表头数据里面定义了这10个表里面的数据项，也有少量的sample 数据

数据里面涉及身份证等个人信息是加密的

讨论后觉得可以从最基本的逻辑入手，也就是说不正常的 理赔一定会关联相同的车或者人以及银行账号和电话号码。

* 车我们可以用车架号，也可以用车牌号（可能选一个即可）
* 人的话，基于保单角色可以使用身份证号码
* 理赔入账的银行账号容易理解，但是目前看可能信息不全，也许考虑之后加上
* 电话号码目前数据中只有报案电话号码，可能用处不大，以后考虑可以加上

如果就是 基于理赔号码、车和人建图，应该可以建立一个强关联的图，那么理赔之间的最短路径就是2，有问题的理赔一定会有较强的管理关系（比如通过车或者人可以之间关联多个理赔）。

另外，做了一些初步的分析工作。针对理赔表，利用excel 提供的功能（需要的话问下亚亮），计算了一下在表中相同车架号出现的次数，也就是说可以 看出来sample 数据中哪些车出现的次数较多，通过看理赔事故描述是可以说明一些问题的。

另外，csv 文件里面的数据是通过tab 键隔离的，excel 表文件是导入数据后excel处理过的。