

온라인 고객 대상 맞춤형 모델 개발을 통 한 매출액 증대 방안

A반 4조(책과콩나물)

강동호 김애니 손필언 엄송근 유정근 한유리

01. 추진배경

02. 현황 및 개선기회

03. 분석계획

04. 분석결과

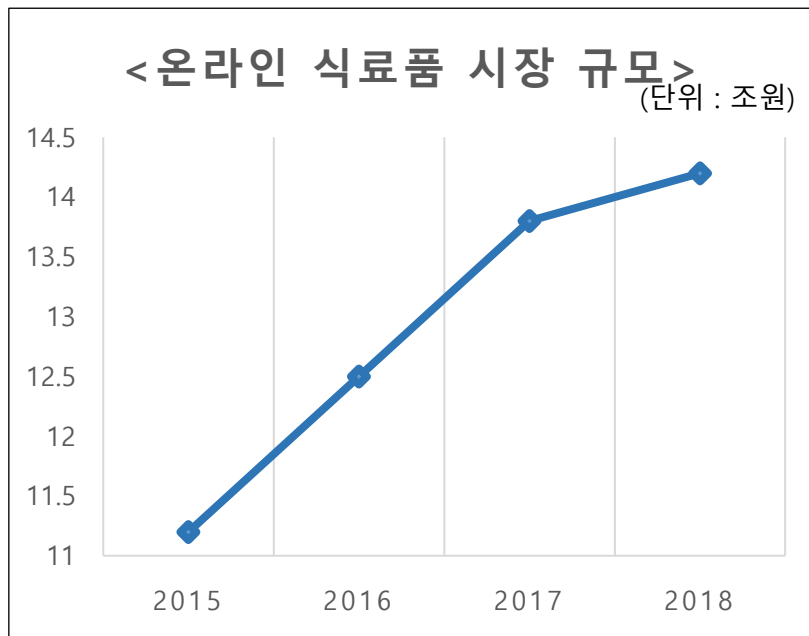
05. 개선안 및 적용방안

06. 소감

▶ 1. 추진배경

온라인 식품 시장 규모 확대에도 불구하고 대형 경쟁사의 시장 참여로 기존 고객이 이탈해 자사 매출액 감소

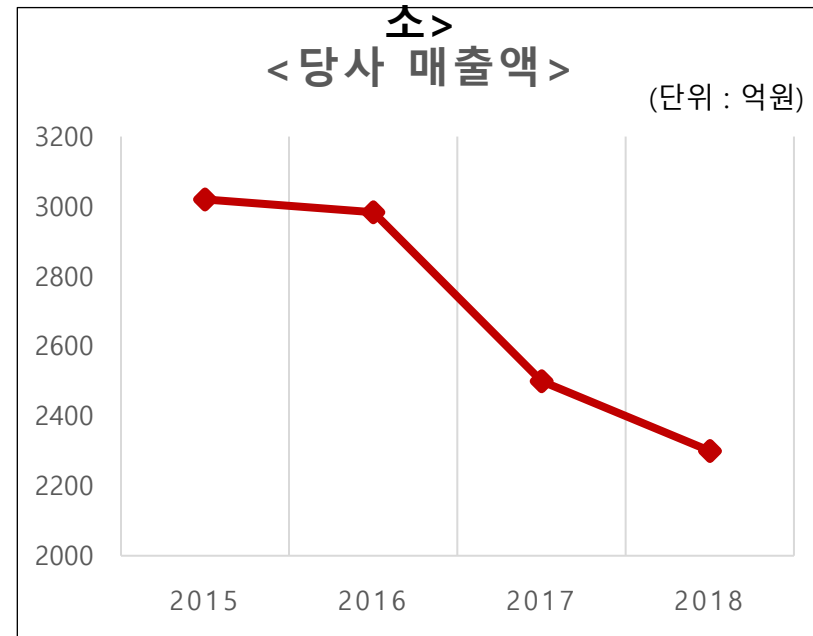
<온라인 식료품 시장 4년간 3조 6천억 원 증가>



통계청(2018)

- 소규모 가구 증가, 이용의 편의성, 저렴한 가격 등으로 **온라인 식료품 시장 성장**

<자사 매출액 4년간 720억 원 감소>



- 옴니채널을 가진 대형 경쟁사의 시장 참여와 충성도가 낮은 온라인몰 고객으로 **경쟁 심화로 매출 감소**

기존 고객 이탈 방지와 신규 고객 유치를 위한 판매 전략 수립 필요

▶ 2. 현상 및 개선기회

제품 재주문율 현 59%에서 70%로 향상

<당사의 고객관리와 데이터 활용 부족>

<경쟁사의 차별화된 마케팅과 시스템>

단순한 제품구매
시스템

고객에게 맞춤형 제품 추천 불가능



S사 고객 및 주문 특성을 활용한 추천 시스템
사용

“경쟁사들과 차별화할 수 있는 **추천 시스템 도입**”

고객 데이터 활용
미흡

고객 주문 특성과 상품 판매 특성을
알 수 없음



L사 고객과 제품 데이터를 활용한 할인
프로모션 진행

“고객의 구매패턴을 분석해 **재주문율 향상**”

모든 고객 대상
일관된 마케팅

모든 고객에게 동일한 혜택을 주어 충성
고객 이탈



E사 고객을 등급으로 분류시켜 등급이 높은
고객일수록 높은 혜택 제공

“고객층별 세분화된 **마케팅 시행**”

3. 분석계획

목적	분석방법	주요내용
고객 정보별 재주문율에 대한 영향 파악	막대그래프	요일별 구매 수와 재주문율의 관계 파악
		시간별 구매 수와 재주문율의 관계 파악
		주문 주기별 구매 수와 재주문율의 관계 파악
	산점도	카트에 추가한 상품순서별 구매 수와 재주문율의 관계 파악
		대분류, 중분류별 구매 수와 재주문율의 관계 파악
세분화된 고객별 마케팅을 위한 특성 파악	주성분 분석	주문 텀, 재주문율, 주문 당 제품 수, 요일, 시간, 고객 당 제품 구매수를 변수로 군집분석을 하여 특성 파악
	K평균	
	Boxplot	
다음 재주문에 미치는 고객, 상품, 주문별 영향 파악	의사결정나무	다음 재주문에 영향을 미치는 변수 중요도를 분석해 영향인자 확인
	KNN	
	그래디언트 부스팅	

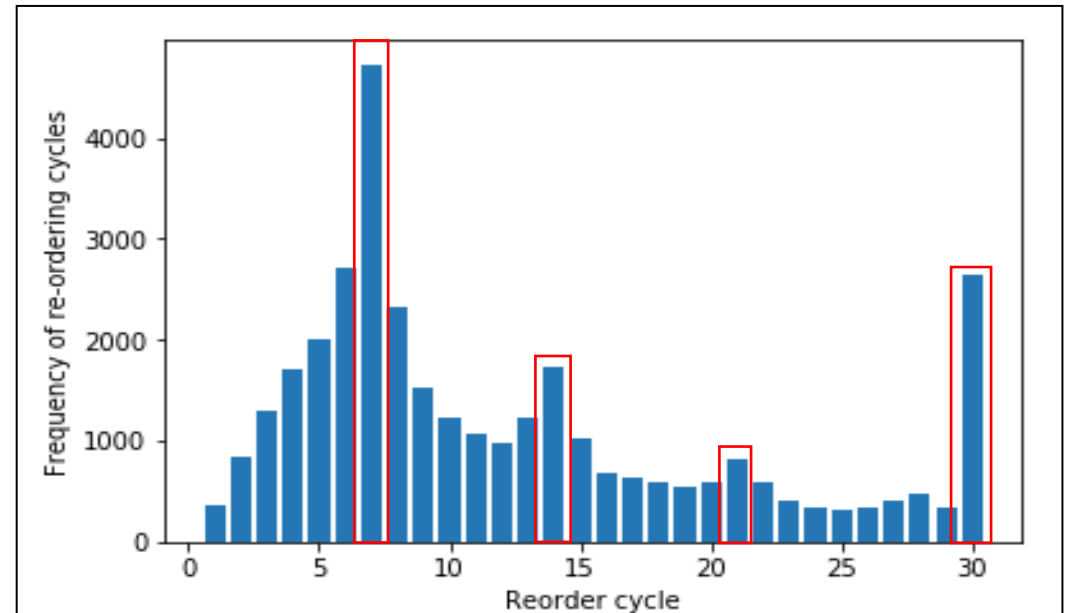
▶ 4. 분석 결과 : 탐색적 분석을 통한 재주문율

주기에 따라 낮아지는 재주문율

<재주문율>

상품명 (상위 제품 3개)	재주문율	주기에 따른 제품 재주문율
Banana	0.841	0.291
Bag of Organic Bananas	0.833	0.315
Organic Strawberries	0.775	0.247
<ul style="list-style-type: none">• 전체 상품의 재주문율 59%• 상위 3개 상품의 주기에 따른 재주문율은 대략 30%• 식품류의 상품의 재구매율이 높음		

<제품 재주문 주기>



<Organic Strawberries 주기 별 주문량>

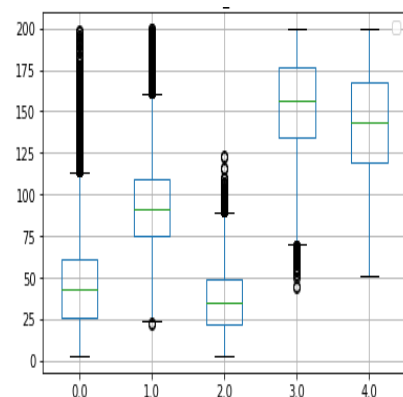
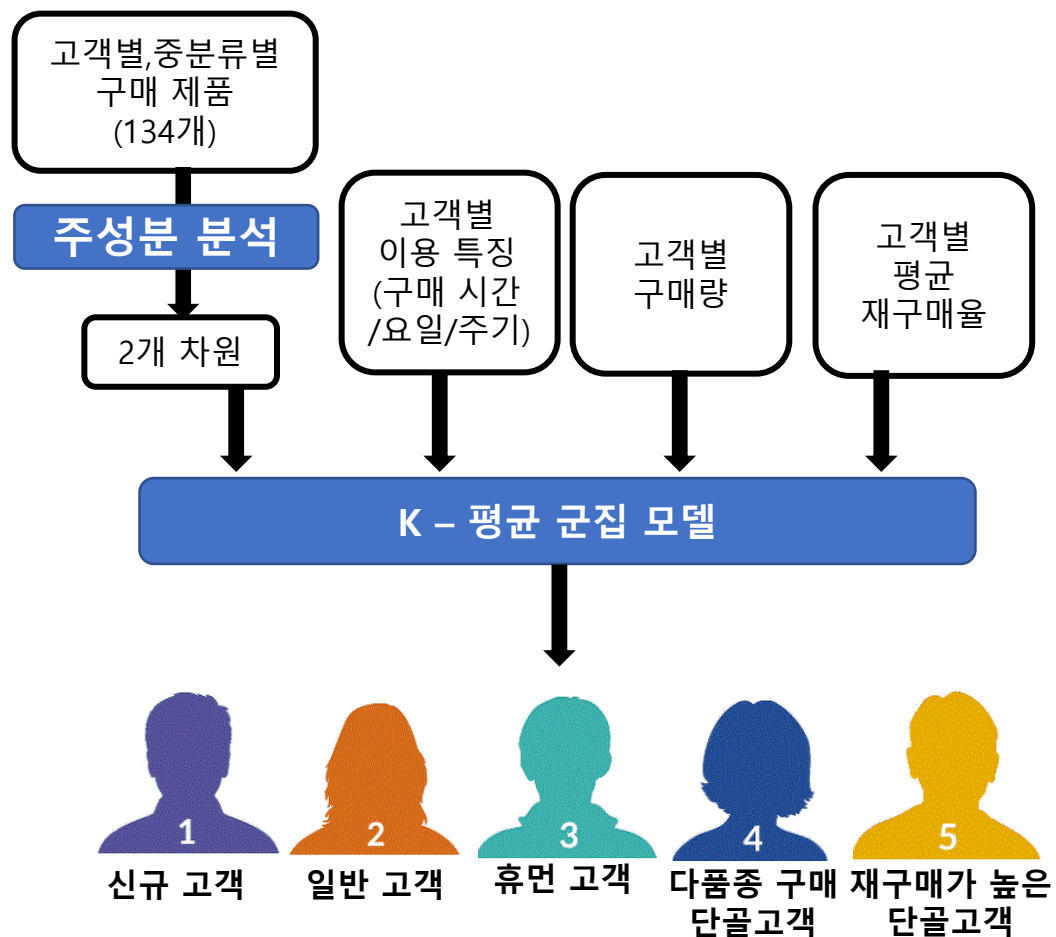
- 주문 주기에 따른 재주문율이 크게 차이남
- 약 7일 주기로 재주문량이 증가

주기에 따라 달라지는 고객 재주문율 향상을 위한 방안 필요

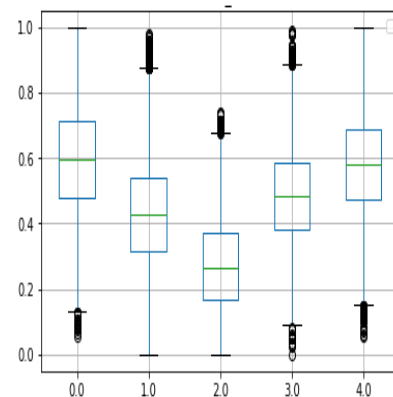
4. 분석 결과 : 고객 세분화를 통한 고

고객 군집별 마케팅 전략을 위한 고객 세분화 결과 구매 주기, 재주문율, 제품구매수의 특성이 뚜렷하게 나타남

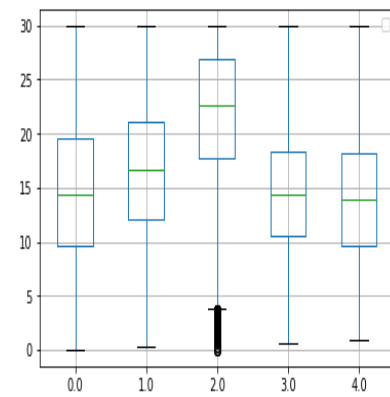
<고객 군집화 모델>



<그림> 군집별 제품구매 수



<그림> 군집별 재구매율



<그림> 군집별 구매 주기

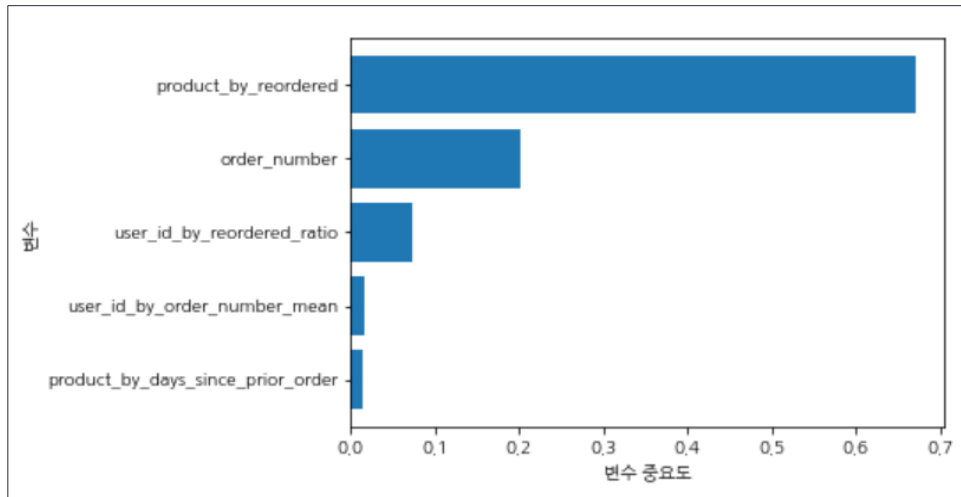
- 군집 1 : 총 제품구매 수 小, 재구매율 高 → '신규 진입 고객'
- 군집 2 : 총 제품구매 수, 재구매율, 구매 주기 모두 평균 → '일반적인 고객'
- 군집 3 : 총 제품구매 수 小, 구매 주기 長 → '휴먼 고객'
- 군집 4 : 총 제품구매 수 多, 재구매율 小, 구매 주기 短 → '다품종 구매 단골 고객'
- 군집 5 : 총 제품구매 수 多, 재구매율 高, 구매 주기 短 → '재구매가 높은 단골 고객'

도출된 결과를 사용해 전체 고객을 세분화하고 세분화된 고객별 마케팅 전략 필요

4. 분석 결과 : 모델 구현을 통한 제품

추천 모델 구현 과정에서 제품별 재구매율, 주문 횟수, 고객별 재주문율이 큰 영향을 미침

<제품 추천 모델에 미치는 영향요소>



<그림> 제품 추천 모델변수 중요도

1순위 영향인자 : 제품별 재주문율

2순위 영향인자 : 누적 주문 횟수

3순위 영향인자 : 고객 ID별 재주문율

<제품 추천 모델>

<div>주문</div> <div>고객</div> <div>제품</div>	모델명	F1-score
	Decision Tree	0.18
	KNN	0.19
	Gradient Boosting	0.25

고객과 제품별 구매 이력과 주문정보를 통해
가장 재구매 확률이 높은 상품 추천
(정밀도와 재현율 0.25)

```
In [18]: df_predict_product(y_pre,y_act)
```

```
실제 상품 재주문 여부:[1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0]  
예측 상품 재주문 여부:[0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0]  
상품 일치 여부 :[0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0]  
상품 일치 개수:총 13개 중 3개 일치
```

제품 추천 모델을 활용한 고객 맞춤형 서비스 제공 필요

▶ 5. 개선안 및 적용방안 수립 : 주기를 재구매율을 고려한 마

구매 주기에 따라 유동적으로 상품 할인 적용



주기에 따른 유동적인 할인

일정 주기내 상품의 재주문 횟수에 따른
할인율 1% 증가



- 할인율을 높이려는 소비자 심리를 이용한 방문 유도로 재구매율 향상
- 주문 주기가 긴 상품에 적용해 구매 주기 단축






주기가 길고 재주문율이 낮은 상품
기존 대비 5% 높은 할인율 적용



- 주기가 길고 재주문율이 낮은 상품의 판매 유도
- 고객 방문 유도로 타 상품 구매 유도

▶ 5. 개선안 및 적용방안 : 고객층별 세분화된 마케팅

군집 분석 결과 신규 진입 고객, 단골 고객, 휴면 고객에 대한 세분화된 고객 관리 마케팅 시행

대상	전략	추진계획
<div>휴면 고객</div> <div>  </div>	<div>휴면 고객 이탈방지 전략</div>	<ul style="list-style-type: none"> • 휴면 고객의 구매 주기를 높이기 위한 쿠폰 제공 • 휴면 고객 대상 기간 한정 이벤트
<div>신규 고객</div> <div>  </div>	<div>패키지 상품 판매 전략</div>	<ul style="list-style-type: none"> • 연관분석 결과를 사용해 자주 주문하는 상품과 연관된 상품 추천 • 패키지 상품 프로모션을 통해 다른 상품 구매 유도
<div>일반 고객</div> <div>  </div>	<div>VIP 전환 전략</div>	<ul style="list-style-type: none"> • 구매량에 따른 고객 등급 시스템을 구축과 등급에 따른 혜택을 통해 구매욕 증감 • 유효기간이 짧은 쿠폰을 제공하여 구매 주기를 줄임
<div>다품종 구매 단골고객</div> <div>  </div>	<div>마일리지 전략</div>	<ul style="list-style-type: none"> • 전체 제품 중 재구매율이 높은 상품 추천 • 마일리지 적립을 통한 지속적인 주문 유도하여 이탈 방지
<div>재구매가 높은 단골고객</div> <div>  </div>	<div>마일리지 전략</div>	<ul style="list-style-type: none"> • 연관분석 결과를 사용해 자주 주문하는 상품과 연관된 상품 추천 • 마일리지 적립을 통한 지속적인 주문 유도하여 이탈 방지

5. 개선안 및 적용방안 수립 : 고객 맞춤형 추천 시스

신규 고객과 기존 고객에 대한 개별화된 추천 서비스 적용

<웹/모바일 어플리케이션 제품 추천 서비스>



- 고객 ID 입력 시, 고객 구매 이력 기반 개인별 상품 추천 서비스
- 이전 구매 내역이 없는 고객의 경우, 전체 구매 내역에서 구매량이 높은 제품 노출

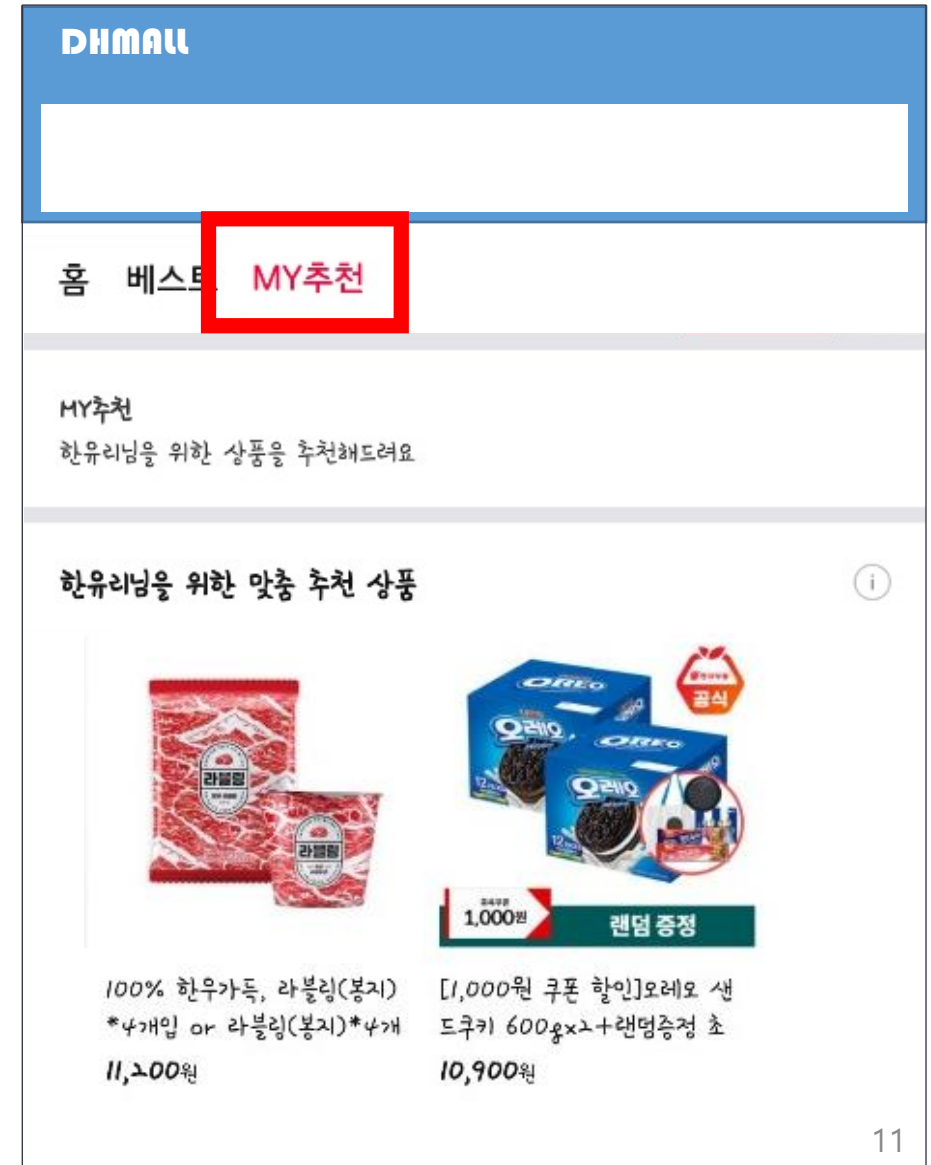
<제품 추천 서비스 기대효과>



상품 노출 빈도 증가로
구매욕구 상승



상품 접근성 증가로
고객 만족도 상승



▶ 6. 결론 및 소

- 대량의 데이터를 활용해 탐색적 분석을 하고 전처리를 하면서 전체적인 데이터 분석의 프로세스를 배울 수 있었습니다.
- 수업시간에 배운 다양한 분석방법을 실제 데이터에 적용하여 보고 이를 통해 유의미한 결과를 도출해 추천 모델을 생성하고 마케팅 방식을 도출하면서 데이터 분석에 대한 흥미를 느낄 수 있었습니다.
- 특히 분석 과정에서 데이터 특성에 알맞은 모델을 선정하는 과정, 그리고 분석 결과가 나온 후 결과를 해석하는 과정에서 통계적 지식의 중요성을 느끼고 빅데이터 분석에 대해 더욱 깊이 배워나가고 싶습니다.
- 분석과정에서 분석 대상에 대한 배경지식 파악이 분석과정, 결론 만큼 중요하다는 것을 알았습니다. 탐색적분석에서 유의미한 패턴이 나오더라도 배경지식에 의해 배제해야하는 경우도 있었고, 또 잘못된 분석방법을 찾아낸 경우도 있었습니다. 아무리 데이터분석을 잘하더라도 배경지식을 무시한다면 제대로 된 분석이 되지 않는다는 것을 깨달았습니다.

▶ 7. 첨부 : 잠자요인 도출

잠재원인	조사대상	자료출처
제품 신선도 유지 x	제품 배송시간	물류 부서 협조요청
	제품별 표면온도	수작업조사
TV 광고에만 치중한 마케팅 전략	자사/경쟁사 SNS 광고여부 횟수 조사	홍보부서, SNS
경쟁사 대비 높은 제품 가격	자사/경쟁사의 유통 마진	내부DB 활용
	자사/경쟁사의 광고비	내부DB 활용
고객별 재구매 데이터 활용 미활용	경쟁사의 고객 재구매 정보	내부 DB 활용, 인터넷 조사, 타사 협력 요청
	자사 제품 노출 대비 구매 정보	내부 DB 활용, 인터넷 조사, 타사 협력 요청
	자사 1회 구매 고객 정보	내부 DB 활용, 인터넷 조사, 타사 협력 요청
품목분류가 되어있지 않은 모바일 UI	한해 월별 웹사이트 대비 모바일 이용객 현황	CS 부서
	모바일 이용 평점 집계	CS 부서

7. 첨부 : 데이터 수집계획

잠재원인	데이터수집계획						
	데이터명	속성	속성	수집방법	담당자	수집 가능성	주요특성
신선식품 배송의 신선 도 하락	품목 아이디	고객 정보	ID	내부 DB	아이유	O	자동측정
	측정 시간	고객 정보	범주형	내부 DB	이성경	O	자동측정
	고객 나이	고객 정보	연속성	내부 DB	금새록	O	자동측정
	가입 일자	고객 정보	DATE	내부 DB	방민아	O	자동측정
	주문 번호	주문 정보	ID	내부 DB	설리	O	자동측정
	제품 아이디	주문 정보	ID	내부 DB	김고은	O	자동측정
	주문 일자	주문 정보	범주형	내부 DB	아이린	O	자동측정
모바일앱 개선 필요	상품 구매 매체 (모바일/웹)	구매 정보	명목형	내부 DB	배슬기	O	자동측정
	주문 번호	구매 정보	DATE	내부 DB	정은지	O	자동측정
	모바일/앱 사용 일시	앱 사용 정보	DATE	내부 DB	헤이즈	O	자동측정
	모바일/앱 사용 기능	앱 사용 정보	범주형	내부 DB	설현	O	자동측정
	모바일/앱 접속 기간	앱 사용 정보	범주형	크롤링	모모	O	자동측정
	사용자 아이디	앱 리뷰 정보	ID	내부 DB	배슬기	O	자동측정
	아이유	앱 리뷰 정보	TEXT	온습도계	아이린	O	자동측정

7. 첨부 : 데이터 특성확인

식품 대분류 (grocery-department)

항목명	의미	속성	특이사항
department-id	대분류 고유 ID	ID	21(missing)
department	대분류	범주형	

식품 중분류 (grocery-aisles)

항목명	의미	속성	특이사항
aisle-id	중분류 고유 ID	ID	6(other)
aisle	중분류	명목형	

식품 명 (grocery-product)

항목명	의미	속성	특이사항
Product-id	상품 고유 ID	ID	
Product-name	상품 명칭	명목형	
Aisle-id	중분류 ID	ID	
Department-id	대분류 ID	ID	

주문 데이터 (grocery-order)

항목명	의미	속성	특이사항
Order-id	주문 번호	ID	
User-id	고객 id	ID	
Eval-set	Prior, test, train로 구분됨	명목형	
Order-number	User-id별 누적주문 횟수	연속형	
Order-dow	주문한 요일	명목형	0 - 6 요일 기준 불분명
Order-hour-of-day	주문한 시간	연속형	
Day-since-prior-order	이전 주문과 현재 주문의 기간	연속형	결측치 약 20만개

prior 주문 데이터 (grocery-order-products-prior)

항목명	의미	속성	특이사항
Order-id	주문 번호	ID	
product-id	제품 id	ID	
Add-to-cart-order	카트 추가 순서	연속형	
reordered	재주문한 상품	이산형	

마지막 주문 데이터 (grocery-order-products¹⁵-train)

▶ 7. 첨부 : 결측치 및 이상치 확인

항목별	의미	유형	이상치(건)	결측치(건)	확인결과	정제방안
day-since-prior-order	이전 주문과 현재 주문의 주기	연속형	-	206229 (1. 206209 2. 20)	1. 1회 주문 시, 발생한 자연 결측치(206209건) 2. 다회 주문 시, 발생한 결측치(20건)	1번의 경우, 데이터를 분리해 처리 2번의 경우, 고객 개인별 평균치로 대체
order_number	사용자 ID별 누적주문 횟수	연속형	24	-	입력된 값이 실수값 등으로 표시된 점을 들어 입력상 오류로 예측	이전 순서를 고려해 값 대체
order_dow	주문한 요일	명목형	-	-	숫자로 표시되어 있어 요일을 알 수 없음	근거를 바탕으로 요일 추정 후 문자로 변환
department_id	상품 대분류	범주형(ID)	548	1258	각 항목에서 예외분류인 'Missing', 'Other'에 기존항목에 포함 발견	상품명 유추 또는 other로 변환
Aisl_id	식품 중분류	범주형(ID)	548	1258	각 항목에서 예외분류인 'Missing', 'Other'에 기존항목에 포함 발견	상품명 유추 또는 other로 변환

7. 첨부 : 결측치/이상치 세부사항

누적구매횟수가 2이상임에도 불구하고(재주문 고객)
현재 주문 과 이전 주문의 텀이 없는 경우(= 이상치)

```
nan=order[(order['days_since_prior_order'].isnull())&(order['order_number']!=1)]
nan
#firstbuy customer strange
```

order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
4859	1148970	301	prior	6.0	2	15
10992	2698174	686	prior	2.0	6	16
12786	451490	802	prior	3.0	0	17
15985	417792	992	prior	8.0	5	11
22411	1263820	1401	prior	20.0	4	16
26609	269316	1656	prior	4.0	2	13
28170	201351	1746	prior	7.0	6	9
30369	2627284	1866	prior	28.0	0	14
37614	2423804	2329	prior	3.0	0	7
46876	255890	2903	prior	47.0	4	16
52399	2404847	3257	prior	2.0	1	12
53556	731572	3337	prior	2.0	0	10
57137	1070838	3532	prior	9.0	5	6
57884	551439	3579	prior	68.0	2	13
61931	885304	3818	prior	11.0	3	20
65332	1852603	4057	prior	7.0	6	12
66768	1956752	4144	prior	7.0	0	12
71787	3089983	4424	prior	56.0	5	10
80641	2530778	4928	prior	4.0	1	18
80964	2140346	4949	prior	37.0	5	9

이전에 구매한 기록이 없고 누적 구매횟수
가 1인 고객(= ONESHOT)

```
oneshot=order[(order['days_since_prior_order'].isnull())&(order['order_number']==1)]
oneshot
#firstbuy customer strange
```

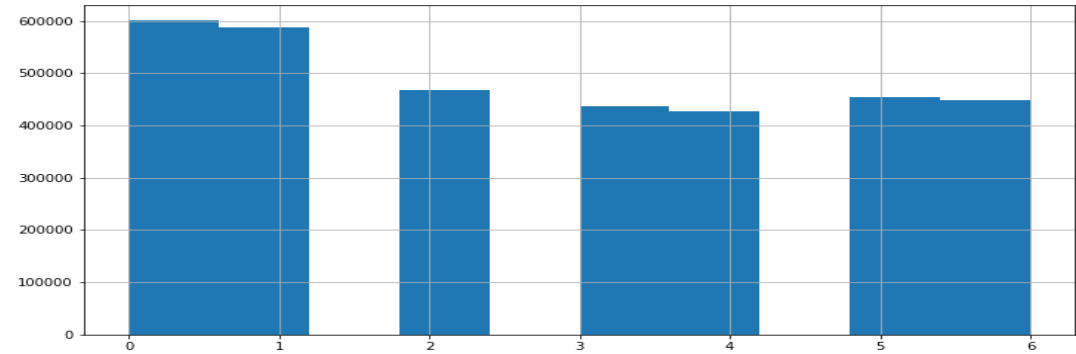
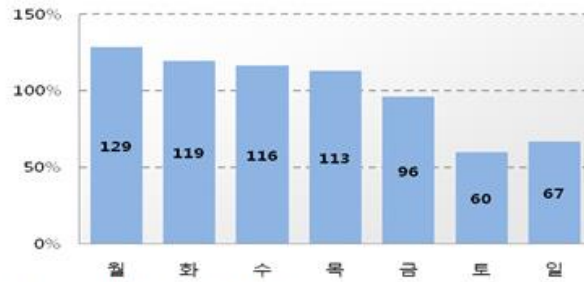
order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
0	2539329	1	prior	1.0	2	8
24	2168274	2	prior	1.0	2	11
37	1374495	3	prior	1.0	1	14
39	3343014	4	prior	1.0	6	11
48	2717275	5	prior	1.0	3	12
50	2086598	6	prior	1.0	5	18
73	2565571	7	prior	1.0	3	9
75	600894	8	prior	1.0	6	0
81	280530	9	prior	1.0	1	17
83	1224907	10	prior	1.0	2	14
93	2411567	11	prior	1.0	0	11
102	1483446	12	prior	1.0	1	12
114	2618231	13	prior	1.0	6	12
128	2906263	14	prior	1.0	2	18
151	2514324	15	prior	1.0	5	13
158	1273676	16	prior	1.0	2	8
182	1737705	17	prior	1.0	2	13
206	2780889	18	prior	1.0	6	18

7. 첨부 : 결측치/이상치 세부사항

인터넷 쇼핑, 월요일 점심시간 전후에 가장 많이 이용

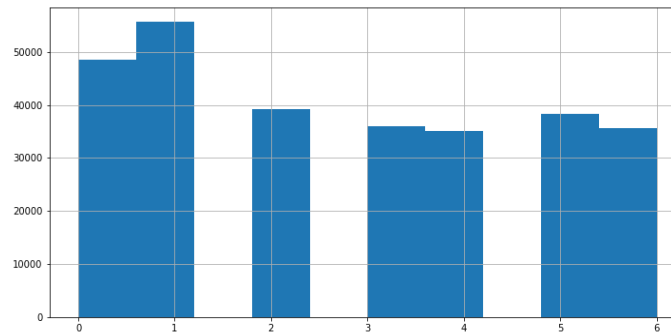
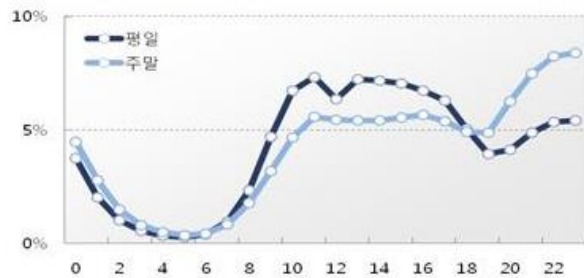
조선비즈 | 손덕호 기자

요일별 인터넷쇼핑몰 매출

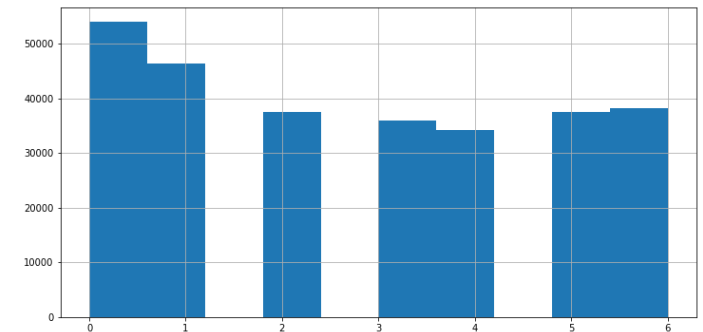


Raw data

시간대별 인터넷쇼핑몰 매출



10시



15시

An isometric illustration showing a hand pointing towards a building. The building has a green roof and a white shopping cart icon on it. Two blue trucks with yellow stripes are parked in front of the building. The text 'QnA' is centered in the middle of the image.

QnA

An isometric illustration featuring a large hand on the left side, with the index finger pointing towards a green building. In front of the building, two blue trucks with yellow and white stripes are parked. A white shopping cart is visible on the green surface of the building. The text '감사합니다.' is centered in the middle of the image.

감사합니다.