

# 『Marketing Science』

---

**Week 8 : [Customer Lifetime Value with Pymc Marketing]**

고객 생애 가치 : Customer Lifetime Value

기업이 한 고객과의 관계를 유지하는 동안 얻는 총 가치를 측정하는 지표

$$CLV_{\text{Nick}} = M \times \frac{r}{(1+r) - p}$$

$$CLV_{\text{Nick}} = 350 \times \frac{0.1}{(1+0.1) - 0.9} = 1750$$

M (평균 마진) : 초기 구매가 발생했다고 한다면

1을 +

r (할인율) : 미래 수익을 현재 가치로 환산한  
비율

p (이탈율) : 유저가 이탈할 가능성

“우리는 광고비를 얼마나 써야 될까요?”

“한 명의 유저를 획득하기 위해서 얼마나 비용을 써야 할까요?”

## 고객 생애 가치 : Customer Lifetime Value

기업이 한 고객과의 관계를 유지하는 동안 얻는 총 가치를 측정하는 지표

“우리는 광고비를 얼마나 써야 될까요?”

“한 명의 유저를 획득하기 위해서 얼마나 비용을 써야 할까요?”

$$1. CLV > CAC$$

$$2. \underbrace{CLV_{prospect}}_{\text{잠재고객의 고객생애가치}} = a(\text{전환율}) * \underbrace{CLV_{customer}}_{\text{기존고객의 고객생애가치}}$$

최대 비용이 정해졌다면, 달성해야 하는 전환율도  
구할 수 있다

## 고객 생애 가치 계산식의 문제점

1. 실제 CLV 는 고객 관계 종료 후에야 알 수 있다.
2. 기간(T)를 인위적으로 설정하는 문제 : 고객과의 관계는 언제까지 지속될까요?
3. 기간 0 또는 1부터 합산하는 차이 : 세그먼트 따라서 분리해서 봐야 하지 않을까요?
4. 이탈율은 일정하지 않습니다 : 유지율은 평균치가 아닙니다
5. 구독 VS 소매 : 진짜 떠났는지, 구매를 쉬고 있는 지, 행동만 하는 지 에 따라 달라집니다.

**HISTORICAL CLV**

VS

**Predictable CLV**

$$E(CLV) = \sum_t \text{expected net cashflow in period } t | \text{alive} \times P(\text{alive in period } t) \times \text{discount factor for period } t$$

예측 CLV / 특정 시점에 순 현금 흐름 X 특정 시점에 생존해 있을 확률 X 특정시점에 적용할 할인율

“우리는 해당 확률에 대해 예측하기가 어렵다”

BUY TILL YOU DIE

## BUY TILL YOU DIE

구매한 시점을 기준으로 그 이후의 행동을 확률적으로 예측하는 모델

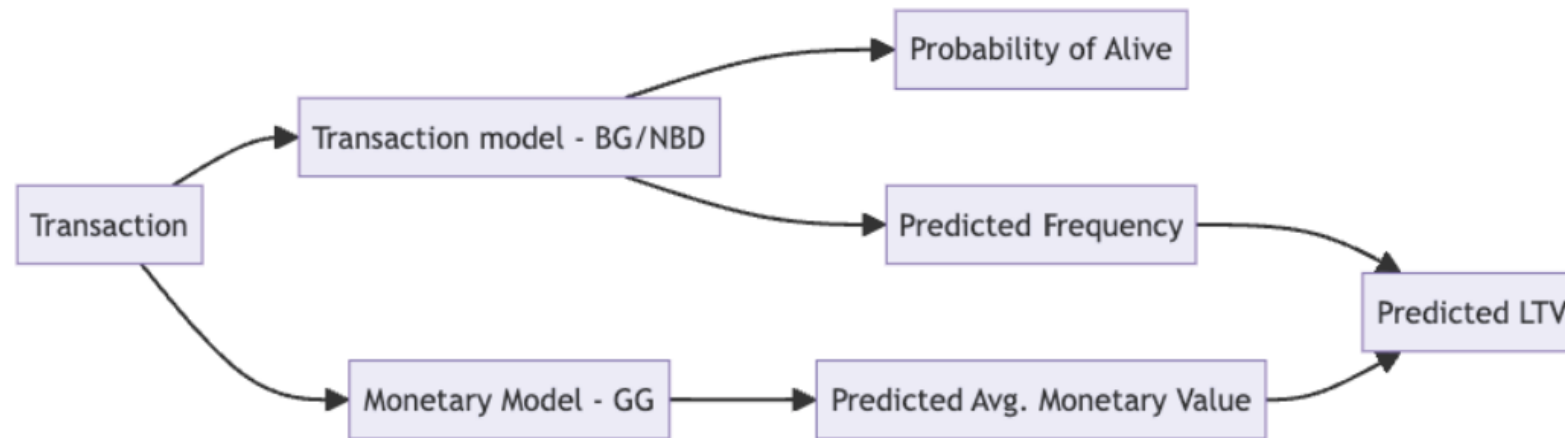


Figure 8.1 – The family of BTYD models to estimate the predicted LTV

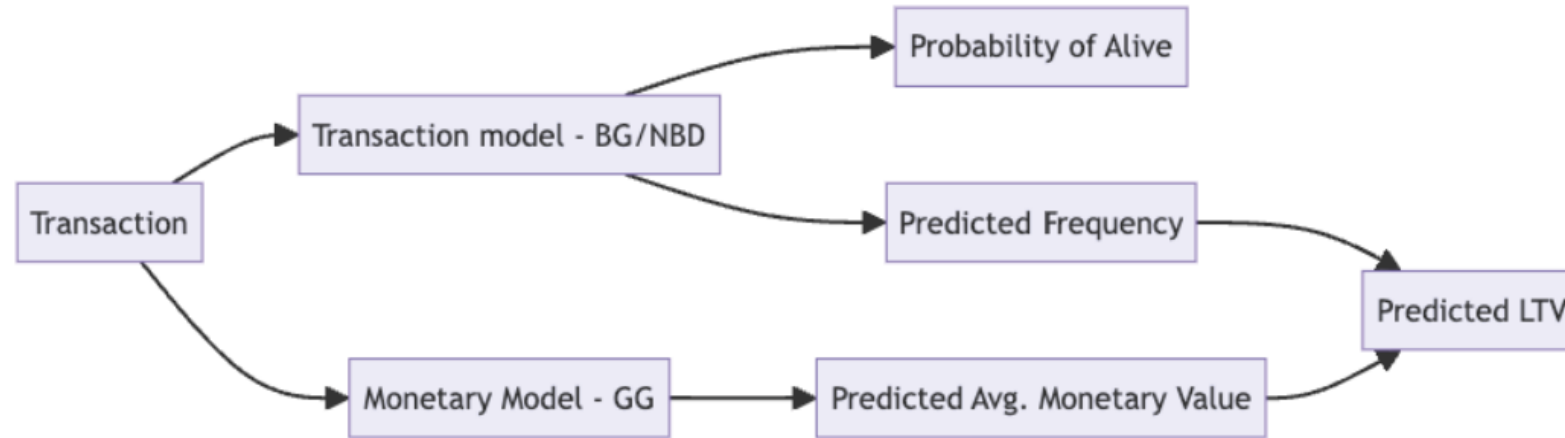


Figure 8.1 – The family of BTYD models to estimate the predicted LTV

## BG / NBD 모델

고객의 재방문 – 재구매 가능성 예측

## Gamma-Gamma 모델

각 구매의 평균 예상 금액 예측

## BG / NBD 모델

이탈 확률이 거래 직후 (즉시) 발생하는 모델

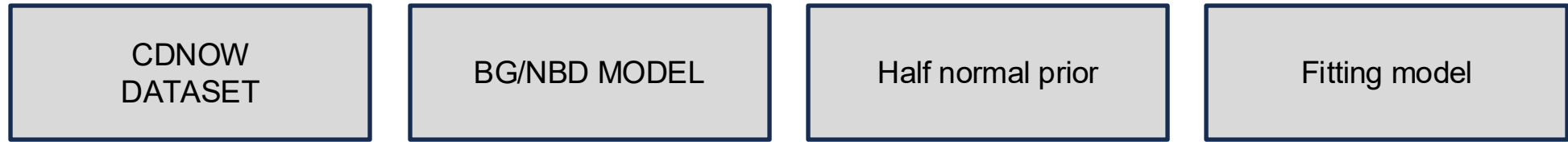
### 주요 핵심 가정 (수학적)

1. 거래율은 감마분포를 따른다  $f(\lambda|r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)}, \lambda > 0$
2. 거래 간 시간은 지수분포로 가정함  $f(t_j|t_{j-1}, \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}, t_j \geq t_{j-1} \geq 0$
3. 이탈 확률은 베타분포로 모델링  $f(p|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1-p)^{b-1}, 0 < p < 1$



## PYMC 실험 프로세스

PyMC Marketing + Arviz + Matplotlib 패키지 사용



향후 구매 고객 예측  
특정 고객 개별 분석  
향후 행동 예측  
생존 확률 추정  
추가 구매 시 생존 확률 변화

## PYMC 실험 프로세스

Estimating the CLV with Gamma – Gamma Model / BG-NBD

최소 1회 이상 반복 구매한 고객 대상  
모델이 학습 된 후, 평균 구매 금액의 분포를 추정할 수 있음

	mean	sd	hdi_ 3%	hdi_ 97%	mcse_ mean	mcse_ sd	ess_ bulk	ess_ tail	r_hat
p	6.408	1.408	4.110	9.113	0.078	0.057	380.0	363.0	1.0
q	3.787	0.301	3.255	4.365	0.015	0.011	394.0	452.0	1.0
v	16.189	4.518	7.986	24.685	0.236	0.167	347.0	334.0	1.0

고객 간 평균 지출 수준 차이 모델링  
개별 고객 내 구매금액 변동성  
스케일 조정

고객 별 평균 구매 금액 예측  
신규 고객의 평균 구매 금액 예측  
CLV 최종 예측

## 질문 1. WHY PYMC?

- A. Lifetime 패키지는 빈도론적 접근 -> 고객 수가 적을 시 불안정 + 도메인 지식 반영 어려움  
-> PYMC 베이지안 확률 추론 방식을 통하여, 불확실성을 포함한 예측이 가능

## 질문 2. 업무 적용방안?

- A. CLV 도출 -> 세그먼트 모델링 (ex.Kmeans) -> 의사결정 기준을 사후 확률로 표현  
ex) CLV 상위 확률 10% 이상일 확률이 80% 이상이면 리타겟팅 진행 (광고 연관)  
ex) 시나리오 모델링 : CLV [30000, 34000] 고객과 [20000, 25000] 고객 중 ROI test

## 1. NC : [게임 고객 LTV 추정하기](#) (2020)

### - 주요 방법론

- : 게임이라는 도메인이 주말/방학 등 외부 요인에 의하여 민감하게 반응함
- : ARPU 단위 이동평균을 통하여, 안정화 (비용이슈 포함) + 리텐션 추정 (잔존율 추정 및 근사치 방식)
- : sBG 함수 사용 but, over-estimate 될 가능성이 존재 -> 기간(t)를 무한대로 설정하는 것이 맞는 것인가?

## 2. [DATABRICKS NOTEBOOK](#) (2023)

- BYTD 주요 실행 방법
- CLV 예측 및 전체 시각화를 세부적으로 보여줌

## 3. [HOW Airbnb Measures Listing Lifetime Value](#) (2025)

- 발상의 전환 : 개별 고객이 아닌 숙소 단위로 LTV를 추정함
- : Baseline LTV / Incremental LTV / Marketing included Incremental LTV
- : 예측 정확도 / 세그먼트별 생산함수 추정 / 실시간 업데이트

4. [AIRBRIDGE : PLTV 탐구](#) with 딜라이트 룸

5. [META \\_LTVIsion](#)

\*\*\*\* PLTV 모델링을 보다 쉽게 구현하고 활용할 수 있게 함 \*\*\*\*  
고객별 INSIGHT 도출 및 PLTV 규모 예측 -> 고가치 유저 식별 / ROI 기반  
예산 할당

“무엇이 최선일까?”

Tradeoff – 어떻게 분석할 수 있을까?

Action Plan 연결성을 디테일하게 만들까?

# 감사합니다.

---