# 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	В	С	D	E	F	G	Н	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。

	P	/thon数据挖	掘建模平台	ì						¢			0	
◆ 首页		我的数据源	共享数据源											
数据源		十 新增数据》	₹ ▲			请输入表名	选择状态 🔻	请选择创建时	间		(	Э	搜索	
「工程		数据来源于5	之件	创建人	数据来源	同步状态	创建时间		操作					
⑦ 个人组件		■ 数据来源于数	<b>牧居库</b>											
₩														
日子						暂无数据								

#### 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。

	<b>第</b>	建数据源		×
1 文件属性		2 预览数据		3 字段设置
1. 上传文件	air_data.csv等待上传… 删除			
2. * 新建目标表名	air_data			
列分隔符	逗룩 (,) 🔻	文件编码	UTF-8	
存储有效期 (天)	180 — +	预览设置	分页显示	
				重置下一步

#### 图 3 本地选择文件上传

				新	主数据	源					×
1 文件属性				1	- 2 <b>页</b> 览数:	据					3 字段设置
注意: 【字段名】另 原字段	<b>只能是以字母开头,由小国</b> 字段名	弓英文字母、数字、下划 类型	线组成	¥	度		1	精度		备注	
member_no	member_no	数值	-	255	-	+	0		+		
ffp_date	ffp_date	日期	-	255	-	+	0		+		
load_time	load_time	日期	-	255	-	+	0		+		
flight_count	flight_count	数值	-	255	-	+	0		+		
sum_yr_1	sum_yr_1	数值	-	255	-	+	0		+		
sum_yr_2	sum_yr_2	数值	-	255	-	+	0		+		
											上一步 确定

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。

	Python数据	挖掘建模型	平台					Ĺ <b>₽</b>		0	
◆ 首页	我的数据源	共享数据源									
数据源	十 新增数	据源 ▼			请输入表名	选择状态 🔻	请选择创建时间		O	搜索	
「五程	表	络	创建人	数据来源	同步状态	创建时间	操作				
⑦ 个人组件	air_	data	teacher	结构化文件	同步完成	2018-12-06 11:41:2	11 💿 🛍				
₩											
日子											

#### 图 5 单击预览数据按钮

	预览数据(分页加载) ×												
member_no	ffp_date	load_time	flight_count	sum_yr_1	sum_yr_2	seg_km_sum	last_to_end						
54993	2006-11-02	2014-03-31	210	239560	234188	580717	1						
28065	2007-02-19	2014-03-31	140	171483	167434	293678	7						
55106	2007-02-01	2014-03-31	135	163618	164982	283712	11						
21189	2008-08-22	2014-03-31	23	116350	125500	281336	97						
39546	2009-04-10	2014-03-31	152	124560	130702	309928	5						
56972	2008-02-10	2014-03-31	92	112364	76946	294585	79						
44924	2006-03-22	2014-03-31	101	120500	114469	287042	1						
22631	2010-04-09	201/4-03-31	73	82440	11/1971	287230	3						
		共 62988 条 100 条/页	হ 👻 < 1 2	3 4 5 6 63	30 〉 前往 1 页								

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。

	Python数据挖掘建构	等于台	4 8 9 0
▼ 谷 首页	Ⅰ程	📩 🔍 Q 💶 120% % 🖹 🖸	工程信息
● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	<ul> <li>● 新建工程</li> <li>▲ 导入工程</li> </ul>		未选择工程
工程	よ 导出工程 ■ 添加文件夹		
⑦ 个人组件	组件		
₩	输入内容进行过滤 ▶ 系统组件		
任务	▶ 个人组件		
	模型		

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。

	创建工程	×
* 工程名称	航空公司客户价值分析	
工程描述	航空市场竞争激烈,某航空公司面临着常旅客流失、竞争力下降、航空资源未充分利用等经营危机。通过积累的大量的会员档案信息和其乘坐航班记录,建立合理的客户	
工程位置	▼ 我的工程	
	重置确定	

#### 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。

	Python数据挖掘建模平台			A   8   G	0
▲ 首页		🛓 🔍 🔍 120%) % 🖺 🗅 🔿		> 字段属性	
<b>劉</b> 据源	▼ 我的工程 1. ▲ 航空公司客户	3. (三 输入源	4.	数据表 air_data	0
工程				字段信息	
			5.	2	
个人组件	组件			字段	类型
<b>₩</b> 模型	输入内容进行过滤 ▼ 系统组件			member_ no	数值
	▼ 输入/输出			ffp_date	日期
任务	2. 二输入源			load_time	日期
	<ul> <li>ご 輸出源</li> <li>● 预处理</li> </ul>			flight_cou nt	数值
	▶ 统计分析				
	▶ 分类				
	模型			> 组件描述	

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。

	Python 数据挖掘建模半台		4
合 首页	I程 O	🛓 Q Q 120% % 🖹 D	◇ 字段属性
2	▼ 我的工程		特征 😡
据源	▲ 航空公司客户		□ 3.
但		2.	添加字段过滤字符串
-108	tildt.		▼ 字段
e ant	A STATE		
			Member_no
<b>₽</b>	● ATLI 力好 ● 因子分析		Mp_date
	© 卡方检验		Ioad_time
昌	€ 相关性分析		flight_count
	€ 单样本T检验		
	1. ● Wilcoxon秩和		
	Wilcoxon付号		
	▶ 2903 TUTA22		
	▶ 正态性检验		
	Mann-Whitney.		
	▶ K-S检验		1
	€ 全表统计		
	横型		> 组件描述

#### 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。

	预览数据										
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。

			预览数据			×
member_no	ffp_date	load_time	flight_count	sum_yr_1	sum_yr_2	seg_km_sum
54993	2006-11-02	2014-03-31	210	239560	234188	580717
28065	2007-02-19	2014-03-31	140	171483	167434	293678
55106	2007-02-01	2014-03-31	135	163618	164982	283712
21189	2008-08-22	2014-03-31	23	116350	125500	281336
39546	2009-04-10	2014-03-31	152	124560	130702	309928
56972	2008-02-10	2014-03-31	92	112364	76946	294585
44924	2006-03-22	2014-03-31	101	120500	114469	287042
22631	2010-04-09	2014-03-31	73	82440	114971	287230
共62300条 25条页 💌 < 1 2 3 4 5 6 … 2492 > 前往 1 页						

#### 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。

	Python数据挖掘建模	平台		<b>G O E</b>
▲	工程	📩 IQIQ 120%) % I 🖺 I O	>> 字段	定属性
数据源	<ul> <li>▼我的工程</li> <li>▲航空公司客户</li> <li>毎件</li> </ul>	金入原	特征 こ 添加字	② 3. ◎ ◎ □ ○
上程 个人组件	¥ 记录去重 ¥ 新增序列 ¥ 主键合并	◆ 全炭统计 ★ 全炭统计 ★ 由失值处理		学段 member_no ffo_date
♥型	X SOL脚本 X 排序 1. X 数据筛选 X 分组聚合	2. 数据标选		load_time flight_count
	XX 修改列名 XX Python脚本 XX 缺失值处理 XX 数学关函数			
	<b>X</b> 特征构造		<ul> <li>&gt; 参数</li> <li>&gt; 组件</li> </ul>	2.22 2.12 2.12 2.12 2.12 2.12 2.12 2.12

图 14 数据筛选组件\_字段属性

	Python数据挖掘建模	平台	A 🛔 🗭 🛛 🖾
谷首页	工程 〇 ▼我的工程	🛃 Q Q 120% % 🖺 🔘	> 字段屬性
数据源	▲ 航空公司客户 分割		✓ 参数设置 条件 4.1
	<b>國</b> 记录去重 國 新增序列	全表统计         4.2           【】 缺失值处理         4.3	and         sum_:         >         0         0           and         sum_:         >         0         0
♥	X 主键合并 X SOL即本 X 排序	又 数据筛选	and v seg_k v > v 0 f
任务	■ 数据筛选 ■ 分组聚合		sum_yr_2>0, seg_km_sum>0
	X Python脚本 X 缺失值处理		
	<b>X</b> 数学关函数 X 特征构造		
	模型		> 组件描述

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。

预选数据						
member_no	ffp_date	load_time	flight_count	sum_yr_1	sum_yr_2	
54993	2006-11-02	2014-03-31	210	239560	234188	
28065	2007-02-19	2014-03-31	140	171483	167434	
55106	2007-02-01	2014-03-31	135	163618	164982	
21189	2008-08-22	2014-03-31	23	116350	125500	
39546	2009-04-10	2014-03-31	152	124560	130702	
56972	2008-02-10	2014-03-31	92	112364	76946	
44924	2006-03-22	2014-03-31	101	120500	114469	
22631	2010-04-09	2014-03-31	73	82440	114971	
共41516条 25 船页 💌 🤇 1 2 3 4 5 6 🚥 1661 〉 前往 1 页						



## 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。

	Python数据挖掘建模平台	Ĺ	] & 🗭 🛛 🗖
▲ 前 )))))))))))))))))))))))))))))))))))	12     ●     ▲     ●     ●       ●     秋台工程     ▲     ●     ●       ▲     ●     ●     ●     ●       ●     ★約1程     ●     ●     ●       ▲     ●     ●     ●     ●       ▲     ●     ●     ●     ●       ▲     ●     ●     ●     ●       ●     ★約2所     ●     ●     ●       ■     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●     ●     ●       ●     ●     ●	120%	<ul> <li>&gt; 字段電性</li> <li>特征</li> <li>③</li> <li>③</li> <li>ぶ加字段过述字符串</li> <li>②</li> <li>学段</li> <li>○</li> <li>●</li> <li>○</li> <li>●</li> <li>○</li> <li>○</li></ul>
	▶ 病计分析 ▶ 分类 ▶ 回归		> 参数设置 > 组供描述

图 17 特征构造组件\_字段属性

	Python数据挖掘建模型	平台	4 2 6 0 2
▼ 合 首页	I程 O	초 🛛 Q 📄 Q 💶 120% 🖉 🖓 🖄 🖸	> 字段属性
● 数据源	▼ 我的工程 ▲ 航空公司客户	🔁 輸入源	✓ 参数设置 4.
工程			
<b>行</b> 个人组件	金田住	X 缺失值处理	表达式 🕑 load_time-ffp_date
₩	XX 修改列名	<b>文</b> 特征[145]道	
日子	<b>X</b> 缺失值处理 X 数学关函数		
	<b>X</b> 特征构造 X 数据离散化		
	<ul> <li>▶ 统计分析</li> <li>▶ 分类</li> </ul>		
			> 组件描述

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。

	预览数据							
count	sum_yr_1	sum_yr_2	seg_km_sum	last_to_end	avg_discount	long		
	239560	234188	580717	1	1	2706		
	171483	167434	293678	7	1	2597		
	163618	164982	283712	11	1	2615		
	116350	125500	281336	97	1	2047		
	124560	130702	309928	5	1	1816		
	112364	76946	294585	79	1	2241		
	120500	114469	287042	1	1	2931		
	82440	114971	287230	3	1	1452		
	共11516条 25条/页 💌 🤇 1 2 3 4 5 6 … 1681 〉 前往 1 页							

#### 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。

	Python数据挖掘建模型	平台				B 🕞 🛛 🖾
▲	Ⅰ程 ●	🛃 Q Q 120% % 🖺 O	~ 字段	调性		
9	▼ 我的工程		特征			0
数据源	▲ 航空公司客户	😑 輸入源	2	3.		
工程			添加字	段过滤字符串		
				字段	类型	取值范围
个人组件		₩ 缺失值处理		member_no	数值	•
$\bigotimes$				ffp_date	日期	2011-01-13
模型	X 纯随机性检验 X 平稳性检验	▼ 特征构造		load_time	日期	2014-03-31
日子	1. 🛛 数据标准化	2. 数据标准化		flight_count	数值	2-22
	XK步差分				44 /m	0.0400
	■ 数据集划分					
	🗙 修改类型					
	⋈ 表堆叠					
	🗙 记录去重		U			
	🗙 新增序列		> 参数	设置		
	模型		> 组(	+描述		

#### 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。

		预览数据			
flight_count	seg_km_sum	last_to_end	avg_discount	long	
14.055230506971206	26.799973857809867	-0.9476829635970925	0.30159700783351123	1.4374009532386682	
9.087691175964295	13.147025806592625	-0.9145840819808607	0.30159700783351123	1.3088293497199535	
8.73286693803523	12.67299516867084	-0.8925181609033728	0.30159700783351123	1.3300613576404752	
0.7848040084241713	12.55998124174909	-0.4181008577373831	0.30159700783351123	0.6600735521484566	
9.93926934699405	13.919953547534726	-0.9256170425196046	0.30159700783351123	0.38759611716842796	
5.68137849184527	13.190167065733208	-0.5173975025860786	0.30159700783351123	0.8889074152918574	
6.320062120117587	12.831385899583898	-0.9476829635970925	0.30159700783351123	1.7028010522451897	
4.333046387714822	12.840328078986799	-0.9366500030583486	0.30159700783351123	-0.04176226522434451	
共 62300 祭 25 祭(页 💗 < 1 2 3 4 5 6 2492 > 前往 1 页					

#### 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。

	Python数据挖掘建模	平台					
▼ 谷 首页	工程 〇	土 Q Q 120% % 🖺 D	→ 字段	3属性			
● 数据源	▼ 我的工程 ▲ 航空公司客户	🔁 輸入源	数据	3.			0
工程		<b>全</b> 表统计	添加字	段过滤字符串	M/ #11		
⑦ ↑人组件	⇒	其 缺失值处理		- <del>7</del> ⊧x flight_count	<u>奥型</u> 数值	4以1直;已围 2-22	
₩	X 纯随机性检验	<b>区</b> 特征构造		seg_km_sum	数值	746-48928	
III 任务	X 平稳性检验 X 数据标准化 X 长步差分	又 数据标准化		avg_discount	数值	0-2	
	X 数据集划分 1. X 修改类型	2. 修改典型	修改规则				0
	翼 表堆叠 翼 记录去重		字段名 flight_c	类型 count numeric	新类型数值	参数 ▼ 6	
	<b>风</b> 新增序列 模型	4.	> 组作	井描述		<u> </u>	

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。

		预览数据		
flight_count	seg_km_sum	last_to_end	avg_discount	long
12.364561	23.665217	-1.02501	0.249814	1.358337
7.909647	11.496058	-0.963704	0.249814	1.231314
7.591438	11.073544	-0.922833	0.249814	1.252291
0.463576	10.972812	-0.04411	0.249814	0.590378
8.673346	12.184984	-0.984139	0.249814	0.321184
4.854848	11.53451	-0.228029	0.249814	0.816453
5.427623	11.214721	-1.02501	0.249814	1.620538
3.645657	11.222691	-1.004574	0.249814	-0.103
	共 41516 条 25 条/页 🔍	< 1 2 3 4 5 6	6 *** 1661 〉 前往 1 页	

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

	Python数据挖掘建模	平台	4 8 9
▲	工程 〇	🕹   Q   Q 🚺 120% 💊   🖺   O 🔷	字段属性
<b>劉</b> 数据源	▼ 我的工程 ▲ 航空公司客户	<b>Ξ</b> \$\\$\\$	基础参数 4.
工程			<u>数</u>
⑦ 个人组件			送代次数 🕑
₩	组件 → → → → → → → → → → → → → → → → → → →		
日子	▶ 分类 ▶ 回归		
	▼ 聚类 ① K-Means		
	<ul> <li>□ 层次聚类</li> <li>□ 高斯混合模型</li> </ul>		
	✿ DBSCAN密度 ✿ K-中心点聚类	C K-Means	高级参数
	模型	>	组件描述

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。

petable								
	flight_count	seg_km_sum	last_to_end	avg_discount	long	cluster_id		
	12.364561	23.665217	-1.02501	0.249814	1.358337	3		
	7.909647	11.496058	-0.963704	0.249814	1.231314	3		
	7.591438	11.073544	-0.922833	0.249814	1.252291	3	I	
	0.463576	10.972812	-0.04411	0.249814	0.590378	3		
	8.673346	12.184984	-0.984139	0.249814	0.321184	3		
	4.854848	11.53451	-0.228029	0.249814	0.816453	3		
	5.427623	11.214721	-1.02501	0.249814	1.620538	3		
	3.645657	11.222691	-1.004574	0.249814	-0.103	3		
		共 41516 条	25条页 💌 🤇 1 2	3 4 5 6 1661 >	前往 1 页			

图 26 K-Means 聚类算法的结果

				K-Mea	ans算法结果			
模型参数								
創出配置的参数以	及参数的取值。							
参数名称	参数值							
聚类个数	5							
最大迭代次数	100							
聚类中心:								
cluster id	flight count	seg km sum	last_to_end	avg_discount	long			

图 27 K-Means 聚类算法的报告