

图 52

3.4.2 预处理

3.4.2.1 缺失值处理

图标: 

描述: 缺失值处理是数据预处理的一部分, 由于采集的数据存在一些属性值的缺省, 如果不做处理, 将直接影响后续算法的挖掘效果, 严重时甚至得到错误的结果。处理方法有删除缺失值、中位数插补、众数插补、均值插补、线性插值、多项式插值。

字段属性

特征列: 必选。选用中位数插值法、众数插值法、均值插值法时, 请选择数值型数据, 如果勾选了非数值类型数据, 则会自动过滤, 下个组件可能无法获取所有列。勾选的列将传入下一个组件。

参数设置

处理方式: 选择对该缺失数据列的处理方式, 可以选择删除缺失值、中位数插补、众数插补、均值插补、线性插值、多项式插值。

输出

表结果: 缺失值处理结果。

报告: 无。

示例

下面对某数据进行缺失值处理。

- 勾选需要进行缺失值处理的数据, 将会在勾选的数据内查找缺失值, 并进行相应的处理。如图 53、**错误!未找到引用源。** 所示。
- 选择处理方式为【删除法】, 如图 54 所示。
- 运行该组件后可通过查看数据查看结果。

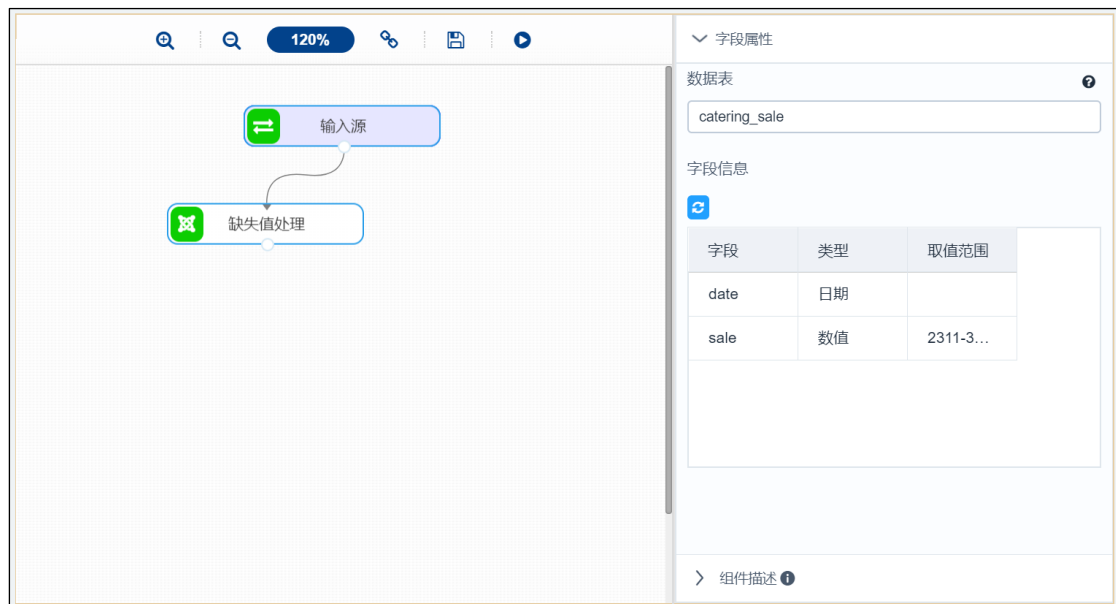


图 53

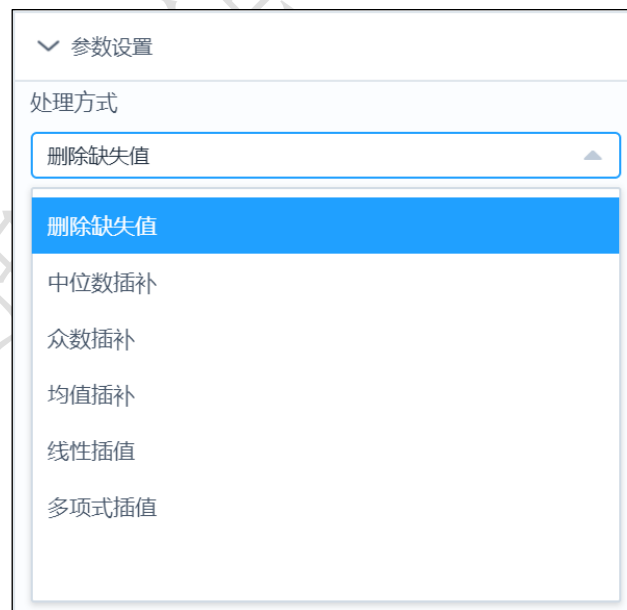


图 54

3.4.2.2 记录选择

图标:  记录选择

描述: 记录选择是对数据表的行进行筛选, 只留下满足条件的数据行。

字段属性