

# 航空公司客户价值分析



# 目录

---



# 背景

---

- 目前，全球经济环境和市场环境正在悄然发生改变，企业的业务也在逐步由产品为主导向客户需求为主导转型。一种全新的“以客户为中心”的业务模式正在形成并被提升到前所未有的高度。
- 随着中国社会经济的发展，我国民航已经从卖方市场转变为买方市场，再加上铁路提速和全国公路网日益完善，不同运输方式间的竞争日趋激烈，航空公司面临的压力越来越大。
- 航空公司属于典型的服务行业，其目的是获取更多的利润，然而当今航空公司产品同质化现象严重，并且竞争激烈，客户资源成为航空公司最为短缺的资源，谁拥有的客户资源多，谁的核心竞争力就强，获得的利润就会更多。

# 背景

---

客户营销战略的倡导者Jay Curry& 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%的客户带来的利润率能够达到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%。

- 虽然这些经验也许并不完全准确，但是客户作为航空公司的主要利润来源，要求航空公司不断地认识、发现、开发和满足客户的需求，与客户建立一种稳定的客户关系。这一现实情况揭示了新时代客户分化的趋势，而这种趋势说明了进行客户价值分析的必要性。目前各航空公司都已认识到这一关键因素，并且将客户价值分析作为公司发展战略之一。
- 本案例将在航空公司客户数据的基础上，建立合理的客户价值模型，对客户进行分群，分析比较不同客户群的客户价值，为企业提供更精准的策略依据，帮助企业制定更加符合市场行情和企业现状的营销策略，为企业带来更多的利润。

# 数据说明

- 目前该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年3月31日为结束时间抽取两年内有乘机记录的所有客户的详细数据。
- 数据包含**会员卡号、入会时间、性别、年龄、会员卡级别、工作地城市、工作地所在省份、工作地所在国家、观测窗口结束时间、观测窗口乘机积分、飞行公里数、飞行次数、飞行时间、乘机时间间隔、平均折扣率**等特征。

	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

# 数据说明

	特征名称	特征说明
客户基本信息	WORK_PROVINCE	工作地所在省份
	WORK_COUNTRY	工作地所在国家
	AGE	年龄
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

# 数据说明

	特征名称	特征说明
乘机信息	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



# 分析目标

---

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，对客户进行分群。
- 对不同的客户类别进行特征分析，比较不同客户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。

# 分析目标

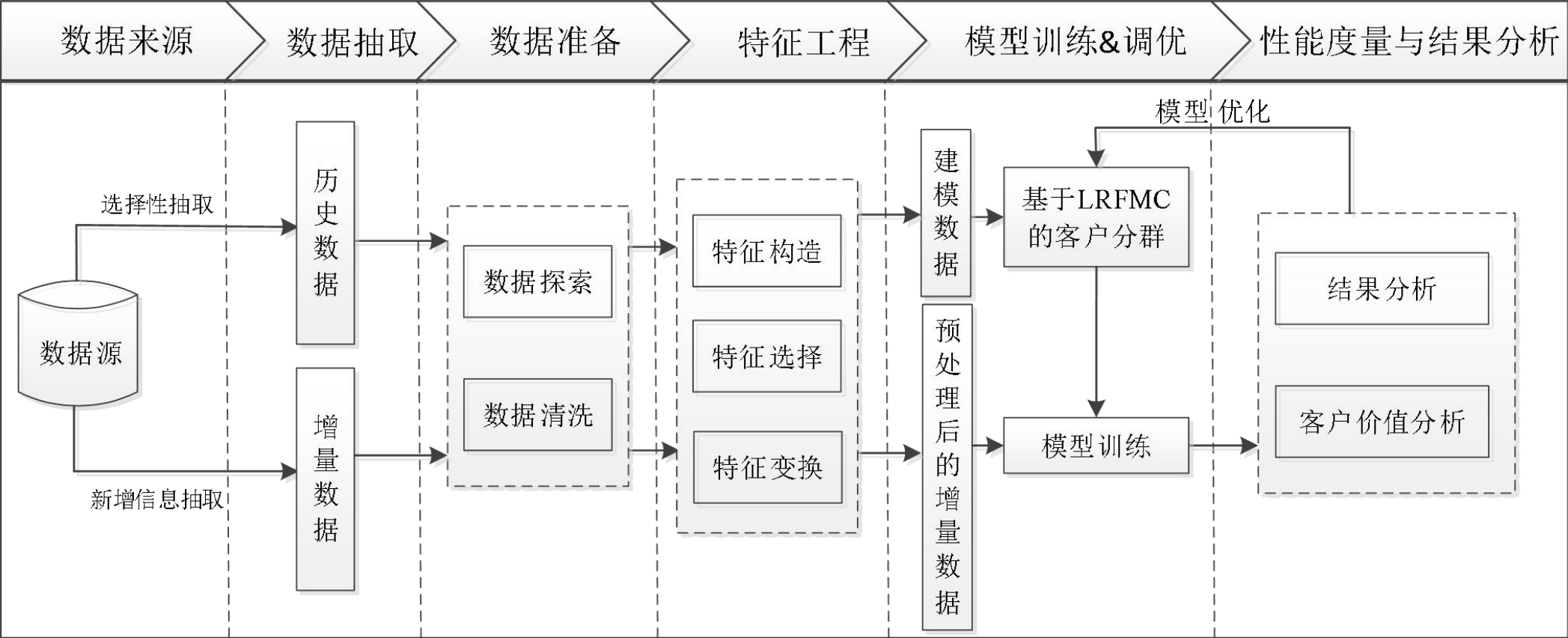
---

本案例的总体流程主要包括以下4个步骤。

- 抽取航空公司2012年4月1日至2014年3月31日的数据。
- 对抽取的数据进行数据清洗、特征构建和标准化等操作。
- 基于RFM模型，使用K-Means算法进行客户分群。
- 针对模型结果得到不同价值的客户，采用不同的营销手段，提供定制化的服务。

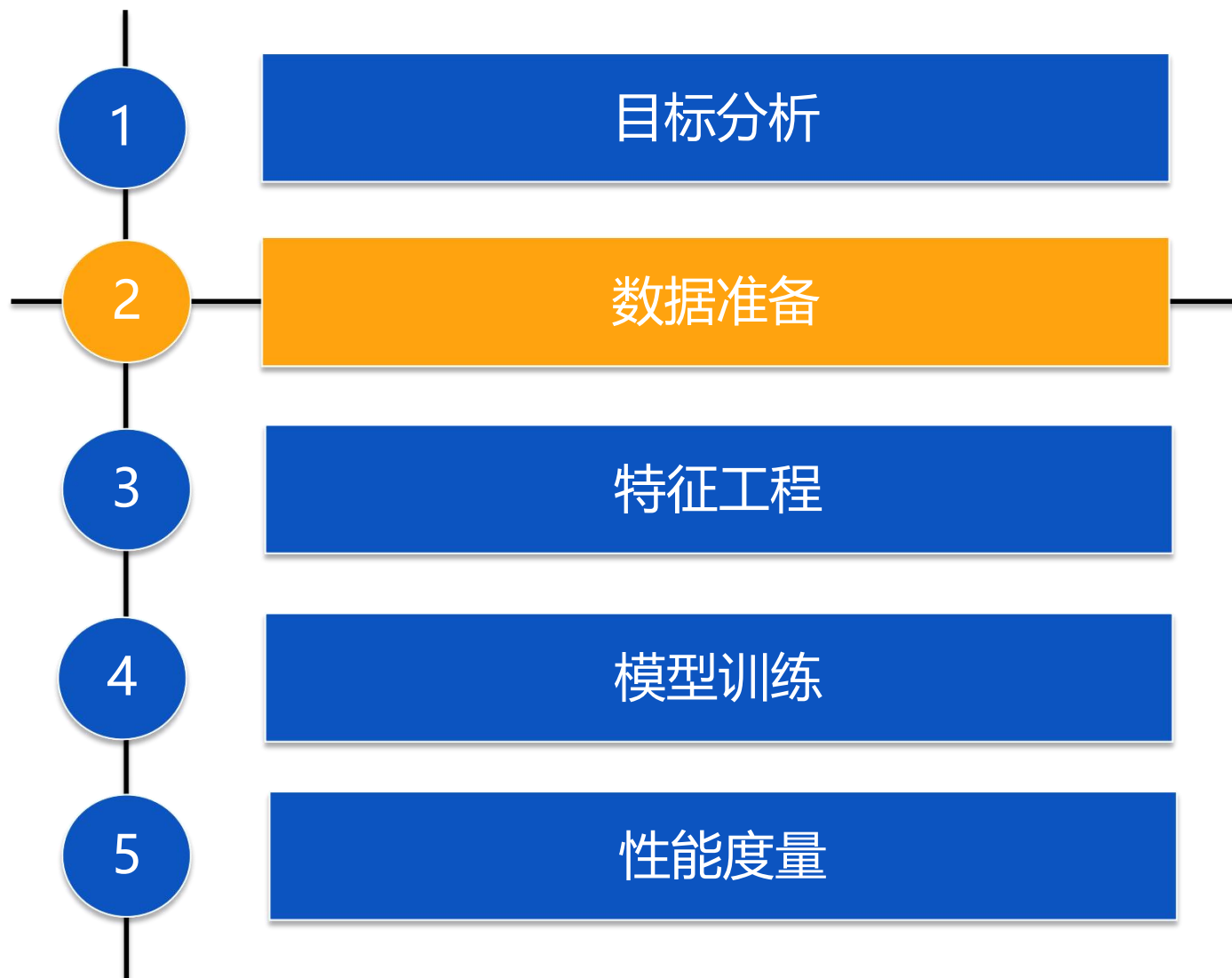
# 分析目标

航空公司客户价值分析总体流程如下图所示。



# 目录

---



# 数据准备

---

国内某航空公司提供的原始数据尚未经过处理，数据质量可能尚未达到可以直接用于建模的程度，可能存在缺失值、异常值等问题，这些问题会导致建立的模型不够精确，为尽可能地排除干扰因素、保证模型的可靠性，需要进行必要的**数据准备**。

# 数据探索

---

- 由于航空公司客户乘机记录信息数据量很大，因此在获取数据时，对原始数据进行截取，以2014年3月31日为结束时间，选取宽度为两年（2012年4月1日~2014年3月31日）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，总共62988条记录。
- 截取数据后对数据进行探索，在数据探索过程中，发现数据中存在缺失值和异常值，具体表现为票价为空值、或票价为0、或折扣率为0、或总飞行公里数为0的记录。这些值的存在会对模型的建立产生不可忽视的影响，如模型的不确定性会变得更加显著，其蕴含的规律更难把握，甚至导致模型输出的结果不可靠，因此在数据清洗过程中要重点对这两类数据进行处理。

# 数据清洗

---

在数据清洗过程中，因为原始数据量大，缺失值和异常值在数据集中占比较小，所以需要对缺失值和异常值均进行删除处理，即丢弃票价为0，或平均折扣率为0，或总飞行公里数为0的记录。

# 目录

---





# 特征工程

---

- 在建模之前需要对原始数据特征进行处理，筛选出更好的特征，才能获取更好的训练数据，让建立的模型能够得到更加精确的结果。
- 特征构造是指从原始数据中人工的构建一些具有实际意义的特征，本案例借助RFM模型进行特征构造。

# 特征构造

---

## 1. RFM模型

RFM模型是识别客户价值应用较为广泛的模型，RFM模型具体的含义如下。

### ➤ R (Recency)

- R (Recency) 指的是**最近一次消费时间与截止时间的间隔**，简称**时间间隔**。通常情况下，最近一次消费时间与截止时间的间隔越短，对即时提供的商品或是服务也最有可能感兴趣。这也是消费时间间隔0至6个月的顾客收到的沟通信息多于1年以上的顾客的原因。
- 最近一次消费时间与截止时间的间隔不仅能够为确定促销客户群体提供依据，还能够从中得出企业发展的趋势。如果分析报告显示最近一次消费时间很近的客户在增加，则表示该公司是个稳步上升的公司。反之，最近一次消费时间很近的客户越来越少，则说明该公司需要找到问题所在，及时调整营销策略。

# 特征构造

---

## ➤ F (Frequency)

- F (Frequency) 指顾客在**某段时间内所消费的次数**，简称**消费频率**。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。增加顾客购买的次数意味着从竞争对手处争取市场占有率，赚取营业额。商家需要做的是通过各种营销方式，去不断地刺激顾客消费，提高他们的消费频率，提升顾客的复购率。

# 特征构造

---

## ➤ M (Monetary)

- M (Monetary) 指**顾客在某段时间内所消费的金额**，简称**消费金额**。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。而这批顾客也必然是商家在进行营销活动时需要特别照顾的群体，尤其是在商家前期资源不足的时候。不过需要注意一点，不论采用哪种营销方式，以不对顾客造成骚扰为大前提，否则营销只会产生负面效果。
- 在RFM模型理论中，时间间隔、消费频率、消费金额是判别客户价值最重要的特征，这3个特征对营销活动的具有十分重要的意义，其中，时间间隔是最有力的特征。

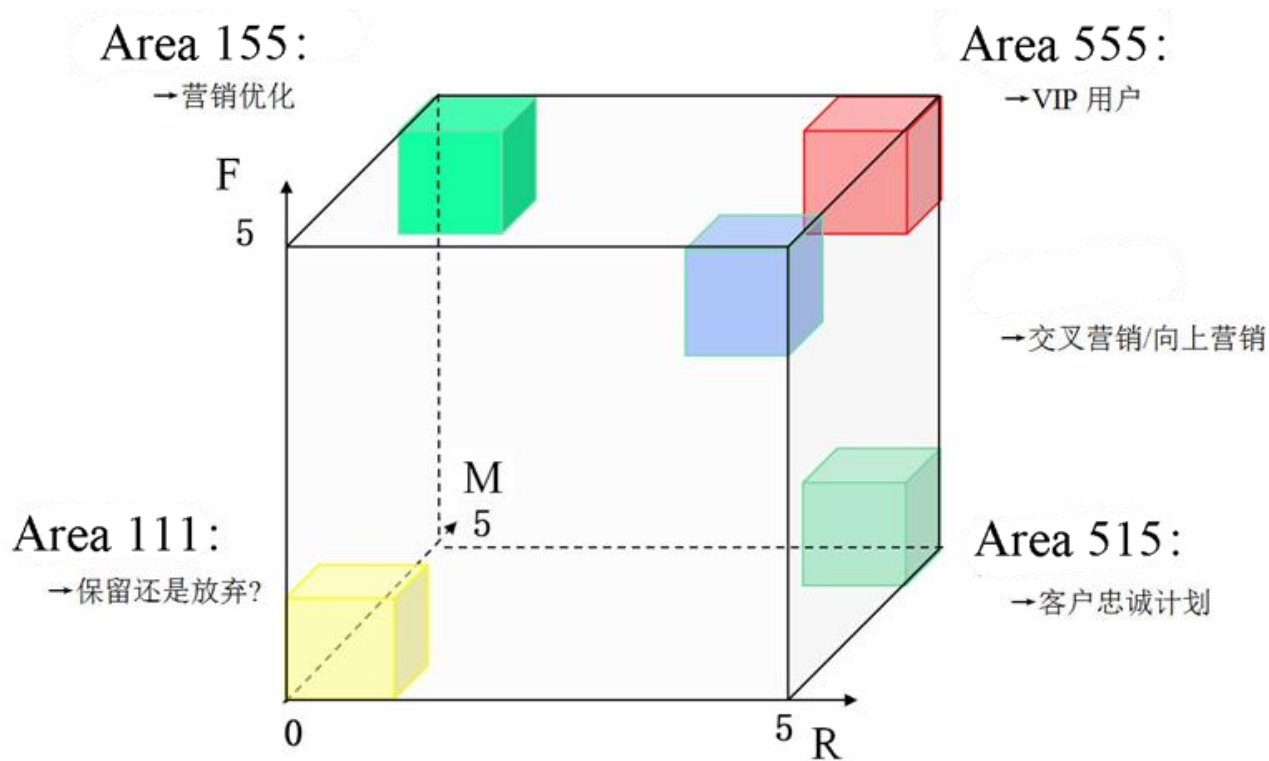
# 特征构造

## 2. RFM模型解读

- RFM模型包括3个特征，无法用平面坐标图来展示，所以这里使用三维坐标系进行展示，X轴表示Recency，Y轴表示Frequency，Z轴表示Monetary，每个轴一般会分成5级表示程度，1为最小，5为最大。需要特别说明的是，R特征，在X轴上R值越大代表该类客户最近一次消费与截止时间的消费间隔越短，客户R维度上的质量越好。
- X轴表示R特征，Y轴表示F特征，Z轴表示M指标，每个轴上划分5等级，等价于将客户划分为 $5 \times 5 \times 5 = 125$ 。这里划分为5级并不是严格的要求，一般是根据实际研究需求和顾客的总量进行划分，对于是否等分的问题取决于该维度上客户的分布规律。

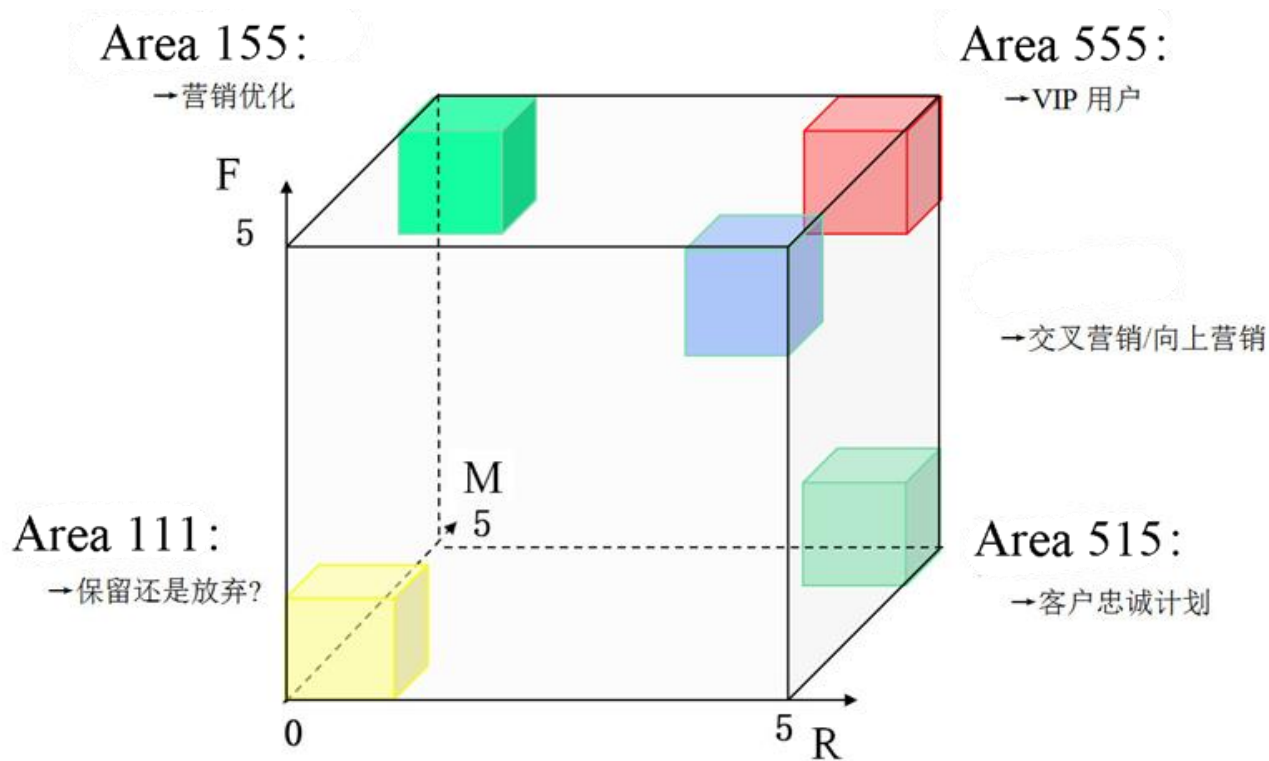
# 特征构造

➤ 通过右图可以看出，左上角方框的客户RFM特征取值为155。消费的近度R值是比较小的，说明该类客户最近都没有来店消费，原因可能是因为最近比较忙或对现有的产品或服务不满意，或是找到了更好的商家。R特征数值变小需要企业管理人员引起重视，说明该类客户可能流失，对企业造成损失。消费频率F很高，说明客户很活跃，经常到商家消费。消费金额M值很高，说明该类客户是具备一定的消费能力，为店里贡献了很多的营业额。这类型客户总体分析比较优质，但是R特征时间近度值较小，其往往是需要针对进行营销优化的客户群体。



# 特征构造

- 同理，若客户RFM特征取值为555。则可以判定该客户为最优质客户，即该类客户最近有到商家消费，消费频率很高，消费金额很大，该类客户往往是企业利益的主要贡献者，是需要重点关注与维护的客户。



# 特征构造

---

## 3. 特征构造

- 在RFM模型中，消费金额表示客户在一段时间内购买该企业产品金额的总和。然而航空票价受到多种因素（如距离、舱位等级）的影响，因此消费同样金额的不同客户对航空公司的价值可能是不同的，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RFM模型中的消费金额这一特征并不适用于航空公司客户价值分析。
- 本案例在RFM模型的基础上，选择客户在一定时间内累积的飞行里程M、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。同时，因为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。



# 特征选择

本案例选择**客户关系长度L**、**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特征，记为LRFMC模型。

L	R	F	M	C
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

# 特征选择

根据航空公司客户价值LRFMC模型，选择与LRFMC特征相关的6个特征：FFP\_DATE、LOAD\_TIME、FLIGHT\_COUNT、avg\_discount、SEG\_KM\_SUM、LAST\_TO\_END。

LOAD_TIME	FFP_DATE	LAST_TO_END	FLIGHT_COUNT	SEG_KM_SUM	AVG_DISCOUNT
2014/3/31	2013/3/16	23	14	126850	1.02
2014/3/31	2012/6/26	6	65	184730	0.76
2014/3/31	2009/12/8	2	33	60387	1.27
2014/3/31	2009/12/10	123	6	62259	1.02
2014/3/31	2011/8/25	14	22	54730	1.36
2014/3/31	2012/9/26	23	26	50024	1.29

# 特征变换

---

由于选取的L、R、F、M、C特征在原始数据中并没有直接给出，因此需要根据原始数据特征进行特征变换得到需要的特征。

- $L = \text{观测窗口的结束时间} - \text{入会时间} = \text{LOAD\_TIME} - \text{FFP\_DATE}$
- $R = \text{最后一次乘机时间至观察窗口末端时长} = \text{LAST\_TO\_END}$
- $F = \text{观测窗口的飞行次数} = \text{FLIGHT\_COUNT}$
- $M = \text{观测窗口总飞行公里数} = \text{SEG\_KM\_SUM}$
- $C = \text{平均折扣率} = \text{avg\_discount}$

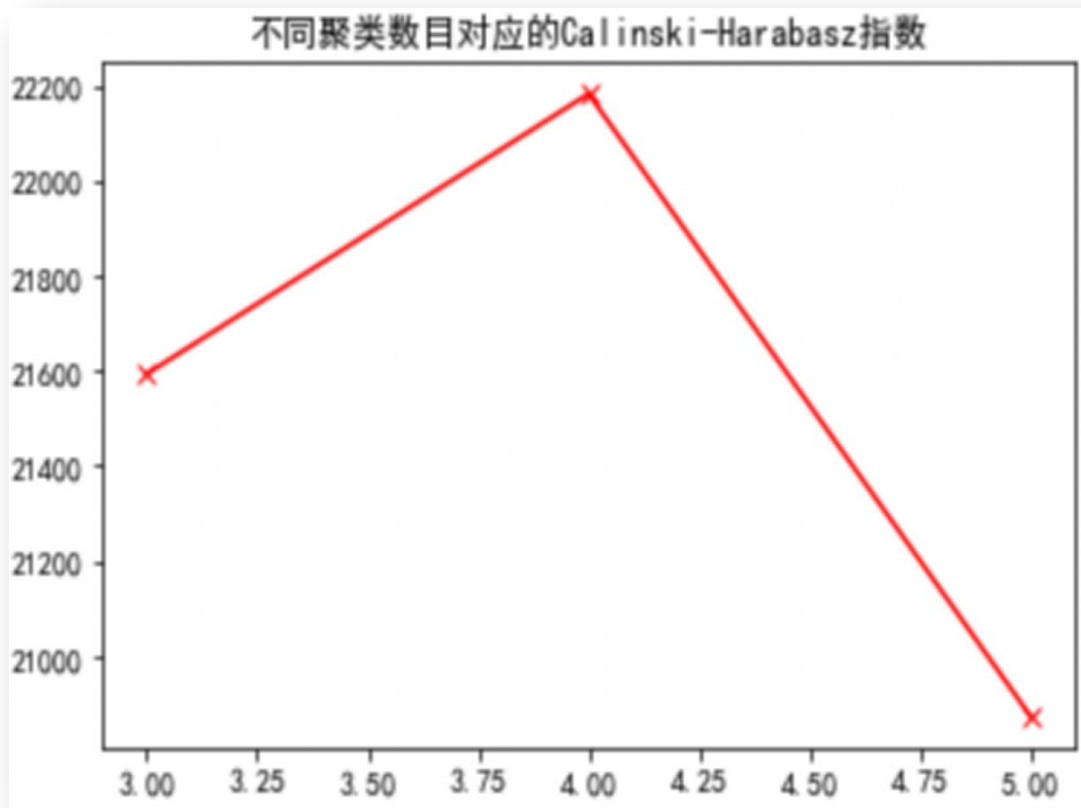
# 目录

---



# 模型训练

采用K-Means聚类算法对航空公司客户进行分群，K-Means聚类需要预先给出 $k$ 值，即需要事先指定聚类数目，本案例根据对业务的理解与分析结合Calinski-Harabasz指数确定聚类数目 $k$ ，Calinski-Harabasz指数越大表示聚类效果越好，在建模过程中 $k$ 取Calinski-Harabasz指数最大值对应的聚类数。



# 目录

---



# 性能度量

---

- 本案例中K-Means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在变化，航空公司客户的数据信息也在变化。
- 因此，考虑业务的实际情况，建议每个月运行一次该模型，通过聚类判断新增加的客户所属的客户群，同时分析新增客户特征的价值。
- 如果新增加客户数据的实际情况与判断结果差异较大，那么需要业务部门重点关注，查看出现差异的原因并确认模型的稳定性，如果模型稳定性变化较大，那么需要重新训练聚类模型。

# 结果分析

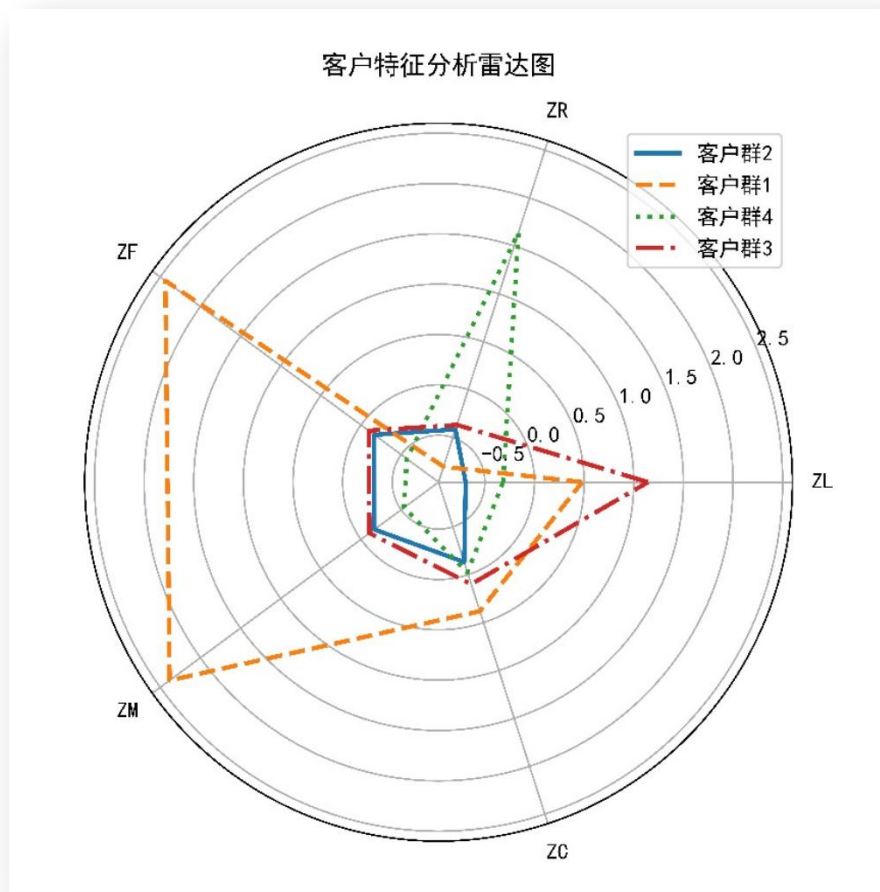
客户分群聚类结果如表所示。

聚类类别	聚类中心				
	ZL	ZR	ZF	ZM	ZC
客户群1	0.4793	-0.7960	2.4384	2.3832	0.3819
客户群2	-0.3092	1.6590	-0.5721	-0.5359	-0.0183
客户群3	1.1388	-0.3667	-0.0957	-0.1049	0.0957
客户群4	-0.6961	-0.4085	-0.1696	-0.1698	-0.1343



# 结果分析

根据聚类结果绘制雷达图，如图所示。



## 结果分析

---

- 客户群1在特征F、M处的值最大，在特征R处的值最小，说明客户群1消费频率和累积飞行里程较大，消费时间间隔较小，这类客户需要航空公司重点关注并保持良好关系，称F、M、R特征为客户群1的优势特征。
- 客户群2在L、C特征处的值最小，说明客户群2入会时间较短、享受的平均折扣系数较小。
- 客户群3在R特征处的值最大，在F、M特征处的值最小，说明客户群3消费频率和累积飞行里程较小，消费时间间隔较大，这类客户可能只有在机票打折的时候才会乘坐航空公司航班，称F、M、R特征为客户群3的弱势特征。
- 客户群4在特征L处的值最大，说明客户群4入会时间较长。

# 结果分析

每个客户群的优势特征和弱势特征总结如表所示，其中正常字体表示最大值，加粗字体表示次大值，斜体字体表示最小值。

群类别	优势特征			弱势特征		
客户群1	F	M	<i>R</i>			
客户群2				<i>L</i>	<i>C</i>	
客户群3				<i>F</i>	<i>M</i>	R
客户群4	L	<b>F</b>	<b>M</b>			

# 客户价值分析

---

## ➤ 重要保持客户

- 平均折扣系数 $C$ 较高，时间间隔 $R$ 较低，消费频率（乘机次数） $F$ 或累积飞行里程 $M$ 较高。这类客户是航空公司最理想的客户，他们为航空公司带来了大部分的利润，但是这类客户占的比例比较小，航空公司应该优先考虑将营销资源投入到这类客户，尽量使这类客户能够保持在公司的高质量消费。

# 客户价值分析

---

## ➤ 重要挽留客户

- 在过去的时间里平均折扣系数 $C$ ，消费频率（乘机次数） $F$ 或飞行里程 $M$ 较高，但是时间间隔 $R$ 较大或消费频率（乘机次数） $F$ 逐渐变小。这类客户存在较高的不确定性因素，航空公司应该提高警惕，采取一定的营销手段，延长客户在航空公司的消费周期，否则，这些不确定因素可能导致这类客户的流失。

# 客户价值分析

---

## ➤ 一般与低价值客户

- 平均折扣系数C较低，时间间隔R较高，消费频率（乘机次数）F或飞行里程M较低，会员入会时间L较低。对于这类客户，航空公司不需要过多地关注，因为他们可能只有在机票打折的时候才会乘坐航空公司航班。

# 客户价值分析

- 根据客户类别特征描述，对模型得出的客户群进行客户价值排名，如表所示。
- 针对不同类别的客户群航空公司应该采取不同的营销策略，对不同类型的客户群提供不同的产品和服务，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。

客户群	排名	排名含义
客户群1	1	重要保持客户
客户群2	3	一般客户
客户群3	2	重要挽留用户
客户群4	4	低价值客户

# 小结

---

- 本案例结合航空公司客户的会员乘机记录信息，重点介绍了K-Means聚类算法在客户价值分析中的应用。
- 首先对原始数据进行探索，寻找数据特点，对原始数据进行清洗，处理缺失值和异常值。
- 然后对RFM模型进行改进，构造LRFMC模型，进而构建特征集合。
- 最后根据对业务的理解和分析结合Calinski-Harabasz指数确定最佳聚类数目，利用K-Means聚类算法对航空公司客户分群，对聚类得出的客户群进行特征分析，划分客户类别，并给出一定的策略建议。







# Thank you!