运用的library和函数： pandas os numpy pathlib path openpyxl load\_workbook requests

问题：上游 下游分别代表的什么

如何输出多表sheet

未掌握的：

a. os.listdir (用于返回指定的文件夹包含的文件os.listdir(path）

b. tolist() (numpy库)将数组或者矩阵转换成列表

c. groupby 进行数据的分组以及分组以后的组内运算

d. (arrow.now().shift(months=-1).strftime('%Y%m')) 获取当前时间然后改变

e. at data.at['a'，'A'] 取data中行名为a 列名为A的值

iat data.iat['a','A'] 取data中第一行第一列的值

f. df.shape len(df) df.columns 得到每列的标题 df.value nan if none

df.loc[0] 第一行 df.loc[0:2, 'A'] iloc函数为Selection by Position

df.iloc[:,0] 第一列的所有数据 drop\_drop\_duplicates()

g：os.path.join()函数：连接两个或更多的路径名组件1.如果各组件名首字母不包含’\’，则函数会自动加上

h. glob() 的使用 文件操作相关模块， 用它查找符合自己目的的文件 （文件搜索） 和find不同！

i: 引入 load\_workbook 模块读取 excel 文件中对应的"修改"表中的数据，封装为列表并返回

J.需要及时熟悉pandas和numpy 在工作中的运用 和对xlsx和csv文件的读取处理和输出

k. tnow = datetime.datetime.now().strftime(**"%Y%m%d"**)

M. DataFrame.drop\_duplicates(subset=None, keep='first', inplace=False)

Subset. 列标签 可选 keep{“first”,”last”, False) 默认first first 删除第一次出现的重复项 last删除重复项除了最后一次出现 False 删除所有重复项 inplace 是否删除重复项或返回副本

质检月数据：

LEO AND YX

1.数据/数据处理情况

2. 每日mapping记录

a. read\_data 读取当天历史收集的mapping记录表 建立表用requests库到指定数据库中获取数据

b. 准备数据：准备历史数据 处理营销的历史数据 and 处理；leo的历史数据 数据汇总

c. 数据清洗： 正 负 数据 clear() yu() yu1() 正的里面吧负数的减掉 + 负数里面的把正的减掉 完成数据清洗

质检所需配置文件

(A).check.py

a. Get\_Delivered\_Dealer\_list 根据以往的交付物SM文件、白名单(二次交付)、不交付名 单返回一个经销商列表 这些经销商将不会出现在本次的质检中

b. check\_t1tot2: 检查T2经销商买到T1的经销商流向

c. check\_black 黑名单匹配关系

d. check\_me\_tome 检查是否自己卖自己

e. check\_repeat 经销商本身的重复

. check\_sm\_repeat 检查sm重复

g. 判断经销商与经销商之间的重复

h. check\_Nowhistory\_sum 当月与历史销量的对比

I . renameDF 改名用

j. 检查主列

k. 检查是否为null pdisnull 判断是否漏产品

chech\_CustNnemandcode 检查客户编码

Read\_data() 分别列出error 情况并尝试再读取

Check\_IM\_date： 质检IM文件是否只有上个月最后两的数据

LEO\_Check\_InMarket： 质检LEO的SM文件中inMarket标志的正确性

LEO\_codedifferentByNamelist： 质检LEO的SM文件与打单名单中同一客户的编码不同 的数据

(B)data\_clear.py

正 负 数据 clear() yu() yu1() 正的里面吧负数的减掉 + 负数里面的把正的减 掉 完成数据清洗

(C）DayDateAndMonthlydis.py

* fold\_scanner 文件扫描
* read\_data 文件读取 并给出ERROR 和解决方案
* DMKDay\_month 读取DMK的数据 并在xlsx中输出DMK日月数据对比结果
* LEODay\_month

1. do\_table

* 从总控表取出所需数据 制作一个汇总模板

1. duibi.py

* 对比 根据给定的经销商代码 拿当前月份流向和历史月份流向进行对比 （下游新增消失匹）

(F） formattable.py

* 汇总到最后的结果集 OutputQualityReport
* 读写到excel write\_excel 数据写到excel 可以写入不同的sheet
* 输出为一个总表，写明本次报告的标题，数据源，和数据文件的创建时间。并通过数据分析输出经销商是否存在以下情况（重复，自己卖自己， 跨省，是否漏产品，销量是否为negative， 断日期） 如果存在断日期 根据sort（）排出经销商的断日期天数。
* 输出下游新增消失重视程度 （消失客户数和新增客户数的总共统计 并对sum进行分析）
* 输出销售差异较大的公司 （各个月的数据）
* 输出跨省top5 通过sort 跨省总距离

主要负责的部分：字段标准化程序， 处理系统 质检系统 报表制定等服务

明天要做的事情： 设置好git 上传到github 接着看代码， 看怎么能优化代码 还有文件的输出