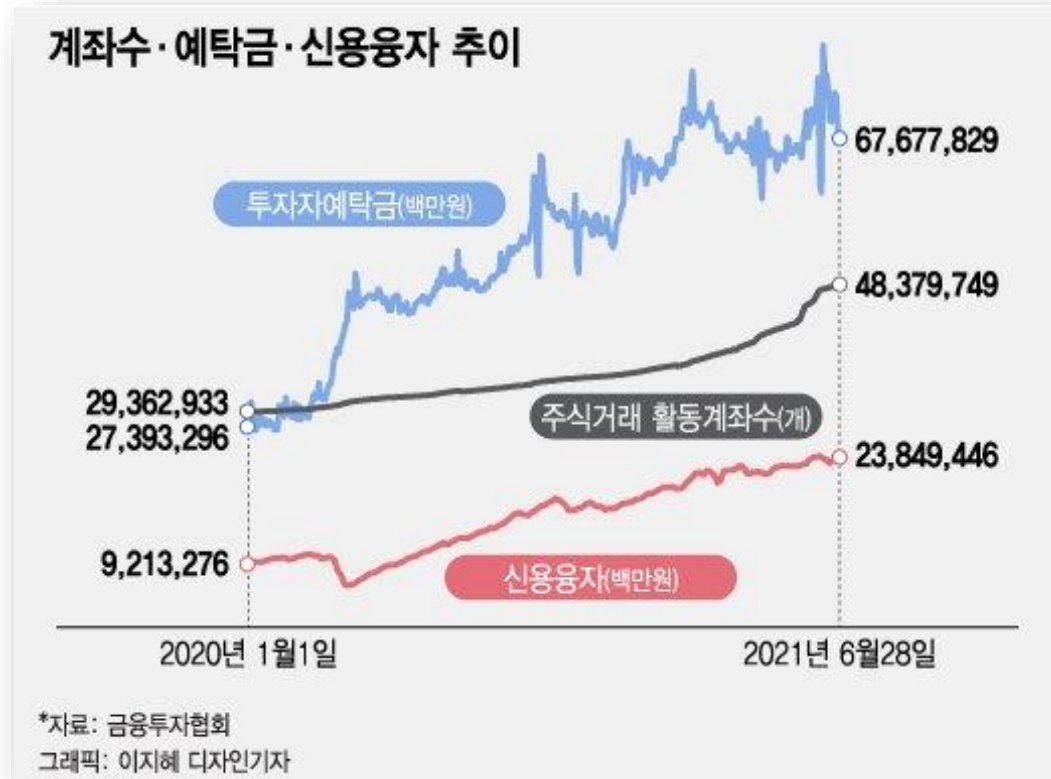


2022 NH투자증권 빅데이터 경진대회

투자, 건강한 문화가 되다

팀 Versatile 강지민 박소연

기획 배경 문제인식



투자가 문화가 된 대한민국
국민 1인당 1주식 계좌 시대

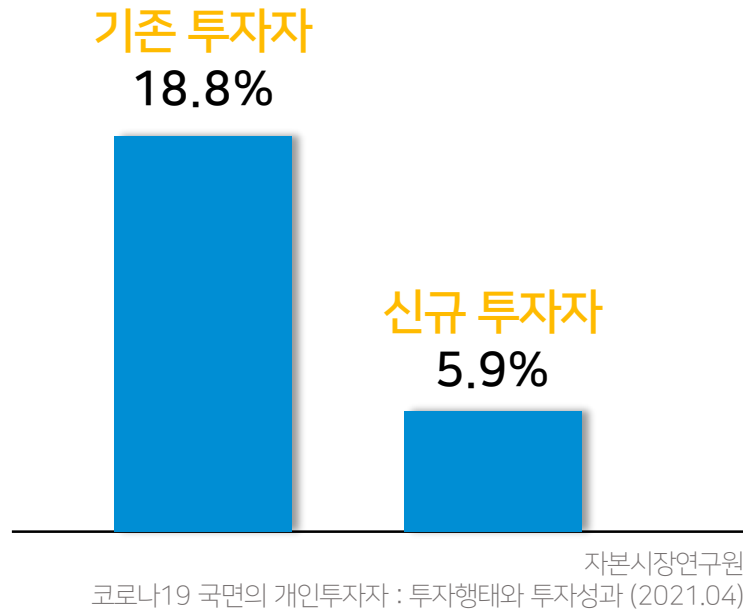


‘빚투’로 불리는 신용융자규모 또한
꾸준히 증가하는 추세

기획 배경

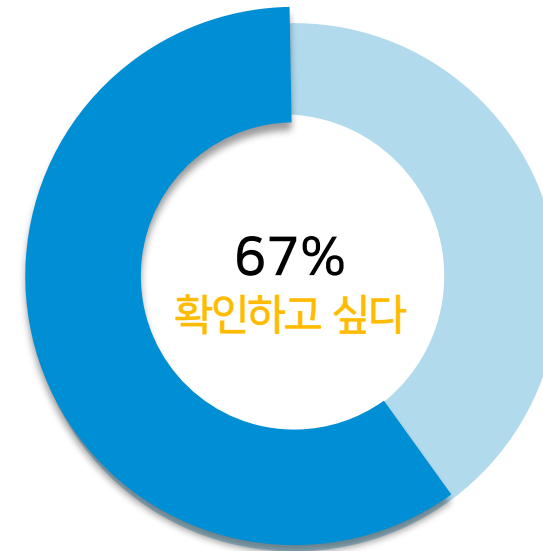
문제인식

기존 투자자와 코로나19 이후 유입된
신규 투자자의 수익률 격차



기존 투자자들과 달리, 신규 투자자들은
단기투자성향을 보이면서 투자성고가 현저히 떨어짐

고수 투자자들의 투자 현황을 확인하고
싶다고 응답한 비율



삼성증권
우수고객 25만명 대상 고객 설문조사 (2022.09)

프라이빗 뱅커상담(PB), 고수 투자자들의 투자 현황 등에
대한 수요 존재하며 투자 가이드를 필요로 함

주식시장에 직접 참여하는 투자자들의 과잉확신

개인투자자가 과도하게 거래하는 원인에 대한 검증된 설명은 크게 두 가지다. 첫 번째는 과잉확신 (overconfidence)의 영향이다.⁵⁾ 스스로의 능력이 뛰어나다는 믿음, 자신이 가진 정보가 더 정확하다는 믿음이 빈번한 거래를 유발한다는 것이다. 물론 기관투자자도 과잉확신 경향을 보이나 개인투자자에서 더 현저한 것으로 평가된다. 불확실성이 높은 주식시장에 직접 참여하는 투자자들은 기본적으로 과잉확신을 가지고 있을 가능성이 높은 데다, 우수한 투자성과는 자신의 능력에 따른 것, 저조한 투자성과는 운이 없었기 때문이라고 생각하기 때문에 주식시장에서 과잉확신은 쉽게 사라지지 않는다고 설명한다. 실증적으로 과잉확신 경향을 가진 투자자는 거래빈도가 높고, 과거의 수익률이나 거래량을 추종하는 거래행태를 띠며, 결과적으로 투자성과는 저조하다. 과잉확신 경향은 주가의 과대평가와 버블을 일으키는 원인으로 지적되기도 한다.

자본시장연구원
코로나19 국면의 개인투자자 : 투자행태와 투자성과 (2021.04)

이러한 근거로, 금융 고객은 **추천 상품 수용도는 비교적 낮은 편**이며 추천보다는 자기주도형 검색 등을 통한 **주도적 투자를 선호하는 경향**이 나타남

자기주도형 금융 공부 문화 형성이 필요한 시점



1. 고객을 투자성향에 따라 분류



2. 투자성향과 유사한
스터디, 멘토 추천



3. 투자 큐레이션 검색 서비스

주제 선정 및 목표

변수 선정

투자 성향과 유의미한 상관관계가 있는

5개의 변수

90일 포트폴리오 베타 계수

월평균 일일 거래 빈도

우량주 투자 비율

파생상품 보유 개수

정보 수집 성향 지수



목표

위험 감수 성향 파악

트레이딩 성향 파악

안전 투자 성향 파악

위험 투자 성향 파악

정보 수집 성향 파악

6개월 간의 시계열 고객 정보 데이터



매월 개별 고객을 분류 갱신
매달 변화하는 개인 투자자의 성향을 파악하고 예측 가능

데이터 수집 및 전처리

데이터 수집

1. NH 투자증권 데이터셋 (내부)

파일명	데이터 파일 설명	건수
cus_ifo.csv	고객 정보	59,081
kr_stk_orr.csv	국내 주식 거래 내역	503,586
oss_stk_orr.csv	해외 주식 거래 내역	51,902
cus_tot_aet_ifo.csv	자산 정보	10,000
cus_itg_sct_bnc.csv	잔고 정보	366,797
GA.csv	MTS 종목 연관 화면 조회 데이터	3,040,921

2. FinanceDataReader 패키지 - 국내, 해외 주식 시세 데이터 (외부)

3. Investing.com - 해외 주식 시가총액에 대한 업종 정보 (외부)

4. 공공데이터포털 : 금융위원회 _주식시세정보 - 국내 주식 종목에 대한 시가총액 정보 (외부)

데이터 수집 및 전처리

데이터 전처리

각 변수들의 단위가 다르기 때문에 정규화 후 모델링 진행

```
[56]: # 정규화 진행
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
scaler.fit(final_scaled)
final_scaled = scaler.transform(final_scaled)

final_scaled = pd.DataFrame(final_scaled, columns=['정보수집지수', '베타계수', '파생상품 유무', '우량주 비율', '일일 거래평균'])
final_scaled
```

```
[56]:
```

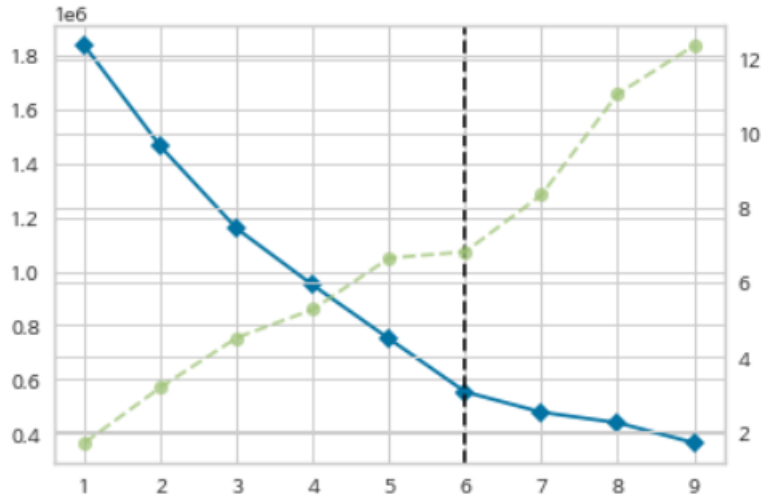
	정보수집지수	베타계수	파생상품 유무	우량주 비율	일일 거래평균
0	1.077831	-0.159915	3.385510	-1.302904	1.640279
1	1.077831	-0.159915	3.385510	-1.302904	1.640279
2	1.077831	-0.159915	3.385510	-1.302904	1.640279
3	1.542168	-0.014303	3.385510	-1.299999	0.623196
4	1.542168	-0.014303	3.385510	-1.299999	0.623196

Elbow Method를 이용하여 **최적의 K 값** 파악
K = 6

```
[57]: from yellowbrick.cluster import KElbowVisualizer
      from sklearn.cluster import KMeans

      model = KMeans()
      visualizer = KElbowVisualizer(model, k=(1,10))
      visualizer.fit(final_scaled)
```

```
[57]: KElbowVisualizer(ax=<AxesSubplot:>, estimator=KMeans(n_clusters=9), k=(1, 10))
```



변수가 모두 수치형이므로
K-Means 알고리즘을 통해
클러스터링 진행

```
[58]: k = 6

      # 그룹 수, random_state 설정
      model = KMeans(n_clusters = k, random_state = 10)

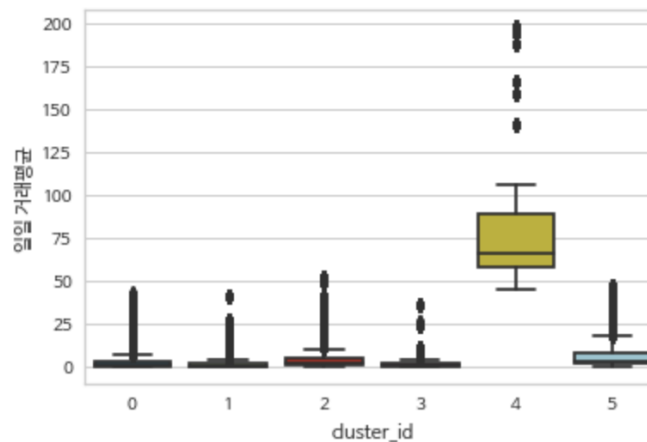
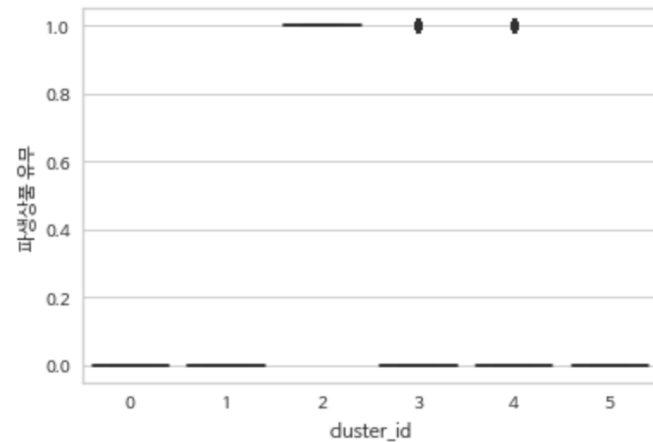
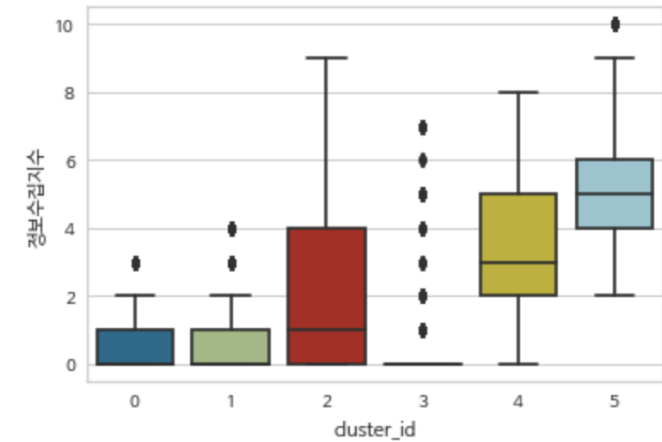
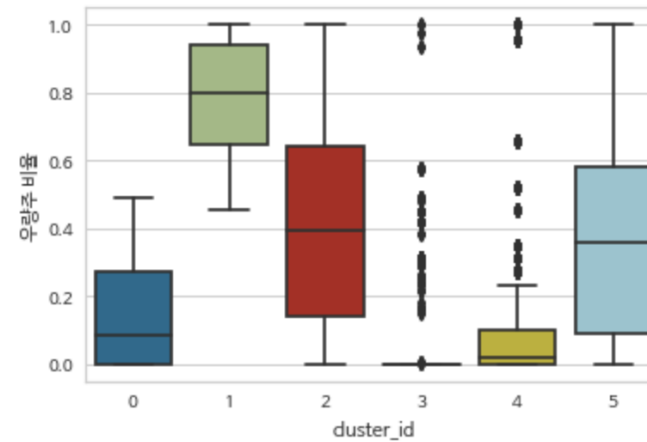
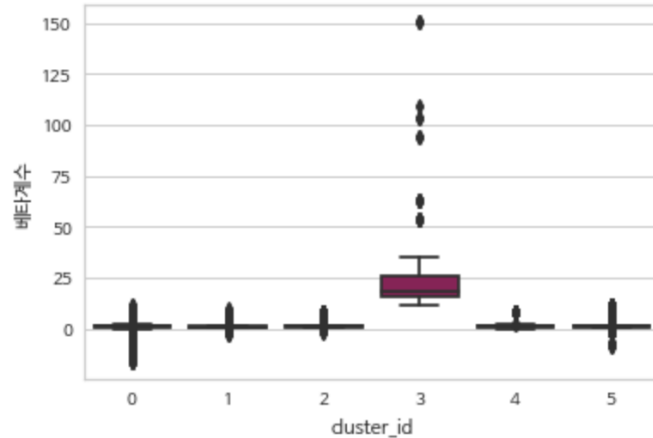
      # 정규화된 데이터에 학습
      model.fit(final_scaled)

      # 클러스터링 결과 각 데이터가 몇 번째 그룹에 속하는지 저장
      final['cluster_id'] = model.fit_predict(final_scaled)
```

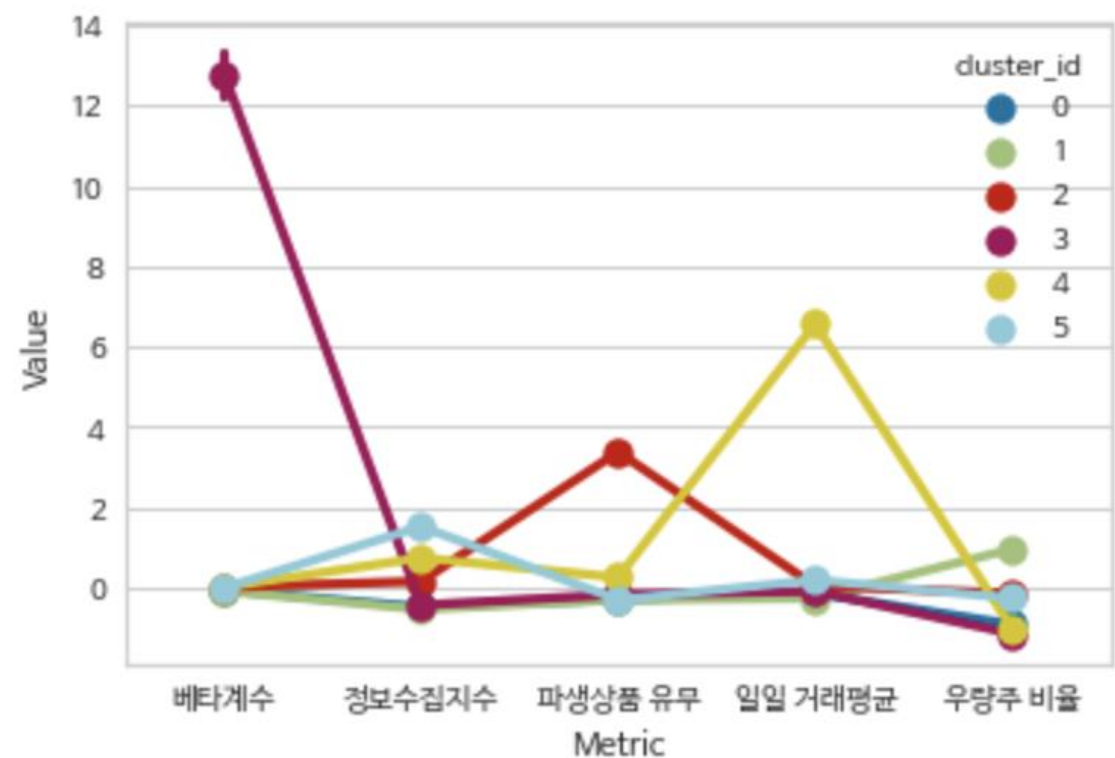
데이터 분석

모델링 결과

6개의 군집으로 분류 후, boxplot을 이용하여 5개의 변수에 대한 클러스터의 특징을 확인



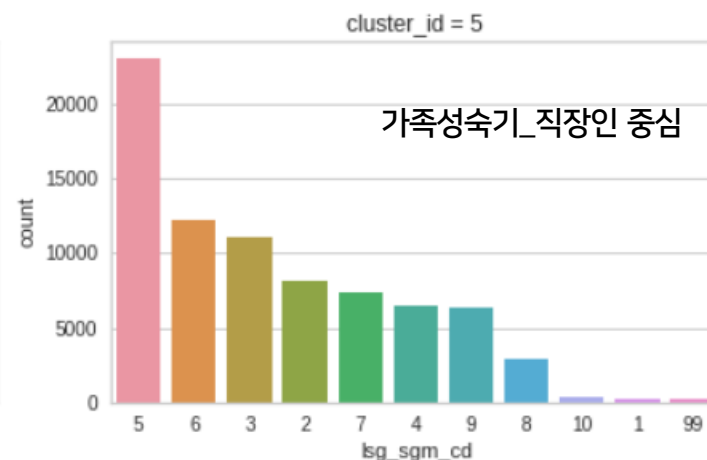
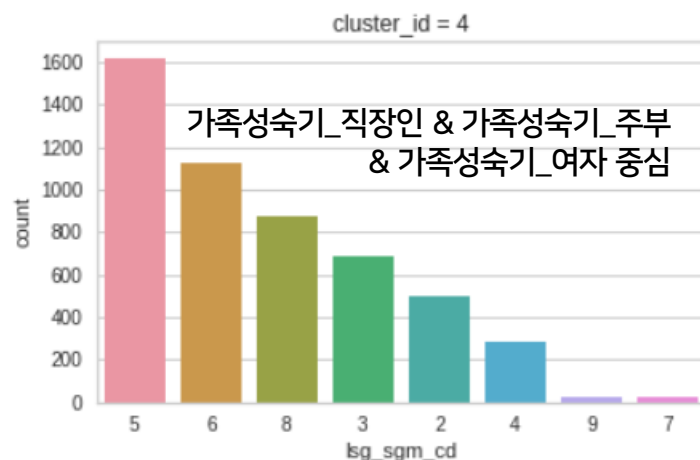
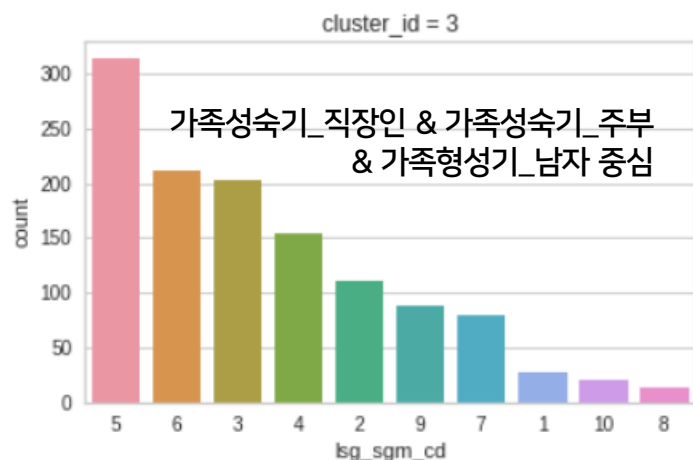
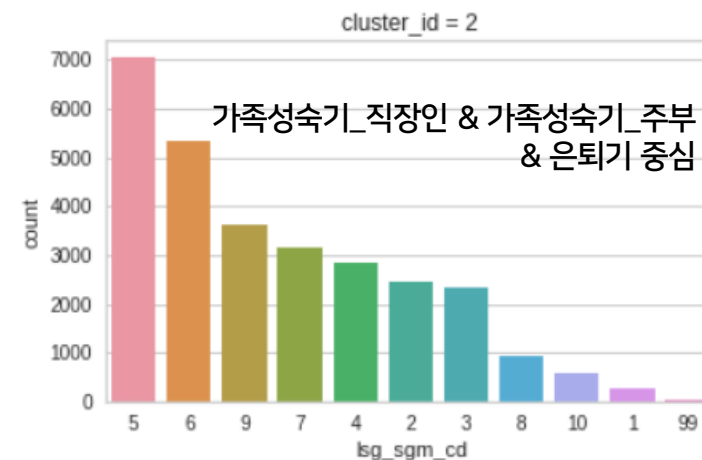
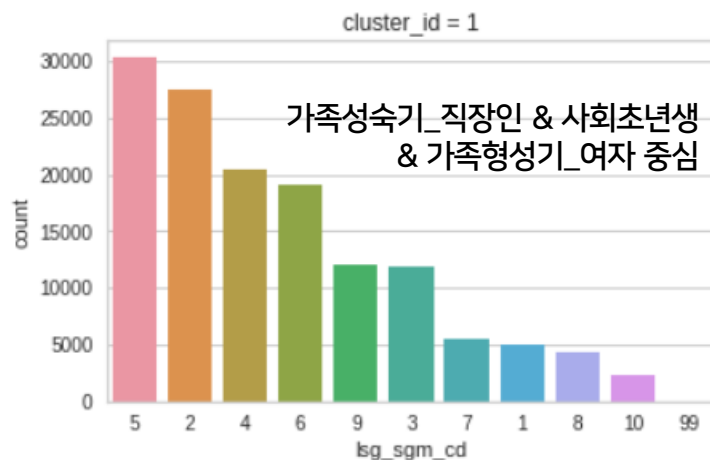
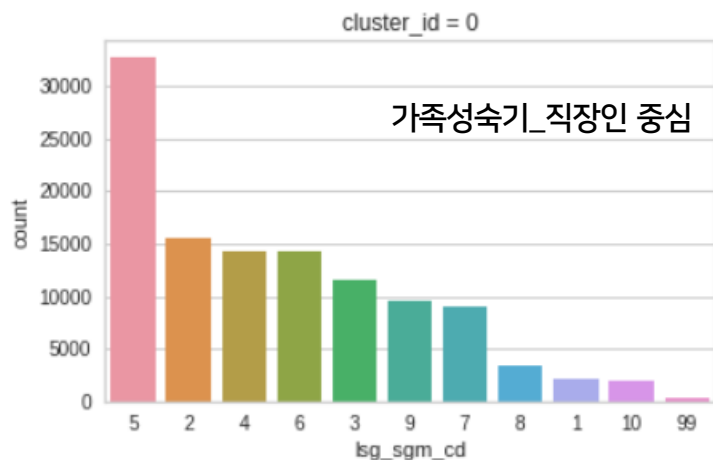
Sankey Plot으로 변수에 따른 클러스터 특성 확인



변수값 분포에 따른 클러스터명 부여

Id	Cluster_name
0	정보 한정형 가치투자자
1	우량주 중심 안전투자자
2	위험 수익 추구형 공격 투자자
3	높은 변동성을 가진 적극 투자자
4	정보 민감형 트레이더
5	정보 수집형 가치투자자

군집별 라이프 스테이지 코드 분포



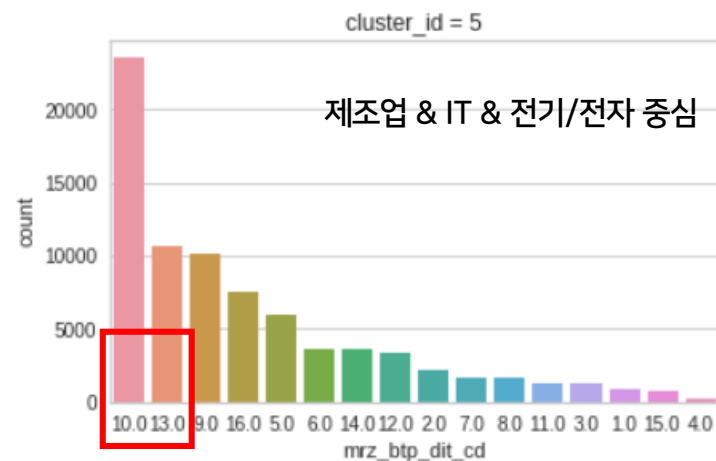
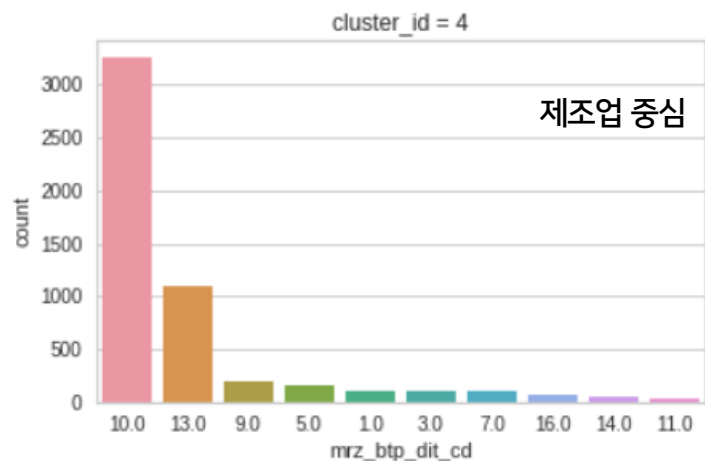
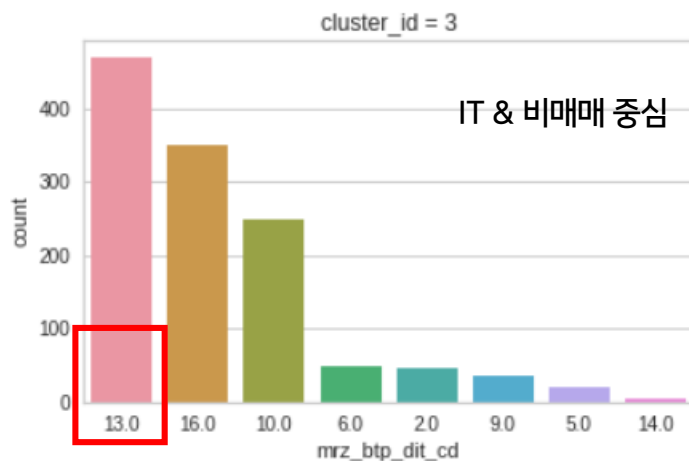
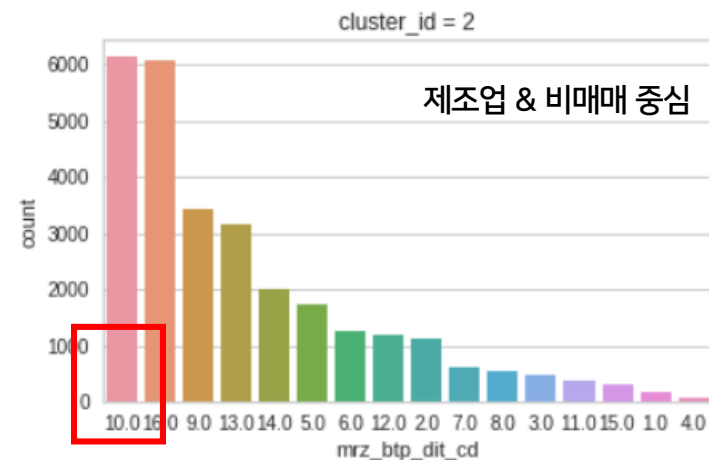
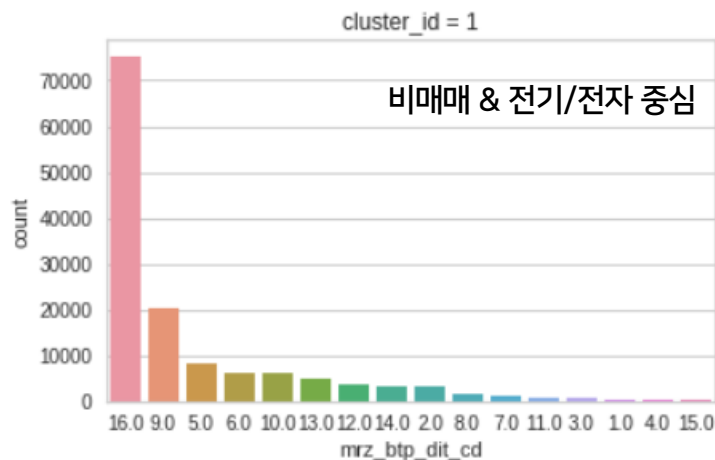
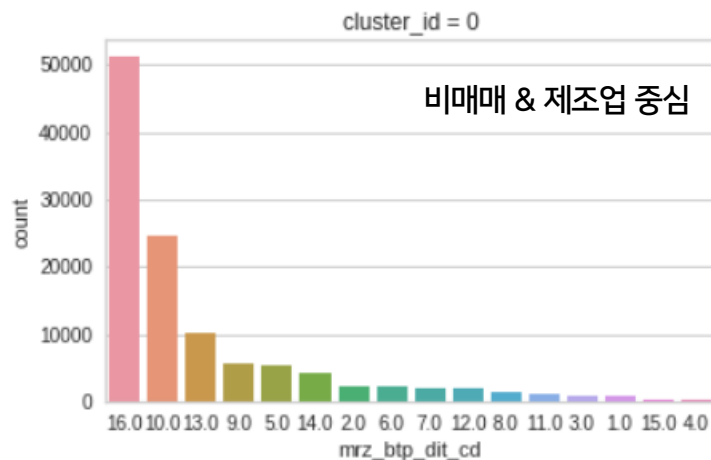
데이터 분석 결과 해석

군집별 주거래 시장구분 코드 분포



데이터 분석 결과 해석

군집별 주거업종 코드 분포



데이터 분석 결과 해석

군집별 고객자산구간 분포



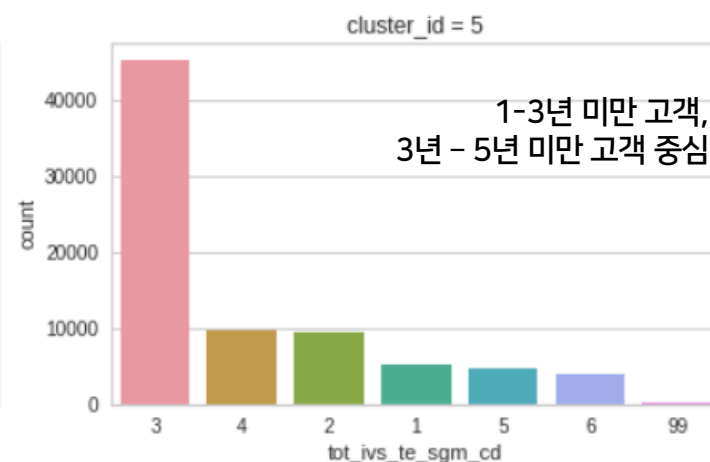
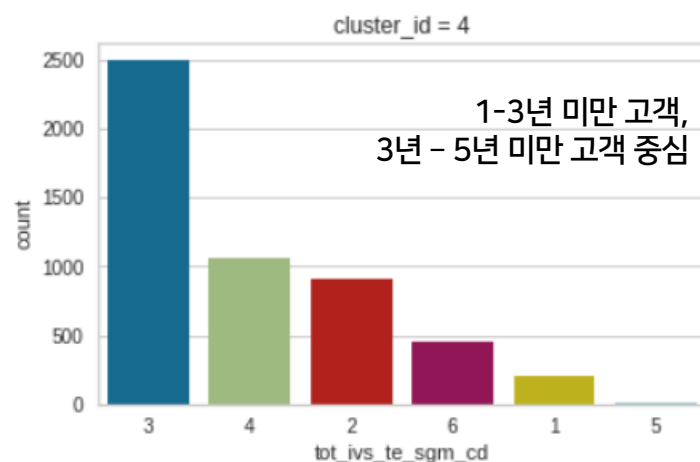
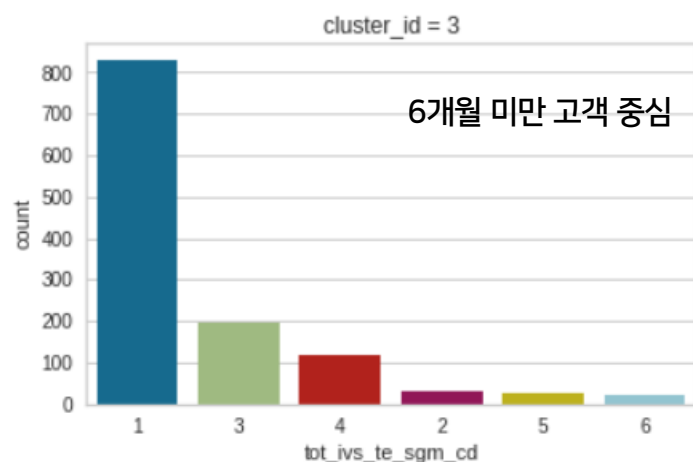
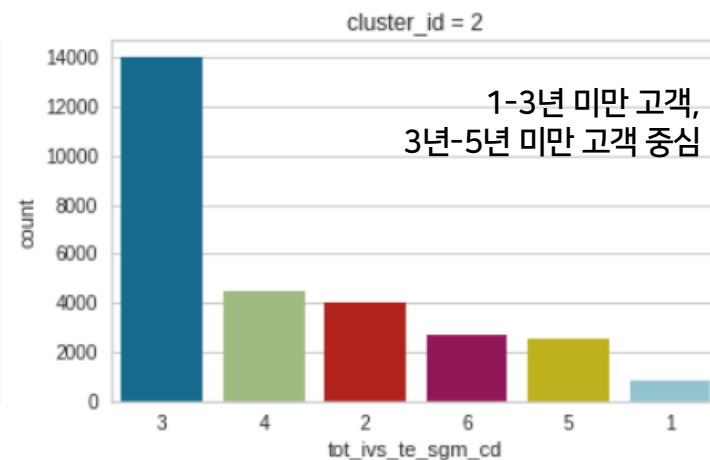
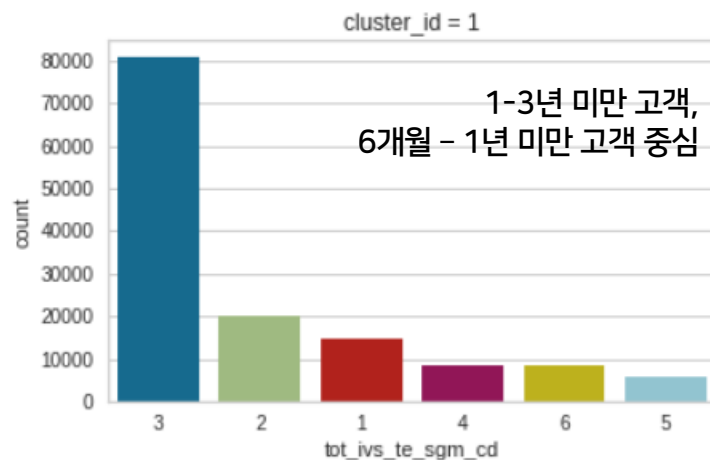
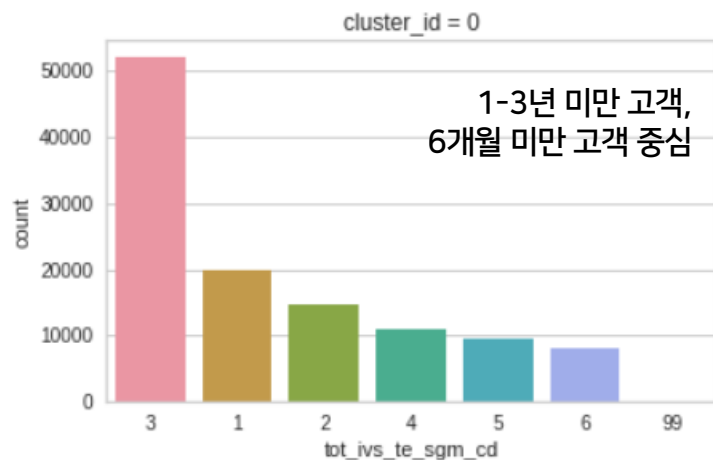
데이터 분석 결과 해석

군집별 수익구간금액 분포



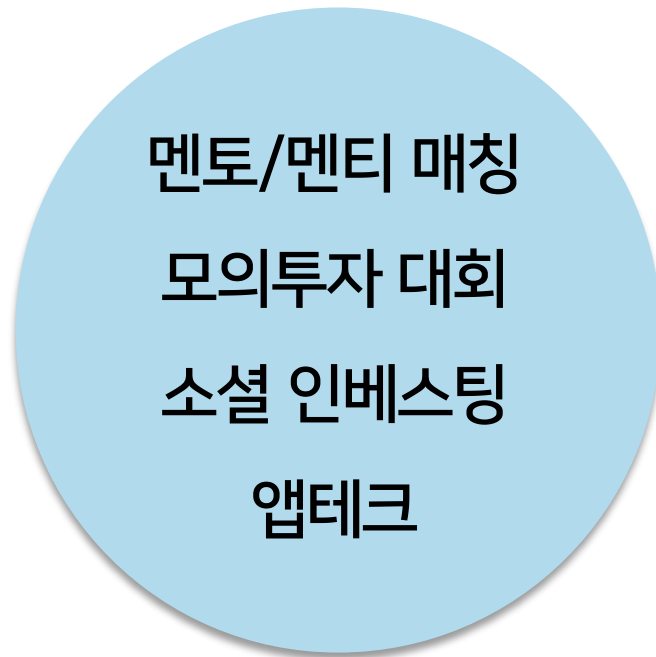
데이터 분석 결과 해석

군집별 총 투자기간 분포



활용방안 및 기대효과

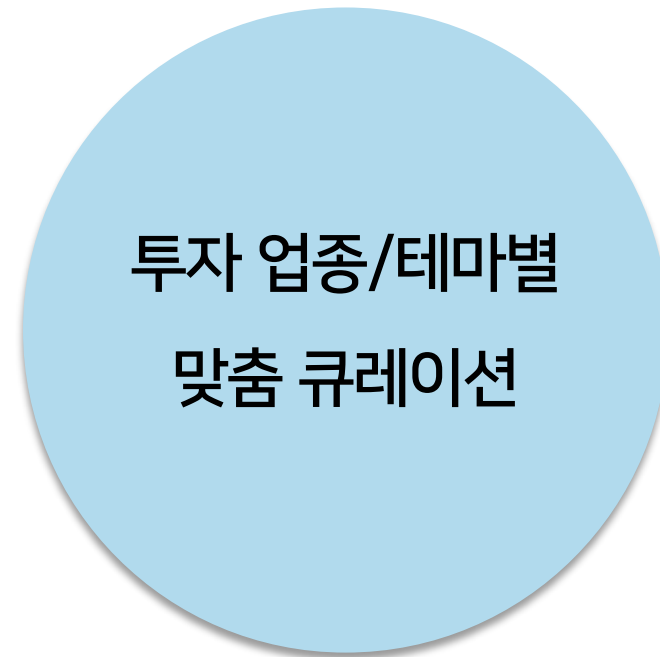
투자 스터디 큐레이션



건강한 투자 문화 형성에 기여



맞춤 큐레이션 검색



투자 성과 증진 기대

활용방안 및 기대효과

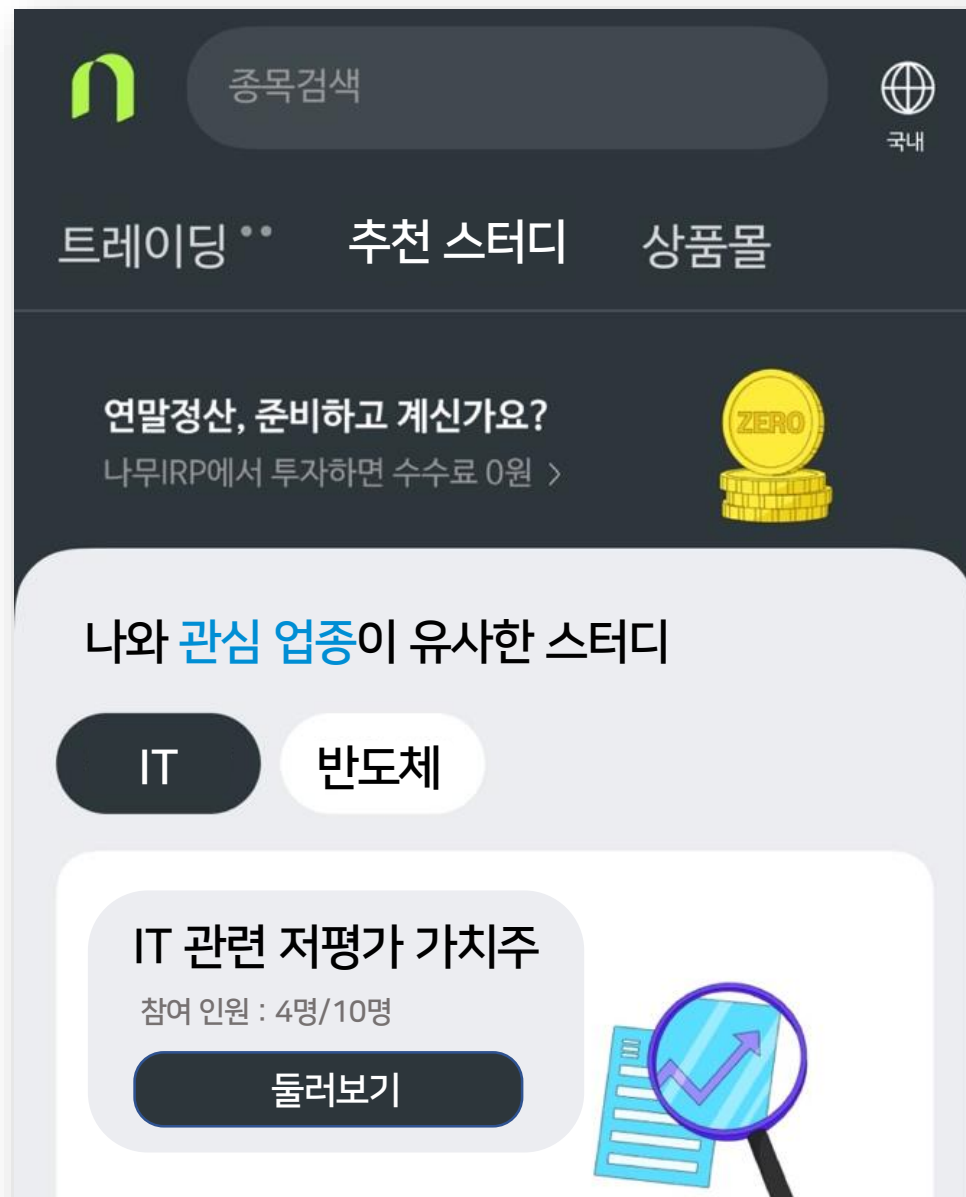
투자 스터디 큐레이션 콘텐츠 01

주식 온라인 교육 멘토링 서비스



고객 분류를 통한 관심 성향별 스터디 추천

2022 NH투자증권 빅데이터 경진대회



투자성향 군집별 모의투자 대회 개최

내가 속한 그룹 확인하고 대회 참여해서 상금 받자!

투자성향 그룹별 모의투자 대회

순위	그룹명	평균 수익률(%)	참가자 수
1등	하이리스크	29.1(%)	147명
2등	우량주짱	28.8(%)	175명
3등	같이가치	25.2(%)	98명

비슷한 투자 성향을 가진 군집끼리 모의투자 대회를
개최하여 직접 투자를 경험

소셜 인베스팅

고수의 포트폴리오 엿보기



닉네임 : 올라가작

한 줄 소개 : 안녕하세요. 미국 IT 기업에 집중 투자합니다.

연평균 수익률 : + 15.8%

포트폴리오 확인

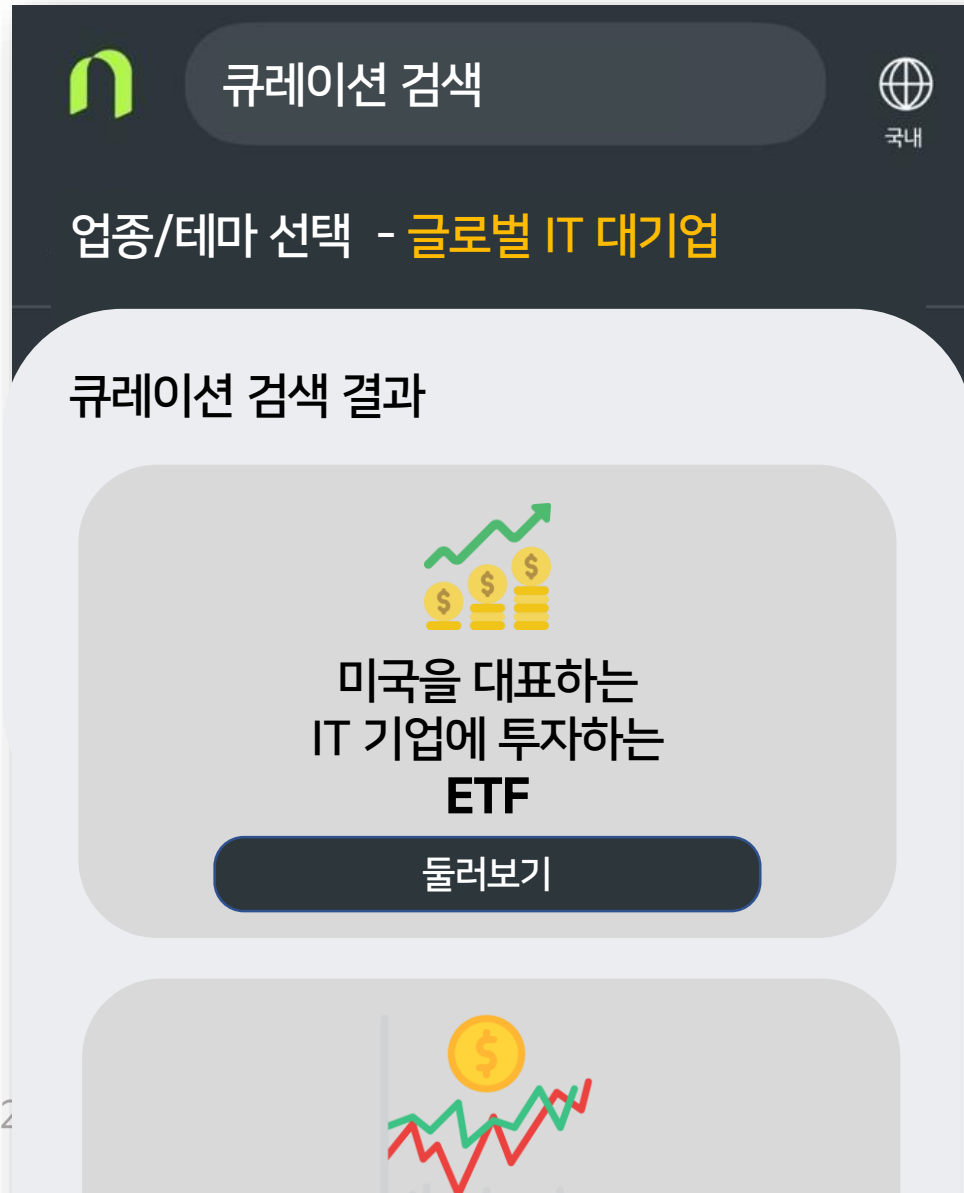


비슷한 투자 성향을 지닌 **고수의 프로필과 포트폴리오를 노출시켜**
투자 고수의 실거래 포트폴리오를 확인 가능

앱테크 서비스

- 멘토링에 대한 리워드 포인트 지급
- 프로필 추천에 따른 리워드 포인트 지급
- 교육 내용과 관련된 퀴즈를 제공하여,
정답자 대상 포인트 지급





큐레이션 검색 서비스

- 클러스터별 특징을 파악
- 해당 고객이 속한 클러스터의 성향을 반영하여, 클러스터 내 인기 있는 종목 혹은 상품을 맞춤 큐레이션

예시 - cluster_id = 3인 군집에 속한 고객인 경우

감사합니다