

# 온라인 행동 데이터 분석을 통한 개인화 마케팅 전략 제안



고객행동기반 맞춤형 서비스와 온오프라인 통합을 중심으로

다다fam  
류재희 이태훈 정은지  
한양대학교 정보융합학과

# Contents

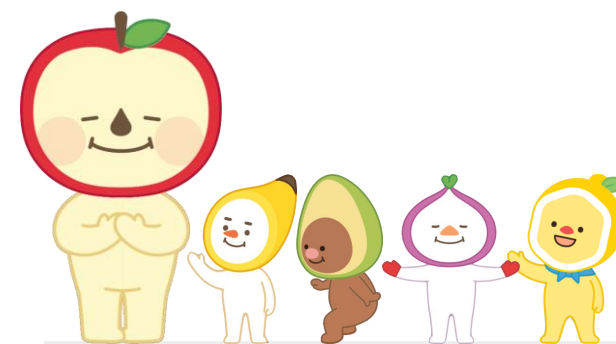
---

	<b>분석개요 및 데이터 탐색</b> 분석 목표 및 전략 · 데이터 탐색 및 시각화 · 데이터 분석 로드맵	03
	<b>고객 및 상품군 분석</b> 온라인 행동기반 고객 세분화 · 고객 가치 분류 · 고객 가치 분류별 마케팅 전략	09
	<b>상품군별 수요트렌드 예측 및 인사이트 도출</b> 수요트렌드 개요 및 모델링 · 수요트렌드 예측 구현	20
	<b>주요 상품군별 온라인 선호지수 생성</b> 선호지수 개요 및 모델링 · 선호지수 모델 개선	28
	<b>신규 서비스 제안</b> 고객행동기반 맞춤형 서비스 “Customer Profiling” 온오프라인 통합을 위한 제안 “Real Smart Store”	32
		37

---

# 01 분석개요 및 데이터 탐색

- 분석 목표 및 전략
- 데이터 탐색 및 시각화
- 데이터 분석 로드맵



## Purpose

## 분석목표

고객행동기반 리타겟팅을 통한

효율적 마케팅

온오프라인 통합

마케팅 아이디어

수요트렌드와 선호지수

결합 사용

고객행동기반

개인 맞춤형 서비스

## Strategy

## 분석전략

## 데이터 전처리

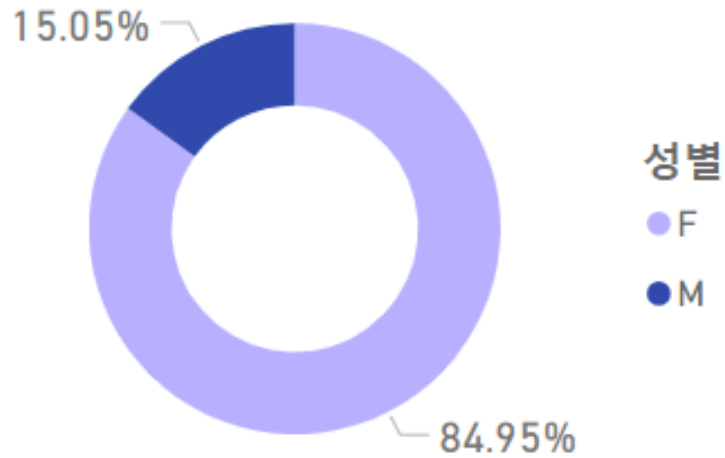
- 다양한 고객의 데이터 중 사용 변수 內 이상치 제거
- 제공된 데이터를 변형하여 분석에 용이한 파생변수 생성
- 외부 데이터 수집 최대화

## 세부 과제별 모델링

- 고객 및 상품군 분석 : 패턴 분석
- 선호지수 : 베이지안 확률
- 수요트렌드 : 시계열 계절 분해법 & RNN LSTM(인공신경망)

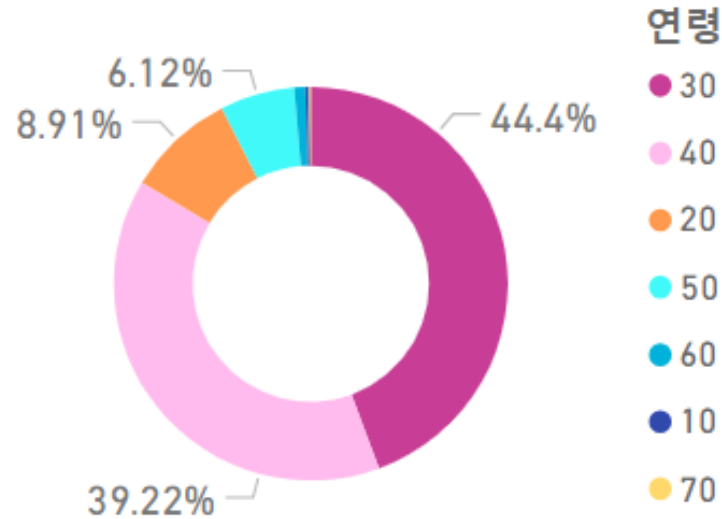
### 고객 성별 분포

※ 04.Custom 클라이언트 별 성별 기준



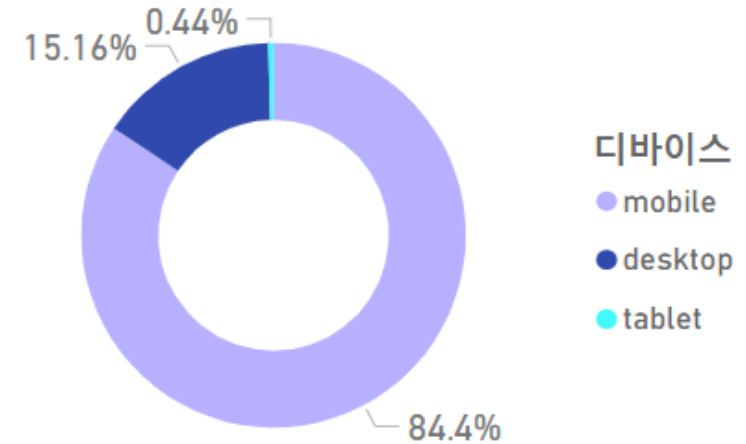
### 고객 구매 연령 분포

※ 구매 발생 건 기준



### 접속 디바이스 분포

※ 05.Session 접속 디바이스 기준



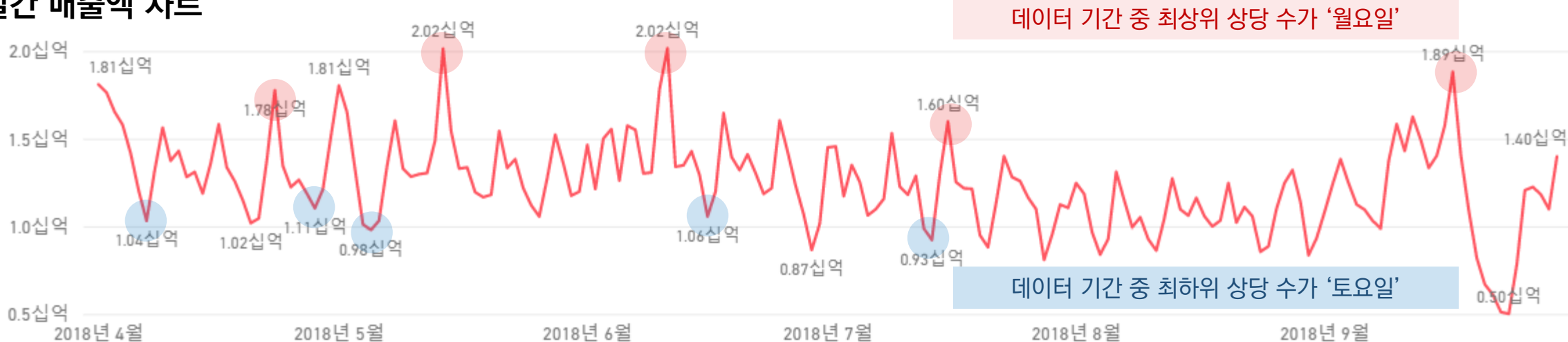


# 01 분석개요 및 데이터 탐색

## 데이터 탐색 및 시각화

7

### 일간 매출액 차트

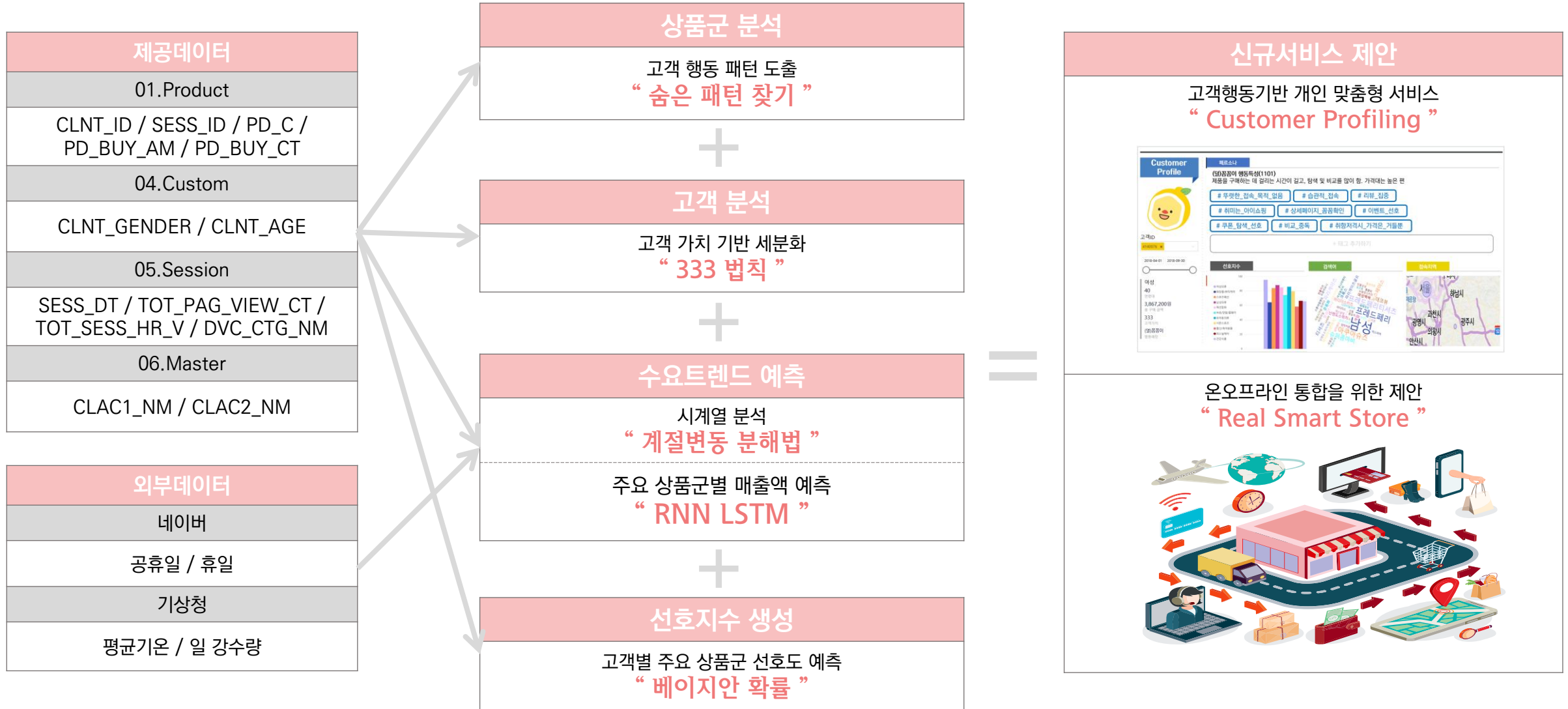


### 상품군별 구매횟수 분포



# 01 분석개요 및 데이터 탐색

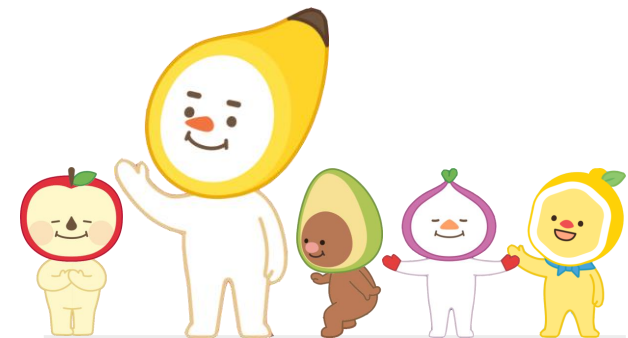
## 데이터 분석 로드맵





## 02 고객 및 상품군 분석

- 온라인 행동기반 고객 세분화
- 고객 가치 분류
- 고객 가치 분류별 마케팅 전략



### What

## 상품군별 온라인 행동 패턴

### 온라인 행동 데이터 정의

고객 온라인 행동 추적을 통해 유입부터 전환까지 과정을 분석할 수 있는 데이터  
사용 디바이스, 제품 정보를 확인하는 횟수, 고려 시간, 검색 횟수 등

### 온라인 행동으로 사용할 데이터

#### ▪ 메인변수

SEARCH\_CNT : 세션 내 총 검색량

TOT\_PAG\_VIEW\_CT : 세션 내 총 페이지 조회 건수

TOT\_SESS\_HR\_V : 세션 내 머무르며 구매하기까지 고민한 시간

#### ▪ 서브변수

PD\_BUY\_AM : 구매한 상품의 금액

### Why

## 온라인 행동 기반 고객 세분화

리타겟팅을 위한 온라인 행동 기반 고객 세분화 및

도출 결과를 활용한 맞춤형 마케팅 전략 제안

(기존 인구통계적 기준 고객 세분화에서 **온라인 행동 기준으로 변경**)

### How

## 온라인 행동 기반 고객 세분화 방법

### 주 고객군의 상품군별 온라인 행동 차이 확인

- 데이터 전처리에서 확인한 7개 고객층을 주 고객 군으로 설정
- 상품군별 각 고객군의 메인변수 평균값 산출
- 각 고객군의 상품군별 상대적 차이 확인

여기서, 평균은 각 고객군의 메인변수 평균값을 의미

**평균 대비 상대적으로 높으면 1 부여, 낮으면 0 부여**

연령	성별	여성의류	화장품 /뷰티케어	건강식품	평균
20	F	807.1	708.6	706.7	726.9
	M	1059.5	656.5	672.5	729.4
30	F	802.5	728.3	787.7	730.0
	M	760.9	691.0	770.4	693.5
40	F	906.7	799.0	898.2	803.9
	M	861.5	829.5	804.3	763.4
50	F	1052.1	914.2	937.3	939.7

How

### 온라인 행동 기반 고객 세분화 방법

#### 행동 패턴 군집화 및 정의

- 데이터 분석 결과 상품군별로 비슷한 온라인 행동 양상을 보이는 고객들을 3개의 집단으로 군집화 함
- 각 상품군 특성을 분석하여 집단을 정의함
- 고객 세분화 기준 변경으로 **효율적 리타겟팅 가능**

(인구통계적 기준 세분화 : 7개 → 행동 패턴 기준 세분화 : 3개)

\* ) 스포츠패션과 유아동의류는 각 2가지 행동 패턴을 고르게 보임

꼼꼼이		매니아	털털이	
덜 꼼꼼이 (1101)	더 꼼꼼이 (1111)	(0011)	더 털털이 (0000)	덜 털털이 (0010)
검색어 입력 無	검색어 입력 有	검색어 입력 有	검색어 입력 無	검색어 입력 有
여성의류, 패션잡화, *스포츠패션1	*스포츠패션2, 건강식품	화장품/뷰티케어	남성의류, 속옷/양말/홈웨어, *유아동의류1, 출산/육아용품	*유아동의류2, 퍼스널케어
제품을 구매하는 데 걸리는 시간이 길고, 탐색 및 비교를 많이 함 가격대는 높은 편		제품을 구매하는 데 걸리는 시간이 짧고, 탐색 및 비교를 적게 함 가격대는 높은 편	제품을 구매하는 데 걸리는 시간이 짧고, 탐색 및 비교를 적게 함 가격대는 낮은 편	

여기서 행동변수 순서는 고려시간, 페이지 조회수, 검색 횟수, 구매단가 순이며, 이 중 구매단가는 행동데이터 해석을 돕는 서브변수로 활용되었음

### Marketing

### 꼼꼼이 공략법

#### 비교쟁이 소비자를 타겟팅하자

- 비교를 좋아하는 소비자 특성을 활용하여 **장바구니에 제품 비교 기능** 추가  
나와 **동일 패턴의 고객**들이 구매한 제품을 **한 눈에 볼 수 있게** 비교 페이지 제공  
(한국 소비자들이 바라는 사항 중 49% 비교도구 필요, 조사기관 : SAP, N=1,000명)
- 연결 네트워크** 상의 제품을 같이 추천하여 **거부감 없이** 자연스러운 노출 진행

1장바구니
주문하실 상품을 선택해주세요.

1장바구니
2주문결제
3주문완료

고객님! 장바구니상품은 30일간 보관됩니다.  
상품을 장기간 보관하시려면 [위시리스트 담기]를 클릭해주세요.

가격인하 알림메일, 받고 계세요?  
위시리스트/장바구니에 담은 상품의 가격이 할인되면 메일을 보내드립니다. [매일수신여부/주소 확인 >](#)

다른 고객들이 많이 구매한 상품입니다.

선택	상품정보	판매가	최대 할인 / 혜택	수량	소계
<input checked="" type="checkbox"/>	[오코볼링] [3,120원할인쿠폰] [오코볼링] 레옹소프트 - cd • Color볼링 옵션변경	39,000원 [상품권혜택가] 35,880원	즉석쿠폰: 3,120원 할인	1 변경	39,000원
<input checked="" type="checkbox"/>	[제이스타일] 빅사이즈/제이스타일 [라 오던 반포털 기모 원피스 • 색상:베이지,사이즈:13 100 110 옵션변경]	32,200원 [상품권혜택가] 28,120원	즉석쿠폰: 7,080원 할인	1 변경	32,200원
<input checked="" type="checkbox"/>	[제이스타일] 빅사이즈/제이스타일 [라 엘르 피자기모 베이직 셔츠 • 색상:베이지,사이즈:13 100 110 옵션변경]	36,800원 [상품권혜택가] 28,710원	즉석쿠폰: 6,090원 할인	1 변경	36,800원
<input checked="" type="checkbox"/>	[스토리아스] [2,450원할인쿠폰] [STORY US] 365 SHIRTS 버튼 다운 옥스포드 셔츠 9578305002051 • 색상:블루,사이즈:100 옵션변경]	49,000원 [상품권혜택가] 46,550원	즉석쿠폰: 2,450원 할인	1 변경	49,000원

다른 고객들이 이 상품군과 함께 구매한 상품입니다.

선택	상품정보	판매가	최대 할인 / 혜택	수량	소계
<input checked="" type="checkbox"/>	[스토리아스] [2,450원할인쿠폰] [STORY US] 365 SHIRTS 버튼 다운 옥스포드 셔츠 9578305002051 • 색상:블루,사이즈:100 옵션변경]	49,000원 [상품권혜택가] 46,550원	즉석쿠폰: 2,450원 할인	1 변경	49,000원

#### 고민쟁이 소비자를 타겟팅하자

- 고민과 탐색이 비교적 많은 꼼꼼이 특성 상 **‘고객의 선택이 틀리지 않았다’**는 **인지부조화 최소화**가 필요
- 구매 고객이 직접 제품을 개봉하는 **1분 이내의 개봉기 영상**을 리뷰에 업로드하도록 이벤트 진행

자세히보기

☐ 사진/영상 상품평만 보기

고객추천순



★★★★★ 2019.01.02



277명이 이 상품평을 추천했습니다.

추천할래요

출처 : Youtube, MK

- 고객혜택 : 고객 등급 한 단계 상향, 감사쿠폰
- 자사혜택 : 독특하고 신뢰성 높은 리뷰 콘텐츠 확보로 다른 고민쟁이 꼼꼼이의 구매율 향상 가능

### Marketing

### 매니아 공략법

#### 재구매 욕구를 타겟팅하자

- 소모품인 화장품/뷰티케어 특성 상 **재구매 가능성**이 높으므로, 매니아 고객이 이전 구매한 제품의 **소모기간 경과 시** 구매한 **제품정보와 할인 쿠폰을 Push**
- 동시에 사용하면 좋은 제품(ex. 립스틱과 립밤)과 양성 제품을 취급하는 브랜드의 경우 **‘남편(아내)를 위한 추천’**을 제시하여 다른 제품도 탐색하도록 유인

**고객님**  
**구매하신 제품은 다 사용하셨나요~?**

■ 6개월 전 구매하신 상품



[조르지오 아르마니]립 마그넷 립 프리즈 스프링 셔벗-틴트 리미티드 컬렉션 3.9

판매가 46,000원  
참좋은예매가 43,700원

■ 같이 구매하면 좋은 상품



[조르지오 아르마니]립 마그넷 립 프리즈 스프링 셔벗-틴트 리미티드 컬렉션 3.9

판매가 46,000원  
참좋은예매가 43,700원

■ 내 남편의 뷰티



[조르지오 아르마니]립 마그넷 립 프리즈 스프링 셔벗-틴트 리미티드 컬렉션 3.9

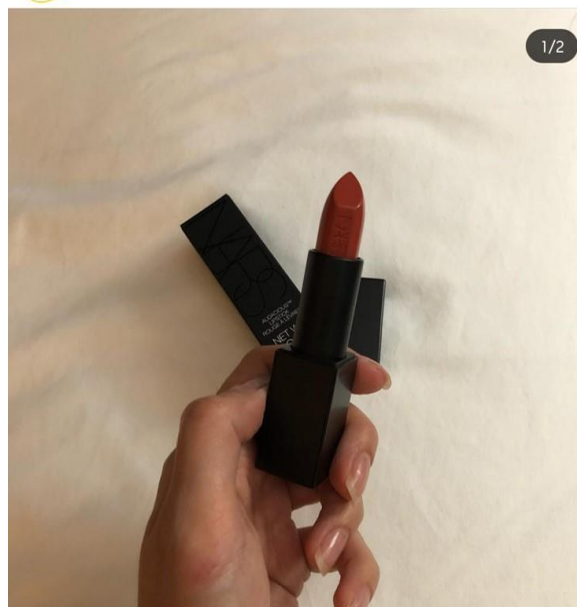
판매가 75,050원  
참좋은예매가 71,500원

**7% 추가 할인**

유효기간  
발행일로부터 1주일

#### 소비자의 지인도 타겟팅하자

- 특정 브랜드 선호도가 높은 매니아 고객의 특성을 활용하여 **지인 추천** 진행
- 고객의 **개인 SNS 계정**에 추천글을 올리고 인증하면 **구매액의 1.5%**를 적립금으로 지급(네이버페이 기준 : 평균 500원 추세보다 조금 높게 측정)



좋아요 50개  
엘포이 발색최고! 써보면 후회 안함  
가을에는 나스립스틱  
#나스 #나스립스틱 #제인 #나스제인 #절친이름도제인 #말린장미립스틱 #의도한전데

### Marketing

### 털털이 공략법

#### 저렴이를 활용하여 타겟팅하자

- 털털이는 **저렴한 가격대**를 선호하며, **깊은 고민 없이** 구매하는 성향이므로 2+1, 1+1 전략이 특효인 소비자임
- 연결 네트워크 상의 **공동 구매 제품**을 같이 추천하여, 타 제품 노출 및 쇼핑몰 내 탐색 시간 증대



#### 귀차니즘 소비자를 타겟팅하자

- 고민과 탐색, 비교가 적은 편인 소비자에게 **코디 페이지 제공**  
오프라인 매장에 DP된 옷을 **‘이대로 주세요’** 하는 소비자 특성을 **온라인에 적용**
- **트렌디한 연예인**을 모델로 활용하여 코디를 제안  
귀차니즘 특성을 가진 털털이 집중 공략

#### 고객님을 위한 ‘박보검이 입은’ 코디 제안



사진출처 : TNGT

### What 고객 가치 분류 및 고가치 고객 추출

#### 고가치 고객 정의

- 당사 온라인 몰에 높은 충성도와 수익성을 보이며, 재구매 가능성이 높을 것
- 분류기준 : 고객 수익성, 고객 충성도

고객 수익성 = 고객별 구매 금액(매출액)

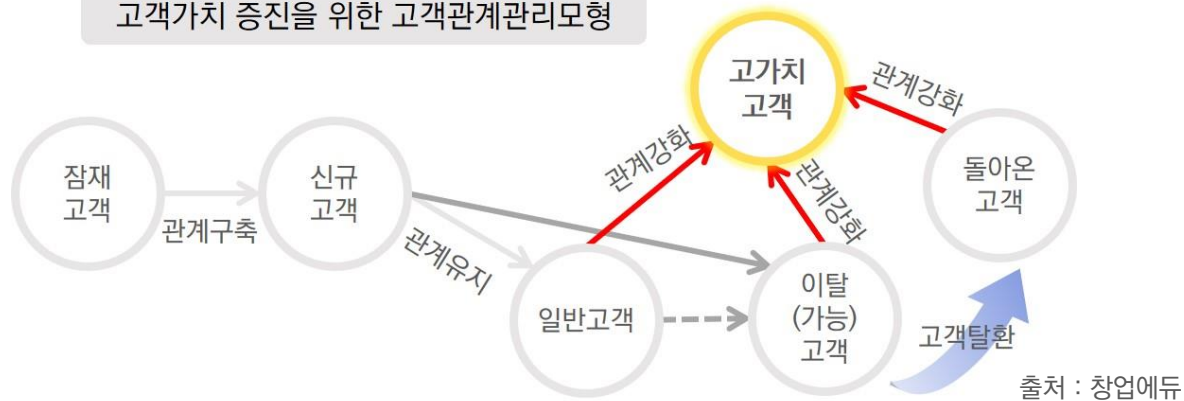
고객 충성도 = 고객별 구매 빈도, 고객별 최근 구매 여부

### Why 효율적 고객 관계 강화

고객 구매 패턴 분석을 통한 **고객 가치 서열화**

서열에 따른 맞춤형 관리방안 적용 및 **단계적인 등급 상승 가능**

고객가치 증진을 위한 고객관계관리모형



### How 고객 가치 분류 모델

#### 분석 모델 추진 방향

##### AS-IS

#### 단순 고객 세분화

- 소득, 연령, 재산 등에 의한 세분화
- 상품 구매에 의한 세분화
- 구매 금액에 의한 세분화

##### TO-BE

#### 고객 서열에 따른 세분화

- 최근성, 최빈성, 구매액에 의한 세분화
- 통계 데이터에 기반한 세분화
- LTV(Life Time Value)에 의한 세분화



### How 고객 가치 분류 모델

#### 분류 모델 사용 변수

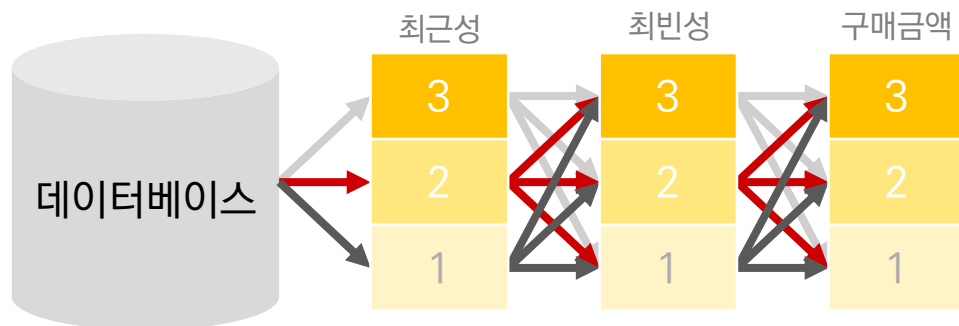
- 기존변수
  - CLNT\_ID : 고객 분류를 위한 ID
- 파생변수
  - MEAN\_DAY : 고객의 평균 구매일 간격  

$$(SESS\_DT \text{ 최대값}) - (SESS\_DT \text{ 최소값}) / (SESS\_DT \text{ 개수} - 1)$$
  - LAST\_MONTH : 고객이 마지막으로 구매한 달  

$$(SESS\_DT \text{에서 추출})$$
  - TOT\_PRICE : 고객이 자사 물에서 구매한 총 구매액  

$$\sum (PD\_BUY\_AM * PD\_BUY\_CT)$$

#### 최근성, 최빈성, 구매금액 채점 방법



### 고객 서열 요소별 분류기준

참고문헌을 바탕으로 자사 데이터에 맞게 재해석한 분류 기준 도출

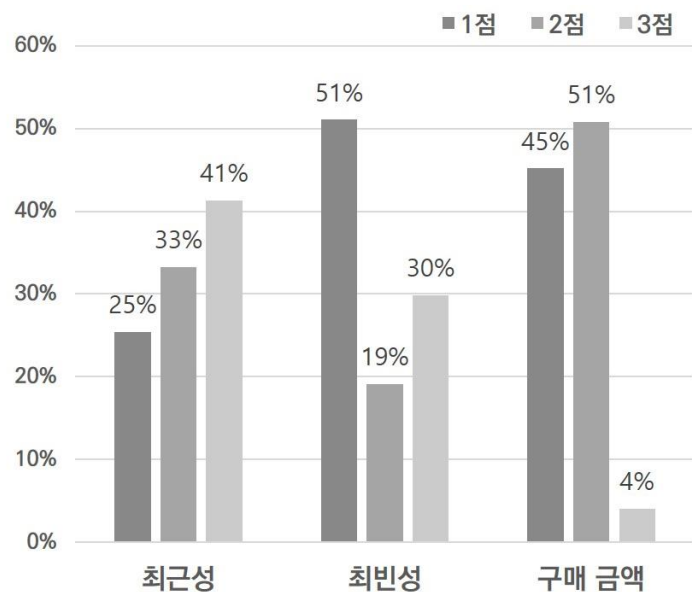
참고문헌 : 최근성, 최빈성, 기부액 기반의 기부자 리모델(2010, 김언옥)

구분	분류기준
최근성	3 : 최근 구매일이 8월 혹은 9월인 고객 2 : 최근 구매일이 6월 혹은 7월인 고객 1 : 최근 구매일이 4월 혹은 5월인 고객
최빈성	3 : 재구매 간격이 15일 이내인 고객 2 : 재구매 간격이 15일 초과인 고객 1 : 1회(하루) 구매 후 재구매가 없는 고객
구매금액	3 : 총 구매액 100만원 이상 2 : 총 구매액 100만원 미만, 10만원 이상 1 : 총 구매액 10만원 미만



### How 고객 가치 분류 모델

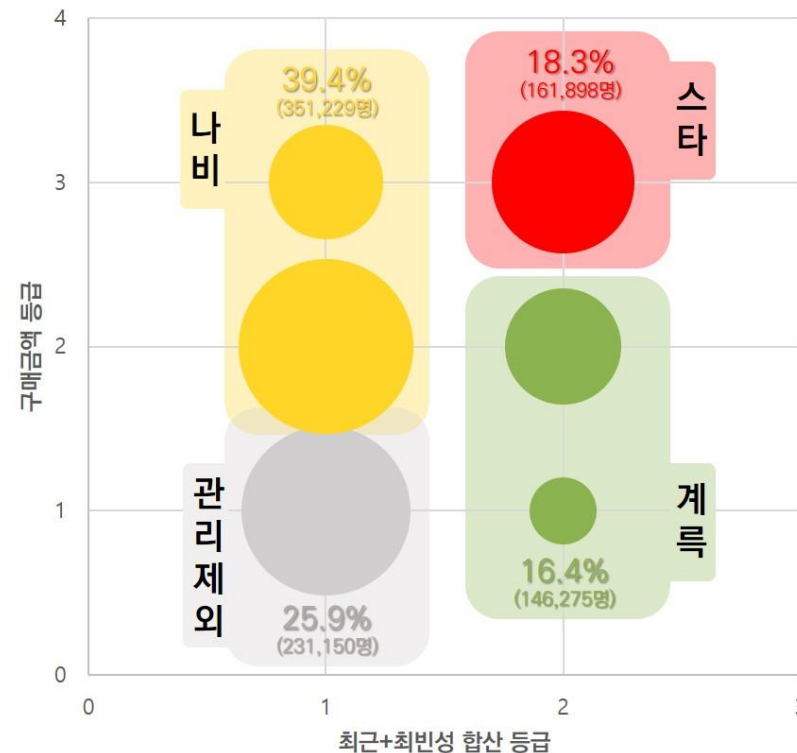
#### 고객 서열 요소별 인원 분포



- 최근성** 최근 2달 이내 구매 고객이 가장 많으며 고루 분포
- 최빈성** 재구매가 없는 고객이 대체로 많으나, 재구매 경험이 있는 고객의 수도 적지 않은 추세임
- 구매금액** 100만원 미만 구매 고객이 전체의 96%를 차지, 100만원 이상 구매고객은 4% 수준

#### 고객 서열 군집화

고객 서열별 분류 분포



- Y축** 구매금액 등급 1, 2, 3순
- X축** 등급 1(최근성 + 최빈성  $\leq 4$ )  
등급 2(최근성 + 최빈성  $> 4$ )

# 02 고객 및 상품군 분석

## 고객 가치 분류별 마케팅 전략

18

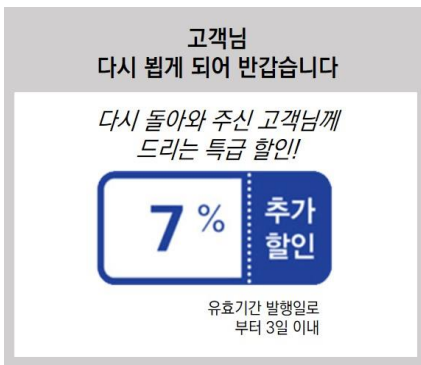
### 나비

당사에 많은 관심을 보이지만,  
큰 금액을 소비하는 소비자

■ 나비 등급 고객은 최근성과 최빈성을 높여 **스타 등급으로 조정**해야 함

■ 재 로그인 기록이 발생하면,  
기한이 짧은 쿠폰 발송으로 구매 촉진 진행

■ 재 로그인 기록이 발생하지 않는다면,  
기존 구매한 제품 정보를 담은 **메일 발송**으로  
로그인 유도



LOTTE.COM

안녕하세요, 롯데닷컴입니다.

**기존 구매 고객님의께만 드리는 할인 정보**

구매하신 제품의 할인 정보가 있어  
안내드립니다.



네퓨어/크리넥스/모나리자/코디 커피

### 스타

당사에 많은 관심을 보이고,  
큰 금액을 소비하는 소비자

■ 스타 등급 고객은 당사의 **고가치 고객**으로 특급 프리미엄 전략 필요

■ 매분기 스타 고객 중 **1,000명**을 뽑아

**감사패키지 발송** 이벤트 진행

고객이 구매한 상품을 위주로 패키지 구성

■ 구매액 대비 **0.8% 수준**으로 포인트 지급  
(L.point 기준 롯데 쇼핑 계열  
포인트 적립 0.1% ~ 0.5% 수준)



스타	나비	계류	관리제외
0.8%	0.5%	0.3%	0.1%



### 관리제외

당사에 많은 관심을 보이지 않고,  
적은 금액을 소비하는 주로 일회성 소비자

- 관리제외 등급 고객은 일회성 구매가 대다수이므로 마케팅을 최소화하고,  
**타 고객군에 선택과 집중 필요**
- 첫 구매 고객의 경우,  
첫인상에서 엄청난 고객 가치를 제공해야 재구매 가능성이  
증대되므로 **웰컴 기프트를 제시**(ex. 화장품 샘플)
- 첫 구매 고객이 아닌 경우,  
별도의 쿠폰을 제시하지 않고 **회원 등급제 화면으로 자동 링크 진행**  
계류 단계 조건으로 유도 후, 유입되면 마케팅 서비스 제공, 유입되지  
않으면 관리 대상에서 제외

### 계류



당사에 많은 관심을 보이지만,  
적은 금액을 소비하는 소비자

- 계류 등급 고객은 구매 금액을 높여 **스타 등급으로 조정**해야 함
- 스타 등급으로 유입 진행. 고객의 현 구매현황에서 구매액만 만족하면  
등급 상향이 가능하다는 것을 적극 강조
- 계류 등급 고객이 로그인하면, 첫 **랜딩페이지**에 등급 조건을 안내 후  
'BIG딜' 페이지로 자동 연결. 할인가에 약한 고객 특성과 구매 금액 충족  
필요 상황이 맞물려 **구매율 향상 기대**

안녕하세요. 엘포이 고객님의.  
이번 달 구매액에서 00만원만 채우시면  
프리미엄 고객 등급으로 상향 조정됩니다.

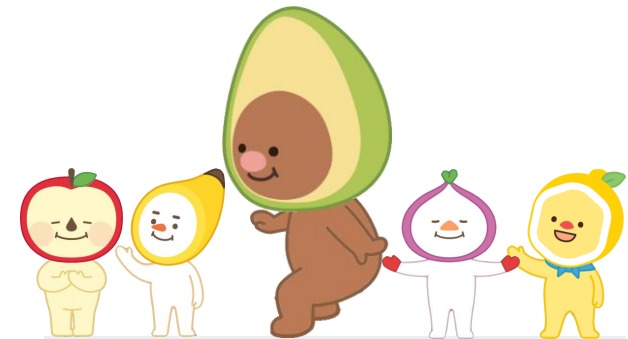
프리미엄 고객이 되시면 롯데에서 제공하는  
프리미엄 고객서비스를 받으실 수 있습니다.

LOTTE.COM

.....

# 03 상품군별 수요트렌드 예측 및 인사이트 도출

- 수요트렌드 개요 및 전처리
- 수요트렌드 예측 구현



### Idea 수요에 영향을 미치는 시계열 요인 탐색

#### 분석 준비

- 시간의 흐름에 따라 변화하면서도, 매출액에 영향을 미치는 요인 탐색
- 날씨, 휴일 등의 데이터가 온라인 쇼핑몰 매출에 영향이 있음을 확인, 수집 가능한 모든 데이터 탐색

#### 데이터 수집

##### 네이버 캘린더

- 데이터 기간(4~9월) 내 공휴일
- 주 타겟 고객층(직장인) 고려, 휴무일 추가 수집(근로자의 날 등)

날짜	구분	날짜	구분
2018-05-01	근로자의날	2018-08-15	광복절
2018-05-05	어린이날	2018-09-23	추석
2018-05-07	대체휴일	2018-09-24	추석
2018-05-22	석가탄신일	2018-09-25	추석
2018-06-06	현충일	2018-09-26	대체휴일
2018-06-13	지방선거		

##### 기상청

- 날씨누리 - 관측자료 - 과거자료
- 데이터 기간(4~9월)내 평균기온, 일 강수량 수집

지점 | 서울(유) | 선택
년도 | 2018 | 선택
월 | 4 | 선택
요소 | 기온/강수량 | 선택

날씨달력 서울(유) / 2018년 4월

일요일	월요일	화요일	수요일	목요일	금요일	토요일
1일	2일	3일	4일	5일	6일	7일
평균기온:15.6℃ 최고기온:18.5℃ 최저기온:11.7℃ 평균운량:9.1 일강수량: -	평균기온:18.4℃ 최고기온:23.7℃ 최저기온:14.9℃ 평균운량:6.1 일강수량: -	평균기온:17.8℃ 최고기온:21.2℃ 최저기온:14.9℃ 평균운량:9.1 일강수량:0.1mm	평균기온:13.4℃ 최고기온:18.2℃ 최저기온:9.5℃ 평균운량:6.1 일강수량:18.5mm	평균기온:7.6℃ 최고기온:9.5℃ 최저기온:6.7℃ 평균운량:9.9 일강수량:10.5mm	평균기온:5.8℃ 최고기온:8.6℃ 최저기온:3.0℃ 평균운량:8.9 일강수량:6.5mm	평균기온:4.2℃ 최고기온:8.6℃ 최저기온:1.4℃ 평균운량:5.1 일강수량:0.2mm

### Idea

### 요일과 매출의 상관관계?



※ 여성의류 상품군의 18. 04. 14. ~ 18. 05. 06.의 매출액(3주간)

- 대개 주말과 월요일의 매출액이 높고, 금요일의 매출액이 낮음
- 요일에 따른 소비 패턴이 존재한다고 가정
- 따라서 각각의 요일이 하나의 계절이고 계절의 수는 7로 설정함
- 이 패턴을 수치화하여 계절 지수를 생성

### Method

### 계절변동 분해법(Seasonal Variation)

- 성수기/비수기 혹은 잘 팔리는/잘 안 팔리는 시기 등과 같이 주기적으로 반복하여 나타나는 경우에 이러한 시계열 자료는 “계절변동이 있다” 라고 판단

기간	날짜	요일	매출액 (Y)	1 7개항 중심이동 평균 (CMA)	2 특정계절지수 (Y/CMA)
1	2018-04-01	일요일	238,179,870	198,390,755	1.08649
	2018-04-02	월요일	241,964,178		
	2018-04-03	화요일	218,457,560		
	2018-04-04	수요일	215,549,465		
	2018-04-05	목요일	184,853,040		
	2018-04-06	금요일	156,493,150		
	2018-04-07	토요일	133,238,020		
2	2018-04-08	일요일	156,956,595	168,931,075	0.92912

요일구분	3 특정계절지수 평균	계절지수
일요일	1.070	1.0727
월요일	1.086	1.08882
화요일	0.989	0.99089
수요일	1.019	1.02074
목요일	0.954	0.95626
금요일	0.923	0.92538
토요일	0.943	0.9452
총합계	6.985	7

1 계절의 크기가 7이므로 7개의 매출액 평균 값을 구함

2 특정계절지수 = 매출액(Y)/중심이동평균(CMA)

3 계절지수 = ((계절의 크기/계절별 총 평균의 합) \* 특정계절지수 평균값)

### Method RNN LSTM(Long Short Term Memory)

순환신경망의 일종으로 **Gradient Vanishing Problem**을 해결하고자 메모리를 도입한 모델  
LSTM은 시계열 데이터 분석에 있어 **기존 알고리즘보다 좋은 성능**을 낸다고 알려져 있음

### Strategy 모델링 전략

4  
April

일	월	화	수	목	금	토
26	27	28	29	30	31	1
2	3	4	5	6	7	8

#### 일주일 단위로 다음 날 매출을 예측

- 입력 : 직전 일주일의 매출 유관 데이터
- 출력 : 익일의 매출액 예측값

Iteration1 : 4/1~4/7 데이터로 4/8의 매출액 예측  
Iteration2 : 4/2~4/8 데이터로 4/9의 매출액 예측  
⋮

#### 학습 과정

- 직전 일주일을 학습하여 하루씩 예측하여 실제 값과 비교하여 정확도를 확인함
- 18년 4월 1일 부터 18년 9월 30일 까지(총 176개의 샘플)

### Input 매출 유관 데이터

사용 입력 속성					Loss율
매출액	강수량	평균기온	공휴일	계절지수	
					1.01%
					0.88%
					0.84%
					0.77%
					0.74%
					0.57%
					0.47%
					0.21%

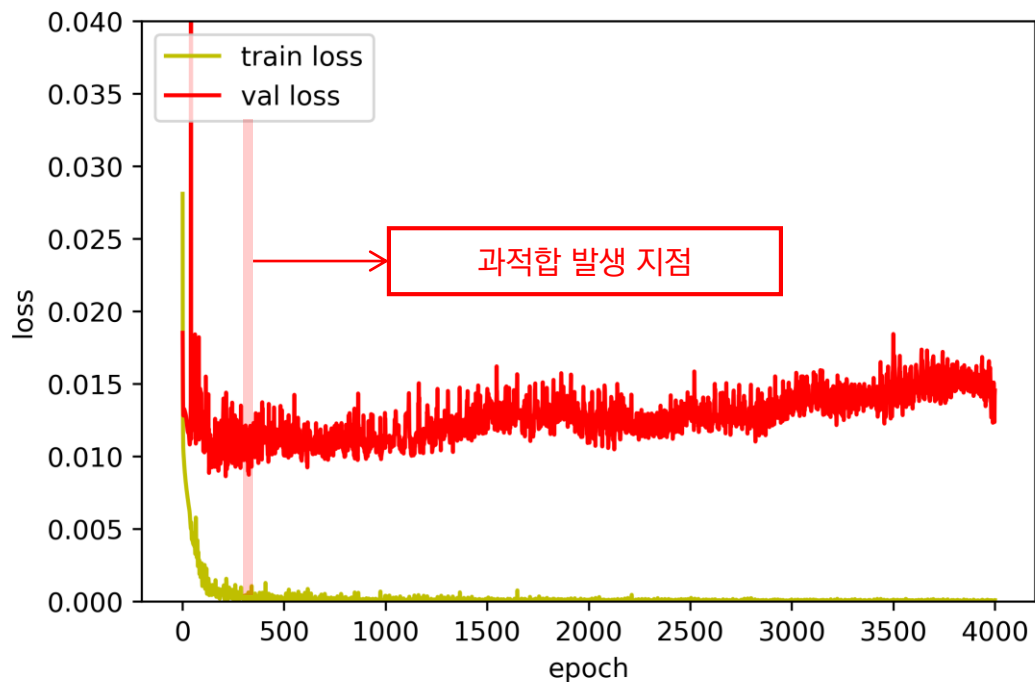
#### 데이터 조합 선정

- 매출액을 기본으로 두고 입력 속성을 하나씩 추가하여 오차가 가장 적은 조합을 선택
- 입력으로 속성을 추가할수록 Loss율이 줄어드는 것을 확인
- 따라서, 모델 입력으로 매출액, 계절지수, 공휴일, 평균기온, 강수량 모두 사용함

※ 입력 속성의 유효성을 빨리 확인하기 위해 Epoch를 150으로 설정하여 진행

### Modeling

### Epoch값 설정



※ Train loss : 훈련데이터를 학습하며 발생하는 Loss (훈련 손실 값)

Val loss : 검증 과정 중 발생하는 Loss (검증 손실 값)

### Epoch란?

- 반복 학습의 횟수를 Epoch라고 함
- 많이 학습할수록 Training set에 대한 손실 값이 줄고 정확도가 올라감
- 하지만, Epoch가 필요 이상으로 커지게 되면 과적합이 발생하여 새로운 data set의 손실 값이 커지게 됨 -> 적정 수준의 Epoch값을 찾아야 함

### 적정 수준의 Epoch값 설정

- 훈련 데이터와 검증 데이터를 랜덤으로 7:3 비율로 나누고 Epoch에 따른 각각의 loss 값을 확인
  - Val loss가 감소하다가 다시 증가하기 시작하는 지점이 과적합이 발생하는 지점임
  - 우리가 가진 data set에서는 대략 300 정도에서 과적합이 발생함
- 따라서 Epoch값을 300으로 설정하고 예측을 진행함



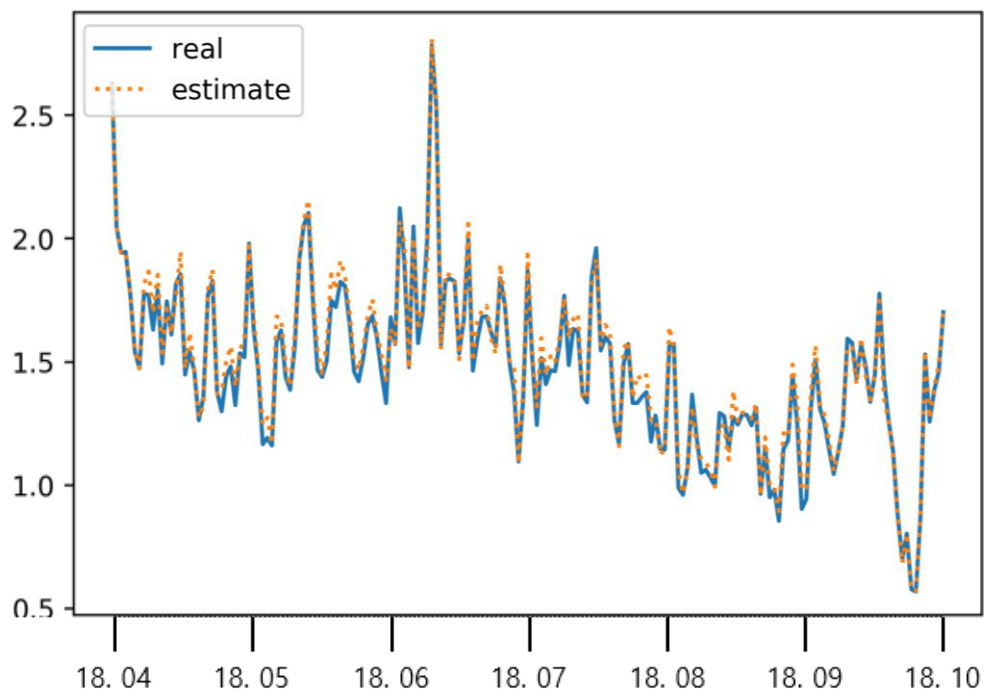
### Modeling

### 1Step 예측 모델

■ 일주일 단위로 다음 날 매출을 예측

- 입력 : 직전 일주일의 매출 유관 데이터
- 출력 : 익일의 매출액 예측 값

여성의류 매출액



### 분석 조건

- 상품군 : 여성의류
- 입력 속성 : 매출액, 강수량, 계절지수, 공휴일, 평균기온
- Epoch : 300

### 분석 결과

- Loss율 : 0.05%
- RMSE : 4,201,454.17
- “ 평균 에러가 약 420만원 ”
- 여성의류 상품군의 하루 평균 매출(약 1억 5천만원)을 감안하면 상당히 높은 예측력을 보임을 알 수 있음

### 분석 결과의 활용

- 과거 자료를 활용하여 익일의 대략적인 매출액을 예측하고 대비할 수 있음
- 그러나, 단기간(하루)의 데이터만 예측 가능하다는 한계가 있음

예측 기간이 확장 되도록 Model modifying

### Idea 예측 기간을 확장하려면?

#### 모델 입력 속성 수정

예측 해야 하는 값 & 입력으로 사용할 값

구분	매출액	강수량	평균기온	공휴일	계절지수
자료 수집	불가	기상예보	기상예보	고정	고정 (주기적 변동)

일	월	화	수	목	금	토
26	27	28	29	30	31	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15

일	월	화	수	목	금	토
26	27	28	29	30	31	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15

8일 예측 매출액이 9일의 입력으로 사용됨

#### 모델 확장

- 1 step 예측 모델로는 익일의 매출액만 예측 가능하므로 **Full step 예측 모델**을 사용
- 다만 Full step 예측은 오류가 누적되므로 **짧은 기간의 트렌드**를 확인하는 것이 적절함

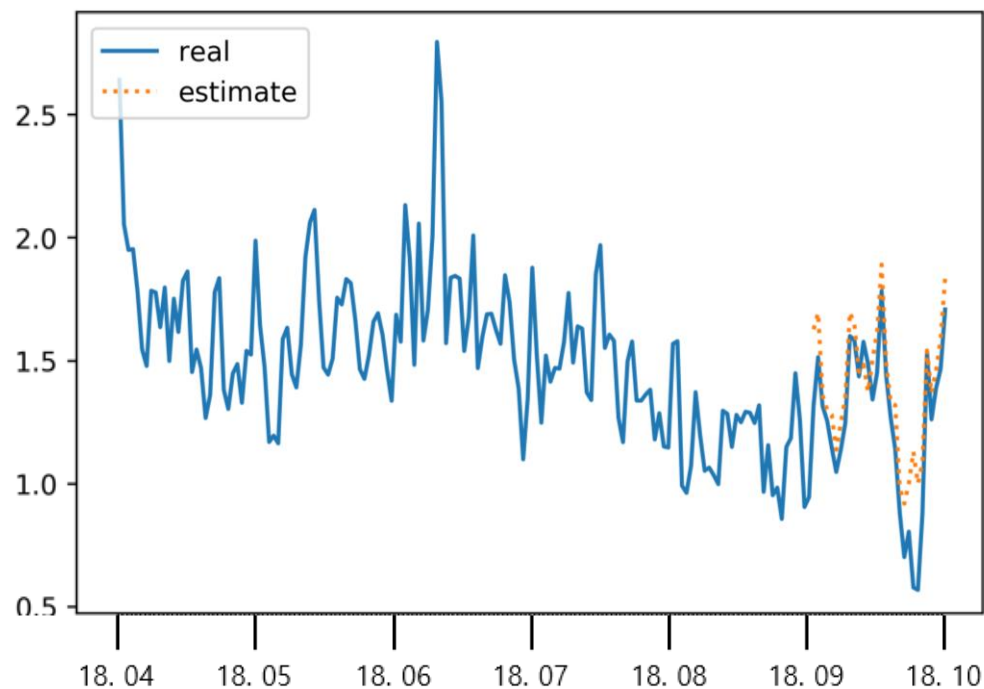
### Modify

### Full Step 예측 모델

직전 학습데이터로 주 단위, 월 단위의 매출을 예측

- 입력 : 예측 월(주) 직전까지의 매출 유관 데이터
- 출력 : 익월(차주)의 매출액 예측 값

여성의류 매출액



### 분석 조건

- 상품군 : 여성의류
- 입력 속성 : 예측한 매출액, 강수량, 계절지수, 공휴일, 평균기온
- Epoch : 300

### 분석 결과

- RMSE : 16,261,713.16
- “ 평균 에러가 약 1,600만원 ”
- 일 단위의 예측에서 주 단위, 월 단위로 확장하여도 높은 예측력을 유지

일, 주, 월 단위 모델의 RMSE (여성의류 외)

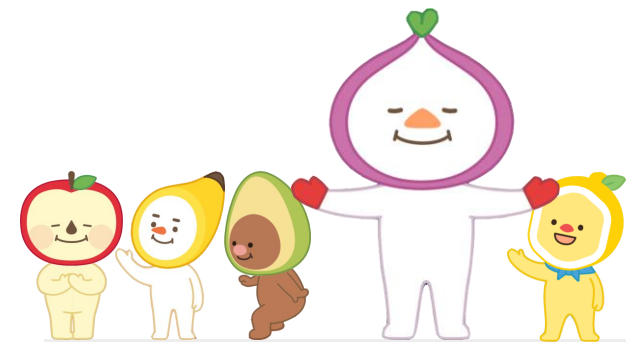
(단위 : 천원)

구분	화장품 /뷰티케어	스포츠 패션	남성의류	패션잡화	속옷/양말 /홈웨어	아동의류	시즌스포츠	출산 /육아용품	퍼스널 케어	건강식품
Loss	0.10%	0.04%	0.33%	0.04%	0.05%	0.08%	0.04%	0.06%	0.11%	0.04%
RMSE	1day	6,152	3,394	6,816	5,857	1,143	2,790	1,707	1,049	3,969
	1week	9,073	10,779	14,324	4,551	2,911	16,271	3,829	2,642	7,772
	1month	13,198	7,673	13,022	4,363	3,468	19,377	11,081	2,055	7,727



## 04 주요 상품군별 온라인 선호지수 생성

- 선호지수 개요 및 모델링
- 선호지수 모델 개선



### Idea 선호지수란?

#### 선호지수 정의

- 선호지수란 고객이 특정 상품군에 관심이 있는 정도(주로 상대적 개념)

#### 선호지수 모델링

- Step1** 선호지수를 해당 상품군을 구매할 확률이라 정의
- Step2** 구매를 했을 때, 특정 제품군을 구매할 확률을  $P(A)$ 라 가정
- Step3**  $P(A)$ 는 사후 확률 개념과 동일하므로 베이지안 확률 활용

### Method 베이지안 확률

#### 베이즈 정리

사전확률로부터 사후확률 도출 가능  
사후확률은 사전확률과 가능도의 곱에 비례

$$P(H|D) = \frac{P(D|H) P(H)}{P(D)}$$

사후확률      가능성도      사전확률  
 ↓                    ↓                    ↓  
 $P(H|D)$        $P(D|H)$        $P(H)$

#### 베이즈 정리를 모델에 적용

##### 고객 1의 구매 기록

여성의류

남성의류

- 사전확률 : 고객 1의 구매기록
- 설명 편의를 위해 2가지 상품군만 고려

사전확률

$P(\text{여성의류})$  : 고객1이 여성의류를 구매할 확률  
 $P(\text{남성의류})$  : 고객1이 남성의류를 구매할 확률

사후확률

$P(\text{여성의류} | \text{구매})$  : 고객 1이 구매를 했을 때 여성의류일 확률  
 $P(\text{남성의류} | \text{구매})$  : 고객 1이 구매를 했을 때 남성의류일 확률  
 $P(\text{여성의류} | \text{구매}) = P(\text{여성의류}) * P(\text{구매} | \text{여성의류}) / P(\text{구매})$   
 $P(\text{남성의류} | \text{구매}) = P(\text{남성의류}) * P(\text{구매} | \text{남성의류}) / P(\text{구매})$

가능도

$P(\text{구매} | \text{해당 상품군})$  : 해당 상품군을 고객1이 구매할 가능성

### 가능도 수정

#### ■ 고객 1의 특성을 반영하여 가능도 확장

- 특정 상품군의 전체 구매기록 중 고객 1 개인의 구매기록은 매우 경미
- 고객 1의 특성(연령대, 성별) 추가하여 가능도를 수정

$$\text{수정가능도} = P(20\text{대} \mid \text{해당 상품군}) * P(\text{남성} \mid \text{해당 상품군})$$

- 해당 상품군을 20대 남성이 구매한 확률

(연령대와 성별은 독립이므로 확률을 서로 곱함)

### 사후확률 계산식 간소화

#### ■ 목적

다수 상품군 중 어떤 상품군을 구매할 확률이 상대적으로 높은 지 확인

$$P(\text{여성의류} \mid \text{구매}) = \frac{P(\text{여성의류}) * P(20\text{대} \mid \text{여성의류}) * P(\text{남성} \mid \text{여성의류})}{P(\text{구매})}$$

$$P(\text{남성의류} \mid \text{구매}) = \frac{P(\text{남성의류}) * P(20\text{대} \mid \text{남성의류}) * P(\text{남성} \mid \text{남성의류})}{P(\text{구매})}$$

(분모는 동일하므로 계산 제외)

### 예시 적용

#### ■ 가정

고객 1은 10번의 구매 기록이 있고, 그 중 7번은 남성의류 3번은 여성의류를 구매

고객 1의 구매기록	여성의류 30% / 남성의류 70%
여성의류 연령대별 구매 비율	20대 40% / 30대 60%
	남 20% / 여 80%
남성의류 연령대별 구매 비율	20대 50% / 30대 50%
	남 50% / 여 50%

※ 설명 편의를 위해 연령대는 20대와 30대의 구매기록만 있다고 가정

- 사전확률 :  $P(\text{여성의류}) = 0.3$ ,  $P(\text{남성의류}) = 0.7$
- 사후확률 :  $P(\text{여성의류} \mid \text{구매}) = P(\text{여성의류}) * P(20\text{대} \mid \text{여성의류}) * P(\text{남성} \mid \text{여성의류})$   
 $= 0.3 * 0.4 * 0.2 = 0.024$   
 $P(\text{남성의류} \mid \text{구매}) = P(\text{남성의류}) * P(20\text{대} \mid \text{남성의류}) * P(\text{남성} \mid \text{남성의류})$   
 $= 0.7 * 0.5 * 0.5 = 0.175$
- $P(\text{여성의류} \mid \text{구매}) < P(\text{남성의류} \mid \text{구매})$  이므로,  
 고객1은 남성의류를 구매할 확률이 더 큼(여성의류보다 남성의류를 더 선호)

### 모델의 문제점

구매 기록이 없는 고객일 경우에 사전확률이 0이 되면 사후확률도 0이 돼서 선호지수를 구할 수 없음

또한 확률은 곱할수록 값이 0에 수렴하고 소수점 이하의 숫자라서 한눈에 알아보기 어려운 면이 있어 모델의 개선이 필요함

### 구매 기록이 없는 고객일 경우에는?

#### ■ 사전확률 계산식 수정

- 모든 상품군 구매 건수에 1을 더하여 계산

#### ■ 사전확률 계산식 수정 예시

구분	여성의류	남성의류	유아동의류	합계
구매건수	0	1	0	1
사전확률	0	1	0	1

+1

구분	여성의류	남성의류	유아동의류	합계
구매건수	1	2	1	4
사전확률	0.25	0.5	0.25	1

### 확률을 곱할수록 작아지는 문제?

#### ■ 사후확률에 자연로그를 취함

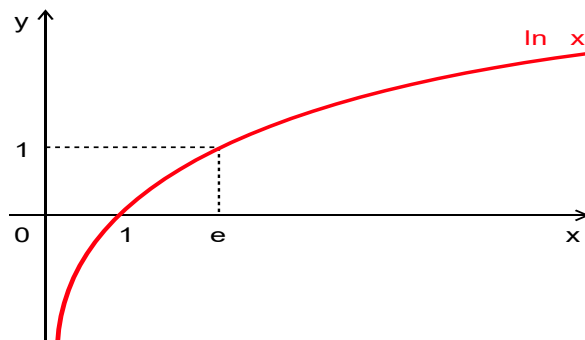
$$P(\text{여성의류} \mid \text{구매}) = 0.024$$

$$P(\text{남성의류} \mid \text{구매}) = 0.175$$

log

$$\text{LN}(P(\text{여성의류} \mid \text{구매})) = -3.73$$

$$\text{LN}(P(\text{남성의류} \mid \text{구매})) = -1.74$$



### 0~100 사이의 숫자로 스케일링

- 선호지수가 음수로 산출되면 직관적이지 않음
- 따라서, 최대값이 100이 되게 스케일링
- 로그를 취한 사후확률의 최소값이 5가 되는 x를 더함

상품군	선호지수	상품군	선호지수
여성의류	-3.73	여성의류	5
남성의류	-1.74	남성의류	6.99

+x

÷ MAX

상품군	선호지수	상품군	선호지수
여성의류	71.53	여성의류	0.71530
남성의류	100	남성의류	1

× 100



## 05 신규 서비스 제안

- 고객행동기반 개인 맞춤형 서비스 “ Customer Profiling ”
- 온오프라인 통합을 위한 “ Real Smart Store ”





## Idea 고객 프로파일링

### 서비스 정의

#### ■ 고객 프로파일링이란?

고객 개개인 맞춤형 경험 제공과 불규칙한 수요 시간패턴 조율이 가능하도록  
고객에 대한 모든 정보와 마케팅 전략을 한 눈에 보여주는 서비스

### 서비스 플로우(예시)

해당 서비스는 앞선 분석에서 도출된 행동 패턴별 군집과 고객 가치별 모든  
서열에 분류 적용 가능하나, 설명 편의를 위해 가상의 특정 고객을 설정함

#### Step1

수요 트렌드 예측 모델에서 도출된 특정 상품군의  
수요 급감 포인트 분석

#### Step2

선호지수 기반 타겟팅 고객 선정

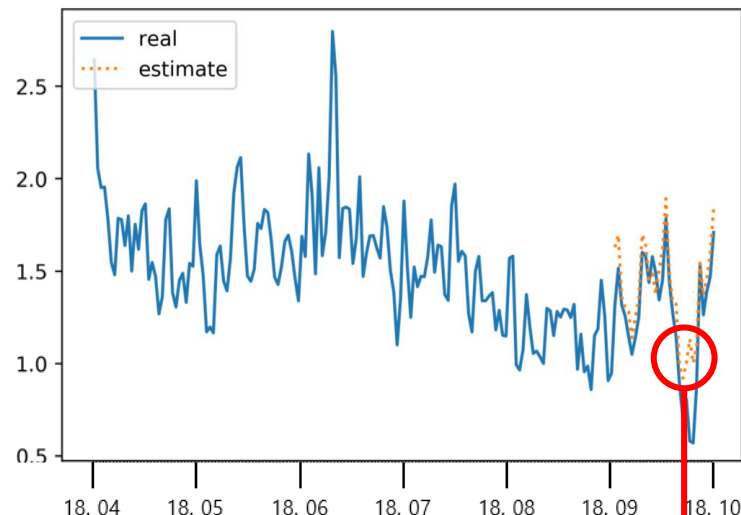
#### Step3

선정된 고객의 프로파일을 확인  
행동 패턴별 군집과 고객 가치 분류에 따른 마케팅 전략 실행

## Step1

## 수요 급감 포인트 분석

여성의류 매출액



- 여성의류 상품군 수요 예측  
모델에 의하면 9.23 ~ 9.25  
수요가 급감할 것으로 예측됨
- 불규칙 수요 조절을 위해 여성의류  
선호지수가 1위인 고객 명단 추출

## Step2

## 선호지수 기반 타겟팅 고객 선정


#### ■ 고객 리스트 중 여성의류 선호지수가 1순위인 고객 선정

선호지수	엘포이	오니	아보	애피	바나미
1순위	여성의류	화장품 /뷰티케어	스포츠패션	여성의류	여성의류
2순위	남성의류	여성의류	패션잡화	패션잡화	건강식품

### Step3

### 선정된 고객의 프로파일 확인

#### Customer Profile



고객ID  
4140076 x

2018.04.01 ~ 2018.09.30

연령대  
40

총 구매 금액  
3,867,200원

333  
고객가치

(덜)꼼꼼이  
행동패턴

#### 페르소나

(덜)꼼꼼이 행동특성(1101)  
제품을 구매하는 데 걸리는 시간이 길고, 탐색 및 비교를 많이 함. 가격대는 높은 편

# 뚜렷한\_접속\_목적\_없음

# 습관적\_접속

# 리뷰\_집중

# 취미는\_아이쇼핑

# 상세페이지\_꼼꼼확인

# 이벤트\_선호

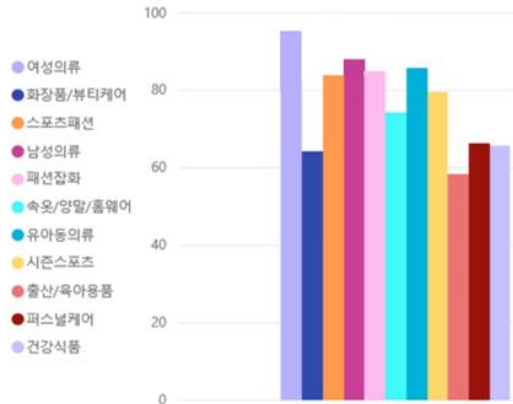
# 쿠폰\_탐색\_선호

# 비교\_중독

# 취향저격시\_가격은\_거들뿐


+ 태그 추가하기

#### 선택지수

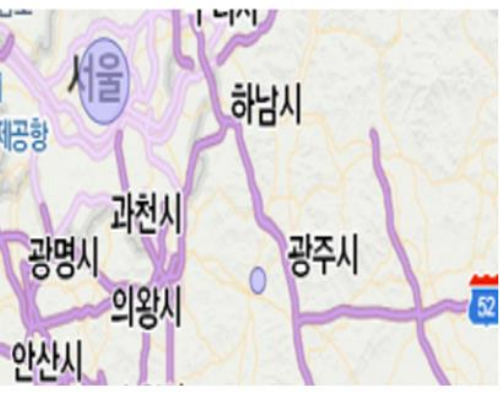


- 여성의류
- 화장품/뷰티케어
- 스포츠패션
- 남성의류
- 패션잡화
- 숙소/양말/홈웨어
- 유아동의류
- 시즌스포츠
- 출산/육아용품
- 퍼스널케어
- 건강식품

#### 검색어



#### 접속지역



필요시 태그를 추가/삭제하여  
Customizing 가능

고객 가치 분류 서열,  
행동 패턴별 군집 정보 제공

### Step3

### 마케팅 전략 실행

고객 정보와 마케팅 전략 내용을 한 눈에 볼 수 있도록 구성

#### Customer Profile



고객ID

4140076 x

2018-04-01

2018-09-30



여성

40

연령대

3,867,200원

총 구매 금액

333

고객가치

(월)꼼꼼이

행동패턴

#### 고객가치 분류 공략법

#### “비교쟁이 소비자”

비교를 좋아하는 소비자 특성을 활용하여 장바구니에 제품 비교 기능 추가

나와 동일 패턴의 고객들이 구매한 제품을 한 눈에 볼 수 있게 비교 페이지 제공

연결 네트워크 상의 제품을 같이 추천하여 거부감 없이 자연스러운 노출 진행



#### 행동 패턴별 공략법

#### “스 타”

스타는 당사의 고가치고객으로 특급 프리미엄 전략 필요

매분기 스타 고객 중 1,000명을 뽑아 감사패키지 발송 이벤트 진행.

고객이 구매한 상품을 위주로 패키지 구성

구매액 대비 0.8% 수준으로 포인트 지급



### Idea Real Smart Store 구성

#### Real Smart Store 정의

Real Smart Store란, ‘고객이 말하는 건 무엇이든, 고객이 말하지 않는 것도 찾아서 알려주는 스토어’임



#### 롯데 ‘Smart Store’ 확장

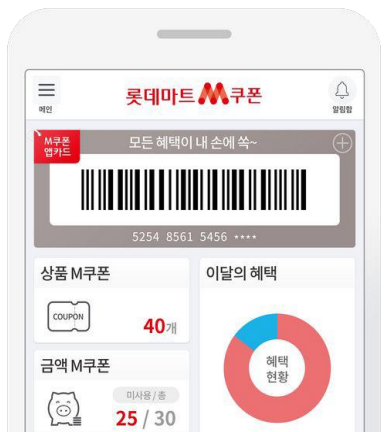
최근 오픈한 ‘롯데 인천터미널점 Smart Store’를 벤치마킹 및 온오프라인 고객 정보 통합 방향으로 확장

##### ▪ M쿠폰앱 바코드 스캔

매장 입구에 위치한 스캐너에 M쿠폰앱 바코드를 스캔하면 고객 맞춤 쿠폰 전송

##### ▪ QR 코드 스캔 구매

매장 내 진열 상품의 QR코드를 스캔하면, 온라인 어플이 연동되면서 온오프라인에서 동일 가격으로 구입 가능



### Method 온오프라인 통합 정보 공유 진행

#### 고객 정보 정의 및 저장 방법

##### 고객 요구사항

##### ▪ 정의

고객이 직접 말하는 정보

##### ▪ 저장 방법

고객 가입 시 고객의 성향과 요구사항을 설문지 형식으로 조사 후 저장

〈고객 성향 분석 설문지 예시〉

1. 당사 물에서 가장 즐겨 구매하는 제품군은 무엇인가요?  
① 의류 ② 화장품 ③ 생활용품 ④ 건강식품
2. 그 이유는 무엇인가요?  
① 저렴해서 ② 편리해서 ③ 필수품이라서 ④ 종류가 다양해서
3. 가장 선호하는 구매 형태는 무엇인가요?  
① 온라인 ② 오프라인 ③ 오프라인으로 확인 후 온라인 구매  
④ 온라인으로 확인 후 오프라인 구매종류가 다양해서

LOTTE.COM

##### 온오프라인 행동 정보

##### ▪ 정의

고객이 말하지 않는 정보로 수집 및 분석 필요

##### ▪ 저장 방법

온라인 : 고객 프로파일링 정보를 활용

오프라인 : 고객 응대 후 판매원이 직접 정보 입력

- 입력정보 : 판매물품, 가격, 고객 고려시간, 특이사항 등

## Scenario

## 온오프라인 통합 고객 관리



고객이 말하는 건 무엇이든, 그리고 말하지 않는 것까지 ‘**LOTTE REAL SMART STORE**’



감사합니다