

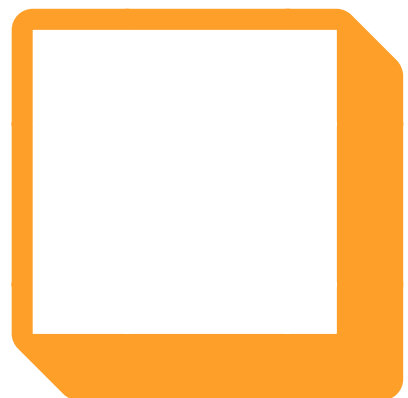
투자성향 기반 개인화 주식 종목 추천

고객 및 시장 데이터를 활용한 대고객 금융 서비스 제안

KSK팀 - 고건호, 송창용, 김도연

MIRAE ASSET
미래에셋증권

Contents.



01

개요

서비스 도입 배경

02

데이터 설명

고객데이터
산업 분류 데이터

03

데이터 전처리

군집분석에 사용한 5개 항목에 대한 설명
추천시스템에 활용할 모델 설명

04

고객 투자 성향 분석

군집분석 모델 설명
고객별 특징 해석

05

Flow Chart 모델 설명

06

서비스 설명

프로젝트 개요



2020년 신종 코로나 감염증(코로나19) 사태 이후 대거 유입된
2030세대의 젊은 투자자층은 위험을 적극 추구하는 성향이 짙다.
가장 높은 개인투자자 비중을 기록한 미래에셋운용은 ETF 상품 전략에서 차별성을 내세우며
코로나19 사태 이후 빠르게 유입된 2030 투자자를 확보

"대표 지수를 쫓는 ETF뿐 아니라 기술 기반 해외 혁신 기업에 집중 투자하는
테마형 ETF를 대거 늘려온 전략이 통한 것으로 보인다"고 설명했다.

출처: MoneyS | 미래에셋 ETF 질주... 2년반만에 삼성운용 독주 무너뜨린 비결은

투자 성향에 기반한 개인화 ETF 추천 서비스 제안

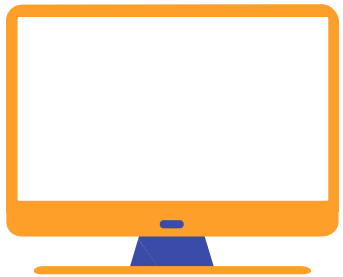
“미래에셋증권의 WM 비즈니스의 궁극적인 목적은
투자자들이 한국 자산에 편중된 투자에서 벗어나
글로벌 우량 자산에 대한 투자를 적극적으로 돕는 가이드 역할을 하는 것”

미래에셋증권은 글로벌 우량 자산 투자를 유도하겠다는 목표도 세웠다.
고금리와 글로벌 국가 간 경제 블록화, AI 혁신이 이뤄지는 글로벌 트렌드 속에서
자산 배분 기반의 포트폴리오 분산투자를 강화한다는 것이다.

출처: 매거진 한경 | [WM리더] 허선호 미래에셋증권 사장 "글로벌 투자에 역점... AI 투자비서 등 활용"

종목 추천을 통해 고객의 분산 투자에 도움을 주는 서비스 제안

데이터 설명



미래에셋에서 제공한 **고객데이터** 중

- **자본수준**(현금성 자산액 + 국내 주식 평가액 + 해외주식 평가액)
 - **주식관심도**(거래체결횟수 + 거래액 + 주거래매체)
 - **투자분산도**(매수액 1,2,3위 주식 합/ 전체 주식 매수액)
 - **주식보유기간**(거래기간에 따라 비율 차등 적용)
 - **국내주식투자비율**(국내 주식 평가액 / 보유 주식 평가액)
- 를 군집분석에 이용

Python library **FinanceDataReader**의

GICS 글로벌산업분류기준을 이용하여

해외 주식의 **산업 분류**

GICS 기반 **wiseindex**의 산업 구분 데이터를 이용하여

국내 주식의 **산업 분류**

데이터 전처리 (자본수준)



자본수준 : 고객의 자본 수준을 수치화

1. 결측치 데이터 -> 0으로 처리
2. 현금성 자산액 + 국내 주식 평가액 + 해외 주식 평가액의 4달 평균 금액 계산
3. rank 함수를 사용하여 전체 고객 대비 본인의 위치를 비율로 계산
4. 자본수준을 5개 등급으로 분류

```
# 등급이 높을수록 자본수준 높음
# 4달 평균 보유 자산 합
def degree(x):
    if x >= 0.8:
        return 1
    elif x >= 0.6:
        return 2
    elif x >= 0.4:
        return 3
    elif x >= 0.2:
        return 4
    else:
        return 5

custom['CAPITAL_avg'] = ((custom['CASH_AST_M1'] + custom['CASH_AST_M2'] + custom['CASH_AST_M3'] + custom['CASH_AST_M4']) +
                        (custom['DMST_AST_EVAL_M1'] + custom['DMST_AST_EVAL_M2'] + custom['DMST_AST_EVAL_M3'] + custom['DMST_AST_EVAL_M4'])) +
                        (custom['OVST_AST_EVAL_M1'] + custom['OVST_AST_EVAL_M2'] + custom['OVST_AST_EVAL_M3'] + custom['OVST_AST_EVAL_M4']))/4

custom['CAPITAL_percentile'] = custom['CAPITAL_avg'].rank(pct=True)
custom['CAPITAL_degree'] = custom['CAPITAL_percentile'].apply(degree) # 등급 매기기
```

데이터 전처리 (주식관심도)



주식관심도 : 주식에 대한 고객의 관심을 수치화

1. 결측치 데이터 -> 0으로 처리
2. HTS 사용 비중이 80% 이상인 고객 분류(6%)
3. 거래횟수, 거래금액, 접속일수 **가중평균**(주거래 매체가 HTS인 경우 가중치 가산)
4. rank 함수를 사용하여 전체 고객 대비 본인의 위치를 **비율**로 계산
5. 자본수준을 **5개 등급**으로 분류

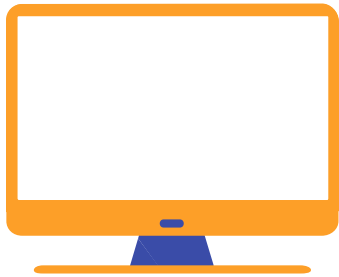
```
# 주거래 매체
# HTS 접속비중이 0.8이상인 경우 주거래매체를 HTS로 판단(비중이 0.8 이상인 고객 비중은 6%)
custom['PLATFORM HTS'] = custom['HTS_DYS_ratio'].apply(lambda x : 1 if x >= 0.8 else 0)

# 거래체결횟수, 거래액, 접속일수(MTS, HTS)
# 등급이 높을수록 관심도가 높음
# 가중평균 거래횟수 0.3, 거래금액 0.3, 접속일수 주거래가 HTS일 경우 0.4, MTS일 경우 0.3

custom['STOCK_INTEREST'] = (custom['TRANSACTION_CNT'].rank(pct=True) * 0.3 +
                           custom['TRANSACTION_AMT_sum'].rank(pct=True) * 0.3 +
                           custom['CONN_DYS_sum'].rank(pct = True) * custom['PLATFORM HTS'].apply(lambda x : 0.4 if x == 1 else 0.3))

custom['STOCK_INTEREST_percentile'] = custom['STOCK_INTEREST'].rank(pct=True)
custom['STOCK_INTEREST_degree'] = custom['STOCK_INTEREST_percentile'].apply(degree) # 등급 매기기
```

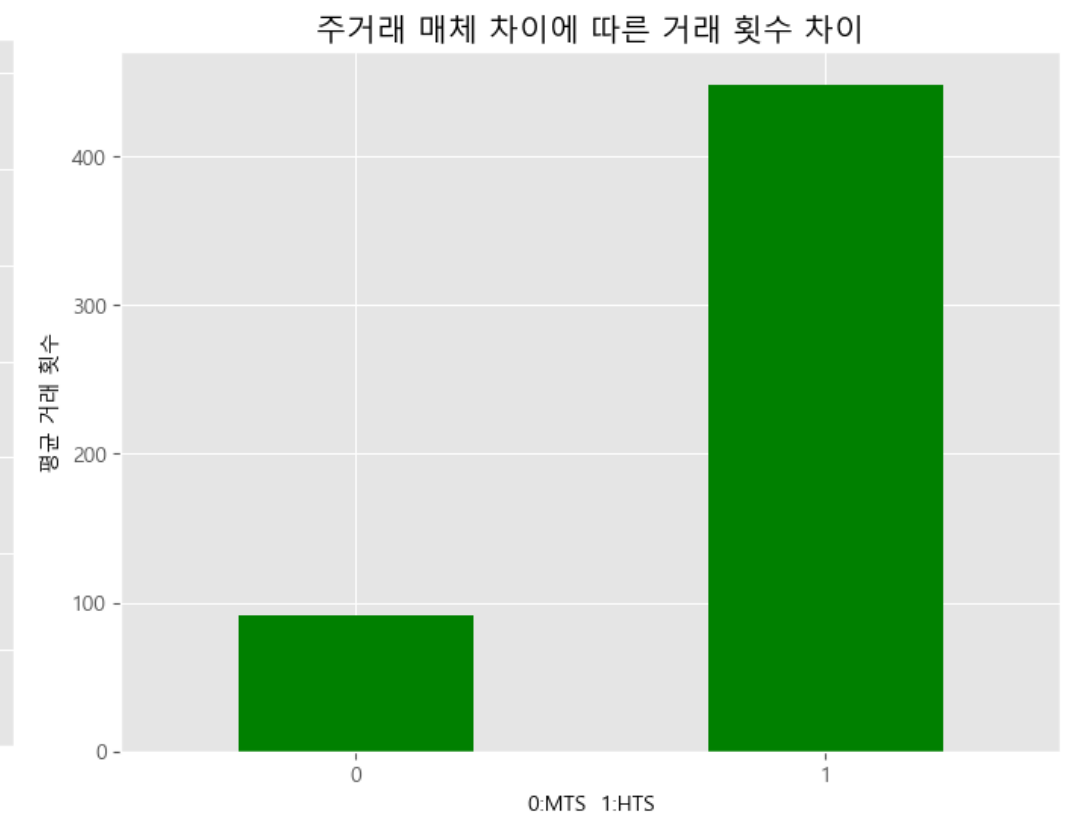
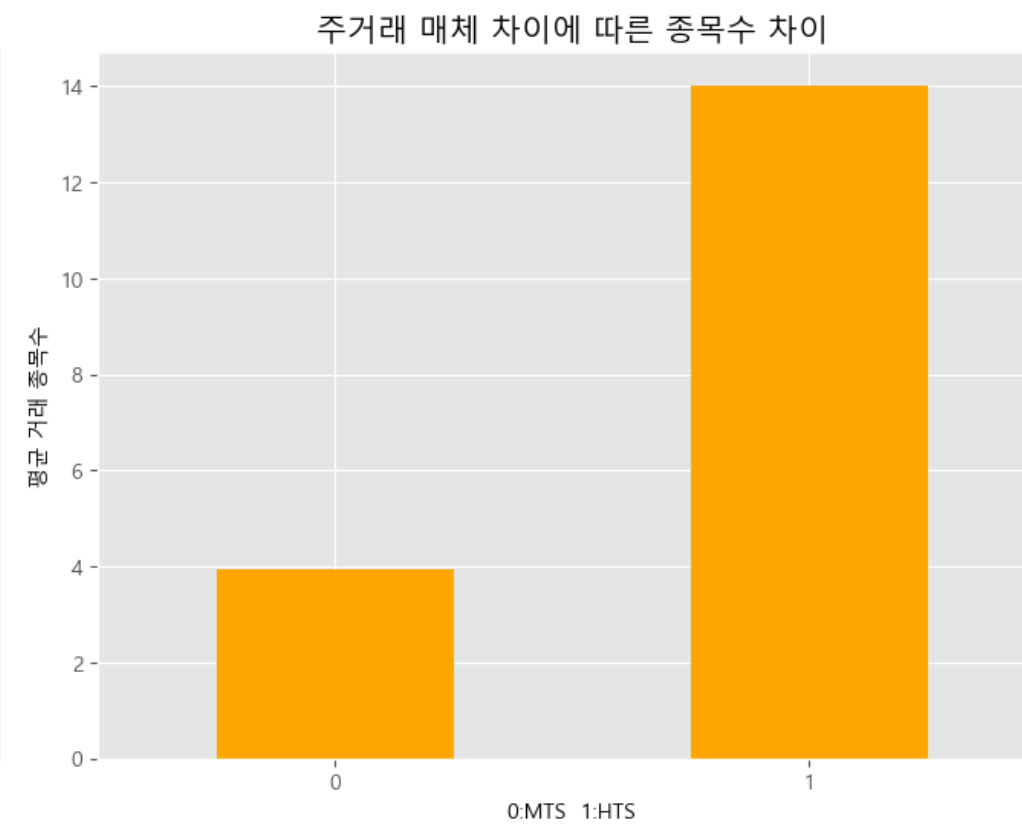
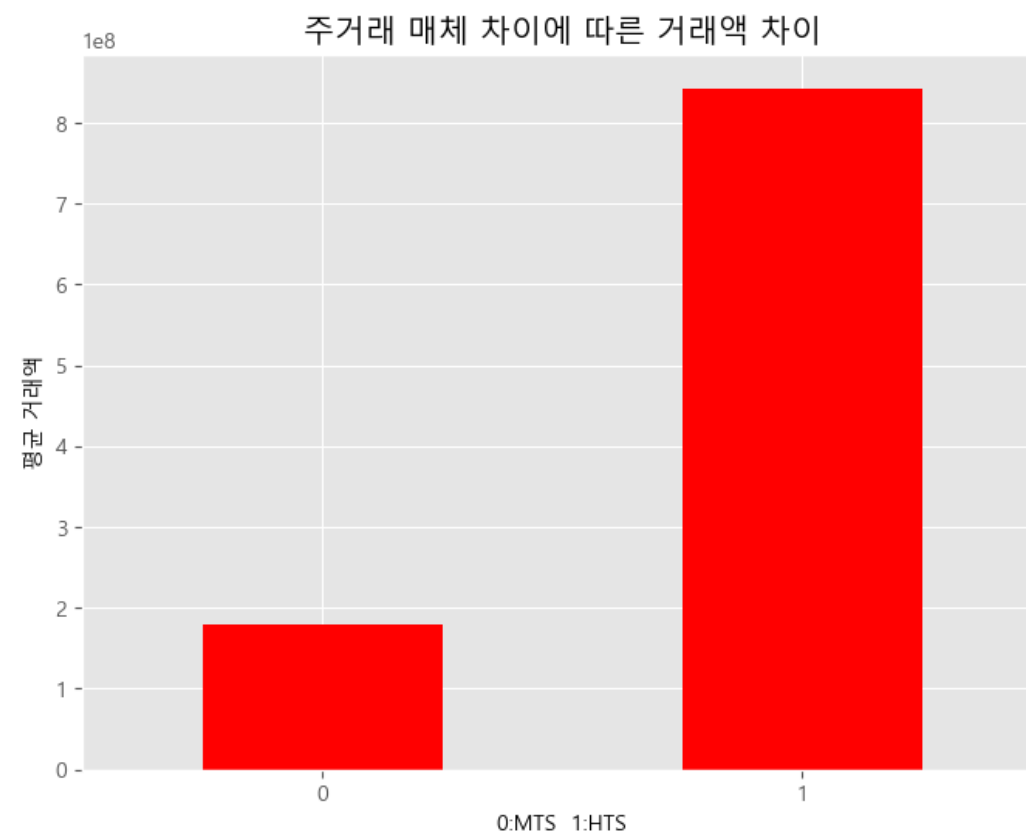
데이터 전처리 (주식관심도)



주식관심도 : 주식에 대한 고객의 관심을 수치화

거래 매체별 가중치 차등 이유

HTS가 주거래인 고객이 거래 횟수, 종목 수, 거래액 모두 높음
→ 주식에 대한 관심이 높다고 판단



데이터 전처리 (국내투자비율)



국내투자비율 : 전체 주식 평가액 대비 해외 주식 투자액의 비율

1. 결측치 데이터 -> 0으로 처리
2. 국내 주식 평가액 합 / 국내, 해외 주식 평가액 합
3. rank 함수를 사용하여 전체 고객 대비 본인의 위치를 비율로 계산
4. 거래기록이 없는 고객으로 인해 분모에 0.001 가산
5. 자본수준을 5개 등급으로 분류

```
custom['DMST_OVST_Ratio'] = (  
    (custom['DMST_AST_EVAL_M1'] + custom['DMST_AST_EVAL_M2'] + custom['DMST_AST_EVAL_M3'] + custom['DMST_AST_EVAL_M4']) /  
    (custom['DMST_AST_EVAL_M1'] + custom['DMST_AST_EVAL_M2'] + custom['DMST_AST_EVAL_M3'] + custom['DMST_AST_EVAL_M4'] +  
     custom['OVST_AST_EVAL_M1'] + custom['OVST_AST_EVAL_M2'] + custom['OVST_AST_EVAL_M3'] + custom['OVST_AST_EVAL_M4'] + 0.001))  
  
custom['DMST_OVST_percentile'] = custom['DMST_OVST_Ratio'].rank(pct=True)  
custom['DMST_OVST_degree'] = custom['DMST_OVST_percentile'].apply(degree) # 등급 매기기
```


데이터 전처리 (투자분산도)



투자분산도 : 전체 평가 금액 대비 상위 3개 종목의 비율

1. 결측치 데이터 -> 0으로 처리
2. 국내, 해외 주식 1,2,3위 평가액 합 / 국내, 해외 주식 평가액 합
3. 제공 데이터 처리 과정으로 인해 발생한 비율이 1이 넘는 값들을 1로 대체
4. rank 함수를 사용하여 전체 고객 대비 본인의 위치를 비율로 계산
5. 자본수준을 5개 등급으로 분류

```
# 1,2,3위 매수액 / 전체 매수액
# 등급이 높을수록 집중투자
custom['DISPERSION'] =
    ((custom['DMST_AST_PCHS_SUM_M1'] + custom['DMST_AST_PCHS_SUM_M2'] + custom['DMST_AST_PCHS_SUM_M3'] + custom['DMST_AST_PCHS_SUM_M4']
    custom['OVST_AST_PCHS_SUM_M1'] + custom['OVST_AST_PCHS_SUM_M2'] + custom['OVST_AST_PCHS_SUM_M3'] + custom['OVST_AST_PCHS_SUM_M4']))
    (custom['DMST_AST_PCHS_M1'] + custom['DMST_AST_PCHS_M2'] + custom['DMST_AST_PCHS_M3'] + custom['DMST_AST_PCHS_M4'] +
    custom['OVST_AST_PCHS_M1'] + custom['OVST_AST_PCHS_M2'] + custom['OVST_AST_PCHS_M3'] + custom['OVST_AST_PCHS_M4']))

# 집중도가 1보다 큰 값이 있어서 1 이상의 값을 1로 변환
custom.loc[custom['DISPERSION'] > 1, 'DISPERSION'] = 1
custom['DISPERSION_percentile'] = custom['DISPERSION'].rank(pct=True)
custom['DISPERSION_degree'] = custom['DISPERSION_percentile'].apply(degree) # 등급 매기기
```

데이터 전처리 (주식보유기간)

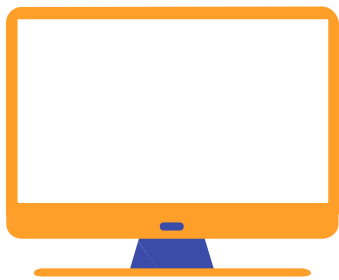


주식보유기간 : 주식 보유 기간을 수치화

1. 결측치 데이터 -> 0으로 처리
2. 매도액 중 기간별 매도액 비중을 가중 평균(-1 ~ 1이 되도록)
3. 당일, 5일 30일, 180일, 1년, 2년 이상을 각각 -0.6, -0.3, -0.1, 0.1, 0.3, 0.6으로 계산
4. rank 함수를 사용하여 전체 고객 대비 본인의 위치를 비율로 계산
5. 자본수준을 5개 등급으로 분류
6. 4달간 거래기록이 없는 고객은 6등급으로 분류

```
# 0.6, 0.3, 0.1 순으로 가중치  
# 높은 등급일수록 장기투자, 결측치는 0으로 처리  
  
custom['TIMELINE'] = (custom['DAY_TR_RATIO'] * -0.6 + custom['SWING_TR_RATIO'] * -0.3 + custom['MONTHS_TR_RATIO'] * -0.1 +  
    custom['MID_TR_RATIO'] * 0.1 + custom['HLD_TR_RATIO'] * 0.3 + custom['YEARS_TR_RATIO'] * 0.6)  
  
custom['TIMELINE_percentile'] = custom['TIMELINE'].rank(pct=True)  
custom['TIMELINE_degree'] = custom['TIMELINE_percentile'].apply(timeline) # 등급 매기기
```

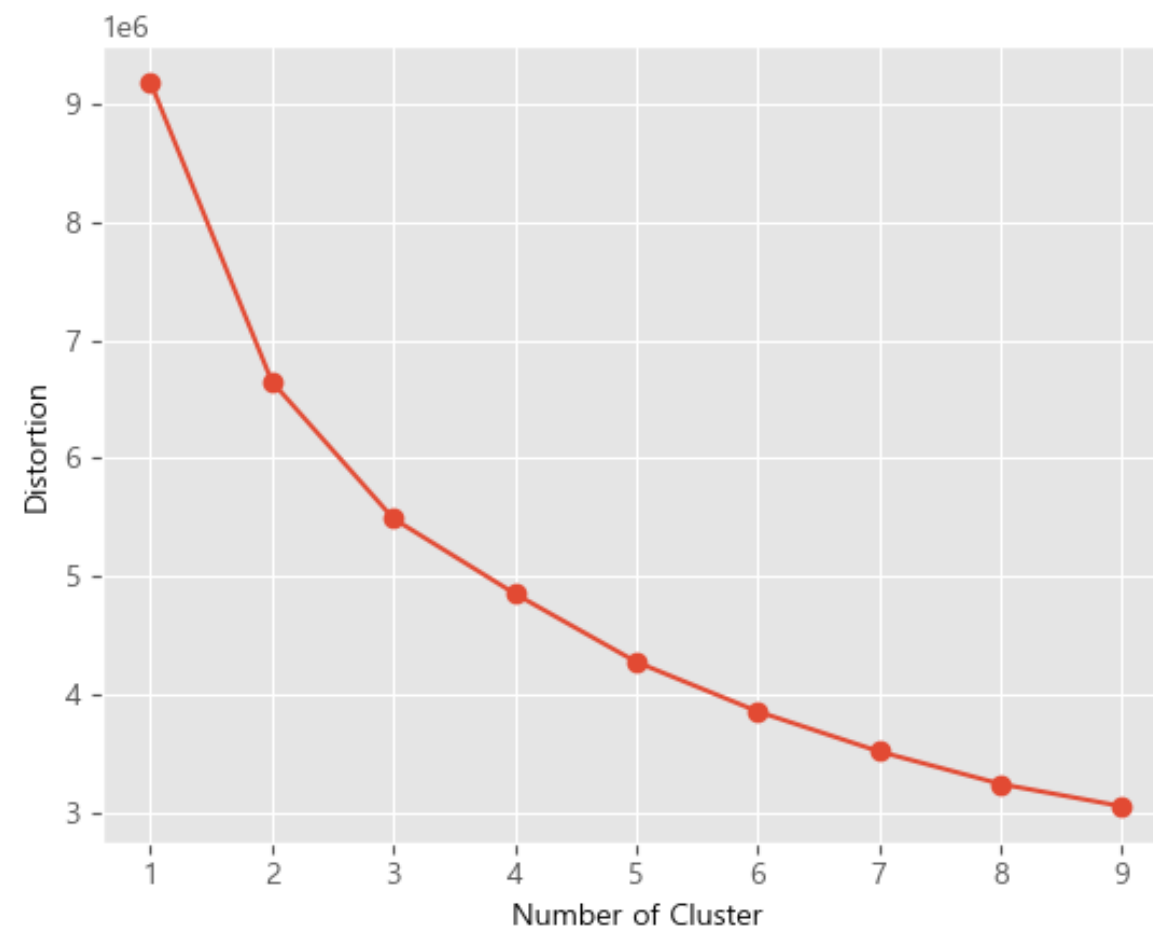
투자 성향 분석 (K-Means)



K-Means란?

K-Means 클러스터링 알고리즘은 n개의 중심점을 찍은 후에,
이 중심점에서 각 점간의 거리의 합이 가장 최소화가 되는 중심점 n의 위치를 찾고,
이 중심점에서 가까운 점들을 중심점을 기준으로 묶는 클러스터링 알고리즘

→ 거래 데이터를 기반으로 수치화한 5개 항목을 기준으로 고객의 투자 성향을 파악

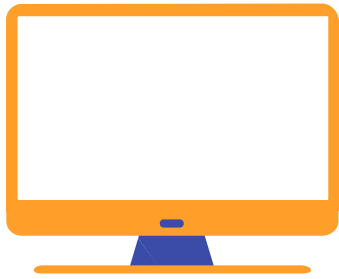


Elbow-Method

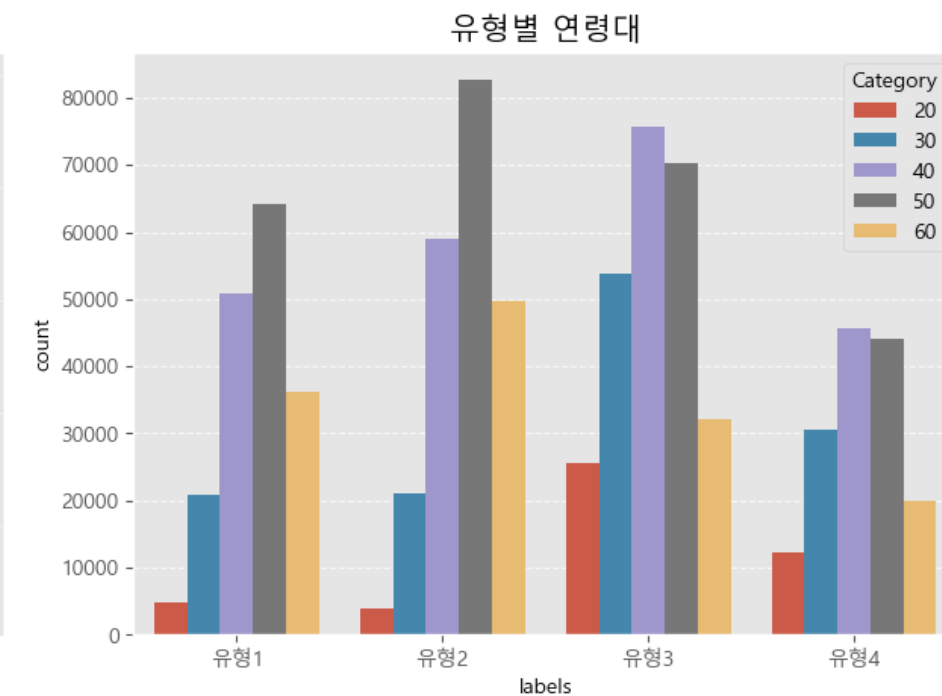
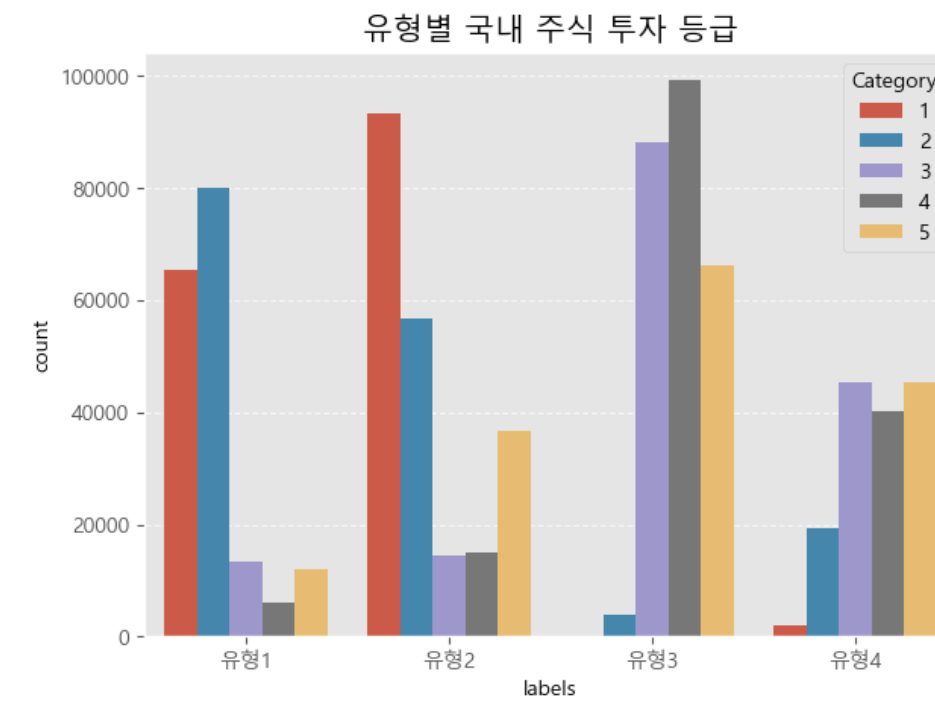
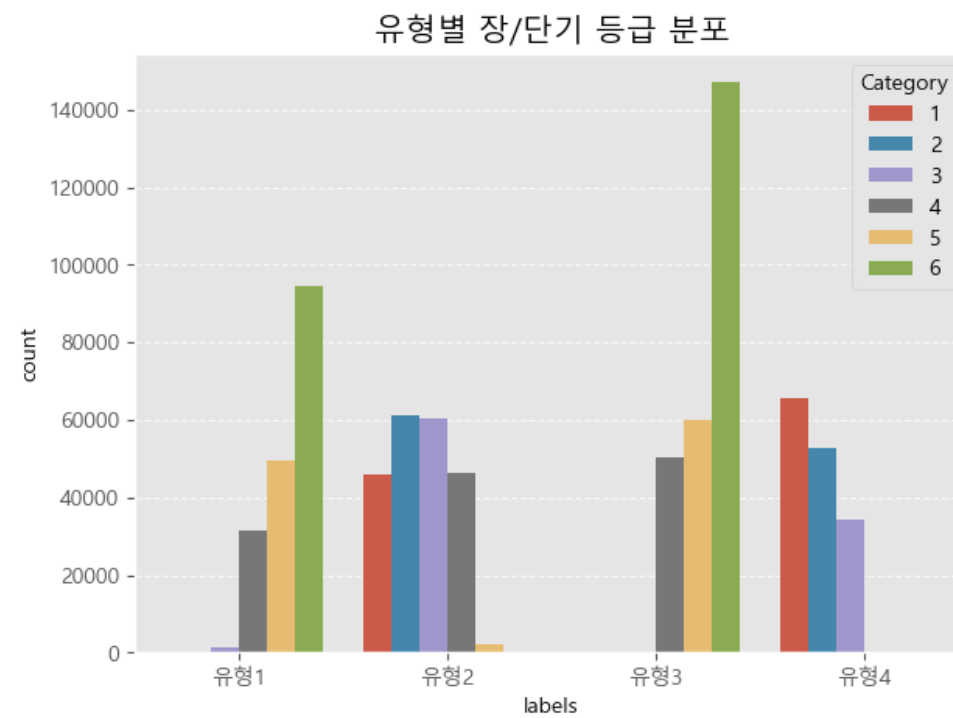
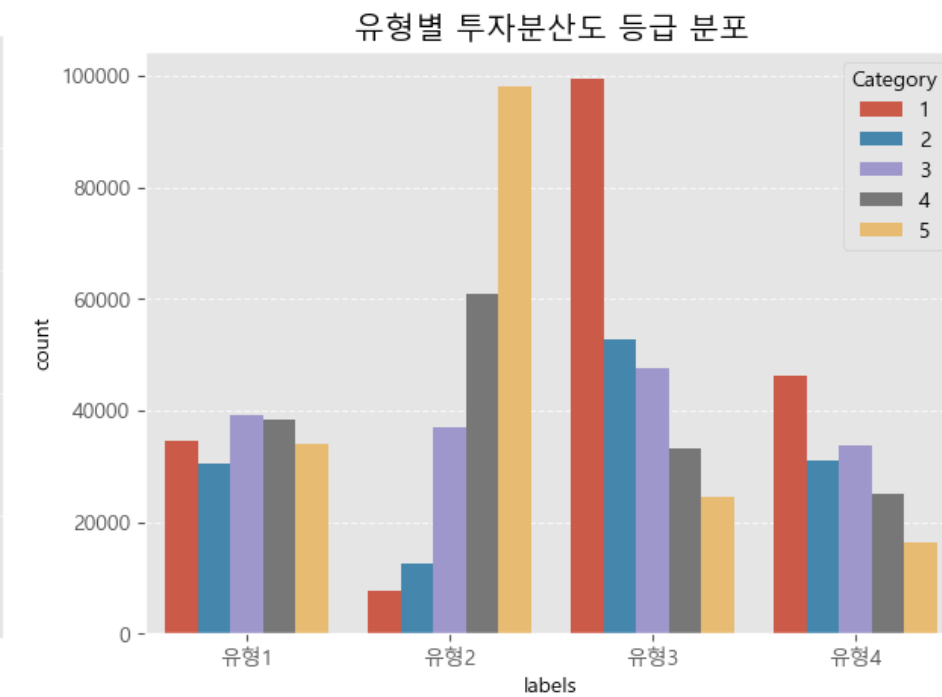
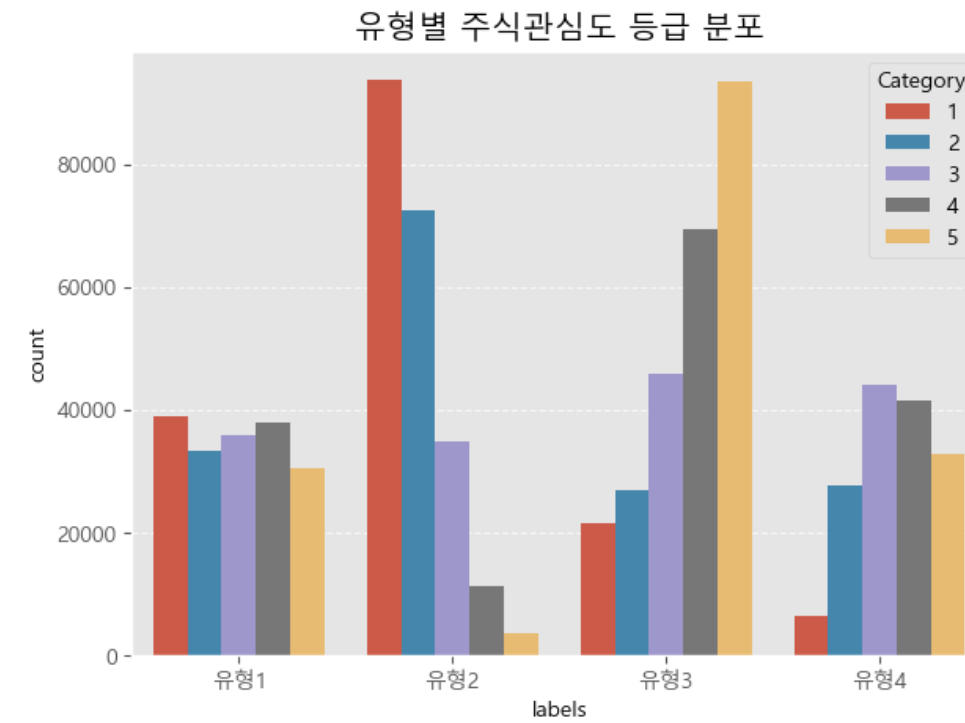
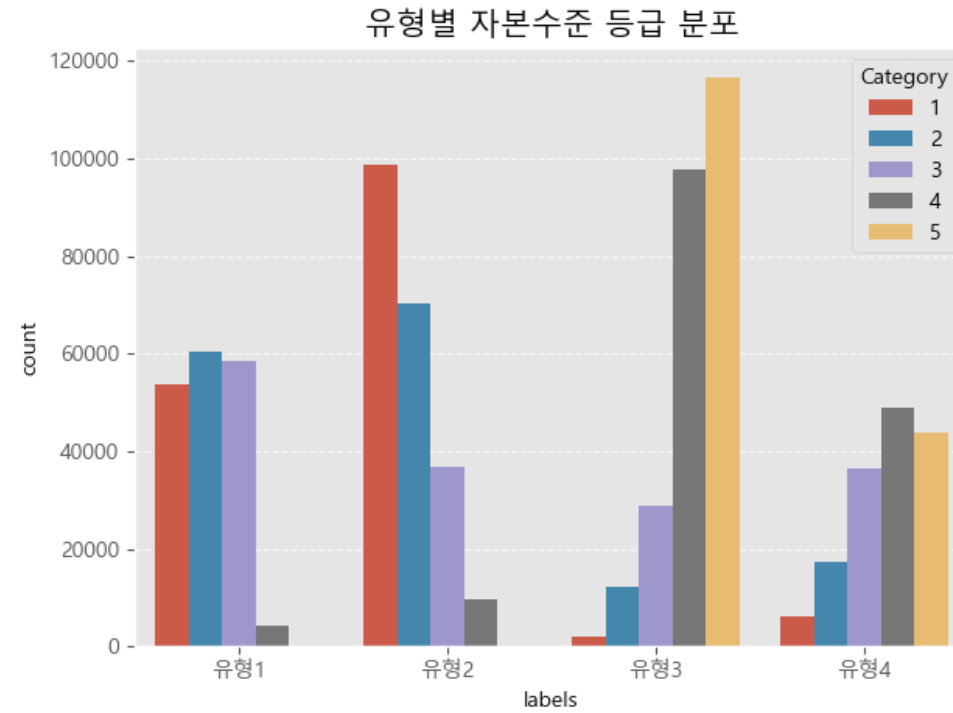
Elbow-Method는 군집의 수를 결정하기 위한 기법으로
SSE가 급격히 감소하는 지점인 elbow point를 기준으로
최적의 k 값을 찾아내는 알고리즘

시각화 결과 군집 갯수가 4~5개일 때 최적
→ 분석에 4개 군집 사용

투자 성향 분석 (K-Means)



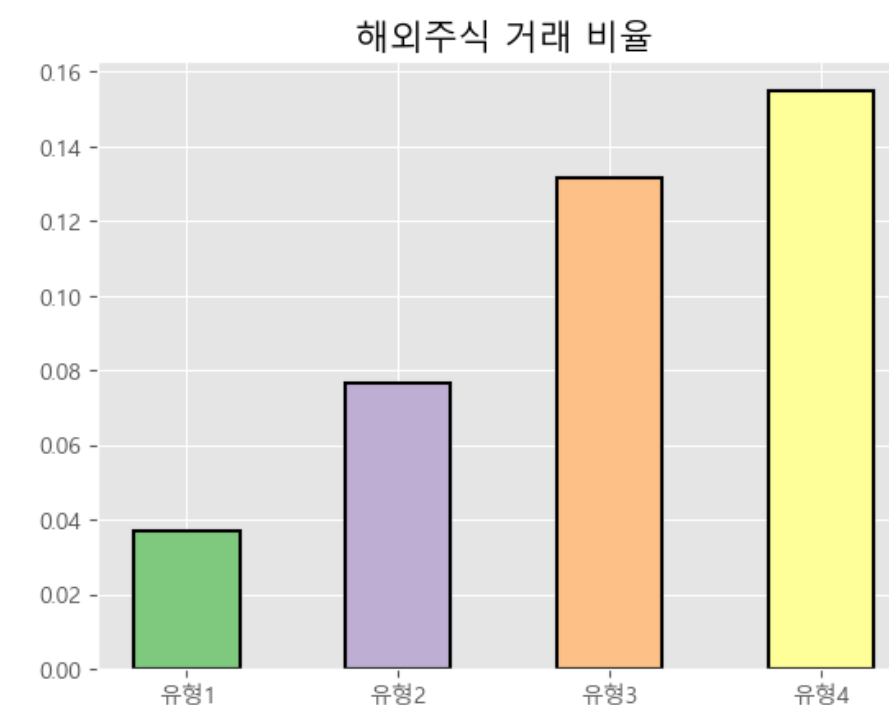
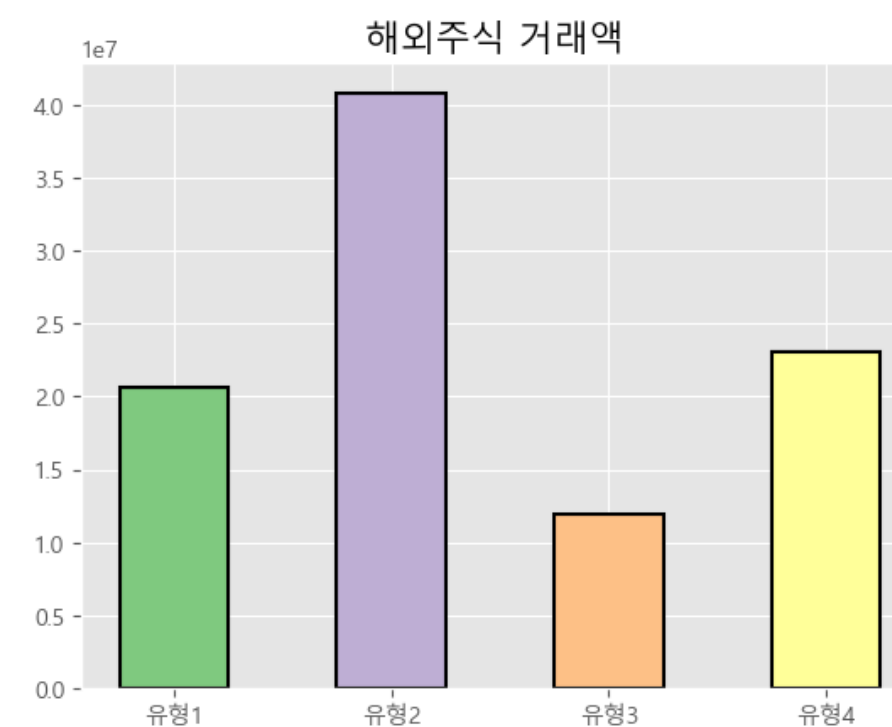
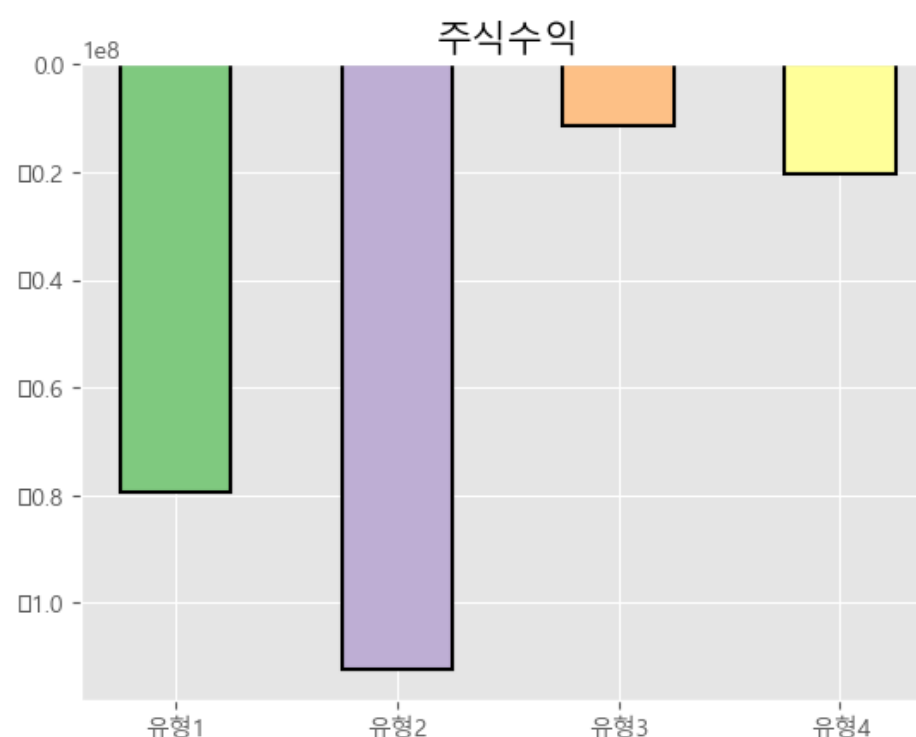
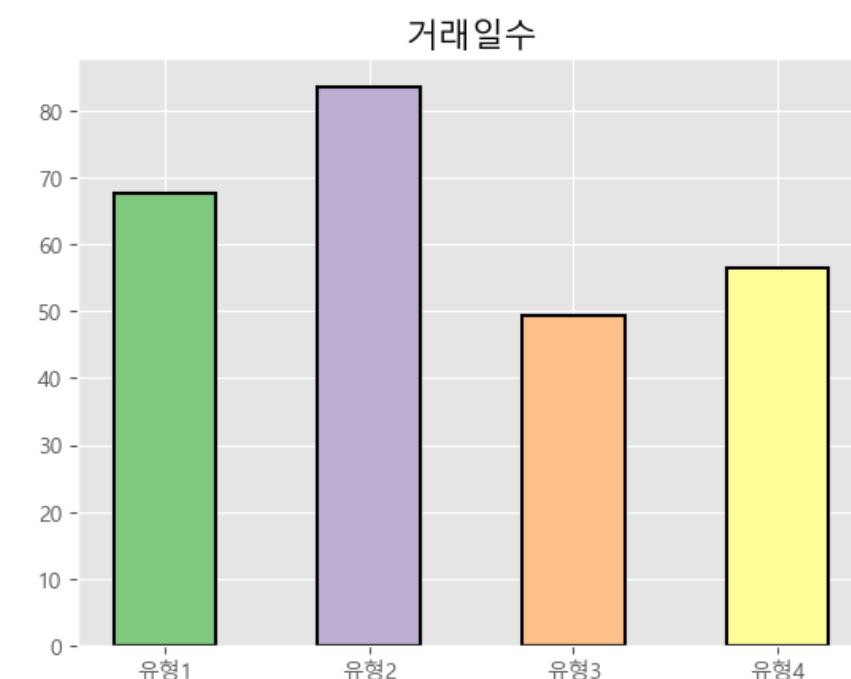
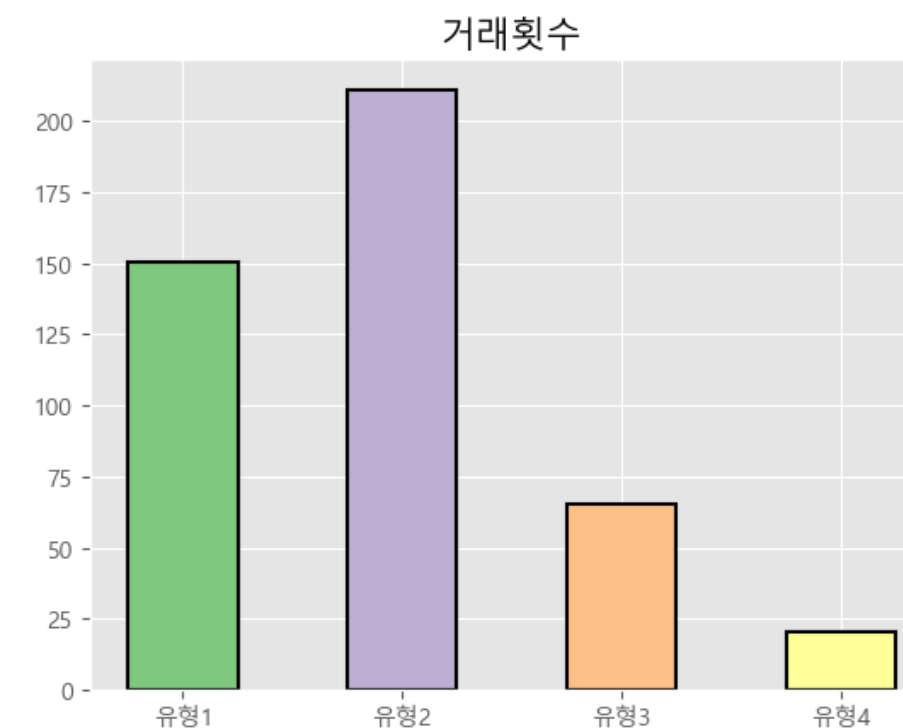
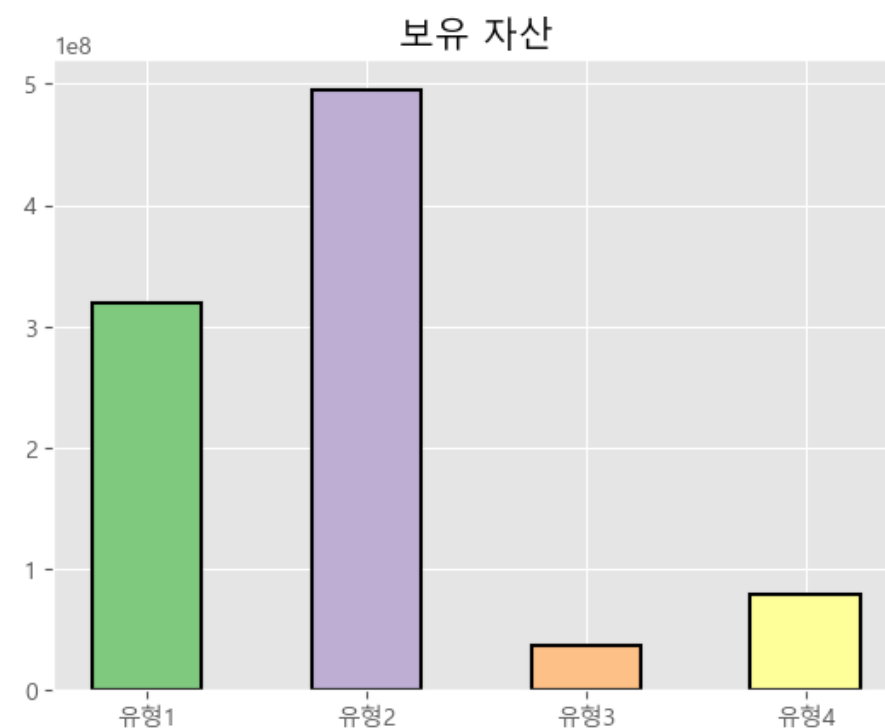
유형별 거래 성향 분석



투자 성향 분석 (K-Means)



유형별 거래 성향 분석



유형별 고객 선행 분석

유형 1: 경제 수준 평균 이상, 주식에 대한 관심도 높으나 해외 주식에 대한 투자가 부족

콘텐츠 기반 추천모델로 관심 산업의 해외 종목 추천

→ 해외 투자 확대와 수익 상승 기대

유형 2: 경제 수준도 높고 주식에 대한 관심도 높으나 수익률이 매우 낮음

→ 협업 필터링 추천모델을 통해 높은 수익률을 기록한 고객들의 종목 추천

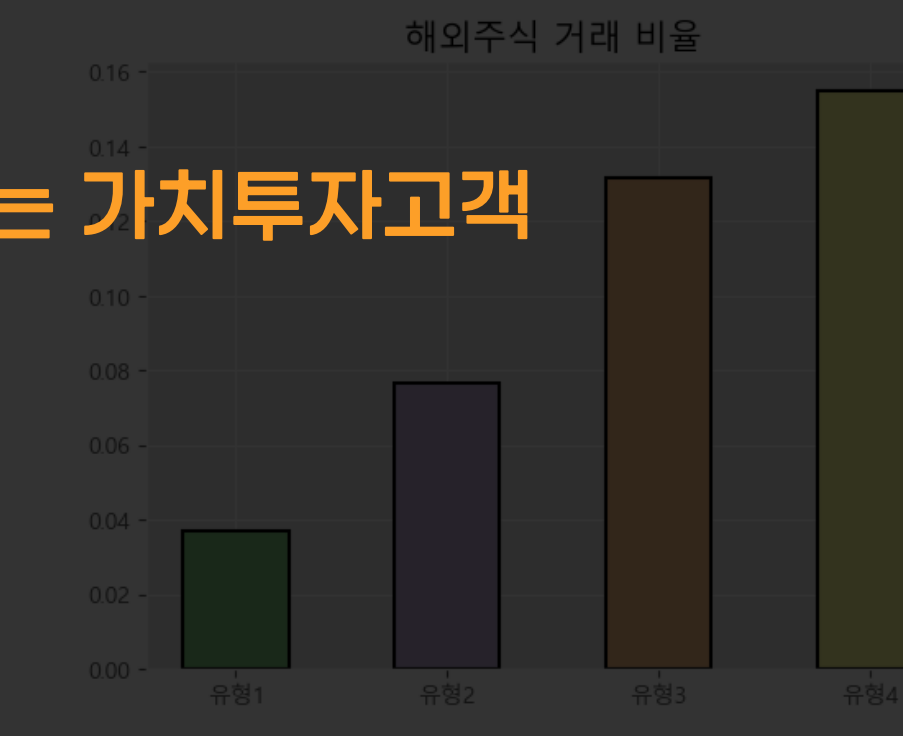
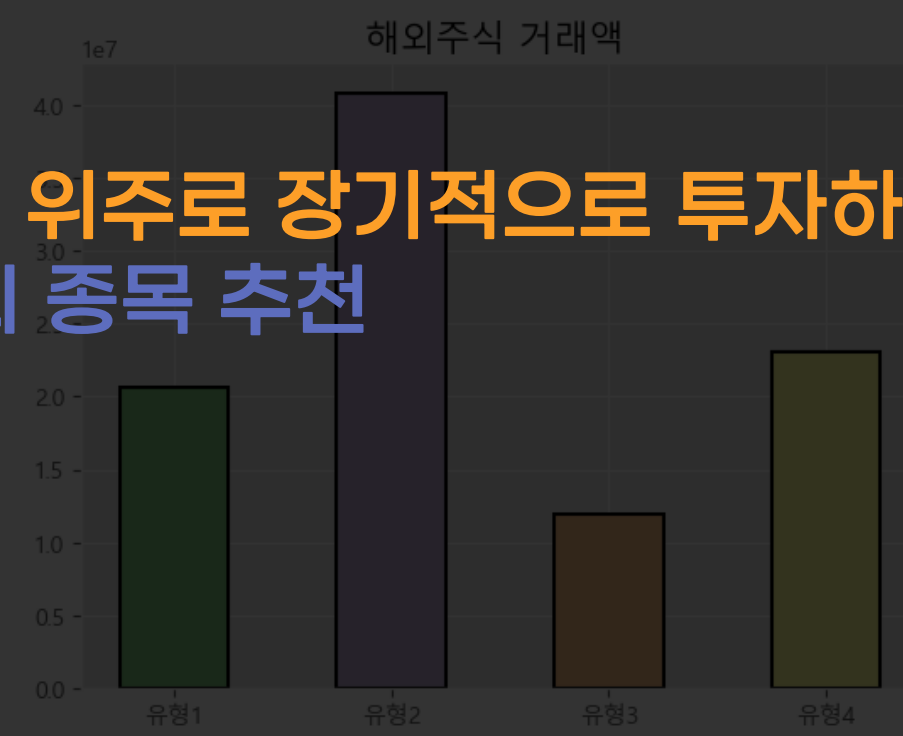
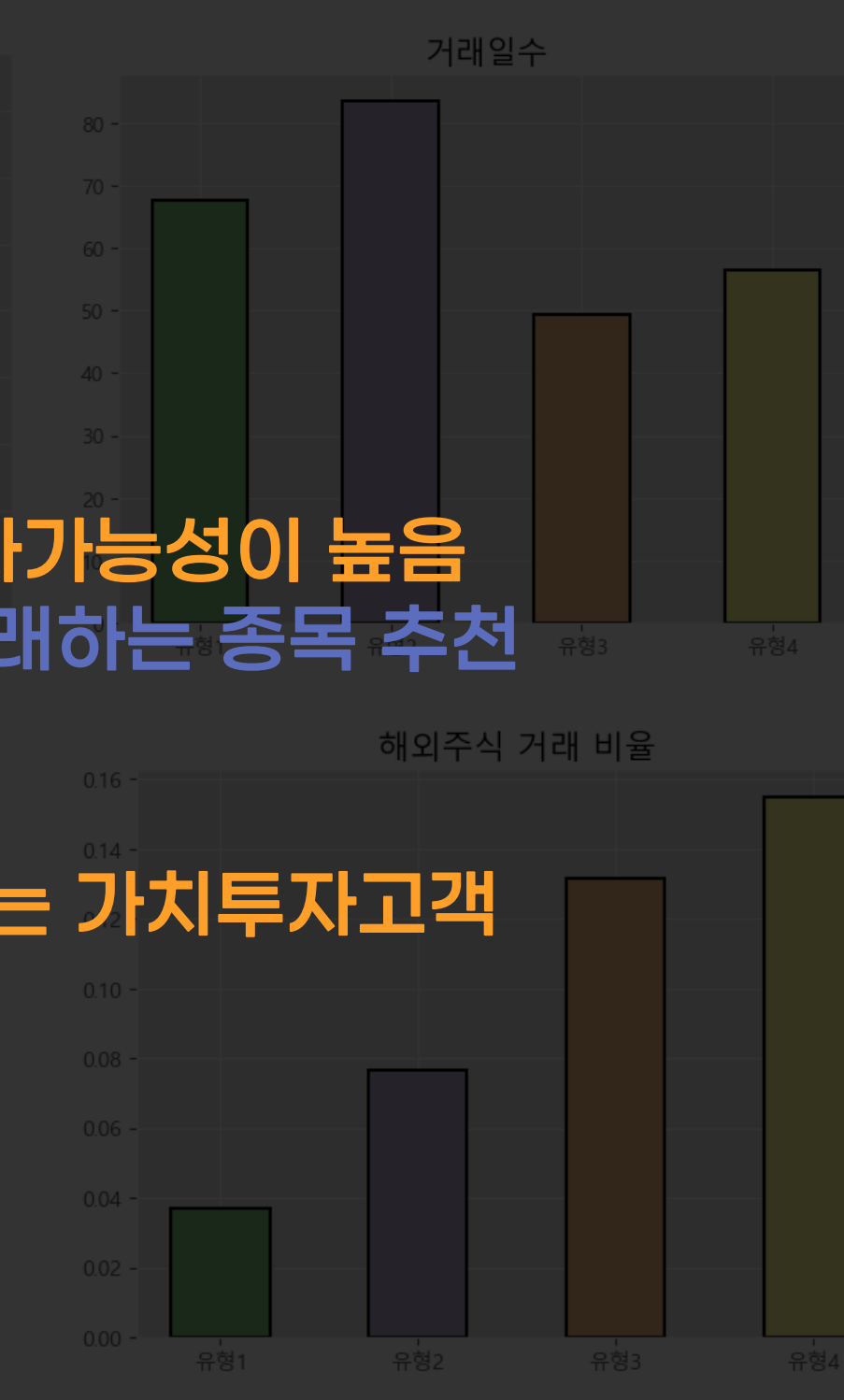
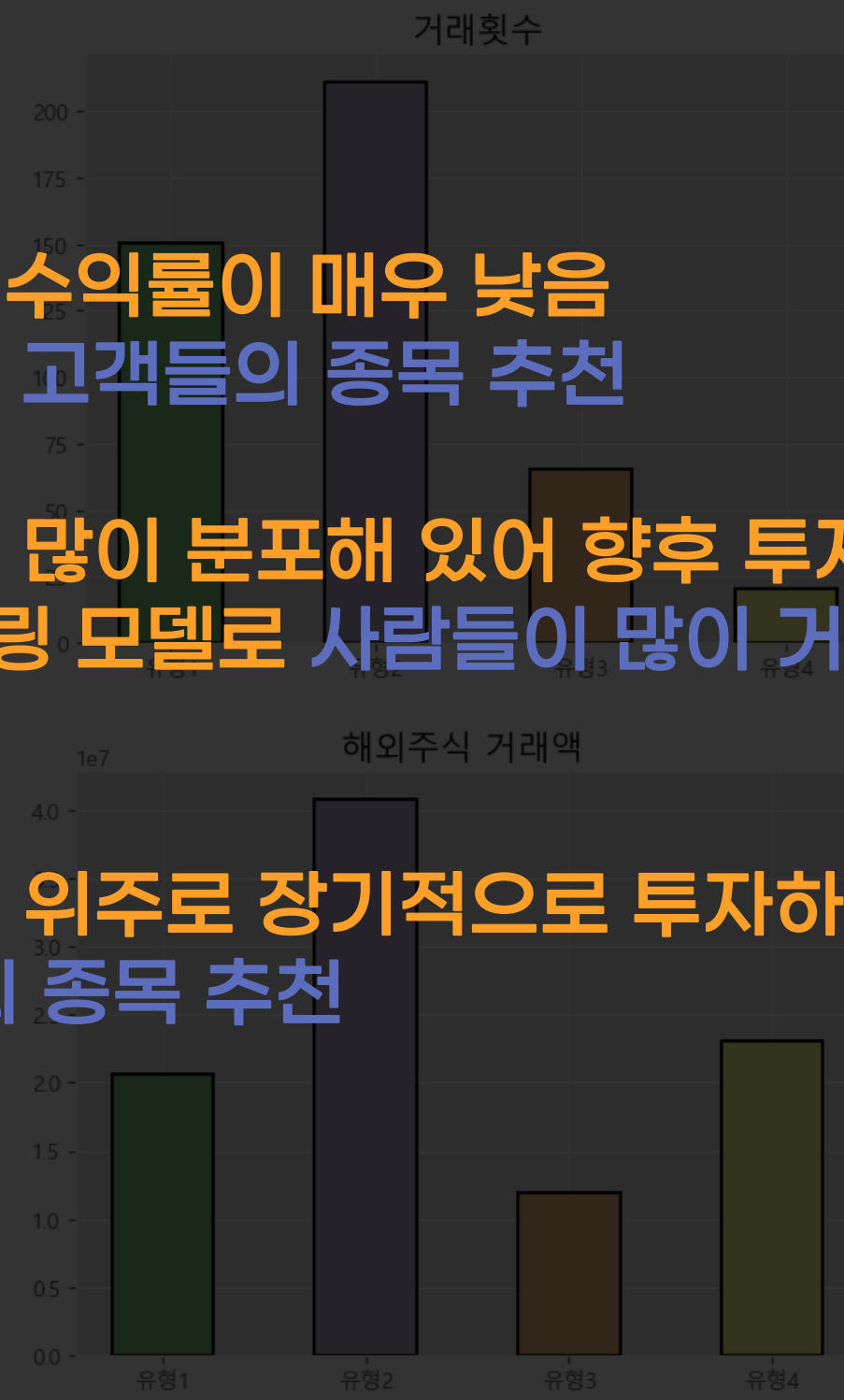
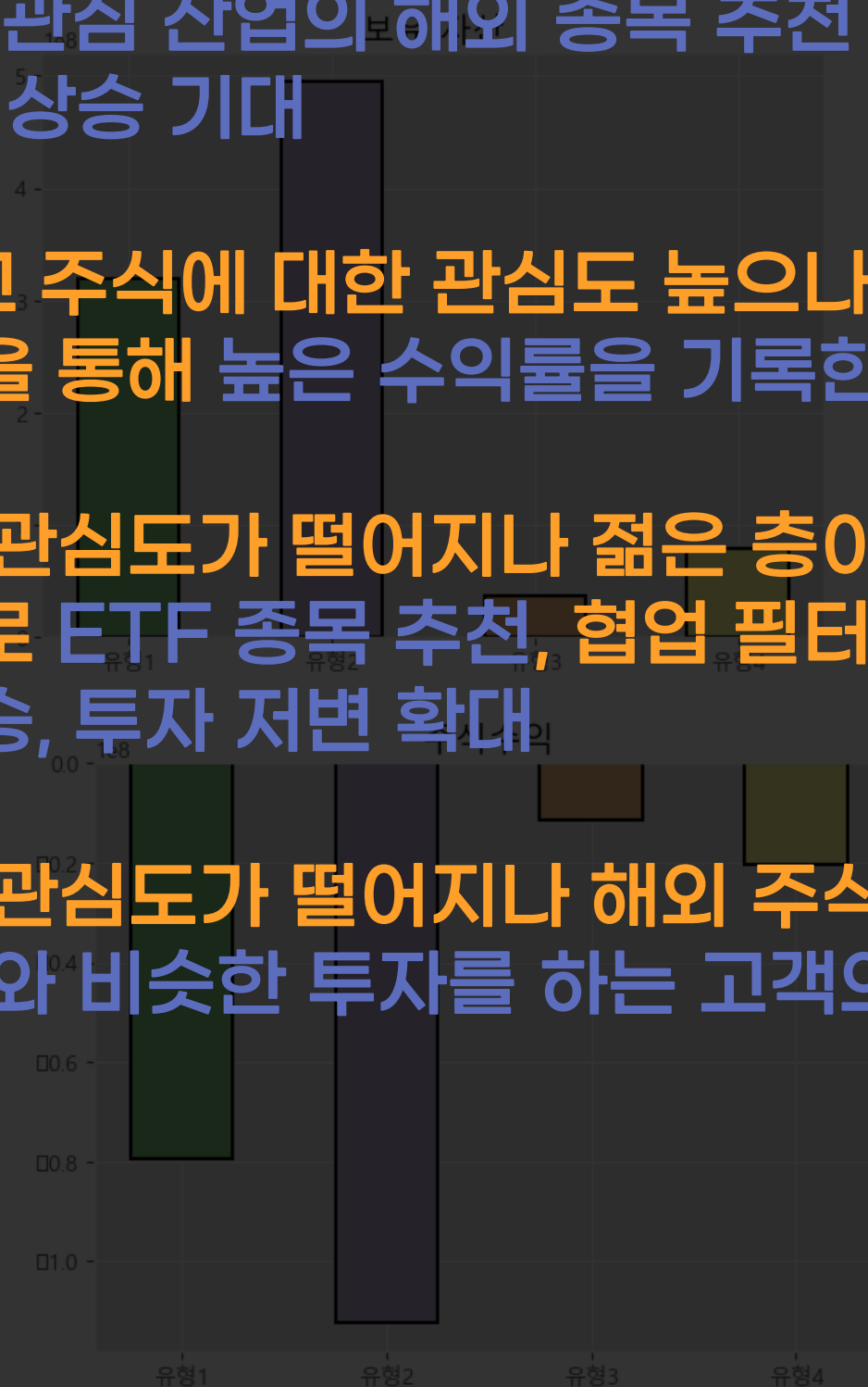
유형 3: 경제 수준, 주식 관심도가 떨어지나 젊은 층이 많이 분포해 있어 향후 투자가능성이 높음

→ 콘텐츠 기반 추천모델로 ETF 종목 추천, 협업 필터링 모델로 사람들이 많이 거래하는 종목 추천

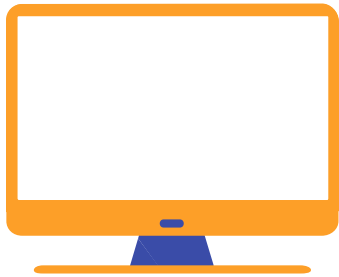
→ 주식에 대한 관심도상승, 투자 저변 확대

유형 4: 경제 수준, 주식 관심도가 떨어지나 해외 주식 위주로 장기적으로 투자하는 가치투자고객

→ 협업 필터링 모델로 나와 비슷한 투자를 하는 고객의 종목 추천



사용 예정 모델 (추천시스템)



콘텐츠 기반 필터링

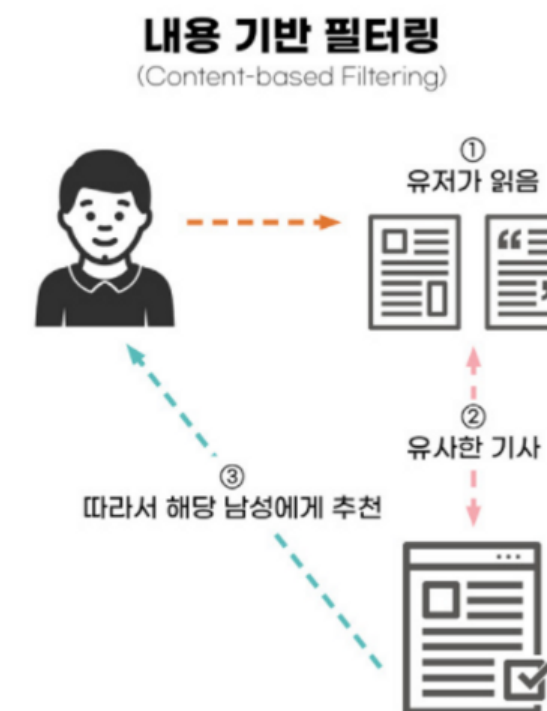
콘텐츠만을 활용하여 추천해주는 알고리즘

사용자가 좋아하는 콘텐츠를 분석하여 그와 유사한 콘텐츠를 추천해주는 기술

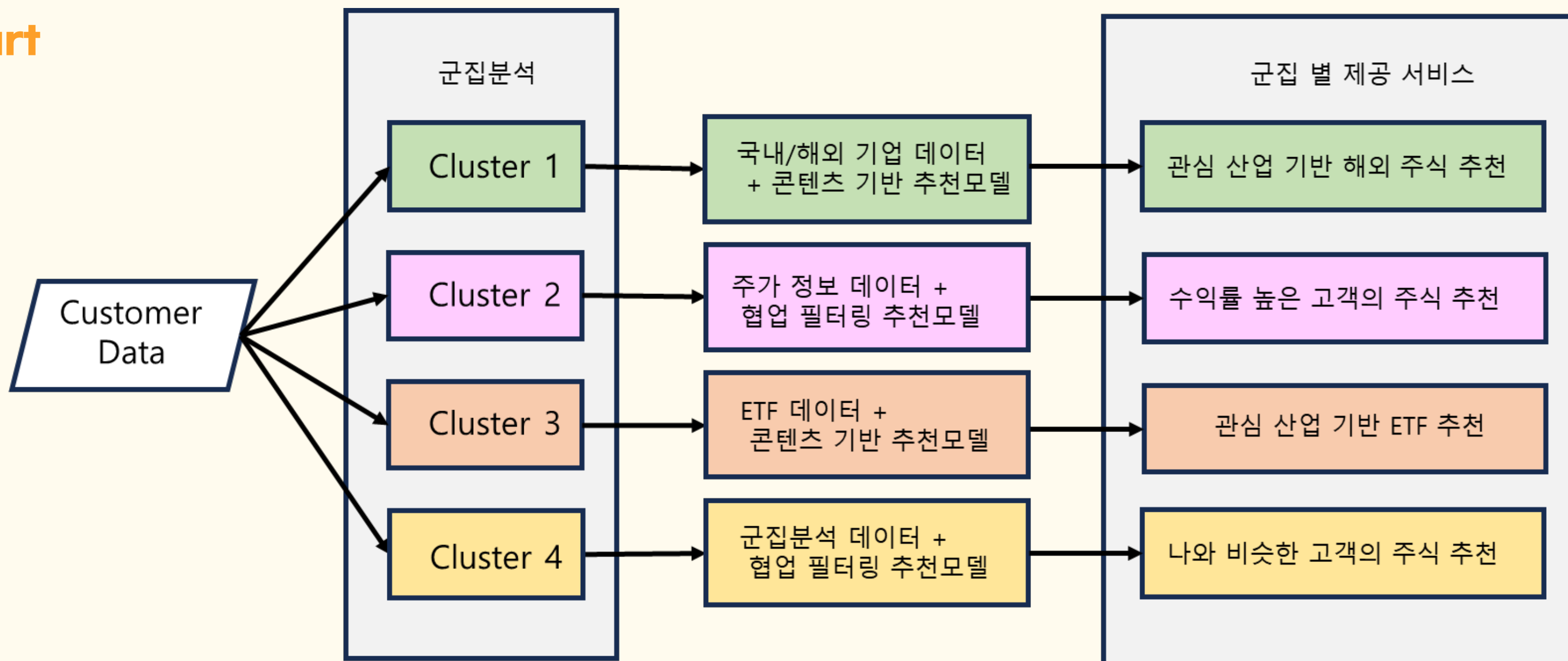
협업 필터링

다른 사용자들로부터 취향 정보들을 모아 사용자의 관심사를 예측하는 알고리즘

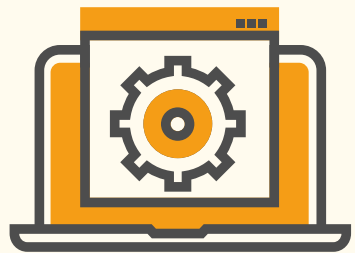
협업 필터링 모델은 비슷한 취향을 가진 사용자들은 어떠한 아이템에 대해 비슷한 선호도를 가질 것이라는 가정 하에 사용자와 아이템 간 상호 작용 데이터를 활용



알고리즘 Flow Chart



서비스 설명



앞선 알고리즘을 통해

1. 국내주식거래만 하던 고객에게는
기존에 거래하던 산업과 **비슷한 해외 주식 종목**을 추천,
해외 주식 거래를 이용하던 고객에게는 더 **다양한 해외 종목**들을 추천
2. 단일주식거래만 하던 고객에게 **관심있는 테마의 ETF** 추천
3. 주식 고수와 유사 성향 고객이 선택한 **협업 필터링 모델 기반 종목** 추천

하여 고객들의 해외자산 투자를 늘리는 것은 물론, 관심도를 높혀 투자 저변 확대를 유도할 수 있습니다.