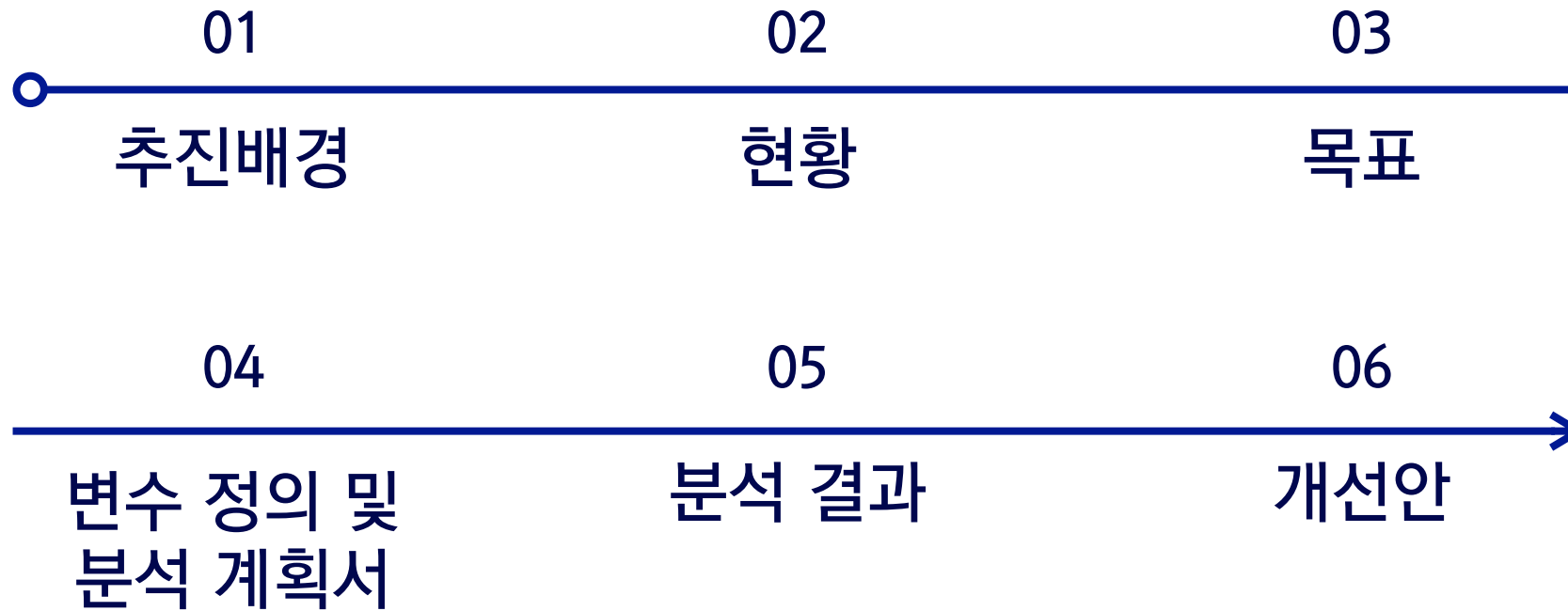


소비자 맞춤형 판매자 추천 기능과 거래성사 가능여부 분류모델을 통한 서비스 차별화 및 거래성사율 증대

1단도전한팀

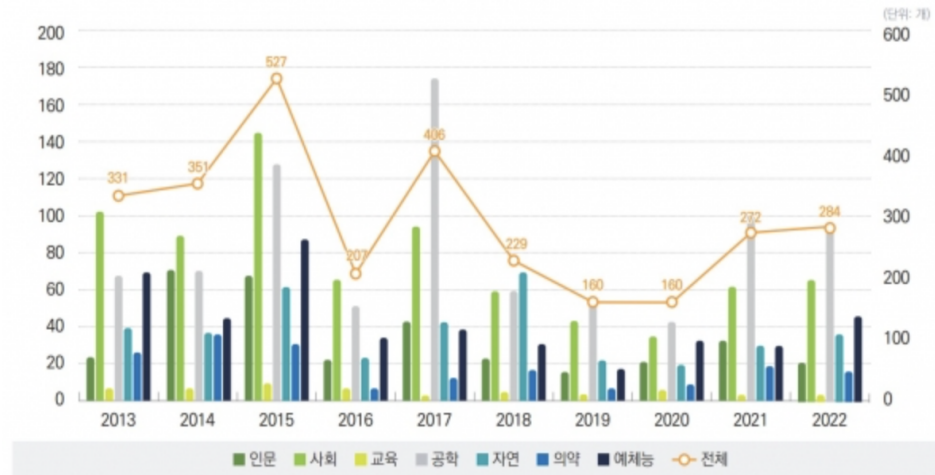
목차



01. 추진배경

- 최근 공학 분야 신설학과 수 증가와 국내 IT 서비스 시장 확대로 IT 서비스 수요 증가
- IT 아웃소싱 업계 경쟁 심화에 따라 경쟁력 확보를 위해 **차별화된 +a 서비스가 필요**

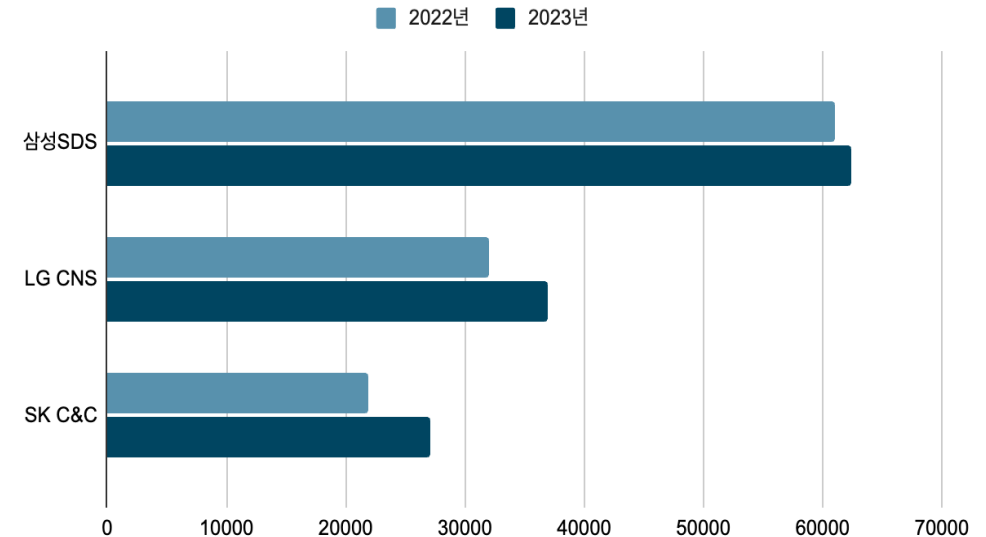
[공학분야 신설 학과/전공 수 변화 추이]



[그림 2] 전공계열별 신설 학과-전공 수 변화 추이(2013~2022)

* 출처 : 한국교육개발원(KEDI)

[주요 IT 서비스사 매출 추이 (단위 : 억 원)]



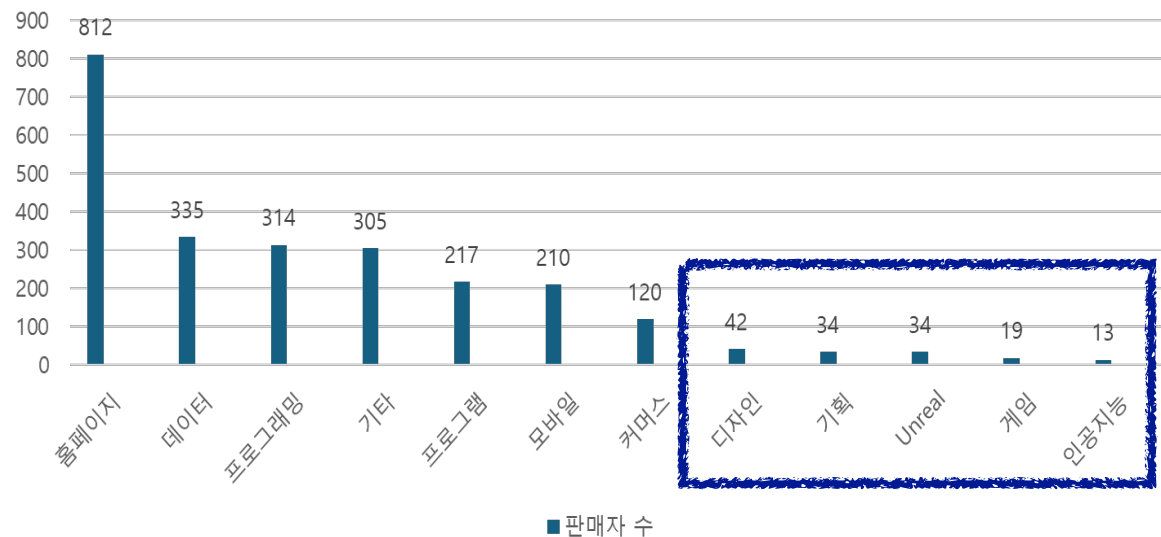
* LG CNS, SK C&C '23년 Q3 누적매출 기준

* 출처: 디지털데일리(www.ddaily.co.kr)

02. 현황분석

- 대분류별 평균 전문가 수는 205명으로 총 12개 중 6개의 카테고리에서 **전문가 풀이 다양함**
- 그러나 디자인, 인공지능 등은 총 거래건수가 높은 인기 분야임에도 불구하고 전문가 수는 현저히 낮은 수준임
- 따라서, 현재 확보된 **전문가의 다양성(강점)**은 살리면서 **대분류별 심한 편차(약점)**은 살릴 방안이 필요함

[대분류별 전문가 수]

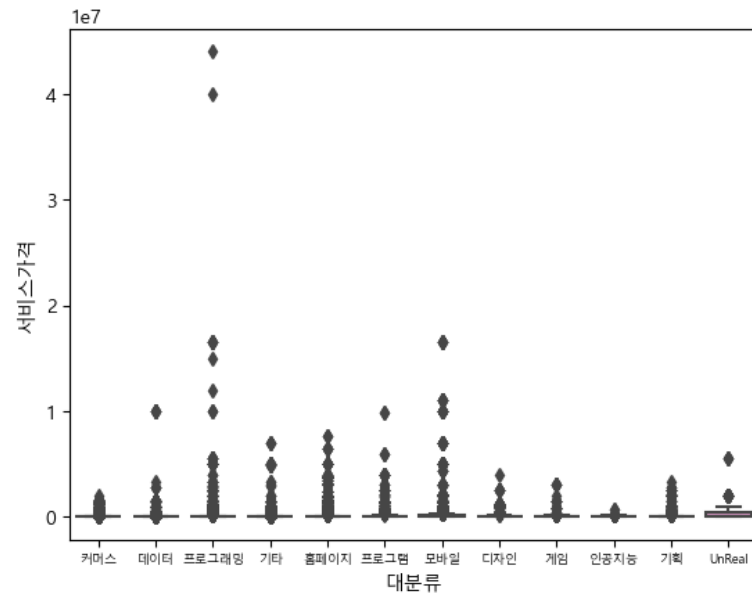


대분류	총거래건수
홈페이지	122812
데이터	61166
디자인	4161
기획	3894
Unreal	707
게임	3013
인공지능	1399

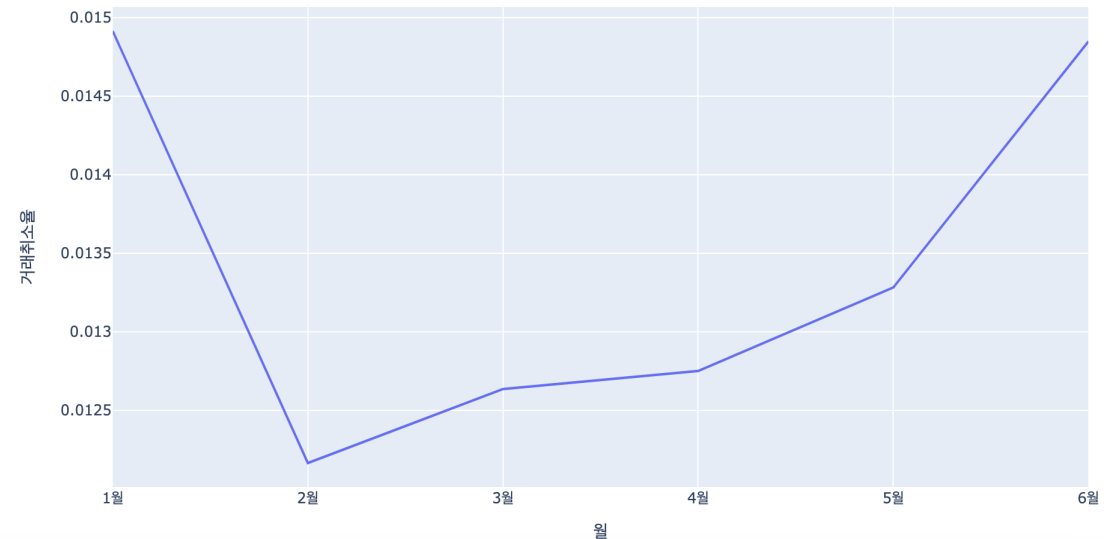
02. 현황분석

- 대분류별 서비스 가격의 분산이 크게 나타나면서 서비스별 표준 가격의 부재로 고객 VOC 증가
- 이로 인해 월별 거래취소율이 상승하는 중
- 따라서, 거래 성사로 이어지는 **적정 가격 수준을 파악**하여 거래 성사 비중을 높일 필요가 있음

[대분류 별 서비스가격 추이]



[월 별 거래취소율]



03. 목표

- ① 소비자 맞춤형 전문가 추천 서비스를 통해 신규고객 유입 유도 및 **판매 기회를 높이고**
- ② 거래성사 가능 여부 모델로 소비자와 판매자의 가격 조정을 유도함으로써 **거래취소율 개선**을 꾀하려 함

측정지표	가중치	현 수준	'24 목표	'25 목표	'26 목표
총 판매기회 건수	40%	56,852건	62,537건 (현 수준+10%)	71,918건 ('24목표+15%)	86,301건 ('25목표+20%)
거래취소율	60%	1.2%	1.0%	0.8%	0.5%

- 총 판매기회 건수 : 거래성사 건수 + 거래취소 건수
- 상기 데이터는 이상치, 결측치 제거 후 정상데이터를 기준으로 했을 때의 수치

04. 데이터 분석 계획서 작성

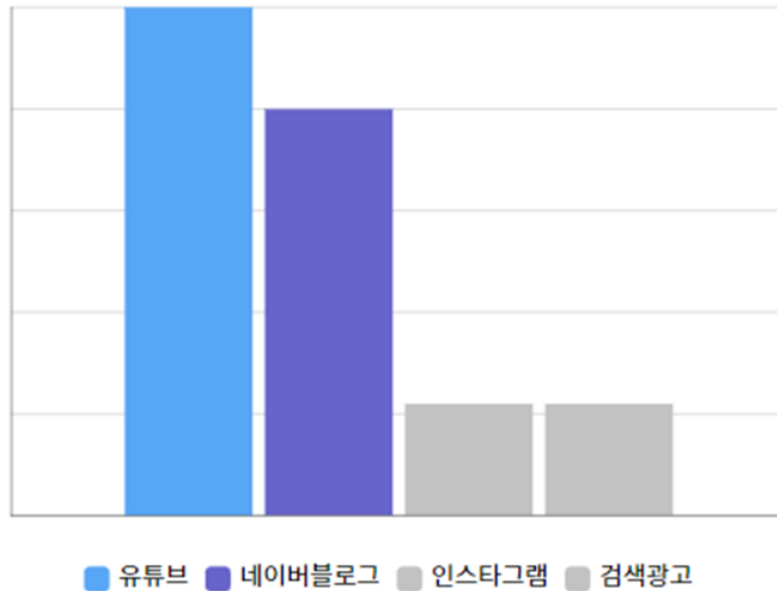
목적		분석방법	분석내용	
전체 데이터 분포 및 이상치 확인		전처리 작업	전체 변수의 구조, 타입, 이상치 확인	
			대치 불가능한 결측치 제거, 중복값 제거	
현황분석	카테고리별 전문가 수	막대그래프	카테고리별 전문가 분포 확인	
	카테고리별 서비스가격 분포	Box-plot	동일한 카테고리 내 서비스 거래가격의 IQR 확인	
	카테고리별 거래수준/피벗테이블 거래건수 확인	라인그래프	월별 매출액(판매금액*수수료율) 확인	
	거래취소율 추이	라인그래프	월별 거래취소율(거래취소건수/전체 거래시도건수)확인	
소비자 맞춤형 서비스 개발		피벗테이블	고객 입력 조건에 따른 추천 판매자 프로필 제공을 위한 피벗테이블	
			input	output
			서비스 카테고리, 희망 최소평점, 지불 용의 최대금액	판매자명/아이디, 서비스명, 평균 서비스금액 및 추가결제금액, 신속알람 서비스 이용여부

04. 데이터 분석 계획서 작성

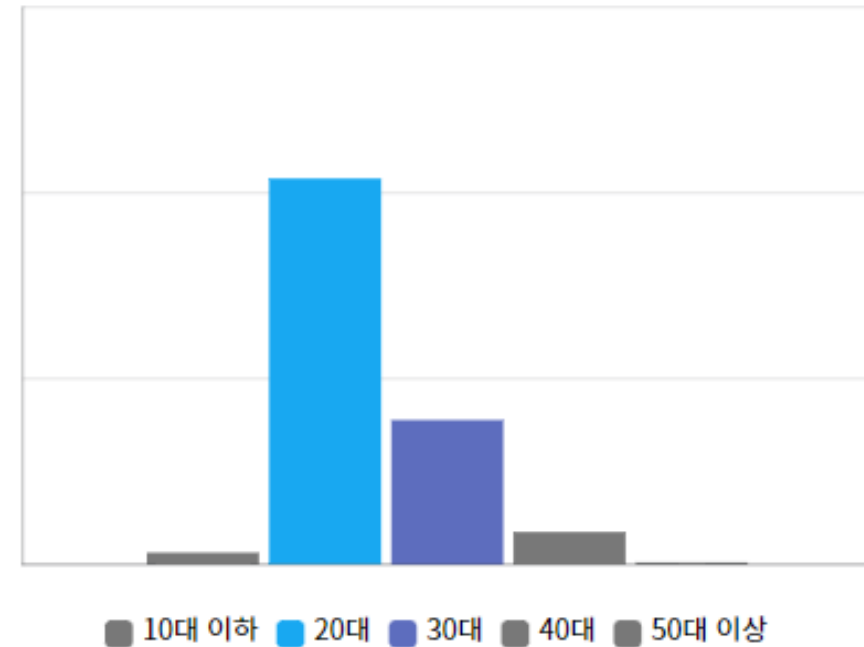
목적	변수	상관관계 예측	분석내용
거래성사율 정체 문제 해결	서비스 평점	(+) 거래성사 Y (변수 값 0)	'거래성사여부'와 상관성이 높은 변수를 추출하기 위해 가설검정 시행
	판매자의 총 판매서비스 수		
	서비스의 누적 판매수		
	판매자의 '프리미어 서비스' 가입여부 (Y)		
	판매자의 '신속알람서비스' 가입여부 (Y)	(-) 거래성사 N (변수 값 1)	
	서비스 가격		
	추가 결제금액 여부 (Y)		
	수수료율		

05. 분석결과(1)

- 고객 유입경로 빈도수 분석 결과 ‘유튜브’와 ‘네이버 블로그’를 통한 유입이 가장 높았음
- 따라서 **유튜브와 네이버 블로그**를 통한 전략적 마케팅 필요



- 연령대 별 서비스 총구매수 분석 결과 20,30대의 비율이 가장 높았음
- 따라서 **20,30대를 공략**하는 마케팅 필요



05. 분석결과(2)



‘거래 성사여부(목표변수) 분류 모델’의 설명변수 선별을 위해 항목별 가설 검정을 진행.

거래 취소 여부에 따른 <평점> 가설검정	
귀무가설	거래 성사 집단과 취소 집단 사이 평점의 차이가 없다.
대립가설	거래 성사 집단과 취소 집단 사이 평점의 차이가 있다.
[Wilcoxon Rank Sum Test] P.value (0.03) < 0.05	

거래 취소 여부에 따른 <서비스가격> 가설검정	
귀무가설	거래 성사 집단과 취소 집단 사이 서비스 가격의 차이가 없다.
대립가설	거래 성사 집단과 취소 집단 사이 서비스 가격의 차이가 있다.
[Wilcoxon Rank Sum Test] P.value (5.26e-19) < 0.05	



평점과 **서비스가격**은 거래 취소 여부와 **상관성이 있으므로** 설명변수로 활용함.

05. 분석결과(2)



‘거래 성사여부(목표변수) 분류 모델’의 설명변수 선별을 위해 항목별 가설 검정을 진행.

거래 취소 여부에 따른 <추가결제금액> 가설검정	
귀무가설	거래 성사 집단과 취소 집단 사이 추가결제금액의 차이가 없다.
대립가설	거래 성사 집단과 취소 집단 사이 추가결제금액의 차이가 있다.
[Wilcoxon Rank Sum Test] P.value (0.0) < 0.05	

거래 취소 여부에 따른 <수수료율> 가설검정	
귀무가설	거래 성사 집단과 취소 집단 사이 수수료율의 차이가 없다.
대립가설	거래 성사 집단과 취소 집단 사이 수수료율의 차이가 있다.
[Wilcoxon Rank Sum Test] P.value (0.60) > 0.05	



추가결제금액은 거래 취소 여부와 **상관성이 있으므로** 설명변수로 활용함.

수수료율은 거래 취소 여부와 **상관성이 없으므로** 설명변수로 활용하지 않음.

05. 분석결과(3)



‘거래 성사여부(목표변수) 분류 모델’을 위해 주어진 데이터의 전처리 진행.

```
# SMOTE를 적용할 때는 반드시 학습 데이터 세트만 오버 샘플링을 해야함

from imblearn.over_sampling import SMOTE

smote = SMOTE(random_state=0)
X_train_over, Y_train_over = smote.fit_resample(X_train, Y_train)

print('SMOTE 적용 전 학습용 피쳐/레이블 데이터 세트:', X_train.shape, Y_train.shape)
print('SMOTE 적용 후 학습용 피쳐/레이블 데이터 세트:', X_train_over.shape, Y_train_over.shape)
print('SMOTE 적용 후 레이블 값 분포: \n', pd.Series(Y_train_over).value_counts())
```

SMOTE 적용 전 학습용 피쳐/레이블 데이터 세트: (255835, 3) (255835,)

SMOTE 적용 후 학습용 피쳐/레이블 데이터 세트: (504808, 3) (504808,)

SMOTE 적용 후 레이블 값 분포:

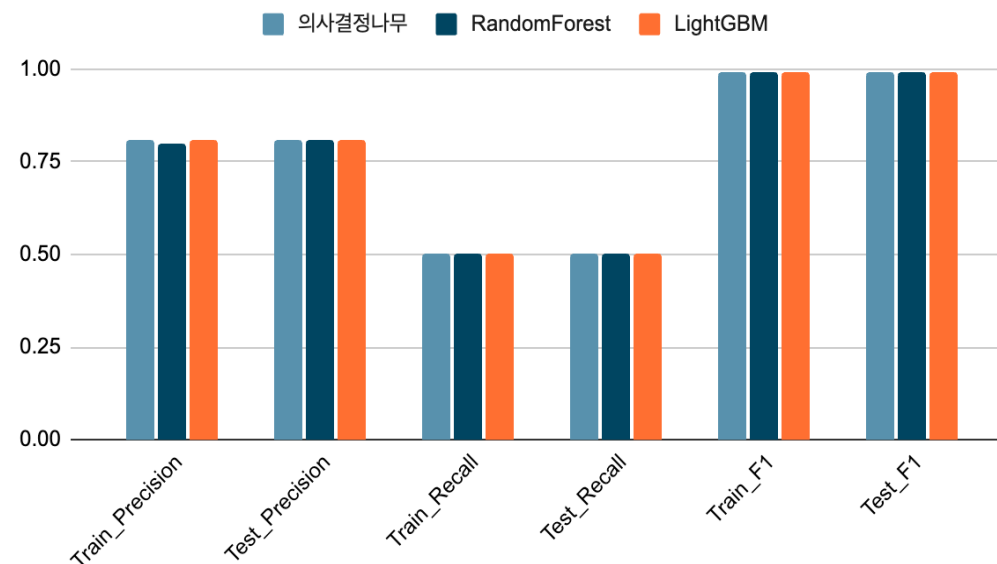
거래취소여부

0.0 252404

1.0 252404

Name: count, dtype: int64

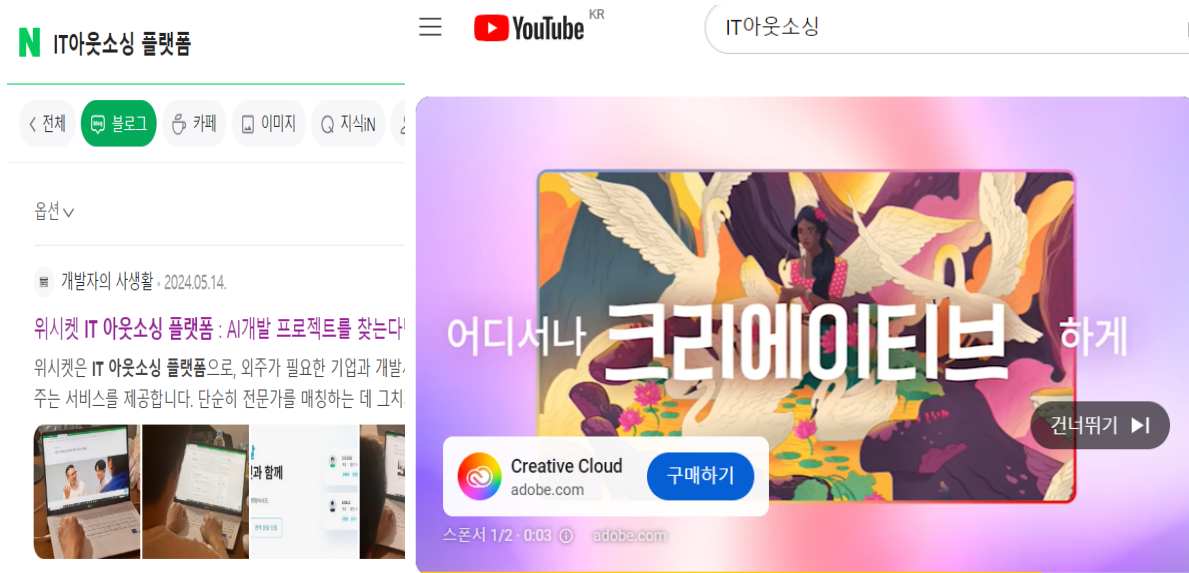
주어진 로그 데이터는 거래성사 336,523개,
거래취소 4,591개로 불균형이 심하므로
SMOTE 기법을 이용해 Over-Sampling 시도



	의사결정나무	RandomForest	LightGBM
Train_Precision	0.81	0.8	0.81
Test_Precision	0.81	0.81	0.81
Train_Recall	0.5	0.5	0.5
Test_Recall	0.5	0.5	0.5
Train_F1	0.99	0.99	0.99
Test_F1	0.99	0.99	0.99

06. 개선안(1)

💡 ‘유튜브’와 ‘네이버 블로그’를 통한 전략적 마케팅



유튜브는 삽입 광고를 통해,
네이버 블로그는 블로그 체험단을 운영하여,
L사의 인지도를 높일 것을 제안하고자 함.

💡 플랫폼 이용자 중 상위 연령대 맞춤 공략

홈페이지 개설이 필요한
00~90년생 초보 사장님들에게
서비스 이용 10% 쿠폰을 드립니다!



플랫폼을 가장 많이 이용하고 있는
20-30대를 맞춤 공략한 마케팅 실시를
제안하고자 함.

06. 개선안(2)



소비자 맞춤형 판매자 추천 프로그램

```
# 사용자 입력 받기
input_cate = input("원하는 서비스 카테고리를 입력하세요: ")
input_rating = float(input("희망하는 최소 평점을 입력하세요 (5점 만점): "))
input_price = float(input("희망하는 최대 가격을 입력하세요: "))

# 조건에 맞는 데이터 필터링
filtered_a2 = a2[
    (a2['대분류'] == input_cate) &
    (a2['평점'] >= input_rating) &
    (a2['판매금액'] <= input_price)
]

# 중복된 판매자 제거
filtered_a2 = filtered_a2.drop_duplicates(subset='판매자')

# 필요한 컬럼 선택 및 정렬
output_a2 = filtered_a2[['판매자', '판매자ID', '서비스명', '신속 알람 서비스 사용여부', '판매금액', '평점']]
output_a2 = output_a2.sort_values(by=['신속 알람 서비스 사용여부', '평점', '판매금액'], ascending=[False, False, True]).head(5)

# 결과 출력
print(output_a2)
```

원하는 서비스 카테고리를 입력하세요: 홈페이지
 희망하는 최소 평점을 입력하세요 (5점 만점): 3
 희망하는 최대 가격을 입력하세요: 150000

판매자	판매자ID	서비스명	신속 알람 서비스 사용여부	₩	사용
2217	규루리 V2_899	PHP JSP 웹사이트 기능개선 및 수정해드립니다	사용		
2808	함께나무 V2_1343	웹사이트 간단 수정 및 퀄리티 업그레이드해 드립니다.	사용		
2028	Tabago V2_813	웹사이트 오류 버그 수정해 드립니다.	사용		
2218	그래그래그래돌라 V2_1828	내 마음에 꼭 맞는 홈페이지를 신속하게 작업해 드립니다.	사용		
2240	품품한훈이 V2_1276	카페24,그누보드,영카트 홈페이지 제작 수정 유지보수해 드립니다.	사용		

판매금액	평점
2217 7392.591687	5.0
2808 7475.290698	5.0
2028 7488.224299	5.0
2218 7560.416667	5.0
2240 7603.613861	5.0

고객이 질문에 따라 조건을 입력하면
 피벗 결과에 따라 상위 판매자를
 추천하는 프로그램 제안

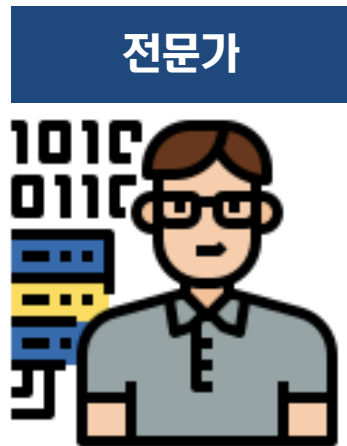
고객 입력 조건	판매자 추천결과
서비스 카테고리 희망 최소평점 희망 최대가격	판매자명/아이디, 서비스명을 신속 알람 서비스 이용여부에 따라, 평균 서비스가격은 낮은 순, 추가결제금액은 낮은 순으로, 평점은 높은 순으로 상위 5명의 판매자를 추천

- 현재 다양한 판매자 pool과 프로파일 요소를 보유하고 있으므로 소비자 니즈에 따라 적합한 판매자를 추천하는 차별화 서비스를 제안하고자 함.
- 바이럴 유도 및 신규고객 유입, 판매기회 확대 등의 효과를 기대할 수 있음.

06. 개선안(3)

💡 거래 성사여부 분류모델

‘평점, 서비스가격, 추가 결제금액’ 이 설명변수(X), ‘거래성사 가능 여부’ 가 목표변수(Y)인 <분류모델>을 개발하여 판매자와 고객의 가격조정을 유도하고 거래성사율을 개선하고자 함.



“ 거래가 성사되지 않을 것 같네?
서비스 가격을 낮춰봐야겠군. ”



“ 이 서비스를 이용하고 싶은데 ...
더 높은 가격을 지불해야 하나? ”

프로젝트 소감

이론강의에서 배웠던 내용을
프로젝트에 다 쏟아내야했을 때
기억이 나지 않아 매우 힘들었지만,
배웠던 내용들을 다시 꺼내
실제 프로젝트에 적용하는 과정들이
모두 의미있었다고 생각합니다.
(신현호)

4주라는 짧은 시간동안 배운 내용을
프로젝트에 담아내는 과정이 결코 쉽진 않았지만
직접 데이터 분석을 해볼 수 있어 좋았습니다.
특히 하나의 제품/서비스가 만들어지기까지
데이터 분석이 미치는 영향력에 대해
다시 한 번 생각해볼 수 있었던 보람찬 시간이었습니다.
(박예슬)

파이썬의 '파'도 몰랐던 몇 주 전이지만,
실제 데이터를 가지고 이렇게 인사이트를
도출해 낼 수 있다는 것이 놀라웠습니다.
데이터분석의 재미를 알게 된 기회였습니다.
(조수연)

짧은 시간 동안 데이터 분석에 대해
알차게 배워갈 수 있어서 좋았습니다.
실제 데이터들을 가지고 프로젝트를 진행해 볼
수 있었다는 점에서도 좋았습니다. 데이터분석에
대한 역량을 기를 수 있었던 시간이었습니다.
(김승현)

강의를 들을 때에는 쉽게만 느껴지던
것들이 실제 데이터를 활용할 때에는
전혀 다른 내용처럼 느껴졌습니다.
혼자 찾아보고 공부해야 할 내용들이 많았지만,
그만큼 배우고 체화한 것들이 많다고 생각합니다.
(장윤경)