

자녀 월령에 따른 유아용품 고객 세분화 모델



INDEX

1. 프로젝트 팀 구성 및 역할
 2. 프로젝트 개요
 3. 프로젝트 수행 절차 및 방법
 - 데이터 전처리
 - 데이터 분석 (EDA)
 - 모델 생성
 4. 프로젝트 수행 결과
 5. 자체 평가 의견
-

1. 프로젝트 팀 구성 및 역할

프로젝트 팀 구성 및 역할

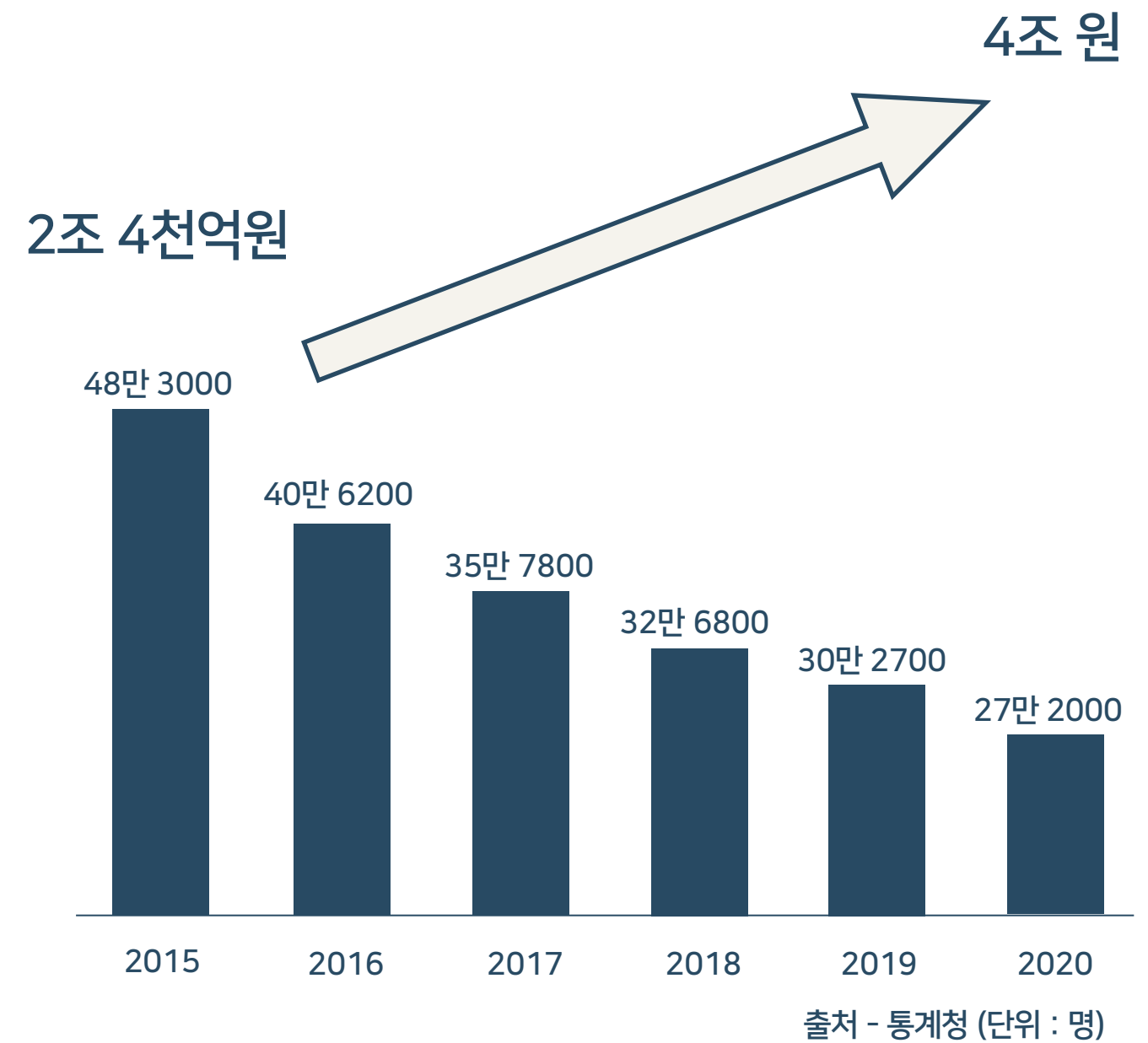
김성민	————	데이터 전처리, 머신러닝 모델 생성
김여민	————	데이터 전처리, 발표
박종서	————	데이터 전처리, 도메인 자료 조사
이송이	————	데이터 전처리, PPT 작성
주효성	————	데이터 전처리, 머신러닝 모델 생성

2. 프로젝트 개요

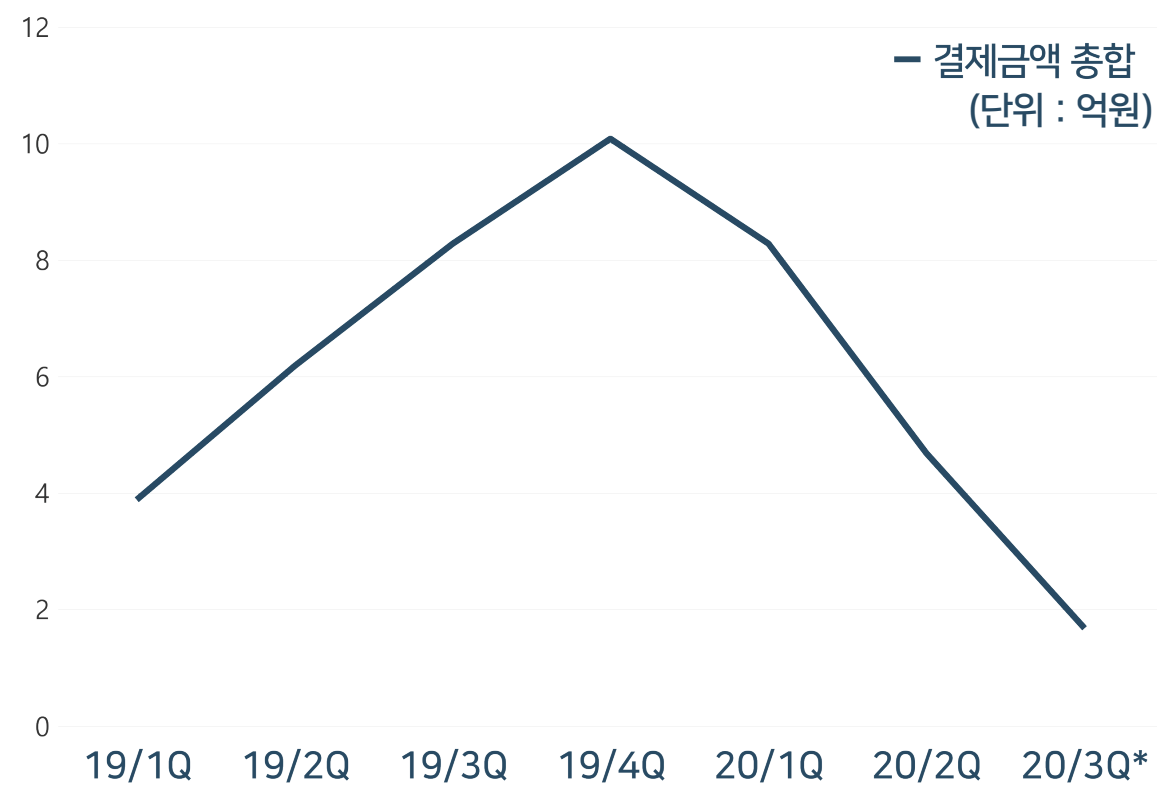
프로젝트 개요

출생아 수는 지속적 감소 추세
반면, 육아용품 시장 규모는 **2배 증가**

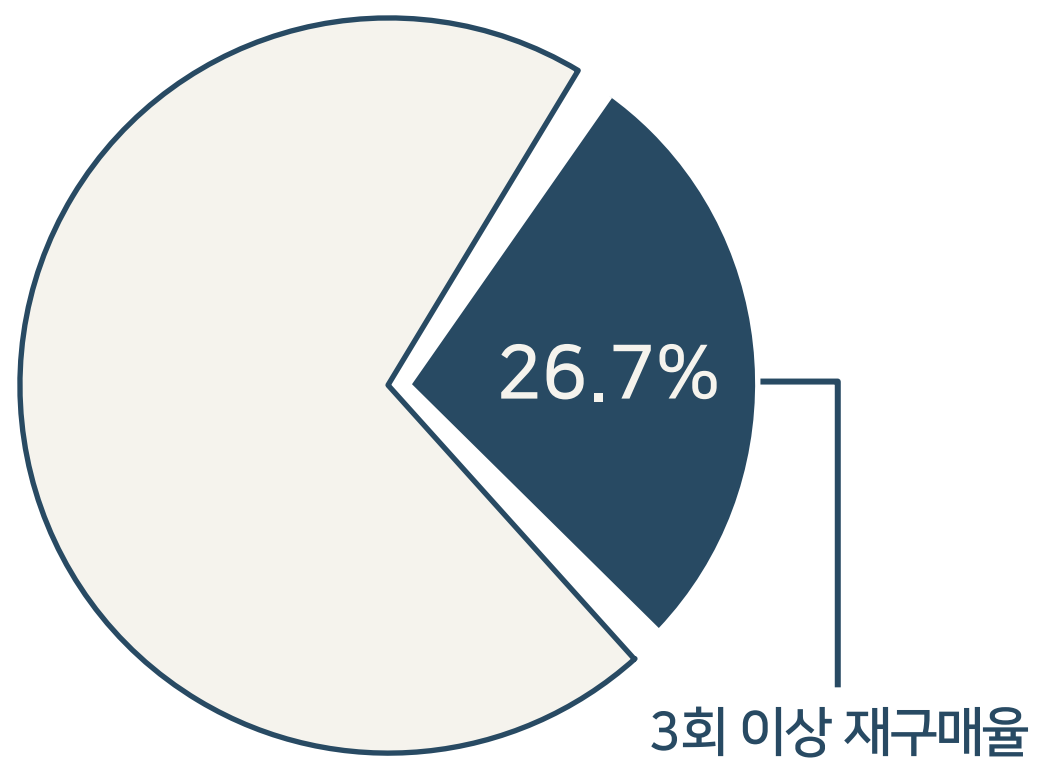
[국내 육아용품 시장 규모 및 출생아 수]



프로젝트 개요

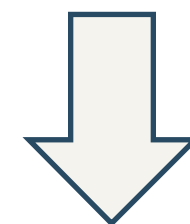


분기별 결제 금액의 **급격한 감소**



판매 품목 특성에 비해
낮은 재구매율

구매 데이터를 바탕으로
고객 세분화

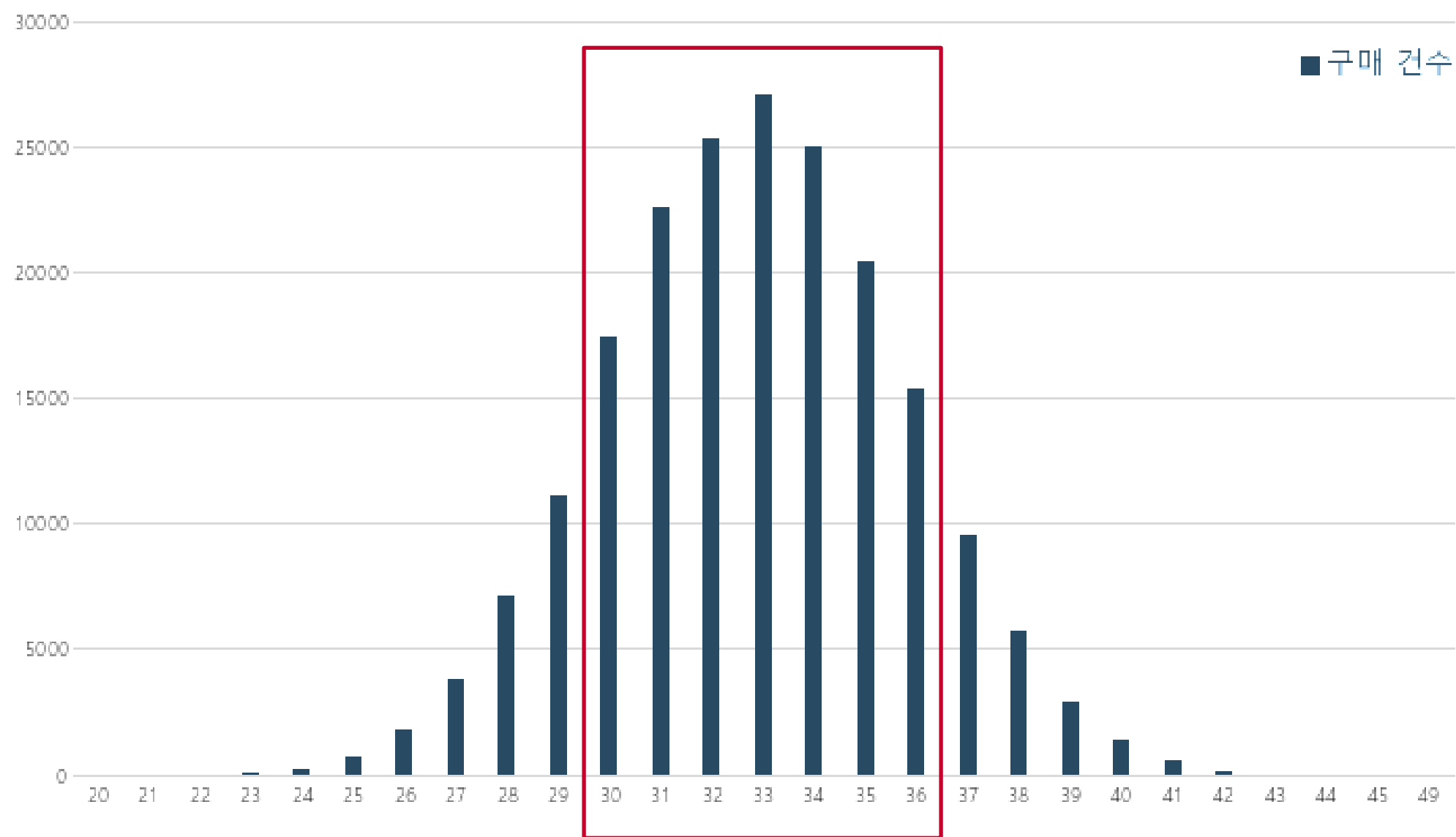


고객 맞춤형 서비스 제공

3. 프로젝트 수행 절차 및 방법

데이터 전처리

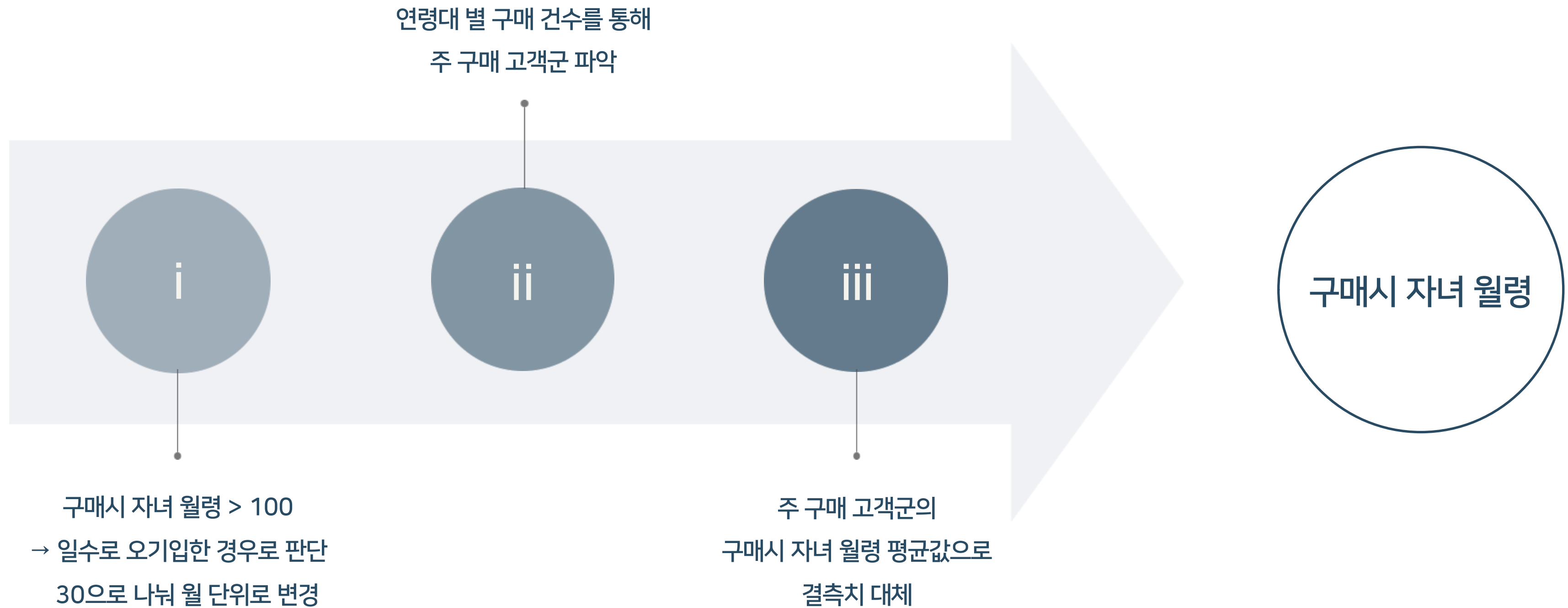
[고객 연령 별 구매 건수]



고객 연령의 대부분이 30대에 분포

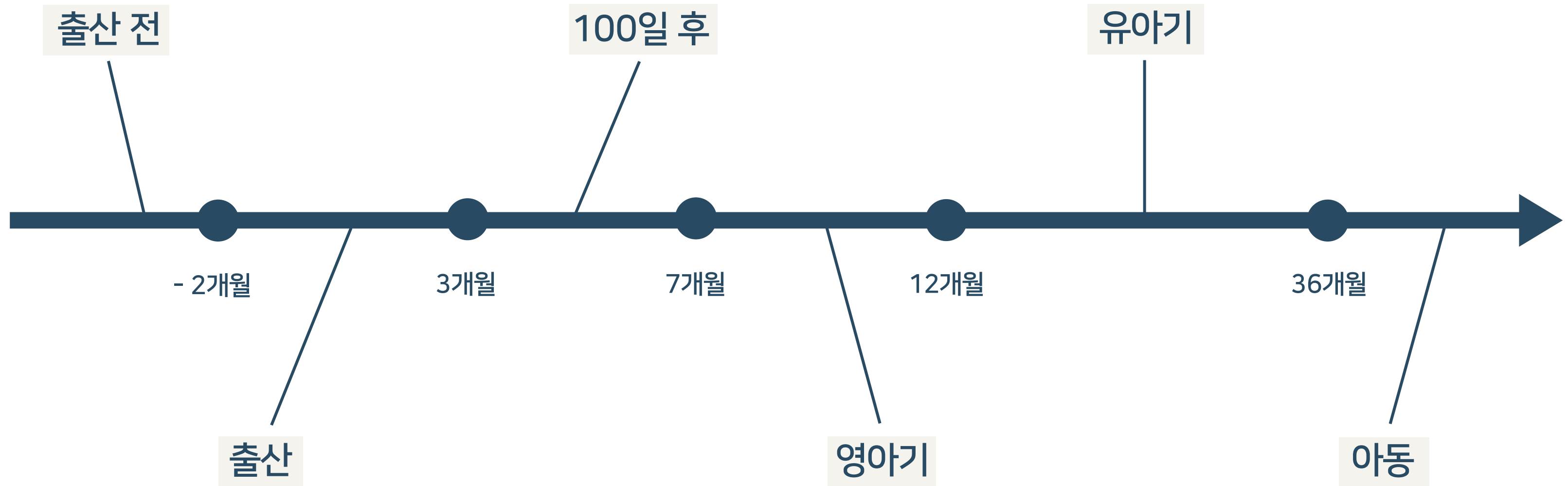
특히 30세 ~ 36세가
높은 비중을 차지

데이터 전처리

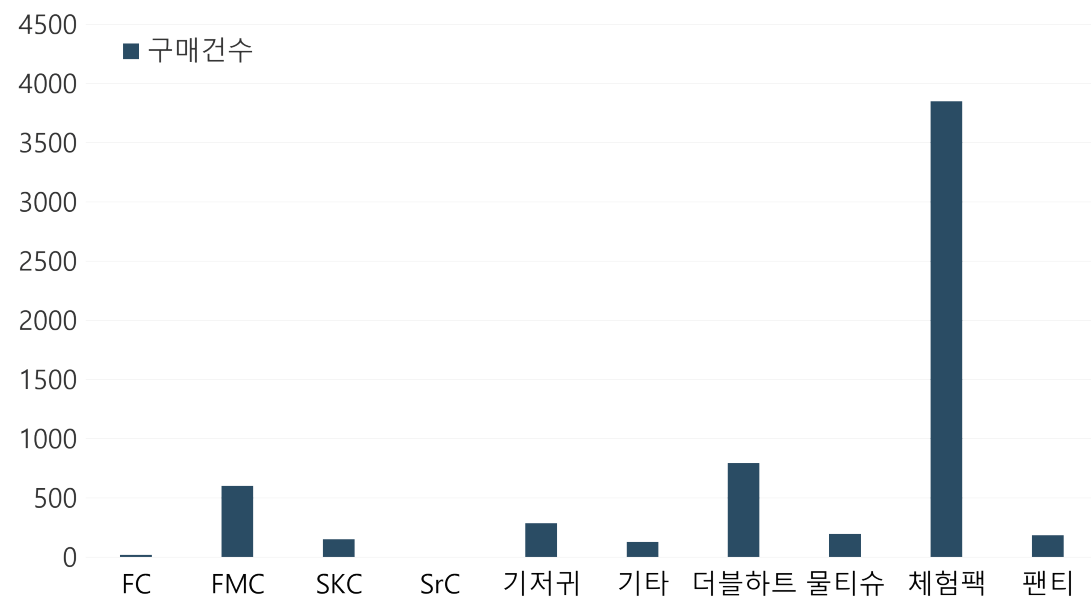


데이터 분석 (EDA)

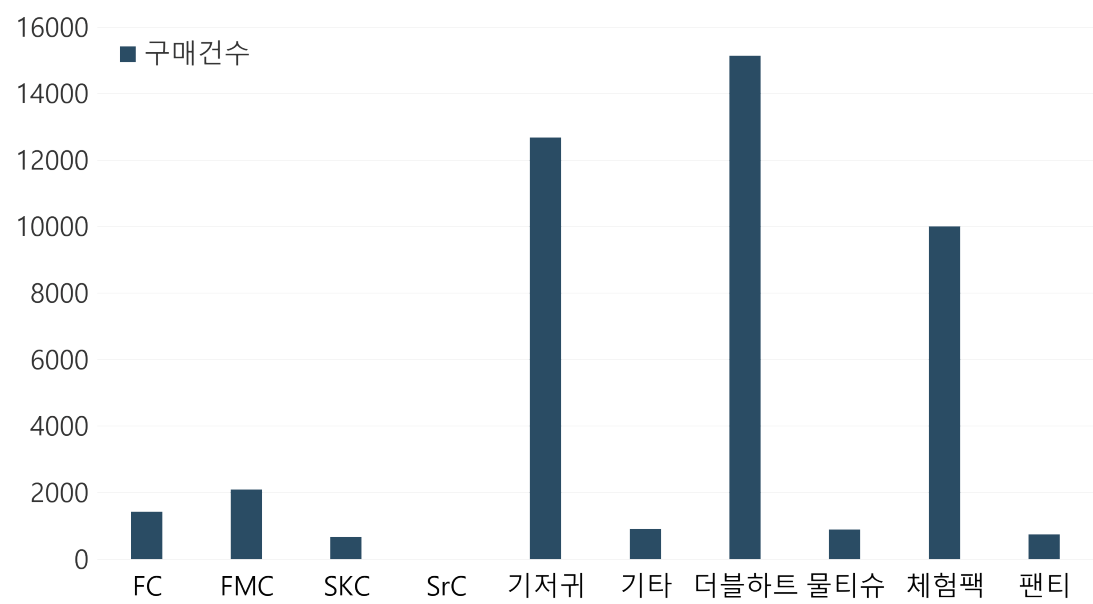
[구매 시 자녀 월령을 생애 주기로 구분]



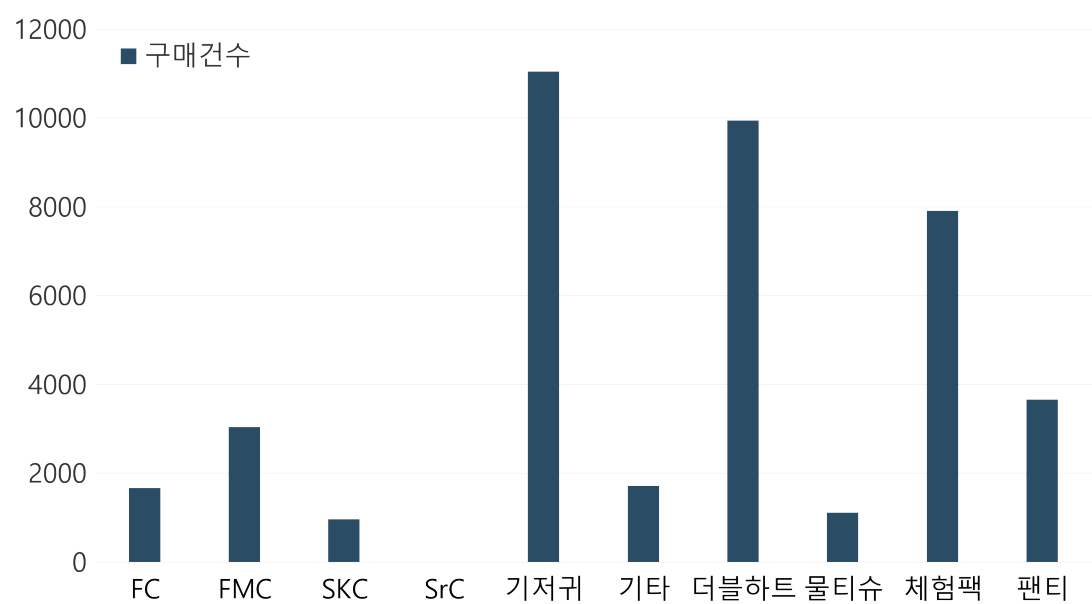
데이터 분석 (EDA)



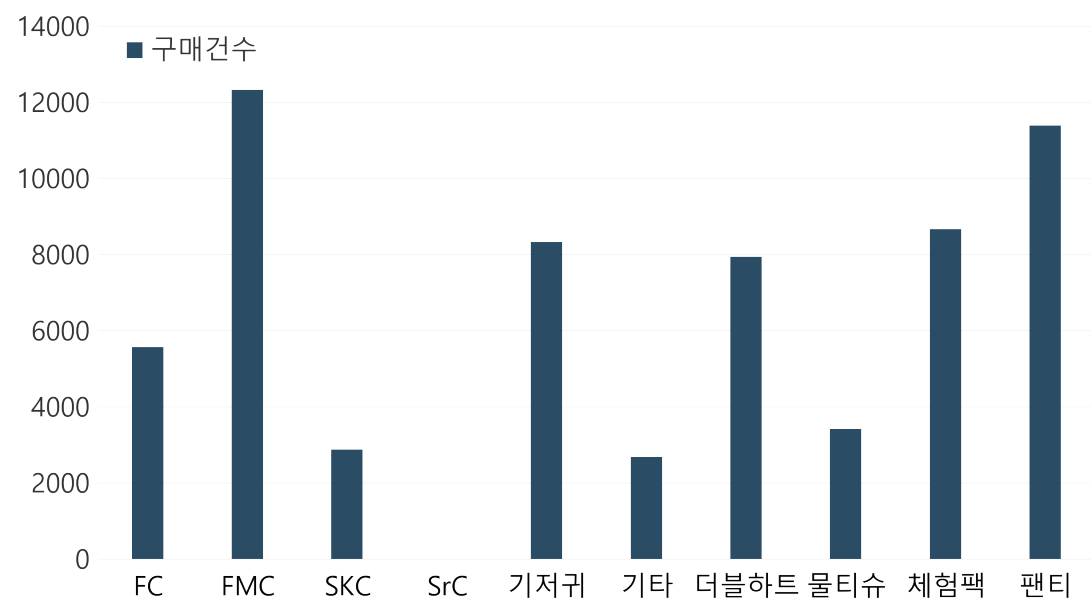
출산전



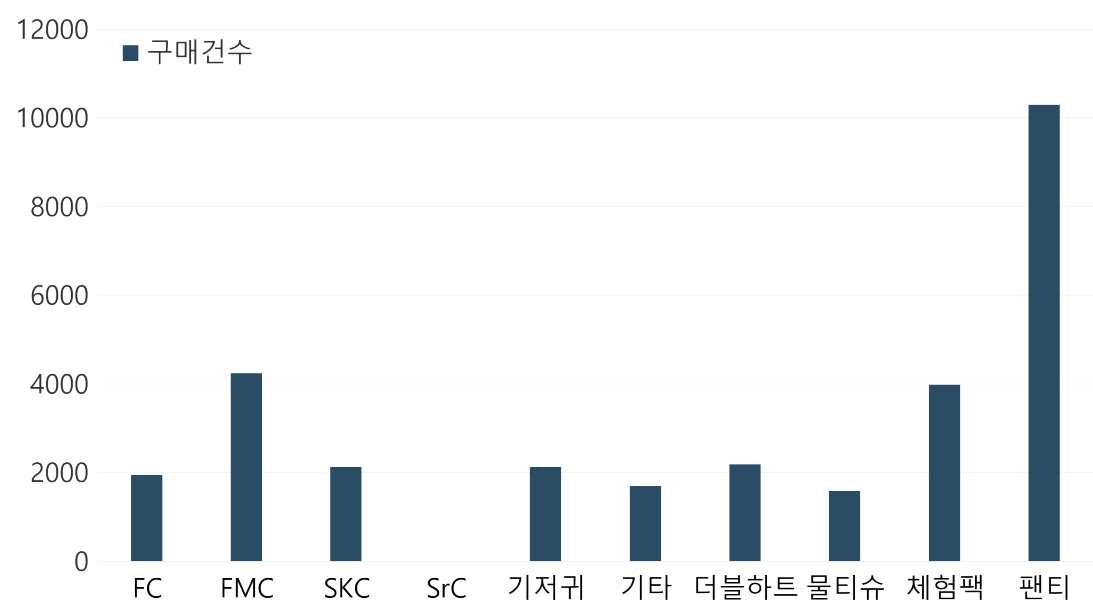
출산



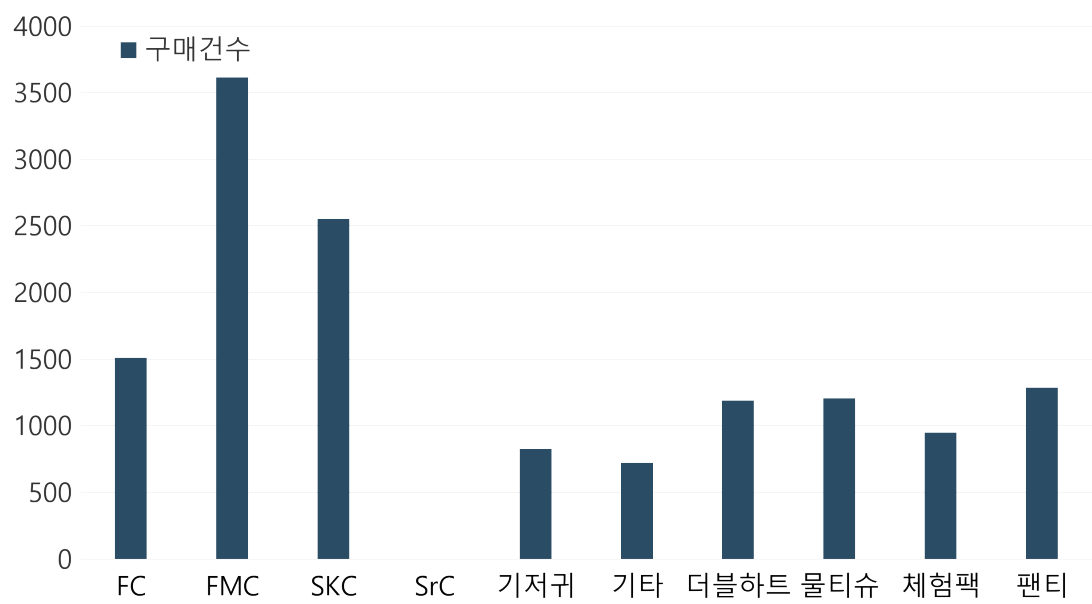
100일 후



영아기

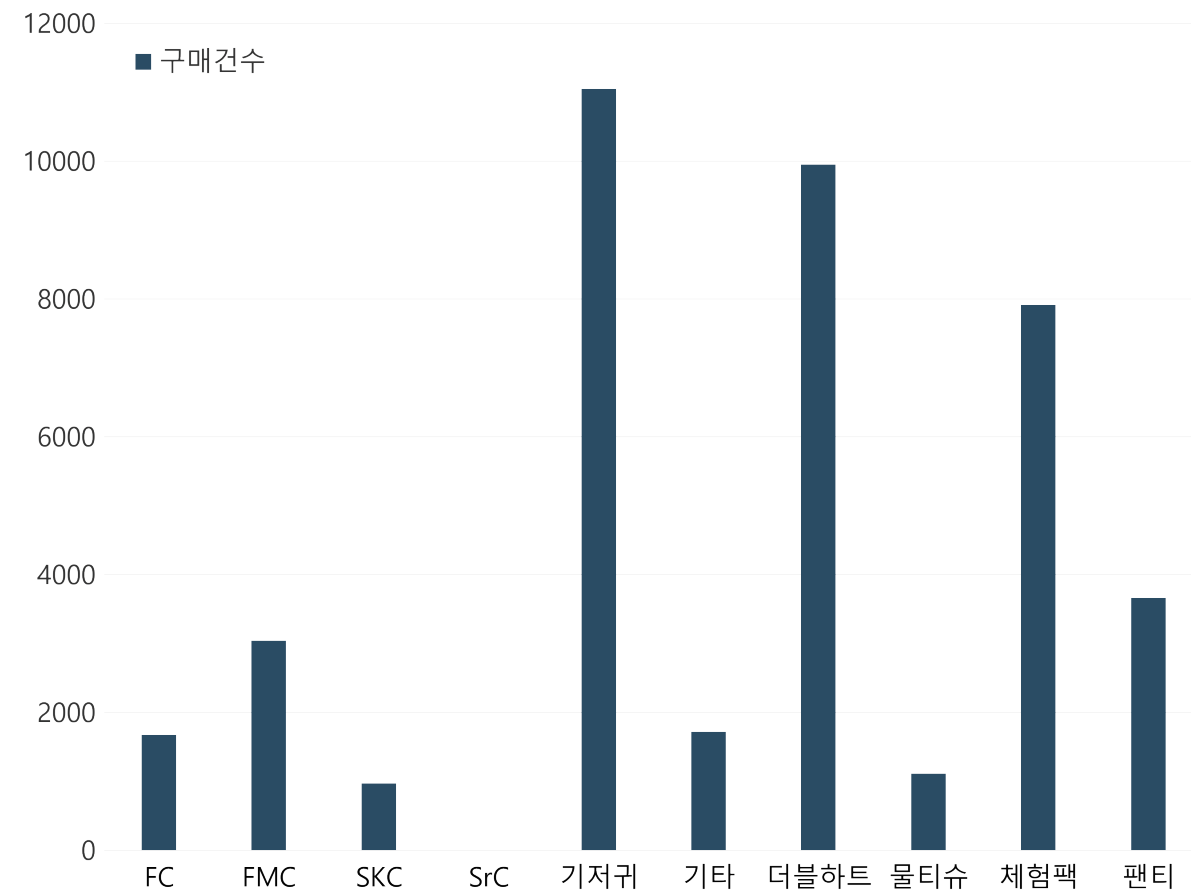


유아기



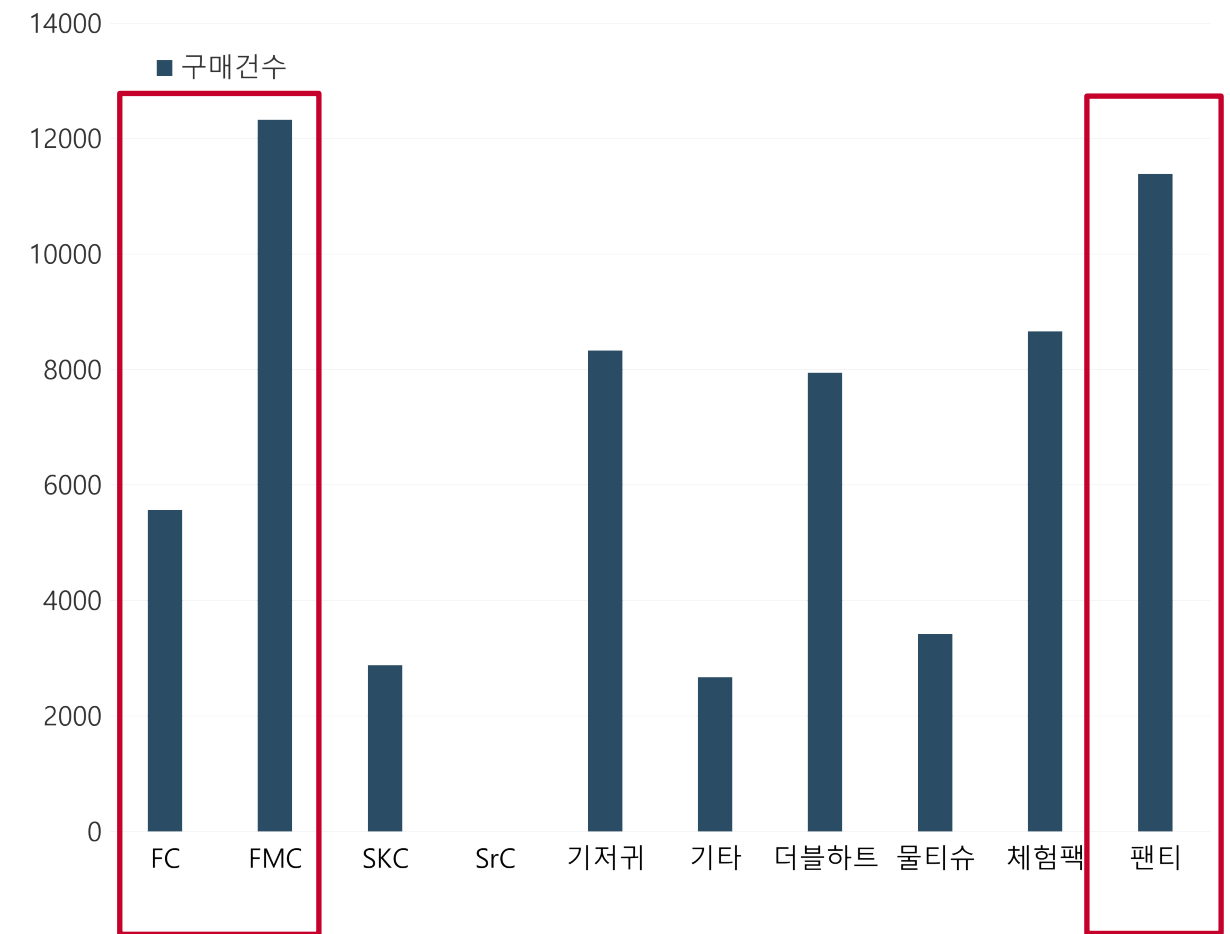
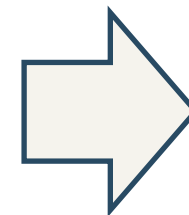
아동

데이터 분석 (EDA)



[100일 후]

기저귀, 더블하트, 체험팩의
구매 건수가 가장 많으며,
대부분의 비율을 차지



[영아기]

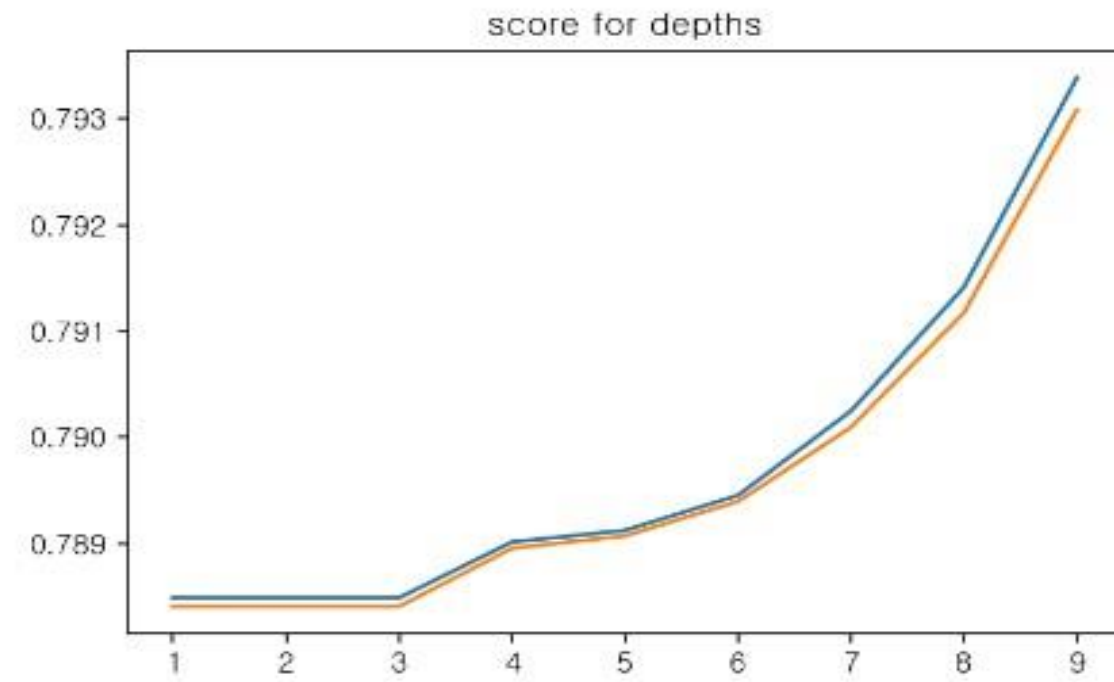
FMC, 팬티, FC 구매 건수 증가
이전 주기에 구매 건수 상위 물품들은
주기 이동 후 구매 건수 하락

모델 생성 1 - 등급별 재구매 예측 모델

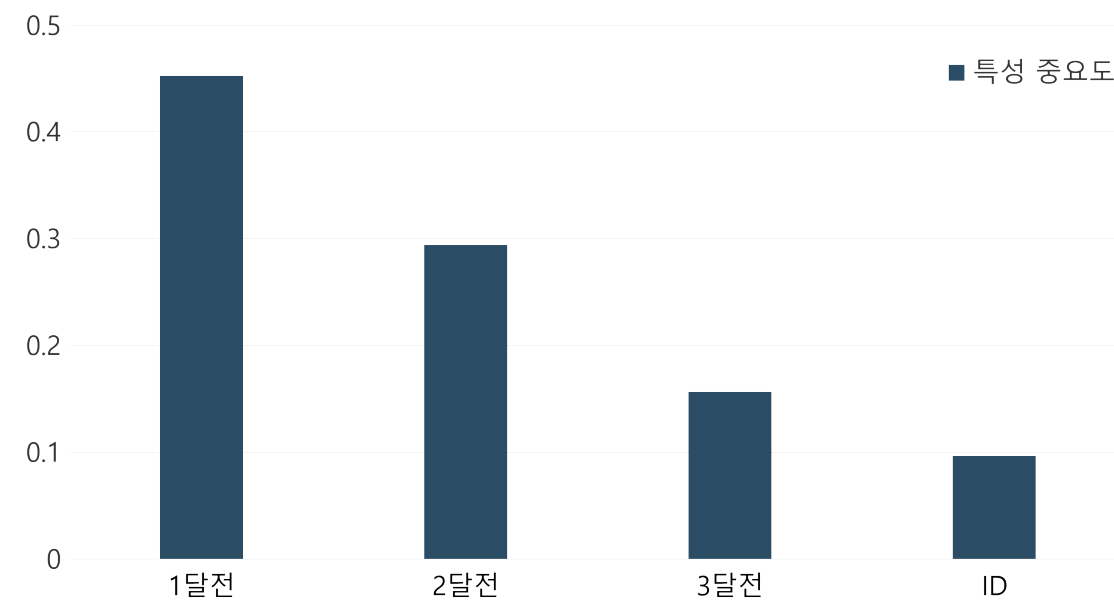
모델 생성

Random Forest 모델 적용

- test size = 0.3
- random state = 50
- max depth = 4
- n_estimators = 150



- 훈련 정확도 : 0.79
- 테스트 정확도 : 0.79

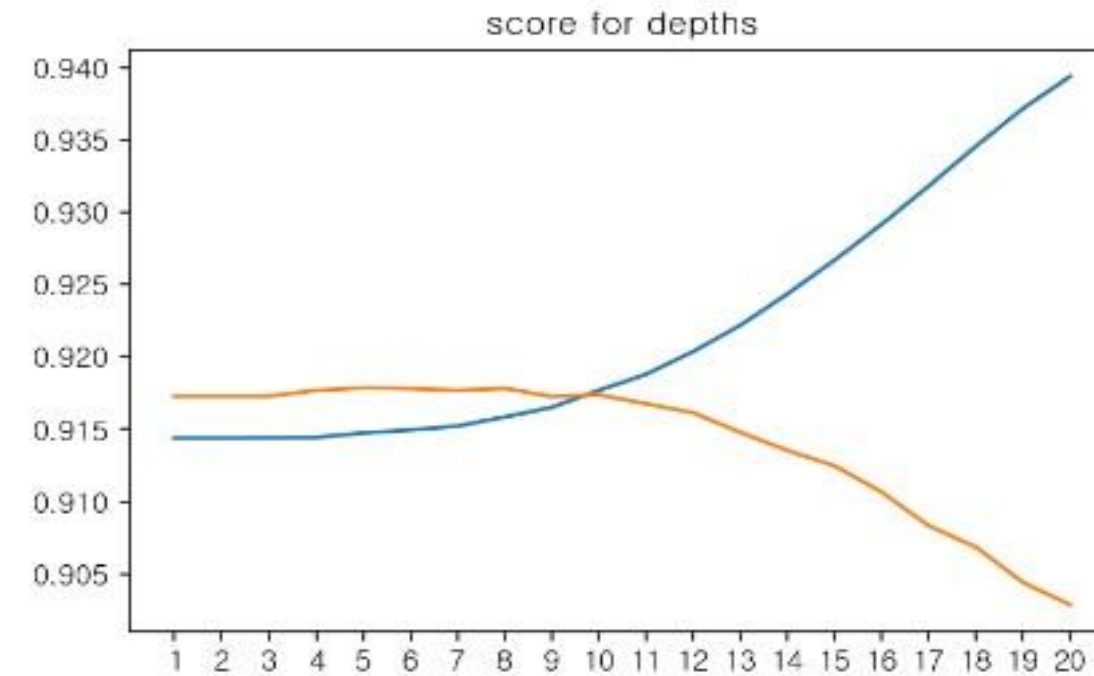


모델 생성 2 - 등급별 구매금액 예측 모델

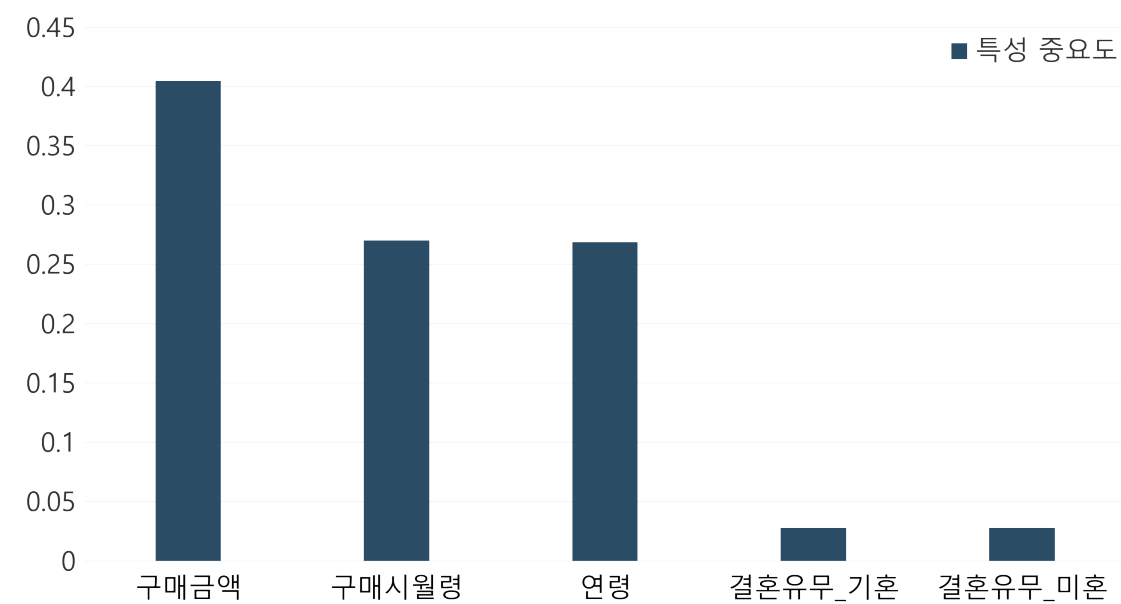
모델 생성

Decision Tree 모델 적용

- test size = 0.3
- random state = 0
- max depth = 8



- 훈련 정확도 : 0.92
- 테스트 정확도 : 0.92



모델 생성 3 - 물품대분류 별 수요 예측 모델

변수 선택

Y 값	물품대분류
X 값	구매일
	ID
	구매금액
	결혼유무
	거주지역
	연령
	결제등록카드
	유입경로
	자녀여부
	구매시월령

모델 생성

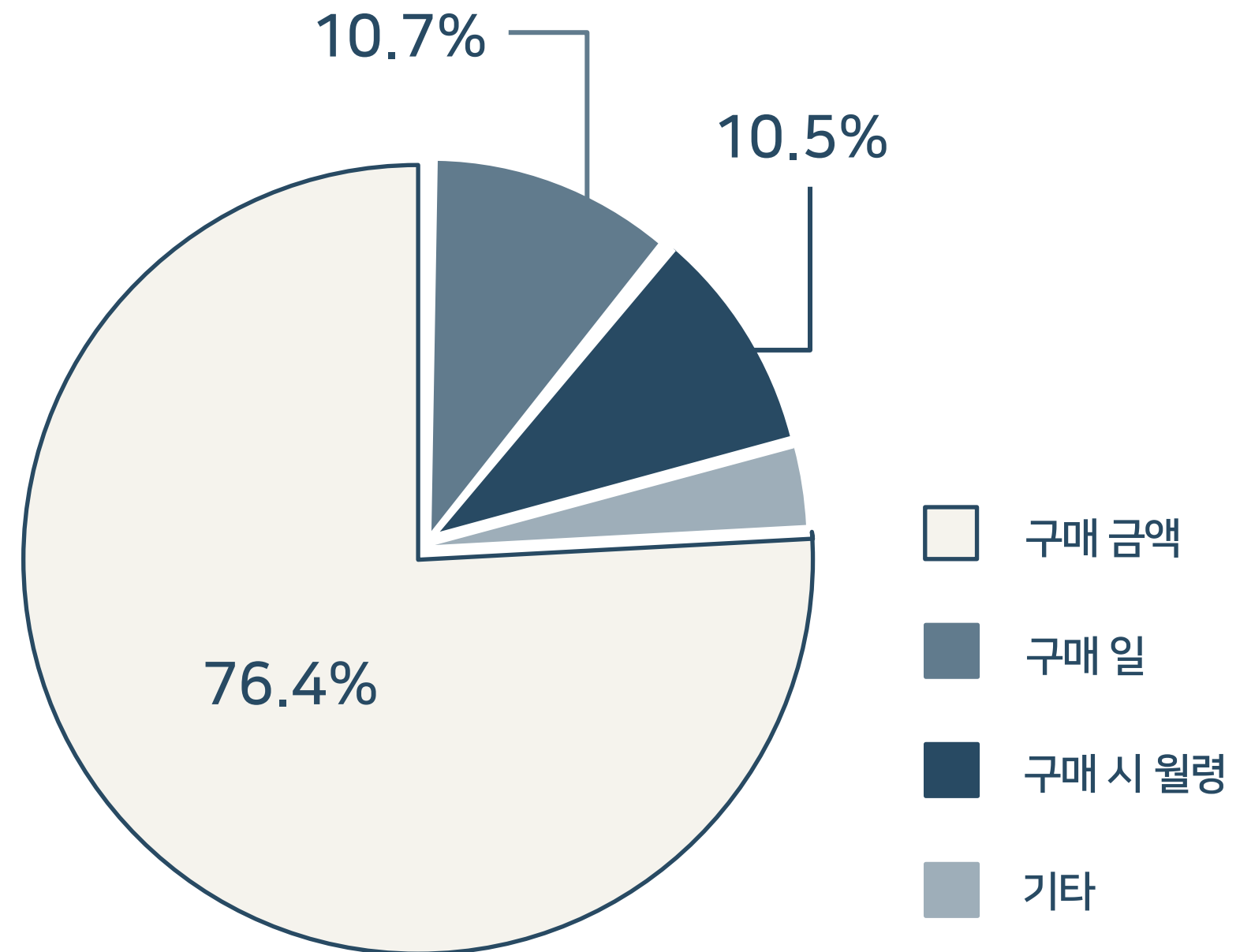
Decision Tree Regressor 모델 적용

- max depth = 30
- min samples leaf = 10
- random state = 100

모델 생성 3 - 물품대분류 별 수요 예측 모델

모델 중요도

'구매 금액'이 76.4%로 중요도가 가장 높음
2순위 '구매 일', 3순위 '구매 시 월령'은
비슷한 정도의 중요도를 보임



4. 프로젝트 수행 결과

서비스 제안 1 - 고객 맞춤 메시지

[100일 후]

물품대분류	상품명
더블하트	모유실감 젖꼭지 L (2입)
기저귀	하기스 네이처메이드 여아 132매 (밴드형)
기저귀	하기스 네이처메이드 남아 132매 (밴드형)
더블하트	모유실감 젖꼭지 M (2입)
체험팩	하기스 맥스드라이 3단계 공용 20매

[영아기]

물품대분류	상품명
더블하트	모유실감 젖꼭지 LL (2입)
체험팩	맘큐 허그박스
더블하트	모유실감 젖꼭지 L (2입)
기저귀	하기스 네이처메이드 여아 132매 (밴드형)
FMC	여행용티슈 카카오프렌즈 70매 X 3개입

주기 별 구매 건수 상위 5가지 상품(물품대분류) 파악

서비스 제안 1 - 고객 맞춤 메시지

안녕하세요 ooo 고객님의 :)

100일 후 아이들은 요즘 □□□ 상품을 사용하고 있습니다.

(상품 링크)

(주기 별 육아 노하우 링크)

.....

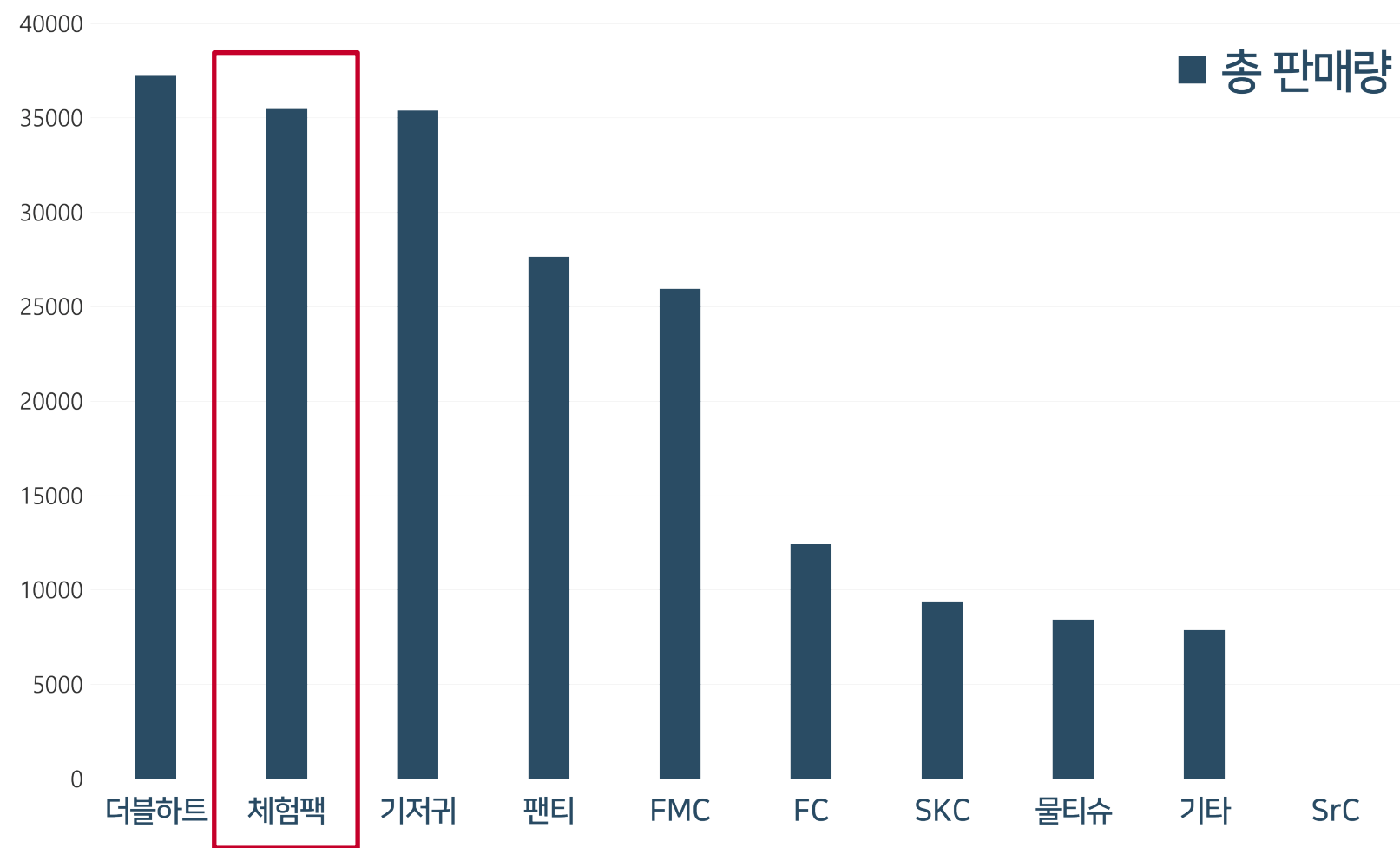
.....

다음 주기인 영아기 아이들은
△△△ 상품을 준비하고 있습니다.
(상품 링크)

‘100일 후’ 주기의 고객에게

- ☑ 같은 주기의 다른 고객은 어떤 상품들을 구매하고 있는지
- ☑ 다음 주기에는 어떤 상품들이 필요하게 될지

서비스 제안 2 - 고객 등급 개편



물품 대분류 별 총 판매량 분석

체험팩이 2순위로 높은 비중 차지

→ 실 구매 전 상품을 체험하고자 하는
고객의 니즈가 높다고 판단

서비스 제안 2 - 고객 등급 개편



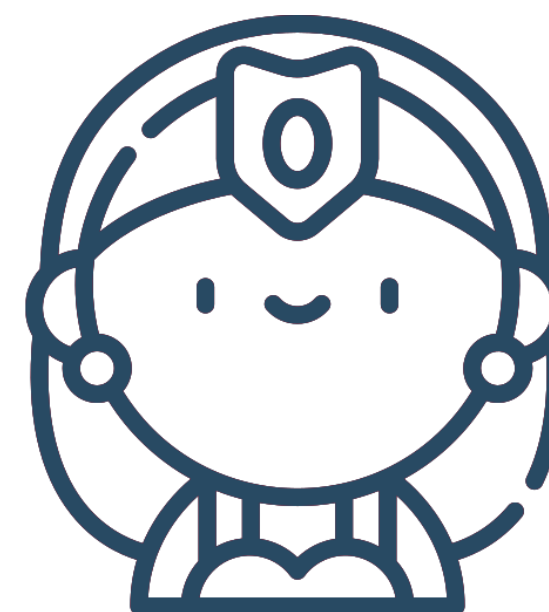
예비맘

- 3개월 누적 결제금액 5만원 이상
- 적립금 1 %
- 1만원 상당의 체험팩 제공



숙련맘

- 3개월 누적 결제금액 7만원 이상
- 적립금 3 %
- 3만원 상당의 체험팩 제공



슈퍼맘

- 3개월 누적 결제금액 10만원 이상
- 적립금 5 %
- 5만원 상당의 체험팩 제공

서비스 제안 2 - 고객 등급 개편

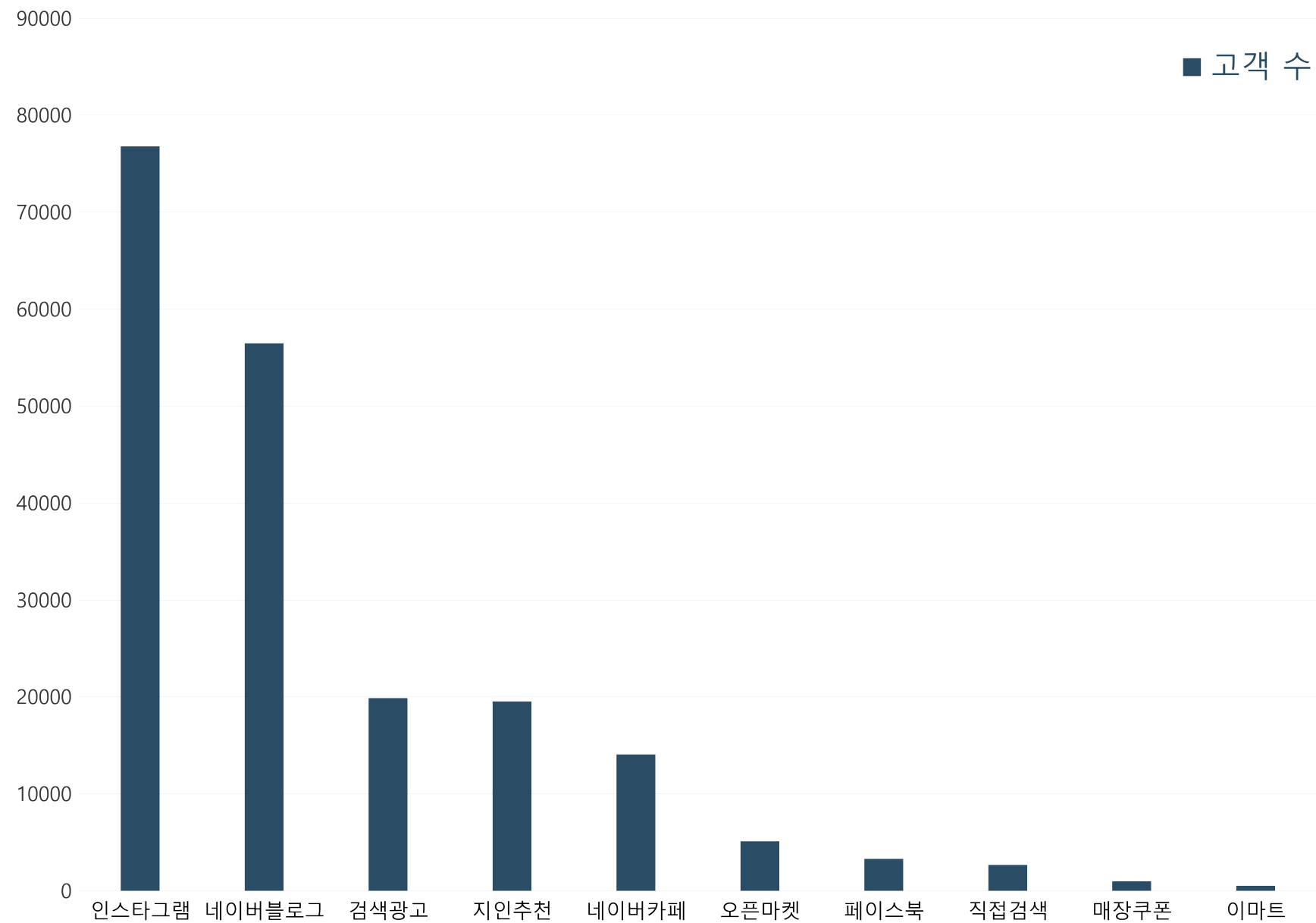


물품 대분류 별 수요예측 모델
→ 물품 대분류 선호도 파악



체험팩 상품 구성

서비스 제안 3 - SNS 광고



유입 경로 별 고객 수

인스타그램과 네이버블로그가 약 67%

→ 세분화된 고객 맞춤형 정보를 토대로
SNS 광고 진행

5. 자체 평가 의견

자체 평가 의견

김성민

고객 정보를 기반으로 하여 고객의 재구매 유도에 초점을 두어 프로젝트를 진행하였습니다. 재구매율을 높이는 방법은 최종적으로 이 쇼핑몰이 갖는 '고객의 이탈 방지'와 '수요량 조절'의 문제점도 잡아줄 수 있습니다. 팀원들과 소통하여 목적에 맞게 데이터를 활용하였고, 이에 대한 모델을 수립하여 솔루션을 제공할 수 있었습니다. 이번 프로젝트를 진행하면서 유아용품 시장을 대략적으로 알게 되었고, 데이터의 중요성과 이 데이터를 분석함에 있어 다방면의 배경지식이 중요한 것을 깨달았습니다.

김여민

유아용품 데이터 중에서 일용품만으로도 고객들의 기준을 나눌 수 있다는 것이 인상 깊었다. 대부분의 소비자가 일용품을 구매한 기록만 남아서 시장의 프리미엄화를 도입하기가 어려웠지만, 데이터를 나누고 합쳐서 머신러닝 모델 도입까지 전부 결과값이 다르게 나와서 흥미로웠다. 이 모든 전처리 과정을 팀원들 없이는 해낼 수 없었고 이 시간이 매우 유익했다

자체 평가 의견

박종서

고객 데이터 중에서 중요하다고 생각했던 부분(구매시 월령)에 초점을 맞추어 프로젝트를 진행하였습니다. 재구매율을 높이기 위한 방법으로 고객 등급제도를 만들었고, 고객 이탈 방지를 위해서 고객 세분화를 통해 맞춤형 메시지 발송 시스템을 만들었습니다. 이 과정을 통해 머신러닝의 중요성을 깨달았습니다.

이송이

러프한 데이터를 사용해 프로젝트를 진행해보니 생각보다 전처리나 데이터 분석에 많은 고민과 시간이 필요하다는 것을 알 수 있었습니다. 데이터 분석 내용이나 문제 해결 방향에 따라 어떤 흐름으로 머신러닝을 활용하면 좋을지 조금 더 고민해 보는 기회가 되었습니다.

주효성

고객 정보 데이터를 토대로 판매기록 데이터와 함께 가공하여 프로젝트를 진행하였다. 머신 러닝의 활용이 문제점 및 해결법을 도출하는데 도움을 주었다. 전체적인 데이터 전처리 및 가공과 여러 머신러닝을 만질 수 있어서 좋은 공부였다.

THANK YOU