

航空客户价值分析

航空客户价值分析

一、项目基本情况介绍

1. 项目背景
2. 项目需求
3. 项目的输出
4. 项目数据说明

二、项目分析方法与过程

1. 分析方法
 - 1.1 目标
 - 1.2 方法
 - 1.3 模型修改
 - 1.4 LRFMC 模型
2. 挖掘步骤
3. 数据抽取
4. 探索性分析
5. 数据预处理
6. 模型构建
7. 模型应用
8. 小结

一、项目基本情况介绍

1. 项目背景

某航空公司需要以客户为中心，按照客户的需求，在对客户的特点了解上使用不同的营销手段，目的是争取更多新客户，降低客户流失率，降低服务成本，提高业务收入，增加 ARPU 值(average revenue per user 每个用户的平均收益，一般以月为单位)，进行精准的市场营销策略制定。

2. 项目需求

本项目的目标客户是公众客户(客户分为公众客户、商业客户即公司、大客户)，因而只对公众客户进行分群。初步的目标是中高端用户、中端用户、低端用户、其他需求用户。但这是经验而言，最终结果需要看模型的运行结果，不能主观臆断。

3. 项目的输出

根据客户信息，对客户进行分类。针对不同类型客户进行特征提取，分析不同类型客户的价值。采取个性化服务，根据客户类型，制定相应营销策略。

4. 项目数据说明

序号	属性名称	说明
1	会员卡号	
2	入会时间	办理会员卡的开始时间
3	第一次飞行日期	
4	性别	
5	会员卡级别	
6	工作地城市	
7	工作地所在省份	
8	工作地所在国家	
9	年龄	
10	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
11	飞行次数	频数

12	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
13	第1年精英资格积分	
14	第2年精英资格积分	
15	第1年总票价	
16	第2年总票价	
17	观测窗口总飞行公里数	
18	观测窗口总加权飞行公里数	$\sum(\text{舱位折扣} \times \text{航段距离})$
19	末次飞行日期	最后一次飞行时间
20	观测窗口季度平均飞行次数	
21	观测窗口季度平均基本积分累积	
22	观察窗口内第一次乘机时间至MAX时长	观察窗口时段，入会时长
23	最后一次乘机时间至观察窗口末端时长	
24	平均乘机时间间隔	
25	观察窗口内最大乘机间隔	
26	观测窗口中第1年其他积分	合作伙伴、促销、外航转入等
27	观测窗口中第2年其他积分	合作伙伴、促销、外航转入等
28	积分兑换次数	
29	平均折扣率	
30	第1年乘机次数	
31	第2年乘机次数	
32	第1年里程积分	
33	第2年里程积分	
34	观测窗口总精英积分	
35	观测窗口中其他积分	合作伙伴、促销、外航转入等
36	非乘机积分总和	
37	第2年非乘机积分总和	

38	总累计积分	
39	第2年观测窗口总累计积分	
40	第2年乘机次数比率	
41	第1年乘机次数比率	
42	第1年里程积分占最近两年积分比例	
43	第2年里程积分占最近两年积分比例	
44	非乘机的积分变动次数	

二、项目分析方法与过程

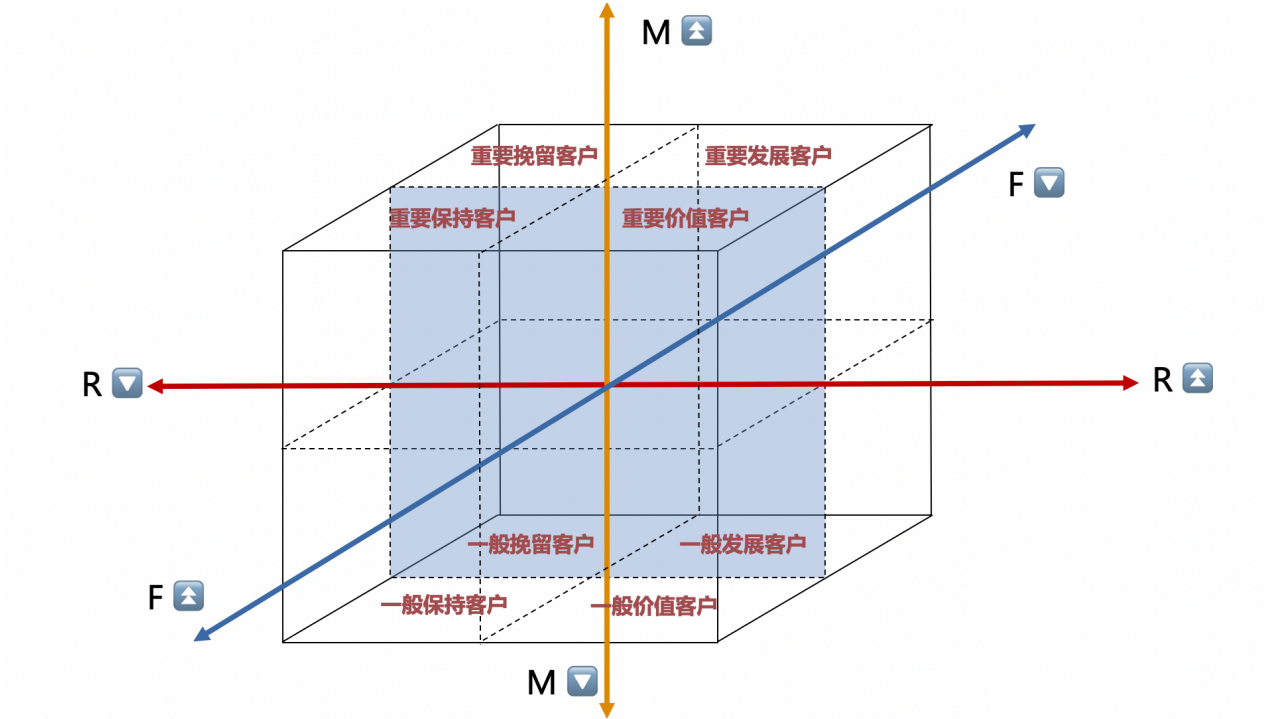
1. 分析方法

1.1 目标

首先，明确目标是客户价值识别。客户价值分析是以客户为中心，从客户需求出发，搞清楚客户需要什么，他们有怎样的一个特征，需要什么样的产品，然后设计相应的产品，通过对客户分群从而满足客户的需求。进行客户价值分析，可以避免商家闭门造车以及主观臆断客户的需求。

1.2 方法

识别客户价值，应用最广泛的模型是三个指标(消费时间间隔(Recency)，消费频率(Frequency)，消费金额(Monetary))。以上指标简称 RFM 模型，作用是识别高价值的客户。如果采用传统的 RFM 模型，如下图，它是依据各个属性的平均值进行划分，但是，细分的客户群太多，精准营销的成本太高。



客户类型	最近一次交易距离观测窗口的天数（黏性）	累计交易频次（忠诚度）	累计交易成交额（消费能力）	对应场景
重要价值客户	+	+	+	RFM都很大，优质客户
重要召回客户	-	+	+	成交量和成交额都很大，但是最近没有交易，需要召回
重要发展客户	+	-	+	成交额大，最近有交易，需要重点识别
重要挽留客户	-	-	+	成交额大，潜在的价值客户，需要挽留
潜力客户	+	+	-	成交量大，且最近有交易，需要挖掘
新客户	+	-	-	最近有交易，是新客户，需要推广
一般维持客户	-	+	-	成交量大，但是贡献不大，黏性也不高，一般维持
流失客户	-	-	-	已经流失的客户

1.3 模型修改

R 为最近乘坐航班时间间隔，F 为飞行次数。消费金额需要修改，因为航空票价收到距离和舱位等级的影响，同样金额对航空公司价值不同，因此，需要修改指标。

引入新的因素：

- 舱位因素 = 舱位所对应的折扣系数的平均值 = C
- 距离因素 = 一定时间内积累的飞行里程 = M

再考虑到，航空公司的会员系统，用户的入会时间长短能在一定程度上影响客户价值，所以增加指标 L = 入会时间长度

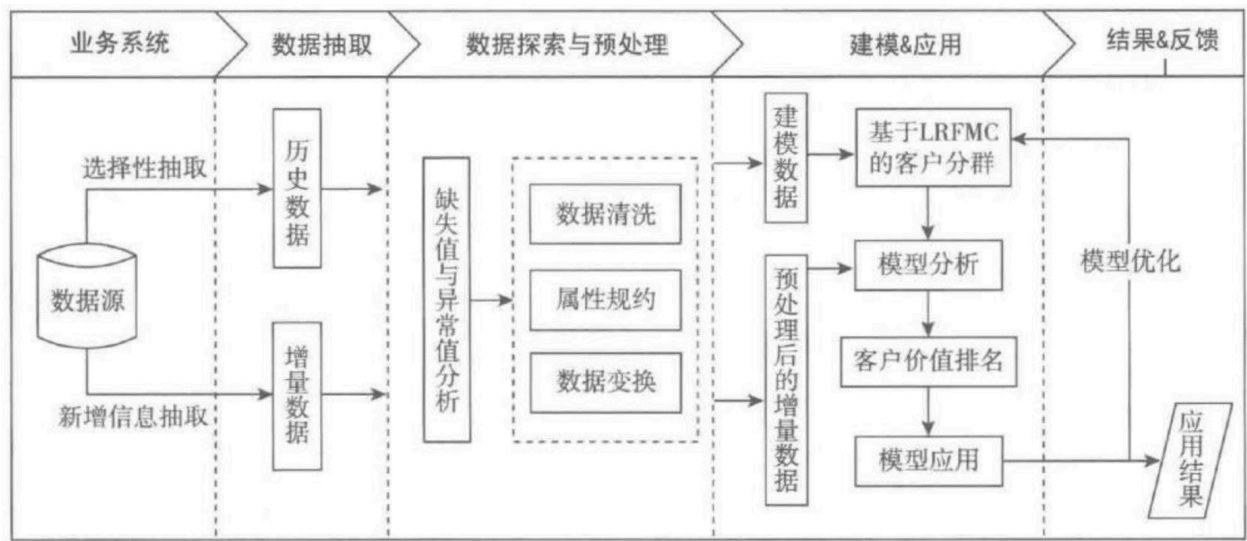
因此，总共确定了五个指标:

- L: 入会时间
- R: 最近乘坐航班
- F: 飞行次数
- M: 累积飞行里程
- C: 平均折扣率

1.4 LRFMC 模型

作为航空公司识别客户价值指标，记为 LRFMC 模型。因为数据没有标签信息，采用聚类的办法进行识别客户价值，以 LRFMC 模型为基础。

总体流程如下图：



2. 挖掘步骤

- 步骤1：从航空公司，选择性抽取与新增数据抽取，形成历史数据和增量数据；
- 步骤2：对步骤1的两个数据，进行数据探索性分析和预处理，主要有缺失值与异常值的分析处理、属性规约、清洗和变换；
- 步骤3：利用步骤2中的已处理数据作为建模数据，基于旅客价值的LRFMC模型进行客户分群，对各个客户群再进行特征分析，识别有价值客户；
- 步骤4：针对模型结果得到不同价值的客户。采用不同的营销手段，指定定制化的营销服务，或者针对性的优惠与关怀。

3. 数据抽取

- 选取2014-03-31为结束时间，选取宽度为两年的时间段，作为观测窗口，抽取观测窗口内所有客户的详细数据，形成历史数据；
- 对于后续新增的客户信息，采用目前的时间作为重点，形成新增数据。

4. 探索性分析

- 本案例的探索分析，主要对数据进行缺失值和异常值分析；
- 我们可以发现，存在票价为空值，折扣率为0，飞行公里数为0。票价为空值，可能是不存在飞行记录，其他空值可能是飞机票来自于积分兑换等渠道；
- 查找每列属性观测值中空值的个数、最大值、最小值。

5. 数据预处理

- 数据清洗
 - 丢弃票价为空记录
 - 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
- 属性归约
 - 原始数据中属性太多，根据航空公司客户价值LRFMC模型，选择与模型相关的属性
 - 删除其他无用属性，如会员卡号等
- 数据变换
 - 将原始数据转换成“适当”的格式，用来适应算法和分析等等的需要
 - 主要采用数据变换的方式为属性构造和数据标准化
 - 需要构造LRFMC的五个指标
 - $L = \text{LOAD_TIME} - \text{FFP_DATE}$ （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））
 - $R = \text{LAST_TO_END}$ （客户最近一次乘坐公司距观测窗口结束的月数=最后一次）
 - $F = \text{FLIGHT_COUNT}$ （观测窗口内的飞行次数）
 - $M = \text{SEG_KM_SUM}$ （观测窗口的总飞行里程）
 - $C = \text{AVG_DISCOUNT}$ （平均折扣率）

6. 模型构建

- **客户聚类**：利用 K-Means 聚类算法对客户数据进行客户分群（根据业务理解和需要，分析与讨论后，确定客户类别数量）
- **模型优化**

7. 模型应用

- **会员的升级与保级(积分兑换原理相同)**

会员可以分为，钻石，白金，金卡，银卡...。部分客户会因为不了解自身积分情况，错失升级机会，客户和航空公司都会有损失，在会员接近升级前，对高价值客户进行促销活动，刺激他们消费达到标准，双方获利。

- **交叉销售**

通过发行联名卡与非航空公司各做，使得企业在其他企业消费过程中获得本公司的积分，增强与本公司联系，提高忠诚度。

- **管理模式**

- 企业要获得长期的丰厚利润，必须需要大量稳定的、高质量的客户。
- 维持老客户的成本远远低于新客户，保持优质客户是十分重要的。
- 精准营销中，也有成本因素，所以按照客户价值排名，进行优先的、特别的营销策略，是维持客户的关键。

8. 小结

本文结合航空公司客户价值案例的分析，重点介绍了数据挖掘算法中 K-Means 聚类算法的应用。针对传统 RFM 模型的不足，结合案例进行改造，设定了五个指标的 LRFMC 模型。最后通过聚类的结果，选出客户价值排行，并且制定相应策略。