

作为一种对客户分类的方法，RFM分析模型起初主要用于直效营销(Direct Marketing)领域，目的是提高老客户交易的次数。

一个例子：

如果给新顾客送米，王永庆就细心记下这户人家米缸的容量，并且问明家里有多少人吃饭，几个大人、几个小孩，每人饭量如何，据此估计该户人家下次买米的大概时间，记在本子上。到时候，不等顾客上门，他就主动将相应数量的米送到客户家里。

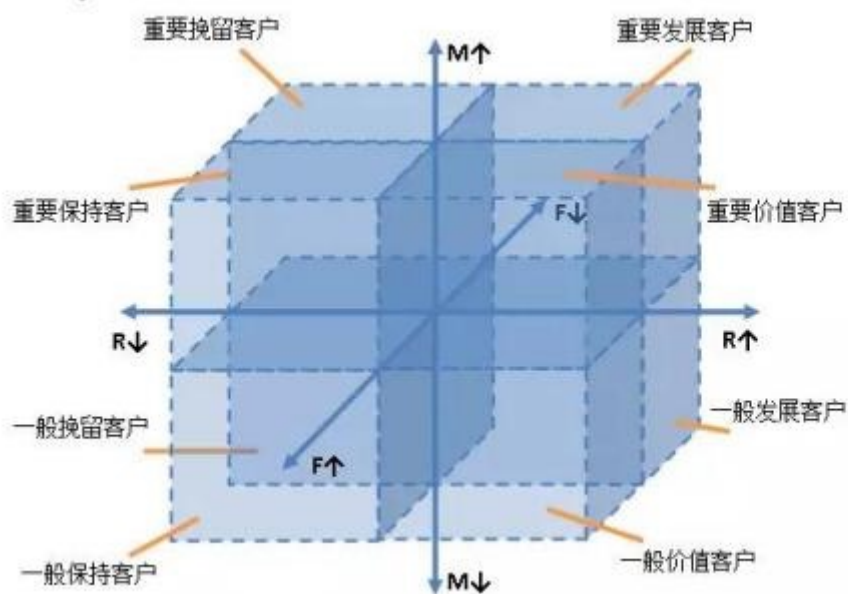
重要概念：RFM

重要概念：挽留，发展，保持，价值

1. 简介

根据美国数据库营销研究所 Arthur Hughes 的研究，客户数据库中有 3 个神奇的要素，这 3 个要素构成了数据分析最好的指标：

- 最近一次消费 (Recency)
- 消费频率 (Frequency)
- 消费金额 (Monetary)



R 分类	F 分类	M 分类	客户类型
高	高	高	高价值客户
低	高	高	重点保持客户
高	低	高	重点发展客户
低	低	高	重点挽留客户
高	高	低	一般价值客户
低	高	低	一般保持客户
高	低	低	一般发展客户
低	低	低	潜在客户

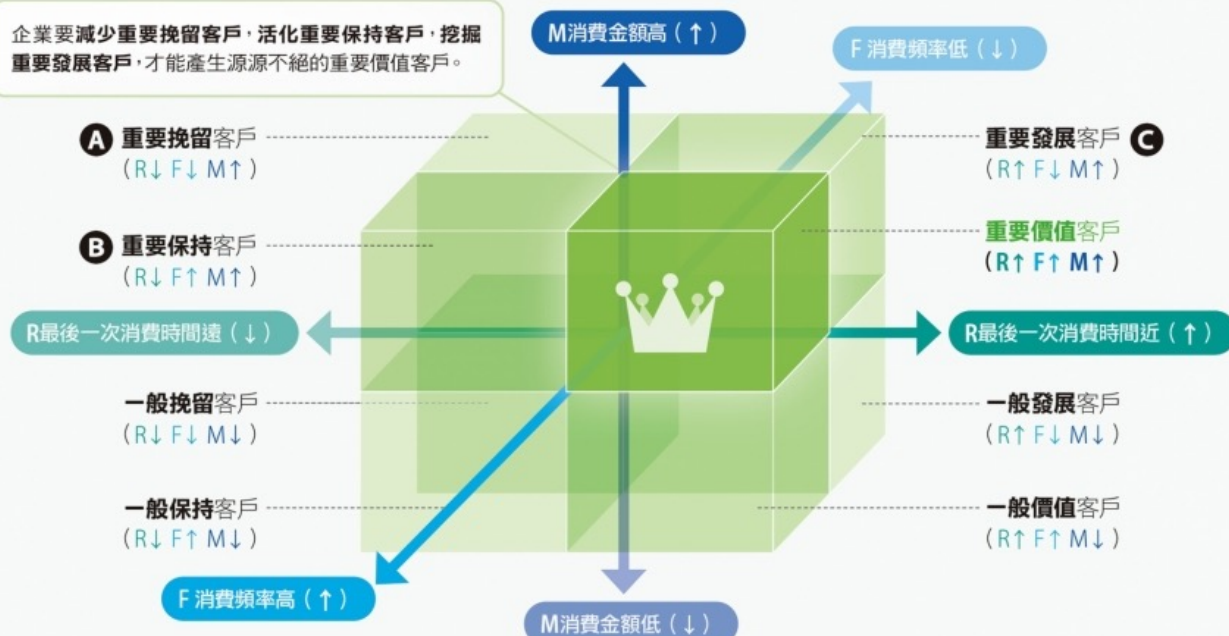
RFM 模型是衡量客户价值和客户创利能力的重要工具和手段。在众多的客户关系管理 (CRM) 的分析模式中，RFM 模型是被广泛提到的。该机械模型通过一个客户的 [近期购买行为](#)、[购买的总体频率](#)以及[花了多少钱](#) 3 项指标来描述该客户的价值状况。

客户分类	客户特征
重要价值客户（111）	最近消费时间近、消费频次和消费金额都很高，必须是VIP啊！
重要发展客户（101）	最近消费时间较近、消费金额高，但频次不高，忠诚度不高，很有潜力的用户，必须重点发展。
重要保持客户（011）	最近消费时间交远，消费金额和频次都很高
重要挽留客户（001）	最近消费时间较远、消费频次不高，但消费金额高的用户，可能是将要流失或者已经要流失的用户，应当基于挽留措施。
一般价值客户（110）	最近消费时间近，频率高但消费金额低，需要提高其客单价
一般发展客户（100）	最近消费时间较近、消费金额，频次都不高。
一般保持客户（010）	最近消费时间较远、消费频次高，但金额不高。
一般挽留客户（000）	都不高

RFM模型

喚醒沉睡客、拓展忠誠客戶

企業要減少重要挽留客戶，活化重要保持客戶，挖掘重要發展客戶，才能產生源源不絕的重要價值客戶。



2. 最近一次消费

2.1 定义

最近一次消费意指上一次购买的时间

顾客上一次是几时来店里、上一次根据哪本邮购目录购买东西、什么时候买的车，或在你的超市买早餐最近的一次是什么时候。

思考题：买车是一个高频的需求吗？适合 RFM 吗？

思考题：在淘宝里你看过的类似的商品会不断的出现，所谓的推荐算法是靠 RFM 模型吗？

理论上，上一次消费时间越近的顾客应该是比较好的顾客，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 3 个月的顾客收到营销人员的沟通信息多于 3 至 6 个月的顾客。

2.2 描述

消费的过程是持续变动的。

在顾客距上一次购买时间满一个月之后，在数据库里就成为消费为两个月的客户。反之，同一天，消费为 3 个月前的客户作了其下一次的购买，他就成为消费为一天前的顾客，也就有可能在很短的期间内就收到新的折价信息。

2.3 功能

消费的功能不仅在于提供的促销信息而已，营销人员的消费报告可以监督事业的健全度。

优秀的营销人员会定期查看消费分析，以掌握趋势。

月报告如果显示上一次购买很近的客户，(消费为 1 个月) 人数如增加，则表示该公司是个稳健成长的公司；反之，如上一次消费为一个月的客户越来越少，则是该公司迈向不健全之路的征兆。

2.4 重要指标

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。

3. 消费频率

3.1 定义

消费频率是顾客在限定的期间内所购买的次数。

思考题：如何增加系统或者平台中的复购、重复消费

我们可以说最常购买的顾客，也是满意度最高的顾客。如果相信品牌及商店忠诚度的话，最常购买的消费者，忠诚度也就最高。增加顾客购买的次数意味着从竞争对手处偷取市场占有率，由别人的手中赚取营业额。

3.2 分类

根据这个指标，我们又把客户分成五等分(这只是其中一种分法)，这个五等分分析相当于是一个“忠诚度的阶梯”(loyalty ladder)，如：

- 购买一次的客户为**新客户**，
- 购买两次的客户为**潜力客户**，
- 购买三次的客户为**老客户**，
- 购买四次的客户为**成熟客户**，
- 购买五次及以上则为**忠实客户**。

其诀窍在于让消费者一直顺着阶梯往上爬，把销售想象成是要将两次购买的顾客往上推成三次购买的顾客，把一次购买者变成两次的。

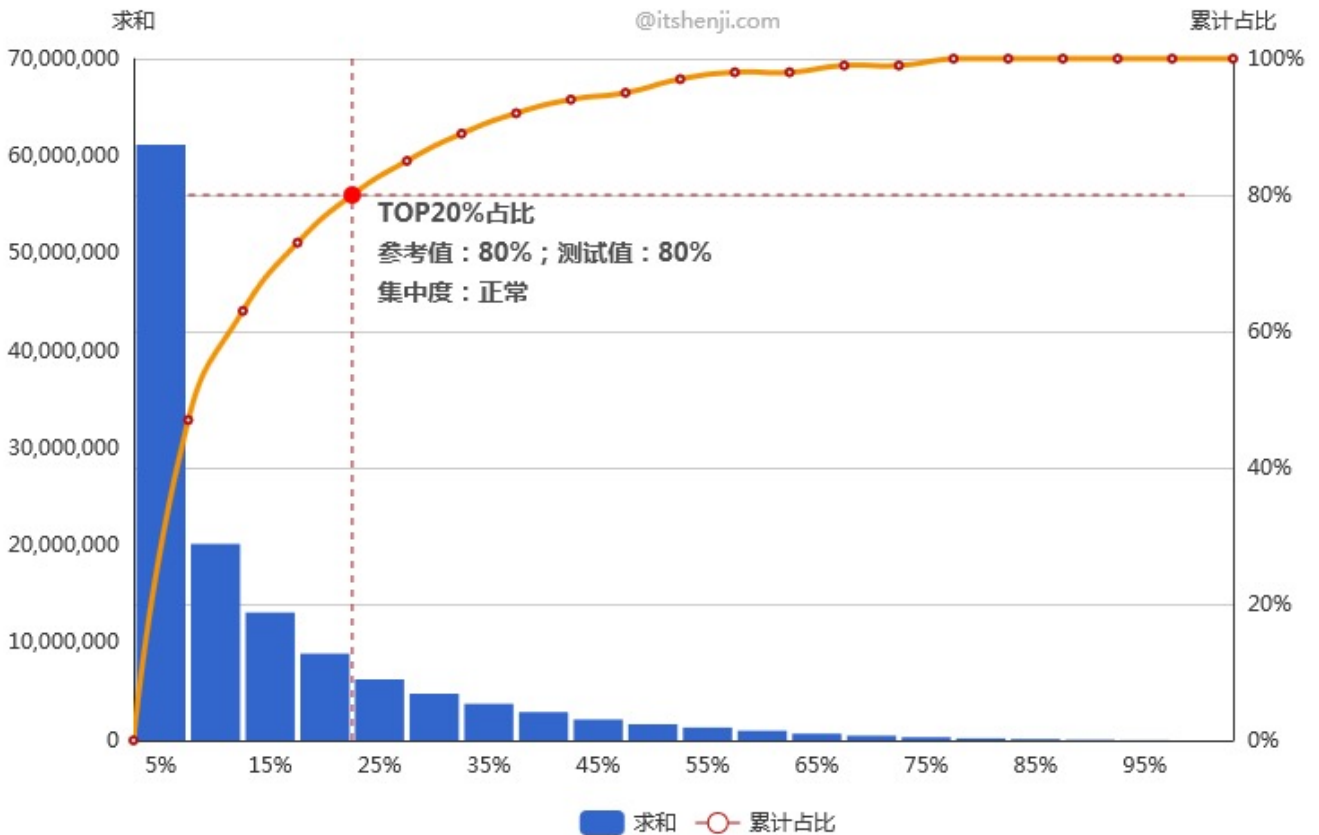
3.3 数据分析

影响复购的核心因素是商品，因此复购不适合做跨类目比较。

比如食品类目和美妆类目：食品是属于“半标品”，产品的标品化程度越高，客户背叛的难度就越小，越难形成忠实用户；但是相对美妆，食品又属于易耗品，消耗周期短，购买频率高，相对容易产生重复购买，因此跨类目复购并不具有可比性。

4. 消费金额

帕累托图（求和）



了解帕雷托法则 (Pareto Principle, 即 80/20 法则)

k7xGWq

根据帕雷托法则，许多结果 (大约 80% 的后果) 是源自 20% 的成因。换句话说，小比例的成因就能带来巨大的影响。了解这个概念很重要，因为它有利于辨别应该优先执行哪些计划，以便发挥最大的影响力。

帕雷托法则从何而来？

帕雷托法则是由义大利经济学家帕雷托 (Vilfredo Pareto) 于 1896 年所发展出来的。据帕雷托观察，意大利有 80% 的土地是由仅 20% 的人口所持有。他还在自己的花园里见证了此现象：20% 的植株结了 80% 的果实。透过数学最能够完美描述这个关系：两个数量的幂次分布，其中一个数量改变会造成另一个数量相关的改变。

此现象还有其他几个不同的名称：

- 帕雷托法则
- 80/20 法则 (最常见)
- 关键少数法则
- 因子稀疏法则

80/20 法则并非正式的数学公式，而更像是归纳后的现象，在经济学、商业、时间管理、甚至运动的领域中都可以观察的到。

理论上 M 值和 F 值是一样的，都带有时间范围，指的是一段期间（通常是 1 年）内的消费金额，在工作中我认为对于一般店铺类目而言，产品的价格带都是比较单一的，比如：同一品牌美妆类，价格浮动范围基本在某个特定消费群的可接受范围内，加上单一品类购买频次不高，所以对于一般店铺而言，M 值对客户细分的作用相对较弱。

4.1 Generating Pareto distribution in Python

Pareto distribution can be replicated in Python using either **Scipy.stats** module or using **NumPy**. Scipy.stats module encompasses various probability distributions(概率分布) and an ever-growing library of statistical functions.

Scipy is a Python library used for scientific computing and technical computing.

NumPy is a Python library used for scientific computing that apart from its scientific uses can be used as a multi-dimensional container for generic data.

4.2 Using Scipy.stats

PYTHON

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import pareto
x_m = 1 #scale
# list of values of shape parameters
alpha = [1, 2, 3]
samples = np.linspace(start=0, stop=5, num=1000)

for a in alpha:
    output = np.array([pareto.pdf(
        x=samples,
        b=a,
        loc=0,
        scale=x_m)])
    plt.plot(samples,
              output.T,
              label='alpha {0}'.format(a))

plt.xlabel('samples', fontsize=15)
plt.ylabel('PDF', fontsize=15)
plt.title('Probability Density function(概率密度函数)', fontsize=15)
plt.grid(b=True, color='grey',
        alpha=0.3, linestyle='-.',
        linewidth=2)
plt.rcParams["figure.figsize"] = [5, 5]
plt.legend(loc='best')
plt.show()
```



4.2 Using Numpy

PYTHON

```
import numpy as np
import matplotlib.pyplot as plt
x_m, alpha = 1, 3.
#drawing samples from distribution
samples = (np.random.pareto(alpha, 1000) + 1) * x_m
count, bins, _ = plt.hist(samples, 100, normed=True)
fit = alpha*x_m**alpha / bins**(alpha+1)
plt.plot(bins,
         max(count)*fit/max(fit),
         linewidth=2,
         color='r')
plt.xlabel('bins', fontsize=15)
```



```
plt.ylabel('probability density', fontsize=15)
plt.title('Probability Density Function', fontsize=15)
plt.grid(b=True, color='grey',
        alpha=0.3, linestyle='-.',
        linewidth=2)
plt.rcParams['figure.figsize'] = [8, 8]
plt.show()
```



4.3 Verifying Pareto distribution

Q-Q plot(Quantile-Quantile plot) is used to determine whether the continuous random variables follow Pareto distribution.

PYTHON

```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
x_m = 10
alpha = 15
# the size of the sample(no. of random samples)
size = 100000
samples = (np.random.pareto(alpha, size) + 1) * x_m
stats.probplot(samples, dist='pareto',
               sparams=(15, 10),
               plot=pylab)
plt.show()
```

5. 案例 1

面试题，如何为公司指定发短信的策略

这个RFM模型在实操时有什么用呢？举个例子：

比如对圈用户群发短信转化只有不到1%时，你可以用RFM做个分析，只选取R值高的用户(最近2周到最近一个月内消费的用户)，转化率可以由1%提升到10%。

这也意味着，以往6元/订单将下降到0.6元/订单。掌柜们是愿意花600元给10000个用户发短信，得到100个订单，还是愿意花48元给800人发短信得到80个订单，相信大家一定会选后者。

而整体的RFM区分，则能够帮掌柜们针对不同的用户发不同的短信，短信的开头是用“好久不见”、还是用“恭喜你成为VIP”，就得看时重要保持客户还是重要价值用户了。只有能区分用户，才能走向精细化运营

6. 案例 2

基于RFM的电信客户市场细分方法

对于电信企业而言,不同的客户具有不同的内在价值,企业的首要问题就是采取有效方法对客户进行分类,发现客户内在价值的变化规律与分布特征,并以此制定客户的差别化服务政策,通过政策的实施将客户分类的结果作用于企业实践。

针对电信行业提出一种基于改进RFM模型的客户分类方法。应用层次分析法来确定RFM模型中每个变量的权重,在此基础上,应用K均值聚类法来对客户进行分类,之后分析每一类客户的行为特征和价值,并且对不同的顾客类别采取不同的策略。

7. 实际操作

通过RFM模型能得到什么信息

- 谁是最佳用户？
- 哪些用户即将流失？
- 谁有潜力成为有价值用户？
- 哪些用户可以留存？



假设有如下的样例数据：

客户名称	日期	消费金额	消费数量
上海****有限公司	2020-05-20	76802	2630

需要将数据集加工成如下格式：

客户名称	客户平均消费金额	所有客户平均消费金额	客户总消费频次	所有客户消费频次平均值	客户最近消费日期	客户最近消费距离当前时间的时差	所有客户最近一次消费距离当前时间的平均时差	消费金额向量化M	消费频次向量化F	最近消费向量化R	RFM	客户类型
	当前客户的平均消费金额	所有客户的平均消费金额	当前客户总的消费频次	所有客户平均消费频次	客户最近一次消费的日期	客户最近一次消费距离当前日期的时差	所有客户最近一次消费距离当前日期的平均时差	消费金额向量化M	消费频次向量化F	最近消费向量化R	concat(RFM)	

SQL

```
SELECT
customer_name, -- 客户名称
customer_avg_money, -- 当前客户的平均消费金额
customer_frequency, -- 当前客户的消费频次
total_frequency, -- 所有客户的总消费频次
total_avg_frequency, -- 所有客户平均消费频次
customer_recency_diff, -- 当前客户最近一次消费日期与当前日期差值
total_recency, -- 所有客户最近一次消费日期与当前日期差值的平均值
monetary, -- 消费金额向量化
frequency, -- 消费频次向量化
recency, -- 最近消费向量化
rfm, -- rfm
CASE
WHEN rfm = "111" THEN "重要价值客户"
WHEN rfm = "101" THEN "重要发展客户"
WHEN rfm = "011" THEN "重要保持客户"
WHEN rfm = "001" THEN "重要挽留客户"
```



```

        WHEN rfm = "110" THEN "一般价值客户"
        WHEN rfm = "100" THEN "一般发展客户"
        WHEN rfm = "010" THEN "一般保持客户"
        WHEN rfm = "000" THEN "一般挽留客户"
    END AS rfm_text
FROM
    (SELECT
        customer_name, -- 客户名称
        customer_avg_money, -- 当前客户的平均消费金额
        customer_frequency, -- 当前客户的消费频次
        total_avg_money, -- 所有客户的平均消费总额
        total_frequency, -- 所有客户的总消费频次
        total_frequency / count(*) over() AS total_avg_frequency, -- 所有客户平均消费频次
        customer_recency_diff, -- 当前客户最近一次消费日期与当前日期差值
        avg(customer_recency_diff) over() AS total_recency, -- 所有客户最近一次消费日期与当前日期差值的平均值
        if(customer_avg_money > total_avg_money,1,0) AS monetary, -- 消费金额向量化
        if(customer_frequency > total_frequency / count(*) over(),1,0) AS frequency, -- 消费频次向量化
        if(customer_recency_diff > avg(customer_recency_diff) over(),0,1) AS recency, -- 最近消费向量化
        concat(if(customer_recency_diff > avg(customer_recency_diff) over(),0,1),if(customer_frequency > total_frequency / count(*) over(),1,0),if(customer_avg_money > total_avg_money,1,0)) AS rfm
    FROM
        (SELECT
            customer_name, -- 客户名称
            max(customer_avg_money) AS customer_avg_money, -- 当前客户的平均消费金额
            max(customer_frequency) AS customer_frequency, -- 当前客户的消费频次
            max(total_avg_money) AS total_avg_money, -- 所有客户的平均消费总额
            max(total_frequency) AS total_frequency, -- 所有客户的总消费频次
            datediff(CURRENT_DATE,max(customer_recency)) AS customer_recency_diff -- 当前客户最近一次消费日期与当前日期差值
        FROM
            (SELECT
                customer_name, -- 客户名称
                avg(money) over(partition BY customer_name) AS customer_avg_money, -- 当前客户的平均消费金额
                count(amount) over(partition BY customer_name) AS customer_frequency, -- 当前客户的消费频次
                avg(money) over() AS total_avg_money, -- 所有客户的平均消费总额
                count(amount) over() AS total_frequency, -- 所有客户的总消费频次
                max(sale_date) over(partition BY customer_name) AS customer_recency -- 当前客户最近一次消费日期
            FROM customer_sales) t1
        GROUP BY customer_name)t2)t3

```

8. 总结

什么是RFM模型



Recency

距离最近一次交易



Frequency

交易频率



Monetary

交易金额

RFM模型被广泛运用于众多客户关系管理(CRM)的分析模式中, 该模型对于客户价值和客户创利能力的衡量有着重要作用。该模型能够整合客户的近期交易行为、交易的总体频率以及交易金额三项指标, 通过三项指标的数据反馈能够将客户加以区别。

1.重要价值客户

2.重要唤回客户

3.重要深耕客户

4.重要挽留客户

5.潜力客户

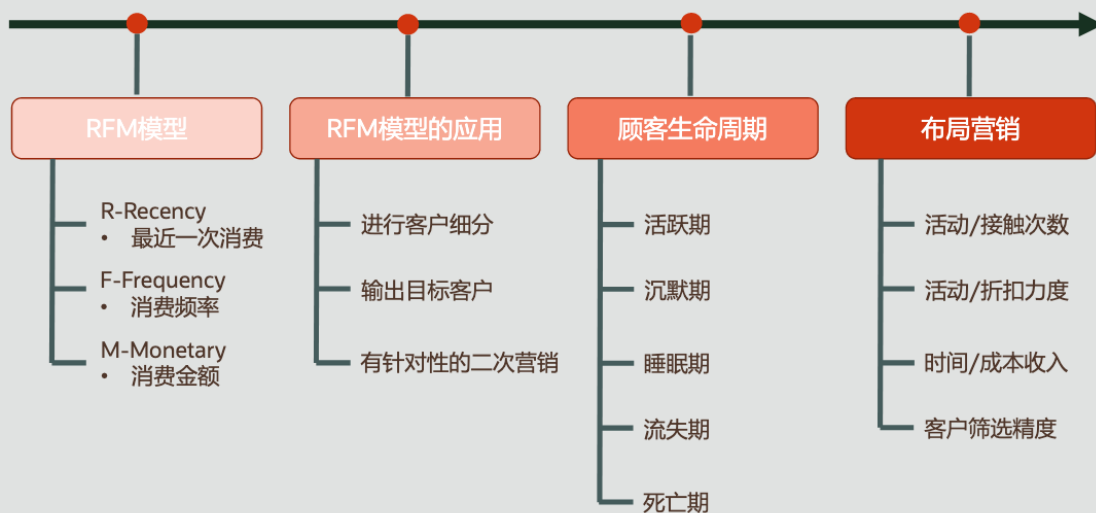
6.新客户

7.一般维持客户

8.流失客户



RFM模型目的



依据RFM模型划分用户层级

客户类型	最近交易距离当前天数(黏性)	累计单数(忠诚度)	累计交易金额(收入)	对应货运圈的成员场景
重要价值客户	↑	↑	↑	RFM都很大, 优质客户, 需要保持
重要唤回客户	↓	↑	↑	交易金额和交易次数大, 但最近无交易。需要唤回
重要深耕客户	↑	↓	↑	交易金额大贡献度高, 且最近有交易。需要重点识别
重要挽留客户	↓	↓	↑	交易金额大, 潜在的有价值客户。需要挽留
潜力客户	↑	↑	↓	交易次数大, 且最近有交易。需要挖掘
新客户	↑	↓	↓	最近有交易, 接触的新客户, 有推广价值
一般维持客户	↓	↑	↓	交易次数多, 但是贡献不大, 一般维持
流失客户	↓	↓	↓	FM值均低过平均值, 最近也没再发货相当于流失



RFM模型的价值



Oracle 敏捷业务数据平台

