航空公司客户价值分析

1 项目背景

信息时代的来临使得企业营销焦点从产品中心转变为客户中心，客户关系管理成为企业的核心问题。客户关系管理的关键问题是客户分类，通过客户分类，区分无价值客户、高价值客户，企业针对不同价值的客户制定优化的个性化服务方案，采取不同营销策略，将有限营销资源集中于高价值客户，实现企业利润最大化目标。准确的客户分类结果是企业优化营销资源分配的重要依据，客户分类越来越成为客户关系管理中亟待解决的关键问题之一。

面对激烈的市场竞争，各个航空公司都推出了更优惠的营销方式来吸引更多的客户，国内某航空公司面临着常旅客流失、竞争力下降和航空资源未充分利用等经营危机。通过建立合理的客户价值评估模型，对客户进行分群，分析比较不同客户群的客户价值，并制定相应的营销策略，对不同的客户群提供个性化的客户服务是必须的和有效的。

2 项目目标

根据航空公司客户的会员档案信息及其乘坐航班记录数据，建立合理的客户价值评估模型对客户进行分群，为航空公司对不同价值的客户类别提供个性化服务、并制定相应的营销策略提供方向与依据。

3 项目步骤

3.1 工程前期准备

3.1.1 导入数据

（1） 介绍航空公司客户会员档案信息及其乘坐航班记录数据

航空公司客户会员档案信息及其乘坐航班记录数据介绍如图 1 所示。

	A	B	C	D	E	F	G	H	I
1	member_no	ffp_date	load_time	flight_count	sum_yr_1	sum_yr_2	seg_km_sum	last_to_end	avg_discount
2	54993	2006/11/2	2014/3/31	210	239560	234188	580717	1	0.961639043
3	28065	2007/2/19	2014/3/31	140	171483	167434	293678	7	1.25231444
4	55106	2007/2/1	2014/3/31	135	163618	164982	283712	11	1.254675516
5	21189	2008/8/22	2014/3/31	23	116350	125500	281336	97	1.090869565
6	39546	2009/4/10	2014/3/31	152	124560	130702	309928	5	0.970657895
7	56972	2008/2/10	2014/3/31	92	112364	76946	294585	79	0.967692483
8	44924	2006/3/22	2014/3/31	101	120500	114469	287042	1	0.965346535
9	22631	2010/4/9	2014/3/31	73	82440	114971	287230	3	0.962070222
10	32197	2011/6/7	2014/3/31	56	72596	87401	321489	6	0.828478237

图 1 航空公司客户会员档案信息及其乘坐航班记录数据

因为业务数据的安全原因，客户会员档案信息数据集的数据已做了脱敏处理，只保留部分重要属性，其各属性及说明如表 1 所示。

表 1 航空公司客户会员档案信息及其乘坐航班记录数据属性及其说明

属性名称	属性说明
member_no	会员卡号
ffp_date	入会时间
load_time	观测窗口的结束时间
flight_count	观测窗口内的飞行次数
sum_yr_1	第一个观测窗口的票价收入
sum_yr_2	第二个观测窗口的票价收入
seg_km_sum	观测窗口的总飞行公里数
last_to_end	最后一次乘机时间至观测窗口结束时长
avg_discount	平均折扣率

（2） 上传数据到 Python 数据挖掘建模平台

在新增数据源上，选择本地上传数据，如图 2 所示。

图 2 本地上传数据源

在本地路径上选择文件，填写在平台新建的目标表名，如图 3 所示。

图 3 本地选择文件上传

根据文件的数据，可以修改文件的字段名和类型，如图 4 所示。

图 4 字段设置

上传成功，可以在平台的数据源上查看 air_data 的数据，单击数据源操作的查看按钮如图 5 所示，平台上 air_data 航空公司客户会员档案信息及其乘坐航班记录数据预览，如图 6 所示。

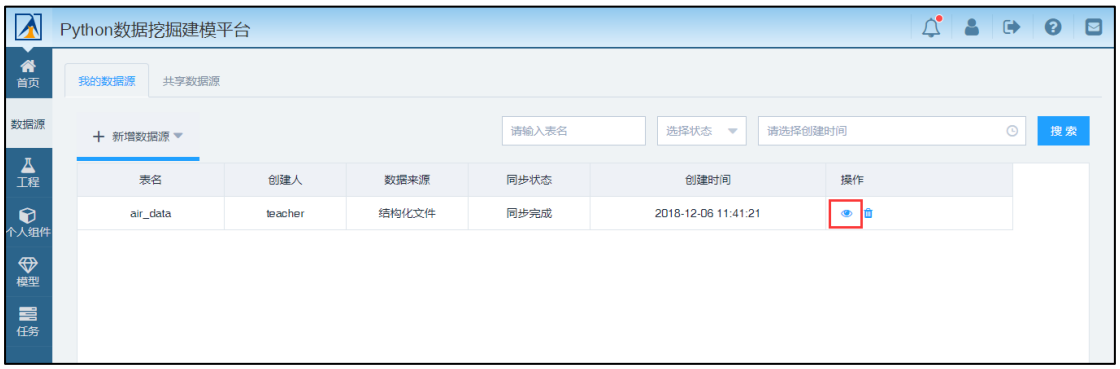

图 5 单击预览数据按钮

图 6 air_data 航空公司客户会员档案信息及其乘坐航班记录数据预览

3.1.2 新建空白工程

右击我的工程，新建一个空白的工程，如图 7 所示。

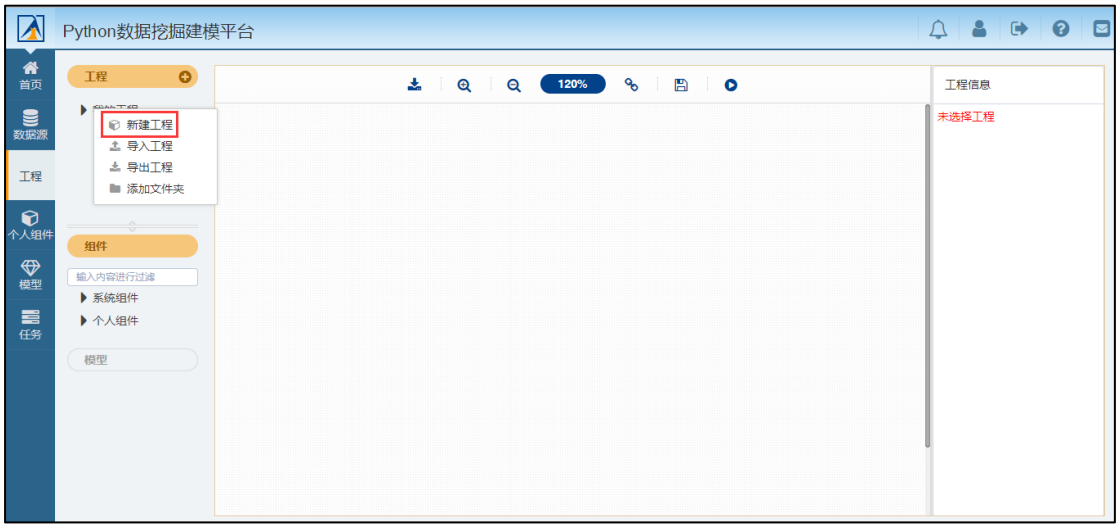

图 7 新建工程

填写工程的信息，包括工程名称和工程描述，如图 8 所示。

图 8 填写工程信息

3.2 数据预处理

读取 air_data 数据，步骤如图 9 所示。

(1) 选择航空公司客户价值分析工程。

(2) 选择输入源组件。

(3) 拖入输入源组件。

(4) 填写数据表名。

(5) 单击更新按钮，更新出航空公司客户会员档案信息及其乘坐航班记录数据。

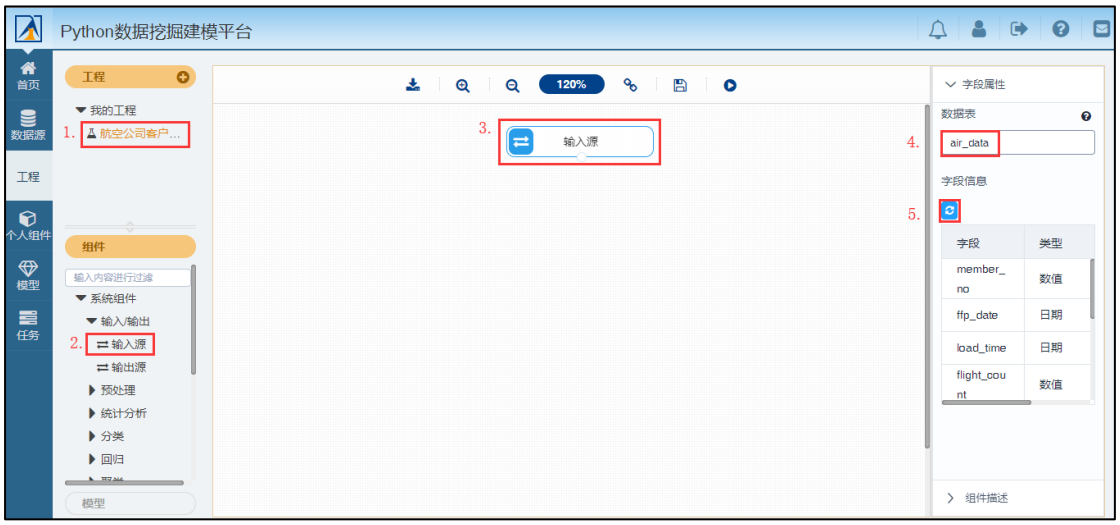

图 9 输入源组件

3.2.1 全表统计

了解数据整体情况，先对数据进行全表统计，分析统计结果。步骤如图 10 所示。

(1) 找到统计分析→全表统计组件。

(2) 拖入全表统计组件，并将数据源和全表统计组件连接。

(3) 单击更新按钮，勾选全部航空公司客户会员档案信息及其乘坐航班记录数据的字段作为输出字段。

(4) 对全表统计组件右键，选择运行该节点。

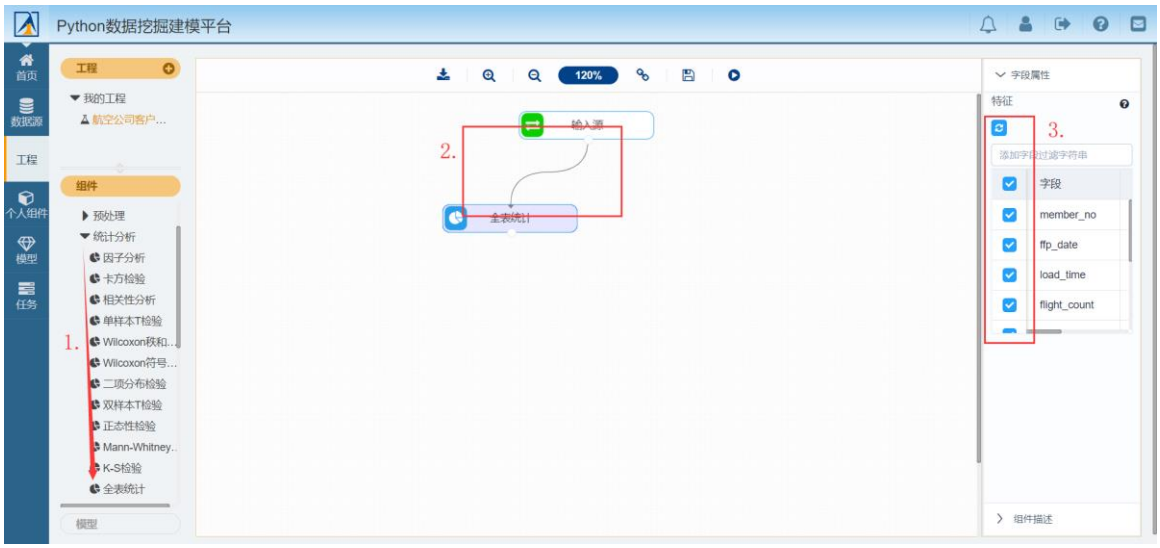

图 10 全表统计组件

(5) 运行完成后，对全表统计组件右键，选择查看数据。查看、分析得到各属性的统计结果：数据量、均值、方差、最大值、最小值等，如图 11 所示。

col	count	mean	std	min	upper_quartile
member_no	62988	31494.5	18183.21	1	15747.75
flight_count	62988	11.84	14.05	2	3
sum_yr_1	62438	5355.29	8109.41	0	1003
sum_yr_2	62850	5604.03	8703.36	0	780
seg_km_sum	62988	17123.88	20960.84	368	4747
last_to_end	62988	176.12	183.82	1	29
avg_discount	62988	0.91	0.29	0	1

图 11 各属性的统计结果

3.2.2 缺失值处理

数据可能存在缺失值，先对数据进行缺失值处理，步骤如图 12 所示。

(1) 找到预处理→缺失值处理组件。

(2) 拖入缺失值处理组件，并将数据源和缺失值处理组件连接。

(3) 单击更新按钮，勾选全部菜品数据的字段作为输出字段。

(4) 对缺失值处理组件右键，选择运行该节点。

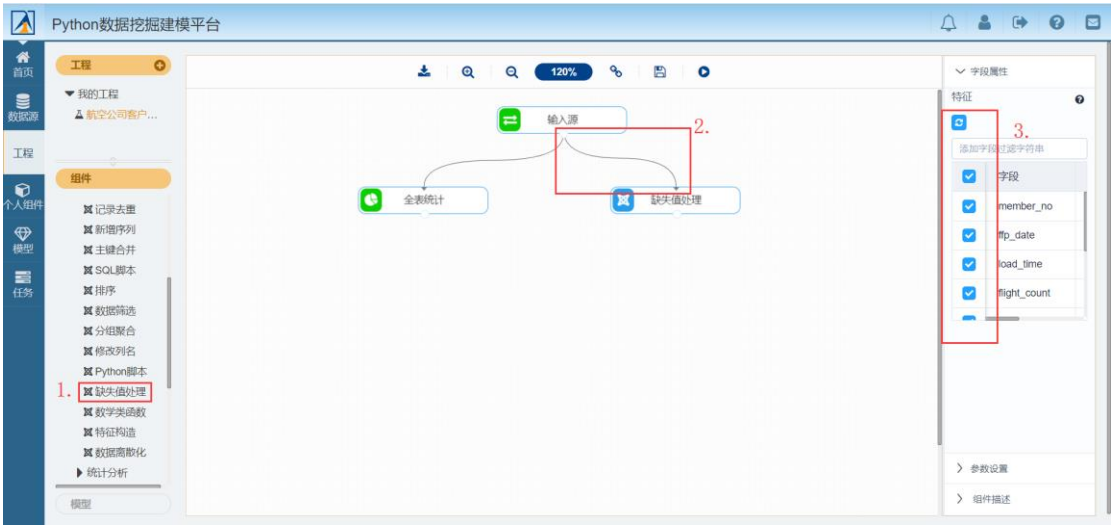

图 12 缺失值处理组件

(5) 运行完成后，对缺失值处理组件右键，选择查看数据，如图 13 所示。

图 13 缺失值处理结果

3.2.3 数据筛选

将航空公司客户会员档案信息及其乘坐航班记录数据进行数据筛选，步骤如图 14、图 15 所示。

(1) 找到预处理→数据筛选组件。

(2) 拖入数据筛选组件，并将数据缺失处理和数据筛选连接。

(3) 选择字段属性，单击更新数据，选择全部字段输出。

(4) 选择参数设置，设置条件为 sum_yr_1>0，sum_yr_2>0，seg_km_sum>0 进行数据筛选。

(5) 对数据筛选组件右键，选择运行该节点。

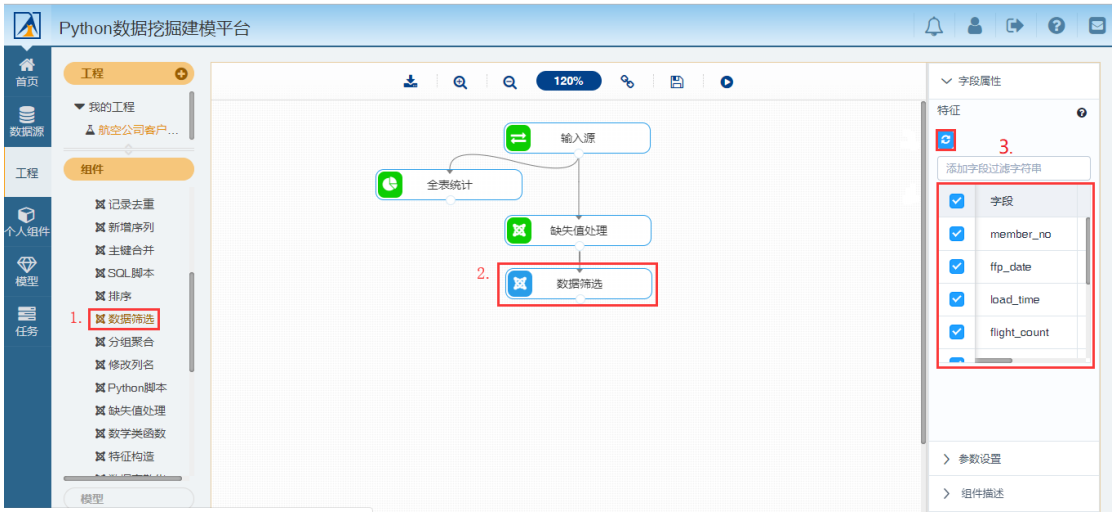

图 14 数据筛选组件_字段属性

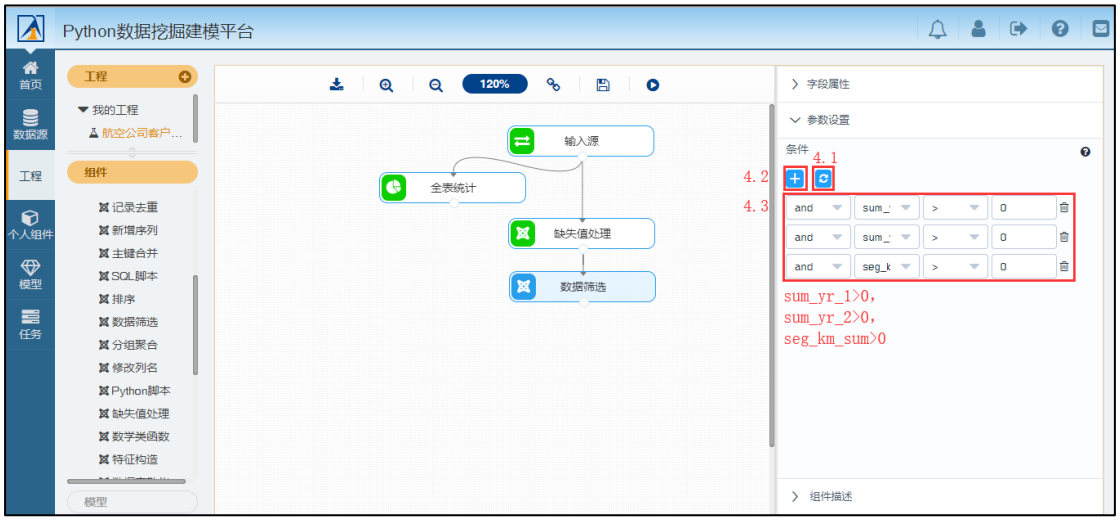

图 15 数据筛选组件_参数设置

(6) 运行完成后，对数据筛选组件右键，选择查看数据，数据筛选的输出表结果，如图 16 所示。

图 16 数据筛选的输出表结果

3.2.4 特征构造

原属性中含有描述相同意义的特征时，这些稀疏特征可通过变换、合并减少数据维度、提高效率，同时还保留原特征的主要信息，这个过程就是特征构造的过程。步骤如图 17、图 18 所示。

(1) 找到预处理→特征构造组件。

(2) 拖入特征构造组件，将缺失值处理和特征构造组件连接。

(3) 选择字段属性，单击更新数据，勾选数据的全部字段。

(4) 选择参数，输入新构造的特征名：long，及其生成的表达式：load_time-ffp_date，即

观测窗口的结束时间-入会时间，新特征名表示了客户在入会时间到观测窗口的时间间隔。

(5) 对特征构造组件右键，选择运行该节点。

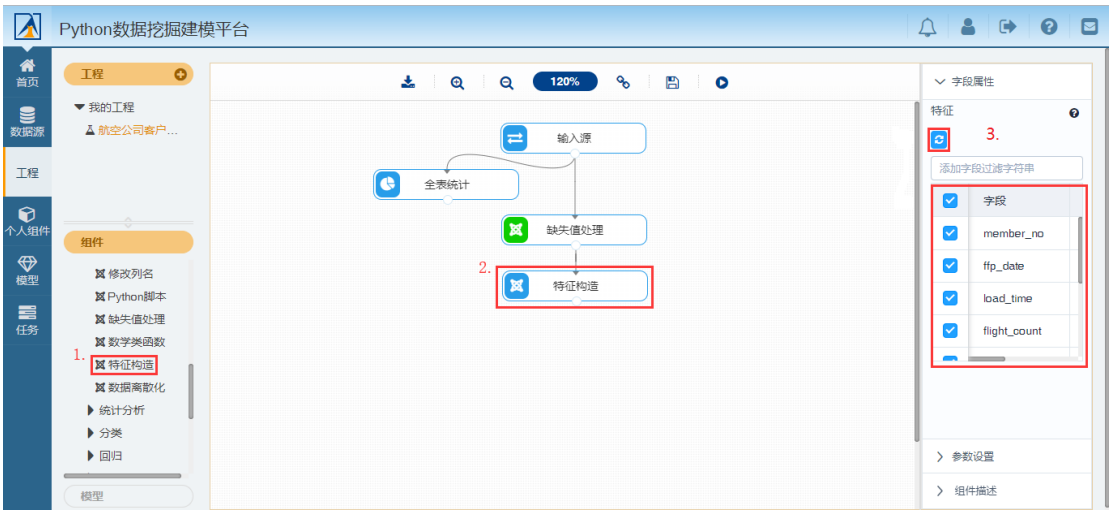

图 17 特征构造组件_字段属性

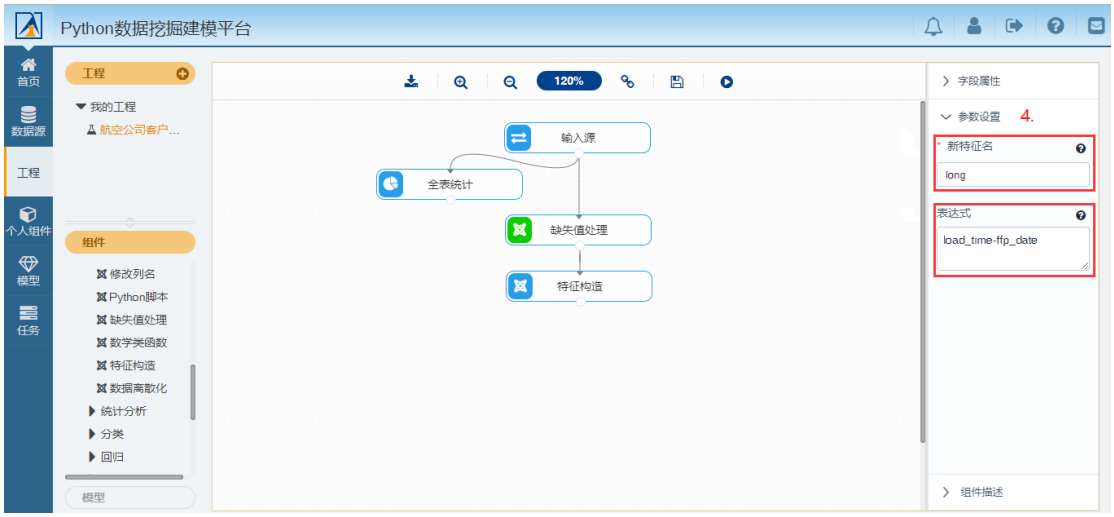

图 18 特征构造组件_参数设置

(6) 运行完成后，对特征构造组件右键，选择查看数据，特征构造的输出表结果如图 19 所示。

图 19 特征构造结果

3.2.5 数据标准化

当属性间的量级相差较大时，如 seg_km_sum 和 avg_ discount，容易造成取值较大的特征决定输出的结果。数据标准化将数据统一映射到特定的区间，消除数据的量纲，步骤如图 20 所示。

(1) 找到预处理→数据标准化组件。

(2) 拖入数据标准化组件，将特征构造和数据标准化组件连接。

(3) 选择字段属性，单击更新数据，勾选 flight_count，seg_km_sum，last_to_end，avg_discount，long 字段。

(4) 对数据标准化组件右键，选择运行该节点。

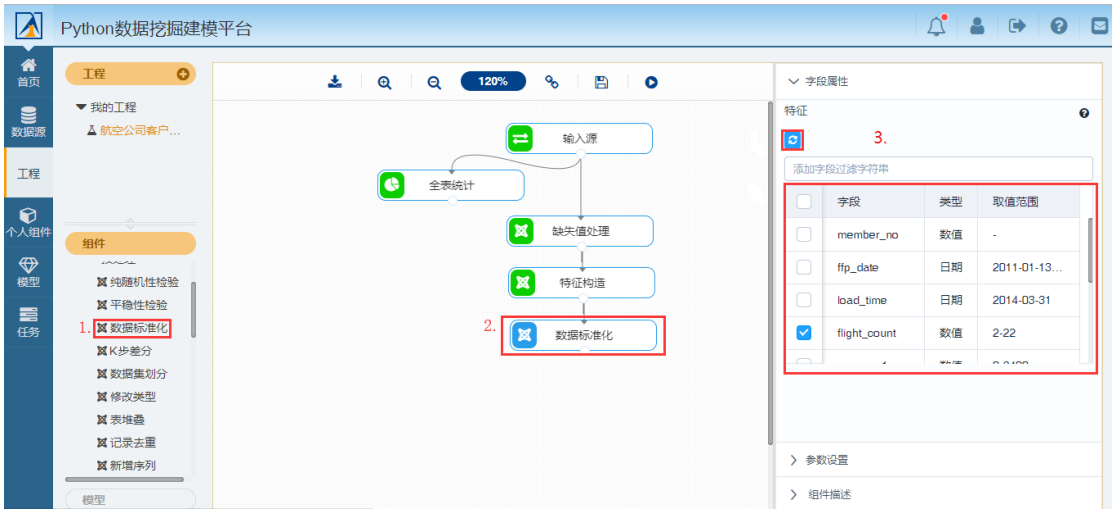

图 20 数据标准化组件

(7) 运行完成后，对数据标准化组件右键，选择查看数据，数据标准化的输出表结果如图 21 所示。

图 21 数据标准化结果

3.2.6 修改类型

新构造的特征 long 为非数值型数据，无法直接进行模型聚类，需要修改数据类型。步骤如所示。

(1) 找到预处理→修改类型组件。

(2) 拖入修改类型组件，将数据标准化和修改类型组件连接。

(3) 选择字段属性，单击更新数据，勾选数据的全部字段。

(4) 设置修改规则，选择 long 字段，设置新类型为数值，同时根据需求，设置参数保留小数点位数。

(5) 对修改类型组件右键，选择运行该节点。

图 22 修改类型组件

(6) 运行完成后，对修改类型组件右键，选择查看数据，修改类型的输出表结果如图 23

所示。

图 23 修改类型结果

3.3 模型构建

3.3.1 K-Means 聚类算法

选择 K-Means 聚类算法模型，步骤如如图 24、图 25 所示。

(1) 找到聚类→K-Means 组件。

(2) 拖入 K-Means 组件，将修改类型和 K-Means 组件连接。

(3) 选择字段属性，单击更新数据，勾选数据的全部字段。

(4) 选择参数设置，设置聚类数（n_clusters）的值为 5，其他的参数都设置为默认值。

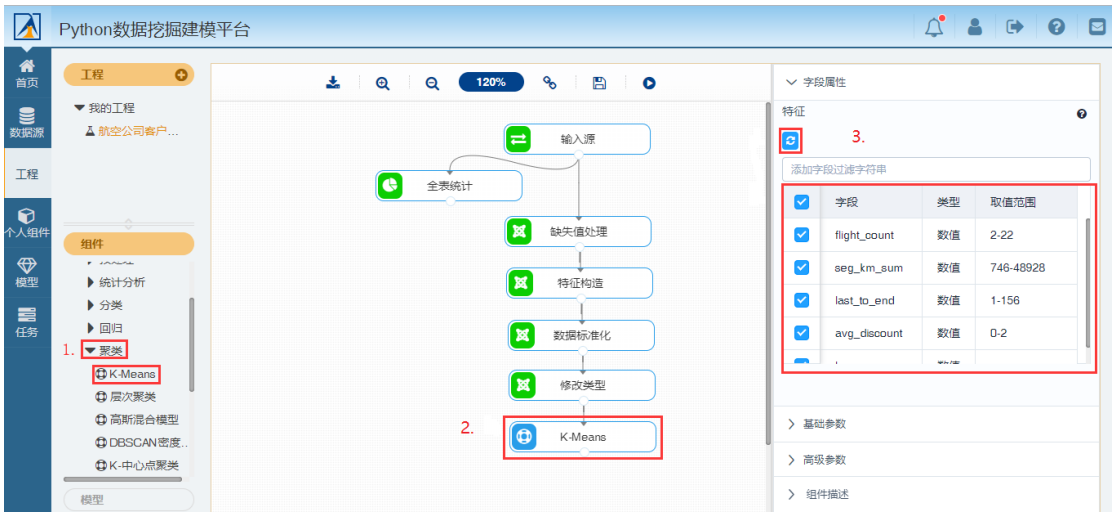

图 24 M-Means 聚类组件_字段属性

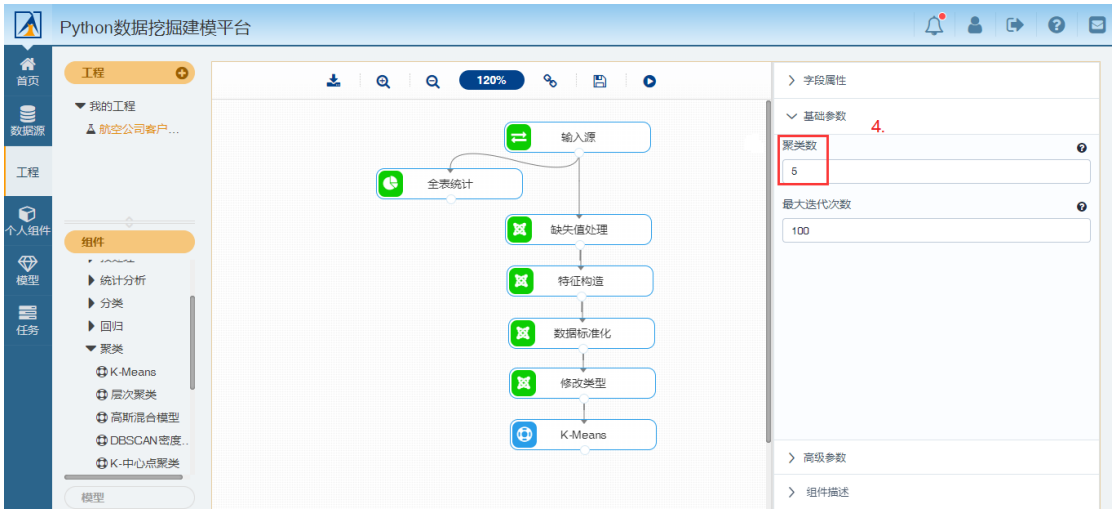

图 25 K-Means 组件_参数设置

(5) 运行完成后，对 K-Means 组件右键，选择查看数据，K-Means 的输出表结果如图 26 所示。选择查看报告，K-Means 的报告如图 27 所示。

图 26 K-Means 聚类算法的结果

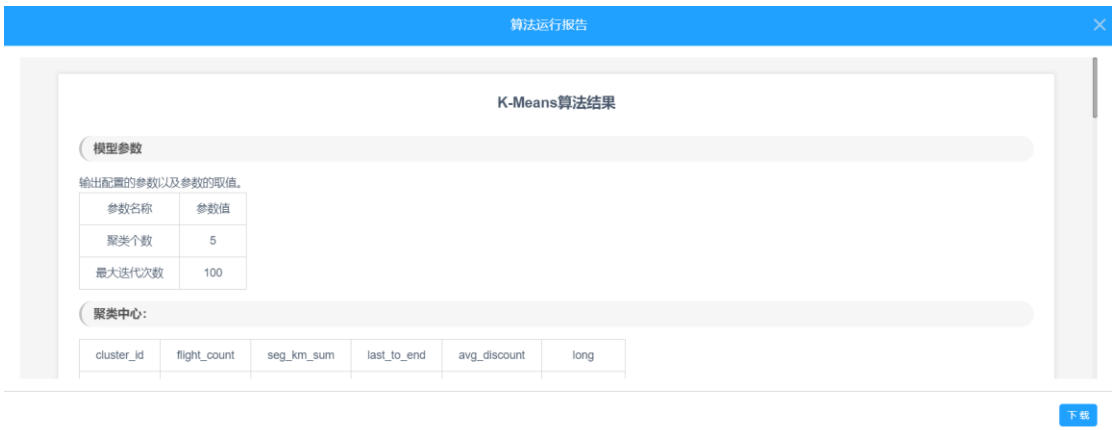

图 27 K-Means 聚类算法的报告