



大数据，成就未来



航空公司客户价值分析

2018/10/23

前言

- 信息时代的来临使得企业营销焦点从产品转向了客户，客户关系管理(Customerrelationship management, CRM)成为企业的核心问题。
- 客户关系管理的关键问题是客户分群。通过客户分群，区分无价值客户和高价值客户。企业针对不同价值的客户制订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。
- 准确的客户分群结果是企业优化营销资源分配的重要依据，客户分群越来越成为客户关系管理中亟待解决的关键问题之一。
- **本章将使用航空公司客户数据，结合RFM模型，采用K-Means聚类算法，对客户进行分群，比较不同类别客户的客户价值，从而制定相应的营销策略。**

本章学习目标

- (1)熟悉航空客户价值分析的步骤和流程。
- (2)了解RFM模型的基本原理。
- (3)掌握K-Means算法的基本原理与使用方法。
- (4)比较不同类别客户的客户价值，制定相应的营销策略。

目录



了解航空公司现状与客户价值分析

任务描述

- 面对激烈的市场竞争，各个航空公司都推出了更多的优惠来吸引客户。国内某航空公司面临着常旅客流失、竞争力下降和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，分析及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服务。

任务分析

- (1)了解航空公司现状。
- (2)认识客户价值分析。
- (3)熟悉航空公司客户价值分析的步骤与流程。

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。



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，总共62988条记录，44个特征。
- 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

航空公司客户数据说明

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

续表

思考

原始数据中包含40多个特征，利用这些特征做些什么呢？我们又该从哪些角度出发呢？



项目目标

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

- 全球经济环境和市场环境已经悄然改变，企业的业务逐步从**以产品为主导**转向**以客户需求为主导**。
- 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。
- 然而与客户保持关系需要花费成本，企业所拥有的客户只有一部分能为企业带来利润。企业的资源也是有限的，忽视高潜力的客户而对所有客户都提供同样的服务，将使企业的资源无法发挥其最大效用去创造最大化的利润。
- 任何企业要想生存和发展，都必须获得利润，追求利润最大化是企业生存和发展的宗旨之一。所以，企业不可能也不应该和所有的客户都保持同样的关系。

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%。

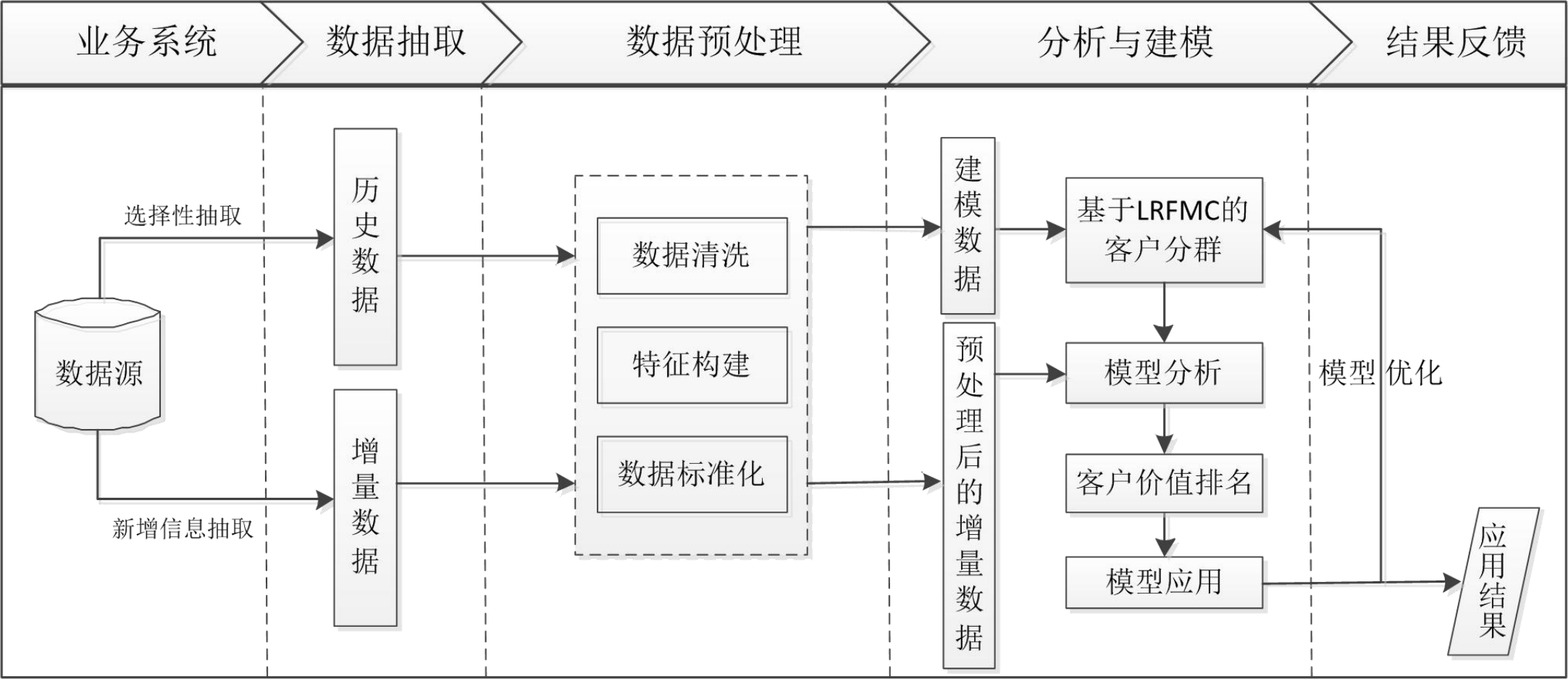
这些经验也许并不完全准确，但是它揭示了新时代客户分化的趋势，也说明了对客户价值分析的迫切性和必要性。

了解客户价值分析

- 如果把客户的盈利性加以分析，就会发现客户盈利结构已发生了重大变化，只有特定的一部分客户给企业带来了利润。企业如果想获得长期发展，必须对这类客户做到有效的识别和管理。如果用同样的方法应对所有与企业有业务往来的客户，必然不会获得成功。
- 众多的企业管理者显然知道客户价值分析的重要性，但对如何进行客户价值分析却知之甚少。如何全方位、多角度地考虑客户价值因素，进行有效的客户价值分析，这是摆在所有企业面前需要认真思考的一个问题。只有甄选出有价值的客户并将精力集中在这些客户身上，才能有效地提升企业的竞争力，使企业获得更大的发展。
- 在客户价值分析领域，最具影响力并得到实证检验的理论与模型有客户终生价值理论、客户价值金字塔模型、策略评估矩阵分析法和RFM客户价值分析模型等。本章将运用改进的客户价值RFM模型进行分析。

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。



目录



预处理航空客户数据

任务描述

- 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。
- 同时，由于原始数据的特征过多，不便直接用于客户价值分析，因此需要对特征进行筛选，挑选出衡量客户价值的关键特征。

对航空客户数据进行预处理可以分为3个步骤

- (1) 处理数据缺失值与异常值
- (2) 结合RFM模型筛选特征
- (3) 标准化筛选后的数据

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。

处理方法：丢弃票价为空的记录。

- 其他的数据可能是客户乘坐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。

处理数据缺失值与异常值

丢弃票价为空的记录

```
import numpy as np
import pandas as pd
airline_data = pd.read_csv("data/air_data.csv",encoding="gb18030")    #导入航空数据
print('原始数据的形状为：',airline_data.shape)
## 去除票价为空的记录
exp1 = airline_data["SUM_YR_1"].notnull()
exp2 = airline_data["SUM_YR_2"].notnull()
exp = exp1 & exp2
airline_notnull = airline_data.loc[exp,:]
print('删除缺失记录后数据的形状为：',airline_notnull.shape)
```

原始数据的形状为： (62988, 44)
删除缺失记录后数据的形状为： (62299, 44)

处理数据缺失值与异常值

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

#只保留票价非零的，或者平均折扣率不为0且总飞行公里数大于0的记录。

```
index1 = airline_notnull['SUM_YR_1'] != 0
```

```
index2 = airline_notnull['SUM_YR_2'] != 0
```

```
index3 = (airline_notnull['SEG_KM_SUM'] > 0) & (airline_notnull['avg_discount'] != 0)
```

```
airline = airline_notnull[(index1 | index2) & index3]
```

```
print('删除异常记录后数据的形状为：',airline.shape)
```

```
删除异常记录后数据的形状为： (62044, 44)
```

构建航空客户价值分析的关键特征

1. RFM模型介绍

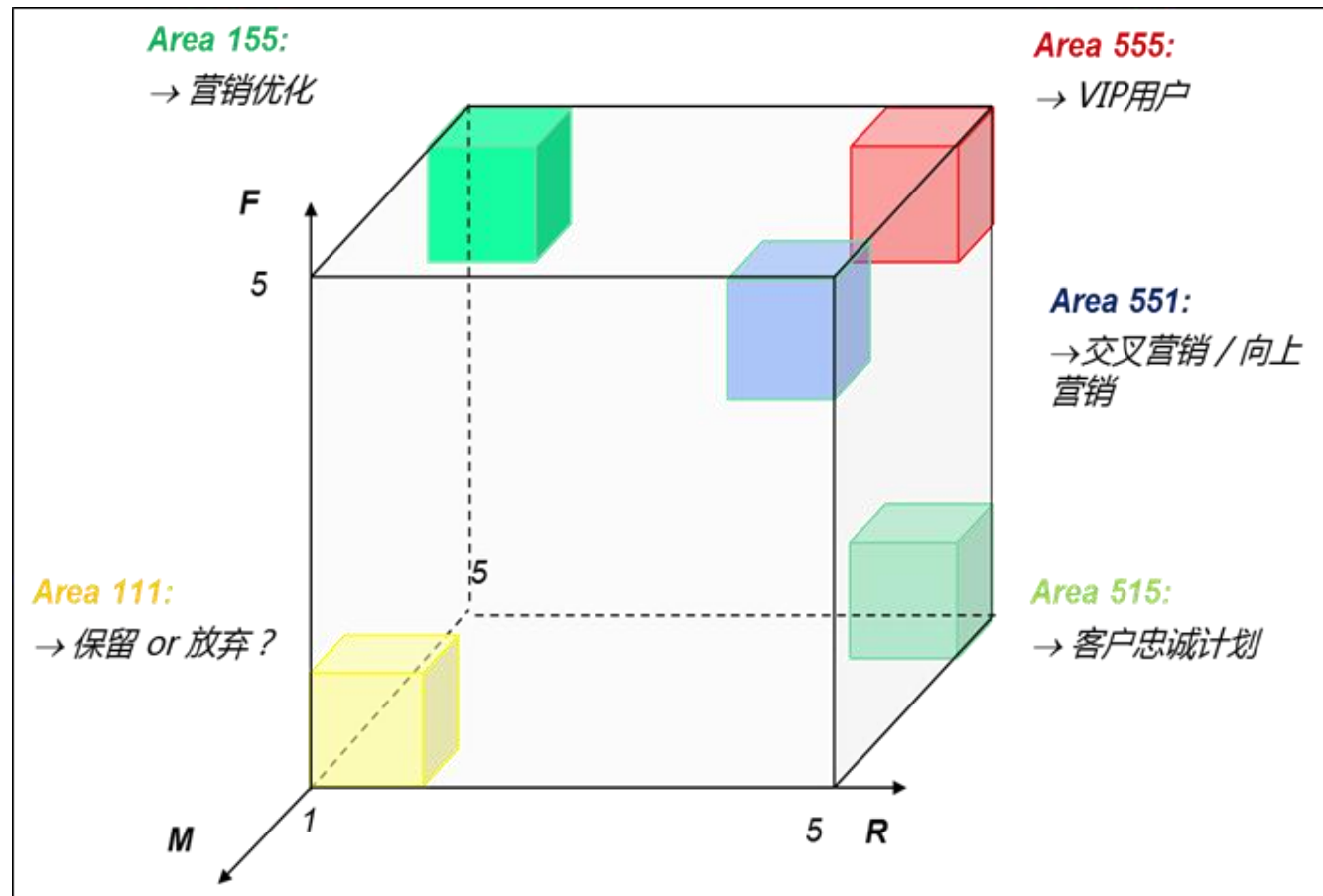
本项目的目标是客户价值分析，即通过航空公司客户数据识别不同价值的客户，**识别客户价值应用最广泛的模型是RFM模型。**

- R (Recency) 指的是最近一次消费时间与截止时间的间隔。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。
- 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

2. RFM模型结果解读

- RFM模型包括三个特征，使用三维坐标系进行展示，如图所示。
- 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。

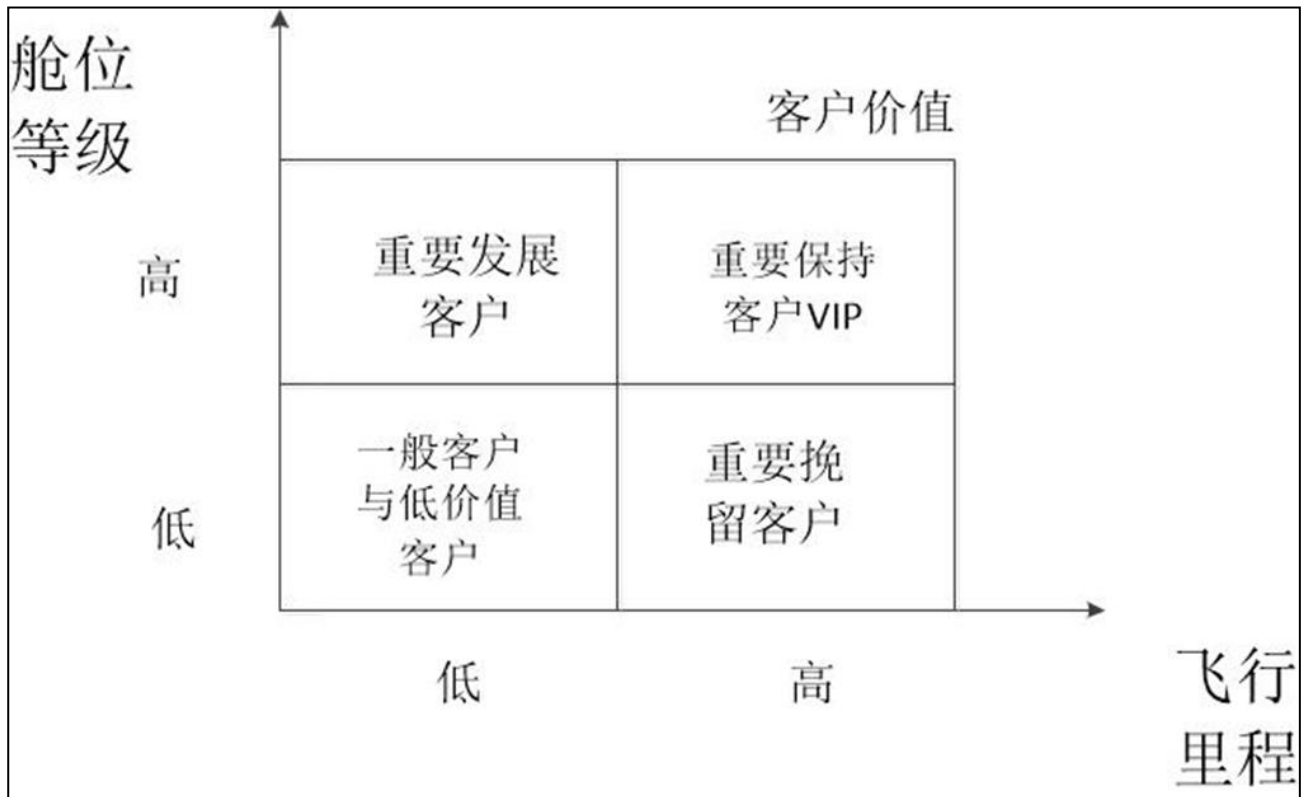


构建航空客户价值分析的关键特征

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。
- 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**这5个特征，作为航空公司识别客户价值的关键特征（如表 3 2所示），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

构建航空客户价值分析的关键特征

表 名	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
乘机信息	AGE	年龄
	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
积分信息	AVG_INTERVAL	平均乘机时间间隔
	MAX_INTERVAL	最大乘机间隔
	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

构建航空客户价值分析的关键特征

表 名	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
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

➤ 由于原始数据中并没有直接给出 LRFCM模型的5个特征，需要通过原始数据提取。

(1) $L = \text{LOAD_TIME} - \text{FFP_DATE}$

(2) $R = \text{LAST_TO_END}$

(3) $F = \text{FLIGHT_COUNT}$

(4) $M = \text{SEG_KM_SUM}$

(5) $C = \text{AVG_DISCOUNT}$

构建航空客户价值分析的关键特征

选取并构建LRFMC模型的5个特征

选取需求特征

```
airline_selection = airline[["FFP_DATE","LOAD_TIME","FLIGHT_COUNT","LAST_TO_END","avg_discount","SEG_KM_SUM"]]
```

构建L特征

```
L = pd.to_datetime(airline_selection["LOAD_TIME"]) - pd.to_datetime(airline_selection["FFP_DATE"])
```

```
L = L.astype("str").str.split().str[0]
```

```
L = L.astype("int")/30
```

合并特征

```
airline_features = pd.concat([L,airline_selection.iloc[:,2:]],axis = 1)
```

```
airline_features.columns = ['L','F','R','C','M']
```

```
print('构建的LRFMC特征前5行为：\n',airline_features.head())
```

```
airline_features = airline_features[['L','R','F','M','C']]    ##调整列的顺序
```

```
print('构建的LRFMC特征前5行为：\n',airline_features.head())
```

构建航空客户价值分析的关键特征

选取并构建LRFMC模型的5个特征

构建的LRFMC特征前5行为：

	L	F	R	C	M
0	90.200000	210	1	0.961639	580717
1	86.566667	140	7	1.252314	293678
2	87.166667	135	11	1.254676	283712
3	68.233333	23	97	1.090870	281336
4	60.533333	152	5	0.970658	309928

	L	R	F	M	C
0	90.200000	1	210	580717	0.961639
1	86.566667	7	140	293678	1.252314
2	87.166667	11	135	283712	1.254676
3	68.233333	97	23	281336	1.090870
4	60.533333	5	152	309928	0.970658

标准化LRFMC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，其数据的取值范围如表所示。从表中数据可以发现，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。

特征名称	L	R	F	M	C
最小值	12.17	0.03	2	368	0.14
最大值	114.57	24.37	213	580717	1.5

```
from sklearn.preprocessing import StandardScaler
airline_scale= StandardScaler().fit_transform(airline_features)
print('标准化后LRFMC五个特征为：\n',data[:5,:])
```

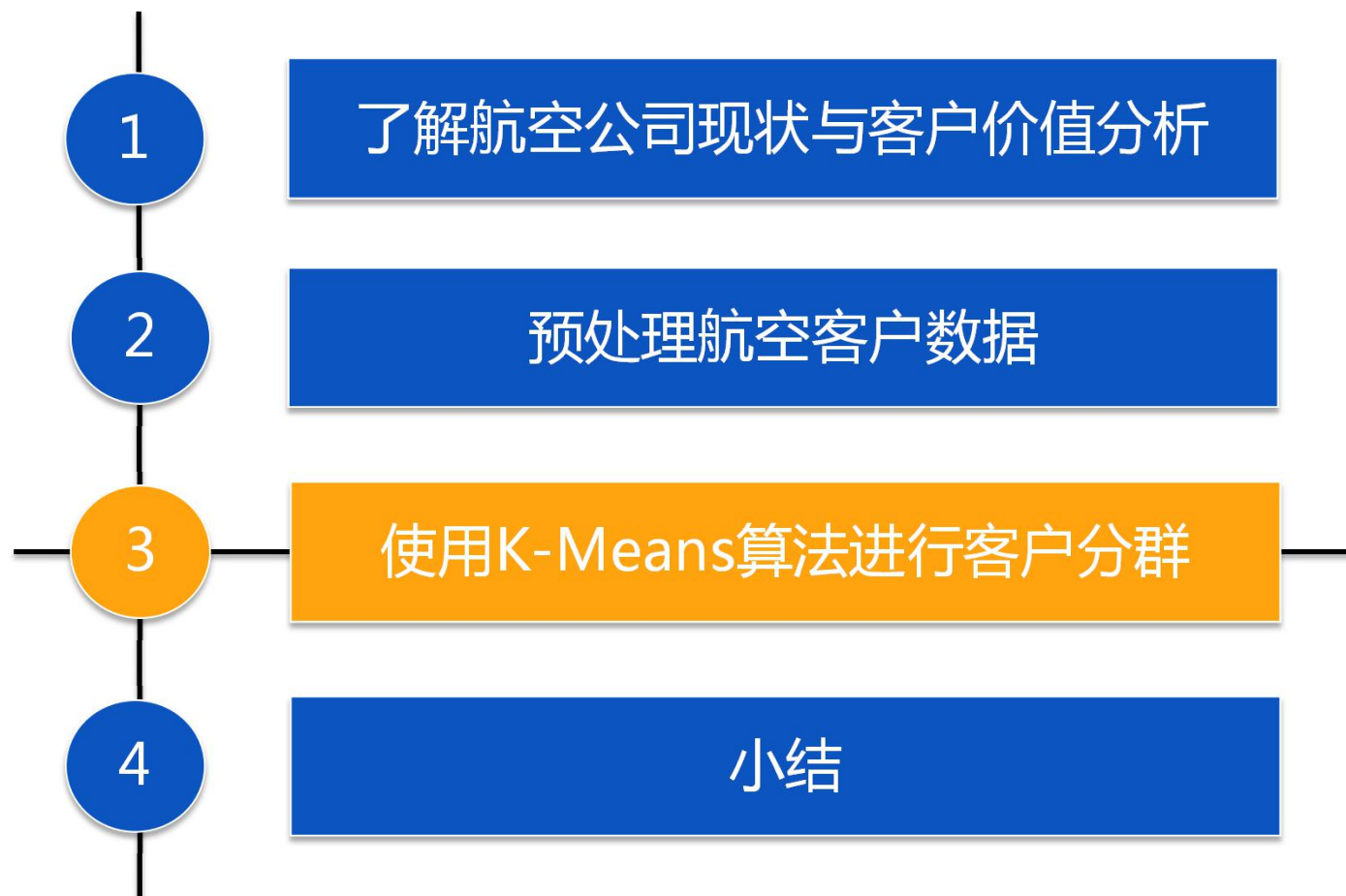
标准化LRFMC五个特征

L、R、F、M和C五个特征的数据示例，上图为原始数据，下图为标准差标准化处理后的数据。

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

ZL	ZR	ZF	ZM	ZC
1.44	-0.95	14.03	26.76	1.30
1.31	-0.91	9.07	13.13	2.87
1.33	-0.89	8.72	12.65	2.88
0.66	-0.42	0.78	12.54	1.99
0.39	-0.92	9.92	13.90	1.34

目录



了解K-Means聚类算法

1. 基本概念

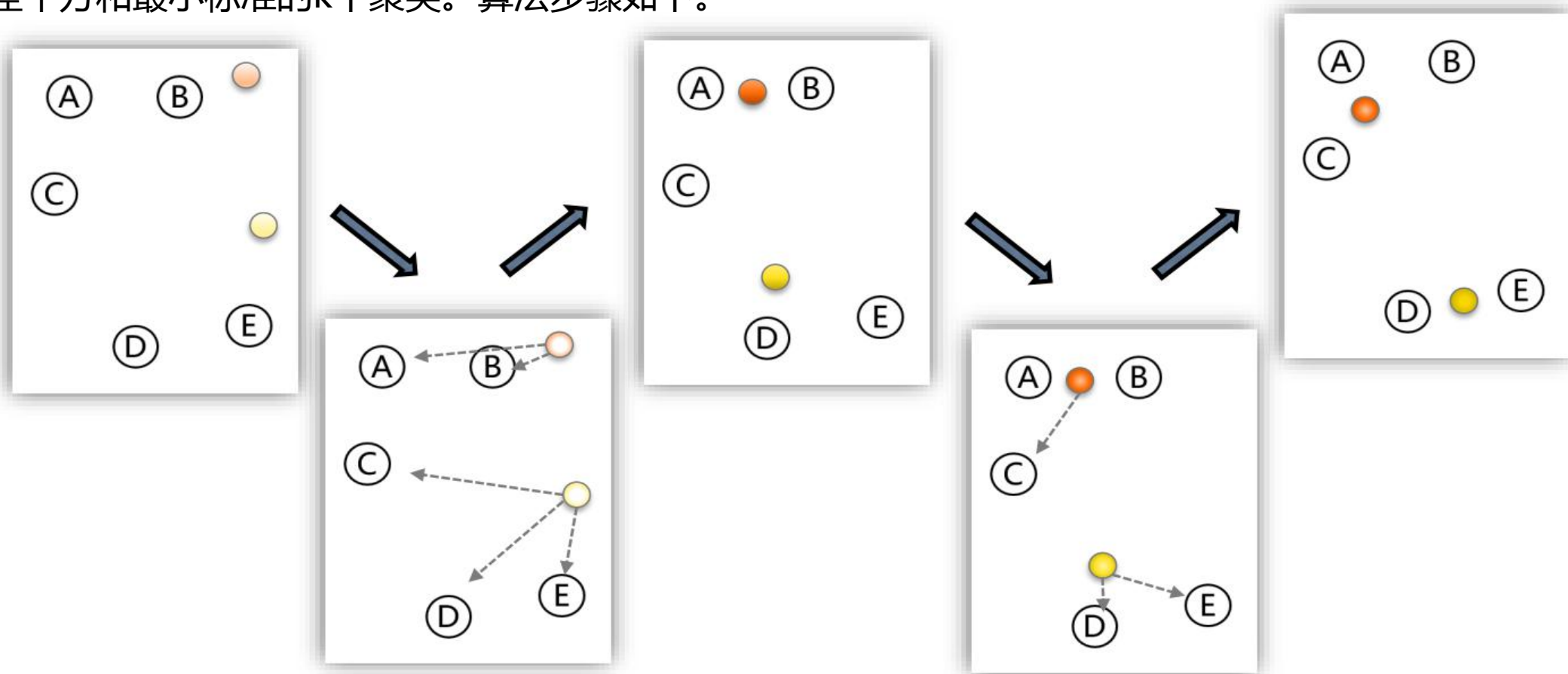
K-Means聚类算法是一种基于质心的划分方法，输入聚类个数 k ，以及包含 n 个数据对象的数据库，输出满足误差平方和最小标准的 k 个聚类。算法步骤如下。

- ① 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。
- ② 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。
- ③ 所有样本分配完成后，重新计算 k 个聚类的中心。
- ④ 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。
- 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

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。



了解K-Means聚类算法

1. 基本概念

- **K-Means聚类算法的结果依赖于初始聚类中心的随机选择。**
- 在实践中，为了得到较好的结果，通常以不同的初始聚类中心多次运行K-Means算法。在所有样本分配完成后，重新计算k个聚类中心时，对于连续数据，聚类中心取该簇的均值。
- K-Means聚类算法的**优点**：原理简单、容易实现，计算复杂度较小。
- **缺点**也较为明显：过分依赖初始值的选择，一旦初始值选择的不好，可能无法得到有效的聚类效果；需要预先给定k值，这往往是最困难的一步；对离群点非常敏感，容易使质心偏移。

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

3. kmeans函数及其参数介绍

➤ sklearn的cluster模块提供了kmeans函数来构建K-Means聚类模型，其基本语法如下。

```
class sklearn.cluster.KMeans(n_clusters=8, init='k-means++', n_init=10, max_iter=300, tol=0.0001,
precompute_distances='auto', verbose=0, random_state=None, copy_x=True, n_jobs=1, algorithm='auto')
```

➤ 常用参数及其说明如表所示。

参数名称	说 明
n_clusters	Int，可选，默认值：8，要形成的簇数以及要生成的质心数。
init	初始化方法，默认为 'k-means++'，即以智能方式选择初始聚类中心，以加速收敛。还可以选择 'random' 和 'ndarray'
n_init	接收int，默认10。使用不同质心种子运行k-means算法的次数。最终结果将是n_init连续运行的最佳输出。。
max_iter	单次运行的k-means算法的最大迭代次数。默认300
tol=0.0001	接收float，表示算法收敛的阈值，默认时0.0001

了解K-Means聚类算法

3. kmeans函数及其参数介绍

➤ sklearn的cluster模块提供了kmeans函数来构建K-Means聚类模型，其基本语法如下。

```
class sklearn.cluster.KMeans(n_clusters=8, init='k-means++', n_init=10, max_iter=300, tol=0.0001,
precompute_distances='auto', verbose=0, random_state=None, copy_x=True, n_jobs=1, algorithm='auto')
```

➤ 续表

参数名称	说 明
precompute_distances	接收boolean或auto，表示是否提前计算好样本之间的距离。
verbose	接收int，缺省0，表示不输出日志信息，1表示每隔一段时间打印一次日志，如果大于1，则打印日志更频繁。
random_state	接收int，表示随机数生成器的种子。默认为none
copy_x	boolean，缺省为True，不修改原始数据
n_jobs	接收int，表示任务使用的CPU数量，默认为1
algorithm	"auto"，"full" or "elkan"，default=" auto"

了解K-Means聚类算法

航空客户价值分析K-Means聚类

```
from sklearn.cluster import KMeans #导入kmeans算法
```

```
k = 5          ## 确定聚类中心数
```

```
#构建模型
```

```
kmeans_model = KMeans(n_clusters = k,n_jobs=4,random_state=123)
```

```
fit_kmeans = kmeans_model.fit(airline_scale) #模型训练
```

了解K-Means聚类算法

3. kmeans函数及其参数介绍

➤ K-Means模型构建完成后可以通过属性查看不同的信息，如表所示。

属性	说明
cluster_centers_	返回ndarray。表示分类簇的均值向量
labels_	返回ndarray。表示每个样本所属的簇的标记。
inertia_	返回ndarray。表示每个样本距离它们各自最近簇中心之和

`kmeans_model.cluster_centers_`

#查看聚类中心

`kmeans_model.labels_`

#查看样本的类别标签

`r1 = pd.Series(kmeans_model.labels_).value_counts()`

#统计不同类别样本的数目

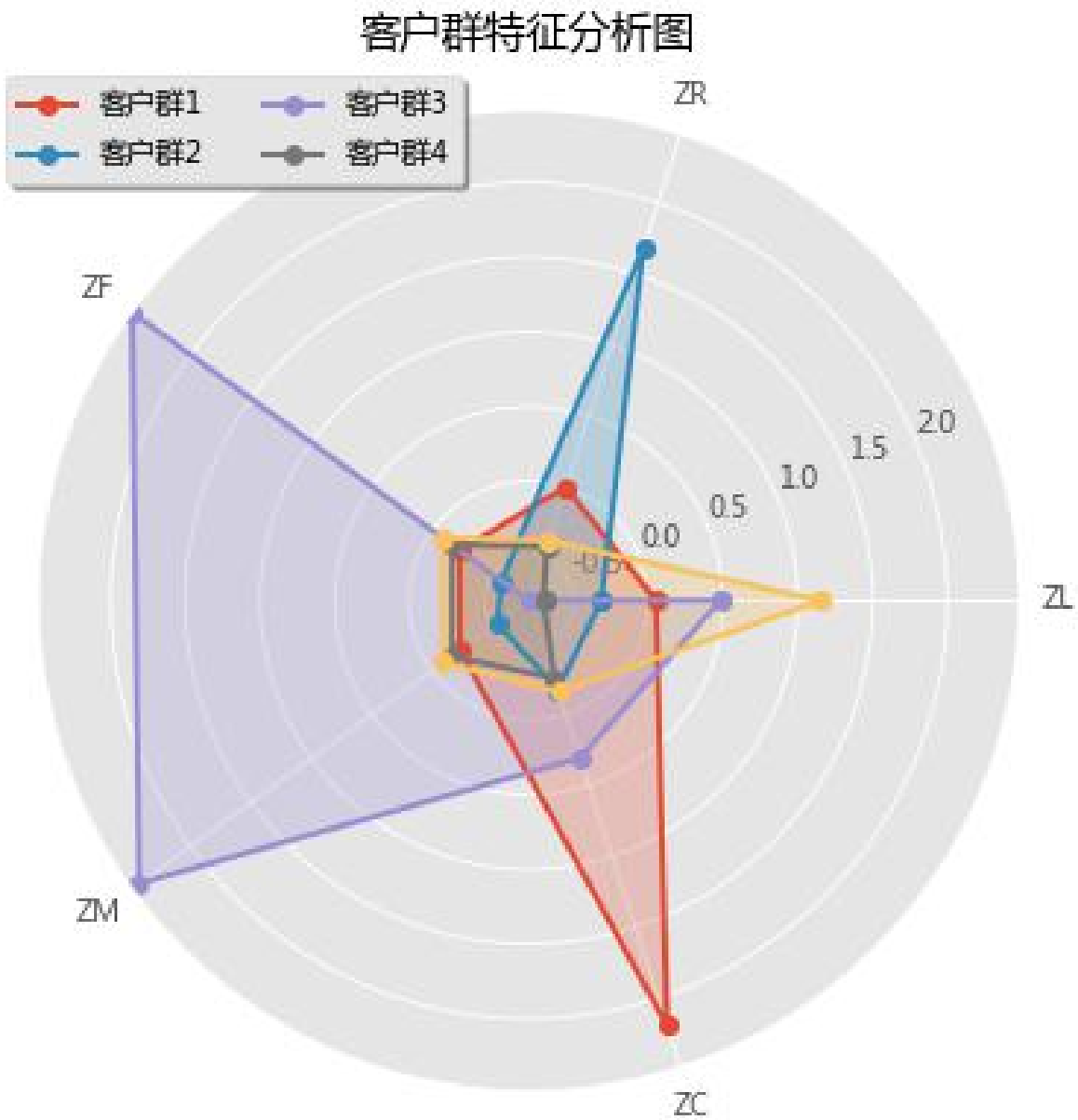
分析聚类结果

对数据进行聚类分群的结果如表所示。

聚类类别	聚类个数	聚类中心				
		ZL	ZR	ZF	ZM	ZC
客户群1	5336	0.483	-0.799	2.483	2.425	0.309
客户群2	4171	0.056	-0.003	-0.226	-0.229	2.200
客户群3	15742	1.160	-0.377	-0.087	-0.095	-0.156
客户群4	24663	-0.700	-0.415	-0.161	-0.161	-0.254
客户群5	12132	-0.313	1.686	-0.574	-0.537	-0.173

分析聚类结果

针对聚类结果进行特征分析，如图所示



了解K-Means聚类算法

航空客户价值分析K-Means聚类模型评价

```
from sklearn.metrics import calinski_harabaz_score    ##Calinski-Harabasz指数评价法
for i in range(2,15):                                ##分别评价K-Means聚类模型聚成2-14类的情况
    kmeans_model = KMeans(n_clusters = i,random_state=123).fit(airline_scale) ##构建并训练模型
    score = calinski_harabaz_score(airline_scale,kmeans_model.labels_)
    print('iris数据聚%d类calinski_harabaz指数为： %f'%(i,score))
```

```
iris数据聚2类calinski_harabaz指数为： 21876.878155
iris数据聚3类calinski_harabaz指数为： 21181.024716
iris数据聚4类calinski_harabaz指数为： 21802.079781
iris数据聚5类calinski_harabaz指数为： 20568.195147
iris数据聚6类calinski_harabaz指数为： 20192.717575
iris数据聚7类calinski_harabaz指数为： 19460.445093
iris数据聚8类calinski_harabaz指数为： 18436.795344
iris数据聚9类calinski_harabaz指数为： 17817.725799
iris数据聚10类calinski_harabaz指数为： 17314.546361
iris数据聚11类calinski_harabaz指数为： 16930.231209
iris数据聚12类calinski_harabaz指数为： 16585.614392
iris数据聚13类calinski_harabaz指数为： 16017.715400
iris数据聚14类calinski_harabaz指数为： 15524.889613
```

分析聚类结果

画出雷达图

```
import matplotlib.pyplot as plt
plt.rcParams['font.sans-serif'] = 'Microsoft YaHei'      # 中文和负号的正常显示
plt.rcParams['axes.unicode_minus'] = False
plt.style.use('ggplot')  ## 使用ggplot的绘图风格
feature = np.array(['ZL','ZF','ZR','ZC','ZM'])
values = kmeans_model.cluster_centers_                  # 构造数据，每个类的类中心
# 使用极坐标方式，设置雷达图的角度，用于平分切开一个圆面
angles=np.linspace(0, 2*np.pi, feature.size, endpoint=False)
# 为了使雷达图一圈封闭起来，需要下面的步骤
values=np.concatenate((values.T,[values.T[0]]))
angles=np.concatenate((angles,[angles[0]]))
```

分析聚类结果

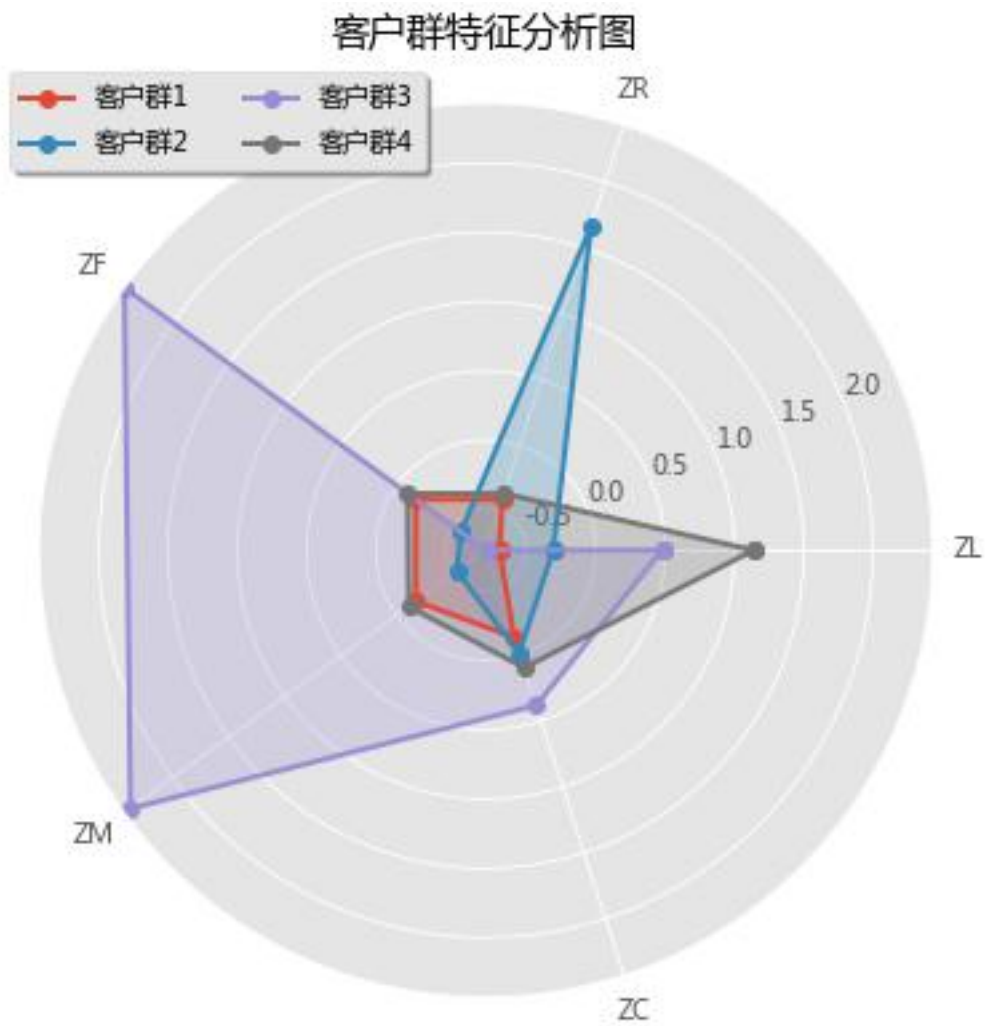
画出雷达图

```
fig=plt.figure(figsize = (6,10))                # 绘图
ax = fig.add_subplot(111, polar=True)            # 这里一定要设置为极坐标格式
ax.plot(angles, values, 'o-', linewidth=2)        # 绘制折线图
plt.legend(['客户群1','客户群2','客户群3','客户群4'],\
           loc='upper center',bbox_to_anchor=(0.2,1.1),ncol=2,fancybox=True,shadow=True)
ax.fill(angles, values, alpha=0.25)              # 填充颜色
ax.set_thetagrids(angles * 180/np.pi, feature)  # 添加每个特征的标签
ax.set_ylim(values.min(),values.max())           # 设置雷达图的范围
plt.title('客户群特征分析图')                   # 添加标题
ax.grid(True)                                    # 添加网格线
plt.show()                                       # 显示图形
```

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。

群类别	优势特征				弱势特征			
客户群1				R	L			C
客户群2					R	M	C	L
客户群3	F	M	C	R				
客户群4	L	M		C				R



分析聚类结果

基于特征描述，本项目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。每种客户类别的特征如图所示。

	重要保持客户	重要发展客户	重要挽留客户	一般客户与低价值客户
平均折扣系数 (C)				
最近乘机距今的时间长度 (R)				
飞行次数 (F)				
总飞行里程 (M)				
会员入会时间 (L)				

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。

群类别	优势特征				弱势特征			
客户群1			R		L		C	
客户群2					R	M	C	L
客户群3	F	M	C	R				
客户群4	L	M	C				R	

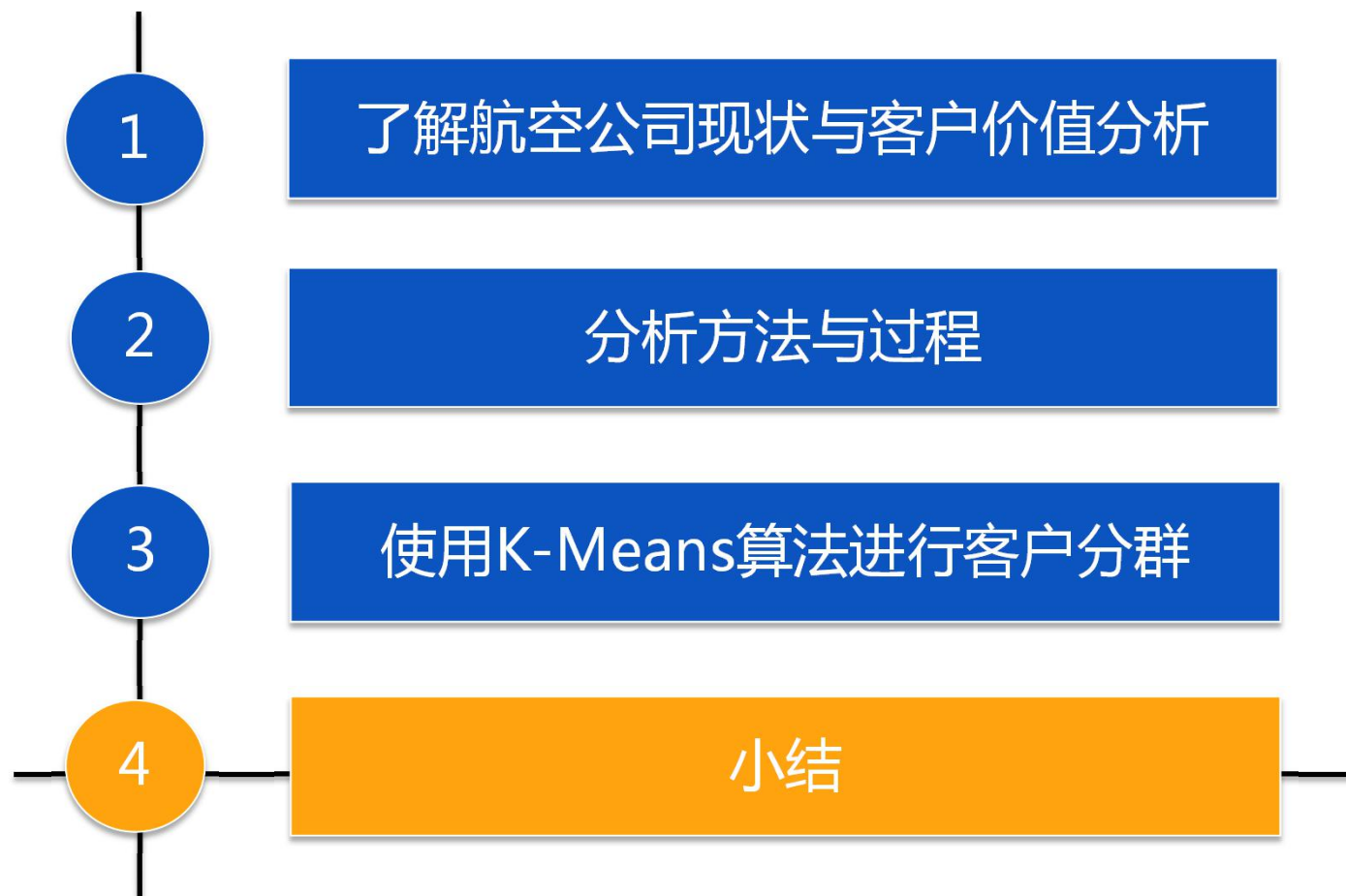
- 客户群3：折扣高，飞行里程高，最近乘机距今时间短，是高价值客户，也就是**重要保持客户**
- 客户群4：折扣高，飞行里程高，入会时间长，但较长时间未乘机，是**重要挽留客户**
- 客户群2：折扣低、飞行里程少，较长时间未乘机，属于**一般客户或低价值客户**
- 客户群1：入会时间短，但最近乘机时间较近，属于**重要发展客户**

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。

- **会员的升级与保级**：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。
- **首次兑换**：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在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。
- **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。

目录



小结

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。





大数据，成就未来

Thank you!

