# 第一阶段 业务分析师

# 建模数据分析师

## 网校相关视频

SPSS数据分析

## 明确本次课知识点，明确重点难点

### 【知识点目标】

* 记录选项
* 记录选项——选择
* 记录选项——样本
* 记录选项——排序
* 记录选项——汇总
* 记录选项——区分
* 记录选项——合并
* 记录选项——追加
* 字段选项
* 字段选项——类型、过滤器
* 字段选项——导出
* 字段选项——填充
* 字段选项——转置
* 字段选项——分区
* 字段选项——字段重排
* 范例——清洗会员信息

## 复习巩固作业讲解

无

## 本次课程任务讲解

### 【知识点1】记录选项

**“记录选项”选项卡**

“记录选项”选项卡可用于对**数据行**进行转换，包含选择、样本、汇总、排序、合并、追加、区分等。

“选择”节点是选出符合我们条件的数据；

“汇总”节点是将数据按照特定条件进行汇总统计；

“排序”节点是将数据按照一定的规则进行排序；

“合并”节点是将两个及以上的文件按照关键字等进行整合；

“追加”节点是将两个及以上的文件进行数据的累加；

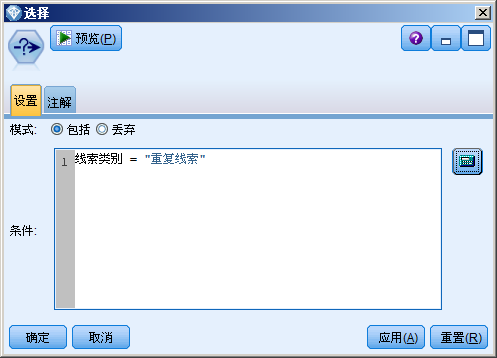
“区分”节点是按照条件将重复数据删除。



### 【知识点2】记录选项——选择

选择节点根据特定的条件选择或丢弃数据流中的部分记录，对数据源的数据起到筛选的作用。

数据：学生成绩.xlsx



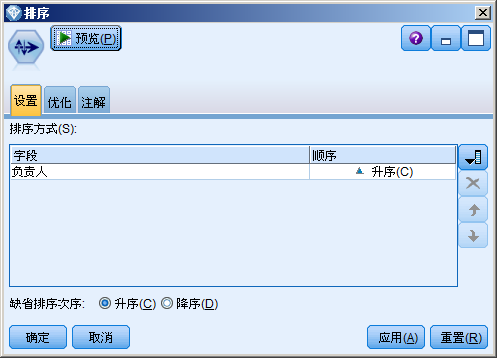
### 【知识点3】记录选项——样本

使用“样本”节点来选择记录的子集进行分析，或指定要废弃的记录的比例。

### 【知识点4】记录选项——排序

排序节点根据一个或多个字段的值，按照升序或者降序对记录进行排序。

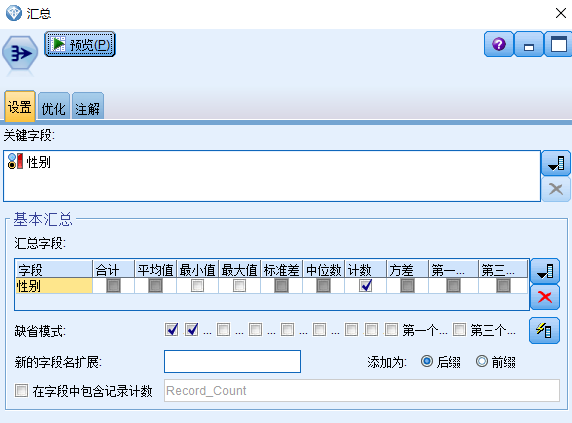
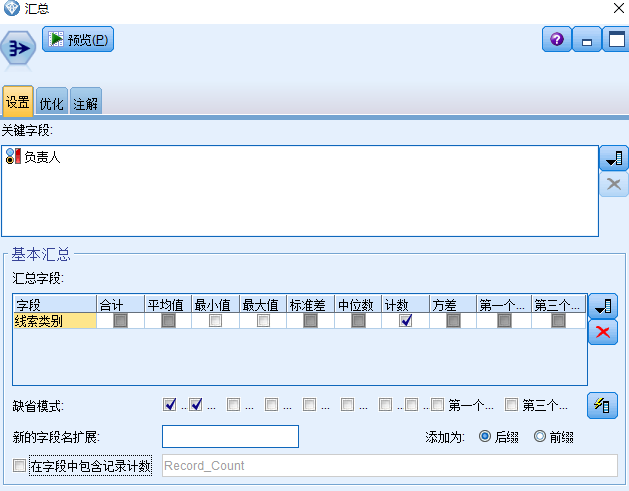
 

### 【知识点5】记录选项——汇总

汇总节点使用一个或多个关键字段作为汇总类别，并使用汇总字段为其计算汇总值，汇总结果会取代原有的数据集。数据源：历史数据.xlsx、客服咨询数据1710.xlsx

历史数据：统计不同性别的数据数量

客服咨询数据：统计不同负责人在这段时间的线索量

求客服咨询数据不同负责人的重复线索量

### 【知识点6】记录选项——区分

区分节点用于去除数据集中的重复记录。根据设置的关键字段确定何种记录属于重复，并可根据需要对其余字段创建组合记录

### 【知识点7】记录选项——合并

合并节点用于将多个数据集合并成一个数据集（在列方向），通过指定合并方式与合并条件，可以实现不同的合并效果。使用关键字合并时，要注意连接的方式（内，外等）

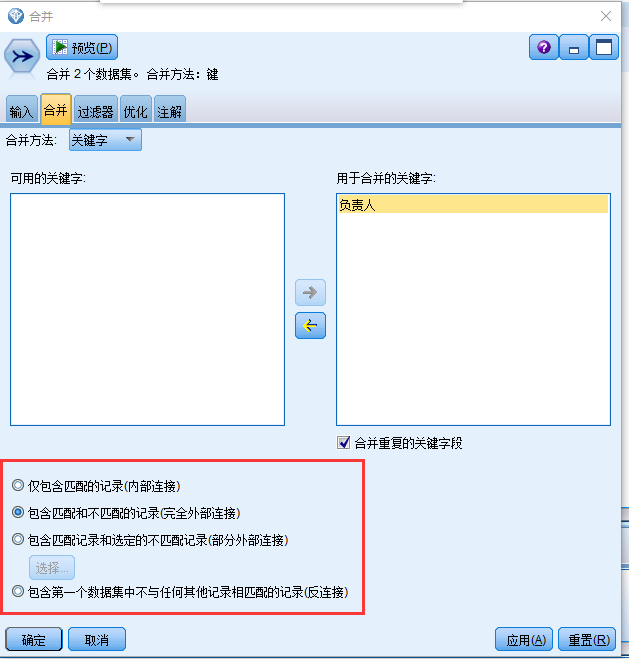
数据源：客服咨询数据1710.xlsx

目的：得到不同负责人的无效量和线索量

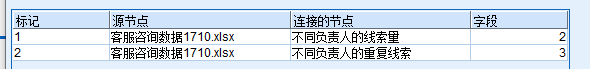
操作：获得无效的线索量——先选择再汇总

获得总的线索量——汇总

把不同负责人对应的无效的和总的线索量放到一张表里——合并



1. 内部连接：两张表或者多张表的共有数据显示在结果
2. 完全外部：所有表的数据全部显示出来
3. 部分外部连接：比如有三张表合并，只显示其中的一个或两个表的所有数据
4. 反连接：返回标记为1的（去除两张表共有的剩下的）数据



### 【知识点8】记录选项——追加

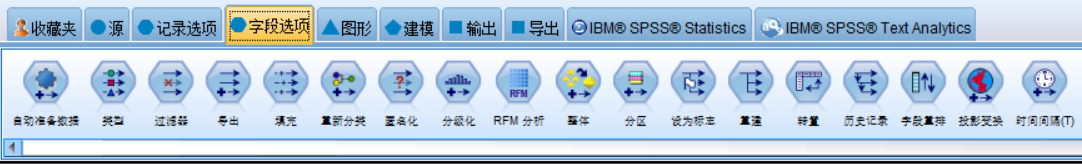
追加节点用于将多个数据集合并成一个数据集（在行方向），通常用于处理结构类似的多个数据源。使用时注意匹配字段的方法。数据源：1月会员信息和2月会员信息和3月会员信息

### 【知识点9】字段选项

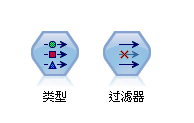
**“字段选项”选项卡**

“字段选项”选项卡可用于对列进行转换或生成新列，包含类型、过滤、导出、填充、转置、字段重排、分区等。



### 【知识点10】字段选项——类型、过滤器

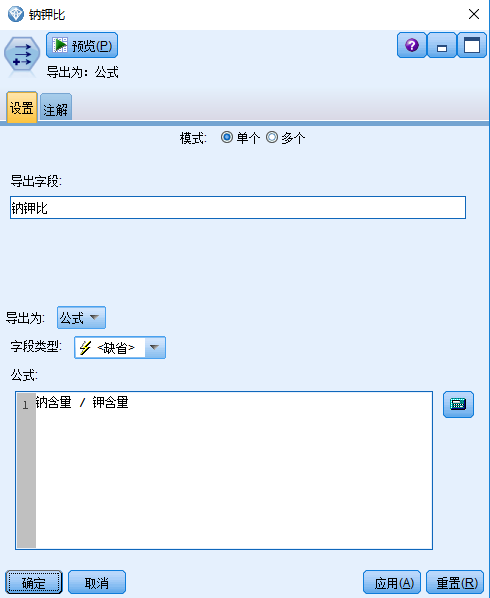
类型与过滤器节点设置字段属性及对字段进行过滤，与源节点中的设定类似。



### 【知识点11】字段选项——导出

导出节点将修改数据值或根据一个或多个现有字段创建新字段，主要利用CLEM表达式完成。

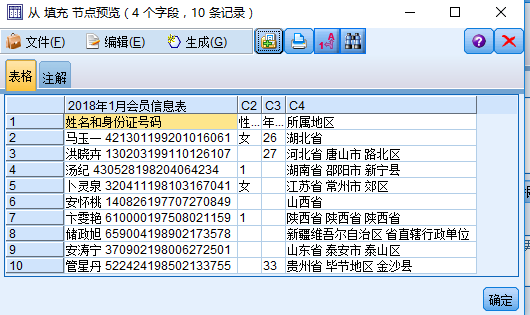
比如对历史数据.xlsx来求钠钾比



### 【知识点12】字段选项——填充

填充节点用于替换字段值和更改存储类型。可以选择基于指定的 CLEM 条件替换值。或者，也可以选择将所有空白值或空值替换为特定值。“填充”节点通常与“类型”节点结合使用，用于替换缺失值。

比如：对追加后的性别男替换为1



### 【知识点13】字段选项——转置

转置节点交换行和列中的数据，使字段变为记录、记录变为字段。



注意：转置的读取字段值必须是唯一的，不能是重复的，否则会报错。

### 【知识点14】字段选项——分区

“分区”节点用于生成分区字段，将数据分割为单独的子集或样本，以供模型构建的训练、测试和验证阶段使用。

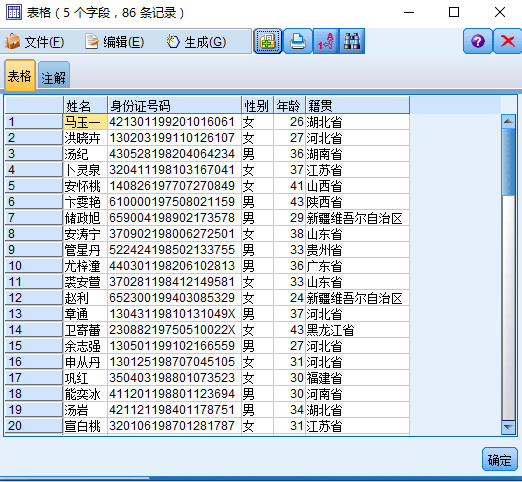
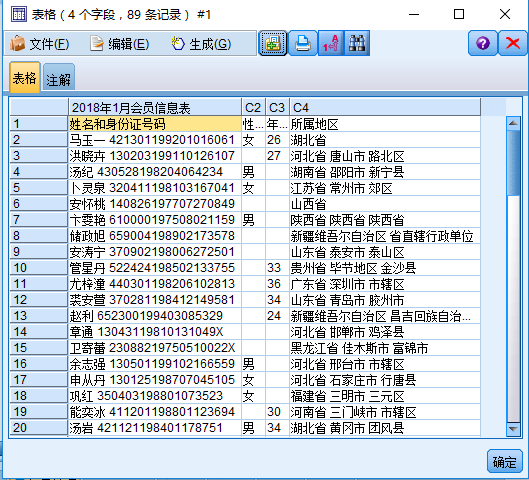


### 【知识点15】字段选项——字段重排

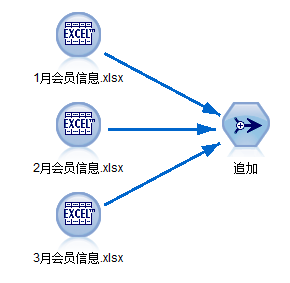
字段重排节点可以自定义数据集中字段的顺序。



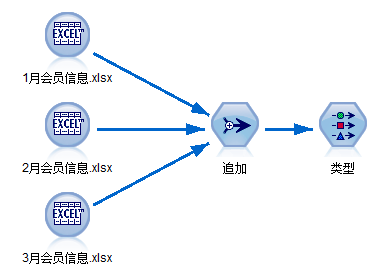
### 【知识点16】范例——清洗会员信息



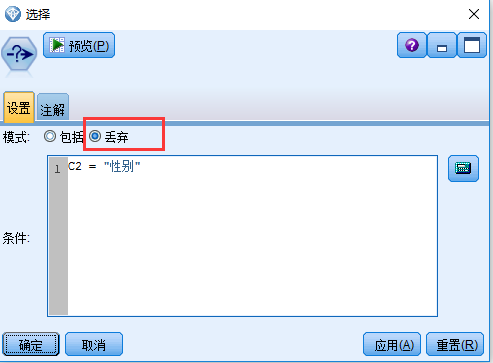
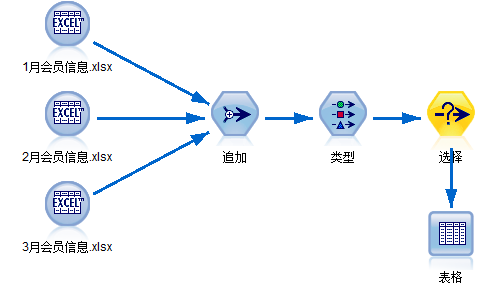
1、追加三份源数据



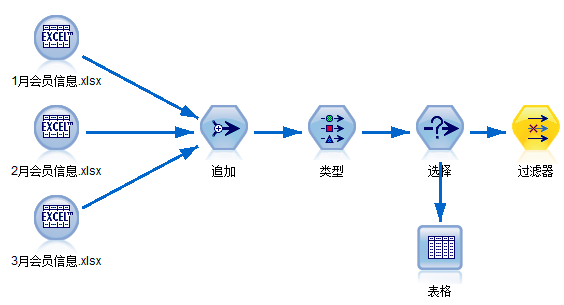
2、添加类型节点，读取具体值



3、把性别字段里的值“性别”丢弃

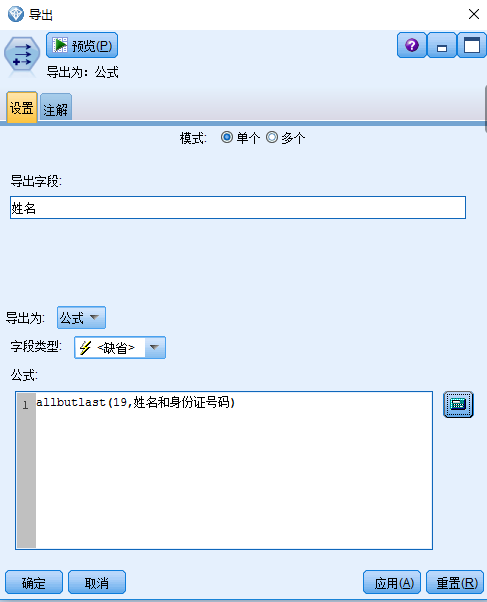
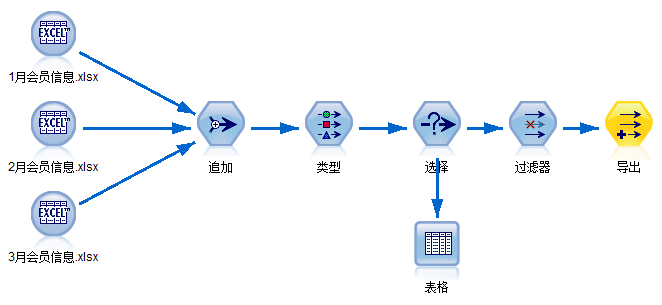


4、更改字段名称：过滤器

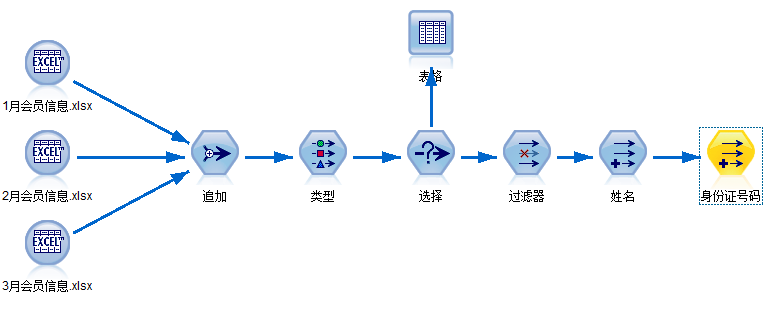


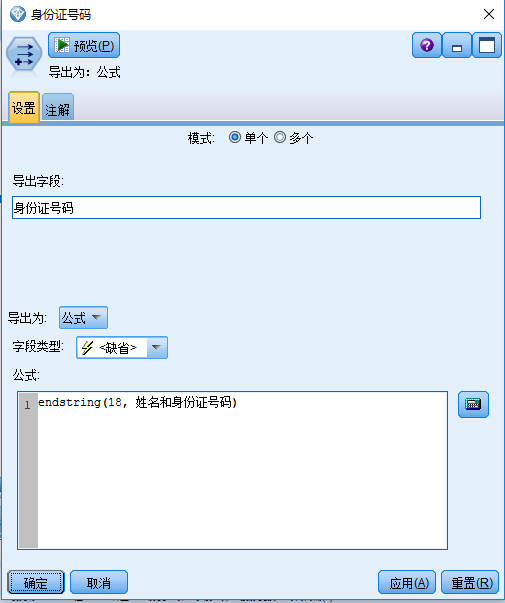


5、分开姓名和身份证号码：函数拆分，导出节点来新建字段

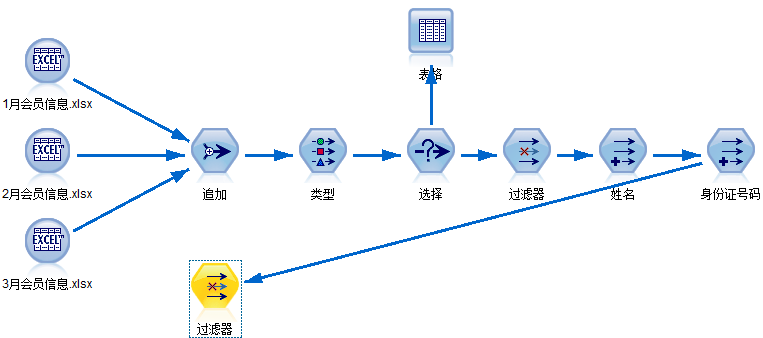


6、分开姓名和身份证号码：函数拆分，导出节点来新建字段





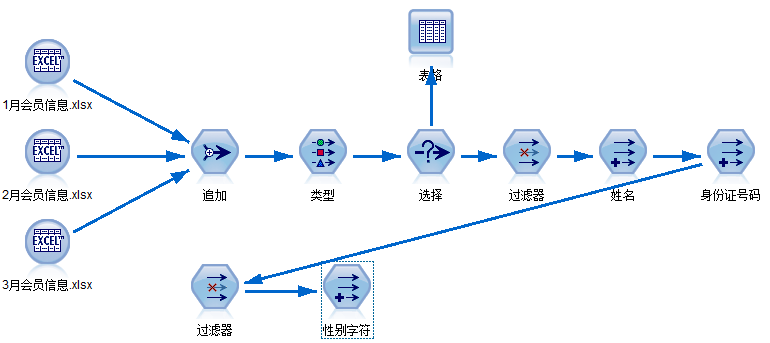
7、去掉姓名和身份证号码字段：过滤器

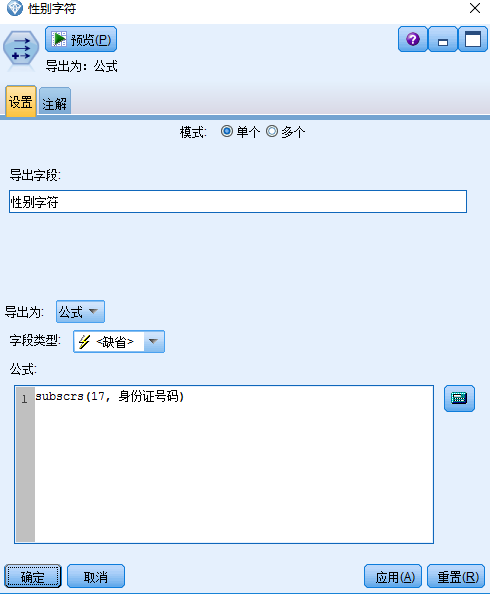




8、通过身份证号码来判断性别

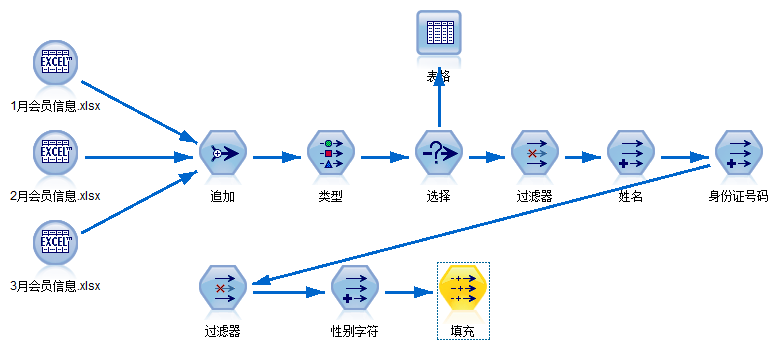
1）取出身份证号码的倒数第二位





8、通过身份证号码来判断性别

2）通过性别字符的奇偶来判断性别，取余数为1为奇数为男

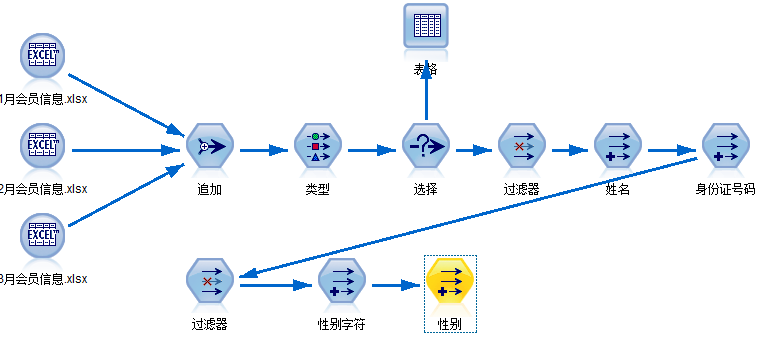




8、通过身份证号码来判断性别

2）通过性别字符的奇偶来判断性别，取余数为1为奇数为男

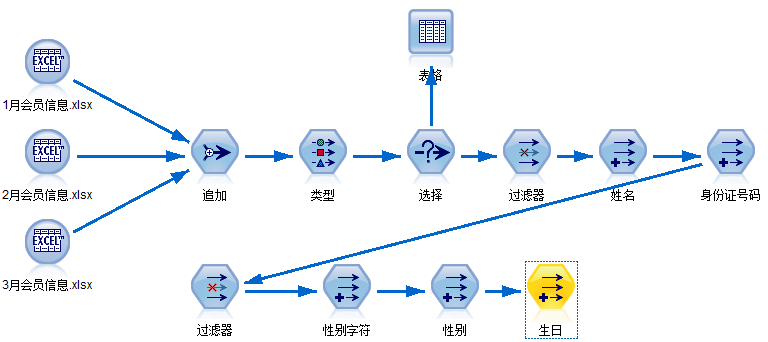
第二种方法：导出

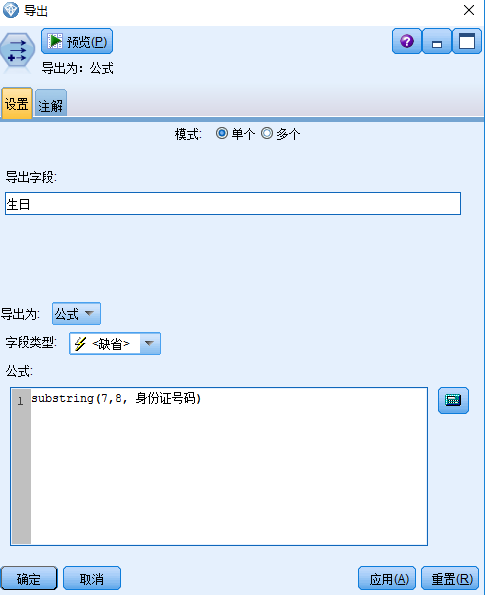




9、获取年龄——从身份证号码中获取生日，并用当前时间减去生日得到年龄

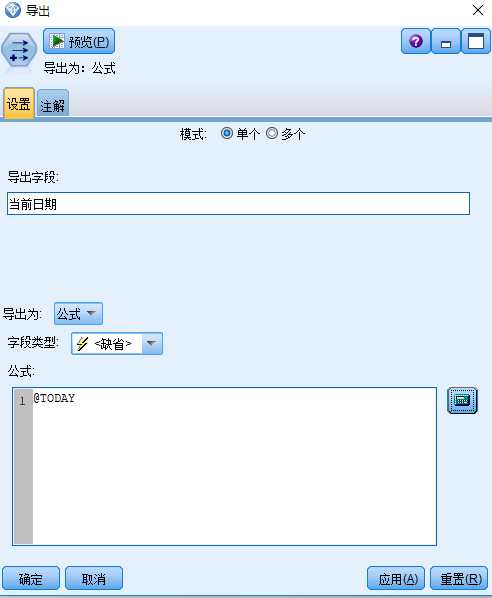
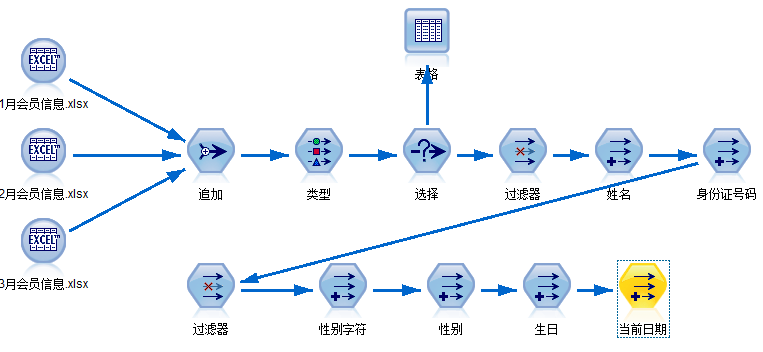
1）获取生日：字符串的substring函数



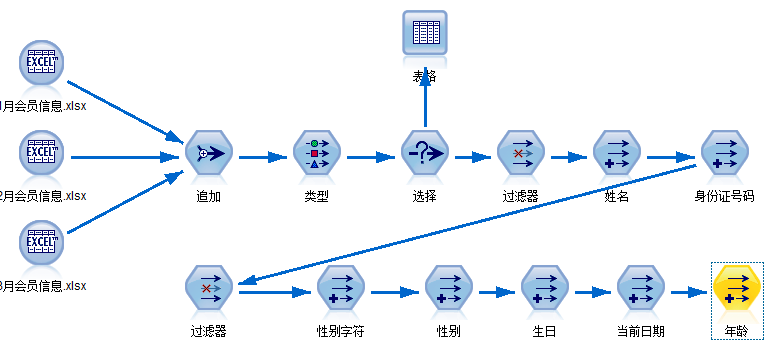


9、获取年龄——从身份证号码中获取生日，并用当前时间减去生日得到年龄

2）获取当前日期：日期和时间里有@today函数

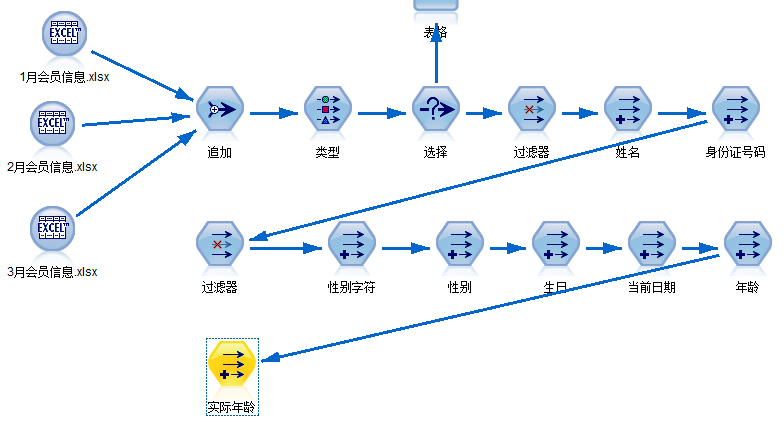


3）获取年龄：转换——to\_date函数，日期和时间——date\_years\_difference函数



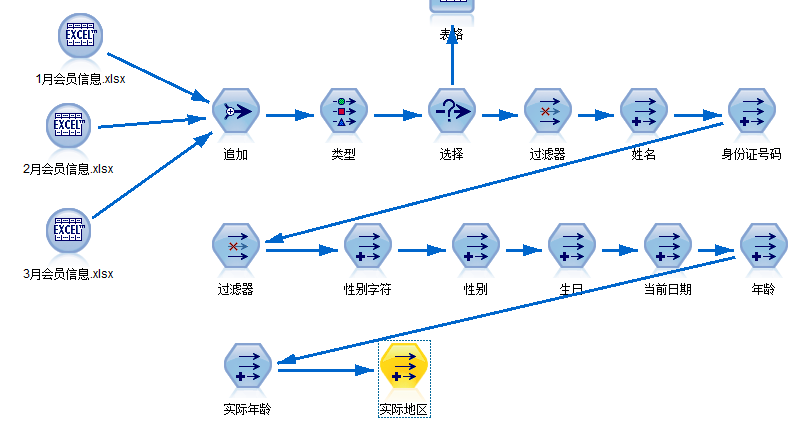


4）获取实际整数年龄：数值——intof函数



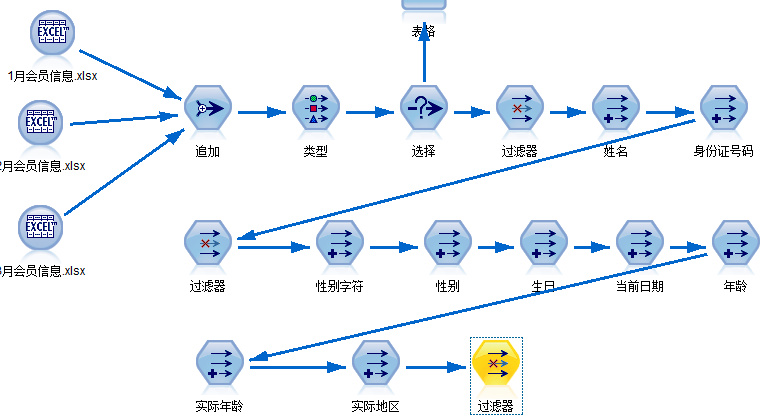


10、处理地区：字符串——textsplit函数



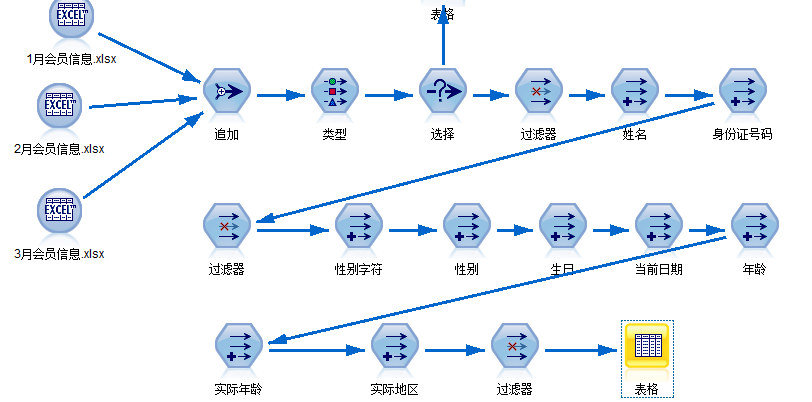


11、过滤不要的字段





12、添加表格查看最终效果



## 自主学习作业讲解

一、根据学生成绩表做以下操作

1、找出语文、数学成绩都在80分以上的同学

2、找出语文和数学都在前十名的同学

二、根据客服咨询数据1710做以下操作

1、利用SPSS Modeler找出不同负责人的重复量、无效量、线索量

2、绘制负责人和重复量的条形图，并且按照降序排序

三、使用提供的数据文件（销售信息），利用SPSS Modeler找出购买总金额前10位的顾客

1、根据订单明细求金额

2、合并订单和订单明细表，把金额放过去

3、求订单表里不同客户的总订单金额

4、对汇总以后的数据中添加客户信息：根据客户表合并



## 课程总结

* 记录选项
* 记录选项——选择
* 记录选项——样本
* 记录选项——排序
* 记录选项——汇总
* 记录选项——区分
* 记录选项——合并
* 记录选项——追加
* 字段选项
* 字段选项——类型、过滤器
* 字段选项——导出
* 字段选项——填充
* 字段选项——转置
* 字段选项——分区
* 字段选项——字段重排
* 范例——清洗会员信息

## 下次自主学习任务布置

#### 观看预习视频

数据建模分析师——SPSS数据分析

#### 课后作业

1. 参见直播平台
2. Qq群课后习题