

# 연합 학술제 프로젝트

## 1. 분석 개요

### 1.1. 배경

트렌드 코리아 2023 도서에 따르면 평균 실종, 선제적 대응 기술이라는 키워드가 등장합니다.

- 평균실종이란 평균이 사라지고 양극화가 심화된다는 말로, 개인주의성향이 강해짐에 따라서 초다극화된 시장이 트렌드가 된 현상입니다.
- 선제적 대응 기술이란 고객이 불편함을 깨닫기도 전에 판매자가 먼저 고객의 불편함을 해결해주는 기술을 의미합니다.

따라서 저희는 해당 트렌드에 맞게 고객들에게 필요한 시기에 맞춤형 상품을 추천하는 방법을 제안합니다.

### 1.2. 분석 목표

- 현재 회사가 가지고 있는 고객들을 RFM 기반으로 분류하여 세그먼트별 특성 분석을 통한 마케팅 전략을 수립한다.
- 각 개인의 소비 데이터를 가지고 고객별 맞춤형 추천 시기와 추천 물품을 정의하여 CRM 마케팅 활동을 스케줄링 할 수 있도록 한다.

## 2. 고객 세분화 과정

### 2.1. RFM + T + CLV 변수 생성

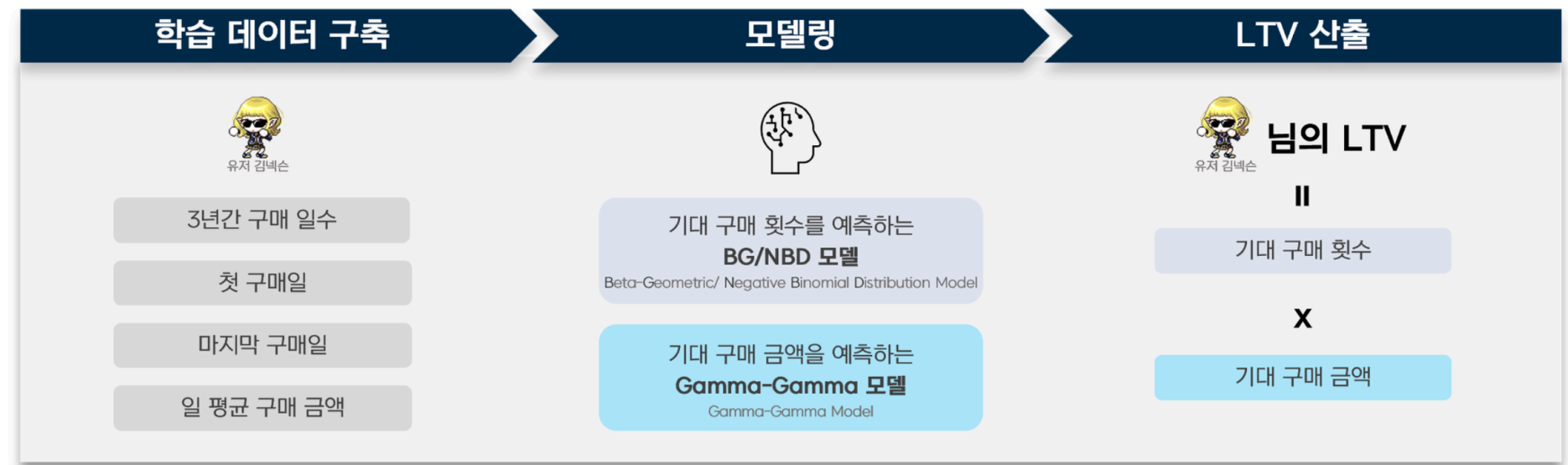
#### 2.1.1. RFM + T 변수 생성

CLV(고객생애가치) 예측을 위해 lifetimes 라이브러리를 사용하기에 해당 라이브러리에서 원하는 형태로 RFM + T 변수를 생성

- **Recency** : 마지막 구매일 이후 경과 일수
- **Frequency** : 총 거래 횟수 - 첫 구매 1회 제외
- **Monetary** : 개별 거래당 평균 금액으로 변환
- **Time** : 최초 구매 이후 현재까지 경과 일수

이후 구매 빈도와 구매 금액이 너무 높은 고객을 이상치로 판단하여 제거하였다.(상위 99.5% 위로만 제거, 16명의 고객 제거됨)

#### 2.1.2. CLV 변수 생성



(참고 : [인텔리전스랩스](#))

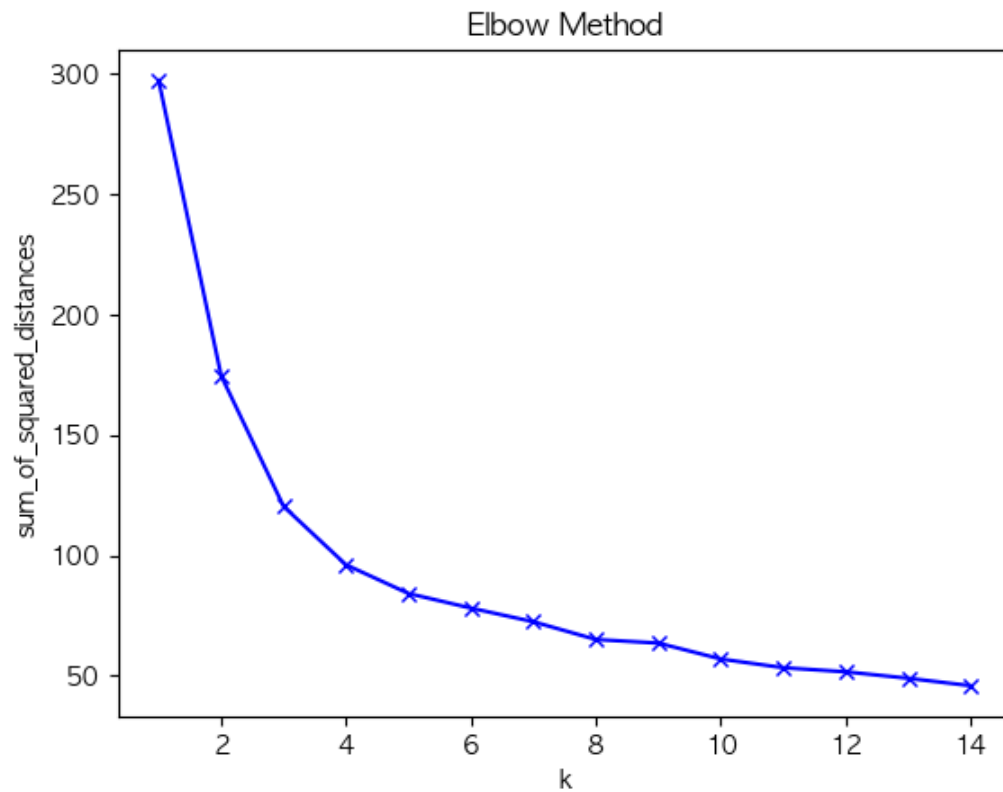
✓ **CLV = 기대 구매 횟수 \* 기대 구매 금액**

- 기대 구매 횟수 : BG/NBD 모델 사용
  - BG : “미래에 유저가 몇 번 구매할까?”에 대한 모델
    - 베타 분포 (확률에 대한 분포) + 기하 분포 (사건이 1번 발생할 때까지의 시행 횟수에 대한 분포)
  - NBD : “미래에 유저가 언제 구매를 중단할까?”에 대한 모델
    - 포아송 분포 (단위 시간 동안의 성공 횟수에 대한 분포) + 감마 분포 (사건을 n번 시행할 때까지의 총 시간에 대한 분포)
- 기대 구매 금액 : Gamma - Gamma 모델 사용

- “미래에 유저가 얼마씩 구매할까?”에 대한 모델
  - 감마 분포 : 0 이상의 범위를 가지고 비대칭적인 분포를 가지는 무언가에 모델링 할 때 좋은 분포

## 2.2. 클러스터링

KMEANS 클러스터링 적용 (변수 : RFM + T + CLV)

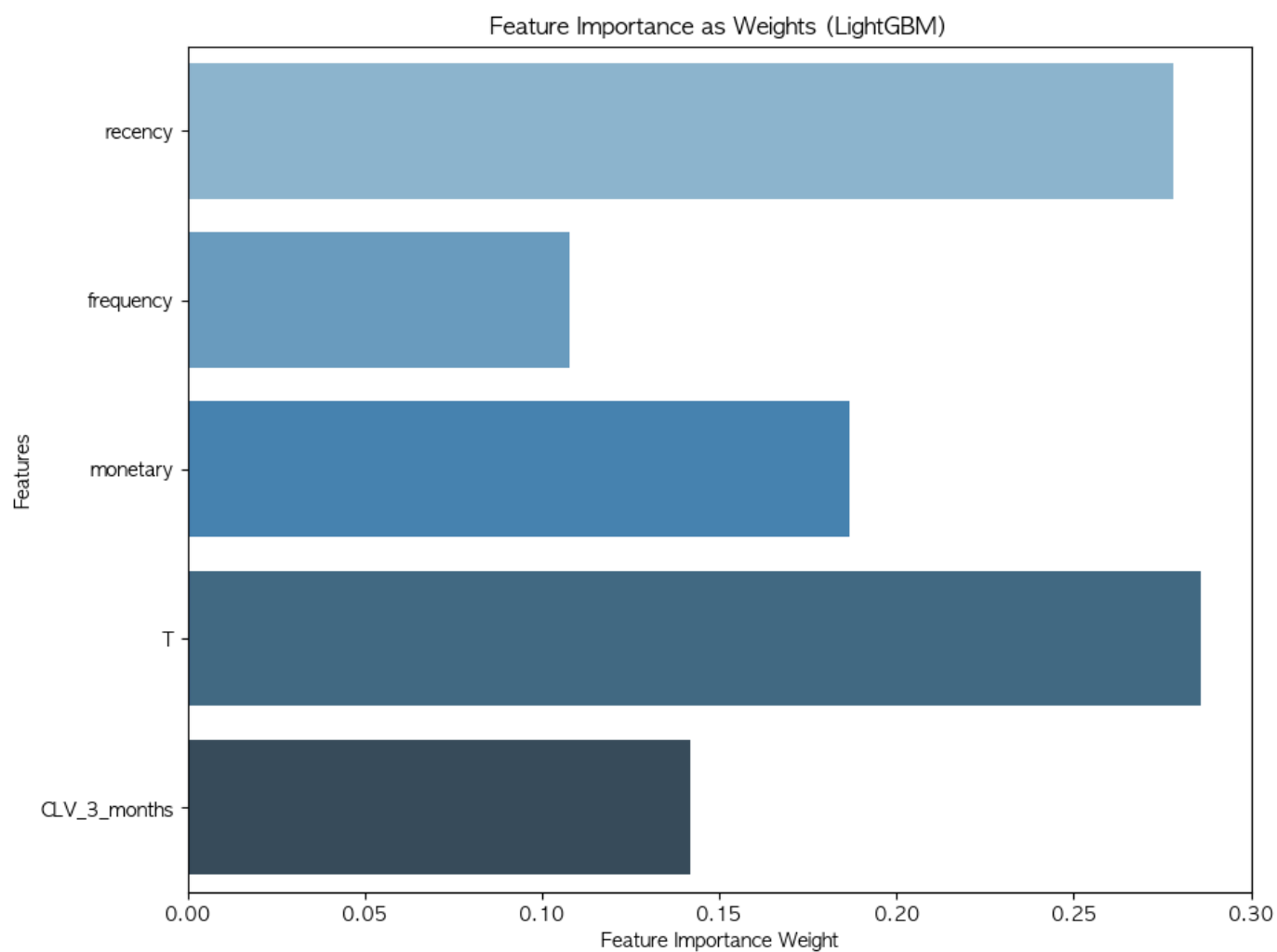


클러스터 개수는 4개로 결정하였음.

## 2.3. 클러스터링 분류 모델을 통한 변수 가중치 설정

고객들을 분류하는데 중요했던 변수에게 더 가중치를 주기 위해 진행

- LightGBM 모델 적용 : 다중 클래스 예측 가능, 빠른 속도, 검증된 성능 ( 정확도 0.986 )



- recency: 0.2782
- frequency: 0.1076
- monetary: 0.1868
- T: 0.2858
- CLV\_3\_months: 0.1417

## 2.4. 가중치를 반영한 최종 점수로 고객 등급 확정

- RFM + T + CLV 변수들을 4분위 구간별로 1~4까지 스코어링
- 이후 가중치 반영하여 final\_score 계산 → segment 분류
- segment : VIP(최상위 고객), Gold(충성 고객), Silver(우수 고객), Bronze(일반 고객), At Risk(이탈 위험 고객)
- segment별 고객 수

Customer Segment Distribution



- segment별 평균 사용 금액 비율
  - VIP, Gold 그룹이 절반 이상의 금액 비율을 차지하고 있음.

Average Monetary Contribution by Customer Segment



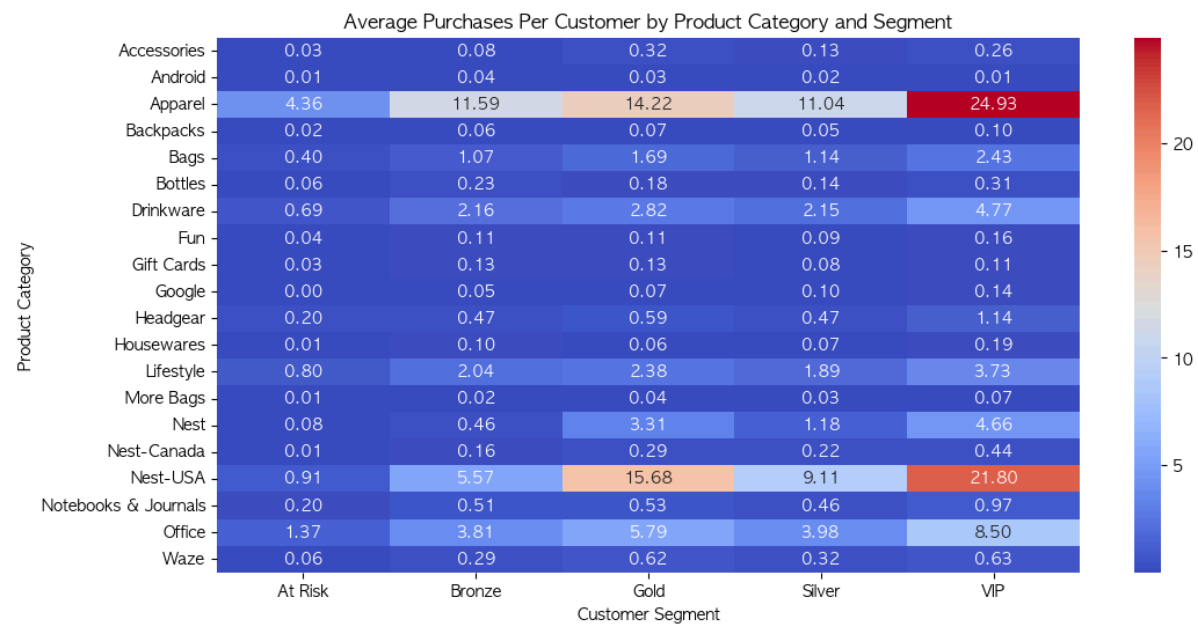
## 3. 고객 등급별 특징 분석 및 전략 제안

### 가설 설정

- 높은 세그먼트 고객일수록 비싼 카테고리의 상품을 많이 구매할 것이다.
- 높은 세그먼트 고객일수록 쿠폰과 같은 혜택을 적극적으로 탐색 및 활용할 것이다.
- 세그먼트에 상관없이 판매량이 높은 시즌이 있을 것이다.
- 세그먼트에 상관없이 평일보다 주말에 구매를 많이 할 것이다.

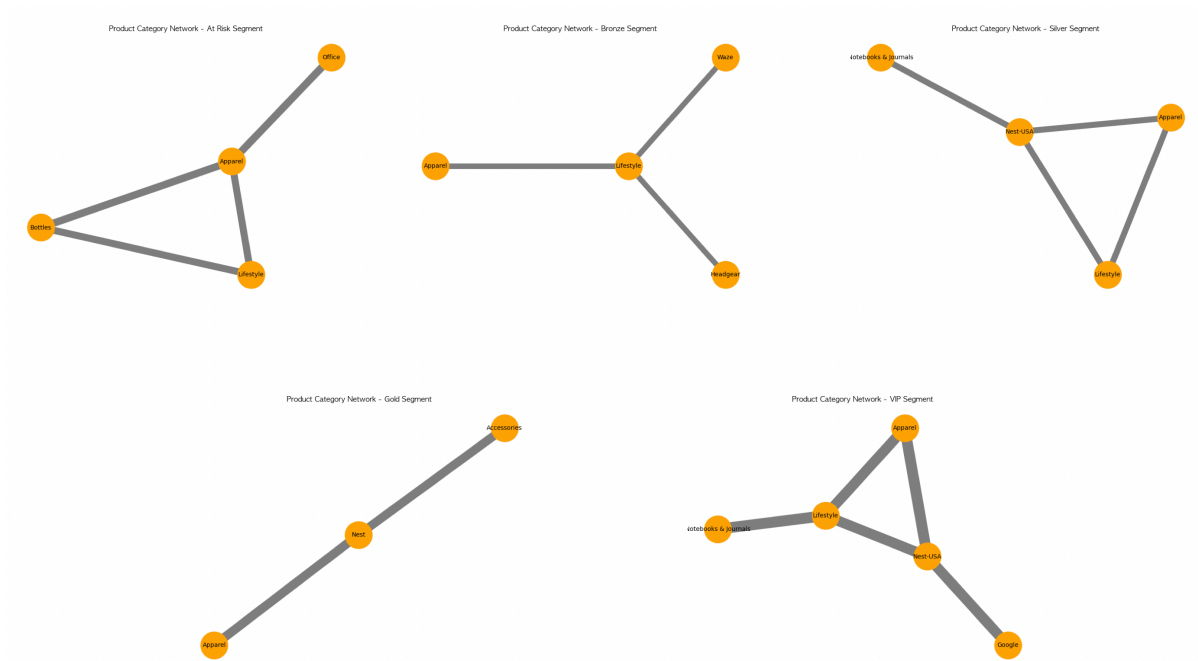
### 3.1. 높은 세그먼트 고객일수록 비싼 카테고리의 상품을 많이 구매할 것이다.

[세그먼트별 각 카테고리 평균 구매량 히트맵]



- Apparel, Nest-USA, Office는 전체적으로 판매량이 높은 카테고리
- 고비용 카테고리로 확인한 Nest-USA, Nest, Bags는 세그먼트가 높아질수록 구매량이 증가하는 것을 확인할 수 있다.

[ 세그먼트별 연관 구매 상품 네트워크 그래프 ]

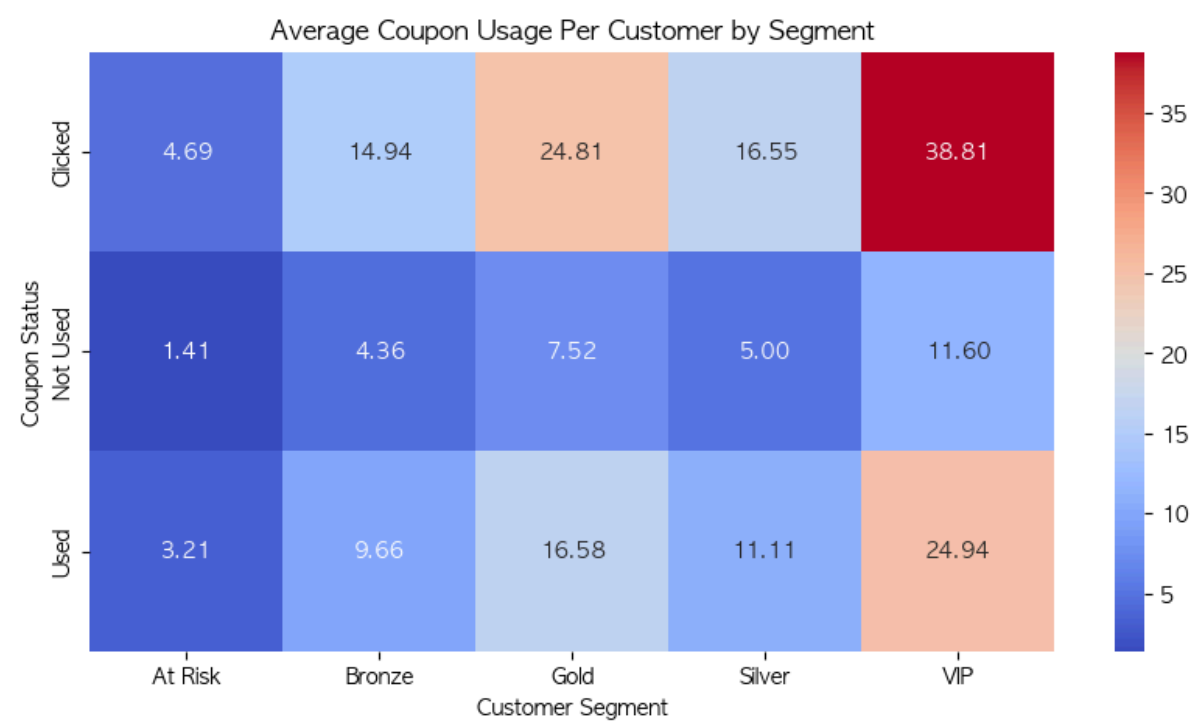


- 높은 세그먼트로 갈수록 향상도(Lift) 자체가 높음(=엣지의 굵기)
  - 향상도 : A가 주어지지 않았을 때 B의 확률 대비 A가 주어졌을 때 B의 확률 증가 비율(연관성)
- 높은 세그먼트로 갈수록 연관성이 있는 카테고리에 Nest-USA, Nest 등 고비용 카테고리가 등장

⇒ 높은 세그먼트 고객일수록 비싼 카테고리의 상품을 많이 구매한다.

### 3.2. 높은 세그먼트 고객일수록 쿠폰과 같은 혜택을 적극적으로 탐색 및 활용할 것이다.

[세그먼트별 평균 쿠폰 사용상태 히트맵]

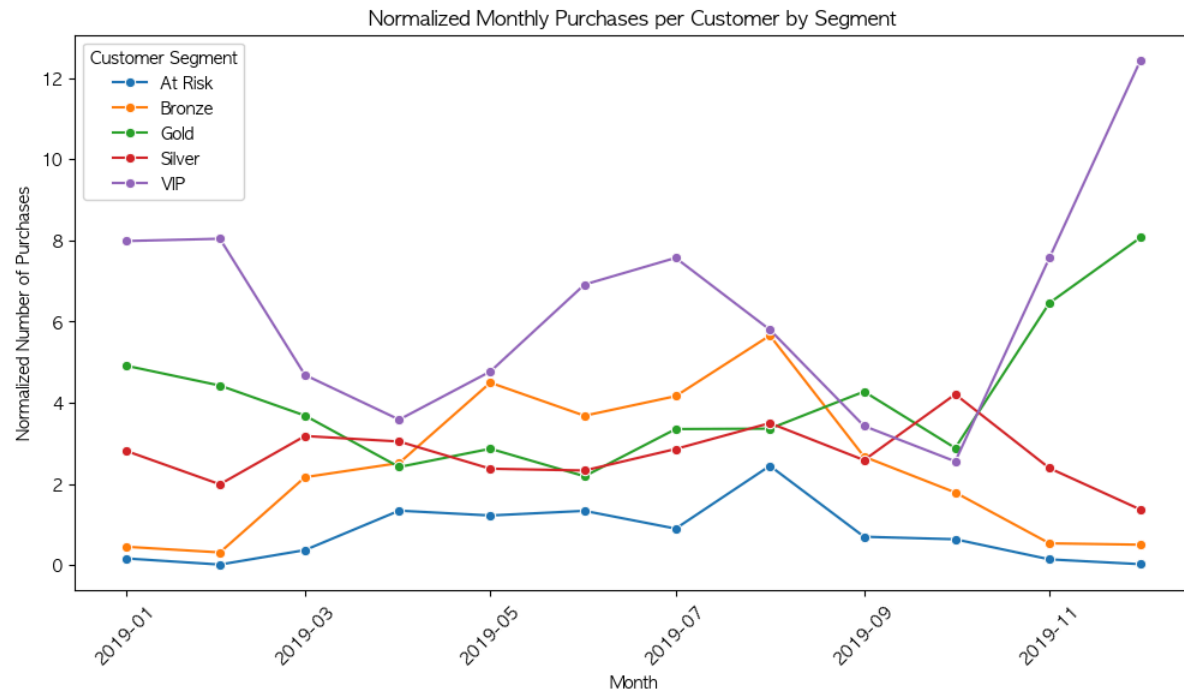


- 높은 세그먼트로 갈수록 모든 쿠폰 상태(사용/미사용/클릭)에서 평균 값이 훨씬 높은 것을 확인할 수 있다.
- 쿠폰 사용 횟수 자체가 거래량에 의존하다보니 더욱 많은 거래를 하는 고세그먼트로 갈수록 평균값이 높을 수 밖에 없기에 비율로 계산하여 다시 확인함.
- 비율로 봤을 때는 모든 세그먼트에서 사용/미사용/클릭 비율이 큰차이없이 유사하였다.

⇒ 쿠폰과 같은 혜택을 적극적으로 탐색 및 활용하는 것은 세그먼트와 큰 상관이 없다.

### 3.3. 세그먼트에 상관없이 판매량이 높은 시즌이 있을 것이다.

[세그먼트별 평균 월별 구매량 그래프]

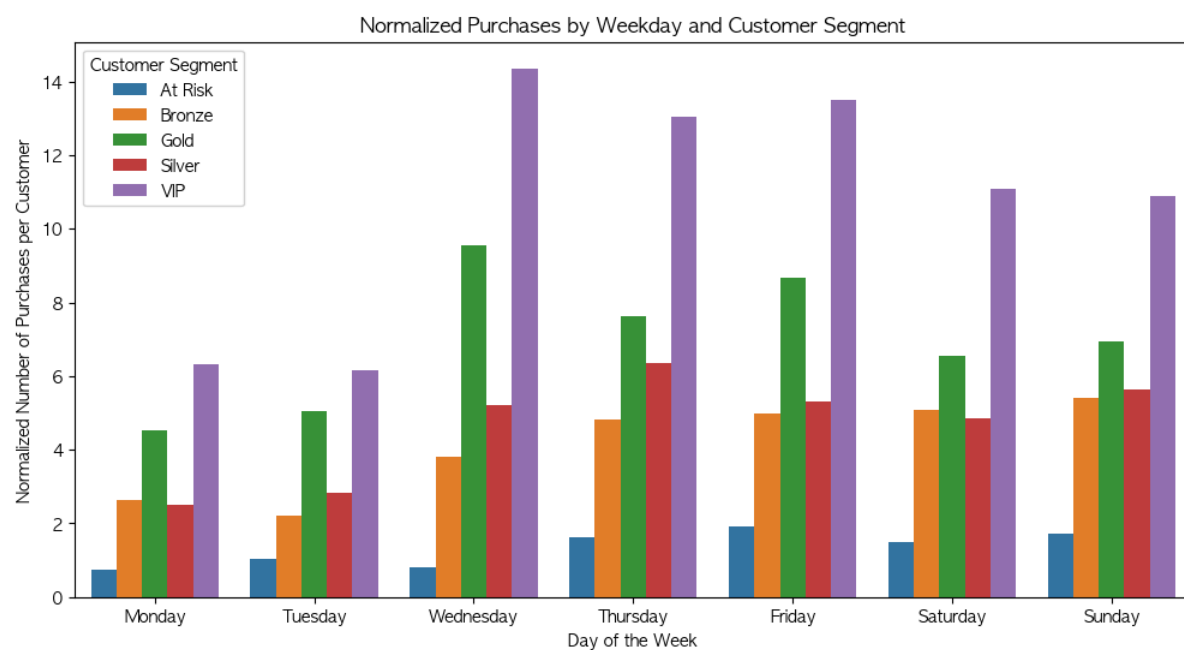


- 시즌별로 구매량이 튀는 세그먼트 : VIP, Gold, Bronze
  - VIP : 연초/여름/연말에 구매량이 매우 높게 오른다.
  - Gold : 연말에만 구매량이 매우 높게 오른다.(최신 활동이 많아 높은 세그먼트 할당 받은 것으로 보임)
  - Bronze : 5월, 8월에 구매량이 VIP와 비슷할정도로 매우 높게 오른다.(가족 행사와 관련 가능성)
- Silver, At Risk 세그먼트는 시즌에 상관없이 일정한 구매량을 보이는 것으로 보임.
- 세그먼트별로 판매량이 높은 시즌이 조금씩 다르지만, 여름 시즌은 확실히 모든 세그먼트에서 높은 판매량을 보임

⇒ 세그먼트에 상관없이 판매량이 높은 시즌이 있으며, 여름 시즌이다.

### 3.4. 세그먼트에 상관없이 평일보다 주말에 구매를 많이 할 것이다.

[세그먼트별 평균 요일별 구매량 그래프]



- 평일과 주말에 큰 차이를 보이지 않는다.
- 월, 화요일에는 모든 세그먼트에서 구매량이 저조하다.
- 주말보다 오히려 수,목,금에 구매량이 많다.

⇒ 세그먼트에 상관없이 평일보다 주말에 구매를 많이 하는 것은 아니다.

3.5. 세그먼트별 분석 결과에 따른 페르소나 정의

페르소나 정의



At Risk (이탈 위험 고객)



"가끔 필요할 때만 쇼핑하는 실용적인 고객"

- 가성비를 중시하는 최소 구매 고객
- 저렴한 제품만 제한적으로 구매
- 여름 시즌에만 집중적으로 구매
- 주로 목요일, 금요일에 구매



Bronze (일반 고객)

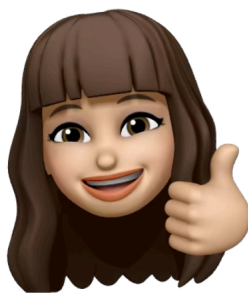


"일상생활에 필요한 제품을 주로 구매하는  
실속형 소비자"

- 생활 필수품을 구매하는 고객.
- 가전제품 및 생활에 필요한 제품을 주로 구매
- 5월, 8월까지 구매가 활발함 → 가족 행사 시기
- 일요일에 구매량이 높음



Silver (우수 고객)



"합리적이면서도 꾸준히 소비하는 스마트 고객"

- 꾸준히 구매하며 다양한 제품군을 체험하는 고객.
- 특정 시즌과 관계없이 일정한 구매 패턴을 유지.
- 주중과 주말 모두 일정한 구매 패턴을 보임.

페르소나 정의



Gold (충성 고객)



"브랜드에 대한 신뢰도가 높고 다양한 카테고리에서  
활발히 구매하는 핵심 고객"

- 브랜드 충성도가 높고, 다양한 제품군을 소비하는 고객.
- 꾸준한 구매 패턴을 유지하며, 고비용 상품과 연관 구매가 많음.
- 연말에 구매량이 증가하는 경향이 있음.
- 주로 수요일, 금요일에 구매량이 집중됨.



VIP (최상위 고객)



"품질과 브랜드 가치를 중시하는 프리미엄 소비자"

- 구매 빈도와 금액이 가장 높으며, 제품군이 매우 다양.
- 반드시 잡아야 할 고객층
- 연초/여름/연말 시즌에 구매량이 급증.
- 수 ~ 금요일에 구매량이 집중됨.

3.6. 세그먼트별 마케팅 전략

1 At Risk (이탈 위험 고객)

- 여름 시즌 할인 프로모션 & 무료 배송 혜택 제공

2 Bronze (일반 고객)

- 세트/묶음 판매 프로모션 진행
- 생활 필수품 정기 배달 서비스 제공

3 Silver (우수 고객)

- 정기 구독 서비스 제공

4 Gold (충성 고객)

- 프리미엄 제품 추천
- 연말 프로모션 집중 혜택

5 VIP (최상위 고객)

- 인기 상품 사전 예약 서비스
- 연초/여름/연말 프로모션 집중 혜택

# 4. 개인 고객별 맞춤형 상품과 추천 시기 예측

- 언제 살건데? (WHAT?)
  - 뭘 살건데? (WHEN?)
- ⇒ 언제, 무엇을 살지 조합하여 개인 맞춤형 마케팅 활동 계획 가능

## 4.1. 고객들의 평균 구매 주기 예측하기

**데이터** [1452 rows x 11 columns]

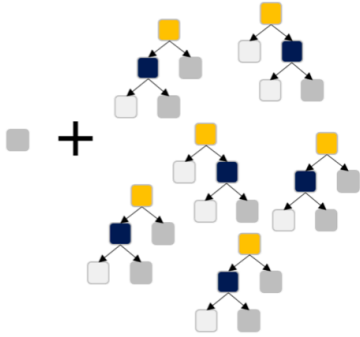
독립변수		종속변수
범주형	수치형	수치형
성별  고객 지역	recency	평균구매간격
	frequency	
	monetary	
	T	
	CLV	
	쿠폰사용비율	
	월평균구매횟수	
	가입기간	

StandardScaler

**모델**

Gradient Boosting

- A leaf and restricted tress
- With equal learning rate



- 범주 + 수치형을 모두 수용 가능한 트리 모델에서 선택하고자 함
- 행이 적은 편이므로, 트리의 깊이가 깊지 않은 모델을 고려
- 실제 적용시 성능이 우수

- 수치형 변수 StandardScaler 적용
- Gradient Boosting Regressor 사용
- 평가 지표 : RMSE(큰 오차를 민감하게 잡음), MAE(전반적인 오차 수준 확인)

결과 : RMSE: 2.12, MAE: 0.72

## 4.2. 고객들의 연관 구매 상품 추천하기

- 고객별 거래 날짜를 기준으로 구매한 제품ID를 그룹화
- 이후 Apriori 알고리즘을 사용하여 연관 규칙 생성
- 연관성의 지표인 Lift(향상도)가 높은 상위 5개의 상품을 출력하도록 설정

결과 (유저별 추천제품 5개 )

고객ID		추천제품
38120	USER_0000	[Product_0992, Product_0990, Product_0969, Pro...
10667	USER_0001	[Product_0880, Product_0916, Product_0916, Pro...
23561	USER_0002	[Product_0880, Product_0916, Product_0880, Pro...
50740	USER_0003	[Product_0880, Product_0916, Product_0916, Pro...
38084	USER_0004	[Product_0880, Product_0880, Product_0880, Pro...
...	...	...
12216	USER_1463	[Product_0880, Product_0985, Product_0908, Pro...
23496	USER_1464	[Product_0916, Product_0916, Product_0916, Pro...
12242	USER_1465	[Product_0994, Product_0880, Product_0976, Pro...
43154	USER_1466	[Product_0880, Product_0880, Product_0916, Pro...
29342	USER_1467	[Product_0880, Product_0916, Product_0880, Pro...

## 4.3. 평균 구매 간격 예측 모델 & 연관 상품 추천을 활용한 솔루션

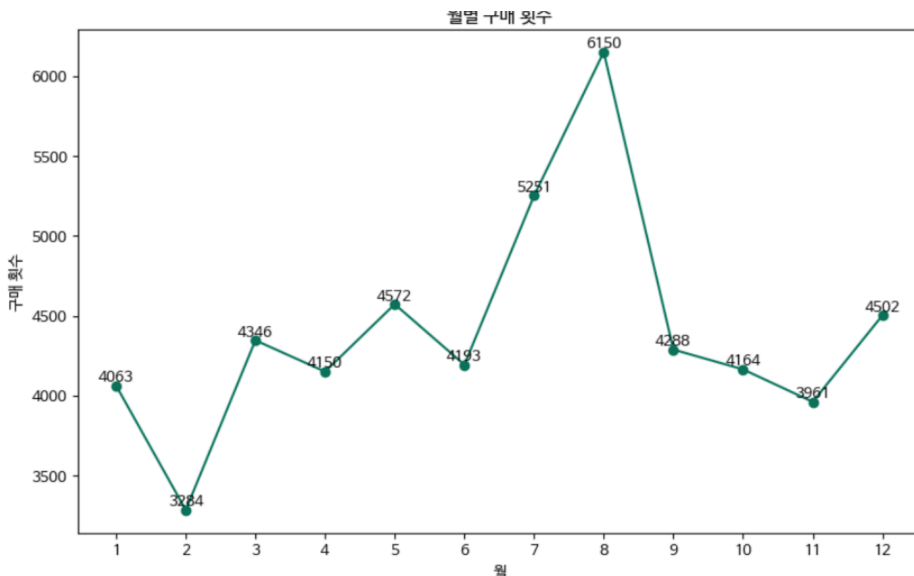
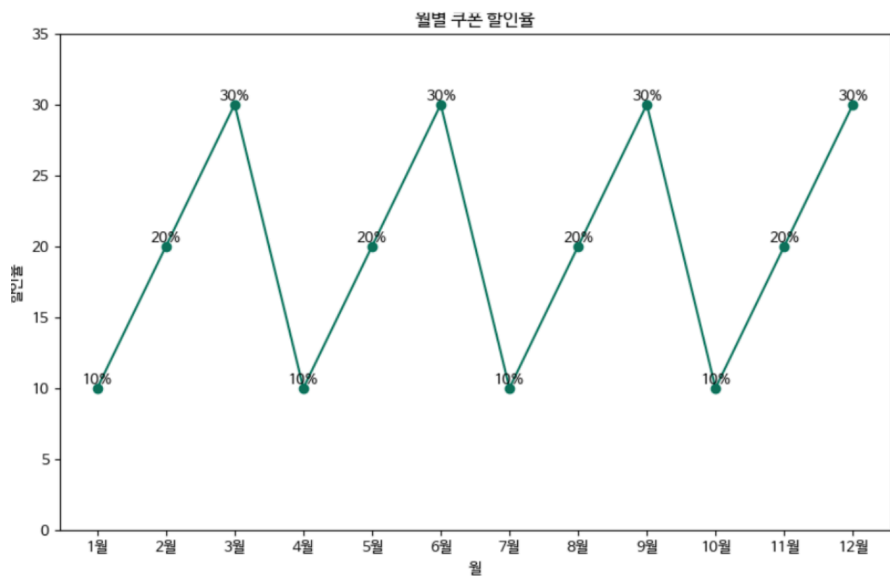
구매 주기 예측 기반 제품 추천 자동화

- 각 고객별로 예측된 구매 주기에 따라 FC 메시지 자동화 세팅 가능
- 메시지 내용 예시
  - “OO고객님, XX 제품을 찾고 계신가요? N% 할인 중”



# Appendix

## [쿠폰 발행 솔루션]



- 월별로 모든 제품에 같은 할인율의 쿠폰을 발행
- 고객들의 구매 패턴 파악 및 시즌별 수요 높은 제품을 식별하여, 해당 제품에 대한 집중 할인 쿠폰
- 제품의 수요와 재고 상황을 실시간으로 모니터링하여, 상황에 맞게 할인율 유연하게 조정 추천

## [ 추가 데이터 솔루션 ]

- 고객들의 행동 로그 데이터는 제공되지 않아 더 딥다이브한 분석이 어려움
  - 트래커를 설정하여 행동 데이터 분석 데이터 추가 적재 추천
- 제품 선호도와 같은 데이터를 수집한다면 고도화된 추천 시스템 모델 개발 가능