

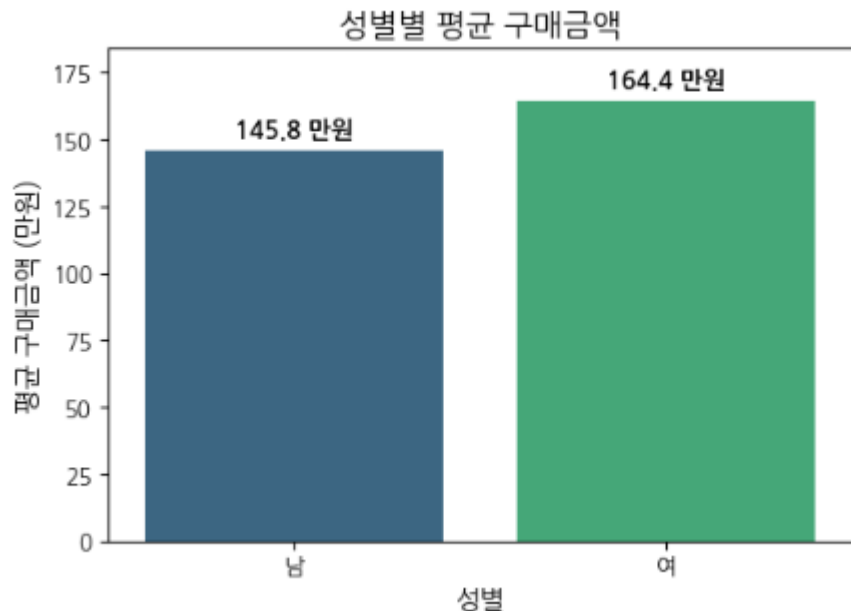
# AI 도구를 활용한 데이터분석

# 데이터 소개

- 항목: 이름, 성별, 나이, 구매금액(만원), 가입연도, 회원등급
- 데이터 수: 200명 (가공 고객 데이터)
- 주요 분석 대상: VIP vs 일반 고객

# 막대그래프

-목적: 성별별 평균 구매금액



성별과 나이에 따라 VIP가 될 확률을 로지스틱 회귀 분석을 활용하여 모델을 생성해서 계수로 비교해봤다.

먼저 성별을 모델을 통해서 계수를 계산했을 때 -0.419가 나왔고, 음의 값이 나왔기에 여성일수록 VIP일 가능성이 높은 것을 확인할 수 있었다.

하지만 p-value = 0.184 → 통계적으로 유의하지 않았다.

그 다음 나이를 모델을 통해서 계수를 계산했을 때 0.023으로 나이가 많을 수록 VIP가 될 확률이 조금 높은 것을 볼 수 있다. p-value = 0.073 → 유의수준 0.1에서 약간 유의한 결과라고 볼 수 있다.

모델

$\text{logit}(\text{VIP}) = 0.144 - 0.419 \cdot \text{성별(숫자)} + 0.023 \cdot \text{나이}$

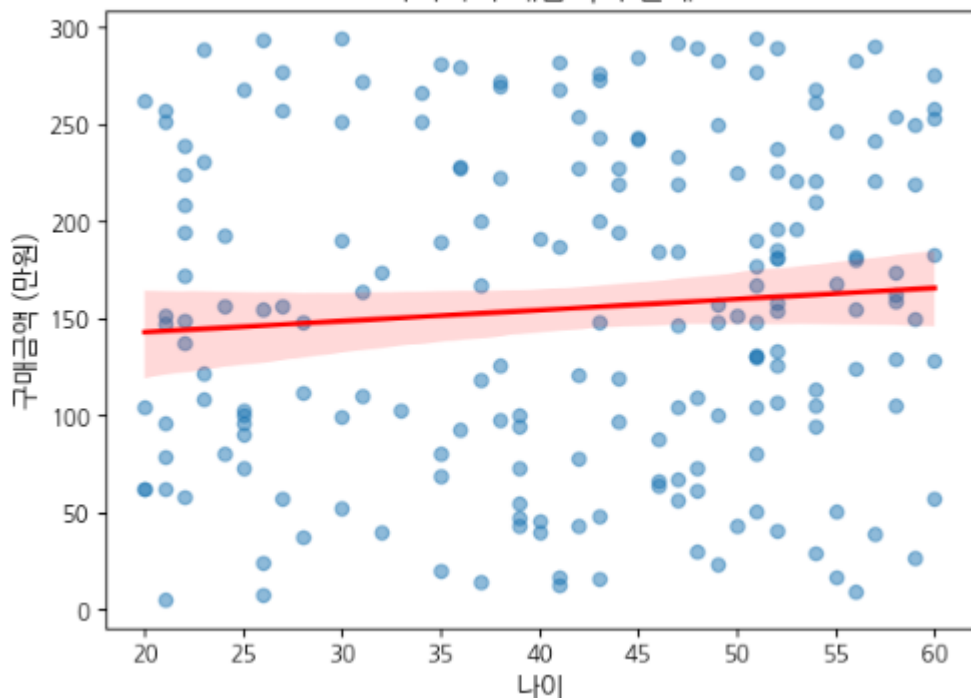
성별(숫자): 남성=1, 여성=0

나이: 연속 변수 (20~60세 범위)

# 산점도

-목적: 나이 vs 구매금액관계

나이와 구매금액의 관계



나이와 구매금액 사이에 상관관계가 있는지 분석해봤다.

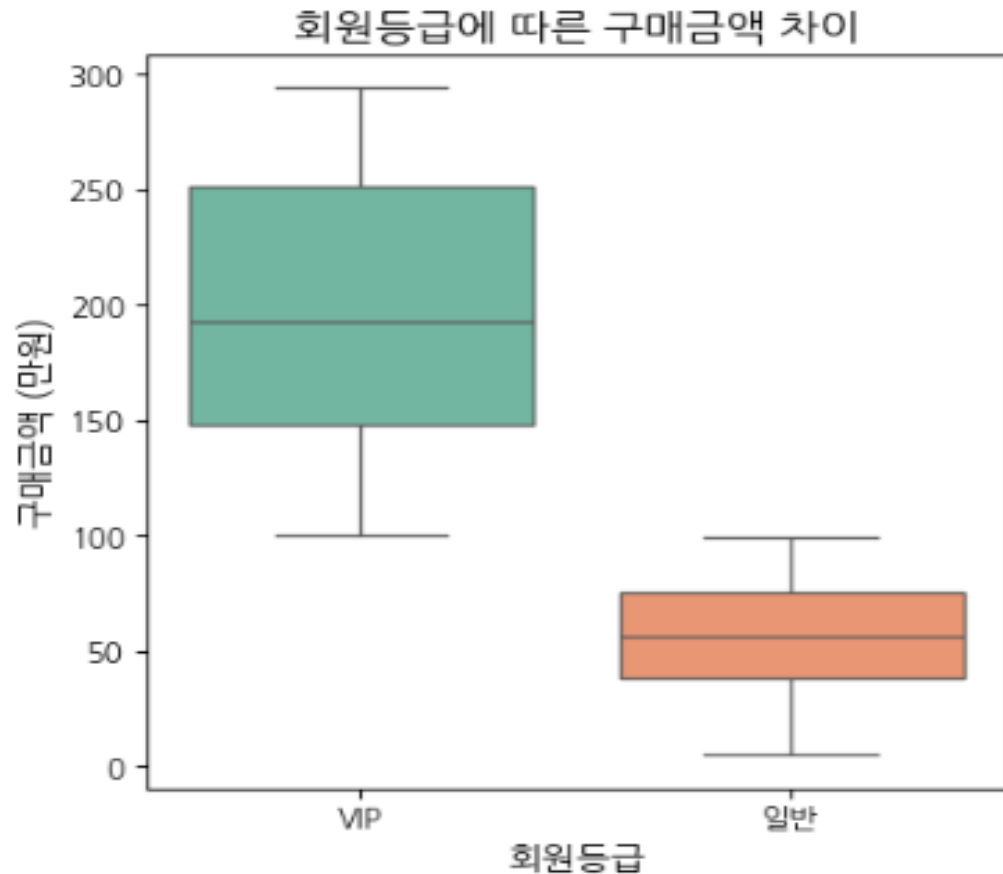
먼저 시각적인 분석을 했을 때 점들은 전체적으로 넓게 퍼져 있기에 데이터 간 분산이 큰 것을 알 수 있다.

빨간선은 회귀선이고, 살짝 우상향 하기에 나이가 많아질수록 구매금액도 약간 증가하는 경향이 있다는 걸 알 수 있다. 그러나 기울기가 크지 않고, 분산도 크기에 강한 관계가 있는 것은 아니다.

상관계수인 0.0828을 통해 수치적 분석을 했을 때 거의 0에 가까운 수이고, 양의 값이므로 약한 양의 상관관계이다.

# 박스플롯(Box plot)

-목적: VIP vs 일반 고객의 구매금액  
비교



회원등급에 따른 구매금액 차이를 분석해봤다.

VIP회원의 경우

중앙값(50%): 약 190만 원

범위: 대략 100만 원 ~ 295만 원

IQR (중간 50%): 약 150만~250만 원 수준 → 변동성 있음

전반적으로 높은 구매금액에 분포되어 있는 것을 볼 수 있다.

일반 회원의 경우

중앙값(50%): 약 60만 원

범위: 대략 10만 원 ~ 100만 원

IQR: 약 35만~80만 원 → 낮은 소비 수준

전체적으로 낮은 구매력을 가진 것을 볼 수 있다.

# 결과 해석 및 인사이트 제시

## • 성별별 평균 구매금액

- 여성이 남성보다 약 18.6만 원 더 많이 지출하고 있음

-> 여성 고객층이 전체 매출에 더 크게 기여하기에 여성 대상 마케팅 효율이 더 높을 가능성이 존재한다.  
예를 들어 여성의 쇼핑 선호도, 제품 카테고리 분석을 통해 여성 전용 추천 시스템을 강화할 수 있다.

## • 나이와 구매금액의 관계

- 전체적으로 나이에 따라 구매금액이 소폭 증가하지만, 상관계수는 약 +0.08로 매우 약한 양의 상관관계를 가지고 있기에 나이가 많다고 무조건 지출이 큰 것은 아니다.

->나이에 따른 구매력 차이는 크지 않다. 하지만 일부 연령대(예: 50대)에서는 고액 지출자가 자주 보이기에 라이프스타일 기반 타겟팅이 더 적절해 보인다.

## • 회원등급(VIP vs 일반)별 구매금액 차이

-VIP 고객은 150~250만 원 수준의 지출, 일반 고객은 40~80만 원 사이의 지출을 가지기에 VIP의 평균 및 중앙값이 일반 고객보다 훨씬 높다.

->VIP 고객이 구매금액을 이끌어가는 핵심 집단이다. 그에 따라 VIP 고객 유지를 위해 리워드, 전용 혜택, 빠른 응대 제공하고 일반 고객 대상으로 승급 유도 캠페인 (ex: 누적 구매금액 기반 VIP 전환 제안)을 활성화 시킨다.