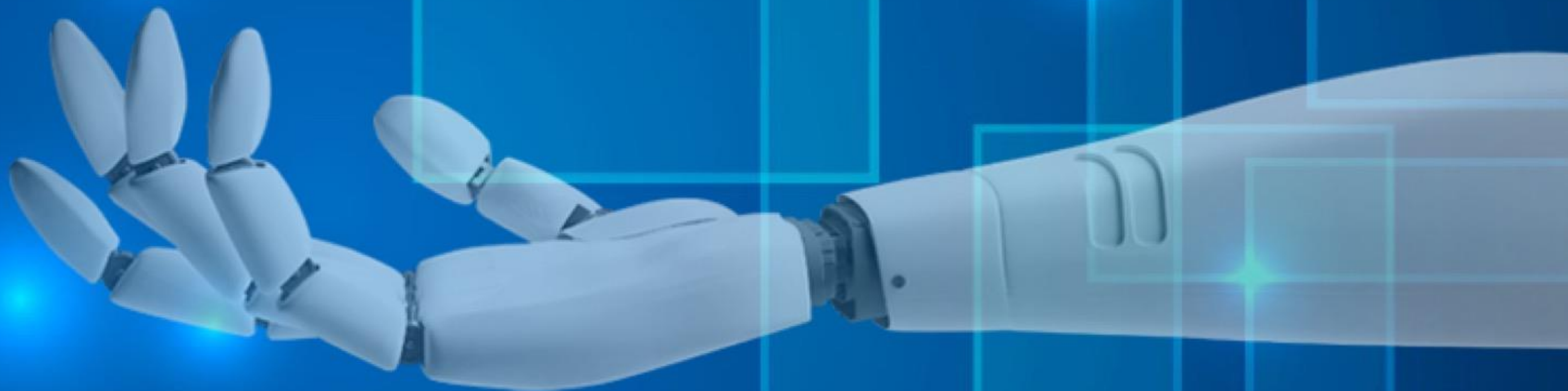


航空公司客户价值分析

需求分析



航空公司客户 价值分析

需求分析

达内教育研究院

目录

了解航空公司现状与客户价值分析

预处理航空客户数据

使用K-Means算法进行客户分群

小结

分析航空公司现状

1. 行业内竞争

民航的竞争除了三大航空公司之间的竞争之外，还将加入新崛起的各类小型航空公司、民营航空公司，甚至国外航空巨头。

航空产品生产过剩，产品同质化特征愈加明显，于是航空公司从价格、服务间的竞争逐渐转向对客户

的竞争。



分析航空公司现状

2. 行业外竞争

随着高铁、动车等铁路运输的兴建，航空公司受到巨大冲击。



- 信息时代的来临使得企业营销的角点从产品中心转向客户中心，客户关系管理称为企业的核心问题，客户关系管理中的关键问题就是客户分类，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，采取不同的营销策略，将有限的营销资源用在高价值的客户身上，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

分析航空公司现状

航空公司数据特征说明

- 目前航空公司已积累了大量的会员档案信息和其乘坐航班记录。
- 以2014-03-31为结束时间，选取宽度为两年的时间段作为分析观测窗口，抽取观测窗口内有乘机记录的所有客户的详细数据形成历史数据，44个特征，总共62988条记录。数据特征及其说明如右表所示。

	特征名称	特征说明
客户基本信息	MEMBER_NO	会员卡号
	FFP_DATE	入会时间
	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

分析航空公司现状

开门见山

表 名	特征名称	特征说明
乘机信息	FLIGHT_COUNT	观测窗口内的飞行次数
	LOAD_TIME	观测窗口的结束时间
	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
	SUM_YR	观测窗口的票价收入
	SEG_KM_SUM	观测窗口的总飞行公里数
	LAST_FLIGHT_DATE	末次飞行日期
	AVG_INTERVAL	平均乘机时间间隔
	MAX_INTERVAL	最大乘机间隔
积分信息	EXCHANGE_COUNT	积分兑换次数
	EP_SUM	总精英积分
	PROMOPTIVE_SUM	促销积分
	PARTNER_SUM	合作伙伴积分
	POINTS_SUM	总累计积分
	POINT_NOTFLIGHT	非乘机的积分变动次数
	BP_SUM	总基本积分

分析航空公司现状

- 结合目前航空公司的数据情况，可以实现以下目标。
 - 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，比较不同类别客户的客户价值。
 - 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。

了解客户价值分析

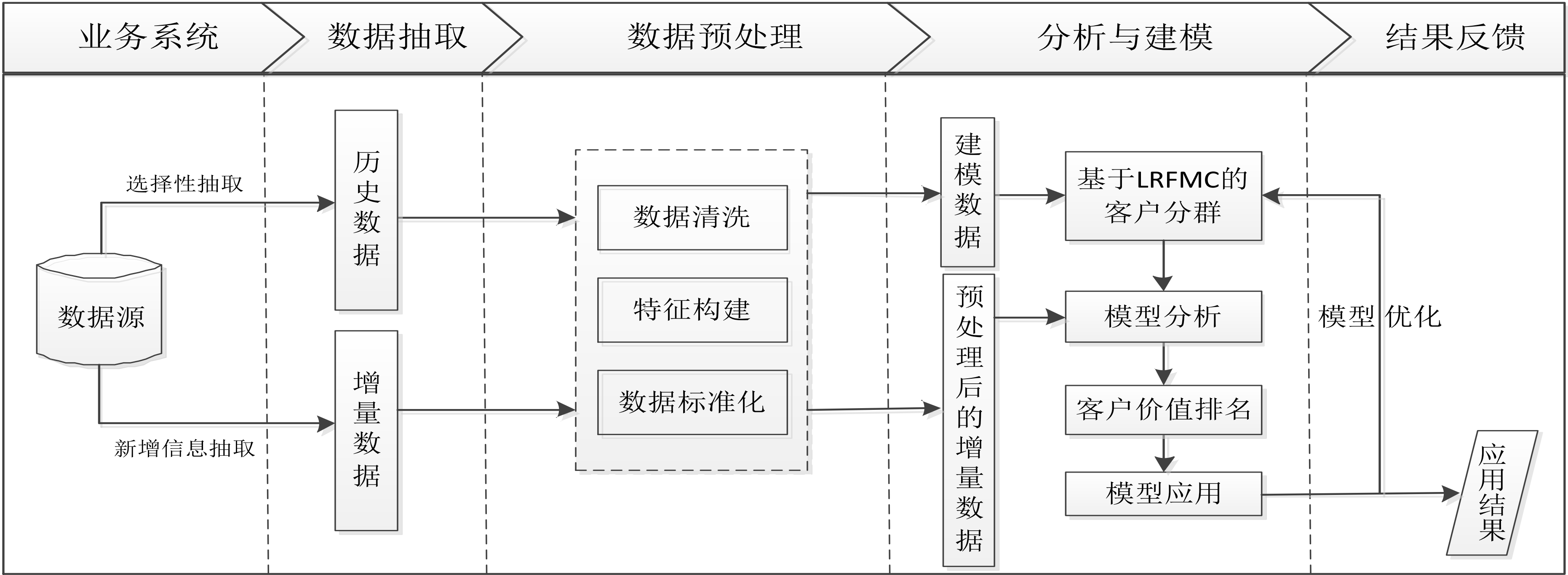
客户营销战略倡导者Jay & Adam Curry从国外数百家公司进行了客户营销实施的经验中提炼了如下经验

- 公司收入的80%来自顶端的20%的客户。
- 20%的客户其利润率100%。
- 90%以上的收入来自现有客户。
- 大部分的营销预算经常被用在非现有客户上。
- 5%至30%的客户在客户金字塔中具有升级潜力。
- 客户金字塔中客户升级2%，意味着销售收入增加10%，利润增加50%。

这些经验也许并不完全准确，但是它揭示了新时代客户分化的趋势，也说明了对客户价值分析的迫切性和必要性。

熟悉航空客户价值分析的步骤与流程

航空客户价值分析项目的总体流程如图所示。



构建航空客户价值分析的关键特征

1. RFM模型介绍

本项目的目标是客户价值分析，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RFM模型。

- R (Recency) 指的是最近一次消费，意指上一次购买的距离截止时间的的时间间隔
- 理论上，上一次消费时间越近的顾客应该比较好的顾客，对提供即时的商品或是服务也最有可能会有反应。营销人员若想业绩有所成长，只能靠偷取竞争对手的市场占有率，而如果要密切地注意消费者的购买行为，那么最近的一次消费就是营销人员第一个要利用的工具。历史显示，如果我们能让消费者购买，他们就会持续购买。这也就是为什么，0至6个月的顾客收到营销人员的沟通信息多于31至36个月的顾客。

构建航空客户价值分析的关键特征

1. RFM模型介绍

消费报告是维系顾客的一个重要指标。买过你的商品、服务或是曾经光顾你商店的消费者，是最有可能再向你购买东西的顾客。再则，要吸引一个几个月前才上门的顾客购买，比吸引一个一年多以前来过的顾客要容易得多。营销人员如接受这种强有力的[营销哲学](#)——与顾客建立长期的关系而不仅是卖东西，会让顾客持续保持往来，并赢得他们的忠诚度。

构建航空客户价值分析的关键特征

1. RFM模型介绍

本项目的目标是客户价值分析，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RFM模型。

- F (Frequency) 指顾客在某段时间内所消费的次数。可以说消费频率越高的顾客，也是满意度越高的顾客，其忠诚度也就越高，顾客价值也就越大。
- M (Monetary) 指顾客在某段时间内所消费的金额。消费金额越大的顾客，他们的消费能力自然也就越大，这就是所谓“20%的顾客贡献了80%的销售额”的二八法则。

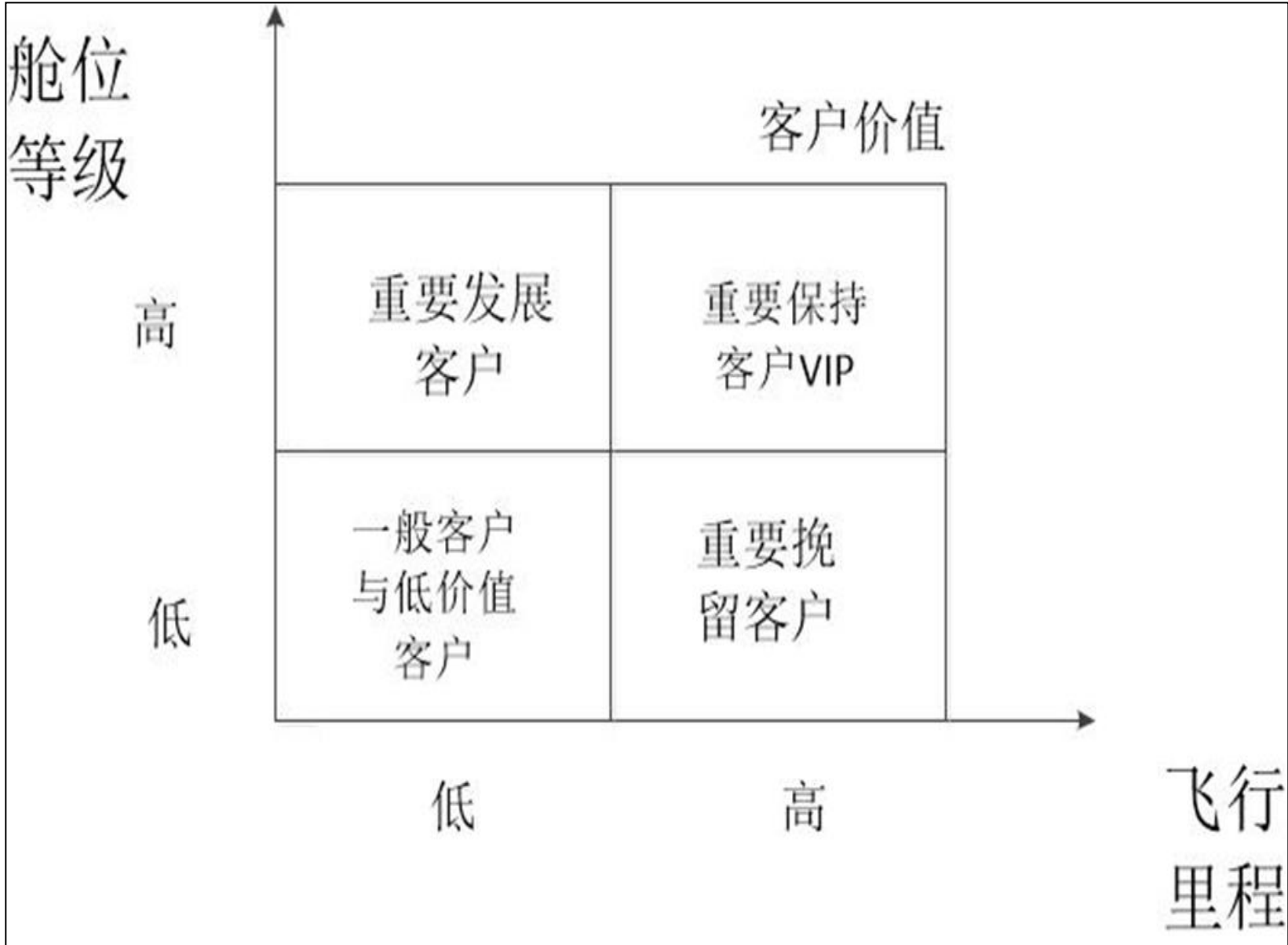
构建航空客户价值分析的关键特征

3. 传统RFM模型在航空行业的缺陷

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，

eg：一位购买长航线，低等级舱位的旅客，和一位短航线，高等级舱位的旅客，后者客户价值更大

因此这个特征并不适合用于航空公司的客户价值分析。



构建航空客户价值分析的关键特征

4. 航空客户价值分析的LRFMC模型

本项目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两个特征代替消费金额。此外，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特征。

本项目将客户关系长度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作为航空公司识别客户价值的关键特征，记为LRFMC模型。

模型	L	R	F	M	C
航空公司 LRFMC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

处理数据缺失值与异常值

- 航空公司客户原始数据存在少量的缺失值和异常值，需要清洗后才能用于分析。
- 通过对数据观察发现原始数据中存在：
 - 1. 票价为空值，票价为空值的数据可能是客户不存在乘机记录造成。
 - 处理方法：丢弃票价为空的记录。
 - 2. 票价为0，折扣率不为0，总飞行公里数大于0的记录。 可能是客户乘坐0折机票或者积分兑换造成。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。
 - 处理方法：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。

标准化LRPMC五个特征

完成五个特征的构建以后，对每个特征数据分布情况进行分析，五个特征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，需要对数据做标准化处理。

了解K-Means聚类算法

1. 基本概念

K-Means聚类算法是一种基于质心的划分方法，输入聚类个数 k ，以及包含 n 个数据对象的数据库，输出满足误差平方和最小标准的 k 个聚类。算法步骤如下。

- 从 n 个样本数据中随机选取 k 个对象作为初始的聚类中心。
- 分别计算每个样本到各个聚类质心的距离，将样本分配到距离最近的那个聚类中心类别中。
- 所有样本分配完成后，重新计算 k 个聚类的中心。
- 与前一次计算得到的 k 个聚类中心比较，如果聚类中心发生变化，转(2)，否则转(5)。
- 当质心不发生变化时停止并输出聚类结果。

了解K-Means聚类算法

2. 数据类型

K-Means聚类算法是在数值类型数据的基础上进行研究，然而数据分析的样本复杂多样，因此要求不仅能够对特征为数值类型的数据进行分析，还要适应数据类型的变化，对不同特征做不同变换，以满足算法的要求。

了解K-Means聚类算法

```
array([[ 0.04656257, -0.00199113, -0.23014218, -0.23446473,  2.1786552 ],
       [ 0.48365797, -0.79940021,  2.4831749 ,  2.42445945,  0.30923797],
       [-0.31336805,  1.68669161, -0.57393574, -0.53678267, -0.17460843],
       [ 1.16084862, -0.37737722, -0.0866405 , -0.09455513, -0.15659965],
       [-0.70031334, -0.41503559, -0.16089813, -0.16064611, -0.25672317]])
```

```
# 统计不同类别样本数目
r1=pd.Series(kmeans_model.labels_).value_counts()
r1
```

```
4    24638
3    15735
2    12119
1     5337
0     4215
dtype: int64
```

针对聚类结果进行特征分析，如图所示。

图表标题

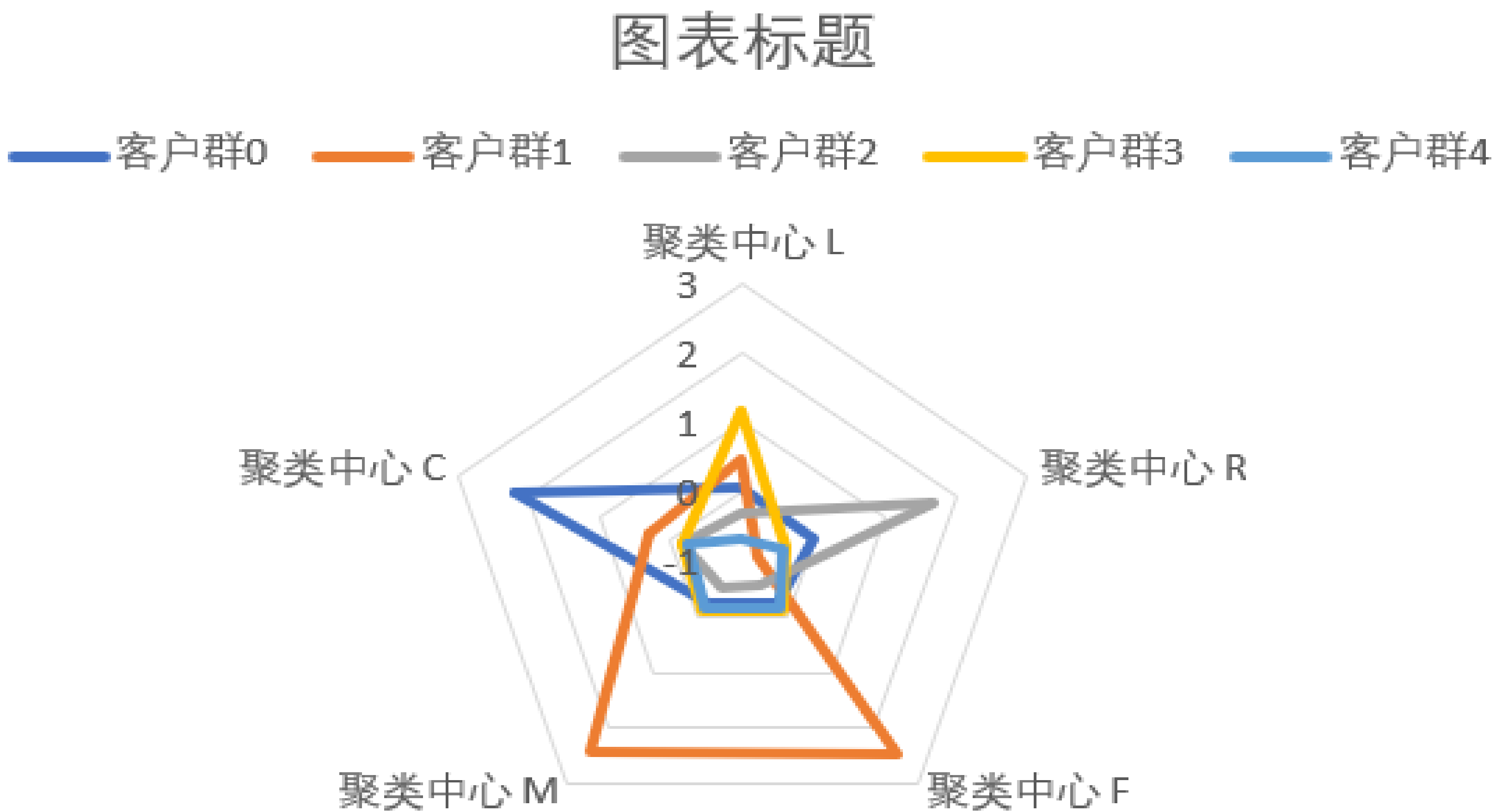
客户群0 客户群1 客户群2 客户群3 客户群4

聚类中心 L 聚类中心 R 聚类中心 F 聚类中心 M 聚类中心 C

客户群	聚类中心 L	聚类中心 R	聚类中心 F	聚类中心 M	聚类中心 C
客户群0	0.5	1.5	1.0	1.0	2.0
客户群1	0.5	1.0	3.0	3.0	0.5
客户群2	0.5	2.0	0.5	0.5	0.5
客户群3	1.0	0.5	0.5	0.5	0.5
客户群4	0.5	0.5	0.5	0.5	0.5

分析聚类结果

- 针对聚类结果进行特征分析，如图所示。



分析聚类结果

- 结合业务分析，通过比较各个特征在群间的大小对某一个群的特征进行评价分析，从而总结出每个群的优势和弱势特征，具体结果如表所示。

客户群体	优势特征	客户群体	营销策略
客户群0	C , R	重要挽留	潜在大客户，需要挽留
客户群1	F,M,R	重点价值客户	优质客户需要保持
客户群2	无	一般客户	贡献不大，维持即可
客户群3	L , R	重要发展	进一步挖掘客户价值
客户群4	R	一般保持	进一步推广

以理服人

小节

本项目结合航空公司客户价值分析的案例，重点介绍了数据分析算法中K-Means聚类算法在客户价值分析中的应用。针对RFM客户价值分析模型的不足，使用K-Means算法构建了航空客户价值分析LRFMC模型，详细描述了数据分析的整个过程。

谢谢