현대자동차
데이터분석 해카톤
- Task 1

(재구매 유형 분리 로직 설명서 & EDA 결과서)

960723 김동욱

목차

과제 1

과제1 개요

재구매유형 분리/추정

EDA

1-1. 문제 정의

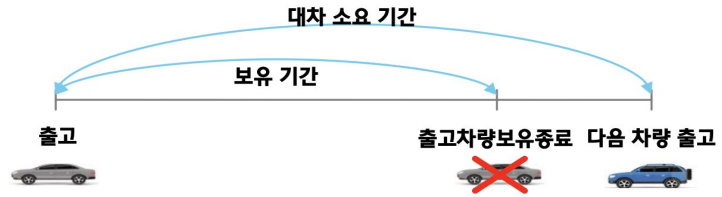
1. 주어진 가상 고객 구매 **Data**를 활용하여 “대차”와 “추가구매”를 분리/추정하는 로직을 개발

2. “대차”, “추가구매” 구분 결과에 대해 각각 **EDA**를 수행하여 고객과 차량 관점에서 특징/차이점을 도출

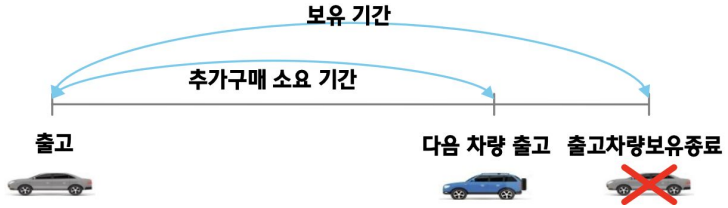
고객별 차량 구매 유형은 크게 **3**가지로 분류됨.

- 최초구매: 처음으로 차량을 구매하는 경우
  → 주어진 데이터에서 고객별 “출고일자”를 기준으로 가장 첫번째 구매 내역

- 대차: 기존에 보유하던 차량을 처분하고 새로운 차량을 구매하는 경우
  → 대차 소요 기간과 보유 기간이 유사하다는 특징이 있음.



- 추가구매: 기존에 보유하던 차량을 처분하지 않고 새로운 차량을 구매하는 경우
  → 추가구매 소요 기간과 보유 기간 사이에 일정 수준의 차이가 있음.



이 때, “재구매 고객”은 “대차”와 “추가구매” 고객을 의미.

과제 1

2-1. 대차/추가구매 정의

"대차 소요 기간"과 "추가구매 소요 기간"은 모두 기존 차량들의 보유 종료일과 신차의 출고일 간의 차이를 기준으로 파악할 수 있음.

이 차이를 "재구매 판단값"이라는 변수로 정의.

- t: **재구매 판단값** (단위: 일)
- t_{new}: **신차의 출고일**
- t_{prev}: **기존 차량의 보유 종료일**

$$\rightarrow t = t_{prev} - t_{new}$$

이렇게 정의된 "재구매 판단값"이 일정한 범위에 속하면 신차의 재구매 유형을 대차, 그렇지 않으면 추가구매로 분류.

$$f(t) = \begin{cases} \text{대차}, & \text{if } a < t < b \\ \text{추가구매}, & \text{otherwise} \end{cases} \qquad where \quad a, b \quad are\ time\ in\ days$$

이 때, 구간 **[a, b]**를 "대차 인정 기간"으로 정의하고, **a**를 "대차 하한값", **b**를 "대차 상한값"으로 정의.

과제 1

과제1 개요

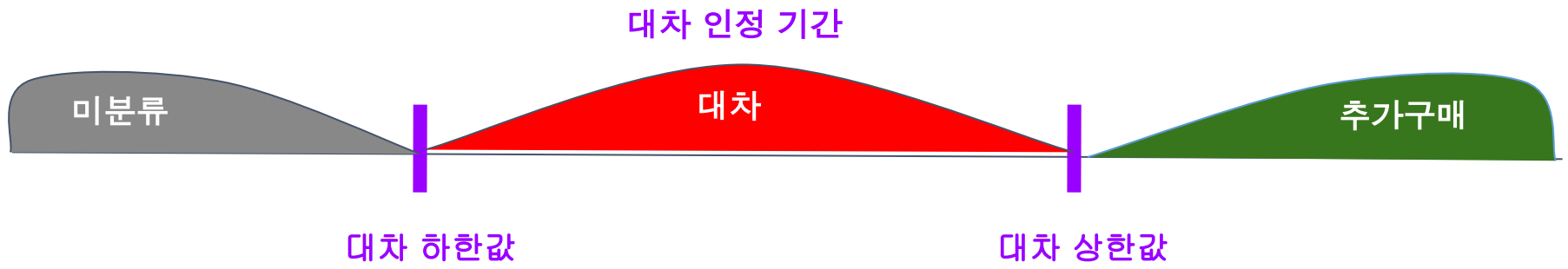
재구매유형 분리/추정

EDA

2-2. 대차/추가구매 분리/추정 알고리즘 설명

1. “최초구매”로 라벨링된 데이터는 제외.
2. “대차 하한값”과 “대차 상한값”을 구체적인 상수값으로 설정한 다음, “대차 인정 기간”을 정의.
3. 보유중인 차량들로부터 신차의 “재구매 판단값”을 전부 계산한 다음, 하나 이상의 산출값이 “대차 인정 기간”에 포함되면 해당 신차에 대한 재구매 유형을 “대차”로 라벨링.
4. (3)단계에 해당하지 않으며, 산출된 “재구매 판단값” 중 하나 이상이 “대차 상한값” 밖에 존재하는 경우 해당 신차에 대한 재구매 유형을 “추가구매”로 라벨링.
5. 그 외는 전부 “미분류”로 라벨링.

<예시>



과제 1

과제1 개요

재구매유형 분리/추정

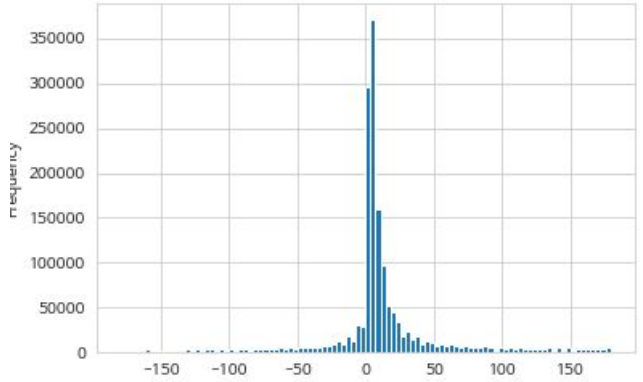
EDA

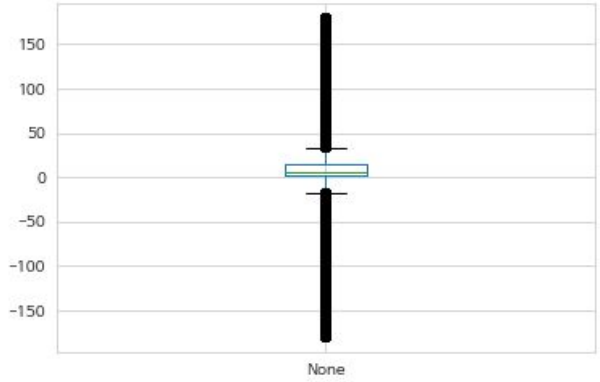
2-3. 대차 하한값 / 상한값 설정

"대차 하한값"과 "대차 상한값"을 구체적인 상수값으로 설정하기 위해 다음과 같은 통계량 (T) 을 정의한 다음, 부트스트랩 샘플링 (Bootstrap Sampling)으로 통계량의 분포를 파악.

- t_{new}: 신차의 출고일
- t_{prev}: 기존 차량의 보유 종료일

$$T = absmin(\mathbf{t}_{prev} - \mathbf{t}_{new}) \qquad (bold\ font\ indicates\ vector)$$





- 히스토그램과 상자그림을 기준으로 할 때, 약 [−17.5, 34.5] 사이에 대부분의 값들이 분포.
- 히스토그램의 분포를 고려해 상기 구간에 일주일을 가감하여 대차 인정 기간을 설정
  - 대차 하한값: −25 (단위: 일)
  - 대차 상한값: 41 (단위: 일)

과제 1

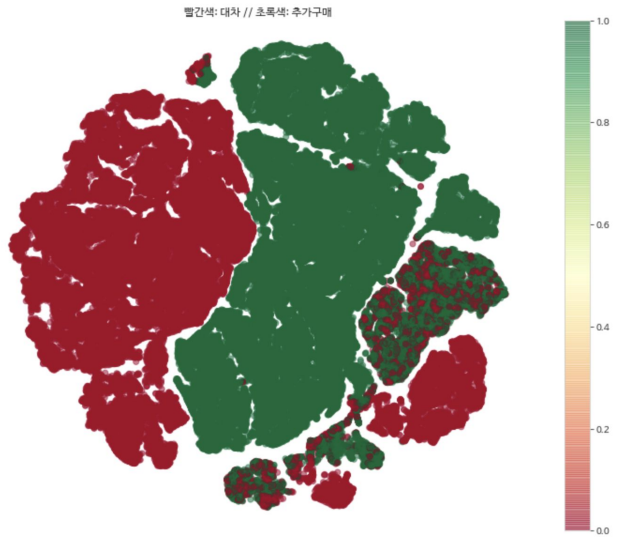
과제1 개요

재구매유형 분리/추정

EDA

2-4. 재구매유형 분류 결과

(T-SNE 차원축소 알고리즘을 이용한 대차 / 추가구매 시각화)




```
# 재구매 유형별 엔트리 수
df.TYPE_PURCHASES.value_counts()
```

```
최초구매    602223
추가구매    297652
대차        251529
미분류      181093
```

- 추가구매로 분류된 엔트리: 약 **30**만개
- 대차로 분류된 엔트리: 약 **25**만개

(특정 고객의 차량 구매내역을 바탕으로 한 재구매유형 구분 예시)

	(고객 ID) CUS_ID	(출고일) WHOT_DT	(보유종료일) CAR_HLDG_FNH_DT	(차량명) CAR_NM	(재구매 유형) TYPE_PURCHASES
183	AONEEOT120000144	2006-09-12	2017-10-24	아반떼	최초구매
184	AONEEOT120000144	2007-02-16	2008-12-18	클릭	추가구매
185	AONEEOT120000144	2013-03-22	2017-02-03	엑센트	추가구매
186	AONEEOT120000144	2017-10-10	2017-12-30	싼타페 DM	대차

→ 최초 구매차량

→ “아반떼”를 보유한 상태로“클릭” 추가구매

→ “아반떼“를 보유한 상태로 “엑센트” 추가구매

→ “아반떼”로부터의 대차

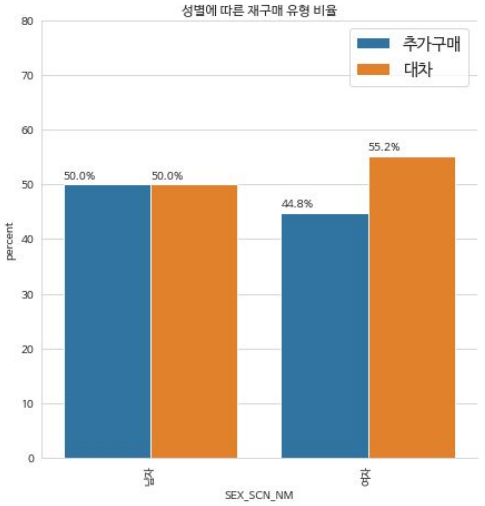
과제 1

과제1 개요

재구매유형 분리/추정

EDA

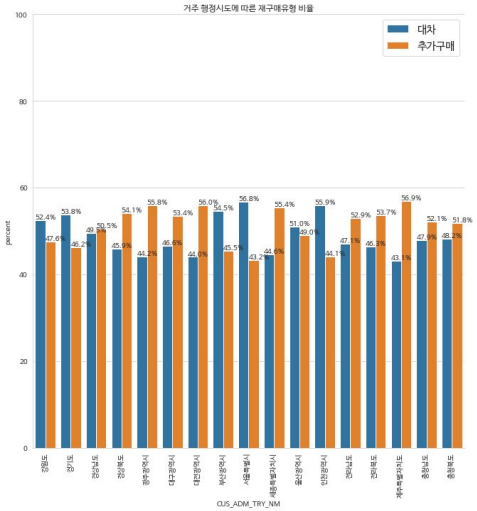
3-1. 고객 관점에서의 분석



<성별별>

ㅁ 남성: 차이없음
ㅁ 여성: 대차비율이 약 **10.4%p** 높음.

→ 여성 고객들의 경우 남성 고객들보다 실수요에 의한 차량 구입 성향이 큰 것으로 판단됨.

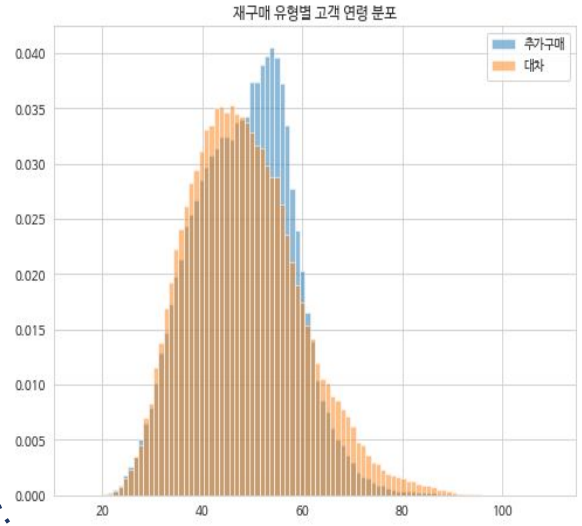


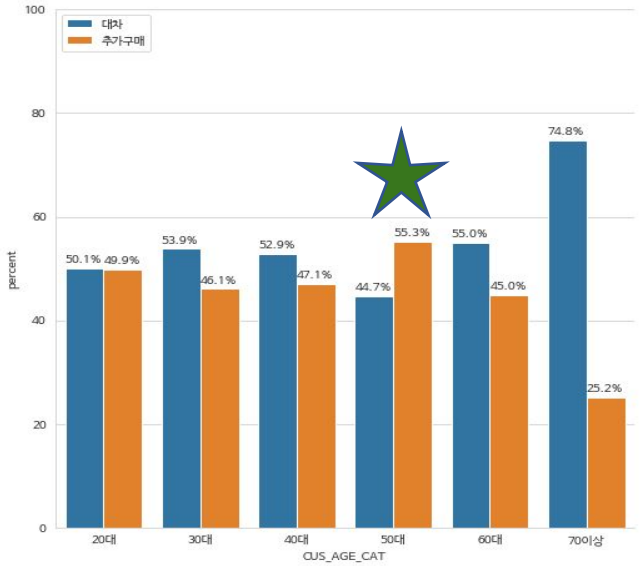
<거주지역별>

ㅁ 대차의 비율이 더 높은 지역:
서울, 부산 및 수도권

ㅁ 추가구매의 비율이 더 높은 지역:
지방 행정시도

→ 교통인프라 수준과 추가구매 수요간에 음의 상관관계가 있다고 판단됨.





<연령별>

ㅁ **50대를 제외**한 전 연령층에서 추가구매의 비율보다 대차의 비율이 높은 것으로 나타남.

→ **50**대 고객들의 경제력과 사회경제적 위치로 인해 추가구매의 수요가 높은 것이라고 판단됨.

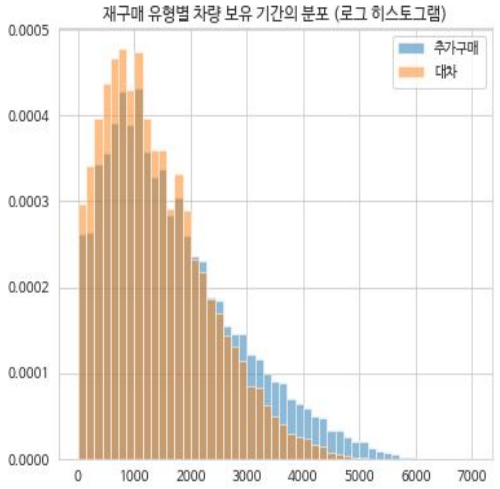
과제 1

과제1 개요

재구매유형 분리/추정

EDA

3-2. 차량 관점에서의 분석

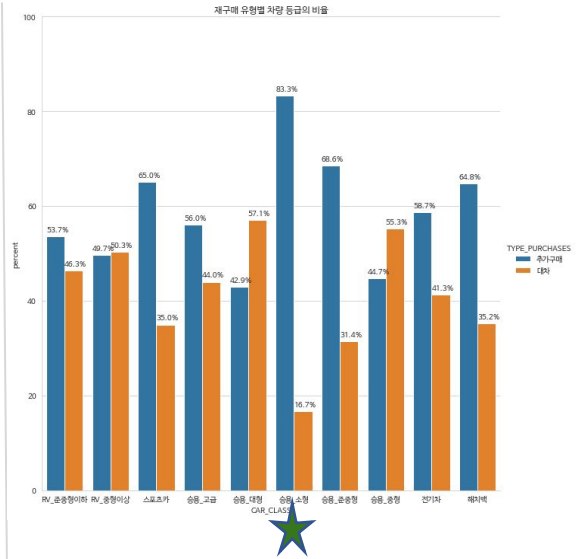


<차량 보유 기간별>

ㅁ 추가구매 집단 > 대차 집단

→ 추가구매 차량들은 주행거리가 연식에 비해 짧을 가능성이 높음.

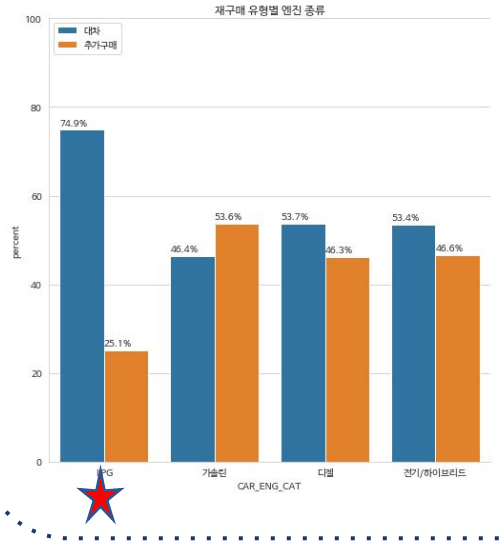
→ 대차 차량은 추후 또 다른 차량으로 대차될 가능성이 높음.



<차량 등급별>

ㅁ <u>소형 승용차 그룹</u>에서 추가구매의 비율이 압도적으로 높음

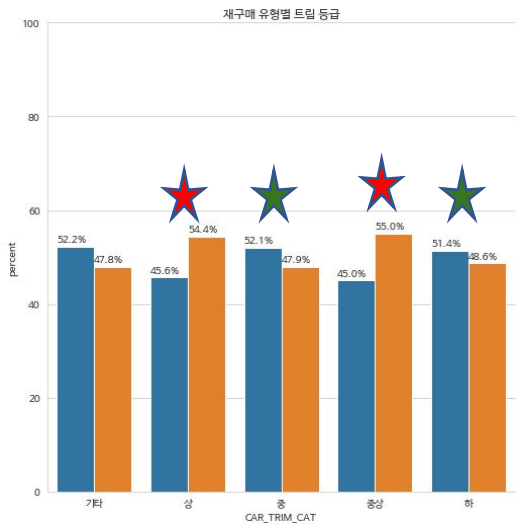
→ 추가구매는 서브 차량으로써의 니즈를 기반으로한 소비일 가능성이 높으므로, 비교적 부담없이 탈 수 있는 저렴한 차량들이 선호되는 것으로 해석됨.



<엔진 종류별>

ㅁ <u>LPG 그룹</u>에서 대차의 비율이 눈에 띄게 높음

→ LPG 차량 (택시, 렌트카, SUV, 트럭 등) 의 특성을 고려했을 때 효율성을 중시하는 고객의 성향이 반영된 결과라고 해석됨.



<트림 등급별>

ㅁ 대차: 중상급 이상 ↑
ㅁ 추가구매: 중급 이하 ↑

→ 대차 차량은 구매자가 1순위로 타고 다니는 차량일 가능성이 높기에 옵션에 많은 투자.

→ 반면, 추가구매 차량은 특수 목적 (ex. 근거리 반복 이동) 으로만 탑승할 가능성이 높기 때문에 옵션에 크게 투자 X

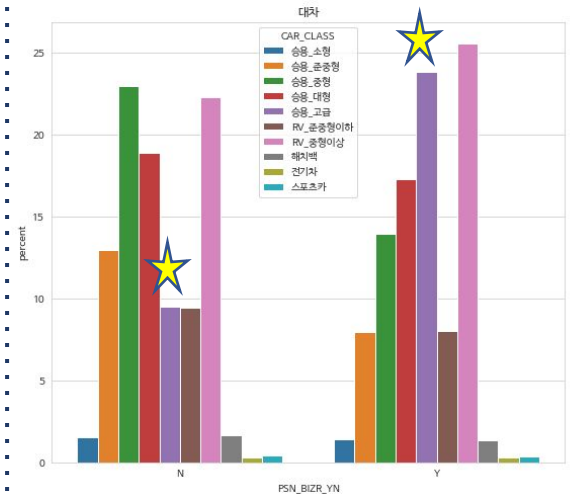
과제 1

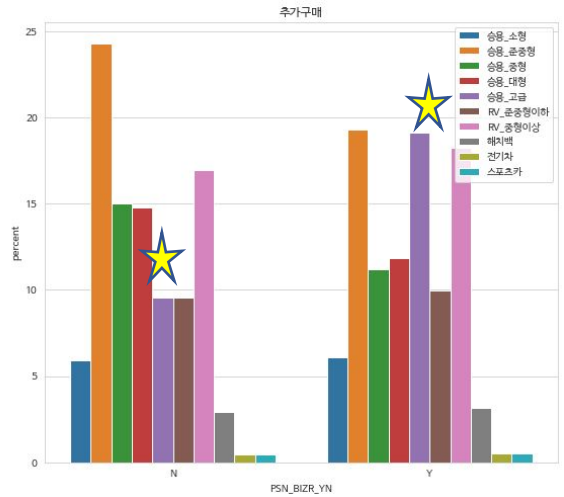
과제1 개요

재구매유형 분리/추정

EDA

3-3. 다변량 분석

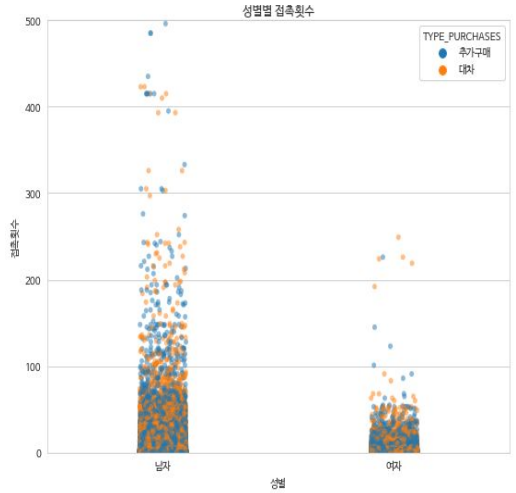




<개인사업자 여부 & 차량 등급별>

□ 개인사업자 고객 → 고급 승용차의 비율이 크게 높음. (대차/추가구매 모두)

개인 사업자 고객들에게 차량은 단순한 이동 수단을 넘어, 경제적 능력과 사회적 지위를 과시하는 용도로 사용될 수 있음.
→ 비교적 가격대가 높은 모델을 선호.

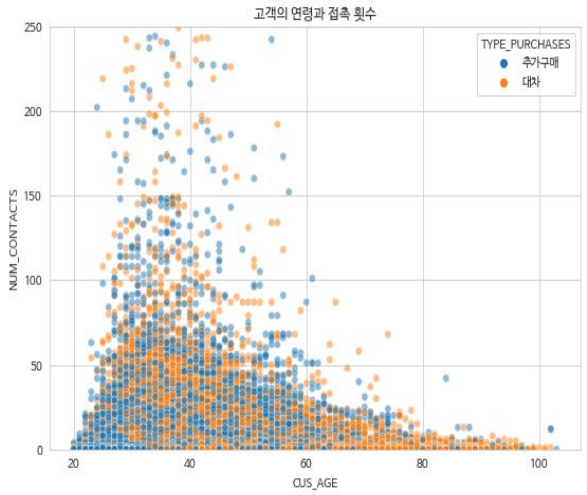


<성별 & 접촉 횟수별>

□ 남성 > 여성

→ 차량에 대한 관심도 차이라고 판단됨.

→ 최근 구매력 있는 여성 1인 가구가 증가하고 있는 점을 감안할 때, 여성 고객들에 대한 적극적인 타겟 마케팅이 필요.



<나이 & 접촉 횟수>

□ 연령대가 높아질 수록 접촉횟수가 감소함.

→ 구매력이 높은 중장년 고객들의 자발적 접촉을 유도할 수 있는 방안이 필요.