

NH투자증권 Y&Z세대 투자자 프로파일링

투자자 개인 맞춤형 금융상품 제안

밀레니얼 세대 & 코로나 장 유입을 기준으로

윤희상

CONTENTS

01. 주제 선정

분석 배경, 분석 목적

02. 분석 과정

데이터 설명, 분석 개요, 데이터 전처리

03. 분석 결과

ANOVA, 클러스터링, 집단 분석

04. 결론

집단 별 맞춤형 금융상품 추천

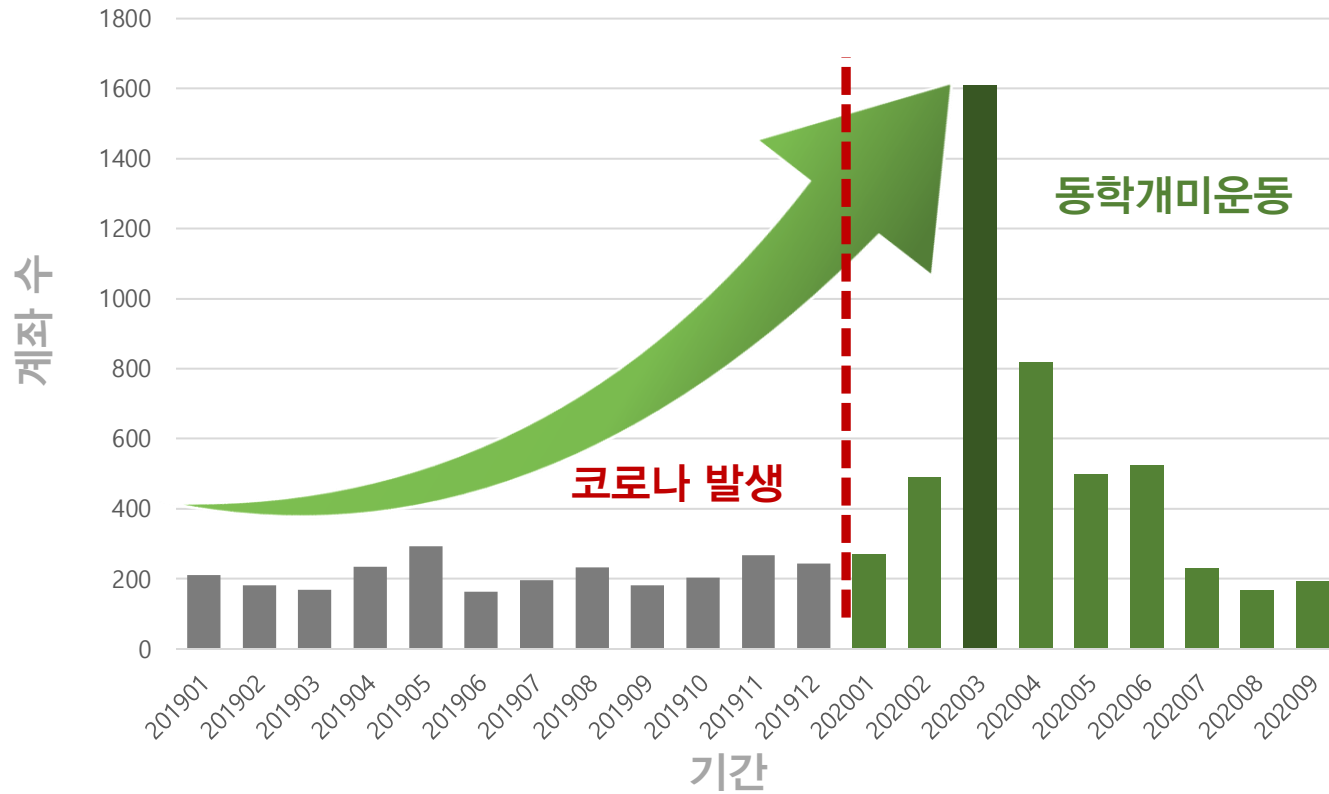
01. 주제 선정

분석 배경, 분석 목적

01 주제 선정

ㄱ. 분석 배경

월별 개설된 계좌 수

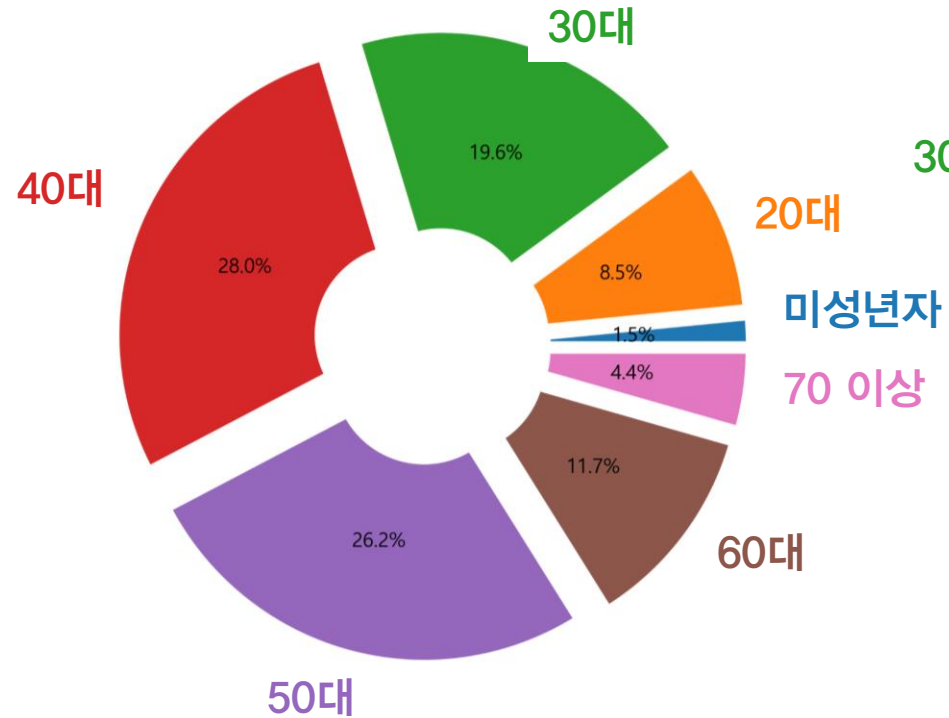


코로나 장 동학개미운동이 발발하면서
한 축이 된 새로 유입된 수많은 '주린이'들

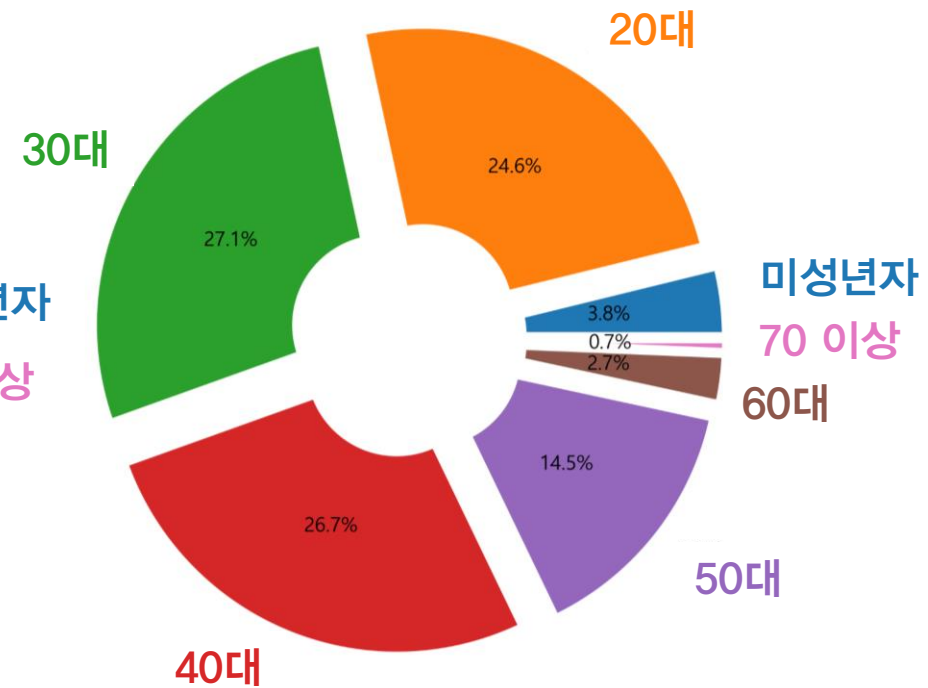
01 주제 선정

ㄱ. 분석 배경

2020년 이전 계좌 개설 연령대 비율



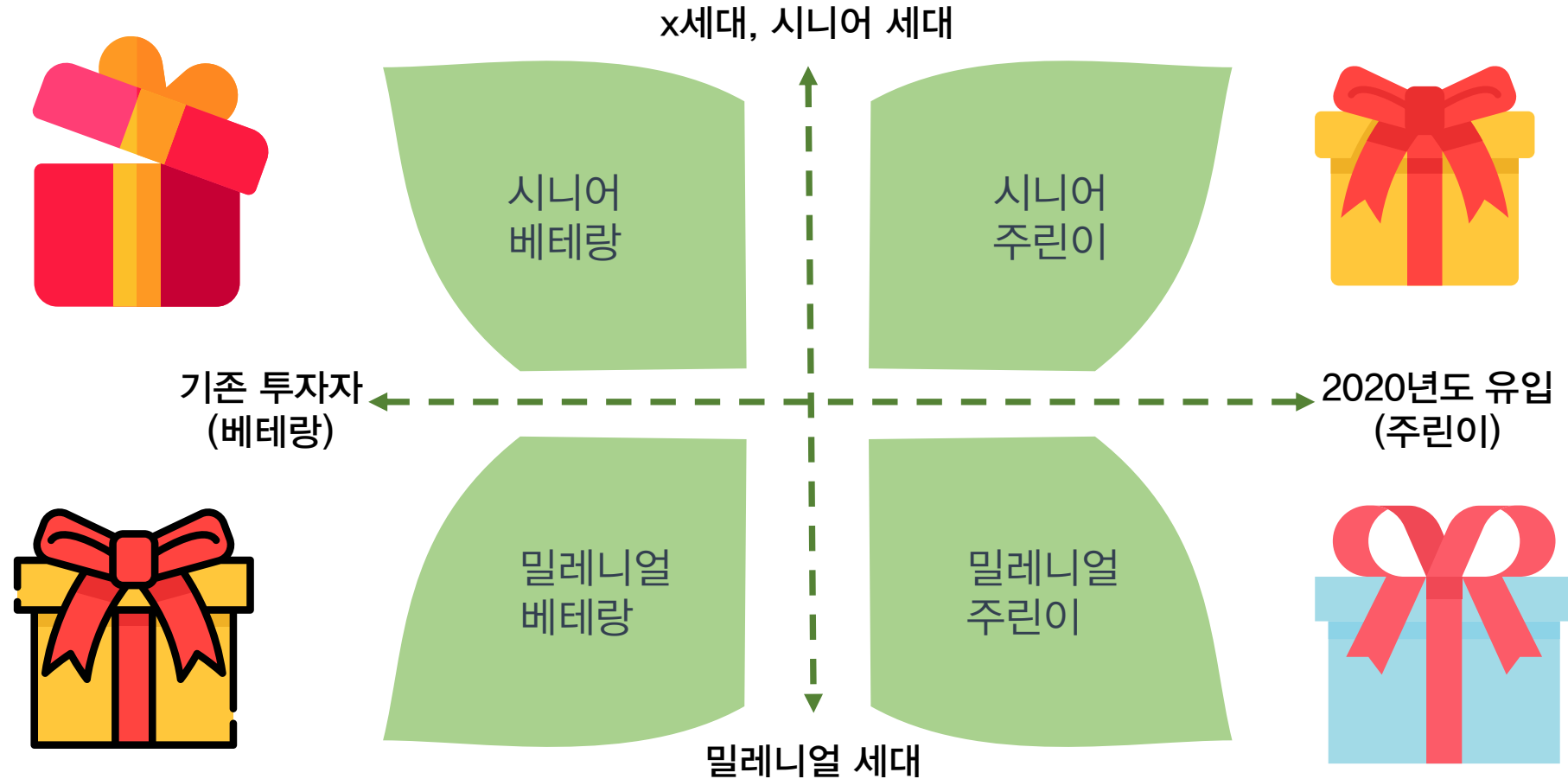
2020년 계좌 개설 연령대 비율



특히 20 ~ 30대에서 큰 폭으로 유입되었고 주식에 대한 관심도 커져갔다

01 주제 선정

└. 분석 목적



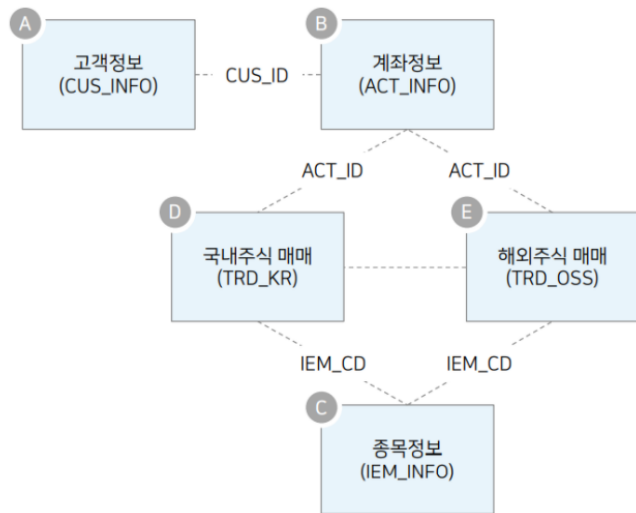
하지만 당장 유입은 됐지만 어떻게 운용해야 하는지 모르는 이들이 대부분..
각 유형의 투자성향에 맞는 금융상품 제안이 필요!!

02. 분석 과정

데이터 설명, 분석 개요, 데이터 전처리

02 분석 과정

ㄱ. 데이터 설명



※ NH투자증권 제공 데이터

파생변수

1. 체결비용 : 체결수량(CNS_QTY) * 체결가격(ORR_PR)
2. 세대 : 20세 미만 : YOUNG, 20~30 : Y, 30~40 : Z, 40세 이상 : OLD
3. 년월 : 주문 날짜(ORR_DT)에서 년,월만 추출
4. 이탈 기간: 2020년 7월 1일 기준 각 투자자별 마지막 거래일로부터 지난 시간

	계좌번호	고객번호	종목코드	매매구분코드	체결수량	체결가격	체결비용	종목 한글명	세대	지난 시간	주문날짜	년월
0	6caf4de40251d37cbba310f0c6dd0ed8f4c7ca4ccfa675...	8eafaa5dc6eb52ada6b2d93ae1a0544f2c91e71667c945...	A215600	2	35.00	74200.00	2597000.00	신라젠	old	546	2019-01-02	201901
1	6caf4de40251d37cbba310f0c6dd0ed8f4c7ca4ccfa675...	8eafaa5dc6eb52ada6b2d93ae1a0544f2c91e71667c945...	A001060	2	20.00	41100.00	822000.00	JW중외제약보통주	old	546	2019-01-02	201901
2	effb34815e1b7efb34beb550f604ede8ac71a5a71edb30...	1d6cbe3709fcfb798268d27058a2c89e191c7893060a66...	A281820	2	20.00	9460.00	189200.00	케이씨텍보통주	old	546	2019-01-02	201901
3	fea6c038cb7236a02d1e779262698a36219b4c9b527ab1...	f682d5d3e7a44cfc687e78366d638ae324d79b3dcf8247...	A006650	2	1.00	148500.00	148500.00	대한유화보통주	y	546	2019-01-02	201901
4	fea6c038cb7236a02d1e779262698a36219b4c9b527ab1...	f682d5d3e7a44cfc687e78366d638ae324d79b3dcf8247...	A051360	2	10.00	6550.00	65500.00	토비스	y	546	2019-01-02	201901
5	fea6c038cb7236a02d1e779262698a36219b4c9b527ab1...	f682d5d3e7a44cfc687e78366d638ae324d79b3dcf8247...	A051360	2	10.00	6560.00	65600.00	토비스	y	546	2019-01-02	201901
6	fea6c038cb7236a02d1e779262698a36219b4c9b527ab1...	f682d5d3e7a44cfc687e78366d638ae324d79b3dcf8247...	A051360	2	10.00	6570.00	65700.00	토비스	y	546	2019-01-02	201901
7	fea6c038cb7236a02d1e779262698a36219b4c9b527ab1...	f682d5d3e7a44cfc687e78366d638ae324d79b3dcf8247...	A018260	2	1.00	205000.00	205000.00	삼성에스디에스보통주	y	546	2019-01-02	201901
8	6dee1607a1b1c6ae828eb8c84bef0ef963776eb06163a7...	4fdc8ee4c3be41e47366e4b792f9b5d4f39c7942acf6fc...	A074610	2	5.00	5480.00	27400.00	이엔플러스보통주	old	546	2019-01-02	201901
9	6dee1607a1b1c6ae828eb8c84bef0ef963776eb06163a7...	4fdc8ee4c3be41e47366e4b792f9b5d4f39c7942acf6fc...	A074610	2	5.00	5450.00	27250.00	이엔플러스보통주	old	546	2019-01-02	201901

※ 편의를 위해 한글로 컬럼명 전환

02 분석 과정

└. 분석 개요

데이터 전처리

베테랑/주린이 구분
투자자금 도출



투자자 타겟팅



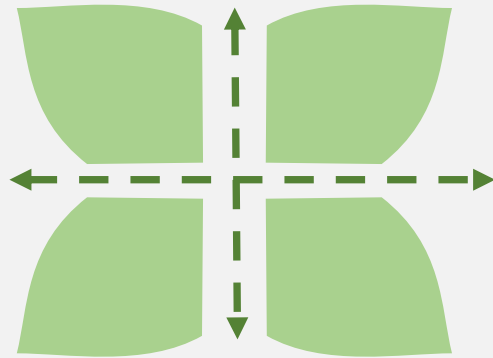
수익률

매매회전율

이탈기간



요소 별 데이터 분석

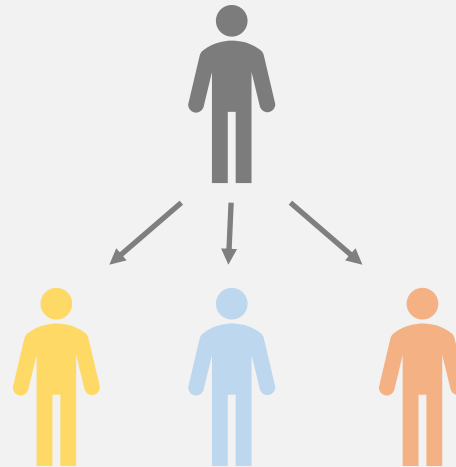


ANOVA분석을 통한
집단 간 유의미한 차이 도출



클러스터링

K-means clustering



집단 분석 및 상품 추천



02 분석 과정

ㄷ. 데이터 전처리

■ 투자자 타겟팅

ㄱ. 베테랑 / 주린이 구분

- 베테랑 : 2020년 이전부터 투자해온 숙련된 투자자
- 주린이 : 2020년 코로나 장에 유입된 초보 투자자

ㄴ. 투자자 중 아래 조건에 맞는 경우

- 매매거래내역을 분석했을 때, 특정 종목의 보유 수량이 0 이상인 경우
- 2020년에 거래한 내역이 있는 경우
- 계좌를 하나만 보유한 경우

ㄷ. 투자자금이 100만원 ~ 5천만원인 투자자만 취급



02 분석 과정

ㄷ. 데이터 전처리

투자자 타겟팅

ㄱ. 베테랑 / 주린이 구분

2020년 개설 or 이전 개설한 계좌 수로 구분

```
act['2020년 개설'] = 0
act['2020년 이전 개설'] = 0
```

```
for i in range(len(act)):
    if act.loc[i, 'act_opn_ym'] >= 202001:
        act.loc[i, '2020년 개설'] += 1
    else:
        act.loc[i, '2020년 이전 개설'] += 1
```

```
act_g = act.groupby(['cus_id'], as_index = False).agg({'2020년 개설' : 'sum', '2020년 이전 개설' : 'sum'})
```

주린이의 경우 : 20년 이전 개설 x 20년에 첫 개설

```
new_cus = act_g[(act_g['2020년 개설'] == 1) & (act_g['2020년 이전 개설'] == 0)]
```

베테랑의 경우 : 20년 이전 개설한 계좌 수 1개

```
old_cus = act_g[(act_g['2020년 이전 개설'] == 1) & (act_g['2020년 개설'] == 0)]
```

	cus_id	2020년 개설	2020년 이전 개설
0	0005485133f45a0074dc4588abe1ae0fb6b3dfebded28b...	1	0
1	000ffabe461b9b0f5360913aa1b340e84e44446617a8dc...	0	1
2	001b04c9eec677e2278040c51e01d262e72ce32ef25f00...	0	1
3	001ee721662751620282570981b2b4e2c1d42a8ef9329e...	0	2
4	001f5ef83673487f2abf25efdd61e284adcc1953001e69...	1	0

베테랑

2020년 이전에 개설한 계좌를 보유한 사람

주린이

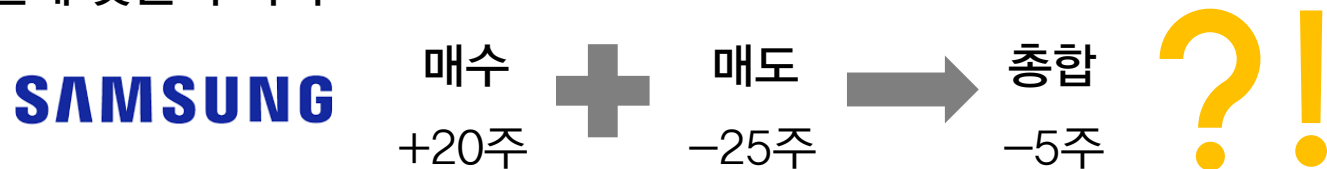
2020년에만 개설한 계좌를 보유한 사람

02 분석 과정

ㄷ. 데이터 전처리

투자자 타겟팅

ㄴ. 조건에 맞는 투자자



- ① 제공된 데이터 중 어떤 투자자는 특정 종목의 체결 수량 합이 음수인 경우가 존재
-> 이는 일반적인 케이스가 아니므로 제외시킴(증권사 이전 etc.)

```
# 2020년 거래내역만 가져옴  
clean_trd = clean_trd[clean_trd['년'] == 2020]
```

- ② 코로나 장의 거래를 비교하고자 하므로 2020년 거래내역이 존재하는 투자자만 추출

	cus_id	2020년 개설	2020년 이전 개설
0	0005485133f45a0074dc4588abe1ae0fb6b3dfebded28b...	1	0
1	000ffabe461b9b0f5360913aa1b340e84e44446617a8dc...	0	1
2	001b04c9eec677e2278040c51e01d262e72ce32ef25f00...	0	1
3	001ee721662751620282570981b2b4e2c1d42a8ef9329e...	0	2
4	001f5ef83673487f2abf25efdd61e284adcc1953001e69...	1	0

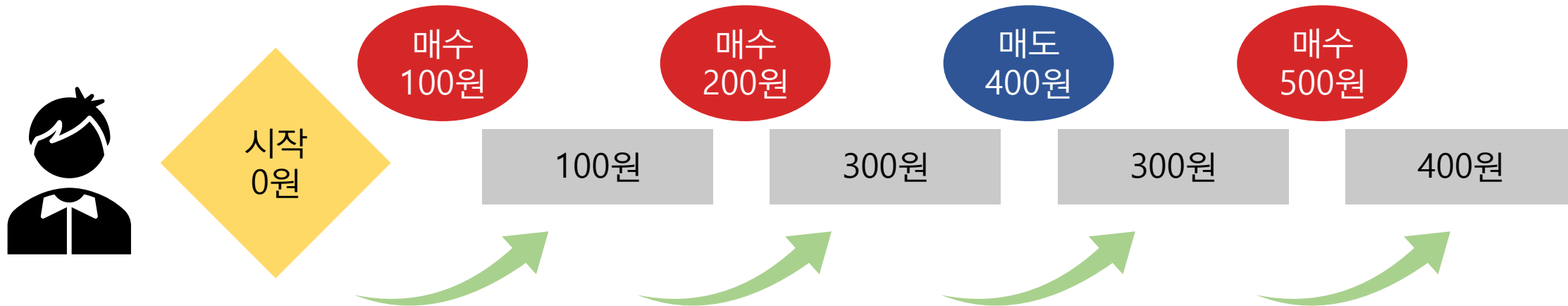
- ③ 분석 편의를 위해 여러 계좌를 보유한 투자자는 제외시킴

02 분석 과정

ㄷ. 데이터 전처리

투자자 타겟팅

ㄷ. 투자자금이 100만원 ~ 5천만원인 투자자만 취급



③ 운용하는 투자자금의 규모가 투자에 영향을 미치므로 알고리즘을 통해 투자자금 유추

※단, 운용 자금을 100% 사용한다는 가정

1117	5ba37ca6d553c7ba7c1df3efa3ca38dc00543a6c8e32b7...	51996420.00	04	04
1118	a638551d75c193e8d3331d8e8ae2fd89c1ec7c18c1a1fe...	279648950.00	03	03
1119	f305bf95c6cc5cd48e048a49e2b87a9ed97f6fa3552c15...	597090000.00	02	02
1120	d6fd4427fda40fe44a213286d6ec58e8b42fa1c0d990cf...	700179800.00	02	02
1121	36fb6f2fe2f0570cc169d6fd537ed64d3f1168bf90e370...	2984527415.00	03	01

```
from sklearn.metrics import accuracy_score
print(accuracy_score(y['tco_cus_grd_cd'], y['class_label']))

0.6666666666666666
```

제공된 투자자의 고객등급과 유추한 투자자금으로 도출해낸 고객등급을 비교해본 결과 정확도 67%로 관측은 적중률 보임

02 분석 과정

ㄷ. 데이터 전처리

투자자 타겟팅

ㄷ. 투자자금이 100만원 ~ 5천만원인 투자자만 취급



투자자금 분포를 보면
아웃라이어 값들이 발견됨

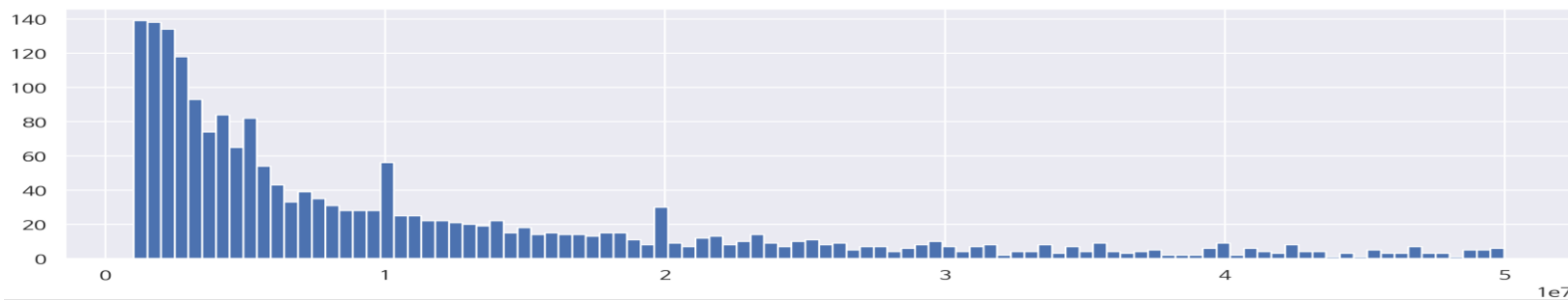
연령별 1인당 총자산 현황 (단위 : 원)



모 은행에서 발표한 20대 30대 평균자산
그러나 증여, 상속재산이 반영된 결과로 신뢰도가 크게 떨어짐

※출처 : 신한은행

Q3 + 1.5*IQR 값이 4610만원이라는 점과 20대 미혼의 평균 자산이 9700만원이라는 보고를 반영하여
100만원 이상 5천만원 이하의 투자자만 분석



추려낸 투자자의
투자자금 분포

03. 분석 결과

ANOVA, 클러스터링, 유형 분석

03 분석 결과

ㄱ. ANOVA

수익률

전체 매매비용 / 투자자금 * 100(%)

※아직 매도하지 않고 보유하고 있는 수량은 전체 거래내역중 가장 최근 거래 가격으로 계산

매매회전율

거래 금액 / (투자자금*2) * 100(%)

이탈 기간

2020년 7월 1일 기준 마지막 거래한 내역으로부터 지난 기간

세 가지 요인을 중점적으로 유입/세대 간 유의미한 차이점 도출

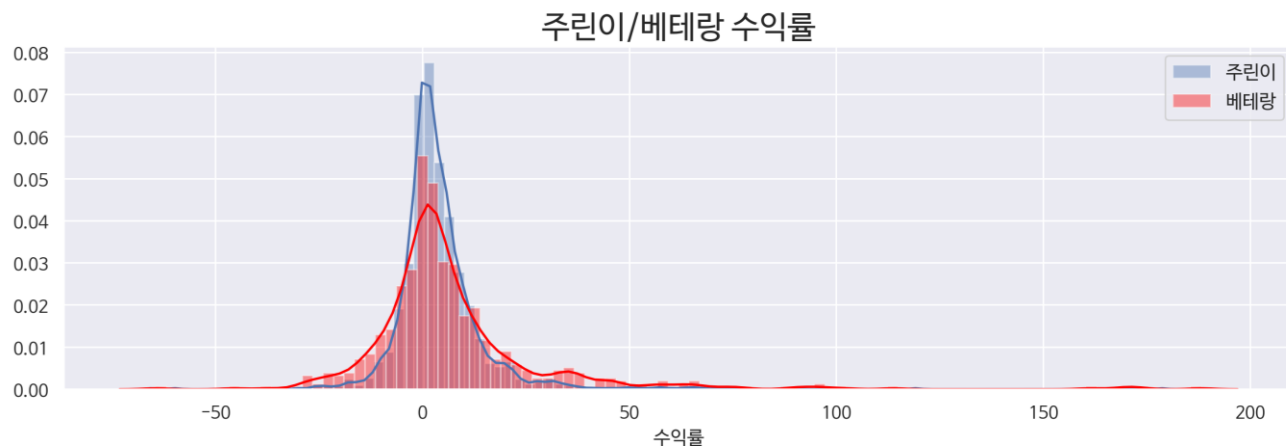
03 분석 결과

1. ANOVA

수익률

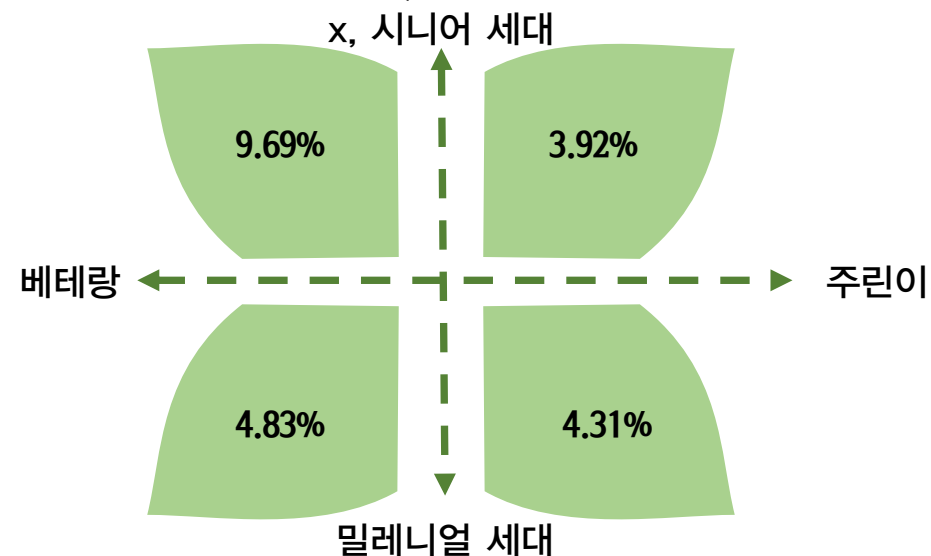
가설1 : 베테랑의 수익률이 주린이의 수익률보다 높을 것이다

<주린이/베테랑 수익률>



주린이의 수익률 편차가 더 적고 0%대에 몰려있는 반면,
베테랑의 경우 상대적으로 수익률 편차가 크다

<집단 별 수익률 비교>



전반적으로 베테랑의 수익률이 주린이의 수익률보다 높다
특히 시니어 세대에서 베테랑의 수익률이 약 2.5배 정도 차이가 난다

두 집단의 수익률 차이를 보기 위해 T-test 시행

```
from scipy import stats
s, p = stats.levene(new_profit['수익률'], old_profit['수익률'])
print('leven 등분산 검정 결과 p_value값이 {0:.2f}로 0.05보다 작아 기존 귀무가설을 기각, 이분산을 가정한다'.format(p))
s, p = stats.ttest_ind(new_profit['수익률'], old_profit['수익률'], equal_var=False)
print('이분산 가정 T검정 결과 pvalue 값이 {0:.2f}로 0.05보다 작으므로 두 집단간의 수익률에 유의미한 차이가 있다'.format(p))
```

leven 등분산 검정 결과 p_value값이 0.00로 0.05보다 작아 기존 귀무가설을 기각, 이분산을 가정한다
이분산 가정 T검정 결과 pvalue 값이 0.00로 0.05보다 작으므로 두 집단간의 수익률에 유의미한 차이가 있다

➡ T검정에 앞서 levene 방법으로 등분산 검정을 해본 결과,
등분산성을 기각하고 표본의 개수가 충분하므로 이분산 가정 T 검정 시행

가설 1 검정 완료

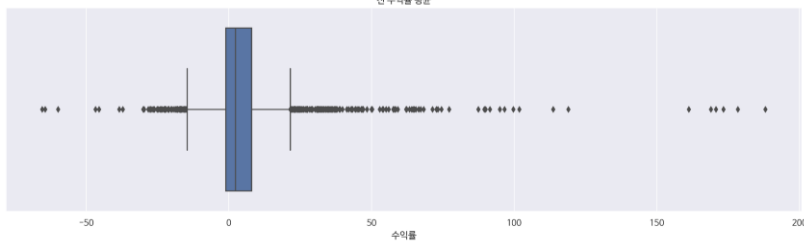
03 분석 결과

7. ANOVA

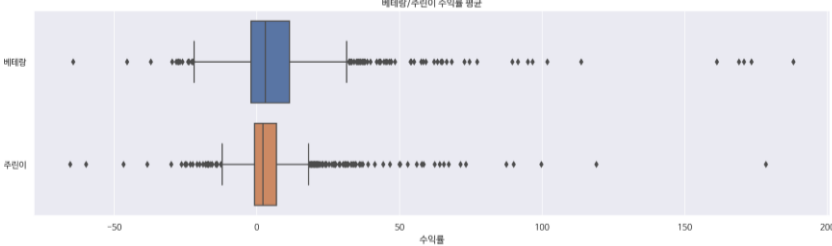
수익률

가설2 : 세대 간 수익률 차이가 존재할 것이다

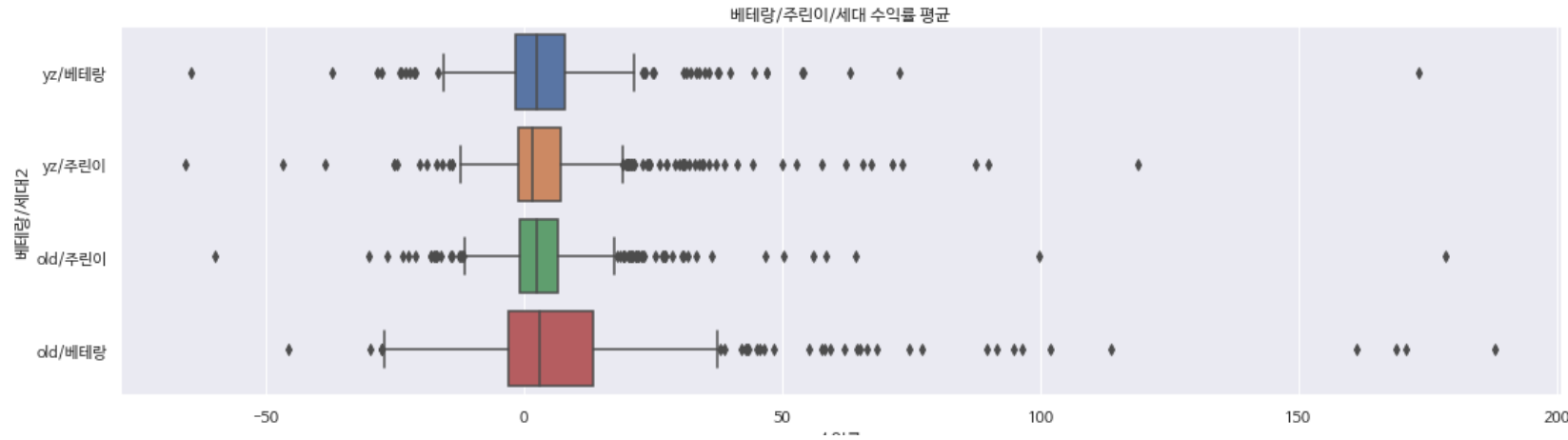
〈모든 투자자 수익률 평균〉



〈베테랑/주린이 투자자 수익률 평균〉



〈베테랑/주린이/세대 투자자 수익률 평균〉



가설 2 검정 완료

03 분석 결과

ㄱ. ANOVA

수익률

가설2 : 세대 간 수익률 차이가 존재할 것이다

네 집단의 수익률 차이를 보기 위해 ANOVA 시행

```
s, p = stats.levene(old_v_gr, yz_v_gr, old_j_gr, yz_j_gr)
print('네 집단의 등분산 검정을 해본 결과 pvalue 값이 {:.2f}로 유의수준 5%에서 등분산성이 기각되었다'.format(p))
```

네 집단의 등분산 검정을 해본 결과 pvalue 값이 0.00로 유의수준 5%에서 등분산성이 기각되었다

ANOVA 분석에 앞서 등분산성 검정 -> 결과 등분산성 기각

```
group = [old_v_gr, yz_v_gr, old_j_gr, yz_j_gr]
for g in group:
    print('shapiro 정규성 검정 pvalue 값 ', stats.shapiro(g)[1])
```

```
shapiro 정규성 검정 pvalue 값 8.408178482368096e-25
shapiro 정규성 검정 pvalue 값 5.66763153088594e-20
shapiro 정규성 검정 pvalue 값 1.8789083336444212e-34
shapiro 정규성 검정 pvalue 값 1.1328583441984977e-33
```

ANOVA 분석에 앞서 정규성 검정 -> 결과 정규성 기각

```
from statsmodels.sandbox.stats.multicomp import MultiComparison
import scipy.stats

comp = MultiComparison(real_all_profit['수익률'], real_all_profit['베테랑/세대2'])
result = comp.allpairtest(scipy.stats.ttest_ind, method='bonf')
result[0]
```

Test Multiple Comparison ttest_ind FWER=0.05 method=bonf
alphacSidak=0.01, alphacBonf=0.008

group1	group2	stat	pval	pval_corr	reject
베테랑/old	베테랑/yz	2.4758	0.0136	0.0814	False
베테랑/old	주린이/old	4.6933	0.0	0.0	True
베테랑/old	주린이/yz	4.5491	0.0	0.0	True
베테랑/yz	주린이/old	0.8495	0.3958	1.0	False
베테랑/yz	주린이/yz	0.4871	0.6263	1.0	False
주린이/old	주린이/yz	-0.592	0.554	1.0	False

```
from pingouin import welch_anova

aov = welch_anova(dv='수익률', between='베테랑/세대2', data=real_all_profit)
aov
```

	Source	ddof1	ddof2	F	p-unc	np2
0	베테랑/세대2	3	724.87	5.11	0.00	0.02

등분산성, 정규성이 기각되었으므로 ANOVA 방법 중 하나인 Welch test를 수행

-> pvalue가 0.00으로 네 집단 중 수익률의 차이가 존재함을 증명

Bonferonni 방법을 채택하여 사후검정 수행

베테랑/old 와 주린이/old & 주린이/yz

간의 유의미한 수익률 차이를 검정할 수 있었다.

가설 2 검정 완료

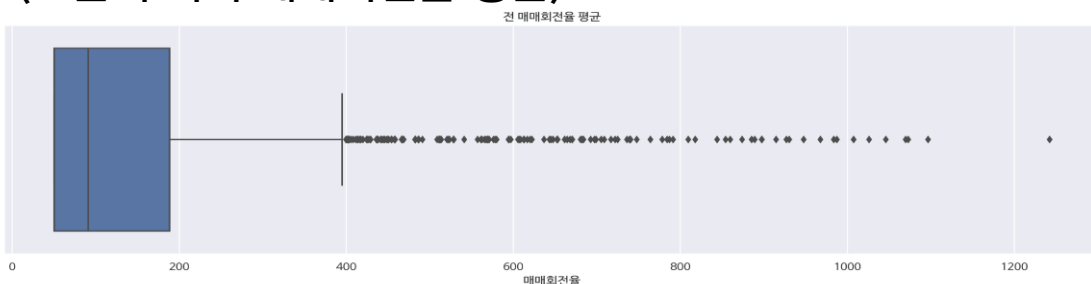
03 분석 결과

1. ANOVA

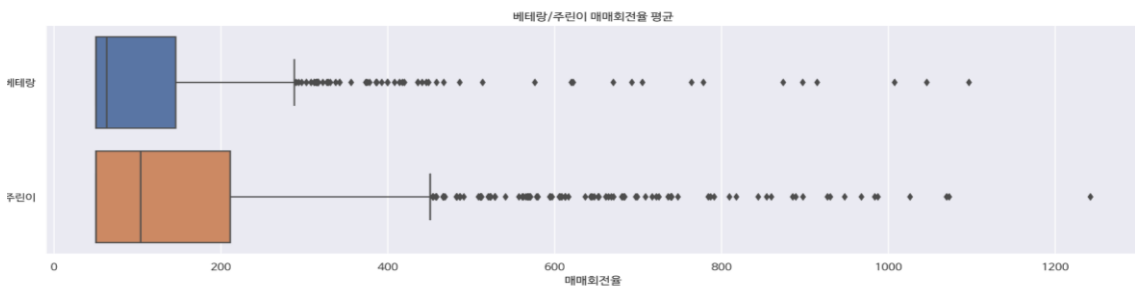
매매회전율

가설3 : 주린이/베테랑 간 매매회전율 차이가 존재할 것이다

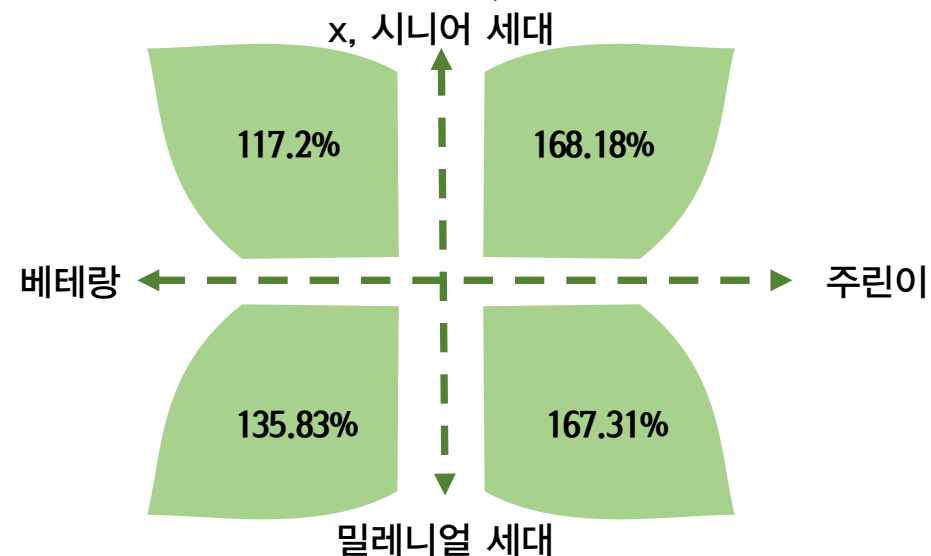
<모든 투자자 매매회전율 평균>



<베테랑/주린이/세대 투자자 매매회전율 평균>



<집단 별 매매회전율 비교>



전반적으로 주린이의 매매회전율이 베테랑보다 높다

두 집단의 매매회전율 차이를 보기 위해 T-test 시행

```
s, p = stats.levene(real_all_profit[real_all_profit['베테랑'] == '베테랑']['매매회전율'], real_all_profit[real_all_profit['베테랑'] == '주린이']['매매회전율'])
print('leven 등분산 검정 결과 p_value값이 {0:.2f}로 0.05보다 작아 기존 귀무가설을 기각, 이분산을 가정한다'.format(p))
s, p = stats.ttest_ind(real_all_profit[real_all_profit['베테랑'] == '베테랑']['매매회전율'], real_all_profit[real_all_profit['베테랑'] == '주린이']['매매회전율'])
print('이분산 가정 t검정 결과 pvalue 값이 {0:.2f}로 0.05보다 작으므로 두 집단간의 수익률에 유의미한 차이가 있다'.format(p))
```

leven 등분산 검정 결과 p_value값이 0.00로 0.05보다 작아 기존 귀무가설을 기각, 이분산을 가정한다
이분산 가정 t검정 결과 pvalue 값이 0.00로 0.05보다 작으므로 두 집단간의 수익률에 유의미한 차이가 있다

➡ T검정에 앞서 levene 방법으로 등분산 검정을 해본 결과, 등분산성을 기각하고 표본의 개수가 충분하므로 이분산 가정 T 검정 시행

가설 3 검정 완료

03 분석 결과

ㄱ. ANOVA

매매회전율

가설4 : 주린이/베테랑/세대 간 매매회전율 차이가 존재할 것이다

네 집단의 매매회전율 차이를 보기 위해 ANOVA 시행

```
s, p = stats.levene(old_v_gr, yz_v_gr, old_j_gr, yz_j_gr)
print('네 집단의 등분산 검정을 해본 결과 pvalue 값이 {0:.2f}로 유의수준 5%에서 등분산성이 기각되었다'.format(p))
```

네 집단의 등분산 검정을 해본 결과 pvalue 값이 0.00로 유의수준 5%에서 등분산성이 기각되었다

ANOVA 분석에 앞서 등분산성 검정 -> 결과 등분산성 기각

```
group = [old_v_gr, yz_v_gr, old_j_gr, yz_j_gr]
for g in group:
    print('shapiro 정규성 검정 pvalue 값 ', stats.shapiro(g)[1])
```

```
shapiro 정규성 검정 pvalue 값 5.437241483717188e-29
shapiro 정규성 검정 pvalue 값 6.4044055187906876e-24
shapiro 정규성 검정 pvalue 값 4.4538173215191184e-33
shapiro 정규성 검정 pvalue 값 9.25536497379467e-33
```

ANOVA 분석에 앞서 정규성 검정 -> 결과 정규성 기각

Test Multiple Comparison ttest_ind FWER=0.05 method=bonf
alphacSidak=0.01, alphacBonf=0.008

group1	group2	stat	pval	pval_corr	reject
베테랑/old	베테랑/yz	-1.6177	0.1062	0.6375	False
베테랑/old	주린이/old	-4.8577	0.0	0.0	True
베테랑/old	주린이/yz	-5.1926	0.0	0.0	True
베테랑/yz	주린이/old	-2.5354	0.0114	0.0684	False
베테랑/yz	주린이/yz	-2.6745	0.0076	0.0456	True
주린이/old	주린이/yz	0.1309	0.8959	1.0	False

```
aov = welch_anova(dv='매매회전율', between='베테랑/세대2', data=real_all_profit)
aov
```

	Source	ddof1	ddof2	F	p-unc	np2
0	베테랑/세대2	3	832.72	13.57	0.00	0.02

등분산성, 정규성이 기각되었으므로 ANOVA 방법 중 하나인 Welch test를 수행

-> pvalue가 0.00으로 네 집단 중 매매회전율의 차이가 존재함을 증명

Bonferonni 방법을 채택하여 사후검정 수행

베테랑/old 와 주린이 전 세대
베테랑/yz 와 주린이/yz

간의 유의미한 매매회전율 차이를 검정할 수 있었다.

가설 4 검정 완료

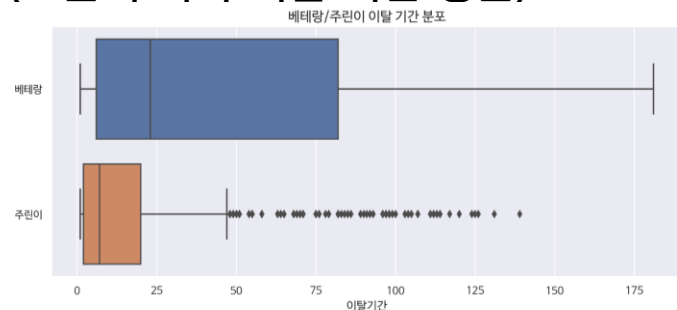
03 분석 결과

7. ANOVA

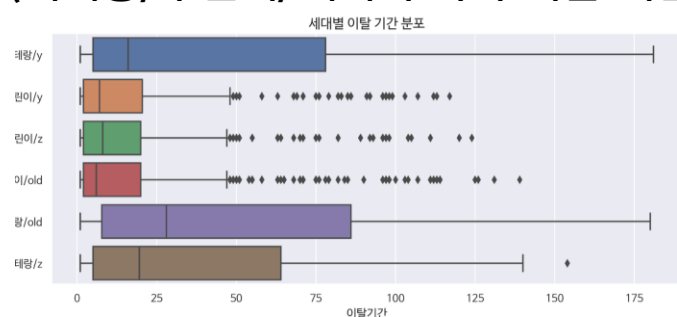
이탈 기간

가설5 : 주린이/베테랑 간 이탈기간 차이가 존재할 것이다

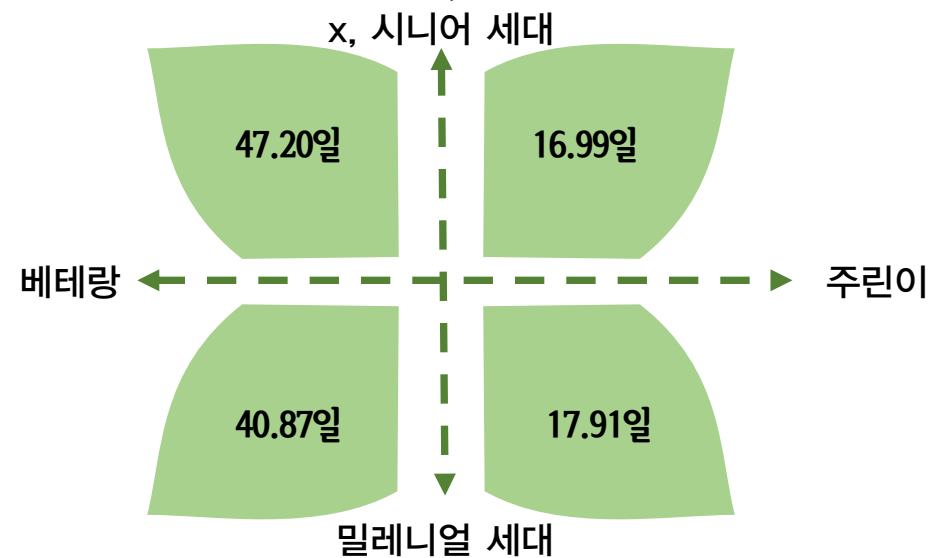
<모든 투자자 이탈 기간 평균>



<베테랑/주린이/세대 투자자 이탈 기간 평균>



<집단 별 이탈기간 비교>



베테랑과 주린이 간의 이탈 기간 차이가 약 2~3배로 매우 크다

두 집단의 이탈기간 차이를 보기 위해 T-test 시행

```
s, p = stats.levene(real_all_profit[real_all_profit['베테랑'] == '베테랑']['이탈기간'], real_all_profit[real_all_profit['베테랑'] == '주린이']['이탈기간'])
print('leven 등분산 검정 결과 p_value값이 {0:.2f}로 0.05보다 작아 기존 귀무가설을 기각, 이분산을 가정한다'.format(p))
s, p = stats.ttest_ind(real_all_profit[real_all_profit['베테랑'] == '베테랑']['이탈기간'], real_all_profit[real_all_profit['베테랑'] == '주린이']['이탈기간'])
print('이분산 가정 T검정 결과 pvalue 값이 {0:.2f}로 0.05보다 작으므로 두 집단간의 이탈기간에 유의미한 차이가 있다'.format(p))
```

leven 등분산 검정 결과 p_value값이 0.00로 0.05보다 작아 기존 귀무가설을 기각, 이분산을 가정한다
이분산 가정 T검정 결과 pvalue 값이 0.00로 0.05보다 작으므로 두 집단간의 이탈기간에 유의미한 차이가 있다

➡ T검정에 앞서 levene 방법으로 등분산 검정을 해본 결과, 등분산성을 기각하고 표본의 개수가 충분하므로 이분산 가정 T 검정 시행

가설 5 검정 완료

03 분석 결과

ㄱ. ANOVA

이탈 기간

가설 6 : 주린이/베테랑/세대 간 이탈기간 차이가 존재할 것이다

네 집단의 이탈기간 차이를 보기 위해 ANOVA 시행

```
s, p = stats.levene(old_v_gr, yz_v_gr, old_j_gr, yz_j_gr)
print('네 집단의 등분산 검정을 해본 결과 pvalue 값이 {0:.2f}로 유의수준 5%에서 등분산성이 기각되었다'.format(p))
```

네 집단의 등분산 검정을 해본 결과 pvalue 값이 0.00로 유의수준 5%에서 등분산성이 기각되었다

ANOVA 분석에 앞서 등분산성 검정 -> 결과 등분산성 기각

```
group = [old_v_gr, yz_v_gr, old_j_gr, yz_j_gr]
for g in group:
    print('shapiro 정규성 검정 pvalue 값 ', stats.shapiro(g)[1])
```

```
shapiro 정규성 검정 pvalue 값 1.0016492164804673e-17
shapiro 정규성 검정 pvalue 값 4.9568801515962634e-17
shapiro 정규성 검정 pvalue 값 2.408454074599217e-34
shapiro 정규성 검정 pvalue 값 5.730813910578231e-35
```

ANOVA 분석에 앞서 정규성 검정 -> 결과 정규성 기각

Test Multiple Comparison ttest_ind FWER=0.05 method=bonf
alphacSidak=0.01, alphacBonf=0.008

group1	group2	stat	pval	pval_corr	reject
베테랑/old	베테랑/yz	1.633	0.103	0.6179	False
베테랑/old	주린이/old	13.3836	0.0	0.0	True
베테랑/old	주린이/yz	13.6309	0.0	0.0	True
베테랑/yz	주린이/old	9.7243	0.0	0.0	True
베테랑/yz	주린이/yz	9.7853	0.0	0.0	True
주린이/old	주린이/yz	-0.6872	0.4921	1.0	False

```
aov = welch_anova(dv='이탈기간', between='베테랑/세대2', data=real_all_profit)
aov
```

	Source	ddof1	ddof2	F	p-unc	np2
0	베테랑/세대2	3	714.48	61.93	0.00	0.13

등분산성, 정규성이 기각되었으므로 ANOVA 방법 중 하나인 Welch test를 수행

-> pvalue가 0.00으로 네 집단 중 매매회전율의 차이가 존재함을 증명

Bonferonni 방법을 채택하여 사후검정 수행

베테랑 모든 세대 와 주린이 모든 세대

간의 유의미한 이탈기간 차이를 검정할 수 있었다.

가설 6 검정 완료

03 분석 결과

ㄱ. ANOVA

■ 소론

- 수익률, 매매회전율, 이탈기간에서 모두 집단 간 유의미한 차이가 존재
- 수익률의 경우 베테랑이 주린이보다 3.5% 가량 높았으며 특히 베테랑 중에서도 old 세대의 수익률이 다른 주린이들에 비해 큰 차이로 높음
- 매매회전율 같은 경우에는 주린이가 베테랑보다 40% 가량 더 높은 수치를 보였으며 주목할만한 점은 베테랑/old의 경우 큰 차이로 수익률이 높았지만 반대로 매매회전율 측면에서는 낮은 경향을 보임
- 이에 두 요인간에 상관관계분석을 해본 결과 -0.11 로 작지만 음의 상관관계를 확인할 수 있었음
- 마지막으로 이탈기간의 경우, 베테랑이 주린이보다 20일 가량 더 길었으며 이것이 장투의 영향인지 정말 이탈한 것인지는 추후 추가적인 분석 必

03 분석 결과

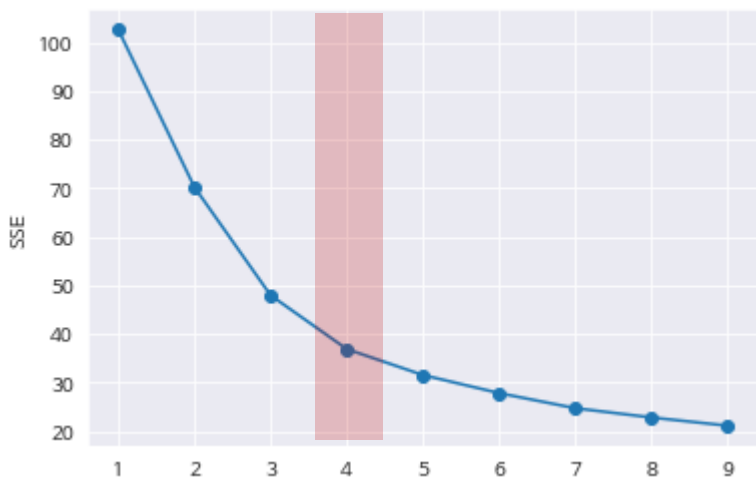
└. 클러스터링

주린이/베테랑 세대 간 K-means 군집분석

〈사용 변수〉

- 수익률(%) = (매매비용 결산) / (매수비용) * 100
- 매매회전율(%) = (매매비용 총합)/(투자금액*2) * 100
- 투자자금(원)
- 매수 종목 수(개)
- 이탈기간(일)(2020년 7월1일 기준 마지막 거래일과 비교)

고객번호	수익률	지난 시간	투자자금	매매회전율	이탈기간	매수 종목 수
f16de38ef7fc3b1fb67773cedcf7c913...	17.80	75 days	9994800.00	78.71	75	2
1b52a3e96d4695e4f99e5e254269c1...	1.52	2 days	4978800.00	50.00	2	1
d59c3f0af1fe875aa5b41cf353f10010...	10.10	51 days	2134000.00	50.00	51	1
aaaa67e9ceed4610e7c3316ddf1f544...	1.52	1 days	31176046.00	967.75	1	97
db8b0fcd7a77f7ad8268501577e80d...	0.23	5 days	3916800.00	349.16	5	18



※ Elbow 방식

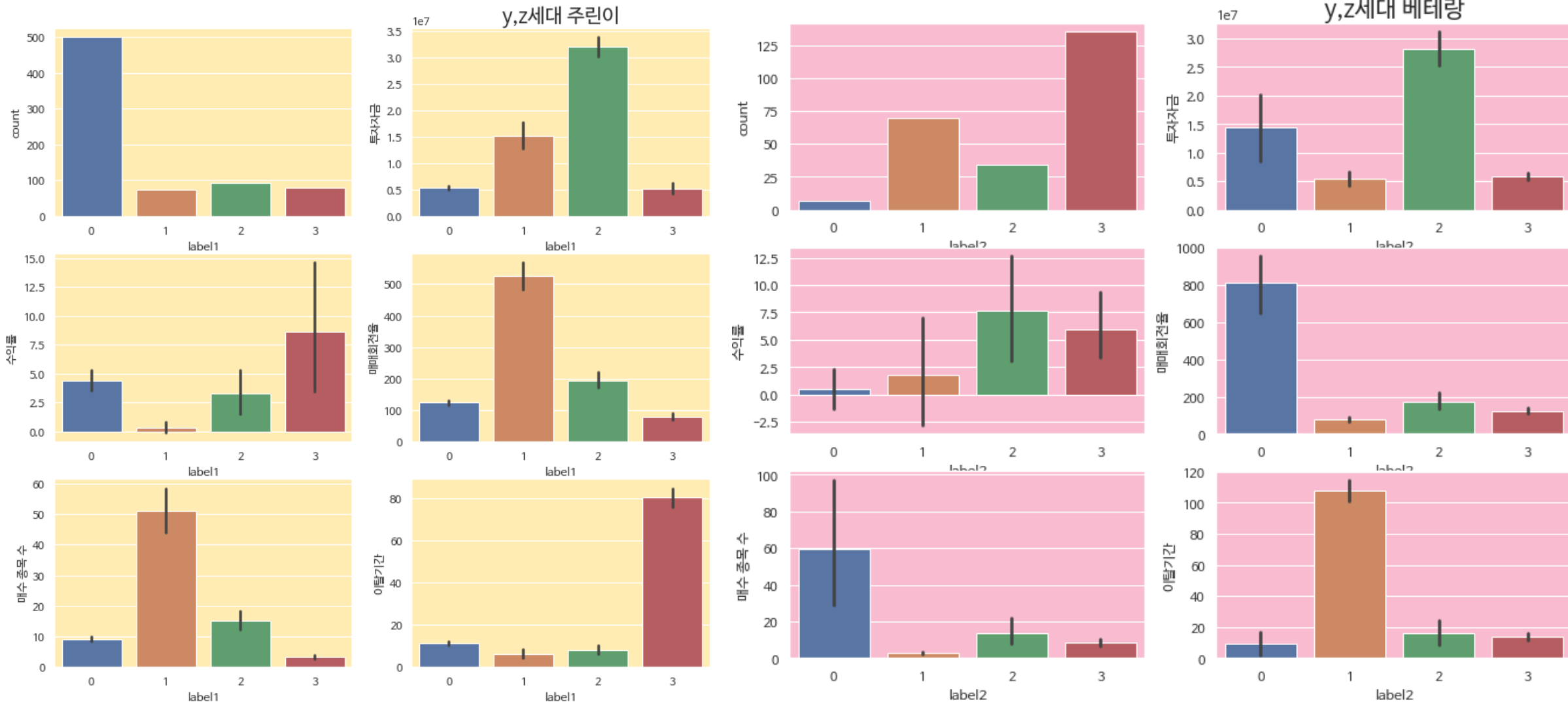
〈군집분석 이전 선행작업〉

- (원), (%), (개) 등의 수준을 맞춰주기 위한 스케일링 작업(MinMaxScaler 사용)
- elbow를 통한 적정 군집수 책정 (4개 집단 모두 4개의 군집을 적정수로 정함)

03 분석 결과

└. 클러스터링

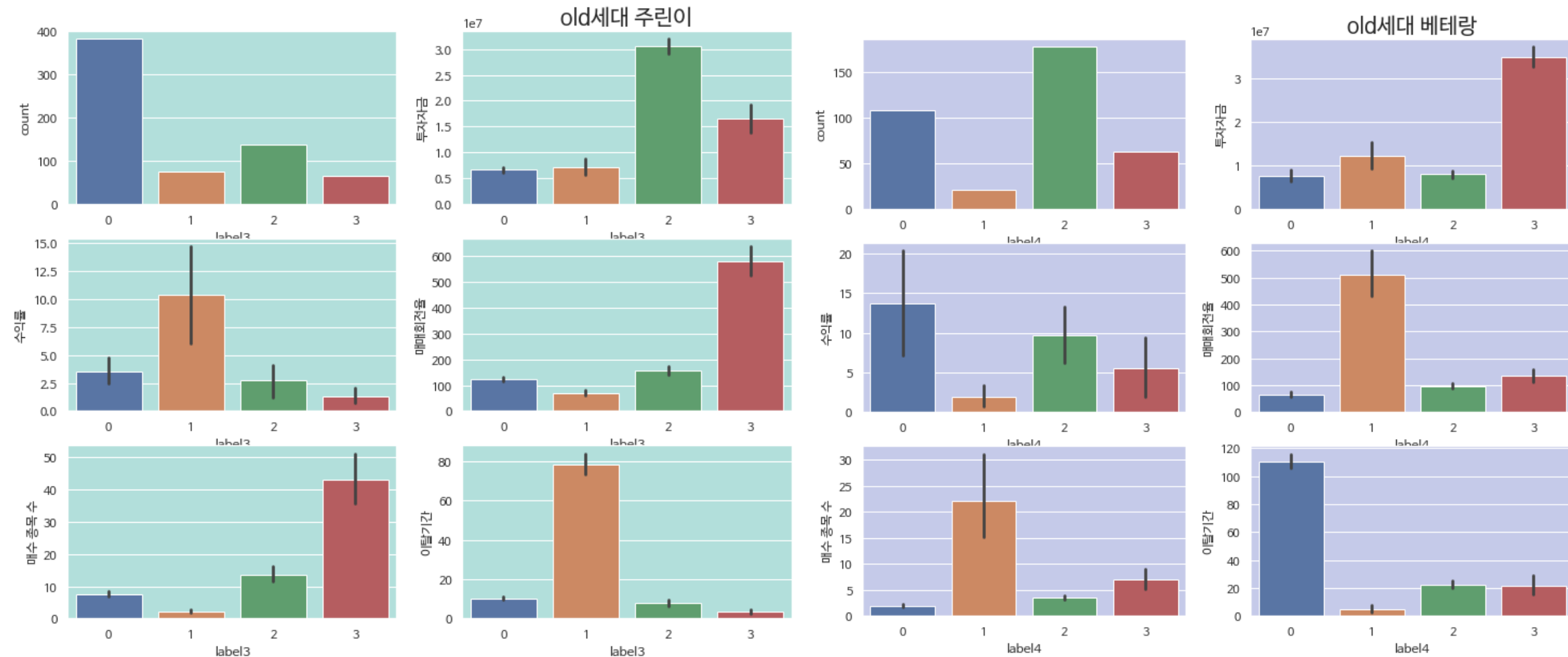
주린이/베테랑 세대 간 K-means 군집분석



03 분석 결과

└. 클러스터링

■ 주린이/베테랑 세대 간 K-means 군집분석



03 분석 결과

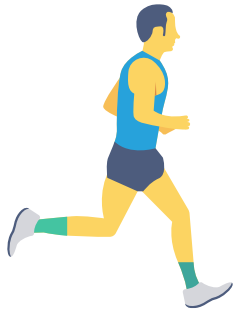
↳ 클러스터링

소론

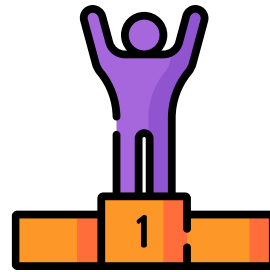
〈기본적으로 네 집단 모두 비슷한 양상으로 군집화 됨을 확인〉



A : 부자 투자자
자본금 3천 언저리.
2~5% 수익률



B : 부지런한 중형 투자자
자본금 1천 ~ 1천5백
엄청 돈을 굴리지만 수익률은 꼴찌



C : 승리자 소시민
자본금 500 언저리
적은 시간 투자, 높은 수익률
최고 승리자



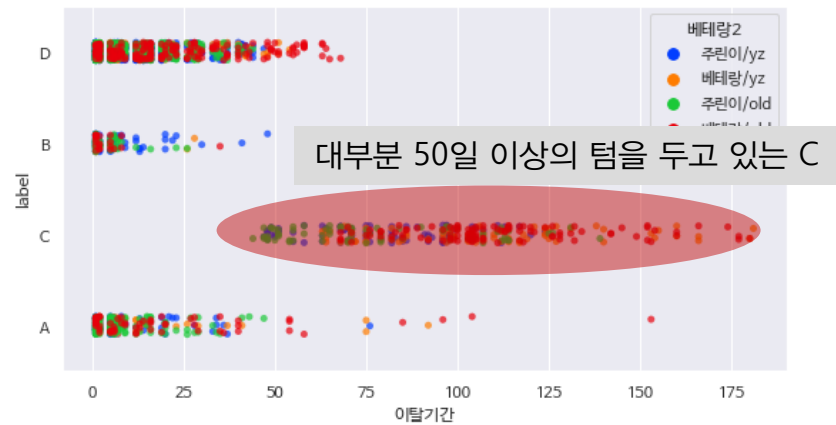
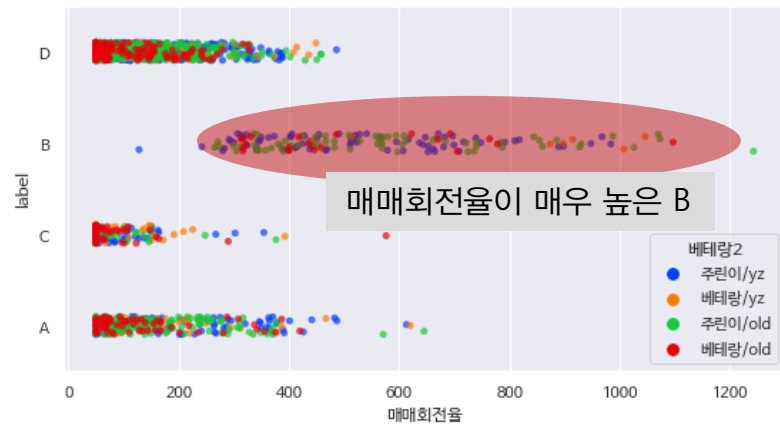
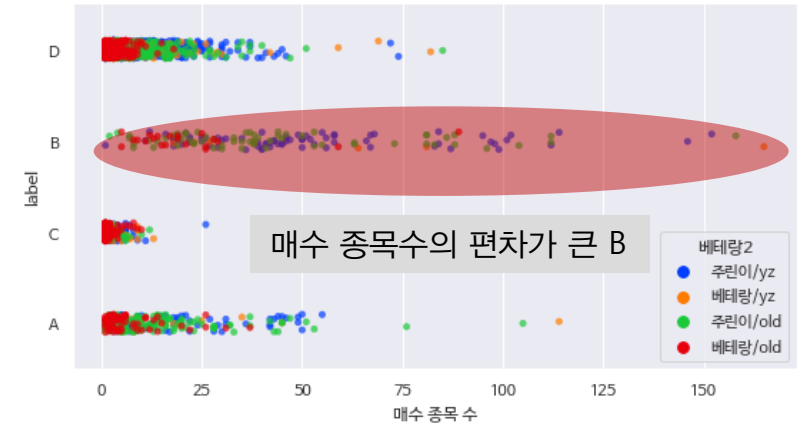
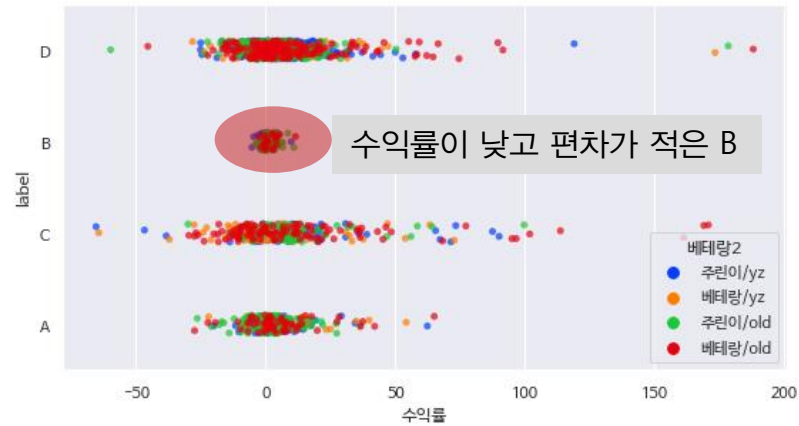
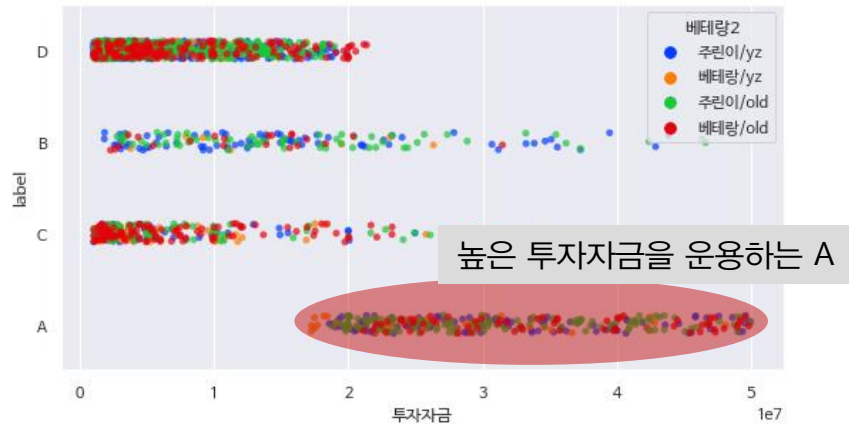
D : 꾸준한 소시민
자본금 500 언저리 여전히 투자 中
코로나장에 흐름을 타
돈을 벌고 싶은 소시민인 우리들.

집단	군집	수익률(%)	매매회전율(%)	매수종목수(개)	이탈기간(일)	집단	군집	수익률(%)	매매회전율(%)	매수종목수(개)	이탈기간(일)
주린이/ YZ세대	A	2~6	200	15~20	10	주린이/ OLD세대	A	1~4	170	12~15	10
	B	0~1	500	40~50	10		B	2	600	35~50	3
	C	4~14	100	3~5	80		C	7~15	80	3	80
	D	3~7	100	10	10		D	2~5	100	8	10
베테랑/ YZ세대	A	2.7~15	150~220	8~22	25	베테랑/ OLD세대	A	2.5~10	120	7	20~25
	B	0~4	450~700	30~65	20		B	1~3	420~600	15~30	3
	C	-2.5~7	80	3	100		C	7.5~20	80	3	100
	D	2.5~10	100	10	20		D	6~12.5	100	4	20

03 분석 결과

ㄷ. 유형 분석

유형 별 요인 분포



03 분석 결과

ㄷ. 유형 분석

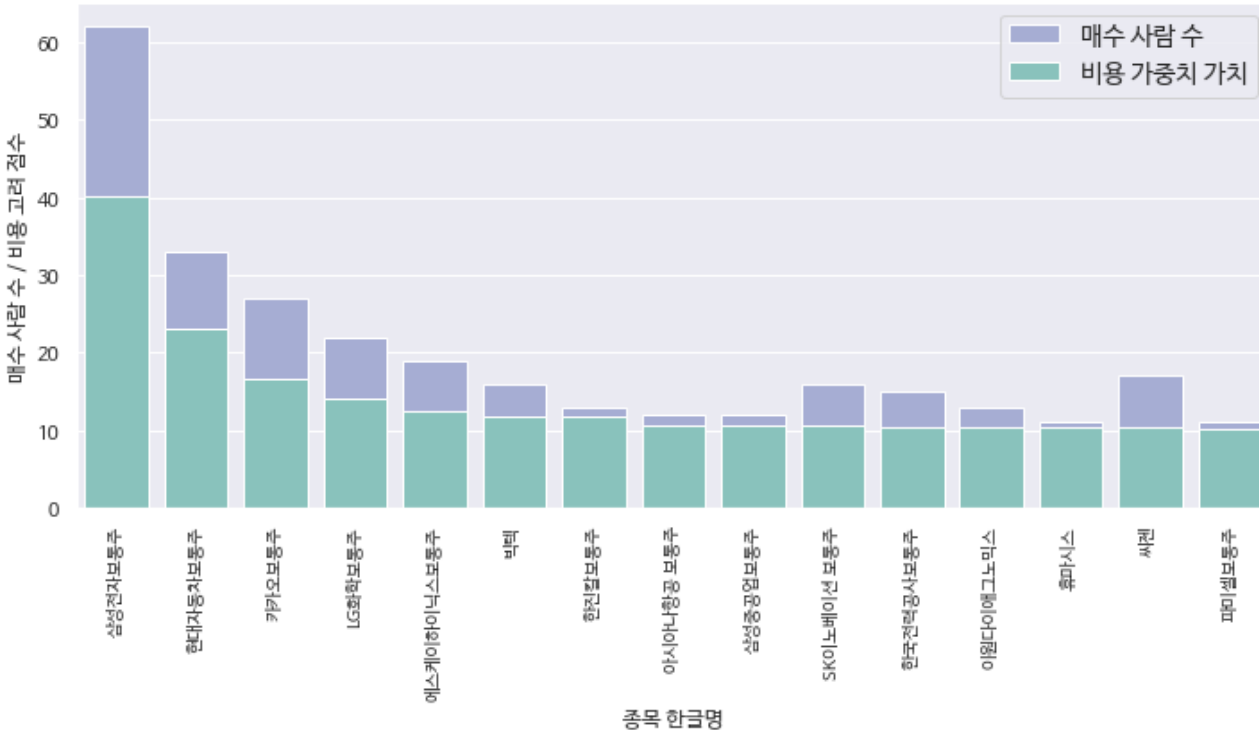
■ 유형 별 선호 종목 분석



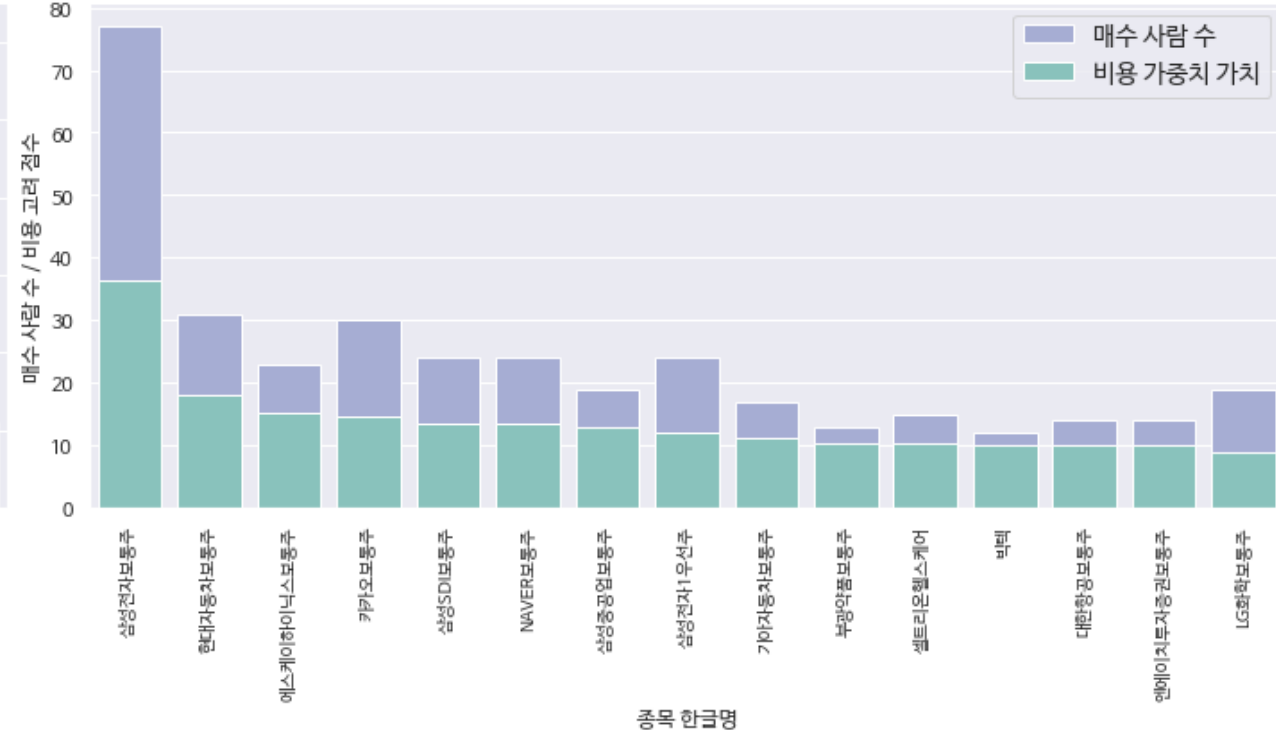
〈두 가지 측면에서 분석〉

- ㄱ. 단순 매수 여부(하늘색) : 해당 종목을 매수한 적이 한 번이라도 있으면 +1
- ㄴ. 비용 가중치 점수(녹색) : 사람마다 특정 종목에 투자한 비중이 다르므로 각 종목별 매수에 쓴 비용을 정규화(MinMax) 작업 후 더해준 값(운용 투자자금의 편차를 줄여주기 위함)

부자 투자자(A) 성향 yz세대 투자자들의 선호종목



부자 투자자(A) 성향 old세대 투자자들의 선호종목



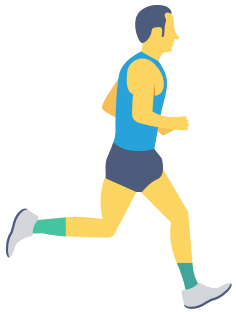
03 분석 결과

ㄷ. 유형 분석

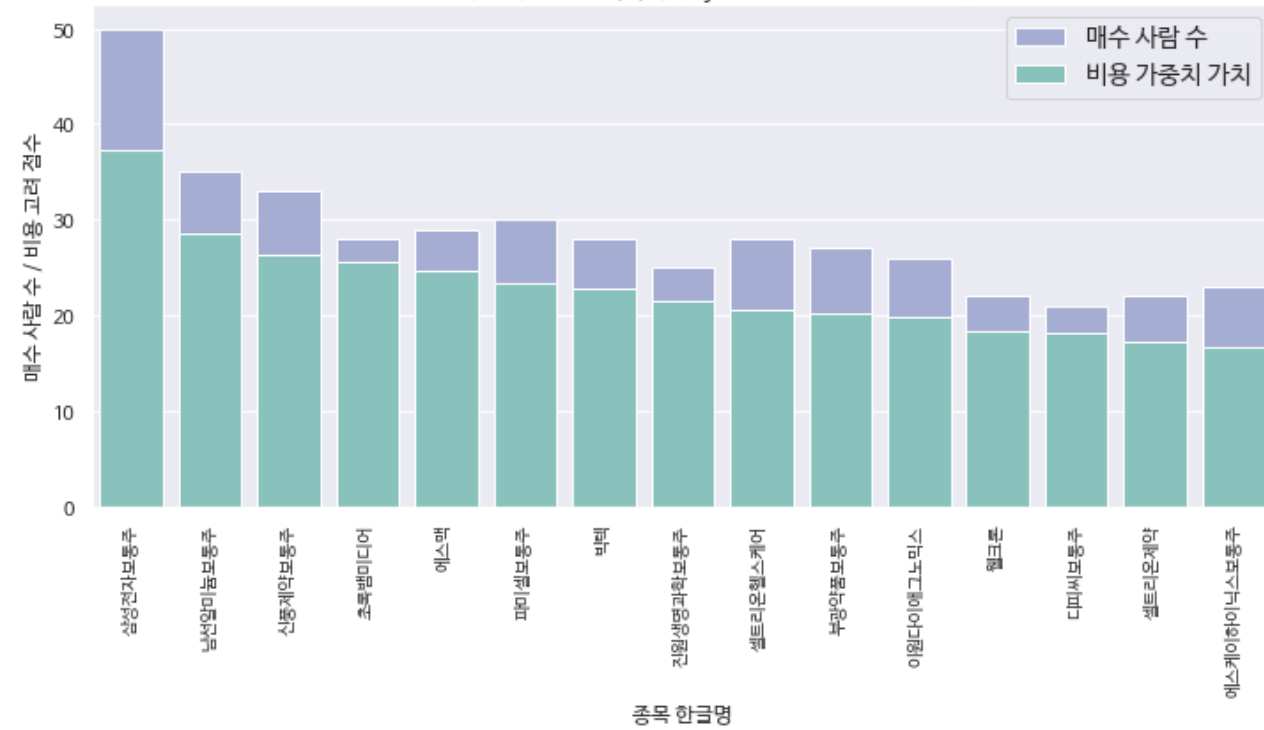
유형 별 선호 종목 분석

〈두 가지 측면에서 분석〉

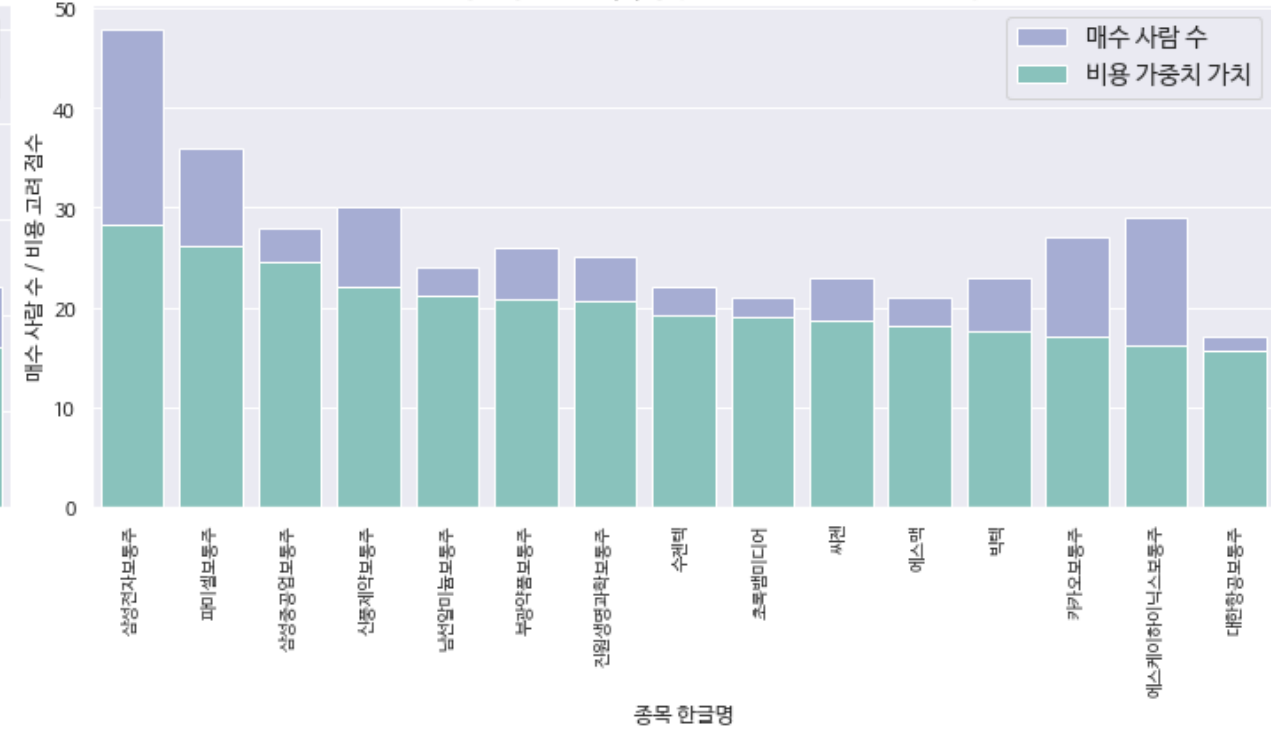
- ㄱ. 단순 매수 여부(하늘색) : 해당 종목을 매수한 적이 한 번이라도 있으면 +1
- ㄴ. 비용 가중치 점수(녹색) : 사람마다 특정 종목에 투자한 비중이 다르므로 각 종목별 매수에 쓴 비용을 정규화(MinMax) 작업 후 더해준 값(운용 투자자금의 편차를 줄여주기 위함)



부지런한 중산층 투자자(B) 성향 yz세대 투자자들의 선호종목



부지런한 중산층 투자자(B) 성향 old세대 투자자들의 선호종목

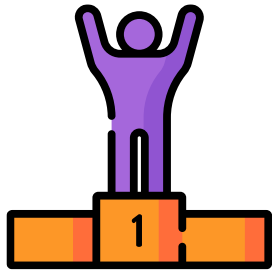


03 분석 결과

ㄷ. 유형 분석

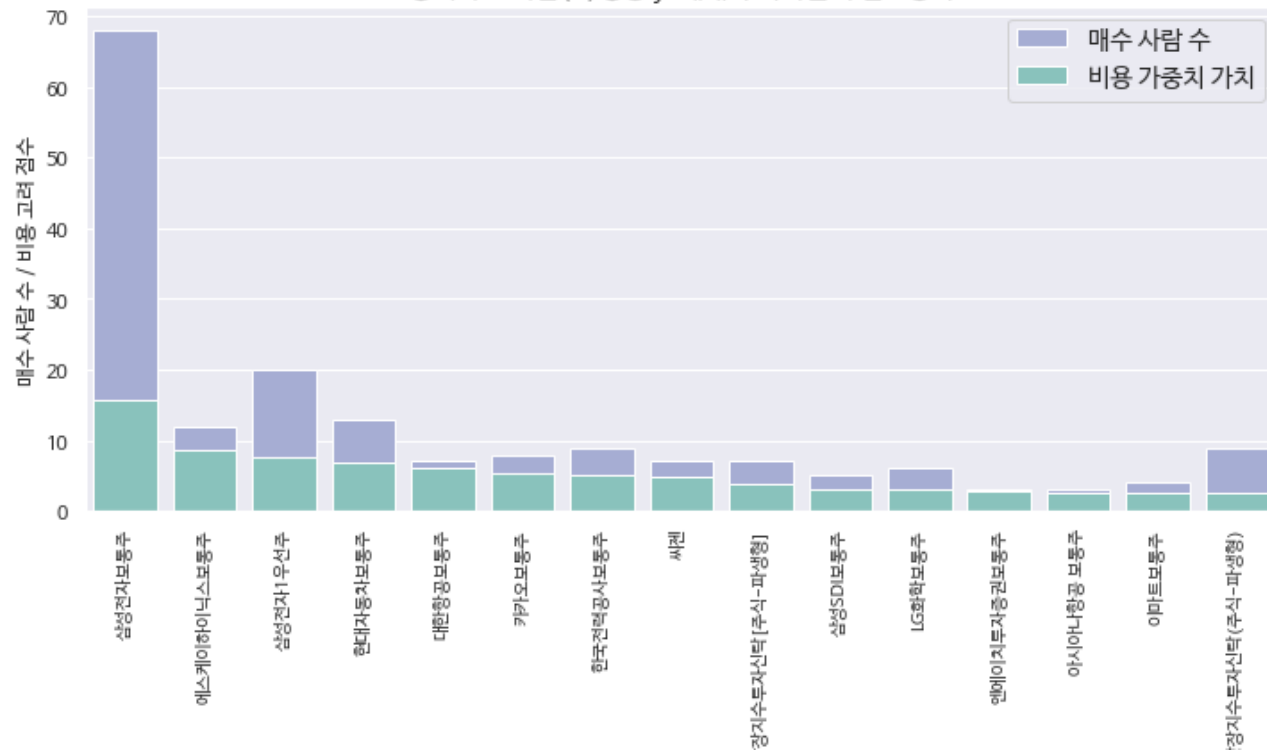
유형 별 선호 종목 분석

〈두 가지 측면에서 분석〉

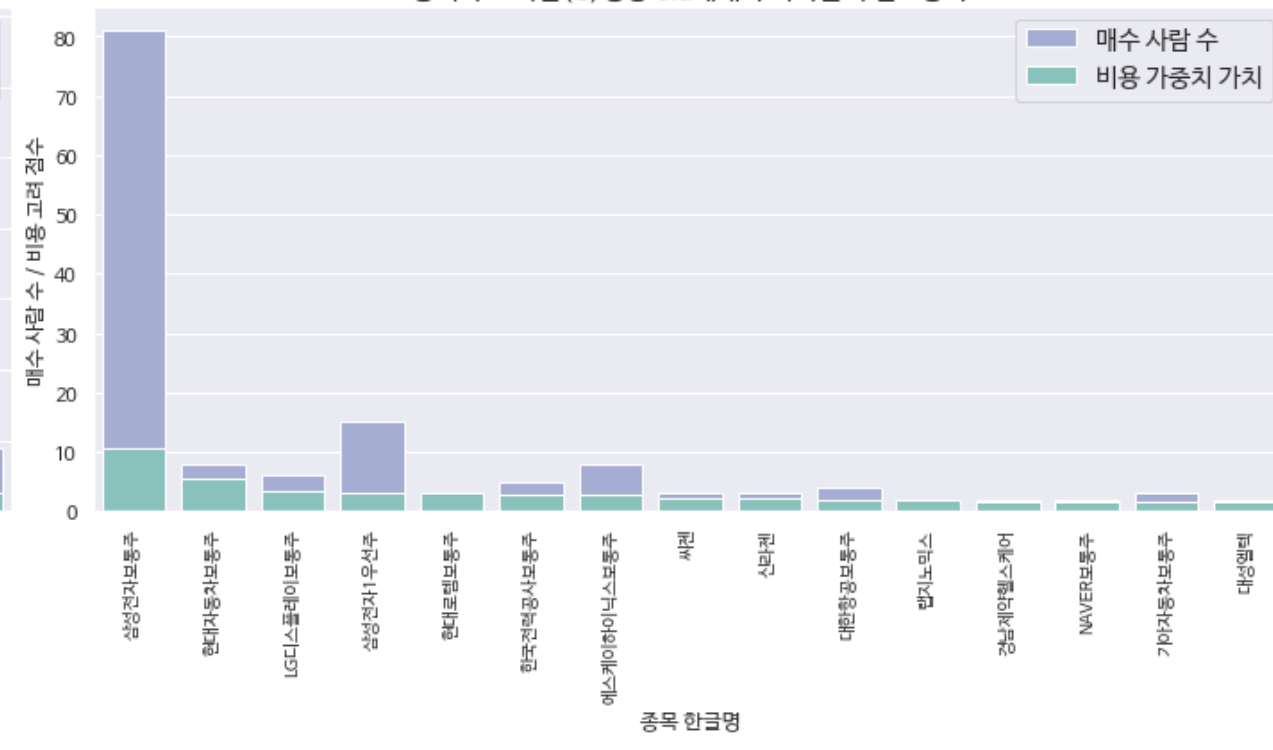


- ㄱ. 단순 매수 여부(하늘색) : 해당 종목을 매수한 적이 한 번이라도 있으면 +1
- ㄴ. 비용 가중치 점수(녹색) : 사람마다 특정 종목에 투자한 비중이 다르므로 각 종목별 매수에 쓴 비용을 정규화(MinMax) 작업 후 더해준 값(운용 투자자금의 편차를 줄여주기 위함)

승리자 소시민(C)성향 yz세대 투자자들의 선호종목



승리자 소시민(C)성향 old세대 투자자들의 선호종목



03 분석 결과

ㄷ. 유형 분석

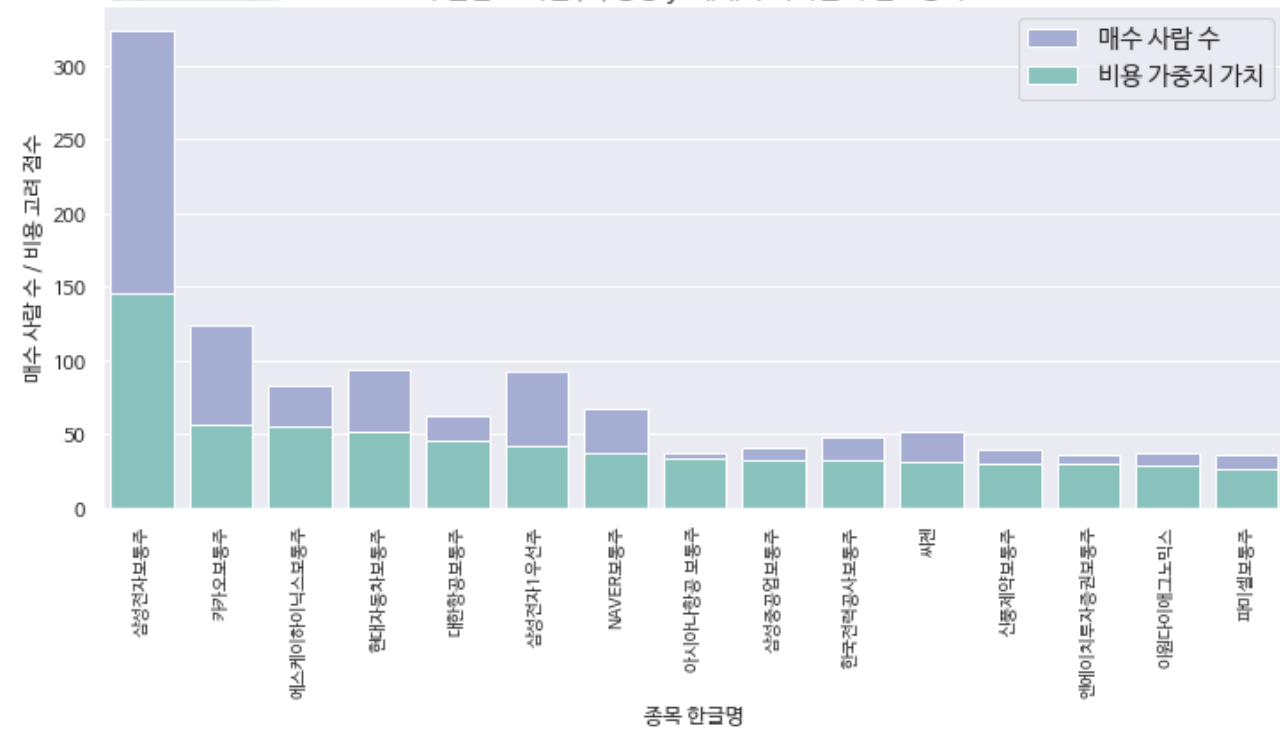
유형 별 선호 종목 분석

〈두 가지 측면에서 분석〉

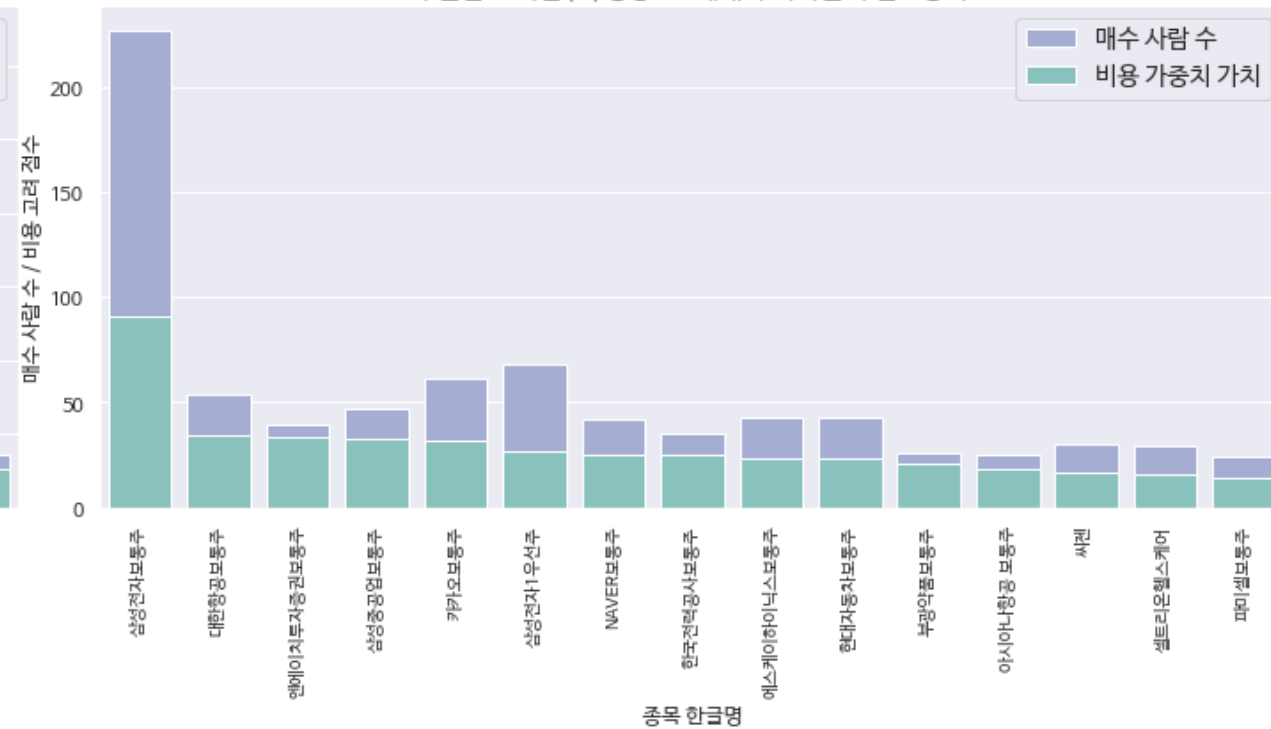
- ㄱ. 단순 매수 여부(하늘색) : 해당 종목을 매수한 적이 한 번이라도 있으면 +1
- ㄴ. 비용 가중치 점수(녹색) : 사람마다 특정 종목에 투자한 비중이 다르므로 각 종목별 매수에 쓴 비용을 정규화(MinMax) 작업 후 더해준 값(운용 투자자금의 편차를 줄여주기 위함)



중복 한금명
꾸준한 소시민(D)성향 yz세대 투자자들의 선호종목



꾸준한 소시민(D)성향 old세대 투자자들의 선호종목



03 분석 결과

ㄷ. 유형 분석

■ 소론

- 주목할만한 부분으로는 가장 높은 수익률을 보였던 C그룹의 선호 종목이 다른 그룹들과 차별됨
- 물론 다른 그룹들도 [삼성전자]의 선호도가 1순위였지만 유독 C그룹의 경우 [삼성전자]만(only) 매수한 것을 확인
- 또한 유일하게 C의 yz세대 경우, 인버스 레버리지 상품이 순위에 등극
- 반면 B의 경우 높은 매매회전율과 매수 종목 수를 보인 그룹이었는데 그에 따라 종목간 선호도 차이가 크게 없음
이들은 다양한 종목에 단타로 치고 빠지는 식의 투자자로 보임

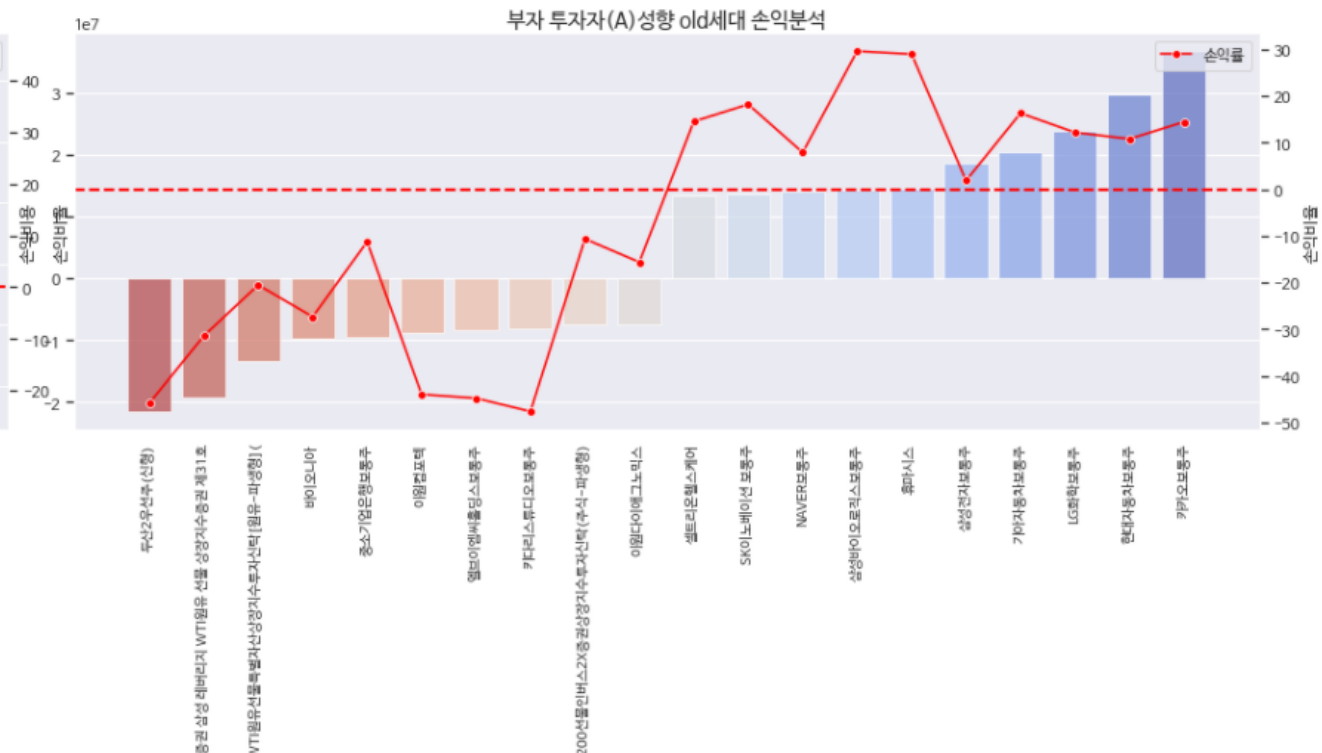
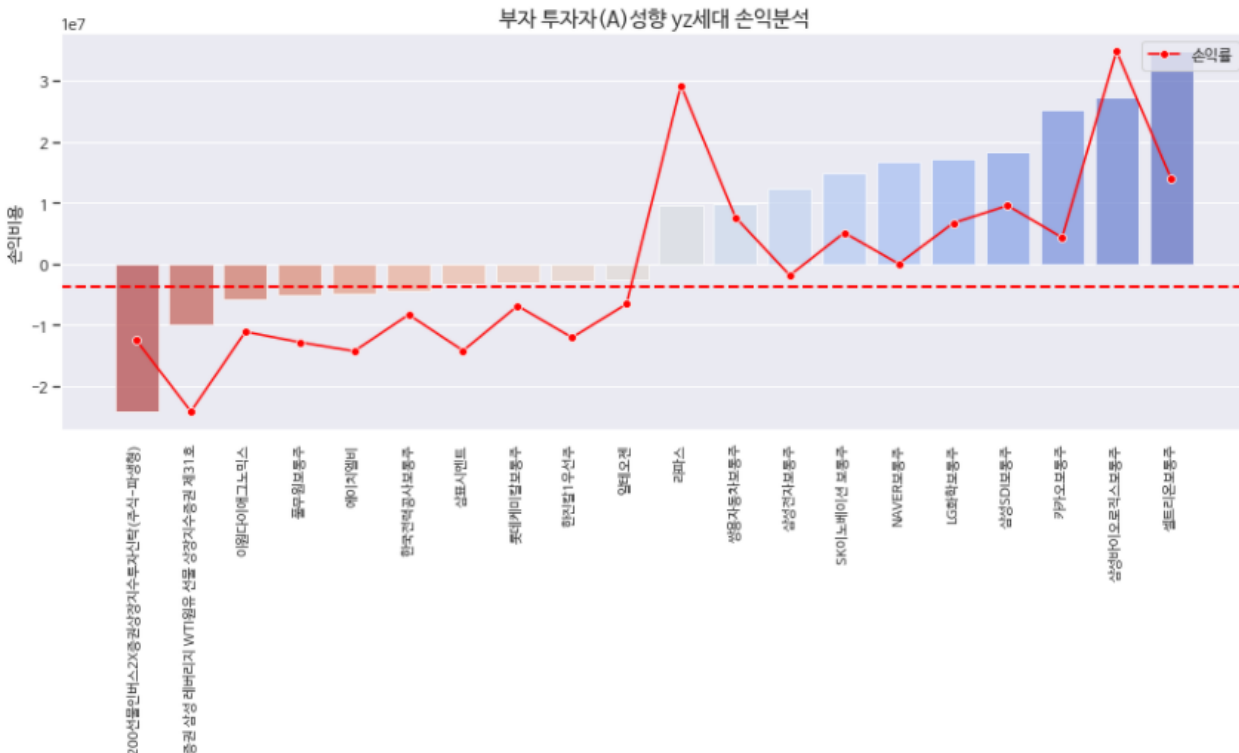
ㄷ. 유형 분석

유형 별 선호 종목 분석



〈두 가지 측면에서 분석〉

- ㄱ. 매매비용 결산한 종목별 전체 손익을 계산하여 각 유형에 큰 영향을 준 종목들 비교
ㄴ. 같은 유형 중, 세대별로 비교하여 중복되지 않는 종목을 차별되는 종목으로 판단하고 비교분석



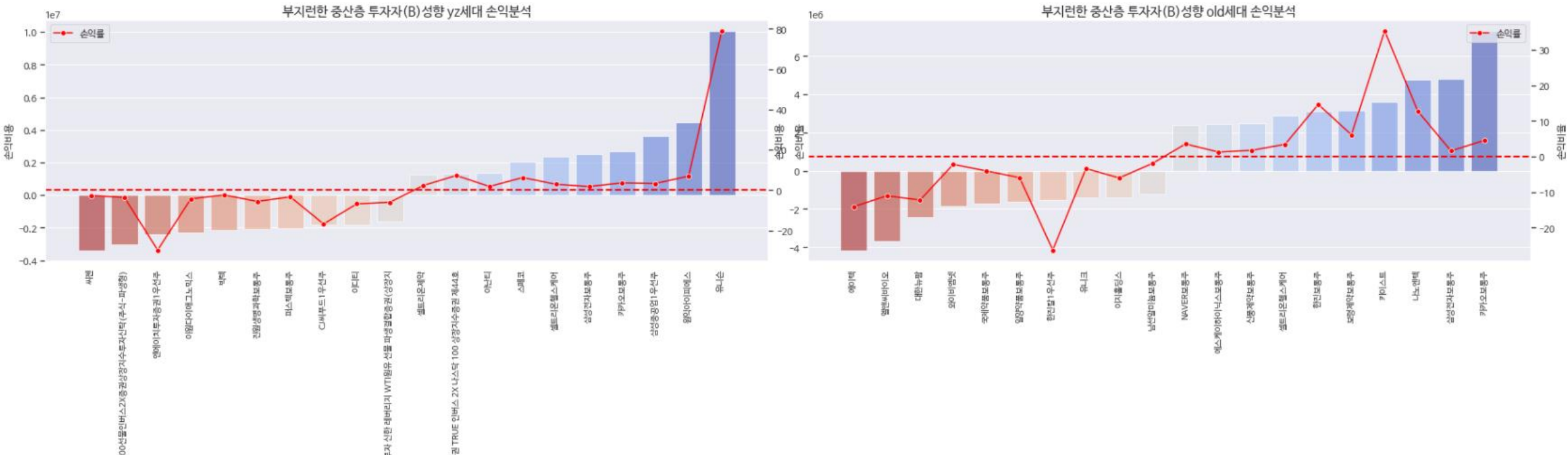
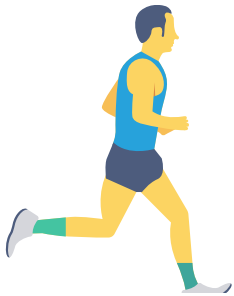
03 분석 결과

ㄷ. 유형 분석

유형 별 선호 종목 분석

〈두 가지 측면에서 분석〉

- ㄱ. 매매비용 결산한 종목별 전체 손익을 계산하여 각 유형에 큰 영향을 준 종목들 비교
- ㄴ. 같은 유형 중, 세대별로 비교하여 중복되지 않는 종목을 차별되는 종목으로 판단하고 비교분석



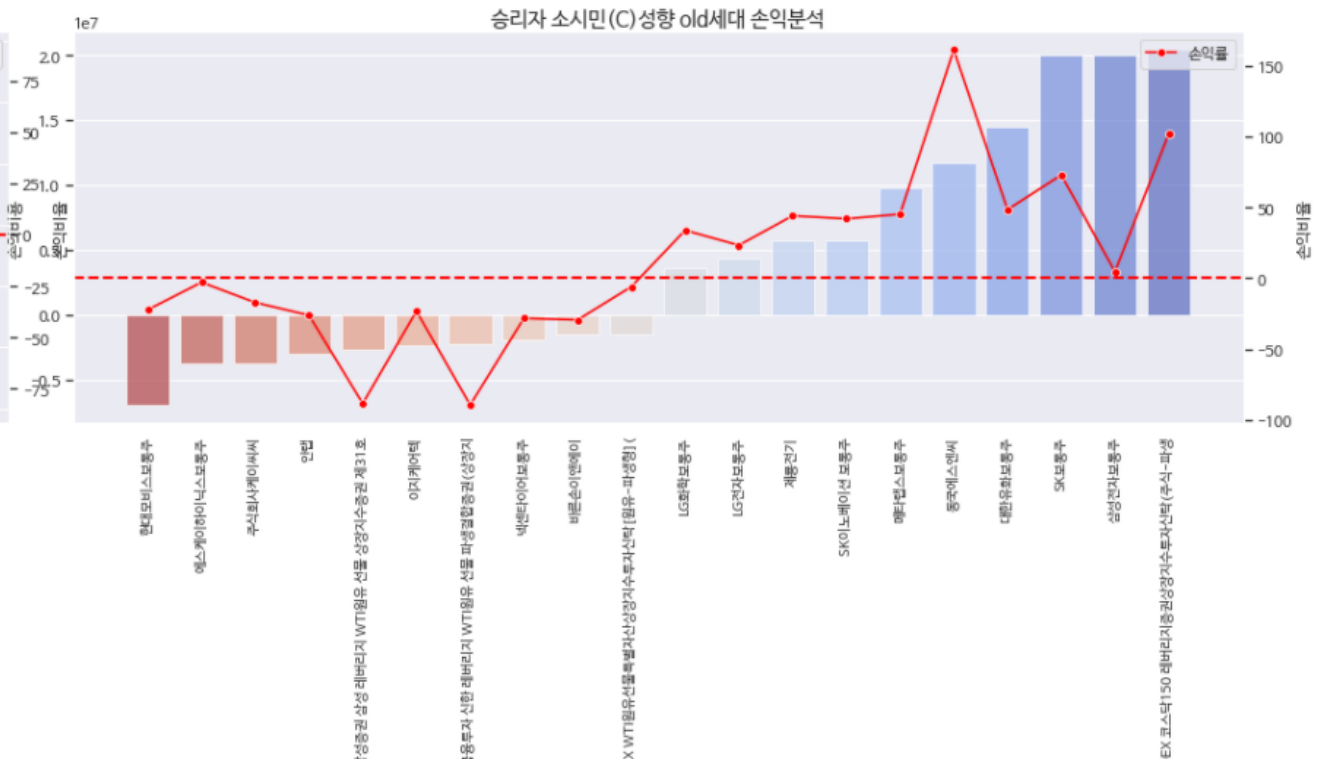
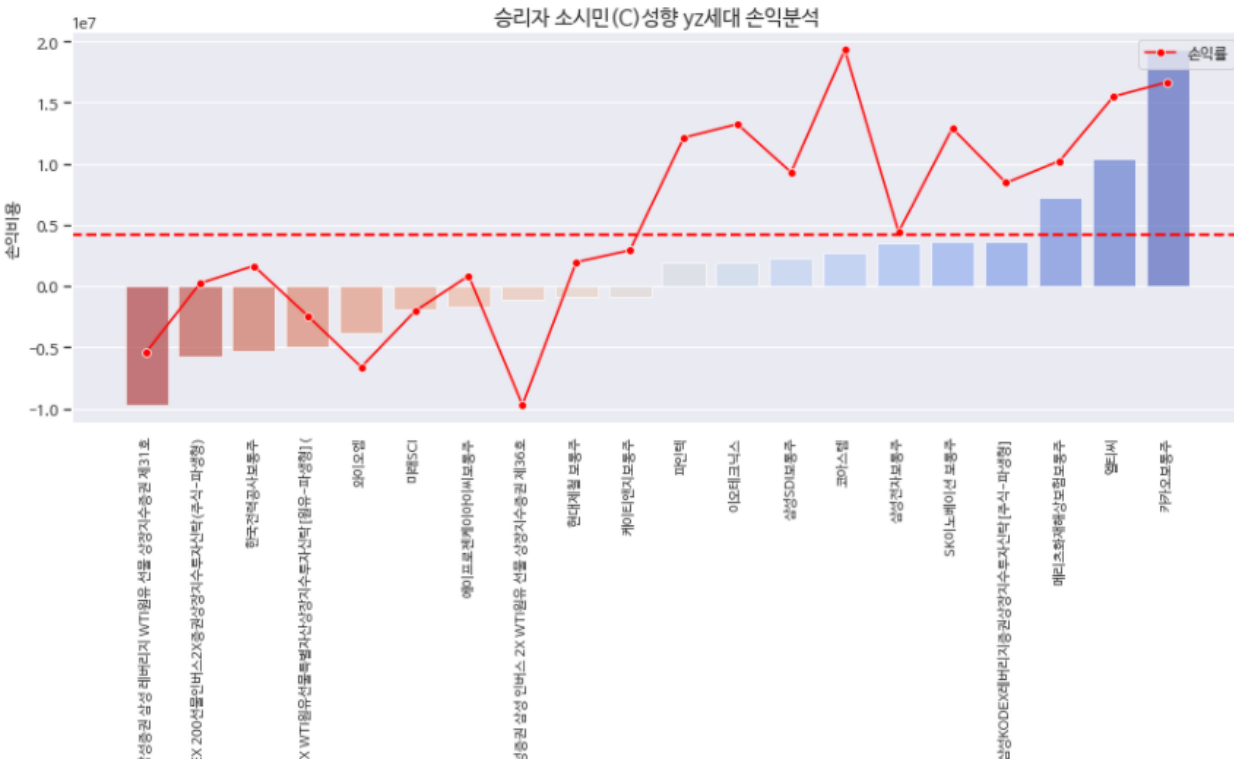
03 분석 결과

ㄷ. 유형 분석

유형 별 선호 종목 분석

〈두 가지 측면에서 분석〉

- ㄱ. 매매비용 결산한 종목별 전체 손익을 계산하여 각 유형에 큰 영향을 준 종목들 비교
- ㄴ. 같은 유형 중, 세대별로 비교하여 중복되지 않는 종목을 차별되는 종목으로 판단하고 비교분석



03 분석 결과

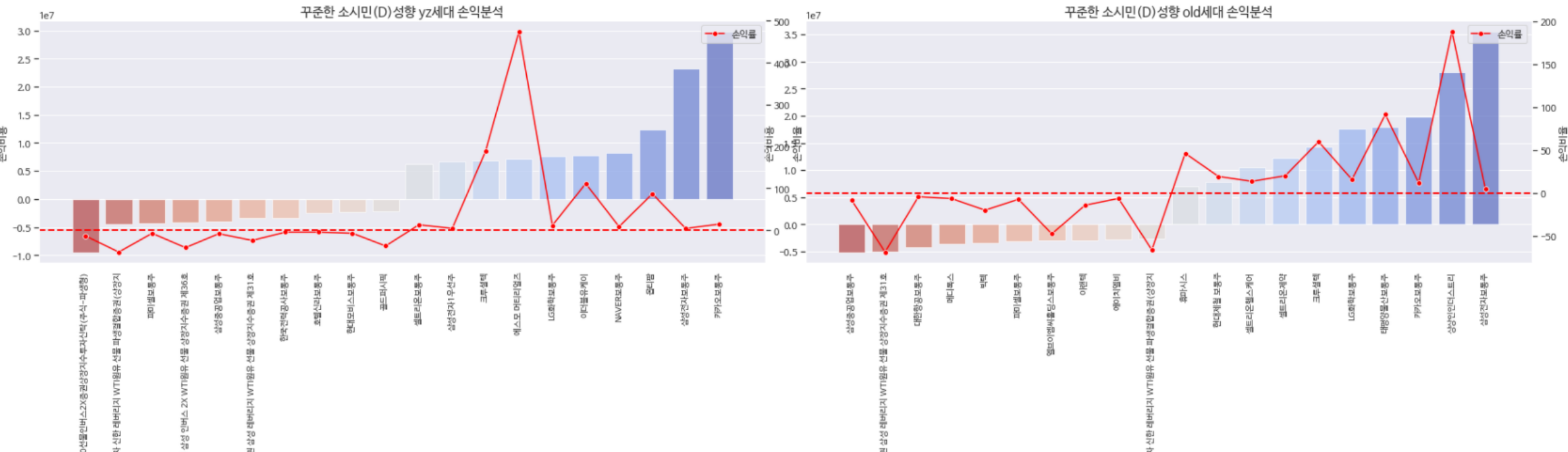
ㄷ. 유형 분석

유형 별 선호 종목 분석

〈두 가지 측면에서 분석〉



- ㄱ. 매매비용 결산한 종목별 전체 손익을 계산하여 각 유형에 큰 영향을 준 종목들 비교
ㄴ. 같은 유형 중, 세대별로 비교하여 중복되지 않는 종목을 차별되는 종목으로 판단하고 비교분석



03 분석 결과

ㄷ. 유형 분석

■ 소론

- A의 YZ와 OLD 모두 (-)의 대부분이 ETF 종목에서 발생하였다. YZ는 이에 대한 부분을 셀트리온, 삼바, 삼성SDI와 같은 우량주로 만회하였고 OLD도 카카오, 현대자동차, 기아자동차와 같은 우량주로 만회하였다.
- B의 경우 YZ가 특히 ETF 그 중에서도 인버스/레버리지 2배 상품에 손을 많이 댔고 손해를 봤다. 반면 OLD의 경우, 에이텍, 대한뉴팜과 같은 소형주에서 손해를 봤다. YZ가 유니슨이라는 종목에서 많이 이득을 봤는데 추가분석 결과, 소수의 개인들이 크게 이득을 본 것이라 큰 의미를 둘 순 없을 것 같다
- C의 경우 YZ에서 가장 ETF에 손해를 많이 본 집합인데 레버리지,인버스가 하위 2개이고 원유 선물도 큰 비용을 차지하고 있다. 앞서 선호 종목에서 삼전이 압도적으로 1순위였음에도 불구하고 YZ에서는 큰 비중을 차지않고 있음으로 보아 뒤늦게 매수하였거나 단타로 산 것으로 예상된다. OLD의 경우 인버스가 가장 큰 비중을 차지하고 있다는 점이 재밌는 포인트다
- 마지막으로 D의 경우 C와 마찬가지로 YZ에서 ETF로 큰 손해를 본 집단이다. 반면 카카오와 삼전과 같은 우량주로 큰 이득을 보았는데 다른 집단에 비해 빠른 타이밍에 들어간 것으로 보인다.

04. 결론

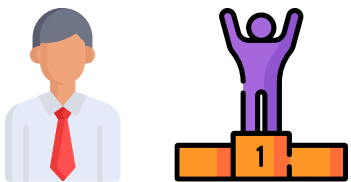
집단 별 맞춤형 금융상품 추천, 한계

04 결론

ㄱ. 집단 별 맞춤형 금융상품 추천

결론

C,D / yz세대



인버스/레버리지를 비롯한 ETF 상품이 그들의 투자에 있어서 큰 pain이었다. 이들을 위한 추가적인 교육과 추천 시스템이 필요할 것으로 보인다

또한 ETF뿐 아니라 주린이들이 하이리스크 하이리턴의 인버스/레버리지 상품에 많은 비중을 차지하고 있는데 기본적인 주식에 대한 공부부족한 것으로 해석된다.

슈카, 단군, 똥똥똥이와 같은 젊은 세대에게 친숙한 경제 스트리머와 협업하여 재밌게 주식을 배울 수 있는 기회를 제공하는 것도 필요하다고 생각된다.



(위)스트리머 슈카



(위)스트리머 침착맨

04 결론

└. 한계

■ 한계

- 세대, 베테랑, 주린이 등 군집화를 거듭할수록 적어지는 표본의 수
- 처음 접하는 종목들에 대한 이해 부족
- 투자자금을 유추하는 부분에 있어서 왜곡의 가능성
- 투자자의 투자 종목으로 성향을 파악하는데 어려움을 겪음
- 해외주식 거래내역을 반영하지 못함

감사합니다