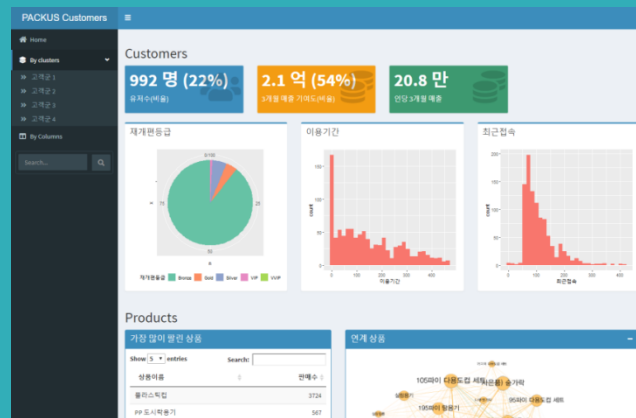
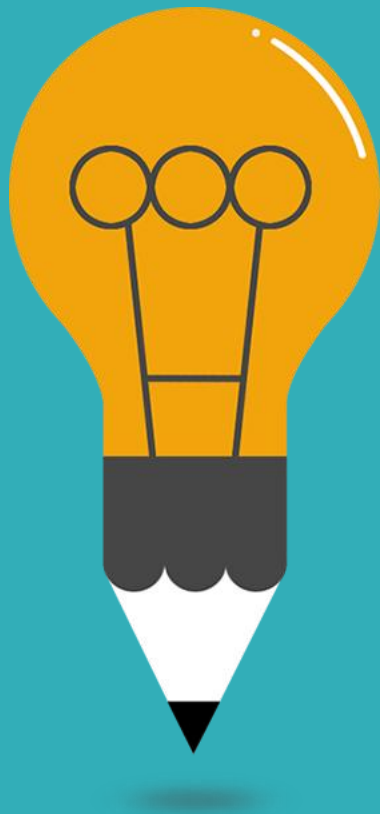


PACKER board

(PACKER : PACKUS의 고객)

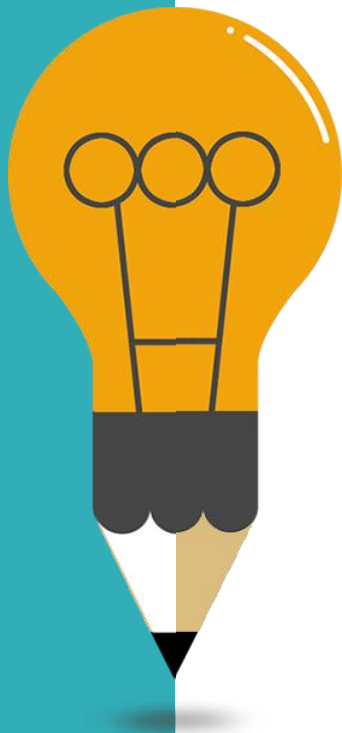
PACKER를 효과적으로 관리하는 고객관리 대시보드



조 이름 : 히스토그램

조원 : 김연강(응통), 신은수(응통), 이승한(경영), 정민지(응통)

목차



01

PACKUS의 현 상황
기업 내부 & 외부 상황

서론

02

SOLUTION 1. 세분화된 고객 관리

1) 고객군 별 전략 / 2) 장바구니 분석 / 3) 고객등급 재개편

본론 1

03

SOLUTION 2. 제품 별 수요예측

1) 목적 / 2) 데이터 전처리 / 3) 모델링 / 4) 결과

본론 2

04

PACKUS의 미래

"PACKER board"를 통한 효과적 고객 관리

결론



1. PACKUS의 현 상황

기업 내부 & 외부 상황

[내부 상황] 미흡한 고객관리 체계

1. 성장 단계에 있는 PACKUS

(Active User & 매출액이 증가하고 있는 추세)

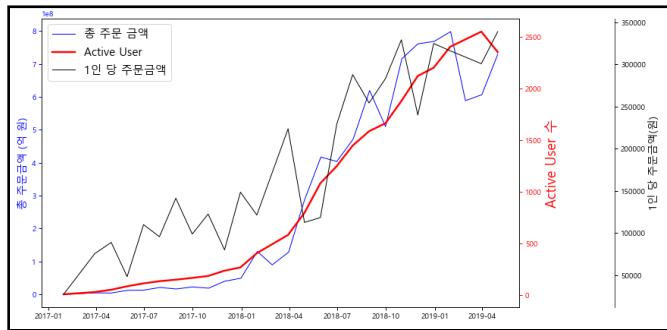
(Active User : 회원가입 +1, 최종접속 -1)

2. 이를 받쳐주지 못하는 '섬세하지 못한 고객 관리 체계'

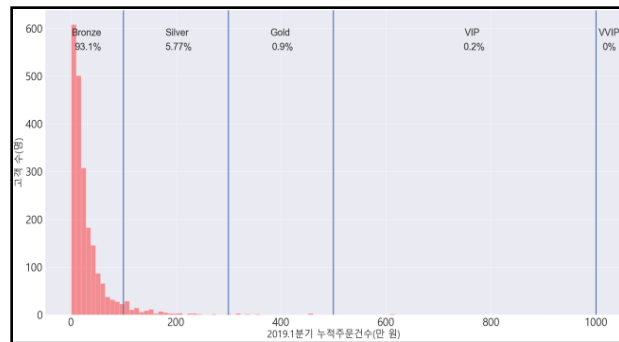
(불균형적인 고객 등급 제도)

[표 1] 현재 팩커스의 고객 등급 체계

회원 구분	BRONZE	SILVER	GOLD	VIP	VVIP
분기별 누적 결제	~100 만원	100 만원~	300 만원~	500 만원~	1 천만원~
2019.1 분기 구매 기준	2017 명 (93.1%)	125 명 (5.77%)	20 명 (0.9%)	4 명 (0.2%)	0 명 (0%)



총 주문 금액 / Active User / 1인당 주문금액의 추이



(현 고객등급 체계에 따른) 등급별 고객 분포

“성장 단계에 있는 PACKUS의 세부적 고객관리의 필요성!”

[외부 상황] 다양해지는 배달 음식 종류

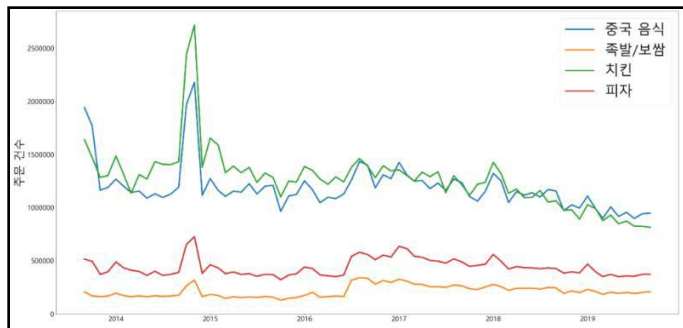
1. 거의 모든 종류의 음식이 배달 가능 + 1인분 배달의 등장

(빙수, 마라탕, 양꼬치, 분짜, 디저트...)

PACKUS에서 판매하는 다양한 종류의 용기들 >

2. 다양해지고 있는 배달 용기 종류

(기존의 대표적인 배달 음식들 (TOP4)의 비중 감소! 새로운 종류의 배달 음식 등장)



[SKT 고객 주문] TOP 4 배달음식 배달건수 변화 추이



[요기요] 아직까지 낯선 배달음식 종류들

☰ 카테고리 전체보기

식자재

초특가 SALE

NEW

BEST상품

한식(분식) 용기

중식 용기

일식 용기

양식 용기

실링제품

베이커리/카페 용기

농수산물 용기

주방용품/소모품

스티커 제작

소량

“배달 용기 ‘종류’별로 각각의 수요 예측 필요”

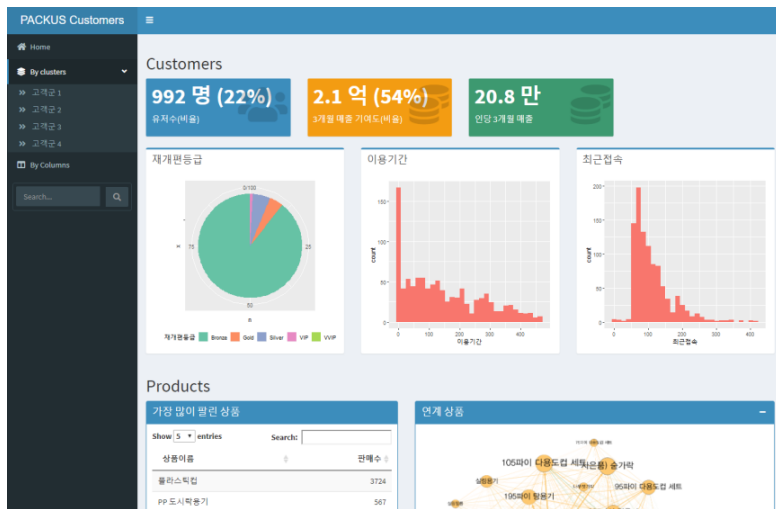
기업 내부

“성장 단계에 있는 PACKUS의 세부적 고객관리의 필요성!”

기업 외부

“배달 용기 ‘종류’별로 각각의 수요 예측 필요”

How
To
Solve?



PACKER board

[1] 세부적으로 고객 군을 나눠서,
각각에 맞는 최적의 방식으로 “고객 관리”

[2] 제품 별로 수요량을 예측하여
효과적인 “재고 관리”



SOLUTION 1. 세부적인 고객관리

1) 고객군 별 전략 2) 장바구니 분석 3) 고객등급 재개편

SOLUTION 1. 세부적인 고객관리



How to Solve?

3가지 방법을
통한 고객 관리

01

고객군 별 전략

고객군 별로 “구매력 / 이용 행태 / 혜택에 따른 반응” 분석
이에 알맞은 방식으로, 고객군 별 전략 수립

02

장바구니 분석

고객 군 별로 ‘구매 물품’ 분석
연관성 분석을 통해 ‘함께 구매하는 물품’ 파악
고객군 별 구매 물품 추천

03

고객등급 재개편

기존의 불균형적인 고객 등급 재개편
각각의 등급에 알맞은 방식의 혜택 제공

1-1. 고객군 별 전략

(SOLUTION 1. 세부적인 고객 관리)

데이터 전처리

고객군 별 비교

1

2

3

4

5

데이터 소개

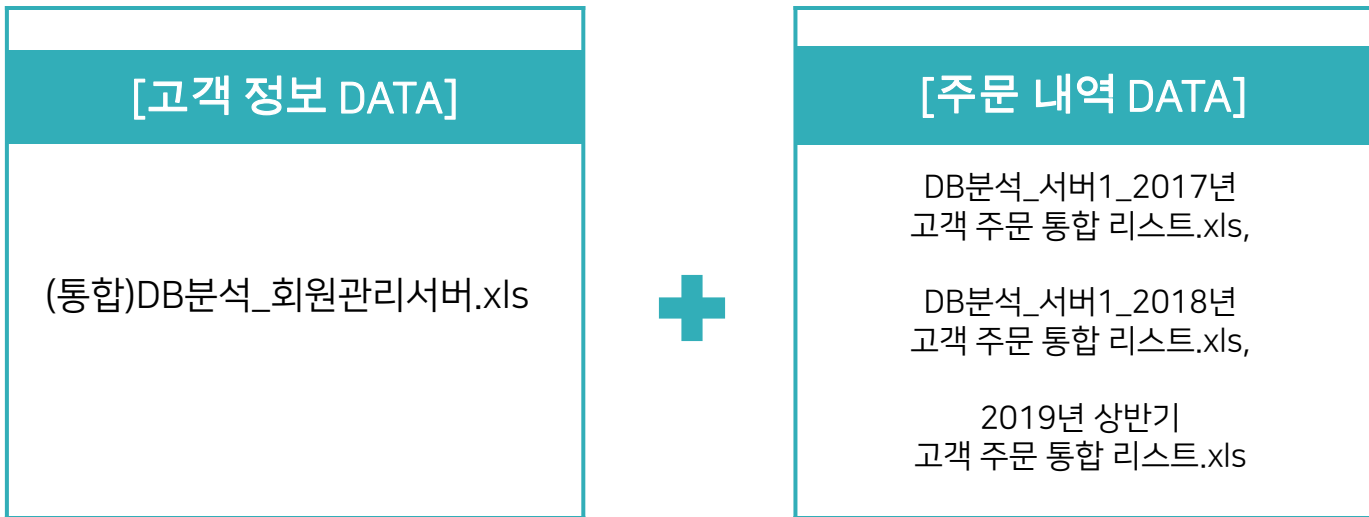
고객군 생성 (Clustering)

고객군 별 제언

1. 데이터 소개

1-1. 고객군 별 전략

(SOLUTION 1. 세부적인 고객 관리)



(base) 고객 정보 DATA + (추가) 주문 내역 DATA

최종: "총 12가지"의 고객 특징을 담고 있는 "4583명의 고객 정보"

1. 새로운 변수 추가

(1) 최근 접속까지의 일 수

- 2019.10.10 - 최근 접속일
- “가장 최근에 언제 활동했는지 알 수 있는 변수”

(2) 이용 기간

- 최종 접속일 - 회원 가입일
 - “고객의 충성도를 파악할 수 있는 변수”
- (휴면 기간이 있을 수 있으나, 휴면 이후 복귀할 정도면 충성도 유지한다고 가정)

(3) 주문 경로

- PC를 통한 주문과 Mobile을 통한 주문의 비율
 - “고객의 주문 방식을 파악할 수 있는 변수”
- (추후, 고객 별 프로모션을 진행할 경우 적절한 수단에 대한 정보 제공)

(새로운 3개의 변수를 포함하여)

총 12개의 변수 선정!

약 30%의 값이 NA인 **배송지역** 변수 ('주소(동/읍/면)')와,
'누적 주문건수'와 상관 관계(0.99)가 높은 '**총 실주문 건수**' 제외

2.데이터 전처리

1-1. 고객군 별 전략

(SOLUTION 1. 세부적인 고객 관리)

2. 범주형 변수의 Dummy화

범주형 변수 : '회원 구분', '회원 등급', '회원 가입 경로'

- 1) 회원 구분 : 개인 / 사업자
- 2) 회원 등급 : 일반 / 전화 주문 / 사업자 / 기타
- 3) 회원 가입 경로 : 잡지 / 인쇄물 / 인터넷 / 지인 추천 / 기타 / 없음

3. 이상치 (Outlier) 제거

Clustering에 있어서, 고객군 별 특징을 일반화 하는데 지장!

수치형 변수들의 Z-score가 (-3,3)을 벗어나는 데이터들 제거



"중복 ID, 결제 정보 없는 ID, 일자 정보 없는 ID,
이상치" 제거 후, **4583명의 고객 data**

2.데이터 전처리

1-1. 고객군 별 전략


(SOLUTION 1. 세부적인 고객 관리)

4. 수치형 변수의 정규화

변수들 간의 규모 차이가 큰 경우 0

'표준 정규분포'로 scaling

	누적 주문 건수	총 사용 적립금	총구매금 액	추천 인 아 이 디	평 생 회 원	PC_ 주문서 쿠폰 할인금 액	이 용 기 간	최근 접속	회 원 구 분_ 개 인	...
아이디										
cgh1385	1.0	0.0	70500.0	0	0.0	0.666667	0.0	1	1032	1 ...
pjki177	3.0	0.0	63000.0	0	0.0	0.555556	8000.0	43	977	0 ...
bear011	37.0	31900.0	2870970.0	0	0.0	0.867925	0.0	701	316	0 ...
kim0750	1.0	0.0	77700.0	0	0.0	1.000000	5000.0	0	1016	0 ...
nonane9979	3.0	0.0	238900.0	0	0.0	0.875000	0.0	26	990	0 ...



	누적주문 건수	총 사용 적 립금	총구매금 액	추천 인 아 이 디	평 생 회 원	PC_ 주문서 쿠폰 할인금 액	이용기간	최근접속	회 원 구 분 _ 개 인	...
아이디										
cgh1385	-0.449808	-0.187955	-0.396559	0	0.0	0.666667	-0.056221	-0.724735	4.162994	1 ...
pjki177	-0.330645	-0.187955	-0.401915	0	0.0	0.555556	0.175104	-0.424708	3.844411	0 ...
bear011	1.695123	1.532370	1.603136	0	0.0	0.867925	-0.056221	4.275711	0.015623	0 ...
kim0750	-0.449808	-0.187955	-0.391418	0	0.0	1.000000	0.088357	-0.731879	4.070315	0 ...
nonane9979	-0.330645	-0.187955	-0.276312	0	0.0	0.875000	-0.056221	-0.546148	3.919712	0 ...

생성된 최종 DATASET

총 4가지 측면의 고객 데이터 정보!

(1. 고객 개인 정보 / 2. 고객 주문 정보 / 3. 혜택 관련 정보 / 4. 기타 정보)

4 가지 측면의 고객 데이터

고객의 특성을 대변할 수 있는 네 분류의 변수들 (총 12개)

1. 고객 개인 정보	(1) 회원 구분, (2) 회원 등급, (3) 이용 기간, (4) 최근 접속까지 일 수, (5) 평생회원 여부
2. 주문 관련 정보	(6) 누적 주문 건수, (7) 총 구매 금액
3. 혜택 관련 정보	(8) 총 사용 적립금, (9) 쿠폰 할인 금액
4. 그 외 정보	(10) 주문 경로, (11) 추천 아이디 유무, (12) 회원 가입 경로

범주형 변수는 더미화하여, 총 22개의 Column!

2.데이터 전처리

1-1. 고객군 별 전략

(SOLUTION 1. 세부적인 고객 관리)

	변수 명	변수 설명
개인 정보	1. 회원 구분	개인 / 사업자
	2. 회원 등급	일반 / 전화주문 / 사업자 / 기타
	3. 이용 기간	회원 가입일 ~ 최근 접속일 까지의 일 수
	4. 최근 접속까지 일 수	최근 접속일 ~ 2019.10.10(분석일) 까지의 수
	5. 평생회원 여부	평생회원이면 1, 아니면 0
주문 정보	6. 누적 주문건수	(가입 이후) 구매한 총 건수
	7. 총 구매금액	(가입 이후) 구매한 총 금액
혜택 정보	8. 총 사용 적립금	총 사용한 적립금
	9. 쿠폰 할인 금액	쿠폰으로 인해 할인 받은 금액
기타 정보	10. 주문 경로	PC와 Mobile을 통한 주문 비율
	11. 추천 아이디	타인에게 추천을 받아 가입했으면 1, 아니면 0
	12. 회원 가입 경로	잡지 / 인쇄물 / 인터넷 / 지인 추천 / 기타 / 없음

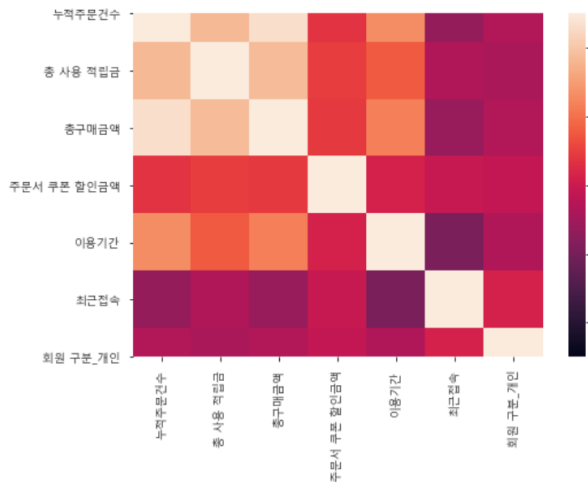
2.데이터 전처리

1-1. 고객군 별 전략

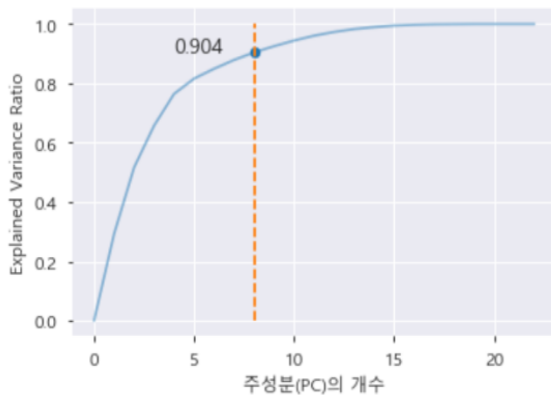
(SOLUTION 1. 세부적인 고객 관리)

5. PCA를 통한 차원 축소

총 22개의 column을, 8개의 PC(주성분)으로 차원 축소!



다중공산성의 문제
(주문 건수 & 적립금 & 주문 금액)



8개의 주성분으로 90.4% 설명 가능

원본 변수들: 주문건수, 주문금액, 주문서 수. PC_1, PC_2, PC_3, PC_4, PC_5, PC_6, PC_7, PC_8.

아이디	주문건수	주문금액	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수	주문서 수
cg11385	-0.449808	-0.187955	-0.396559	0.000000	0.000000	0.000000	-0.056221	-0.724735	4.162994	1
pjki77	-0.330645	-0.187955	-0.401915	0.000000	0.000000	0.000000	0.175104	-0.424708	3.844411	0
bear011	1.695123	1.532370	1.603136	0.000000	0.000000	0.000000	-0.056221	4.275711	0.015623	0
kim0750	-0.449808	-0.187955	-0.391418	0.000000	0.000000	0.000000	0.088357	-0.731879	4.070315	0
nonane9979	-0.330645	-0.187955	-0.276312	0.000000	0.000000	0.000000	-0.056221	-0.546148	3.919712	0



아이디	pc1	pc2	pc3	pc4	pc5	pc6	pc7	pc8
91559380	0.359893	-0.733428	2.474140	1.423490	-0.700689	-0.488936	-0.797622	0.034800
dambi531	-1.431374	-0.144208	2.212528	1.117344	-0.552615	-0.132613	-0.687181	0.071707
scsodamae	-0.783358	1.390950	2.603797	0.260452	-0.460697	0.464107	0.615058	-0.225362
ehfd1108	-0.981171	1.298937	2.386230	0.006509	-0.530484	-0.543432	-0.107830	0.013430
wanya	-1.928318	0.072628	1.954450	1.096761	-0.551976	0.148392	-0.409228	-0.031211

3.고객군 생성 (Clustering)

1-1. 고객군 별 전략

(SOLUTION 1. 세부적인 고객 관리)

1. 사용한 Clustering 기법

- 1) K-Means
- 2) DBSCAN
- 3) Mean-Shift
- 4) SOM (Self-Organizing Map)

2. Clustering 기법 간의 성능 비교

	사용 기법	Silhouette Score
1	K-Means (k=5)	0.256
2	DBSCAN	0.01
3	Mean-Shift	0.341
4	SOM (2x2)	0.347
5	SOM (3x3)	0.335

Why "SOM"?

- 1) 다른 기법들보다 **높은 'silhouette score'**
- 2) 격자 형식으로 생성되는 cluster ! **인접한 cluster끼리 비슷한 특징** 가짐을 알 수 있음
- 3) **3차원**으로 보아도 잘 구분된 Cluster

3.고객군 생성 (Clustering)

1-1. 고객군 별 전략

(SOLUTION 1. 세부적인 고객 관리)

3. Self-Organizing Map을 통해 생성된 고객 군

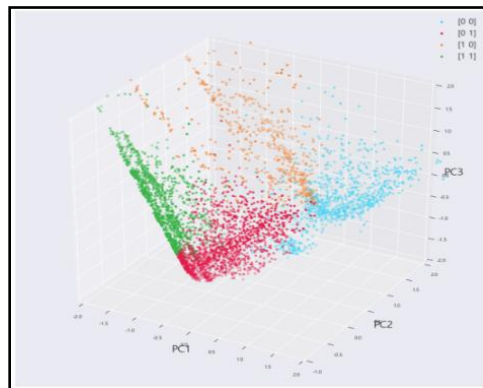
SOM : 인공 신경망을 통한 clustering 방법

총 4개(2x2)의 cluster 생성

(Cluster : [0 0], [0 1], [1 0], [1 1])



인접한 cluster끼리 유사



3차원으로 나타낸 Cluster

고객군 별로 속한 고객 수는?

Cluster	고객 수
고객군 1 ([0 0])	992명 (21.6%)
고객군 2 ([0 1])	1484명 (32.4%)
고객군 3 ([1 0])	650명 (14.2%)
고객군 4 ([1 1])	1457명 (31.8%)

4.고객군 별 비교

1-1. 고객군 별 전략

(SOLUTION 1. 세부적인 고객 관리)

[표 4] 고객 군별(클러스터 별) 비교 분석

		고객군 1	고객군 2	고객군 3	고객군 4
	고객 수	992 명(21.6%)	1484 명(32.4%)	650 명(14.2%)	1457 명 (31.8%)
	총 주문금액	56 만 원	45 만 원	22 만 원	17 만 원
구매력	건당 주문금액	6.4 만 원	6.3 만 원	6.2 만 원	5.8 만 원
	주문건수	8.2 건	6.7 건	3.3 건	2.6 건
	전화주문고객 비율	5.2 %	52.3 %	5.7 %	49.4 %
이용 행태	이용기간	146 일	131 일	27 일	12 일
	최근 접속	55 일	51 일	322 일	274 일
혜택 활용	적립금 사용률	15 %	2 %	1 %	0.1 %
	쿠폰 할인액	965 원	359 원	950 원	488 원

3가지 측면 ('구매력', '이용 행태', '혜택 활용')에서 비교

- 기업 입장에서

- 1) 높은 수익을 낼 수 있는 항목에 '파란색',
- 2) 그렇지 않은 항목에 '분홍색'

이 표의 정보를 토대로, 고객 군에 따라

어떠한 측면에 맞춰서 전략을 펼쳐야 할 지 예상 가능!

4.고객군 별 비교

1-1. 고객군 별 전략

(SOLUTION 1. 세부적인 고객 관리)

고객군 간의 비교

[1] 1&2와 3&4는 '**구매력**'과 '**이용 기간**' 측면에 의해 구분

고객군 1&2 : 장기간 이용 + 현재도 이용 중 + 구매력 높음

고객군 3&4 : 장기간 동안 접속 X + 이용기간 짧음 + 구매력 낮음

[2] 1&3와 2&4는 '**주문 방식**'에 의해 구분

고객군 1&3 : 전화 주문 고객 비율이 5% 정도

고객군 2&4 : 전화 주문 고객 비율이 50% 정도

[3] 1과 2&3&4는 '**적립금**'이용 측면에 의해 구분

고객군 1 : 적립금 사용률 15%

고객군 2&3&4 : 적립금 사용률이 각각 2, 1, 0.1%

[표 4] 고객 군별(클러스터 별) 비교 분석

		고객군 1	고객군 2	고객군 3	고객군 4
	고객 수	992 명(21.6%)	1484 명(32.4%)	650 명(14.2%)	1457 명 (31.8%)
	총 주문금액	56 만 원	45 만 원	22 만 원	17 만 원
구매력	건당 주문금액	6.4 만 원	6.3 만 원	6.2 만 원	5.8 만 원
	주문건수	8.2 건	6.7 건	3.3 건	2.6 건
	전화주문고객 비율	5.2 %	52.3 %	5.7 %	49.4 %
이용 행태	이용기간	146 일	131 일	27 일	12 일
	최근 접속	55 일	51 일	322 일	274 일
혜택 활용	적립금 사용률	15 %	2 %	1 %	0.1 %
	쿠폰 할인액	965 원	359 원	950 원	488 원

5.고객군 별 제언

1-1. 고객군 별 전략

(SOLUTION 1. 세부적인 고객 관리)

고객군 1

1. 3가지 측면(구매력, 이용 행태, 혜택 활용)에 있어서 모두 기업에게 “가장 필요한 고객 군”
2. 어떠한 고객군 보다 이들의 **Retention에 신경써야.**
by 맞춤형 MKT! (뒤에서 언급할 고객군 별 구매 물품 추천)

고객군 2

1. 인터넷이 아닌 “**전화를 통해서 주문**”을 하는 고객이 많음
(이에 따라, 적립금 사용률도 함께 낮음)
2. 고객군 1 못지 않은 “**구매력을 갖춘 고객**”
3. 전화를 통해 주문할 때, “**모바일 가입 유도**”를 통해 혜택 정보들을 알림으로써 구매 방식의 전환을 유도

고객군 3

1. “**구매력은 낮지만**”, 전화 보단 대부분 “**인터넷**”을 통해 주문
2. 모바일 이용도가 높기 때문에 **홍보나 프로모션을 펼치기 용이**
이러한 고객 특성을 활용하여, **적극적인 온라인 MKT** 실시 가능

고객군 4

1. 모든 측면에 있어서 가장 기업 입장에서 “**비선호적인 고객군**”
2. “**낮은 구매력**” 뿐만 아니라, 대부분 인터넷 보다 “**전화로 주문**”
(온라인 MKT이 용이하지 않음)
3. 기업의 한정된 자원을 이 고객 군에게 사용하기 보다는,
다른 세 고객 군에게 사용하는 것이 더 나은 판단일 수도 있다.

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

장바구니 분석

고객군 별 제언

1

2

3

4

물품 데이터 전처리

Gephi 를 활용한 시각화

1. 물품 데이터 전처리

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

1. 사은품 & 구매품 구분

- 같은 종류라도, “사은품 & 구매품” 간에는 큰 의미의 차이
- 텍스트 처리 과정 중, 두 개가 서로 섞일 것 방지!

- 총 667 종류의 물품 中

“구매품 : 650 종류”

“사은품 : 17 종류”

사은품

(‘사은품’ 단어 포함)

0	
0	160mm 스포키 검정 100개 사은품
1	사은품) 위생 장갑 50매
2	사은품) 위생 롤백 5호
3	사은품) 고무장갑
4	사은품) 친환경 종이호일 2호
5	사은품) 위생 랍 4호
6	사은품) 위생 롤백 2호
7	사은품) 친환경 종이호일 1호
8	사은품) 칼라 3중 위생 면행주
9	사은품) 다용도 스탠드 지퍼백M 15매
10	사은품) 위생 랍 2호
11	사은품) 다용도 스탠드 지퍼백XL 15매
12	사은품) 위생 청 수세미(3개)
13	사은품) 독일정품 칼라 행주(3개입)
14	사은품) 맛잇술갈 손가락 100개(검정)
15	사은품) 반짝이 수세미
16	사은품) 위생 장갑 150매

구매품

(‘사은품’ 단어 미포함)

0	
0	105파이 다용도컵 세트 5호
1	소량 105파이 다용도컵 세트 5호
2	170파이 원형 샐러드 3칸 150ml x 3 투명 40
3	PS 95mm 세발 미니 포크 (투명) 10
4	170파이 샐러드 1칸 (550ml x 1)/투명/5
5	170파이 샐러드 1칸 (550ml x 1)/검정/5
6	300파이 원형 점용기
7	사각 즉용기
8	95파이 다용도컵 세트 4호
9	70파이 다용도컵 세트 2호
10	PP 6칸 도시락용기 02 20
11	75파이 다용도컵 세트 3호(대/165ml)_백색/PS-10
12	170mm 수저 투명 1,50
13	PP 5칸 도시락용기 10
14	92파이 아이스컵 투명 동형 구멍있음 PET 100
15	92파이 아이스컵 무지 (9온스/250ml)_투명/PET-1,00
16	PET_투명컵16온스_92파이_0922mm_480ml_100
17	75파이 다용도컵 세트 3호(대/165ml)_투명/PS-10
18	95파이 다용도컵 세트 4호(소/200ml)_백색/PS-10

1.물품 데이터 전처리

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

2. 비슷한 물품 묶기

- 일부는 '소량 & 다품종' 의 물품들이라 묶어서 처리
- "색상, 호, 칸 수, 개입 등" 을 고려 대상에서 배제
- "띄어 쓰기" 무시 (ex. 고무 장갑 = 고무장갑)
- 그 외의 비슷한 단어 묶기
(ex. 맛있숟갈 = 숟가락, 반짝이 수세미 = 청 수세미)

- 물품 종류 축소

"구매품 : 650 종류 -> 266종류"

"사은품 : 17 종류 -> 9 종류"

3. 사은품 + 구매품 합치기

- 나누어서 전처리 한 두 데이터 다시 Merge
- 장바구니 분석 시, 구매품-사은품 pair 파악 위해!

주문상품명	
210파이 중화면용기	4798
PP 도시락용기	3469
사은품) 위생 장갑	3156
195파이 탕용기	2746
105파이 다용도컵 세트	2533
실링용기	1861
사각 죽용기	1812
사은품) 수세미	1639
195파이 미니탕용기	1629
사은품) 행주	1511

가장 많이 거래된 물품 TOP 10

1.물품 데이터 전처리

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

4. 주문자 ID & 물품 & 고객군 묶기

	주문자ID	주문상품명	cluster
4190	mhlee0213	아이스컵	[0 0]
4191	mhlee0213	샤은품) 위생 장갑	[0 0]
4192	mhlee0213	샤은품) 위생 장갑	[0 0]
4193	mhlee0213	샤은품) 수세미	[0 0]
4222	kinamio	95파이 다용도컵 세트	[0 0]
4223	kinamio	사각 족용기	[0 0]
4224	kinamio	PP 도시락용기	[0 0]

5. 적게 등장하는 물품 버리기

각각의 cluster 내에서 10회 미만 등장하는 물품 제거!
(이후 장바구니 분석에서 더 엄격한 기준)

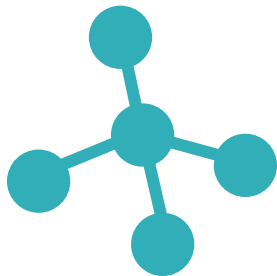
* Cluster 별 물품 수

1	c00.shape, c01.shape, c10.shape, c11.shape ((8149, 4), (9990, 4), (2125, 4), (3706, 4))
1	c00.shape, c01.shape, c10.shape, c11.shape ((7958, 4), (9791, 4), (1906, 4), (3441, 4))



“Cluster별로” 많이 구매하는 물품 / 함께 구매하는 물품 파악하기 위해!

고객군(Cluster)별로 장바구니 분석



2가지 지표를 통한
장바구니 분석

01

[Support]

- 얼마나 자주 등장하는지!
- 가장 많이 구매한 물품 파악!

02

[Confidence]

- 특정 조건 하에, 다른 사건이 얼마나 자주 발생?
- confidence(A->B) : A를 산 사람이 B도 샀을 확률... 함께 사는지 물품 파악

2.장바구니 분석

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

1. Support : 어떤 물품을 가장 많이 샀는가?

(사은품) 위생장갑이 4 고객군에서 모두 1위 차지!

(%)는 구매한 사람 수 기준 (물품 등장 횟수 X)

	고객군 1	고객군 2	고객군 3	고객군 4
2위	105파이 다용도 컵 세트 (23.5%)	195파이 탕 용기 (20.7%)	PP 도시락 용기 (17.7%)	PP 도시락 용기 (14.3%)
3위	195파이 탕 용기 (19.1%)	(사은품) 수세미 (19.1%)	105파이 다용도 컵 세트 (12.5%)	(사은품) 수세미 (13.6%)

고객군 1,2는 “**탕 요리**”를 많이 하는 요식업자

고객군 3,4는 “**도시락**”을 많이 만드는 요식업자

로 짐작 가능!

2.장바구니 분석

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

2. Confidence : 어떤 물건과 어떤 물건을 함께 사는가?

1. 소비자의 “구매” 행동을 파악하기 위해, **사은품은 제외!**

2. **당연한 사항 제외**

(ex. 중화면 용기 & 중화면 용기 뚜껑)

(ex. 플라스틱 컵 & 아이스 컵, 95파이 컵 & 70파이 컵)

3. 해당 물품이 **최소 5%이상의 고객에게 구매된 경우만 포함!**

(거래 건수가 미미한 것으로 고객 구매패턴 일반화 Hard)



특정 물품 구매 시, “**추천 물품**” / “**함께 구매하기 좋은 물품**” 등으로

광고를 할 경우, 구매 유도 효과를 가져올 수 있음

2.장바구니 분석

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

2. Confidence : 어떤 물건과 어떤 물건을 함께 사는가?

빨간색 %가 confidence

고객군 1	250파이 원형 접시 미니 (6.9%)	➡	210파이 중화면 용기 (<u>15.1%</u>)	83.6%
고객군 2	수저 (11.8%)	➡	105파이 다용도 컵 세트 (<u>18.7%</u>)	42.3%
고객군 3	95파이 다용도 컵 세트 (5.0%)	➡	PP 도시락 용기 (<u>17.7%</u>)	26.7%
고객군 4	사은품이 대부분의 물품을 차지하는 고객군!			

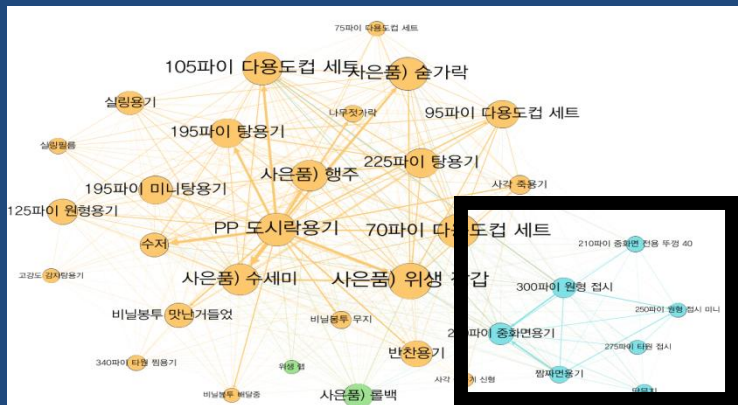
3.Gephi를 활용한 시각화

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

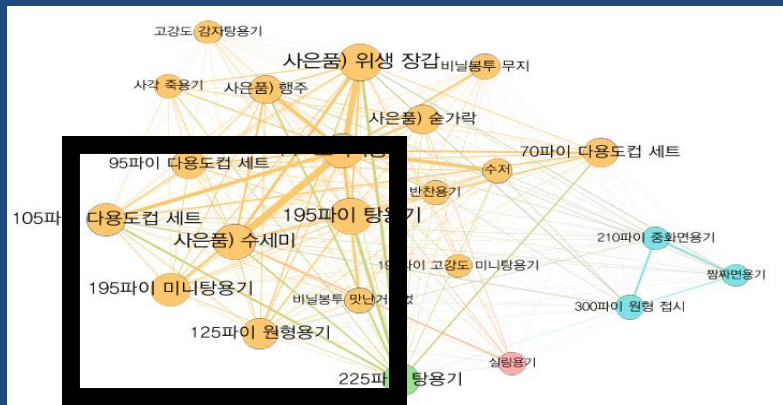
Cluster 별 시각화

Cluster 1



Focus할 물품 : "중화 용기"

Cluster 2



Focus할 물품 : "탕 용기"

3.Gephi를 활용한 시각화

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

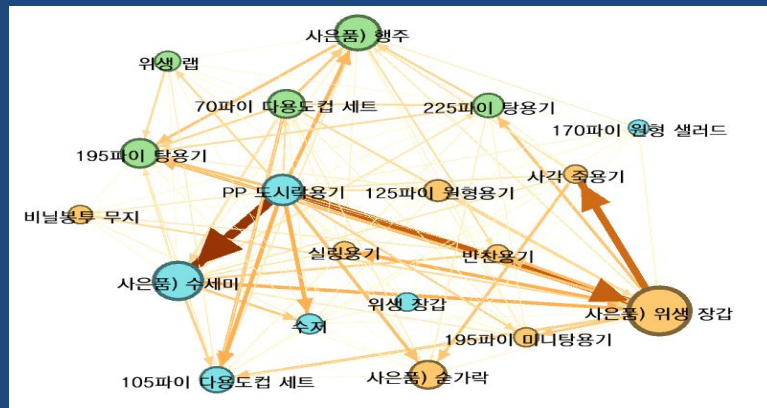
Cluster 별 시각화

Cluster 3



Focus할 물품 : "도시락 용기"

Cluster 4



Focus할 물품 : "X" (사은품)

4.고객군 별 제언

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

고객군 1

[중화 용기]

중화 용기와 관련된 물품이 그 어느 고객군 보다 비중이 높은 군집

방안 1) 나무 젓가락이나, 단무지 등을 끼워 팔기를 통해 구매를 유도

‘도시락 용기’ 구매자에게는 사은품이 지나칠 정도로 많이, ‘중화 용기’ 구매자에게는 사은품을 거의 X

방안 2) 중화용기에도 이에 알맞은 사은품을 제공하여 더욱 더 고객을 끌어올 필요가 있다.

고객군 2

[탕 용기]

3개의 물품 (도시락,짬짜면,중화면용기) 제외하고, 전부 “탕 용기” (탕, 감자탕, 원형 용기)일 정도로 높은 비중
탕 용기를 구매 시, ‘다용도 컵’(70,95,105 파이)을 함께 자주 구매!

방안 1) 탕 용기 구매 시, 추천 구매 물품으로 ‘다용도 컵’ 노출 시켜 구매 유도

4.고객군 별 제언

1-2. 장바구니 분석

(SOLUTION 1. 세부적인 고객 관리)

고객군 3

[도시락 용기]

어느 집단에서나 모두 도시락 용기가 많이 등장.

BUT 고객군 3만큼 "도시락 용기가 다른 상품과의 연관성이 높은 군집은 없음"

도시락 용기와 '다용도 컵'이 자주 함께 구매가 이루어짐

방안 1) 도시락 용기 구매 시, 추천 구매 물품으로 '다용도 컵' 노출 시켜 구매 유도

고객군 4

[X]

앞의 1-1분석에서도 보았듯, 가장 '돈이 안 되는 고객군'

장바구니 분석에서도 볼 수 있듯, 여러 사은품들이 가장 높은 degree를 보임

방안 1) 기업의 한정된 자원을 고려하면, 이들보다는 다른 고객군을 더 신경 쓰기

1-3. 고객등급 재개편

(SOLUTION 1. 세부적인 고객 관리)

1

현재 등급 제도의
상황과 문제점

2

개편 등급 제도

고객 등급 재개편

1-3. 고객 등급 재개편

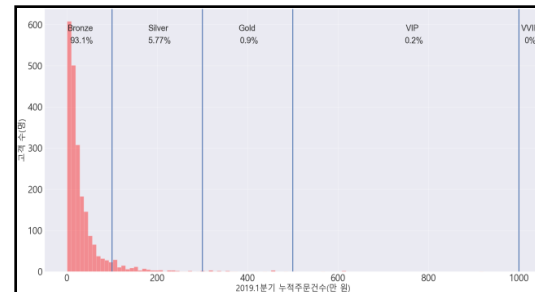
(SOLUTION 1. 세부적인 고객 관리)

1. 현재 등급 제도의 상황과 문제점

현재 등급 제도

- 지난 7월에 처음 도입한 PACKUS의 고객등급 체계
- Bronze부터 VVIP까지 총 5개의 등급 (구매가 이루어지지 않은 '새내기' 제외)
- **매우 불균형적인 구조**

회원 등급	Bronze	Silver	Gold	VIP	VVIP
분기별 누적 결제	~100만	100만~	300만~	500만~	1000만~
2019.1분기 구매 기준	2017명 (93.1%)	125명 (5.77%)	20명 (0.9%)	4명 (0.2%)	0명 (0%)



2. 개편 등급 제도

“등급을 나누는 기준” 수정이 불가피!

* 등급을 나누는 기준 :

- 1) **단위 기간** (현 : 1분기 당)
- 2) **구매 금액** (현 : xx원 이상)



How to Solve?

[개편 시 유의 점]

- 1) 기존 고객층의 혼란 최소화를 위해, '**등급 개수/명칭**' **현행 유지**
- 2) 등급별로 **고른 사람 수**가 들어가도록

2. 개편 등급 제도

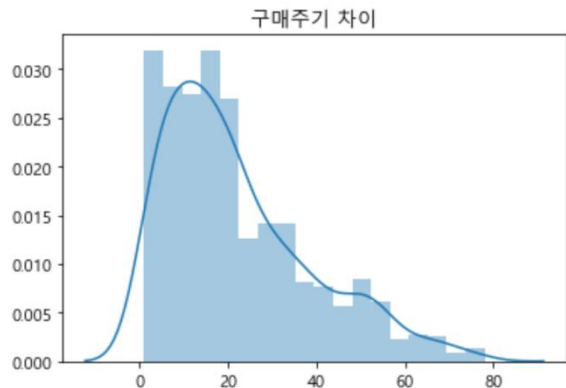
1. 단위 기간

우선, 평균 구매주기 파악!

- 2건 이상 거래가 있는 사람 대상 (995명)
- 고객 별, 가장 최근 2개의 거래의 기간 차이 평균 : **약 22일** (즉, 1분기 = 약 4~5회의 구매)

2019.1분기 전체 고객 수 : 2166

2019.1분기 2건 이상 구매 고객 수 : 995



Conclusion

구매 1회당 등급이 자주 바뀌면 불안정한 고객 등급!

단위 기간은 '1분기' 로 현행 제도 유지

고객 등급 재개편

1-3. 고객 등급 재개편

(SOLUTION 1. 세부적인 고객 관리)

2. 개편 등급 제도

2. 구매 금액

현재 '너무 높은 기준치. 적정 기준치는?

- 수익 극대화 위한 최적의 기준은 파악하기는 Hard
- 안정적인 고객 등급 위해 "고객 등급 별로 일정 %의 인원이 포함되도록"

(Target)

상위 등급으로 갈 수록 인원수가
줄어드는 구조의 산업 benchmarking

bench marking

업권별 신용등급 분포 현황(인원 기준)						
	은행	카드	캐피탈	보험	저축은행	대부업체
1등급	41.6%	0.1%	0.0%	0.5%	0.0%	0.0%
2등급	18.4%	11.0%	2.9%	11.8%	0.1%	0.0%
3등급	12.5%	7.2%	2.8%	5.5%	0.1%	0.0%
4등급	7.0%	20.3%	9.5%	21.5%	1.8%	0.4%
5등급	6.3%	22.5%	16.7%	20.6%	10.8%	4.3%
6등급	4.6%	17.1%	20.8%	16.4%	30.4%	19.6%
7등급	2.6%	7.9%	13.4%	8.7%	19.8%	20.8%
8등급	2.1%	4.2%	9.0%	5.4%	15.0%	18.8%
9등급	2.7%	5.2%	13.1%	5.4%	13.2%	23.3%
10등급	2.0%	4.3%	11.4%	3.8%	8.1%	12.7%

(나이스 평가정보, 2017년 5월 기준)

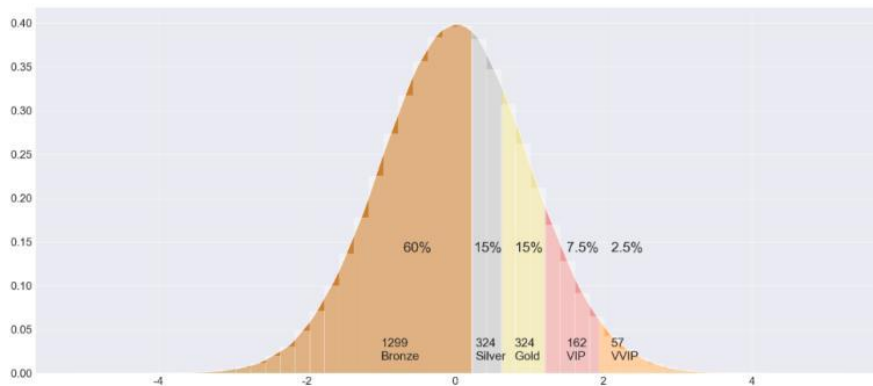
	Bronze	Silver	Gold	VIP	VVIP
등급별 고객 %	60%	15%	15%	7.5%	2.5%

고객 등급 재개편

1-3. 고객 등급 재개편

(SOLUTION 1. 세부적인 고객 관리)

2. 개편 등급 제도



SOLUTION

기준 금액	Bronze	Silver	Gold	VIP	VVIP
기존	~100만	100만~	300만~	500만	1000만~
개편	~25만	25만~	40만~	80만~	180만~

등급	속하게 될 고객 수
Bronze	1299 (60%)
Silver	324 (15%)
Gold	324 (15%)
VIP	162 (7.5%)
VVIP	52 (2.5%)

* 각 등급별 비율을 목표와 같이 유지하려면,
실제 정확한 등급 컷은 차례대로 248,840원(25만), 391,250원(40만), 771,900원(80만), 1,743,620원(180만) 이다.

고객 등급 재개편

1-3. 고객 등급 재개편

(SOLUTION 1. 세부적인 고객 관리)

3. 등급 별 적정 혜택

패커스 등급별 혜택		새내기	BRONZE	SILVER	GOLD	VIP	VVIP
회원 구분		새내기	BRONZE	SILVER	GOLD	VIP	VVIP
본기별 누적 결제	회원가입시	100만원 미만	100만원 이상	300만원 이상	500만원 이상	1천만원 이상	
적립률		0.1%	0.2%	0.3%	0.4%	0.5%	0.6%
마일리지	유효 기간	유효기간 6개월					
할인쿠폰 매월 발급/유효기간 1달	1,000원	1,000원	1,000원	1,000원	1,000원	1,000원	
	1,500원	1,500원	1,500원	1,500원	1,500원	1,500원	
	2,000원	2,000원	2,000원	2,000원	2,000원	2,000원	
	3,000원				3,000원	3,000원	
	5,000원					5,000원	
추가할인혜택						10%할인	
기타 혜택	가입 축하금			2,000원			
	SMS 수신			500원			
	생일 쿠폰			3,000원			
	추천인 등록			5,000원(최대 10회)			



분기당 1000만원 이상 구매한VVIP 고객에게
고작1000원~5000원 쿠폰을?

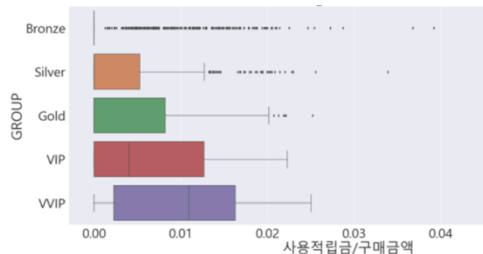
Bronze, Silver, Gold :

- 50%이상이 적립금 사용 X

- 높은 쿠폰이용률

(VIP(91%)보다 높음 (각각 97%, 97%, 93%))

등급 별 구매금액 대비 적립금 사용 비율



Bronze, Silver, Gold : “**쿠폰**” 위주의 혜택

VIP, VVIP : “**적립율**” 위주의 차등적 혜택

1. 고객 세분화 요약

1-1. 고객군 별 전략

어떠한 “방식”의 MKT 방식이 적절한가!

- Clustering 통한 고객군 분류
- 주요 지표들을 통해 본 고객군 별 특징

1-2. 장바구니 분석

어떠한 군집에 “어떠한 물품”을 추천해줘야 하는 가?

- 군집별 ‘많이 사는 물품’ & ‘함께 사는 물품’

1-3. 고객 등급 재개편

현재 PACKUS의 고객등급 체계를 어떻게 개편할 것인가?

- 등급 별 ‘기준 금액’ 과 ‘등급에 따른 혜택’ 수정



SOLUTION 2. 제품 별 수요예측

1) 데이터 소개 2) EDA 3) 모델링 4) 활용방안

SOLUTION 2. 상품별 수요예측



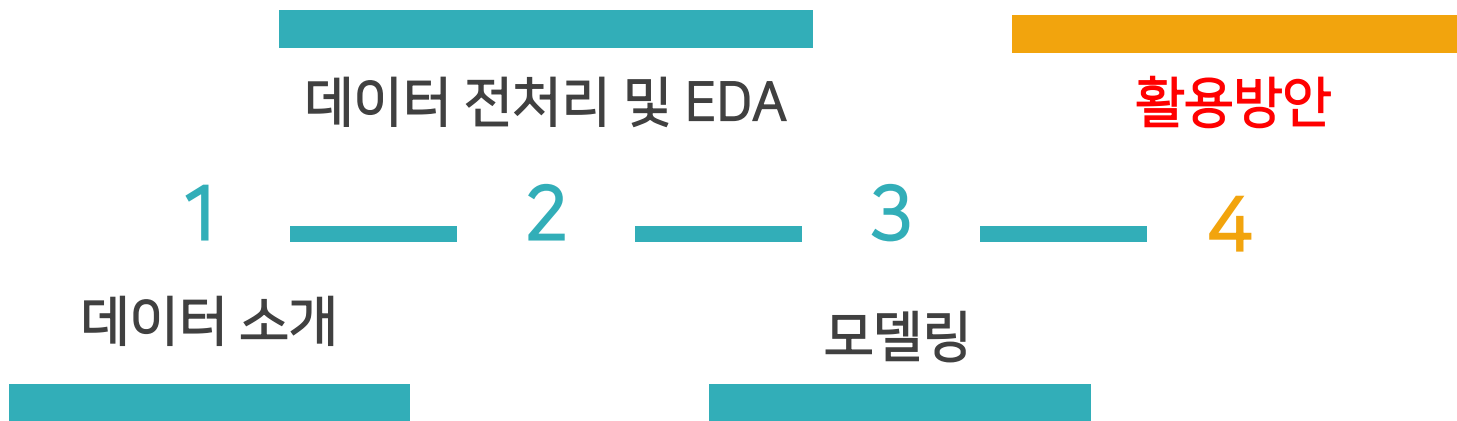
How to Solve?

예측한 판매량을 토대로 적절한 수량의 재고 유지



상품별로 다음 주기에 팔릴 판매량을 미리 예측

SOLUTION 2. 제품 별 수요예측



1. 데이터 소개

01 주문 내역 DATA



02 상품 정보 DATA



03 외부 DATA

DB분석_서버1_2017년_고객 주문 통합 리스트.xls

DB분석_서버1_2018년_고객 주문 통합 리스트.xls

DB분석_서버1_2019년 상반기 고객 주문 통합 리스트.xls

DB분석_서버2_상품리스트.xls

상품 카테고리(패커스 홈페이지에서 수집)

삼릉물산 주가 데이터 (네이버 금융 사이트에서 수집)

SKT 서울시 배달업종별 이용통화량

2. 상품별 수요예측

2. 데이터 전처리 및 EDA

2. 상품별 수요예측

01 주문 내역 DATA

- 주어진 고객 주문 통합 리스트 데이터 3개를 합침
- 데이터가 온전치 못한 2019년 6월 제외
 - ➔ 분석 대상 기간 : 2017.01.02 ~ 2019.05.31
- 상품번호를 기준으로 **상품 구매금액**과 **판매 수량**을 파악

(사용한 변수 : '결제일시(입금확인일)', '상품번호', '주문상품명', '상품구매금액', '수량', '취소구분', '판매가')

	결제일시(입금확인일)	상품번호	주문상품명	상품구매금액	수량	취소구분	판매가
0	2017-01-02 16:38:58	45.00	105파이 다용도컵 세트 5호 대 370ml 백색 PS 1,000개 1set(2박스)	66500	1	취소안함	66500
1	2017-01-15 01:24:43	67.00	소량 105파이 다용도컵 세트 5호 대 370ml 투명 PS 100개	17000	1	취소안함	17000
2	2017-01-18 18:38:50	46.00	105파이 다용도컵 세트 5호 특대 400ml 백색 PS 1,000개 1set(2박스)	73700	1	취소안함	73700

2. 데이터 전처리 및 EDA

01 주문 내역 DATA

[주문 내역 filtering]

4종류의 주문 내역 제외!

1. 취소된 주문
2. 사은품
3. 수령인 주소 X
4. 상품번호 X

38190건 대상으로 분석

2. 상품별 수요예측

[물품 filtering]

201905	258
201904	58
201901	27
201806	25
201902	21
201812	21
201801	20
201903	18
201804	18
201811	17
201712	16
201906	15

총 상품 : 1231개

실질적으로 판매 : 628개

최근까지 잘 팔리는 상품 : 331개

마지막 판매일이 2019 년 4,5,6 월인 상품

→ 최근까지도 잘 팔리고 수요가 많은 제품

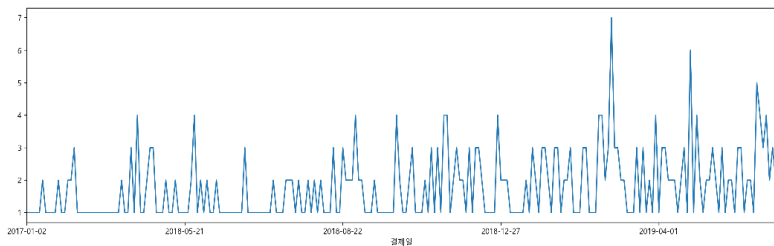
"최종 : 331 개의 상품"

2. 데이터 전처리 및 EDA

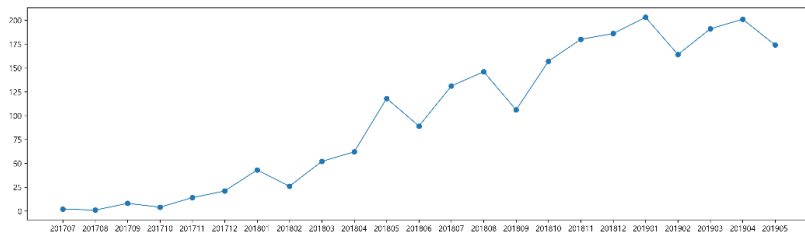
2. 상품별 수요예측

01 주문 내역 DATA

상품번호 45번의 **일별** 매출 현황



상품번호 45번의 **월별** 매출 현황



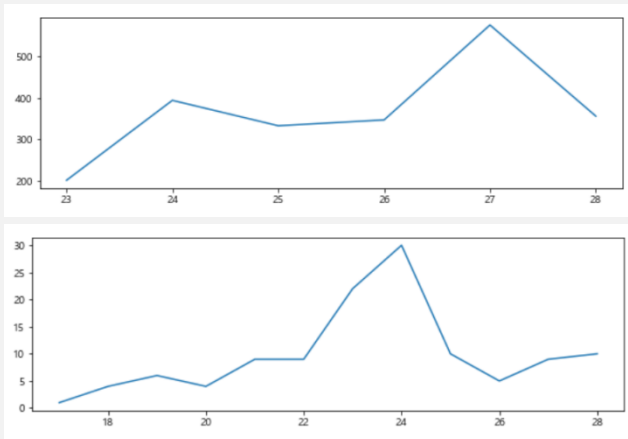
다수의 상품이 하루 판매량이 0인 날들이 많아
하루 단위 예측 보다는 **“월 단위 예측”**이 더 의미 있을 것으로 판단
➔ 월별 상품별로 묶은 이후 분석을 진행

2. 데이터 전처리 및 EDA

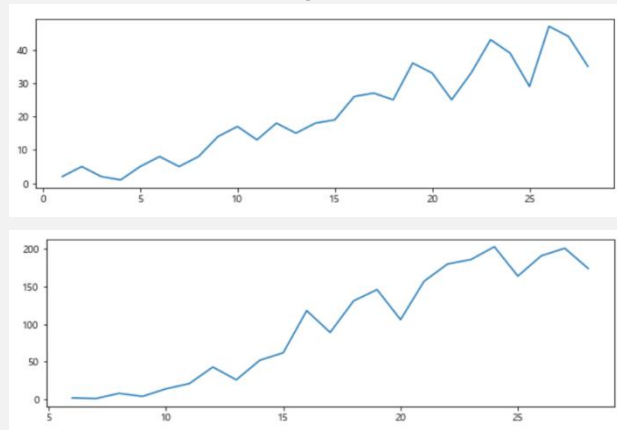
2. 상품별 수요예측

01 주문 내역 DATA

신상품 판매추이



기존 상품 판매추이

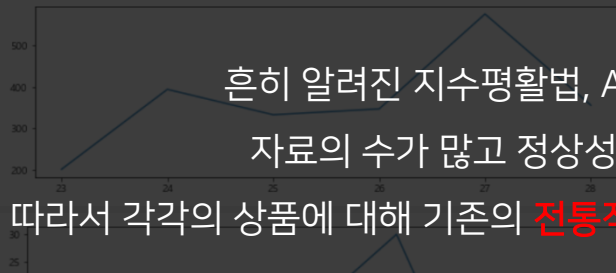


팔리기 시작한 지 얼마 되지 않은 신상품들은 자료의 수가 너무 적음
그러나 대부분 증가하는 경향을 보임 & 이전부터 팔렸던 상품의 초기 판매량과 비슷한 경향
➔ 앞으로도 비슷한 수요를 보일 것으로 예상

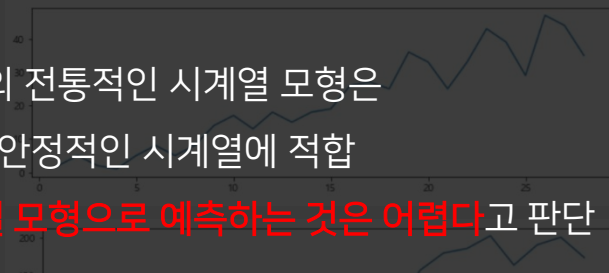
01 주문 내역 DATA



신상품 판매추이



기존 상품 판매추이



흔히 알려진 지수평활법, ARIMA 등의 전통적인 시계열 모형은
자료의 수가 많고 정상성을 보이는 안정적인 시계열에 적합
따라서 각각의 상품에 대해 기존의 **전통적인 시계열 모형으로 예측하는 것은 어렵다**고 판단

다양한 변수를 더 추가하여 하나의 모델로
여러 개 상품에 대한 분석이 가능한 **Machine Learning 모델**을 사용

그러나 대부분 증가하는 경향을 보임 & 이전부터 팔렸던 상품의 초기 판매량과 비슷한 경향
➔ 앞으로도 비슷한 수요를 보일 것으로 예상

2. 데이터 전처리 및 EDA

02 상품정보 DATA

1. 패커스 홈페이지에서 상품 카테고리 정보 크롤링



2. 기본적으로 한 상품에 여러 개의 카테고리가 있음 → 카테고리 이항변수화

주문상품명	best	chinese	japanese	korean	western	sealing	bakerycafe	agri	kitchen	sticker	small
원형 210mm 전자렌지 직화 냄비 1,450ml 200개 1박스	0	0	1	1	1	0	0	0	0	0	0
알루미늄 일회용 캠핑용기 세트 1 SET	0	0	1	1	1	0	0	0	0	0	0
198mm (7치) 대나무젓가락 1,500개 골내젓갈	0	0	0	0	0	0	0	0	1	0	0
반찬용기 D-1616 1칸 투명 200개 1set(2박스)	0	1	1	1	1	0	0	0	0	0	0

배달용기 전문쇼핑몰
PACKUS

≡ 카테고리 전체보기

식자재

초특가 SALE

NEW

BEST상품

한식(분식) 용기

중식 용기

일식 용기

양식 용기

실링제품

베이커리/카페 용기

농수산물 용기

주방용품/소모품

스티커 제작

소량

2. 데이터 전처리 및 EDA

03 외부 DATA

거시 경제 및 외부 환경 변화 등을 반영할 수 있는 data를 추가로 수집

- 삼릉물산 주가 (출처 : 네이버 금융 사이트)

패커스의 모회사인 삼릉물산의 주가를 추가

- SKT 배달업종별 이용 통화량

포장용기의 주 사용처인 배달업종의 이용통화량

2. 상품별 수요예측



새로운 변수 추가

(1) Target lags

n 시차전의 판매량을 변수로 추가

최근 경향과 장기 경향을 반영하기 위해 각각 (1, 2, 3 시차), (6, 12 시차)를 선택(='item_cnt_lag_#')

(2) 3 개월 이동 평균 및 3 개월 이동 표준편차

(='rolling_mean_3', 'rolling_std_3')

(3) Trend feature

각 상품의 판매가 변화율을 반영하기 위한 변수

각 상품의 전체기간 평균판매가에서 **한 시차 전의 판매가가 얼마나 변했는지**(='delta_price_lag')

한 시차 전의 판매가로부터 **현재 판매가가 얼마나 변했는지**(='delta_price')

(4) 주가

패커스의 모회사인 삼릉 물산의 **한 시차전의 주가**. (='stock_lag_1')

2. 데이터 전처리 및 EDA

2. 상품별 수요예측

새로운 변수 추가

- (5) 카테고리 별 평균판매량
해당 카테고리의 한 시차 전 평균 판매량(='date_cat_(카테고리명)_lag_1')
- (6) 처음판매가 된 기준년월
(='item_first_sale')
- (7) 해당월 / 해당월의 일수
(='month', 'days')

생성된 최종 DATASET

item_cnt_lag_1	item_cnt_lag_2	item_cnt_lag_3
2.00	5.00	3.00
3.00	6.00	3.00
1.00	1.00	0.00

...

rolling_std_3	rolling_mean_3	delta_price_lag	delta_price	month	days	item_first_sale
1.53	3.33	0.07	0.00	0	31	0
1.73	4.00	-0.03	0.00	0	31	0
0.00	1.00	-0.02	0.00	0	31	1



판매량

item_cnt
9
6
2

3. 모델링

회귀 문제로 접근!

목적

다음 달 “판매량” 예측

2. 상품별 수요예측

사용한 모델

Boosting

1. Light GBM
2. XGBoost

Bagging

3. Random Forest
4. Extra tree

3. 모델링

시계열 Cross-Validation



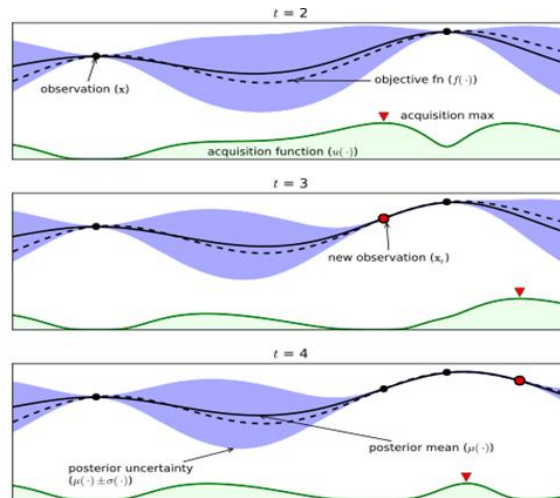
Train set Valid set Test set

test data : 2019년 5월의 판매량
cross validation score 구하기

2. 상품별 수요예측

Hyperparameter Tuning

조정할 수 있는 Tree Model에 대하여
Bayesian Optimizer를 이용하여 최적화



3. 모델링

2. 상품별 수요예측

모델 평가

- cross-validation 결과 (2019.02 ~ 04 예측 평균)

	LGBM	XGB	RF	Extra
RMSE	20.94	20.35	15.67	15.71
R2	0.77	0.77	0.87	0.87

- test set 예측 결과 (2019.05 예측)

	LGBM	XGB	RF	Extra
RMSE	10.04	7.74	14.68	13.07
R2	0.93	0.96	0.86	0.89

Boosting 모델(LGBM, XGB)은 outlier에 취약

➔ 아직은 안정적이지 않은 PACKUS의 수요 변화에 부적절

Bagging 모델(RF, Extra)은 일반화 성능이 좋음

➔ CV score & Test score 결과 차이가 크지 않은

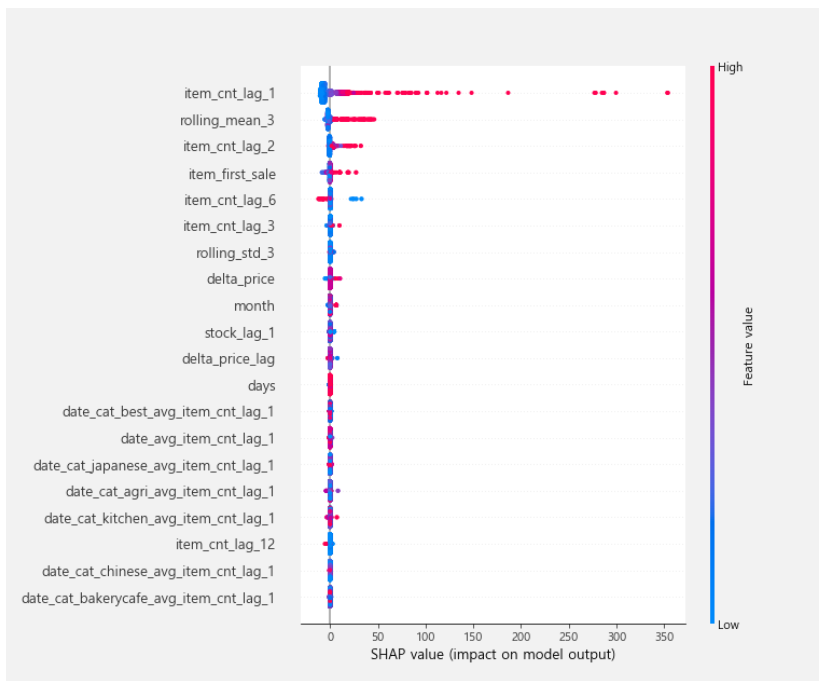
Random Forest / Extra Tree model을 사용하는 것이 타당

➔ 앞으로의 예측에 있어서도 좋은 일반화 성능을 보일 가능성 높음

3. 모델링

중요 변수 - SHAP

Random forest



2. 상품별 수요예측

해석방법

- 빨간색 : 해당 변수 값이 큼, 파란색 : 해당 변수값이 작음
- 왼쪽 : 예측값에 - 영향을 줌, 오른쪽 : 예측값에 + 영향을 줌

기업 입장에서 통제 가능한 변수들(판매가격의 변화) 보다,

통제 가능하지 않은 변수들(한, 두 시차 전의 판매량)이 높은 중요도!

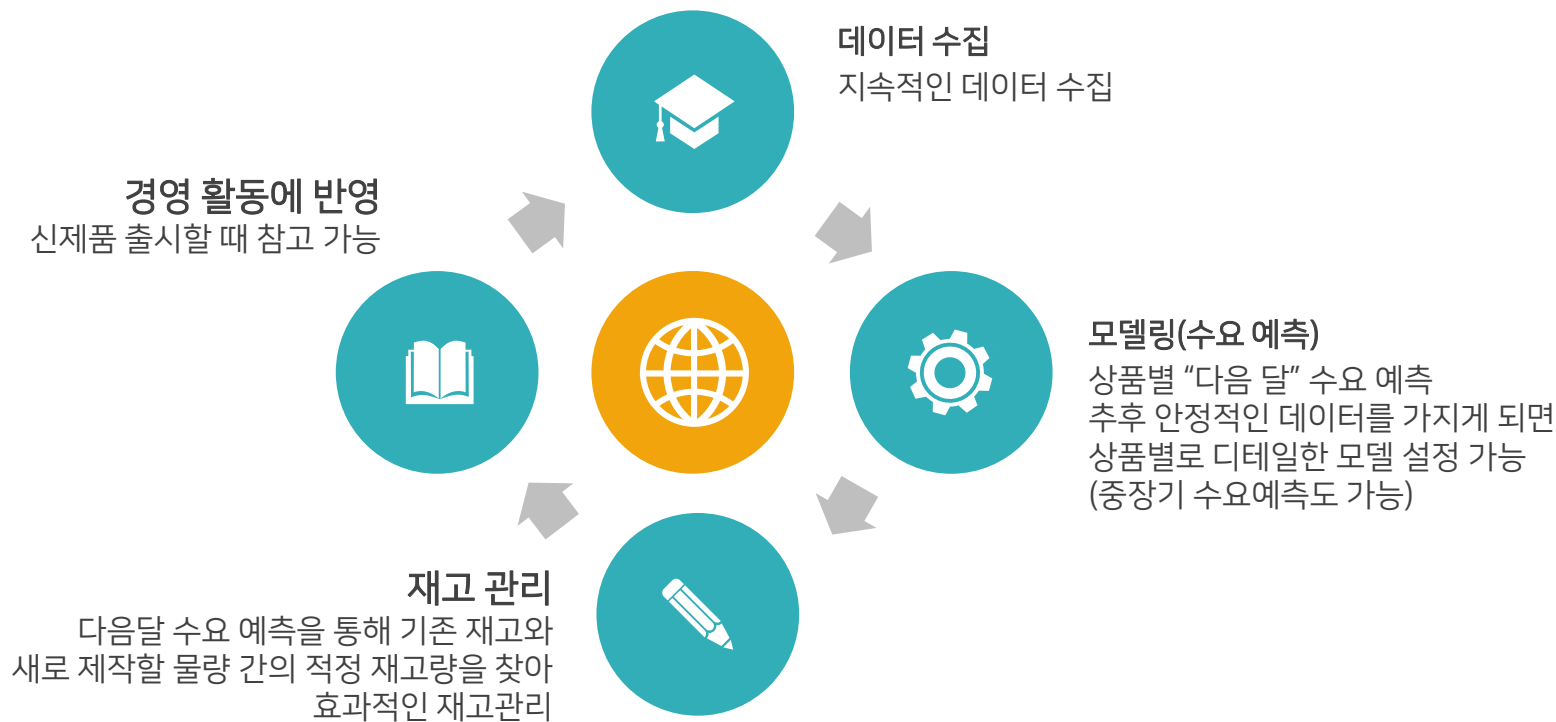
[결론] 다른 변수들의 변화를 유도하여 판매량을 늘리는 전략보다,

각각 품목의 변화 추이를 파악하여 이를 토대로 추후 판매량을 예측

하여 재고비용을 최소화!

4. 활용방안

2. 상품별 수요예측





4. PACKER board

SOLUTION 1&2를 효과적으로
관리하기 위한 고객관리 대시보드

PACKER board

PACKER : PACKUS의 고객

[SOLUTION 1 : 고객 관리]

고객군 분류를 통한 세부적 고객 관리

[SOLUTION 2 : 재고 관리]

상품별 수요예측을 통한 적정재고 유지





Thank you