

구매감소 고객 예측기반 매출관리 솔루션

K-digital 3기

2021.10.18 | 3조

박민아 : 김진 : 이민찬 : 이슬

INDEX

- 01 유통업계 동향 파악 및 주제선정
- 02 EDA를 통한 과제 도출
- 03 변수 선정 및 모델링
- 04 시나리오 및 솔루션 제안

3조를 소개합니다



박민아
팀장

기술 및 기획 총괄
parkmina365@gmail.com



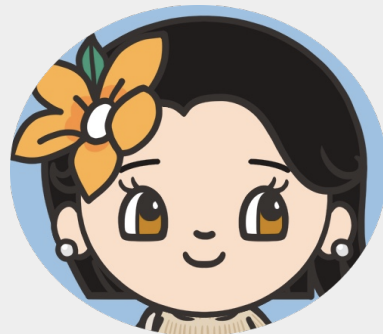
김진
팀원

기획, 데이터 분석, 솔루션 도출
camuscheer333@gmail.com



이민찬
팀원

기획, 데이터 탐색, 분석, 모델링
leemc9955@gmail.com



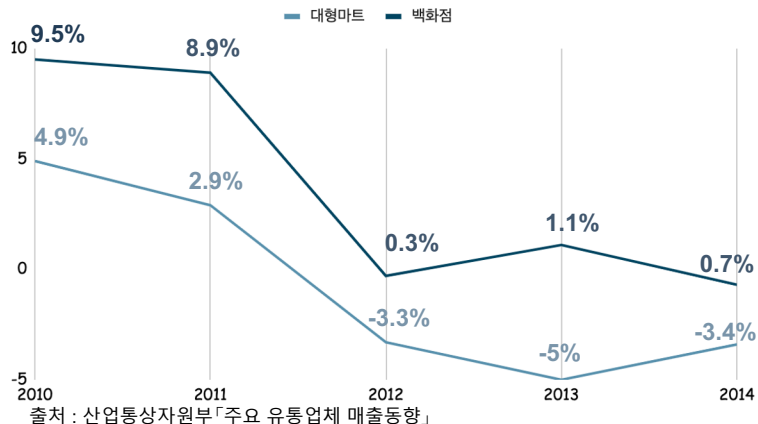
이슬
팀원

기획, 데이터 탐색, 분석, 모델링
seuly1203@gmail.com

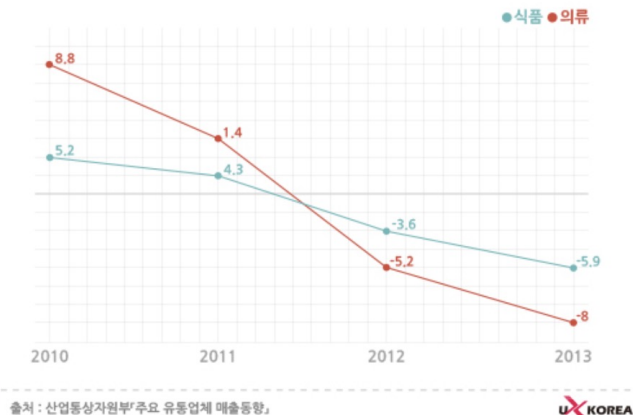
1.유통업계 동향 파악 및 주제 선정

1. 유통업계 동향 파악 및 주제 선정 - 업계 현황

주요 유통업체 매출 증감률(대형마트, 백화점)



주요 유통업체 식품 의류 매출 동향(전년 동원대비 증감률, %)



- 2011년 이후 주요 유통업체 매출 증감률, 마이너스 성장 기록함
- 동일한 시기, 유통업체 주요 품목인 식품 및 의류 매출, 지속적으로 하락
- 주요 업체 및 주요 상품의 매출 하락으로 인한 유통업계 불황 도래

1. 유통업계 동향 파악 및 주제 선정 - 동향 분석

업체 매출액

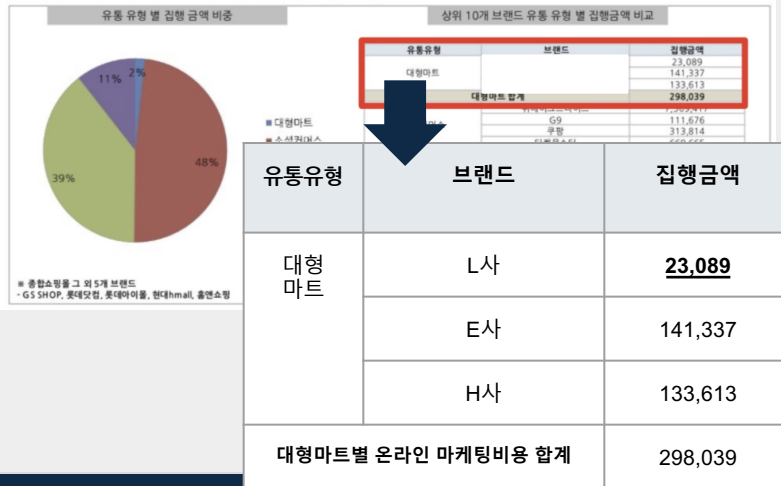
업체	국내 점포수	국내 매출액 (2013)	영업이익	비고(2012년대비)
E사	150	10조 7801억	7592억	매출 -1.5% 감소, 영업이익 2% 감소
H사	139	8조 9298억	4476억	영업이익 -24.4% 감소
L사	108	8조 8365억	3160억	매출 -0.1%, 영업이익은 -12% 감소
C사	11	2조 5371억	1308억	영업이익 6.3% 증가

*2013년 기준

UX KOREA

3-1. 2015 온라인 마케팅 분석

■ 온라인 분석 - 유통 유형별 마케팅 비용 분석



- 2013년 유통업체 중 C사를 제외한, E사, H사, L사 매출 감소
- L사의 온라인 마케팅 집행비용 분석 결과, 국내 업계 매출 1위인 E사의 16%에 불과함

1. 유통업계 동향 파악 및 주제 선정 - 동향 분석

E사

E사 : 소비자 니즈 트렌드를 반영한 상품 개발

- E사에서만 구매 가능한 상품 및 브랜드 출시
- LED 전구 : 2015년 1분기 전구 매출 40% 점유
- 홍삼정 : 홍삼 제품 매출의 90% 점유
- E사 자체 식품(PB상품) P제품 : 간편 가정식 부문 매출 55.7% 점유

H사

H사 : 소비자 니즈를 반영한 인기 상품 할인 진행

- 고객 구매 빈도가 높은 상위 500개 상품, 연중 상시 10%-30% 할인
- 해당 프로모션 홍보를 위해 2015년 3월, 1000억원 자체투자, 온라인 마케팅 비용 집중 투자

- 타사의 경우 매출 방어를 위해 온라인 마케팅을 강화하며 고객 중심 마케팅을 시도함
- L사는 온, 오프라인 채널 동시 활성화를 도모 했으나 고객 니즈 파악이 미흡하여 활성화로 이어지지 않음
- 실제 고객 데이터를 활용한 고객 중심 마케팅 필요

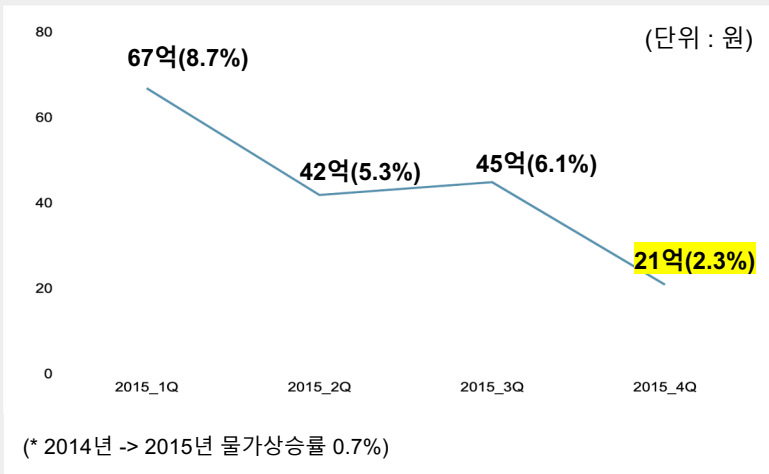
2. EDA를 통한 과제 도출

2. EDA를 통한 과제 도출

◆ 2014-2015 각 분기별 구매액

[26] grouped_2014_1 = outer_frame['구매금액_2014_1']	[35] grouped_2015_1 = outer_frame['구매금액_2015_1']
[27] grouped_2014_1.sum()	[36] grouped_2015_1.sum()
77694762743.0	84454475955.0
[28] grouped_2014_2 = outer_frame['구매금액_2014_2']	[37] grouped_2015_2 = outer_frame['구매금액_2015_2']
[29] grouped_2014_2.sum()	[38] grouped_2015_2.sum()
80545967397.0	84834674190.0
[30] grouped_2014_3 = outer_frame['구매금액_2014_3']	[39] grouped_2015_3 = outer_frame['구매금액_2015_3']
[31] grouped_2014_3.sum()	[40] grouped_2015_3.sum()
74949714657.0	79536551000.0
[32] grouped_2014_4 = outer_frame['구매금액_2014_4']	[41] grouped_2015_4 = outer_frame['구매금액_2015_4']
[34] grouped_2014_4.sum()	[42] grouped_2015_4.sum()
96411395784.0	98591615215.0

◆ 2015년 분기별, 2014년 대비 구매액 증감률



- 2015년 1분기부터 2015년 4분기, **지속적으로 구매액 증감률이 하락하는 이슈 발견**
- 특히 가장 높은 매출을 기록하는 4분기, **구매액 증감률 2.3%에 그침**

2. EDA를 통한 과제 도출 - 분기별 구매고객 감소비율

◆ 2014년 - 2015년 분기별 구매고객 감소 비율(물가상승률 0.7% 반영)

```
# 2015년 구매액 보정치 추가
# 물가상승률 0.7%로 반영하여 2015년 구매액에 1000/1007 곱하는 연산하여 만들
C = 1000/1007

outer_frame['구매금액_2015_1_n'] = outer_frame.구매금액_2015_1*C
outer_frame['구매금액_2015_2_n'] = outer_frame.구매금액_2015_2*C
outer_frame['구매금액_2015_3_n'] = outer_frame.구매금액_2015_3*C
outer_frame['구매금액_2015_4_n'] = outer_frame.구매금액_2015_4*C

outer_frame = outer_frame.fillna(0)
# 1분기 구매금액 비교 (물가상승률 보정치)
df_temp = outer_frame.구매금액_2015_1_n - outer_frame.구매금액_2014_1
print('1분기 구매 감소고객 백분율 (보정) : ', f'{len(df_temp[df_temp < 0])/len(df_temp)+100:.1f}%')
# 2분기 구매금액 비교 (물가상승률 보정치)
df_temp = outer_frame.구매금액_2015_2_n - outer_frame.구매금액_2014_2
print('2분기 구매 감소고객 백분율 (보정) : ', f'{len(df_temp[df_temp < 0])/len(df_temp)+100:.1f}%')
# 3분기 구매금액 비교 (물가상승률 보정치)
df_temp = outer_frame.구매금액_2015_3_n - outer_frame.구매금액_2014_3
print('3분기 구매 감소고객 백분율 (보정) : ', f'{len(df_temp[df_temp < 0])/len(df_temp)+100:.1f}%')
# 4분기 구매금액 비교 (물가상승률 보정치)
df_temp = outer_frame.구매금액_2015_4_n - outer_frame.구매금액_2014_4
print('4분기 구매 감소고객 백분율 (보정) : ', f'{len(df_temp[df_temp < 0])/len(df_temp)+100:.1f}%')

1분기 구매 감소고객 백분율 (보정) : 43.7%
2분기 구매 감소고객 백분율 (보정) : 44.0%
3분기 구매 감소고객 백분율 (보정) : 46.4%
4분기 구매 감소고객 백분율 (보정) : 52.6%
```

- 구매 감소 고객 파악을 위해, 물가 상승률을 반영한 분기별 구매 감소 고객 백분율 비교
- **4분기 구매 감소 고객 백분율이 52.6%**로 가장 높음
- 지속적으로 구매 감소 고객이 증가, 점진적인 매출하락의 원인이 됨

2. EDA를 통한 과제 도출 - 분기별 구매금액 통계표

◆ 2014년 - 2015년 각 분기별 구매금액 통계표

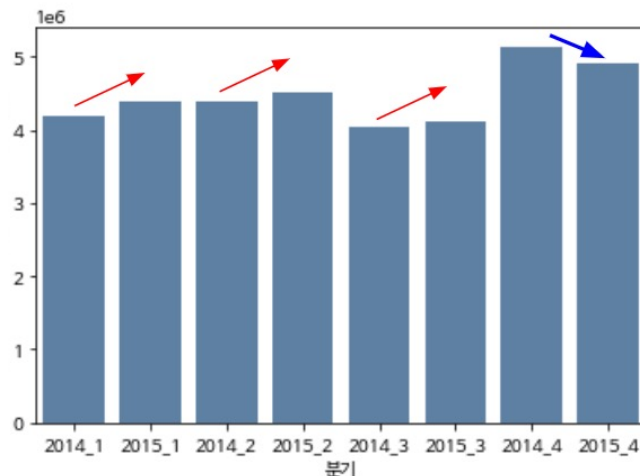
분기별 평균 구매금액
ppI_pv1.describe()

분기	2014_1	2014_2	2014_3	2014_4	2015_1	2015_2	2015_3	2015_4
count	1.905400e+04	1.913900e+04	1.921200e+04	1.926800e+04	1.929200e+04	1.930600e+04	1.930500e+04	1.931900e+04
mean	4.077609e+06	4.208473e+06	3.901193e+06	5.003705e+06	4.377694e+06	4.394213e+06	4.119997e+06	5.103350e+06
std	9.909987e+06	1.067154e+07	8.563611e+06	1.416751e+07	9.045264e+06	9.563959e+06	9.115610e+06	1.342569e+07
min	3.000000e+03	1.480000e+03	1.000000e+03	1.700000e+03	3.800000e+02	1.000000e+03	1.700000e+03	9.900000e+02
25%	1.141361e+06	1.186367e+06	1.226416e+06	1.321172e+06	1.287495e+06	1.304736e+06	1.298370e+06	1.231070e+06
50%	1.928462e+06	1.993484e+06	1.996125e+06	2.210862e+06	2.104906e+06	2.120692e+06	2.083830e+06	2.108012e+06
75%	4.197174e+06	4.381028e+06	4.036430e+06	5.137120e+06	4.379638e+06	4.512851e+06	4.102356e+06	4.915456e+06
max	6.248265e+08	8.364896e+08	6.115875e+08	1.319230e+09	4.011433e+08	6.441495e+08	6.173036e+08	8.805502e+08

- 고객별 구매경향을 파악하기 위해 2014년 1분기 부터 2015년 4분기까지 구매 금액표 산출
- 매출 상위 25%의 전체매출 기여도가 4분위 그룹 중 가장 높음
- 2015년 4분기의 경우, 매출 상위 25%의 1인당 평균 구매액이 감소하는 이슈 발생

2. EDA를 통한 과제 도출 - 상위 25% 고객 분기별 매출변화

◆ 상위 25% 고객의 분기별 매출 변화



- 매출 상위 25% 고객의 1인당 평균 구매액, 4분기 제외 모두 증가함
- 매출 하락의 주 요인은 위 고객층의 구매액 감소에 있는 것으로 판단

→ 모델링을 통해 주요 구매 감소 영향 요인에 기반한 구매 패턴 유형화 후
매출 방어 솔루션을 제안하고자 함

3. 변수 선정 및 모델링

3. 변수 선정 및 모델링 - 변수 설정

◆ 변수 리스트

인구 통계학적 변수	구매액 관련 변수	상품 구매개수 관련 변수	방문횟수 관련 변수
성별 멤버십 가입 여부	회당 구매금액 증감율 총 구매금액 평균 총 구매금액 증감율 A사 구매 금액 증감율 B사 구매 금액 증감율 C사 구매 금액 증감율	A사 상품 구매 개수 평균 B사 상품 구매 개수 평균 C사 상품 구매 개수 평균	A사 방문 횟수 평균 B사 방문 횟수 평균 C사 방문 횟수 평균

- 총 컬럼개수 14개
- 구매감소고객 예측 정확도를 높이기 위한 구매액, 구매개수, 방문횟수 관련 변수 및 인구 통계학적 변수로 선정

3. 변수 선정 및 모델링 - Feature 선정

◆ Lightgbm Feature Importance

Weight	Feature
0.1997 ± 0.0058	총구매금액증감율
0.0371 ± 0.0023	B방문횟수평균
0.0340 ± 0.0044	총구매금액평균
0.0305 ± 0.0022	A방문횟수평균
0.0244 ± 0.0022	C방문횟수평균
0.0182 ± 0.0020	B상품구매개수평균
0.0162 ± 0.0009	회당구매금액증감율
0.0114 ± 0.0014	A상품구매개수평균
0.0112 ± 0.0018	B구매금액증감율
0.0105 ± 0.0012	C상품구매개수평균
0.0080 ± 0.0012	A구매금액증감율
0.0078 ± 0.0012	C구매금액증감율
0.0009 ± 0.0005	성별
0.0006 ± 0.0005	멤버십가입여부

◆ XGBoost Feature Importance

Weight	Feature
0.1808 ± 0.0114	총구매금액증감율
0.0241 ± 0.0019	B방문횟수평균
0.0211 ± 0.0025	A방문횟수평균
0.0162 ± 0.0022	C방문횟수평균
0.0149 ± 0.0025	총구매금액평균
0.0036 ± 0.0013	A상품구매개수평균
0.0027 ± 0.0007	회당구매금액증감율
0.0018 ± 0.0018	B구매금액증감율
0.0009 ± 0.0014	B상품구매개수평균
0.0008 ± 0.0005	C상품구매개수평균
0.0004 ± 0.0002	멤버십가입여부
0.0003 ± 0.0008	C구매금액증감율
0.0000 ± 0.0002	성별
-0.0000 ± 0.0009	A구매금액증감율

- 적합한 Feature 선정을 위해 Lightgbm, XGBoost의 Feature Importance 확인
- 총 구매금액 변동이 종속변수인 구매감소 고객예측과 연관성이 높은 것을 파악할 수 있었음

3. 변수 선정 및 모델링 - 활용 모델

활용 모델

XGBoost



LightGBM



CatBoost

AdaBoost

*Logistic
Regression*



Decision tree

- 총 7개의 모델 활용
- 부스팅 모델(XGBoost, LightGBM, CatBoost, AdaBoost) 및 Logistic Regression, Random Forest, Decision Tree 사용

3. 변수 선정 및 모델링 - 학습 및 검증 데이터

* 학습용 데이터

Train_Data (X_train)

	2014년 상반기 정보	2014년 하반기 정보
---	-----------------	-----------------

Train_Target (y_train)

= 2014년 상반기 대비, 2015년 상반기 구매감소여부 (0, 1)

* 테스트용 데이터

Test_Data (X_test)

2014년 상반기 정보	2014년 하반기 정보	2015년 상반기 정보
-----------------	-----------------	-----------------

Test_Target (y_test)

= 2014년 상반기 대비, 2015년 하반기 구매감소여부(0, 1)

3. 변수 선정 및 모델링 - 모델링 결과

```
x_train = x_train_1[['성별', '멤버십가입여부',  
                    '총구매금액평균', '총구매금액증감율',  
                    '회당구매금액증감율',  
                    'A구매금액증감율', 'B구매금액증감율', 'C구매금액증감율',  
                    'A상품구매개수평균', 'B상품구매개수평균', 'C상품구매개수평균',  
                    'A방문횟수평균', 'B방문횟수평균', 'C방문횟수평균']]  
  
x_test = x_test_1[['성별', '멤버십가입여부',  
                   '총구매금액평균', '총구매금액증감율',  
                   '회당구매금액증감율',  
                   'A구매금액증감율', 'B구매금액증감율', 'C구매금액증감율',  
                   'A상품구매개수평균', 'B상품구매개수평균', 'C상품구매개수평균',  
                   'A방문횟수평균', 'B방문횟수평균', 'C방문횟수평균']]  
  
model1 = XGBClassifier()  
model2 = LGBMClassifier()  
model3 = CatBoostClassifier(silent=True)  
model4 = AdaBoostClassifier()  
model5 = LogisticRegression()  
model6 = DecisionTreeClassifier()  
model7 = RandomForestClassifier()
```

XGB train 정확도: 0.7424

XGB test 정확도: 0.7394

LightGBM train 정확도: 0.7873

LightGBM test 정확도: 0.7392

CatBoost train 정확도: 0.8024

CatBoost test 정확도: 0.7370

AdaBoost train 정확도: 0.7323

AdaBoost test 정확도: 0.7342

LogisticRegression train 정확도: 0.7246

LogisticRegression test 정확도: 0.7303

DecisionTree train 정확도: 1.0000

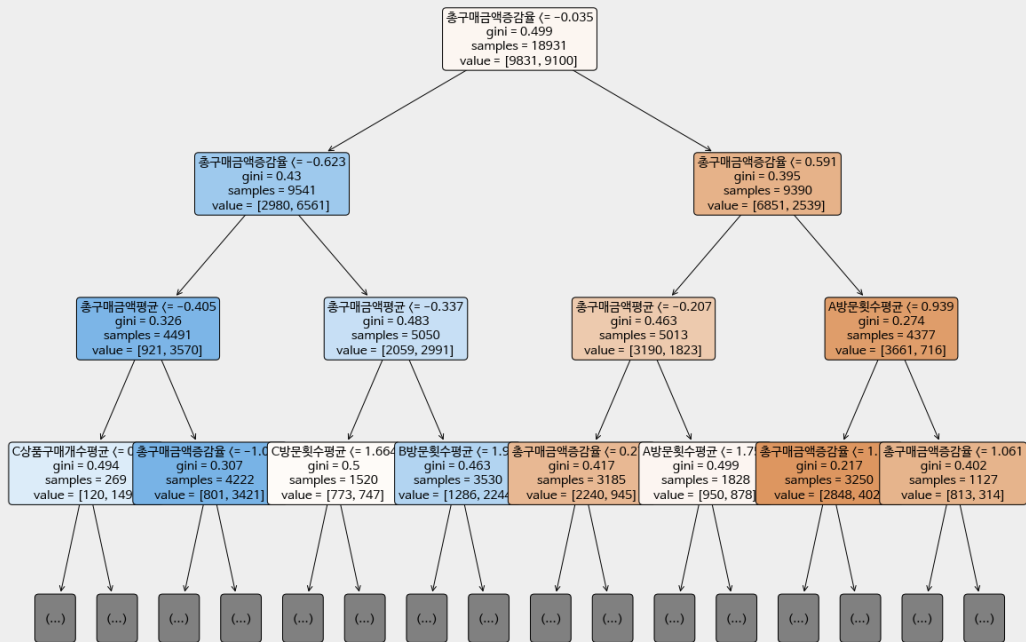
DecisionTree test 정확도: 0.6443

RandomForest train 정확도: 1.0000

RandomForest test 정확도: 0.7349

- 총 7개의 모델을 사용하여 14개의 변수를 사용하여 모델링 진행
- 정확도를 더 향상 시키기 위하여 하이퍼 파라미터 튜닝 진행

3. 변수 선정 및 모델링 - Decision tree 결과



- Decision tree 결과, '총 구매금액 증감율', '총 구매금액 평균', 'A방문횟수평균'을 중심으로 나눔
- 해당 변수들이 중요 변수임을 확인

3. 변수 선정 및 모델링 - 하이퍼 파라미터 튜닝

◆ 하이퍼 파라미터 튜닝

```
model = LGBMClassifier(learning_rate = 0.1, max_depth = 10, num_leaves = 15)
model.fit(x_train, y_train)
print('train score :', model.score(x_train, y_train))
print('test score :', model.score(x_test, y_test))
```

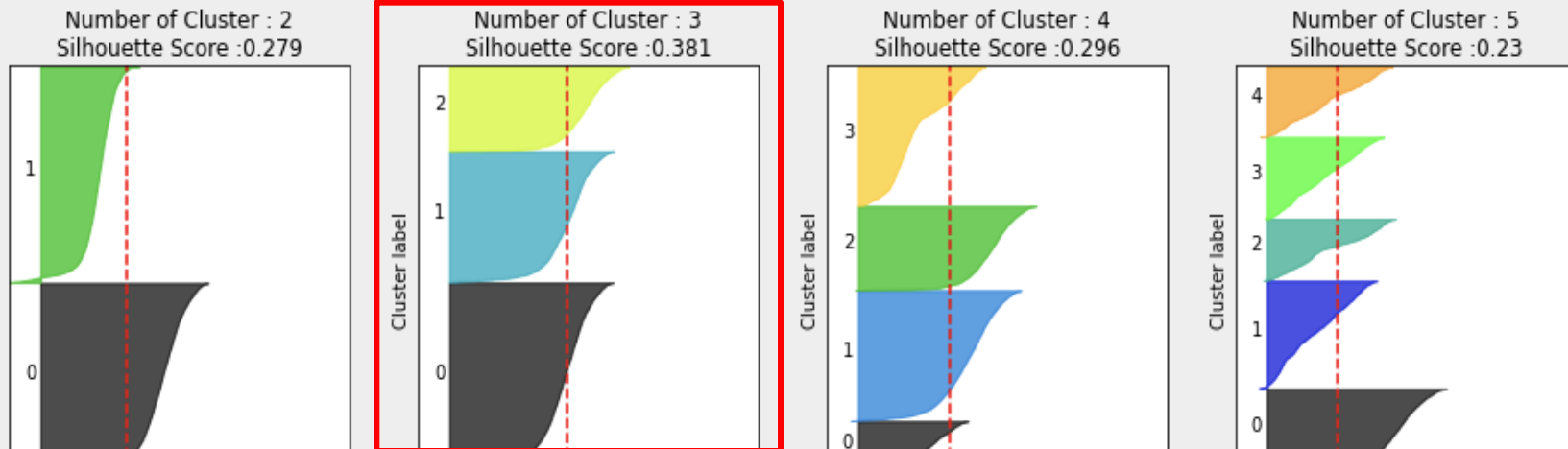
```
train score : 0.7556388991601077
```

```
test score : 0.7401174789649151
```

- GridSearch CV 활용, 하이퍼 파라미터 튜닝을 통한 정확도 향상 진행
- Test Score 0.74대로 향상

3. 변수 선정 및 모델링 - 군집화(실루엣 분석)

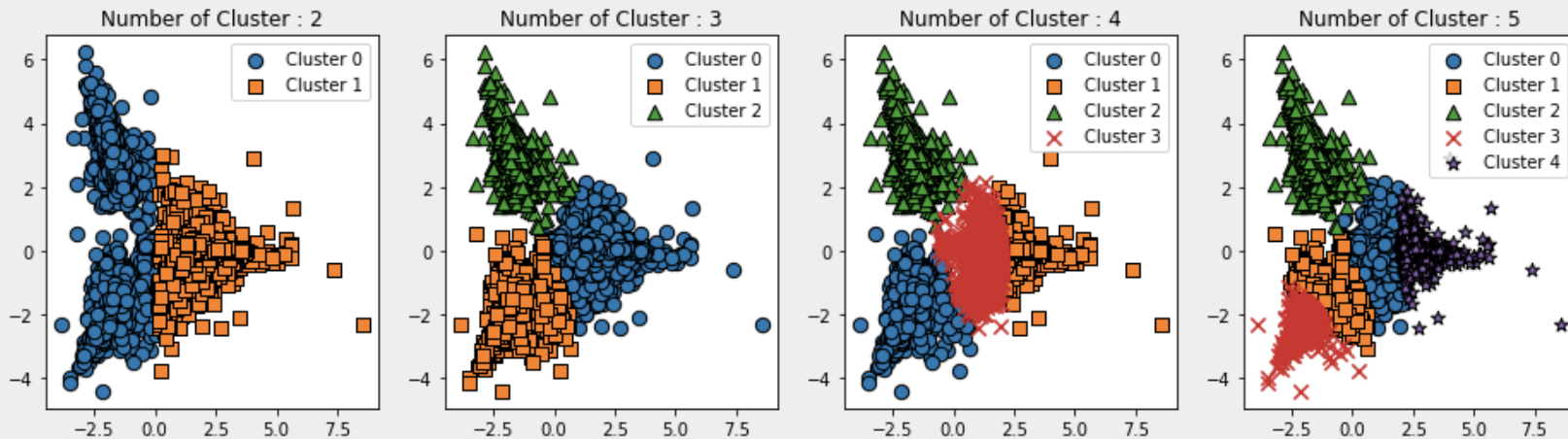
◆ 실루엣 분석 결과



- 2개 - 5개의 그룹으로 분류한 Silhouette Score 분석 중 군집을 3개로 분류한 점수가 0.381로 가장 높게 나옴
- 실루엣 분석 결과, 3개의 군집으로 분류되는 것을 확인

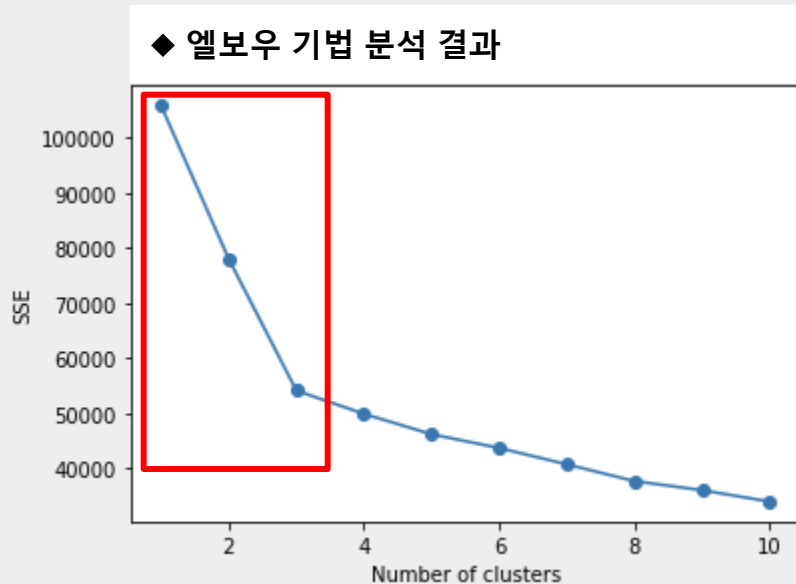
3. 변수 선정 및 모델링 - 군집화(PCA 분석)

◆ PCA 분석 결과



- PCA 분석 결과 3개의 군집으로 나눈 결과가 가장 명확하게 분류되는 것을 파악할 수 있었음

3. 변수 선정 및 모델링 - 군집화(엘보우 기법)



- SSE의 급격한 감소가 둔화되는 지점인 3회차 이후를 적절한 군집의 개수로 판단
- 즉, 앞선 분석 결과와 마찬가지로 3개의 군집으로 분류되는 것을 파악할 수 있었음

3. 변수 선정 및 모델링 - 결과

◆ 군집결과 분석 데이터 프레임

	성별	멤버십가입 여부	회당구매금액증 감률	총구매금액평 균	총구매금액증 감률	A구매금액증 감률	B구매금액증 감률	C구매금액증 감률	A상품구매개수 평균	B상품구매개수 평균	C상품구매개수 평균	A방문횟수평 균	B방문횟수평 균	C방문횟수평 균	고객 수
0	0.197626	0.181631	-0.097872	3.790680e+06	-0.202216	-0.179792	-0.118538	-0.133491	8.215514	21.346577	559.367217	7.676815	5.732886	171.408153	1938
1	0.166324	0.211727	-0.206298	1.644313e+07	-0.324274	-0.342863	0.026365	0.008730	176.768576	54.581489	26.756103	136.926991	11.341623	6.095749	4383
2	0.181095	0.391376	-0.118139	4.557501e+06	-0.238591	-0.216175	-0.183186	-0.017059	13.736761	431.637479	15.553123	12.721061	115.074959	4.447098	3015

*성별 : 0 - 여성 / 1 - 남성

*멤버십 가입여부 : 0 - 미가입 / 1 - 가입

- 변동구매 감소로 예측되는 고객들의 변수별 평균값을 나타낸 데이터 프레임 생성
- 방문횟수, 구매 금액 증감률, 각 계열사별 구매 개수 분석 결과
A사 주 이용 고객, B사 주 이용 고객, C사 주 이용 고객, 총 3가지 그룹으로 나뉘는 것을 확인
- 각 그룹을 위한 솔루션 제안이 필요

4. 시나리오 및 솔루션 제안

4. 시나리오 및 솔루션 제안 - 군집화 결과

	A사 주 이용 고객	B사 주 이용 고객	C사 주 이용 고객
고객수	4,383명	1,938명	3,015명
성별 비중	여성 : 82% 남성 : 18%	여성 : 84% 남성 : 16%	여성 : 81% 남성 : 19%
멤버십 가입 여부	가입 : 21% 미가입 : 79%	가입 : 39% 미가입 : 61%	가입 : 18% 미가입 : 82%

- 군집 분석 결과 A사 4,383명, B사 1,938명, C사 3,015명 고객으로 분류됨
- 구매감소 고객의 여성 고객 비중이 3사 모두 80% 이상을 점유
- 3사 중 B사의 멤버십 가입률이 가장 높음 -> 가입률 39%

4. 시나리오 및 솔루션 제안 - 타겟 고객 군집별 제안

◆ 군집 1

계열사	A사 주 위주 그룹
고객수	4,383명
성별 비중	여성 : 82%, 남성 : 18%
멤버십 여부	가입 : 21%, 미가입 : 79%
고객 특징	의류 및 잡화를 구매하는 고객 비율 높음

◆ 군집 1 대상 제안 솔루션

단골고객 대상 프로모션 및 할인 쿠폰 제공

특정 횟수 및 금액 이상 구매 고객 대상,
앱 내 의류 브랜드 신규 상품 홍보 및 코디 정보
콘텐츠 배포 의류 할인 쿠폰 제공

기대효과 :

- 1) 의류 제품 매출 상승
- 2) 신규 상품 홍보 효과
- 3) 잡화 및 신발상품과 연계 판매, 추가 매출 발생

- 군집 1 고객 대상, 의류 특화 코디 정보 콘텐츠 및 할인 쿠폰을 담은 콘텐츠 제공
- 의류상품 매출 향상 및 해당 시즌 신규 의상 홍보,
잡화 및 신발 상품과 연계 판매로 인한 추가 매출 발생 기대

4. 시나리오 및 솔루션 제안 - 타겟 고객 군집별 제안

◆ 군집 2

계열사	B사 주 위주 고객
고객수	1,938명
성별 비중	여성 : 84%, 남성 : 16%
멤버십 여부	가입 : 39%, 미가입 : 61%
고객 특징	타 계열사에 비해 멤버십 가입률이 높음

◆ 군집 2 대상 제안 솔루션

개인 맞춤 모바일 쿠폰 배포

종이 쿠폰을 모바일 쿠폰으로 전환 후,
고객의 장바구니 분석 후
재구매율이 높은 상품 위주로 할인 쿠폰 제공

기대효과 :

- 1) 쿠폰 제공으로 추가 구매 유도 가능
- 2) 고객별 개인화 마케팅 가능

- 군집 2 고객의 경우 전체 고객 중 **멤버십 가입률 39%**로, 타 계열사 보다 가입률이 높음
- 지류쿠폰에서 모바일쿠폰으로 전환, 고객의 장바구니 분석 후 재구매율이 높은상품 위주로 할인 쿠폰 제공
- 고객별 개인화 마케팅 및, 쿠폰 제공을 통한 추가 구매 유도 기대

4. 시나리오 및 솔루션 제안 - 타겟 고객 군집별 제안

◆ 군집 3

계열사	C사 주 이용 고객
고객수	3,015명
성별 비중	여성 : 81%, 남성 : 19%
멤버십 여부	가입 : 18%, 미가입 : 82%
고객 특징	식재료 구매를 위해 일상적으로 방문하는 고객 비율이 높음

◆ 군집 3 대상 제안 솔루션

PB상품 별 무료 멤버십 제도

구매율이 높은 PB상품 위주로 무료 멤버십 개설
가입자들에게 매월 멤버십 별 할인 쿠폰 제공
구매 횟수가 많을 수록 추가혜택 증정
예) 특정 구매 횟수 (예:3,6,9회)시 추가 할인 제공

기대효과 :

- 1) 품목별 고객 관리
- 2) PB상품 홍보
- 3) 다른 PB 연계 판매

- 군집 3 고객의 경우 **식재료 구매를 위해 일상적으로 매장에 방문하는 고객 비중이 높음**
- 구매율이 높은 PB(Private Brand) 상품을 위주로 무료 멤버십 제도를 개설
- 멤버십 가입자들에게 할인 쿠폰을 제공하며 구매 유도 및 PB상품 홍보
다른 PB상품 연계 판매를 통한 추가 매출 발생 기대



Q & A

감사합니다