

투자패턴에서 미래자산을 찾다

데이터 분석을 통한
일반금융소비자의 행태별 맞춤 팩터 탐색과
은퇴 자금 목표 달성을 위한 투자 전략 평가

After Learner

강수연
장현진

 NAVER
CLOUD PLATFORM

 MIRAE ASSET
미래에셋증권

1 기획의도

당신의 발자취를 찾아 **자산배분**을 해결하는 명탐정 **팩터**
당신의 **포트폴리오**를 끝까지 책임지는 동반자 **리밸런싱**

2 시계열 데이터 분석

5가지 주가지수 데이터에서 주목할 이벤트 시점

2022.12.01-2023.04.30 중 이벤트 시점 추출

(1) 8가지 이벤트 시점별 A종목 뉴스 데이터의 **연관규칙 분석**

(2) 연관규칙과 금융소비자 매수[매도] 종목 데이터 간 **네트워크 시각화**

3 서비스 기획

투자자 행태와 관련한 **팩터** 설정 및 발굴
팩터 구성·**생애주기**·**자산규모**를 고려한 투자 전략 설정

1

기획의도

당신의 발자취를 찾아 자산배분을 해결하는 명탐정 팩터



- 코로나 19 발생 이후 크게 증가한 **개인투자자** 비중
- 그러나 신규 투자자 중 60%(2020년 기준)는 경험하는 **투자손실**
- 가장 많이 증가하는 투자 인구는 20대 · 30대 · 40대
- 개인투자자의 연간 거래회전을 **1,600%**
- 단순 증권 도메인 지식에 대한 부족한 이해이라기에는 실패의 **원인**을 찾아봐야 할 때!

미래에셋 금융소비자 데이터,
지수 데이터, 뉴스 데이터를
이용하여 주식성과에
영향을 미칠 수 있는
투자자 행태를 발견

행태적 특성과 관련이 있는
팩터를 찾아 투자자의
비합리적인 투자행태로
인한 포트폴리오 개선

당신의 포트폴리오를 끝까지 책임지는 동반자 리밸런싱



- 세계에서 가장 빠른 **고령화** 비율, 노인빈곤율 1위
- 퇴직 후 필요한 **노후자금** 최소 6억 이상
- 성공적인 은퇴 생활을 보장하는 시기는 은퇴 후 10년
- 은퇴 후에도 같은 생활 수준을 유지하기 위한 **금퇴** 준비
- 나이대별, 자산별 다양한 **요인**을 함께 고려하는 투자 전략 필요

수익을 얻는 시기가
전체 은퇴시기의 영향을 미치는
시퀀스 리스크 완화를 위해
멀티팩터 선정 시
생애주기별, 자산규모별
투자전략 수립

일반 퇴직자보다
일찍 노후를 대비할수록
투자 금융자산을 적극적으로 활용할수록
금퇴확률이 증가하므로
투자자들의 포트폴리오 구성 단계에서
개인에게 맞는 은퇴준비가 필요

2

시계열 데이터 분석

시계열 지수 데이터 전처리와 차원 축소

증권시장별 휴장일 차이에 따른 결측값 처리



지수 시계열 데이터의 노이즈 탐색



시간의 흐름에 따른 지수 데이터
누적으로 인한 추세와 계절성

1차 차분하여 지수의 평균변화를
일정하도록 전처리



5가지 주가지수의 1차원 축소 끝에
생성한 새로운 시계열

시간정보 세분화과정에서의 시행착오

■ 5-NN 결측값 보간법

최근접한 5개의 이웃 지수 데이터의 평균값으로
비어있는 값을 채움

■ 정상성 검정

관측된 시간과 지수 데이터의 특징이 관련되어
있다는 문제를 차분으로 완화

■ 시계열 오차의 자기상관성 검정

1일, 5일, 10일, 50일 전 시점이 현재 시점의 지수
데이터에 영향을 미친다는 문제를 차분으로 완화

■ 주성분분석

원본 데이터의 정보[누적분산 기여율]을 **약 92.44%**
보존하면서 5차원 지수 데이터를 1차원 선형 결합

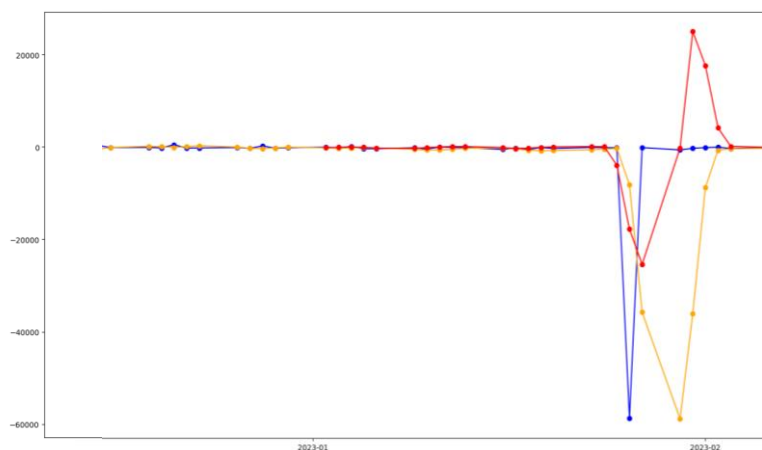
DOW NASDAQ S&P500 KOSPI KOSDAQ

5가지 주가지수 데이터에서 주목할 이벤트 시점

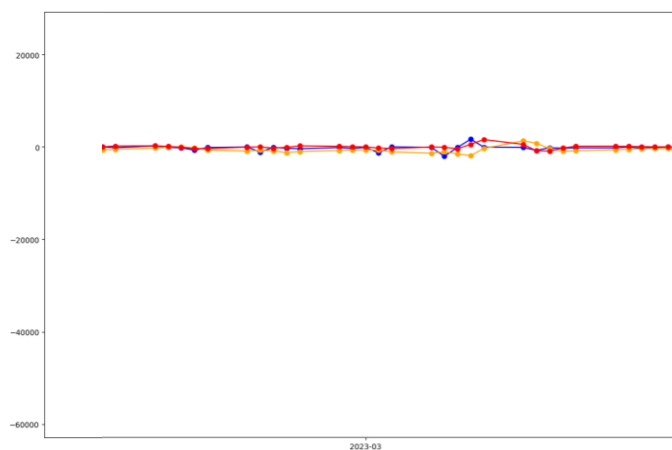
금융소비자 데이터와 병합하였을 때 의사결정에 극명한 변화가 드러나는 일자를 알아보기 위해
지수의 변동이 크게 나타나는 지점들을 정해
Dramatic/Stable(변동 크기에 따라*), Change/Not Change(변동의 증감에 따라**)로 날짜 지정

*) 2계 미분계수 절댓값의 상위 10% 여부

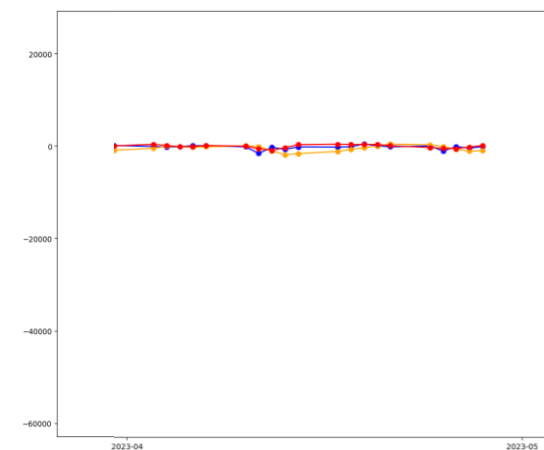
**) 기존 단위 변동률 함수에서 첨점에 해당하는 구간 여부



>



>



2023-01-25

2023-01-26

2023-01-27

2023-01-31



2023-02-01

2023-02-02



2023-03-10



2023-04-12



가장 극명한 변화를 보이는 10%를 지정하여 8개의 시점 추출

연관규칙 분석

- AI종목뉴스의 종목 데이터와 뉴스태그 데이터가 함께 등장하는 빈도 확인
- 어떤 종목과 뉴스태그가 연관성이 있는지 규칙 탐색
- 금융소비자의 매수[매도]종목 데이터와 비교하기 위한 네트워크 시각화 수행

이벤트 시점	주가지수의 변동	연관규칙 있는 빈발 뉴스태그 집합
2023-01-25 ~ 2023-01-27	지수의 변동폭이 크고 변동의 증감률이 적은 시점	IT: AI, 인공지능 경제 : 은행, 금리, 대출
2023-01-31 ~ 2023-02-02	지수의 변동폭이 크면서 변동의 증감률이 큰 시점	
2023-03-10	지수의 변동폭이 크고 변동의 증감률이 적은 시점	국제: 이머징 마켓, 아시아, 미국 경제 : 채권, 금리
2023-04-12	지수의 변동폭이 크고 변동의 증감률이 적은 시점	국제: 미국, 아시아, 이머징 마켓 사회 : 미세먼지, 황사, 중국

연관규칙과 금융소비자 매수[매도]종목 데이터 간 네트워크 시각화

네트워크 시각화[이분 그래프]

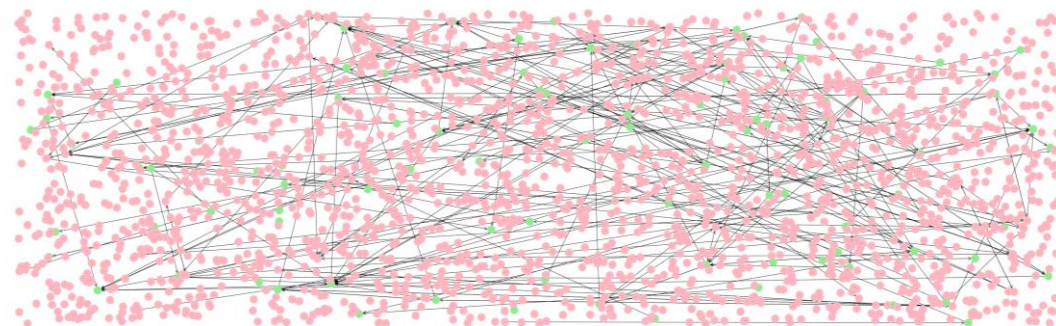
- 가정
시뉴스가 소비자 매수[매도]종목에 영향을 미침
- 점의 구분[노드]
각 네트워크의 점에 색상 차이를 두어 소비자의 매수[매도]종목과 뉴스종목

카잔알고리즘

- 강한 연관관계를 찾는 알고리즘을 활용하여
투자의사결정에서 소비자의 시종목뉴스
의존도를 추정

- 선의 방향[엣지]
소비자 거래종목에서 연관규칙 분석결과의 뉴스종목,
뉴스종목에서 소비자 거래종목으로 중복될 때

뉴스태그와 뉴스종목 사이의 선은 얼마나 영향을
미치는지 따라 큰 쪽에서 작은 쪽으로 향하게 함



- 시뉴스데이터 = 연두색
소비자의 매수[매도]종목 = 연분홍색
- 그래프의 방향과 역방향으로 갈 수 있는 경로를 탐색
할 때 중복되는 경로에서 나타나는 점을 분석

금융소비자의 투자심리지수 컬럼 산출

“

[illegible]

금융소비자의 투자심리지수 파생변수 생성

1. 전환율 | 거래회전율 | 주식수익률

전환율

- 매수[매도] 전환율

$$= \frac{\text{특정기간 동안의 주식 매수[매도]체결횟수}}{\text{특정기간 동안의 HTS 또는 MTS 접속일수}}$$

**트레이딩 시스템 유입빈도
대비 실제 주식매매거래
체결 여부 확인**

거래회전율

- 주식매수[매도]대금 회전율

$$= \frac{\text{3개월간 주식 매수[매도]액 합계}}{\text{특정 기간 동안의 주식 매수[매도]액 합계}}$$

- 주식매수[매도]거래 회전율

$$= \frac{\text{3개월간 주식 매수[매도]종목 수 합계}}{\text{특정 기간 동안의 주식 매수[매도]종목 수 합계}}$$

주식수익률

- 국내[해외]주식 수익률
- $$= \frac{\text{월말일 기준 국내주식 자산평가액}}{\text{월말일 기준 국내주식 자산매입액}}$$

매수매도 불균형

- 매수-매도 불균형(IBSI)
- $$= \frac{\text{특정기간 동안의 주식 매수액 합계}}{\text{특정기간 동안의 주식 매도액 합계}}$$

금융소비자의 투자심리지수 파생변수 생성

2. 분산투자지표 | 과거 최고자산 대비 현재 시점의 지표

과거 최고자산 대비 현재 시점의 지표

- 과거 최고자산 대비 등락률

$$= \left(\frac{\text{월말일 기준 예수금 등 현금성 자산액} + \text{월말일 기준 국내[해외]주식 자산 평가액} - \text{과거 최고자산액}}{\text{과거 최고자산액}} \right) \times 100$$

- 과거 최고자산 대비 변동기간

현재 총자산과
이전 최고 자산액을 비교



자산이 증가(감소)한 정도와
그렇게 되기까지의 기간을 확인

분산투자지표

- 분산투자지표_매수[매도]기준

$$\left(\frac{\begin{array}{l} \text{특정기간 동안의} \\ \text{주식매수[매도]액 1위} \\ \text{종목매수[매도]액 합계} \end{array} + \begin{array}{l} \text{주식매수[매도]액 2위} \\ \text{종목매수[매도]액 합계} \end{array} + \begin{array}{l} \text{주식매수[매도]액 3위} \\ \text{종목매수[매도]액 합계} \end{array}}{\text{특정기간 동안의 주식매수[매도]액 합계}} \right)$$

투자자의 매수[매도] 1,2,3위
종목 매매거래액의 비율이

전체 주식 매수[매도]액에서
차지하는 비중을 확인



투자 종목이 분산되어
있는 정도 확인

금융소비자의 투자심리지수 파생변수 생성

3. 대체입출고비중 | 이전 6개월 매도액 중 매수 후 특정 기간 이내의 매도액 비중

진입률[이탈률]

- 유가증권대체입고[출고]비중

$$= \frac{\left(\begin{array}{c} \text{특정기간 동안의} \\ \text{주식입고[출고]액 합계} \end{array} + \begin{array}{c} \text{특정기간 동안의} \\ \text{주식입금[출금]액 합계} \end{array} \right)}{\left(\begin{array}{c} \text{월말일 기준 예수금 등} \\ \text{현금성 자산액} \end{array} + \begin{array}{c} \text{월말일 기준 국내[해외]} \\ \text{주식 자산평가액} \end{array} \right)}$$

타 증권사로부터 순입금고액이
얼마나 들어오거나 나갔는지 확인



특정 기준을 정하여 분석할 때
고객의 서비스 진입[이탈] 여부 추정

이전 6개월 매도액 중 매수 후 특정 기간 이내의 매도액 비중

- 이전6개월 매도액 중 당일매수매도발생 매도액 비중
- 이전6개월 매도액 중 매수 후 5일, 30일, 180일,
1년 이내 매도액 비중

주성분분석을 통해
투자자의 매수 후 매도가 얼마나 빈번하게
이루어지는지 금액의 비중[가중치]을 채택

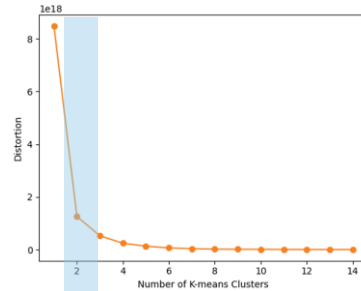


개별 투자자의 거래빈도를 확인

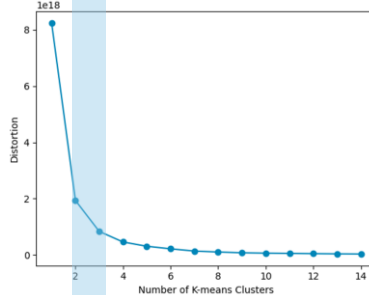
3가지 기간별 금융소비자 군집 세분화

새로운 투자심리지수 컬럼에 기존 CS_DATA의
연령대, 관리사원 존재 여부를 추가하여 3개의 분석용 데이터프레임 생성

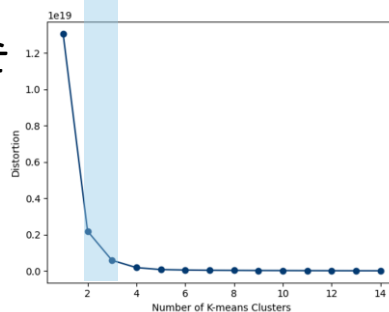
1월 하순



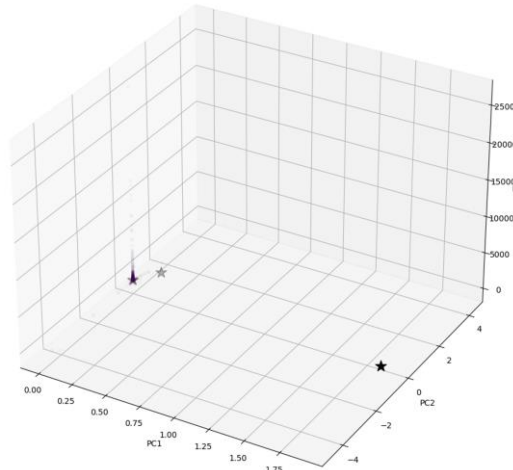
2월 하순



3월 하순



중심경향치와의
거리 유사도 기반의
K-Means 클러스터링



Elbow Method에 의해
최적의 군집 수 = 3으로 설정

K-평균 군집분석 과정에서
데이터 내 결측값 또는 0의
비율에 영향을 받아

3개 군집의 중심경향치가
0에 가까웠으며,
3차원 축소한 값이 0에
밀집된 문제가 있었다.

밀도 기반의
이상치 탐지 클러스터링
(DBSCAN)

Epsilon값을 조정하더라도
지속적으로 군집이 1개만
형성되는 문제가 있었다.

결론 : 군집으로 범주형 변수인 관리사원 유무, 연령대 등과 투자의 시기, 규모 등을 기준으로 구분

3

서비스 기획

투자자 행태와 관련한 팩터 설정 및 발굴

일반 금융소비자의 투자심리지수가
국내주식 수익률 및 해외주식 수익률에
대해 가지는 설명력을 확인

특정기간 동안의 국내주식 수익률과
해외주식 수익률을 다변량 분산분석,
다변량 시계열 선형 확률과정 등을
이용하여 가설을 수립한 후 검정

집단의 특성으로부터 도출되는
분석결과를 종합적으로 고려하여
해당 집단의 증권수익률에
가장 영향력 있는 행태적 특성을 파악

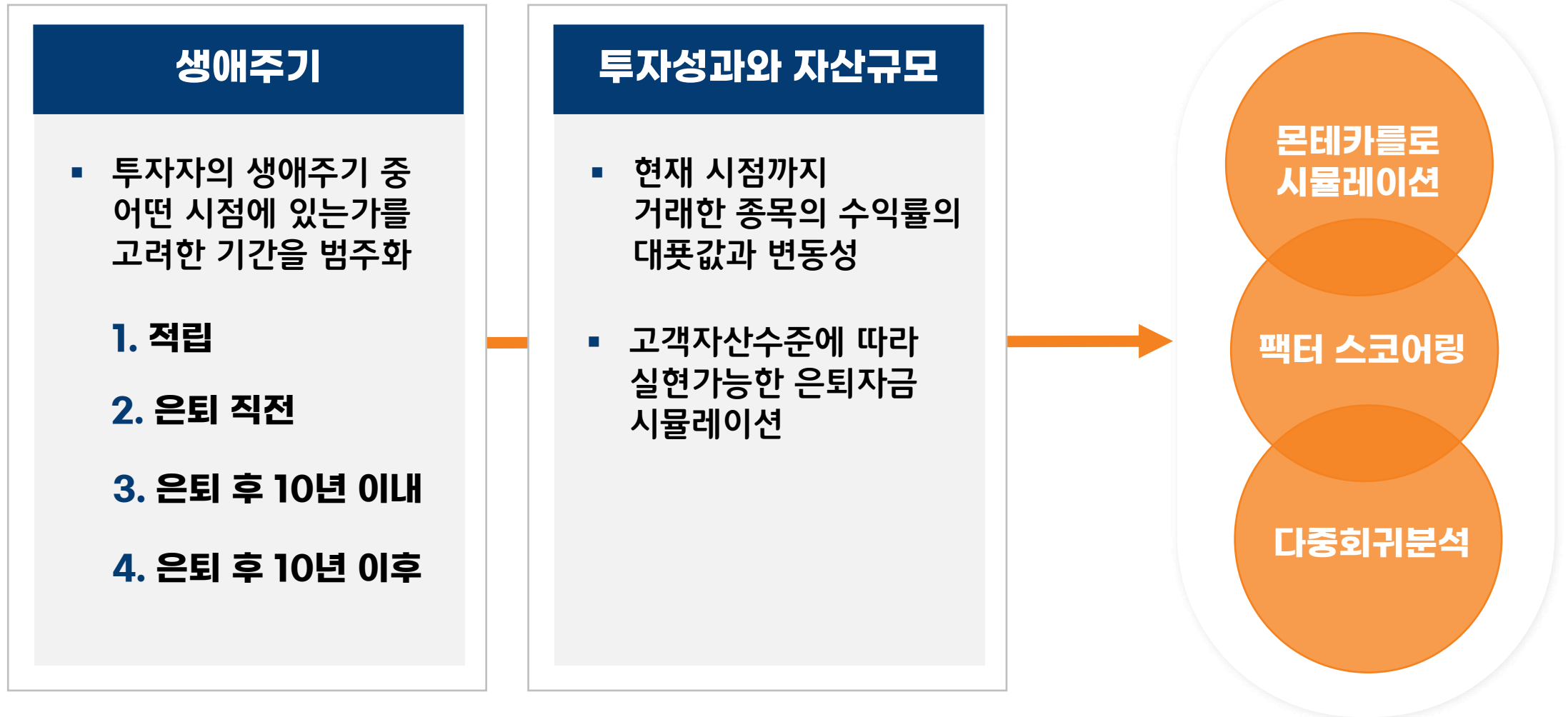
행태적 특성에 근거하여 금융소비자 집단
간 차이를 살펴볼 때 통계적으로 유의한
행동패턴과 관련한 팩터를 찾을 수 있다.

투자자의 특성을 고려하여 찾을 수 있는 팩터 예시

팩터 유형	활용 목적
밸류에이션[가치] 팩터	특정 내재가치 정도에서 투자하였을 때 좋은 투자 성과가 나타나는지를 보는 팩터
사이즈[시가총액] 팩터	기업의 규모 를 고려하여 투자 성과를 측정하는 팩터
성장성 팩터	회사의 재무 데이터 를 고려해 성장 가능성 이 높은 기업의 투자 성과를 측정하는 팩터 (이익의 퀄리티를 계량적으로 판단)
모멘텀[시간민감도] 팩터	회사의 실적 추이나 과거의 주가 상승 트렌드 가 주가에 반영되는지를 보는 팩터 (최근 상승세를 기록한 종목)

팩터구성·생애주기·자산규모를 고려한 투자전략 설정

성공적인 은퇴계획을 위하여 고려해야 할 투자자의 특성에 맞추어 팩터 포트폴리오의 리밸런싱을 추천



- 변진호 · 김근수, 「주식시장 투자심리지수의 유용성」, 2013.
- 김민기 · 김준석, 「국내 개인투자자의 행태적 편의와 거래형태」, 2022.
- Mukesh Kumar 외 1인, 「Association Rules Mining in the Stock Data」
- 전익진 · 이학연, 「연관규칙 기반 동시출현단어 분석을 활용한 기술경영 주제 네트워크분석」, 2016.
- 유재필 · 신현준, 「연관분석을 이용한 금융 상품 거래 동향의 빅데이터 분석」, 2021.
- 김다윤 · 신용현, 「The effects of pre-/post-retirement borrowing constraints on optimal consumption, investment, and retirement」, 2023.
- 미래에셋투자자와 연금센터, 「전략적 인출 설계와 은퇴 포트폴리오의 과학」, 2022.
- 미래에셋투자자와 연금센터, 「초장수 시대의 노후 자산 운용」, 2019.