

广电大数据客户价值分析

1 项目背景

随着经济的不断发展，人民的生活水平显著提高，对生活质量的要求也在提高。同时互联网技术的高速发展适应了时代的需求，为人们提供了许多娱乐的渠道。其中“三网融合”为人们在信息化时代利用网络等高科技手段获取所需的信息提供了极大的便利。

“下一代广播电视网（NGB；Next Generation Broadcasting）”即广播电视网、互联网、通信网“三网融合”，有线无线相结合、全程全网的广播电视网络。它不仅可以为用户提供高清晰的电视、数字音频节目、高速数据接入和语音等三网融合业务，也可为科教、文化、商务等行业搭建信息服务平台，使信息服务更加快捷方便。在三网融合的大背景下，广播电视运营商与众多的家庭用户实现信息实时交互。这使我们利用大数据分析手段为分析客户价值成为可能。

2 项目目标

1. 构建 RFMCL 指标
2. 构建客户价值分析模型

3 项目步骤

3.1 新建数据源

进入 PB 平台，点击左侧“数据源”导航栏，点击“新建数据源”，在弹出的下拉窗口中选择“数据来源于数据库”，如图 3-1 所示。



图 3-1 新建输入源

在弹出的窗口中填入如下信息：

1. 新建目标表：case_tv_f
2. 数据库类型选择 PostgreSQL
3. URL 链接：jdbc:postgresql://192.168.0.95:5432/case_tv
4. 用户名：postgres
5. 密码：root
6. SQL 语句：select phone_no, count(phone_no) as count_phone from media3 GROUP BY phone_no

如图 3-2 所示。

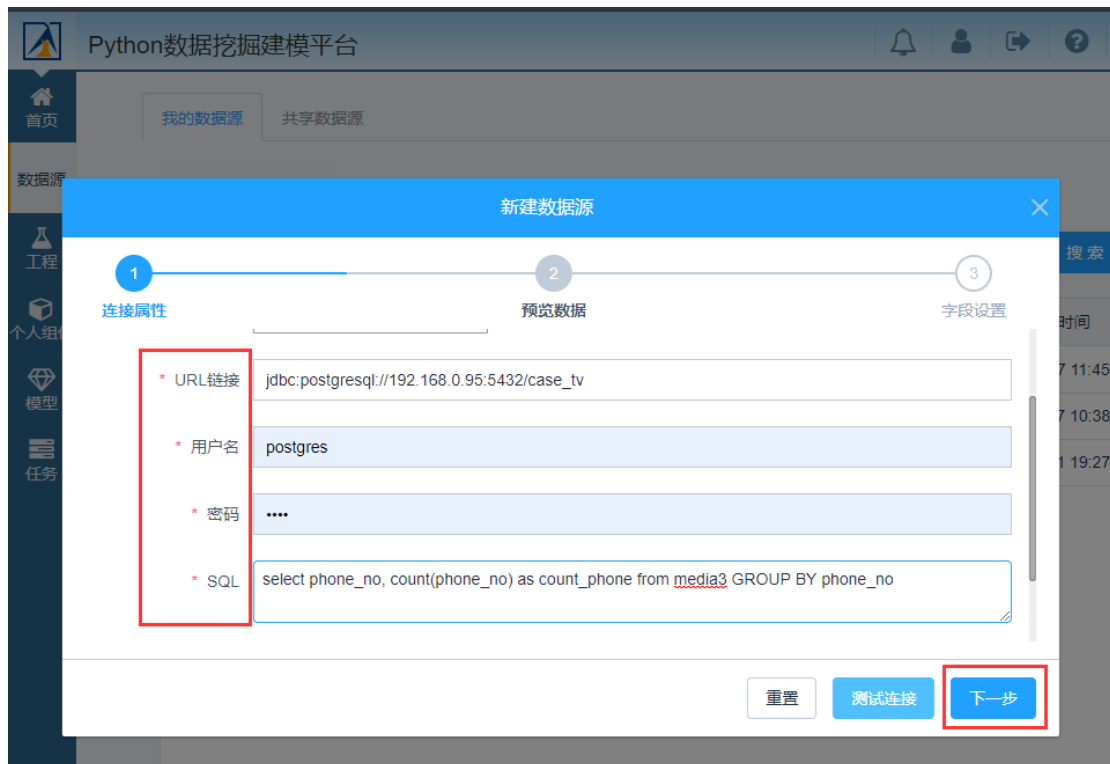


图 3-2 case_tv_f 连接属性

点击测试连接，提示连接成功，点击下一步进入数据预览，如图 3-3 所示。

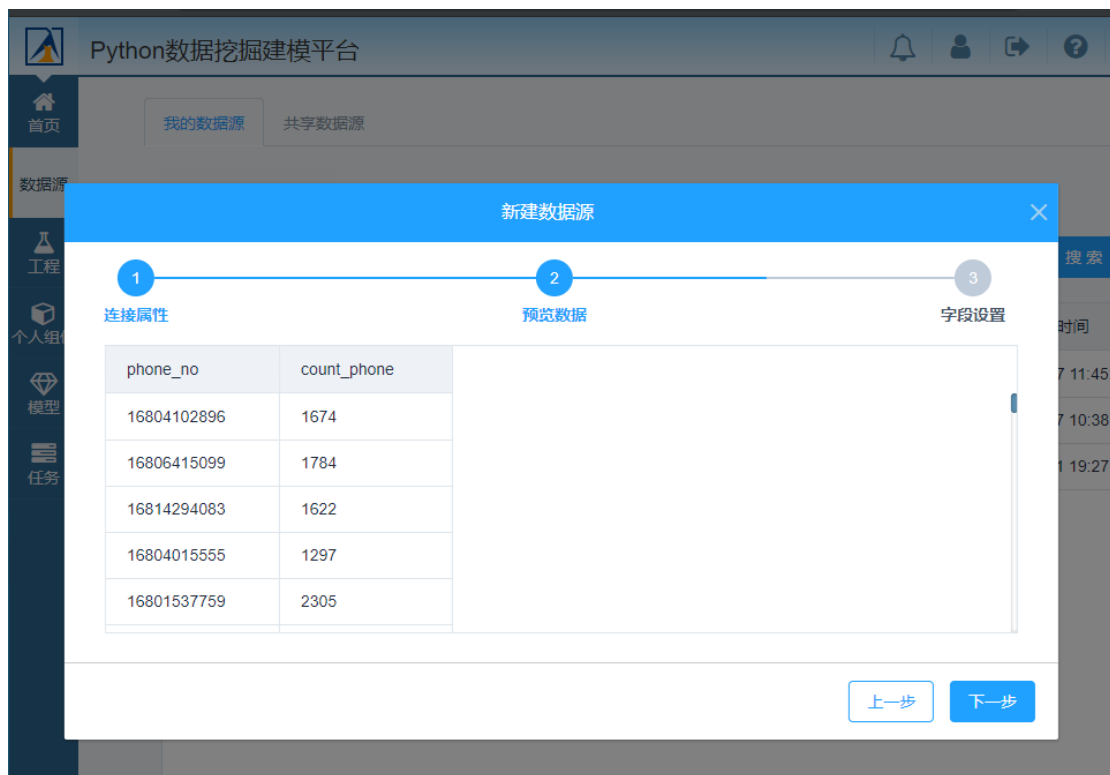


图 3-3 case_tv_f 数据预览

点击下一步，进入字段设置界面，将 phone_no 的类型改为字符型，点击确定，创建数

据源成功。

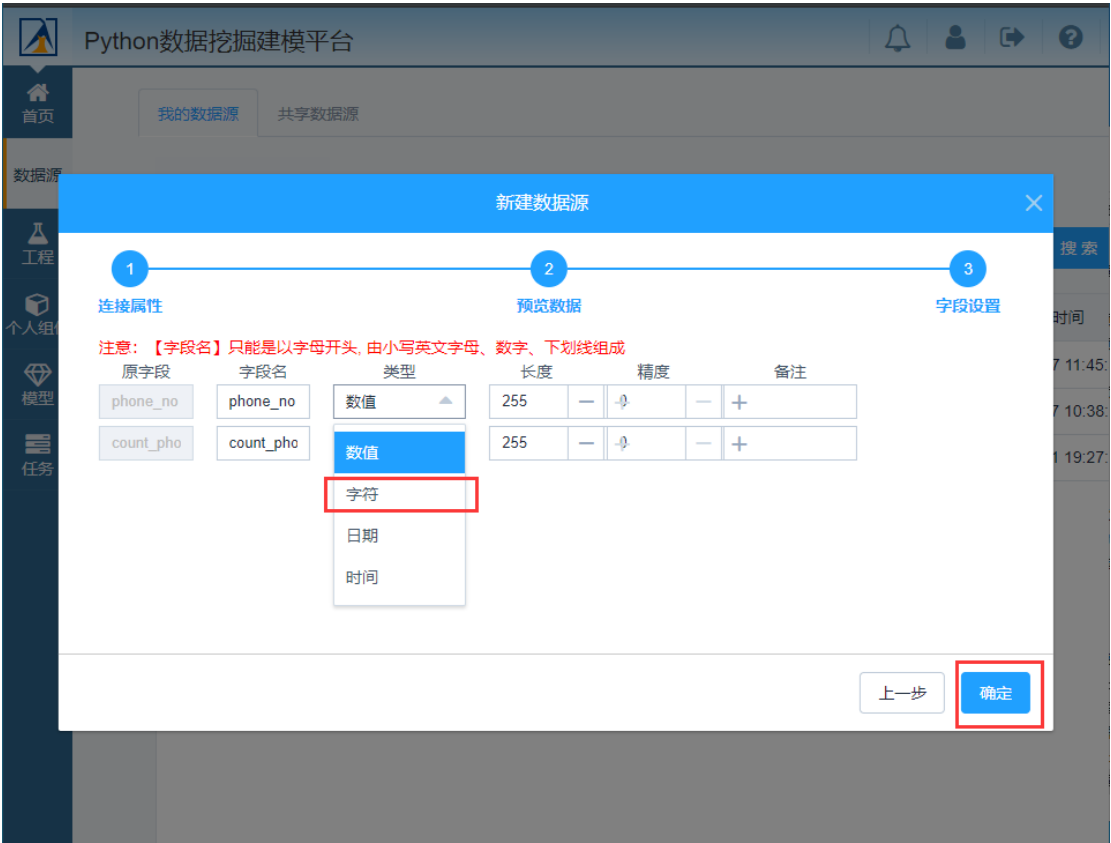


图 3-4 case_tv_f 字段设置

创建成功后，点击同步按钮，将数据同步至平台，成功后可点击消息铃，查看同步信息，如图 3-5 所示。

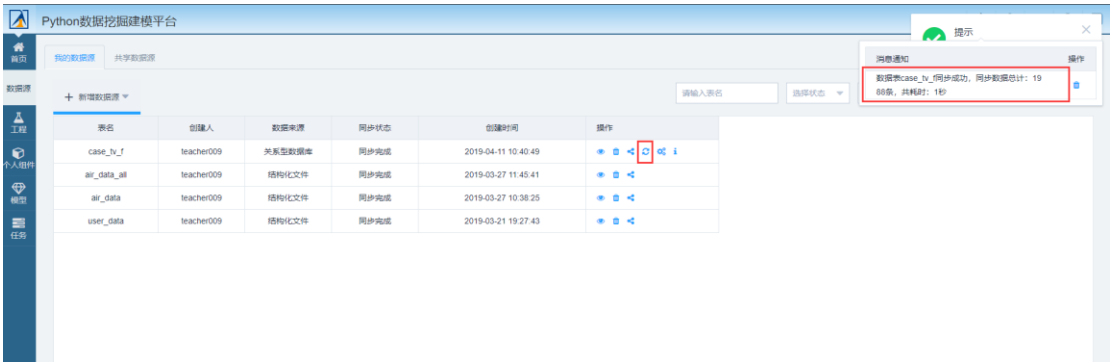


图 3-5 同步成功

重复 3.1 中的上述操作，分别导入 case_tv_c，case_tv_r，case_tv_billevents，case_tv_userevents，各部分信息如所示。

表 3-1 数据源信息

数据源名称	参数
-------	----

case_tv_c	1. 新建目标表: case_tv_c 2. 数据库类型选择 PostgreSQL 3. URL 链接: jdbc:postgresql://192.168.0.95:5432/case_tv 4. 用户名: postgres 5. 密码: root 6. SQL 语句: select phone_no, sum(wat_time)as sum_wat from media3 GROUP BY phone_no
case_tv_r	1. 新建目标表: case_tv_r 2. 数据库类型选择 PostgreSQL 3. URL 链接: jdbc:postgresql://192.168.0.95:5432/case_tv 4. 用户名: postgres 5. 密码: root 6. SQL 语句: SELECT phone_no, EXTRACT (epoch FROM now() - MAX (to_timestamp(media3.origin_time, 'yyyy-MM-dd hh24:mi:ss'))) AS now_origin FROM media3 GROUP BY phone_no
case_tv_billevents	1. 新建目标表: case_tv_billevents 2. 数据库类型选择 PostgreSQL 3. URL 链接: jdbc:postgresql://192.168.0.95:5432/case_tv 4. 用户名: postgres 5. 密码: root 6. SQL 语句: select * from billevents2
case_tv_userevents	1. 新建目标表: case_tv_userevents 2. 数据库类型选择 PostgreSQL 3. URL 链接: jdbc:postgresql://192.168.0.95:5432/case_tv 4. 用户名: postgres 5. 密码: root 6. SQL 语句: select * from userevents2

注意：所有数据源字段设置过程中均需将 **phone_no** 改为字符型，创建完成后需要进行同步操作。

3.2 工程配置

3.2.1 输入源

点击左侧导航栏中的“工程”，找到“我的工程”，右键新建目录“广电大数据营销推荐”，在新建的目录上右击，点击新建工程“客户价值分析”，如图 3-6 所示。

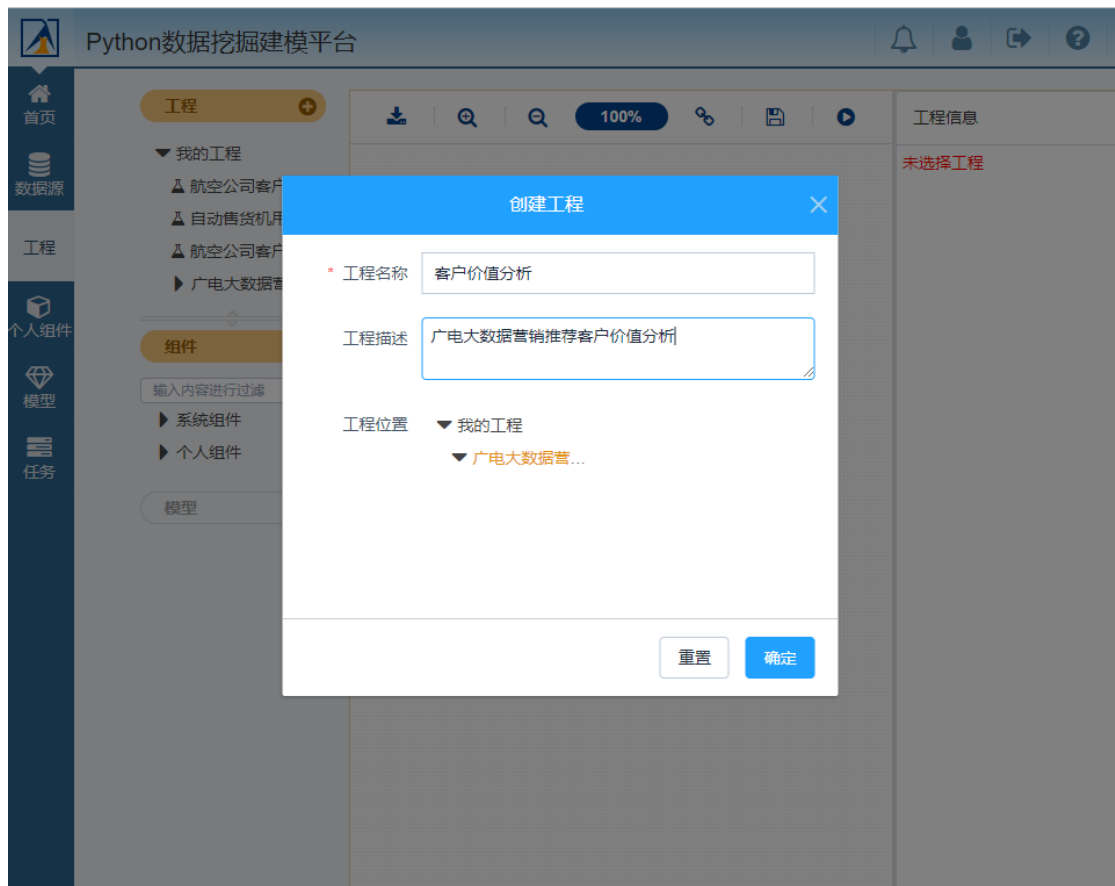


图 3-6 新建工程

点击右侧系统组件，拖入 5 个“输入源”组件，依次单击组件，在右侧参数设的“数据表”中分别填入“case_tv_c”，“case_tv_f”，“case_tv_r”，“case_tv_billevents”，“case_tv_userevents”，为方便区分，将其重命名为“C 指标”，“F 指标”，“R 指标”，“M 指标”，“L 指标”如图 3-7 所示。



图 3-7 拖入输入源

3.2.2 主键合并

点击右侧“预处理”，拖动两个“主键合并”，重命名为“主键合并-1”和“主键合并-2”，并将其相连，如图 3-8 所示。

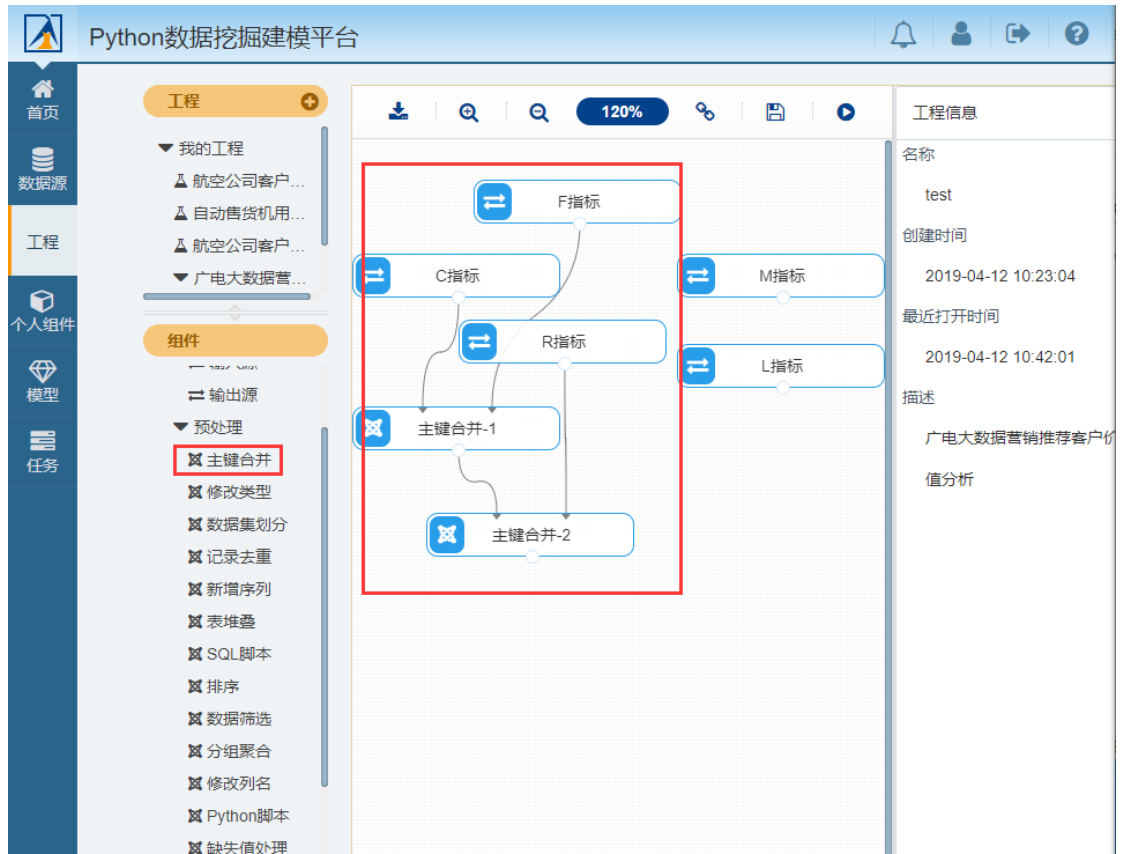


图 3-8 拖入表输出

单击“主键合并-1”，参数配置，如图 3-9 所示。

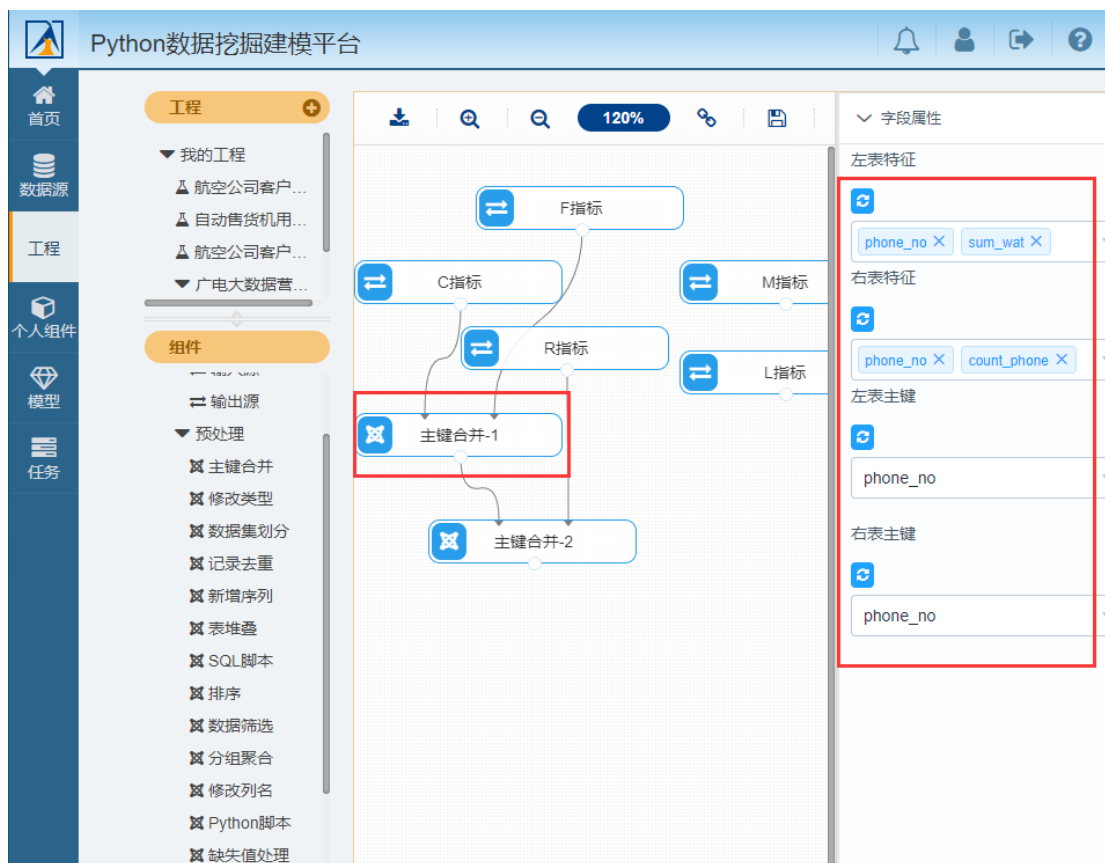


图 3-9 主键合并-1 设置

右键单击“主键合并-1”，运行到此节点，单击“主键合并-2”，参数配置，如图 3-10 示。

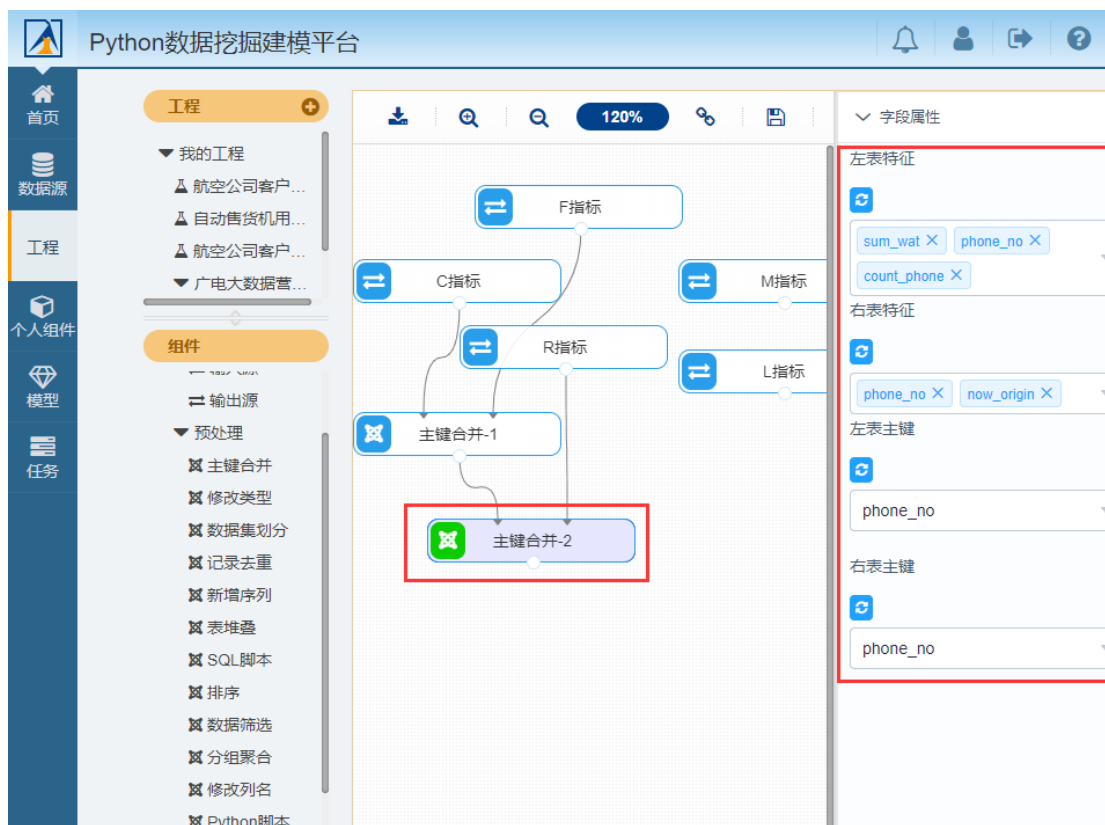


图 3-10 主键合并-1 设置

3.2.3 特征构造

点击右侧“预处理”，拖动“特征构造”，并将其与“M 指标”相连，如图 3-11 所示。

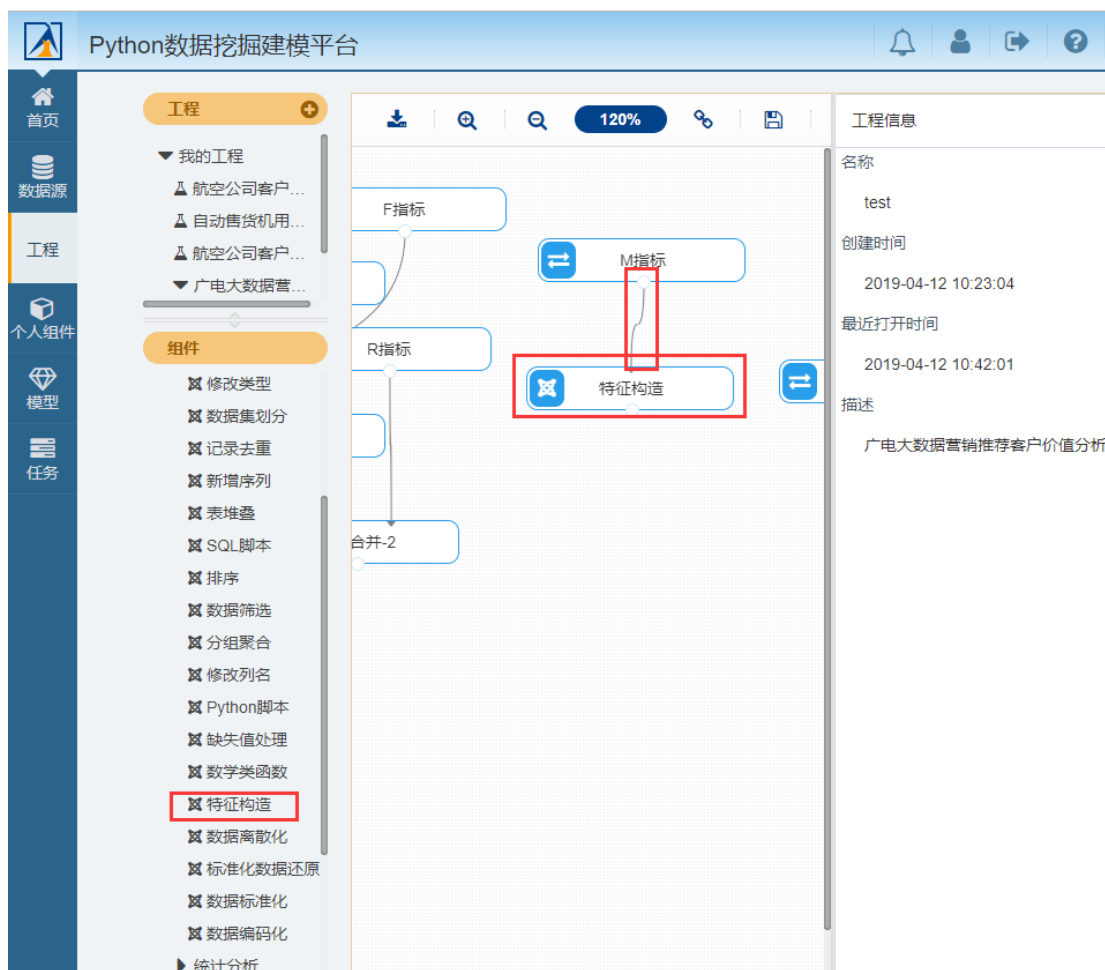


图 3-11 特征构造

在“特征构造”的“字段属性”中，选中所有特征，如图 3-12 所示。

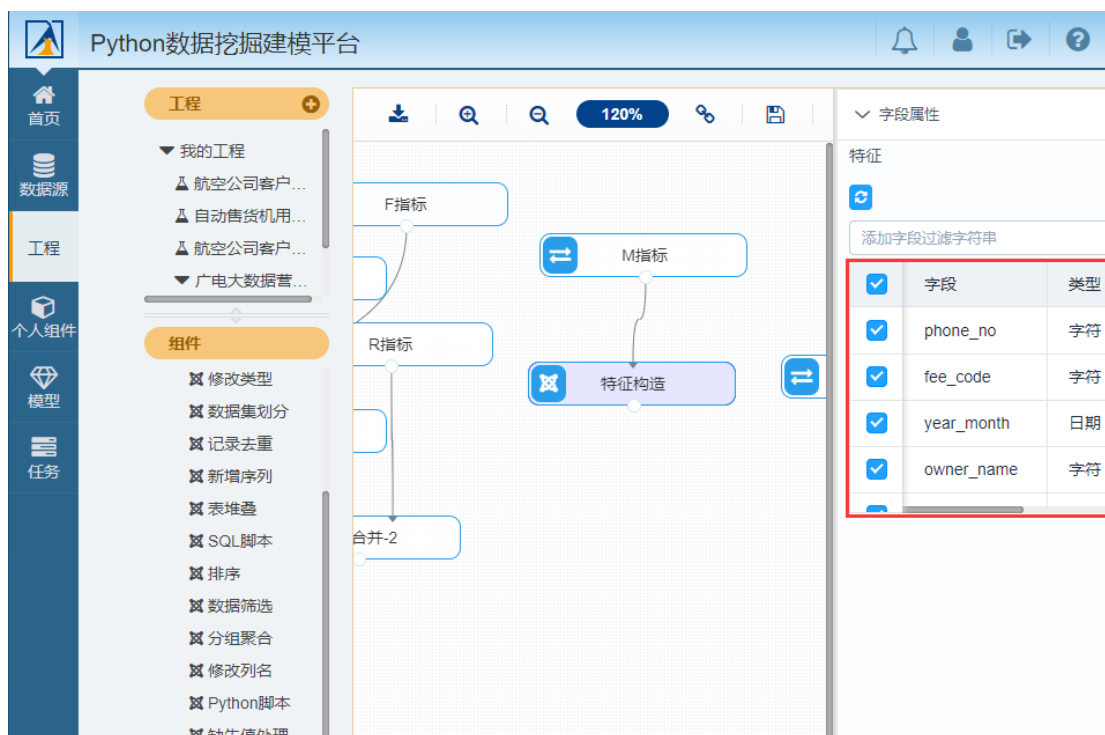


图 3-12 特征选择-字段属性

点击参数设置，填入参数如图 3-12 所示。

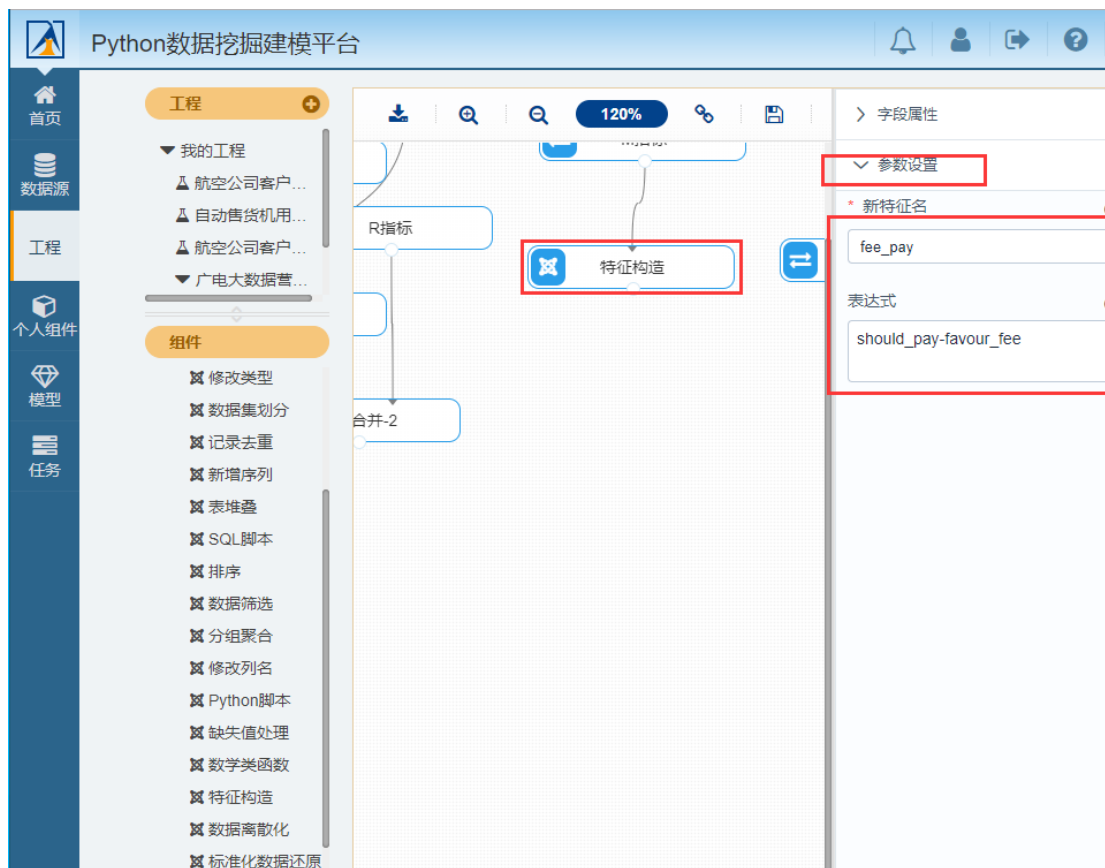


图 3-13 特征构造-参数设置

3.2.4 分组聚合

点击右侧“预处理”，拖动“分组聚合”，并将其与“特征构造”相连，设置“字段属性”，如图 3-14 所示。

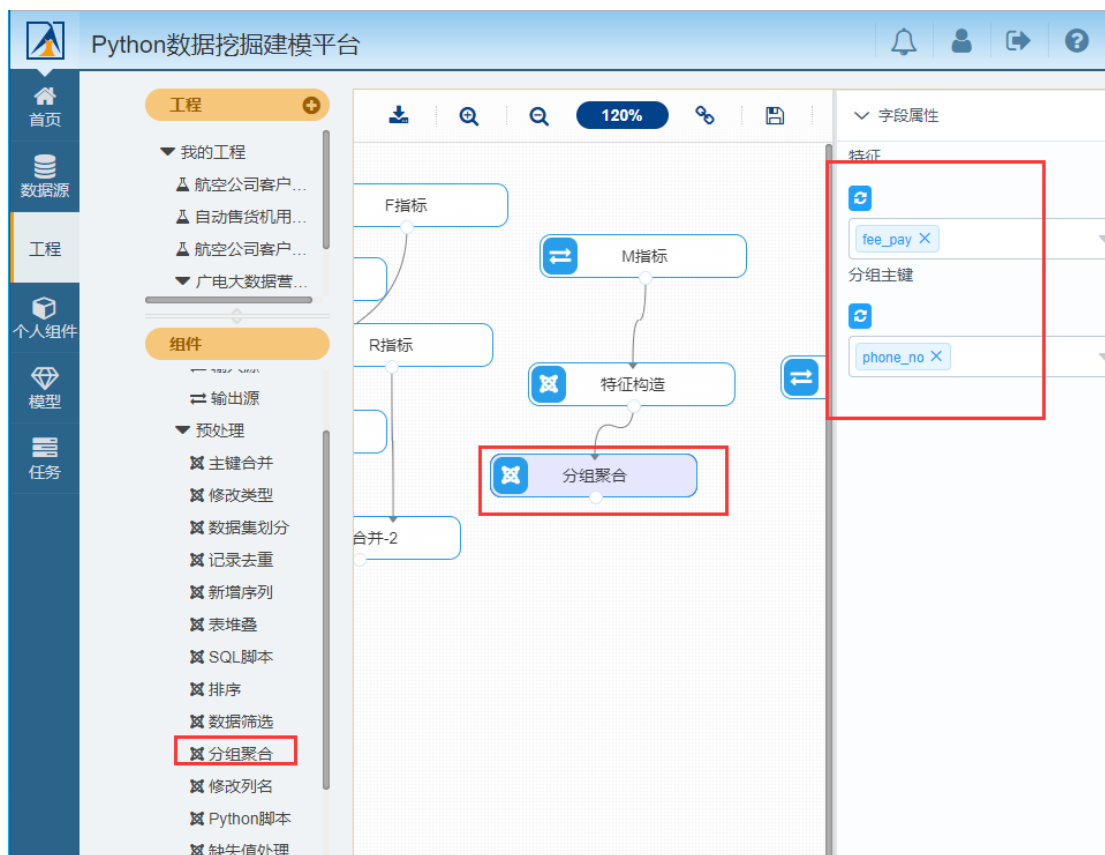


图 3-14 分组聚合-字段属性

在“特征构造”的“参数设置”如图 3-15 所示。

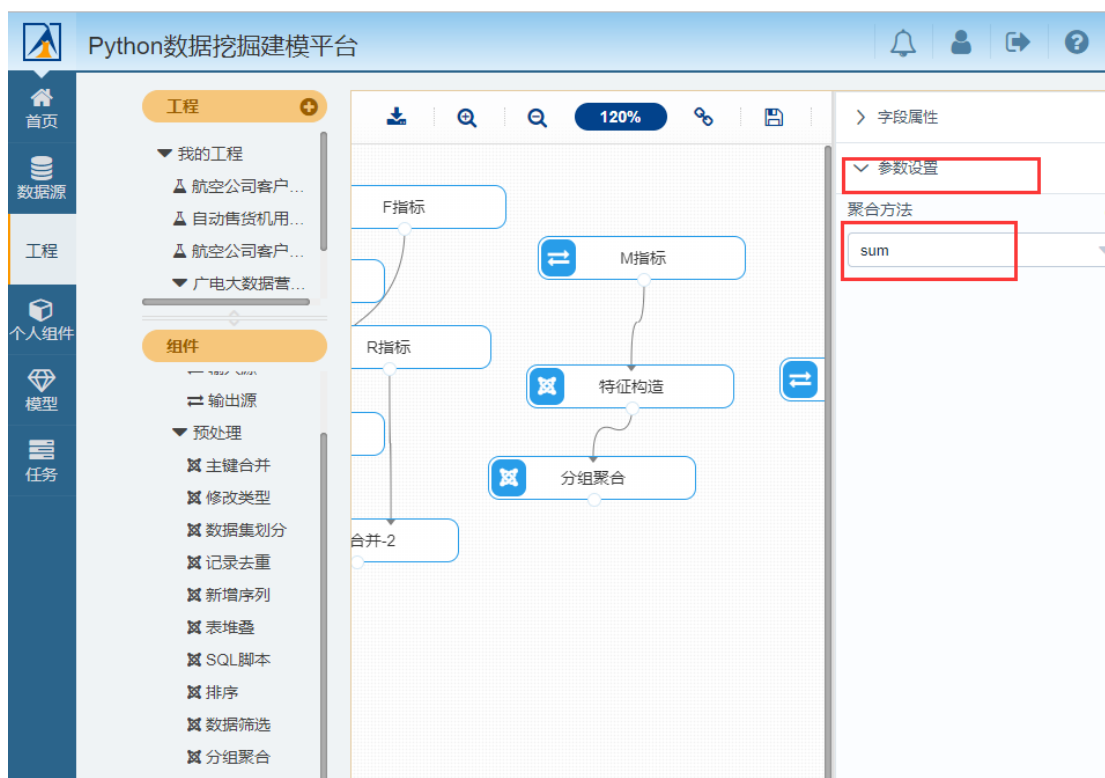


图 3-15 分组聚合-参数设置

3.2.5 Python 脚本

点击右侧“预处理”，拖动“Python 脚本”，并将其与“L 指标”相连，如图 3-14 所示。

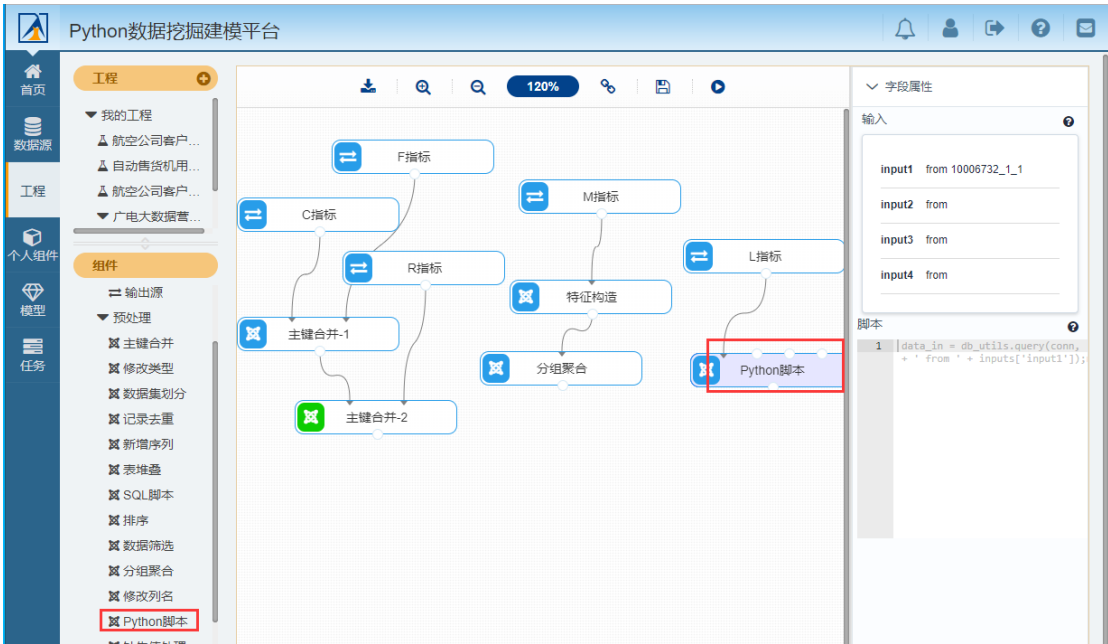


图 3-16 Python 脚本

单机 Python 脚本，在“脚本”中填入代码，如代码 3-1 所示。

代码 3-1 Python 脚本

```
import pandas as pd
data_in = db_utils.query(conn, 'select * from ' + inputs['input1'])
data_in['now'] = pd.datetime.now()
data_in['recenty'] = (data_in['now'] - data_in['run_time']).astype(int)
return(data_in)
```

3.2.6 主键合并

点击运行，待运行完成后，点击右侧“预处理”，拖动“主键合并”，并将其与“分组聚合”和“Python 脚本”相连，重命名为“主键合并-3”，并设置“字段属性”，如图 3-17 所示。



图 3-17 主键合并-3

右键点击运行“主键合并-3”，待运行完成后，拖动“主键合并”，重命名为“主键合并-4”，将其与“主键合并-2”与“主键合并-3”相连，并设置“字段属性”，如图 3-18 所示。

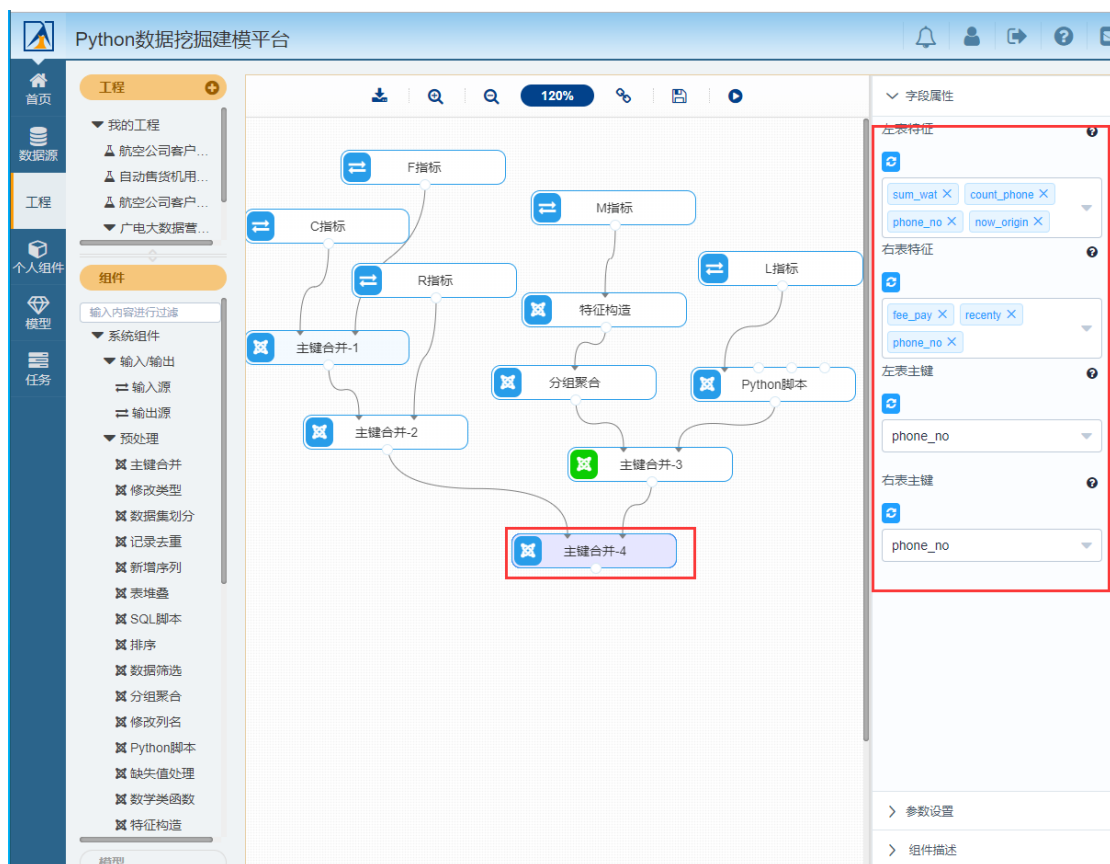


图 3-18 主键合并-4

3.2.7 数据标准化

右击“主键连接-4”选择“运行此节点”，待运行完成后，点击右侧“预处理”，拖动“数据标准化”，“主键合并”相连，重命名为“主键合并-3”，并设置“字段属性”，如图 3-19 所示。

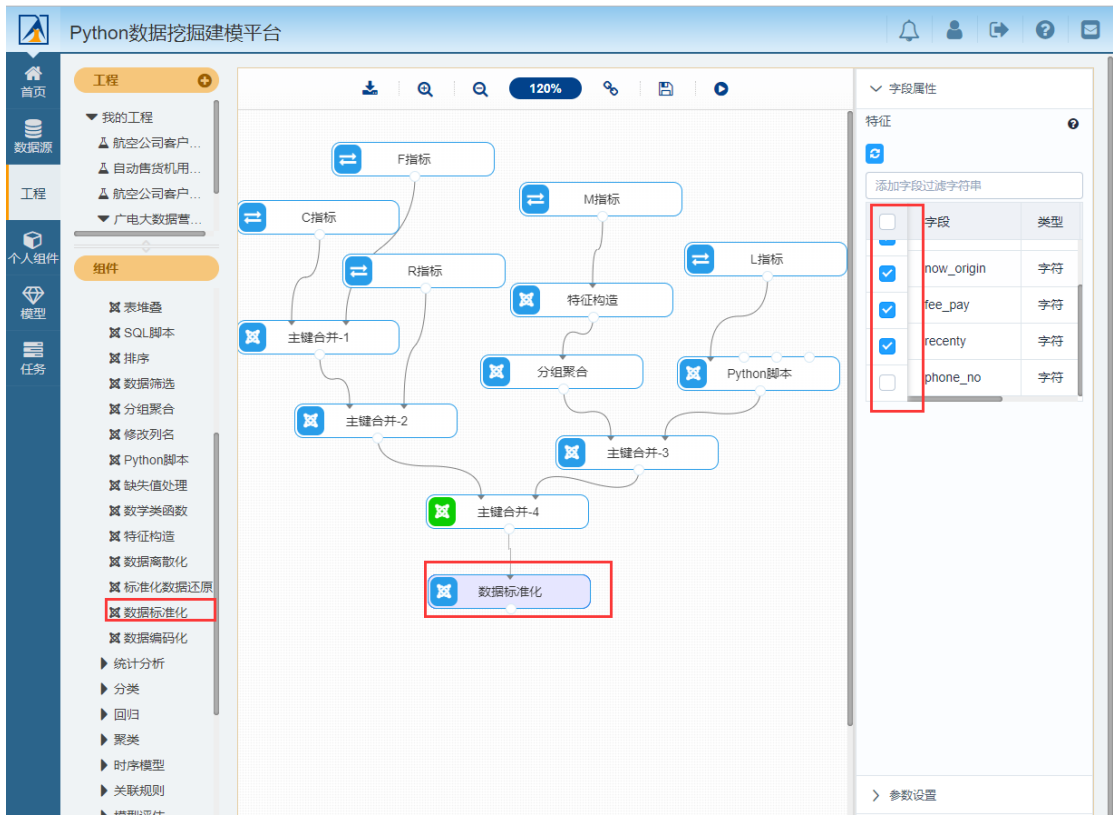


图 3-19 数据标准化-字段属性

点击“参数设置”，设置参数如图 3-20 所示。



图 3-20 数据标准化-参数设置

3.2.8 K-Means 聚类

点击右侧“聚类”，拖动“K-Means”，与“数据标准化”相连，并设置“字段属性”，如图 3-21 所示。

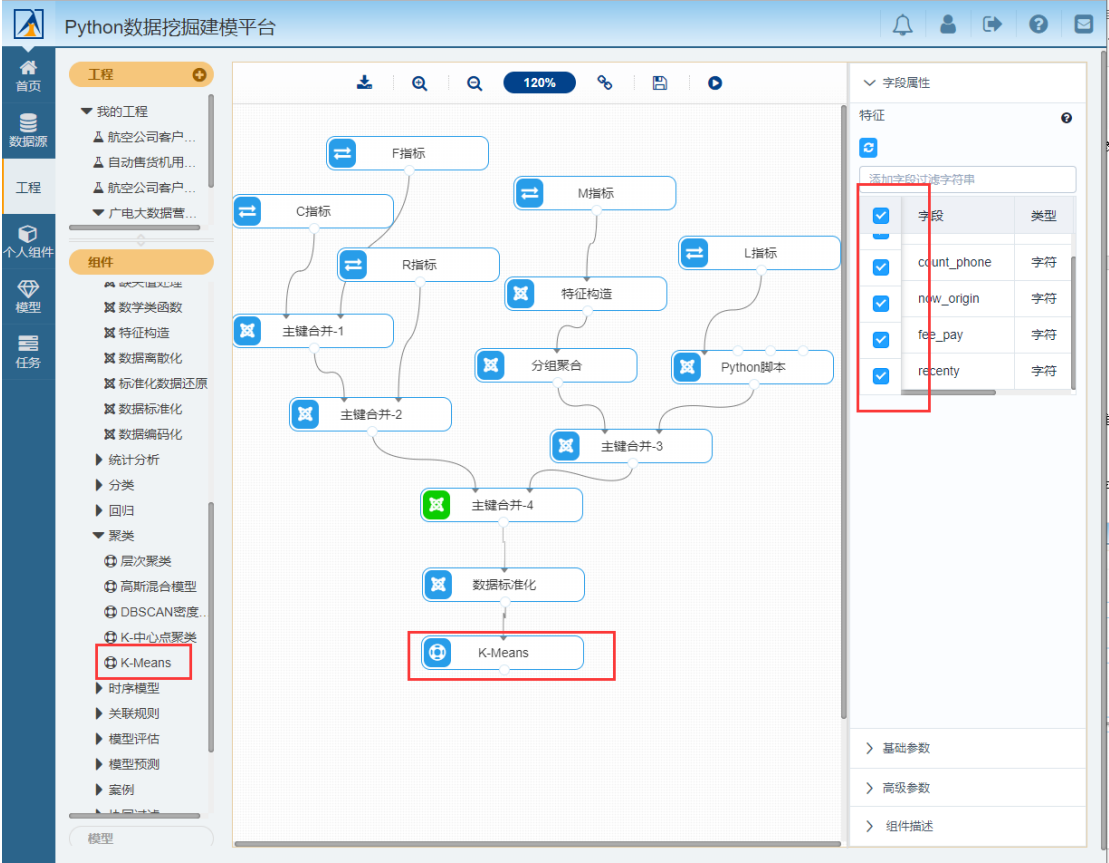


图 3-21 K-Means-字段属性

点击“基础参数”，设置参数如图 3-22 所示。



图 3-22 KMeans--参数设置

点击运行，运行完成后，右键点击“K-Means”，查看运行报告，如图 3-23 所示。



图 3-23 KMeans 运行报告