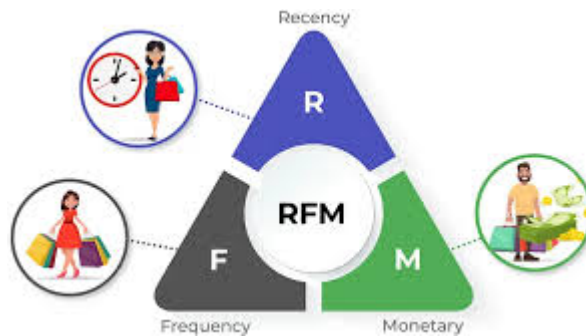




# DataScience\_CA01\_RFM과 LTV

|          |                   |
|----------|-------------------|
| ▼ Status | Customer Analysis |
|----------|-------------------|



## Outline

- RFM
- LTV

## 학습목표

이 수업을 통해 아래의 학습목표를 이루고자 한다.

- | 고객분석의 기초인 RFM 분석을 이해한다.
- | RFM 피쳐들의 추정을 통한 LTV 분석을 이해한다.
- | 그외 추가 고객 분석에 확장성의 기초를 이해한다.

## RFM(Recency Frequency Monetary)

### RFM 분석이란?

RFM 분석은 고객의 과거 거래 데이터를 기반으로

현재 고객 가치를 정량적으로 추정하고, 유사한 행동을 보이는 고객 집단을 분류하는 분석 기법입니다.

- **R (Recency):** 고객이 마지막으로 구매한 이후 경과 시간
- **F (Frequency):** 특정 기간 동안 고객의 총 구매 횟수
- **M (Monetary):** 총 구매 금액

RFM은 마케팅의 고전적 세분화 전략이자, LTV(고객 생애 가치) 예측의 근사 지표로 간주됩니다.

## RFM 분석의 이론적 기초

RFM은 다음과 같은 이론적 개념들과 연관됩니다:

| 구성 요소     | 이론적 기반                                  |
|-----------|---|
| Recency   | 행동경제학의 recency effect, 재방문 가능성 추정       |
| Frequency | 반복 행동 모델 (Markov Chain, Renewal theory) |
| Monetary  | 고객 가치, 매출 기여도 추정, LTV 구성 요소             |

RFM은 비모수(non-parametric) 방식의 행동요약이며, 확률적 고객 생애 가치 추정모형(BGNBD, Pareto/NBD 등)의 기초 정보로도 활용됩니다.

## 수학적 구조와 정의

각 고객  $i$ 에 대해:

- $R_i$ : 기준 시점으로부터 마지막 구매까지의 경과 시간
- $F_i$ : 해당 기간 동안의 총 구매 횟수
- $M_i$ : 누적 구매 금액

이를 벡터 형태로 나타내면 다음과 같습니다.  $\text{RFM}_i = (R_i, F_i, M_i)$

RFM은 3차원 유클리드 공간 내 고객 분포로 해석될 수 있으며, 각 고객은 이 공간상의 점(point)으로 표현

## RFM의 통계적 해석 및 계산

### Recency

$$R_i = T_{\text{now}} - T_{\text{last purchase}}$$

시간 간격 데이터로서 지수 분포(exponential distribution)의 성질을 가짐

### Frequency

$$F_i = \sum_t 1_{\text{purchase}_t}$$

카운트 데이터이며 포아송 분포(Poisson distribution)나 음이항 분포로 근사 가능

### Monetary

$$M_i = \sum_t \text{purchase amount}_t$$

연속적인 금액 데이터이며 감마(Gamma) 분포 또는 로그 정규 분포로 표현 가능

## RFM과 LTV(Lifetime Value)와의 연결

고객 생애 가치는 다음과 같이 모델링됩니다:

$$LTV_i = \mathbb{E} \left[ \sum_{t=1}^T m_{i,t} \cdot P_{i,t} \right]$$

- $m_{i,t}$ : 시점 t에서의 예상 구매 금액
- $P_{i,t}$ : 시점 t에서의 구매 확률

RFM에서:

- Recency와 Frequency는  $P_{i,t}$  추정의 주요 정보
- Monetary는  $m_{i,t}$  추정의 기반

즉, RFM은 전체적인 LTV를 근사하는 3차 요약 통계 벡터로 해석할 수 있습니다.

## RFM 점수화 방식

### 방법 1: 분위수 기반 등급화 (Quantile Ranking)

각 지표에 대해 상위 20%를 5점, 하위 20%를 1점 등으로 등급화:

- Recency: 값이 작을수록 점수 높게 (최신 고객 우대)
- Frequency, Monetary: 값이 클수록 점수 높게 (많이, 비싸게 구매한 고객 우대)

$$\text{RFM Score}_i = (R_i^{(q)}, F_i^{(q)}, M_i^{(q)})$$

### 방법 2: Z-score 정규화 후 KMeans 등 군집화

$Z(x_i) = \frac{x_i - \mu}{\sigma}$  정규화된 RFM 벡터로 클러스터링 수행:

$$\text{RFM}_i = (Z(R_i), Z(F_i), Z(M_i))$$

## 고객 세그먼트 분류 예

| RFM 점수  | 고객 유형     | 특성                  |
|---------|-----------|---------------------|
| (5,5,5) | 핵심 VIP 고객 | 최근, 자주, 많이 구매       |
| (5,1,1) | 신규 저가 고객  | 최근 방문했지만 아직 충성도 낮음  |
| (1,5,5) | 이탈 고액 고객  | 과거에는 우수하지만 최근 이탈 조짐 |
| (1,1,1) | 장기 이탈 고객  | 오래 전, 적게, 작게 구매한 고객 |

이러한 점수 기반 세그먼트는 마케팅 캠페인의 정밀 타겟팅에 매우 유용

## 군집 기반 분석 확장

RFM 벡터는 3차원 입력이므로, 다음과 같은 방법으로 비지도 학습 기반 세분화가 가능합니다:

- K-means: 수치적 중심 기반 클러스터링
- DBSCAN: 밀도 기반 이상 고객 탐지
- GMM: 혼합 분포 기반 세분화

또한, RFM 공간에서의 거리 기반 유사도(예: 유클리드 거리, 코사인 거리)를 이용해 유사 고객을 추출할 수 있습니다

# LTV (Customer Lifetime Value)

## LTV (Customer Lifetime Value)

### 기본 개념

LTV는 고객 한 명이 기업과 관계를 유지하는 전체 생애 동안 **얼마나 많은 수익을 창출할 것으로 기대되는지를 나타내는 지표**로, 마케팅, CRM, 전략 재무관리에서 핵심적인 개념

- Gupta & Lehmann (2003)에 따르면, LTV는 다음과 같이 정의됩니다.

$$LTV_i = \sum_{t=1}^T \frac{\mathbb{E}[R_{i,t} - C_{i,t}]}{(1+d)^t}$$

- $R_{i,t}$ : 고객 i의 t시점 수익
- $C_{i,t}$ : 고객 유지 및 지원에 드는 비용
- d: 할인율
- T: 고객의 생애 기간 (retention 기간)

미래 현금 흐름을 할인하여 현재 가치(Present Value)로 환산하는 전형적인 DCF(Discounted Cash Flow) 구조

## 고전적 모델: BG/NBD + Gamma-Gamma (Fader, Hardie & Lee, 2005/2010)

이 모델은 고객의 생존 여부와 재구매 행동, 그리고 구매 금액을 모두 반영하여 비계약적 (non-contractual) 환경에서 개별 고객의 LTV를 추정하는 가장 널리 쓰이는 통계 기반 접근법

### BG/NBD 모델 (Beta-Geometric/Negative Binomial Distribution)

#### 모델 목적

- 고객의 구매 횟수와 생존 여부를 동시에 추정하여, 향후 구매 수를 예측

#### 가정

1. 고객은 생존 상태일 때만 구매를 한다.
2. 고객은 생존 상태일 때 포아송(Poisson) 분포를 따라 구매한다.
3. 고객은 확률적으로 이탈(churn)하며, 이탈 이후에는 구매하지 않는다.
4. 고객 간 이탈률 및 구매율은 서로 다르며, 베이지안 사전분포를 따른다.

#### 확률적 구조

- 고객의 구매율  $\lambda \sim \text{Gamma}(r, \alpha)$
- 고객의 이탈 확률  $p \sim \text{Beta}(a, b)$

즉, 구매는  $\text{Poisson}(\lambda)$ , 이탈은  $\text{Geometric}(p)$  과정을 따른다고 본다.

### 수식 예시 (생존 상태에서의 구매 확률 예측)

고객 i가 t 기간 내 x회 구매했다면,

향후 고객이 여전히 생존해 있고 향후에도 구매할 확률은 다음과 같이 표현

$$P(X(t) = x \mid r, \alpha, a, b) = \text{BG/NBD의 확률질량함수(PMF)}$$

→ 이 수식은 고객의 생존 여부까지 고려하여 미래의 재구매 가능성과 기대 횟수를 함께 추정

### Gamma-Gamma 모델 (Monetary Value 예측)

#### 모델 목적

- 구매 횟수와는 별개로, 고객의 평균 구매 금액을 추정하기 위함

## 모델 가정

1. 고객당 평균 구매금액은 감마분포(Gamma)를 따름
2. 개별 거래의 금액은 해당 고객의 평균 구매금액을 중심으로 변동하며, 이것도 감마 분포를 따름
3. 구매 횟수와 금액은 서로 독립임 (조건부 독립)

## 수식 구조

$$p(\bar{x}_i | p, q, \nu) = \text{Gamma-Gamma likelihood}$$

- $\bar{x}_i$ : 고객  $i$ 의 평균 구매 금액
- $p, q, \nu$ : 하이퍼파라미터
  - $p$ : shape parameter (개인 간 이질성 조절)
  - $q$ : rate parameter (집단 평균 수준 조절)
  - $\nu$ : 각 고객의 거래 수 (관측 횟수 반영)

이 모델은 고객이 얼마나 자주, 그리고 얼마나 많이 지출할지를 분리하여 추정

## BG/NBD + Gamma-Gamma 통합 LTV 추정

최종적으로 고객의 LTV는 아래와 같이 구성

$$LTV_i = \underbrace{\mathbb{E}[\text{구매 횟수}_i]}_{\text{BG/NBD}} \times \underbrace{\mathbb{E}[\text{구매당 금액}_i]}_{\text{Gamma-Gamma}}$$

각 모델은 개별 고객별로 예상 재구매 횟수와 평균 구매 금액을 추정하여,

두 값을 곱함으로써 개별 고객의 미래 가치(미래 매출 또는 수익)를 통합적으로 계산