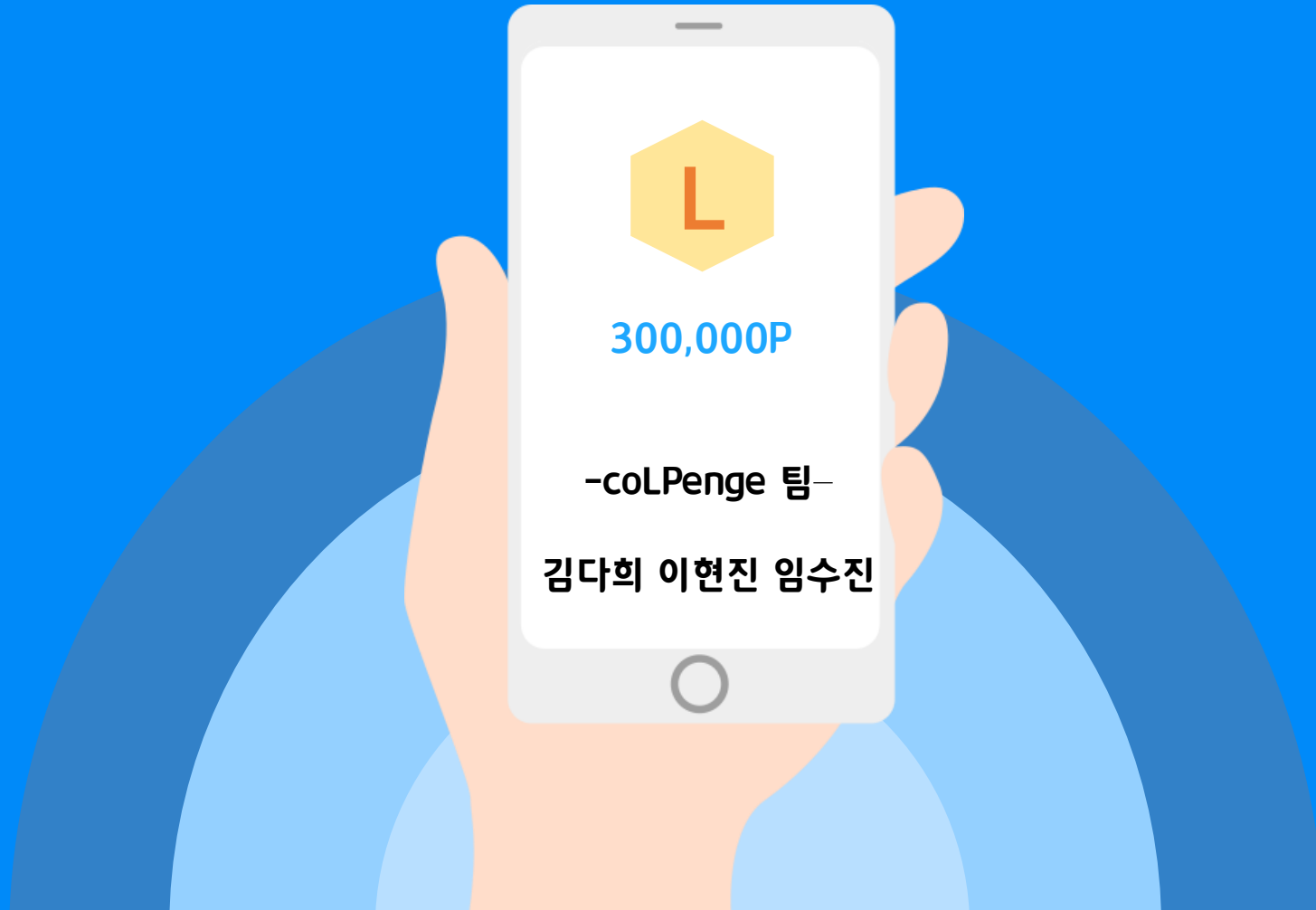


L.Point | L.Pay 고객 인사이트 도출 및 상품 추천과 재구매 예측 대시보드 구현



CONTENTS

1



주제 선정 배경

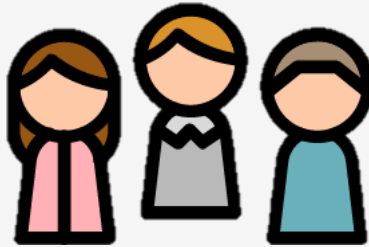
2



데이터 및 EDA

- CLTV 추정
- 코호트 분석

3



고객군 군집화

- RFM 분석
- 클러스터링

4



고객별
상품 추천

- 인기 기반 추천
- ALS 기반 추천

5



고객별
재구매 예측

- LSTM
- ARIMA
- Prophet

6



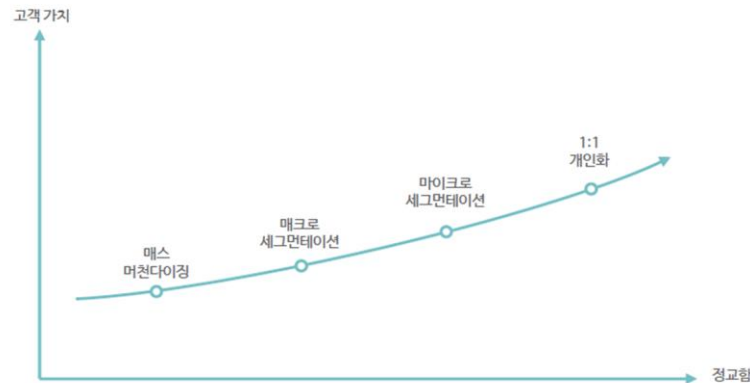
마케팅 전략 도출
및 결과 활용 방안



1. 주제선정 배경

초개인화 마케팅 시대의 도래

마케팅 트렌드의 변화

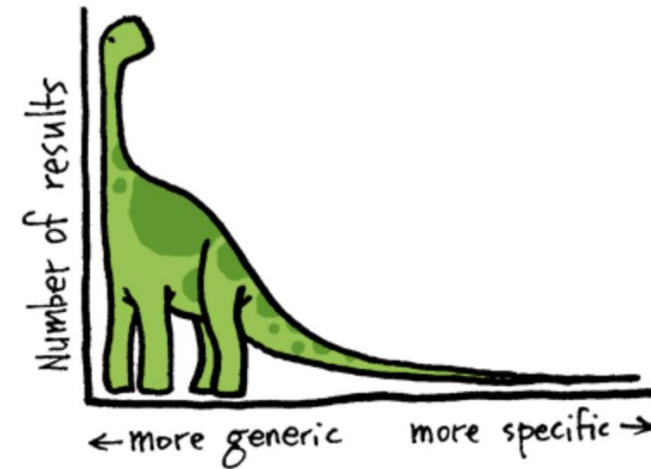


소비습관의 변화와 고객 가치 변화로 인한
1:1 초개인화 마케팅 필요성의 증가



개인의 L.pay 구매 기록 데이터를 통해
상품 추천과 재구매 예측 모델을 개발하고
개인 고객 관리를 위한 대시보드 구현

우리에게 버릴 고객은 없다



개인화가 대두되며 단순히 파레토 법칙 및 롱테일 법칙 등
일부 고객에 대한 한정성이 아닌 모든 잠재 고객의 가치가 중요



고객 전체를 대상으로 군집별 특징 자체를 파악하여
개인을 위한 맞춤형 마케팅 전략 도출

1. 주제선정 배경

초개인화 마케팅 시대의 도래

우리에게 버릴 고객은 없다

L.Point | L.Pay 고객 인사이트 도출 및 상품 추천과 재구매 예측 대시보드 구현

출처: Boxever & Tnooz(2015,6), A Brief History of Personalization

소비습관의 변화와 고객 가치 변화로 인한
1:1 초개인화 마케팅 필요성의 증가



개인의 L.pay 구매 기록 데이터를 통해
상품 추천과 재구매 시기 예측 모델을 개발하고
개인 고객 관리를 위한 대시보드 구현

← more generic more specific →

개인화가 대두되며 단순히 파레토 법칙 및 롱테일 법칙 등
일부 고객에 대한 한정성이 아닌 모든 잠재 고객의 가치가 중요



우선순위를 떠난 고객 군집별 특징 자체를 파악하여
군집 내 개개인을 위한 맞춤형 마케팅 전략 도출

2-1. 데이터 설명

Raw Data



고객 데모 정보



상품 구매 정보



제휴사 이용 정보



상품 분류 정보



점포 정보



엘페이 이용



products.csv

점포 종류가 유통사(A)인 상품 거래 데이터

L.pay 거래를 한 번이라도 한 고객을 표시하는 Lpay_client 열 추가

L.pay 거래를 여부를 표시하는 Lpay 열 추가

고객별 상품 추천 및 재구매 시기 예측과 final.csv 생성에 활용



services.csv

점포 종류가 유통사가 아닌(B~E) 거래 데이터

L.pay 거래를 한 번이라도 한 고객을 표시하는 Lpay_client 열 추가

L.pay 거래를 여부를 표시하는 Lpay 열 추가

final.csv 생성에 활용



final.csv

products.csv와 services.csv의 공통 열만 추출하여 생성

고객 생애 가치 계산 및 고객군 세분화에 활용

2-1. 데이터 설명

products.csv

영수증번호	고객번호	성별	연령대	거주지대분류코드	상품코드	대분류명	중분류명	소분류명	구매일자	구매시간	구매수량	구매금액	제휴사(유동사만 A01~06)	점포코드	점포지역대분류	점포지역중분류	채널구분(1:오프, 2:온)	Lpay_olient	Lpay
A04000035215	M355489415	남성	20대	Z10	PD1173	생수	음료	생수	20210101	0	3	2850.0	A04	A043757	Z10	Z10001	1	1	0
A04000035215	M355489415	남성	20대	Z10	PD0223	사탕/캔디	과자	젤리	20210101	0	1	1000.0	A04	A043757	Z10	Z10001	1	1	0
A04000035215	M355489415	남성	20대	Z10	PD1918	스킨케어	화장품/뷰티케어	페이셜클렌저	20210101	0	2	25800.0	A04	A043757	Z10	Z10001	1	1	0
A04000035215	M355489415	남성	20대	Z10	PD0223	사탕/캔디	과자	젤리	20210101	0	1	1000.0	A04	A043757	Z10	Z10001	1	1	0

4381743 rows x 20 columns

services.csv

영수증번호	고객번호	성별	연령대	거주지대분류코드	방문일자	이용일자	이용시간	이용금액	제휴사(B-E: 숙박/엔터/F&B/렌탈)	점포코드	점포지역대분류	점포지역중분류	채널구분(1:오프, 2:온)	Lpay_olient	Lpay	
21123122C023392290	M481349058	여성	40대	Z17	20211231	20211231	22	94000		C02	C020002	Z17	Z17018	1	0	0
21123122D013417945	M421637803	여성	20대	Z10	20211231	20211231	22	11100		D01	D010717	Z10	Z10046	1	0	0
21123122E012916542	M716433837	남성	30대	Z17	20211231	20211231	22	17770		E01	E010001	Z17	Z17001	1	0	0
21123123D010326849	M449095124	여성	20대	Z10	20211231	20211231	23	25300		D01	D010717	Z10	Z10046	1	1	0

248304 rows x 16 columns

final.csv

고객번호	성별	연령대	거주지대분류코드	구매일자	구매시간	구매금액	제휴사	점포코드	점포지역대분류	채널구분(1:오프, 2:온)	Lpay	Lpay_client
M355489415	남성	20대	Z10	20210101	0	2850.0	A04	A043757	Z10	1	0	1
M355489415	남성	20대	Z10	20210101	0	1000.0	A04	A043757	Z10	1	0	1
M355489415	남성	20대	Z10	20210101	0	25800.0	A04	A043757	Z10	1	0	1
M355489415	남성	20대	Z10	20210101	0	1000.0	A04	A043757	Z10	1	0	1

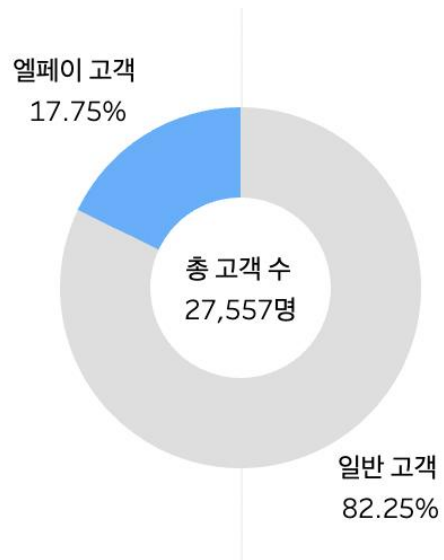
4630047 rows x 14 columns

2-2. EDA

L.pay 고객 비율

엘페이 고객 : 한 번이라도 L.pay를 이용한 거래를 한 고객

일반 고객 : L.pay거래를 한 번도 하지 않은 고객



L.pay 거래 비율

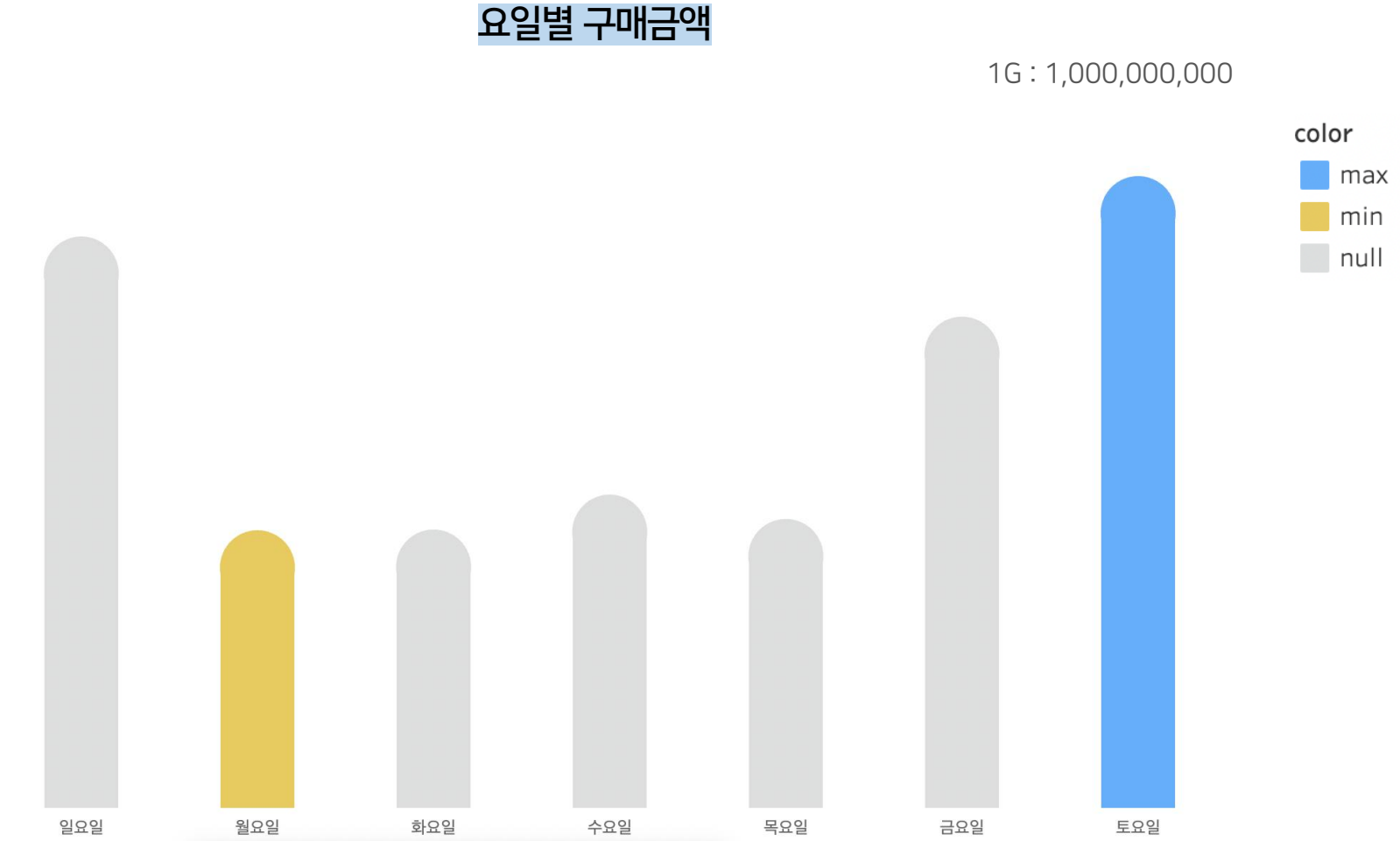
엘페이 거래 : L.pay를 이용한 거래

일반 거래 : L.pay를 이용하지 않은 거래



L.pay 거래 건수가 매우 적고, 엘페이 고객들 중에서도 엘페이 거래보다는 일반 거래를 많이 한다는 것을 알 수 있음

2-2. EDA



토요일의 구매금액 총 합이 가장 크고 월요일에 가장 적음. 주말이 시작하는 요일과 새로운 주가 시작하는 요일의 차이가 있는 것 확인 가능

2-2. EDA

구매 시간에 따른 구매 금액과 거래 건수

1G : 1,000,000,000

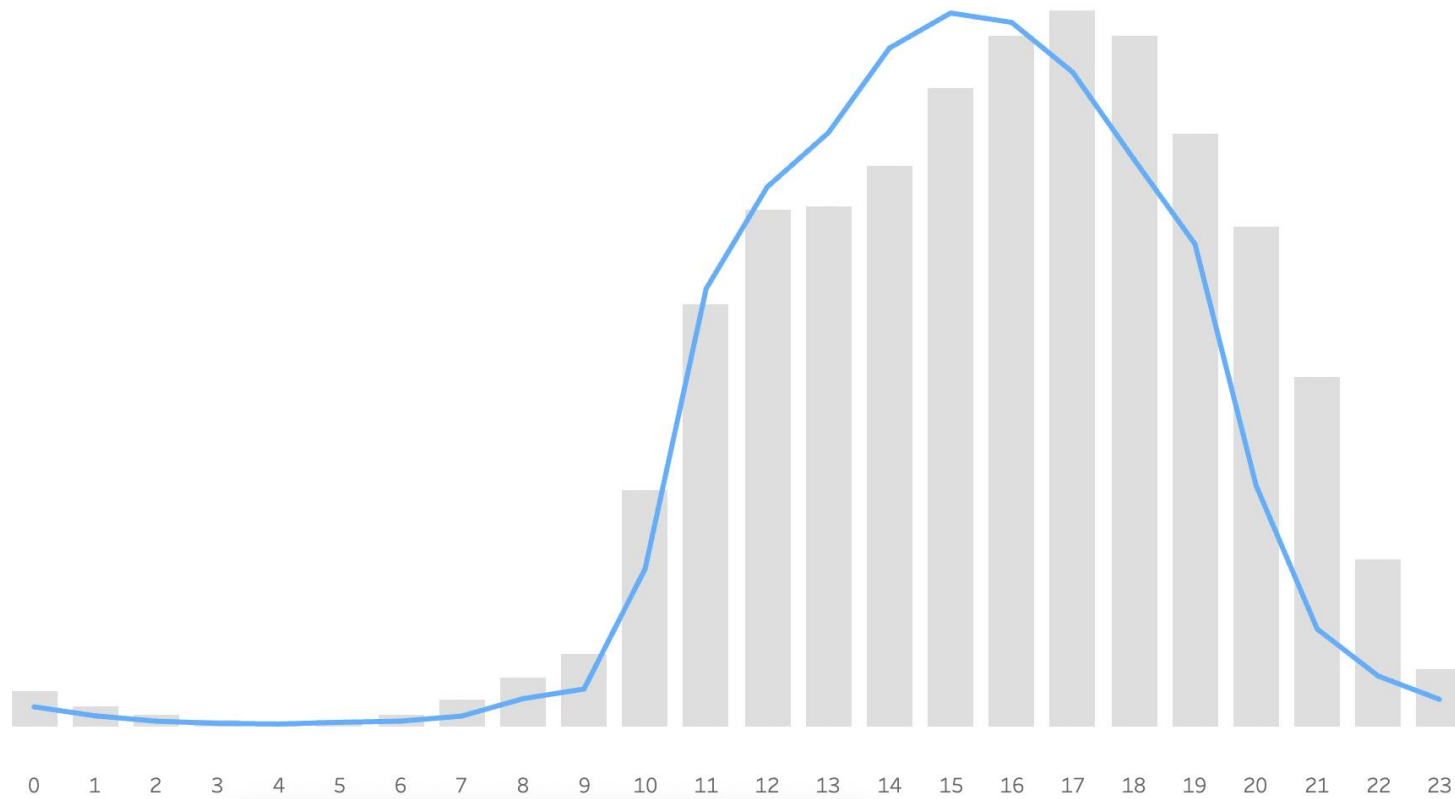
측정값 이름

■ 거래건수

■ 구매금액

거래 건수

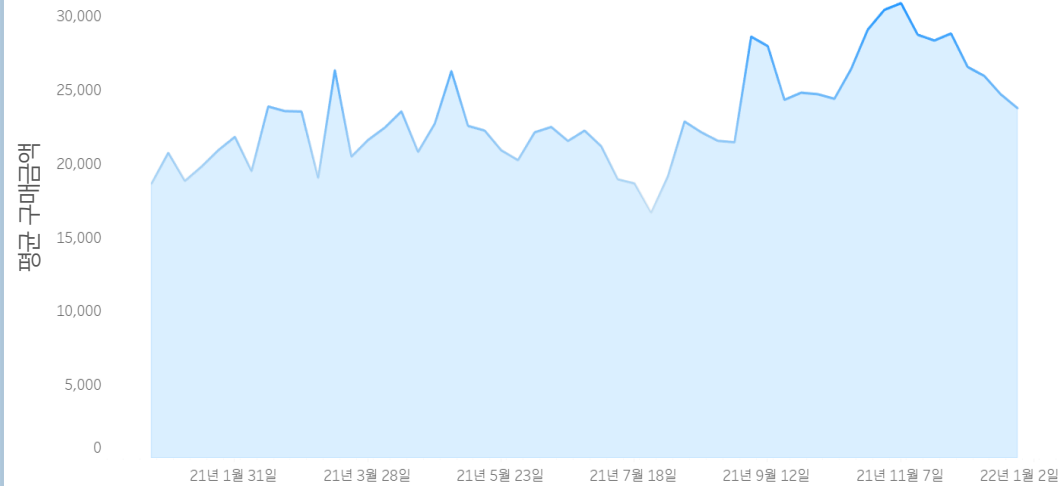
구매 금액



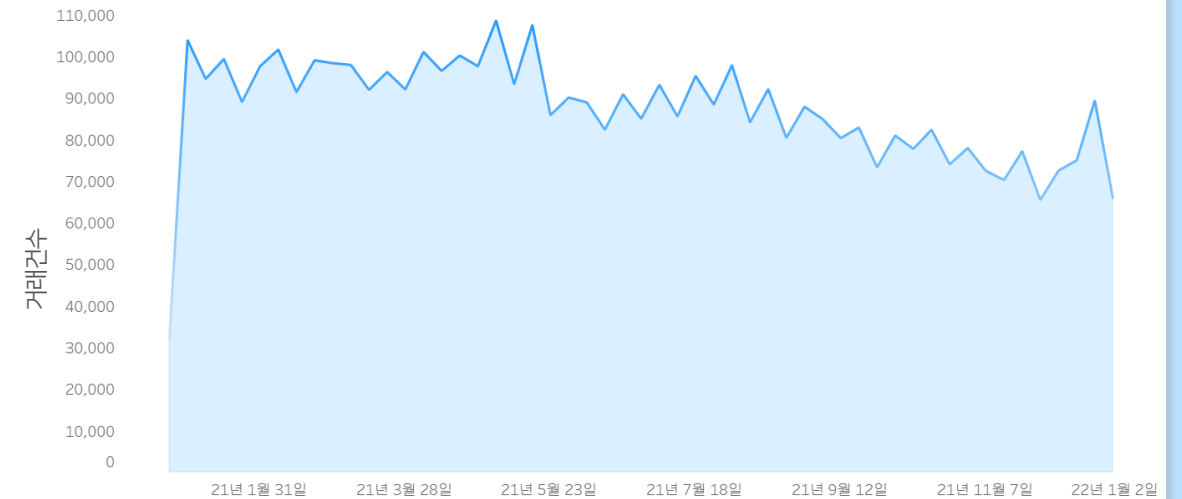
오후 2시 ~ 오후 6시 사이에 거래가 가장 활발함. 구매금액은 오후 2시에 가장 크고 거래 건수는 오후 5시에 가장 많음

2-2. EDA

주별 구매 금액



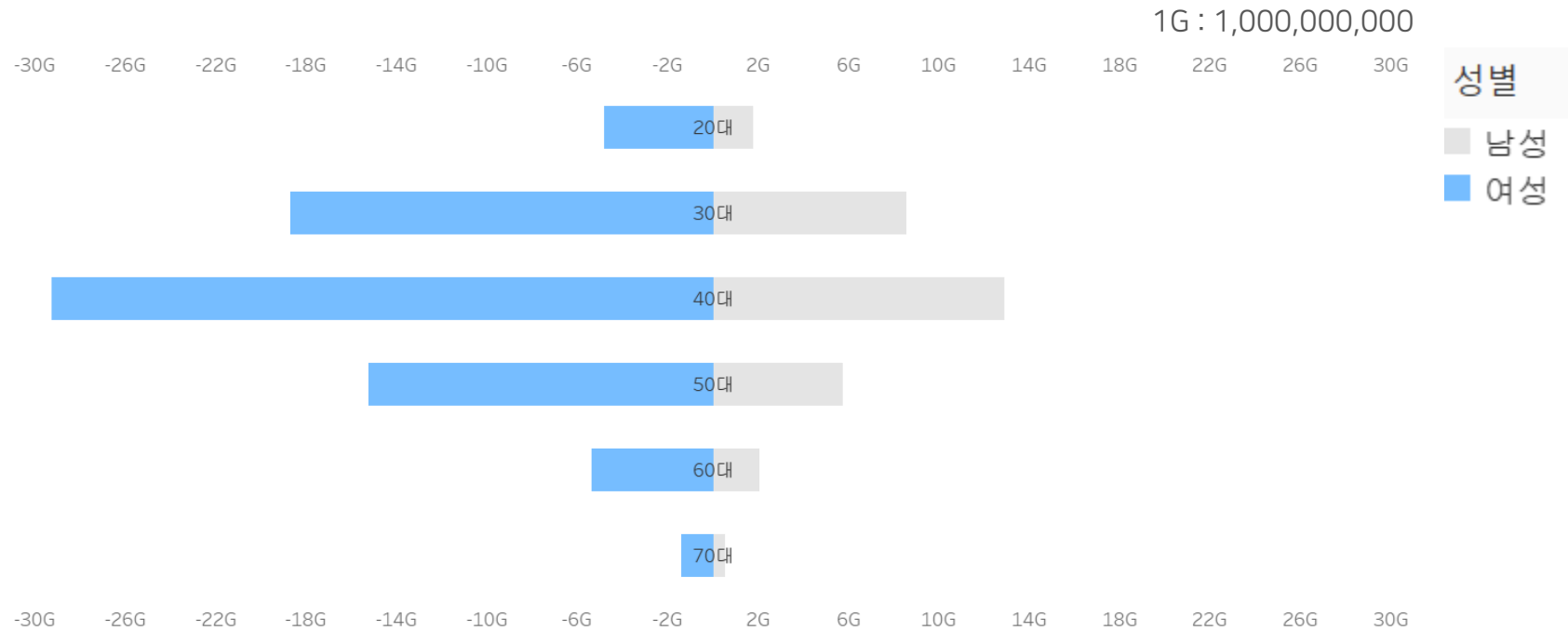
주별 거래 건수



주별 평균 구매 금액은 증가하는 추세이지만 거래 건수는 감소하는 추세
특히 21년 7월 18일 주의 평균 구매 금액이 다른 주에 비해 매우 적음
이 날의 거래 건수에 대한 특이사항은 없으므로 저가의 상품이 많이 구매되었을 확률이 높음

2-2. EDA

연령대와 성별에 따른 구매금액 합계



주로 여성의 구매가 활발하고, 40대의 구매가 두 성별 모두 가장 활발함을 알 수 있음

2-2. EDA

온라인 거래와 오프라인 거래 건수 비교



채널구분

오프라인

온라인



제휴사별 구매 금액 합계 비교

평균(구매금액)

3,296

329,103

A05

B01

A01

A06

E01

C02

D01

A02

C01

D02

오프라인 구매가 대부분이며, 서비스 업종보다는 상품(A)의 구매 비중이 큼

상품 중에서도 제휴사 A05 구매 금액이 가장 크고 서비스 업종 중에서는 숙박업종 B01의 구매 금액이 가장 큼

2-3. 고객 생애 가치(Customer Lifetime Value; CLV) 추정

고객 생애 가치 (CLV)

고객이 비즈니스에 기여한 금전적 가치로, customer relationship이 수익성이 있는지 판별하고자 사용

보편적인 CLV 계산 방법

= (평균 구매 금액 × 총 마진 × 구매 빈도 × 고객 수명) - 고객획득비용

+



CLV 추정을 위한 Lifetimes library (python)

final.csv 데이터 이용

RFM 분석

'기업 매출에 가장 중요한 요소는 최근성, 행동 빈도, 구매금액이다'라는 가정을 바탕으로 3가지의 관점에서 고객의 가치를 분석하는 방법

Recency, Frequency, Monetary

+

Recency : 고객의 첫 구매로부터 마지막 구매 사이의 기간

T : 고객의 첫 구매로부터 연구 기간 끝 (2021.12.31) 사이의 기간

Frequency : 고객의 재구매 빈도 (상품 구매 횟수)

Monetary : 고객의 평균 구매 금액

2-3. 고객 생애 가치(Customer Lifetime Value; CLV) 추정

BG-NBD 모델 생성

개별 고객의 이탈 확률은 기하 분포를 따르고, 고객 간 이탈의 이질성은 베타 분포를 따른다고 가정하는 BG(Beta-Geometric mix) 모델을 사용
생존 기간 동안 개별 고객의 구매 빈도는 포아송 분포를 가정하며, 고객 간 재구매 차이는 감마 분포를 가정

GAMMA-GAMMA 모델 생성

고객별 CLV를 추정하기 위해 사용한 모델

실제로 금전적 가치(Monetary)와 구매 빈도(Frequency)간에 아무런 관계가 없다고 가정

[상관관계 확인]

	Monetary	Frequency
Monetary	1	- 0.032
Frequency	- 0.032	1

→ Monetary와 Frequency 간 상관 계수가 0에 가까우므로 두 변수 사이의 관계가 없다고 판단

2-3. 고객 생애 가치(Customer Lifetime Value; CLV) 추정

BG-NBD, GAMMA-GAMMA 모델 기반 CLV 추정

'재구매 고객'에 대한 분포 파악을 목적으로 Frequency가 0보다 큰 고객들을 대상으로 선정 (총 27,553명)

파라미터 조정을 통해 일주일간의 CLV를 추정

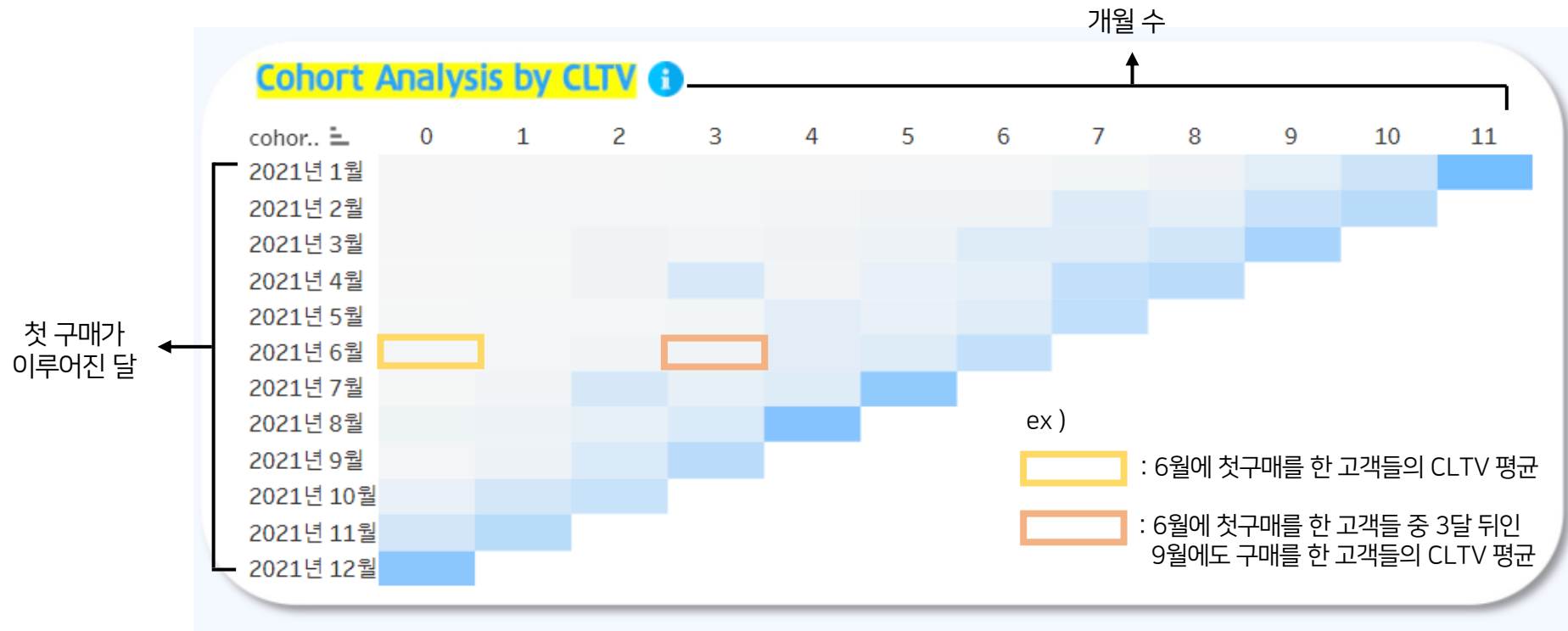
MinMaxScaler를 통해 0부터 1사이의 값으로 CLV 분포 확인

	고객번호	frequency	recency	T	monetary_value	clv	scaled_clv
9531	M343264688	116.0000	360.0000	361.0000	3532795.4741	223705756.3888	1.0000
10655	M384121563	88.0000	359.0000	364.0000	4471356.6818	213370507.4982	0.9538
1564	M055769504	36.0000	325.0000	344.0000	10016366.6667	208202549.8953	0.9307
8019	M288629527	69.0000	357.0000	362.0000	5341590.7246	201324260.7645	0.9000
7433	M268374518	88.0000	343.0000	364.0000	4240110.6818	197871857.8603	0.8845

2-3. 고객 생애 가치(Customer Lifetime Value; CLV) 추정

코호트 분석 (Cohort Analysis)

사용자를 그룹으로 분류하여 그룹의 행동과 유지율을 분석할 때 활용하는 기법



2-3. 고객 생애 가치(Customer Lifetime Value; CLV) 추정

코호트 분석 (Cohort Analysis)

사용자를 그룹으로 분류하여 그룹의 행동과 유지율을 분석할 때 활용하는 기법

Cohort Analysis by CLTV ⓘ

cohor..	0	1	2	3	4	5	6	7	8	9	10	11	
2021년 1월	14,042	24,548	46,430	68,922	34,588	141,960	99,270	164,790	316,357	745,220	1,395,432	4,264,370	→ 감소하는 구간이었지만 증가하는 경향, 8월부터 급격히 증가
2021년 2월	28,171	135,236	132,603	117,412	174,871	244,406	258,111	870,058	662,055	1,548,886	2,111,867		→ 감소하는 구간이었지만 증가하는 경향, 8월부터 급격히 증가
2021년 3월	14,886	78,717	263,917	181,942	264,539	345,350	861,423	820,453	1,267,270	2,609,244			→ 감소하는 구간이었지만 증가하는 경향, 4,5월 급격히 증가, 8월부터 급격히 증가
2021년 4월	19,062	83,627	277,753	1,053,258	219,195	456,921	591,166	1,746,568	2,055,504				→ 7월까지 급격히 증가 / 8월에 급격히 하락 / 다시 9월부터 증가
2021년 5월	67,533	111,700	154,330	170,492	688,911	488,269	774,244	1,814,519					→ 8월부터 급격히 증가
2021년 6월	141,093	120,968	204,917	215,662	716,376	839,346	1,733,175						→ 9월부터 급격히 증가
2021년 7월	55,728	222,171	1,111,039	545,037	845,418	3,366,616							→ 9월까지 급격히 증가 / 10월에 급격히 하락 / 12월에 급격히 증가
2021년 8월	302,335	330,227	528,958	958,991	3,759,765								→ 10월부터 급격히 증가
2021년 9월	152,349	347,037	938,856	2,048,700									→ 급격히 증가하는 경향
2021년 10월	497,319	1,185,555	1,588,502										→ 급격히 증가하는 경향
2021년 11월	1,246,411	2,069,158											→ 급격히 증가하는 경향
2021년 12월	3,538,064												

[정리]

전반적으로 CLTV는 증가하는 경향을 보임

대부분의 그룹에서 8월부터 12월까지 CLTV가 급격히 증가하는 경향을 보임

일부 그룹에서 8월, 10월에 급격히 CLTV가 하락하는 경향을 보임

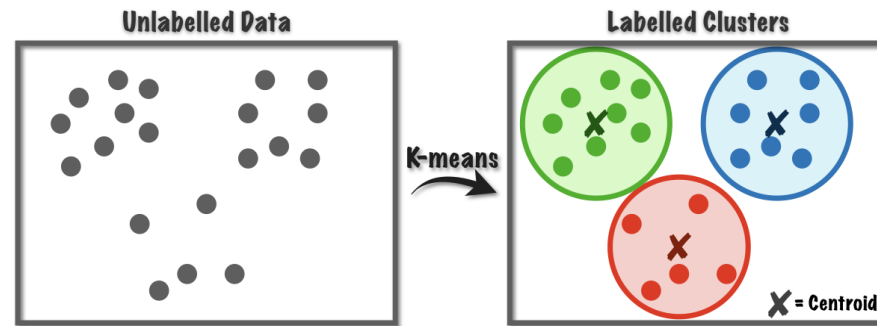
특히 2021년 4월 첫 구매 집단의 3개월 뒤 CLTV가 급격히 증가함

3. 고객군 군집화

K-means Clustering

주어진 데이터를 k개의 클러스터로 묶는 비지도학습 알고리즘

고객 정보를 이용하여 고객들을 유사한 특성을 가진 그룹으로 군집화하기 위해 Clustering을 사용



products.csv 파일 이용

3. 고객군 군집화

파생변수 생성

Lpay 거래 내역과 관련된 파생변수 생성

- Lpay : 엘페이 결제 여부
(엘페이로 결제 : 1, 그 외 수단으로 결제 0)
- Lpay_clients : 엘페이 고객 여부
(엘페이 사용경험 0 : 1, 엘페이 사용경험 X : 0)

RFM 변수 생성

- R : 고객의 첫 구매로부터 마지막 구매사이의 기간
- M : Lpay로 결제한 금액과 그렇지 않은 금액을 나누어서 생성
 - M1 : Lpay로 구매한 상품에 대한 금액
 - M2 : Lpay로 구매하지 않은 상품에 대한 금액

거래별 데이터셋을 고객별 데이터셋으로 변환

집계함수를 이용한 변환

- R : R mean
- F : 고객의 상품 구매 횟수 count
- M1 : M1 mean
- M2 : M2 mean
- Lpay : Lpay sum

고객 정보 → 성별, 연령대는 범주형 변수이므로 Label Encoding 적용

- 성별
- 연령대

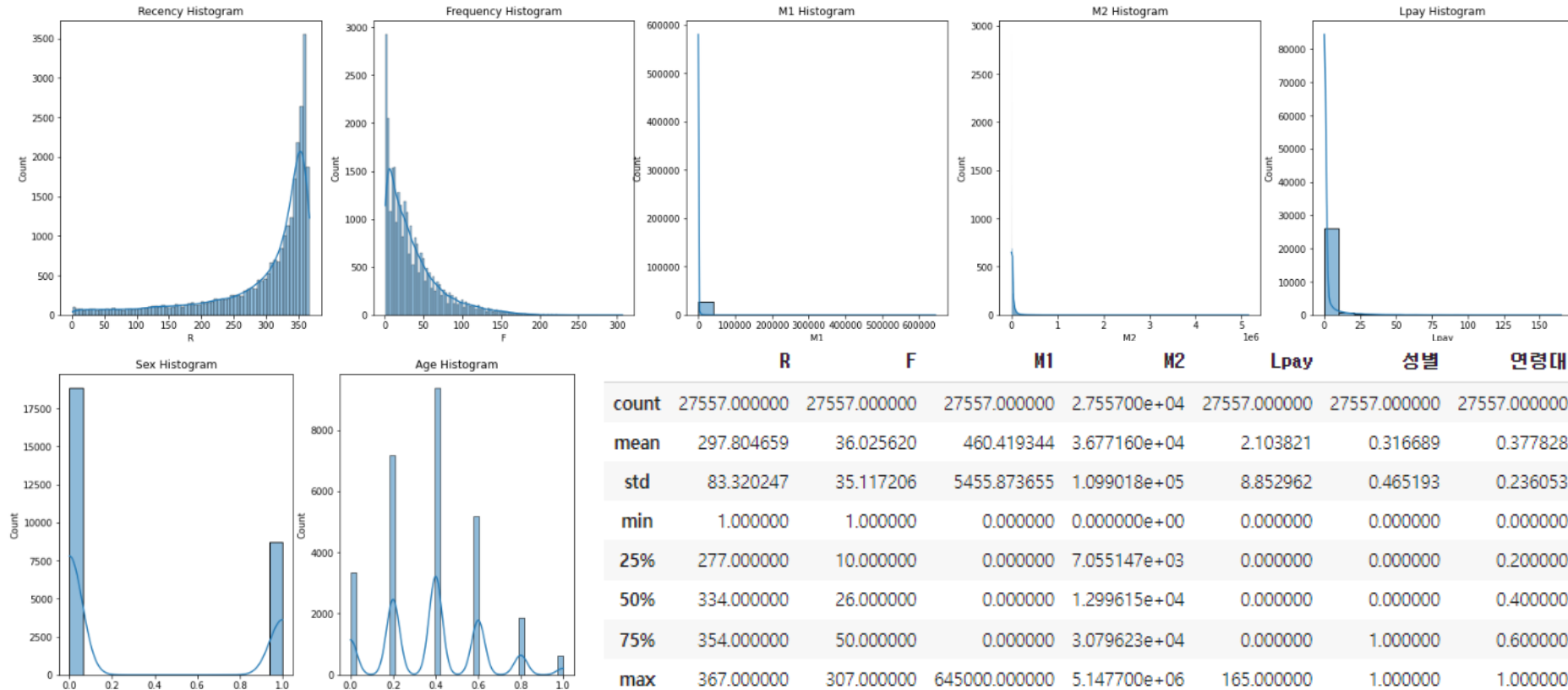
고객번호	R	F	M1	M2	Lpay	성별	연령대
M000034966	341.0	52	0.000000	12654.230769	0	0	0.4
M000059535	0.0	1	0.000000	46000.000000	0	0	0.2
M000136117	360.0	124	0.000000	222669.596774	0	0	0.2
M000201112	126.0	21	0.000000	2529.523810	0	0	0.6
M000225114	304.0	169	0.000000	13336.686391	0	0	0.4

[최종 데이터셋]

3. 고객군 군집화

변수 분포 확인

R, F, M1, M2, Lpay, 성별, 연령대



변수 간 스케일에 많은 차이가 존재해 표준화 진행

3. 고객군 군집화

K-means clustering 적용

R, F, M1, M2, Lpay, 성별, 연령대 정보를 활용하여 K-means clustering 진행

- 1) Raw 데이터 (아무런 처리 진행하지 않음)
- 2) StandardScaler 적용
- 3) 로그변환 적용
- 4) 로그변환 & StandardScaler 적용



* StandardScaler

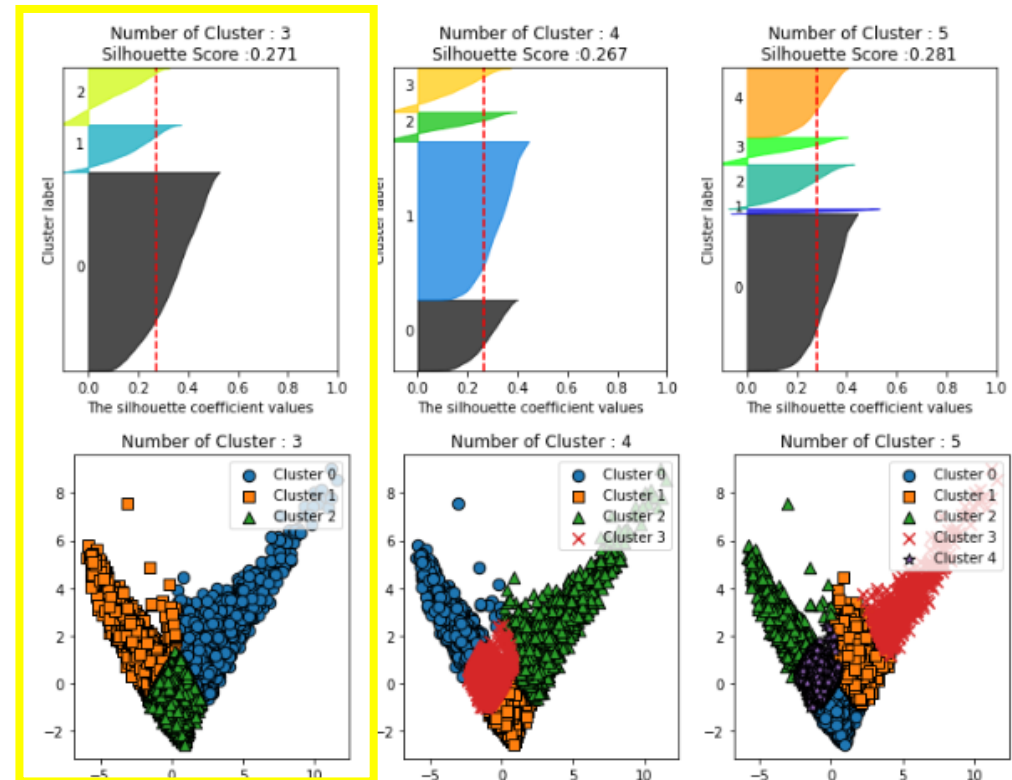
: 입력된 x들의 정규 분포를 평균이 0이고 분산이 1인 표준 정규 분포로 변환

$$\rightarrow Z = \frac{X - \mu}{\sigma}$$

* Silhouette Coefficient

: 각 데이터 포인트와 주위 데이터 포인트들과의 거리 계산을 통해 값을 구하며, 군집 안에 있는 데이터들은 잘 모여있는지, 군집끼리는 서로 잘 구분되는지 클러스터링을 평가하는 척도로 활용

→ 실루엣 계수의 평균값이 1에 가까울수록 군집화가 잘 되었다고 생각

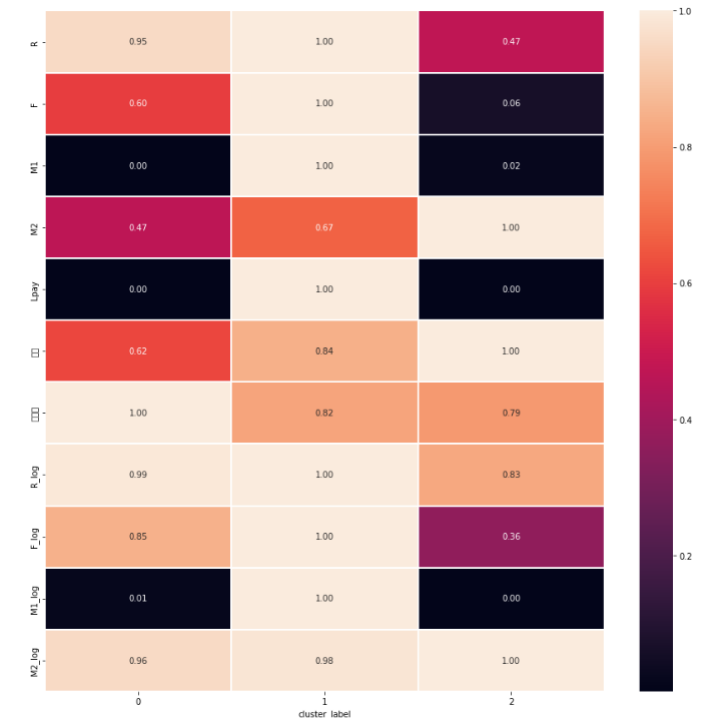
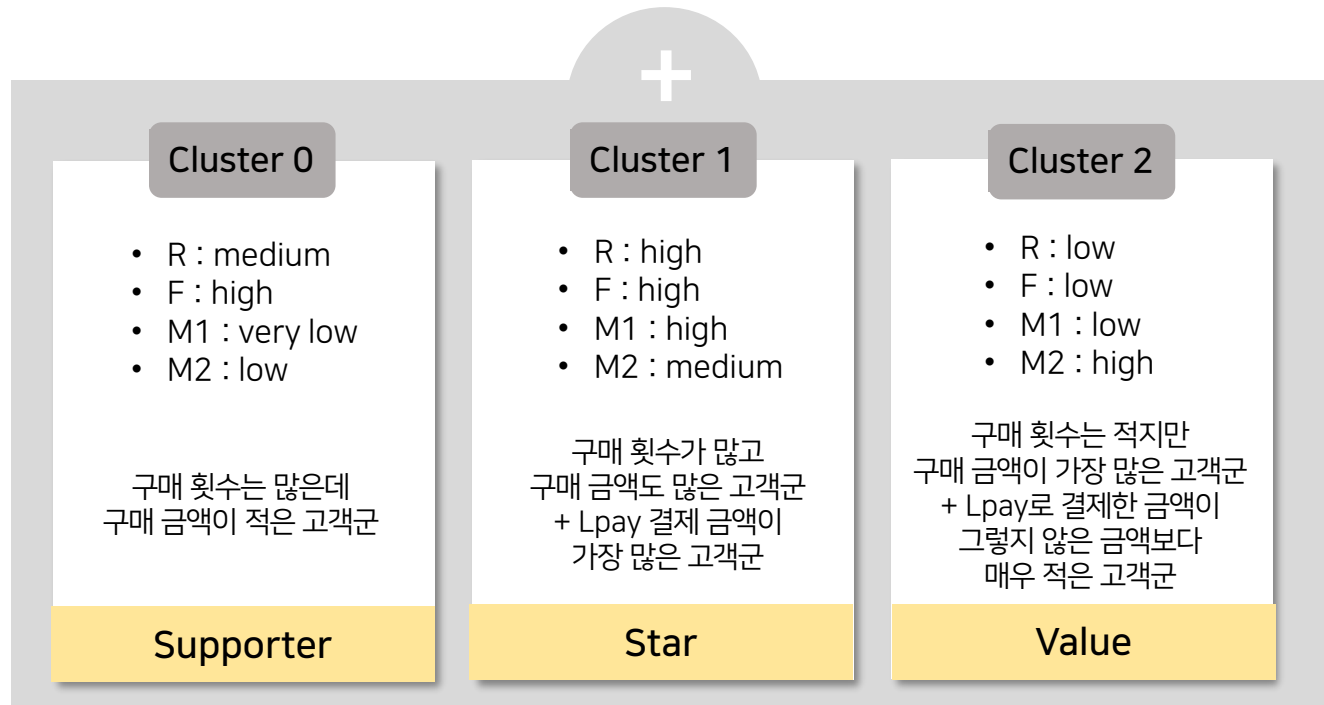


→ 군집을 3개로 나누었을 때 군집별 분포가 비슷하고 실루엣 계수도 0.271로 상대적으로 높아 적합한 군집의 개수라고 판단

3. 고객군 군집화

K-means clustering 결과 해석

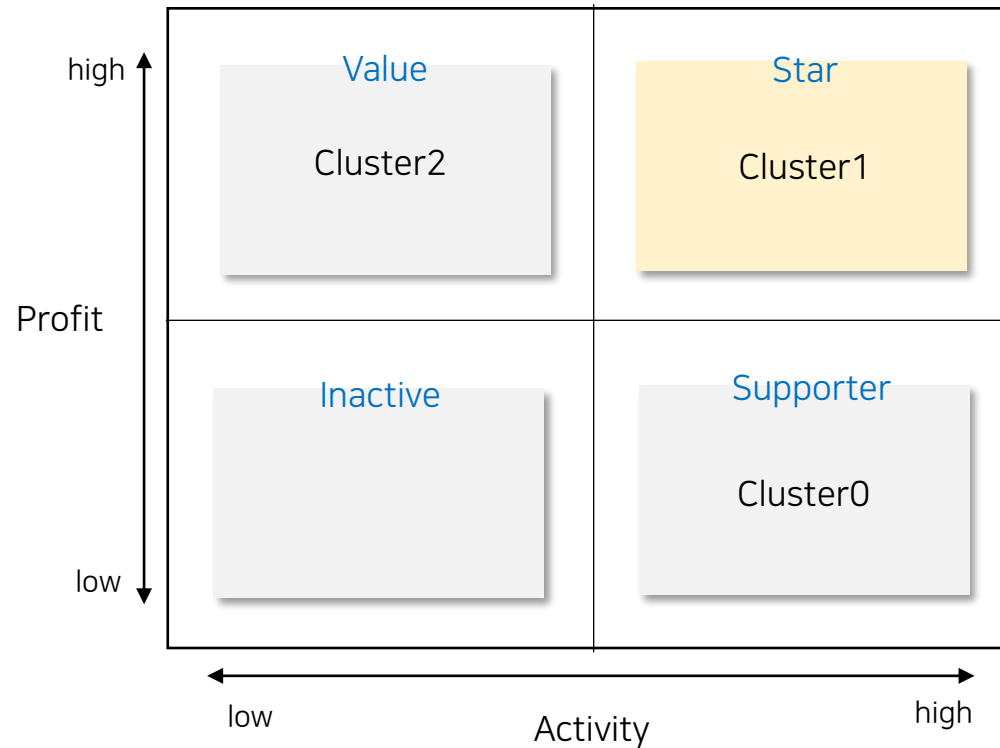
	R	F	M1	M2	Lpay	성별	연령대	R_log	F_log	M1_log	M2_log	고객 수
cluster_label												
0	326.172855	38.354562	0.919149	28689.454357	0.041498	0.269518	0.405389	5.781282	3.385137	0.095848	9.559977	→ 18,073
1	341.710557	64.404019	2862.284379	41186.650087	13.213444	0.369600	0.331070	5.831520	3.989647	6.588939	9.806684	→ 5,155
2	161.477401	4.029292	54.381183	61399.317614	0.004656	0.437633	0.320466	4.826341	1.436835	0.026360	9.993205	→ 4,329



3. 고객군 군집화

K-means clustering 결과 해석

[Customer Switching strategy]



L.pay VIP 고객

- 구매 횟수가 많고 구매 금액도 많은 특징을 가짐

Activity ↑ Profit ↑

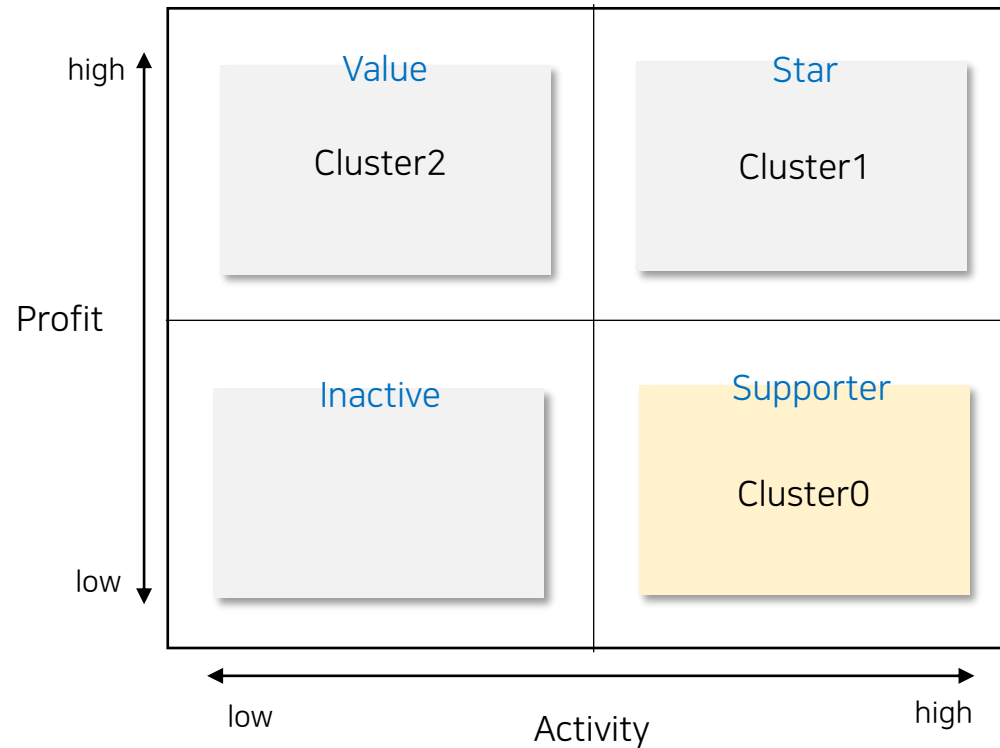
- Lpay 결제 금액이 가장 많은 특징을 가짐

→ 활동량이 많고 기업에게 많은 이익을 가져다 주며
Lpay를 가장 잘 활용하는 핵심 고객군

3. 고객군 군집화

K-means clustering 결과 해석

[Customer Switching strategy]



L.pay 잠재고객

구매 횟수는 많은데
구매 금액이 적은 특징을 가짐

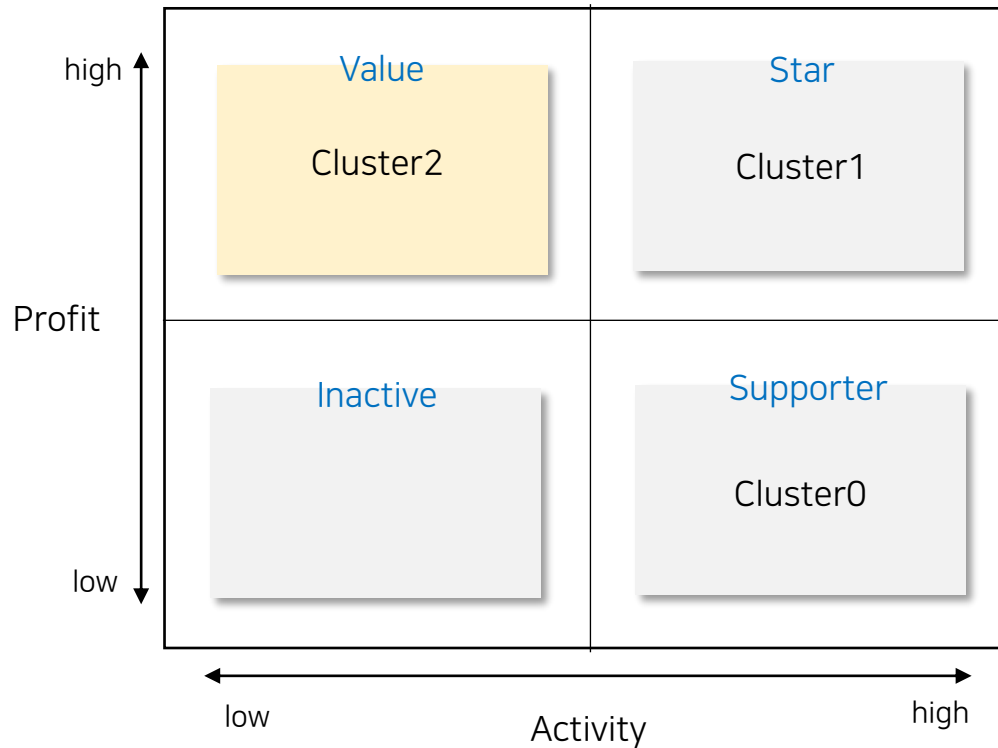
Activity ↑ Profit ↓

→ 활동량이 많지만 기업에게 많은 이익을
가져다 주지 않는 잠재 고객군

3. 고객군 군집화

K-means clustering 결과 해석

[Customer Switching strategy]



L.pay 유력고객

- 구매 횟수는 적지만 구매 금액이 가장 많은 특징을 가짐

Activity ↓ Profit ↑

- Lpay로 결제한 금액이 매우 적은 특징을 가짐

→ 기업에게 많은 이익을 가져다 주지만 Lpay를 잘 사용하지 않고 활동량이 적은 가치 고객군

4. 고객별 상품 추천

클러스터별 가장 거래가 활발한 고객에게 구매 확률이 높은 상품 추천(대표 3인 대상)



추천 시스템

정보 필터링(IF) 기술의 일종으로, 특정 사용자가 관심을 가질만한 정보 (영화, 음악, 책, 뉴스, 이미지, 웹 페이지 등)를 추천하는 것. 협업 필터링 기법을 주로 사용

ALS(Alternating Least Square)

협업 필터링의 한 종류인 Matrix Factorization(MF)의 두 행렬 중 하나를 고정시키고 다른 하나의 행렬을 순차적으로 반복하면서 최적화하는 방법

• 사용 방법

1. 초기 아이템, 사용자 행렬을 초기화
2. 아이템 행렬을 고정하고 사용자 행렬을 최적화
3. 사용자 행렬을 고정하고 아이템 행렬을 최적화
4. 위의 2, 3 과정을 반복해가면서 두 행렬 모두 수렴의 근접한 값을 찾기

• 특징

- 분산처리가 가능하여 학습 속도 매우 빠름
- Sparse한 데이터에 대해 강건(Robust)하여 Implicit Feedback에 유리

$$\begin{array}{c} \hat{R} \\ m \times n \end{array} = \begin{array}{c} \text{User} \\ \text{Latent} \\ \text{Matrix} \\ P^T \\ m \times k \end{array} \cdot \begin{array}{c} \text{Item} \\ \text{Latent} \\ \text{Matrix} \\ Q \\ k \times n \end{array}$$

m : User 수
 n : Item 수
 k : 잠재 벡터 크기

*P와 Q는 정규분포 기반의 랜덤 값들로 초기화

4. 고객별 상품 추천

고객번호 : M057015266

1. User-Item 테이블 생성

고객번호	상품코드	구매수량
M000034966	PD0116	2
M000034966	PD0169	1
M000034966	PD0178	1



scipy.sparse.csr_matrix 라이브러리 이용한
Compressed Sparse matrix by Row(CSR) 변환



압축희소행렬

Client : 27,553명

Product : 1,933개

2. 희소성 확인

- 협업필터링을 구축하기 위한 희소성은 99.5%까지 가능
- 생성된 압축희소행렬의 희소성은 약 96.85%로 추천시스템 구축 가능

3. 모델 생성 및 훈련

Factors : 20
Regularization : 0.1
Iterations : 30
Alpha : 15

4. 모델 적합성 확인

ALS 기반 추천 시스템은 평균 AUC : 0.897

인기있는 아이tem 기반 알고리즘의 평균 AUC : 0.876

-> ALS 기반 추천 시스템이 가장 인기있는 아이tem 기반 알고리즘보다 나은 성능

5. 고객 당 추천 지수가 높은 상위 10개의 상품 제시

	상품코드	대분류명	중분류명	소분류명	구매확률	고객번호	클러스터
0	PD0045	건강식품	영양제	일반비타민	1.494372	M057015266	1
1	PD1536	축산물	닭고기류	닭정육	1.307551	M057015266	1
2	PD0054	건강식품	영양제	기타영양제	1.305899	M057015266	1
3	PD1348	주류	전통주	기타전통주	1.251865	M057015266	1
4	PD0501	병통조림	수산물통조림	기타수산물통조림	1.241449	M057015266	1
5	PD1302	조미료	장류	된장	1.194650	M057015266	1
6	PD1918	화장품/뷰티케어	스킨케어	페이셜클렌저	1.190217	M057015266	1
7	PD0369	냉장식품	포장반찬	절임반찬	1.170852	M057015266	1
8	PD1503	축산물	국산돼지고기	국산돼지부산물	1.168248	M057015266	1
9	PD0051	건강식품	영양제	유산균/프로바이오틱스	1.159787	M057015266	1

4. 고객별 상품 추천

Cluster0

고객 번호 : M124357021

잠재고객

	상품코드	대분류명	중분류명	소분류명	구매확률
0	PD0045	건강식품	영양제	일반비타민	1.057568
1	PD1269	조리식품	온장조리	기타온장조리	1.048656
2	PD1853	퍼스널케어	핸드/풋케어	핸드워시/손세정제	1.032695
3	PD0186	과일	국산과일	기타국산과일류	1.014869
4	PD1246	조리식품	냉장조리	샐러드	1.006727
5	PD1261	조리식품	온장조리	치킨류	0.994582
6	PD1332	주류	맥주	무알콜맥주	0.987812
7	PD1645	커피/차	원두커피	캡슐원두커피	0.981469
8	PD0397	대용식	레토르트	즉석죽	0.972314
9	PD1839	퍼스널케어	바디케어	바디워시	0.964023

고객 번호 : M282063613

	상품코드	대분류명	중분류명	소분류명	구매확률
0	PD0716	수산물	어류	조기류	1.161803
1	PD1297	조미료	식용유류	포도씨유	1.106532
2	PD1364	주방잡화	조리도구	기타조리도구	1.100786
3	PD0937	양곡		쌀	1.099442
4	PD0886	식기/조리기구	밀폐/보관용기	반찬통/밀폐용기	1.099117
5	PD0708	수산물	어류	가자미	1.077906
6	PD1377	주방잡화	주방정리용품/소모품	비닐장갑	1.075472
7	PD1315	조미료	조미료류	역상조미료	1.068706
8	PD1344	주류	전통주	청주	1.063706
9	PD0947	양곡	잡곡	혼합잡곡	1.061516

고객 번호 : M576689847

	상품코드	대분류명	중분류명	소분류명	구매확률
0	PD0054	건강식품	영양제	기타영양제	1.318370
1	PD1644	커피/차	원두커피	티백/드립원두커피	1.236829
2	PD0644	속옷/양말/휴웨어	여성양말류	스타킹	1.195596
3	PD0068	건강용품	건강관리용품	디퓨저	1.186211
4	PD0233	과자	스낵류	영양바	1.170586
5	PD1686	테넌트/음식점	식당	중식	1.148579
6	PD1922	화장품/뷰티케어	스킨케어	에센스/세럼	1.144057
7	PD1889	화장품/뷰티케어	메이크업	BB/파운데이션/컨택트	1.141590
8	PD1340	주류	와인	레드와인	1.136884
9	PD1919	화장품/뷰티케어	스킨케어	스킨/토너	1.098038

Cluster1

고객 번호 : M057015266

VIP 고객

	상품코드	대분류명	중분류명	소분류명	구매확률
0	PD0045	건강식품	영양제	일반비타민	1.494372
1	PD1536	축산물	닭고기류	닭정육	1.307551
2	PD0054	건강식품	영양제	기타영양제	1.305899
3	PD1348	주류	전통주	기타전통주	1.251865
4	PD0501	병통조림	수산물병통조림	기타수산물병통조림	1.241449
5	PD1302	조미료	장류	된장	1.194650
6	PD1918	화장품/뷰티케어	스킨케어	페이셜클렌저	1.190217
7	PD0369	냉장식품	포장반찬	절임반찬	1.170852
8	PD1503	축산물	국산돼지고기	국산돼지부산물	1.168248
9	PD0051	건강식품	영양제	유산균/프로바이오틱스	1.159787

고객 번호 : M596502154

	상품코드	대분류명	중분류명	소분류명	구매확률
0	PD0247	구기/필드스포츠	골프	골프필드용품	1.437724
1	PD1170	음료	기능성음료	기타기능성음료	1.422678
2	PD0797	스포츠패션	여성골프의류	여성골프티셔츠/탑	1.371040
3	PD0768	스포츠패션	남성골프의류	남성골프의류세트	1.346315
4	PD0727	수산물	어류	기타어류	1.287015
5	PD0760	스포츠패션	남성골프의류	남성골프티셔츠	1.254650
6	PD1411	채소	뿌리채소	기타뿌리채소	1.253746
7	PD0250	구기/필드스포츠	골프	골프패션잡화	1.249544
8	PD0807	스포츠패션	여성골프의류	여성골프의류세트	1.246358
9	PD1268	조리식품	온장조리	즉석어묵	1.192444

고객 번호 : M378415176

	상품코드	대분류명	중분류명	소분류명	구매확률
0	PD1244	조리식품	냉장조리	조밥류	1.244849
1	PD0727	수산물	어류	기타어류	1.169768
2	PD0080	건강용품	일반의약품	안전상비의약품	1.094229
3	PD1247	조리식품	냉장조리	샌드위치	1.093494
4	PD0612	세제/위생	화장지/티슈	각티슈/마용티슈	1.053864
5	PD0363	냉장식품	디저트	빵/케이크	1.041995
6	PD0601	세제/위생	세탁세제	섬유유연제/항기취제	1.036101
7	PD1847	퍼스널케어	여성위생용품	생리대	1.028635
8	PD1487	축산물	계란류	가공계란	1.026767
9	PD1498	축산물	국산돼지고기	국산돼지뒷다리	1.022940

Cluster2

고객 번호 : M408936009

유력고객

	상품코드	대분류명	중분류명	소분류명	구매확률
0	PD0708	수산물	어류	가자미	1.082634
1	PD1171	음료	두유	일반두유	1.064340
2	PD0937	양곡		쌀	1.028074
3	PD0936	양곡		쌀	1.019379
4	PD0611	세제/위생	화장지/티슈	롤티슈	1.013959
5	PD1384	채소	나물류	생나물류	1.006776
6	PD1422	채소	열매채소	오이	1.005252
7	PD0613	세제/위생	화장지/티슈	물티슈	1.001771
8	PD1437	채소	임채소	시금치	0.998491
9	PD0603	세제/위생	주방세제	일반주방세제	0.997686

고객 번호 : M652598612

	상품코드	대분류명	중분류명	소분류명	구매확률
0	PD1688	테넌트/음식점	식당	분식	1.220898
1	PD1028	완구	실외놀이/스포츠완구	실외놀이완구	1.131016
2	PD0362	냉장식품	디저트	푸딩/젤리	1.096156
3	PD1033	완구	여아완구	패션인형	1.082836
4	PD1252	조리식품		떡	1.077285
5	PD1379	주방잡화	주방정리용품/소모품	장바구니	1.062881
6	PD1689	테넌트/음식점	식당	디저트	1.054919
7	PD0239	과자	초콜릿	헬초콜릿	1.020968
8	PD1164	음료	과채음료	어린이음료	1.020316
9	PD0161	과일	견과류	혼합견과	1.014589

고객 번호 : M865603201

	상품코드	대분류명	중분류명	소분류명	구매확률
0	PD0282	기타(비상품)	기타	봉투보정금	0.986144
1	PD0354	냉장식품	냉장간편식	냉장핫도그/핫바류	0.898364
2	PD1243	조리식품	냉장조리	삼각김밥	0.896023
3	PD1178	음료	즉석음료류	파우치/즉석음료	0.879751
4	PD1241	조리식품	냉장조리	김밥	0.862682
5	PD1248	조리식품	냉장조리	햄버거	0.861986
6	PD0232	과자	스낵류	일반스낵	0.857050
7	PD0382	담배	일반담배	담배관련용품	0.849157
8	PD0218	과자	비스킷류	쿠키	0.830748
9	PD1173	음료	생수	생수	0.823759

5. 고객별 재구매 예측

Cluster별 가장 구매 건수가 많은 고객을 대상으로 가장 거래 건수가 많은 상품에 대한 구매수량 예측(대표 3인 대상)



시계열 분석

어떤 현상에 대하여 과거에서부터 현재까지의 시간에 흐름에 따라 기록된 데이터를 바탕으로 미래의 변화에 대한 추세를 분석하는 방법

ARIMA

- 과거의 관측 값과 오차를 사용해서 현재의 시계열 값을 설명하는 ARMA(Auto-regressive Moving Average) 모델을 일반화 한 것
- 분기/반기/연간 단위로 다음 지표를 예측하거나 주간/월간 단위로 지표를 리뷰하며 트렌드에 이상치가 없는지를 모니터링 하는 데 사용
- 안정적 시계열 뿐만 아니라 다소 불안정적인 시계열의 특징을 보여도 적용이 가능

LSTM

- LSTM은 RNN(순환신경망)의 변형 버전으로 자신의 아웃풋을 다시 인풋으로 사용
- RNN의 기울기 소실과 기울기 폭발의 문제로 인한 아주 긴 시퀀스 데이터 성능 문제 보완
- 시계열처럼 긴 자료를 가지고 모델링을 할 때 사용

Prophet

- Facebook의 핵심 데이터 사이언스 팀에서 출시한 오픈 소스 소프트웨어
- 계절적 효과가 강하고 여러 시즌의 과거 데이터가 있는 시계열에서 가장 잘 작동
- 누락 데이터와 추세 변화에 강하며 이상값 처리에 용이

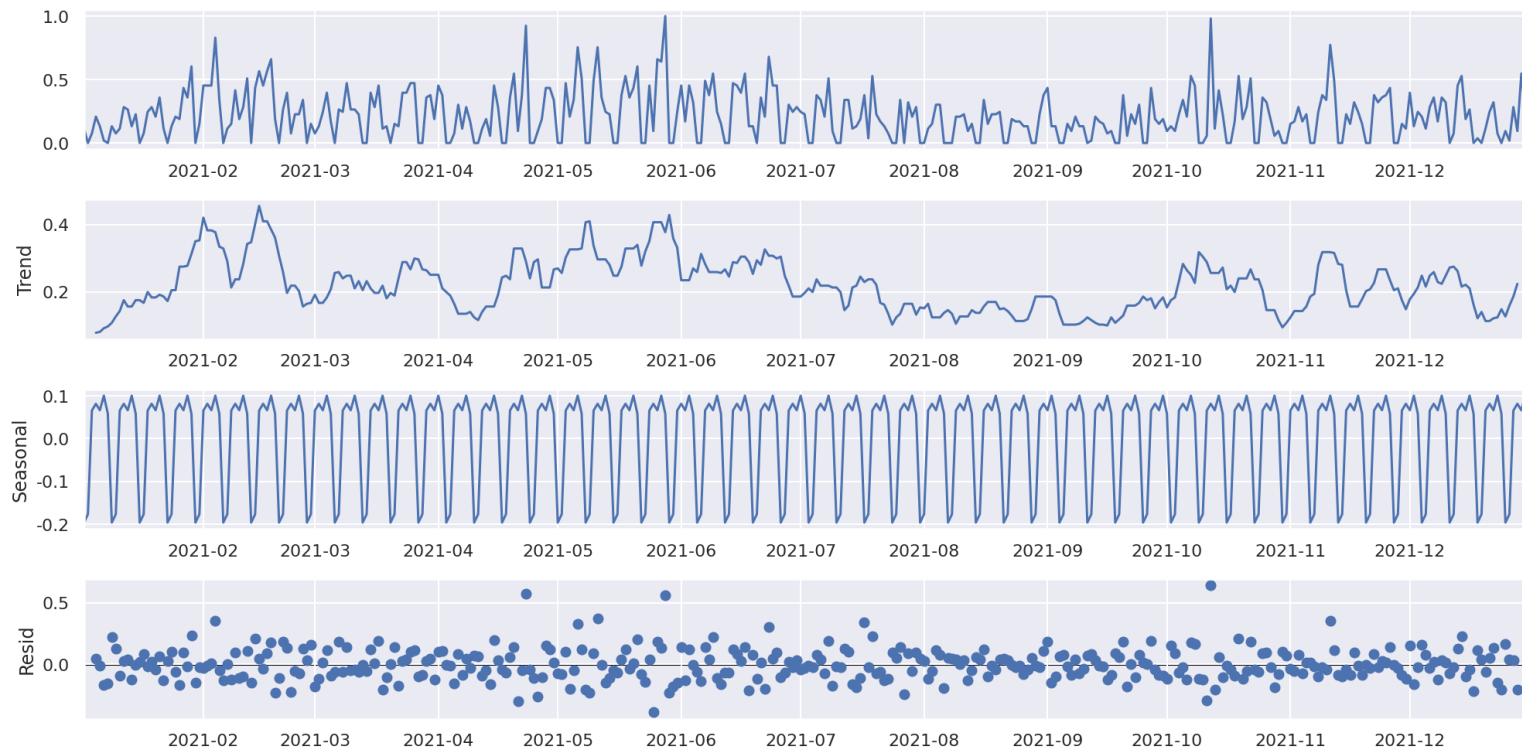
5. 고객별 재구매 예측(ARIMA)

1. 첫 구매일부터 마지막 구매일까지 구매수량 데이터프레임 생성

고객번호 : M057015266, 대상 상품 : 과자

구매수량	
2021-01-02	6
2021-01-03	0
2021-01-04	4
2021-01-05	11

2. MinMaxScaler를 변수 값을 0~1의 범위로 한 후 raw data의 추세, 잔차, 계절성 확인

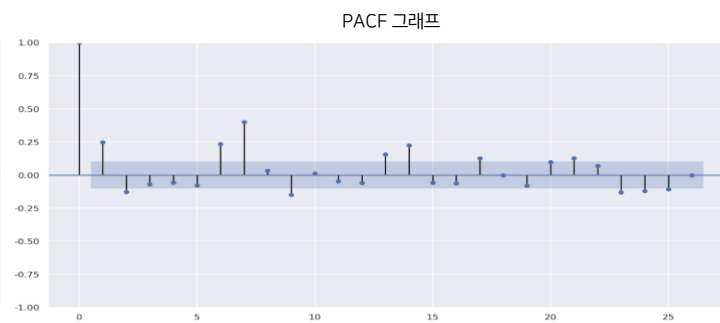
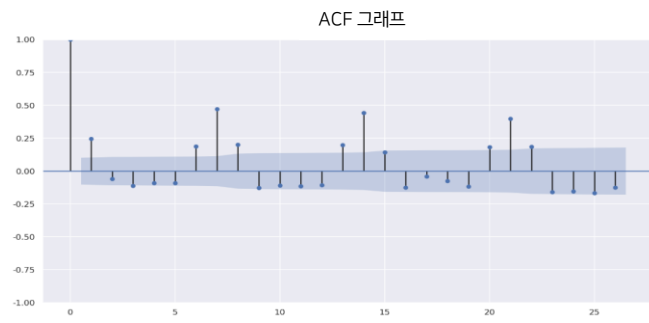


5. 고객별 재구매 예측(ARIMA)

3. 데이터의 정상성 확인

- 정상성 : 평균, 분산이 시간에 따라 일정한 성질. 추세나 계절성이 없는 시계열.

정상성을 나타내지 않는 데이터는 복잡한 패턴을 모델링하여 분석하기 어렵기 때문에 정상성을 갖도록 전처리



정상성 만족

고객번호 : M057015266, 대상 상품 : 과자

- ACF

자기상관함수.
시차에 따른 일련의 자기상관으로 시차가 커질수록 0에 가까움.
정상 시계열은 상대적으로 빠르게 0에 수렴.
비정상 시계열은 천천히 감소

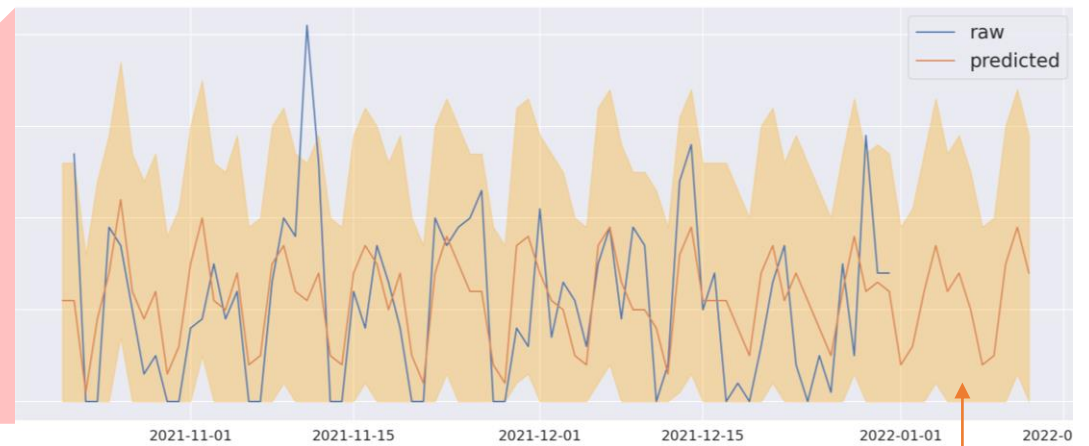
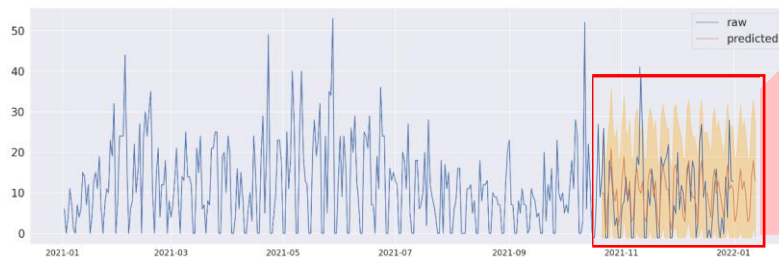
- PACF

편자기상관함수.
시차에 따른 일련의 편자기상관으로 시차가 다른 두 시계열 간의 순수한 상호 연관

4. pmdarima의 auto_arima를 이용하여 tuning 후 결과값 출력

order : (6, 0, 4),

seasonal_order: (2, 1, 1, 12)



MAE : 0.112075, RMSE : 0.141436

95% 신뢰구간

5. 고객별 재구매 예측(LSTM)

고객번호 : M057015266, 대상 상품 : 과자

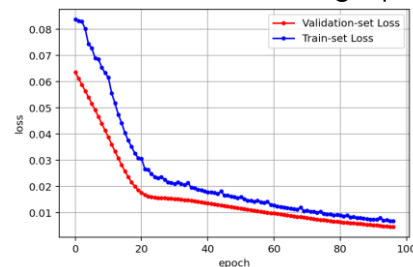
1. 첫 구매일부터 마지막 구매일까지 거래건수, 온오프 평균, Lpay 거래건수, Lpay 구매금액 총합, 구매시간 평균, 구매금액 총합, 구매수량 데이터프레임 생성

	거래건수	온/오프_평균	Lpay거래건수	Lpay구매금액_총합	구매시간_평균	구매금액_총합	구매수량
2021-01-02	6	1.0	0	0.0	22.000000	11040.0	6
2021-01-03	0	0.0	0	0.0	0.000000	0.0	0
2021-01-04	2	1.0	0	0.0	19.000000	2622.0	4
2021-01-05	8	1.0	0	0.0	18.500000	12400.0	11

2. MinMaxScaler를 통해 각 변수 값의 범위 통일

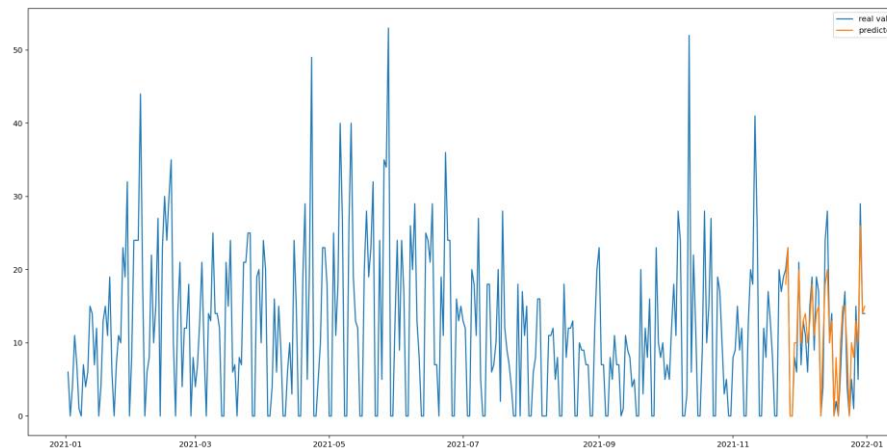
3. LSTM 모델 생성 및 훈련

Train/Validation Loss graph



Optimizer : Adam, learning_rate_0.0001
 Loss : mse
 Epoch = 100

4. 실제 값과 예측 값의 그래프



성능 측정

MAE : 0.050800853302015565
 MSE : 0.0045351597451546485
 RMSE : 0.06734359468542386
 MAPE : 67.772783

X값(거래건수, 온오프 평균, Lpay 거래건수, Lpay 구매금액 총합, 구매시간 평균, 구매금액 총합)들을 통해 y값(구매수량)을 예측하는 구조로, X값에 대한 정보가 없는 미래에 대한 예측 불가능

5. 고객별 재구매 예측(Prophet)

1. 첫 구매일부터 마지막 구매일까지 구매수량 데이터프레임 생성

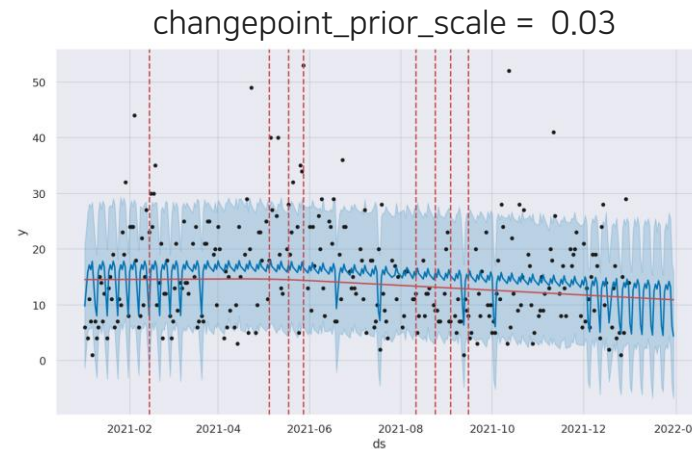
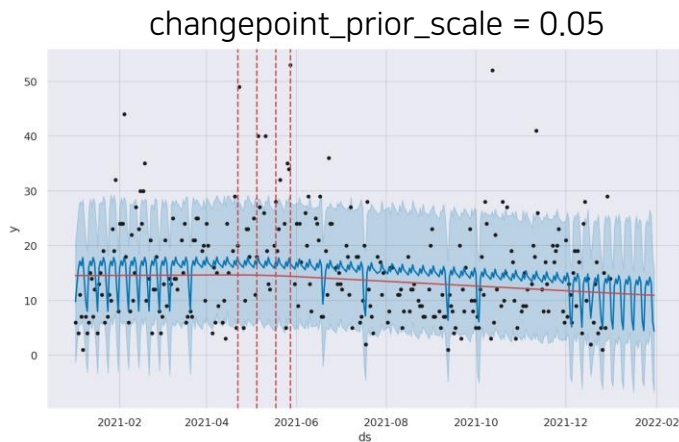
구매수량	
2021-01-02	6
2021-01-03	0
2021-01-04	4
2021-01-05	11

고객번호 : M057015266, 대상 상품 : 과자

2.. Prophet 패키지 이용

changepoint : 트렌드 변화시점을 명시한 리스트 값

changepoint_prior_scale : changepoint(trend)의 유연성 조절. 값을 너무 높여버리면 overfitting의 위험



파란색 선 : 모델이 예측한 값

검정색 점 : 실제 데이터

빨간 실선 : 트렌드

빨간 점선 : 트렌드가 변화하는 changepoint

Changepoint 이외의 전반적인 추세나 구매수량 예측 값 파악하기 어려움

5. 고객별 재구매 예측(Prophet)

1. 첫 구매일부터 마지막 구매일까지 시간 정보를 추가하여 구매수량 데이터프레임 생성

고객번호 : M057015266, 대상 상품 : 과자

구매수량	
2021-01-02 22:00:00	6
2021-01-02 23:00:00	0
2021-01-03 00:00:00	0
2021-01-03 01:00:00	0

2.. Prophet 패키지 이용

파라미터 설명

changepoint : 트렌드 변화시점을 명시한 리스트 값
 changepoint_prior_scale : changepoint(trend)의 유연성 조절
 Weekly_seasonality : 주 계절성
 Daily_seasonality : 일 계절성

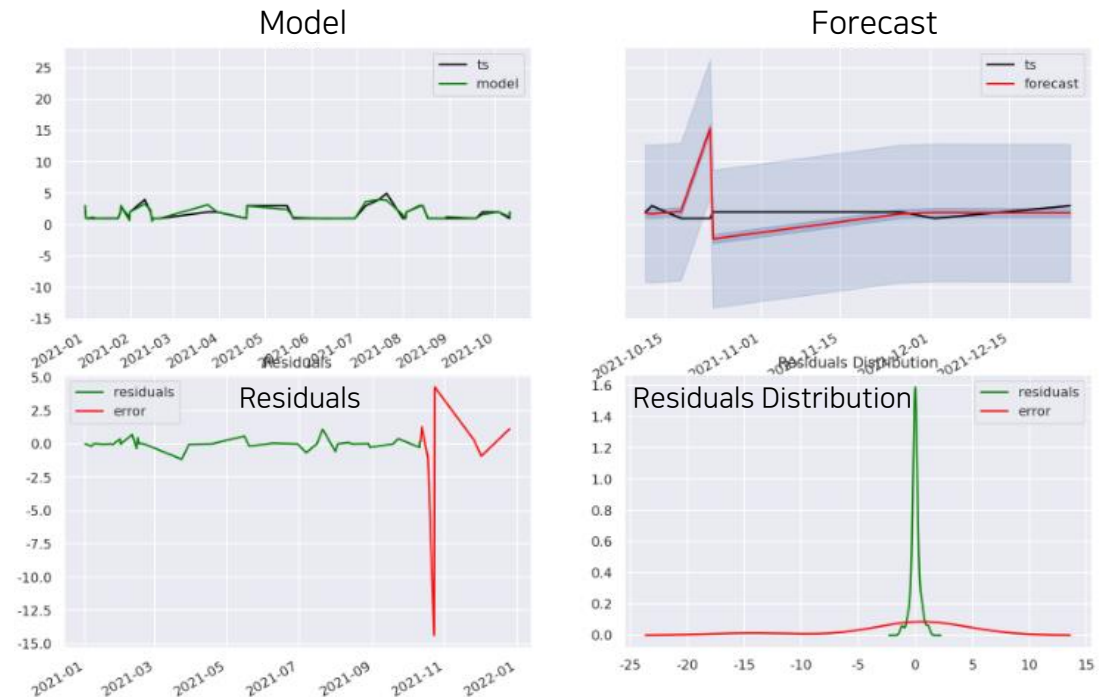
파라미터

changepoint_prior_scale=0.3
 weekly_seasonality=10
 daily_seasonality=10

성능

MAE: 44.0
 MAPE: 3176.0
 MSE: 3891.0
 RMSE: 62.0

결과 그래프



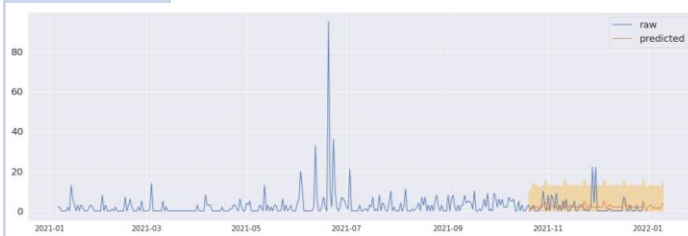
신뢰성 있는 예측값의 판단이 어려움

5. 고객별 재구매 예측

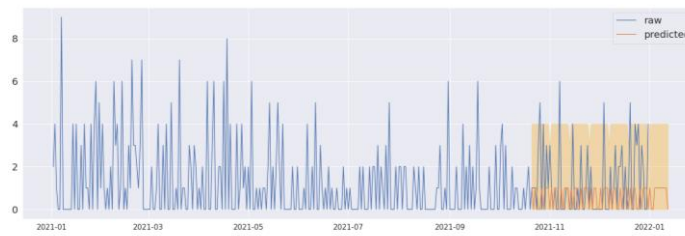
ARIMA로 각 클러스터 별 가장 거래가 활발한 고객 3명씩 가장 많이 구매한 상품의 구매수량 예측

Cluster0

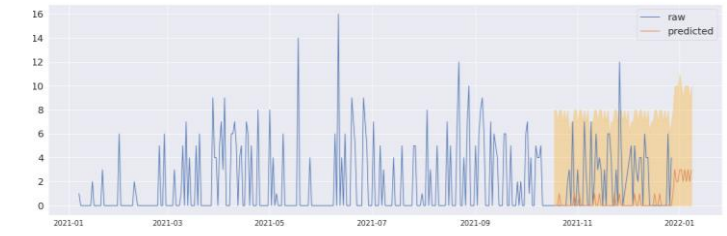
잠재고객



M124357021고객의 음료 구매수량 예측



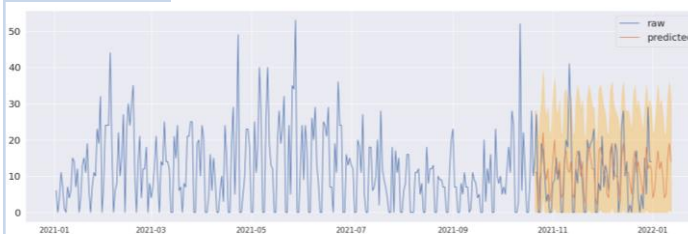
M282063613고객의 채소 구매수량 예측



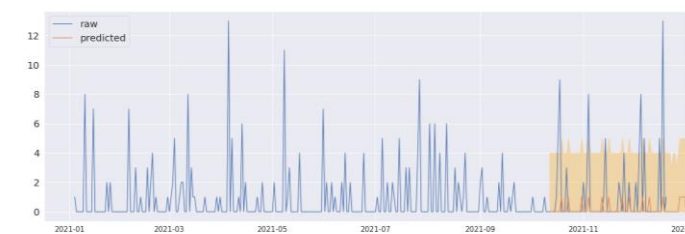
M576689847고객의 조미료 구매수량 예측

Cluster1

VIP 고객



M057015266고객의 과자 구매수량 예측



M378415176고객의 과자 구매수량 예측



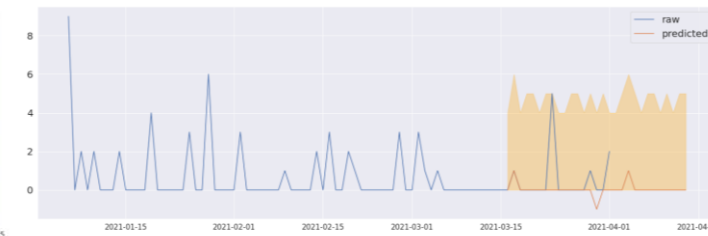
M696502154고객의 식기/조리기구 구매수량 예측

Cluster2

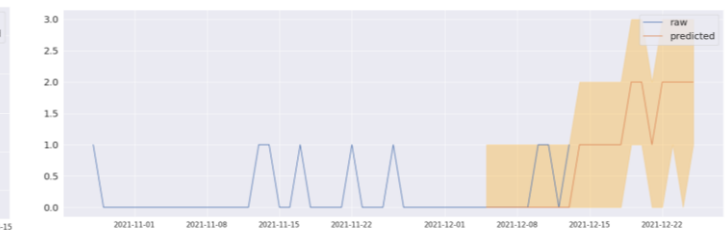
유력고객



M408936009고객의 채소 구매수량 예측



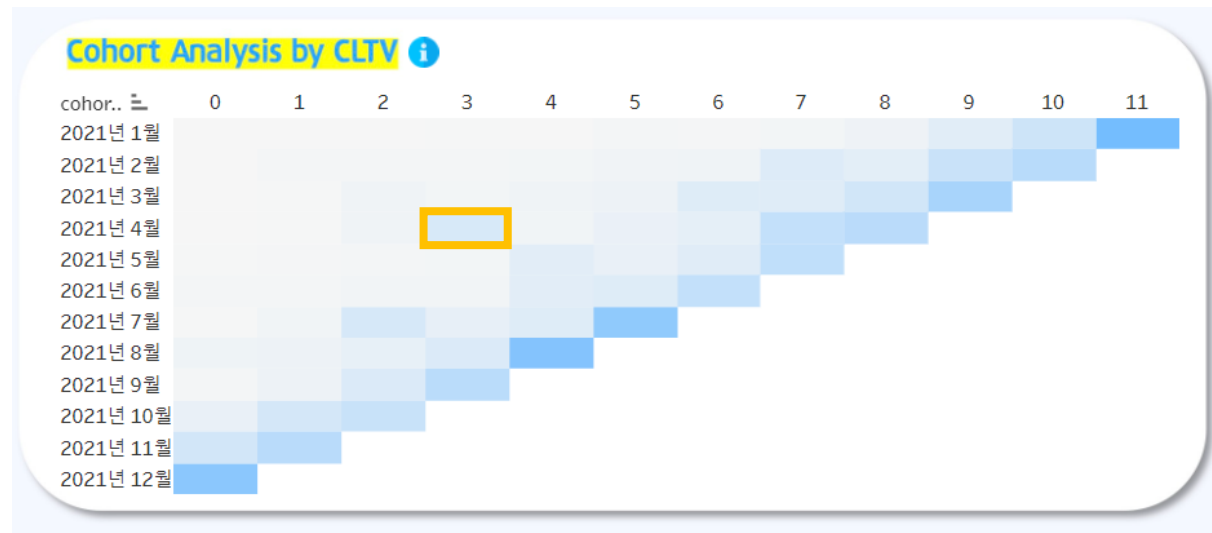
M4652598612고객의 과자 구매수량 예측



M865603201고객의 담배 구매수량 예측

6-1. 마케팅 전략 예시

CLV를 이용한 코호트 분석



CLV가 급격히 증가하는 8월부터 12월까지 진행된 이벤트나 특이사항 파악 및
 일부 그룹에서 8월과 10월에 급격히 CLV가 하락하는 이유에 대한 파악 필요
 4월 첫 구매 고객의 3개월 뒤 급격히 증가하는 CLV로 인해 7월의 이벤트나 내부 요인 파악 필요



내부 데이터 대한 정보 부족으로 인해 명확한 요인 파악보다는
 특이사항을 살펴볼 필요성에 대한 근거 제시 기능

6-1. 마케팅 전략 예시

고객 군집별 CRM 전략 도출

L.pay VIP 고객



Cluster1

활동량이 많고 기업에게 많은 이익을 가져다 주며 Lpay를 가장 잘 활용하는 **핵심 고객군**

CRM 목표

: 핵심 고객의 유지 및 강화

→ 고객 충성도를 높이기 위한
royalty program 실행 및 이탈 방지
(ex. 당일 무료 배송, 멤버십 할인)

L.pay 잠재고객



Cluster0

활동량이 많지만 기업에게 많은 이익을 가져다 주지 않는 **잠재 고객군**

CRM 목표

: 구매금액의 증가를 통한 핵심고객으로의 발전

→ 구매 금액 증가 및 Lpay 사용 유도에 초점
(ex. 일정 금액 이상 구매시 쿠폰 지급, Lpay 결제 시 포인트 적립(차등))

L.pay 유력고객



Cluster2

기업에게 많은 이익을 가져다 주지만 Lpay를 잘 사용하지 않고 활동량이 적은 **가치 고객군**

CRM 목표

: 방문 유도 및 구매 연결을 통한
핵심고객으로의 발전 및 Lpay 사용 유도

→ 방문 유도 뿐만 아니라 구매까지 연결에 초점
(ex. 출석체크 이벤트, Lpay 결제시 포인트 적립(차등))

6-1. 마케팅 전략 예시

상품 추천

고객번호 : M057015266

	상품코드	대분류명	중분류명	소분류명	구매확률	고객번호	클러스터
0	PD0045	건강식품	영양제	일반비타민	1.494372	M057015266	1
1	PD1536	축산물	닭고기류	닭정육	1.307551	M057015266	1
2	PD0054	건강식품	영양제	기타영양제	1.305899	M057015266	1
3	PD1348	주류	전통주	기타전통주	1.251865	M057015266	1
4	PD0501	병통조림	수산물통조림	기타수산물통조림	1.241449	M057015266	1
5	PD1302	조미료	장류	된장	1.194650	M057015266	1
6	PD1918	화장품/뷰티케어	스킨케어	페이셜클렌저	1.190217	M057015266	1
7	PD0369	냉장식품	포장반찬	절임반찬	1.170852	M057015266	1
8	PD1503	축산물	국산돼지고기	국산돼지부산물	1.168248	M057015266	1
9	PD0051	건강식품	영양제	유산균/프로바이오틱스	1.159787	M057015266	1

개인 기준

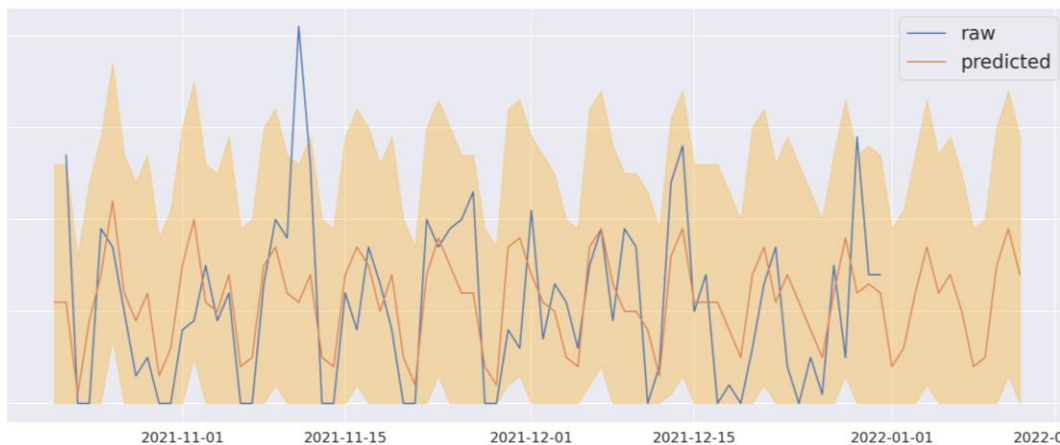
구매 확률이 높은 상품 기반으로 고객에게 추천 상품으로 제시
(일반비타민, 닭정육, 영양제, 전통주, ...)

추천 상품 파악을 통한 고객별 구매 성향 기반으로 고객 관리
(영양제, 된장, 반찬과 같은 영양소 가득한 식품에 대한 소비를 주로 함)

전체 기준

전반적으로 많이 추천되는 상품들의 재고 확보 또는 이벤트 활용 가능
재고에 비해 잘 추천되지 않는 상품들 대상 마케팅 전략 마련 가능

재구매 시기 예측



M057015266 고객의 2022-01-01 부터 2022-01-12까지 “과자”
구매 수량 예측 결과 예측일로부터 12일간 과자의 구매는 꾸준하며
특히 2022-01-11의 구매 수량이 19개로 가장 많을 것임을 알 수 있음



온라인 구매 시 과자 상품 기반 추천을 통해 개인별 페이지 배치에 활용
과자 신상품 제공 또는 앱 알림을 통한 고객 맞춤형 서비스 제공

6-2. 결과 활용 방안

대시보드 구현 - 고객 생애주기 분석 페이지



EDA 파트에서 확인했던 전체 고객 대상의 전반적인 시각화 및 인사이트와 첫 구매 시기별 코호트 집단에 대한 고객 생애주기 가치 등을 확인할 수 있는 페이지

6-2. 결과 활용 방안

대시보드 구현 - 고객 클러스터 분석 페이지



클러스터별 특징과 고객 정보 및 구매 내역 등을 확인함으로써 개별 고객을 효과적 관리할 수 있으며 특히 상품 추천과 재구매 시기 및 수량 예측을 통해 개인 맞춤 마케팅 전략을 한 눈에 파악할 수 있음

THANK YOU



위 QR을 통해 대시보드를
확인하실 수 있습니다.

https://public.tableau.com/app/profile/.74014994/viz/L_PointL_PayCustomerReport_16598720476590/CLTV