2月25日的总结：

删除多余的文件和一些之前的测试文件（还没完成） 理解Pipeline\_system的百度脑图，理解如何把代码的规范性运用到对zhijian\_month\_shuju的旧代码的优化

**存在的py文件**：

* Clear\_data.py 开始做数据清洗 消除红冲的历史数据
* Explain.py: 对于主数据匹配程序的说明
* Main\_all1107.py 准备主数据，并对主数据进行预处理 读当天的历史收集的所有mapping记录表 诺无则建立表 经销商code进行省份匹配
* Prepare\_data.py 准备营销的历史数据
* Read\_data.py 读当天的历史的所有mapping记录表 假如没有则建立该表
* Check.py

Check\_T1toT2 -> df.to\_csv(‘jjjj.csv’) && df.to\_csv(‘aaaaajjjj.csv’) 需要更改csv的输出路径

* Data\_clear.py 仅仅是用于数据清洗的def df1是销量为负的 df2是销量为正的 du和du1分别用于在df2里删除df1和在df2中删除df1
* DayDateAndMonthlydis.py： 运维销售返回情况（日返回情况，返回状态，月返回状态）扫描文件夹 来读取不同文件&&不同sheet
* DisName.py 读取瀚上床历史经销商数清单 读取清单中数据并判断状态 并输出在file\_history下面

* do\_table.py： 从总控表去除所需数据 制作一个汇总表模板，需要7列：1.经销商编码前缀2 经销商名称 3 产品线+当月打单名单 4 经销商级别 5 月度采集预案 6采集结果 7文件名过滤关键字
* duibi.py： 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 可改进的点： 把消失\_新增详情.xlsx 写到input数据folder里面 可以规范代码并做到一些优化
* Format\_table.py： 汇总到最后的结果集
* Full.py

读取补录名单 筛选今年的历史数据 筛选需要的产品 去重

df\_h.drop\_duplicates([‘1’,’2’,’3’], inplace=True) 筛选出今年的补录数据和补录产品

输出到补录名单中的补录情况.xlsx

* Jxs\_cf.py 查询经销商重复 需要修改路径的文件名：xnow\_cp\_qty.xls, 所有经销商信息.xlsx， 有无经销商重复
* ProvinceAndCityMapping.py 输入文件名和路径(无需更改） In:C:\Users\xiaofei.yu\Desktop\月报模板\行政区代码.xlsx'

Out: C:\Users\xiaofei.yu\Desktop\月报模板\省城对应关系表.xlsx'

* SM\_LEO.py 销售月报

In: D:\\质检相关文件\\数据返回情况表\\

In:r'D:\质检相关文件\需要质检的数据\LEO\SM\_202008\_2020090509142333\_LEO8月.csv'

拿取当月数据 筛选日期数据并找出最大日期关联

In: r’./黑名单/黑名单.xlsx

* **质检程序.py 输出一个tkinter 输出一个窗口表格**
* Main\_finally.py 读取历史数据并合并完，将处理完的数据写入到excel，写入不同的sheet. 需要写入excel的有 最大日期，DMP中已有得

数据条款，需要补录的总销量，DMP中已有得总销量.

整理基础配置到config.py 并分别分出 1.变量声明，2全局基本路径配置，3全局参数配置 server配置 全局文件名配置。 4编写字段名转换，或者读取改名文件的形式 代码规范性 ctrl+alt+l

如何正确设置路径 如下面的路径，使用r就防止了\n的转义

path = r'c:\Data\narcies.shp'

* 1

当然，我们也可以用其他方法设置路径

path = 'c:\\Data\\narcies.shp'

path = 'c:/Data/narcies.shp'

**明日规划**： 对zhijian\_month\_shuju（质检系统）的进行进一步理解，查看可优化的地方和优化的实现方法. 参考pipeline的框架。