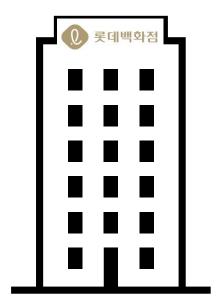
잠재유형 예측 모델을 통한 신규고객 잠재유형 예측 및 고객 유형화를 통한 개인화 마케팅 전략수립

Team, 우리 좀 롯데

(백찬진, 이성규, 지윤혁)









개요

01. 주제선정배경 02. 프로젝트 개요

분석

01. 데이터 전처리 02. 고객행동 분석 및 유형화

01. 외부데이터와 EDA를 활용한 파생변수 02. 신규 고객 잠재유형 예측 모델

전략

01. 유형별 개인화 마케팅 02. 비즈니스 기대효과



잠재유형?

짧은 거래내역만을 가지고 있는 신규 고객의 경우 장기적으로

어떠한 유형의 고객이 될지 알 수 없다.

잠재 유형이란 신규 고객이 귀사의 장기적인 고객이 되었을 때.

기존 고객들을 기준으로 나누어진 유형 중, 속할 것으로 예상되는 유형이다.

ex) 신규 고객 A의 짧은 거래내역을 바탕으로 모델은 A의 잠재유형이 VIP일 것으로 예측







STEP 1.

거래내역을 기준으로 고객을 '신규고객'과 '기존고객'으로 분리

STEP 2.

기존 고객을 특정 기준으로 유형화하여 데이터에 기반한 특징 도출

STEP 3.

신규 고객의 짧은 거래 내역만을 보고 해당 고객이 장기적으로 속할 유형을 예측하는 알고리즘 생성

STEP 4

도출한 유형별 특징을 활용한 개인화 마케팅 및 비즈니스 전략 수립





1. 신규고객



신규 고객의 짧은 거래 내역을 바탕으로 해당 고객이 속할 유형을 미리 예측하여 마케팅

2. 기존고객



"

기존 고객을 유형화하여 각 유형의 특징을 분석하고 유형에 따른 차별화된 마케팅을 진행

"





1. 신규고백의 짧은 개래 내역을 바탕으로 기계 시구고객의 짧은 개래 내역을 바탕으로 기계 시구 고객의 잠자 유형을 이 이 측하는 서비스는 왜 필요한가?

"





기존 고객을 유형화하여 각 유형의 특징을 분석하고 유형에 따른 차별화된 마케팅을 진행





기존 고객과 다르게 신규 고객의 특징은 알아내기까지 일정 수준 이상의 거래가 필요함

	Α	В	С	D	Е	F	G
1	기존고객						
2	고객 ID	나이	성별	거주지	관심 상품	충성도	구매력
3	1	23	남	서울	컴퓨터용품	하	하
4	2	51	여	경기도	식자재	상	중
5	3	42	여	부산	여성의류	상	상
6	4	62	남	경기도	식당	중	상
7							
8							
9	신규고객						
10	고객 ID	나이	성별	거주지	관심 상품	충성도	구매력
11	5	32	남	서울	?	?	?
12	6	25	여	부산	?	?	?

기존 고객



신규 고객





고객ID 5의 짧은 거래내역

고객 ID	나이	성별	거주지	구매품목	구매일자	구매시간	구매금액	구매지점
5	32	남	서울	컴퓨터 모니터	2021-10-12	14:23	150000	Z10
5	32	남	서울	남성 의류	 2021-10-28	16:12	89000	Z10
5	32	남	서울	주류	2021-10-31	18:56	4500	Z10
5	32	남	서울	스낵류	2021-10-31	18:56	2300	Z10



전규 고객 잠재 유형 예측모델



- 신규 고객 5의 잠재유형을 미리 예측
- 잠재 유형에 따른 발빠른 마케팅 실행
- 해당 고객의 잠재유형과 같은 고객들의 평이 좋은 제품 추천
- -> 신규고객의 잠재 유형을 미리 예측할 수 있다면 위와 같은 이점을 통해
 - 귀사에 대한 신규 고객의 호감도를 증가시킬 수 있으며
 - 신규 고객을 충성 고객으로 만드는데 중요한 역할을 할 수 있음

L.PAY with L.POINT



2. 기존 고객을 테이터에 기반하여 유형화하는 것이 왜 필요한가?

기조그개



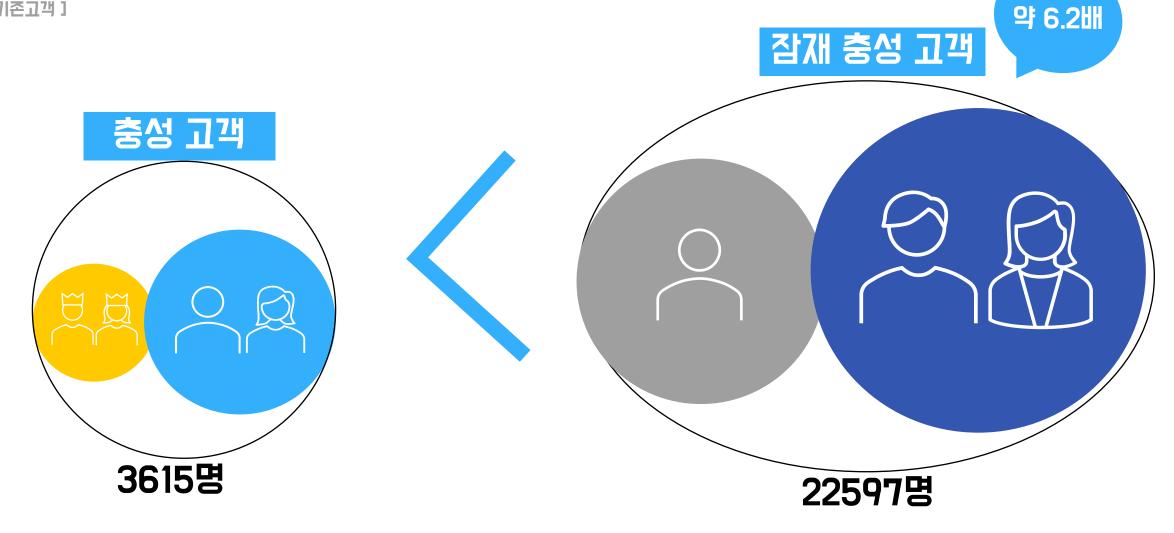
66 기존 고객을 유형화하여 각 유형의 특징을 분석하고

마는 교칙을 유명되어서 각 유명의 특명을 문식하고 유형에 따른 차별화된 마케팅을 진행

77







특정 기준에 따라 기존 고객을 잠재 충성 고객과 충성 고객으로 나눈 후 인원수 비교

L.PAY with **L.POINT**

특정 기준에 따라 기존 고객을 잠재 충성 고객과 충성 고객으로 나는 후 입원수 비교



주제선정배경 2. EDA를 통한 유형별 세부 특징 파악

[기존고객]



각 유형을 데이터를 바탕으로 다양한 각도에서 관찰하고 이해하는 과정을 통해 이전에 알지 못했던 특징을 발견할 수 있음



데이터의 분포를 통해 유형별로 긍정적, 부정적 특징을 도출하여 구체적인 마케팅 방안 수립 가능



01.개요 프로젝트 개요

STEP 1.

거래내역을 기준으로 고객을 '신규고객'과 '기존고객'으로 분리

STEP 2.

기존 고객을 RFM을 기준으로 k-means Clustering을 통해 4가지의 유형으로 분류

STEP 3.

EDA를 통하여 분류한 유형별 특징을 정리하고 통계 기법을 사용하여 검증

STEP 4.

가 유형별 EDA를 바탕으로 파생변수를 만든 후, 신규 고객의 잠재유형을 예측하는 모델을 개발

STEP 5.

가 유형별 특징을 바탕으로 개인화 마케팅 및 비즈니스 전략 수립







O2.분석 전처리 [Flow] Pre-processing 1 [상품 구매 정보 데이터 중복 제거 & 결측치 처리]

	고객번호	영수증번호	채널구분	제휴사	점포코드	상품코드	구매일자	구매시간	구매금액	구매수량
3	M430112881	A01000003148	1	A01	A010039	PD0290	20210101	10	19000.0	1
13	M707991383	A01000012930	1	A01	A010053	PD0900	20210101	11	125300.0	1
32	M625839441	A01000026464	1	A01	A010045	PD0787	20210101	12	39500.0	1
33	M625839441	A01000026464	1	A01	A010045	PD0787	20210101	12	39500.0	1
62	M165312283	A01000046512	1	A01	A010057	PD1255	20210101	13	1600.0	1
4381738	M533286446	E06052119978	2	A06	NaN	PD0507	20211231	23	50000.0	1
4381739	M533286446	E06052119978	2	A06	NaN	PD0507	20211231	23	50000.0	1
4381740	M533286446	E06052119978	2	A06	NaN	PD0507	20211231	23	50000.0	1
4381741	M533286446	E06052119978	2	A06	NaN	PD0507	20211231	23	50000.0	1
4381742	M533286446	E06052119978	2	A06	NaN	PD0507	20211231	23	50000.0	1

중복 제거

4381743개의 전체 데이터에서 모든 컬럼에서 중복이 나타난 약 5.4% 데이터를 확인함

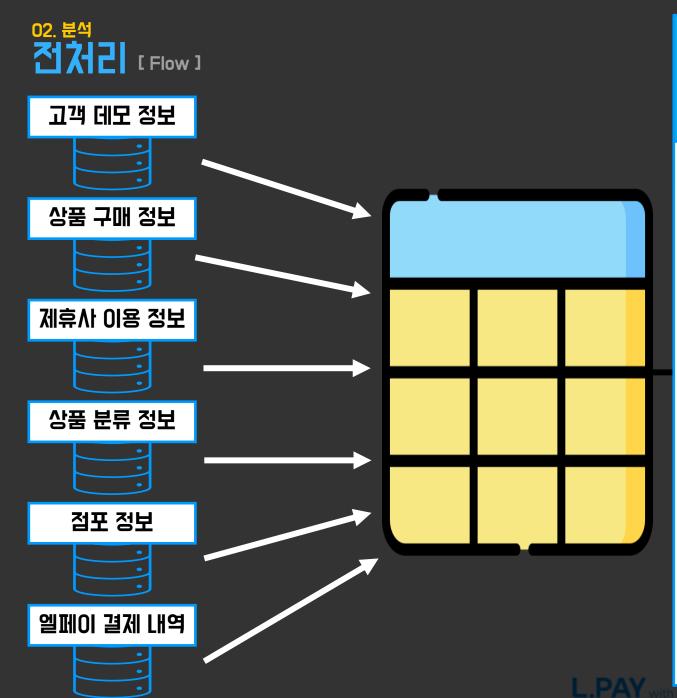
모든 컬럼이 일치하고. 같은 영수증에서 같은 상품코드로 구매를 하였는데, 구매수량이 모두 1개인 것을 확인함

데이터 자체의 오류라 판단하여 중복 데이터를 제거함

결측치 처리

'접포코드'에 결측치 55개가 존재하는 데이터의 '채널구분'에 따르면 결측치가 있는 구매는 모두 <mark>온라인 채널에서의 구매</mark>이므로, 결측치를 '온라인' 텍스트로 대치함





Pre-processing 2 [Total Dataset 구축]

모든 데이터 테이블의 정보를 쓸 수 있도록. 모든 데이터를 결합한 Total Dataset을 구축함

- i) [상품 구매 정보] 데이터를 기준으로 [제휴사 이용 정보], [엘페이 결제 내역] 컬럼을 통일하고, 이 데이터들을 Concat하여 Total DataFrame을 구성
- ii) Total DataFrame에 '고객번호'를 기준으로 [고객 데모 정보] 데이터를 Merge
- iii) Total DataFrame에 '접포코드'와 '제휴사'를 기준으로 [상품 분류 정보] 데이터를 Merge
- iv) Total DataFrame에 '상품코드'를 기준으로 [점포 정보] 데이터를 Merge

=> Total Dataset 구축



Pre-processing 3 [고객별 RFM Table 구축]



데이터셋의 시작시점을 기준으로 가장 최근 구매한 날짜까지의 일수 간격을 구함 -> 가장 최근에 방문한 고객이 가장 높은 점수



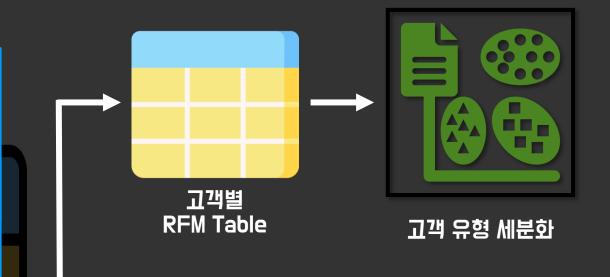
데이터셋에서 '영수증 번호'를 기준으로 도출한 구매횟수

-) 가장 많은 영수증 기록을 가진 고객이 가장 높은 접수



총 구매금액

-) 가장 많은 총 구매금액을 가진 고객이 가장 높은 접수



RFM이란?

사용자별로 얼마나 최근에, 얼마나 자주, 얼마나 많은 금액을 지출 했는지에 따라 사용자들의 분포를 확인 하거나 사용자 그룹(또는 등급)을 나누어 분류 하는 분석 기법

Recency (거래 최근성): 얼마나 최근에 구입했는가?

Frequency (거래빈도): 얼마나 빈번하게 우리 상품을 구입했나?

Monetary (거래규모): 구입했던 총 금액은 어느 정도인가?



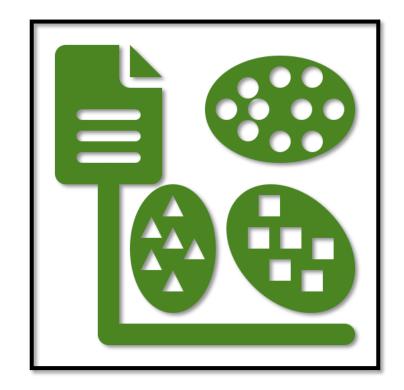
02.분석

고객행동분석 및 유형화

[k-means Clustering]

	Customer_ID	Recency	Frequency	Monetary
0	M000034966	356	13	616240
1	M000136117	363	89	28697031
2	M000201112	331	5	51420
3	M000225114	364	88	2084270
4	M000261625	337	49	6072050
26207	M999673157	351	17	2851370
26208	M999770689	364	216	2078770
26209	M999849895	334	35	1095312
26210	M999926092	335	7	129500
26211	M999962961	364	154	15582194

K-means Clustering



고객별 RFM 테이블

고객 유형 세분화



02.분석 고객행동분석 및 유형화 [k-means Clustering]

Customer Clustering

K-means

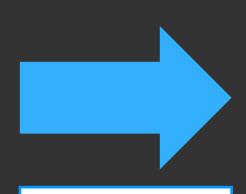
각 데이터로부터 '데이터가 속한 군집의 중심까지의 거리'를 기준으로 서로 유사한 데이터들은 같은 그룹으로, 유사하지 않은 데이터는 다른 그룹으로 분리하는 비지도학습 방법

실루엣 계수

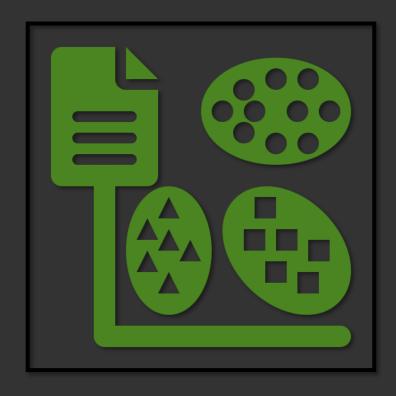
각 데이터 포인트와 주위 데이터 포인트들과의 거리 계산함 군집 안에 있는 데이터들은 잘 모여있는지, 군집끼리는 서로 잘 구분되는지 클러스터링을 평가하는 척도

Clustering을 위해 k-means, k-medoids, DBSCAN, Gaussian Mixture Model, 등 다양한 방법 사용

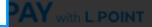
> -) 실루엣 계수를 기준으로 접수가 가장 높았던 K-means를 사용함 (약 0.588)



K-means Clustering



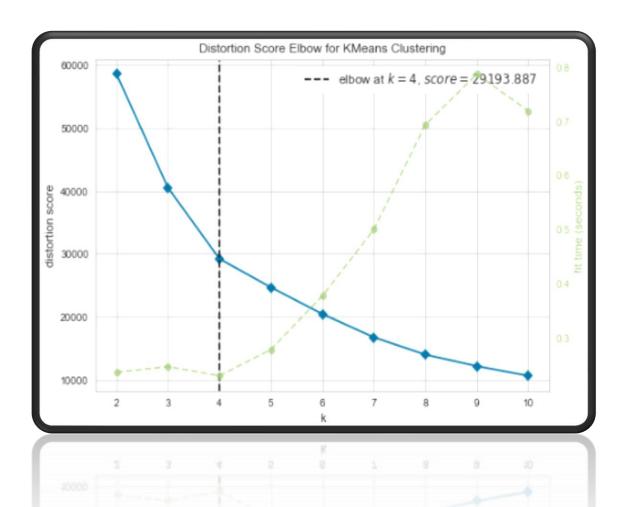
고객 유형 세분화



02.분석

고객행동분석 및 유형화

[k-means Clustering]



Elbow point

KMeans는 주어진 데이터를 K개의 군집으로 그룹화하는 알고리즘 이때 군집의 갯수 K개는 하이퍼파라미터로, 각 군집간의 거리의 합인 inertia가 급격이 떨어지는 구간을 K로 설정해야 함 이를 Elbow Method라고 함

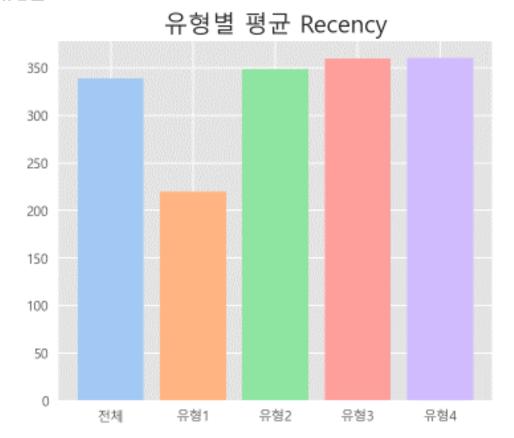
고객별 RFM 테이블의 K-means Clustering 실행 결과.

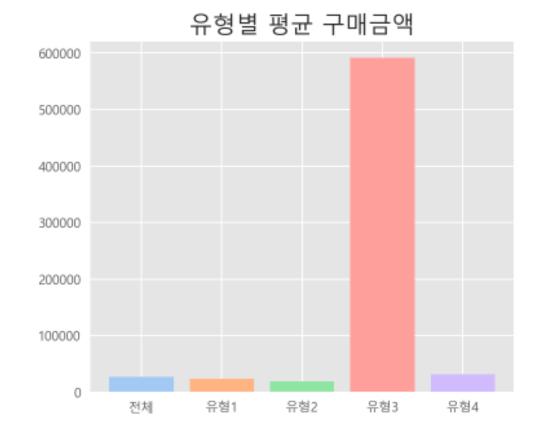
Elbow point가 K=4 구간이므로.

고객 유형을 4개 군집으로 세분화함



[유형별 EDA]

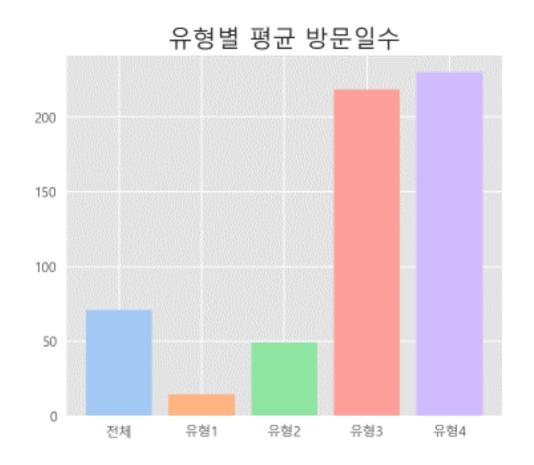




나른 유형은 전체 평균가 큰 차이를 보이지 않으나 유형1의 경우 비교적 최근에 방문을 하지 않음 나른 유형은 전체 평균과 큰 차이를 보이지 않으나 유형3의 경우 평균 구매금액이 높음을 알 수 있음



[유형별 EDA]



"

유형1, 2의 경우 전체 평균과 비교하였을 때. 방문 일수가 적은 모습을 보이고 유형3, 4의 경우 방문 일수가 많음을 알 수 있음

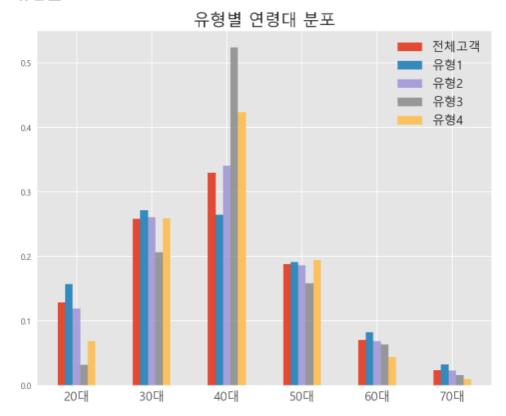


방문 횟수를 기준으로 보았을 때, 유형3, 4의 경우 충성 고객일 확률이 높음



"

[유형별 EDA]



유형 1.

20대의 비율이 조금 높은 경향을 보이고. 40대의 비율은 적은 모습을 보임

유형 2.

연령대 분포상 두드러지는 특징을 찾기 어려움

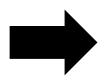
유형 3.

40대의 비율이 매우 높고, 20대의 비율은 매우 낮음

유형 4.

40대의 비율이 매우 높고, 20대의 비율은 낮음

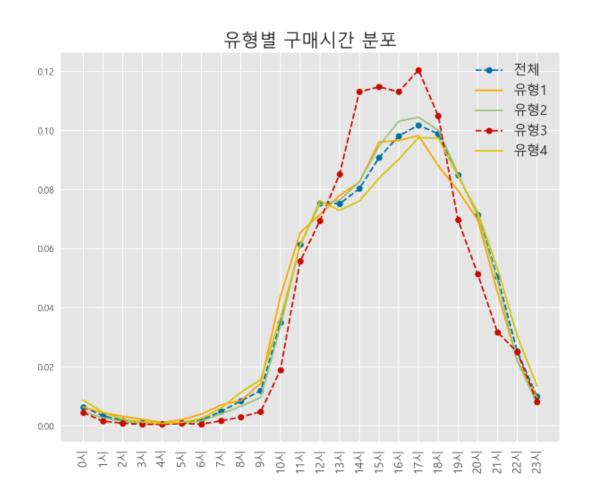
L.PAY with L.POINT



유형3, 4의 경우 구매력이 높은 40대의 분포가 많은 것을 알 수 있으며,

두 유형 모두 구매력이 높은 40대의 분포가 많고, 이용량이 많지만 구매 금액에서 차이를 보임 유형1의 경우 구매력이 비교적 낮은 20대의 분포가 많은 것을 알 수 있다.

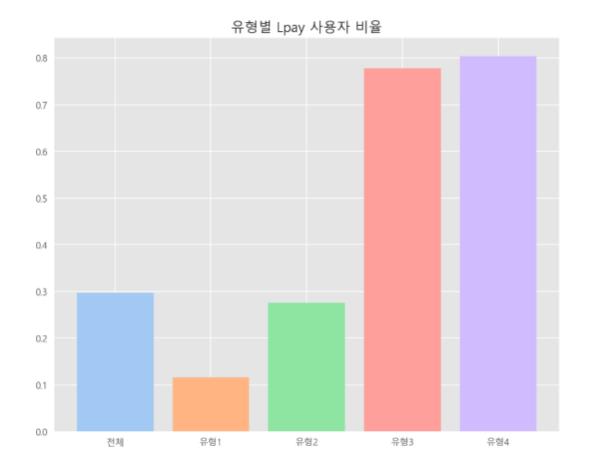
[유형별 EDA]



대부분의 유형들은 비슷한 형태를 보이는 반면 유형 3의 경우 구매시간이 오후 2시 ~ 5시 사이에 많이 분포하는 모습을 보임



[유형별 EDA]



"

전체와 비교해 보았을 때. 유형1의 경우 Lpay이용자의 수가 매우 적고. 유형3.4의 경우 이용률이 높음을 알 수 있음.

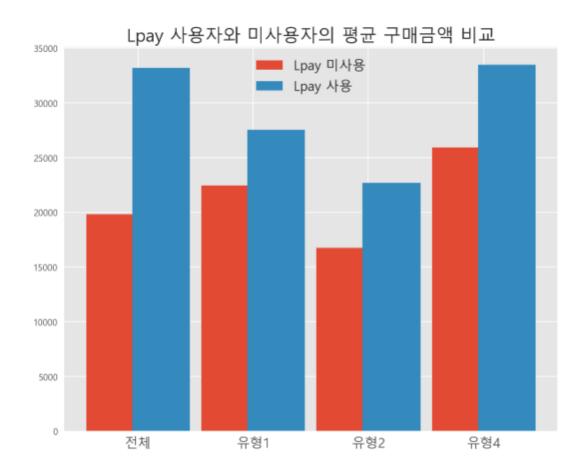


Lpay를 사용하는 사람이 많다는 것은 롯데의 이용률이 높다는 것으로 볼 수 있기 때문에

유형3. 4의 경우 충성 고객일 확률이 높음



[유형별 EDA]



구매금액의 차이가 큰 유형 3을 제외한 나머지 유형들의 Lpay사용자와 미사용자의 구매금액을 비교한 결과이다.

유형 3을 제외한 모든 유형에서 Lpay 사용자의 평균 구매금액이 미사용자보다 높음을 알 수 있다.



Lpay 사용을 장려하는 마케팅 진행시 수의 증대의 효과를 얻을 수 있다고 판단



02.분석

고객행동분석 및 유형화

[유형별 EDA]



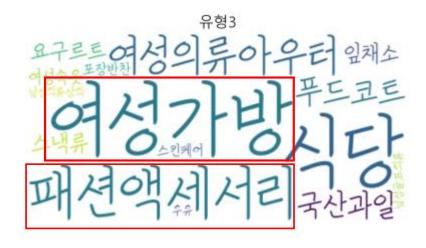


모든 집단에서 많이 나타난 품목을 제외하고 보면 유형1, 2, 4 모두 최다 구매품목이 비슷한 모습을 보임.

-> 평균적으로 구매하는 제품군은 비슷하다고 판단됨.



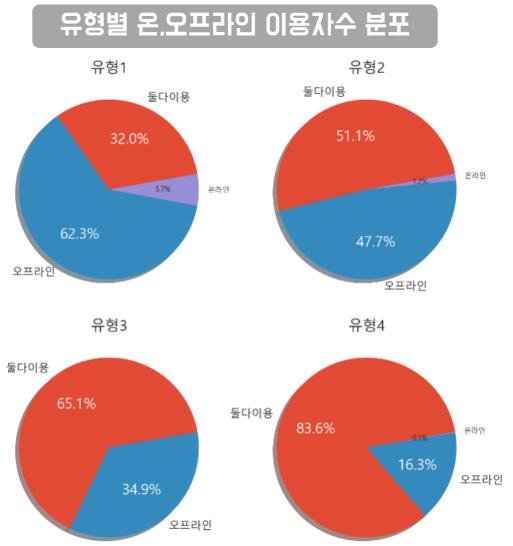
유형3 최다구매품목



유형 3의 경우 '여성가방', '패션액세서리' 등 다른 유형에 비해 비교적 가격대가 있는 제품군을 많이 구매한 것을 알 수 있다. 평균 구매금액이 높은 집단임을 고려해볼 때, 유형 3의 주 거래 제휴사는 백화점이며, 흔히 VIP로 분류되는 고객일 가능성이 높음



[유형별 EDA]



구매 내역을 바탕으로 온라인만 이용한 사람을 '온라인'. 오프라인만 이용한 사람을 '오프라인'.

모두 이용한 적이 있는 사람을 '둘다이용'으로 분류하여 분석.

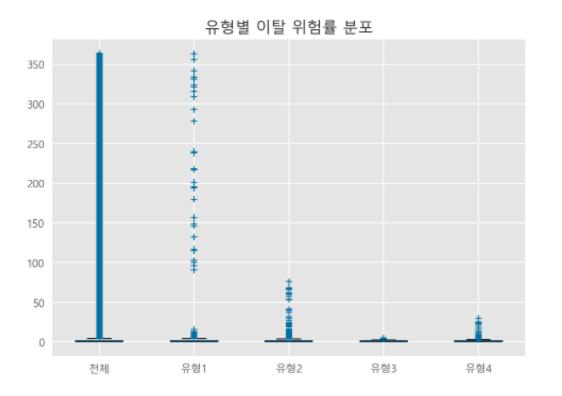
유형 3, 4의 경우 온라인만을 이용한 고객의 분포가 거의 없는 반면, 유형 1, 2의 경우 비교적 오프라인과 온라인의 비율이 높음.



유형 1, 2, 4의 경우 최다 구매품목이 비슷했기 때문에 유형 1, 2의 고객들에게 온라인에 대한 홍보를 한다면 충성고객의 수를 늘려 매출 증대에 기여할 수 있을 것으로 판단



[유형별 EDA]



최종 구매경과일과 평균 구매주기를 통해 고객의 이탈 위험률을 계산한 후 유형별 이탈 위험률 분포 사분위수를 그래프로 나타낸 결과이다.

유형 1, 2의 이탈 위험률 분포가 유형 3, 4보다 높음을 알 수 있다.



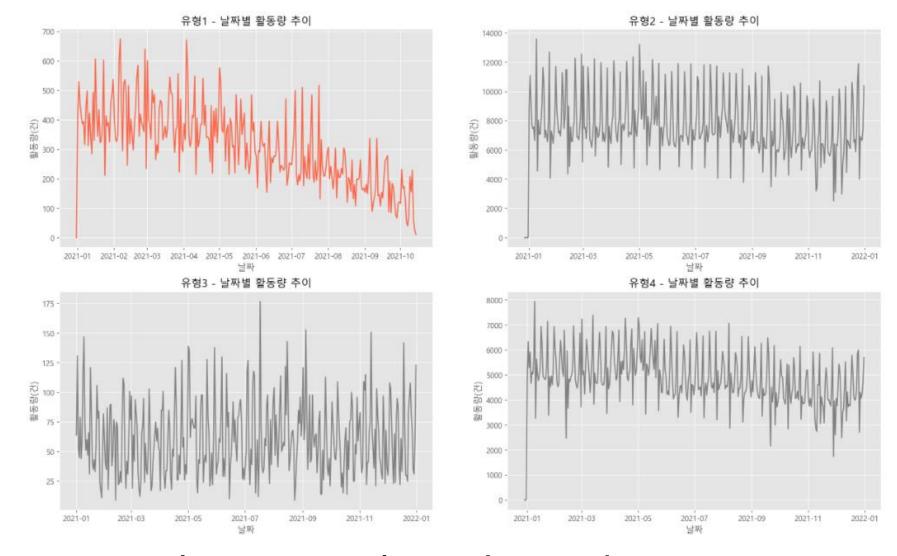
유형 3. 4의 경우 이탈 위험률이 낮음을 보아 충성 고객임을 알 수 있다



02.분석

고객행동분석 및 유형화

[유형별 EDA]



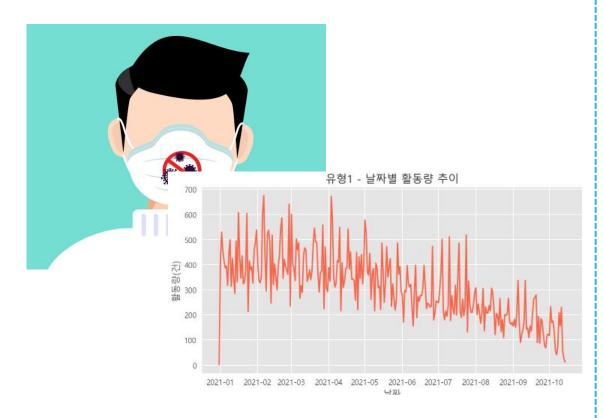
유형 1의 경우 날짜별 활동량의 추이가 감소하는 모습을 보임



02.분석

고객행동분석 및 유형화

[유형별 EDA-외부데이터]



제공받은 데이터는 코로나가 유행하던 시기



유형1의 활동량이 감소하는 이유는 코로나와 관련이 있지 않을까?

대한민국의 코로나19 총 누적 확진자와 사망자

	국가	구분	일자	누적확진(명)	신규확진(명)	인구확진률(%)	인구(명)	누적사망(명)	신규사망(명)	사망률(%)
0	대한민국	일간	2021-01-01	62593	824	0.12	51815810	942	25	1.50
1	대한민국	일간	2021-01-02	63244	651	0.12	51815810	962	20	1.52
2	대한민국	일간	2021-01-03	64264	1020	0.12	51815810	981	19	1.53
3	대한민국	일간	2021-01-04	64979	715	0.13	51815810	1007	26	1.55
4	대한민국	일간	2021-01-05	65818	839	0.13	51815810	1027	20	1.56
	• • •		•••							•••
581	대한민국	일간	2022-08-05	20383621	110610	39,34	51815810	25236	45	0.12
582	대한민국	일간	2022-08-06	20489128	105507	39.54	51815810	25263	27	0.12
583	대한민국	일간	2022-08-07	20544420	55292	39.65	51815810	25292	29	0.12
584	대한민국	일간	2022-08-08	20694239	149819	39.94	51815810	25332	40	0.12
585	대한민국	일간	2022-08-09	20845973	151734	40.23	51815810	25382	50	0.12

586 rows × 10 columns

2021년 코로나 신규 확진자, 사망자, 등 코로나에 관한 데이터를 <mark>외부 데이터</mark>로 사용

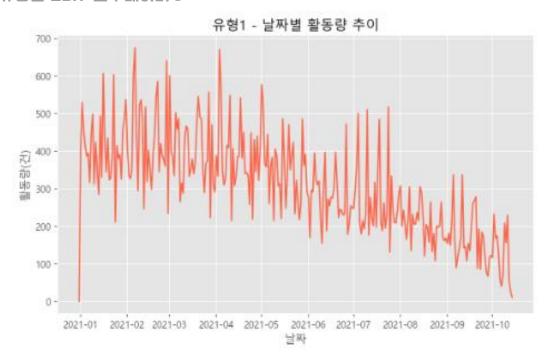
외부데이터출처: O INSFILER
https://insfiler.com/detail/rt_corona19_all-0008

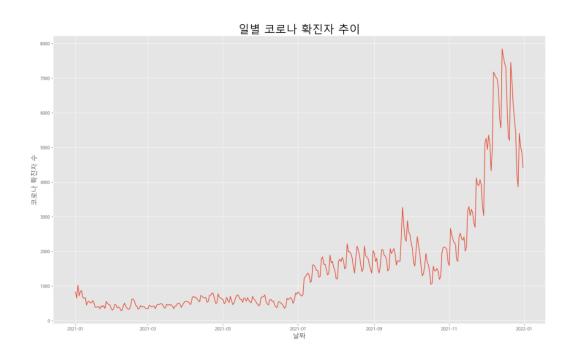


02.분석

고객행동분석 및 유형화

[유형별 EDA-외부데이터]





고객 유형 1의 경우 시간의 흐름에 따라 활동량이 감소하는데, 코로나 확진자의 수는 증가하는 것으로 보아 연관이 있을 것으로 판단



유형 1의 활동량과 코로나 확진자에 정말 유의미한 관계가 있는 것인가?

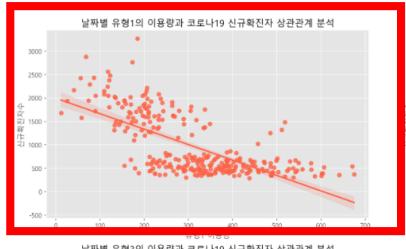


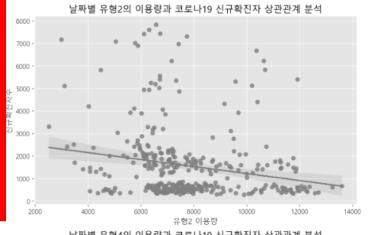
02.분석

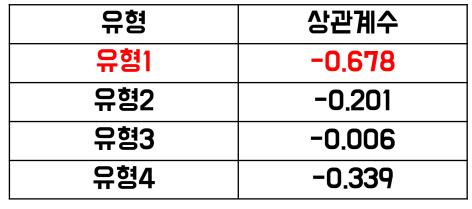
고객행동분석 및 유형화

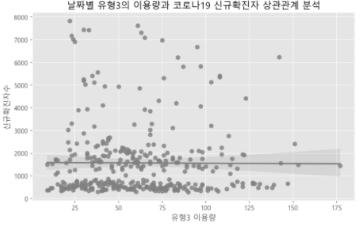
[유형별 EDA-외부데이터]

유형 1의 활동량과 코로나 확진자 사이의 관계가 있는지 피어슨 상관계수를 확인하여 검증











유형1의 경우

-0.678의 강한 상과계수를 보임

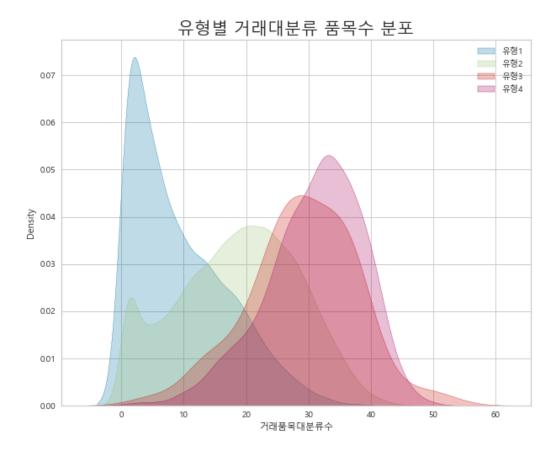
즉, 코로나 확진자수가 증가할수록

유형 1의 활동량은 감소함을 알 수 있음



^{02.분석} 고객행동분석 및 유형화

[유형별 EDA]



유형별로 구매했던 대분류수의 분포를 KDE plot으로 나타낸 결과 이다. X축은 거래 대분류 품목수, y축은 밀도를 나타낸다.

유형 3, 4의 경우

30-35사이의 다양한 품목을 구매한 고객의 분포가 높은 반면 유형 1의 경우 0-10사이의 적은 품목을 구매한 고객의 수가 높다. 유형 2의 경우 위의 두 유형 사이에 가장 많은 고객이 분포한다.



충성도가 높은 유형일수록 거래하는 대분류품목수가 많음을 알 수 있음





소비금액

높음

유형3 <u>롯데고수</u>

95 유형을 <mark>충성도</mark>와 <mark>소비금액</mark>에 따라 분류



매우 낮음 나는 그 사람들이 나는 그 사람들이 나는 그 사람들이 나는 그 사람들이 가는 그 사람들이 되었다면 그 사람들이 되었다면

유형1

"

'크로나 싫어'



유형2 초보주부



유형4 주부9단



충성도

낮음

고객행동분석 및 유형화

[고객 유형 정의]

유형1 '코로나 싫어'

전체 유형중 20대 비율 가장 높음 전체 유형중 방문한지 가장 오래된 평균 방문일수 낮음 평균 구매금액 낮음 Lpay사용비율 가장 낮음 온라인 사용비율 가장 낮음 이탈 위험률 분포 가장 높음



유형2 초보주부

전연령대 고르게 분포 최근 방문 이력 있음 평균 방문일수 낮음 구매금액은 평균과 가까움 Lpay사용비율 낮음 온라인 사용비율 낮음 이탈 위험률 분포 비교적 높음



무덤3

전체 유형중 40대 비율 가장 높음 최근 방문 이력 있음 평균 방문일수 높음 평균 구매금액 매우 높음 Lpay사용비율 비교적 높음 온,오프라인 동시 이용 비율 높음 충성 고객으로 판단



유형4 주부9단

40대 비율 높음 최근 방문이력 있음 전체 중 평균 방문일수 가장 높음 평균 구매금액 평균 Lpay사용비율 가장 높음 온,오프라인 동시 이용 비율 가장 높음 충성 고객으로 판단





외부데이터와 EDA를 활용한 II생변수

[파생변수 생성 기준]

파생변수 생성기준

- 1. EDA를 바탕으로 각 유형을 잘 분류할 수 있는 변수생성
- 2. 신규 고객 유형 예측에 불리하게 작용할 변수 생성 금지 ex) 총구매금액, 총구매건수
 - -> 신규 고객의 경우 총 구매금액과 총구매건수가 기존 고객에 비해 적기 때문에 변수가 생성된다면 올바른 예측이 이뤄질 수 없음

코로나 신규 확진자수 평균, 코로나 신규 사망자수 평균, 주구매시간분류1, 주구매시간분류2, 평균구매금액,

> 평균구매월 평균 낱개 구매금액. 이용채널. Lpay 이용여부



외부데이터와 EDA를 활용한 II생변수

[파생변수 생성-중요변수]

평균 구매 금액



44 고객이 한번 거래할 때, 평균적으로 지출하는 금액

평균 구매 금액을 보았을 때, 유형3의 경우 다른 유형들과 명확한 차이를 보였음



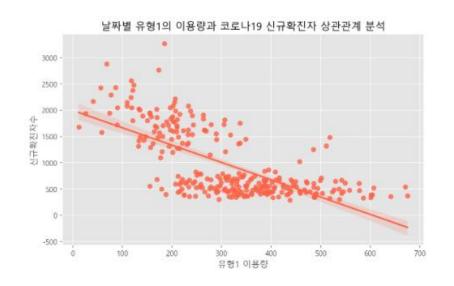
고객의 거래내역 중 <mark>구매금액의 평균</mark>을 보면 고객의 잠재 유형을 예측하는데 도움이 될 것이라 판단함



외부데이터와 EDA를 활용한 파생변수

[파생변수 생성-중요변수]

코로나 관련 변수



유형 1의 경우 코로나 확진자수에 영향을 많이 받음

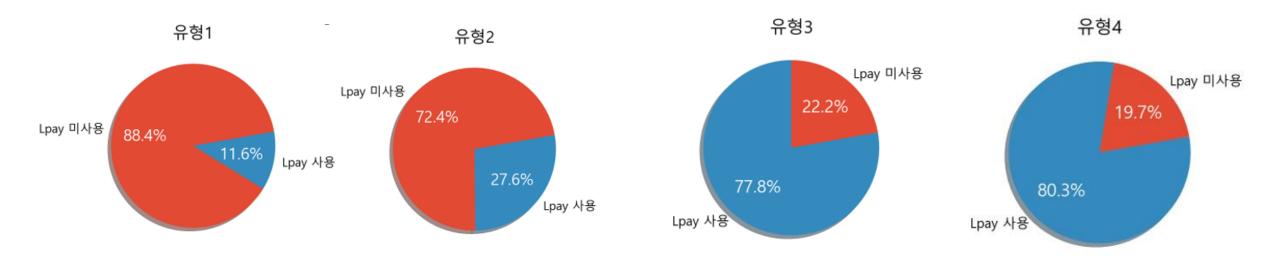


코로나의 정도를 알 수 있는 신규 확진자수와 신규 사망자수를 사용하여 고객이 구매한 날짜들의 평균 신규 확진자수와 평균 신규 사망자수 변수를 생성

외부데이터와 EDA를 활용한 파생변수

[파생변수 생성-중요변수]

Lpay 사용 여부



(유형1, 유형2)에 비해 (유형3, 유형4)가 Lpay를 압도적으로 많이 사용



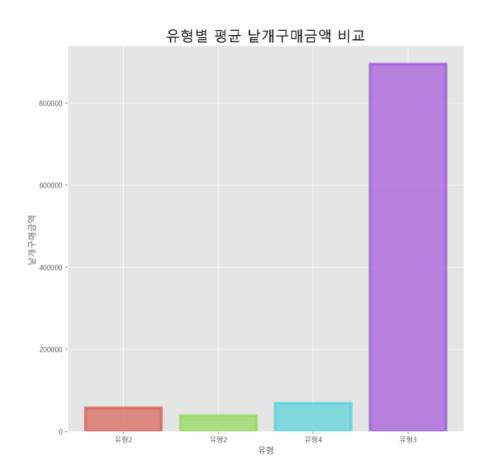
고객의 Lpay 사용 여부 변수를 생성한다면 유형1.2와 유형3.4를 구분할 수 있는 유의미한 변수가 될 것이라고 판단



외부데이터와 EDA를 활용한 파생변수

[파생변수 생성-중요변수]

평균 낱개 구매 금액



46 고객이 한번 거래할 때, 하나의 수량의 평균금액

> 유형3이 다른 유형에 비해 낱개로 구매하는 제품의 금액대가 높은 것을 알 수 있음



유형 3은 다른 유형과 확연한 차이를 보임 유형 3을 제외한 유형들 사이에도 조금의 차이를 보임



03.모델링 외부데이터와 EDA를 활용한 II~생변수

Feature Importance(CART)?

변수의 중요도를 판단하는 방법이다.

특성을 기준으로 샘플들을 spit할 때,

불순도가 감소하는 양을 바탕으로 불순도 감소가 클 경우 중요한 변수라고 판단하는 방법

다른 변수들과 상관관계를 배제한 독립적인 변수의 중요도를 확인하지 못한다는 단점을 가짐

Permutation Importance?

CART 방식이 가진 단점을 보완한 방법으로 변수들 간의 상호 관계를 배제하고 독립 변수의 중요도를 확인하는 방법

-> 위의 두가지 방법을 함께 사용하여 변수의 중요도를 확인하는 것이 가장 바람직함!



외부데이터와 EDA를 활용한 파생변수

[파생변수 생성-변수중요도 검증]

CART Feature Importance



Feature importance

Permutation Importance

Weight	Feature
0.2674 ± 0.0096	평균구매금액
0.1932 ± 0.0052	코로나신규사망자수평균
0.1663 ± 0.0248	최소구매금액
0.1590 ± 0.0115	평균낱개구매금액
0.1092 ± 0.0156	평균구매일
0.1061 ± 0.0021	<u>구매월평균</u>
0.0922 ± 0.0049	Lpay이용여부
0.0806 ± 0.0057	코로나신규확진자수평균
0.0683 ± 0.0073	영수증번호길이평균
0.0432 ± 0.0109	구매요일최대값
	21 more

모든 변수 중 중요한 변수 상위 10개씩을 도출한 결과이다.

EDA를 통해 생성한 변수의 중요도가 높음을 알 수 있음.

외부 데이터를 사용하여 생성한 변수가 두방법 모두에서 관찰됨



외부데이터와 EDA를 활용한 파생변수

[파생변수 생성-일반변수]

일반 파생변수

영수증 거래내역

영수증번호길이평균, 영수증번호길이최대, 영수증 타입 최대값

구매일자

최다구매월,최다 구매 계절,평균구매일, 최다 구매 요일, 주중 주말 선호도, 평균구매월, 구매 요일 최대값, 최다 구매년도

구매시간

평균구매시간, 최다구매시간,주구매시간분류1(새벽,오전,오후,저 역),주구매시간분류2(출근, 퇴근)

구매 금액

최대구매금액, 최소구매금액, 평균 낱개 구매금액

구매 수량

평균구매수량, 최대구매수량, 최소구매수량

7IEI

이용채널, 최다구매경로,최다제휴사타입, 최다제휴사

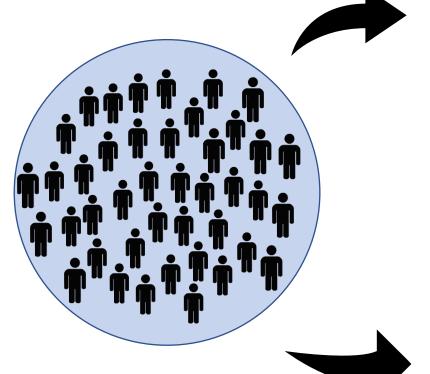


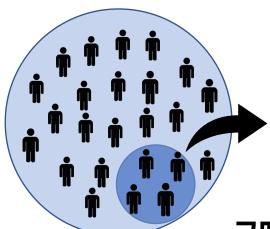


신규고객 군집 예측모델

[모델 검증 과정

데이터 분리

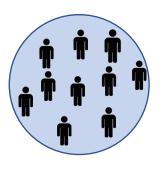




[Train]

[Train] 의 20%를 검증데이터 [Valid]로 활용

구매횟수가 3건이 넘어가는 고객을 추출하여 해당 고객을 대상으로 모델을 학습



구매횟수가 3건 이하인 고객을 대상으로 예측



03.모델링 시규고객 군집 예측모델

사용 모델



Light Gradient Boosting Model (LGBM)

Graident Boosting 프레임워크로

Tree 기반 학습 알고리즘으로 속도가 빠르기 때문에

큰 사이즈의 데이터를 다룰 수 있고. 실행 시 적은 메모리를 차지한다는 장점을 가짐
속도가 빠르고 사용 메모리가 적다는 장점을 가지기 때문에 실제 적용 시 이점을 가짐





신규고객 잠재유형 예측모델

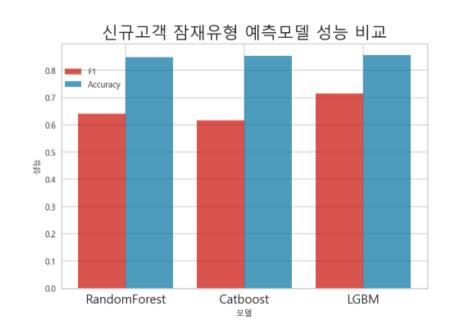
Model	LGBM	속도와 성능면에서 가장 우수		
Encoding	Label Encoding	One Hot Encoding보다 속도가 빠르고 간단함		
Scaling	Standard Scaler	기존 변수의 범위를 정규분포로 변환		
목적	Multi Class	47\지의 유형으로 예측		
평가지표	Accuracy, F1 Score	정확도에 더해 데이터의 불균형까지 고려함		



03.모델링 신규고객 군집 예측모델



	구분	F1 Score	Accuracy Score			
	RandomForest	0.6404	0.8483			
7	LGBM	0.7139	0.8556			
	CatBoost	0.6161	0.8527			



다른 모델들과 성능 비교 결과 LGBM이 속도와 성능 측면에서 가장 좋은 모습을 보임





04.전략 유형별 개인화 마케팅 [기존고객]

유형1 <mark>'코로나 싫어'</mark>

#방문횟수 적은 #온라인 이용률 낮은 #Lpay이용률 낮은 #20대 많은 # 코로나의 영향을 많이 받은

마케팅 전략 1. 온라인 마케팅

코로나의 영향을 많이 받는 집단이기 때문에 비교적 코로나의 영향을 덜 받을 수 있는 '온라인' 마케팅 실시 집단 내 20대의 분포가 많기 때문에 온라인의 이용을 권장할 수 있는 이벤트 기반 마케팅을 한다면 <mark>참여율이 높을 것으로 예상</mark>

마케팅을 통해 온라인 이용률을 높인다면 해당 유형의 약점이었던 방문횟수, 온라인 이용률, 코로나로 인한 이탈 위험률을 보완 가능함

마케팅 전략 2. Lpay 권장 마케팅

온라인에서 사용 가능한 Lpay 제휴사의 홍보, 확장을 통해 Lpay 이용률을 올릴 수 있는 마케팅 실시

Lpay를 이용하는 고객의 수가 올라간다면 '코로나 싫어' 유형의 분포가 '주부9단'의 분포와 비슷하게 변하게 되고 자연스럽게 충성고객으로 전환될 수 있음

마케팅 전략 3. 개인화 Push 알림 마케팅

해당 유형은 일회성 구매를 하는 고객이 많기 때문에 지속적인 접속/방문을 유도할 수 있는 마케팅 진행. 개인화 Push알림 서비스를 진행하여 지속적인 접속/방문을 유도하여 리텐션을 확보



04.전략 유형별 개인화 마케팅 [기존고객]

유혐2 초보주부

#방문횟수 적은 #온라인 이용률 낮은 #Lpay이용률 낮은

'초보주부' 유형은 최근 방문 이력이 있고, 구매 금액 또한 평균과 비슷하지만 방문횟수가 낫다는 문제점이 있다.

-> 자주 방문할 수 있는 마케팅을 통해 방문회수를 높인다면 '주부9단'과 같은 형태의 충성고객으로 발전할 가능성이 있다.

마케팅 전략 1. 할인, 이벤트 등 방문횟수를 늘이기 위한 방안수립

'초보주부' 유형이 주로 구매하는 품목의 할인을 통해 가격 경쟁력을 바탕으로 방문 횟수를 늘릴 수 있는 마케팅을 실시함. 이벤트를 통해 방문횟수를 늘일 수 있는 마케팅 방안 수립. Ex) 특정 기간 동안 방문 고객에 한해 적립율을 높임, 할인을 진행

마케팅 전략 2. 온라인, Lpay 마케팅

'코로나 싫어' 유형과 마찬가지로 온라인과 Lpay 사용률을 높이기 위한 마케팅을 진행함. 온라인 이벤트를 통하여 온라인으로 귀사의 제휴사를 방문하기만 하여도 포인트를 주는 등의 마케팅을 진행한다면 자연스럽게 방문 횟수가 높아지고 수익 창출로 이어질 것으로 예상함

추가로 Lpay로 처음 결제하였을 때의 혜택을 더욱 높인다면 (ex) 첫결제 10% 확정할인) Lpay 사용률 제고에 도움이 될 수 있다고 판단



04.전략 유형별 개인화 마케팅 [충성고객]

유형3 롯데 고수

#방문횟수 많음 #온라인 이용률 높음 #Lpay이용률 높음 #구매력 높음 #40대 비율 많음 #충성고객 '롯데 고수' 유형은 롯데에 대한 충성도도 높고 사용 금액도 많은 유형으로 MVG의 혜택을 높이고 롯데에 대한 긍정적인 이미지를 가질 수 있는 마케팅을 통해 이탈을 방지하는 전략으로 마케팅을 진행

마케팅 전략 1. MVG전용 혜택 고지

해당 유형의 고객에게는 다른 고객에 비해 더욱 고급화된 서비스를 통해 해당 고객들이 MVG라는 느낌을 받을 수 있는 혜택을 제공 Ex) 1. 구매금액을 바탕으로 등급을 정하여 더욱 높은 등급을 달성하고 싶어하는 MVG의 심리를 자극.

마케팅 전략 2. 구매금액에 따른 캐시백

'롯데 고수' 유형의 고객은 재방문 확률이 높은 고객이기 때문에 구매 금액에 따라 '자사 상품권'을 지급하는 마케팅을 진행 -> 100만원 구매고객에게 5만원 상품권을 지급한다면, 5만원의 사용을 위한 재방문시 투입비용 대비 큰 효과를 얻을 수 있음

04.전략 유형별 개인화 마케팅 [충성고객]

유형4 <mark>주부9단</mark>

#방문횟수 많은 #온라인 이용률 높은 #Lpay이용률 높은 #구매금액 평균 #40대 비율 많은 #충성고객 연령대와 구매 품목을 확인해본 결과 '주부9단'의 대표 인물은 '40대 주부'일 것으로 예상, 주부를 겨냥한 마케팅 실행

마케팅 전략 1. '주부'겨냥 마케팅 전략

- 1. 아이와 남편이 집에 없고, 어느정도 집안일이 끝난 한가한 오후 시간대를 공략하여 특정 시간대에 방문을 유도할 수 있는 마케팅을 진행 ex) 백화점, 마트 등 공간에 상품 구매와 별개로 주부들이 주로 이용할 수 있는 프로그램 마련 -> 방문을 통해 구매까지 연결될 수 있음
- 2. 주부모임 단위 공략 마케팅을 통해 기존 고객과 연결된 신규 고객 유입 유도 ex) 여성 3인 이상 이용시 할인 혜택 -> 기존 고객을 통하여 새로운 고객의 유입을 기대할 수 있음

마케팅 전략 2. 아이를 통한 우회 공략 마케팅 전략

아이들을 위한 이벤트를 진행할 경우 마케팅의 대상은 자연스럽게 '주부'가 됨. 이를 이용하여 아이를 통한 지출 유도 ex) 키즈 메뉴에 추가 할인을 줌

마케팅 전략 3. 개인화 Push 알림 마케팅

해당 유형은 충성도가 높은 유형이기 때문에 개인화 Push알림을 통해 이용률을 유지할 수 있는 마케팅 진행. ex) 개인화 Push알림 서비스를 진행하여 '세일 정보'를 알림



^{04.전략} 유형별 개입화 마케팅

[실제 적용 예시]



김00. 26세. 여성 "거래내역이 3건 이하인 신규고객"

	고객번호	영수증번호	채널구분	계휴사	구매일자	구매시간	구매금액	성별	연령대	중분류명
7000	M412637909	A02199868530	1	A02	20210726	19	500	여성	20대	주방정리용품/소모품
7001	M412637909	A02199868530	1	A02	20210726	19	1380	여성	20대	냉장간편식
7002	M412637909	A02199868530	1	A02	20210726	19	1260	여성	20 C H	향신료



신규고객 잠재유형 예측모델



잠재 유형: '코로나 싫어'유형

*특징

-유형의 특징처럼 일회성 거래 형태를 보일 가능성이 높음 -온라인 이용과 Lpay 이용률이 적을 확률이 높음

*마케팅

개인화 Push알림 마케팅 온라인 마케팅 Lpay 권장 마케팅



04.전략 유형별 개인화 마케팅

[비즈니스 기대효과]



충성고객 수의 증가로 인한 수익 창출 기대



신규 고객의 <mark>잠재 유형을</mark> 예측하여 통한 맞춤형 마케팅



개인화 마케팅을 통한기업의 긍정적 이미지 제고



Thank you





