



소비자 맞춤 카드 추천 시스템

머신러닝을 통한 다음달 소비 예측 및 카드 추천

2023.05.10

3조 - Clovacash

조장: 남호제 / 조원: 박세형, 우정연, 이성민

Contents

PART1.

1. 제안 배경
2. 제안 목표

PART2.

1. 사용 데이터 소개
2. 데이터 전처리
3. 데이터 EDA

PART3.

1. 모델링
2. 모델 성능 평가

PART4.

1. 카드 추천 Streamlit 구현
2. 시연

PART5.

1. 결론 및 마무리
2. Q&A

PART1. 제안 배경



고객
다양한 소비 성향의 증가



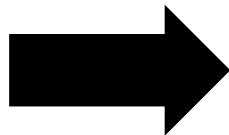
카드 회사
마이 데이터를 통한 개인화



비즈니스
사용자 데이터 기반
맞춤 추천의 효율성 증대



접근성
많은 카드 정보를 일일이 검색



하나의 사이트에서
다양한 소비 성향을 반영한
카드 추천

PART1. 제안 목표



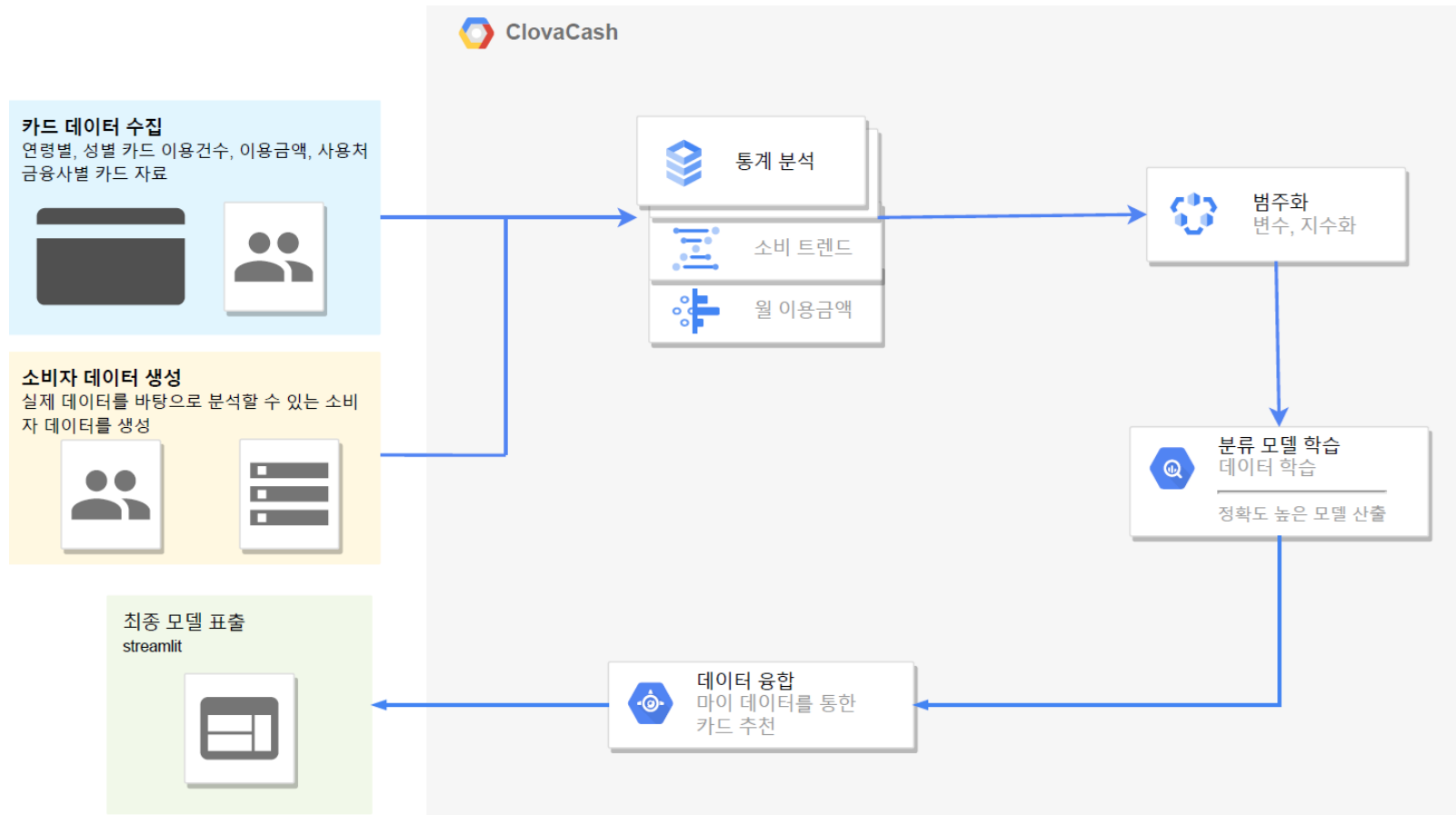
누구나 손쉽게 이용할 수 있는 추천 플랫폼

"마이 데이터" 기반

개인 소비 특성 맞춤형 카드 상품

추천 서비스

PART2. 데이터 및 EDA / 분석 flow



PART2. 사용 데이터 소개 / 고객 데이터 생성

Python의 fakers 라이브러리를 이용해서 약 고객 1만명의 가상의 회원 데이터 (이름, 성별, 나이, 사용처 등)을 생성하고 카드 사용 날짜 등을 난수를 이용해 랜덤하게 생성. 실제 카드 정보를 담은 데이터와 병합

customer	gender	age	card_name	date	merchant	amount	company	installment	location	category	point	
0	C024	F	30	마일앤조이카드(대한항공)	2023-01-19	면세점	17702	IBK기업은행	12	제주	여행/교통	TOP POINT1
1	C018	M	50	삼성 iD SIMPLE 카드	2022-12-21	자동차용품	1148	삼성카드	20	부산	자동차	보너스포인트
2	C045	F	60	K-FIRST (모바일전용)	2023-02-02	할인점/슈퍼마켓	80319	BC카드	13	대전	유통	TOP POINT2
3	C048	M	40	신한카드 Deep Oil	2022-12-17	농수산물	31451	신한카드	16	제주	음/식료품	마이신한포인트
4	C006	F	70	신세계 BC 바로 리워드 플러스	2023-04-10	LG	5251	BC카드	16	경북	주유	네이버 포인트
5	C050	M	10	탄탄대로 올쇼핑 티타늄카드					29	인천	자동차	포인트리
6	C020	M	70	다담카드					6	서울	의료	포인트리
7	C030	M	50	현대카드ZERO Edition2(할인)					6	충북	의료	M-POINT
8	C050	M	10	탄탄대로 올쇼핑 티타늄카드					31	대구	가정생활/서비스	포인트리
9	C015	M	60	현대카드ZERO Edition2(할인)					18	경기	전자상거래	네이버 포인트
10	C038	F	20	별다줄카드	2022-10-29	중식	6942	NH농협카드	17	울산	요식/유통	NH포인트
11	C033	F	40	BLISS.7 카드(포인트)	2022-10-29	자동차용품	1299	IBK기업은행	14	인천	자동차	네이버 포인트
12	C034	F	50	마일앤조이카드(아시아나)	2022-05-11	호텔	13131	IBK기업은행	29	인천	여행/교통	TOP POINT1
13	C025	F	50	신한카드 래플(Raffle)	2022-12-13	주유소	3741	신한카드	32	전남	주유	마이신한포인트
14	C017	M	60	NH20 해봄카드	2022-09-30	의복/의류	1870	NH농협카드	25	전북	의류/잡화	네이버 포인트
15	C002	F	30	카드의정석 POINT	2023-03-23	미용실	6167	우리카드	6	부산	미용	네이버 포인트
16	C004	F	40	THE iD. PLATINUM(포인트)	2022-07-01	통신	34672	삼성카드	17	충북	가정생활/서비스	보너스포인트
17	C016	M	40	NH20 해봄카드	2022-07-06	자동차서비스	1473	NH농협카드	10	전남	자동차	NH포인트

10000건 생성

PART2. 사용 데이터 소개 / 고객 데이터 생성

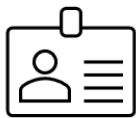
카드 사용 데이터

고객	성별	연령대	카드명	결제일자	사용처	사용금액	업종명
C001	M	20	신한카드 Mr.Life	2023-01-05	편의점	2500	유통
C001	M	20	신한카드 Mr.Life	2023-01-05	세탁소	8000	가정생활/ 서비스
C002	F	30	현대카드 MBOOST	2022-12-13	호텔	320000	여행/교통
C003	F	40	올바른 FLEX카드	2022-05-10	커피	12000	요식/유흥



PART2. 사용 데이터 소개 / 데이터 EDA

데이터 속성 및 변수



고객

사용 고객 구분
C + 3자리 숫자로 표시



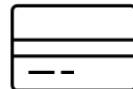
성별

남여 성별 구분
M(남성) / F(여성)



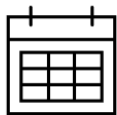
연령대

연령별 구간
10대~70대



카드명

사용 카드 이름
혜택 정보가 있는 카드



결제일자

카드를 사용한 날짜
년-월-일로 표시



사용처

결제한 장소
업종명에서의 소분류



사용금액

결제시 사용한 금액
단위: (원)



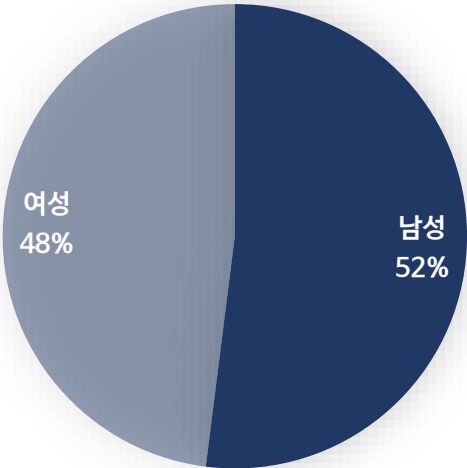
업종명

결제한 업종
사용처의 대분류

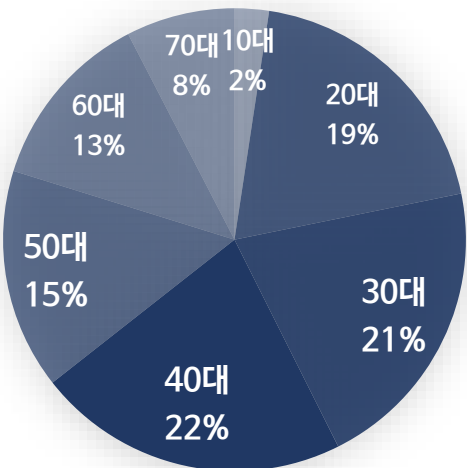
PART2. 사용 데이터 소개 / 고객 데이터 EDA

고객 카드 소비 패턴 EDA

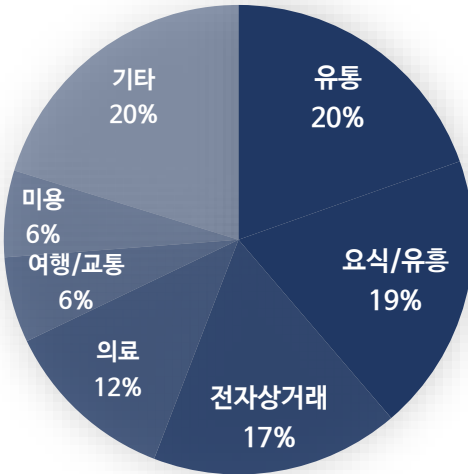
성별 비율



연령별 비율



업종별 소비 비율



PART2. 사용 데이터 소개 / 카드 데이터 생성

‘카드 고릴라’ 사이트 카드 정보





신한카드 Mr.Life

신한카드

자세히 보기

온라인발급 전용 카드

공과금	마트,편의점	식음료
10% 할인	10% 할인	10% 할인

해외겸용 15,000원 전월실적 30만원 이상

신한카드 이용하고 최대 17만원 캐시백 받아가자!

card_data.csv

카드이름	카드혜택	연회비	전월실적
신한카드 Mr.Life	공과금 10% 할인...	해외겸용 15,000원	30만원 이상



웹 크롤링 (Web Crawling)

PART2. 사용 데이터 소개 / 카드 데이터 EDA

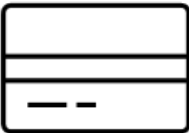
‘카드 고릴라’ 카드 정보 EDA

 신한카드		 삼성카드	
 KB국민카드		 롯데카드	
 하나카드		 우리카드	
 NH농협카드		 IBK기업은행	
 현대카드		 BC 바로카드	



카드회사

10개의 카드사, 은행
신한카드,삼성카드,KB국민카드,롯데카드,
하나카드,우리카드,NH농협카드,IBK기업은행,
현대카드,BC바로카드



카드 정보

1017개의 카드 정보
카드 고릴라에서 카드사와 은행과 협업하여
등록한 1017개의 카드 혜택 정보



카드 사용

성인 평균 2.5개의 카드 사용
* KDI경제정보센터 2022.05.30

PART3. 모델링 및 평가 / 모델링

- 1. 사용모델 :
 - 데이터가 많고 변수가 다양하므로 앙상블(Ensemble) 모델을 활용
 - 모델 최적화를 위해 hyperparameter tuning 진행 (GridSearchCV)
- 2. 사용변수:
 - 고객 기본 속성(성별, 나이대) 및 사용 정보(카드정보, 사용처 등) 으로 예측 모델 학습
 - 문자형 컬럼은 학습을 위해 label encoding 진행

변수 예시 ex)

고객 속성

10대	10~19세
20대	20~29세
30대	30~39세
40대	40~49세
50대	50~59세
60대	60~69세
70대	70~79세

카드 정보

현대카드zero edition2
삼성 id SIMPLE
신한카드 래플(Raffle)
BC 바로 에어 플러스
⋮
롯데백화점 롯데카드
NH20 해봄카드

63개

카드 사용처

할인점/슈퍼마켓
농수산물
약국
스포츠/레저용품
⋮
여행사
생활잡화

70개

모델 학습을 위한 dataset 분리

6	7	8	9	10	11	12	1	2	3	4	5
Train											
									Test		

변수설명:

독립변수 -> 고객 고유 속성 및 카드 사용 데이터 (사용처, 할부개월, 구매 횟수 등)

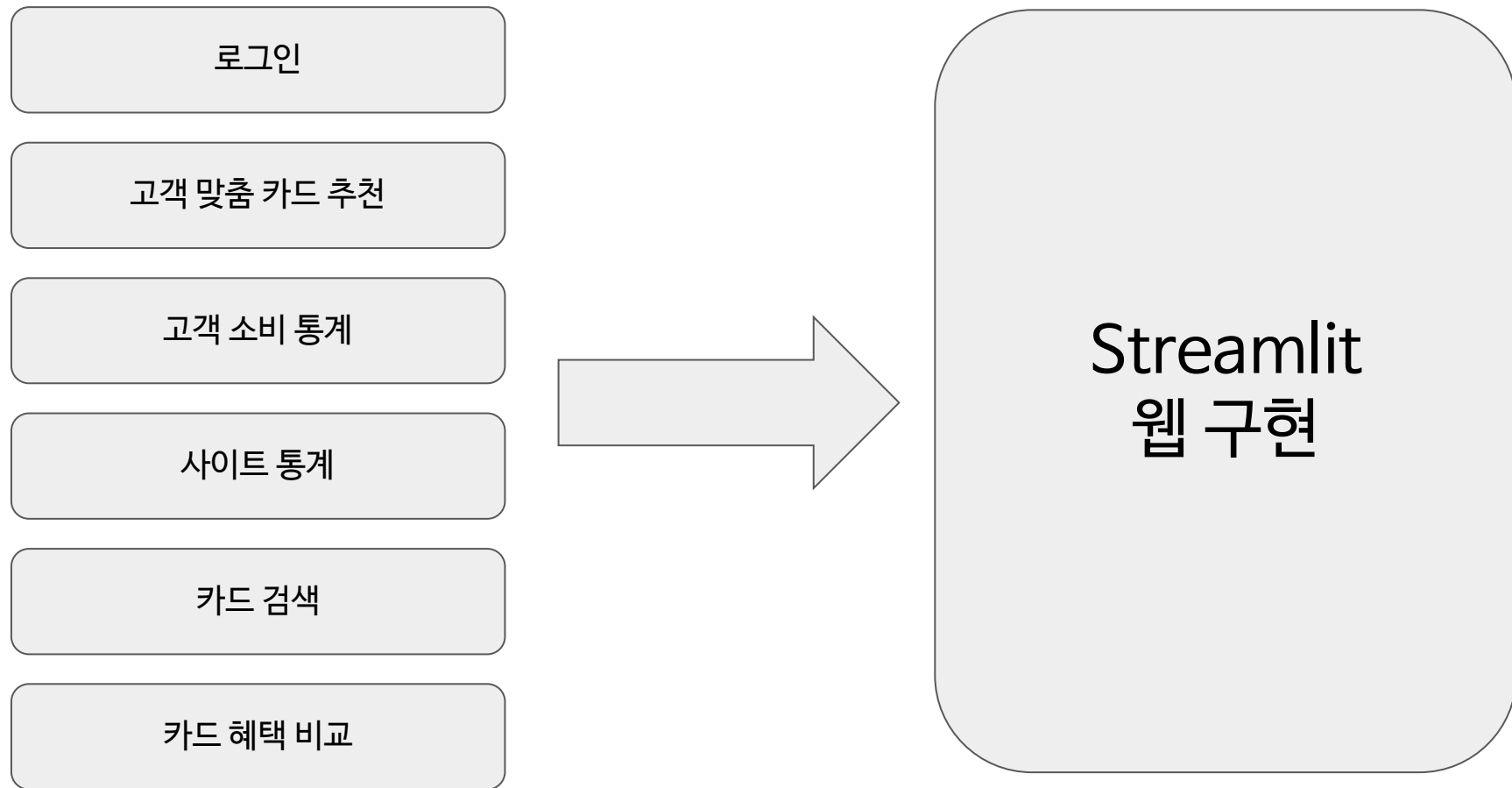
종속변수 -> 카드 사용 금액

PART3. 모델링 및 평가 / 모델 성능 평가

분류 모델별 정확도

	✓ LightGBM	XGBoost	DecisionTreeClassifier	RandomForestClassifier
Accuracy	0.9767	0.9267	0.8200	0.6100
Precision	0.9700	0.9242	0.8062	0.5739
Recall	0.9688	0.9195	0.8129	0.5882
F1 score	0.9677	0.9163	0.8005	0.5635

PART4.카드 추천 Streamlit 구현/ 구현 flow



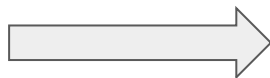
클로바캐시



User ID

C001

Password



자신의 아이디 및 비밀번호 기입
후 사이트 로그인

Log in

PART4. 카드 추천 Streamlit 구현 / 카드추천 page

main

고객님을 위한 카드 추천

고객님의 소비 통계

사이트 이용자들의

다른 카드 검색

다른 카드 혜택 비교

버튼 클릭

추천 카드

고객님의 최근 3개월 소비 패턴을 바탕으로
앞으로 쓰기 적절한 카드입니다.



소비 패턴
기반 추천

LOCA LIKIT Shop

롯데카드



LOCA LIKIT Shop

이 카드는 고객님의 소비 패턴 중
전자상거래에서 기존 카드보다

110,790원 만큼 할인이 됩니다!

	LOCA LIKIT Shop
혜택1	온라인쇼핑 60% 할인
혜택2	미용실 60% 할인
혜택3	편의점 60% 할인

PART4. 카드 추천 Streamlit 구현 / 고객 소비통계 page

- Main
- 고객님을 위한 카드 추천
- 고객님의 소비 통계**
- 사이트 이용자들의 통계
- 다른 카드 검색
- 다른 카드 혜택 비교

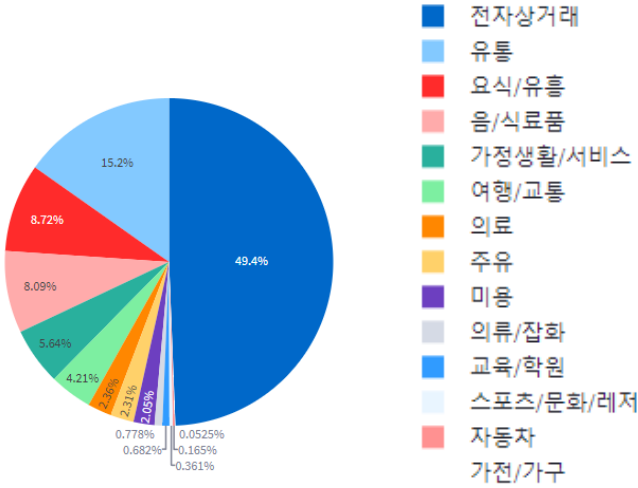
현재 사용하시는 카드는 다음과 같습니다.

현대카드ZERO Edition2(할인형)

1년간 사용한 카테고리 지난달 사용 금액 나만의 통계표

1년간 총 사용한 카테고리: (2022-05-11 ~ 2023-05-07)

사용 통계



고객 사용 통계량 표시

사용 통계 데이터

순위	카테고리	사용금액(원)	사용비중(%)
1위	전자상거래	1,090,400	49.38
2위	유통	335,730	15.21
3위	음식/유흥	192,570	8.72
4위	음/식료품	178,670	8.09
5위	가정생활/서비스	124,560	5.64
6위	여행/교통	92,910	4.21
7위	의료	52,010	2.36
8위	주유	50,930	2.31
9위	미용	45,230	2.05
10위	의류/잡화	17,180	0.78

PART4. 카드 추천 Streamlit 구현 / 사이트 통계 page

Main

고객님을 위한 카드 추천

고객님의 소비 통계

사이트 이용자들의 통계

다른 카드 검색

다른 카드 혜택 비교

클로바캐시에 저장된 고객들의 통계입니다.

총 50명이 이용하고 있어요!

저장된 고객 통계 1년간 사용한 카테고리 지난달 사용 금액

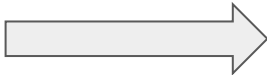
클로바 캐시에 저장된 고객들의 통계입니다.

성별 사용 통계



남성

여성



사이트 이용고객 통계량 표시

성별 사용 통계 데이터

	고객수(명)	비율(%)
남성	26	52.0000
여성	24	48.0000

PART4. 카드 추천 Streamlit 구현 / 카드 검색 page

Main

고객님을 위한 카드 추천

고객님의 소비 통계

사이트 이용자들의 통계

다른 카드 검색

다른 카드 혜택 비교

카드 선택

신한카드 Mr.Life



신한카드 Mr.Life

버튼 클릭

다른 카드가 알고 싶다면? ->
얼마든지 검색 가능!

	혜택1	혜택2	혜택3
0	공과금 10% 할인	마트,편의점 10% 할인	식음료 10% 할인

PART4. 카드 추천 Streamlit 구현 / 카드 혜택 비교 page

- Main
- 고객님을 위한 카드 추천
- 고객님의 소비 통계
- 사이트 이용자들의 통계
- 다른 카드 검색
- 다른 카드 혜택 비교**

카드 혜택 비교

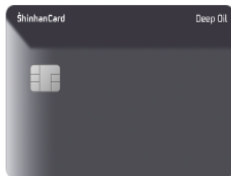
카드 3개를 골라주세요

신한카드 Mr.Life x 신한카드 Deep Oil x 신한카드 Deep ... x



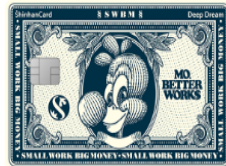
신한카드 Mr.Life

	0
카드이름	신한카드 Mr.Life
연회비	해외경용 15,000원
전월실적	전월실적 30만원 이상
혜택1	공과금 10% 할인
혜택2	마트,편의점 10% 할인
혜택3	식음료 10% 할인



신한카드 Deep Oil

	2
카드이름	신한카드 Deep Oil
연회비	해외경용 10,000원
전월실적	전월실적 30만원 이상
혜택1	4개정유사 10% 할인
혜택2	차량서비스 10% 할인
혜택3	편의점/커피/택시 5% 할인



신한카드 Deep Dream(모베리웍스)

	3
카드이름	신한카드 Deep Dream(모베리웍스)
연회비	해외경용 10,000원
전월실적	전월실적 없음
혜택1	모두드림 0.7% 적립
혜택2	더해드림 2.1% 적립
혜택3	챙겨드림 3.5% 적립



자신의 카드와 다른 카드를 비교해보고 싶다면?

-> 혜택을 비교하고 싶은 카드를 자유롭게 선택 하여 비교 분석

Clovacash

(카드 추천 시스템)

시연

✓ 실제 마이 데이터나 고객의 소비 데이터가 필요

랜덤한 데이터 생성 및 개인적인 소비 성향을 예측 하고자 했지만 실제 진행한 모델에 대해서는 변수가 적었는지 Rule Base(특정 영역에 해당되면 고정된 결과가 나오는 것)형태가 계속해서 나타났습니다. 실제 마이 데이터나 고객의 개인 성향을 좀 더 반영할 수 있었다면 좀 더 소비자에 맞춰서 예측이 가능할 것이라 생각합니다.

✓ 데이터 전처리와 EDA의 중요성

데이터 수집 과정부터 분석, EDA까지 머신 러닝을 동작 시키기 전 사전 작업에 많은 시간과 노력이 필요한 것을 다시한번 느끼게 되었습니다. 앞으로 데이터 분석가 및 프로그래머로 일하더라도 정확한 예측을 위해 지금까지 배웠던 내용을 바탕으로 데이터 분석 및 시각화에 집중할 것 입니다.

Part 5. 결론 및 마무리 / 사용 프로그램 소개

- 사용 언어 : Python (ver.3.7.4)
- 사용 라이브러리 : Pandas, Numpy, Matplotlib, Seaborn, Plotly, Fakers, BeautifulSoup, Selenium
- 학습 모델 : Sklearn(DecisionTreeClassifier, XGBoost, LightGBM, RandomForestClassifier)
- 협업 어플리케이션 : Notion, Microsoft PowerPoint, Microsoft Excel
- 웹 표시 : Streamlit

역할 분담

- 남호제 : 고객 데이터 수집, 보고 및 문서작업
- 박세형 : 카드 정보 데이터 수집, 크롤링 및 카드 데이터 관리
- 우정연 : 데이터 전처리 및 학습 모델링 구상
- 이성민 : Streamlit 제작, 크롤링 및 카드 데이터 관리
- 공동작업 : 발표자료 제작, Streamlit UI 설정

후기

남호제: 6개월 배웠던 것을 바탕으로 데이터를 분석하고 직접 웹상에 표출하기 까지 오랜 시간이 걸렸습니다. 복잡한 작업도 많고 한달이 결코 긴 시간은 아니지만 끝까지 노력해 주신 팀원분들이 있었기에 마무리 할 수 있었습니다. 너무 고생 많으셨습니다!

박세형: 1달이라는 시간 내에서 팀원들과 최선을 다해서 진행한 프로젝트가 완성이 되어서 너무 기쁘지만 짧은 시간내에 구현을 하다 보니 아쉬운 점이나 부족한 점이 있지만 그래도 그 과정에서 많은 것을 배울 수 있었습니다. 모두 고생들 많으셨습니다!!

우정연: 많은 시행착오와 어려움이 있었고, 팀원들마다 분석 및 개발 경험이 달라서 고난이 있었지만 결국 프로젝트를 완성할 수 있게 되어 매우 보람을 느낍니다! 짧지 않은 시간동안 주말 상관없이 너무 고생 많으셨습니다~!!

이성민: 최종프로젝트에서 팀원들과 의견충돌 없이 잘 마무리 하여서 다행이라 생각하고, 프로젝트 진행중에 많은 도움이 되지 못한 것 같아 미안하네요.. 한달이란 시간동안 고생하셨습니다!!

Q&A