

航空公司客户价值分析

背景描述及业务需求

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。

将根据航空公司客户数据（见表1-1）实现以下目标：

- 借助航空公司客户数据，对客户进行分类。
- 对不同的客户类别进行特征分析，比较不同类客户的客户价值。
- 对不同价值的客户类别提供个性化服务，制定相应的营销策略。

这个案例的目标是客户价值识别，即通过航空公司客户数据识别不同价值的客户。识别客户价值应用最广泛的模型是通过三个指标（最近消费时间间隔(Recency)、消费频率(Frequency)、消费金额(Monetary)）来进行客户细分，识别出高价值的客户，简称RFM模型。

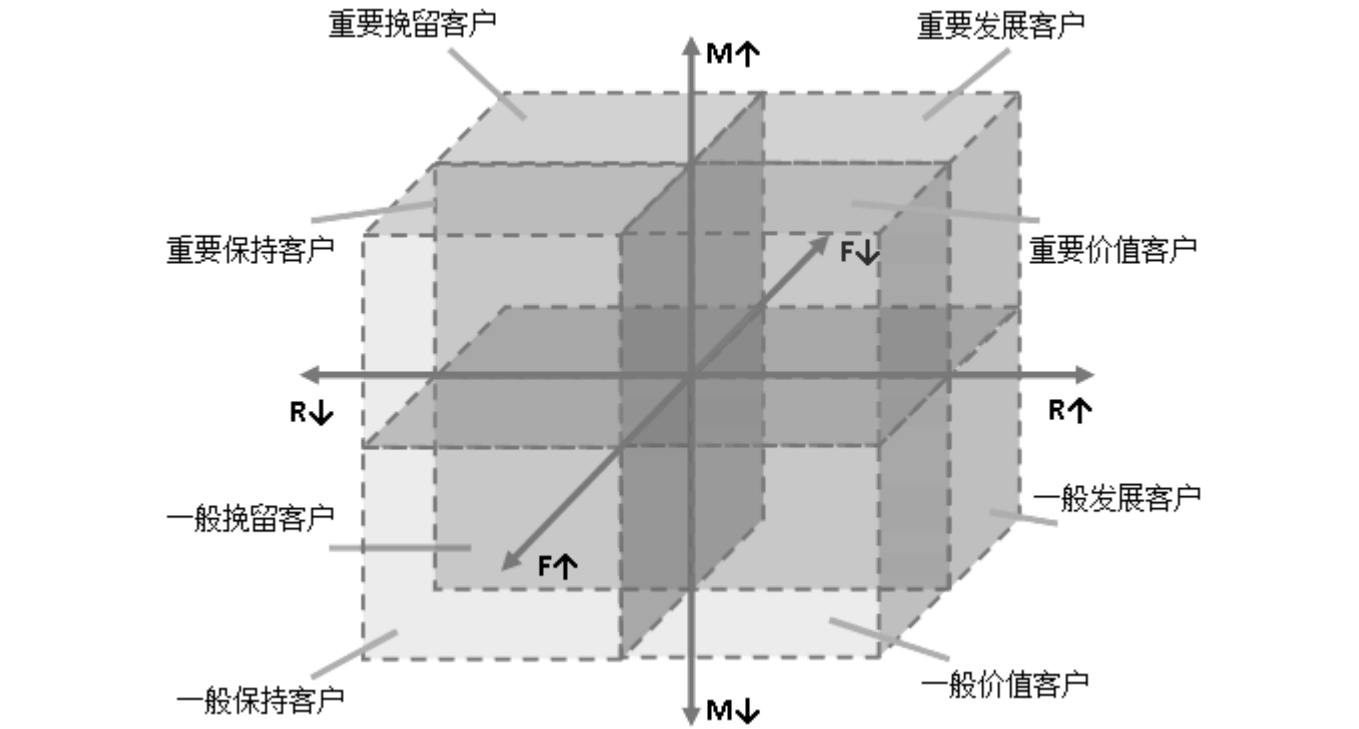
在RFM模型中，消费金额表示在一段时间内，客户购买该企业产品金额的总和。由于航空票价受到运输距离、舱位等级等多种因素影响，同样消费金额的不同旅客对航空公司的价值是不同的，比如一位购买长航线、低等级舱位票的旅客与一位购买短航线、高等级舱位票的旅客相比，后者对于航空公司而言价值可能更高。因此这个指标并不适合用于航空公司的客户价值分析。我们选择客户在一定时间内累积的飞行里程M和客户在一定时间内乘坐舱位所对应的折扣系数的平均值C两个指标代替消费金额。此外，考虑航空公司会员入会时间的长短在一定程度上能够影响客户价值，所以在模型中增加客户关系长度L，作为区分客户的另一指标。

本案例将客户入会时长L、消费时间间隔R、消费频率F、飞行里程M 和折扣系数的平均值C五个指标作为航空公司识别客户价值指标（见表1-1），记为LRPMC模型。

表1-1 指标含义

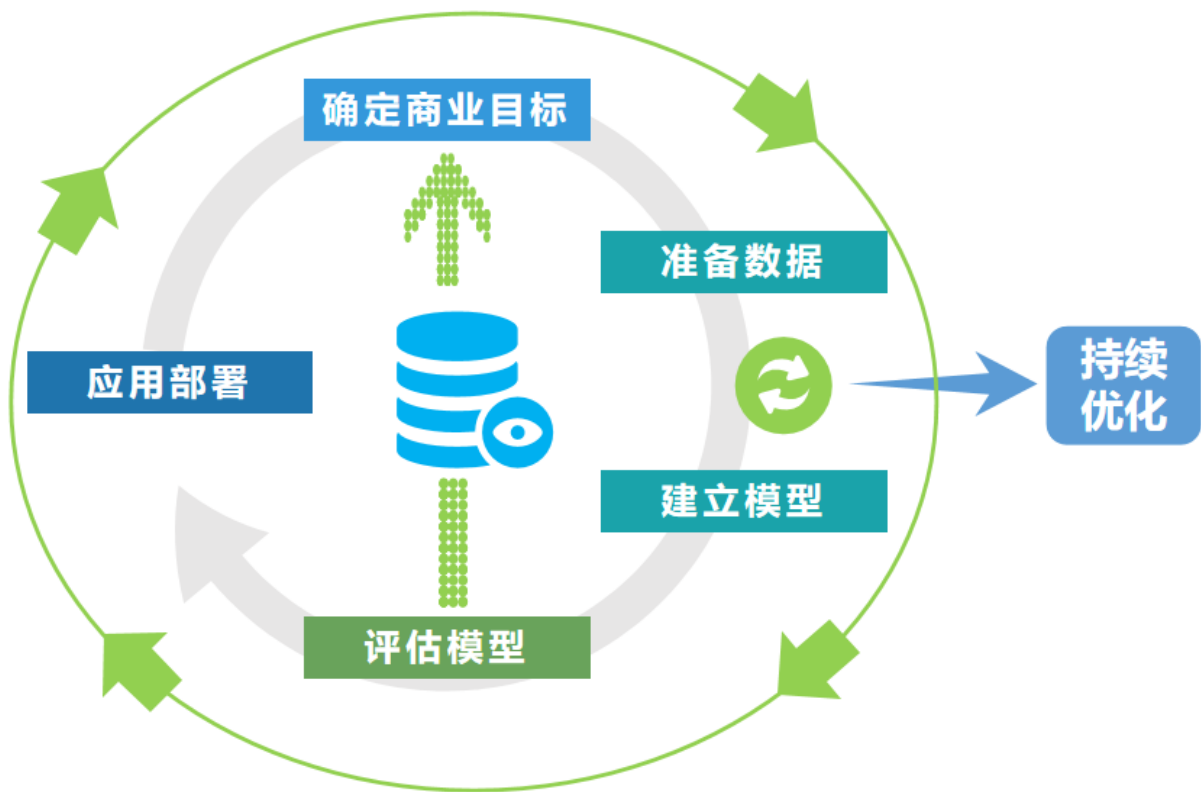
模型	L	R	F	M	C
航空公司LRPMC模型	会员入会时间距观测窗口结束的月数	客户最近一次乘坐公司飞机距观测窗口结束的月数	客户在观测窗口内乘坐公司飞机的次数	客户在观测窗口内累计的飞行里程	客户在观测窗口内乘坐舱位所对应的折扣系数的平均值

针对航空公司LRPMC模型，如果采用传统RFM模型分析的属性分箱方法，如图所示（它是依据属性的平均值进行划分，其中大于平均值的表示为↑，小于平均值的表示为↓），虽然也能够识别出最有价值的客户，但是细分的客户群太多，提高了针对性营销的成本。因此，本案例采用聚类的方法识别客户价值。通过对航空公司客户价值的LRPMC五个指标进行K-Means聚类，识别出最有价值客户。



- 背景描述及业务需求
- 实施过程
 - 数据接入
 - 数据探索
 - 数据预处理
 - 数据清洗
 - 数据规约
 - 数据变换
 - 预处理流程图
 - 建立模型
 - 客户聚类
 - 客户价值
 - 模型应用
- 总结

本案例航空客户价值分析的总体流程如图所示：



实施过程

本案例数据集共计62988条，以下是字段说明：

字段名称	字段说明	备注
MEMBER_NO	会员卡号	
FFP_DATE	入会时间	办理会员卡的开始的时间
FIRST_FLIGHT_DATE	第一次飞行日期	
GENDER	性别	
FFP_TIER	会员卡级别	
WORK_CITY	工作地城市	
WORK_PROVINCE	工作地所在省份	
WORK_COUNTRY	工作地所在国家	
AGE	年龄	
LOAD_TIME	观测窗口的结束时间	选取样本的时间宽度，距离现在最近的时间。
FLIGHT_COUNT	飞行次数	频数
BP_SUM	观测窗口总基本积分	航空公里的里程就相当于积分，积累一定分数可以兑换奖品和免费里程。
EP_SUM_YR_1	第一年精英资格积分	
EP_SUM_YR_2	第二年精英资格积分	
SUM_YR_1	第一年总票价	
SUM_YR_2	第二年总票价	

SEG_KM_SUM	观测窗口总飞行公里数	
WEIGHTED_SEG_KM	观测窗口总加权飞行公里数（ Σ 舱位折扣 $\times$ 航段距离）	
LAST_FLIGHT_DATE	末次飞行日期	最后一次飞行时间
AVG_FLIGHT_COUNT	观测窗口季度平均飞行次数	
AVG_BP_SUM	观测窗口季度平均基本积分累积	
BEGIN_TO_FIRST	观察窗口内第一次乘机时间至MAX（观察窗口始端，入会时间）时长	
LAST_TO_END	最后一次乘机时间至观察窗口末端时长	
AVG_INTERVAL	平均乘机时间间隔	
MAX_INTERVAL	观察窗口内最大乘机间隔	
ADD_POINTS_SUM_YR_1	观测窗口中第1年其他积分（合作伙伴、促销、外航转入等）	
ADD_POINTS_SUM_YR_2	观测窗口中第2年其他积分（合作伙伴、促销、外航转入等）	
EXCHANGE_COUNT	积分兑换次数	
avg_discount	平均折扣率	
P1Y_Flight_Count	第1年乘机次数	
L1Y_Flight_Count	第2年乘机次数	
P1Y_BP_SUM	第1年里程积分	
L1Y_BP_SUM	第2年里程积分	
EP_SUM	观测窗口总精英积分	
ADD_Point_SUM	观测窗口中其他积分（合作伙伴、促销、外航转入等）	
Eli_Add_Point_Sum	非乘机积分总和	
L1Y_ELi_Add_Points	第2年非乘机积分总和	
Points_Sum	总累计积分	
L1Y_Points_Sum	第2年观测窗口总累计积分	
Ration_L1Y_Flight_Count	第2年的乘机次数比率	
Ration_P1Y_Flight_Count	第1年的乘机次数比率	
Ration_P1Y_BPS	第1年里程积分占最近两年积分比例	
Ration_L1Y_BPS	第2年里程积分占最近两年积分比例	
Point_NotFlight	非乘机的积分变动次数	

数据接入

在实验中添加 [数据源](#) 节点，将航空公司客户数据读取进来，部分数据如图所示：

查看输出

当前显示 44 列 / 总共 44 列, 100 条 / 总共有 62988 条数据

列筛选 请选择

表头真名 表头别名

MEMBER_NO	FFP_DATE	FIRST_FLIGHT_DATE	GENDER	FFP_TIER	WORK_CITY
54993	2006/11/02	2008/12/24	男	6	
28065	2007/02/19	2007/08/03	男	6	
55106	2007/02/01	2007/08/30	男	6	
21189	2008/08/22	2008/08/23	男	5	Los Angeles
39546	2009/04/10	2009/04/15	男	6	贵阳
56972	2008/02/10	2009/09/29	男	6	广州
44924	2006/03/22	2006/03/29	男	6	乌鲁木齐市
22631	2010/04/09	2010/04/09	女	6	温州市
32197	2011/06/07	2011/07/01	男	5	DRANCY
31645	2010/07/05	2010/07/05	女	6	
58877	2010/11/18	2010/11/20	女	6	
37994	2004/11/13	2004/12/02	男	6	
28012	2006/11/23	2007/11/18	男	5	CAN

提示: 点击单元格可查看超出的内容。注意: 表头中 表示特征列, * 表示标签列

下载预览数据

为了便于理解字段含义，让字段看起来更加直观，使用 元数据编辑 节点，添加中文字段别名，更改后的输出结果如下：

查看输出

当前显示 44 列 / 总共 44 列, 100 条 / 总共有 62988 条数据

列筛选 请选择

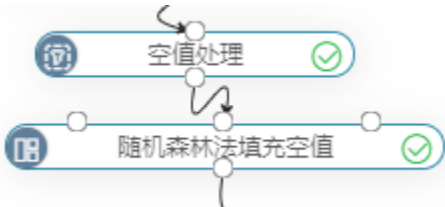
表头真名 表头别名

#	会员卡号	入会时间	第一次飞行日期	性别	会员卡级别	工作地城市	工作地所在省份	工作地所在国家	#	年龄	观测窗口的结束时间
	54993	2006/11/02	2008/12/24	男	6	.	北京	CN		31	2014/03/31
	28065	2007/02/19	2007/08/03	男	6	null	北京	CN		42	2014/03/31
	55106	2007/02/01	2007/08/30	男	6	.	北京	CN		40	2014/03/31
	21189	2008/08/22	2008/08/23	男	5	Los Angeles	CA	US		64	2014/03/31
	39546	2009/04/10	2009/04/15	男	6	贵阳	贵州	CN		48	2014/03/31
	56972	2008/02/10	2009/09/29	男	6	广州	广东	CN		64	2014/03/31
	44924	2006/03/22	2006/03/29	男	6	乌鲁木齐市	新疆	CN		46	2014/03/31
	22631	2010/04/09	2010/04/09	女	6	温州市	浙江	CN		50	2014/03/31
	32197	2011/06/07	2011/07/01	男	5	DRANCY	null	FR		50	2014/03/31

提示: 点击单元格可查看超出的内容。注意: 表头中 表示特征列, * 表示标签列

下载预览数据

流程图如图所示：



数据探索

本案例的探索分析是对数据进行缺失值分析与异常值分析，分析出数据的规律以及异常值。通过对数据观察发现原始数据中存在票价为空值，票价最小值为0、折扣率最小值为0、总飞行公里数大于0的记录。票价为空值的数据可能是客户不存在乘机记录造成。其它的数据可能是客户乘坐0折机票或者积分兑换造成。

为了查看整个数据集数值型数据的情况，我们接入一个 **全表统计** 节点，选中所有数值型字段。

选择列

源数据列表0/9

Q 请输入搜索内容

☐ #会员卡号

☐ Ab 入会时间

☐ Ab 第一次飞行日期

☐ Ab 性别

☐ Ab 工作地城市

☐ Ab 工作地所在省份

☒ 全部

☐ 字符

☐ 数字

到右边>

< 到左边

已选字段列表0/35

Q 请输入搜索内容

☐ #会员卡级别

☐ #年龄

☐ #观测窗口内的飞行次数

☐ #观测窗口总基本积分

☐ #第一年精英资格积分

☐ #第二年精英资格积分

☐ #第一年总票价

确定取消

输出结果如图所示,可以看到部分数据(年龄、第一年总票价、第二年总票价)存在缺失值。

查看全量统计

	方差	总和	行数	唯一值	缺失值	偏度
	0.13976...	258387	62988	3	0	3.87319..
	97.7313...	2657660	62988	84	420	0.62593..
	197.387...	745741	62988	153	0	3.21522..
	2.66978...	6881490...	62988	23449	0	5.66033..
	0.0	0	62988	1	0	NaN
	2708337...	16735258	62988	3101	0	12.1242..
	6.57631...	3343736...	62988	15827	551	5.08791..
	7.57485...	3522130...	62988	16767	138	4.60123..
	4.39357...	1078598...	62988	29081	0	3.88451..
	3.09006...	8.04807...	62988	58631	0	4.94586..
	3.19335...	97137.1...	62988	300	0	3.13140..
	4339394...	8.95336...	62988	29710	0	5.56812..
	25463.4...	7567724	62988	730	0	1.71990..

数据预处理

本案例主要采用数据清洗、数据规约与数据变换的预处理方法。

数据清洗

通过数据探索分析，发现数据中存在缺失值，票价最小值为0、折扣率最小值为0、总飞行公里数大于0的记录。由于原始数据量大，这类数据所占比例较小，对于问题影响不大，因此对其进行丢弃处理。具体处理方法如下：

- 丢弃票价为空的情况。
- 丢弃票价为0、平均折扣率不为0、总飞行公里数大于0的记录。

使用 [空值处理](#) 节点，删除票价为空值的行，选择列如图：

选择列

可选字段42

Q 请输入搜索内容

☐ #会员卡号

☐ Ab入会时间

☐ Ab第一次飞行日期

☐ Ab性别

☐ #会员卡级别

☐ Ab工作地城市

☒ 全部

☐ 字符

☐ 数字

到右边>

< 到左边

已选字段2

Q 请输入搜索内容

☐ #第一年总票价

☐ #第二年总票价

确定

取消

空值处理节点参数空值处理方式选择“过滤整行”。

运行成功：3秒

公司客户数据

元数据编辑

空值处理

参数

属性

帮助

选择列 ?

选择列

空值处理

过滤整行

值 ?

使用 [行选择](#) 节点，保留总票价不为0或者观测光口总飞行公里数和平均折扣率同时为0的情况，根据条件筛选如图所示：

过滤与筛选

☐ 根据行号筛选: 筛选行 行号: 正数topN 负数bottomN

☐ 根据行号区间筛选: 筛选行 输入的筛选区间: ~

☒ 根据条件筛选

筛选行 观测窗口总飞 等于 值 0

【观测窗口总飞行公里数】 = 0

and 【平均折扣率】 = 0

or 【第一年总票价】 != 0

or 【第二年总票价】 != 0

确定 取消

数据规约

原始数据中属性太多，根据航空公司客户价值LRFMC模型，选择与LRFMC指标相关的七个属性：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）、AVG_DISCOUNT（平均折扣率）、SEG_KM_SUM（观测窗口总飞行公里数）、LAST_FLIGHT_DATE（末次飞行日期）、LAST_TO_END（最后一次乘机时间至观察窗口末端时长）。删除与其不相关、弱相关或冗余的属性，例如：会员卡号、性别、工作地城市、工作地所在省份、工作地所在国家、年龄等属性。

使用 列选择 节点，选择需要的属性列如图所示：

选择输出列

源数据列表 0/37

Q 请输入搜索内容

☐ # 会员卡号

☐ A_b 第一次飞行日期

☐ A_b 性别

☐ # 会员卡级别

☐ A_b 工作地城市

☐ A_b 工作地所在省份

全部 字符 数字

到右边 >

< 到左边

已选字段列表 0/7

Q 请输入搜索内容

☐ A_b 入会时间

☐ A_b 观测窗口的结束时间

☐ # 观测窗口内的飞行次数

☐ # 观测窗口总飞行公里数

☐ A_b 末次飞行日期

☐ # 最后一次乘机时间至观察窗口末端时长

☐ # 平均折扣率

确定 取消

经过属性选择后的数据集如下：

查看输出

×

当前显示 7 列 / 总共 7 列, 100 条 / 总共有 62044 条数据

列筛选

请选择

表头真名 表头别名

入会时间	观测窗口的结束时间	# 观测窗口内的飞行次数	# 观测窗口总飞行公里数	未次飞行日期	# 最后一次乘机时间至观察窗口末端时长	# 平均折扣率
2006/11/02	2014/03/31	210	580717	2014/03/31	1	0.961639043
2007/02/19	2014/03/31	140	293678	2014/03/25	7	1.25231444
2007/02/01	2014/03/31	135	283712	2014/03/21	11	1.254675516
2008/08/22	2014/03/31	23	281336	2013/12/26	97	1.090869565
2009/04/10	2014/03/31	152	309928	2014/03/27	5	0.970657895
2008/02/10	2014/03/31	92	294585	2014/01/13	79	0.967692483
2006/03/22	2014/03/31	101	287042	2014/03/31	1	0.965346535
2010/04/09	2014/03/31	73	287230	2014/03/29	3	0.962070222
2011/06/07	2014/03/31	56	321489	2014/03/26	6	0.828478237
2010/07/05	2014/03/31	64	375074	2014/03/17	15	0.708010153
2010/11/18	2014/03/31	43	262013	2014/03/10	22	0.988658044

提示: 点击单元格可查看输出的内容。注意: 表头中 表示特征列, * 表示标签列

下载预览数据

数据变换

为了方便后续的日期处理, 字符串类型的字段FFP_DATE (入会时间) 和LOAD_TIME (观测窗口的结束时间) 格式化成为指定的日期格式 (yyyy/MM/dd), 加入一个 派生列 节点完成此项操作, 参数派生列配置如图:

派生列配置

×

输入关键字搜索列

STRING

入会时间

观测窗口的结束时间

未次飞行日期

DOUBLE

平均折扣率

INTEGER

观测窗口内的飞行次数

函数

累积

字符拼接(拼接符)

字符拼接

字符截取

小数点保留

内容提取(年)

内容提取(月)

添加/编辑表达式

1

派生列名: 请输入或左侧拖入派生列名

操作: 增加 修改

入会时间00

to_date([入会时间], 'yyyy/MM/dd')

观测窗口的结束时间00

to_date([观测窗口的结束时间], 'yyyy/

未次飞行日期00

to_date([未次飞行日期], 'yyyy/MM/d

确定

取消

由于原始数据中并没有直接给出LRFMC五个指标, 需要通过原始数据提取这五个指标, 具体的计算方式如下:

- n L = LOAD_TIME - FFP_DATE: 会员入会时间距观测窗口结束的月数 = 观测窗口的结束时间 - 入会时间[单位: 月]。
- n R = LAST_TO_END: 客户最近一次乘坐公司飞机距观测窗口结束的月数 = 最后一次乘机时间至观察窗口末端时长[单位: 月]。
- n F = FLIGHT_COUNT: 客户在观测窗口内乘坐公司飞机的次数 = 观测窗口的飞行次数[单位: 次]。
- n M = SEG_KM_SUM: 客户在观测时间内在公司累计的飞行里程 = 观测窗口总飞行公里数[单位: 公里]。
- n C = AVG_DISCOUNT: 客户在观测时间内乘坐舱位所对应的折扣系数的平均值 = 平均折扣率[单位: 无]。

通过派生列节点对数据进行提取, 该节点的派生列配置, 表达式如下:

指标	表达式
L	DATEDIFF([观测窗口的结束时间00], [入会时间00])/30
R	[最后一次乘机时间至观察窗口末端时长]/30
F	[观测窗口内的飞行次数]

M	[观测窗口总飞行公里数]
C	[平均折扣率]

整个派生列配置如图：

派生列配置

输入关键字搜索列

STRING

入会时间

观测窗口的结束时间

末次飞行日期

DOUBLE

平均折扣率

INTEGER

观测窗口总飞行公里数

函数

累积

字符拼接(拼接符)

字符拼接

字符截取

小数点保留

内容提取(年)

内容提取(月)

添加/编辑表达式

1

派生列名: 请输入或左侧拖入派生列名 操作: 增加 修改

L	DATEDIFF([观测窗口的结束时间00],	
R	[最后一次乘机时间至观察窗口未端时	
F	[观测窗口内的飞行次数]	
M	[观测窗口总飞行公里数]	
C	[平均折扣率]	

确定 取消

输出结果显示了指标数值分布情况：

🕒 查看全量统计

列名	直方图	箱线图	最大值	最小值	平均值	标准差
L	查看	查看	114.566...	12.1666...	49.6230...	28.262
R	查看	查看	24.3666...	0.03333...	5.75109...	6.0508
F	查看	查看	213	2	11.9713...	14.110
M	查看	查看	580717	368	17321.6...	21052.
C	查看	查看	1.5	0.13601...	0.72218...	0.1848

五个指标的数据提取后，对每个指标数据分布情况进行分析，接入一个 [全表统计](#) 节点看一下指标数值的分布情况，其数据的取值范围如上图所示。从表中数据可以发现，五个指标的取值范围数据差异较大，为了消除数量级数据带来的影响，需要对数据进行标准化处理。接入一个 [PYTHON脚本](#) 节点，通过属性修改名称为标准化处理，脚本代码如下图所示：

```
1 # -*- coding: UTF-8 -*-
2 import numpy as np
3 import pandas as pd
4 from sklearn.preprocessing import StandardScaler
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
10     scaler = StandardScaler()
11     dataframe2 = pd.DataFrame(scaler.fit_transform(dataframe1[['L', 'R', 'F', 'M', 'C']]), columns=['L', 'R', 'F', 'M', 'C'])
12     dataframe2.columns = ['Z'+i for i in dataframe2.columns]
13     dataframe1 = pd.concat([dataframe1, dataframe2], axis=1)
14     # Return value must be a pandas.DataFrame
15     return dataframe1
```

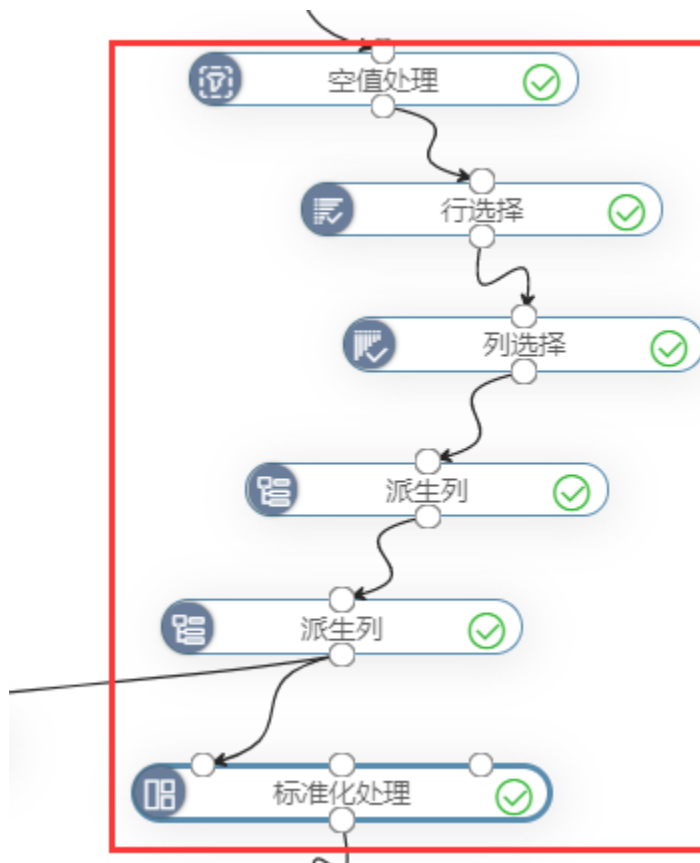
* 不能使用字段别名

确定

取消

预处理流程图

整个预处理流程图如下：



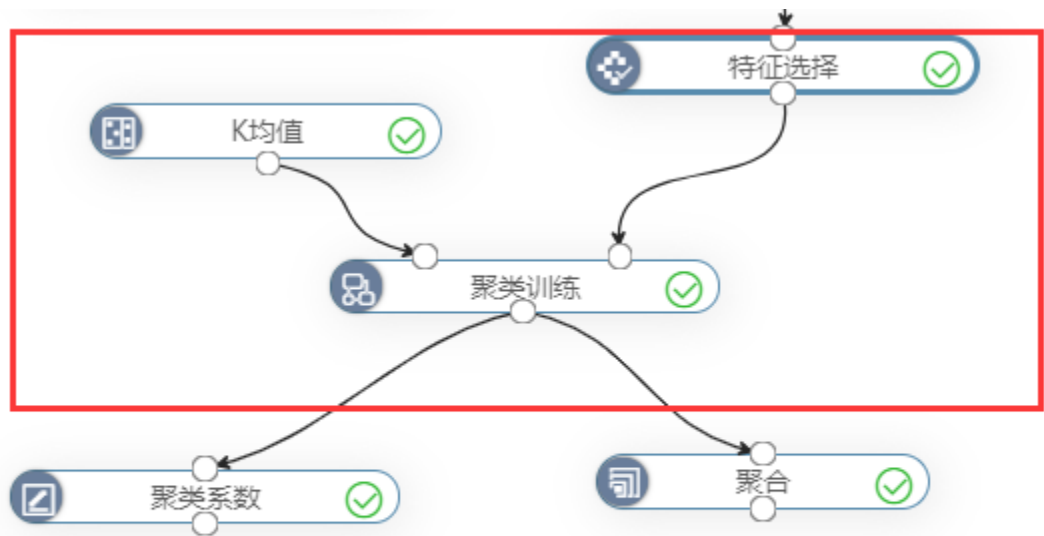
建立模型

客户价值分析模型构建主要由两个部分构成，第一个部分根据航空公司客户五个指标的数据，对客户作聚类分群。第二部分结合业务对每个客户群进行特征分析，分析其客户价值，并对每个客户群进行排名。

客户聚类

采用 [K均值](#) 算法对客户数据进行客户分群，聚成五类（需要结合业务的理解与分析来确定客户的类别数量）。

利用K均值算法进行客户分群的流程如图，聚类类别数为k=5。



接入 [特征选择](#) 节点，选择需要做聚类的特征列：

✦选择特征列

源数据列表0/10

Q 请输入搜索内容

☐ A₀ 入会时间

☐ A₀ 观测窗口的结束时间

☐ # 观测窗口内的飞行次数

☐ # 观测窗口总飞行公里数

☐ A₀ 末次飞行日期

☐ # 最后一次乘机时间至观察窗口末端时长

☒ 全部

☐ 字符

☐ 数字

到右边>

< 到左边

已选字段列表0/5

Q 请输入搜索内容

☐ # L

☐ # R

☐ # F

☐ # M

☐ # C

确定

取消

接入 [K均值](#) 算法节点，参数配置如图，K值取值为5，迭代次数填写50。

» 参数

属性

帮助

归一化

方法选择

标准化

参数

☐ 单位标准差归一化

☐ 平均数据中心化

K值 范围是 ≥ 2 的整数 ?

5

初始化模式

kmeans++

随机种子 范围是任意整数 ?

2

收敛阈值 范围是 ≥ 0 的数 ?

0.0001

迭代次数 范围是 > 0 的整数 ?

50

最后接入 [聚类训练](#) 节点，实现模型的训练。

训练完成后，对聚类结果做进一步分析，接入 [聚类系数](#) 节点，输出各类别的聚类系数指标，如图：

查看输出

当前显示 6 列/总共 6 列, 5 条/总共有 5 条数据

列筛选 请选择

表头真名 表头别名

class	# ZL	# ZR	# ZF	# ZM	# ZC
0	1.154612288954416	-0.3722329778514664	-0.0925097687742423	-0.10117226353151328	-0.1381244566788786
1	0.1485372522570826	-0.04647065819181977	-0.14004384911544676	-0.12601413365906874	2.5386962374685282
2	0.4814788768230204	-0.7988788224617277	2.485158062338137	2.4257863677834934	0.28622789059425524
3	-0.31431951587624674	1.6753150432874173	-0.5736969737411564	-0.5376964503852085	-0.146445173181426
4	-0.6980761623880857	-0.40963068660153956	-0.16666064636034386	-0.16803957230308159	-0.22022437949660578

提示: 点击单元格可查看输出的内容。注意: 表头中 表示特征列, * 表示标签列

下载预览数据

为了想统计分类的分布数量情况，我们接入一个 聚合 节点，对各分类的样本数量进行统计，输出如图：

查看输出

当前显示 2 列/总共 2 列, 5 条/总共有 5 条数据

列筛选 请选择

表头真名 表头别名

# Group_prediction	# Count_prediction
4	25219
3	12342
0	15986
1	3182
2	5315

提示: 点击单元格可查看输出的内容。注意: 表头中 表示特征列, * 表示标签列

下载预览数据

客户价值

将聚类的结果利用Smartbi自助仪表盘进行分析，得到客户群特征分析图，预测有 0、1、2、3、4五种类别。通过 聚合 节点对每个类别的L、R、F、M、C进行统计，求取均值以及每类总数量，如表2-3所示。

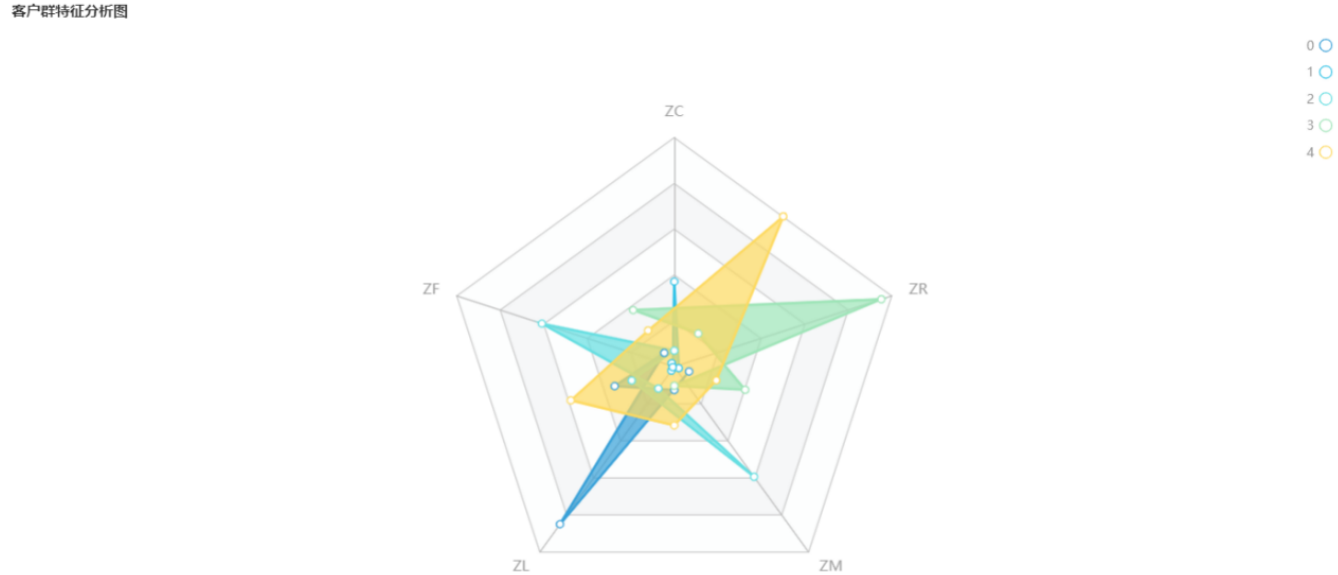


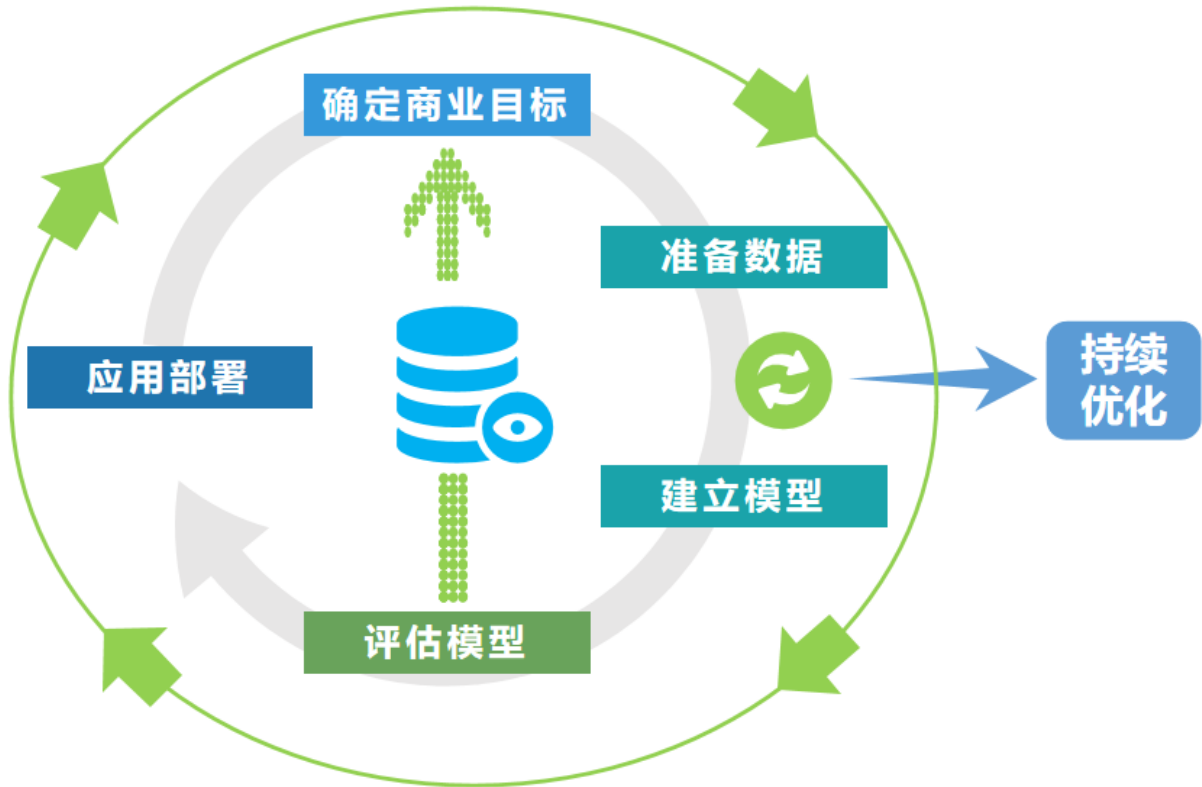
表2-3 客户群特征描述表

群类别	会员入会时间距观测窗口结束的月数 (L)	客户最近一次乘坐公司飞机距观测窗口结束的月数 (R)	客户在观测窗口内乘坐公司飞机的次数 (F)	客户在观测时间内在公司累计的飞行里程 (M)	客户在观测时间内乘坐舱位所对应的折扣系数的平均值 (C)	数量
0	82.26	3.50	10.67	15191	0.70	15986
1	53.79	5.47	9.98	14654	1.19	3182
2	63.23	0.92	47.04	68391	0.78	5315

3	40.74	15.89	3.88	6002	0.70	12342
4	29.89	3.27	9.62	13785	0.68	25219

本案例定义五个等级的客户类别：重要保持客户、重要发展客户、重要挽留客户、一般客户、低价值客户。他们之间的区别如图2-21所示，其中每种客户类别的特征如下：

- 重要保持客户：**这类客户的平均折扣率（C）较高（一般所乘航班的舱位等级较高），最近乘坐过本公司航班（R）低，乘坐的次数（F）或里程（M）较高。他们是航空公司的高价值客户，是最为理想的客户类型，对航空公司的贡献最大，所占比例却较小。航空公司应该优先将资源投放到他们身上，对他们进行差异化管理和一对一营销，提高这类客户的忠诚度与满意度，尽可能延长这类客户的高水平消费。
- 重要发展客户：**这类客户的平均折扣率（C）较高，最近乘坐过本公司航班（R）低，但乘坐次数（F）或乘坐里程（M）较低。这类客户入会时长（L）短，他们是航空公司的潜在价值客户。虽然这类客户的当前价值并不是很高，但却有很大的发展潜力。航空公司要努力促使这类客户增加在本公司的乘机消费和合作伙伴处的消费，也就是增加客户的钱包份额。通过客户价值的提升，加强这类客户的满意度，提高他们转向竞争对手的转移成本，使他们逐渐成为公司的忠诚客户。
- 重要挽留客户：**这类客户过去所乘航班的平均折扣率（C）、乘坐次数（F）或者里程（M）较高，但是较长时间已经没有乘坐本公司的航班（R）高或是乘坐频率变小。他们客户价值变化的不确定性很高。由于这些客户衰退的原因各不相同，所以掌握客户的最新信息、维持与客户的互动就显得尤为重要。航空公司应该根据这些客户的最近消费时间、消费次数的变化情况，推测客户消费的异动状况，并列岀客户名单，对其重点联系，采取一定的营销手段，延长客户的生命周期。
- 一般与低价值客户：**这类客户所乘航班的平均折扣率（C）很低，较长时间没有乘坐过本公司航班（R）高，乘坐的次数（F）或里程（M）较低，入会时长（L）短。他们是航空公司的一般用户与低价值客户，可能是在航空公司机票打折促销时，才会乘坐本公司航班。



其中，重要发展客户、重要保持客户、重要挽留客户这三类重要客户分别可以归入客户生命周期管理的发展期、稳定期、衰退期三个阶段。

根据每种客户类型的特征，对各类客户群进行客户价值排名，其结果如表2-4所示。针对不同类型的客户群提供不同的产品和服务，提升重要发展客户的价值、稳定和延长重要保持客户的高水平消费、防范重要挽留客户的流失并积极进行关系恢复。

表2-4 客户群价值排名

客户群	排名	排名含义
0	3	重要挽留用户
1	1	一般用户
2	2	重要保持客户
3	5	低价值客户
4	4	重要发展客户

本模型采用历史数据进行建模，随着时间的变化，分析数据的观测窗口也在变换。因此，对于新增客户详细信息，考虑业务的实际情况，该模型建议每个月运行一次，对其新增客户信息通过聚类中心进行判断，同时对本次新增客户的特征进行分析。如果增量数据的实际情况与判断结果差异大，需要业务部门重点关注，查看变化大的原因以及确认模型的稳定性。如果模型稳定性变化大，需要重新训练模型进行调整。目前模型进行重新训练的时间没有统一标准，大部分情况都是根据经验来决定。根据经验建议：**每隔半年训练一次模型比较合适。**

模型应用

根据对各个客户群进行特征分析，采取下面的一些营销手段和策略，为航空公司的价值客户群管理提供参考。

（1）会员的升级与保级

航空公司的会员可以分为白金卡会员、金卡会员、银卡会员、普通卡会员，其中非普通卡会员可以统称为航空公司的精英会员。虽然各个航空公司都有自己的特点和规定，但会员制的管理方法是大同小异的。成为精英会员一般都是要求在一定时间内（如一年）积累一定的飞行里程或航段，达到这种要求后就会在有效期内（通常为两年）成为精英会员，并享受相应的高级别服务。有效期快结束时，根据相关评价方法确定客户是否有资格继续作为精英会员，然后对该客户进行相应地升级或降级。

然而，由于许多客户并没有意识到或根本不了解会员升级或保级的时间与要求（相关的文件说明往往复杂且不易理解），经常在评价期过后才发现自己其实只差一点就可以实现升级或保级，却错过了机会，使之前的里程积累白白损失。同时，这种认知还可能导致客户的不满，干脆放弃在本公司的消费。

因此，航空公司可以在对会员升级或保级进行评价的时间点之前，对那些接近但尚未达到要求的较高消费客户进行适当提醒甚至采取一些促销活动，刺激他们通过消费达到相应标准。这样既可以获得收益，同时也提高了客户的满意度，增加了公司的精英会员。

（2）首次兑换

航空公司常旅客计划中最能够吸引客户的内容就是客户可以通过消费积累的里程来兑换免票或免费升舱等。各个航空公司都有一个首次兑换标准，也就是当客户的里程或航段积累到一定程度时才可以实现第一次兑换，这个标准会高于正常的里程兑换标准。但是很多公司的里程积累随着时间会进行一定地削减，例如有的公司会在年末对该年积累的里程进行折半处理。这样会导致许多不了解情况的会员白白损失自己好不容易积累的里程，甚至总是难以实现首次兑换。同样，这也会引起客户的不满或流失。可以采取的措施是从数据库中提取出接近但尚未达到首次兑换标准的会员，对他们进行提醒或促销，使他们通过消费达到标准。一旦实现了首次兑换，客户在本公司进行再次消费兑换就比其他公司进行兑换要容易许多，在一定程度上等于提高了转移的成本。另外，在一些特殊的时间点（如里程折半的时间点）之前可以给客户一些提醒，这样可以增加客户的满意度。

（3）交叉销售

通过发行联名卡等与非航空类企业的合作，使客户在其他企业的消费过程中获得本公司的积分，增强与公司的联系，提高他们的忠诚度。如可以查看重要客户在非航空类合作伙伴处的里程积累情况，找出他们习惯的里程积累方式（是否经常在合作伙伴处消费、更喜欢消费哪些类型合作伙伴的产品），对他们进行相应促销。

客户识别期和发展期为客户关系打下基石，但是这两个时期带来的客户关系是短暂的、不稳定的。企业要获取长期的利润，必须具有稳定的、高质量的客户。保持客户对于企业是至关重要的，不仅因为争取一个新客户的成本远远高于维持老客户的成本，更重要的是客户流失会造成公司收益的直接损失。因此，在这一时期，航空公司应该努力维系客户关系水平，使之处于较高的水准，最大化生命周期内公司与客户的互动价值，并使这样的高水平尽可能延长。对于这一阶段的客户，主要应该通过提供优质的服务产品和提高服务水平来提高客户的满意度。通过对常旅客数据库的数据挖掘、进行客户细分，可以获得重要保持客户的名单。这类客户一般所乘航班的平均折扣率（C）较高，最近乘坐过本公司航班（R 低），乘坐的频率（F）或里程（M）也较高。他们是航空公司的价值客户，是最为理想的客户类型，对航空公司的贡献最大，所占比例却比较小。航空公司应该优先将资源投放到他们身上，对他们进行差异化管理和一对一营销，提高这类客户的忠诚度与满意度，尽可能延长这类客户的高水平消费。

总结

本案例主要参考《Python数据分析与挖掘实战》一书。本案例主要针对客户价值进行分析，通过对航空公司客户价值的LRFMC五个指标进行K-Means聚类，识别出最有价值客户。将指标聚集成五个等级的客户类别：重要保持客户、重要发展客户、重要挽留客户、一般客户、低价值客户。根据对各个客户群进行特征分析，采取下面的一些营销手段和策略，为航空公司的价值客户群管理提供参考。