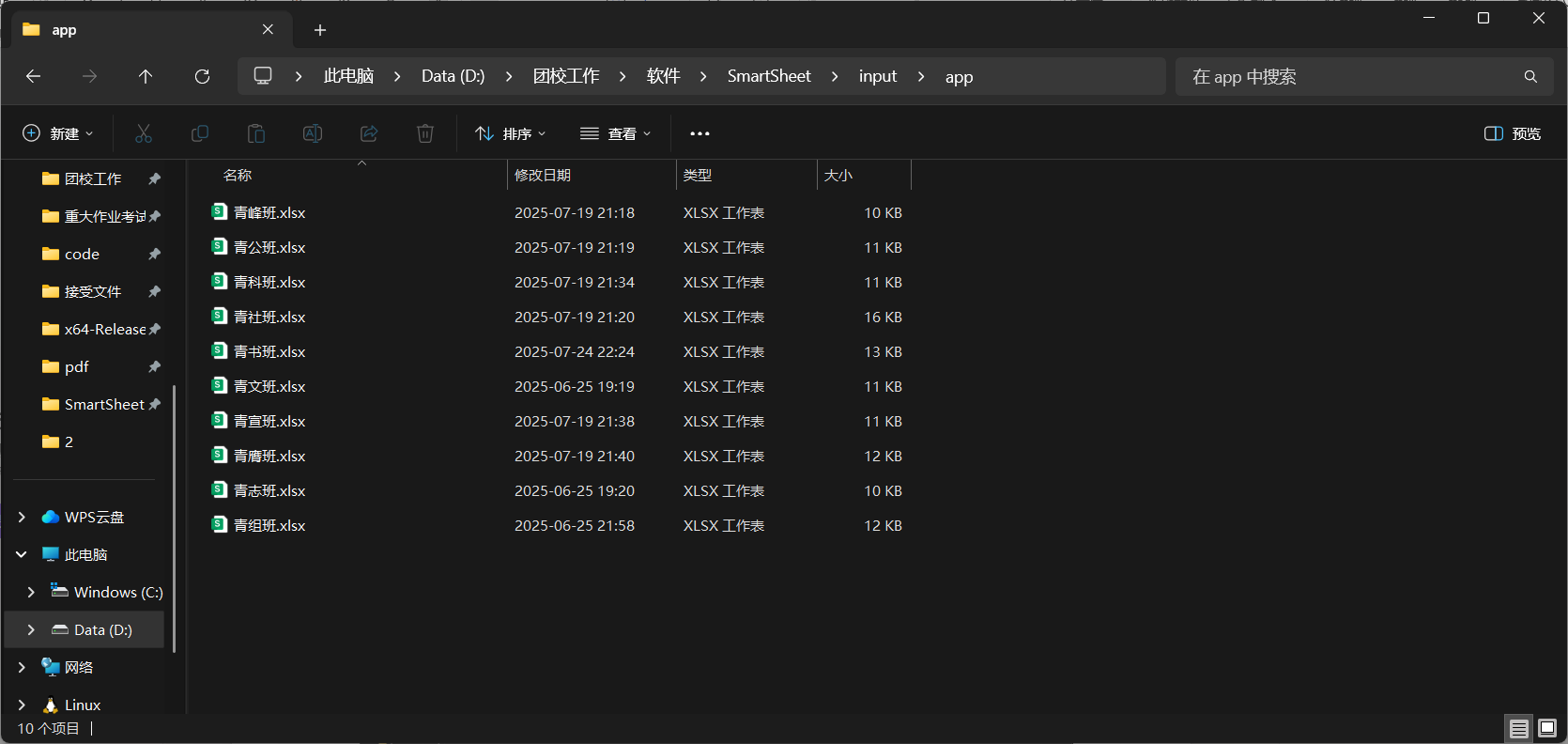




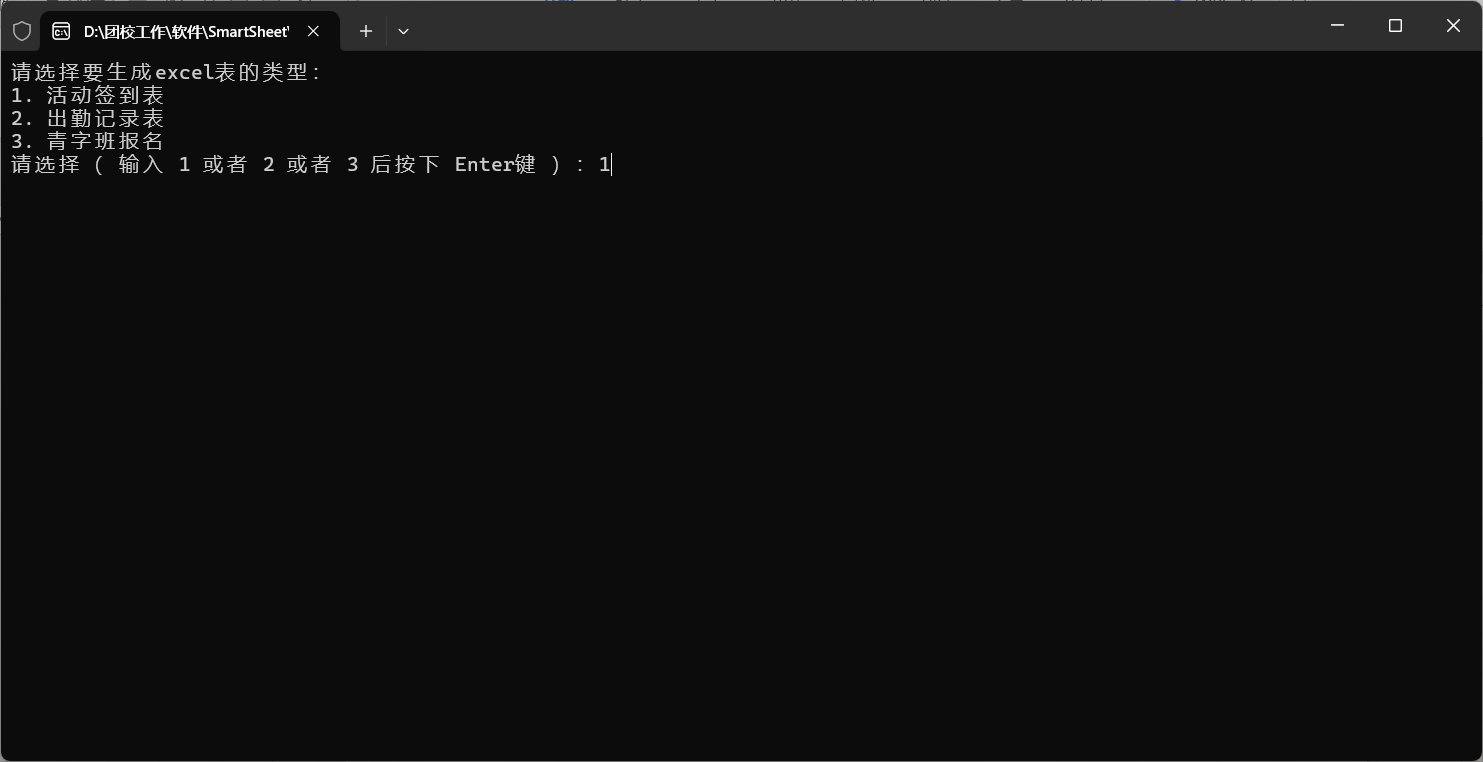
2.如果必要，确保文件夹“input/app/”中，青字班各班班委的报名文件都[已经过简单处理](#根文件夹说明与规范)，按照“班级名+.xlsx”命名，且再次确认没有被其他程序打开；

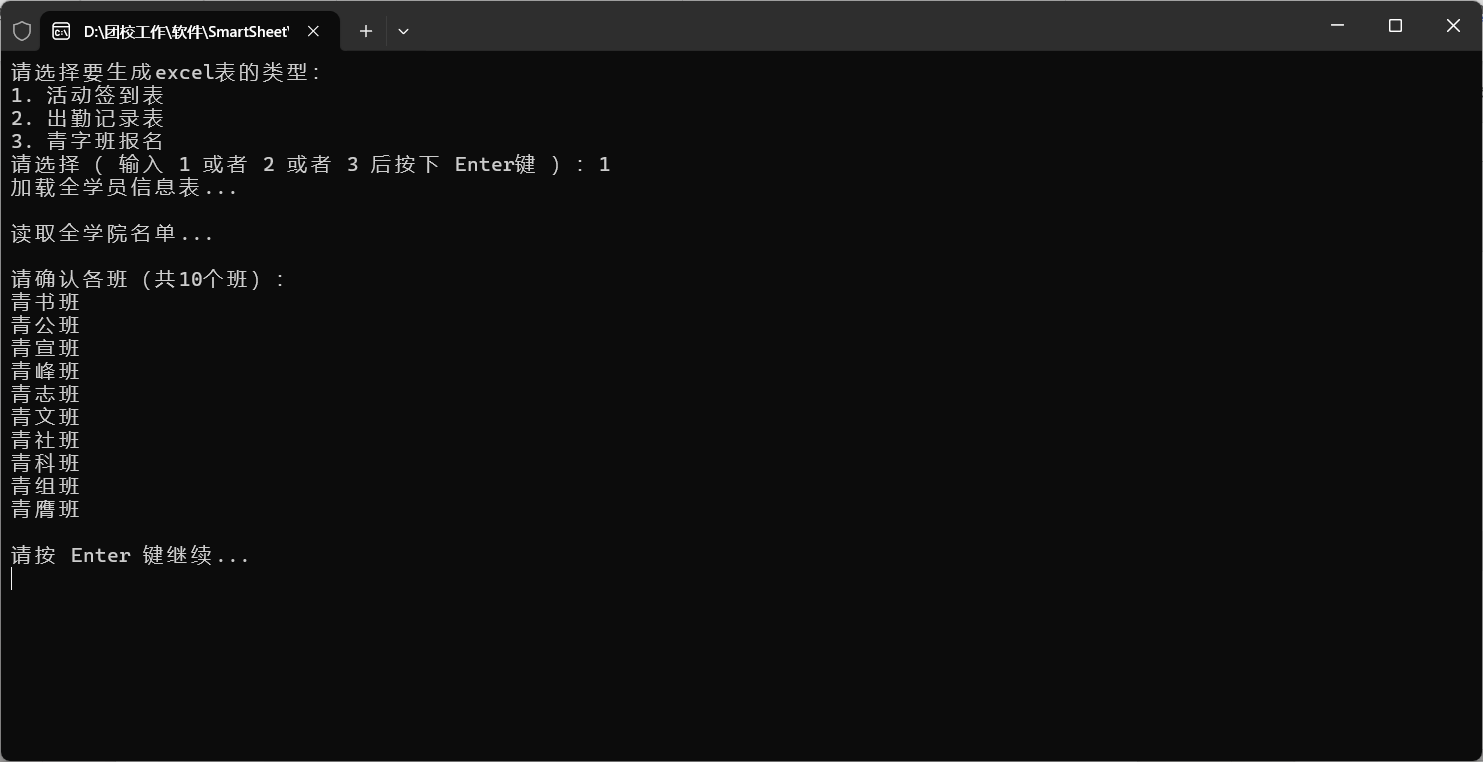




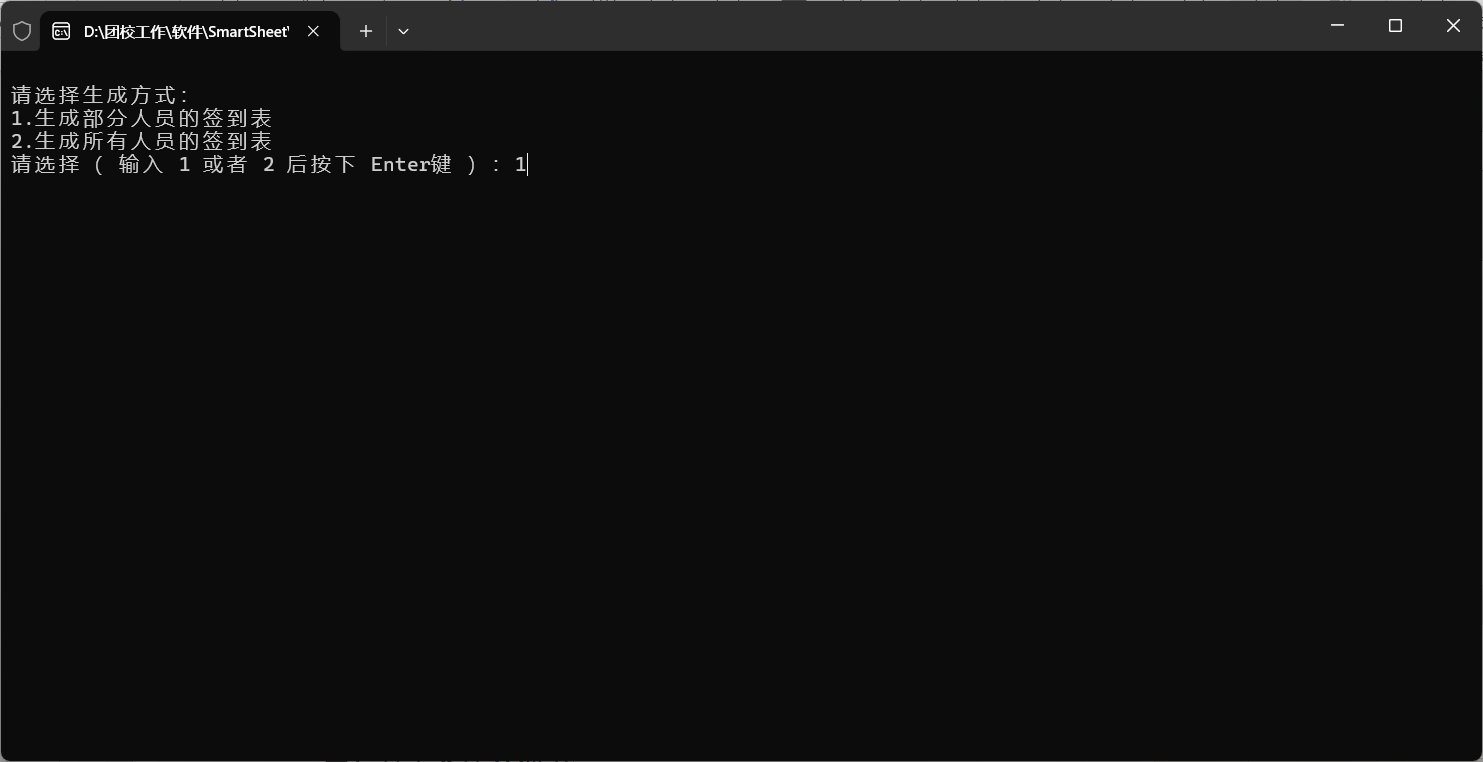


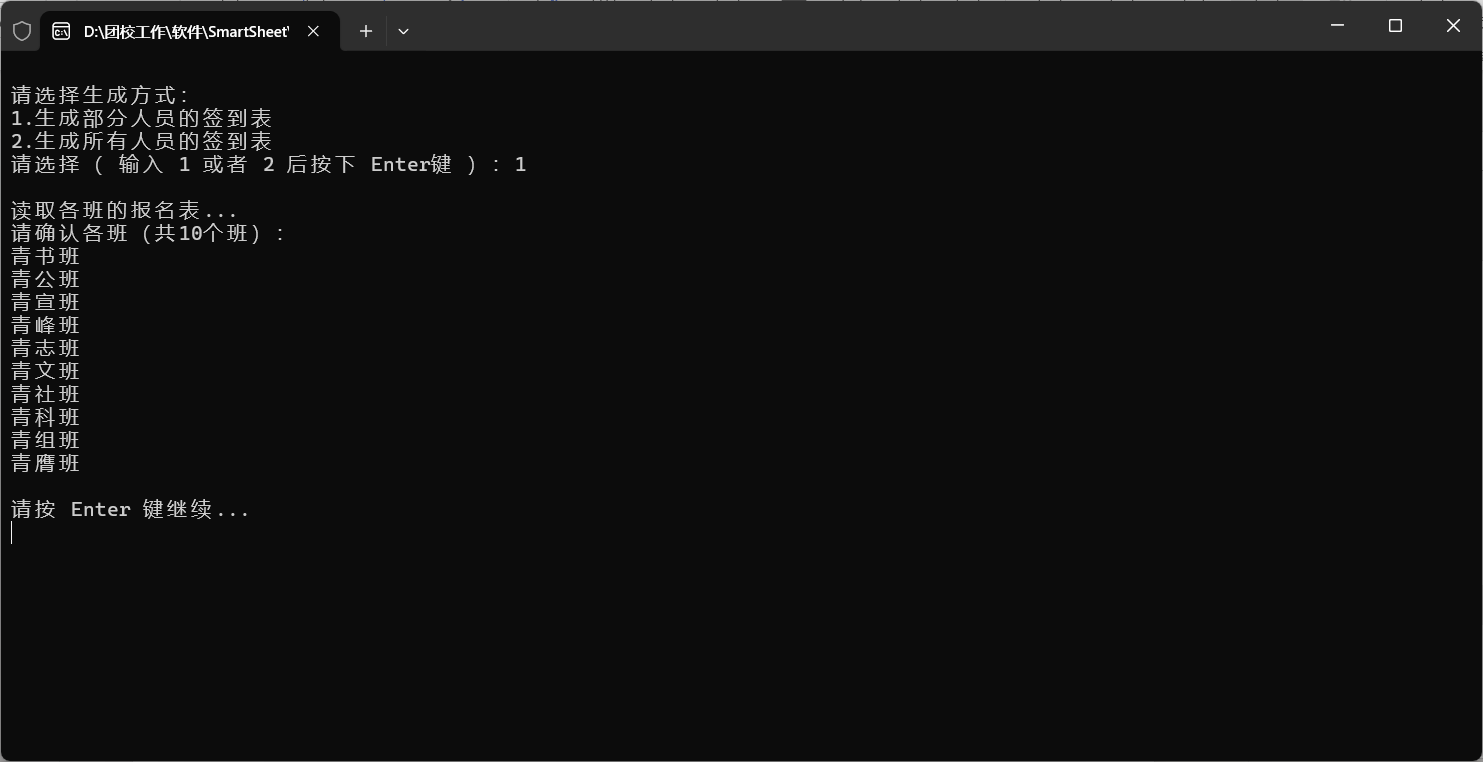
3.在程序中输入“1”并按下回车，确认花名册班级数量与名称都正确后按下回车键，进入签到表输出选择页面；



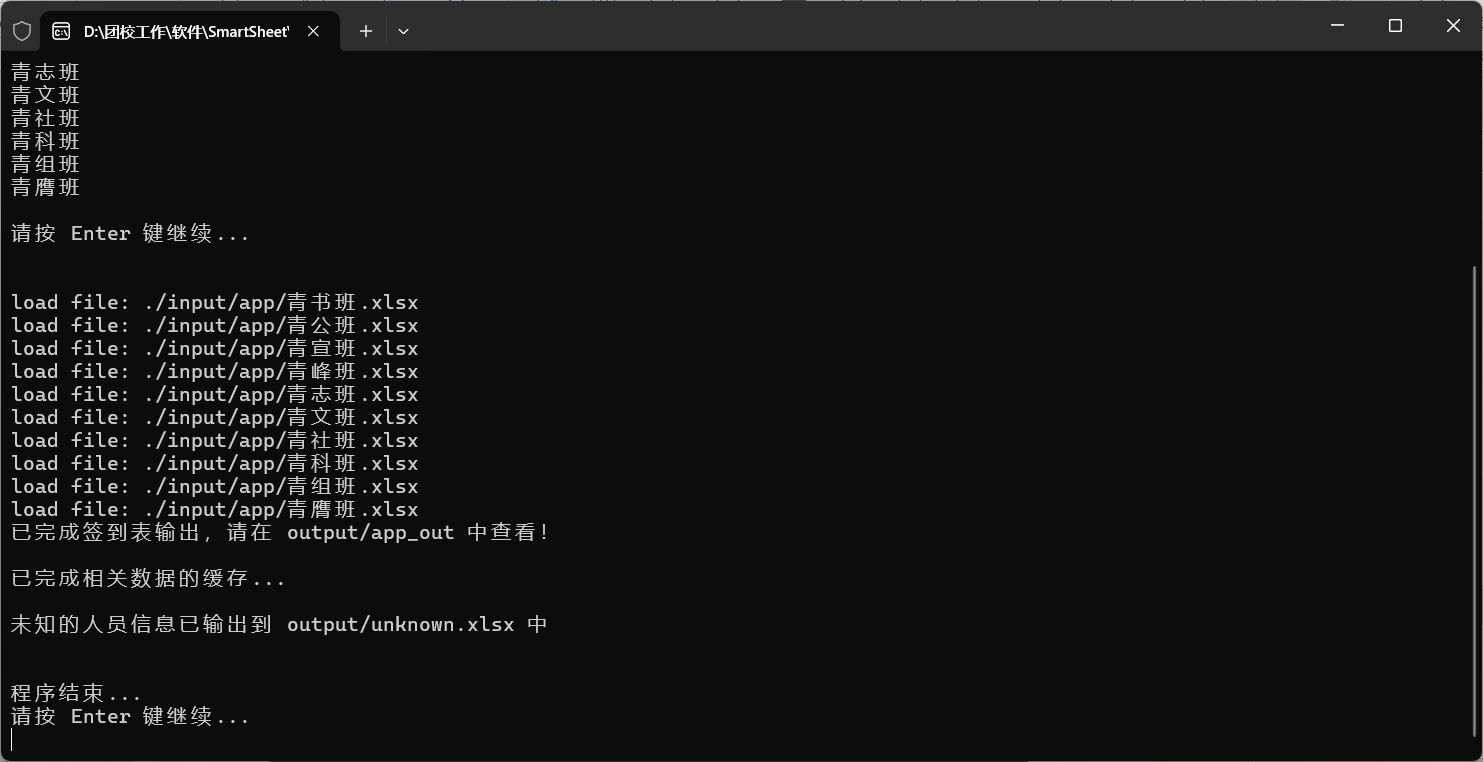


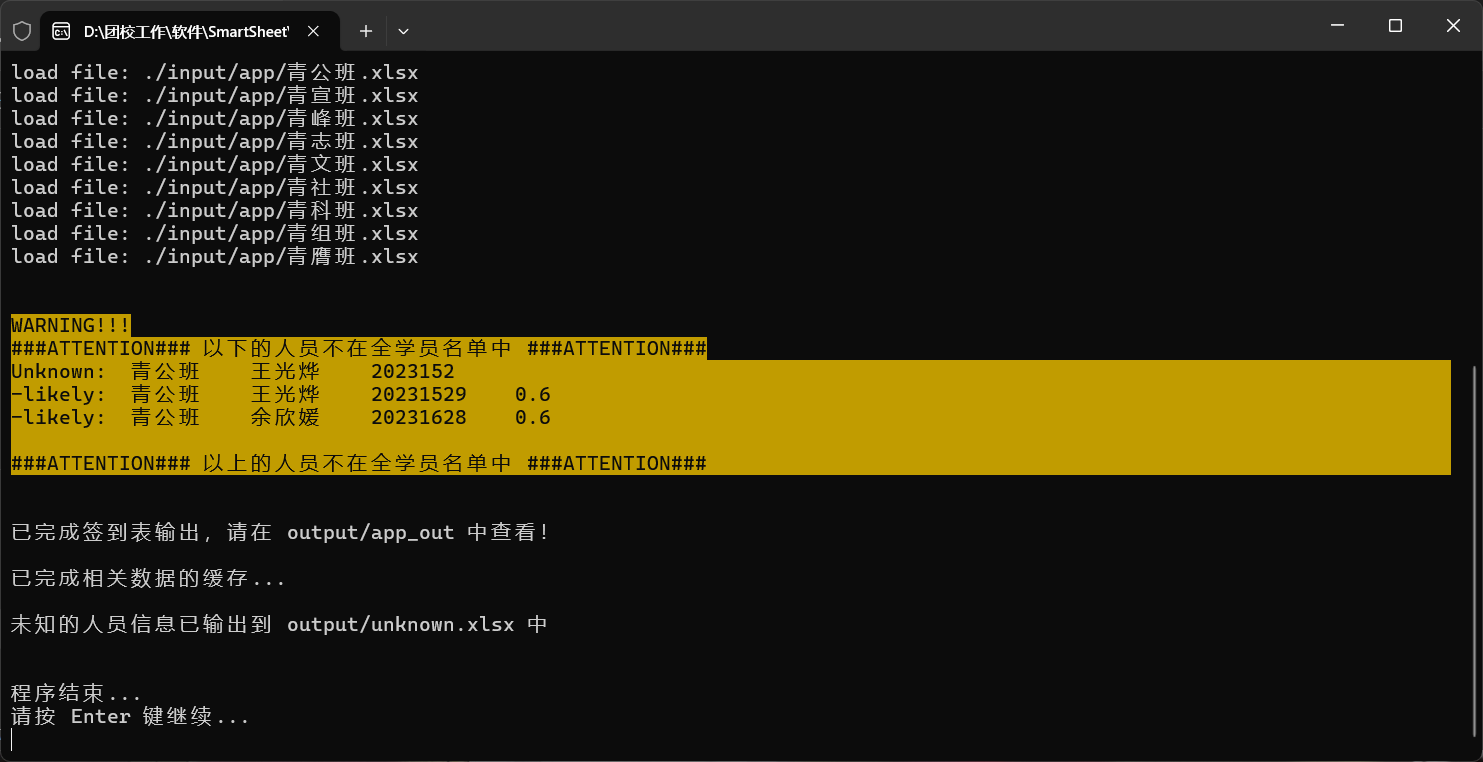
4.如果要生成所有人员报名表，输入“2”按照提示按回车；如果要生成的签到表有班委提供的报名表，输入“1”后回车，按照程序操作接着按下回车即可；

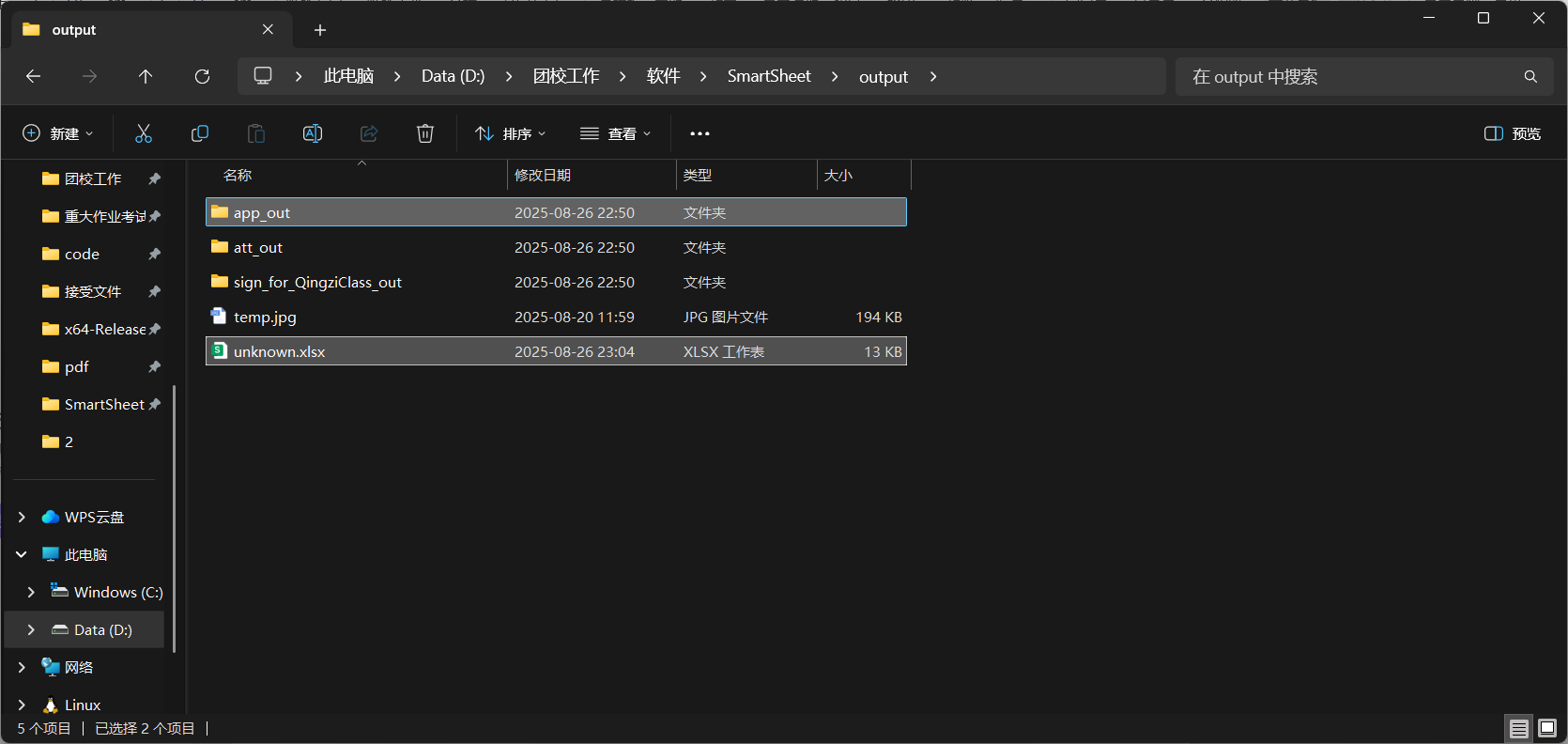




5.输出的签到表在文件夹“output/app\_out”中，如果班委提供的报名表中存在花名册中不存在的人员，程序会将这些人员信息输出到“output/unknown.xlsx”中，需要手动校正“input”中的人员信息，反复运行以上程序，直到黄色的WARNING消失为止，切勿直接修改输出的签到表；

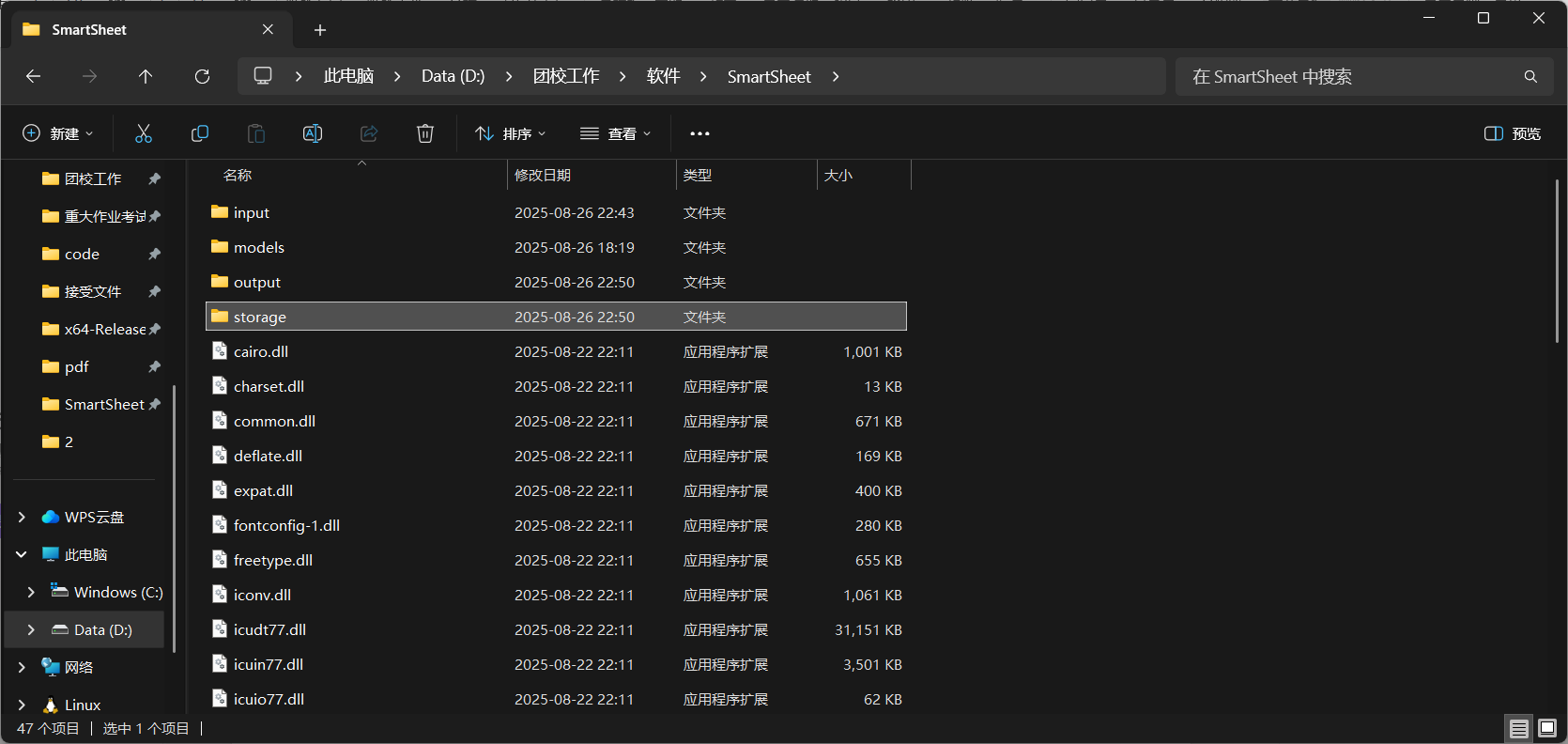


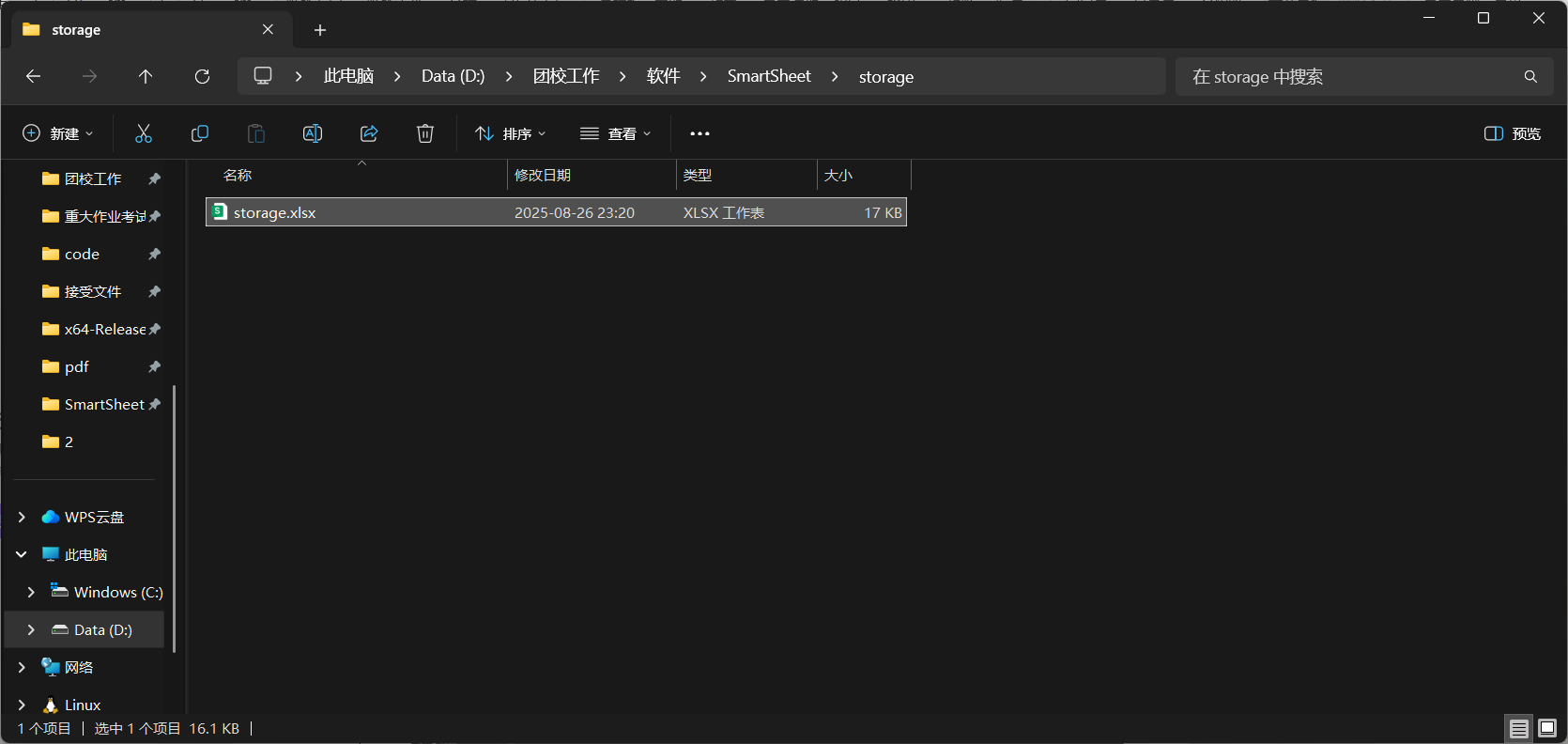




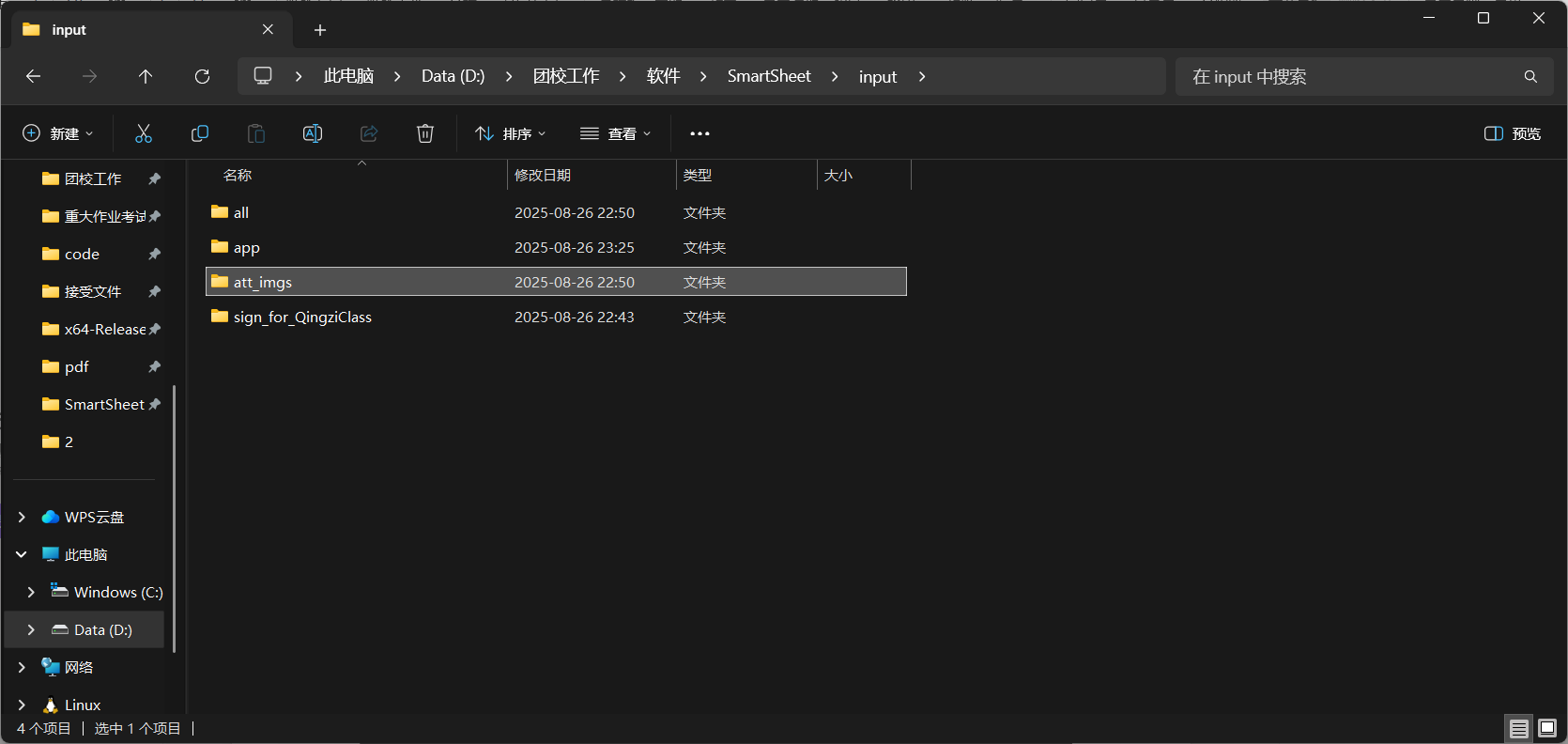
（二）生成签到汇总表

1.首先确保，你的电脑是运行了“生成签到表”功能的电脑，或者是下载了运行了“生成签到表”功能的电脑的缓存文件“storage.xlsx”；



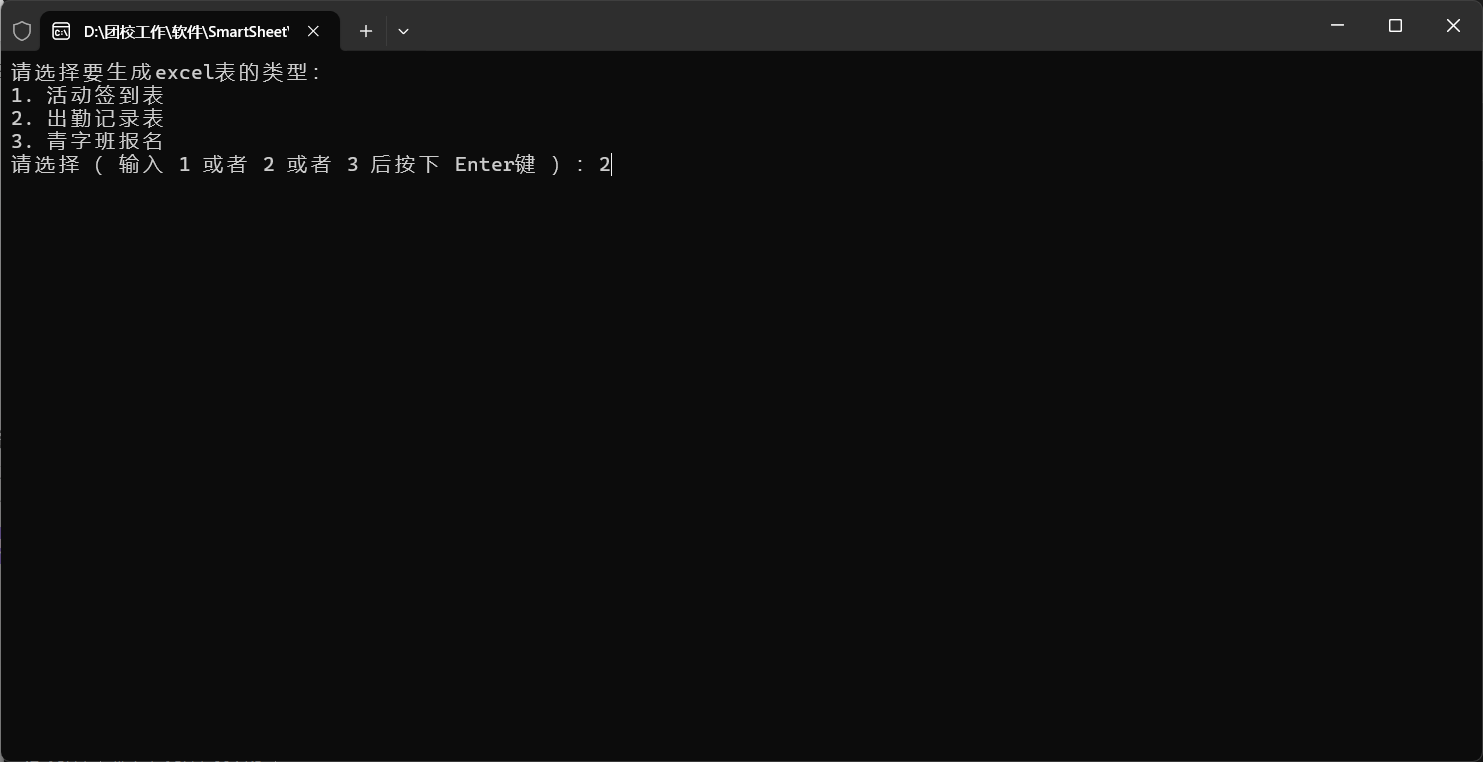


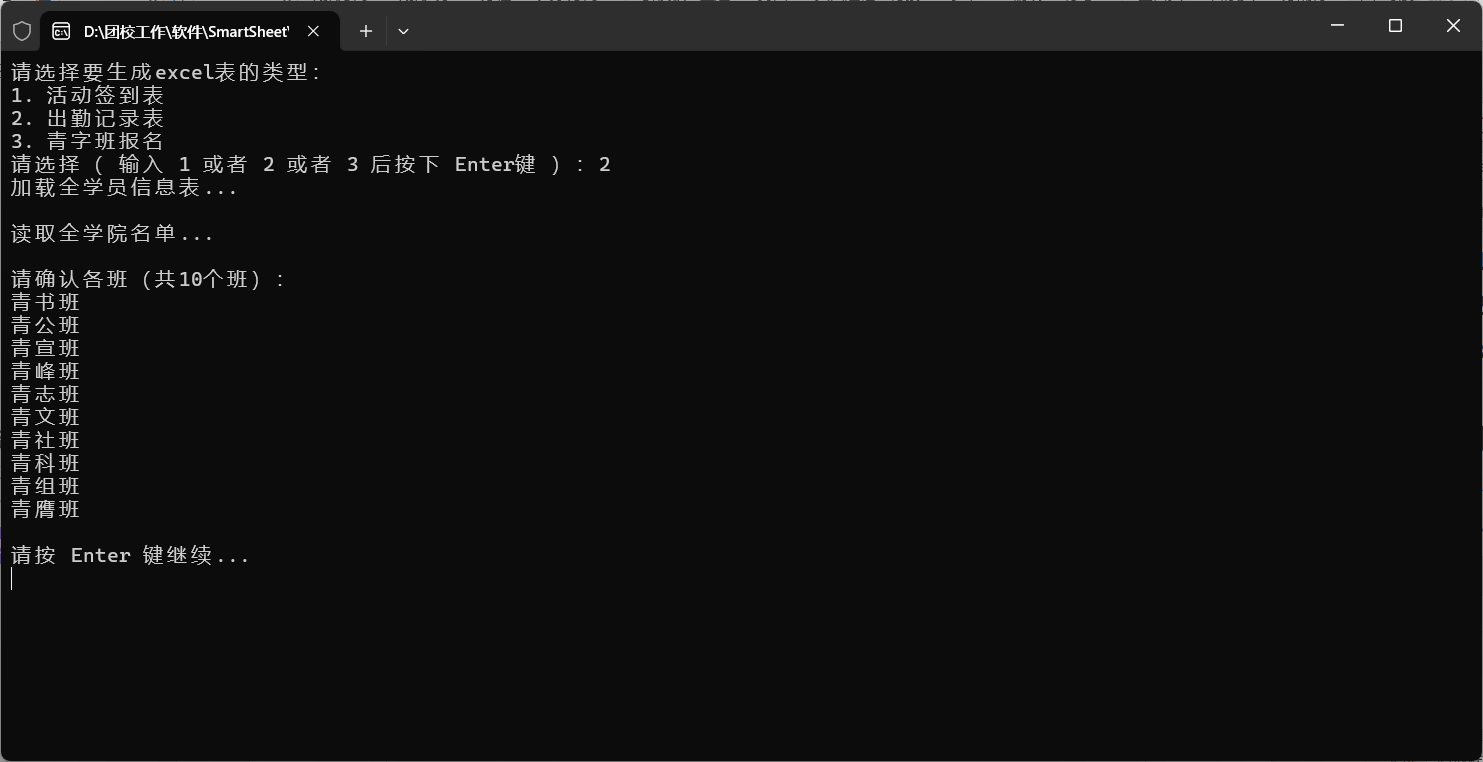
2.确保文件夹“input/att\_imgs/”中，青字班各班线下签到照片按照标准命名，且已经过校正（照片中只能有打印表格的A4纸）与增强，并且没有被打开；

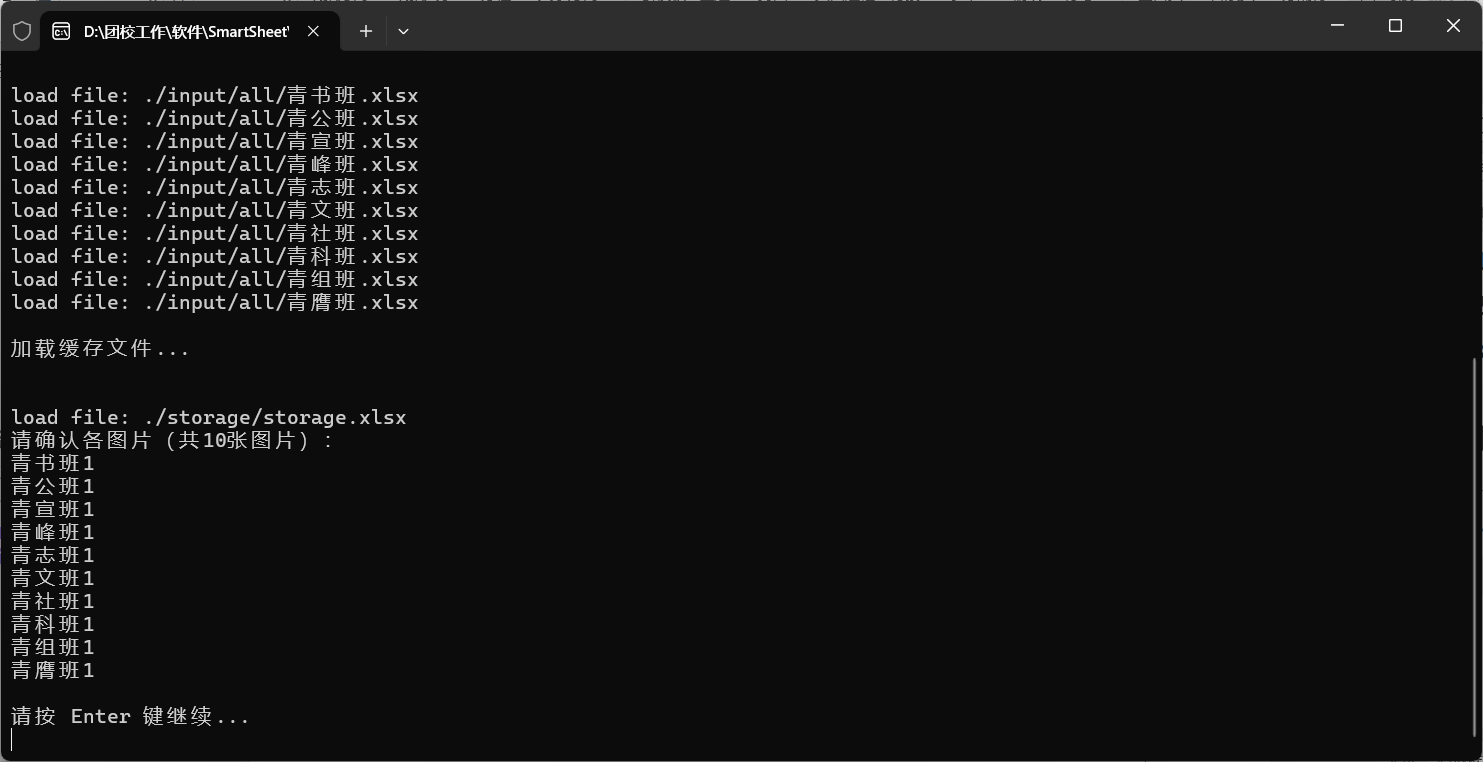


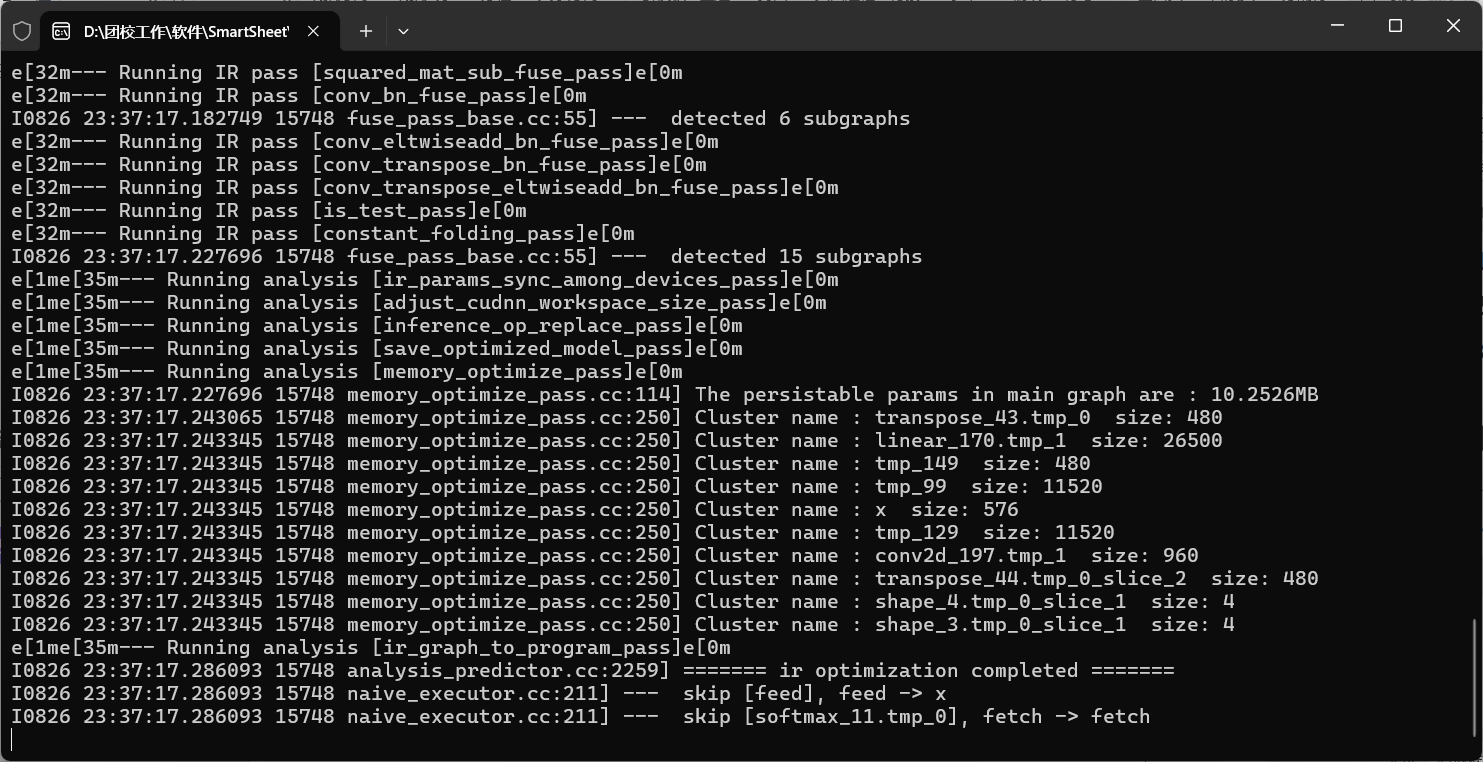


3.输入“2”之后按照回车，按照提示确认文件夹各个文件的数量与名称都是正确的，之后进入到OCR解析图片，如果图片较多，需要等待较长时间；

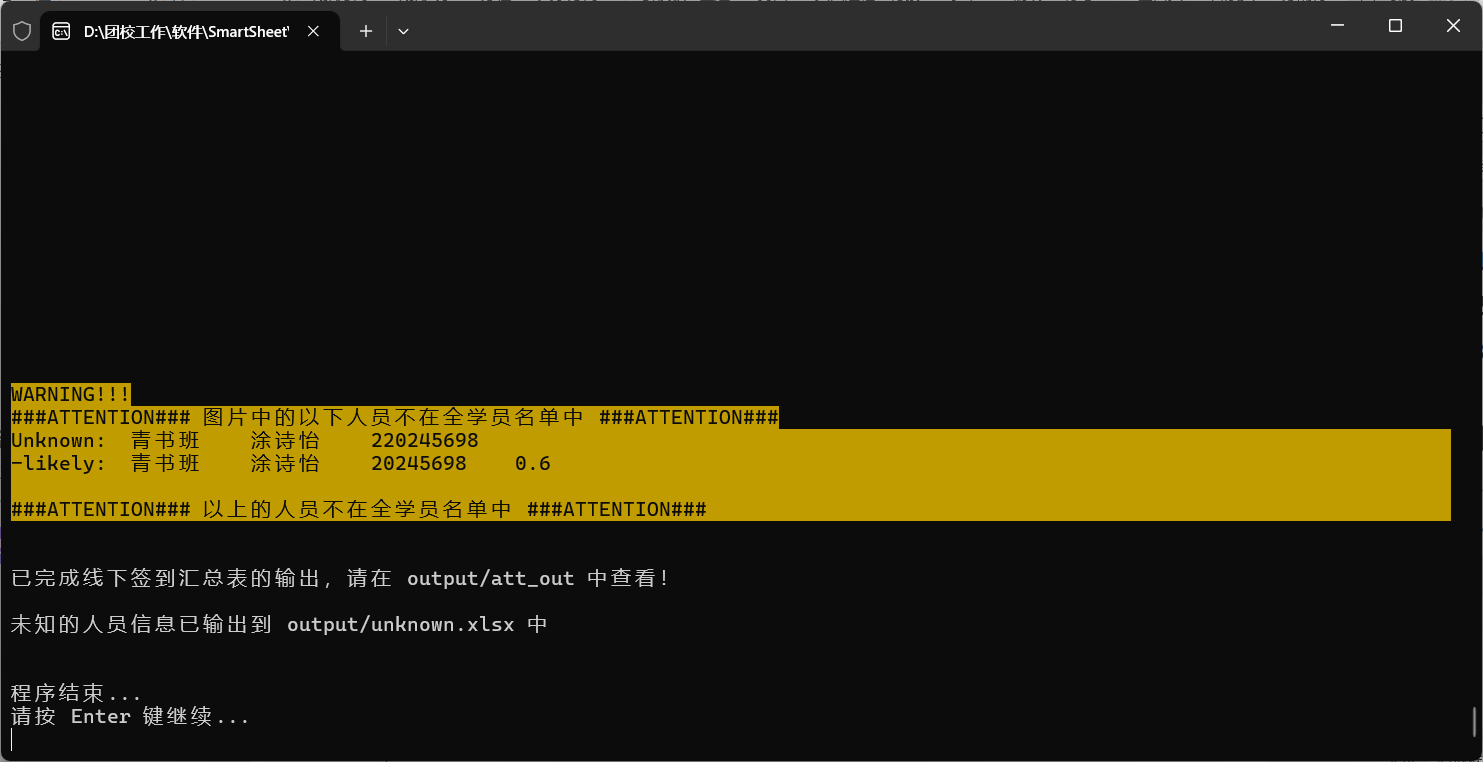






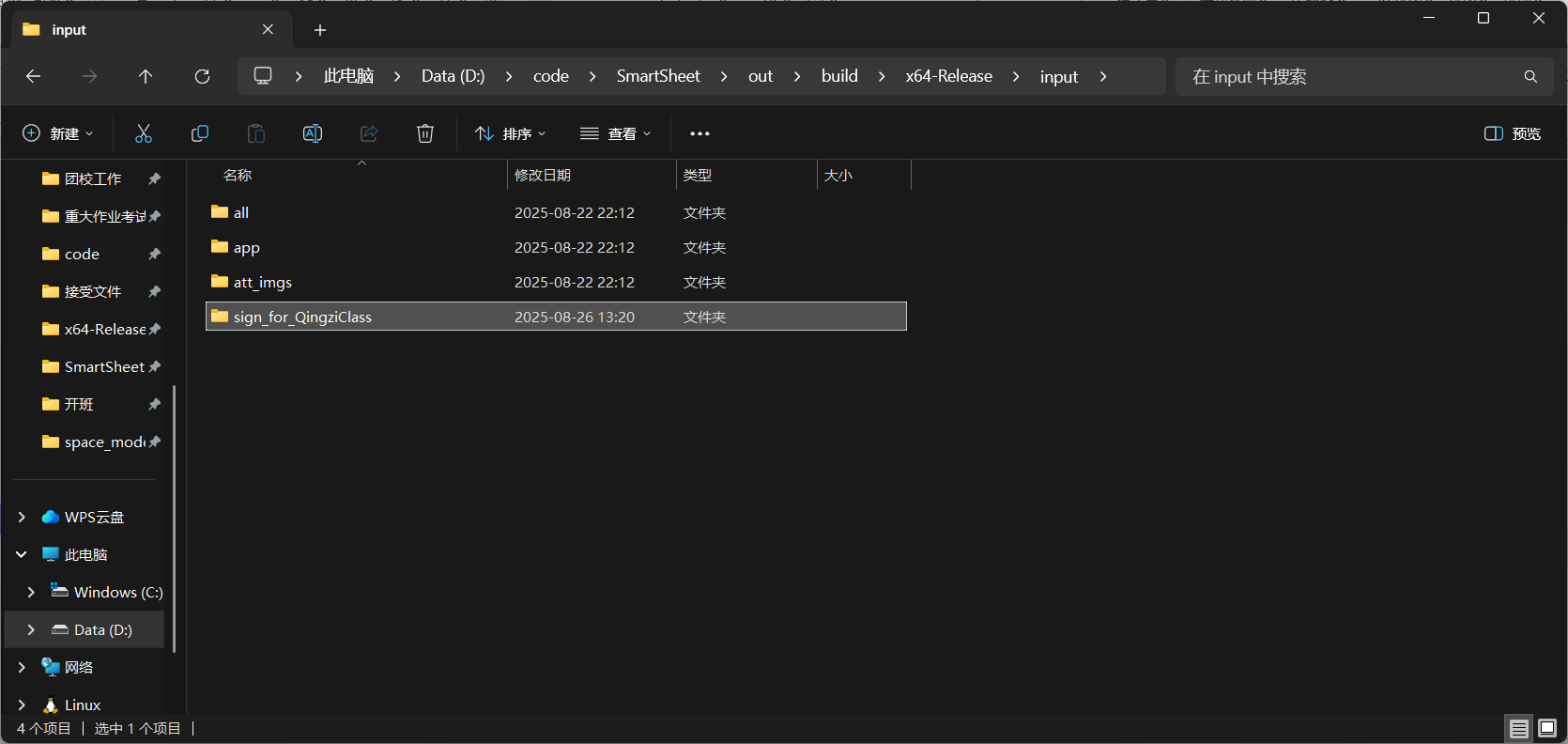


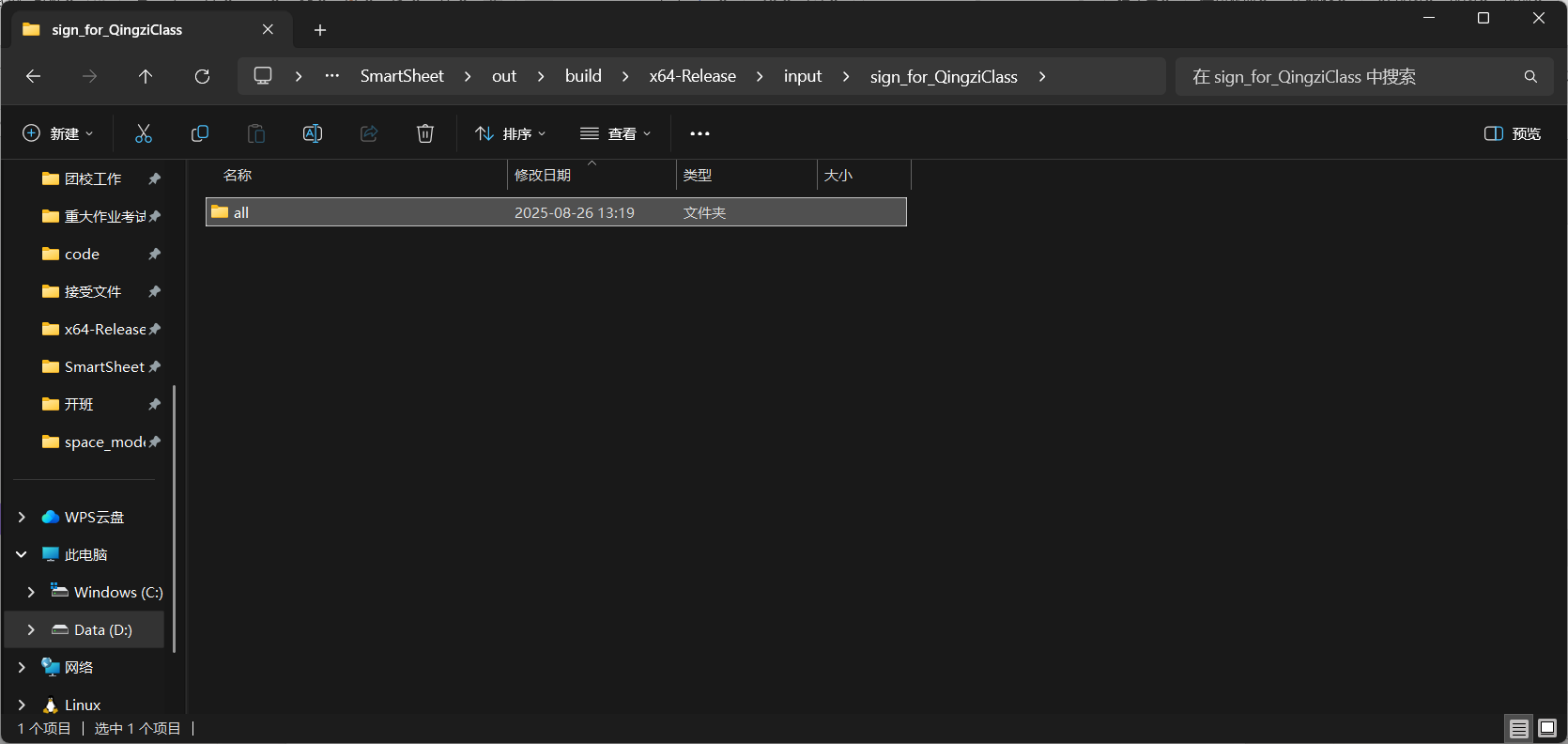
4.程序会把线下签到汇总表输出到“output/att\_out”，未知的人员会输出到“output/unknow.xlsx”，注意校对修改输出的签到表；



**（三）****统计每学期青字班报名的人员**

1.首先确保“input/sign\_for\_QingziClass/all”中，已经将所有的压缩文件解压成文件夹；

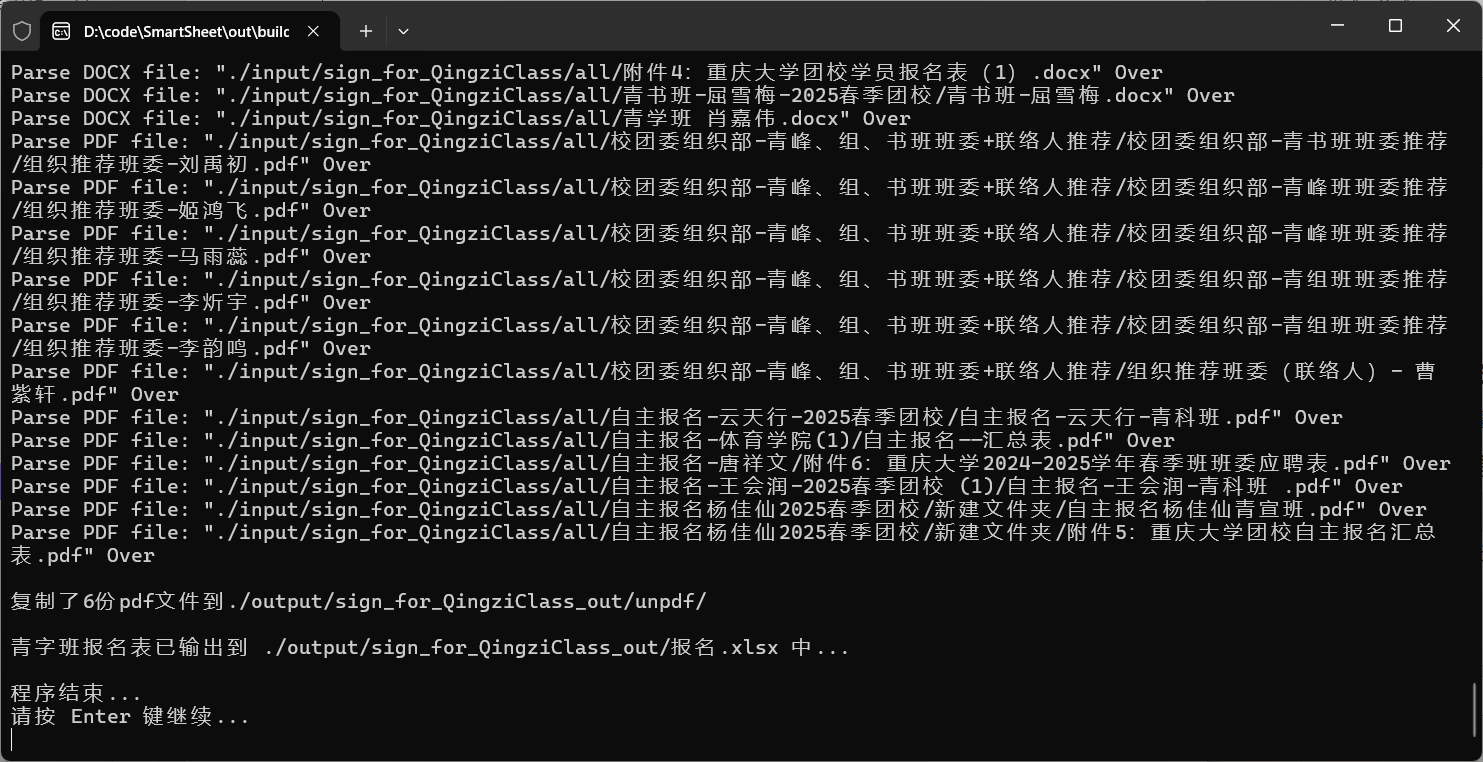




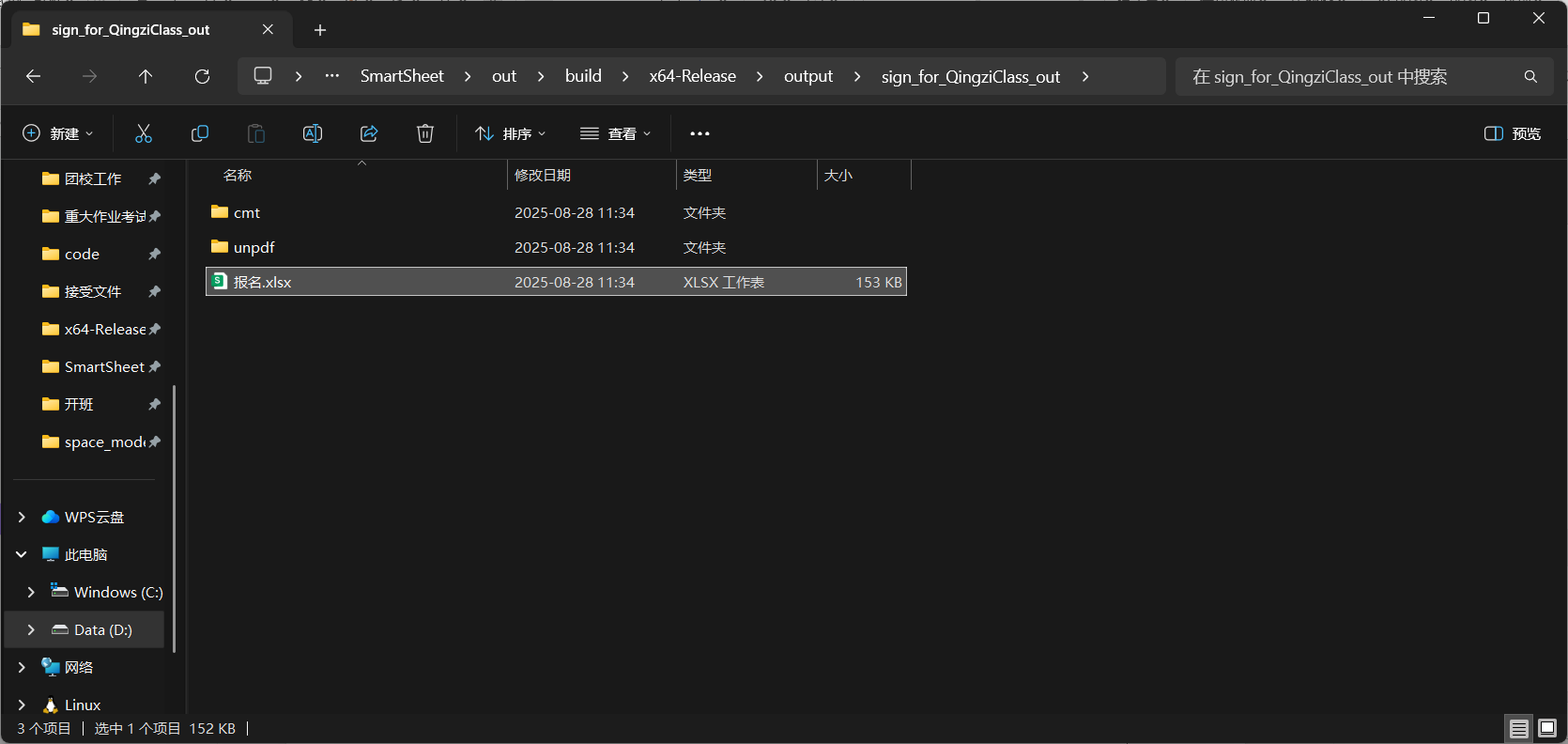


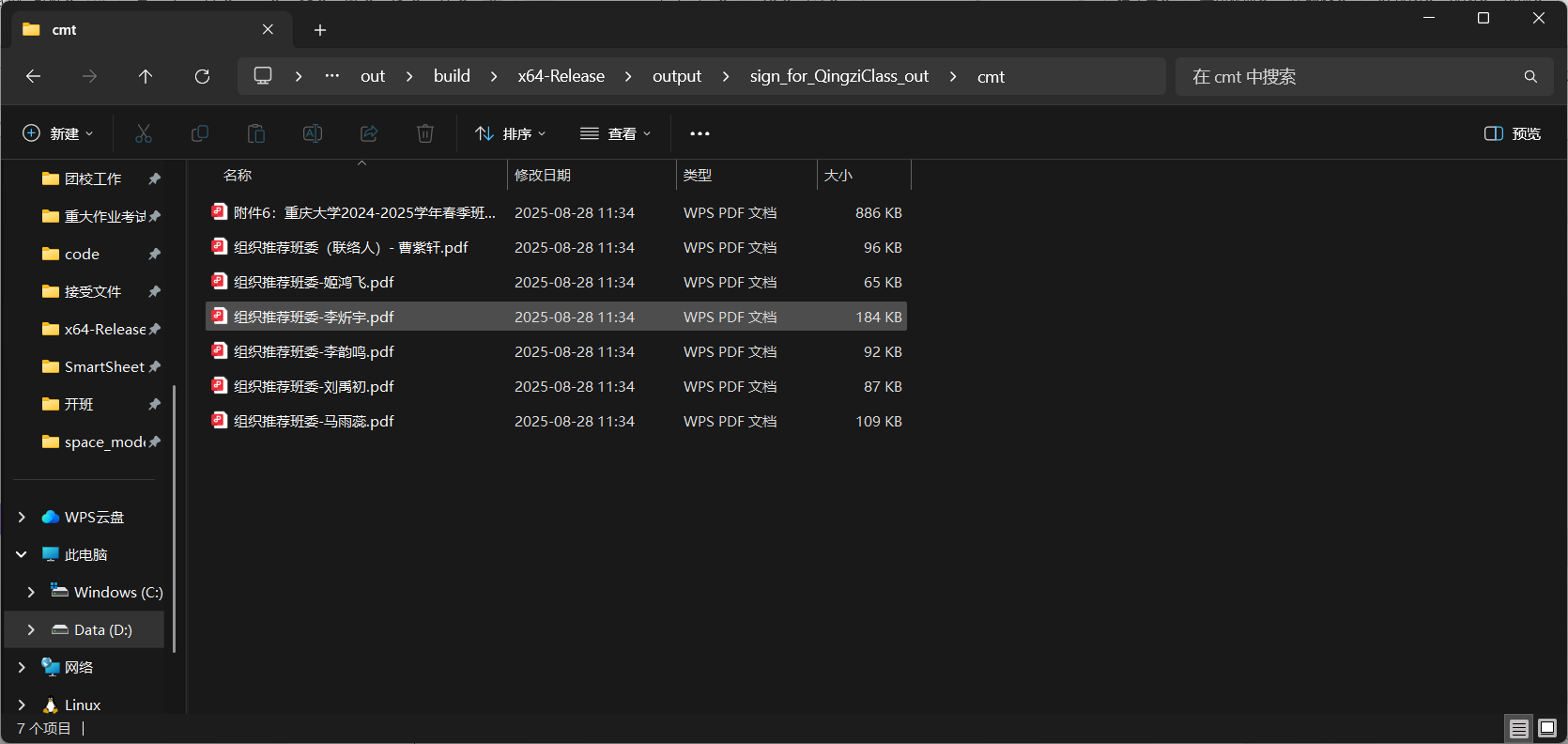
2.输入“3”后按下回车，程序会自动工作，解析所有的docx与pdf文件；

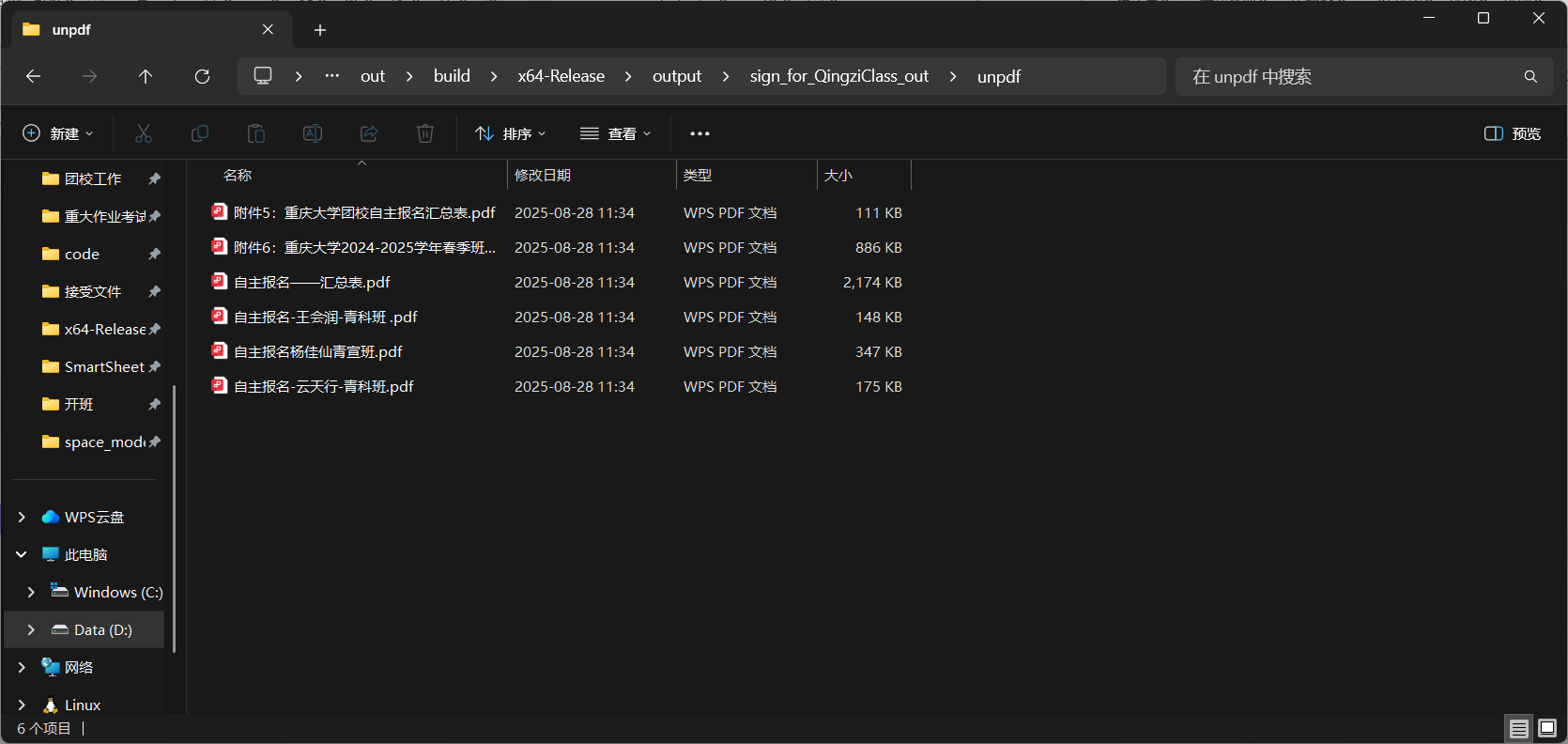




3.输出的统计的信息的excel表在“output/sign\_for\_QingziClass\_out/报名.xlsx”中，班委应聘的pdf输出在“output/sign\_for\_QingziClass\_out/cmt/”中，无法解析的pdf输出在“output/sign\_for\_QingziClass\_out/unpdf/”中；

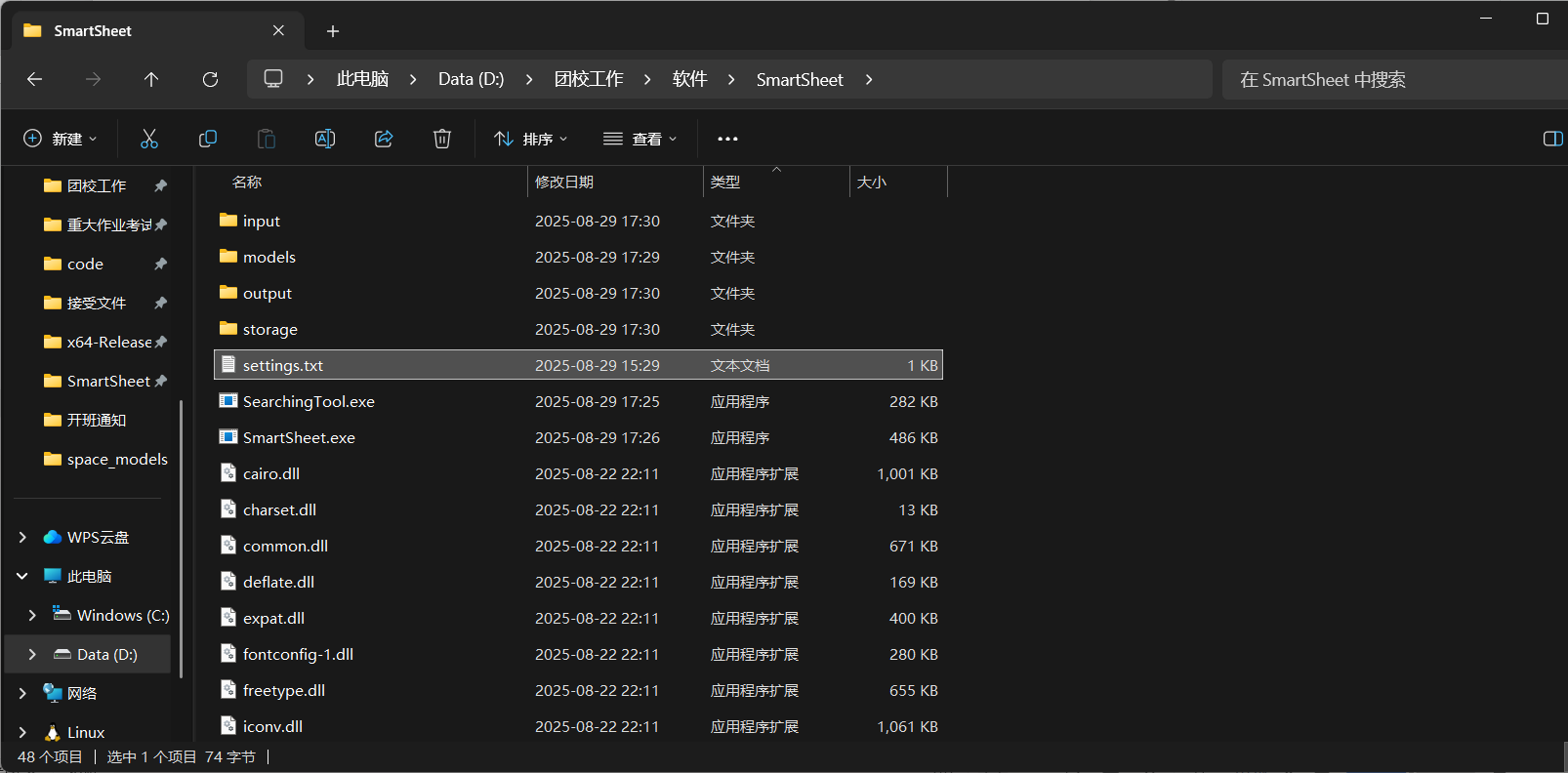






1. **设置文件**

本程序还提供了设置文件“settings.txt”，可以修改功能（二）的PPOCR的模型配置文件，以及启用或者禁用某些功能。文件中的内容可以通过“#”注释“#”之后的内容，被注释的内容不会被识别。





合法的输入：[键]=[路径/修改值] 合法的键值：

|  |  |
| --- | --- |
| 键 | 含义 |
| rec | 文字识别模型的文件夹。 |
| cls | 文本方向分类模型的文件夹。 |
| det | 文字检测模型的文件夹。 |
| dict | 字典库地址。 |
| ManualDocPerspectiveCorrector | true：启用手动透视矫正；  false：禁用手动透视矫正；  手动透视矫正默认是禁用的。  （[使用方式](#六、手动透视矫正)） |
| ImageEnhancement | 图像增强技术：  auto：将自动对图片进行增强（此功能启用后其他ImageEnhancement功能将失效）；  removeshadow：去除阴影，将对照片的阴影去除，同时做白平衡；  sharpen：轻度锐化，对照片锐化，如果图片的有效内容占图片面积过小，有可能识别不到文字和网格线。  推荐使用auto（已经增强的图片可以不用开启）。 |

推荐的文字识别模型：ch\_PP-OCRv4\_rec\_infer

如果不准：ch\_PP-OCRv4\_rec\_server\_infer（使用这个将会花费更多时间）

四、各个文件夹说明与规范

（一）合法的表头关键词

以下关键词是绝对合法的（非法的关键词也不会报错）：

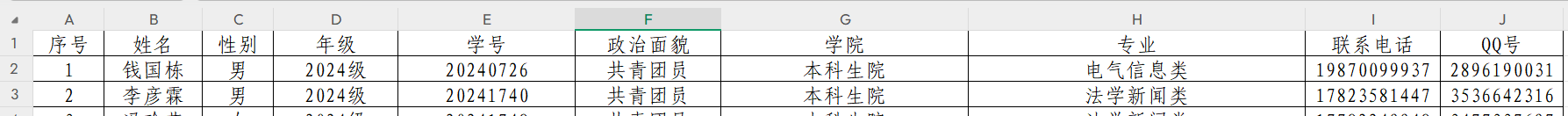
| **姓名** | **性别** | **民族** | **年级** | **学号** | **政治面貌** | **学院** | **专业** | **学生职务** | **社团** | **联系电话** | **QQ号** | **邮箱** | **报名青字班** | **是否报名班委** | **应聘岗位** | **报名方式** | **个人简介** | **个人特长** | **工作经历** | **获奖情况** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

（二）input/all

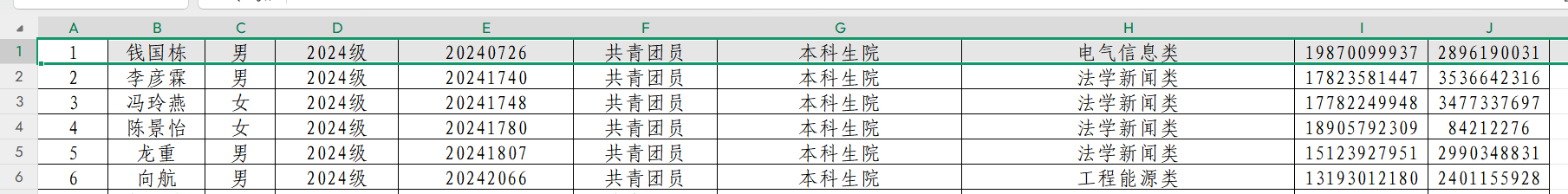
此文件夹用于储存青字班花名册各班的xlsx文件。内部所有文件请**一定**按照“青X班.xlsx”命名。每一份xlsx文件都**必须具有表头**，**不能具有标题**，**不能具有合并的单元格**。工作簿只能有一个，工作区内不能具有空单元格，若有空单元格，请使用“-”占位。

表头应该至少具有以下[关键词](#合法的表头关键词)：“姓名”“学号”“学院”“联系方式”（只能如此命名）。

✔以下的格式是正确的：



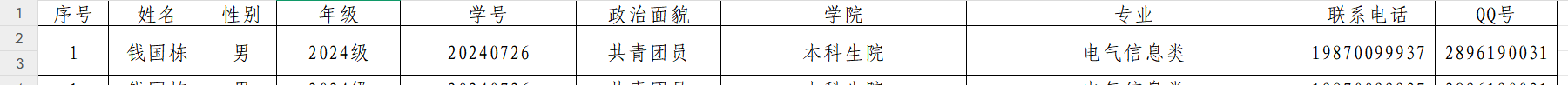
❌没有表头是错误的：



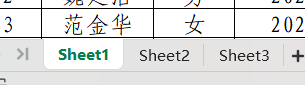
❌具有标题是错误的：



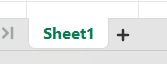
❌具有合并的单元格是错误的：



❌具有多个工作簿是错误的：



✔以下的格式是正确的：



❌工作区内具有空单元格是错误的：



✔以下的格式是正确的：

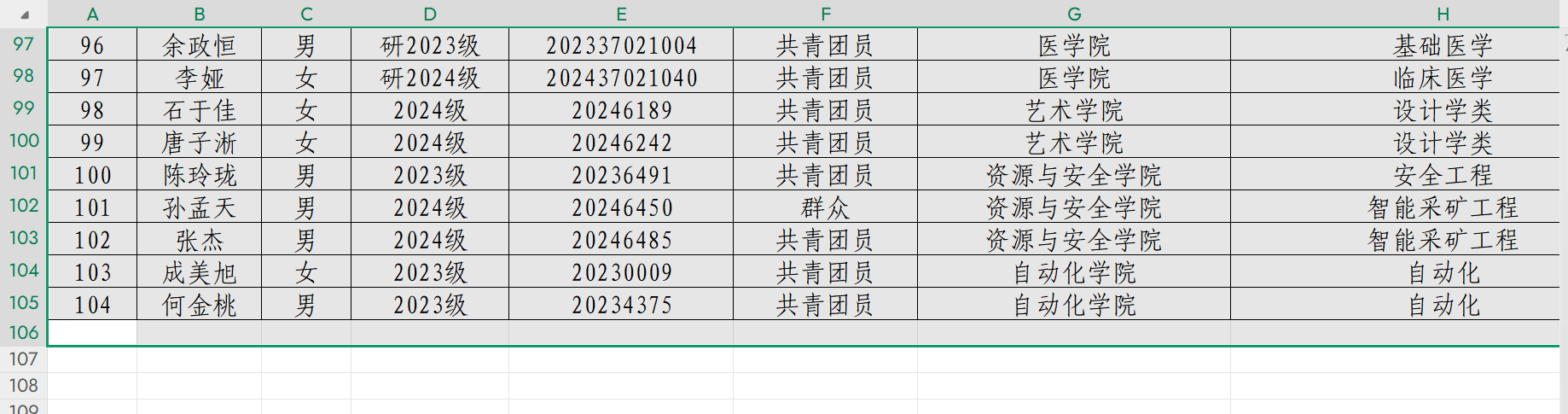


判断工作区是否具有空格，**选中具有文字的单元格**，按下“Ctrl+A”全选工作区。

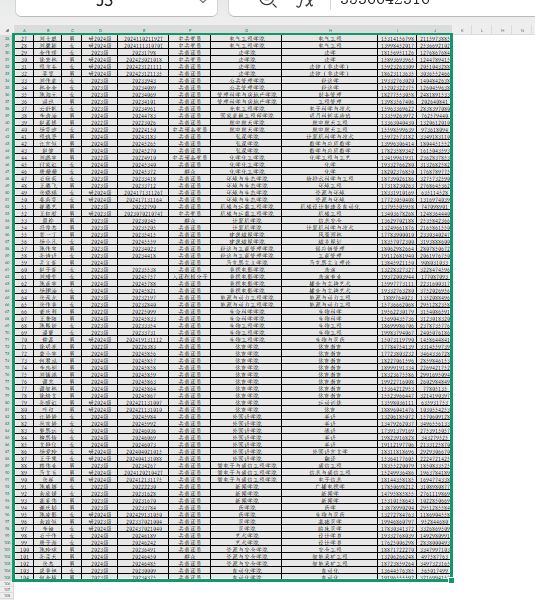
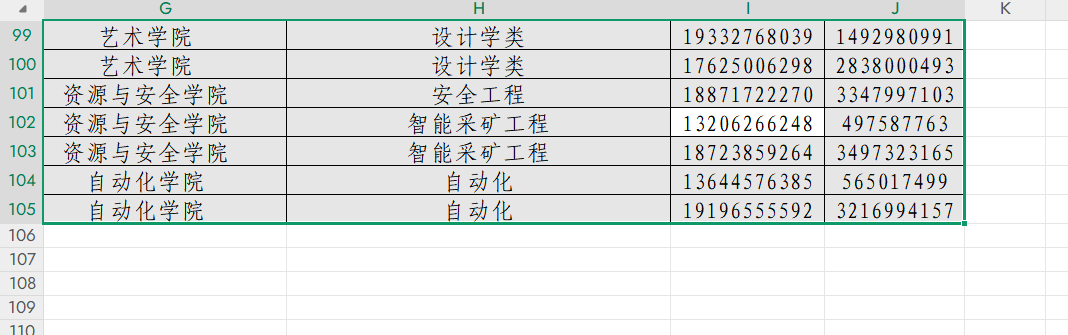
❌全选选中了空单元格是错误的，应当删除K列：



❌全选选中了空单元格是错误的，应当删除106行：



✔全选后，以下的格式是正确的：

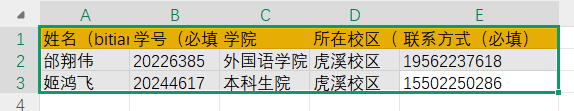
（三）input/app

此文件夹用于储存青字班班委发来的报名表的xlsx文件（若后缀不为xlsx，请另存为xlsx）。内部所有文件请**一定**按照“青X班.xlsx”命名。每一份xlsx文件都**必须具有表头**，**不能具有标题**，**不能具有合并的单元格**。工作簿只能有一个，工作区内不能具有空单元格，若有空单元格，请使用“-”占位。

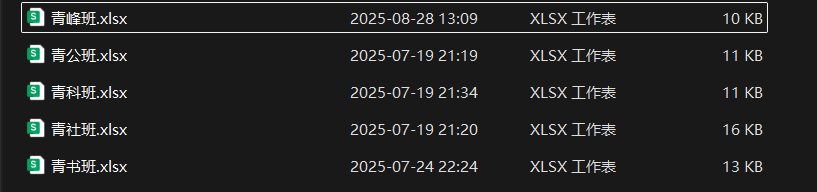
（具体图片与“[（一）input/all中示例](#xlsx标准示例图片)”同理）

表头应该至少具有以下关键词：“姓名”“学号”。（表头关键词中可以具有其他内容，如“姓名（必填）”“学号（必填）”，但是**关键词必须连在一起**）

✔以下的格式是正确的：



✔以下的命名与后缀是正确的：



（四）input/att\_imgs

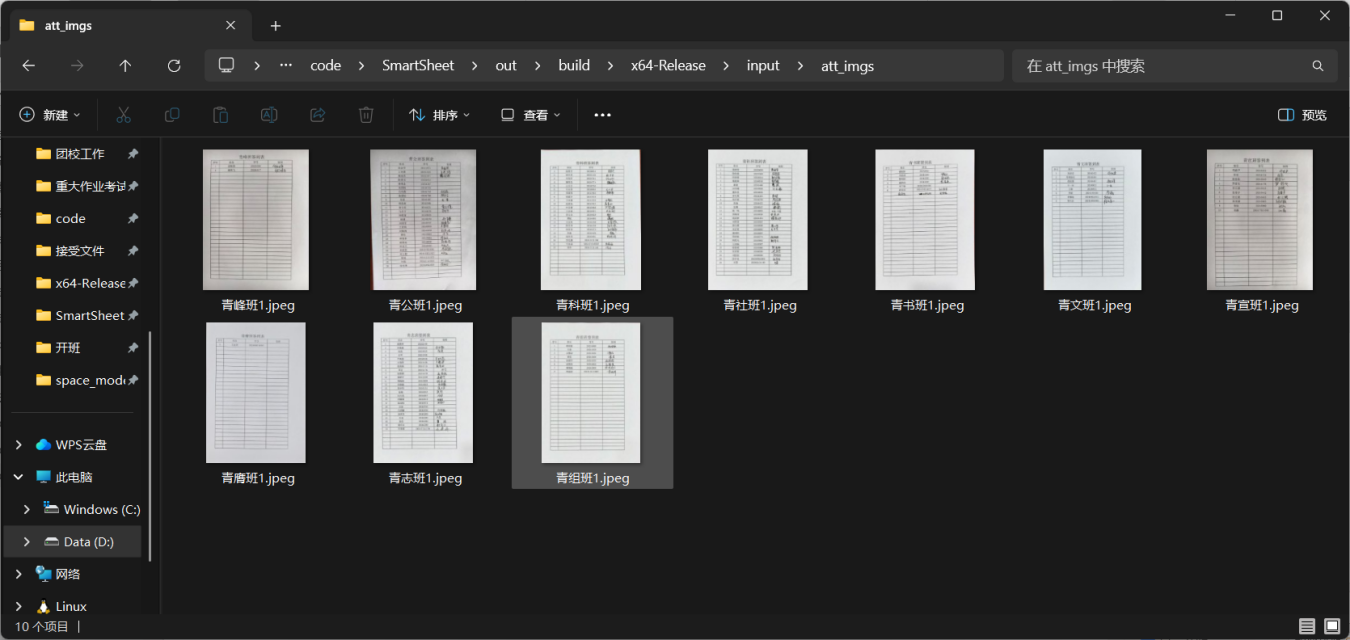
此文件夹用于存放青字班的线下签到的照片。照片按照形如“青X班1.jpeg”“青X班2.jpeg”“青X班3.jpeg”的形式来命名。同一个班的照片用班级名之后加的数字来区分。

图片要保证：**~~表格的各横线水平等长，各竖线竖直等长，无明显的透视现象；纸张要充满图片，不能有明显的边界~~**（[为你内置了手动透视校正技术](#ManualDocPerspectiveCorrector)）（[使用指南](#六、手动透视矫正)）；**~~必须使用对比度增强的图片增强技术~~**（[为你内置了图片增强技术](#ImageEnhancement)）；**除标题外，所有内容必须处于表格线内，超出表格线的内容请自己用尺子补齐表格线**。

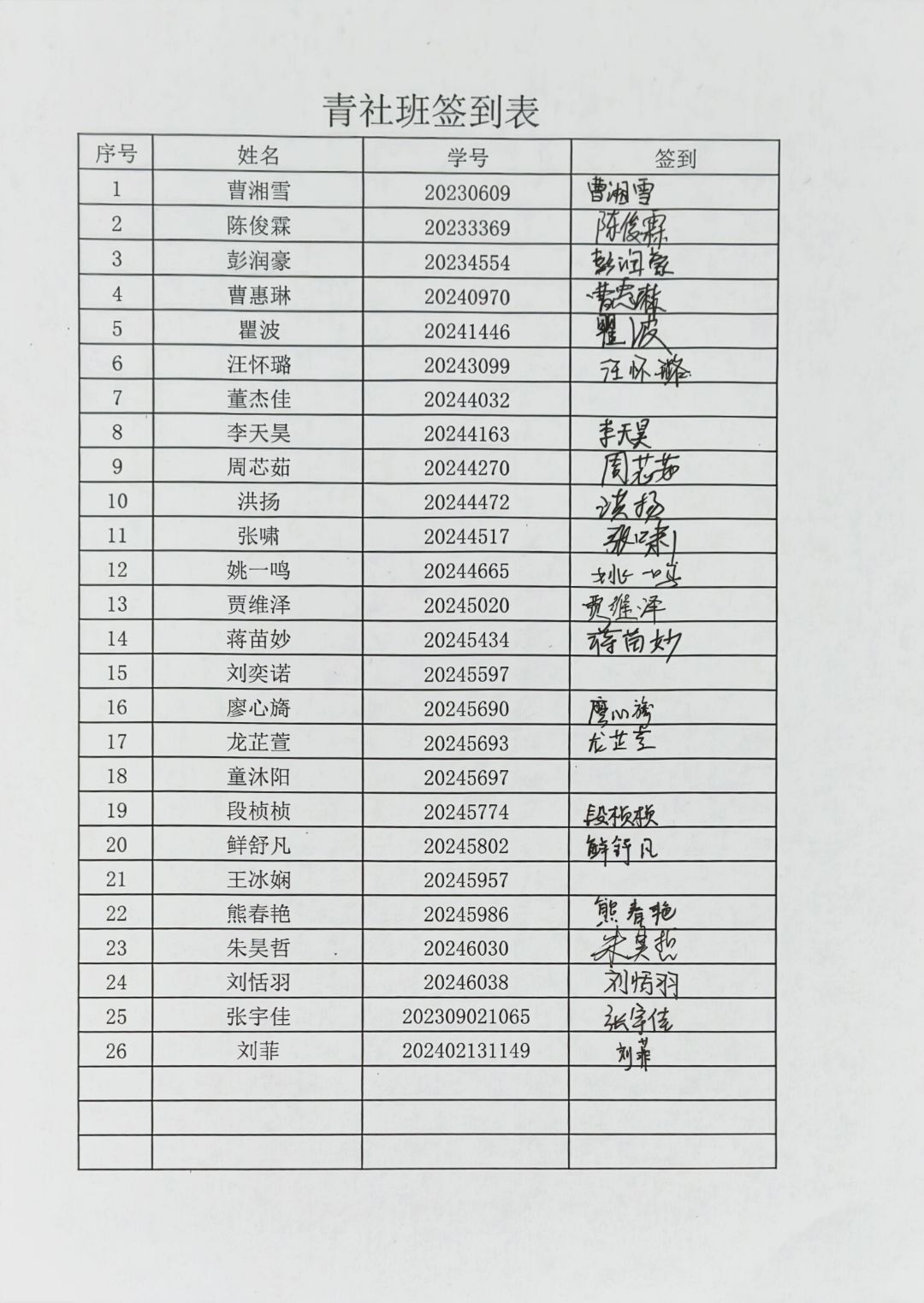
合法的后缀：“.jpg”“.png”“.jpeg”“.tiff”“.tif”“.jpe”“.bpm”“.dib”“.webp”“.raw”。

**推荐**的后缀：“.jpg”“.png”“.jpeg”。

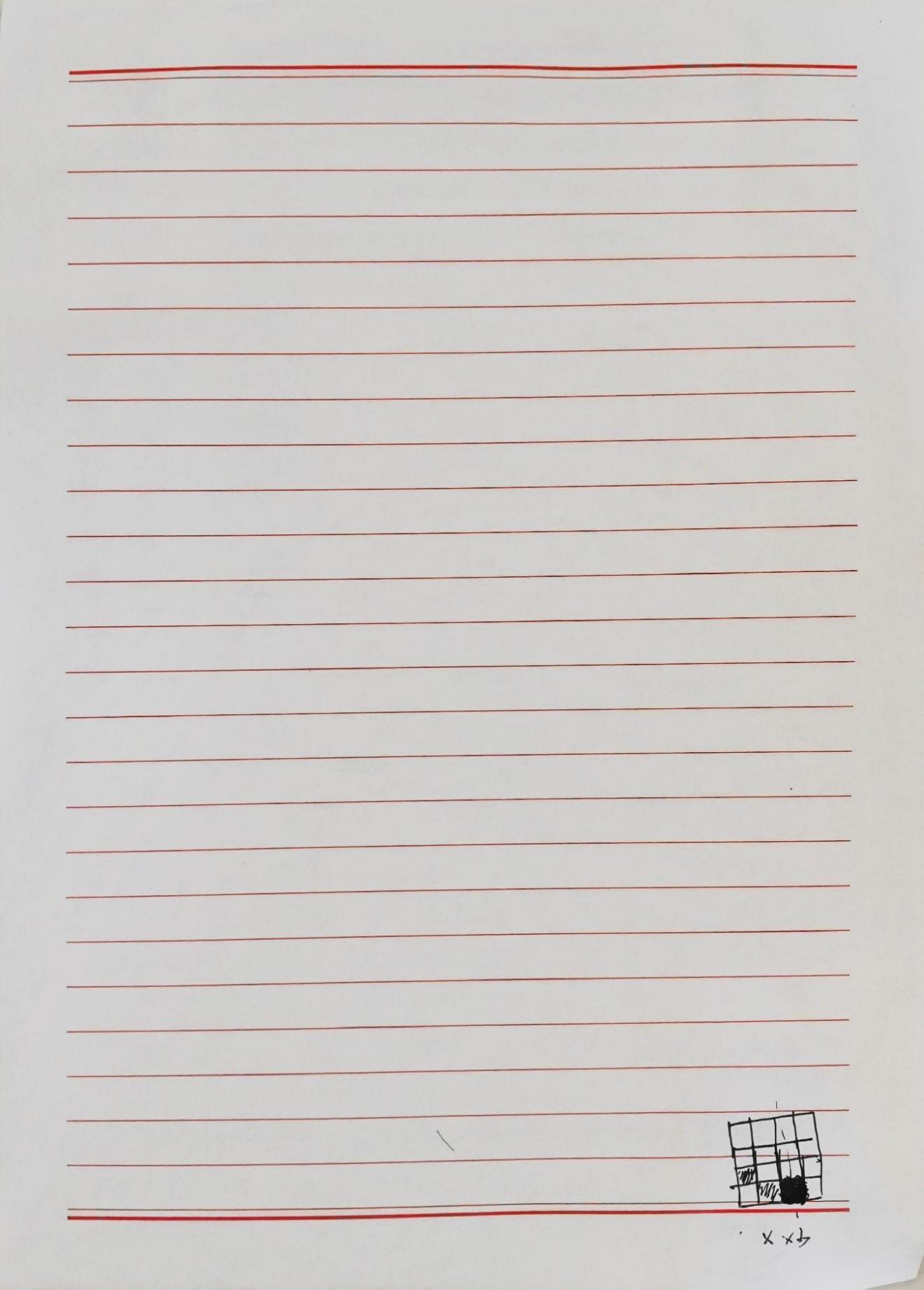
✔以下的命名格式是正确的：



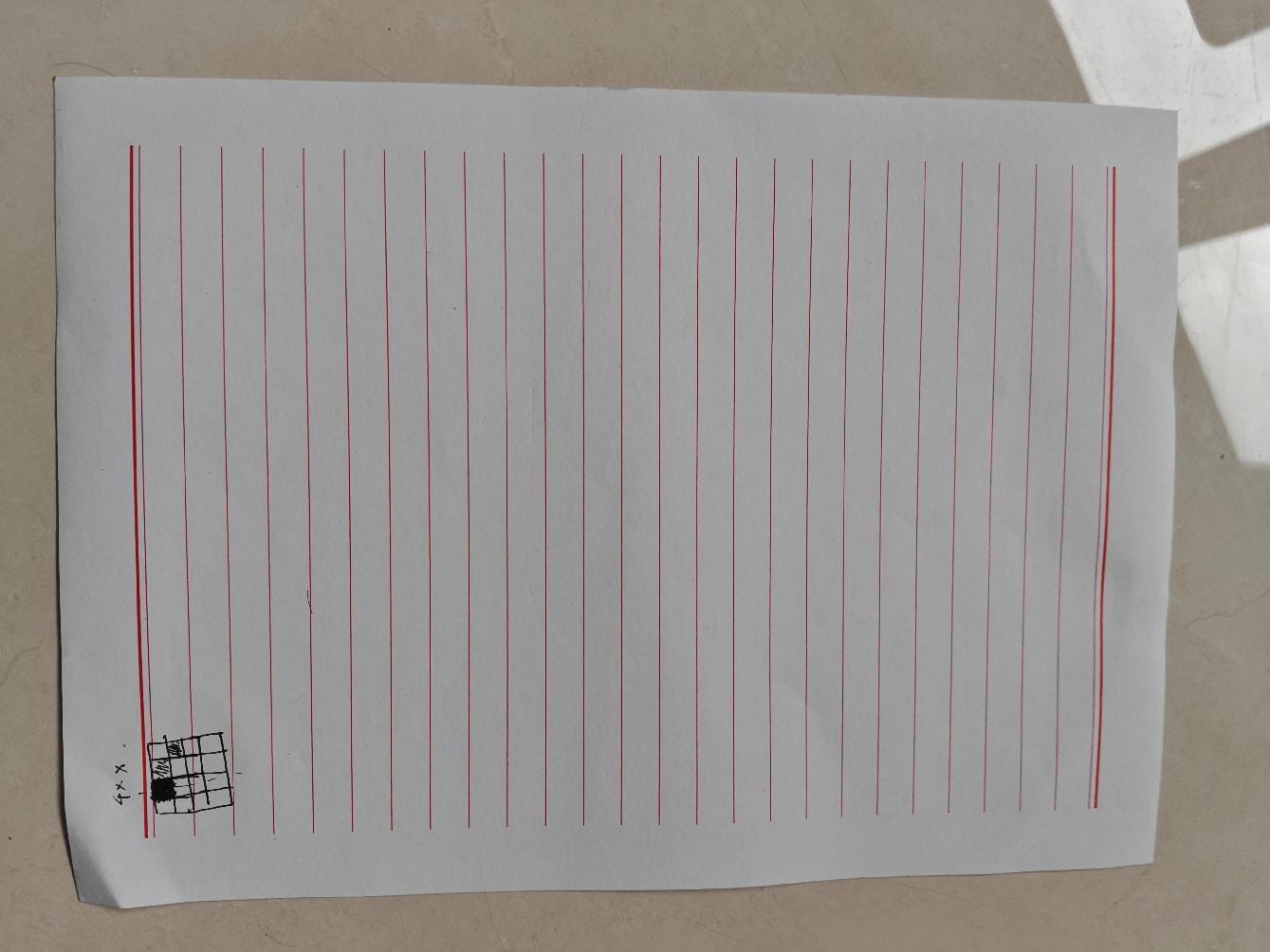
✔以下的图片是正确的：



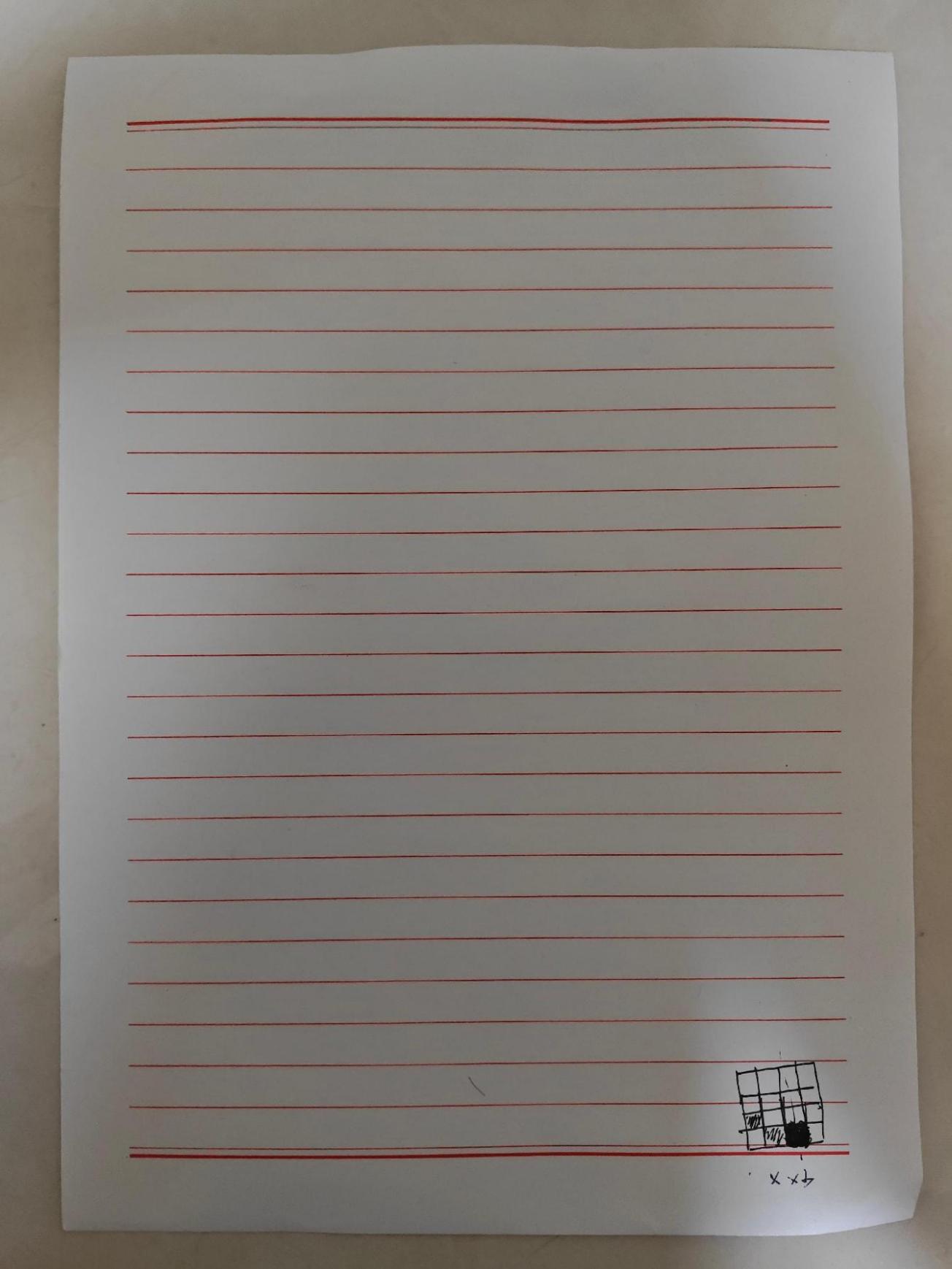
✔以下的图片是正确的（横线略有倾斜，可开启透视矫正）：



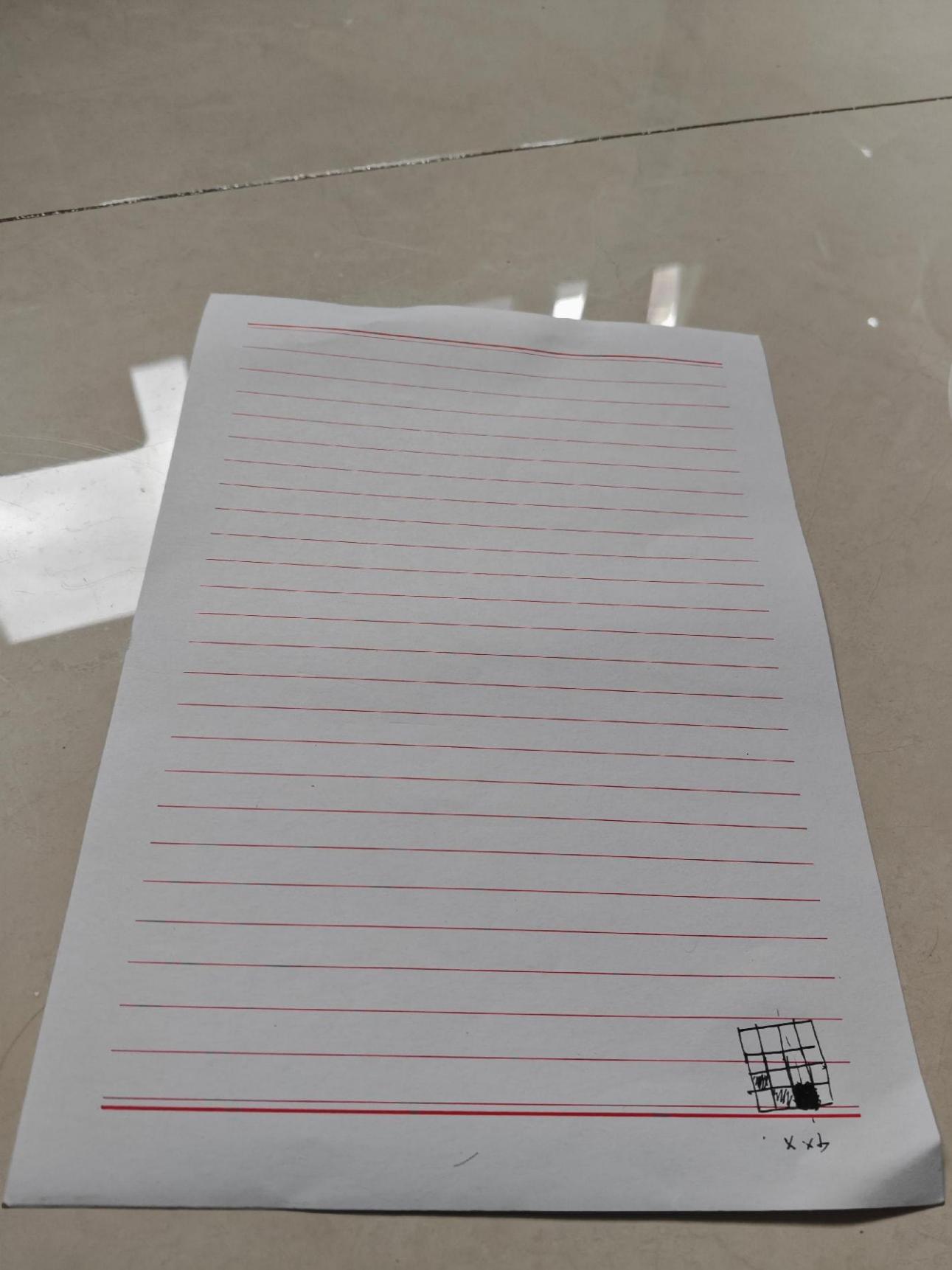
❌没有填充满的照片错误的（如果出现，请开启透视矫正）：



❌没有填充满、没有增强、有阴影的照片错误的（如果出现，请开启透视矫正）：



❌表格的各横线不等长，有明显的透视现象，横线不水平的照片错误的（如果出现，请[开启透视矫正](#ManualDocPerspectiveCorrector)）：

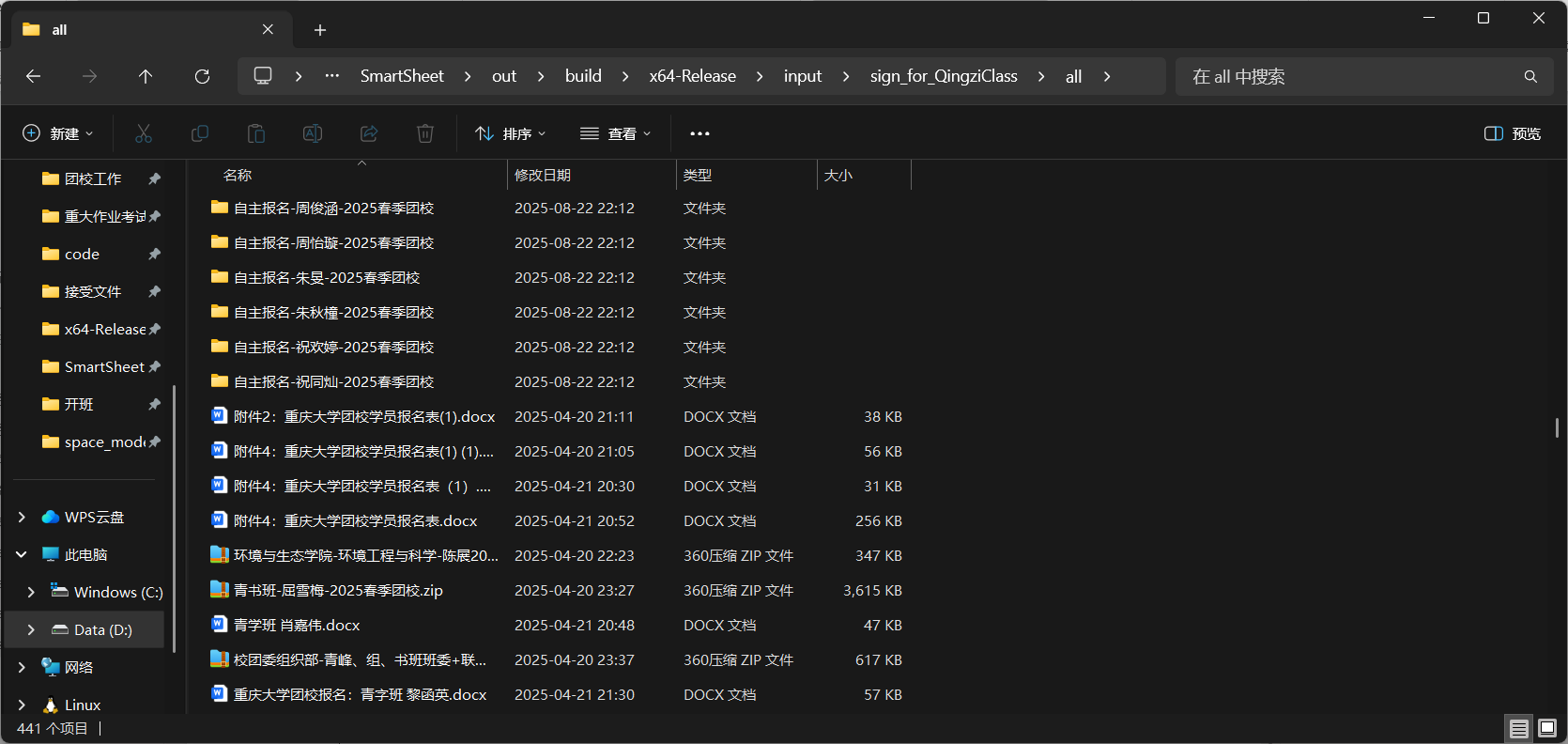


（五）input/sign\_for\_QingziClass/all

此文件存放报名同学发来的**已经解压**后的文件，用于青字班报名。

**注意**：**压缩文件（无论什么格式）无法被识别**。

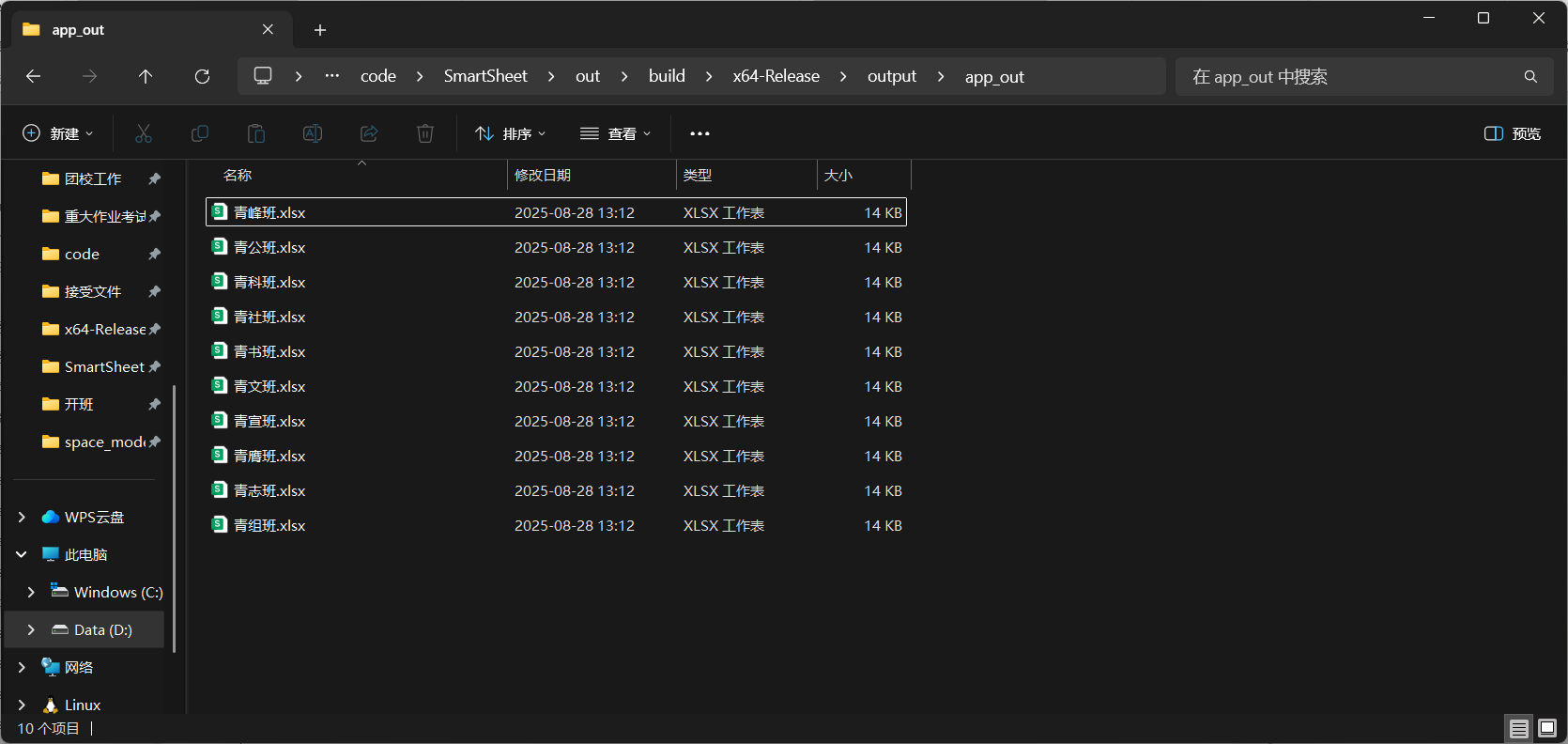
✔文件夹、docx、pdf、zip、rar等文件混合是正确的（但是压缩文件无法被识别）：

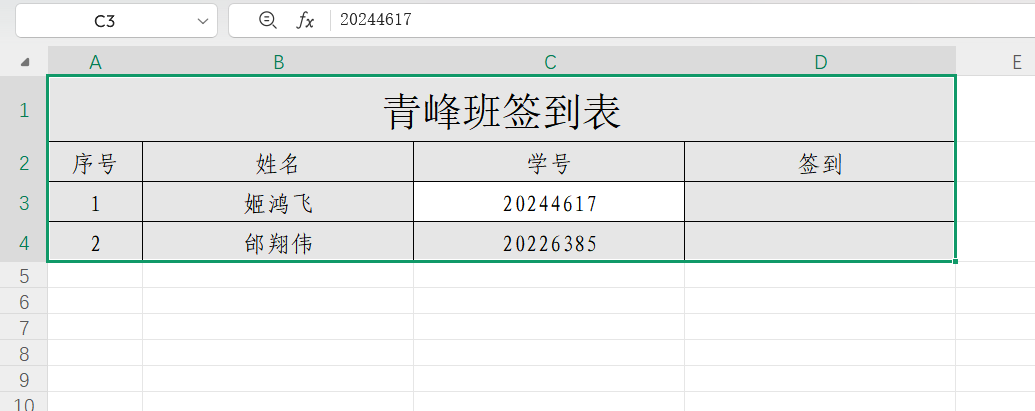


**（**六**）out**put/app\_out

输出线下签到表。

注意：每次运行程序时，此文件夹中的文件都会被删除。

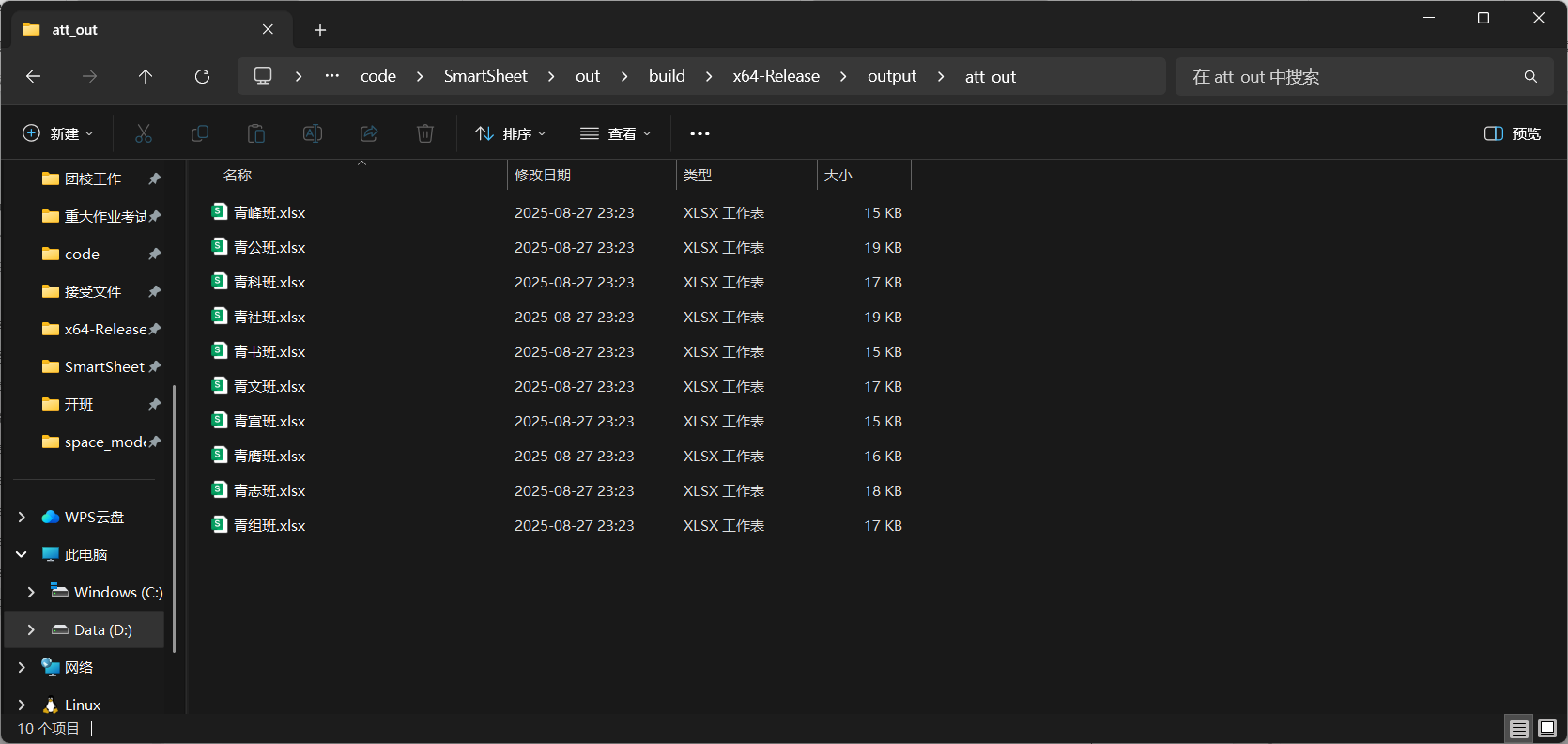




（七）output/att\_out

输出线下签到汇总表。（排序功能还在开发）

注意：每次运行程序时，此文件夹中的文件都会被删除。

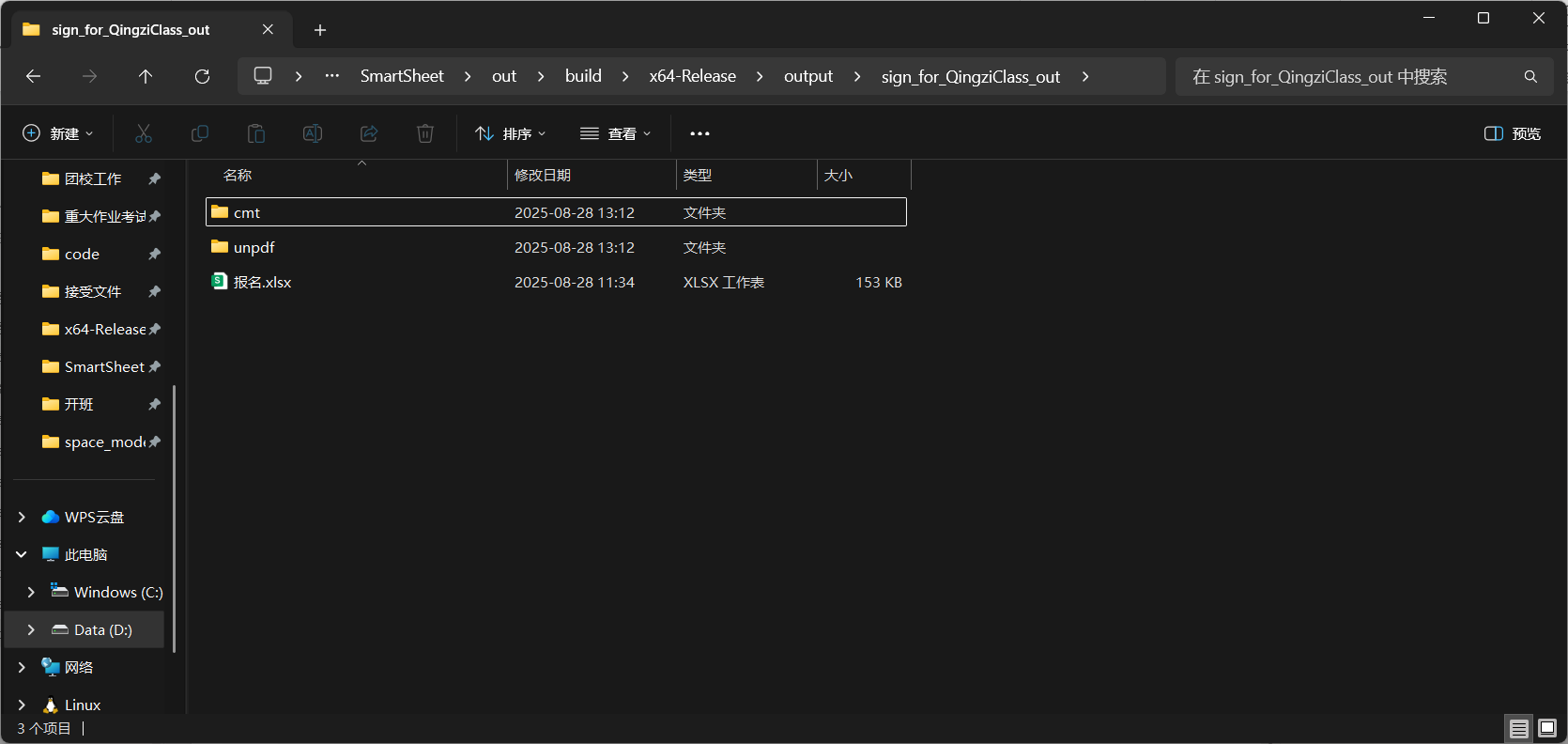




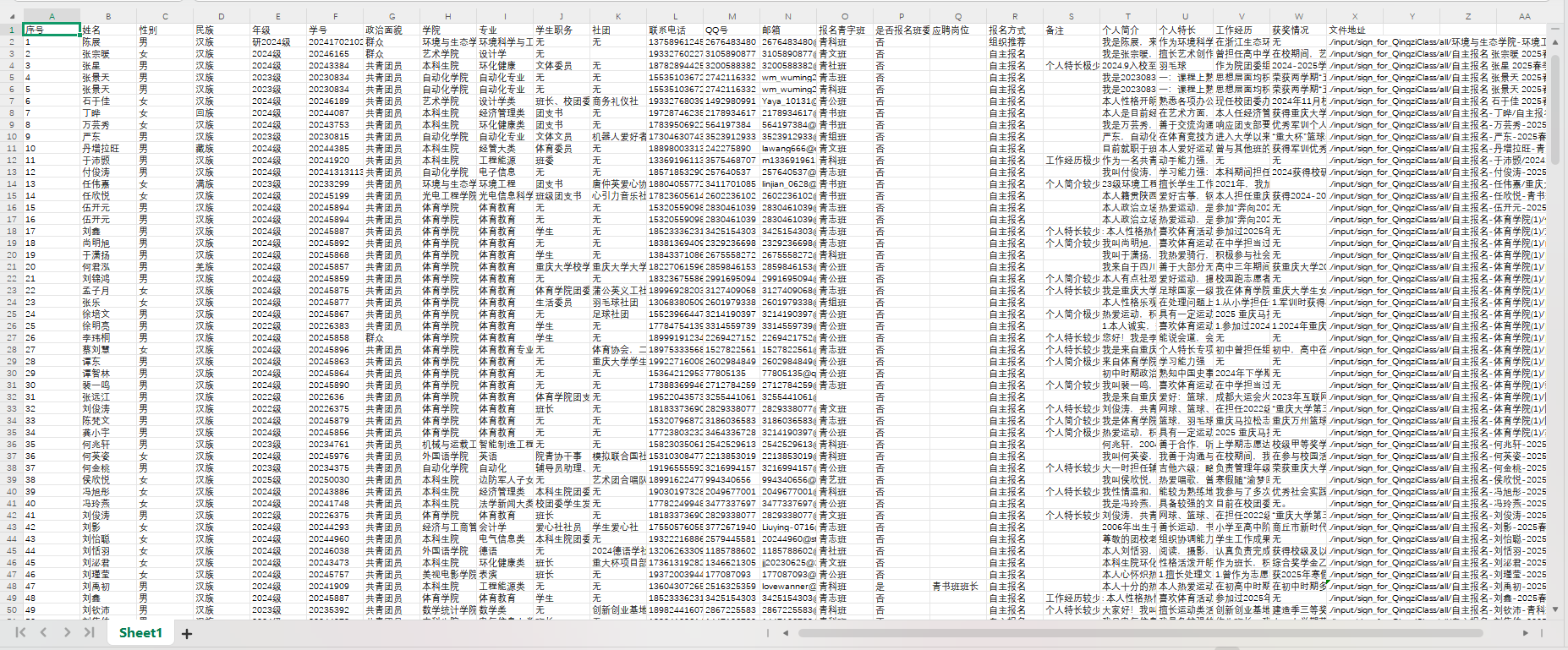
（八）output/sign\_for\_QingziClass\_out

此文件储存检测到的班委报名表“/cmt”、未知的pdf“/unpdf”、以及输出的报名审核表“/报名.xlsx”。

注意：每次运行程序时，此文件夹中的文件都会被删除。



**1./报名.xlsx**



**2./cmt**

识别到的班委报名表储存的位置。

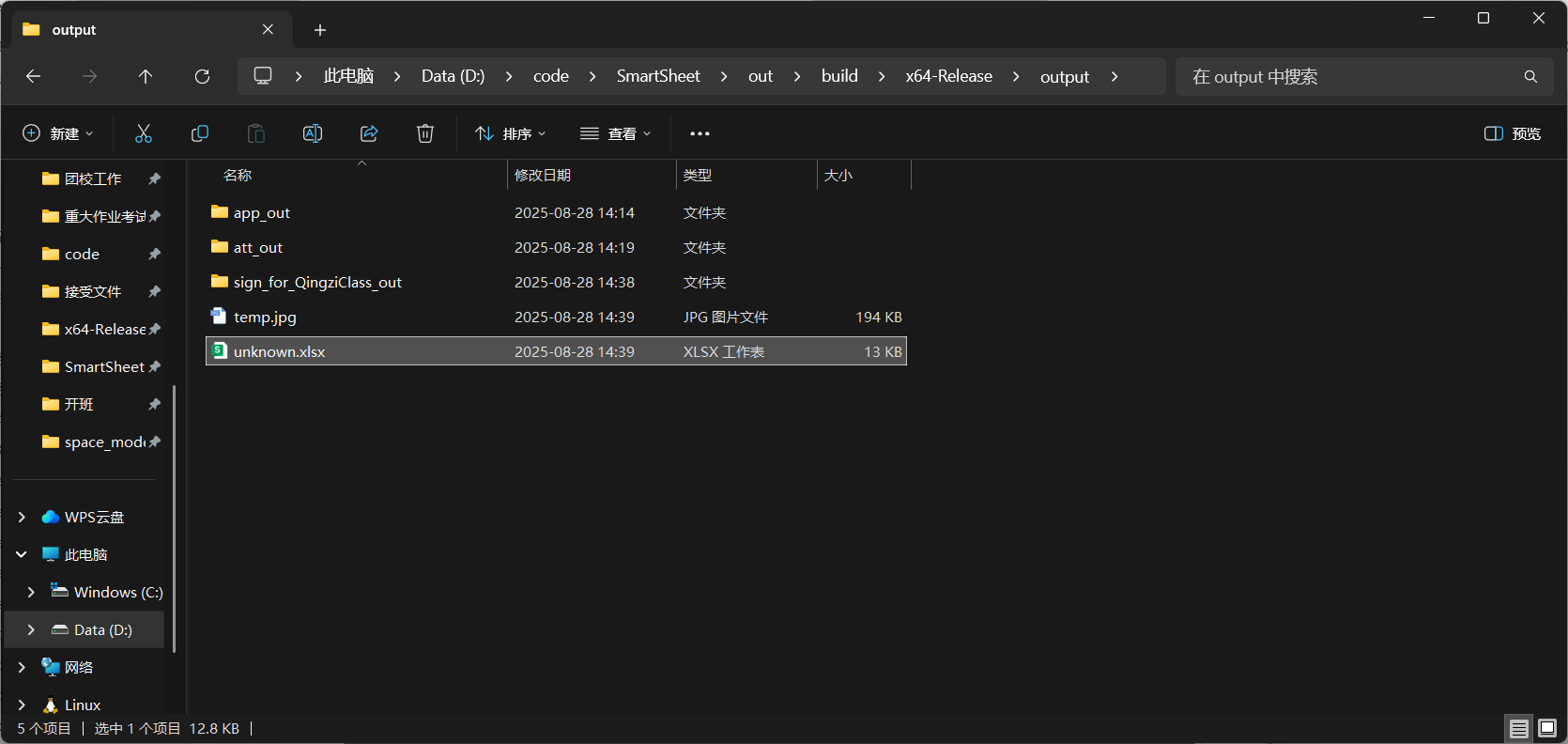
**3./unpdf**

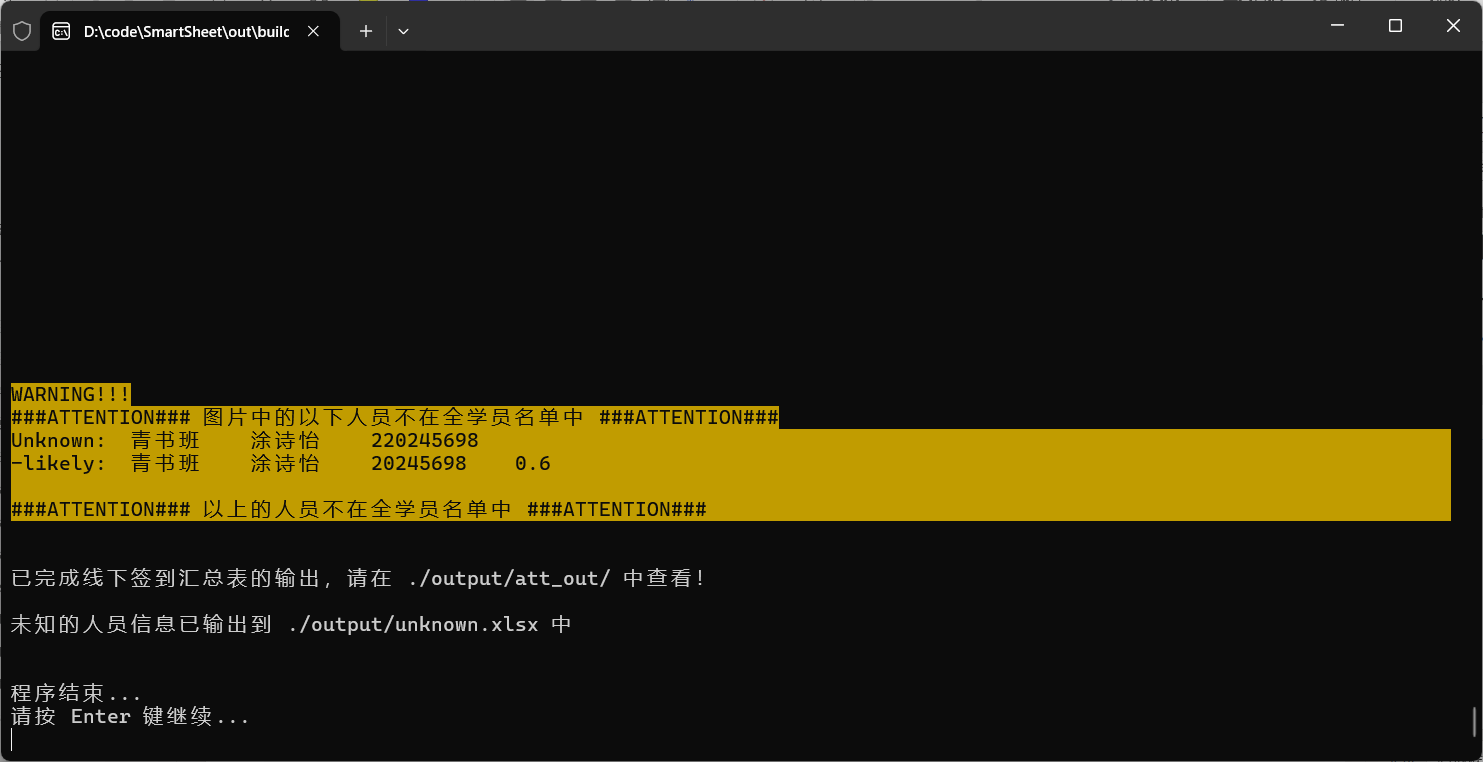
不能识别的pdf文件，需要自己点开查看。

（九）output/unknown.xlsx

此文件用于输出与学员花名册中人员信息比对不同的人员，功能（一）（二）都会输出此文件。

如：在功能（二）中输出了一个错误的人员。“Unknown”表示从图片中识别到的人员；“likely”表示的是花名册中的人员。Excel文件中同理。

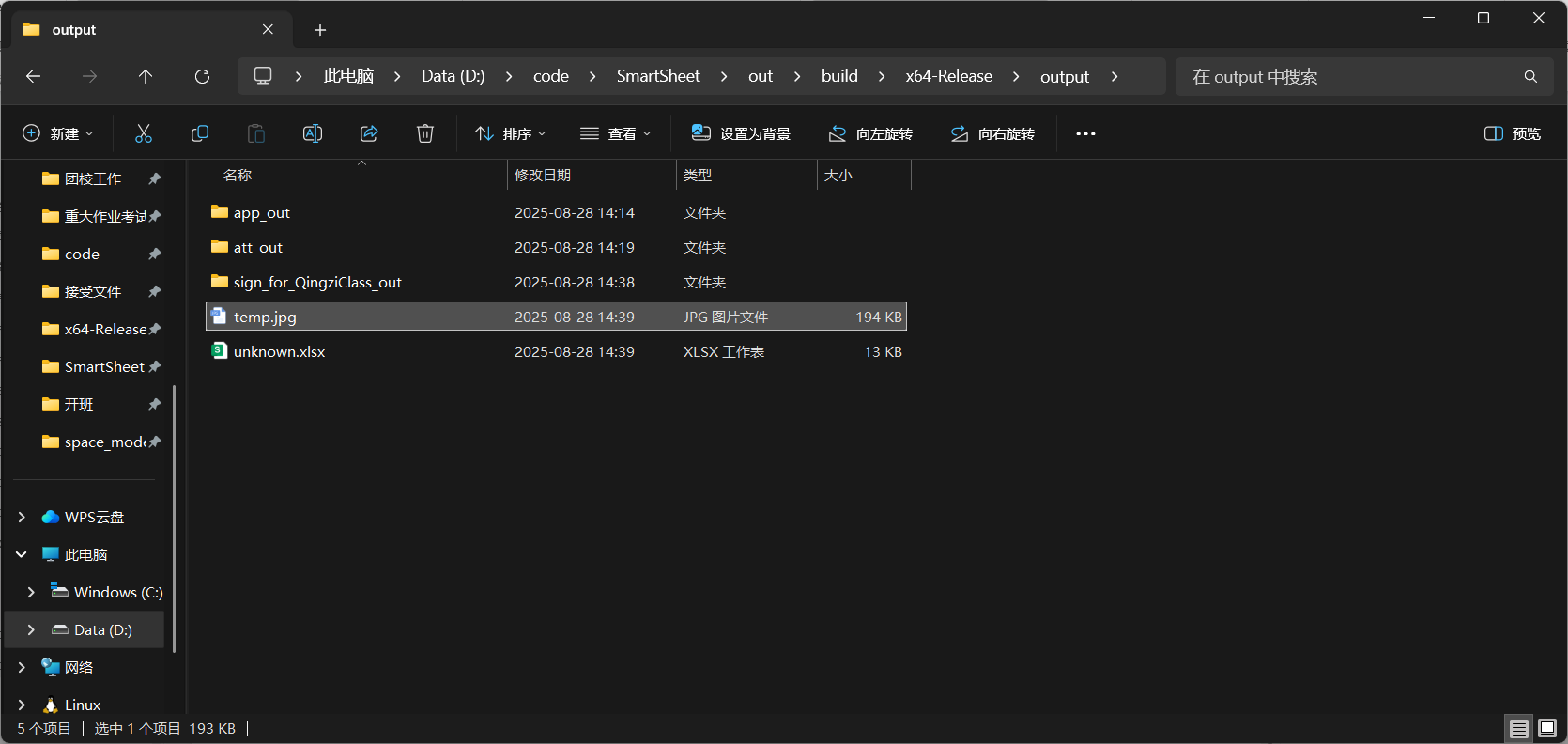






（十）output/temp.jpg

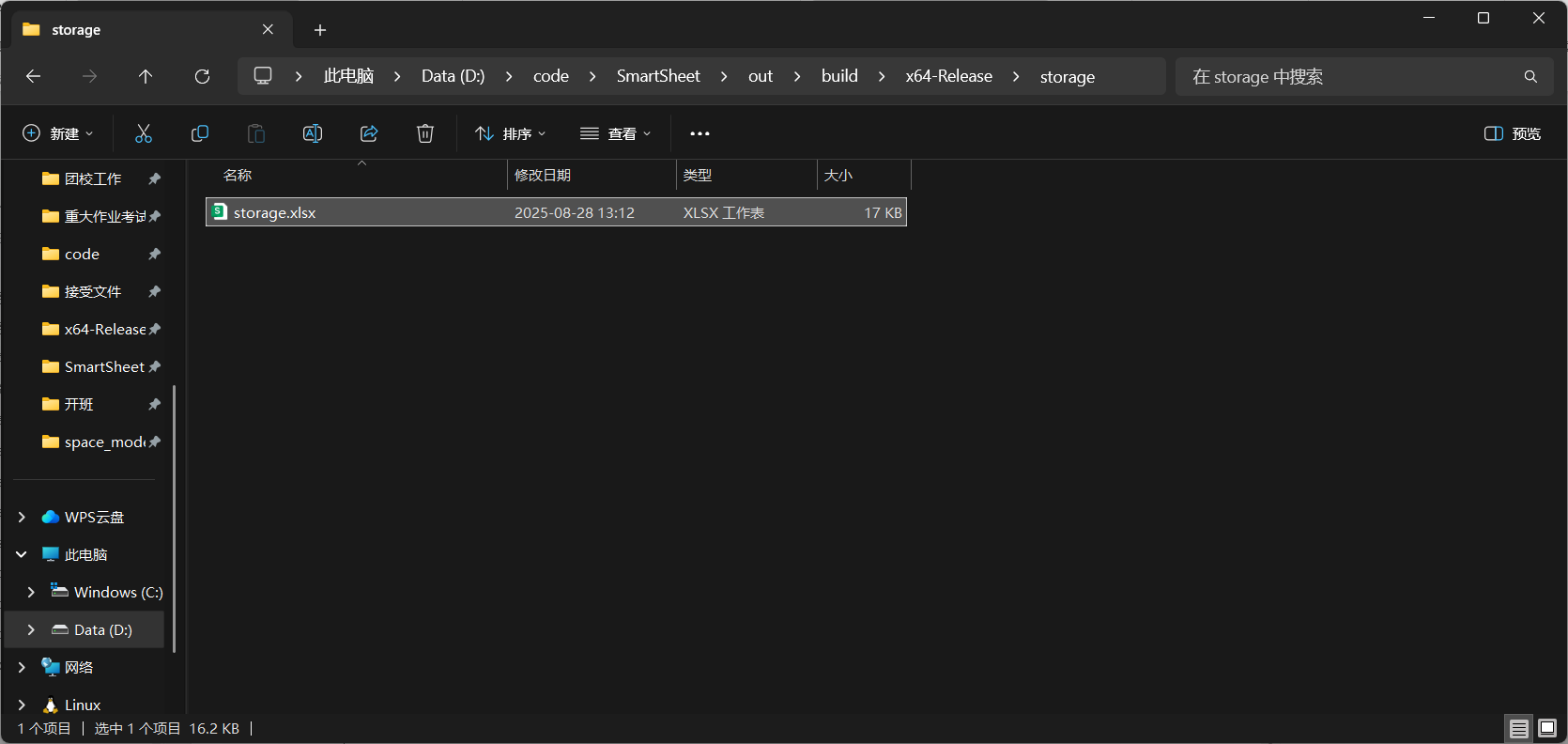
ocr功能调试时未修改的生成。（~~修改了代码之后请删除此说明~~）

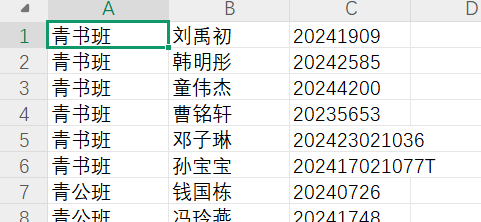


（十一）storage/storage.xlsx

保存所有报名人员的信息的缓存，由功能（一）生成，启动功能（二）所**必须**的文件。

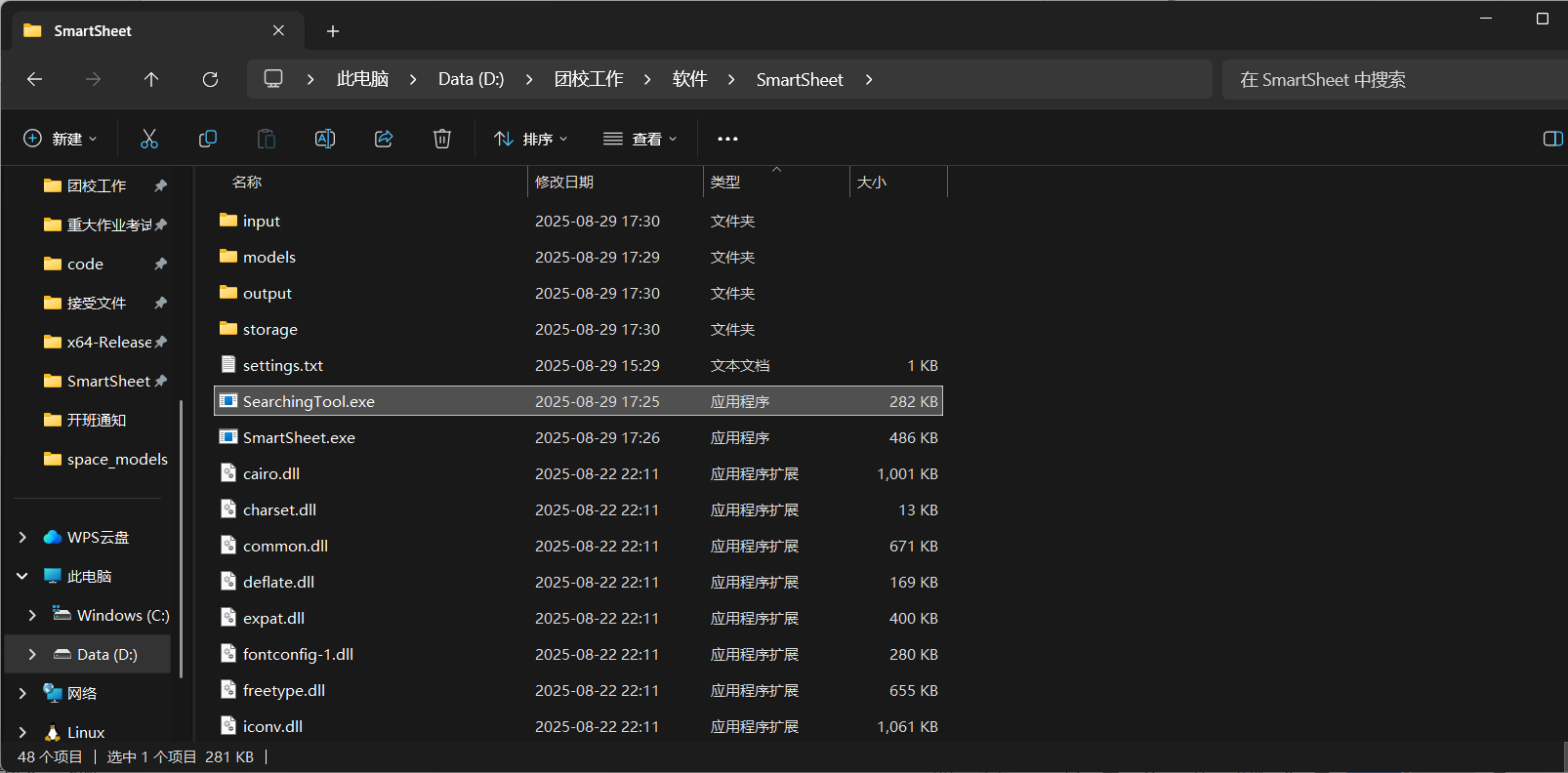
如果功能（一）与功能（二）分别在电脑A与电脑B上分别运行，只用将电脑A上生成的storage.xlsx复制到电脑B的同源处即可。



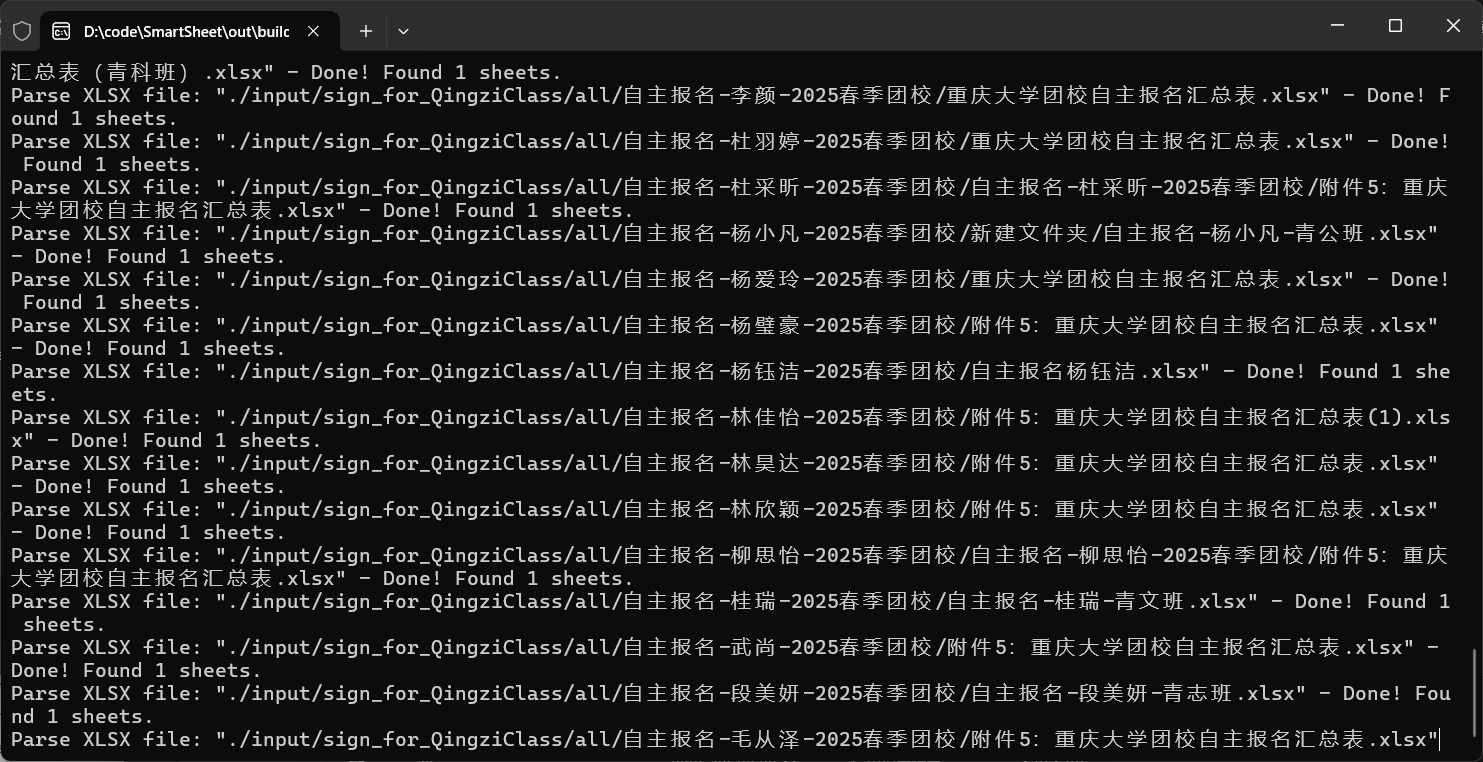


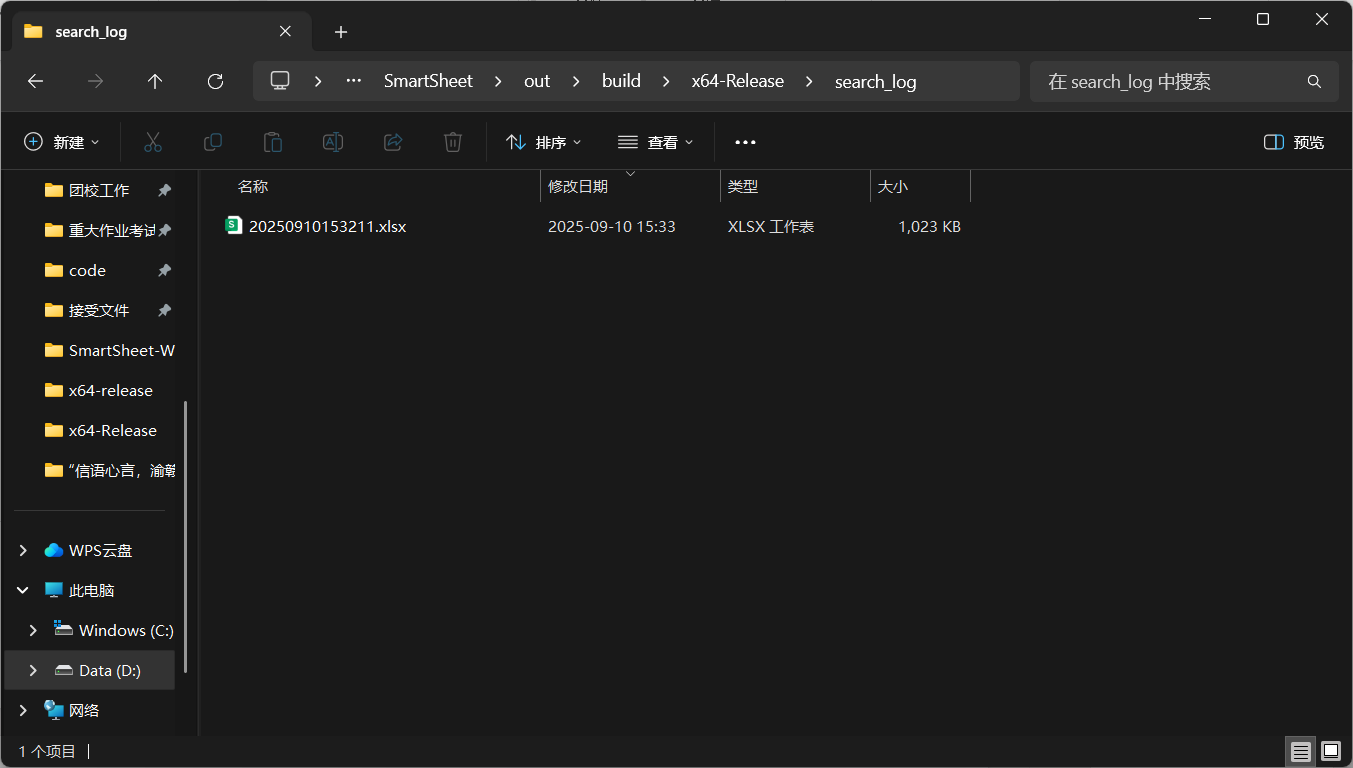
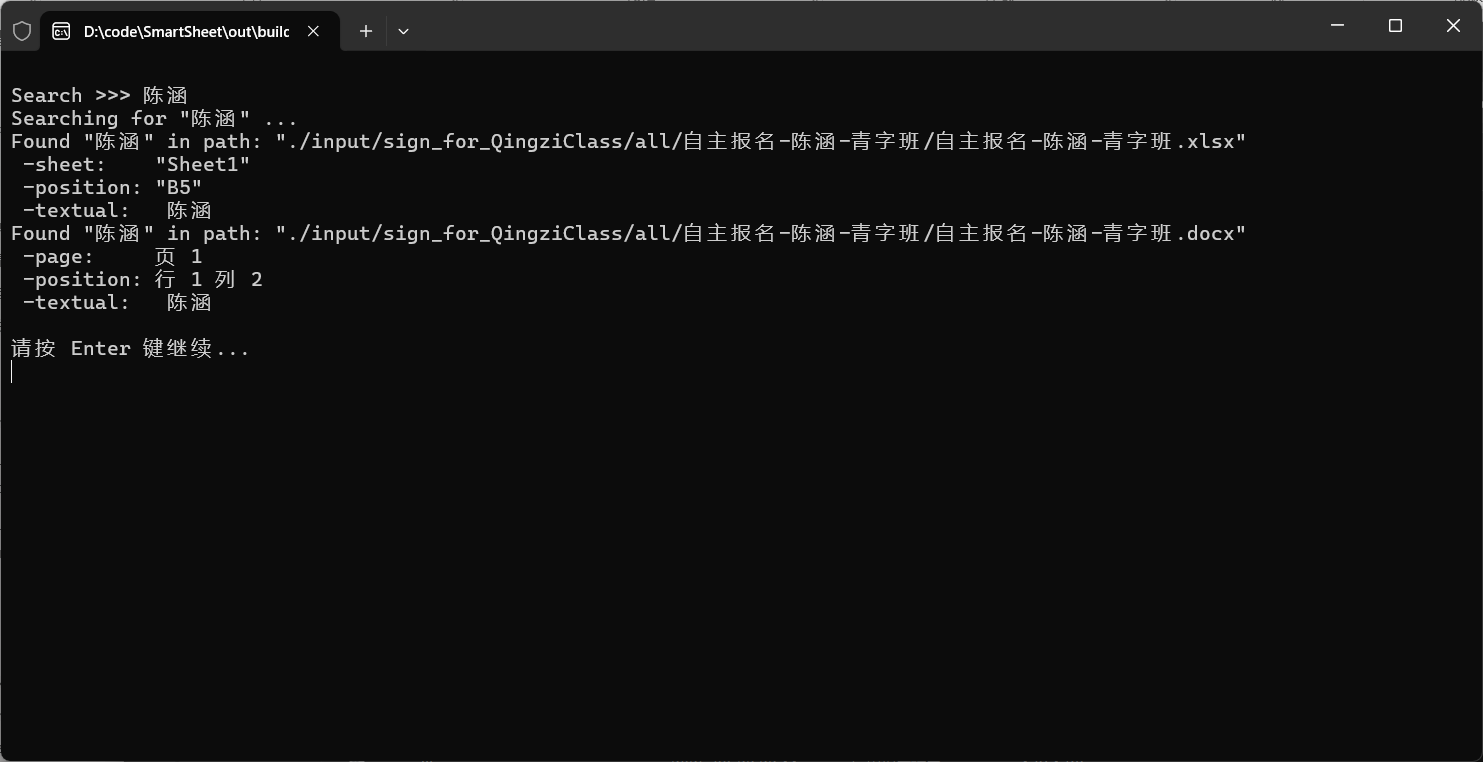
五、搜索工具

此程序内置了一个搜索工具SearchingTool.exe，可以用于搜索“input”文件夹中的所有文件中的内容。



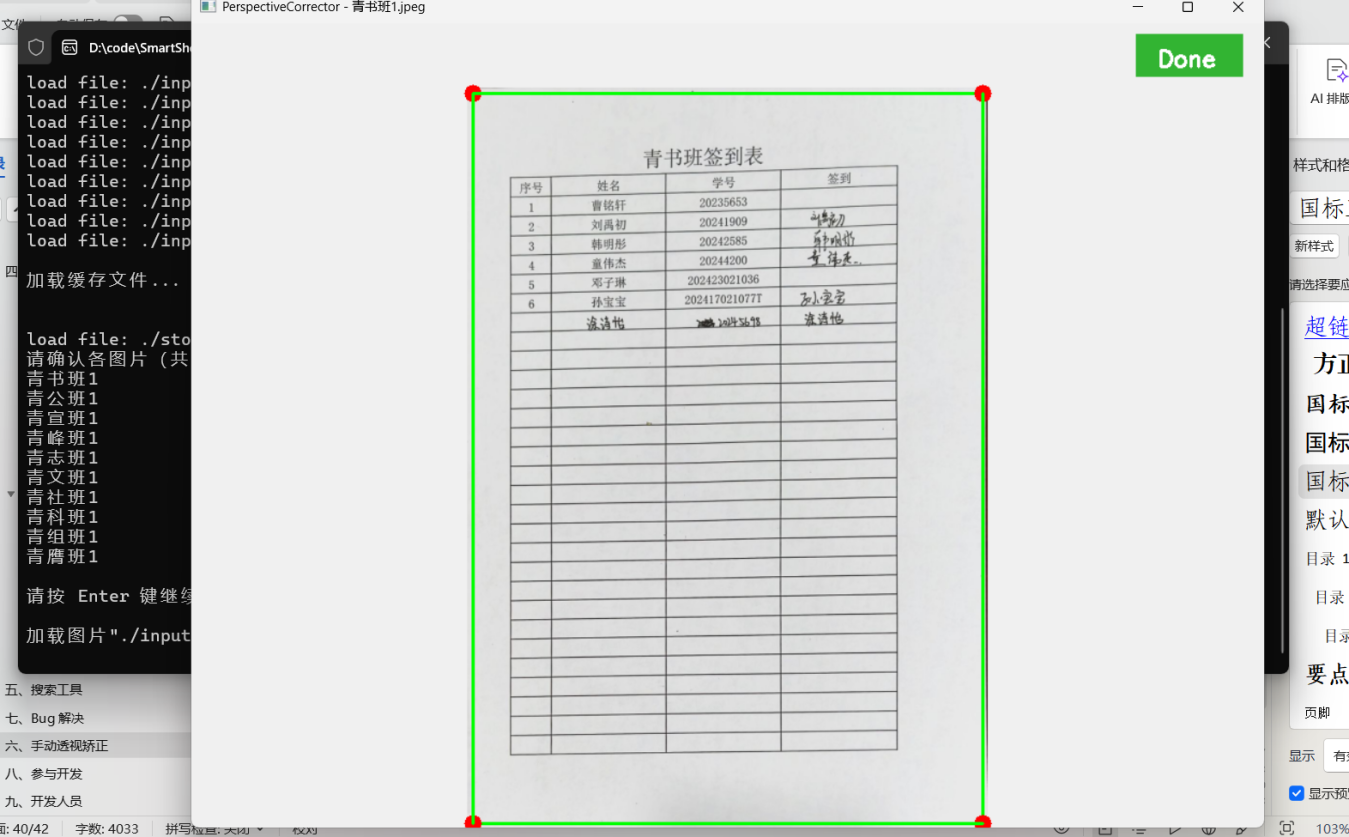
双击运行，待加载所有文件结束之后，进入搜索界面“Searching >>> ”，可以输入要搜索的内容，按下回车可以看到相关内容出现的位置。输入“$exit$”可以退出程序，输入“$save$”可以退出并保存搜索结果到“search\_log/”中。





六、手动透视矫正

受限于开发人员水平，自动的透视矫正技术不可用，因而只能手动矫正。在settings.txt中[开启手动透视矫正](#ManualDocPerspectiveCorrector)之后，会在程序进行[ocr期间](#ocr期间)弹出手动矫正框。



矫正框由四个红色角点，四个绿色边框构成，角点可以拖动（注意：**右上角的角点只能对应A4纸的右上角**）。拖动角点，确保**绿色边框与网格线平行**。确认之后点击右上角的绿色“Done”按钮即可。



七、Bug解决

目前尚未添加解析错误的文件。

如遇到问题请联系：教学部-刘思成[（2561925435@qq.com）。](mailto:（2561925435@qq.com）。)

如果你对此模块感兴趣：[加入我们](https://github.com/lscatfish/SmartSheet)

八、参与开发

[加入我们](https://github.com/lscatfish/SmartSheet)

本项目基于C++20标准以及众多第三方库开发。

很高兴你能参与本项目开发，我们希望此项目最终是一个窗口化的程序，而不是目前的一个控制台窗口。

（一）仓库地址

Github标准仓库：<https://github.com/lscatfish/SmartSheet>

Release版本：<https://github.com/lscatfish/SmartSheet/releases>

国内镜像Gitee：<https://gitee.com/sicheng-liu/SmartSheet>

国内镜像Gitcode：<https://gitcode.com/lscatfish/SmartSheet>

Release版本国内镜像Gicode（保持更新），你总是可以从中下载到最新版：<https://gitcode.com/lscatfish/SmartSheet/releases>

**受限于仓库属性，最好下载zip而不是直接clone。**

已禁用Gitcode与Gitee的push功能，但是代码同步于Github，所有更改请推送至Github仓库。

（二）注意事项

相关事项请读取项目的README.md文件。

新功能加入后，请修改本文档。

九、开发人员

感谢刘思成-lscatfish对此项目做出的贡献。

（有人参与开发就把名字与GithubID加上）