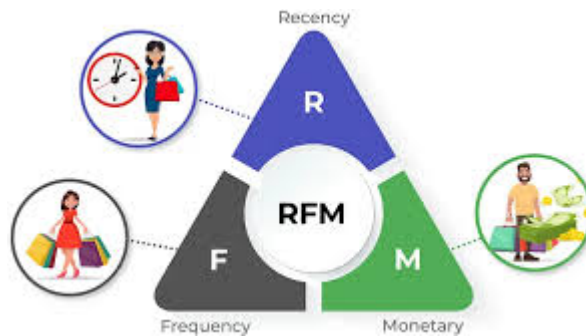




DataScience_CA01_RFM과 LTV

▼ 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:

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

- $m_{i,t}$: 시점 t 에서의 예상 구매 금액
- $P_{i,t}$: 시점 t 에서의 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 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.
