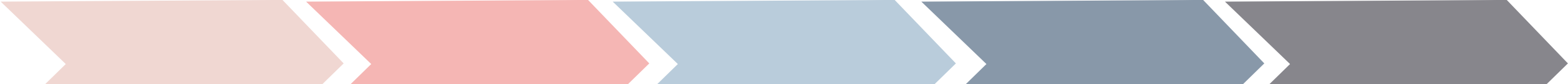


PDC를 사용한 군집 분석 및 추천 방법

2020.06.20

빅데이터경영통계전공

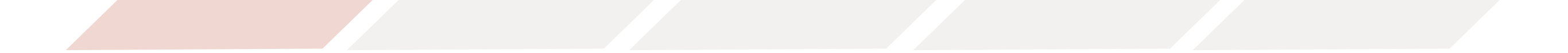
20192784 윤경서



CONTENTS



Data Cleansing	Features	고객 세분화 기준	고객세분화	상품추천
이상치 처리		구매금액에 따른 기준(P)	Customer Segment	clac_nm3 항목정리
잘못된 값		구매일에 따른 기준(D)	Segment별 시각화	segment별 성별
결측치 처리		구매건수에 따른 기준(C)		주요구매상품
				segment별 상품추천



Data Cleansing

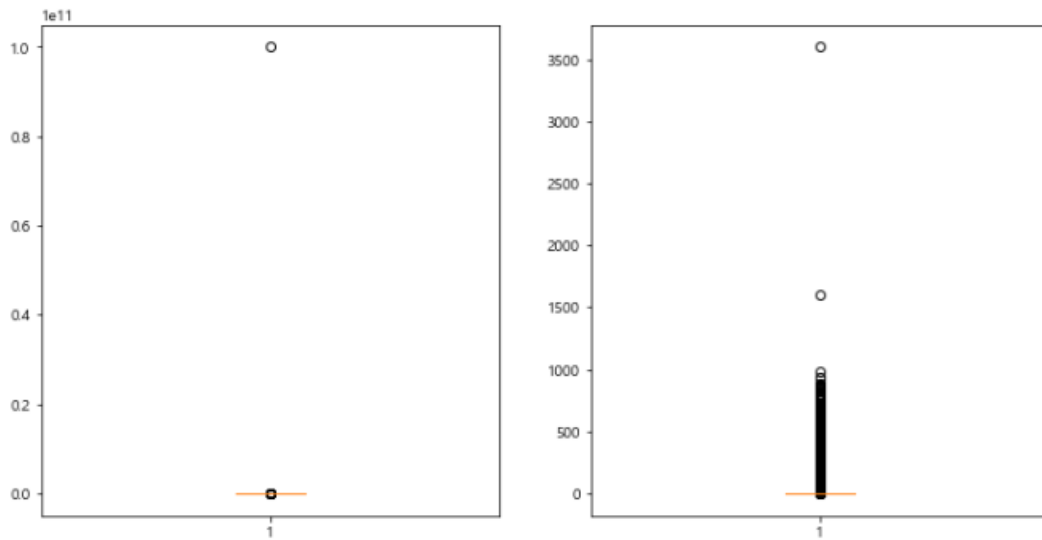
1

이상치 처리

잘못된 값

결측치 처리

1. 이상치 처리



구매금액과 구매건수에 대한 이상치를 확인하기 위해 boxplot을 그려보았다. 그 결과 구매금액에서는 1개, 구매수량에서는 2개의 이상치를 발견했고, 값을 바꾸는 것보다 제거하는 것이 더 정확한 분석을 하도록 도울 것 같아 제거했다.

2. 구매금액이나 구매건수 항목이 0인 경우

구매금액 = 0, 구매수량 = 0

구매금액과 구매건수 모두 0이기 때문에 두 항목 모두 임의대로 설정할 수 없어 제거하였다.

구매금액 = 0, 구매수량 \neq 0

상품이 같아도 고객별로 구매한 금액이 달라서 상품의 평균구매금액으로 구매금액을 채웠다. 해당상품의 구매금액이 모두 0이어서 평균 구매금액이 0인 상품은 제거했다.

구매금액 \neq 0, 구매수량 = 0

상품별 평균구매금액과 고객의 구매금액을 비교했다. 고객의 구매금액 \leq 평균구매금액 * i 라는 식을 사용해 i 를 1~35까지 값을 넣었고 식이 True일 때 i 값을 구매수량으로 채워넣었다.

3. 결측치 처리

df의 결측치를 확인해보니 clac_nm3열에 26개의 결측치가 있는 것을 알 수 있었다.

```
# 결측값인 clac_nm3의 상품번호
df.loc[df.isnull()['clac_nm3'], :]['pd_c'].unique()

array(['0196', '0524'], dtype=object)

# 결측값인 clac_nm3의 clac_nm2
c0196 = list(df.query('pd_c == "0196"')['clac_nm2'].unique())
c0524 = list(df.query('pd_c == "0524"')['clac_nm2'].unique())

c0196, c0524

(['Packaged Side Dishes'], ['Frozen Instant Foods'])

df_0196 = df.query('pd_c == "0196"').fillna('Packaged Food A')
df_0524 = df.query('pd_c == "0524"').fillna('Frozen Food A')
```

결측값인 clac_nm3의 상품번호를 뽑아 '0196', '0524'인 것을 확인했다. 결측값인 clac_nm3 값을 채워주기 위해 중분류 항목을 뽑아 확인하였다. 해당 상품별 clac_nm2는 다음과 같았다.

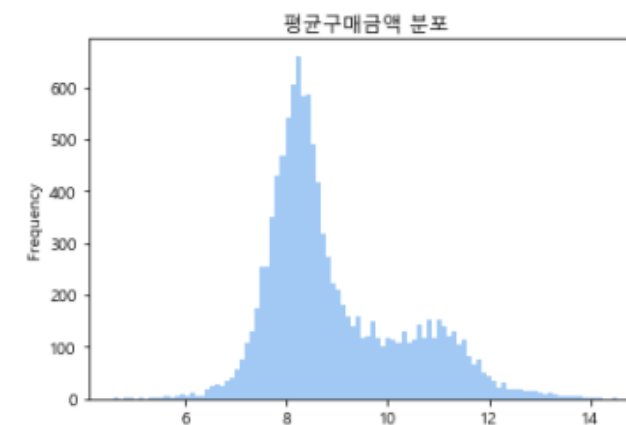
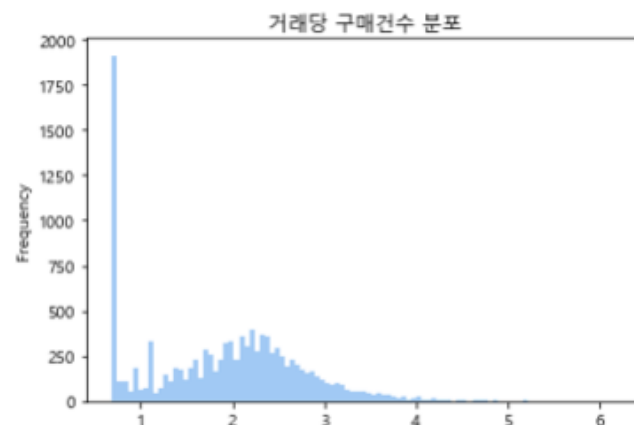
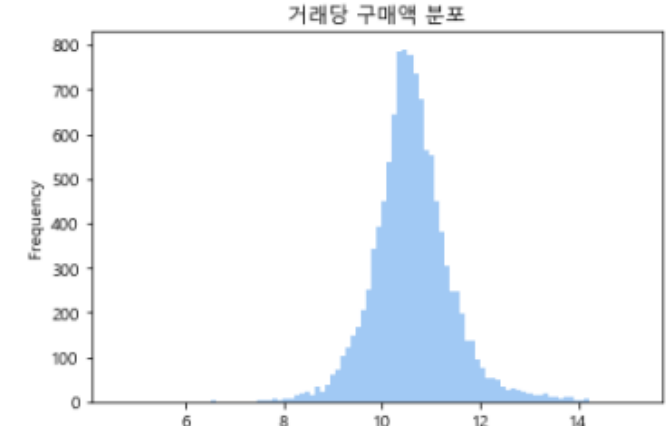
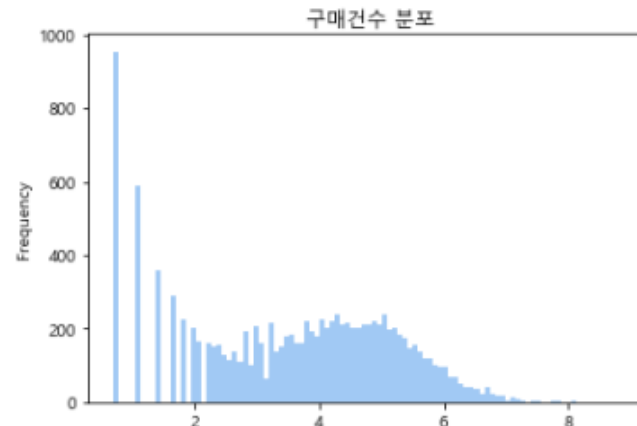
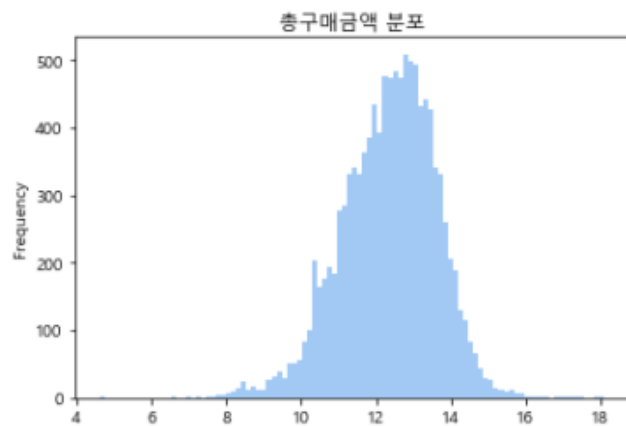
상품번호 0196 : clac_nm2 = Packaged Side Dishes
상품번호 0524 : clac_nm2 = Frozen Instant Foods
clac_nm2에 따라 clac_nm3의 결측값을
상품 0196은 Packaged Food A,
상품 0524는 Frozen Food A로 결측값 처리했다.



Features 2

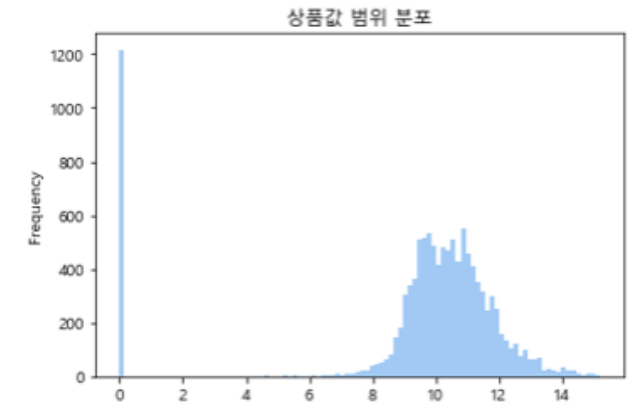
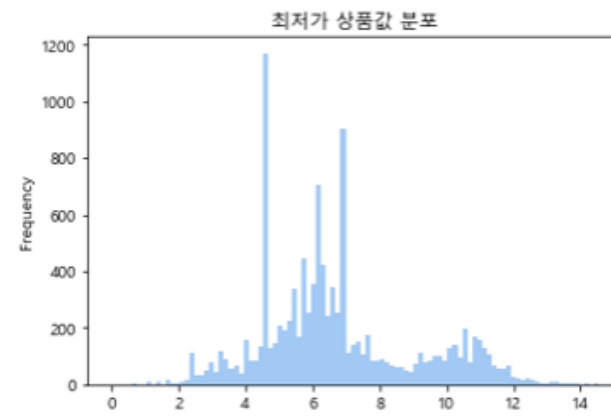
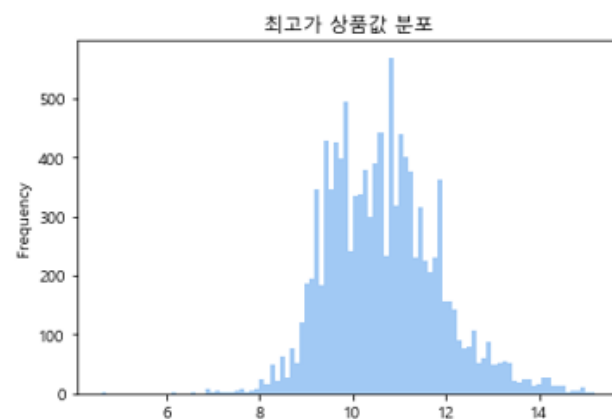
Features

- ▶ **총구매금액** : clnt_id별 buy_am의 합을 구해 고객당 총구매금액을 보기 위한 열이다.
- ▶ **구매건수** : clnt_id별 buy_ct의 합을 구해 고객당 총구매건수를 보기 위한 열이다.
- ▶ **거래당구매액** : clnt_id별 (buy_am의 합 / trans_id의 갯수)를 구해 고객별로 한번 거래할 때의 평균 구매액을 보기 위한 열이다.
- ▶ **거래당구매건수** : clnt_id별 (buy_ct의 합 / trans_id의 갯수)를 구해 고객별로 한번 거래할 때의 평균 구매건수를 보기 위한 열이다.
- ▶ **평균구매금액** : (총구매금액 // 구매건수)를 구해 고객별로 평균 구매금액을 보기 위한 열이다.

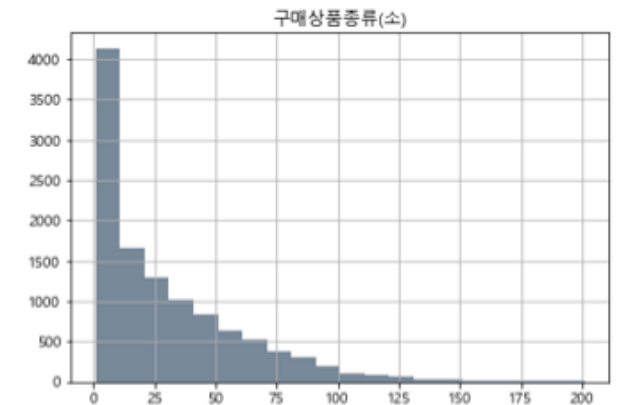
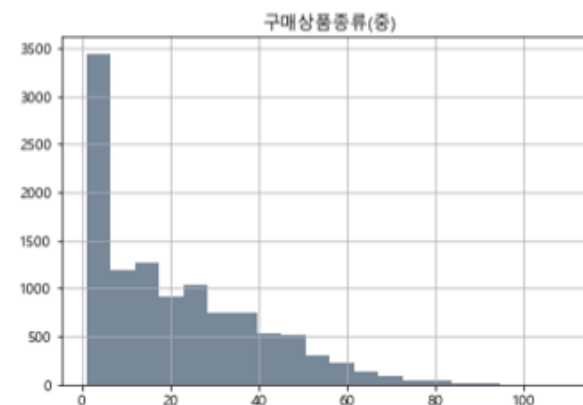
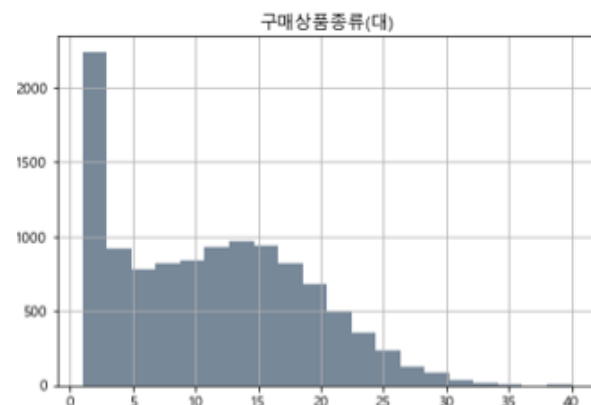


Features

- ▶ **최고가상품값** : 거래마다 구매한 상품을 개당 얼마에 샀는지 ($\text{buy_am} / \text{buy_ct}$)를 통해 구한 후, 최댓값을 뽑아 고객별로 가장 비싸게 주고 산 상품이 얼마였는지 보기 위한 열이다.
- ▶ **최저가상품값** : 거래마다 구매한 상품을 개당 얼마에 샀는지 ($\text{buy_am} / \text{buy_ct}$)를 통해 구한 후, 최솟값을 뽑아 고객별로 가장 싸게 주고 산 상품이 얼마였는지 보기 위한 열이다.
- ▶ **상품값범위** : (최고가상품값 - 최저가상품값)를 구해 고객별로 지출한 금액의 범위를 보기 위한 열이다.

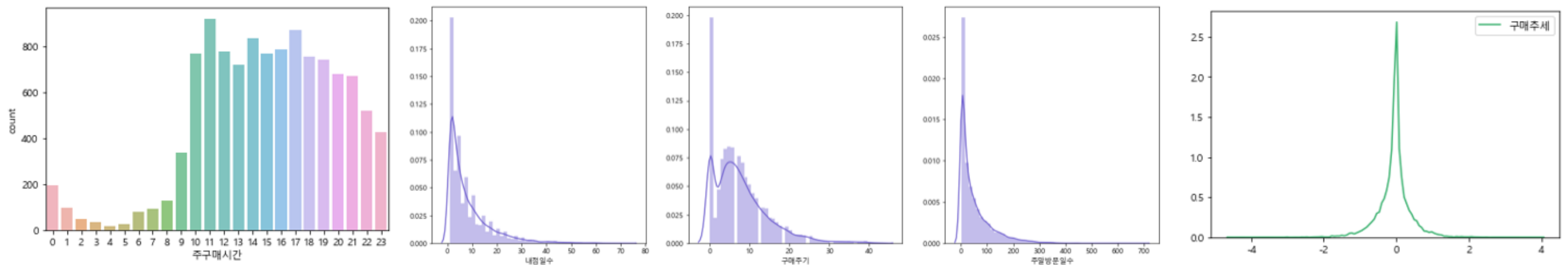


- ▶ **구매상품종류(대)** : clac_nm1의 고유한 값 갯수를 구해 상품 대분류 갯수를 보기 위한 열이다.
- ▶ **구매상품종류(중)** : clac_nm2의 고유한 값 갯수를 구해 상품 중분류 갯수를 보기 위한 열이다.
- ▶ **구매상품종류(소)** : clac_nm3의 고유한 값 갯수를 구해 상품 소분류 갯수를 보기 위한 열이다.



Features

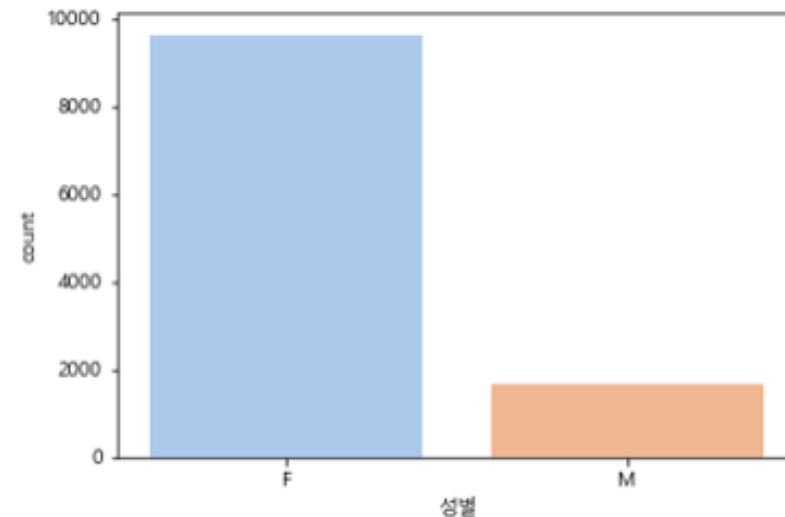
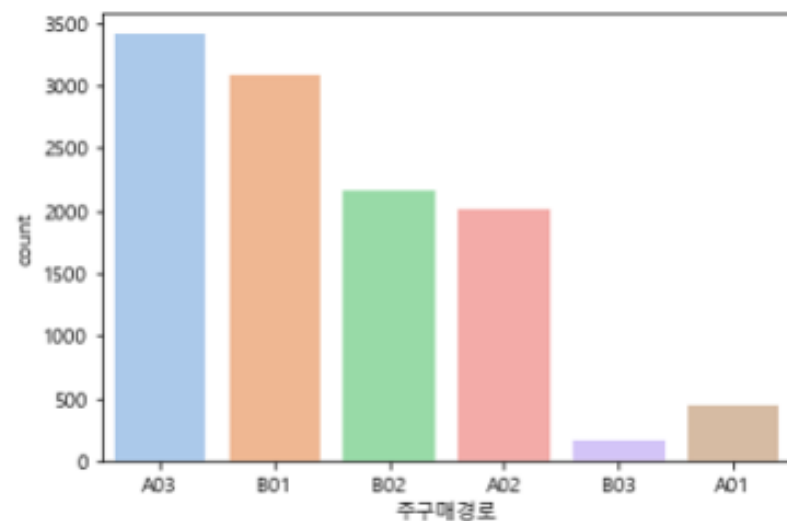
- ▶ **주구매시간** : 고객별 de_tm에서 시간만 뽑고, 가장 빈도가 많은 시간을 뽑아내어 고객별 주요 구매시간을 보기 위한 열이다.
- ▶ **내점일수** : 고객별 de_dt의 고유한 값 갯수를 세어 고객별 내점 일수가 얼마나 되는지 보기 위한 열이다.
- ▶ **구매주기** : 고객별 de_dt에서 (가장 오래된 날짜 - 가장 최근 날짜) / 날짜의 갯수를 구해 고객별 구매주기를 보기 위한 열이다.
- ▶ **주말방문일수** : 고객별로 de_dt의 weekday가 4보다 클 때를 세어 고객별 주말방문 일수를 보기 위한 열이다.
- ▶ **구매추세** : 마지막 날짜부터 91전까지 (7일씩 13번) 매주 구매건수를 계산하여 고객별 구매 추세를 보기 위한 열이다.



Features

앞에서 (buy_am / buy_ct)를 구한 열의 사분위수를 구한 후, 분위수를 기준으로 고객별로 가격대별 구매한 상품의 갯수를 센다. 그 열들을 고객이 구매한 상품의 총갯수로 각각 나누었다.

- ▶ **고가상품구매율** : 고객별로 제3사분위수 보다 큰 가격의 상품 구매율을 보기 위한 열이다.
- ▶ **중저가상품구매율** : 고객별로 제2사분위수 보다 큰 가격의 상품 구매율을 보기 위한 열이다.
- ▶ **저가상품구매율** : 고객별로 제1사분위수 보다 큰 가격의 상품 구매율을 보기 위한 열이다.
- ▶ **최저가상품구매율** : 고객별로 제1사분위수 이하 가격의 상품 구매율을 보기 위한 열이다.
- ▶ **주구매경로** : 고객별로 가장 빈도가 많은 biz_unit을 뽑아내어 전체적인 주된 구매경로를 보기 위한 열이다.
- ▶ **성별** : 고객별로 성별을 보고 전체적인 성비를 보기 위한 열이다.



고가 상품 구매 율	중 저 가 상 품 구 매 율	저 가 상 품 구 매 율	최 저 가 상 품 구 매 율
0.00	0.59	0.20	0.21
0.06	0.38	0.25	0.31
0.00	0.44	0.24	0.32
0.00	0.21	0.28	0.51
0.89	0.05	0.04	0.02
...
1.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00

고객 세분화 기준

3

구매금액에 따른 기준(P)

구매일에 따른 기준(D)

구매건수에 따른 기준(C)

고객 세분화 기준 설정

RFM처럼 항목을 점수화시켜 고객세분화를 했다.

이때 기준을 RFM으로 안하고 새로운 기준을 만들었다.

M(구매금액)은 고객이 구입한 총 구매금액을 나타내기 때문에 저가상품을 대량구매한 고객과 고가상품을 소량구매한 고객을 구분할 수 없다고 생각하여 금액에 따른 기준 P(Price)를 새로 만들었다. R(최근성)과 F(구매빈도)는 고객이 방문한 일수를 가지고 계산한다는 공통성이 있다고 생각되어 방문 일수에 따른 기준 D(Day)를 새로 만들었다.

RFM에서는 고객이 대량구매를 주로 하는 고객인지 소량구매를 주로 하는 고객인지 판단할 수 없다고 생각하여 구매건수에 따른 기준 C(Count)를 새로 만들었다.

구매금액에 따른 기준(P)

구매일에 따른 기준(D)

구매건수에 따른 기준(C)

1. 가격대별 구매율(고가, 중저가, 저가, 최저가)에 따라 점수 부여

- 높은 가격대의 구매율이 높을수록 비싼 제품을 많이 구매하는 것이므로 가격대가 높을수록 높은 점수 부여(1~4점)

2. 상품값 범위를 가격대별 구매율 나누는 기준으로 사용했던 4분위수를 기준으로 나눠 그에 따라 점수 부여

- 상품값의 범위가 넓을수록 구매하는 상품의 범위가 넓은 것이므로 상품값범위가 클수록 높은 점수 부여(1~4점)

고객 세분화 기준 설정

RFM처럼 항목을 점수화시켜 고객세분화를 했다.

이때 기준을 RFM으로 안하고 새로운 기준을 만들었다.

M(구매금액)은 고객이 구입한 총 구매금액을 나타내기 때문에 저가상품을 대량구매한 고객과 고가상품을 소량구매한 고객을 구분할 수 없다고 생각하여 금액에 따른 기준 P(Price)를 새로 만들었다. R(최근성)과 F(구매빈도)는 고객이 방문한 일수를 가지고 계산한다는 공통성이 있다고 생각되어 방문 일수에 따른 기준 D(Day)를 새로 만들었다.

RFM에서는 고객이 대량구매를 주로 하는 고객인지 소량구매를 주로 하는 고객인지 판단할 수 없다고 생각하여 구매건수에 따른 기준 C(Count)를 새로 만들었다.

구매금액에 따른 기준(P)

구매일에 따른 기준(D)

구매건수에 따른 기준(C)

1. 구매주기에 따라 점수 부여

- 구매주기가 짧을수록 자주 구매하는 고객이기 때문에 점수당 범위를(최대 - 최소)/4로 잡은 후 점수 부여(1~4점)

2. 구매추세에 따라 점수 부여

- 구매추세가 증가할수록 최근에 구매한 고객이기 때문에 0을 중간 기준점으로 보고 4개의 범위로 나눈 후 점수 부여

고객 세분화 기준 설정

RFM처럼 항목을 점수화시켜 고객세분화를 했다.

이때 기준을 RFM으로 안하고 새로운 기준을 만들었다.

M(구매금액)은 고객이 구입한 총 구매금액을 나타내기 때문에 저가상품을 대량구매한 고객과 고가상품을 소량구매한 고객을 구분할 수 없다고 생각하여 금액에 따른 기준 P(Price)를 새로 만들었다. R(최근성)과 F(구매빈도)는 고객이 방문한 일수를 가지고 계산한다는 공통성이 있다고 생각되어 방문 일수에 따른 기준 D(Day)를 새로 만들었다.

RFM에서는 고객이 대량구매를 주로 하는 고객인지 소량구매를 주로 하는 고객인지 판단할 수 없다고 생각하여 구매건수에 따른 기준 C(Count)를 새로 만들었다.

구매금액에 따른 기준(P)

구매일에 따른 기준(D)

구매건수에 따른 기준(C)

1. 거래당 구매건수에 따라 점수 부여

- 거래당 구매건수가 많은 고객일수록 한번 거래할 때의 구매가격이 커지므로 거래당 구매건수에서 4분위수를 계산한 후, 4분위수값을 기준으로 점수 부여(1~4점)

2. 구매건수에 따라 점수 부여

- 구매건수가 많은 고객일수록 우리 상품을 많이 구매한다는 것이므로 구매건수에서 4분위수를 계산한 후, 4분위수 값을 기준으로 점수 부여(1~4점)

구매금액에 따른 기준(P)

- 가격대별 구매율에 따른 점수

- ▶ 4점 : 고가상품구매율 > 0.5
- ▶ 3점 : 중저가상품구매율 > 0.5
고가상품구매율 + 중저가상품구매율 > 0.5
- ▶ 2점 : 저가상품구매율 > 0.5
저가상품구매율 + 중저가상품구매율 > 0.5
- ▶ 1점 : 최저가상품구매율 > 0.5
저가상품구매율 + 최저가상품구매율 > 0.5

- 상품값 범위에 따른 점수

- ▶ 4점 : 상품값범위 ≥ 25727
- ▶ 3점 : $8955 \leq$ 상품값범위 < 25727
- ▶ 2점 : $4351 \leq$ 상품값범위 < 8955
- ▶ 1점 : 상품값범위 < 4351

구매금액에 따른 기준(P)

- 가격대별 구매율에 따른 점수

- ▶ 4점 : 고가상품구매율 > 0.5
- ▶ 3점 : 중저가상품구매율 > 0.5
고가상품구매율 + 중저가상품구매율 > 0.5
- ▶ 2점 : 저가상품구매율 > 0.5
저가상품구매율 + 중저가상품구매율 > 0.5
- ▶ 1점 : 최저가상품구매율 > 0.5
저가상품구매율 + 최저가상품구매율 > 0.5

- 상품값 범위에 따른 점수

- ▶ 4점 : 상품값범위 >= 25727
- ▶ 3점 : 8955 <= 상품값범위 < 25727
- ▶ 2점 : 4351 <= 상품값범위 < 8955
- ▶ 1점 : 상품값범위 < 4351

- 최종 P점수 -

- ▶ 4점 : 가격대별 구매율 = 4
가격대별 구매율 = 3, 상품값 범위 = 1 or 2
- ▶ 3점 : 가격대별 구매율 = 3, 상품값 범위 = 3 or 4
가격대별 구매율 = 2, 상품값 범위 = 3 or 4
- ▶ 2점 : 가격대별 구매율 = 2, 상품값 범위 = 1 or 2
가격대별 구매율 = 1, 상품값 범위 = 3 or 4
- ▶ 1점 : 가격대별 구매율 = 1, 상품값 범위 = 1 or 2

구매일에 따른 기준(D)

- 구매주기에 따른 점수

- ▶ 4점 : $1 \leq \text{구매주기} \leq 10$
- ▶ 3점 : $11 \leq \text{구매주기} \leq 20$
- ▶ 2점 : $21 \leq \text{구매주기} \leq 30$
- ▶ 1점 : $31 \leq \text{구매주기} \leq 43$ or $\text{구매주기} = 0$

※ 구매주기의 최소값 = 0, 구매주기의 최대값 = 43

- 구매추세에 따른 점수

- ▶ 4점 : $2.00 \leq \text{구매추세} < 4.00$
- ▶ 3점 : $0.00 \leq \text{구매추세} < 2.00$
- ▶ 2점 : $-2.25 \leq \text{구매추세} < 0.00$
- ▶ 1점 : $-4.55 \leq \text{구매추세} < -2.25$

※ 구매추세의 최소값 = -4.55, 구매주기의 최대값 = 3.96

구매일에 따른 기준(D)

- 구매주기에 따른 점수

- ▶ 4점 : $1 \leq \text{구매주기} \leq 10$
- ▶ 3점 : $11 \leq \text{구매주기} \leq 20$
- ▶ 2점 : $21 \leq \text{구매주기} \leq 30$
- ▶ 1점 : $31 \leq \text{구매주기} \leq 43$ or

※ 구매주기의 최소값 = 0, 구매주

- 구매추세에 따른 점수

- ▶ 4점 : $2.00 \leq \text{구매추세} < 4.00$
- ▶ 3점 : $0.00 \leq \text{구매추세} < 2.00$
- ▶ 2점 : $-2.25 \leq \text{구매추세} < 0.00$
- ▶ 1점 : $-4.55 \leq \text{구매추세} < -2.25$

※ 구매추세의 최소값 = -4.55, 구매주기의 최대값

- 최종 D 점수 -

- ▶ 4점 : 구매주기 = 4, 구매추세 = 4 or 3
구매주기 = 3, 구매추세 = 4
- ▶ 3점 : 구매주기 = 4, 구매추세 = 2
구매주기 = 3, 구매추세 = 2 or 3
구매주기 = 2, 구매추세 = 4
- ▶ 2점 : 구매주기 = 4, 구매추세 = 1
구매주기 = 3, 구매추세 = 1
구매주기 = 2, 구매추세 = 2 or 3
구매주기 = 1, 구매추세 = 4
- ▶ 1점 : 구매주기 = 2, 구매추세 = 1
구매주기 = 1, 구매추세 = 1 or 2 or 3

구매건수에 따른 기준(C)

- 거래당 구매건수에 따른 점수

- ▶ 4점 : 거래당구매건수 > 11
- ▶ 3점 : $6.4 < \text{거래당구매건수} \leq 11$
- ▶ 2점 : $2.1 < \text{거래당구매건수} \leq 6.4$
- ▶ 1점 : 거래당구매건수 ≤ 2.1

※ 거래당구매건수의 사분위수

제1사분위수 = 2.1, 제2사분위수 = 6.4, 제3사분위수 = 11

- 구매건수에 따른 점수

- ▶ 4점 : 구매건수 ≥ 127
- ▶ 3점 : $41 \leq \text{구매건수} < 127$
- ▶ 2점 : $8 \leq \text{구매건수} < 41$
- ▶ 1점 : 구매건수 < 8

※ 구매건수의 사분위수

제1사분위수 = 8, 제2사분위수 = 41, 제3사분위수 = 127

구매건수에 따른 기준(C)

- 거래당 구매건수에 따른 점수

- ▶ 4점 : 거래당구매건수 > 11
- ▶ 3점 : $6.4 < \text{거래당구매건수} \leq 11$
- ▶ 2점 : $2.1 < \text{거래당구매건수} \leq 6.4$
- ▶ 1점 : $\text{거래당구매건수} \leq 2.1$

※ 거래당구매건수의 사분위수

제1사분위수 = 2.1, 제2사분위수 =

- 구매건수에 따른 점수

- ▶ 4점 : 구매건수 ≥ 127
- ▶ 3점 : $41 \leq \text{구매건수} < 127$
- ▶ 2점 : $8 \leq \text{구매건수} < 41$
- ▶ 1점 : $\text{구매건수} < 8$

※ 구매건수의 사분위수

제1사분위수 = 8, 제2사분위수 = 41, 제3사분위수 = 127

- 최종 C점수 -

- ▶ 4점 : 구매건수 = 4, 거래당구매건수 = 1 or 2
구매건수 = 3, 거래당구매건수 = 1
- ▶ 3점 : 구매건수 = 4, 거래당구매건수 = 3 or 4
구매건수 = 3, 거래당구매건수 = 2 or 3
- ▶ 2점 : 구매건수 = 3, 거래당구매건수 = 4
구매건수 = 2, 거래당구매건수 = 1 or 2 or 3
- ▶ 1점 : 구매건수 = 2, 거래당구매건수 = 4
구매건수 = 1



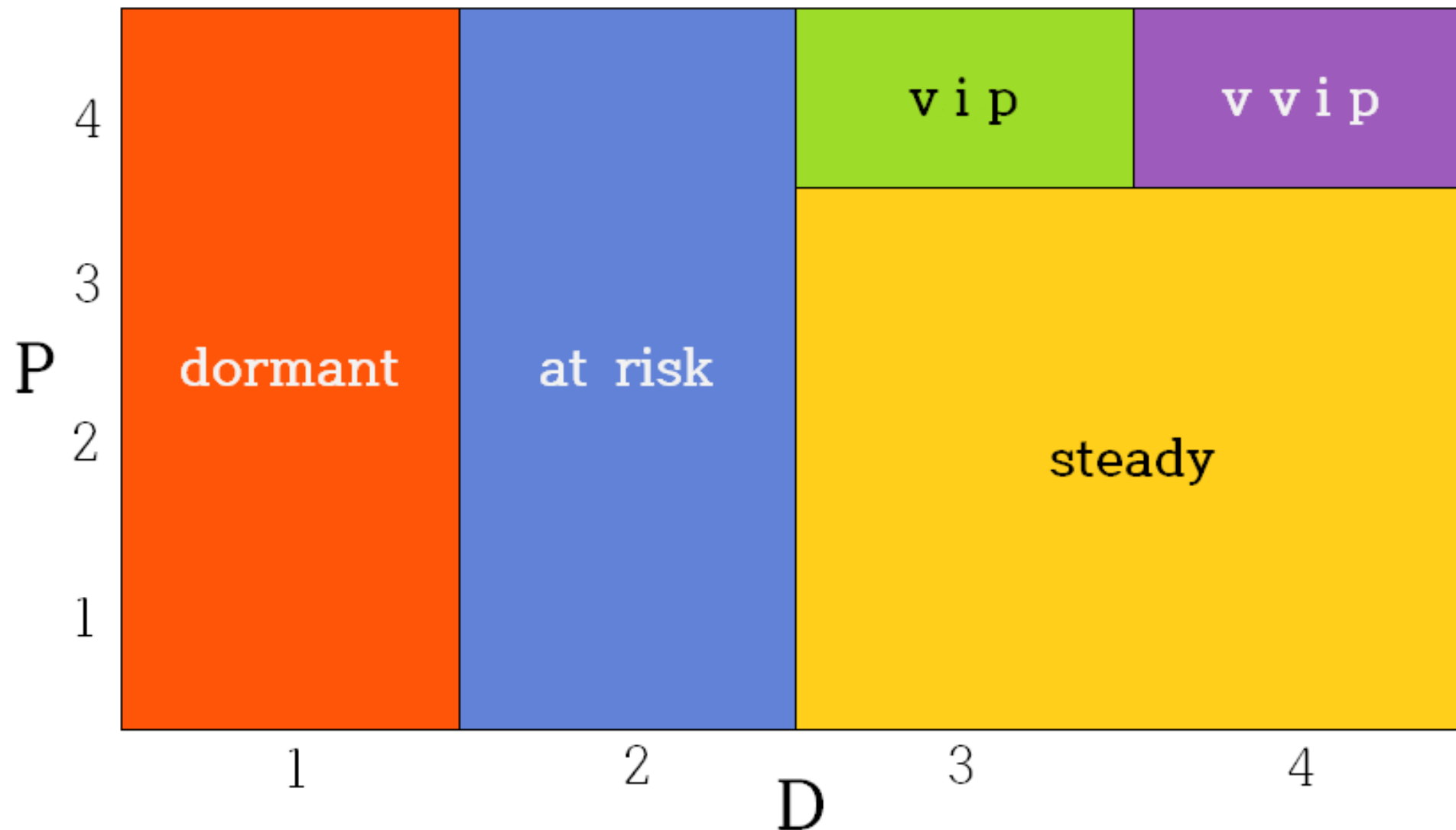
고객세분화

4

Customer Segment

Segment별 시각화

PD기준으로 나눈 최종 Segment

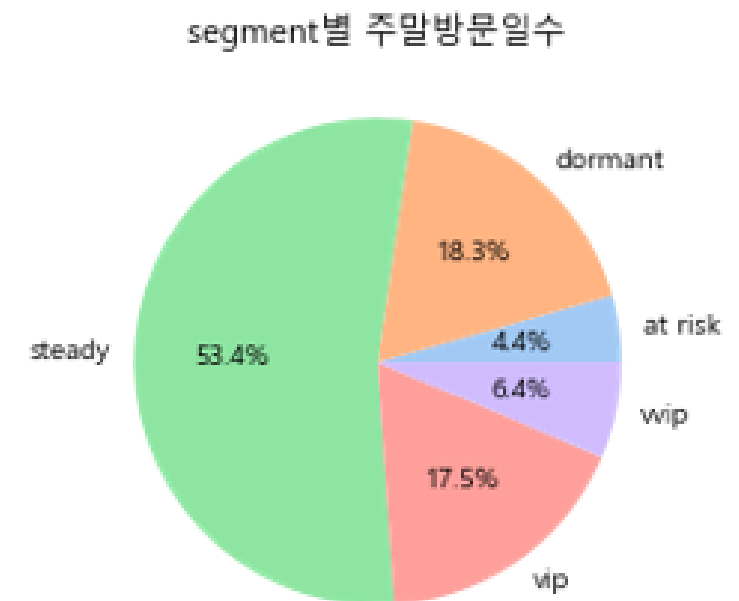
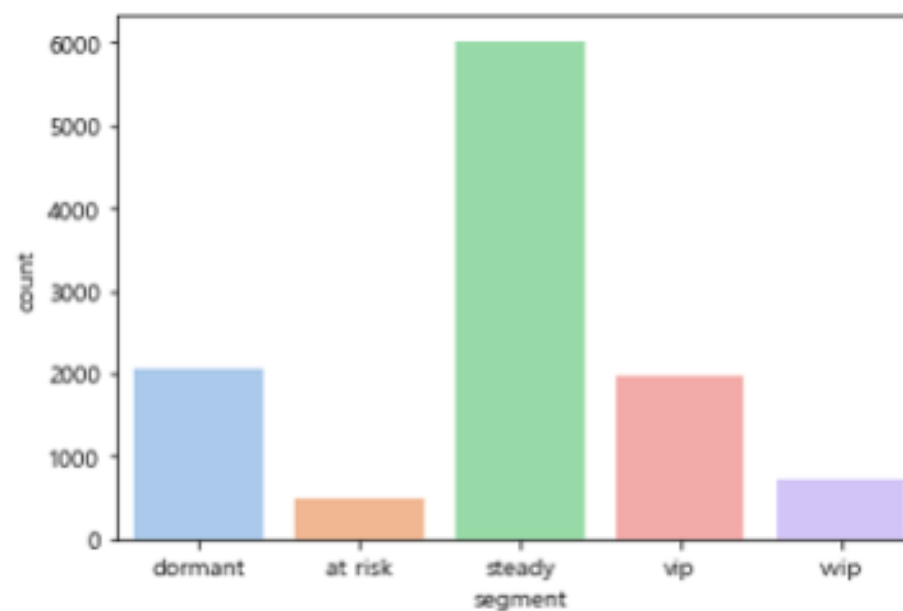
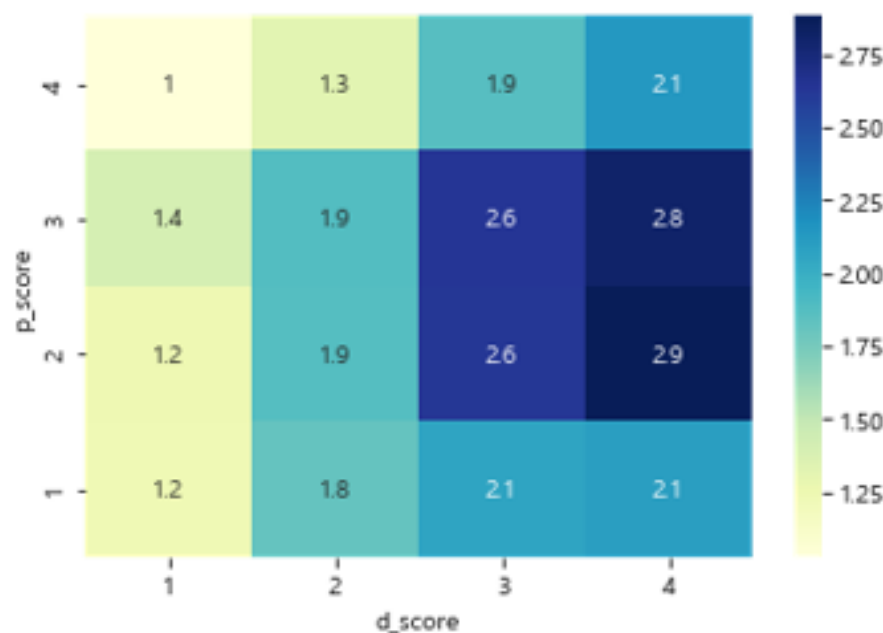


P를 구매금액, D를 구매일수라고 하겠다.

- ▶ **vvip** : 구매금액은 최고가이고, 구매일수는 매우 많은(작은) 고객
- ▶ **vip** : 구매금액은 최고가인데 구매일수는 아주 많지는 않은(중양값보다 조금 높은) 고객
- ▶ **steady** : 구매금액은 고가는 아니지만, 구매일수가 중양값 이상인 고객 (꾸준히 구매하는 고객)
- ▶ **at risk** : 구매금액은 저가부터 고가까지 다양하지만, 구매일수가 중양값 이하로 낮은 고객
- ▶ **dormant** : 구매금액은 저가부터 고가까지 다양하지만, 구매일수가 중양값보다 현저히 낮은 고객

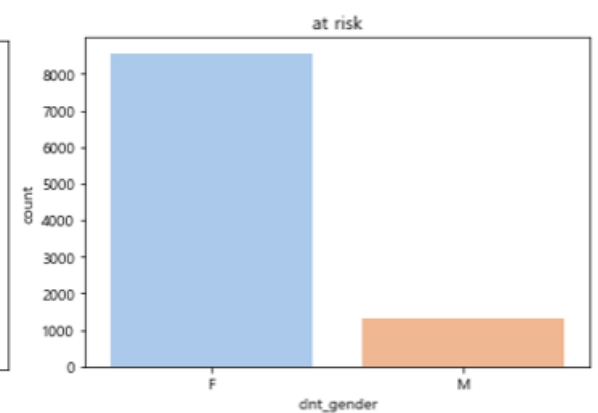
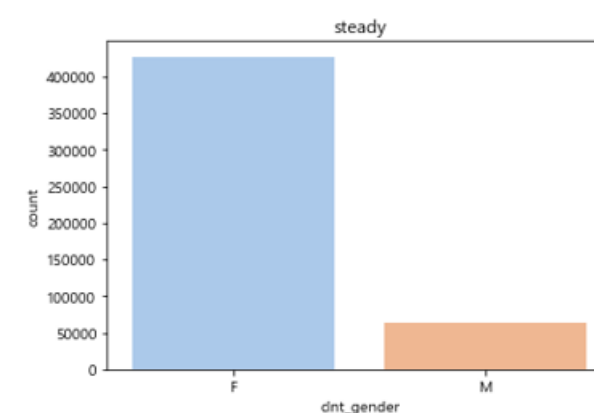
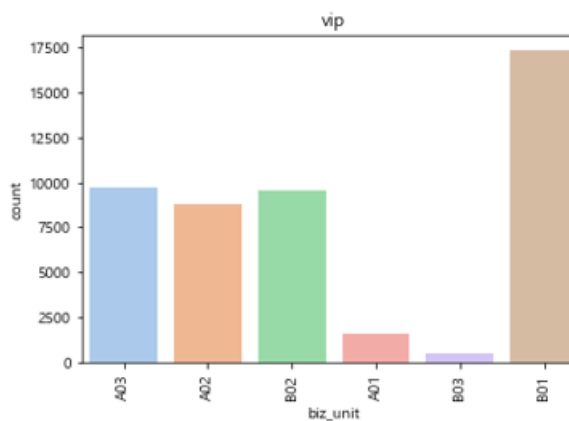
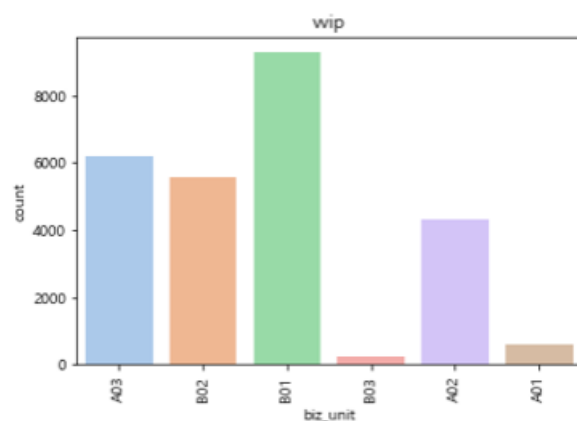
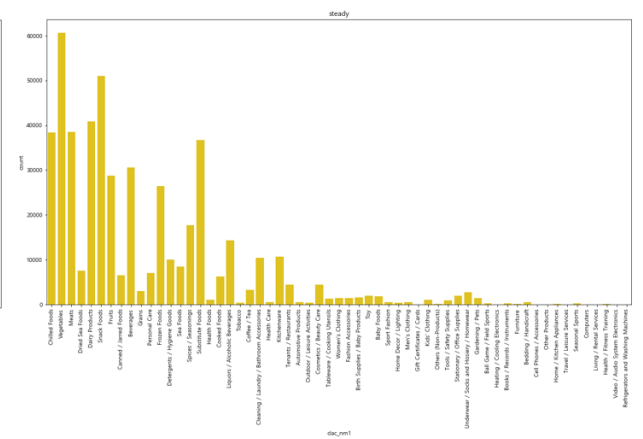
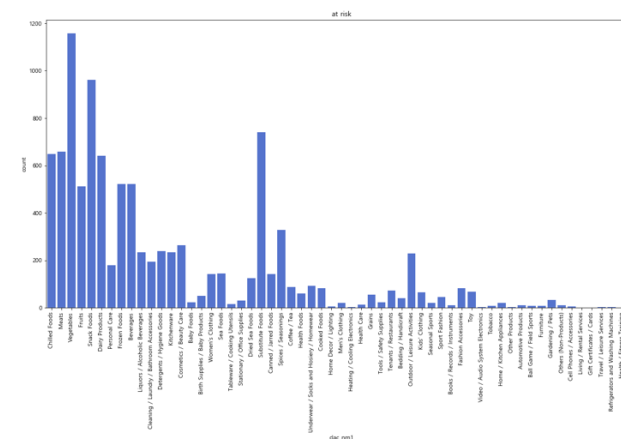
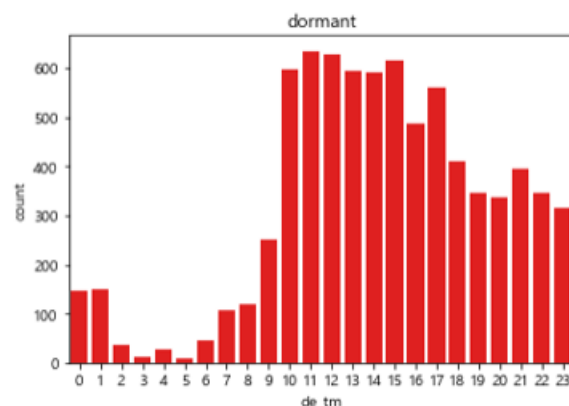
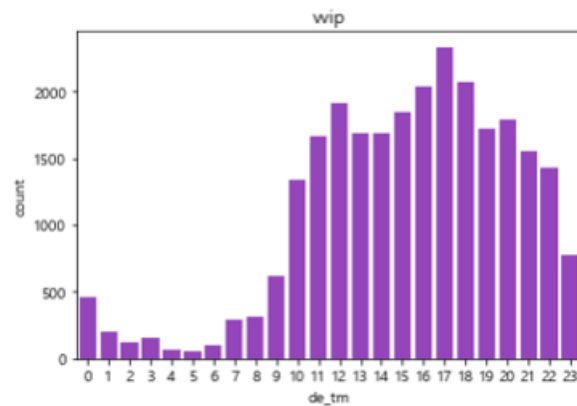
Segment별 시각화

- ▶ PD grid에 C정보 추가 : P(구매금액)이 2등급, 3등급이고, D(구매일수)는 3등급, 4등급일 때의 C(구매건수)가 높은 등급을 보인다는 것을 알 수 있다.
- ▶ segment별 고객 수 : 안정적으로 꾸준한 거래를 보이는 steady가 가장 많고, 구매일수 등급이 평균보다 조금 떨어지는 at risk가 가장 적다는 것을 알 수 있다. segment별 고객수의 분포를 알 수 있다.
- ▶ segment별 주말방문일수 : 어떤 segment가 주말에 가장 많이 방문하는지 알 수 있다. segment별 고객수의 분포와 결과값이 비슷한 것을 보니 segment별 주말방문일수는 고객수에 비례한다.



Segment별 시각화

- ▶ **segment별 구매시간** : 각 segment별로 어느 시간대에 가장 많이 구매 했는 지를 알 수 있다.
대체적으로 저녁시간쯤 구매를 많이 하는데, dormant고객은 점심시간쯤 구매가 많다.
- ▶ **segment별 구매상품종류(대)**: 각 segment별로 대분류상품중 어떤 상품을 가장 많이 구매 했는 지를 알 수 있다.
- ▶ **segment별 업종단위** : 각 segment별로 크게는 온라인 구매가 많은지 오프라인 구매가 많은지를 볼 수 있고, 온라인과 오프라인 안에서 어떤 이용처를 통해 구입했는지 알 수 있다.
- ▶ **segment별 성비** : 각 segment별로 성별 비율이 어떻게 되는지 볼 수 있다. 모든 segment에서 압도적으로 여성이 많은 것을 알 수 있다.



상품추천

5

clac_nm3

항목정리

segment별 성별 주요구매상품

segment별 상품추천

clac_nm3 항목정리

행은 clnt_id, 열은 clac_nm3로 하는 Ratings Matrix 만든다. (구매하면 1, 구매가 없으면 0)
빈발항목(최소 지지도 0.1 이상을 갖는 항목 집합)을 추출하는데 이때 두개 이상인 빈발항목만
추출한다. (최소 신뢰도 0.8 이상인 연관규칙을 도출해 볼 수 있다.)
빈발항목에 들어가지 않은 구매상품종류(소)는 구매상품종류(대)로 상향시킨다.

Segment별 성별 주요구매상품

모든 segment에서 여성이 압도적으로 많기 때문에 성별을 나눠서 분석한다.

▶ 여성고객

	dormant	at risk	vip	vvip	steady
0	Women's Clothing	Vegetables	Women's Clothing	Women's Clothing	Vegetables
1	Cosmetics / Beauty Care	Spices / Seasonings	Cosmetics / Beauty Care	Vegetables	Spices / Seasonings
2	Spices / Seasonings	Meats	Spices / Seasonings	Spices / Seasonings	Meats
3	Vegetables	Frozen Foods	Meats	Meats	General Snacks
4	Underwear / Socks and Hosiery / Homewear	Fruits	Fashion Accessories	Fashion Accessories	Fruits
5	Fruits	General Snacks	Vegetables	Fruits	Frozen Foods
6	Meats	Ramens	Fruits	Cosmetics / Beauty Care	Fresh Milk
7	Ramens	Substitute Foods	General Snacks	General Snacks	Snack Foods
8	Frozen Foods	Detergents / Hygiene Goods	Snack Foods	Frozen Foods	Substitute Foods
9	Fashion Accessories	Chicken Eggs	Frozen Foods	Substitute Foods	Chicken Eggs

▶ 남성고객

	dormant	at risk	vip	vvip	steady
0	Women's Clothing	Outdoor / Leisure Activities	Sport Fashion	Cosmetics / Beauty Care	Spices / Seasonings
1	Spices / Seasonings	Cosmetics / Beauty Care	Meats	Health Foods	Vegetables
2	Ramens	Meats	Fashion Accessories	Meats	Frozen Foods
3	Frozen Foods	General Snacks	Vegetables	Vegetables	Fruits
4	Meats	Imported Beer	Spices / Seasonings	Spices / Seasonings	Meats
5	Vegetables	Fruits	Frozen Foods	Fruits	General Snacks
6	General Snacks	Vegetables	Cosmetics / Beauty Care	Women's Clothing	Snack Foods
7	Fruits	Ramens	Women's Clothing	Ramens	Substitute Foods
8	Underwear / Socks and Hosiery / Homewear	Frozen Foods	Snack Foods	Sea Foods	Fresh Milk
9	Sport Fashion	Detergents / Hygiene Goods	Underwear / Socks and Hosiery / Homewear	Personal Care	Beverages

Segment별 상품추천 - vvip

- vvip 고객은 더 발전할 등급이 없기 때문에 vvip 군집 내에서 유사집단을 뽑아 추천상품 추천
- 주구매시간이 16시 ~ 18시이므로 이 시간대에 추천

▶ vvip 여성고객

vvip의 여성고객들의 주요구매상품을 참고하여 vvip 여성고객을 대표하는 clnt_id를 만든다.
유사도가 가장 높은 이웃의 수는 50으로 설정하고 고객마다 유사집단을 생성한다.

이 때, 유사집단 생성은 vvip 여성고객 내에서만 한다.

추천 상품수는 5개로 하며 고객별로 이미 구매한 상품을 제외하고 5개를 추천하도록 한다.
고객별로 각자 5개씩 추천을 하며 추가로 vvip 여성고객을 대표하는 clnt_id의 추천상품도 추천한다.

고객별 추천상품과 vvip 여성고객을 대표하는 clnt_id의 추천상품은 다음과 같다.

clnt_id		recommend_items
0	40	[Fashion Accessories, Sport Fashion, Health Fo...
1	57	[Snack Foods, Meats, Sea Foods, Fresh Milk, Ch...
2	627	[Substitute Foods, General Snacks, Tofu, Chill...
3	637	[Women's Clothing, Fashion Accessories, Person...
4	696	[Women's Clothing, Cosmetics / Beauty Care, Fa...
...
619	72004	[Meats, Sea Foods, Corn Snacks, Bananas, Dried...
620	72185	[Women's Clothing, Cosmetics / Beauty Care, Pe...
621	72238	[Beverages, Sea Foods, Bananas, Frozen Foods, ...
622	72373	[Spices / Seasonings, Women's Clothing, Deterg...

	purchased	recommend
0	Vegetables	Chicken Eggs
1	Meats	Substitute Foods
2	Spices / Seasonings	Sea Foods
3	Fruits	Ramens
4	Frozen Foods	Women's Clothing

Segment별 상품추천 - vvip

- vvip 고객은 더 발전할 등급이 없기 때문에 vvip 군집 내에서 유사집단을 뽑아 추천상품 추천
- 주구매시간이 16시 ~ 18시이므로 이 시간대에 추천

▶ vvip 남성고객

vvip의 남성고객들의 주요구매상품을 참고하여 vvip 남성고객을 대표하는 clnt_id를 만든다.
유사도가 가장 높은 이웃의 수는 50으로 설정하고 고객마다 유사집단을 생성한다.

이 때, 유사집단 생성은 vvip 남성고객 내에서만 한다.

추천 상품수는 5개로 하며 고객별로 이미 구매한 상품을 제외하고 5개를 추천하도록 한다.
고객별로 각자 5개씩 추천을 하며 추가로 vvip 남성고객을 대표하는 clnt_id의 추천상품도 추천한다.

고객별 추천상품과 vvip 남성고객을 대표하는 clnt_id의 추천상품은 다음과 같다.

	clnt_id	recommend_items
0	353	[Spices / Seasonings, Vegetables, Frozen Foods...
1	521	[Fruits, Personal Care, Chicken Eggs, Sea Food...
2	551	[Spices / Seasonings, Personal Care, Fruits, C...
3	1166	[Bibim Ramens, Peppers, Spoon Type Yogurts, Fu...
4	1483	[Frozen Foods, General Snacks, Ramens, Sea Foo...
...
94	69238	[Meats, Personal Care, Fresh Milk, Spoon Type ...
95	71638	[Vegetables, Spices / Seasonings, Meats, Froze...
96	72107	[Spices / Seasonings, Fruits, Vegetables, Meat...
97	72176	[Fruits, Vegetables, Chicken Eggs, Snack Foods...

	purchased	recommend
0	Vegetables	Substitute Foods
1	Spices / Seasonings	Personal Care
2	Meats	Cosmetics / Beauty Care
3	Fruits	Snack Foods
4	Frozen Foods	Detergents / Hygiene Goods

Segment별 상품추천 - vip

- vip 고객은 vvip 등급으로 상향될 수 있도록 vvip 고객에서만 유사집단을 뽑아 추천
- 주구매시간이 14시 ~ 17시이므로 이 시간대에 추천
- 평균구매건수 이상, 4분위수를 기준으로 한 단계별 특정금액이상 구매시 적립금 추가지급

▶ vip 여성고객

vip의 여성고객들의 주요구매상품을 참고하여 vip 여성고객을 대표하는 clnt_id를 만든다.

유사도가 가장 높은 이웃의 수는 50으로 설정하고 고객마다 유사집단을 생성한다.

이 때, 유사집단 생성은 vvip 여성고객 내에서만 한다.

추천 상품수는 5개로 하며 이미 구매한 상품을 제외하고 5개를 추천하도록 한다.

vip 여성고객을 대표하는 clnt_id의 추천상품을 추천한다.

vip 여성고객을 대표하는 clnt_id의 추천상품은 다음과 같다.

vip 여성고객의 평균 구매수량은 2이고, 구매금액의 제1사분위수는 약14000원, 제2사분위수는 약28000원, 제3사분위수는 약 62000원이다.

	purchased	recommend
0	Women's Clothing	Underwear / Socks and Hosiery / Homewear
1	Cosmetics / Beauty Care	Personal Care
2	Spices / Seasonings	Health Foods
3	Meats	Sport Fashion
4	Fashion Accessories	Sea Foods

구매수량이 2개 이상이고,
상품금액이 각각 15000원 이상이면 300원,
30000원 이상이면 750원,
60000원 이상이면 1800원 적립금 추가지급

Segment별 상품추천 - vip

- vip 고객은 vvip 등급으로 상향될 수 있도록 vvip 고객에서만 유사집단을 뽑아 추천
- 주구매시간이 14시 ~ 17시이므로 이 시간대에 추천
- 평균구매건수 이상, 4분위수를 기준으로 한 단계별 특정금액이상 구매시 적립금 추가지급

▶ vip 남성고객

vip의 남성고객들의 주요구매상품을 참고하여 vip 남성고객을 대표하는 clnt_id를 만든다.
유사도가 가장 높은 이웃의 수는 50으로 설정하고 고객마다 유사집단을 생성한다.
이 때, 유사집단 생성은 vvip 남성고객 내에서만 한다.
추천 상품수는 5개로 하며 이미 구매한 상품을 제외하고 5개를 추천하도록 한다.
vip 남성고객을 대표하는 clnt_id의 추천상품을 추천한다.
vip 남성고객을 대표하는 clnt_id의 추천상품은 다음과 같다.
vip 남성고객의 평균 구매수량은 2이고, 구매금액의 제1사분위수는 약15000원, 제2사분위수는 약31000원, 제3사분위수는 약 80000원이다.

	purchased	recommend
0	Cosmetics / Beauty Care	Underwear / Socks and Hosiery / Homewear
1	Outdoor / Leisure Activities	Personal Care
2	Sport Fashion	Health Foods
3	Meats	Sport Fashion
4	Fashion Accessories	Sea Foods

구매수량이 2개 이상이고,
상품금액이 각각 15000원 이상이면 300원,
30000원 이이면 750원,
80000원 이상이면 2400원 적립금 추가지급

Segment별 상품추천 - steady

- steady 고객은 고객수도 가장 많고, 꾸준히 안정적인 매출을 얻을 수 있도록 도와주는 고객이므로 계속 유지할 수 있게 steady고객 내에서 유사집단의 추천상품 추천
- 주구매시간이 15시 ~ 18시이므로 이 시간대에 추천
- steady 고객내에서의 가격상승을 위해 상위 10%에 해당하는 가격대의 상품 할인권 지급

▶ steady 여성고객

steady의 여성고객들의 주요구매상품을 참고하여 steady 여성고객을 대표하는 clnt_id를 만든다.

유사도가 가장 높은 이웃의 수는 50으로 설정하고 고객마다 유사집단을 생성한다.

이 때, 유사집단 생성은 steady 여성고객 내에서만 한다.

추천 상품수는 5개로 하며 고객별로 이미 구매한 상품을 제외하고 5개를 추천하도록 한다.

고객별로 각자 5개씩 추천을 하며 추가로 steady 여성고객을 대표하는 clnt_id의 추천상품도 추천한다. 고객별 추천상품과 steady 여성고객을 대표하는 clnt_id의 추천상품은 다음과 같다.

steady 여성고객의 상위 10%에 해당하는 가격대는 30000원이다.

clnt_id		recommend_items
0	9	[Chicken Eggs, Beverages, Detergents / Hygiene...
1	20	[Tofu, Substitute Foods, Detergents / Hygiene ...
2	23	[Tofu, Snack Foods, Ramens, Spoon Type Yogurts...
3	24	[Fresh Milk, General Snacks, Chilled Foods, Ca...
4	29	[Beverages, Sea Foods, Frozen Foods, Tofu, Chi...
...
5174	72258	[Spices / Seasonings, Ramens, Vegetables, Frui...
5175	72321	[General Snacks, Frozen Foods, Ramens, Vegetab...
5176	72340	[Substitute Foods, Chilled Foods, Beverages, F...
5177	72356	[Spices / Seasonings, Meats, Substitute Foods,...

purchased		recommend
0	Vegetables	Fruits
1	Spices / Seasonings	Trash Bags
2	Meats	Chicken Eggs
3	General Snacks	Detergents / Hygiene Goods
4	Frozen Foods	Snack Foods

30000원대 상품 10%
할인쿠폰 지급

Segment별 상품추천 - steady

- steady 고객은 고객수도 가장 많고, 꾸준히 안정적인 매출을 얻을 수 있도록 도와주는 고객이므로 계속 유지할 수 있게 steady고객 내에서 유사집단의 추천상품 추천
- 주구매시간이 15시 ~ 18시이므로 이 시간대에 추천
- steady 고객내에서의 가격상승을 위해 상위 10%에 해당하는 가격대의 상품 할인권 지급

▶ steady 남성고객

steady의 남성고객들의 주요구매상품을 참고하여 steady 남성고객을 대표하는 clnt_id를 만든다.

유사도가 가장 높은 이웃의 수는 50으로 설정하고 고객마다 유사집단을 생성한다.

이 때, 유사집단 생성은 steady 남성고객 내에서만 한다.

추천 상품수는 5개로 하며 고객별로 이미 구매한 상품을 제외하고 5개를 추천하도록 한다.

고객별로 각자 5개씩 추천을 하며 추가로 steady 남성고객을 대표하는 clnt_id의 추천상품도 추천한다. 고객별 추천상품과 steady 남성고객을 대표하는 clnt_id의 추천상품은 다음과 같다.

steady 남성고객의 상위 10%에 해당하는 가격대는 30000원이다.

	clnt_id	recommend_items
0	202	[Dried Sea Foods, Fish Cakes, Personal Care, P...
1	336	[Fruits, Vegetables, Chicken Eggs, Frozen Food...
2	416	[Snack Foods, Chilled Foods, Corn Snacks, Froz...
3	456	[Tofu, Meats, Ramens, General Snacks, Snack Fo...
4	464	[Frozen Foods, Chicken Eggs, Substitute Foods,...
...
840	72267	[Vegetables, Substitute Foods, Chicken Eggs, T...
841	72284	[Chilled Foods, Beverages, Vegetables, Fruits,...
842	72333	[Spices / Seasonings, Fruits, Fresh Milk, Meat...
843	72424	[Fruits, Snack Foods, Sea Foods, Chilled Foods...

		purchased	recommend	id
0		0 Spices / Seasonings	Meats	ts
1	Spi	1 Frozen Foods	Ramens	js
2		2 Vegetables	Snack Foods	js
3		3 Fruits	Tofu	js
4		4 General Snacks	Fresh Milk	js

30000원대 상품 10%
할인쿠폰 지급

Segment별 상품추천 - at risk

- at risk 고객은 dormant 고객이 될 위험이 있는 고객이기 때문에 꾸준한 구매가 필요한 제품을 추천. 즉, steady 고객에서만 유사집단을 뽑아 추천상품 추천
- 주구매시간이 11시 ~ 12시, 18시이므로 이 시간대에 추천
- 평균 구매수량 이상 상품 구매시 포인트 적립률 3% 상향

▶at risk 여성고객

at risk의 여성고객들의 주요구매상품을 참고하여 at risk 여성고객을 대표하는 clnt_id를 만든다.

유사도가 가장 높은 이웃의 수는 50으로 설정하고 고객마다 유사집단을 생성한다.

이 때, 유사집단 생성은 steady 여성고객 내에서만 한다.

추천 상품수는 5개로 하며 이미 구매한 상품을 제외하고 5개를 추천하도록 한다.

at risk 여성고객을 대표하는 clnt_id의 추천상품을 추천한다.

at risk 여성고객을 대표하는 clnt_id의 추천상품은 다음과 같다.

at risk 여성고객의 평균구매수량은 약 2개이다.

	purchased	recommend
0	Vegetables	Chicken Eggs
1	Spices / Seasonings	Fresh Milk
2	Meats	Detergents / Hygiene Goods
3	Frozen Foods	Bananas
4	Fruits	Sea Foods

2개 이상 상품을 구매할 시 포인트 적립률
3% 상향

Segment별 상품추천 - at risk

- at risk 고객은 dormant 고객이 될 위험이 있는 고객이기 때문에 꾸준한 구매가 필요한 제품을 추천. 즉, steady 고객에서만 유사집단을 뽑아 추천상품 추천
- 주구매시간이 11시 ~ 12시, 18시이므로 이 시간대에 추천
- 평균 구매수량 이상 상품 구매시 포인트 적립률 3% 상향

▶at risk 남성고객

at risk의 남성고객들의 주요구매상품을 참고하여 at risk 남성고객을 대표하는 clnt_id를 만든다.

유사도가 가장 높은 이웃의 수는 50으로 설정하고 고객마다 유사집단을 생성한다.

이 때, 유사집단 생성은 steady 남성고객 내에서만 한다.

추천 상품수는 5개로 하며 이미 구매한 상품을 제외하고 5개를 추천하도록 한다.

at risk 남성고객을 대표하는 clnt_id의 추천상품을 추천한다.

at risk 남성고객을 대표하는 clnt_id의 추천상품은 다음과 같다.

at risk 남성고객의 평균구매수량은 약 2개이다.

	purchased		recommend
0	0	Meats	Chicken Eggs
1	1	Ramens	Fresh Milk
2	2	General Snacks	Detergents / Hygiene Goods
3	3	Frozen Foods	Bananas
4	4	Fruits	Sea Foods

2개 이상 상품을 구매할 시 포인트 적립률
3% 상향

Segment별 상품추천 - dormant

- dormant 고객은 휴면고객이기 때문에 전체고객에서 유사집단의 추천상품 추천
- 주구매시간이 10시 ~ 15시이므로 이 시간대에 추천
- 온라인으로는 배송비 무료 쿠폰 지급, 오프라인으로는 방문할인 서비스 제공
- 한개 이상 구매시, dormant 고객의 평균구매금액 이상의 제품에 한해 기간이 2주인 10% 할인 쿠폰 지급

▶ dormant 여성고객

dormant 여성고객들의 주요구매상품을 참고하여 dormant 여성고객을 대표하는 clnt_id를 만든다. 유사도가 가장 높은 이웃의 수는 50으로 설정하고 고객마다 유사집단을 생성한다.

이 때, 유사집단 생성은 전체 여성고객 내에서만 한다.

추천 상품수는 5개로 하며 고객별로 이미 구매한 상품을 제외하고 5개를 추천하도록 한다.

고객별로 각자 5개씩 추천을 하며 추가로 dormant 여성고객을 대표하는 clnt_id의 추천상품도 추천한다. 고객별 추천상품과 dormant 여성고객을 대표하는 clnt_id의 추천상품은 다음과 같다. dormant 여성고객의 평균구매금액은 약30000원이다.

clnt_id			recommend_items	
0	2	[General Snacks, Bibim Ramens, Vegetables, Mea...	purchased	recommend
1	9	[Chicken Eggs, Beverages, Chilled Foods, Canne...	0 Women's Clothing	Stationary / Office Supplies
2	12	[Fruits, Cosmetics / Beauty Care, Meats, Chick...	1 Cosmetics / Beauty Care	Sport Fashion
3	20	[Tofu, Substitute Foods, Detergents / Hygiene ...	2 Spices / Seasonings	Apples
4	23	[Tofu, Ramens, Snack Foods, General Snacks, De...	3 Vegetables	Health Foods
...	4 Underwear / Socks and Hosiery / Homewear	Fixed-price Living Products
9616	72373	[Vegetables, Spices / Seasonings, Detergents / ...		
9617	72400	[Women's Clothing, Health Foods, Sanitary Pads...		
9618	72410	[Fruits, Ramens, Fresh Milk, Tofu, Substitute ...		
9619	72423	[Grains, Domestic Porks - Picnics, Fish Cakes,...		

온라인 고객 : 배송비 무료

오프라인 고객 : 방문 할인

30000원대 상품 10%

할인쿠폰 지급 (단, 기간 2주)

Segment별 상품추천 - dormant

- dormant 고객은 휴면고객이기 때문에 전체고객에서 유사집단의 추천상품 추천
- 주구매시간이 10시 ~ 15시이므로 이 시간대에 추천
- 온라인으로는 배송비 무료 쿠폰 지급, 오프라인으로는 방문할인 서비스 제공
- 한개 이상 구매시, dormant 고객의 평균구매금액 이상의 제품에 한해 기간이 2주인 10% 할인 쿠폰 지급

▶ dormant 남성고객

dormant 남성고객들의 주요구매상품을 참고하여 dormant 남성고객을 대표하는 clnt_id를 만든다. 유사도가 가장 높은 이웃의 수는 50으로 설정하고 고객마다 유사집단을 생성한다.

이 때, 유사집단 생성은 전체 남성고객 내에서만 한다.

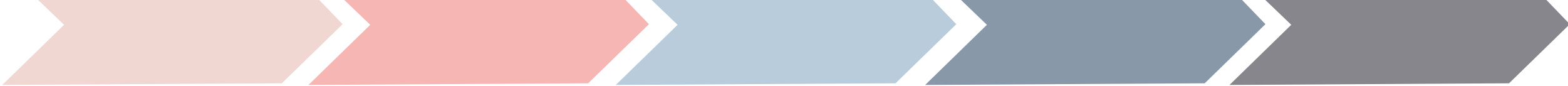
추천 상품수는 5개로 하며 고객별로 이미 구매한 상품을 제외하고 5개를 추천하도록 한다.

고객별로 각자 5개씩 추천을 하며 추가로 dormant 남성고객을 대표하는 clnt_id의 추천상품도 추천한다. 고객별 추천상품과 dormant 남성고객을 대표하는 clnt_id의 추천상품은 다음과 같다. dormant 남성고객의 평균구매금액은 약30000원이다.


clnt_id			recommend_items			purchased			recommend		
0	54	[Fruits, Spices / Seasonings, Chilled Foods, G...	0	0	Women's Clothing	Tofu	Supplies				
1	182	[Ramens, Spices / Seasonings, Chilled Foods, M...	1	1	Spices / Seasonings	Fruits	t Fashion				
2	202	[Dried Sea Foods, Fish Cakes, Onions, Personal...	2	2	Ramens	Canned / Jarred Foods	Apples				
3	269	[Ramens, Spices / Seasonings, Bibim Ramens, Ch...	3	3	Frozen Foods	Frozen Korean Pancakes	lth Foods				
4	282	[Women's Clothing, Cosmetics / Beauty Care, Sp...	4	4	Meats	Chicken Eggs	Products				
...									
1643	72267	[Vegetables, Substitute Foods, Chicken Eggs, F...									
1644	72284	[Beverages, Chilled Foods, Fruits, Vegetables, ...									
1645	72333	[Spices / Seasonings, Fresh Milk, Beverages, F...									
1646	72424	[Fruits, Snack Foods, Canned / Jarred Foods, S...									


온라인 고객 : 배송비 무료
오프라인 고객 : 방문 할인

30000원대 상품 10%
할인쿠폰 지급 (단, 기간 2주)



Q & A





THANK YOU!

2020.06.20

빅데이터경영통계전공

20192784 윤경서

