

# 구독 이탈 원인 분석 및 고객 유지 전략 수립 보고서

마케터: 이진

2025.02.20

## 1. 개요

본 보고서는 한 구독 기반 서비스(Subscription Service) 운영팀이 고객 데이터를 기반으로 구독 유지율을 높이기 위한 KPI를 설정하고 분석하는 데이터 보고서를 작성하려고 함. EDA(탐색적 데이터 분석) 및 데이터 전처리를 수행하여 고객 이탈 패턴을 분석하고, 시각화하여 고객 리텐션 전략을 제안하는 것이 목표.

## 2. 데이터 분석

subscription\_data.csv

총 2,000 명의 구독 고객 데이터를 포함하며, 각 고객의 구독 상태, 서비스 이용 패턴, 프로모션 참여 여부 등의 정보를 포함.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
!pip install koreanize-matplotlib
import koreanize_matplotlib

df=pd.read_csv('subscription_data.csv')
df.isnull().sum()

df["월평균이용시간(시간)"]=df["월평균이용시간(시간)"].fillna(df["월평균이용시간(시간)"].mean())

df["결제금액"]=df["결제금액"].fillna(df["결제금액"].median())

df.isnull().sum()
```

```

#구독유지고객 vs 구독이탈고객 결제금액 분포
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

plt.figure(figsize=(6, 5))
sns.boxplot(x=df["구독상태"], y=df["결제금액"])
plt.xticks([0, 1], ["이탈", "유지"])
plt.title("구독유지고객 vs 구독이탈고객 결제금액 분포")
plt.xlabel("구독여부")
plt.ylabel("월 결제 금액")
plt.show()
t_stat, p_value = stats.ttest_ind(df[df["구독상태"] == 1]["결제금액"],
                                   df[df["구독상태"] == 0]["결제금액"],
                                   nan_policy='omit')

print(f"\nT-검정 통계량: {t_stat}, P-값: {p_value}")

#서비스이용횟수가 높은 고객과 낮은 고객의 이탈률 비교
grouped = df.groupby("서비스이용횟수")["구독상태"].mean().sort_values()
plt.figure(figsize=(15, 4))
sns.barplot(x=grouped.index, y=grouped.values)
plt.title("서비스이용횟수가 높은 고객과 낮은 고객의 이탈률 비교")
plt.xlabel("서비스이용횟수")
plt.ylabel("이탈률")
plt.show()

#월평균이용시간이 높은 그룹과 낮은 그룹의 이탈률 비교
plt.figure(figsize=(6, 4))
sns.boxplot(x=df["구독상태"], y=df["월평균이용시간(시간)"])
plt.xticks([0, 1], ["이탈", "유지"])
plt.title("월평균이용시간이 높은 그룹과 낮은 그룹의 이탈률 비교")
plt.show()

#월평균이용시간이 낮은 그룹과 높은 그룹의 이탈률 비교
print(df.groupby("구독상태")["월평균이용시간(시간)"].nlargest(5).groupby("구독상태").mean()[0])
print(df.groupby("구독상태")["월평균이용시간(시간)"].nlargest(5).groupby("구독상태").mean()[1])
print(df.groupby("구독상태")["월평균이용시간(시간)"].nsmallest(5).groupby("구독상태").mean()[1])
print(df.groupby("구독상태")["월평균이용시간(시간)"].nsmallest(5).groupby("구독상태").mean()[0])

#할인 사용 여부에 따른 구독 여부 비교
plt.figure(figsize=(8, 5))
sns.barplot(x=df["할인사용여부"], y=df["구독상태"])
plt.title("할인 사용 여부에 따른 구독 여부 비교")
plt.xlabel("할인사용여부")
plt.ylabel("구독여부")
plt.show()

```

```

#추천 가입 여부에 따른 구독 여부 비교
plt.figure(figsize=(8, 5))
sns.barplot(x=df["추천가입여부"], y=df["구독상태"])
plt.title("추천 가입 여부에 따른 구독 여부 비교")
plt.xlabel("추천 가입 여부")
plt.ylabel("구독 여부")
plt.show()

#고객 유형별 평균 구독 기간
plt.figure(figsize=(6, 4))
sns.barplot(x=df["고객유형"], y=df["구독기간(개월)"], estimator=np.mean)
plt.title("고객 유형별 평균 구독 기간")
plt.xlabel("고객 유형")
plt.ylabel("평균 구독 기간(개월)")
plt.show()

df.groupby("고객유형")["구독기간(개월)"].mean()

#광고 채널별 평균 구매 금액
print(df.groupby("고객유형")["결제금액"].mean())
plt.figure(figsize=(8, 5))
sns.barplot(x=df["고객유형"], y=df["결제금액"], estimator=np.mean)
plt.title("광고 채널별 평균 구매 금액")
plt.xlabel("광고 채널")
plt.ylabel("평균 구매 금액")
plt.show()

```

## 2.1.구독 유지 고객 vs 이탈 고객 분석

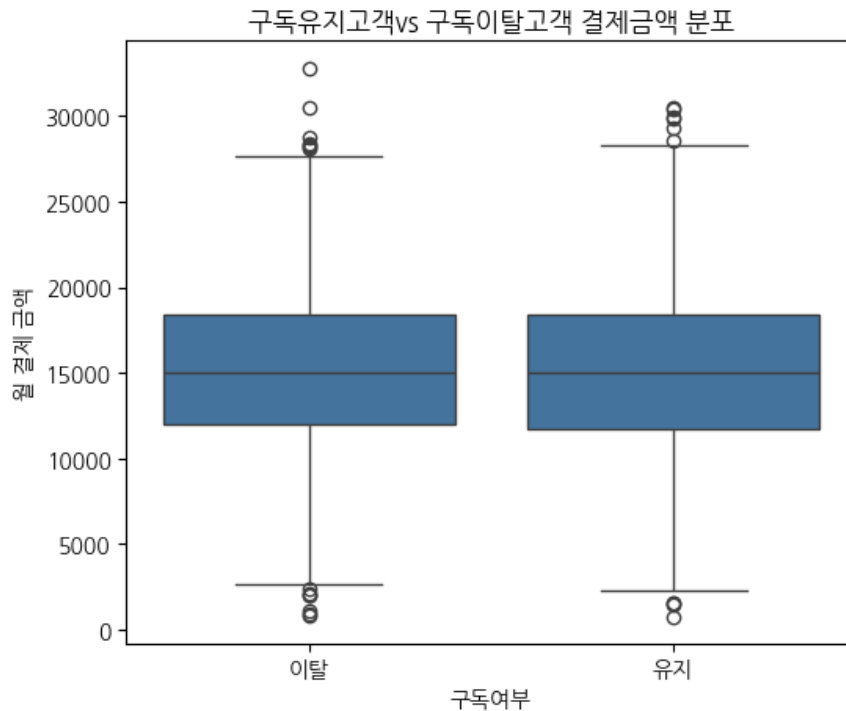
### 분석

<구독 유지 고객과 이탈 고객의 서비스 이용 횟수 평균과 월평균 이용 시간 비교>

서비스이용횟수 월평균이용시간(시간)		
구독상태		
0	25.184943	20.305087
1	24.820014	20.046414

- 구독 유지 고객과 이탈고객의 서비스 이용 횟수와 월 평균 이용시간 차이는 미미함.

### <구독여부에 따른 결제 금액 비교>



- 구독 이탈 고객과 유지 고객 전반적으로 최솟값부터 25%까지의 영역과 최댓값부터 75%까지에 해당하는 영역이 비슷한 것으로 보아 월 결제 금액에 차이가 없는 것으로 보임.
- 구독 여부에 관계없이 이상치가 있지만 이 역시 양쪽 모두 비슷한 양상임.
- T-검정 통계량: -0.9235357544707836, P-값: 0.35583959185176905임.
- p값이 0.05보다 높아 구독 여부와 월 결제금액 사이의 관계는 통계적으로 유의미한 차이를 보이지 않는 것으로 나타남.

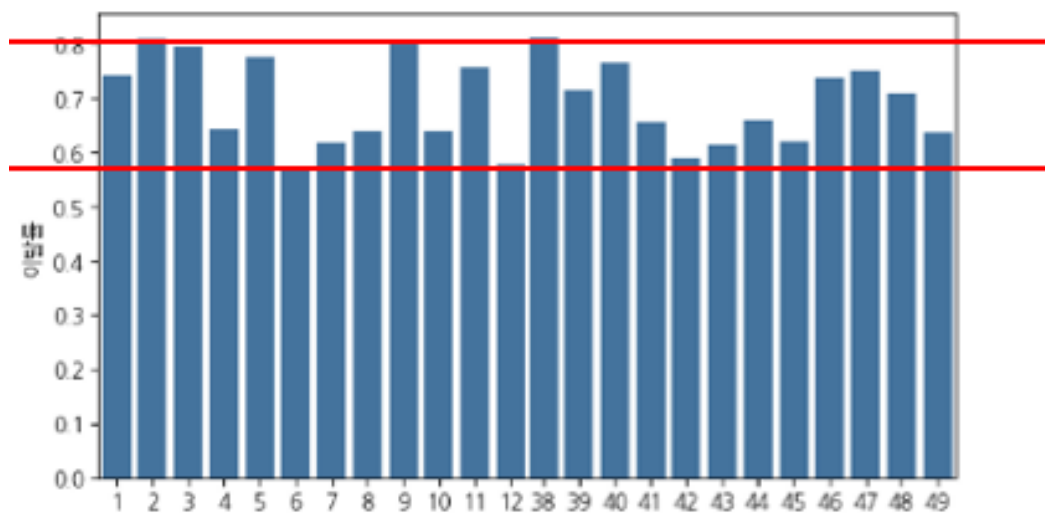
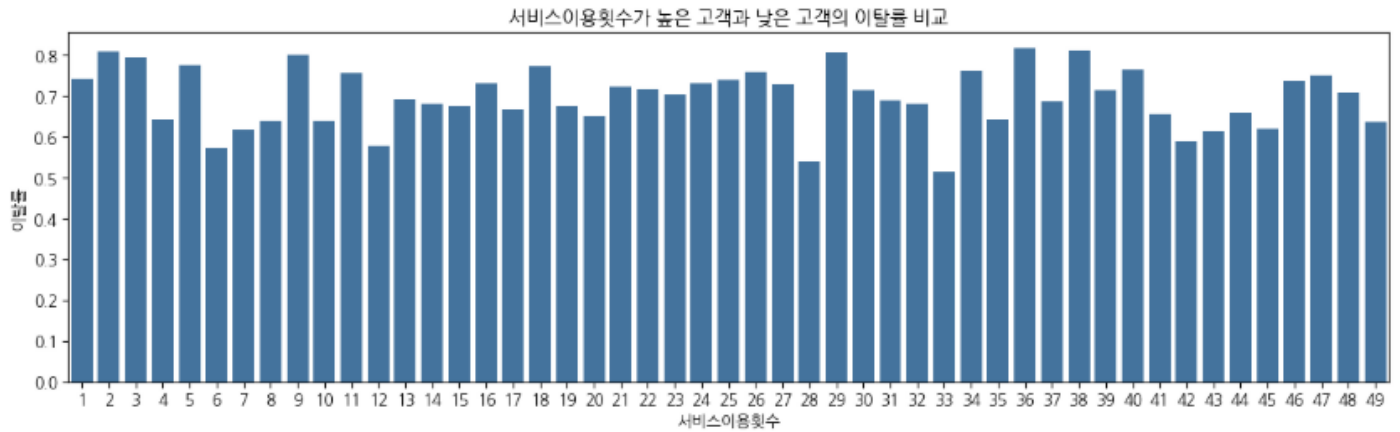
### 개선방향

- 구독 여부에 영향을 미치는 다른 특성(결제금액제외)을 파악 해야함.
- 구독 여부에 관계없이 두 집단의 사용시간이 비슷함.
- 구독자에게 제공되는 개인 맞춤 서비스의 정확도가 높기 때문에 체류시간이 단축되는 것인지, 개인화 서비스에 오류가 발생해 체류시간을 지속하지 못하는 것인지 따져볼 필요가 있음.
- 구독 여부에 상관없이 두 집단의 이용횟수가 비슷함.
- 구독 이탈시 와 유지 시 비슷한 수준의 서비스 제공이 원인.
- 구독자만의 특별한 혜택 프로모션이 필요.

## 2.2. 서비스 이용 패턴과 이탈률 분석

### 분석

<서비스 이용횟수와 고객 이탈률 비교>



<가시성을 위해 서비스 이용횟수 1/4 와 4/4 부분을 붙여 편집한 그래프>

- 서비스 이용횟수가 1~12회로 낮은 그룹에 속하는 고객들의 이탈률이
- 서비스 이용횟수가 38~49회로 높은 그룹에 속하는 고객들의 이탈률보다 높음.

<월평균이용시간이 낮은 그룹과 높은 그룹의 이탈률 비교>

	구독 유지	구독 이탈
월평균이용시간 높은 그룹 (오름차순 정렬 후 하위 5개값 평균)	47.99	45.96
월평균이용시간 낮은 그룹 (오름차순 정렬 후 상위5개값 평균)	0.098	0.64

- 월평균이용시간이 높은 그룹의 이탈률 차이가 작음(1)
- 월평균이용시간이 낮은 그룹의 이탈률 차이는 약간 있음.(0.5)
- 전반적으로 월평균이용시간이 이탈률에 미치는 영향은 크지 않음.

<고객 유형별 이탈률 분석>

VIP	0.717593
일반	0.678748
프리미엄	0.719298

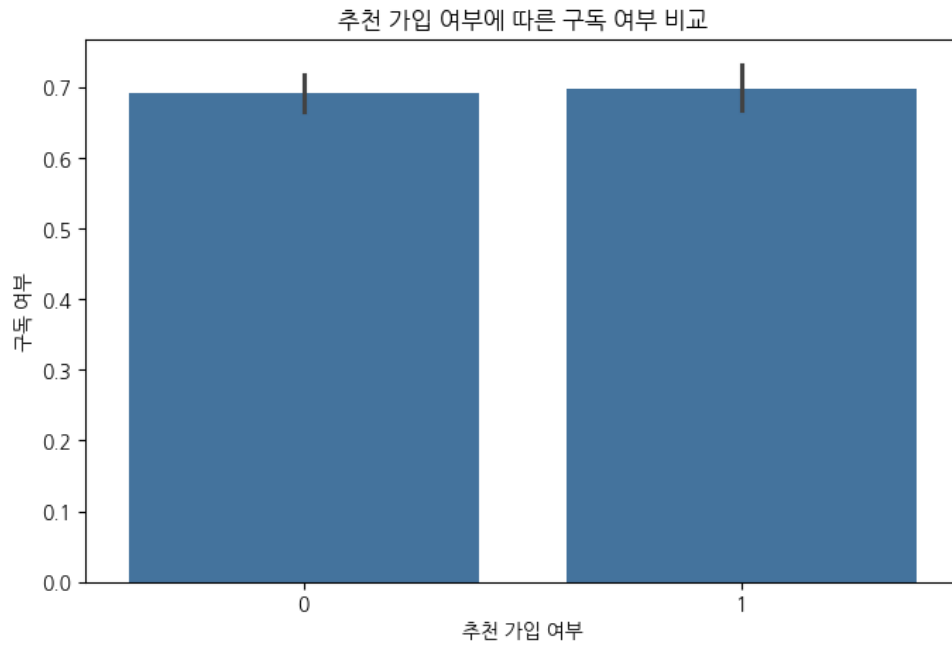
- 이탈률이 가장 높은 고객유형은 프리미엄 고객임.
- VIP고객도 프리미엄 고객만큼 이탈률이 높음
- 일반고객은 비교적 이탈률이 낮음.

**개선방향**

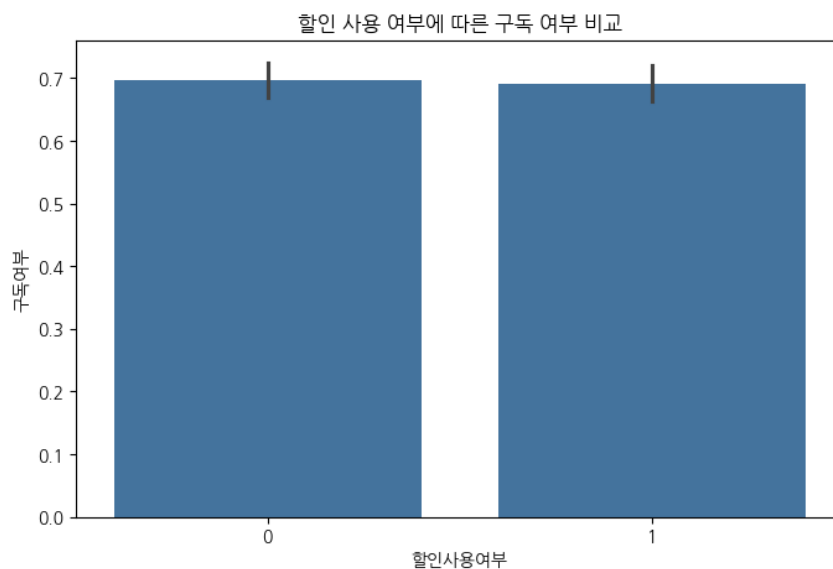
- 서비스이용횟수가 낮은 고객의 이탈률이 높음
- 월평균이용시간이 낮은 고객의 이탈률이 높음
- 프리미엄고객과 vip고객의 이탈률이 높음
- 서비스 이용횟수 증가를 위해 구독자의 개인 맞춤형 알고리즘 정확도 개선 필요
- 서비스 이용 활성화를 위한 맞춤 콘텐츠 추천 푸시 알림
- 고객 체류시간 증가를 위한 CTR과 전환율 분석 후 페이지 개선
- 고객 리텐션의 핵심인 프리미엄과 vip고객을 위한 섬세한 관리 필요
- 프리미엄, vip고객만을 대상으로한 특별 맞춤 서비스 제공
- 상품권, 특정 서비스 무료 체험 제공 등 브랜드 충성에 대한 보답

### 2.3.할인 및 프로모션 효과 분석

#### 분석



- 할인사용이 구독 유지 및 이탈에 주는 영향이 미미함.
- 할인 시용의 영향을 받는 고객 유입과 고객 유지는 다른 문제인 것으로 보임
- 할인률이 적거나 구독료 대비 서비스의 품질 만족도가 떨어지는 것으로 보임



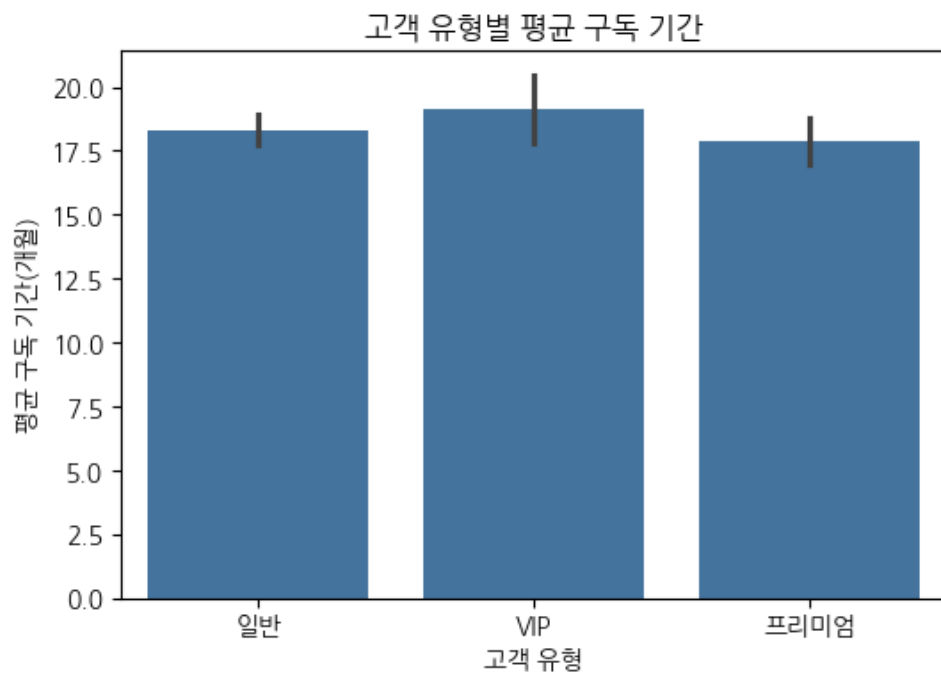
- 추천 가입이 구독 유지 및 이탈에 미치는 영향이 미미함.
- 추천 시 가입자가 받는 혜택이 적은 것으로 보임.

## 개선방향

- 구독 기간 누적에 따른 할인률 상승 및 다양한 서비스 체험 가능 쿠폰 제공
- 추천 시 가입자에게 제공하는 프로모션 개선
- 추천인과 가입자의 관심 서비스 알고리즘 상관관계 분석 필요
- 구독 유입 경로 확대를 위한 광고 채널 마케팅 예산 증가

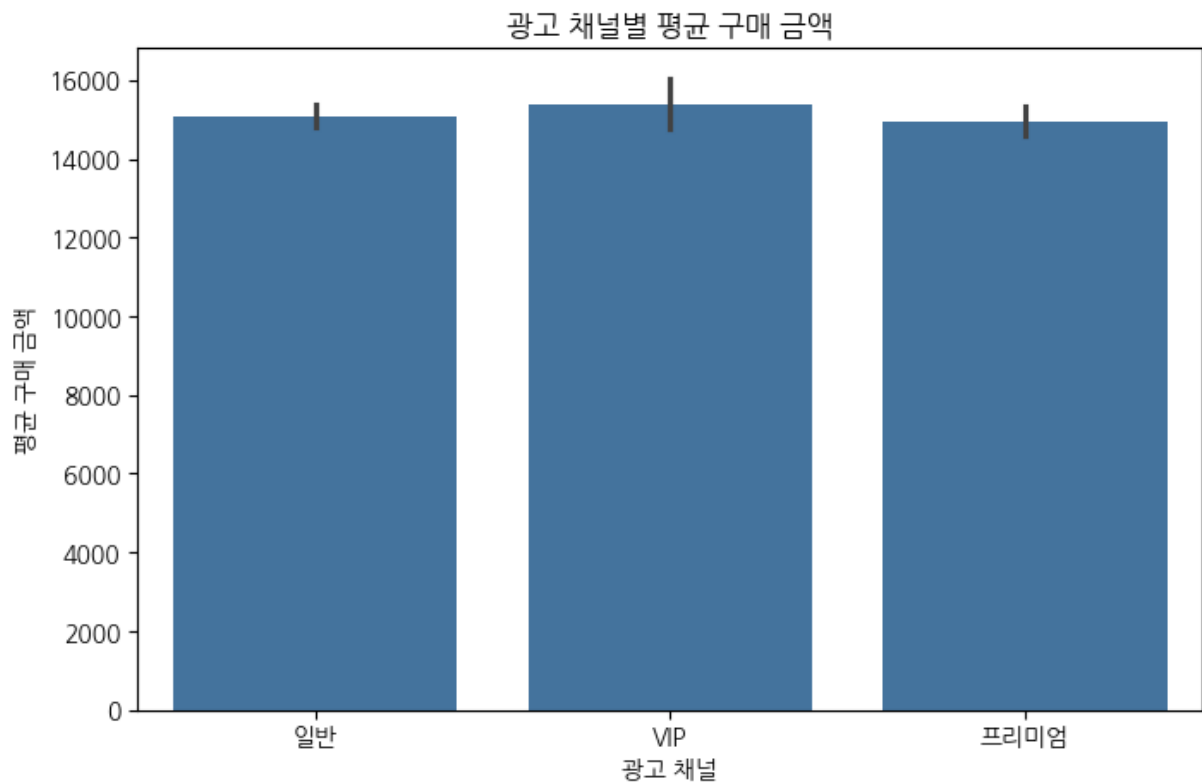
## 5. 고객 유형별 구독 유지 패턴 분석

### 분석



- Vip 고객은 평균 19개월, 일반고객은 18개월, 프리미엄 고객은 17개월 구독을 하며 vip 고객의 구독기간이 가장 김.
- 리텐션의 핵심인 프리미엄 고객의 구독 기간이 비교적 짧은 것으로 보아 개인화 마케팅 전략 수정 필요.





- vip고객은 약15,400원, 일반 고객은 약15,100원, 프리미엄 고객은 약 15,000원. 차례대로 평균 구매가격이 높음.
- 프리미엄 고객의 평균 구매금액이 전체 매출 부진에 영향을 줄 것이라 봄.

### 개선방향

- 프리미엄 고객 서비스 품질 향상은 매출 증가로 이어질 가능성이 높음
- 프리미엄 고객층 대상 설문조사 실시
- 다른 고객층과 차별화된 프리미엄 고객층만의 고유 리텐션 전략 수립
- 새로운 서비스 개시 전 베타 체험권 제공
- 전반적인 고객 구매금액 증가를 위해 타겟별 맞춤화 상품 리스트 추가 및 수정

## 6. 최종 결론

이번 분석을 통해 고객 행동 패턴을 기반으로 고객 리텐션 전략을 최적화할 수 있는 방향을 도출할 수 있었다.

첫 번째 전략으로 구독 유지 고객만을 위한 개인 맞춤화 서비스 및 혜택을 제공한다.

두 번째 전략으로 구독 이탈을 줄이기 위해 서비스 품질 개선 및 알고리즘 정교화를 위한 예산을 집중 투자한다..

세 번째 전략으로 전환율이 높은 광고 채널에 집중 투자하고, 광고 성과를 지속적으로 분석하여 최적화한다.

네 번째 전략으로 충성 고객을 유지하기 위한 프리미엄고객 혜택과 맞춤형 리텐션 캠페인을 운영한다.

이러한 전략을 실행하면 고객 유지율을 증가시키고 구매 전환율을 개선할 수 있다.