

L사 구매관련 데이터를 활용한 구매감소 예측

5조 발표

윤예은 정재훈 정제훈 최지원 허진욱



문제제기

01

02

데이터 처리 방안
및 활용 분석
기법

예측 모델

03

04

군집화 및 솔루션



1. 문제제기

뉴스룸 | 최신기사

백화점고객, 먹는데만 지갑열어...올해 식품매출 11% ↑

송고시간 | 2014-12-29 06:29

전체 매출은 정체...백화점 식품매장·식당은 '대박' 행진

(서울=연합뉴스) 신호경 기자 = 경기 불황과 해외 직접구매(직구), 온라인·모바일 쇼핑 등의 영향으로 올해 백화점 매출은 전반적으로 매우 부진했지만, 유일하게 식품(식당 포함) 부문만은 두 자릿수 성장세를 이어간 것으로 나타났다.

이처럼 백화점내 식품매장과 식당이 쇼핑 후 들르는 '부속 시설'이 아니라 고객이 백화점을 찾는 '첫 번째 목적'이 되면서, 국내외 유명 맛집과 식료품 매장을 유치하기 위한 백화점들 간 경쟁도 뜨거워지고 있다.

◇ 전체 매출은 3%대 저성장...식품 부문은 9~11% 늘어

홈 > 생활경제

롯데백화점, '셀프기프팅족' 겨냥 마케팅 펼쳐

고객 95% '스스로에게 선물할 것'...예상비용 전년비 10만원 ↓

김봉식 기자 | tiger3936@finomy.com



승인 2014.12.16 14:50:56



[현대경제신문 김봉식 기자] 롯데백화점이 '셀프기프팅(Self-gifting)족'들을 겨냥해 '나를 위한 작은 사치'라는 테마로 마케팅을 펼친다.

이는 최근 사회적 피로도 및 개인적인 스트레스가 높아짐에 따라 자기 보상을 위해 자신에게 스스로 선물하는 셀프기프팅족이 늘고 있는 데 따른 것이다.

롯데백화점은 지난달 22일과 23일 양일간 20~60대 방문고객 천명을 대상으로 설문조사를 진행한 결과 '자신을 위한 연말 선물을 준비하겠느냐'라는 질문에 '그렇다'고 답한 고객이 95%였다고 16일 밝혔다.

이는 지난해 96%와 비슷한 수준으로 셀프기프팅이 새로운 소비문화로 자리 잡고 있는 것으로 해석된다.

2014년 트렌드 'DARK HORSES'란?

D Dear, got swag 가벼운 허세, 스웨그

A Answer is in your body 몸으로 말해요

R Read between the ultra-niches 극단적 틈새시장

K Kiddie 40s '어른아이', 40대 남성들

H Hybrid Patchworks 융합

O Organize your platform 플랫폼을 펼쳐라

R Reboot everything 과거의 재해석

S Surprise me, guys! 깜짝 이벤트

E Eyes on you, eyes on me 관음사회

S Say it straight 돌직구 스타일



〈2014□소비트렌드〉

2015년 유통 키워드는

S SPA Living
유명 SPA(유통제조일괄형) 리빙 브랜드 국내 진출

C Chic Wearable Device
세련된 웨어러블 **정보기술(IT)** 기기 확산

A Access From China
중국 소비자의 한국 온라인쇼핑 접근성 확대

L Luxury Sportism
고급 스니커즈 등 명품 속 캐주얼 열풍

E Elder Surfer
온라인·모바일 쇼핑에 적극적인 고령층 등장

자료: 롯데백화점

〈2015□소비트렌드〉

〈3C분석〉

- 고객
- 경쟁사
- 자사

〈전략조정 4C〉

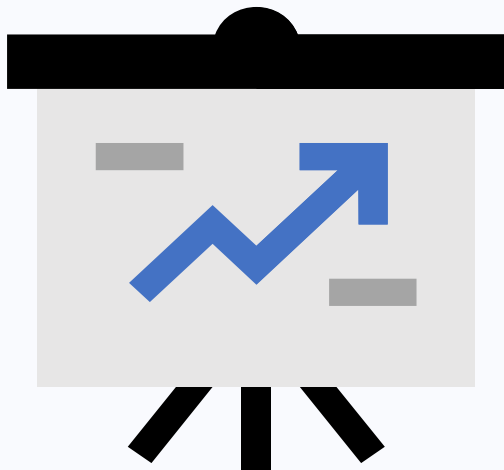
- 고객
(고객의 니즈에 맞게끔)
- 원가
(이익뿐만 아니라 고객의 입장에서 적절한 가격)
- 소통
(소셜네트워크 등 쌍방향 소통)
- 편의
(고객이 편히 제품을 받는것)

〈전략조정 4P〉

- 제품
(제품의 다양성, 품질, 디자인, 브랜드)
- 가격
(제품과 서비스에 맞는 가격)
- 판매
(광고, 할인행사, 이벤트)
- 유통(접근성 등)

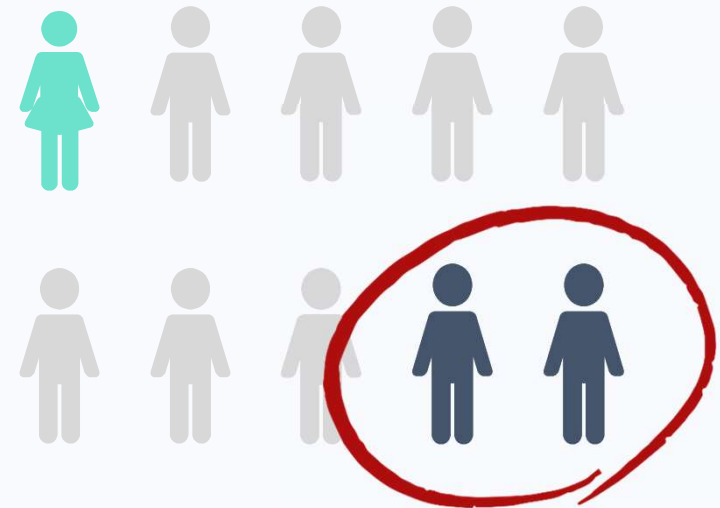
온라인 주문이 매출에 미치는 영향?

상품인지도나 경쟁사 이용이 매장에 미치는 영향?



'나를 위한 소비'를 타겟으로 하는 상품추천 서비스?

소액구매층을 타겟으로 하는 솔루션?





2. 데이터 처리 방안 및 활용 분석 기법

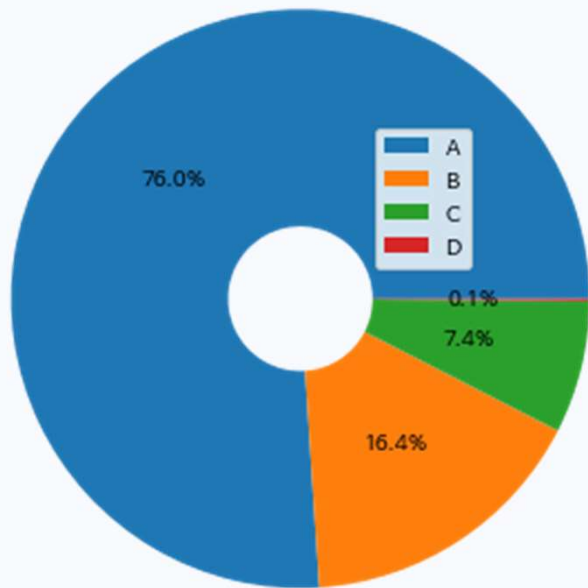


제공된 데이터

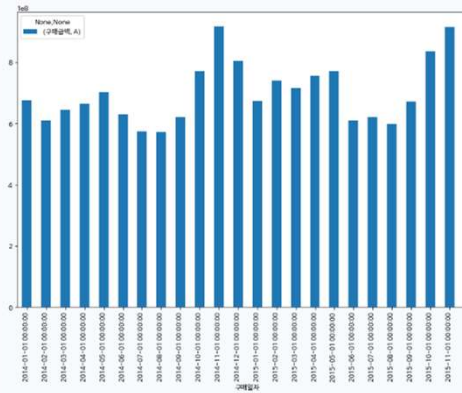


■ 구매상품 TR	□ 19383명.□28593030건의 구매 내역
■ 고객 DEMO	□ 19383명의 데이터
	□ '고객번호'.□'성별'.□'연령대'.□'거주지역'
	□ 거주지역은 178개 null□값 존재
■ 경쟁사이용	□ 경쟁사는 총 'A01'.□'A02'.□'B01'.□'B02'.□'C01'.□'C02'.□'C03'.□'D01'.□'D02' 총 9개 존재
	□ 5881명 /□19383명의 고객이 경쟁사 1회 ~□40회 이용
■ 채널이용	□ 6766명 /□19383명이 모바일 /□온라인을 이용한 적이 있음
	□ B_MOBILE□/□APP□5170명.□A_MOBILE□/□APP□1435명.□C_MOBILE□/□APP□46명.□D_MOBILE□/□APP□227명
	□ B_ONLINEMALL□1328명.□C_ONLINEMALL□618명
■ 멤버십여부	□ 하이마트 4418명.□다동이 2009명.□더영 642명.□롤스 387명
	□ 멤버십을 가지고 있는 회원 숫자 6498명 /□19383명 (약 1/3 정도)
	□ 그중 5601명이 한 개의 멤버십만 보유.□837명은 2개의 멤버십.□59명이 3개의 멤버십.□ 모든 멤버십을 보유한 사람은 1명

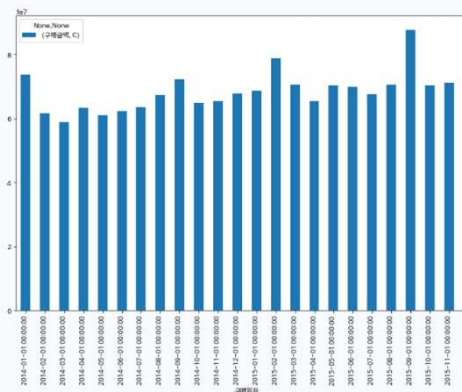
제휴사별 합계 구매금액



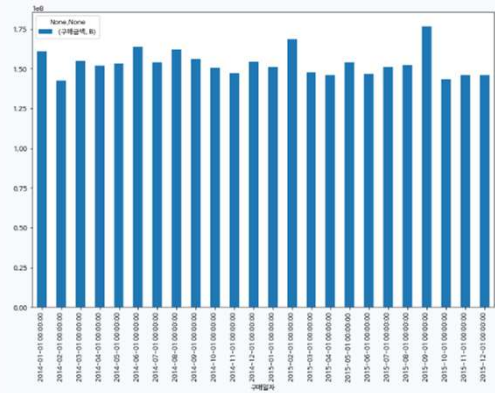
제휴사별 매출액 분포



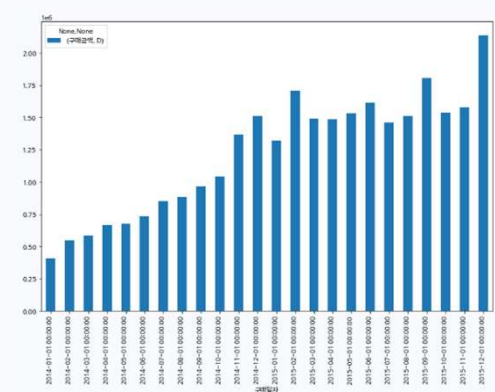
A□제휴사 매출액 시계열



C□제휴사 매출액 시계열

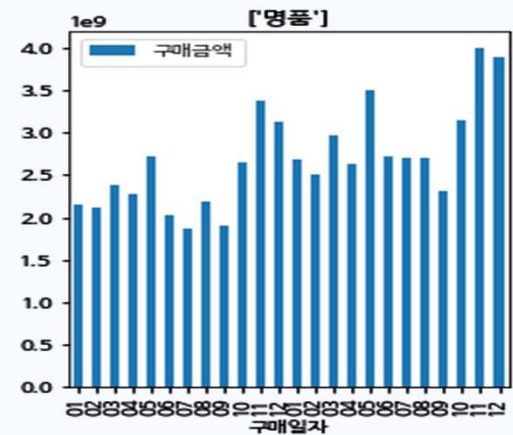
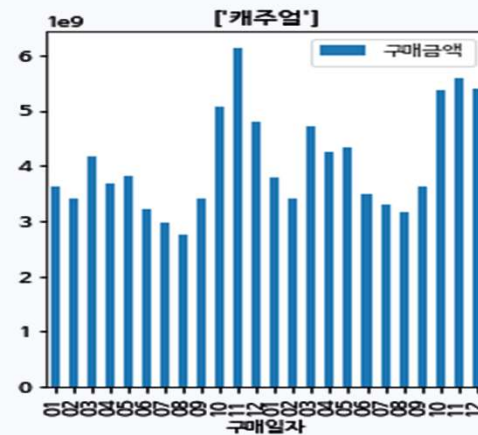
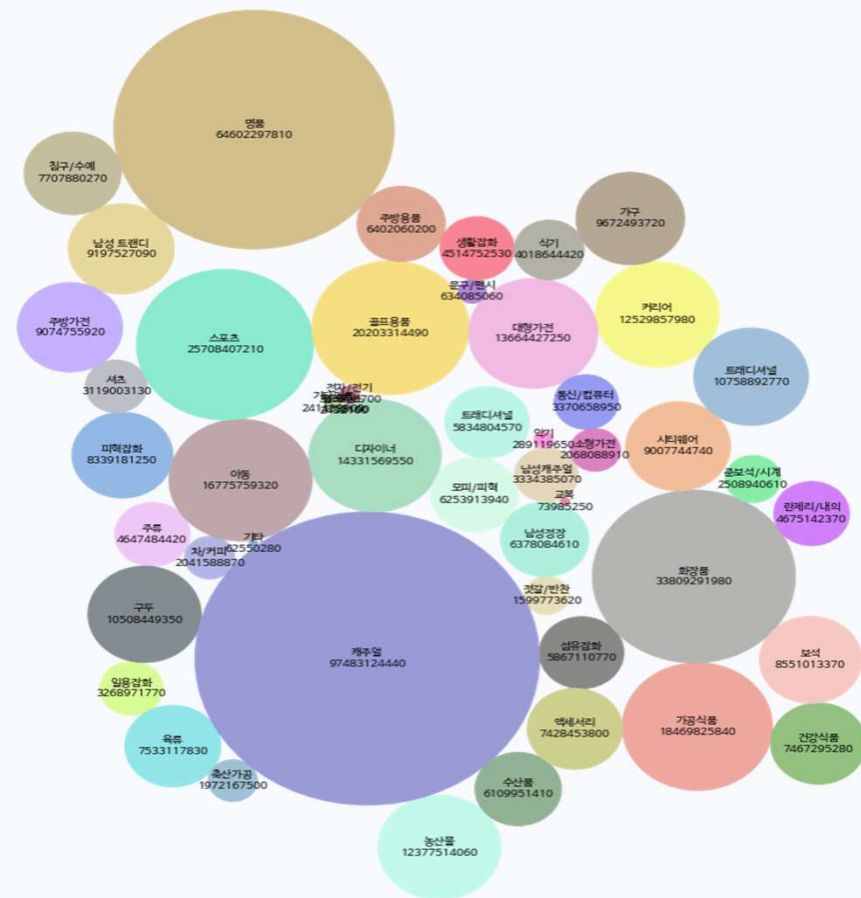


B□제휴사 매출액 시계열

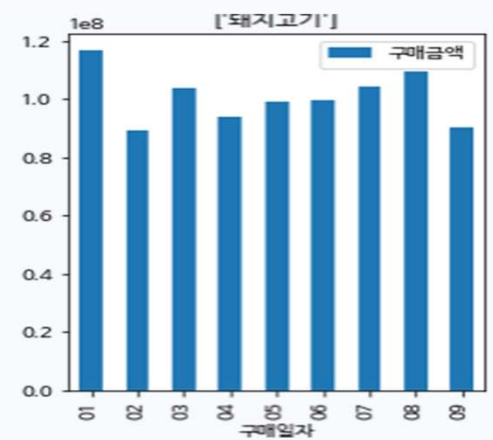
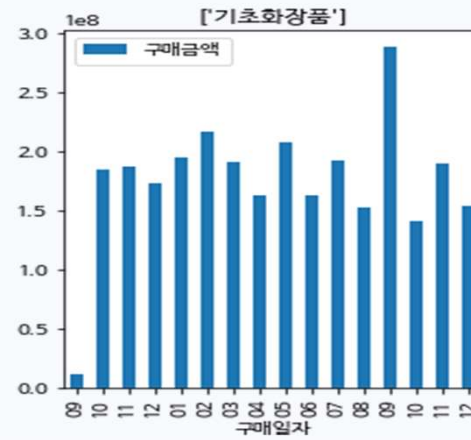
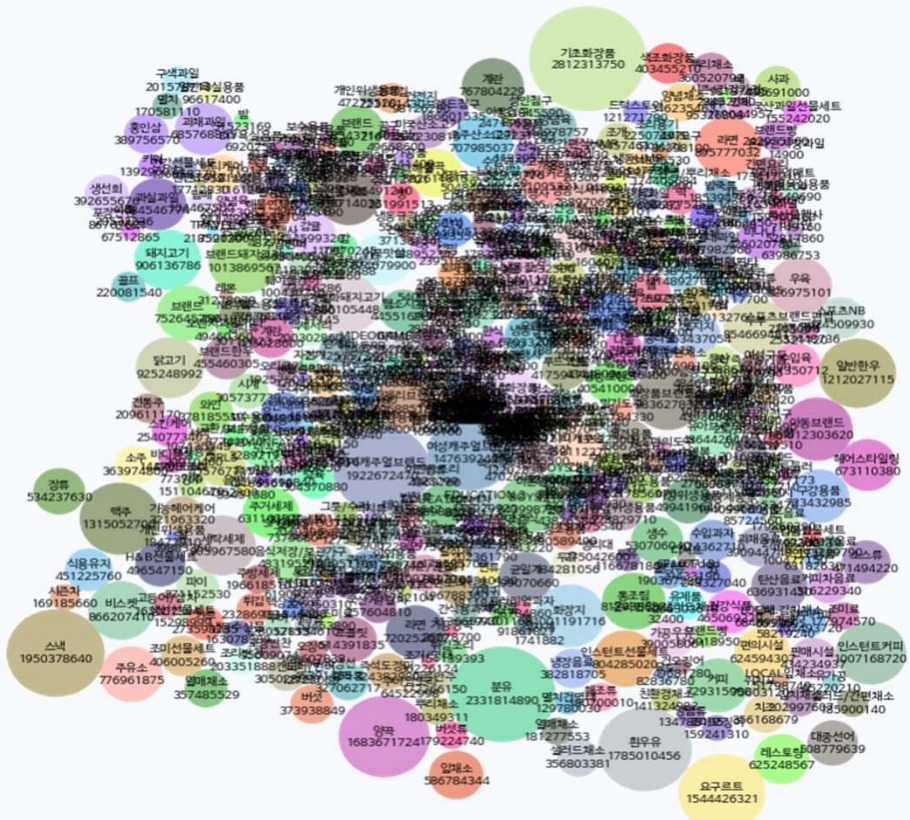


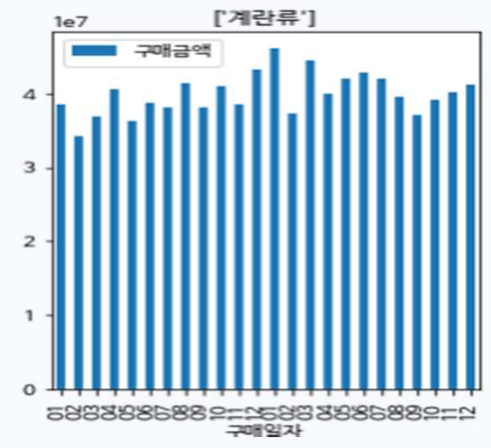
D□제휴사 매출액 시계열

A□제휴사 매출액 구성 & □분포

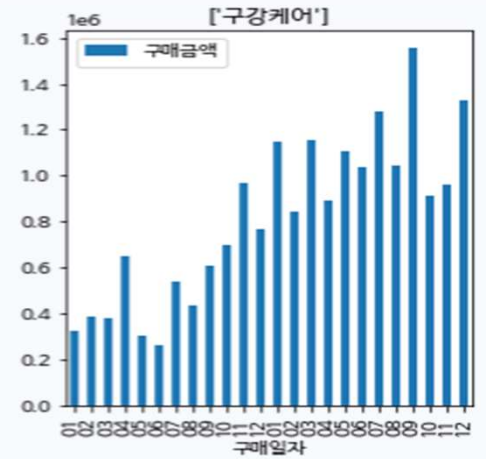
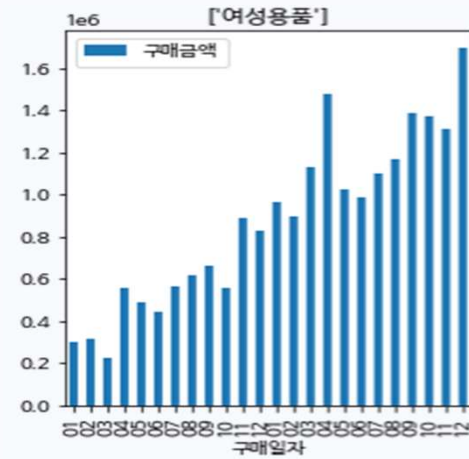
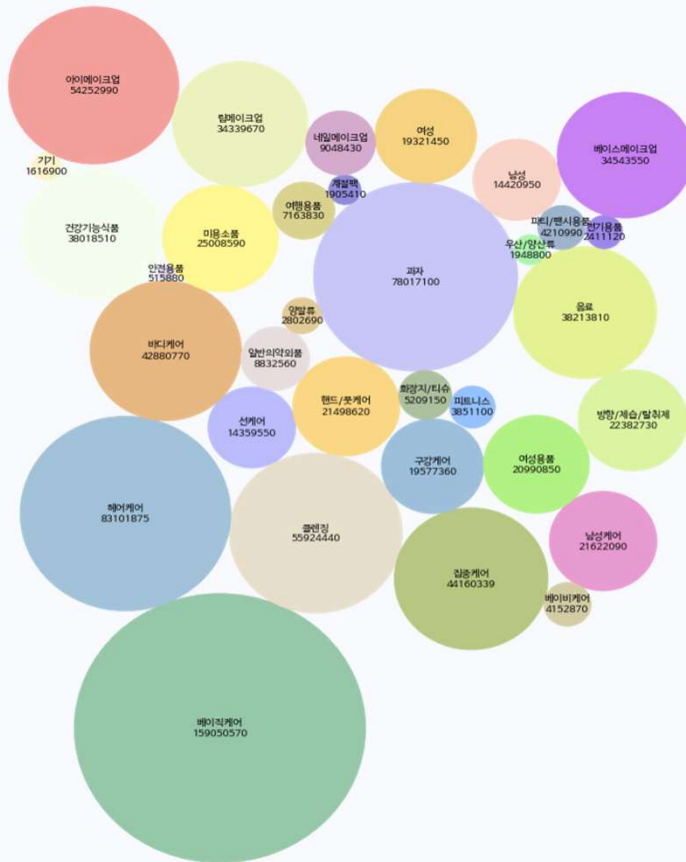


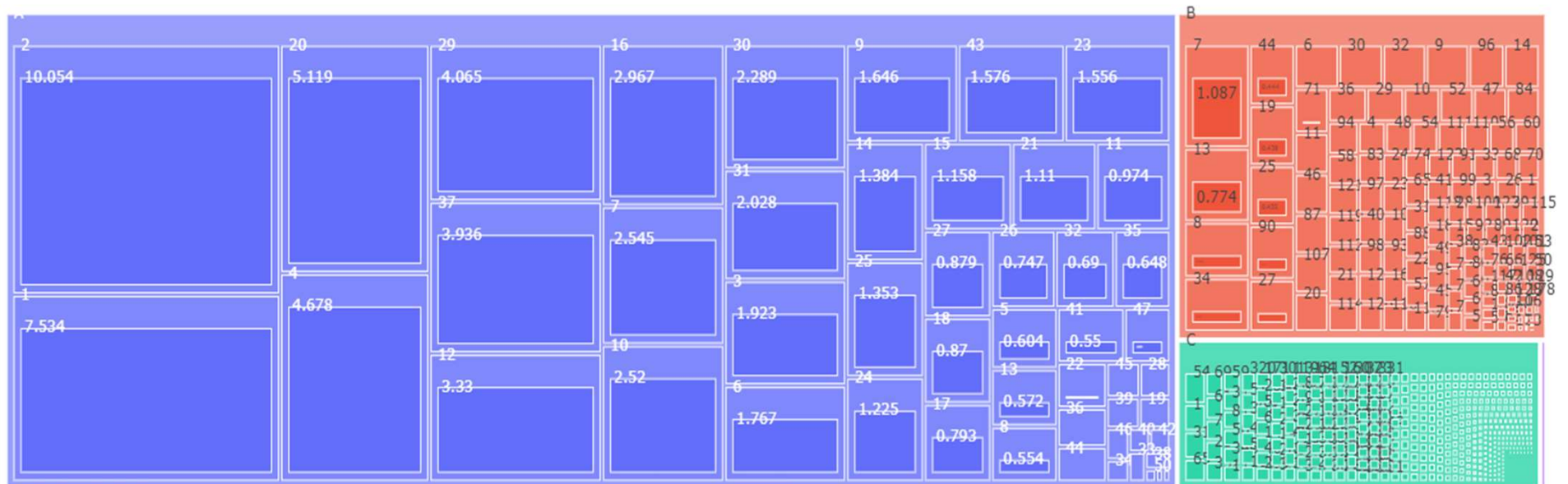
B2B제휴사 매출액 구성 & 분포





DQ제휴사 매출액 구성 & 분포





카테고리 재분류 & 피쳐 생성



<A 제휴사의 중분류 >

['일용잡화', '육류', '축산가공', '섬유잡화', '액세서리', '화장품', '보석', '준보석', '시티웨어', '란제리/내의', '블라우스', '기타의류', '남성 트렌디', '스포츠', '골프용품', '대형가전', '주방용품', '문구/팬시', '생활잡화', '식기', '가전']

<B 제휴사의 중분류 >

['담배', '기저귀', '화장지', '브랜드한우', '닭고기', '계란', '오리고기', '미국산소고기', ...]

<C 제휴사의 중분류 >

['담배', '화장지', '계란류', '수입과일', '건과일', '견과류', '양곡', '잡곡', '양채류', '나물류', '선어류', ...]

<D 제휴사의 중분류 >

['여성용품', '화장지/티슈', '집중케어', '남성케어', '베이비케어', '헤어케어', ...]

중분류로
통합

분류	고급/일반	내구/준내구/ 비내구	편의/선매/전문
N냉동류	일반	비내구	편의
N냉장류	일반	비내구	편의
N도서/음반	고급	비내구	편의
N스낵류	일반	비내구	편의
N음료류	일반	비내구	편의
N자동차용품	일반	비내구	편의
N푸드코트	일반	비내구	편의
가공식품	일반	비내구	편의
가구	고급	내구	선매
건강식품	고급	비내구	전문
골프용품	고급	내구	선매
교복	고급	준내구	선매

등급을 반영한 증감지수 피쳐 생성



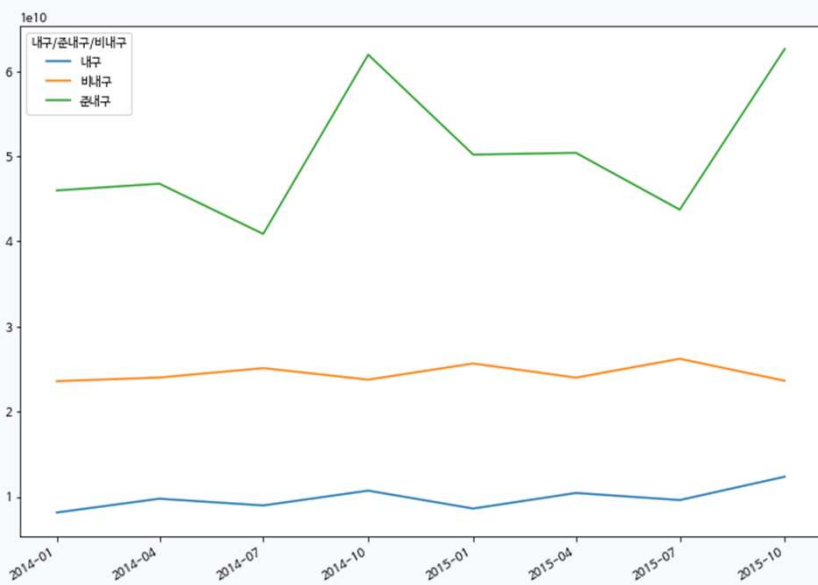
```
def quan_grade(mom_df, quan_start, quan_end, col_name):  
  
    df = mom_df[mom_df['구매일자'].between(quan_start, quan_end)]  
    # groupby  
    df = df.groupby(['고객번호'], as_index=False)[['구매금액']].sum()  
    # pd.cut  
    p_li = [0.04, 0.11, 0.23, 0.40, 0.60, 0.77, 0.89, 0.96] # 퍼센트  
    bins = [0] + [df['구매금액'].quantile(q=p_li[i]) for i in range(8)] +  
            [df['구매금액'].quantile(1)] # 등급별 기준값  
    labels = [1, 2, 3, 4, 5, 6, 7, 8, 9] # 등급  
    df[col_name] = pd.cut(df['구매금액'], bins=bins, labels=labels)  
    df[col_name] = pd.to_numeric(df[col_name])  
    # unvaluable column out  
    df.drop('구매금액', axis=1, inplace=True)  
  
    return df
```

등급 반영

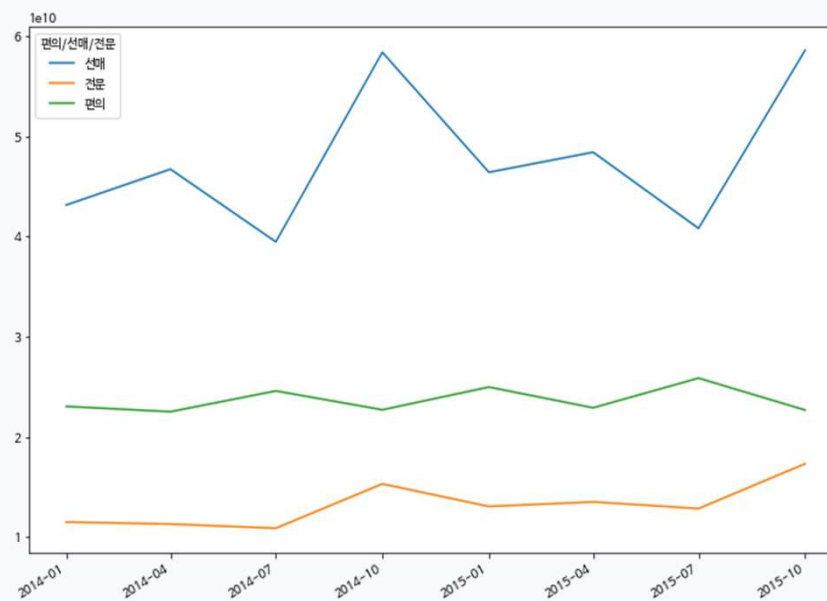
〈증감지수 피쳐
생성〉

고객번호	fluctuation	총구매금액
1	1	71840032
2	0	74036940
3	0	2486049
4	-1	12956553
5	4	10620810
6	-1	17784434
7	0	55434791
8	0	44352982
9	-1	162118481

분류별 매출액

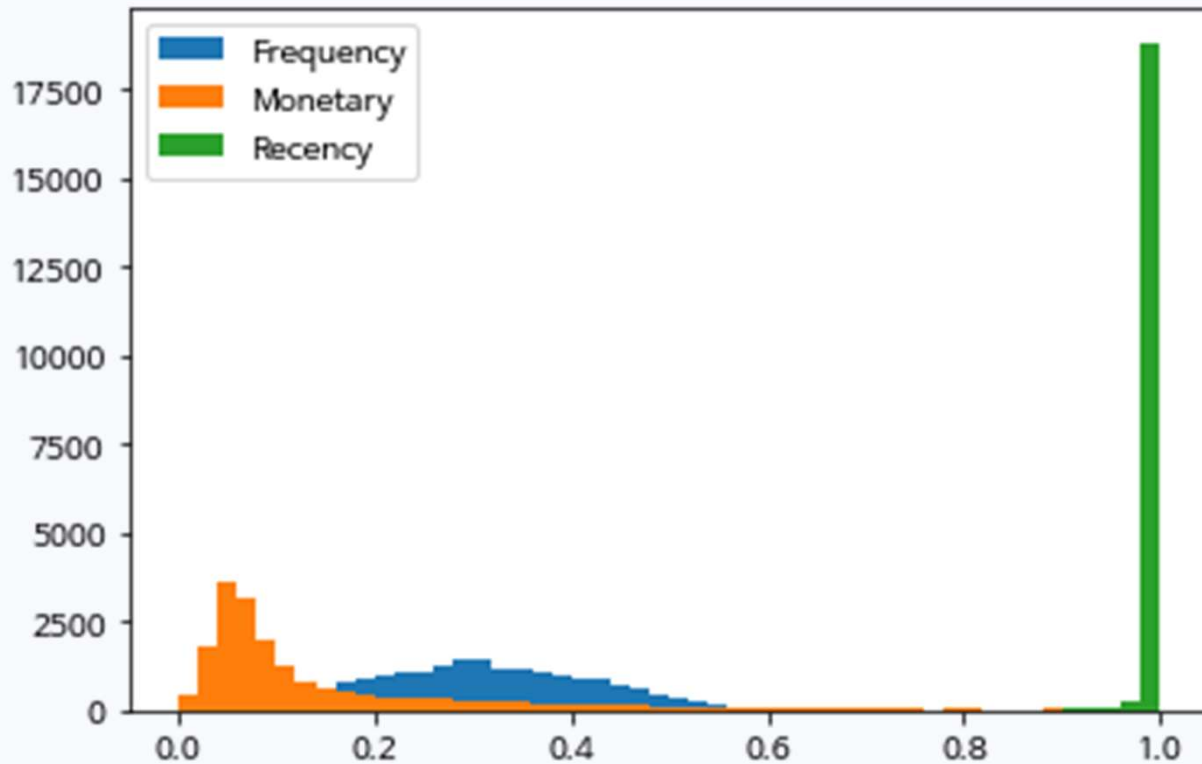


준내구, □선매 제품



일반 상품, □편의, □비내구
제품





- Recency □ □ □

고객이 얼마나 최근에 구입했는가?

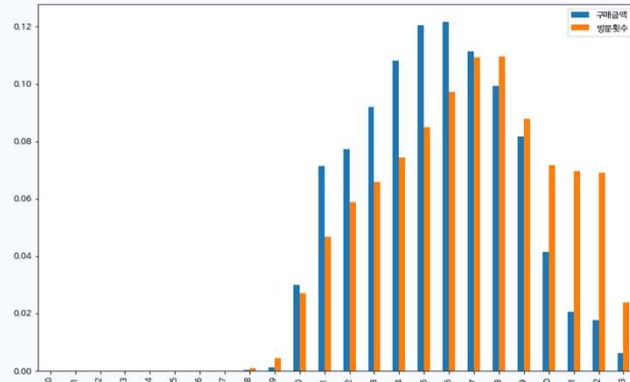
- Frequency □ □

고객이 얼마나 빈번하게 우리 상품을 구매했는가?

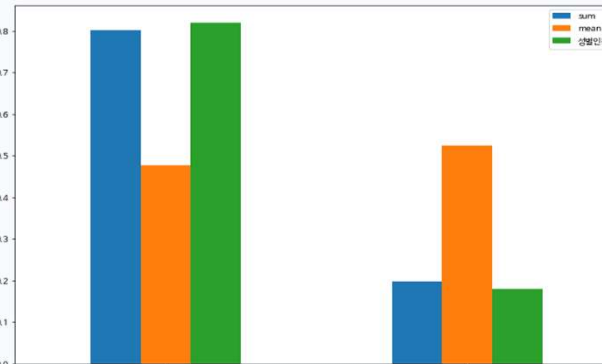
- Monetary

고객이 구입했던 총 금액은 어느정도인가?

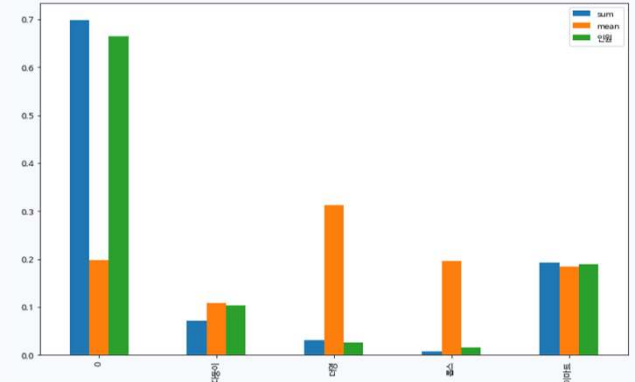
기타 특성들



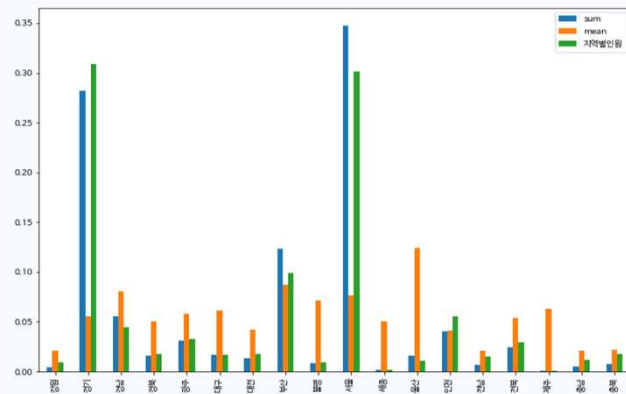
〈구미금액별〉



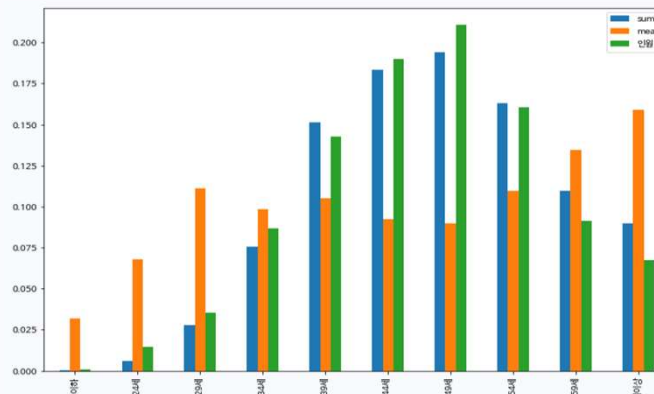
〈구성별〉



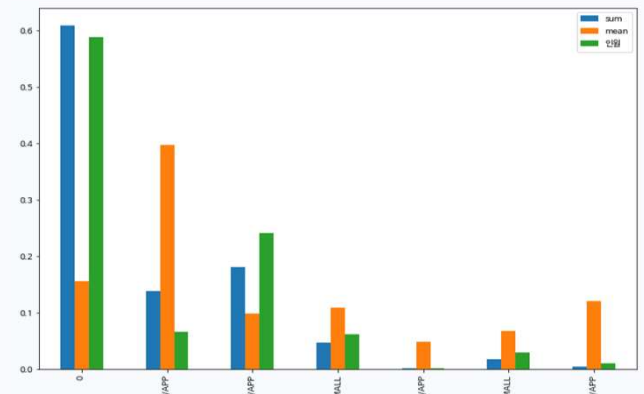
〈구멤버십〉



〈구지역별〉



〈구연령대별〉



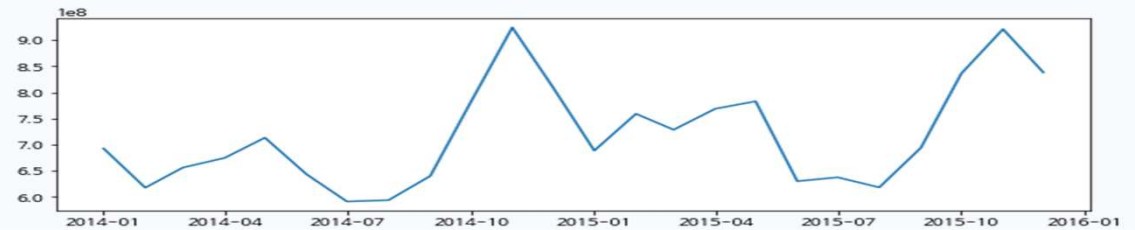
〈구경쟁사별〉

전체 매출액은 증가중인데...

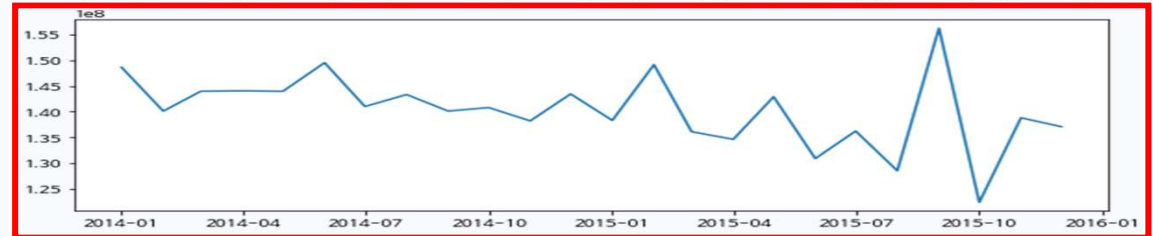


```
A      8933
B      6604
C      3663
D       183
Name: 구매금액이큰제휴사
```

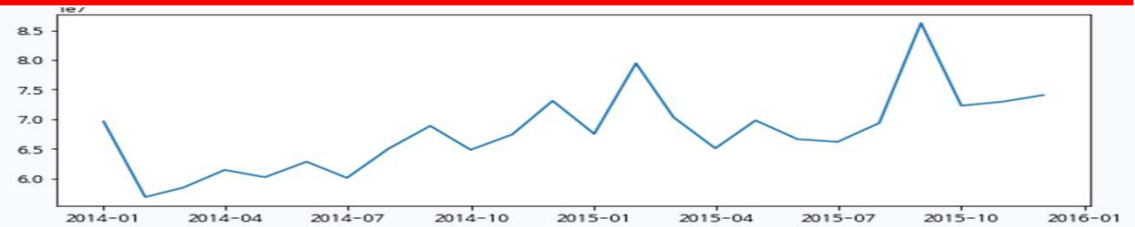
A
제휴사



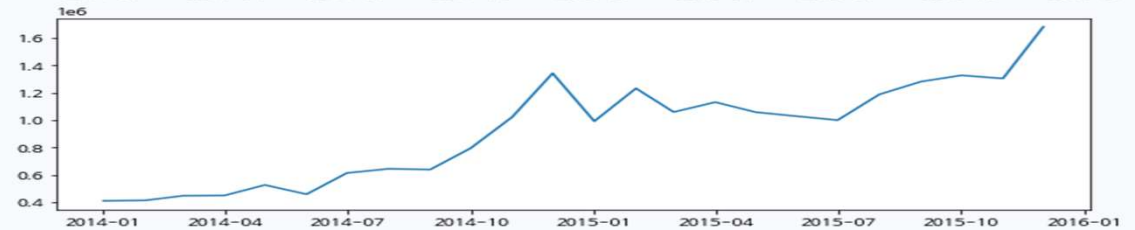
B
제휴사



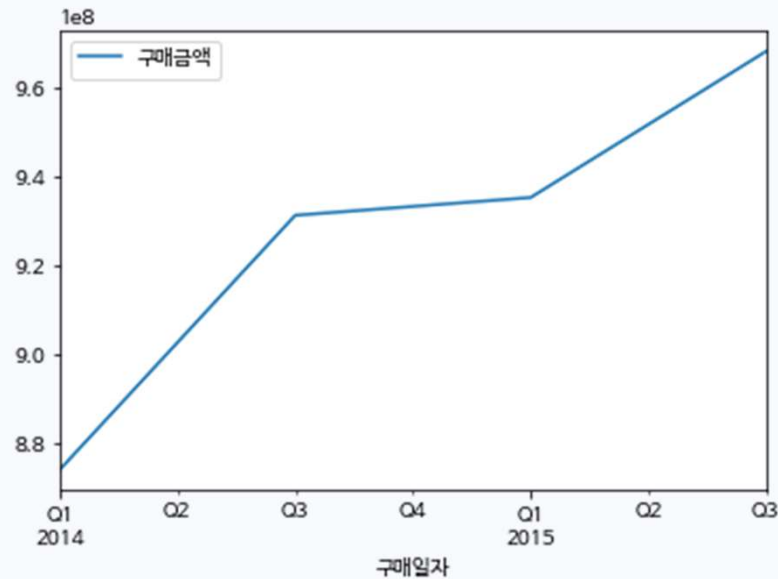
C
제휴사



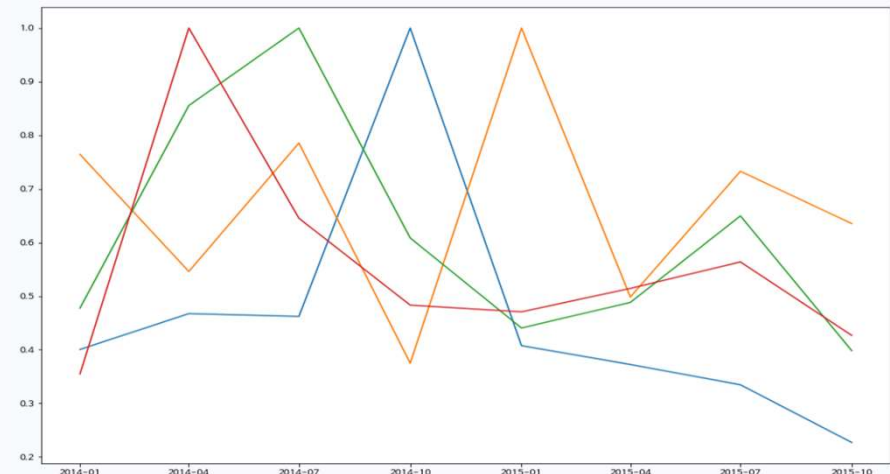
D
제휴사



전체 매출액은 증가중인데...



〈□전체 합계 매출액〉



〈□Random한 일부 고객 매출액〉

전체 합계 매출액은 증가 추세를 보이지만,
오른쪽 시각화 자료인 고객별 매출액을 보면 이탈고객이 적지
않다!



3. □예측 모델 개발

〈각 피쳐별 등급
생성〉

9등급

8등급

7등급

6등급

5등급

4등급

3등급

2등급

1등급

〈분기 마다 변동
반영〉

중 분류명 변동 지수

제휴사별 변동지수

상품 속성별 변동지수

〈고객 속성 정보〉

- 신규고객
- 명품 구매자
- 성별, 연령대 등

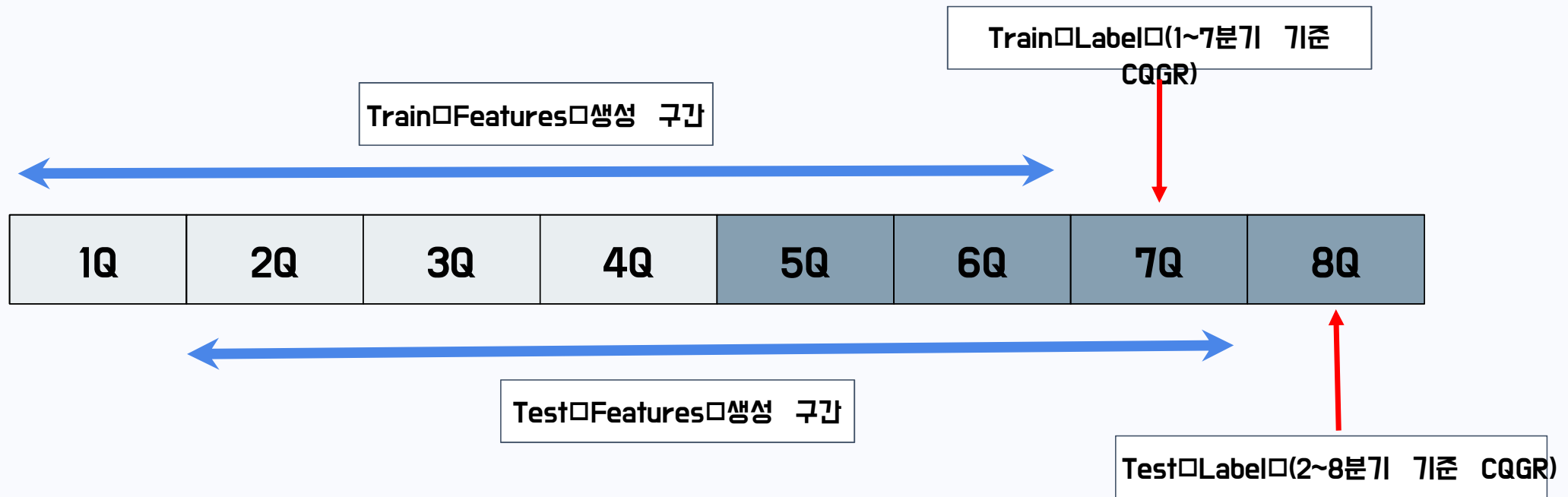
〈구매 패턴〉

- 제휴사 별 지출
- 제휴사 방문횟수
- 제휴사별 빈도
- 돈 비율
- 자주방문하는 제휴사
- 경쟁사 이용횟수 등

〈상품 속성 정보〉

- 내구, 비내구
- 편의, 선택, 전문
- 중분류별 변동지수
- 일반제품 반기 매출증가
- 일반제품 전분기 매출증가
- 일반 품 직전달 매출증가 등

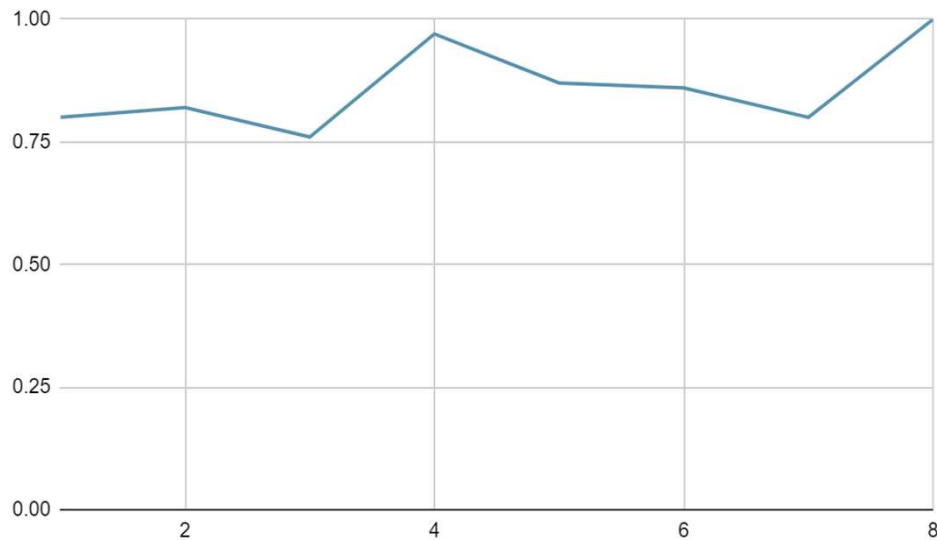
Train & Test



기존 라벨



Points scored



기존라벨 - 각 분기의 추세를 반영

〈모델〉	〈Test정확도〉
Lightgbm	0.516
Xgboost	0.506
DecisionTree	0.511
RandomForest	0.511
LogisticRegression	0.444
SVM	0.509

The diagram shows the CAGR formula:
$$\text{CAGR Formula} = \left[\left(\frac{\text{Ending Value}}{\text{Beginning Value}} \right)^{\frac{1}{\text{No. of years}}} - 1 \right] \times 100\%$$
 A callout box labeled 'Mean' points to the fraction $\frac{\text{Ending Value}}{\text{Beginning Value}}$. Another callout box labeled 'Mean' points to the exponent $\frac{1}{\text{No. of years}}$. There are also icons of a bar chart and a calculator.

〈CAGR〉-복합 연간 성장률

- C: Compound
- A: ANNUAL
- G: GROWTH
- R: RATE

CAGR은 기존 같은 년도로 매출액 합계로 비교하지만, 각 분기는 기간이 다르기에 평균으로 계산

〈전체 피쳐〉

- 내구,비내구
- 편의,선매,전문
- 중분류별 변동지수
- 일반제품 반기 매출증가
- 일반제품 전분기 매출증가
- 일반 품 직전달 매출증가 등 128개

〈모델〉

Lightgbm

Xgboost

DecisionTree

RandomForest

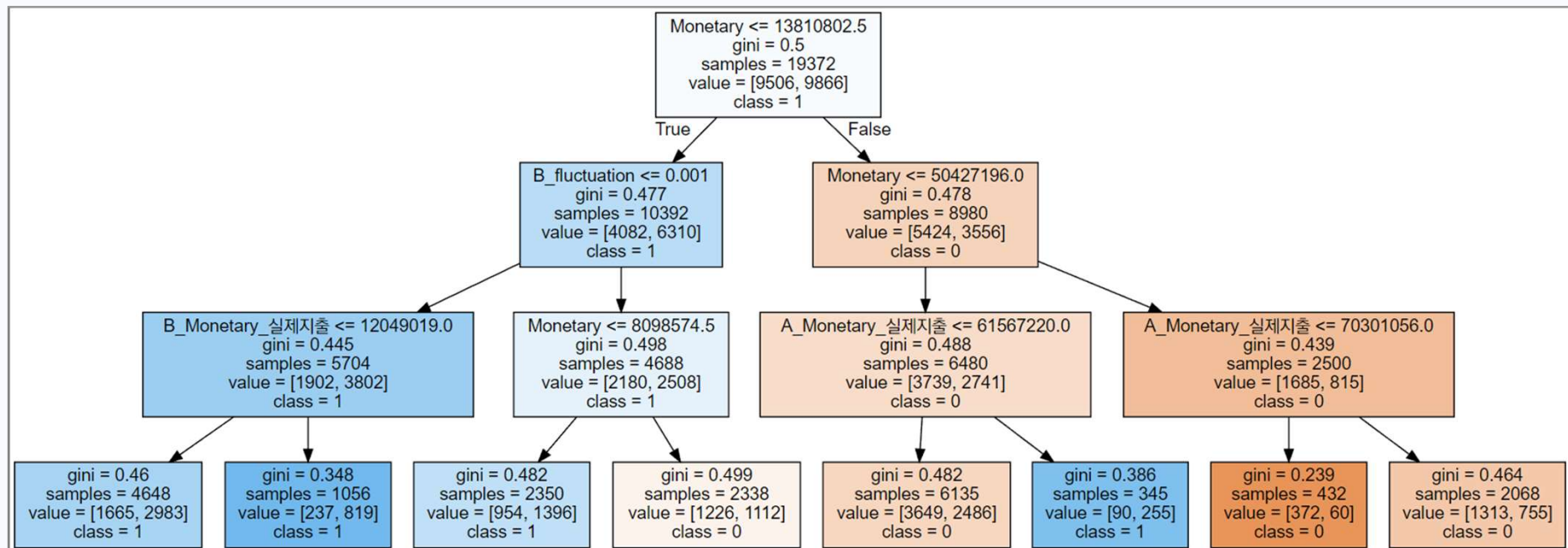
LogisticRegression

SVM

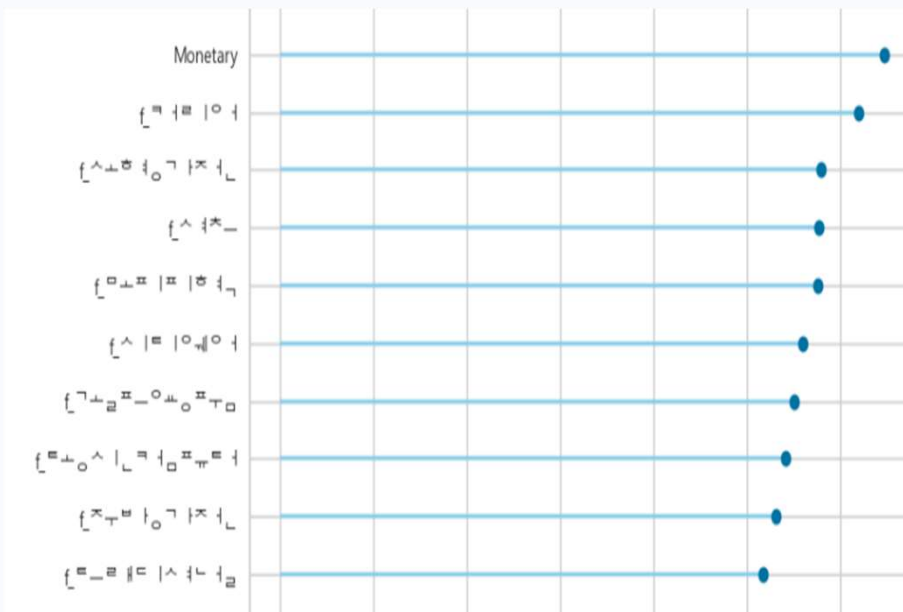


전체 피쳐 Bestmodel
정확도 67%

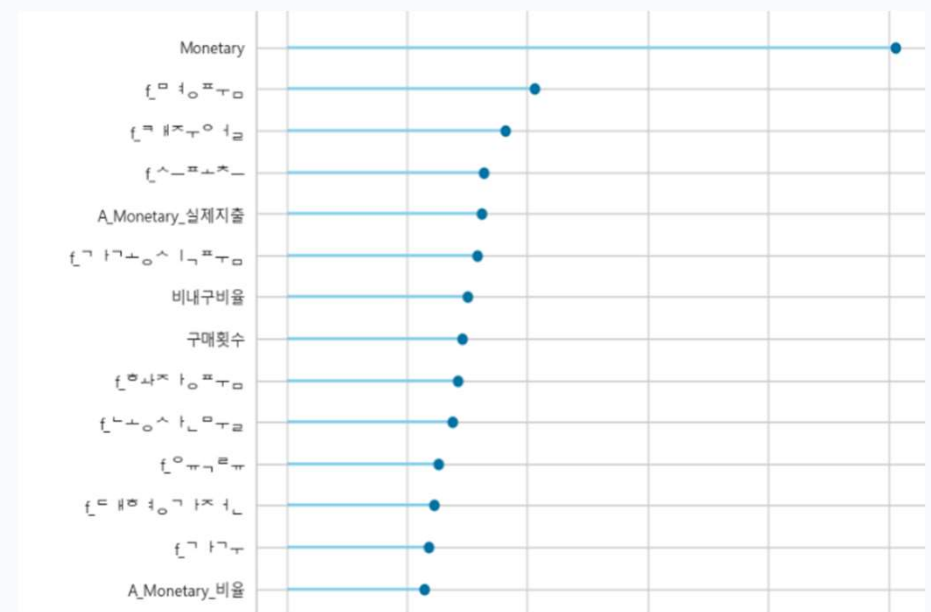




Feature importance - Pycaret



〈SVM〉



〈Lightgbm〉

〈□Feature□Select□〉			
Fluctuation	고급_ fluctuation	A_Monetary_ 실제지출	A_fluctuation
Monetary	편의_ fluctuation	B_Monetary_ 실제지출	B_fluctuation
CAGR	선매_ fluctuation	C_Monetary_ 실제지출	C_fluctuation
일반_ fluctuation	전문_ fluctuation	D_Monetary_ 실제지출	D_fluctuation

〈□Feature□Name□〉	〈□Feature□ Explanation□〉
고급.□일반.□편의.□ 선매.□전문 지수	고객의 상품 속성 구매 변동지수
A.□B.□C.□D□실제 지출	고객의 제휴사별 구매액
A.□B.□C.□D□변동 지수	고객의 제휴사별 구매액 변동 지수
CAGR	연평균 변화율
Monetary	RFM

모델 선정 및 정확도 개선



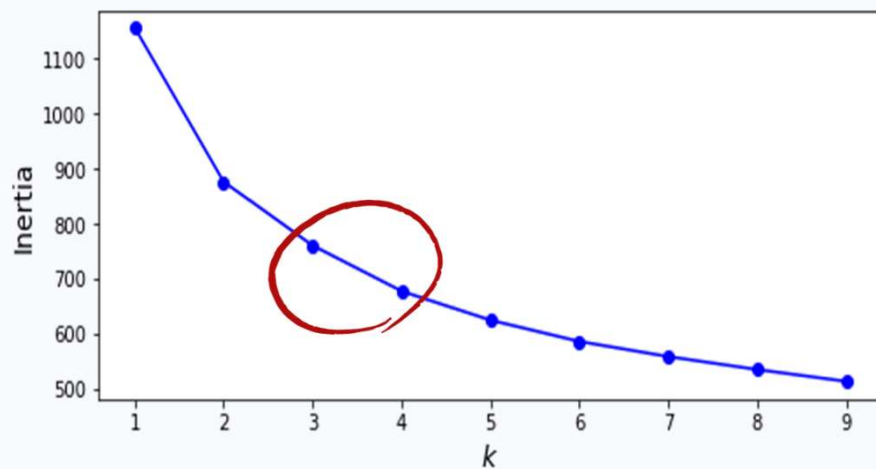
〈모델〉	〈GridSearch파라미터 조정〉	〈Test 정확도〉	〈Test 정밀도〉	〈Test 재현율〉	〈T-F1score〉
Lightgbm	<ul style="list-style-type: none"> ○ max_depth 10 ○ min_sample_leaf 8 ○ n_estimators = 300 	0.727 Best Model	0.646	0.767	0.701
Xgboost	<ul style="list-style-type: none"> ○ max_depth 10 ○ min_sample_leaf 8 ○ n_estimators = 300 	0.723	0.642	0.763	0.697
Decision Tree	<ul style="list-style-type: none"> ○ max_depth 10 ○ min_sample_leaf 8 ○ n_estimators = 300 	0.701	0.614	0.756	0.678
Random Forest	<ul style="list-style-type: none"> ○ max_depth 10 ○ min_sample_leaf 8 ○ n_estimators = 300 	0.703	0.597	0.763	0.663
Logistic Regression	<ul style="list-style-type: none"> ○ c = 0.01 ○ penalty L2 	0.444	0.470	0.797	0.592
SVM	<ul style="list-style-type: none"> ○ loss = squared_hinge ○ alpha 0.01 ○ penalty none 	0.626	0.539	0.731	0.620

〈모델 Lightgbm〉	〈Test 정확도 〉
Scale	0.725
PCA	0.656
LDA	0.724
SVD	0.718
Testset 0.7 : Testval 0.3	0.724

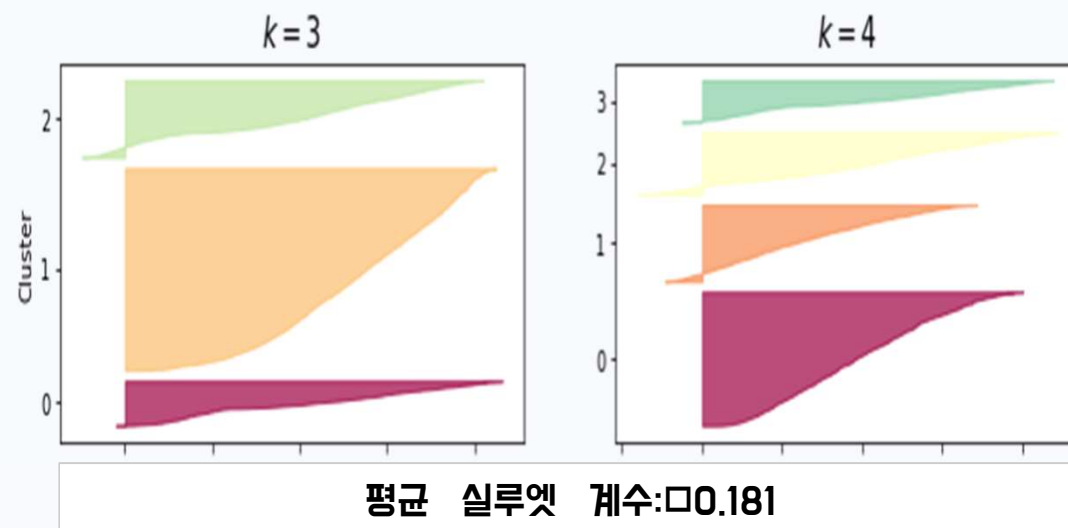


**구매 감소 고객을 대상으로
군집화**

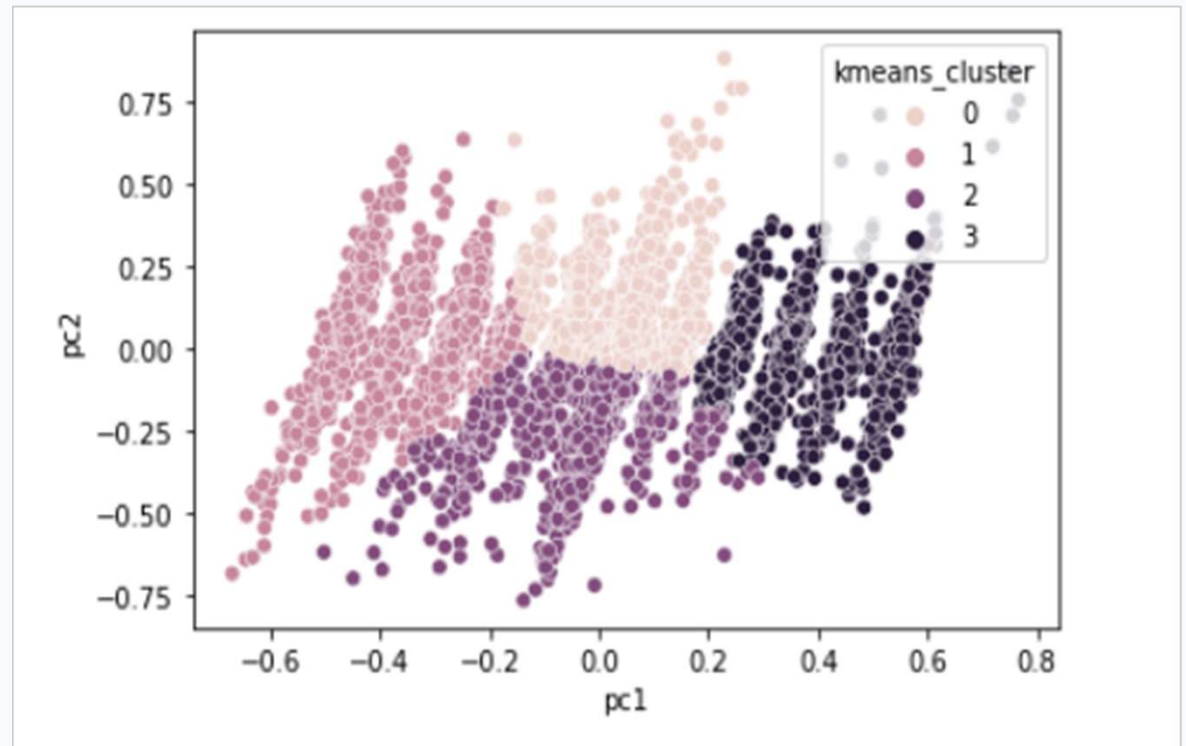
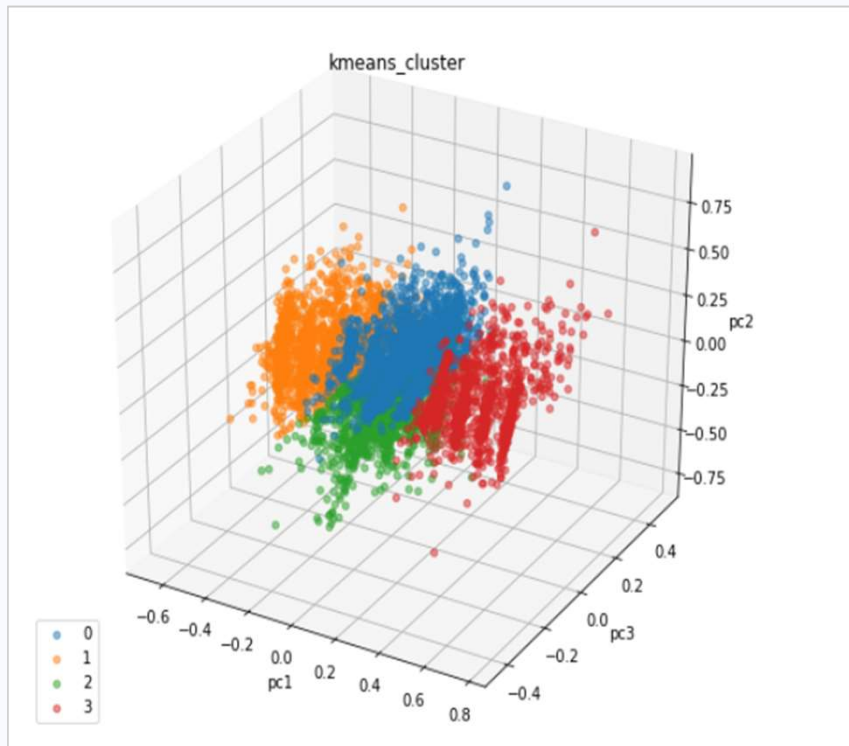
적정 군집 수 찾기



〈kmeans
elbow〉



〈kmeans
silhouette
diagram〉



<div>증감지수</div> <div>Cluster</div>	전체 증감지수	A제휴사 증감지수	B제휴사 증감지수	C제휴사 증감지수	D제휴사 증감지수	고급상품 증감지수	일반상품 증감지수	편의상품 증감지수	선매상품 증감지수	전문상품 증감지수
0군집(1975명)	0	0	-1	0	0	0.461267	-1.267848	-1.207089	0.464305	-0.618671
1군집(1821명)	1	0	0	0	0	0.934102	0.512356	0.426689	0.635365	3.340472
2군집(1290명)	2	0	0	0	0	2.075194	3.501550	3.359690	1.853488	1.139649
3군집(4671명)	0	0	0	0	0	0.384715	1.152003	1.126740	0.381503	-0.569739

0군집 분석

A군제휴사

	평균구매액
고급	175727
일반	24257
내구	271152
준내구	30032
비내구	191501
선매	186855
전문	685824
편의	18500

B군제휴사

	평균구매액
고급	15986
일반	6747
내구	17156
준내구	7125
비내구	14256
선매	21752
전문	40026
편의	7097

3군집 분석

A군제휴사

	평균구매액
고급	150033
일반	21327
내구	184319
준내구	26470
비내구	173968
선매	168650
전문	557905
편의	16959

C군제휴사

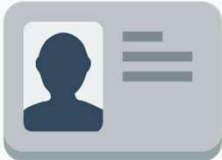
	평균구매액
고급	6523
일반	4919
내구	8944
준내구	4741
비내구	7432
선매	8619
전문	20197
편의	5121



4.0솔루션



0군집 솔루션 (잠재 요인 협업필터링)

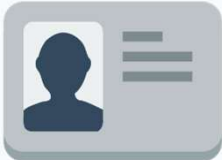


고객번호	10577	구매 제품 수	333
전체 제품 수	39	비 구매 제품 수	32

A군제휴사 전문 상품 Top-10추천 제품과 예상평점		
1	디자이너부틱	5.038579735822618
2	건강식품	4.443690415278115
3	수입인류	3.484216733535685
4	명품잡화	3.2995447581494077
5	global오피집샵	3.0287003241841215
6	디자이너	2.9428955027928247
7	남성해외패션	2.7807699838145115
8	주얼리멀티	2.707262660782355
9	디자이너캐릭터	2.706115237536897
10	엘리튼	2.5777618711749373

지출증가 예상치	
추천제품 (평균구매액)	예상지출증가율
디자이너부틱 (671238.5039329861)	8.827683351762559
건강식품 (109281.85613286949)	1.4372024494741424
수입인류 (909098.6742925948)	11.955862461911664
명품잡화 (1287546.8157574623)	16.932961270070095
global오피집샵 (67323.55917771561)	0.8853947725767597
디자이너 (288174.9502789231)	3.7898857054037904
남성해외패션 (933778.5038167939)	12.280435201613855
주얼리멀티 (169818.181818182)	2.2333360313499213
디자이너캐릭터 (127786.666666666)	1.6805654373235797
엘리튼 (1135744.3203883495)	14.936555591203188

3군집 솔루션 (잠재 요인 협업필터링)



고객번호	8066	구매 제품 수	462
전체 제품 수	39	비 구매 제품 수	38

A□제휴사 전문 상품 Top-10□추천 제품과 예상평점

1	건강식품(홍삼)	5.038579735822618
2	디자이너부틱	4.443690415278115
3	건강식품	4.306516019222889
4	건강식품(비타민)	3.8641965800942626
5	수입인류	3.484216733535685
6	명품잡화	3.2995447581494077
7	어덜트	3.182746855347172
8	보석	3.064520036890415
9	global□편집샵	3.0287003241841215
10	주얼리멀티	2.9428955027928247

지출증가 예상치

추천제품 (평균구매액)	예상지출증가율
건강식품(홍삼)□(273813.0482442559)	32.21807564936165
디자이너부틱 (671238.5039329861)	78.98094352021428
건강식품 (109281.85613286949)	12.858595054428008
건강식품(비타민)□(73202.81793132)	8.613373033099819
수입인류 (909098.6742925948)	106.96864173896304
명품잡화 (1287546.8157574623)	151.49855340408843
어덜트 (783362.0239808153)	92.17390154079493
보석 (1629220.8028357923)	191.70145254894163
global□편집샵 (67323.55917771561)	7.9215929864562895
주얼리멀티 (169818.181818182)	19.98157159981149

0군집 솔루션 (잠재 요인 협업필터링)



고객번호	16370	구매 제품 수	343
전체 제품 수	1514	비 구매 제품 수	1403

B군제휴사 편의 상품 Top-10추천 제품과 예상평점

1	기능성우유	6.798457343866663
2	부침두부	4.991654826383558
3	마시는요구르트	4.875984406563062
4	일반원우유	4.836373327146546
5	국산콩나물	4.602245253879025
6	보네스빠	4.556736613086273
7	봉지면	4.50255856342054
8	NB돼지고기	4.467240249370265
9	일반닭	4.4467971909224016
10	기능성시유	4.431674554697352

지출증가 예상치

추천제품 (평균구매액)	예상지출증가율
기능성우유 (4081.9152959190774)	0.05301388358678594
부침두부 (2996.5155045944443)	0.038917251488641114
마시는요구르트 (5743.746041897393)	0.07459691393438561
일반원우유 (4001.0065905869064)	0.0519630816041996
국산콩나물 (1895.8714306923343)	0.024622634238085084
보네스빠 (10356.818181818182)	0.13450919816224882
봉지면 (8125.781803387213)	0.1055336084526175
NB돼지고기 (14210.848023430799)	0.18456341892655814
일반닭 (7641.727351194586)	0.09924695022532148
기능성시유 (4252.428125303388)	0.05522841932102551

0군집 솔루션 (사용자 기반 협업필터링)



A제휴사 전문상품 지출증가 예상치

추천상품 (평균구매금액)	예상지출증가율
시계 (377397.80535499397)	15.611658980272011
숙녀고정행사 (77651.14293063534)	3.2121627250087954
건강식품(비타민)□(73202.817931326)	3.0281507039073574
명품잡화 (1287546.8157574623)	53.26141679555565
어덜트 (783362.0239808153)	32.40501296763128
global□편집샵 (67323.55917771561)	2.78494583780639
조화 (83669.56521739131)	3.461124311448671
숙녀단기행사 (108317.84615384616)	4.480739558198479
엘리튼 (1135744.3203883495)	46.98186573185143
악기 (827583.3333333334)	34.23429758846589

B제휴사 일반/편의상품 지출증가 예상치

추천상품 (평균구매금액)	예상지출증가율
생굴 (5559.295406879195)	0.31255385529391333
치즈 (6973.867613616485)	0.39208371734447667
대파 (2296.386928104575)	0.12910711431840127
쌈장 (2778.770891555702)	0.1562276316634552
숙주나물 (2109.1684107505835)	0.11858134349695604
하드캔디 (4817.727272727273)	0.27086152518216977
과일/야채음료 (4241.783329606083)	0.23848089298317632
팬이버섯 (1255.8820508031379)	0.07060800839747117
복숭아 (9698.680130597015)	0.5452777095330248
오렌지 (6621.671206896552)	0.3722825848835498

A제휴사 전문상품 지출증가 예상치

추천상품 (평균구매금액)	예상지출증가율
건강식품(홍삼)□(273813.0482442559)	3.379740399973528
시계 (377397.80535499397)	4.658312004480521
직수입 (281399.6884375721)	3.4733841270560504
global□편집샵 (67323.55917771561)	0.830990905225093
컨템포러리 (1006213.046875)	12.419929975251804
잡화멀티샵 (245536.61921708184)	3.030717626358765
피아노/악기 (881778.8778877888)	10.88400905855373
주얼리멀티 (169818.181818182)	2.0961067174160917
음반 (20996.464646464647)	0.2591644199474672
명품잡화행사 (479686.41509433964)	5.920884949816568

5팀 작업로그

파일 수정 보기 삽입 서식 데이터 도구 확장 프로그램 도움말 [몇 초 전에 마지막으로 수정했습니다.](#)

A	K	L	M	N	O	P	Q	R	S
1	5팀 작업								
2									
3	12/27 (화)	12/28 (수)	12/29 (목)	12/30 (금)		1/2 (월)	1/3 (화)	1/4 (수)	1/5 (목)
4	일정	- 문제 분석 (근거, 대안 지) - 문제 분석	- 피드백 - 변동변수 피처 개발	- 변동변수 피처 개발 - 모델 공부		- 피처 정제 (변동변수) - 모델 공부	- 피처 정제, 목록 정리 - 데이터셋 정리작업 - 모델 공부	- 모델 계획 수립 - 피드백	- 피처 정제 및 모
5	공부	- 새로운 피처 생성 회의 - 새로운 피처 생성	- 새로운 피처 생성	- 새로운 피처 생성		- 생성된 피처 분기별로 나누기		- ARIMA 설명 문서화	- 상위 모델 선정

5팀 회의록

파일

수정

보기

삽입

서식

데이터

도구

확장 프로그램

도움말

어제 오후 2:07에 마지막으로 수정했습니다.

100%

기본값 (Arial)

18

B I U A

田

33

≡

↓

↶

↷

↺

↻

↵

Σ

▼

A1:B1	5팀 회의록			
	A	B	D	E
1	5팀 회의록			
2	날짜	안건	결정사항	세부내용
35	1/3 (화)	1. 피쳐 정제 작업과 공유	<ul style="list-style-type: none"> - 최종적으로 피쳐 정보는 담당자가 스프레드시트 [가쓰] - 은아: 중분류 피쳐작업, 레이블값, 최종 데이터프 - 카쓰: 정제: 파이캐럿 공부, 피쳐 정제 마무리 - 공사: 모델 후보 3가지 시도. 	<ul style="list-style-type: none"> - test 데이터셋 total 에서 고객번호 5 정보 다시 처리. fillna(0) 사용해보기. [은아] - 고급/일반, 선행/연의 도 구매등급과 증감지수 뽑아내서 피쳐 만들자. - 피쳐 수정하였음
36		2. 모델 개발 회의		
37		3. 레이블값		
38	1/4	멘토링	<ul style="list-style-type: none"> 1. Pycaret 결과를 바탕으로 세 가지 모델 후보 뽑 2. feature importance 확인 후 피쳐 정제 3. 2~7분기 자료의 data와 target을 val 7 : test 3 4. validation data로 fine tuning. 	<ul style="list-style-type: none"> - Pycaret 결과 보는데 test 로 넣으니 성능이 확 떨어짐. - 성능 저하의 이유를 고민하다가, 혹시 데이터가 가진 계절성의 문제인지 강사님에게 질문. -> A. 전체 매출 관련해서는 계절성이 고려되지
39	1/5 (목)	<ul style="list-style-type: none"> 1. Pycaret 결과를 바탕으로 세 가지 2. feature importance 확인 후 피쳐 	<ul style="list-style-type: none"> - 정제 작업은 내일 바로 돌리도록 1. 상위 모델에서 피쳐 중요도를 추출 2. 상위 피쳐들을 n개를 뽑아서 정제 작업 마무리 - 파이캐럿 애러 잡느라 시간을 많이 소모 	<ul style="list-style-type: none"> < 상위 모델 > 1. lr 2. light gbm 3. svm
40	1/6 (금)		- 1/13 발표 리허설	<ul style="list-style-type: none"> <파이캐럿으로 test 돌린 결과> - 공유폴더에 피쳐중요도 파일 - 가장 정확도가 높은 모델은 SVM - 상식 수준의 기준점보다 정확도가 높긴 하지만, 드라마틱한 차이를 보이지 않는다. - 과대적합이 보인다. => 파이캐럿이 아니라 직접 모델을 구성해서 돌려보자.
41	1/6 (금)	멘토링		<ul style="list-style-type: none"> - 7분기, 8분기 구매값과 고객은 다른 특징을 보이니 공통된 것도 있었다. 전체구매금액과 제휴사별 구매금액이 중요했다. 분기별로 달리 중 - 다음주 초까지: 구매값소예측 -> 그룹(군집화) -> 군집유형의 특징(변수) 탐색 -> 변수 변형 (-> 재군집화) -> 대응방안 -> 발표자료(논리전) - 데이터 만지면서 공부하는 게 프로젝트의 중점
	1/7 (토)	1. 구매값소예측 결과		<ul style="list-style-type: none"> [사용모델] SVM, LR, LightGBM, KNN 정확도 75-, 60, 66, 66

+

≡

시트1

+

태그



윤예은(카쓰)

프로젝트의 시작부터 끝까지 다 함께 계획하고 진행하면서 무엇이 중요한지 배우는 시간이었습니다.



정재훈(훈)

데이터를 분석하고 처리하는 과정과 혼자만의 생각으로는 발견할 수 없던 오류 수정하는 법 등을 배웠습니다.



정제훈(짱제)

프로젝트 기간동안 나에게 무엇이 부족하고 무엇을 채워야 하는지 알 수 있는 의미있는 시간이었습니다.



최지원(공사)

많이 배울 수 있었던 프로젝트였습니다.모두 고생하셨습니다~!



허진욱(은아)

미래 예측은 어려운것 같습니다.팀원분들 모두 수고하셨습니다.

감사합니다.

Q□&□A



[팀장] 김윤예은(카쓰)



정재훈(훈)



정제훈(짱제)



최지원(공사)



허진욱(은아)