# Lpoint 어플리케이션 쇼핑 탭 활성화

고객 군집화 기반 상품 추천 시스템

TEAM INDEX

팀장이충성팀원윤종찬팀원오승용

- INDEX 1. 문제정의
  - **EDA**
  - 3. 전처리
  - 4. 군집화
  - 5. 상품추천
  - 6. 마케팅 전략

# 문제 정의

# 1. 문제 정의

문제정의 EDA 군집화 상품추천 마케팅

### 엘포인트 어플리케이션



### 현재의 엘포인트 어플리케이션

- 2021.04.11을 기준으로 엘포인트 어플리케이션 개편
- 혜택, 쇼핑, 엘페이, 포인트, 금융 5개 탭이 존재
- App Store와 Play Store의 대부분의 리뷰를 확인해 보면, 쇼핑탭에 대한 언급보다 포인트, 룰렛, 엘박스 등 포인트 적립 관련기능들에 대한 관심도가 높음
- 개편된 엘포인트 어플리케이션의 쇼핑 기능의 사용 빈도가 낮으며, 쇼핑 기능의 추천 시스템이 부족함을 파악

● 엘포인트 어플리케이션의 쇼핑 기능 활성화를 위해 알맞은 추천 시스템을 설계하는 것을 목표로 설정

# 1. 문제 정의

#### 문제정의

EDA

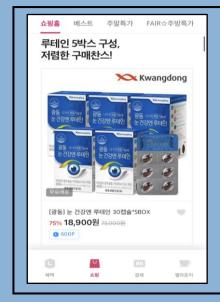
군집화

상품추천

마케팅

### 한계점과 개선 방안

#### 엘포인트 어플 쇼핑 탭의 한계점





- 매스마케팅으로 인한 엘포인트 어플 사용자의 쇼핑 기능 관심 부족 및 활성화 어려움
- '상품 구매'를 바탕으로 하는 Lpoint의 사용을 촉진하지 못하여 어플의 쇼핑 기능에서는 해당 이점 활용 불가
- Lpoint를 '적립'하는 방식에 비해 사용을 촉진하기 위한 방안 미비

### 개선 방안

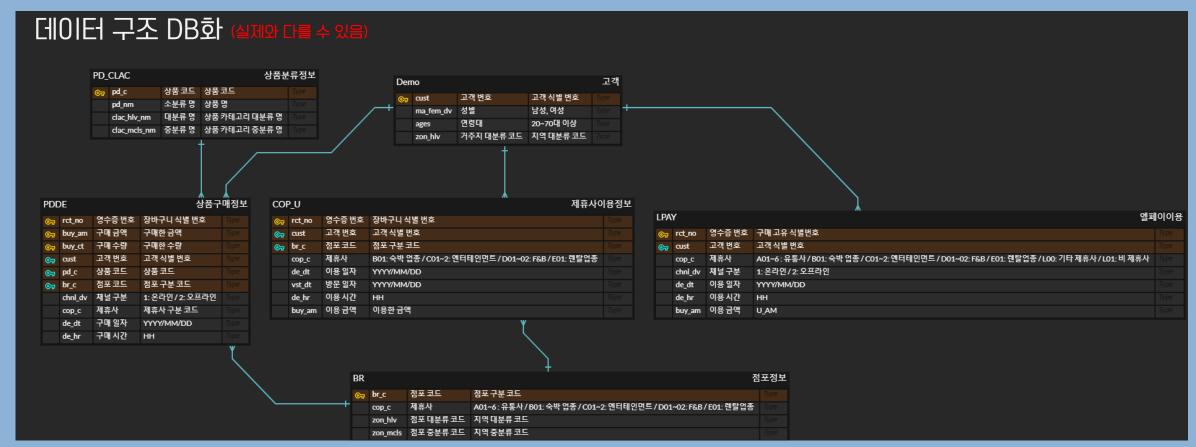
■ 엘포인트 어플에서의 <mark>구매 이력이 없는 고객 대상으로도 개인화 마케팅</mark>을 활용하여 상품 추천

#### 기대 효과

- 개인화 마케팅을 이용하여 각 고객의 관심도 높은 상품 추 천을 통해 엘포인트 쇼핑 탭에 대한 관심 향상 및 쇼핑 기 능 활성화
- 이로 인한 Lpoint 사용 활성화

# **EDA**

# 데이터 기본 구조 파악



- → 실제로는 모집단이 존재하는 데이터 셋에서 복원 추출된 일종의 Sample 데이터 셋
- → PDDE, COP\_U, LPAY의 경우 rct\_no 명명 방식이 다르므로 별도의 시스템에서 수집된 것으로 추측 가능

## 추천 방식 선정

방식	특징	한계점
협업 필터링	고객 <mark>선호도</mark> 에 따라 비슷한 고객을 찾고, 비슷한 선호도를 가진 <mark>고객의 기호를 반영</mark> 한 추천	주어진 데이터 셋에서 고객의 선호도를 명확 히 알 수 없음 (만족도, 리뷰, 평점 등)
연관 규칙 추천	구매 내역 중 장바구니 단위로, A를 구매했을 때 B를 구매할 신뢰도를 통해 연관성이 있는 품목을 찾아 추천	한 번의 구매 단위에 적용되므로 이 다음 구매 에 대한 개인화 추천에 적합하지 않음
군집 기반 추천	비슷한 성향을 갖는 고객을 군집으로 묶고, <mark>군집</mark> 특성에 따른 추천	고객 성향을 추출할 수 있는 특징을 선정, 많 은 군집의 경우 프로파일링이 어려움

- **협업 필터링**은 고객 구매 자체를 선호라고 정의할 수 없기에 **한계점이 존재**
- 연관 규칙은 장바구니 1건을 독립적인 집합으로 처리하기에 다음 구매에 대한 추천으로 적합하지 않음
- → 군집 기반 추천은 기준을 모델 설계 의도에 따라 정의하고, 비슷한 유형의 고객을 통한 추천 가능

### 사용할 데이터 선택

데이터 선택

- 1 상품을 추천해야 하므로 상품 구매 이력이 존재하는 PDDE (유통사 구매 내역) 데이터 선택
- ② 고객정보를 기준으로 선정해야 하므로 DEMO (고객정보) 데이터 선택
- ③ 상품 정보를 통해 상품 기준 또한 반영 되어야 하므로 PD\_CLAC (상품 정보) 데이터 선택



선행 프로세스

- 1 구매가 없는 고객은 군집 대상이 되지 않으므로 제외
- 2 '고객 정보-구매 내역' 을 통해 군집화에 적합한 고객 특징(feature) 선정
- 3 군집화 알고리즘 선택

## 군집화에 적합한 고객 기준 선정(1)

#### 컬럼 기반 Feature 선정

Column Name	Data Type	Column 사용 방식	Encoding 차원
ma_fem_dv		0과 1로 Label Encoding	1차원
ages		One Hot Encoding	6차원
chnl_dv		0과 1로 Label Encoding	1차원
zon_hlv		One Hot Encoding	17차원
zon_mcls	Categorical	One Hot Encoding	257차원
COP_C (유통사 기준)		One Hot Encoding	6차원
pd_c		One Hot Encoding	1933차원
br_c		One Hot Encoding	8808차원
clac_hlv_nm		One Hot Encoding	60차원
mcls_nm		One Hot Encoding	349차원
buy_ct	Numeric	상품 단위 가격	-
de_dt / de_hr	Numenc	주기 형성, 방문 빈도	-
buy_am	Continuous	종합 및 Scaling	_

### Categorical Feature Encoding

- 일반적으로 Categorical Data 처리를 위해 Label Encoding과 One Hot Encoding을 사용
- Label Encoding의 경우 binary하지 않은 데이터에서 처리 시, Label이 갖는 스케일이 있기에 독립된 형태로 Encoding 불가
- One Hot Encoding은 벡터 형태로 데이터를 분리하기 때문에 Column의 unique 값만큼의 차원을 형성하므로 100차원이 넘는 형태의 encoding은 데이터가 매우 희소하게 변환됨
- One Hot Encoding을 수행했더라도, Continuous 한 Feature와 함께 군집화를 이루기 위해 cosine similarity를 활용해야 하는데, cosine similarity는 방향성만 고려하므로 Continuous Feature의 스케일이 무시되는 문제발생

#### Numeric & Continuous Feature

- buy\_ct의 경우, 구매 시 수량은 금액과 연계되어 활용해야 하고, 그 자체만으로는 의미를 가질 수 없음
- de\_dt/de\_hr 은 Time Series로 주기 추출, 시계열 패턴 파악 등에 활용될수 있음
- buy\_am은 고객의 소비를 직접적으로 보여주는 Column으로, Continuous 의 속성을 그대로 활용할 수 있음
- → 즉, 모든 컬럼을 고려하여 군집 Feature를 선정하는 것은 **차원의 저주**에 빠질 수 있고, **스케일 불균형** 초래

## 군집화에 적합한 고객 기준 선정(2)

#### 도메인 기반 Feature 선정

가정 . Online을 한 번도 사용하지 않은 고객과 그렇지 않은 고객의 소비 양상이 다름

특징. 비슷한 고객이라고 생각할 수 있는 기준은 여러 기준이 있으나, 소비에 대한 부분으로 특징 지을 수 있음

● 따라서 군집을 나누는 Feature를 만들 때, 소비 성향을 나타낼 수 있는 형태로 데이터 셋 구성

#### 소비 성향 Feature 활용 계획

- 소비 성향을 이루는 특징에는 크게 소비 금액과 소비 품목으로 생각할 수 있음
- 마케팅 전략이 차별화와 채널에 따른 가정을 확인하여 **채널 정보** 반영
- 고객 별 전체 **소비 금액을 등급으로 변환**하여 새로운 단위로 분할
- 소비 품목까지 고려하여 군집을 만들고, 추천해야 하므로 **구매 품목에 대한 column**을 고려
- 온/오프라인으로 마케팅 채널을 분할하고, 소비 금액이 비슷한 고객으로 2차 분할하여 분할된 각 데이터 셋을 독립된 형태로 군집화 수행
- 정리하자면 고객 01용 채널과 소비 금액으로 데이터를 분할하고, 품목 별 소비에 따라 고객을 군집화

# 2. EDA

문제정의

**EDA** 

군집화

상품추천

마케팅

### 전체적인 모델 설계 흐름

◆ Preprocessing → Clustering → Recommend의 순서로 전체 모델 흐름을 설계

Preprocessing

Clustering

Recommend

Input Data 제작 (소비 성향)

이상치 제거 (DBSCAN)

offline / online 분할

소비 등급별 데이터 분할

output

데이터 분할

0 1







5





각 0 ~ 7 데이터 셋 군집 Input

**Pivoting** 

Scaling

**Kmeans Clustering** 

가장 가까운 이웃 3명 탐색

이웃 구매 이력 취합

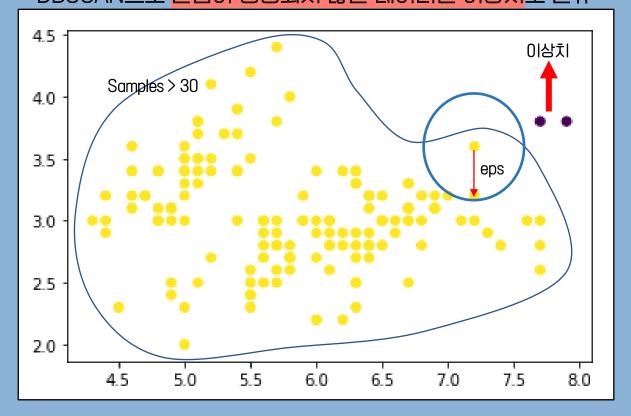
구매하지 않은 상위 5개 품목

# 전처리

## DBSCAN을 이용한 이상치 처리

- Kmeans Clustering의 효과적인 군집화를 위해 수행
- 따라서 index: cust, column: clac\_hlv\_nm, values: buy\_am로 이루 어진 Pivot Table 사용

#### DBSCAN으로 군집이 형성되지 않은 데이터는 이상치로 분류



#### ❖ DBSCAN

- element가 모여 있는 밀도를 기반으로 한 클러스터링 기법
- 주요 파라미터

eps	min_samples
element가 이루는 최소 거리	군집을 이루는 최소 element 수

- 어느 점을 기준으로 반경 eps 내에 점이 min\_samples 개 이상 있으면 하나의 군집으로 인식하는 방식
- 군집을 형성하지 못한 점은 이상치로 분류

#### ❖ 파라미터 설정

- min\_samples = 30(고정 값)으로 설정
- eps 는 탐색을 통해 값 설정
  - \* 순차 탐색을 통해 전체 데이터에서 이상치 5%를 제거하는 거리 값으로 설정
- 군집에서 가장 가까운 점과의 거리가 eps 값보다 큰 점은 군집이 형성
   성 불가하므로 이상치로 판단

# 3. 전처리

#### 문제정의

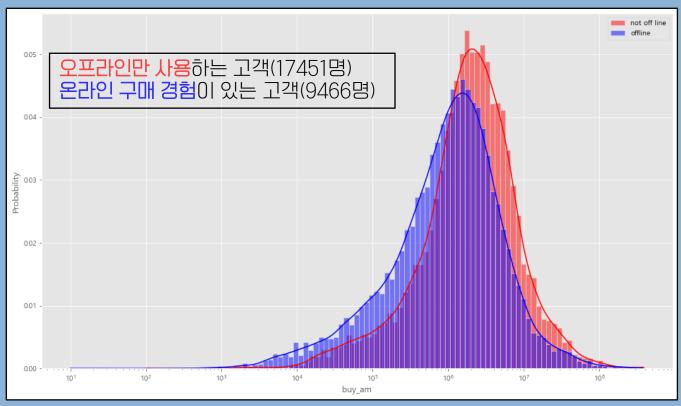
#### 군집화

· 상품추천

마케팅

## 데이터 셋 분할 • On-line / Off-line

#### 오프라인만 사용하는 고객과 온라인 구매경험이 있는 고객의 총 구매 금액 분포



#### 오프라인과 오프라인이 아닌 고객의 통계적 차이

**EDA** 

분포의 차이를 검정하는 KS-Test 수행

#### 1. 가설 설정

- 기무가설 : 두 표본분포는 같은 분포이다.
- 대립가설 : 두 표본분포는 같지 않은 분포이다.

#### 2. 통계량 및 P-value

- KS 통계량 = 0.188
- P-value = 1.121e-190

#### 3. 검정 결과

■ P-value가 매우 작은 값으로 1% 유의수준에서 귀무 가설 기각

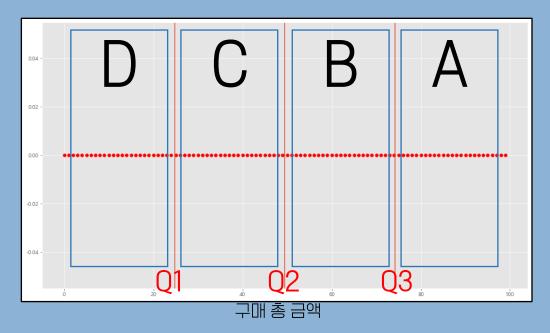
결론: 오프라인만 사용하는 고객과 온라인 구매 경험이 있는 고객의 총 구매 금액 분포는 같지 않음

#### 온라인과 오프라인의 마케팅 관점의 차이

- 시공간의 제약 유무의 차이가 존재
- 채널 별로 다른 마케팅 전략 필요

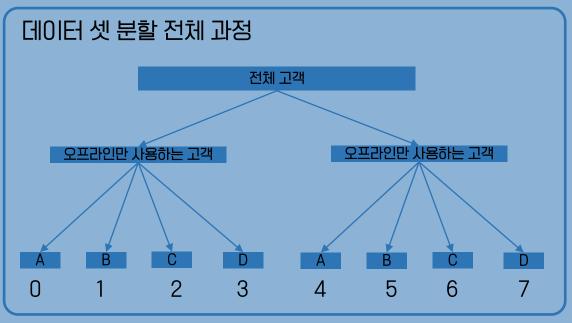
# 3. 전처리

## 데이터 셋 분할 🗗 소비 등급



- 1 오프라인만 사용하는 고객과 온라인 구매경험이 있는 고객으로 분할
- 2 구매 총액을 Q1, Q2, Q3를 기준으로 구매 등급 분류
- ③ quantile을 사용한 분할이기 때문에 각 등급에 속한 고객의 수는 같음





#### 데이터 분할 결과

데이터 셋	분할 데이터 셋 설명
0	오프라인만 사용하면서, 소비 등급이 A인 고객들
1	오프라인만 사용하면서, 소비 등급이 B인 고객들
2	오프라인만 사용하면서, 소비 등급이 C인 고객들
3	오프라인만 사용하면서, 소비 등급이 D인 고객들
4	온라인 구매 경험이 있고, 소비 등급이 A인 고객들
5	온라인 구매 경험이 있고, 소비 등급이 B인 고객들
6	온라인 구매 경험이 있고, 소비 등급이 C인 고객들
7	온라인 구매 경험이 있고, 소비 등급이 D인 고객들

# 군집화

# 4. 군집화

문제정의

EDA

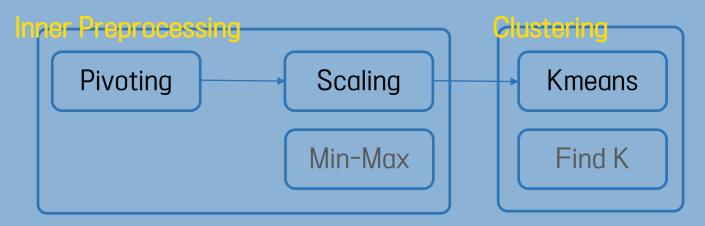
**군집화** 상품추천

마케팅

Kmeans: 상품 구매 기반 군집화

- ◆ K-means clustering
  - 주어진 데이터를 k개의 군집으로 묶는 알고리즘으로 각 군집 거리 차이의 분산을 최소화 하는 방식으로 군집 형성
  - Continuous한 Feature를 다루기 때문에 euclidean metric을 사용하는 K-means Clustering 수행

### ◆ 진행 과정

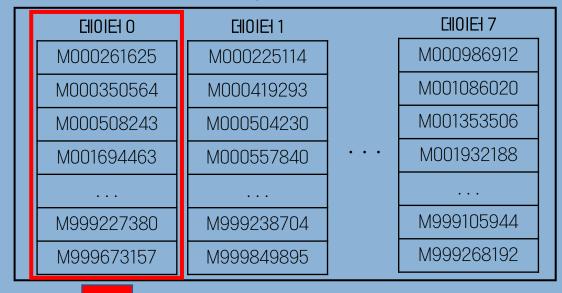


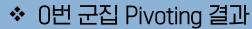
● 위의 과정을 전처리 단계에서 분리된 8개의 데이터 셋에 적용하여 독립적인 군집으로 분류

# **Inner Preprocessing**

- (1) Pivoting
- 전처리 과정에서 나눠진 총 8개의 데이터를 고객-상품구매금액(대분류) 기준으로 각각 Pivoting Table 형성
- 상품 대분류를 column으로 하여, 군집에 속한 모든 고객의 상품 별 총 구매 금액을 확인

#### 〈전처리 후 8개의 데이터 셋〉





CUST	남성의류	주방잡화	테넌트/음식점	속옷/양말/홈웨어	구기/필드스포츠		기타(비상품)	담배
M000261625	0	0	329,600	207,000	0		0	0
M000350564	317,300	0	10,000	0	1,120,000		0	0
M000508243	555,000	0	0	11,700	0	•••	0	0
M999673157	839,600	0	6,500	0	0		7,050	110,000

# 4. 군집화

문제정의

EDA

군집화

상품추천 마케팅

# **Inner Preprocessing**

### (2) Min-Max Scaling

- 품목 별 기본 금액 차이 해소를 위해 Pivoting 결과 테이블별로 Min-Max scaling 진행
- 각 품목의 최대 금액과 최소 금액을 이용하여 품목 별 값을 0에서 1로 scaling
- 0~7번의 8개 데이터 세트에 대해 독립적인 Scaling 적용

.fit()
.transform()

.fit()
.transform()

.fit() .transform()

.fit() .transform()

.fit()
.transform()

.fit()
.transform()

fit()
.transform()

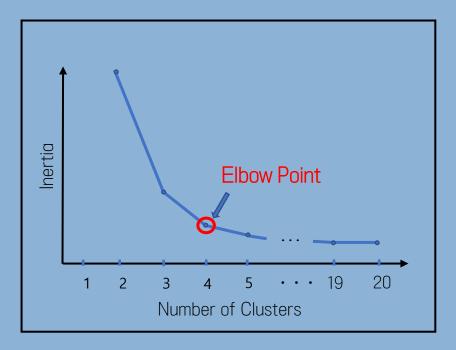
.fit()
.transform()

## ❖ 0번 군집 Scaling 결과

CL	JST	남성의류	주방잡화	테넌트/음식점	속옷/양말/홈웨어	구기/필드스포츠	:	기타(비상품)	담배
M0002	261625	0	0	0.142228	0.094998	0		0	0
M0003	350564	0.013025	0	0.004315	0	0.166024		0	0
M0005	508243	0.022783	0	0	0.005369	0		0	0
M9996	673157	0.034466	0	0.002805	0	0		0.001494	0.057803

# **Kmeans Clustering**

- (3) 최적 군집 개수 선정 (Find K): Elbow Method를 이용해 최적 군집 개수 선정
- ❖ Elbow Method: 군집을 늘려가면서 군집 별 Inertia(군집 내 분산)의 감소율이 크게 작아지는 지점을 최적 군집 개수로 설정하는 방법



• 그래프에서 Elbow Point(크게 꺾이는 지점)에서의 k 값을 최적 군집 개수로 선정

군집 개수 후보 범위: 2-20

 ◆ 일반적으로 Inertia 그래프의 꺾이는 부분을 시각적으로 확인하여 주관적 판단이 개입되는 단점 존재

모든 군집 수 k에 대해 Inertia 감소율의 변화 폭이 가장 큰 k를 최적 군집으로 선정하도록 함수화 하여 객관적인 기준을 세움

# 4. 군집화

문제정의

**EDA** 

군집화

상품추천

마케팅

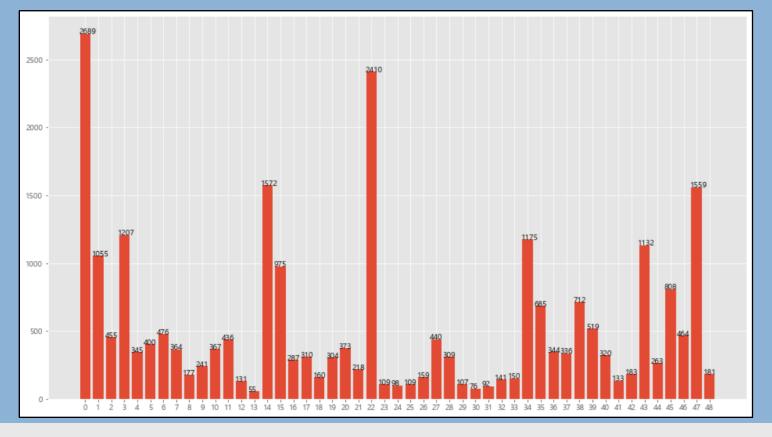
# **Kmeans Clustering**

❖ 군집 내 가장 가까운 3명의 구매 내역으로 추천 해주는 방식 사용했기 때문에 한 군집 내 최소 고객은 3명 이상으로 설정 필요



- 3명 이하의 고객이 포함된 군집 내 고객은 이상치로 판단
- 4명 이상이 포함된 군집만으로 군집화 진행

(4) 군집화 결과: 총 49개의 군집으로 군집화



분할 데이터 No	군집 수
0	3
1	11
2	8
3	12
4	3
5	6
6	3
7	3
총	49

# 상품 추천

문제정의

EDA

군집화

상품추천

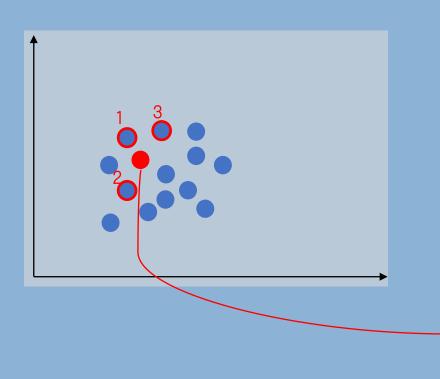
마케팅

# 추천 로직 도식화

(1) 군집 내 가장 가까운 이웃 3명 찾기

(2) 이웃의 구매 이력 가져오기

(3) 구매하지 않은 품목 중 상위 5개 추천



고객	구매 품목
1	[a,b,c]
2	[a,d,e,g]
3	[b,c,d,f,h]

Target	[a,f,g]
--------	---------

후보 품목	구매 수량	Target
<del>- a</del>	3	0
b	2	Χ
С	2	Χ
d	2	X
е	1	Χ
<del>f</del>	1	0
<del>- g</del>	1	0
h	1	Χ

→ 군집 수(49)만큼 반복

Recommend [b, c, d, e, h]

문제정의

**EDA** 

군집화

상품추천

마케팅

## 군집 내 가장 가까운 이웃 3명 찾기

구매 내역 Pivoting

고객-구매 총 금액으로 pivoting 하여 input data 형성

Target 군집 필터링

Target 군집에 대해 전체 데이터에서 고객 필터링

군집 Min-Max Scaling

독립적인 군집으로 인식하여 Min-Max Scaling 수행

군집 내 고객 간 거리 측정

scipy.spatial.distance.cdist() 를 활용하여 고객 간 cosine 유사도 측정

가장 가까운 3명 이웃 추출

argsort() 를 통해 가장 가까운 3명의 이웃 인덱스를 추출하고, 고객ID (cust)로 변환하여 3명의 이웃 도출

문제정의

EDA

군집화 상품추천

마케팅

## 군집 내 이웃 3명 기반 상품 추천하기

이웃 구매 이력

이웃 3명 구매한 상품 수집

이웃 3명이 구매한 상품을 value\_counts()를 통해 구매 개수 별 정리

구매하지 않은 품목 중 상위 5개 추천 Target 군집 필터링

set()을 이용해 이웃이 3명이 구매한 상품에서 고객이 구매한 상품 제외

5가지 상품 추천

인덱싱을 통해 이웃 고객이 많이 구매한 상품 5가지 품목 추천

예외 처리

5가지 상품을 추천하지 못하는 경우 try-except 문을 사용해 예외처리

위 과정을 군집 수(49) 만큼 반복

최종 결과

결과 생성

군집 별 추천 상품 데이터프레임을 결합 후 .csv 형식으로 내보내기

문제정의

EDA

군집화

상품추천

마케팅

# 결과 예시

고객	추천1	추천2	추천3	추천4	이웃 고객의 <sup>추천5</sup> 부족으로 N
M002867247	남성스포츠점퍼/재킷	남성등산점퍼/재킷	NaN	NaN	NaN
M009682567	기타남성의류세트	남성스포츠스웨트셔츠/후드/집업 -	여성속못세트	일반미	냉장핫도그/핫바류
M017524395	여성일반지갑	야구모자	캐쥬얼크로스백	여성등산EI셔츠/탑	남성등산바지
M025168868	기타생활서비스	국물용기라면	감자스낵	일반빵	J+OIC+
M035670981	목수수스낵	봉투보증금	일반스낵	콜라	남아의류세트
M977028904	즉석죽	기타남성의류세트	베이커리	영화/문화모바일상품권	샐러드류
M983457029	기타남성의류세트	스파/워터파크이용권	남성남방셔츠	남성골프티셔츠	남성코트
M995885121	남성골프티셔츠	식빵	국산맥주	일반소주	일반계란
M998129365	유아동런닝/트레이닝화	기타등산용품	남성팬티	성인이불/이불커버	스케이트보드/킥보드
M998600186	남성스니커즈	여성플랫	남성스포츠스웨트셔츠/후드/집업	여성샌들	실내액세서리편의용품

- 고객별 군집에서 이웃 고객 3명의 구매 이력으로 5가지 상품 추천
- M983457029 고객에게 기타남성의류세트, 스파/워터파크이용권, 남성남방셔츠, 남성골프티셔츠, 남성코트 추천←

# 마케팅 전략

# 6. 마케팅 전략

문제정의

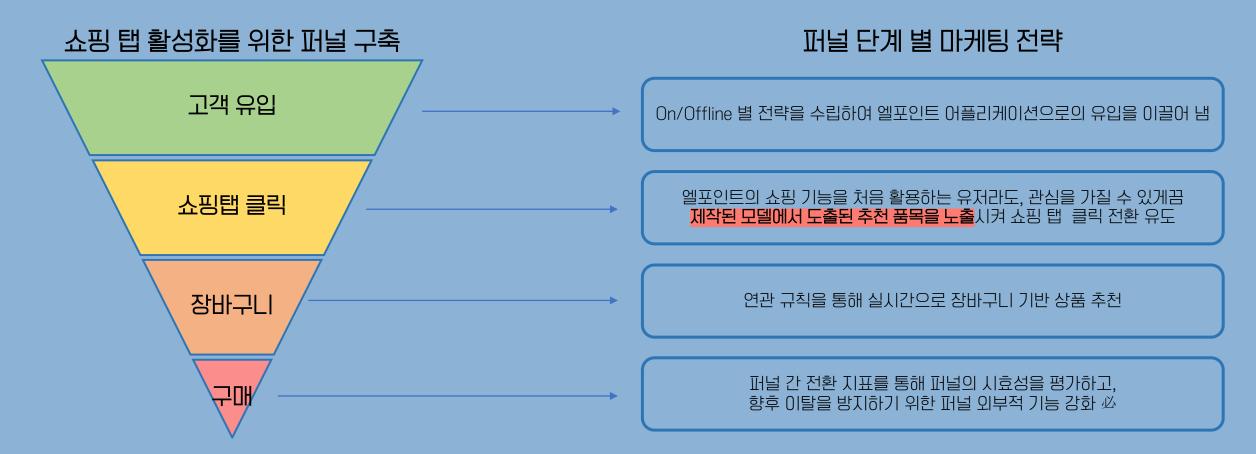
EDA

군집화 상품추천

마케팅

### 퍼널을 통한 마케팅 전략 제시

- → 어플리케이션에서 이루어지는 마케팅은 단계 별 전환을 이루는 것에 목표를 둠
- 따라서 퍼널을 구축하고, 퍼널 간 전환을 이루어 내는 과정을 통해 전환 지표를 측정하는 그로스 마케팅 전략 수립
- 위에서 구축된 군집화 결과는 쇼핑탭 클릭 퍼널에 적용하여 쇼핑 탭 관심도 향상에 기여



# 6. 마케팅 전략

문제정의

EDA

군집화 상품추천

마케팅

### 고객 유입 마케팅 예시

On/Offline 별 전략을 수립하여 엘포인트 어플리케이션으로의 유입을 이끌어 냄

#### 오프라인만 사용하는 고객 유입 전략 예

- 1) 오프라인 결제 시 어플 홍보
- 2) 어플 설치 OR코드 사용으로 어플 사용 유도

### 온라인 구매 경험이 있는 고객 유입 전략 예

- 1) 푸시 알림으로 방문 유도
- 2) SNS 광고를 통한 어플 사용 유도
- 3) 영상 광고를 통한 어플 사용 유도

● 엘포인트 社의 실현 가능한 다양한 전략을 통해 고객 유입 전략 수립 고객 유입 퍼널 단계 구축

# 감사합니다