

数据挖掘-Python脚本

- 功能概述
- 操作说明
- 对Python支持边界

随着大数据时代的到来，Python已经在数据处理、数据可视化、机器学习等领域受到广泛应用。本文将介绍如何在smartbi中编写Python脚本，利用其丰富的科学计算扩展库，满足更多数据处理和分析场景，提高生产效率。

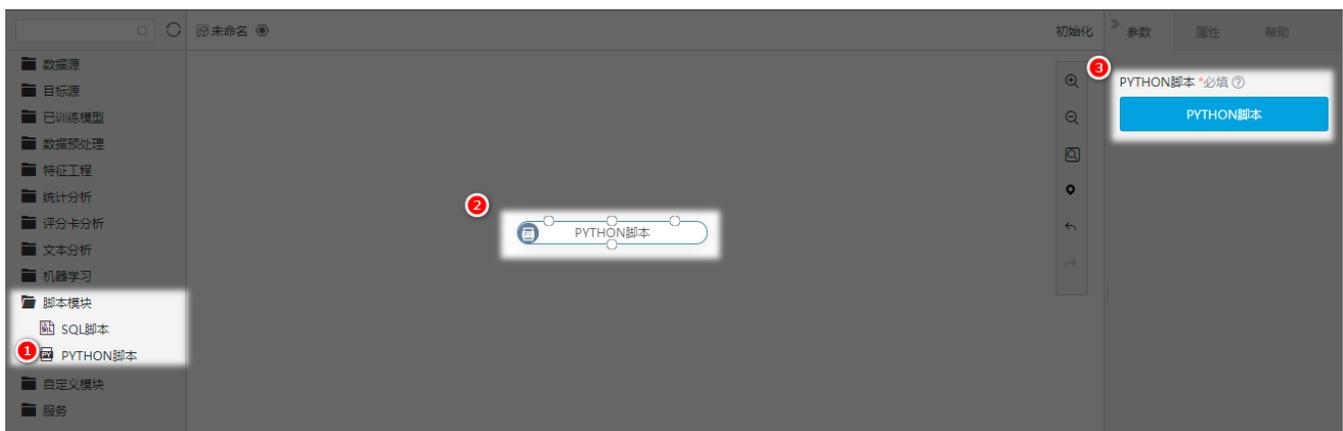
功能概述

Python脚本，该功能支持用户编写Python代码，以便实现更自由、更高效的数据处理、数据分析操作，丰富了数据建模过程。

操作说明

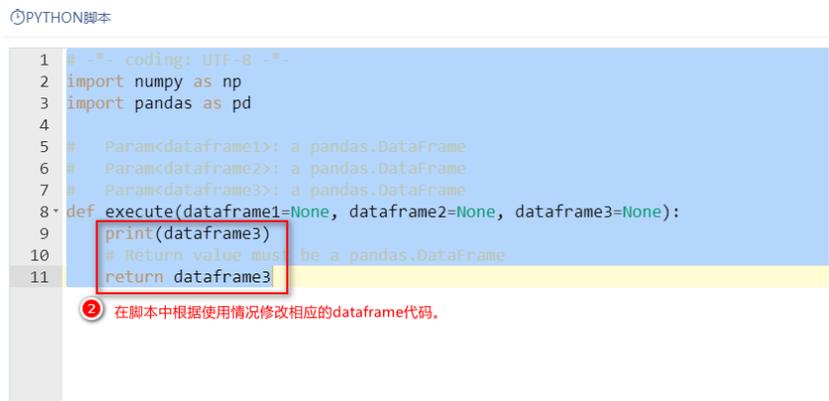
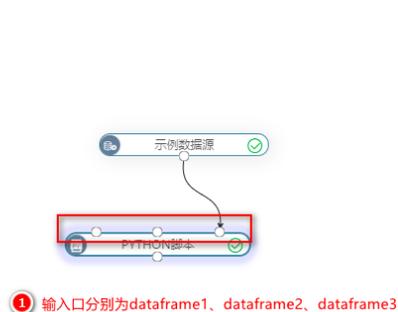
(1) 功能入口

从左侧资源树的 **脚本模块** 拖拽出“Python脚本”节点到画布中。



(2) 数据输入

用户可根据实际情况，对“Python脚本”使用1~3个输入口（dataframe1~3）来获取数据，然后在Python脚本中按需修改dataframe相关代码。



(PS: 用户也可以不使用 **输入**，直接在Python脚本中用代码获取数据。)

(3) 编写脚本

用户编写的代码必须得放在execute函数里面，自定义函数可放在execute函数内外。

```
PYTHON脚本
1 # -*- coding: UTF-8 -*-
2 import numpy as np
3 import pandas as pd
4
5 # Param<dataframe1>: a pandas.DataFrame
6 # Param<dataframe2>: a pandas.DataFrame
7 # Param<dataframe3>: a pandas.DataFrame
8 def execute(dataframe1=None, dataframe2=None, dataframe3=None):
9     print(dataframe3)
10    # Return value must be a pandas.DataFrame
11    return dataframe3
```

(4) 默认Python示例效果

该脚本会直接输出数据，同时在日志打印数据。

[查看日志](#)

Python 3.7.4

| | id | 卡号 | 消费额 | 付款方式 | 性别 | 年龄 | 物品 |
|------|-------|---------|---------|--------|-----|------|------|
| 0 | 505.0 | 50531.0 | 37.4002 | CARD | F | 36.0 | 水果 |
| 1 | 505.0 | 50531.0 | 37.4002 | CARD | F | 36.0 | 鲜肉 |
| 2 | 505.0 | 50531.0 | 37.4002 | CARD | F | 36.0 | 乳制品 |
| 3 | 506.0 | 42653.0 | 14.0502 | CARD | M | 49.0 | 软饮料 |
| 4 | 507.0 | 94685.0 | 31.7561 | CARD | M | 29.0 | 鲜肉 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 2795 | 502.0 | 47598.0 | 42.9427 | CARD | F | 40.0 | 啤酒 |
| 2796 | 502.0 | 47598.0 | 42.9427 | CARD | F | 40.0 | 白酒 |
| 2797 | 502.0 | 47598.0 | 42.9427 | CARD | F | 40.0 | 糖果 |
| 2798 | 504.0 | 20260.0 | 13.1243 | CHEQUE | M | 21.0 | 蔬菜罐头 |
| 2799 | 504.0 | 20260.0 | 13.1243 | CHEQUE | M | 21.0 | 糖果 |

[2800 rows x 7 columns]

对Python支持边界

产品中的Python环境为anaconda 3-2019.10，我们支持的第三方库以anaconda为准，同时支持如下第三方库：

pyarrow、sanic、tensorflow、lightgbm、xgboost、textblob、psycopg2-binary、mysql-connector、hmmlearn、snowlp。