FlySense 数据查重与异常检测工具 · 部署与运维指南

数据查重 · 异常检测（Windows）

适用版本：FlySense 数据查重与异常检测工具 v1.0.0

适用系统：Windows 10 / Windows 11（64 位）

使用方式：一个 `FlySense.exe` 文件，双击就能用。**不用安装、不用联网、不用懂 Excel 函数。**

> **这份指南分两部分，看您是谁**
>
> 第 1 ~ 11 章是给**使用者**看的：怎么激活、图文步骤、大白话，电脑小白也能看懂。
>
> 第 12 ~ 14 章是给**企业管理员 / IT 部门**看的：批量处理模式、命令行静默运行、日志与排错。
>
> 同目录下已经排版好了 `部署与运维指南.pdf`，可以直接发给客户。
>
> 改完本文件后想重新生成 PDF：双击同目录的 `生成PDF.bat` 即可（用的是电脑自带的浏览器，不需要装 Typora、不需要联网）。也可以用 Typora 打开本文件 → 【文件】 → 【导出】 → 【PDF】。
>
> 指南里的截图放在同目录的 `图片` 文件夹里，导出 PDF 时请让这个文件夹和本指南放在一起。

目录

1. 这个软件能帮你做什么
2. 激活与授权（第一次打开必须做）
3. 打开软件
4. 怎么用：就四步
5. 结果怎么看
6. 导出结果（Excel 里有什么）
7. 两个滑块怎么调
8. 什么样的数据能用
9. 常见问题与排错
10. 五分钟试一遍（演示数据）

11. 名词表
12. 批量处理模式（一次跑完一整个文件夹）
13. 命令行静默运行（给 IT 部门用）
14. 如何看日志和排错

一、这个软件能帮你做什么？

表格数据里最常见的两个毛病，它一眼就能找出来：

它能做什么	通俗解释	什么时候用得上
去重	找出表格里**内容几乎一模一样**的行	客户名单重复、订单重复录入、导出数据时串了行
异常检测	找出**跟其它行格格不入**的行（数值离谱的那种）	抓录入错误、抓异常交易、抓“不小心多打了几个零”

举个例子：一张 100 行的客户表，它能告诉你“**有 3 对行是重复的**”，或者“**第 77 行、第 90 行的数据很不对劲**”。

它不会做什么（免得您误会）：

- 它**不会修改**您的原文件，只读不写，原表永远安全。
- 它**不会把数据传到网上**，全程在您自己的电脑里计算。
- 它**不会替代您的判断**，它只负责“把可疑的行挑出来给您看”，删不删由您决定。

二、激活与授权（第一次打开必须做）

FlySense 是收费软件，**一台电脑买一份**。第一次打开时，软件会先弹出一个**激活窗口**，

要您填一次授权码；填对以后就再也不会打扰您了。

> ## 一句话说明白
>
> **您把软件显示的「机器码」发给我们 → 我们回您一串「授权码」 → 您填回去，就激活了。**
>
> 整个过程一分钟，**不需要联网。**

第一步：看到激活窗口



窗口上有三个按钮，按您的情况点其中一个：

按钮	什么情况下点
免费试用	想先试试好不好用。点了就进主界面，**3 天 / 10 次**之内免费用（哪个先到算哪个）。
激活并进入软件	已经买到授权码了，填进去点它。
退出	先不用了，关掉软件。

第二步：点【复制机器码】，把它发给我们

窗口上那串大字（例如 `A1B2-C3D4-E5F6`）就是**这台电脑的机器码**：

1. 点一下【复制机器码】按钮（窗口里会出现绿字「已复制！」）；
2. 打开微信 / QQ / 邮件，**粘贴**进来发给我们；
3. 顺便把您的**公司名称**也一起发过来，我们好对账。

> **机器码是什么？** 它是软件根据您这台电脑的主板、CPU 等信息算出来的一串编号。 它**不包含您的任何数据**，也**不需要联网**，不用担心隐私。 每台电脑的机器码都不一样，所以我们才知道这份授权是发给哪台电脑的。

第三步：把我们回复的授权码填进去

我们收到机器码后，会回您一串**授权码**（例如 `2BG9-4V5W-GYFW-CVP8`）：

1. 把它**粘贴**到激活窗口的第二个框里；
2. 点【激活并进入软件】（在这个框里直接按回车也行）。

填对了就直接进主界面，窗口底部的状态条会显示「**已授权**」：



> **以后还要再填吗？** 不用。授权是您在电脑上当场算出来比对的， 激活一次就记住了，以后每次打开都不会再弹激活窗口。
>
> **当时先点了「免费试用」，后来才买授权码？** 也没问题。 主界面**右下角**有一个**【激活 / 输入授权码】**按钮， 随时点它就能把授权码填进去，不用等试用期结束。

填错了会怎么样？

填错（或者少粘几位）不会出任何问题，软件只会提示一句，让您重新填：



看到 **「授权码无效，请检查后重试」**，请对着我们发的那串码再看一遍。最常见的三种情况：

- **少粘了几位**：授权码一共 16 位（`XXXX-XXXX-XXXX-XXXX`），请整串复制；
- **粘多了别的东西**：比如把聊天记录里的称呼一起粘进去了，只留那串码就行；
- **机器码发错了**：比如把另一台电脑的机器码发给了我们。那就重新点一次【复制机器码】再发一次。

> 放心两点：
>
> 1. **不分大小写**：`2bg94v5wgyfwcvp8` 和 `2BG9-4V5W-GYFW-CVP8` 是一样的，横线打不打都认； 2. 但一台电脑的授权码**只能在这一台电脑上用**，抄到别的电脑上不管用。

试用期是怎么算的?

- 从**第一次打开**那一刻开始算：**3 天**，或者**10 次运行**，哪个先到算哪个；
- 试用期内，主界面底部的状态条会一直提醒您，例如「试用中：还剩 2 天 / 8 次」；
- 试用期结束后，软件会**再次弹出激活窗口**，这时候**必须填授权码**才能继续用，

命令行静默运行也会直接停下（退出码 2）。

换了电脑怎么办？

机器码是跟着电脑走的，所以：

- **换了一台新电脑**：在新电脑上打开 FlySense → 点【复制机器码】→ 把新的机器码发给我们，

我们重新给您一个授权码。（同一家公司换机器，请把**旧的机器码**也一起说明，方便我们对账。）

- **同一台电脑重装了系统**：主板和 CPU 没变，机器码通常也不变，直接填**原来那个授权码**就行。

万一变了，按上面「换新电脑」的办法重新申请。

- **换了主板 / 换了硬盘**：机器码会变，需要重新申请。

> 小建议：把授权码抄在交付包里的 `README.txt` 上（或者存进公司的密码本）。 以后重装系统填一下就行，不用每次来找我们。

授权信息存在哪？安全吗？

激活成功后，软件会在您电脑里**两个地方各存一份**（互为备份，删掉一个也不会失效）：

- 文件：`C:\Users\您的用户名\AppData\Local\FlySense\license.dat`
- 注册表：`HKEY_CURRENT_USER\Software\FlySense`

这两个位置都只属于**当前登录的用户**，不需要管理员权限，也不会影响这台电脑上的其他同事。

命令行静默运行时的授权

给 IT 部门用的静默模式（见本指南《命令行静默运行》一章）**同样要认授权**：

情况	会发生什么
这台电脑已经激活过	照常处理数据，不用管授权的事
没激活，试用也结束了	**不处理任何文件**，直接退出，**退出码 = 2**，日志里写明「授权失败」和本机机器码
只是查 `--help` / `--version`	不需要授权，随时能看

三、打开软件

1. 找到 `FlySense.exe` 这个文件。
2. **双击它**。

3. 等几秒，会出现下面这个窗口：

> **第一次打开会先弹一个「激活窗口」**（要填一次授权码，或者点【免费试用】），激活之后才会进到下面这个主界面。具体怎么弄看上一章 **《激活与授权》**。



> [截图：界面长这样 —— 第一次打开 FlySense 时的窗口]

几个小提示

- 第一次打开会慢 3~10 秒，**这是正常的**，原因见 常见问题 1。
- 如果杀毒软件弹窗拦住了，看 常见问题 2。
- 建议把 exe 单独放在一个文件夹里（例如 `D:\FlySense\`），以后找起来方便。
- 窗口右下角要是被您拖得太小了，把窗口边缘往外拉大就行，界面会自动排版。

四、怎么用：就四步

> **记住这一句就够了：**
>
> **第一步，点这里选文件；第二步，选“去重”或“异常检测”；第三步，点“开始运行”；第四步，点“导出结果”。**

第一步，点【选择数据文件】

点界面上左上角的【选择数据文件】按钮，在弹出的窗口里找到您的表格（支持 `.xlsx` 和 `.csv`），双击它。

选好之后，按钮右边会显示文件名字（例如 `客户名单.xlsx`），表示**您已经选上了。**



> [截图：点【选择数据文件】 → 选中您的 Excel → 按钮右边出现文件名]

注意：这一步只是“告诉软件去看哪个文件”，您的原文件没有被打开、也没有被改动。

第二步，选【去重】或【异常检测】

在第 2 步那里点一下小圆点：

- 想找**重复的行** → 点【去重（找出重复的数据行）】。
- 想找**数值离谱的行** → 点【异常检测（找出不正常的数据行）】。



> [截图：第 2 步的两个选项，点一下小圆点就选中了]

只能选一个。 两个都想做？那就跑两遍：先跑去重导出结果，再切到异常检测跑一遍。

第三步，点【开始运行】

点【开始运行】按钮，然后**稍等几秒到几十秒**（数据越多越慢）。

运行的时候：

- 窗口**不会卡死**，只是按钮变灰，这是防止您重复点的保护。
- 下面“运行结果”的框里会一行行往外冒字，那是它在报告进度。
- 窗口最底下的状态条会显示“正在编码并分析 ...”，跑完了会变成“运行完成”。

跑完之后，“运行结果”框里会出现结论。**去重**跑完长这样：



> [截图：去重跑完的结果 —— 结论会写“发现几组重复、涉及几对、几行属于重复项”]

异常检测跑完长这样（这张图里把比例调到了 2%）：



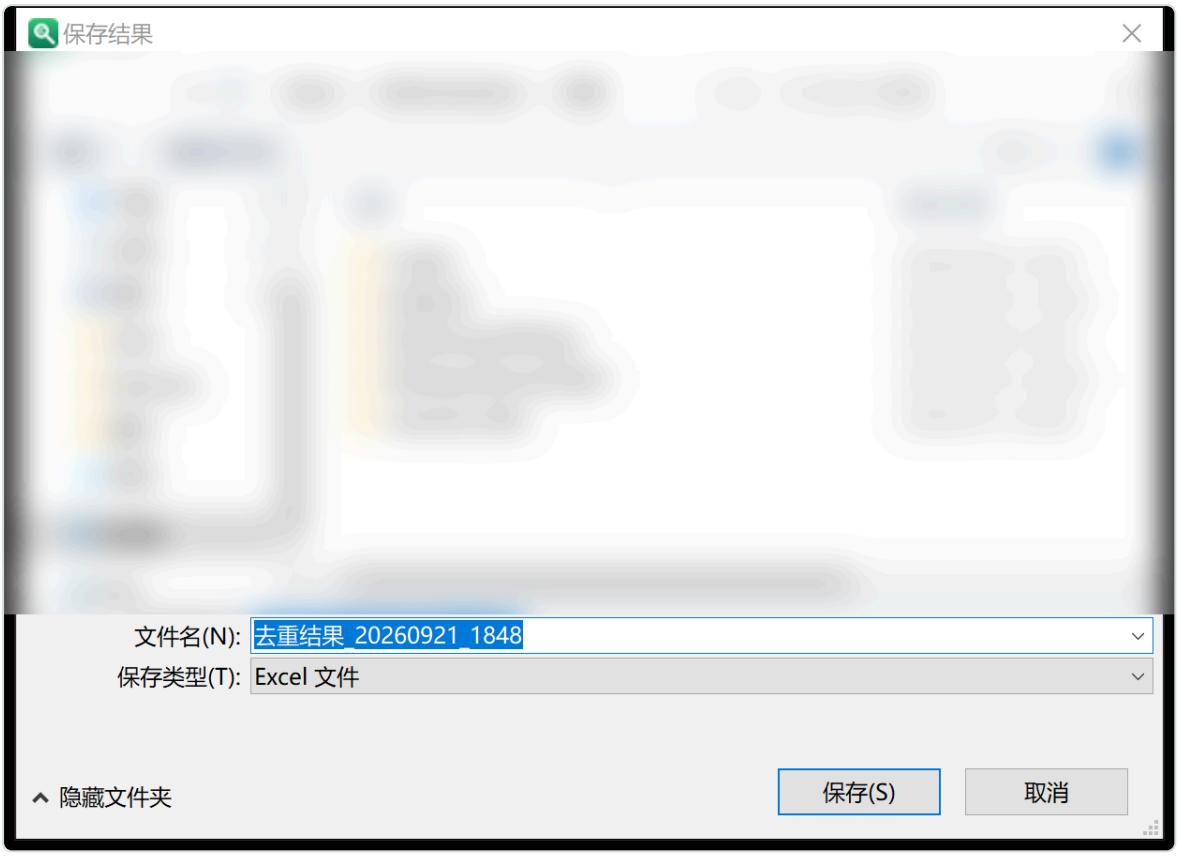
> [截图：异常检测跑完的结果 —— 结论会写“判定异常几条”和“分数阈值”]

第四步，点【导出结果】

结果框最后一行会提示“现在可以点击【导出结果】保存成 Excel。”

这时点【导出结果】，会弹出一个“另存为”窗口：

- 文件名**已经帮您起好了**，例如 `去重结果_20260920_1435.xlsx`（后面那串是日期和时间，不会覆盖上一次的结果）。
- 您只需要选一个**存放位置**（比如桌面），然后点【保存】。
- 保存成功后会弹一个小窗口告诉您文件存到哪儿了，点【确定】即可。



> [截图：点【导出结果】 → 选个文件夹 → 点【保存】 → 弹出“结果已保存到……”]

到这里就完成了。 打开导出的 Excel，就能看到有哪些重复行 / 异常行。

五、结果怎么看

去重跑完，结果框里的每句话是什么意思

以 100 行的示例数据为例，结果框里是这样几行：

结果里的话	大白话解释
`读取成功：100 行 x 7 列`	文件读进来了，一共 100 行数据、7 列
`用于分析的数值列：6 列`	有 6 列是数字，参与比对；文字列（比如客户ID）不参与计算
`【结果】共 100 行数据`	一共分析了 100 行
`发现 3 组重复，涉及 3 对`	有 3 处重复，每处是一对行
`其中 3 行属于重复项（可考虑删除）`	这 3 行是“后面重复的那一行”，可以考虑删掉
`数据背景相似度：0.895（自动测出）`	这份数据本身的行与行之间有多像，是软件自动算出来的
`本次判定阈值：相似度 ≥ 0.979（严格度 0.80）`	相似度超过 0.979 才算重复；这个门槛由滑块决定
`现在可以点击【导出结果】保存成 Excel。`	跑完了，可以导出了

“3 组重复、3 行属于重复项”怎么看？ 一组重复 = 两行内容一样，其中**留一行、另一行标成“重复”**。所以 3 组重复 → 3 行被标成“重复”，通常就是可以删掉的那 3 行。

异常检测跑完，结果框里的每句话是什么意思

结果里的话	大白话解释
`判定异常 2 条（设定比例 2.0%）`	挑出了 2 行“不对劲”的，因为您把比例设成了 2%
`分数阈值：346.51`	每行都有一个“离谱分”，超过 346.51 分才会被挑出来
`提示：异常分越高越“不像其它数据”，可结合业务判断。`	分数越高越可疑，但**要不要处理还得您看一眼**

> **重要提醒**：异常检测挑出来的是“**和大多数行都不一样**”的行，它不等于“错误”。比如某位客户真的是大客户、消费额特别高，它也会被挑出来。**请当成“重点复核名单”，而不是“错误名单”。**

底部状态条

窗口最底下那条细细的状态栏，会显示三种字样：

- `就绪` —— 还没开始跑。
- `运行完成` —— 跑完了，可以导出了。
- `出错了` —— 出问题了，请看 常见问题。

六、导出结果（Excel 里有什么）

导出的 Excel 里有 **3 个工作表**（在 Excel 窗口最下面一排标签里切换）。

去重的结果文件

工作表名	里面是什么
重复明细	所有涉及重复的行，前面加了 4 列提示：`重复组号`（第几组重复）、`角色`（保留 / 重复）、`相似度`（有多像）、`原表行号`（对应原文件第几行）
重复对	一对一对的清单：`行A`、`行B`、`相似度`。想知道“哪两行撞了”，看这张表最快
汇总	本次运行的参数和结论：数据总行数、判定阈值、严格度、重复组数、建议删除行数、导出时间

异常检测的结果文件

工作表名	里面是什么
异常明细	被判定为异常的行，前面加了 3 列：`原表行号`、`异常分`（越高越可疑）、`判定`
全部评分	每一行的异常分都列出来了，方便您自己按分数排序查看
汇总	本次的参数和结论：总行数、异常比例、判定条数、分数阈值、导出时间

小提示：

- `原表行号` 是**按 Excel 的行号算的**（第 1 行是标题行，所以数据从第 2 行开始）。想回原文件核对，直接跳到那一行就行。
- 导出的文件名带日期时间，**不会覆盖**您上一次的结果。
- 如果保存时提示“文件被占用，请先关闭同名 Excel 再试”，意思是**这个文件正被 Excel 打开着**，关掉它再导出一次即可。

七、两个滑块怎么调

在第 3 步那一栏有两个滑块。**平时不用动它们，保持默认就好。** 只有当结果“太多”或“太少”的时候才需要调。

还没轮到的那个滑块会**变灰**（点不动），这是正常的：选了“去重”，就只用得上左边那个；选了“异常检测”，就只用得上右边那个。

滑块一：重复判定严格度（只有“去重”用得上）

- 范围 `0.00` ~ `1.00`，**默认 0.80。**
- 数字**越大越严**：越严 → 只有长得特别像的才算重复 → **挑出来的重复变少**（可能漏）。
- 数字**越小越松**：越松 → 稍微像一点就算重复 → **挑出来的重复变多**（会误报）。

下面是用 100 行示例数据实测出来的结果（示例数据里**故意埋了 3 对完全相同的行**）：

严格度	找到几组重复	说明
0.50	7 组	太松了，多挑出 4 组“长得像但不一样”的 → 误报
0.60	5 组	还是松
0.70	4 组	多出 1 组误报
0.80（默认）	**3 组**	**正好，就是埋进去的那 3 对**
0.90	3 组	也对
1.00	3 组	最严，要求几乎一模一样

怎么调：先跑默认的 0.80。如果结果里明显有“这两行其实不一样”的误报，就往**大**调（0.85、0.90）；如果明明是重复却没被找出来，就往**小**调（0.75、0.70）。

滑块二：判为异常的百分比（%）　（只有“异常检测”用得上）

- 范围 `0.1` ~ `10.0`，**默认 1.0**（意思是“把最不像话的 1% 行挑出来”）。
- 数字**越小越严**：只挑最离谱的那几条 → **条数变少**（可能漏）。
- 数字**越大越松**：挑出来的行更多 → **条数变多**（会包含一些其实正常的行）。

同样用 100 行示例数据实测（示例数据里**故意埋了 2 行离谱数据**）：

异常比例	挑出几条	说明
0.5%	1 条	太严，只挑到最离谱的那一条
1.0%（默认）	**1 条**	100 行的 1% = 1 条，只会挑出最离谱的那一行
2.0%	**2 条**	**两条异常都挑出来了**
5.0%	5 条	多出来的 3 条是“陪跑的”，需要人工看一眼
10.0%	10 条	挑得太多了，翻看起来费劲

怎么调（记住一句话）：行数少的时候要把比例调大一点。

因为“百分比”是按总行数算的：100 行的 1% 就只有 1 条，1000 行的 1% 是 10 条。所以：

- 数据只有几十、一百行 → 想要多挑几条，就把比例调到 **2%~5%**。
- 数据有几千行 → 用 **0.5%~1%** 就已经够挑了。

万能口诀：

您的感受	怎么调
“挑出来的太多了，好多是正常数据”	严格度往**大**调 / 异常比例往**小**调
“明明有问题，却没挑出来”	严格度往**小**调 / 异常比例往**大**调

八、什么样的数据能用

项目	要求
文件格式	`.xlsx`（推荐）或 `.csv`。旧版 `.xls` 请先在 Excel 里另存为 `.xlsx`
第 1 行	必须是**列名**（比如“客户ID、交易金额、登录次数”）。软件要拿它当标题
数据行数	至少 2 行。**超过 5000 行**会先弹窗提醒您“可能要等一会儿”；**超过 50000 行**会拒绝运行，请先抽样
参与计算的列	**只挑数字列**。像“客户ID”“姓名”“地址”这种文字列会被自动忽略（不参与比对）
空单元格	按 0 处理，不会报错
建议	列名写得清楚一些、数字列多一些，结果会更准

为什么行数太多会拒绝？ 这个算法要把每一行和其它的行两两比对，行数翻一倍，计算量大约翻四倍。50000 行以上会算到天荒地老，所以干脆先拦住，让您抽一部分数据再跑。

九、常见问题与排错

1. 第一次打开要等好几秒，是不是卡死了？

不是卡死。 这个软件是“单文件版”（整个软件打包成 1 个 exe），每次双击时，Windows 需要先把里面几百兆的内容解压到临时文件夹，然后再启动界面。

- 一般 3~10 秒，电脑慢一点会更久。
- 这期间**鼠标指针会转圈**，说明它正在忙，不是没反应。
- **不要连续双击好几次**，那会同时启动好几个，反而更慢。

> 如果您想让它开着更快，把它一直开着就行，不用每次都关。

2. 杀毒软件报警了怎么办？

先放心：这是误报。 原因很简单——个人开发者没有购买微软的“代码签名证书”，这类没签名的自制小工具，360、火绒、Windows Defender 经常“看谁都可疑”，会直接拦下来甚至删掉。**几乎所有个人开发者做的小工具都会遇到这个情况。**

怎么处理（按您遇到的提示对号入座）：

Windows Defender / SmartScreen 弹出“Windows 已保护你的电脑”

1. 点【更多信息】；
2. 点【仍要运行】。

Windows Defender 已经把文件删掉了

1. 开始菜单搜“Windows 安全中心” → 【病毒和威胁防护】；
2. 点【保护历史记录】→ 找到被拦的 `FlySense.exe` → 点【还原】；
3. 或者：【病毒和威胁防护】→【管理设置】→【排除项】→【添加排除项】→ 选中 `FlySense.exe`。

360 安全卫士 / 360 杀毒

1. 打开 360 → 【木马查杀】→【恢复区】→ 找到文件 → 点【恢复】；
2. 再到【信任区】把 `FlySense.exe` 添加进去（不加的话下次扫描还会报）。

火绒

1. 打开火绒 → 【安全工具】→【信任区】→【添加文件】→ 选中 `FlySense.exe`。

公司电脑（IT 统一管控）

把 `FlySense.exe` 交给 IT，说明“这是自研的内部工具，需要加进软件白名单”。

还是不放心？您可以自己验证：

1. 打开网站 www.virustotal.com，把 `FlySense.exe` 上传上去扫一遍，自己看结果（70 多个引擎里偶尔有 1~3 家会拦，属于正常现象，重点看微软、卡巴斯基这些主流引擎）。
2. 拿我们提供的 **SHA256 指纹**和您手上的文件核对一下，确认文件没有被别人替换过。核对方法：在 exe 所在文件夹的地址栏输入 `cmd` 回车，然后粘贴下面这行再回车：

```
certutil -hashfile FlySense.exe SHA256
```

（把显示出来的那串字符和我们给您的指纹对一下，一样就说明文件是原版的。）

另一个更省事的办法：把整个软件文件夹（里面就一个 exe）加到杀毒软件的信任区，以后就不会再弹了。

3. 遇到报错，怎么把“病历”发给我们？

程序出错时会在**自己旁边**写一个日志文件，把这个文件发给我们，我们就能知道卡在哪一步。请按下面做：

第一步：找到日志文件。

打开 `FlySense.exe` 所在的文件夹，找这两个文件：

文件名	什么时候会有
`FlySense_error.log`	运行中出错时自动生成
`FlySense_selftest.log`	做完下面的“自检”后会生成

如果没看到文件，可能是 Windows 隐藏了“已知扩展名”——重点看**文件名带 FlySense 的那个文本文件**就行。

第二步（如果连界面都打不开）：做一次“自检”。

1. 打开 `FlySense.exe` 所在的文件夹；
2. 在文件夹**顶部地址栏**里输入 `cmd`，按回车（会弹出一个黑色窗口）；
3. 在黑窗口里粘贴这一行，按回车：

```
FlySense.exe --selftest
```

1. 它会自己检查一遍（不开界面），几秒后文件夹里会出现 `FlySense_selftest.log`。

第三步：把日志发给我们。

把 `FlySense_error.log` 或 `FlySense_selftest.log` 用微信/QQ 当文件发过来（**不要截图**，截图看不到里面的细节），再顺便告诉我们：

1. 您点的是哪个按钮、点了之后发生什么；
2. 您用的数据大概多少行；
3. 报错弹窗上写的那句话（有截图更好）。

有这三样，基本当场就能定位。

4. 双击了没反应，或者窗口一闪就没了

- 先看看杀毒软件的“隔离区 / 恢复区”里是不是有 `FlySense.exe`——十有八九是被拦了（见上面第 2 条）。
- 确认您双击的是 `FlySense.exe`，而不是别的文件。
- 如果还是不行，按上面第 3 条做一次自检，把 `FlySense_selftest.log` 发给我们。

5. 点了【开始运行】没反应?

- 常见原因：**还没选文件**。请先点【选择数据文件】，看到右边出现文件名再运行。
- 如果按钮是灰的，说明正在跑，等它跑完就好；跑完按钮会自己变回可点。

6. 【导出结果】按钮是灰的，点不了?

- 正常现象：**要先把【开始运行】跑完**，有结果了导出按钮才会亮起来。

7. 提示“数据超过 50000 行，请先抽样再运行”

数据太大了，两两比对会非常慢。请在 Excel 里先取一部分（比如随机抽 5000~10000 行）另存成一个新文件，再跑一次。

8. 提示“表格里没有数值列，无法分析”

意思是这张表里**一列数字都没有**（全是文字）。请检查：金额、次数、时长这类列是不是被存成了“文本格式”。在 Excel 里把这一列改成“数字”格式，另存为 `.xlsx` 再试。

9. 提示“旧版 .xls 不支持”

用 Excel 打开这个文件 → 【文件】→ 【另存为】→ 格式选 `Excel 工作簿 (*.xlsx)` → 保存，然后重新选这个新文件。

10. CSV 文件打开是乱码 / 提示编码无法识别

CSV 的编码五花八门。最省事的办法：**用 Excel 打开这个 CSV，另存为 `.xlsx`**，然后用 xlsx 跑，就不会有编码问题了。

11. 结果框里的字看不全 / 窗口太挤

把窗口**边缘往外拖**放大就可以了；太小了拖不动的话，双击窗口标题栏可以让它最大化。

12. 我的数据会被上传吗？需要联网吗？

不需要联网，也不会上传。 软件全程在您自己的电脑里计算，拔了网线照样能用。这也是为什么它可以放心地处理客户名单这类敏感数据。

13. 结果里挑出来的“异常”，是不是就是错的？

不一定是错的，只是“跟大多数行不一样”。 请把它当成一份“重点复核名单”：真正的录入错误通常一眼就能看出来（比如金额多了几个零、次数是负数），而“高消费的大客户”也会被挑出来——那属于正常情况，直接忽略即可。

14. 每次打开都弹出“激活窗口”，或者提示“授权码无效”？

说明这台电脑还没激活成功。照着下面两步查：

1. **先看状态栏**：主界面最底下一行写的是什么？

- 「已授权」= 没问题，不会再弹窗；
- 「试用中：还剩 X 天 / X 次」= 还在免费试用，也正常；
- 「试用期已结束」= 该填授权码了（见本指南《激活与授权》一章）。

1. **授权码是不是给错了**：授权码和机器码是**一一对应**的。如果把 A 电脑的授权码

填到 B 电脑上，就会提示“授权码无效”。确认一下您发给我们的机器码，和这台电脑上显示的**完全一样**。

实在搞不定，就把激活窗口截个图发给我们（要能看到**机器码**和**红色的提示文字**），

我们一眼就能看出是哪一步对不上。

15. 换了电脑 / 重装了系统，软件又要激活?

- **重装系统**：主板和 CPU 没变，机器码通常也不变，把**原来那个授权码**重新填一次就行。
- **换了新电脑**：点【复制机器码】把新机器码发给我们，重新申请一个授权码。
- 建议把授权码抄在交付包的 `README.txt` 上，以后重装系统省得再找我们。

十、五分钟试一遍（演示数据）

如果您想先看看效果，可以用我们提供的演示数据跑一遍（数据里**故意埋好了“标准答案”**，用来验证软件是否正常）。

演示数据里埋了什么：

- **3 对完全一样的重复行**：Excel 第 12 行与第 67 行、第 25 行与第 88 行、第 40 行与第 95 行。
- **2 行数值离谱的异常行**：第 77 行（交易金额 99999999、月消费额 9999999、停留时长 99999）、第 90 行（登录次数 -1、停留时长 88888、交易金额 -8888）。

怎么试：

1. 打开软件，点【选择数据文件】，选中随包的演示数据文件 `演示数据\test_data.xlsx`。
2. 模式选【去重】，严格度保持默认 **0.80**，点【开始运行】。

→ 应该提示“**发现 3 组重复，涉及 3 对，其中 3 行属于重复项**”。

1. 点【导出结果】保存，打开 Excel 核对“重复对”这张表，应该正好是 12+67、25+88、40+95 这三对。
2. 回到软件，模式改选【异常检测】，把**异常比例调到 2.0**，点【开始运行】。

→ 应该提示“**判定异常 2 条**”，导出的明细里应该是**第 77 行和第 90 行**。

如果跑出来和上面写的**完全一致**，说明软件在您的电脑上工作正常。

> （如果没看到「演示数据」这个文件夹，找我们要一份就行。）

十一、名词表

名词	大白话
数据行	表格里的一行，通常代表一个客户 / 一笔订单 / 一条记录
数值列	内容是数字的列（金额、次数、时长……），只有这类列参与计算
重复组	一组内容相同的行（一般是两行）
保留 / 重复	同一组重复行里，软件标“保留”的是留下的那行，标“重复”的是可以删掉的那行
相似度	两行有多像，越接近 1 越像。1.000 就是一模一样
严格度	判定“算不算重复”的门槛，数字越大越严格（挑得更少、更准）
异常分	每行的“格格不入程度”，分数越高越可疑
分数阈值	超过这条线才会被判定为异常；由“异常比例”滑块决定
异常比例	打算把最不像话的百分之几挑出来。设 2% 就是“挑最离谱的 2%”

十二、批量处理模式（一次跑完一整个文件夹）

这一章解决什么问题

以前一次只能选一个文件，100 个客户的表要点 100 次。企业版的【批量处理】就是解决这个的：

勾一下、选一个文件夹，里面的表格**全部自动跑一遍**（去重 + 异常检测一起做），

每个文件的结果分开导出，最后再送您一张汇总报告。

四步搞定

第一步：勾上【批量处理】

界面最上面【第 1 步：选择数据】那一行，最左边有个方框，写着【批量处理（一次跑完整个文件夹，企业版）】。

- 把它**打上勾** → 右边那颗按钮会从【选择数据文件】变成【选择文件夹】。
- 不想批量了，再点一下取消勾选，按钮就变回去。



> [截图：勾上【批量处理】后，按钮变成【选择文件夹】，右边多出【包含子文件夹】]

第二步：确认【包含子文件夹】是勾上的

这个勾选框就在【包含子文件夹】按钮旁边，**默认就是勾上的**：

勾选框状态	软件怎么找文件
勾上（默认，推荐）	连子文件夹一起找。数据按月份、按区域分开存，也能一次跑完
取消勾选	只看您选的那一层，子文件夹里的不管

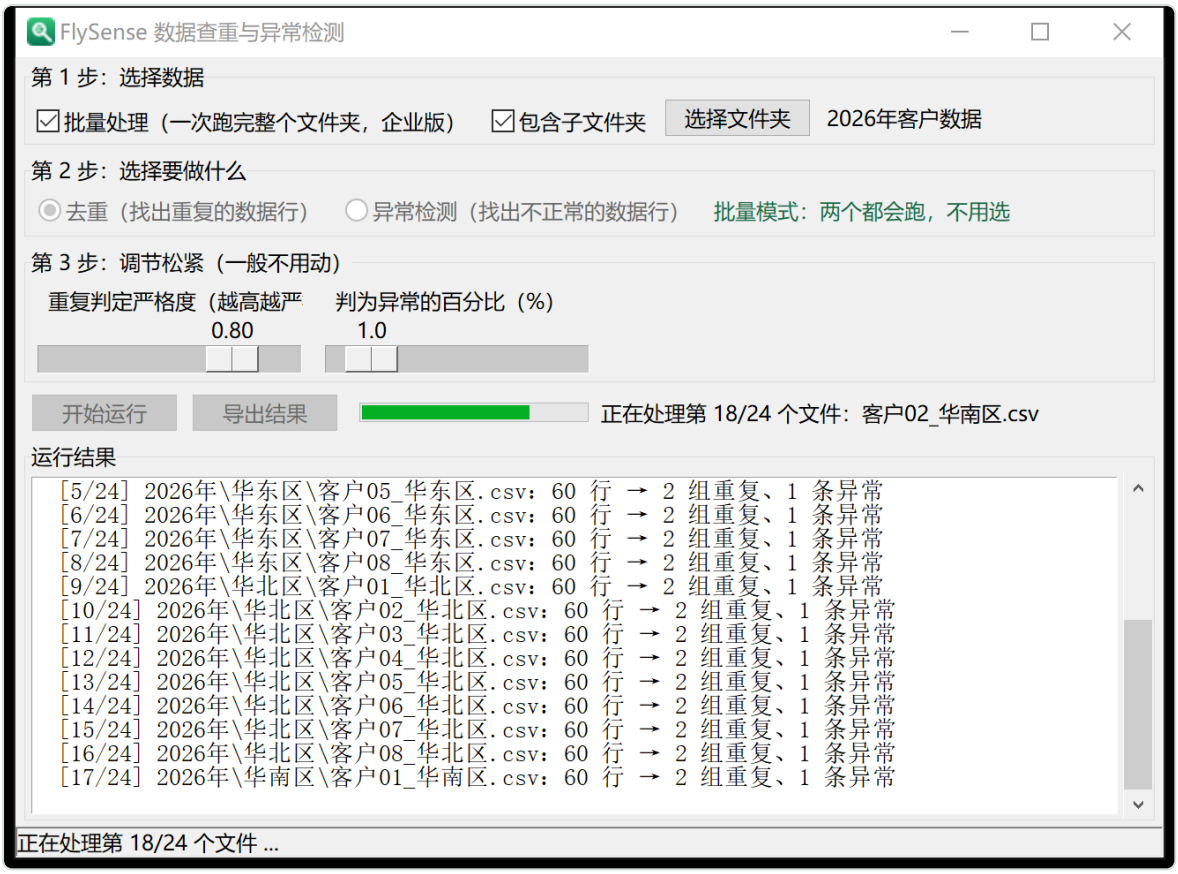
第三步：点【选择文件夹】→ 选好 → 点【开始运行】

选好之后，结果框里会先报个数，例如：

```
里面找到 3 个 .csv / .xlsx 文件（含子文件夹；已跳过隐藏文件和 ~$ 开头的临时文件）
```

第四步：等进度条走完

跑的时候会出现进度条，还会写「正在处理第 X/Y 个文件」：



> [截图：进度条在走，下面写着「正在处理第 18/24 个文件」]

跑完是这样：



> [截图：批量跑完，结果框里写着「成功 3 个，失败 0 个」和重复 / 异常的合计]

结果放在哪？

放在您选的那个文件夹里面，软件会新建一个叫 `FlySense_处理结果` 的文件夹，所有结果都在里面。



> [截图：FlySense_处理结果 文件夹里，带前缀的结果文件 + 汇总报告]

结果文件名前面的「前缀」是什么意思？

前缀是为了**防止不同文件夹里的同名文件互相覆盖。**

原文件（在哪个位置）	导出的结果叫什么
`客户A_华东区.csv`（就在您选的文件夹里）	`客户A_华东区_去重结果.xlsx`、`客户A_华东区_异常结果.xlsx`
`华南区\客户B_华南区.csv`（在子文件夹里）	`华南区_客户B_华南区_去重结果.xlsx`、`华南区_客户B_华南区_异常结果.xlsx`
`2026年\Q1\订单明细_2026Q1.csv`（两层子文件夹）	`2026年_Q1_订单明细_2026Q1_去重结果.xlsx`、`2026年_Q1_订单明细_2026Q1_异常结果.xlsx`

规则一句话：结果名 =【它在哪个子文件夹】+【原文件名】+【_去重结果 或 _异常结果】，

一层一层用下划线连起来。层级再深也一样接下去。

> **原文件一个都不会动**，FlySense 只读不写。上面的前缀只影响导出文件的**名字**， 不会真的在您电脑上建子文件夹。

会自动跳过哪些文件?

- 隐藏文件（名字以 `.` 开头的）
- 隐藏文件夹（以及它里面的所有东西）
- Excel 打开时留下的临时文件（名字以 `~$` 开头的）
- 不是 `.csv` / `.xlsx` 的文件（比如 `.txt`、`.pdf`、老的 `.xls`）

汇总报告：《企业数据汇总报告.xlsx》

跑完，结果文件夹里会多一个 `企业数据汇总报告.xlsx`，里面有三张工作表：

企业数据汇总报告.xlsx

工作表：汇总

项目	数值
处理时间	2026-09-21 18:01:33
数据文件夹	C:\Users\Administrator\Desktop\flysense\批量测试数据
结果文件夹	C:\Users\Administrator\Desktop\flysense\批量测试数据\FlySense_处理结果
文件总数	3
成功处理	3
跳过或失败	0
数据总行数	145
重复组数合计	4
重复行数合计（建议删除）	4
异常条数合计	3
包含子文件夹	是
处理模式	去重 + 异常检测
去重严格度	0.8
异常比例（%）	1
处理引擎	FlySense 1.0.0（4000 位稀疏编码）

工作表：按文件明细

所在子文件夹	文件名	状态	数据行数	列数	参与计算的数值列	重复组数	重复行数（建议删除）	异常条数	去重结果文件	异常结果文件	备注
（根目录）	客户A_华东区.csv	成功	60	9	8	2	2	1	客户A_华东区_去重结果.xlsx	客户A_华东区_异常结果.xlsx	
2026年\Q1	订单明细_2026Q1.csv	成功	40	7	6	1	1	1	2026年_Q1_订单明细_2026Q1_去重结果.xlsx	2026年_Q1_订单明细_2026Q1_异常结果.xlsx	
华南区	客户B_华南区.csv	成功	45	9	8	1	1	1	华南区_客户B_华南区_去重结果.xlsx	华南区_客户B_华南区_异常结果.xlsx	

工作表：使用说明

项目	说明
这份报告是什么	FlySense 批量处理的汇总。每个原始表格的详细结果，在同一个文件夹里各有两个 Excel
重复组数	同一份文件里「内容几乎一样」的行被归成一组，一组通常是两行
重复行数	建议删除的行数：每组保留一行，其余算重复
异常条数	该文件里「跟大多数行都不一样」的行数，需要人工看一眼再决定怎么处理
每个文件的结果	命名规则：<子文件夹名>_<原文件名>_去重结果.xlsx 和 _异常结果.xlsx；文件直接放在第一层时没有子文件夹前缀
子文件夹怎么处理	勾选了「包含子文件夹」时，子文件夹里的表格也会一起处理，层级很深时用下划线连接（例如 2026_Q1_订单明细）
处理模式	「去重 + 异常检测」= 两样都算；命令行静默模式可以只跑其中一样，那时另一项在表里写「—（本次未做）」
异常条数为什么是 0	异常比例是按行数算的：文件越小，1% 可能还不到一条。文件都比较小的时候，把界面上「判为异常的百分比」调到 2%~5% 再看
为什么有文件被跳过	行数少于 2 行、超过 50000 行、或者文件本身读不出来。具体原因写在「备注」列
原始文件动过吗	没有。FlySense 只读不写，你的原始表格一个字节都没有被改动
严格度怎么调	重复挑多了（有误报）就往大调，明明是重复却没挑出来就往小调
命令行静默运行	企业 IT 可以用命令行参数（--input / --mode / --output）在后台跑，不弹窗口；运行日志写在结果文件夹里的 run_log.txt

> [截图：企业数据汇总报告.xlsx 的 汇总 / 按文件明细 / 使用说明 三张表]

工作表	里面是什么
汇总	一共处理了几个文件、成功几个、失败几个、重复多少行、异常多少条
按文件明细	每个文件一行：原始行数、重复行数、异常条数、导出的文件名
使用说明	这份报告怎么看、这次用的是哪套参数（严格度、异常比例）

> 要跟领导汇报的时候，把【汇总】那张表截图发过去就够了。

十三、命令行静默运行（给 IT 部门用）

这一章是给**会敲命令的人**看的。日常用软件的话，看到这里就可以停了。

什么时候用得上

- 每天凌晨自动跑一次，没人守在电脑旁边。
- 服务器上几百个文件，不想点鼠标。
- 已经有调度系统（比如 Windows 任务计划程序），想让 FlySense 当其中一环。
- 脚本需要知道「这次到底成了没有」——靠**退出码**判断。

用法一句话

```
FlySense.exe --input "输入路径" --mode 模式 --output "输出路径"
```

跟双击打开的那个是同一个程序，只是不弹窗口。

不弹窗口的好处：可以挂在后台、塞进定时任务，跑的时候不会有人误点。

参数说明

参数	必填吗	什么意思
`--input <路径>`	**必填**	要处理的**文件**或**文件夹**。给文件夹就整个文件夹一起处理（默认连子文件夹一起）
`--mode <模式>`	选填，默认 `all`	`dedup` = 只去重；`detect` = 只做异常检测；`all` = 两个都做
`--output <路径>`	选填	结果和日志存到哪个文件夹。**不写**就放在「输入所在文件夹\FlySense_处理结果」
`--version`	不用填值	只显示版本号，然后退出
`--help`	不用填值	显示用法说明，然后退出

> **路径写法一定要看这一条**：路径里有空格、中文、或者 `&` 这类符号时，**必须用英文双引号包起来**，例如 `--input "D:\2026 年客户数据"`。不加引号的话，空格后面的内容会被当成别的参数。

六个现成例子（复制就能改）

① 整个文件夹，去重 + 异常检测都做（最常用）

```
FlySense.exe --input "D:\客户数据" --mode all --output "D:\结果"
```

② 只查重复

```
FlySense.exe --input "D:\客户数据" --mode dedup --output "D:\去重结果"
```

③ 只查异常

```
FlySense.exe --input "D:\客户数据" --mode detect --output "D:\异常结果"
```

④ 只处理一个文件（没写 `--output`，结果会出现在 `D:\客户数据\FlySense_处理结果\` 里）

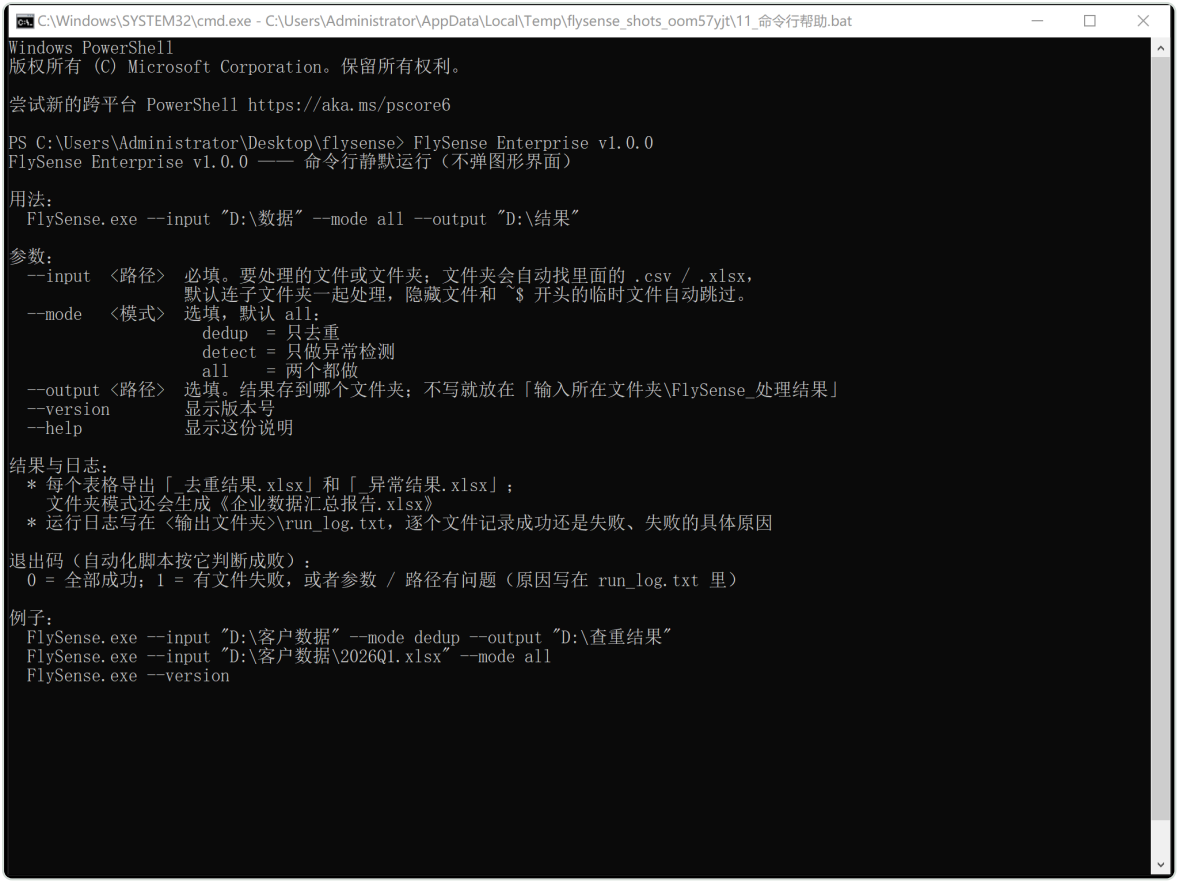
```
FlySense.exe --input "D:\客户数据\2026Q1.xlsx" --mode all
```

⑤ 看版本号

```
FlySense.exe --version
```

⑥ 看帮助

```
FlySense.exe --help
```



> [截图：PowerShell 里敲 --version 和 --help，屏幕上打出来的用法说明]

结果和日志放在哪

- 每个表格导出 `_去重结果.xlsx` 和 `_异常结果.xlsx`，跟图形界面一模一样。
- 文件夹模式还会生成《企业数据汇总报告.xlsx》。
- 运行日志写在 `<输出文件夹>\run_log.txt`。哪个文件成功、哪个失败、失败为什么，全在里面。
- 如果只做 `dedup`（不去异常检测），汇总报告里没做的那一栏会写「—（本次未做）」，不会填个 0 骗人。

退出码（自动化脚本看这个）

退出码	意思	脚本里该怎么办
0	**全部成功**，一个文件都没出问题	继续下一步
1	有文件没处理成功，**或者**参数 / 路径有问题	停下来，去看 `run_log.txt` 里的 `[失败]` 和 `运行失败`
2	这台电脑**没有授权**（免费试用也结束了）	先双击 `FlySense.exe` 完成激活（见本指南《激活与授权》一章）再重跑

> **没授权时什么都不会发生**：软件不会处理任何文件，直接退出，并在 `run_log.txt` 里 写明「**授权失败**」和本机机器码。 `--help` 和 `--version` 不需要授权，随时能看。

Windows 任务计划程序、Jenkins、Python 的 `subprocess`、批处理的 `errorlevel`，认的都是这个码。

几个实用提示

- **全程不弹窗、不需要点确认**，可以放心丢进定时任务。
- 输入路径不存在、或者给的是老 `.xls`，会直接把原因写进 `run_log.txt` 并返回 1，不会静悄悄地失败。
- 输出文件夹写不进去（没权限、盘符不存在）时，日志会自动退回到 `FlySense.exe` 旁边存一份，线索不会丢。
- 命令行用的参数和图形界面上的一样：**去重严格度 0.80、异常比例 1.0%**。想改默认值就改界面上的滑块。

附：配一个「每天凌晨 2 点自动跑」的定时任务

1. 开始菜单搜 `任务计划程序`，打开。
2. 右边点【创建基本任务】，名字写 `FlySense 每日跑数据`。
3. 触发器选【每天】，时间填 `02:00`。
4. 操作选【启动程序】，三栏这样填：

- 程序或脚本：`D:\FlySense\FlySense.exe`
- 添加参数：`--input "D:\客户数据" --mode all --output "D:\结果"`
- 起始于：`D:\FlySense`

1. 第二天早上打开 `D:\结果\run_log.txt`，看最后一行是不是「退出码：0」。

> **「起始于」那一栏一定要填**，填 FlySense.exe 所在的文件夹。不填的话，有些电脑会报「找不到路径」。

十四、如何看日志和排错

日志文件叫 run_log.txt，它在哪？

您是怎么用的	日志在哪
命令行静默运行，**写了** `--output`	您指定的那个输出文件夹里
命令行静默运行，**没写** `--output`	「输入文件夹\FlySense_处理结果\run_log.txt」
图形界面（双击打开的那个）	结果框里的文字就是日志；要存档看【八、常见问题 3】的办法

这是个纯文本文件，双击用记事本就能打开，**不会乱码。**

日志长什么样



> [截图：PowerShell 里跑完，屏幕上显示「退出码 = 0」，下面接着 run_log.txt 的最后 12 行]

一份完整的日志大概是这样（节选）：

```
FlySense 数据查重与异常检测工具 v1.0.0 —— 命令行静默运行
================================================================
开始时间：2026-09-21 17:58:41
输入：D:\客户数据
模式：all（去重 + 异常检测）
输出：D:\结果
参数：去重严格度 0.80，异常比例 1.0%
范围：整个文件夹（含子文件夹）
================================================================
找到 3 个表格文件（含子文件夹；已跳过隐藏文件和 ~$ 开头的临时文件）
  [1/3] 正在处理：客户A_华东区.csv
  [2/3] 正在处理：订单明细_2026Q1.csv
  [3/3] 正在处理：客户B_华南区.csv

每个文件的结果：
  [成功] 客户A_华东区.csv（60 行）；去重 2 行（2 组）→ 客户A_华东区_去重结果.xlsx；异常 1 条 → 客户A_华东区_异常结果.xlsx
  [成功] 2026年\Q1\订单明细_2026Q1.csv（40 行）；去重 1 行（1 组）→ 2026年_Q1_订单明细_2026Q1_去重结果.xlsx
  [成功] 华南区\客户B_华南区.csv（45 行）；去重 1 行（1 组）→ 华南区_客户B_华南区_去重结果.xlsx

================================================================
结果：成功 3 个，失败 0 个
      重复行合计 4 行（4 组）
      异常合计 3 条
结果文件夹：D:\结果
汇总报告：D:\结果\企业数据汇总报告.xlsx
================================================================
结束时间：2026-09-21 17:58:42
耗时：0.5 秒
退出码：0（全部成功）
```

看日志只要盯四个地方：

1. **开头几行**——确认「输入 / 输出 / 模式」是您要的。路径写错了，这里一眼就能看出来。
2. `[成功]` **和** `[失败]` **那几行**——每个文件的结论，**失败的原因就写在冒号后面。**

3. `结果：成功 X 个，失败 Y 个` ——一句话总结。
4. **最后一行「退出码」**——自动化脚本看的就是它。

退出码 0、1、2 分别代表什么

退出码	含义	下一步做什么
0	这次跑得**干净利落，一个文件都没出问题**	不用管，结果和报告都是全的
1	**有文件没处理成功**，或者**参数 / 路径有问题**	打开 `run_log.txt`，搜 `[失败]` 和 `运行失败` 两个词，原因就写在后面
2	这台电脑**没有授权**（免费试用也结束了）	日志里会写「**授权失败**」和本机机器码；先双击 `FlySense.exe` 激活再重跑

> 小提醒：**退出码 1 不等于数据有问题**。多数时候是「某个表正被 Excel 占着打不开」「某个文件其实是空表」「路径写错了一个字」这类小事，处理一下重跑就行了。

常见的失败原因（对着日志看）

日志里出现	一般是	怎么解决
`输入路径不存在：...`	路径打错字、盘符没挂上、文件夹被改名	从文件夹地址栏复制路径，记得加英文双引号
`[失败] ... 另一个程序正在使用`	这个表格正被 Excel 打开着	关掉 Excel 再跑一次
`[失败] ... 表格里没有数值列`	整张表都是文字，没有数字列	属于正常情况，跳过即可
`[失败] ... 只支持 .csv / .xlsx`	文件是老格式 `.xls`	用 Excel 打开，另存为 `.xlsx` 再跑
`[失败] ... 表格是空的`	文件是空表，或者只有表头	属于正常情况，跳过即可
`没有找到任何可处理的表格文件`	文件夹里没有 csv/xlsx，或者都在隐藏文件夹里	确认文件夹选对了，且【包含子文件夹】是勾上的
`参数有误：未知参数 ...`	参数名拼错了（比如写成了 `-input`）	敲 `FlySense.exe --help` 对照一下

遇到报错，怎么把日志发给我们做售后

三步，一分钟：

1. **找到** `run_log.txt`（位置看上面的表）。
2. **把整个文件发过来**：右键 `run_log.txt` → 【发送到】 → 【压缩(zipped)文件夹】，

然后当邮件附件发，或者直接拖进聊天窗口。**别只截一小段图**——

开头几行（这次用的什么参数）和失败那几行，前后文都得有，不然看不出原因。

1. **顺手补三句话**，能让我们一次就定位：

- 您是怎么用的：双击界面，还是命令行（命令行请把那行命令原样复制过来）。
- 输入文件夹大概多少个文件、都是什么格式。
- 您觉得哪里不对，例如「应该有 5 条异常，结果只出来 2 条」。

> **如果是在图形界面上出的错**：鼠标点进结果框 → `Ctrl+A` 全选 → `Ctrl+C` 复制， 粘到记事本里存成 `问题.txt` 发过来就行，效果和 `run_log.txt` 一样。

顺便回答客户 IT 最常问的：数据安全吗？

- FlySense **只在本地算**：不联网、不上传、不装服务、不改系统设置。您的表格数据一字不会离开这台电脑。
- 唯一写入硬盘的东西是**激活信息**：`%LOCALAPPDATA%\FlySense\license.dat` 和注册表

`HKEY_CURRENT_USER\Software\FlySense`，里面只有「未激活 / 试用中 / 已激活」和您的授权码，

不包含任何表格数据。删掉它们只是让软件回到未激活状态，不影响您的原文件。

- **原文件只读不写**，所有结果都是新文件。
- 想验证我们没往外传数据：把网线拔了再跑一遍，结果完全一样。

祝您用得顺手。

用的时候有任何问题、或者希望加什么功能（比如“能不能把重复行自动删掉” “能不能支持多张表一起比对”），随时告诉我们。