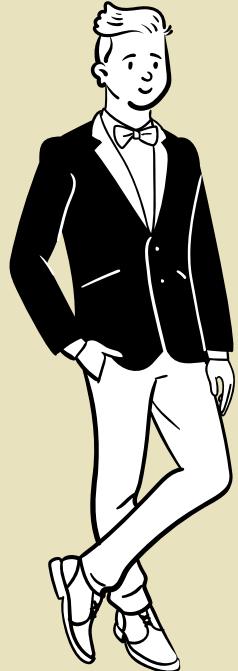


불로소득과 부채가 총 생활비에 미치는 영향



김재현 김정훈 하진
석

저희 팀원들을 소개합니 다 :)



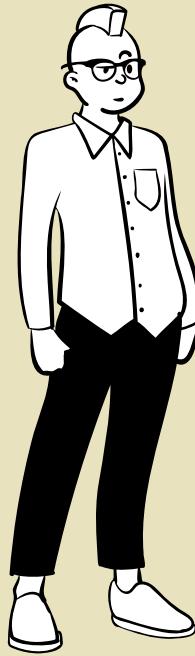
김재현

코드 검수 총괄, 주석 삽입 총괄,
에러 확인



김정훈

데이터분석 FRAME 구성,
PPT Deco



하진석

프로젝트 아이디어 제시,
PPT FRAME 구성

0. 연구 개요

연합뉴스 구독

PICK

한은 "한국, 가계부채 세계 3위...성장 막고 자산불평등 키워"

입력 2023.07.17. 오후 12:00 수정 2023.07.17. 오후 1:42 기사월문

신호경 기자



주: 1) 선진국 - 미국, 프랑스, 독일, 영국, 일본, EU, 덴마크 등 11개국, 신흥국 - 한국, 중국, 태국, 아르헨티나, 브라질, 헝가리 등 20개국

현대사회에서 금융 부채 문제 와 소득 양극화는 심각한 이슈 로 대두되고 있음

연구를 통해 가구별 총생활비와
부채, 불로소득과의 상관관계를
조사한 후 도출된 결론을 바탕으
로 다양한 분야에 도움이 될 수
있는 참고 자료를 제공하고자 함

저소득층 근로소득 줄고 고소득층 불로소득 증가...빈부 격차 '심화'

▲ 박영신 기자 | Ⓛ 입력 2022.09.19 09:37 | Ⓜ 댓글 0

코로나19가 확산됐던 지난 2020년 저소득층의 근로
소득은 오히려 감소했지만 걸었지만, 고소득층의 불
로소득은 빠른 속도로 증가해 소득양극화가 심화된
것으로 나타났다.

18일 진선미 더불어민주당 의원(국회 기획재정위원회)이 국세청으로부터 제출받은 '근로·배당·양도소득 천분위 현황' 자료에 따르면, 2020년 근로소득이 전
년 대비 4.5% 증가하는 동안, 배당소득은 27%, 양도
소득은 49.7% 증가했다.

INDEX

3rd Semi Project

INDEX

1. 데이터 준비

준비한 데이터를 불러오고 데이터 형식
확인 및 데이터 전처리(결측치 / 기초통계량)

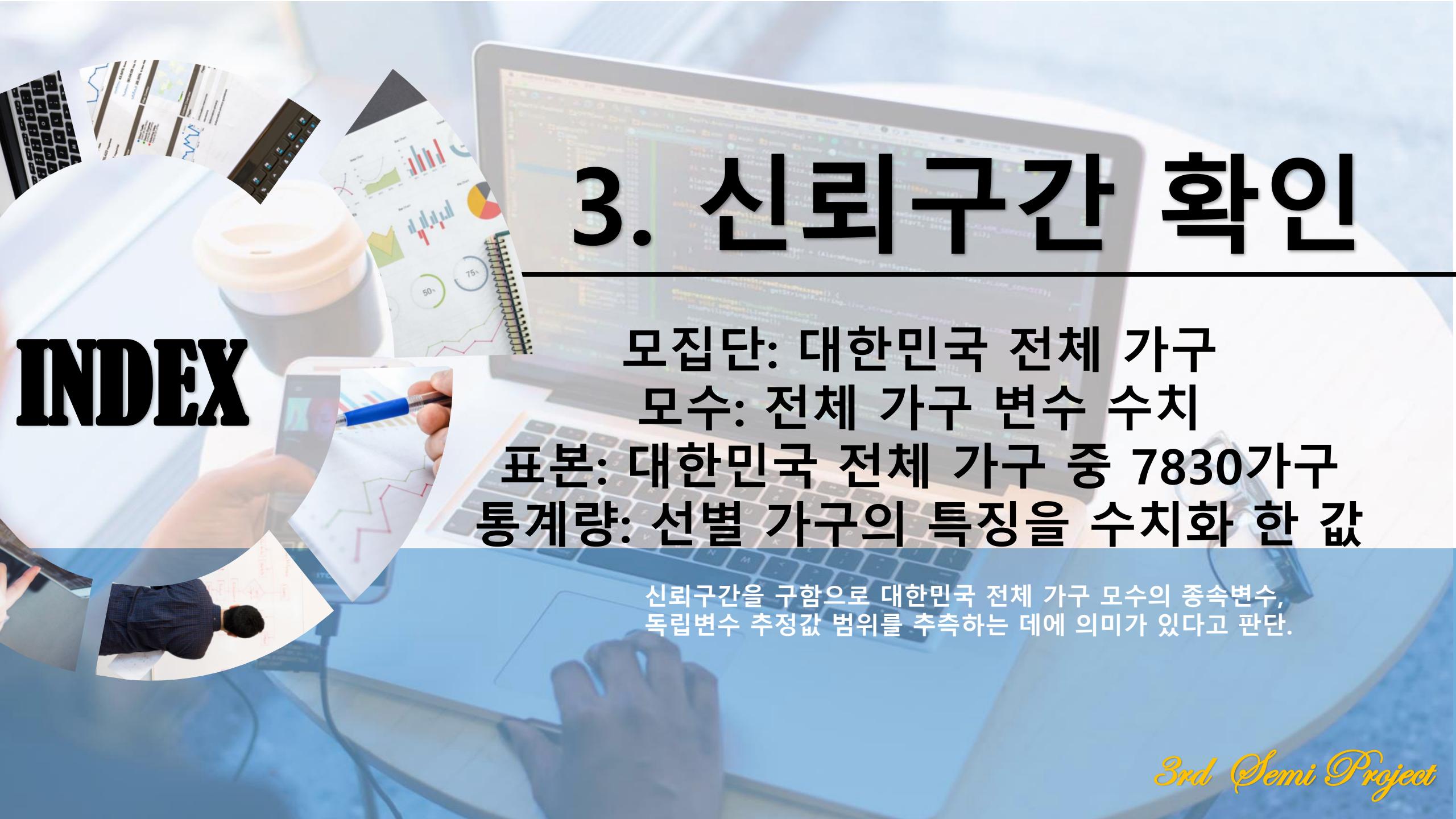
3rd Semi Project

2. 이상치 확인 및 정제

INDEX

이상치가 있는지 그래프 및 히스토그램으로
확인 후 처리 및 정제

3rd Semi Project



3. 신뢰구간 확인

모집단: 대한민국 전체 가구

모수: 전체 가구 변수 수치

표본: 대한민국 전체 가구 중 7830가구

통계량: 선별 가구의 특징을 수치화 한 값

신뢰구간을 구함으로 대한민국 전체 가구 모수의 종속변수, 독립변수 추정값 범위를 추측하는 데에 의미가 있다고 판단.

4. 정규성 및 상관관계 확인

정규분포 가정 확인 및
산점도 행렬, 히트맵으로 상관관계 확인

INDEX

3rd Semi Project

5. 주성분분석

INDEX

데이터 정규화 및
PCA분석을 통해 해석력이 약한 변수 확인

3rd Semi Project



INDEX

6. 다중선형회귀분석

get_df함수를 응용, 경우의 수를 전부 계산하여 비교·분석을 진행해본다.

INDEX



다중선형회귀 분석 결과를 시각화 및
최종 결론 도출

7. 결과

1. 데이터 준비

Data Processing

변수 선정

3rd Semi Project

1. 데이터 준비

Data Processing

데이터 수집

변수	의미	기타	비고
총생활비	한달에 지출되는 총 생활비	월 / 단위:만원	종속변수
이자_배당금_연간소득	은행이자나 주식 배당금으로 얻는 연간 수입	년간 / 단위:만원	불로소득
임대료_연간소득	전세보증금, 월세 등	년간 / 단위:만원	불로소득
기타_소득	상여금, 장려금, 보상금 등	년간 / 단위:만원	불로소득
부모가 보낸 생활비	부모로부터 받은 보조금(생활비개념)	년간 / 단위:만원	불로소득
자녀가 보낸 생활비	자녀로부터 받은 보조금(생활비개념)	년간 / 단위:만원	불로소득
민간보조금	민간 개인 또는 단체로부터 받은 보조금(생활비개념)	년간 / 단위:만원	불로소득
세금환급금	납세자가 내야할 세액보다 더 많이 내서 나오는 환급금	년간 / 단위:만원	불로소득
기초연금	저소득층 노인에게 지급하는 연금	년간 / 단위:만원	불로소득
공적연금	국민연금, 공무원연금, 군인연금, 사립학교교직원연금을 의미	년간 / 단위:만원	불로소득
재난지원금	코로나19에 따른 국민생계 안정을 위해 지원하는 금액	년간 / 단위:만원	불로소득
금융기관대출	금융권에서 받은 대출액	단위:만원	부채
주택관련_부채액	주택 구매, 임대 관련하여 발생한 부채액(대출액)	단위:만원	부채
주거관련_부채의_이자	주택관련 부채액의 이자	년간 / 단위:만원	부채
기타_이자	일반 대출이나 기타 사항에 의해 발생한 이자	년간 / 단위:만원	부채

3rd Semi Project

1. 데이터 준비

Data Processing
데이터 수집

재난지원금이나 공적연금도 불로소득에 포함 되는 걸까? ?



3rd Semi Project

1. 데이터 준비

Data Processing
데이터 수집

아하!



불로소득

말 그대로 일(노동)을 하지 않고 얻는 소득을 의미.

불로소득에 범주로는 은행예금, 채권이자, 주식배당, 용돈, 부동산 임대수익, 골동품과 미술품 같은 공급이 제한된 재화에서 판매하여 나오는 차익, 상속/증여세, 복지제도 혜택, 연금 등이 포함.

여기서 공적연금과 기초연금은 연금속성으로 간주하고 재난지원금은 복지제도의 혜택으로 간주하여 이번 조사 대상에 같이 포함하였음

>출처 : 네이버 지식백과, 나무위키, 위키백과

3rd Semi Project

데이터 전처리

1. 데이터 준비

Data Processing
데이터 전처리

총생활비	이자_배당금_연간소득	임대료_연간소득	기타_소득	부모가보낸생활비	자녀가보낸생활비	민간보조금	세금환급금	금융기관대출	주택관련_부채액	주거관련_부채의_이자	기타이자	기초연금	공적연금	재난지원금
0	275	0.0	0	7.0	0.0	500	360.0	0.0	10000	0.0	0	350.0	0	0
1	239	0.0	0	7.0	0.0	380	10.0	0.0	0	0.0	0	0.0	360	0
2	156	0.0	80	7.0	0.0	0	285.0	0.0	14000	20500.0	720	0.0	360	540
3	541	140.0	0	9.0	0.0	0	20.0	0.0	0	0.0	0	0.0	576	0
4	85	50.0	0	7.0	0.0	420	3.0	0.0	0	0.0	0	0.0	360	0

7865 rows × 15 columns

3rd Semi Project

1. 데이터 준비

데이터 형식 확인

총생활비	Int64
이자_배당금_연간소득	Float64
임대료_연간소득	Int64
기타_소득	Float64
부모가 보낸 생활비	Float64
자녀가 보낸 생활비	Int64
민간보조금	Float64
세금환급금	Float64
기초연금	Int64
공적연금	Int64
재난지원금	Int64
금융기관대출	Int64
주택관련_부채액	Float64
주거관련_부채의_이자	Int64
기타이자	Float64

결측치 확인

총생활비	0
이자_배당금_연간소득	0
임대료_연간소득	0
기타_소득	0
부모가 보낸 생활비	0
자녀가 보낸 생활비	0
민간보조금	0
세금환급금	0
기초연금	0
공적연금	0
재난지원금	0
금융기관대출	0
주택관련_부채액	1079
주거관련_부채의_이자	0
기타이자	3rd Semi Project

1. 데이터 준비

결측치 처리

총생활비	0
이자_배당금_연간소득	0
임대료_연간소득	0
기타_소득	0
부모가 보낸 생활비	0
자녀가 보낸 생활비	0
민간보조금	0
세금환급금	0
기초연금	0
공적연금	0
재난지원금	0
금융기관대출	0
주택관련_부채액	0
주거관련_부채의_이자	0
기타이자	0

Data Processing
데이터 전처리

주택관련_부채액에서 결측치 1079건이 확인되었고
부채액이 없는 것으로 간주하고 수치 값 '0'으로 대체

	총생활비	이자_배당금_연간소득	임대료_연간소득	기타_소득	부모가보낸생활비	자녀가보낸생활비	민간보조금	세금환급금	기초연금	공적연금	재난지원금	금융기관대출	주택관련_부채액	주거관련_부채의_이자
0	275	0.0	0	7.0	0.0	500	360.0	0.0	0	0	25	10000	0.0	0
1	239	0.0	0	7.0	0.0	380	10.0	0.0	360	0	20	0	0.0	0
2	156	0.0	80	7.0	0.0	0	285.0	0.0	360	540	20	14000	20500.0	720
3	541	140.0	0	9.0	0.0	0	20.0	0.0	576	0	125	0	0.0	0
4	85	50.0	0	7.0	0.0	420	3.0	0.0	360	0	35	0	0.0	0
...
7860	384	0.0	0	1.0	50.0	0	0.0	70.0	0	0	100	3000	3000.0	96
7861	715	0.0	660	1.0	40.0	0	0.0	90.0	0	0	100	40000	0.0	0

3rd Semi Project

1. 데이터 준비

0값 확인

```
for column in re_df1.columns:  
    zero_count = re_df1.loc[re_df1[column] == 0].shape[0]  
    number_count = re_df1.loc[re_df1[column] != 0].shape[0]  
    print('0 개수 :', {column : zero_count})  
    print('0 아닌 수 :', {column : number_count})  
    print('-'*30)
```

Data Processing
데이터 전처리

0이 아닌 값들의 수를 확인하여 분석에 용이한 데이터인지 확인

3rd Semi Project

1. 데이터 준비

0값 확인

```
for column in re_df1.columns:  
    zero_count = re_df1.loc[re_df1[column] == 0].shape[0]  
    number_count = re_df1.loc[re_df1[column] != 0].shape[0]  
    print('0 개수 :', {column : zero_count})  
    print('0 아닌 수 :', {column : number_count})  
    print('*'*30)
```

0이 아닌 값들의 수를 확인하여 분석에 용이한 데이터인지 확인



분석에 활용 가능한 데이터로 판단하여 제외하는 변수 없이 진행

Data Processing 데이터 전처리

	변수	0 값인 수	0 값이 아닌 수
0	총생활비	0	7865
1	이자_배당금_연간소득	5489	2376
2	임대료_연간소득	6707	1158
3	기타_소득	1439	6426
4	부모가보낸생활비	6366	1499
5	자녀가보낸생활비	3296	4569
6	민간보조금	4019	3846
7	세금환급금	6275	1590
8	기초연금	4155	3710
9	공적연금	4777	3088
10	재난지원금	239	7626
11	금융기관대출	5500	2365
12	주택관련_부채액	6421	1444
13	주거관련_부채의_이자	6414	1451
14	기타이자	6598	1267

3rd Semi Project

1. 데이터 준비

Data Processing
데이터 전처리

기초 통계량 값 확인

	총생활비	이자_배당금_연간소득	임대료_연간소득	기타_소득	부모가보낸생활비	자녀가보낸생활비	민간보조금	세금환급금	금융기관대출	주택관련부채액	주거관련부채의_이자	기타이자	기초연금	공적연금	재난지원금
count	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000
mean	329.822	55.538	129.336	8.456	45.949	230.697	58.166	14.506	2958.797	1773.448	53.392	47.885	193.396	288.639	94.908
std	1314.994	347.927	644.463	87.830	218.872	411.962	202.689	59.831	8277.347	5527.462	195.466	358.612	219.695	675.031	143.003
min	23.000	0.000	0.000	0.000	0.000	0.000	0.000	-700.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25%	121.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	35.000
50%	228.000	0.000	0.000	7.000	0.000	70.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	60.000
75%	435.000	20.000	0.000	8.000	0.000	300.000	30.000	0.000	1700.000	0.000	0.000	0.000	360.000	288.000	105.000
max	114446.000	22025.000	17640.000	5506.000	5500.000	6537.000	4555.000	1000.000	200000.000	120000.000	7200.000	22616.000	936.000	8796.000	7140.000



- 평균치에 비해 MAX값이 이상치로 판단될 수 있을 정도로 높은 값이 있음을 확인 가능
- 또한 모든 변수의 평균이 중앙값보다 크기 때문에 오른쪽으로 치우쳐 있을것으로 예상 가능

2. 이상치 확인 및 정제

Outlier Check
데이터 탐색 및 시각화

PROACTIVE OUTLIER
MANAGEMENT IN DATA ANALYSIS

Outlier Check

BEST PRACTICES FOR TRUSTWORTHY RESULT

3rd Semi Project

2. 이상치 확인 및 정제

Outlier Check
데이터 탐색 및 시각화

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



BEST PRACTICES FOR TRUSTWORTHY RESULT

3rd Semi Project

2. 이상치 확인 및 정제

Outlier Check
데이터 탐색 및 시각화

이상치란?

관측된 데이터의 범위에서 많이 벗어난 아주 작은 값이나 큰 값



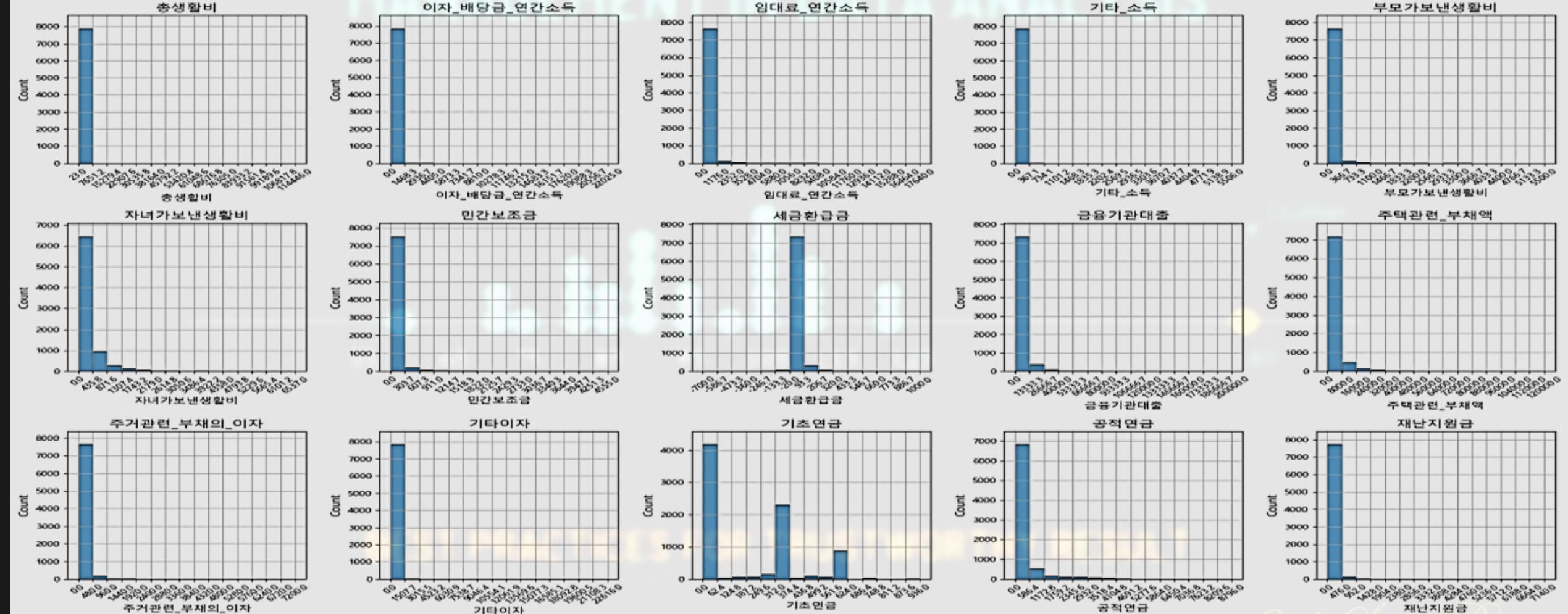
2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

히스토그램 확인

PROACTIVE OUTLIER
MANAGEMENT IN DATA ANALYSIS



3rd Semi Project

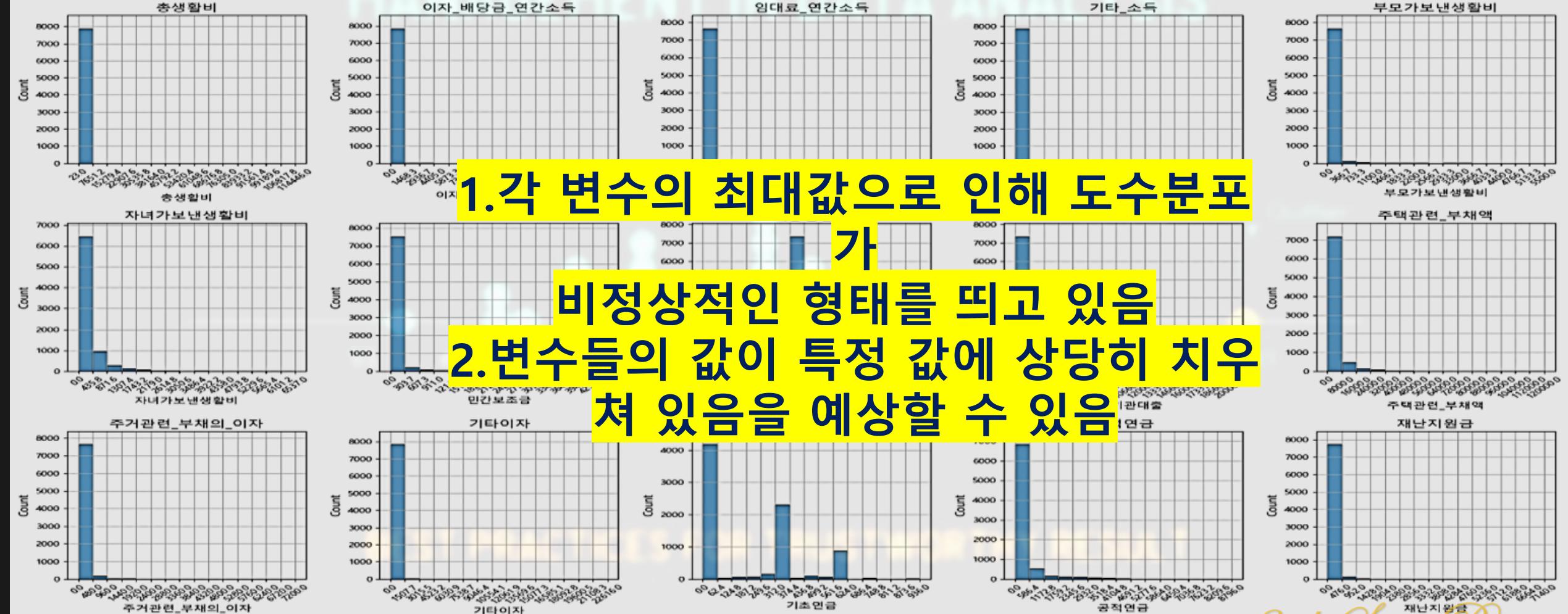
2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

히스토그램 확인

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



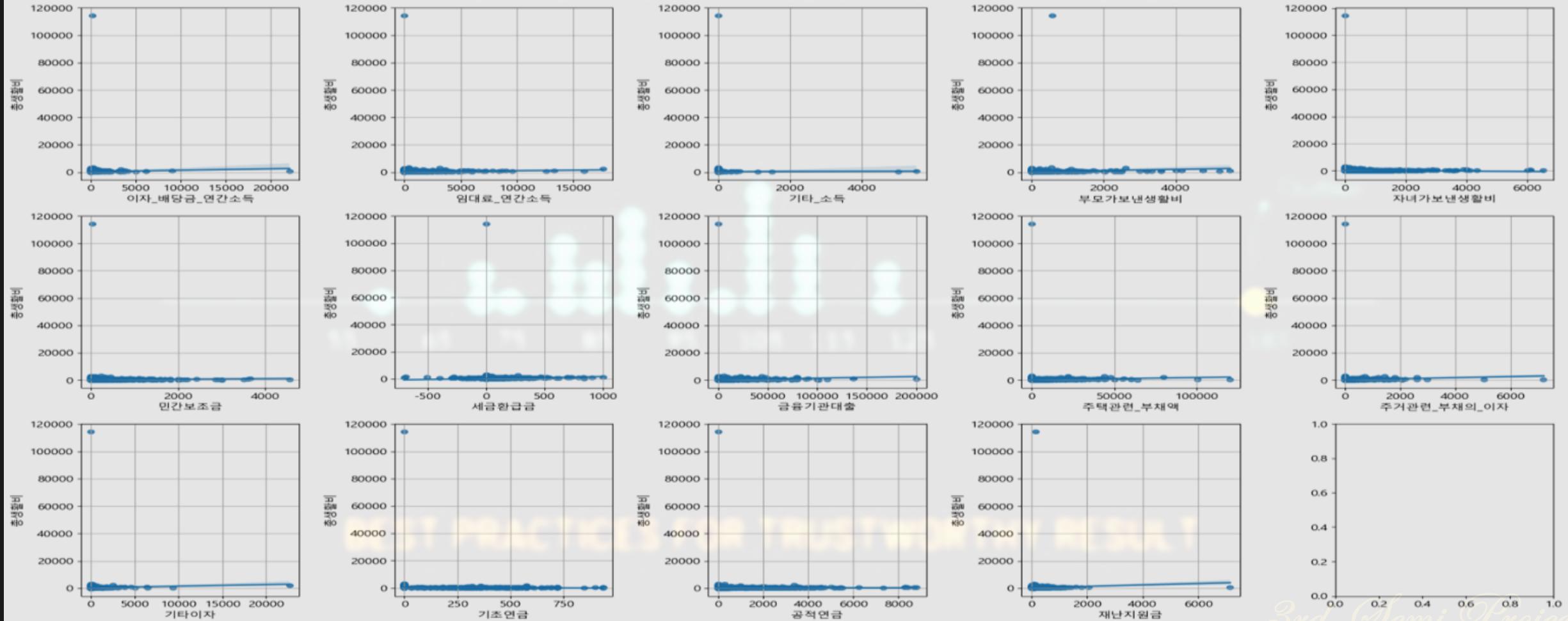
2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

Regplot으로 확인

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



3rd Semi Project

2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

Regplot으로 확인

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



3rd Semi Project

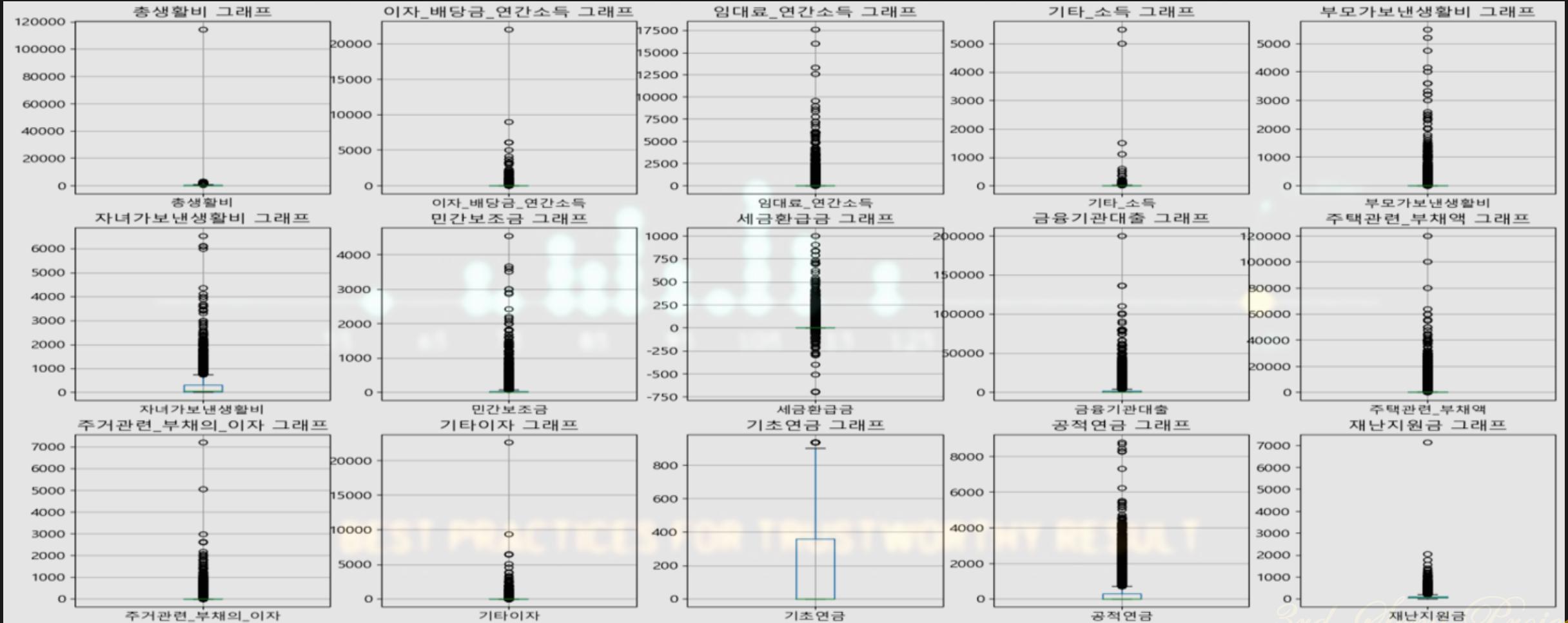
2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

Boxplot으로 확인

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



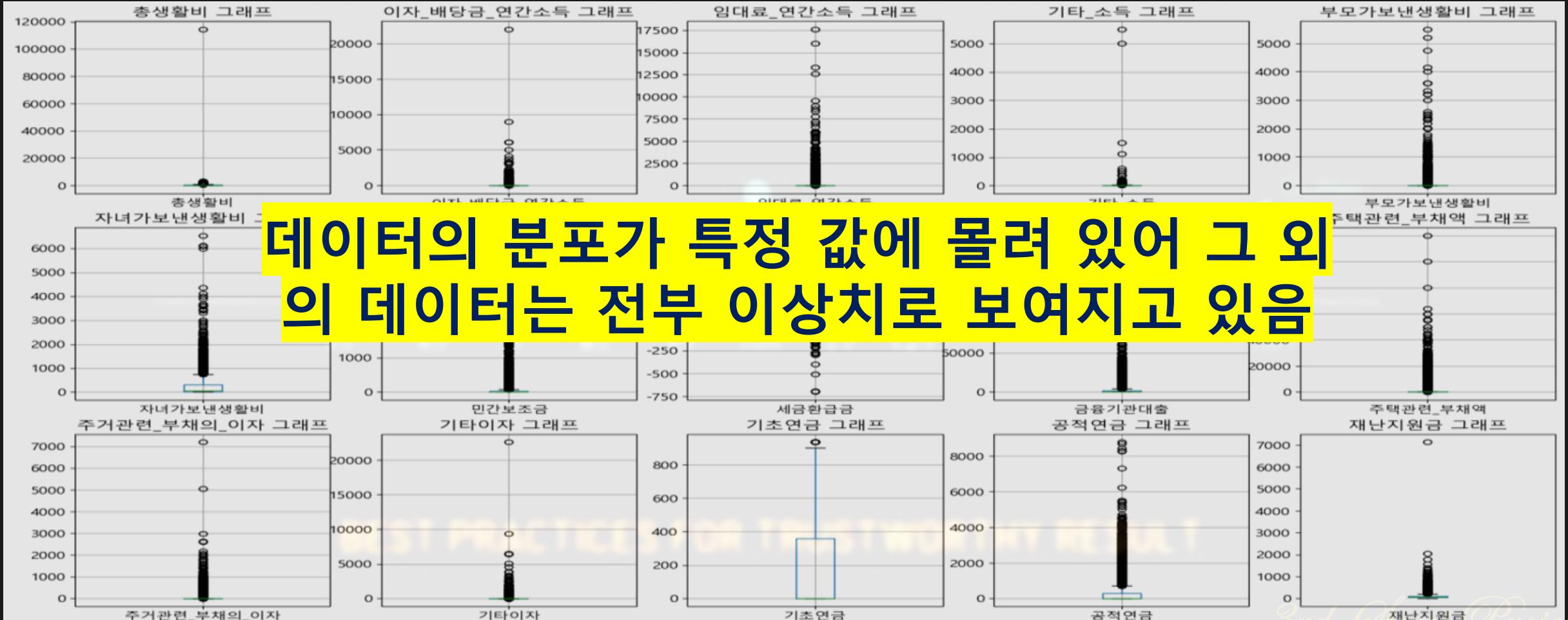
2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

Boxplot으로 확인

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



2. 이상치 확인 및 정제

종합 결과

Outlier Check
데이터 탐색 및 시각화

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



1. 종속변수(총생활비)
에서 1건만 비정상적
으로 높은 수치가 확
인되어 제거하고 진
행



3rd Semi Project

2. 이상치 확인 및 정제

종합 결과

Outlier Check
데이터 탐색 및 시각화

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



2. 몇몇 독립변수들
중 이상치가 발견되
어 프로젝트 팀원들
주관적 판단에 의해
제거하기로 결정

3rd Semi Project

2. 이상치 확인 및 정제

종합 결과

Outlier Check
데이터 탐색 및 시각화

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



3. 종속변수, 독립변수 포함해서 총 6건의 데이터(행) 제거하였음

BEST PRACTICES FOR TRUSTWORTHY RESULTS

3rd Semi Project

2. 이상치 확인 및 정제

이상치 제거

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS

```
df2=re_df1.query('(총생활비 < 100000) and (이자_배당금_연간소득 < 20000) and (기타_소득 < 4900) and \  
                 (금융기관대출 < 190000) and (재난지원금 < 7000)')  
df2
```

BEST PRACTICES FOR TRUSTWORTHY RESULT

Outlier Check
데이터 탐색 및 시각화

3rd Semi Project

2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

이상치 제거

```
df2=re_df1.query('(총생활비 < 100000) and (이자_배당금_연간소득 < 20000) and (기타_소득 < 4900) and \
                  (금융기관대출 < 190000) and (재난지원금 < 7000)')  
df2
```

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS

총생활비	이자_배당금_연간소득	임대료_연간소득	기타_소득	부모가보낸생활비	자녀가보낸생활비	민간보조금	세금환급금	금융기관대출	주택관련부채액	주거관련부채의_이자	기타이자	기초연금	공적연금	재난지원금
0	275	0.0	0	7.0	0.0	500	360.0	0.0	10000	0.0	350.0	0	0	25
1	239	0.0	0	7.0	0.0	380	10.0	0.0	0	0.0	0.0	360	0	20
2	156	0.0	80	7.0	0.0	0	285.0	0.0	14000	20500.0	720	0.0	360	540
3	541	140.0	0	9.0	0.0	0	20.0	0.0	0	0.0	0.0	576	0	125
4	85	50.0	0	7.0	0.0	420	3.0	0.0	0	0.0	0.0	360	0	35
...
7860	384	0.0	0	1.0	50.0	0	0.0	70.0	3000	3000.0	96	0.0	0	100
7861	715	0.0	660	1.0	40.0	0	0.0	90.0	40000	0.0	0	1560.0	0	100
7862	595	0.0	4000	1.0	40.0	0	0.0	0.0	25947	25947.0	67	0.0	0	120
7863	562	40.0	0	0.0	0.0	0	0.0	0.0	12000	12000.0	360	0.0	0	30
7864	514	0.0	0	1.0	0.0	100	0.0	0.0	1880	0.0	0	22.0	0	50

3rd Semi Project

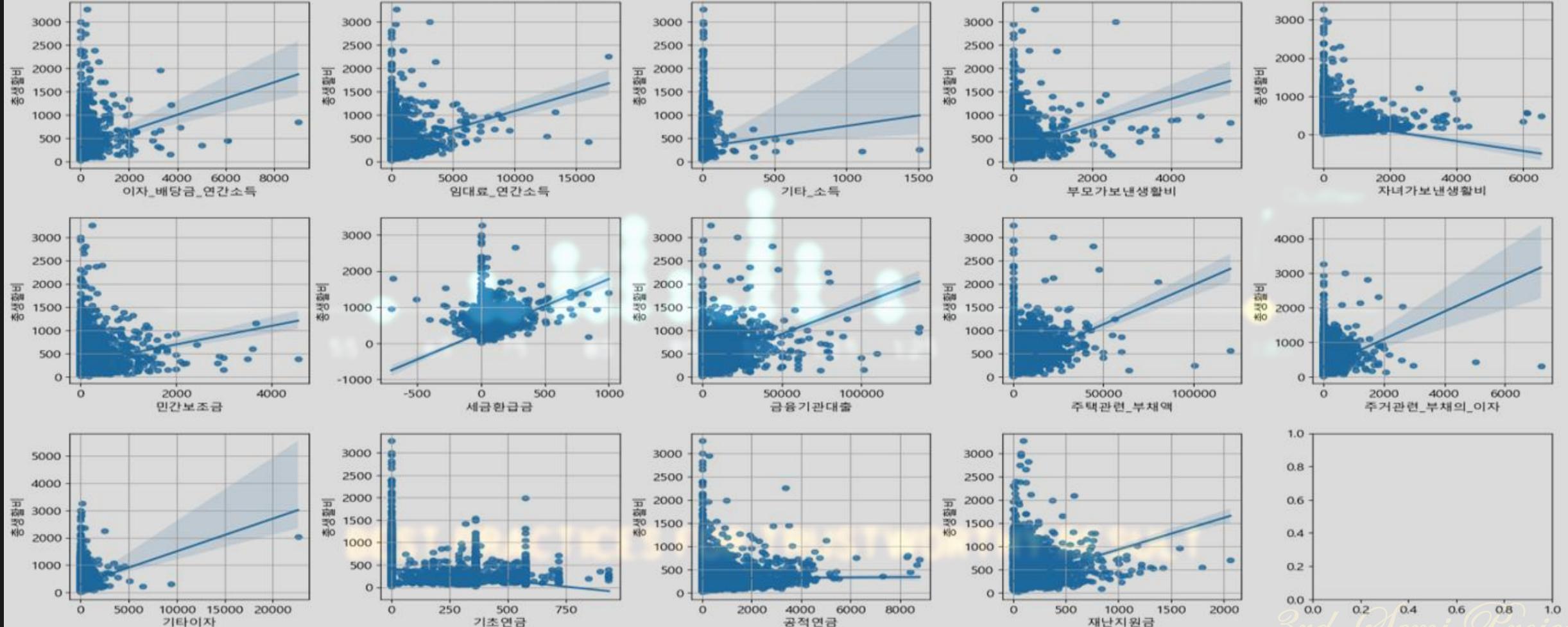
2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

이상치 제거 후 재확인(1)

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



3rd Semi Project

2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

이상치 제거 후 재확인(1)

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



3rd Semi Project

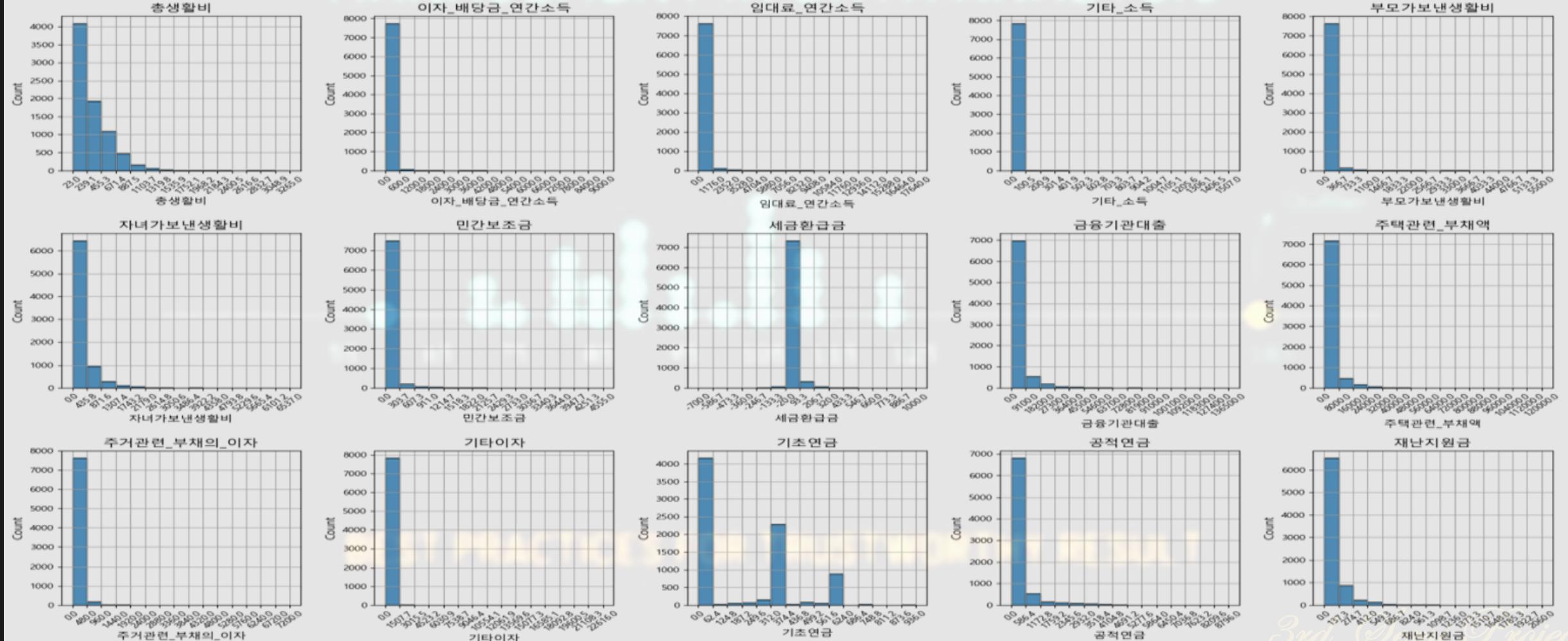
2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

이상치 제거 후 재확인(2)

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



Smart Project

2. 이상치 확인 및 정제

Outlier Check

데이터 탐색 및 시각화

이상치 제거 후 재확인(2)

PROACTIVE OUTLIER MANAGEMENT IN DATA ANALYSIS



3. 신뢰구간 확인

Confidence Interval
데이터 분석



01

- 모집단: 대한민국 전체 가구
- 모수: 대한민국 전체 가구 당 해당 엑셀파일의 종속, 독립 변수들 수치
- 표본: 모집단의 일부인 7859가구
- 통계량: 표본 7859가구의 월 평균 총생활비

02

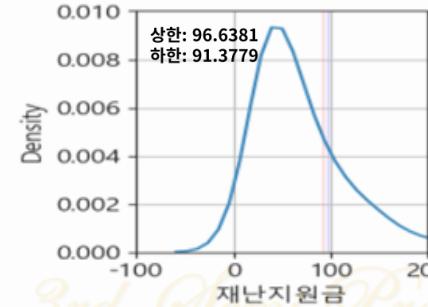
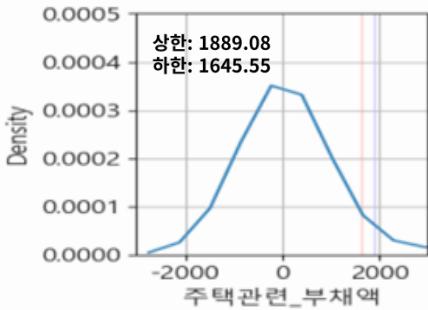
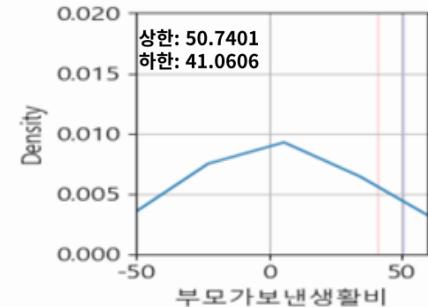
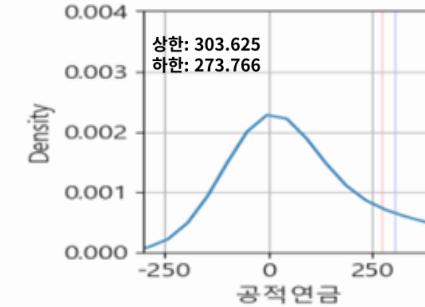
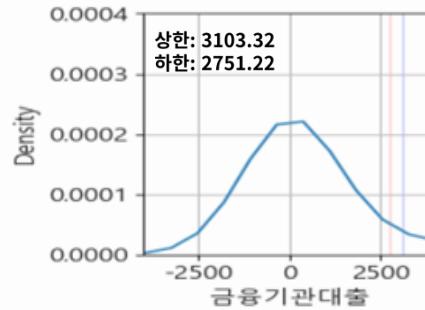
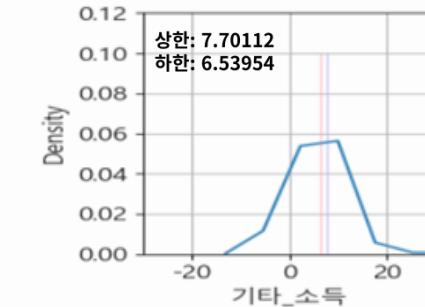
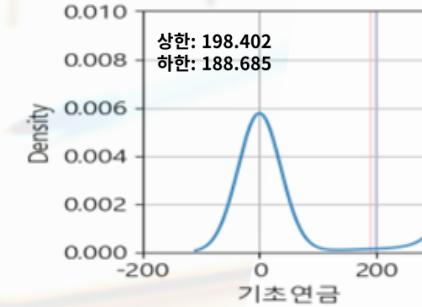
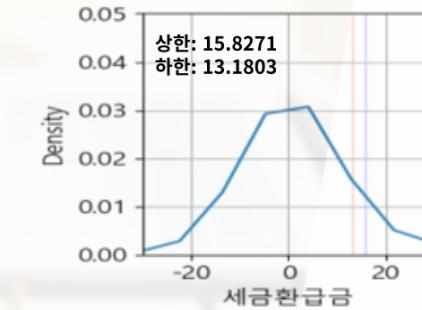
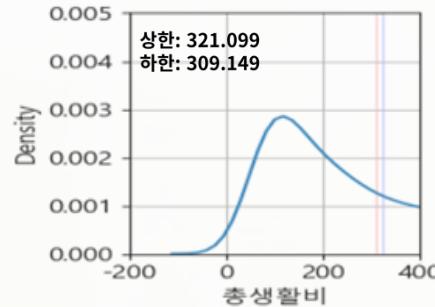
신뢰구간을 구함으로서 표본데이터를 기반으로 모집단의 평균값을 추측해 볼 수 있는 것에 의의를 둠

3rd Semi Project

3. 신뢰구간 확인

신뢰구간 시각화

Confidence Interval
데이터 분석



4. 정규성 및 상관관계 확인

Scatter Plot
데이터 탐색 및 시각화

산점도 행렬

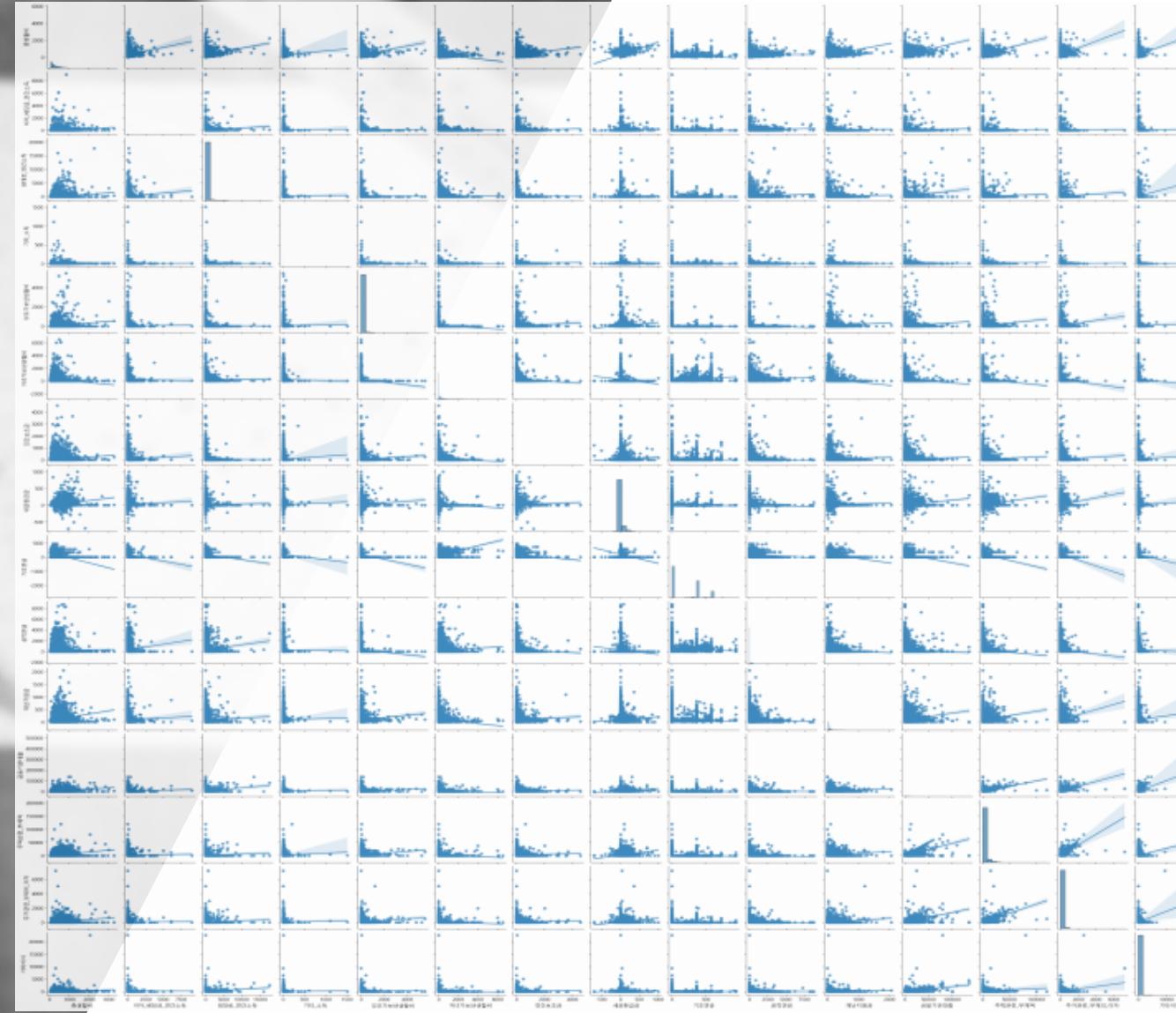
3rd Semi Project

4. 정규성 및 상관관계 확인

Scatter Plot

데이터 탐색 및 시각화

산점도 행렬 확인

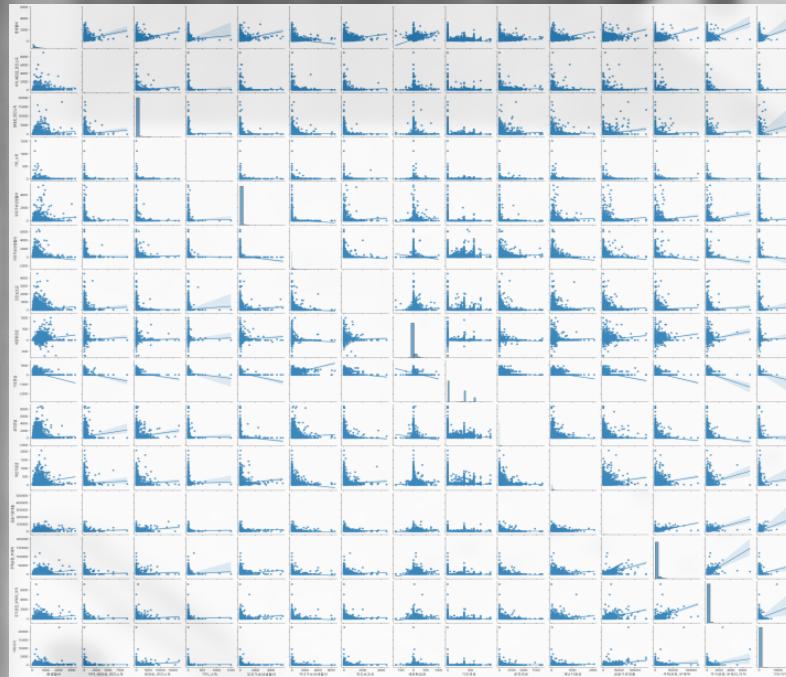


3rd Semi Project

4. 정규성 및 상관관계 확인

Scatter Plot
데이터 탐색 및 시각화

산점도 행렬 확인



01

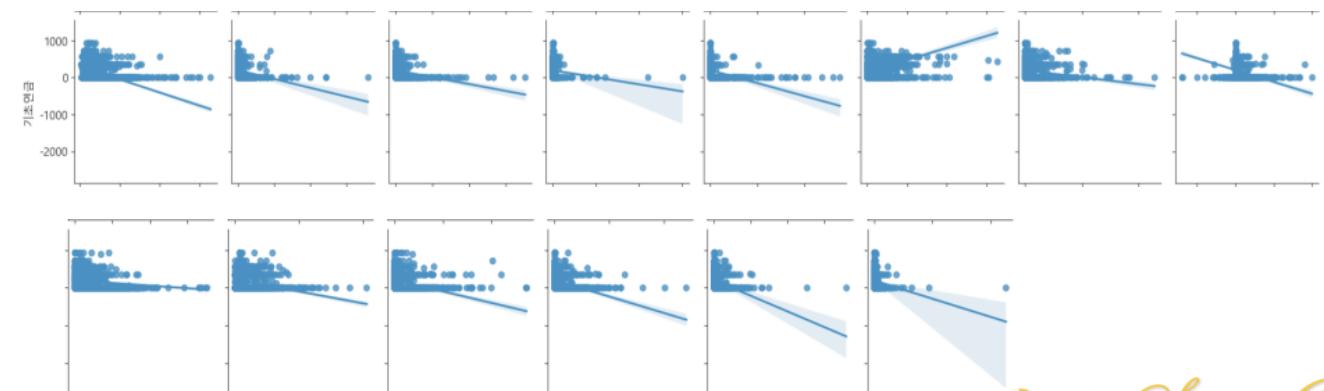
변수가 많아 알아보기 힘들지만 몇몇 요인들 간에 상관관계가 발견된다.

02

금융기관대출-주택관련부채액-주거관련_부채의_이자 간에 상관관계가 있음을 알 수 있다. 다만 이 부분은 대출액이 커질수록 이자도 당연히 커지기 때문에 합리적이라고 볼 수 있다.

03

독립변수 중 하나인 '기초연금'의 경우 대부분의 변수들과 반비례함을 확인 가능



3rd Semi Project

4. 정규성 및 상관관계 확인

Skewness and
Kurtosis

왜도, 첨도 확인

왜도 : 분포의 비대칭도

구분	내용
정규분포	첨도 = 0
위로 뾰족함	첨도 > 0
아래로 뾰족함	첨도 < 0

첨도 : 확률분포의 뾰족한 정도

구분	내용
정규분포	왜도 = 0
왼쪽으로 치우침	왜도 > 0
오른쪽으로 치우침	왜도 < 0

	변수명	정규분포 왜도	정규분포 첨도
0	이자_배당금_연간소득	16.080	398.172
1	임대료_연간소득	11.470	199.716
2	기타_소득	38.127	1832.707
3	부모가보낸생활비	12.077	204.786
4	자녀가보낸생활비	4.408	36.134
5	민간보조금	8.621	110.507
6	세금환급금	5.110	60.736
7	기초연금	0.517	-1.189
8	공적연금	4.467	28.361
9	재난지원금	4.887	38.836
10	금융기관대출	5.733	52.463
11	주택관련_부채액	5.941	64.295
12	주거관련_부채의_이자	12.302	313.154
13	기타이자	38.963	2256.060

규성 및 상관관계 확인

Skewness and
Kurtosis

	변수명	정규분포 왜도	정규분포 첨도
0	이자_배당금_연간소득	16.080	398.172
1	임대료_연간소득	11.470	199.716
2	기타_소득	38.127	1832.707
3	부모가보낸생활비	12.077	204.786
4	자녀가보낸생활비	4.408	36.134
5	민간보조금	8.621	110.507
6	세금환급금	5.110	60.736
7	기초연금	0.517	-1.189
8	공적연금	4.467	28.361
9	재난지원금	4.887	38.836
10	금융기관대출	5.733	52.463
11	주택관련_부채액	5.941	64.295
12	주거관련_부채의_이자	12.302	313.154
13	기타이자	38.963	2256.060

기초연금을
제외한 나머지
변수들은 정규
분포를 따르지
않는 것으로 확인.

4. 정규성 및 상관관계 확인

스피어만 상관분석 & 결과검정

Spearman Analysis
데이터 탐색 및 시각화

데이터가 정규분포를 따르지 않아
스피어만(spearman) 상관분석 진행

4. 정규성 및 상관관계 확인

스피어만 상관분석 & 결과검정

fields	statistic	pvalue	result
총생활비 vs 이자_배당금_연간소득	0.214765	1.136305e-82	True
이자_배당금_연간소득 vs 임대료_연간소득	0.175856	1.292815e-55	True
임대료_연간소득 vs 기타_소득	0.099732	7.854510e-19	True
기타_소득 vs 부모가보낸생활비	0.120537	7.929718e-27	True
부모가보낸생활비 vs 자녀가보낸생활비	-0.467324	0.000000e+00	True
자녀가보낸생활비 vs 민간보조금	-0.153996	6.558783e-43	True
민간보조금 vs 세금환급금	0.110892	6.234430e-23	True
세금환급금 vs 금융기관대출	0.266333	1.000648e-127	True
금융기관대출 vs 주택관련_부채액	0.730055	0.000000e+00	True
주택관련_부채액 vs 주거관련_부채의_이자	0.932065	0.000000e+00	True
주거관련_부채의_이자 vs 기타이자	0.153365	1.437161e-42	True
기타이자 vs 기초연금	-0.203579	2.677688e-74	True
기초연금 vs 공적연금	0.193484	3.711177e-67	True
공적연금 vs 재난지원금	-0.045009	6.565812e-05	True
재난지원금 vs 총생활비	0.529540	0.000000e+00	True

데이터가 정규분포를 따르지 않아
스피어만(spearman) 상관분석 진행

상관분석의 검정 결과
모든 변수들 간의
p_value 값이 0.05이
하로 통계적으로 유
의한 것으로 나타남

4. 정규성 및 상관관계 확인

상관분석 시각화 (Heatmap)

	총생활비	이자_배당금_연간소득	부모가보낸생활비	자녀가보낸생활비	민간보조금	세금환급금	금융기관대출	주택관련_부채액	주거관련_부채의_이자	기타이자	기초연금	공적연금	재난지원금		
총생활비	1	0.21	0.13	0.36	0.42	-0.43	0.18	0.4	0.46	0.39	0.4	0.29	-0.53	-0.072	0.53
이자_배당금_연간소득	0.21	1	0.18	0.13	-0.019	0.083	0.1	0.056	-0.087	-0.088	-0.087	-0.06	-0.099	0.22	0.04
임대료_연간소득	0.13	0.18	1	0.1	-0.056	0.084	-0.0025	-0.0063	-0.0041	-0.046	-0.039	0.042	-0.054	0.14	0.0074
기타_소득	0.36	0.13	0.1	1	0.12	-0.12	0.13	0.16	0.19	0.15	0.16	0.14	-0.12	-0.015	0.3
부모가보낸생활비	0.42	-0.019	-0.056	0.12	1	-0.47	0.11	0.34	0.34	0.34	0.34	0.16	-0.41	-0.33	0.23
자녀가보낸생활비	-0.43	0.083	0.084	-0.12	-0.47	1	-0.15	-0.3	-0.34	-0.31	-0.32	-0.21	0.5	0.36	-0.31
민간보조금	0.18	0.1	-0.0025	0.13	0.11	-0.15	1	0.11	0.11	0.078	0.073	0.098	0.088	-0.093	0.075
세금환급금	0.4	0.056	-0.0063	0.16	0.34	-0.3	0.11	1	0.27	0.27	0.28	0.12	0.28	-0.18	0.19
금융기관대출	0.46	-0.087	-0.0041	0.19	0.34	-0.34	0.11	0.27	1	0.73	0.74	0.58	0.32	-0.19	0.3
주택관련_부채액	0.39	-0.088	-0.046	0.15	0.34	-0.31	0.07	0.27	0.73	1	0.93	0.16	0.28	-0.19	0.26
주거관련_부채의_이자	0.4	-0.087	-0.039	0.16	0.34	-0.32	0.07	0.28	0.74	0.93	1	0.15	0.29	-0.18	0.26
기타이자	0.29	-0.06	0.042	0.14	0.16	-0.21	0.09	0.12	0.58	0.16	0.15	1	-0.2	-0.11	0.18
기초연금	-0.53	-0.099	-0.054	-0.12	-0.41	0.5	-0.088	-0.28	-0.32	-0.28	-0.29	-0.2	1	0.19	-0.18
공적연금	-0.072	0.022	0.14	-0.015	-0.33	0.36	-0.093	-0.18	-0.19	-0.19	-0.18	-0.11	0.19	1	-0.045
재난지원금	0.53	0.04	0.0074	0.3	0.23	-0.31	0.075	0.19	0.3	0.26	0.26	0.18	-0.18	-0.045	1
	총생활비	이자_배당금_연간소득	임대료_연간소득	기타_소득	부모가보낸생활비	자녀가보낸생활비	민간보조금	세금환급금	금융기관대출	주택관련_부채액	주거관련_부채의_이자	기타이자	기초연금	공적연금	재난지원금



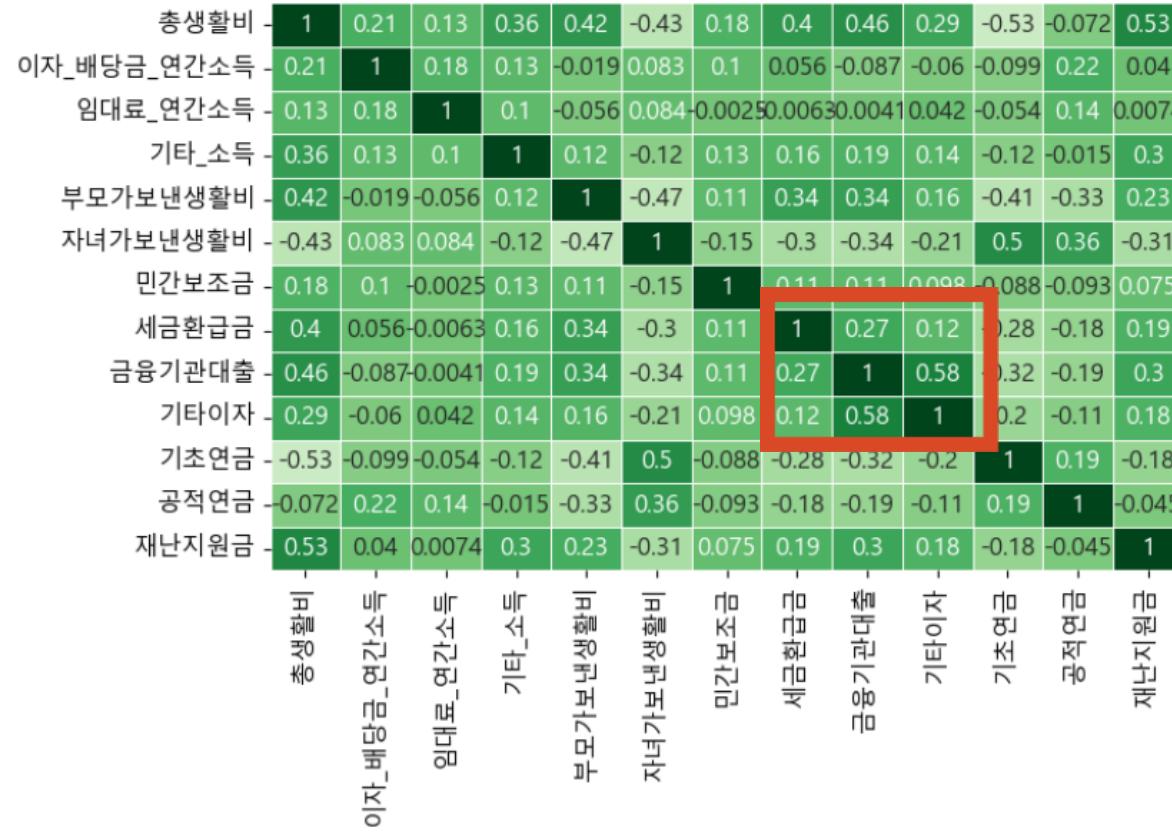
Heatmap Graph
데이터 탐색 및 시각화

'금융기관대출', '주택관련부채', '주거관련_부채의_이자'
· 변수들의 강한 양의 상관
관계 확인

서로 영향을 미치는 변수
들을 제외 후 다시 스피어
만 상관분석 진행

4. 정규성 및 상관관계 확인

상관분석 시각화 (Heatmap)



'주택관련_부채액', '주거관련_부채액'을 제외한 후 스피어만 상관분석 진행

'금융기관대출'도 높은 상관관계를 보였지만 '주택관련_부채액', '주거관련_부채액'을 제외하였더니 상관계수가 감소하여 제외하지 않음

Heatmap Graph
데이터 탐색 및 시각화

3rd Semi Project

5. 주성분(PCA)분석

Data Standardization
통계적분석 및 모델링

주성분 분석을 하기 전 정규화 진행

5. 주성분(PCA)분석

Data Standardization
통계적분석 및 모델링

주성분 분석을 하기 전 정규화
진행

	총생활비	이자_배당금_연간소득	임대료_연간소득	기타_소득	부모가보낸생활비	자녀가보낸생활비	민간보조금	세금환급금	금융기관대출	주택관련부채액	주거관련_부채의_이자	기타이자	기초연금	공적연금	재난지원금
count	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000
mean	329.822	55.538	129.336	8.456	45.949	230.697	58.166	14.506	2958.797	1773.448	53.392	47.885	193.396	288.639	94.908
std	1314.994	347.927	644.463	87.830	218.872	411.962	202.689	59.831	8277.347	5527.462	195.466	358.612	219.695	675.031	143.003
min	23.000	0.000	0.000	0.000	0.000	0.000	0.000	-700.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25%	121.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	35.000
50%	228.000	0.000	0.000	7.000	0.000	70.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	60.000
75%	435.000	20.000	0.000	8.000	0.000	300.000	30.000	0.000	1700.000	0.000	0.000	0.000	360.000	288.000	105.000
max	114446.000	22025.000	17640.000	5506.000	5500.000	6537.000	4555.000	1000.000	200000.000	120000.000	7200.000	22616.000	936.000	8796.000	7140.000

주성분 분석의 해석이 어려워질 수 있기 때문에 각 변수의 단위를 동일한 값으로
변경하여 단순 비교할 수 있도록 유도.

또한 서로 다른 단위나 범위를 가진 '데이터를 동일한 기준으로 맞춤'으로서,
'모델의 성능 향상'을 기대할 수도 있다고 판단.

5. 주성분(PCA)분석

Data Standardization
통계적분석 및 모델링

주성분 분석을 하기 전 정규화
진행

```
model = pca()  
fit = model.fit_transform(x_train_std_df)  
topfeat = fit['topfeat']  
topfeat
```

5. 주성분(PCA)분석

Data Standardization
통계적분석 및 모델링

주성분 분석을 하기 전 정규화
진행

PC	feature	loading	type
0	PC1	금융기관대출	0.491171 best
1	PC2	자녀가보낸생활비	0.497598 best
2	PC3	공적연금	0.567770 best
3	PC4	기타_소득	0.758736 best
4	PC5	세금환급금	-0.454537 best
5	PC6	민간보조금	-0.635869 best
6	PC7	재난지원금	-0.509041 best
7	PC8	공적연금	-0.540182 best
8	PC9	부모가보낸생활비	0.771742 best
9	PC10	이자_배당금_연간소득	0.454611 best
10	PC11	기타이자	0.599527 best
11	PC12	기초연금	-0.762436 best
12	PC3	임대료_연간소득	0.530654 weak
13	PC1	주택관련_부채액	0.483626 weak
14	PC1	주거관련_부채의_이자	0.464062 weak

```
model = pca()  
fit = model.fit_transform(x_train_std_df)  
topfeat = fit['topfeat']  
topfeat
```

임대료 연간소득, 주택관련 부채액, 주거관련 부채의 이자 가 설명력이 낮은 변수로 나타남.

1. 이전 히트맵 그래프에서 상관관계가 높게 나타난 주택관련 부채액, 주거관련 부채의 이자 두 가지 변수를 제외하고 회귀분석 진행

2. 설명력이 낮은 세 가지 변수 모두 제외한 후 회귀분석 진행

3rd Semi Project

6. 다중선형회귀분

MLR
데이터 모델링

석

회귀분석 실시 및 결과보고

6. 다중선형회귀분

MLR

데이터 모델링

회귀분석 실시 첫 결과보고(1)

종속변수	독립변수	비표준화 계수		t	유의확률	공선성 통계량
		B	표준오차			
총생활비	민간보조금	0.0861	0.009	0	9.703*	0
	이자_배당금_연간소득	0.0901	0.009	0	10.112*	0
	기초연금	-0.2715	0.01	0	-27.753*	0
	기타이자	0.0145	0.01	0	1.474*	0.14
	공적연금	0.0141	0.009	0	1.569*	0.117
	기타_소득	0.0154	0.009	0	1.744*	0.081
	금융기관대출	0.2025	0.01	0	19.370*	0
	부모가보낸생활비	0.0967	0.009	0	10.699*	0
	재난지원금	0.1833	0.009	0	20.164*	0
	세금환급금	0.2112	0.009	0	23.159*	0
	자녀가보낸생활비	-0.0258	0.009	0	-2.758*	0.006
	임대료_연간소득	0.0852	0.009	0	9.276*	0

R(0.392), R^2(0.391), F(422.2), 유의확률(0.00), Durbin-Watson(1.876) (*은 0.05 미만을 의미)

이전 히트맵 그래프에서 상관관계가 발견된
주택관련 부채액, 주거관련 부채의 이자
두 가지 변수 제외 후 회귀분석 실시한 결과보고

분석결과 회귀모형은 통계적으로
유의미한 것으로 나타났으며
(F= 422.2, p<0.05)

설명력은 39.2%(수정된 R제곱은
39.1%)로 나타남

Durbin-Watson값은 1.876으로 2에 근
사한 값을 보여 잔차의 독립성 가정

모든 변수들의 VIF값은 100이하로 다중공
선성의 문제 없음

3rd Semi Project

6. 다중선형회귀분

MLR
데이터 모델링

회귀분석 실시 선행 결과보고(2)

종속변수	독립변수	비표준화 계수		t	유의확률	공선성 통계량	
		B	표준오차				
총생활비	민간보조금	0.0844	0.009	0	9.459*	0	1.02785
	이자_배당금_연간소득	0.0956	0.009	0	10.694*	0	1.03494
	기초연금	-0.277	0.01	0	-28.217*	0	1.35581
	기타이자	0.0192	0.01	0	1.948*	0.051	1.23846
	공적연금	0.0213	0.009	0	2.366*	0.018	1.04103
	기타_소득	0.0157	0.009	0	1.766*	0.077	1.00368
	금융기관대출	0.2212	0.01	0	21.439*	0	1.43914
	부모가보낸생활비	0.0943	0.009	0	10.379*	0	1.06943
	재난지원금	0.183	0.009	0	20.024*	0	1.12105
	세금환급금	0.2092	0.009	0	22.827*	0	1.14429
	자녀가보낸생활비	-0.0228	0.009	0	-2.433*	0.015	1.12592

R(0.386), R^2(0.385), F(448.0), 유의확률(0.00), Durbin-Watson(1.869) (*은 0.05 미만을 의미)

PCA 검사에서 설명력이 약한 세가지 변수 제외 후 회귀분석 실시한 결과보고

분석결과 회귀모형은 통계적으로 유의미한 것으로 나타났으며
(F= 448.0, p<0.05)

설명력은 38.6%
(수정된 R제곱은 38.5%)로 나타남

Durbin-Watson값은 1.869으로 2에 근사한 값을 보여 잔차의 독립성 가정

모든 변수들의 VIF값은 100이하로
다중공선성의 문제 발생하지 않음

6. 다중선형회귀분

MLR
데이터 모델링

모든 경우의 수 회~~회~~분석 진행

변수 설정 및 결과	
Case1	weak로 나온 3개의 독립변수를 처리하지 않고 모든 변수를 사용한 MLR 분석 R(0.397), R^2(0.396), F(369.5), 유의확률(0.00), Durbin-Watson(1.880)
Case2	weak 중 하나인 임대료_연간소득만 제외 후 MLR 분석 R(0.389), R^2(0.388), F(384.4), 유의확률(0.00), Durbin-Watson(1.872)
Case3	weak 중 하나인 주택관련_부채액만 제외 후 MLR 분석 R(0.394), R^2(0.393), F(392.3), 유의확률(0.00), Durbin-Watson(1.876)
Case4	weak 중 하나인 주거관련_부채의_이자만 제외 후 MLR 분석 R(0.397), R^2(0.396), F(397.9), 유의확률(0.00), Durbin-Watson(1.880)
Case5	weak 중 임대료_연간소득 및 주택관련_부채액 2가지 weak변수 제외 후 MLR 분석 R(0.387), R^2(0.386), F(412.3), 유의확률(0.00), Durbin-Watson(1.869)
Case6	weak 중 임대료_연간소득 및 주거관련_부채의_이자 2가지 weak변수 제외 후 MLR 분석 R(0.389), R^2(0.388), F(416.5), 유의확률(0.00), Durbin-Watson(1.872)
Case7	weak 중 주택관련_부채액 및 주거관련_부채의_이자 2가지 weak변수 제외 후 MLR 분석 R(0.392), R^2(0.391), F(422.2), 유의확률(0.00), Durbin-Watson(1.876)
Case8	weak 변수 3가지 전부 제외 후 MLR 분석 시행 R(0.386), R^2(0.385), F(448.0), 유의확률(0.00), Durbin-Watson(1.869)



모든 경우의 수를 가정하여 회귀분석을
진행한 결과 독립변수들을 모두 포함하여
회귀분석을 진행한 Case1에서 가장 높은
설명력이 나타남.

스피어만 상관분석을 통해 상관관계가
있는 변수들을 발견했지만 VIF값이
정상범위를 나타내므로 모든 변수를
포함하는 Case1 채택

6. 다중선형회귀분

MLR

데이터 모델링

회귀분석 실시 선택 결과보고(3)

종속변수	독립변수	비표준화 계수		t	유의확률	공선성 통계량
		B	표준오차			
총생활비	민간보조금	0.0856	0.009	9.682*	0	1.02894
	이자_배당금_연간소득	0.0917	0.009	10.333*	0	1.03931
	기초연금	-0.2649	0.01	0	-27.099*	0
	기타이자	0.0277	0.01	0	2.755*	0.006
	공적연금	0.017	0.009	0	1.888*	0.059
	기타_소득	0.0133	0.009	0	1.517*	0.129
	금융기관대출	0.1322	0.014	0	9.705*	0
	부모가보낸생활비	0.0927	0.009	0	10.269*	0
	재난지원금	0.1813	0.009	0	19.978*	0
	세금환급금	0.2052	0.009	0	22.515*	0
	자녀가보낸생활비	-0.0255	0.009	0	-2.744*	0.006
	임대료_연간소득	0.0957	0.009	0	10.359*	0
	주택관련_부채액	0.098	0.015	0	6.676*	0
	주거관련_부채의_이자	0.0006	0.013	0	0.050*	0.96

R(0.397), R^2(0.396), F(369.5), 유의확률(0.00), Durbin-Watson(1.880) (*은 0.05 미만을 의미)

Case1을 기반으로 한
회귀분석 실시한 결과보고

분석결과 회귀모형은 통계적으로
유의미한 것으로 나타났으며
(F= 369.5, p<0.05)

설명력은 39.7%
(수정된 R제곱은 39.6%)로 나타남

Durbin-Watson값은 1.880으로 2에
근사한 값을 보여 잔차의 독립성 가정

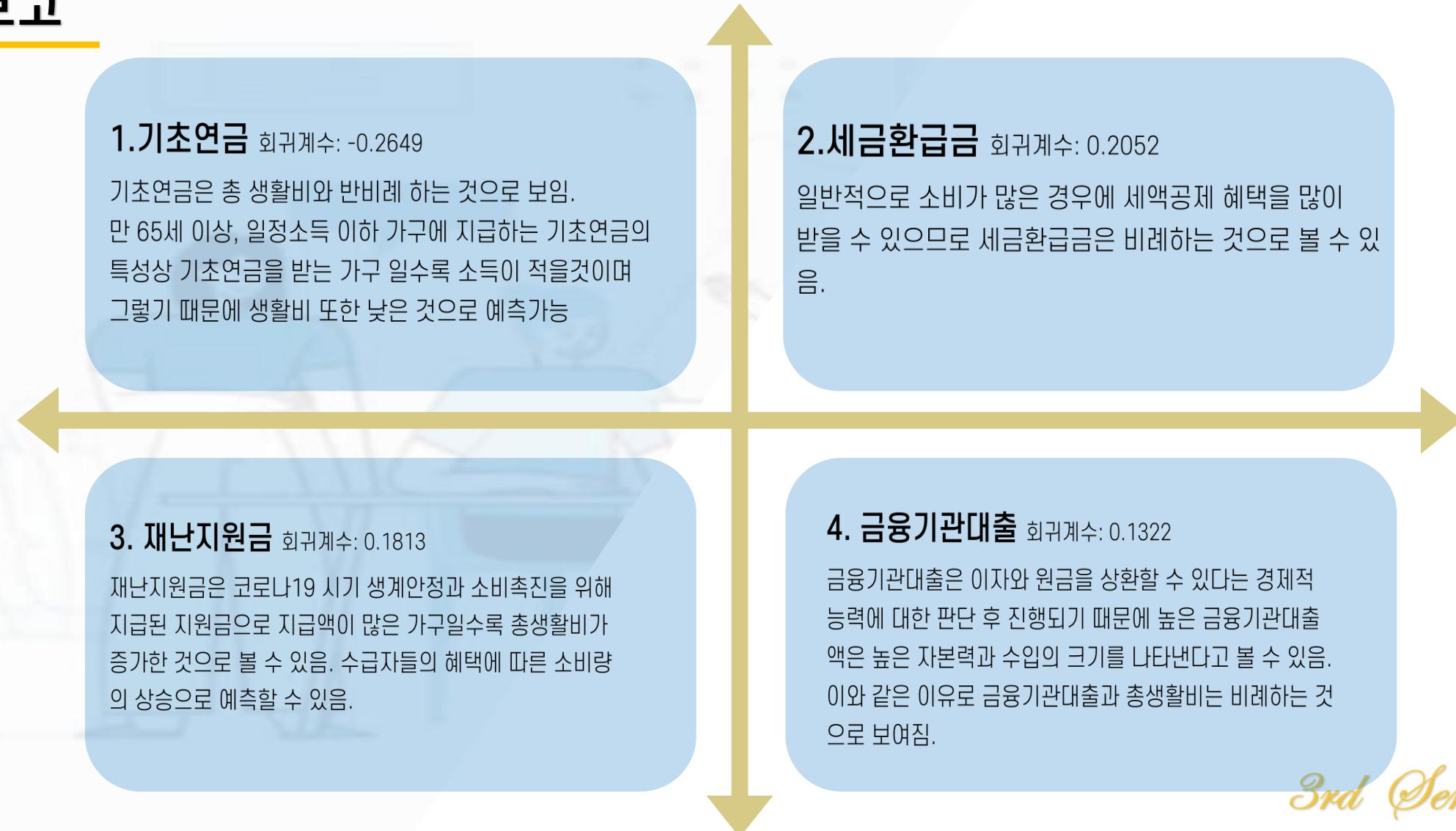
모든 변수들의 VIF값은 100이하로
다중공선성의 문제 발생하지 않음

3rd Semi Project

총 생활비에 가장 큰 영향을 미치는 4가지 요인 확인 및 집중 보고

7. 결과 보고

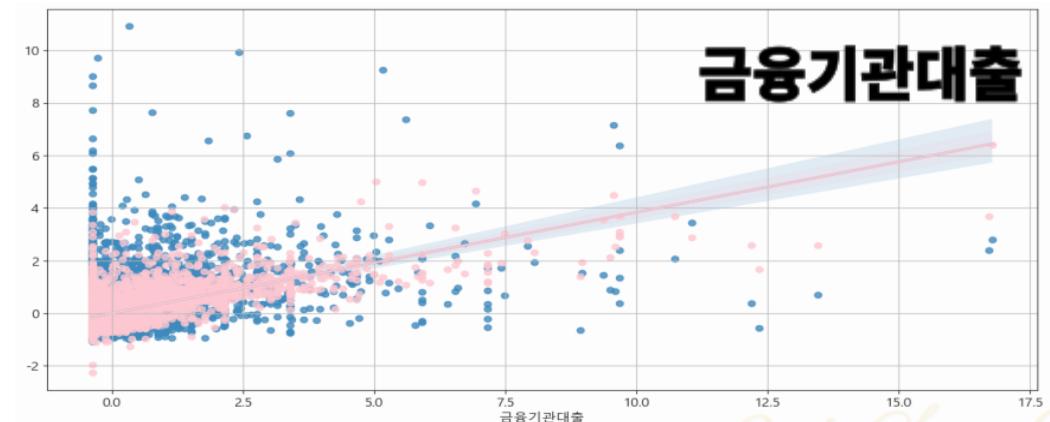
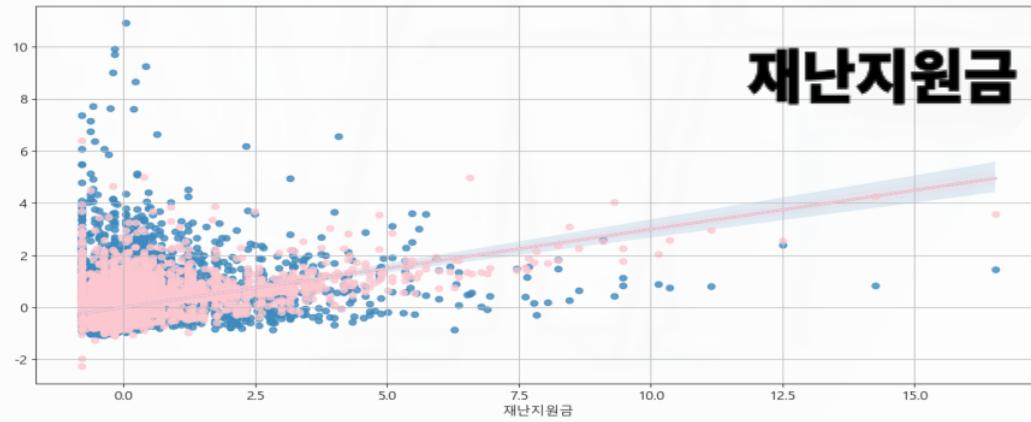
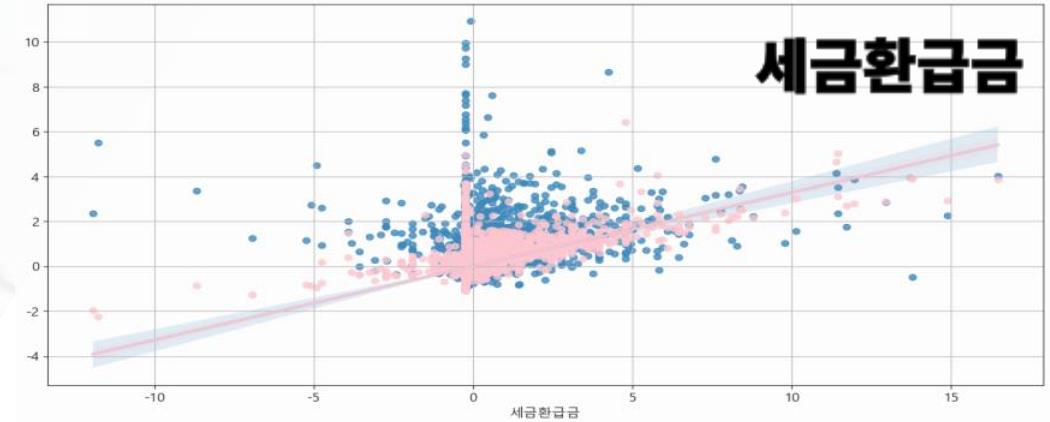
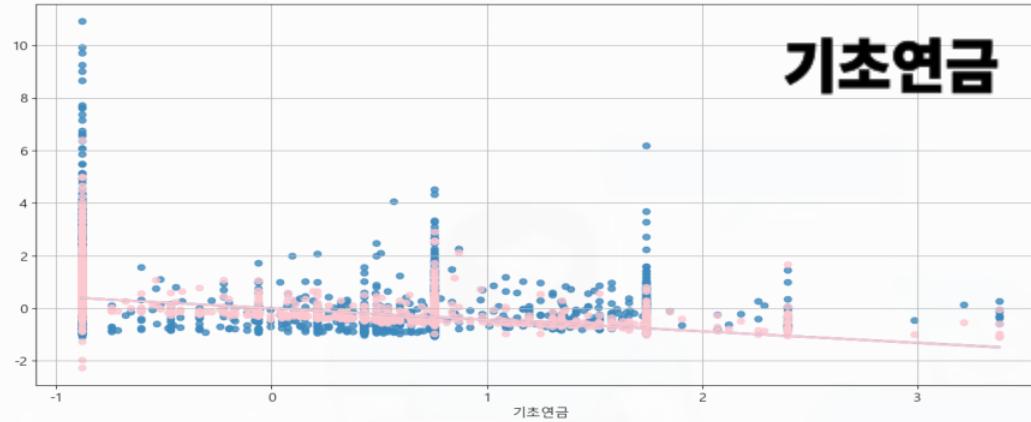
집중 보고



7. 결과 보고

Result
결과 해석

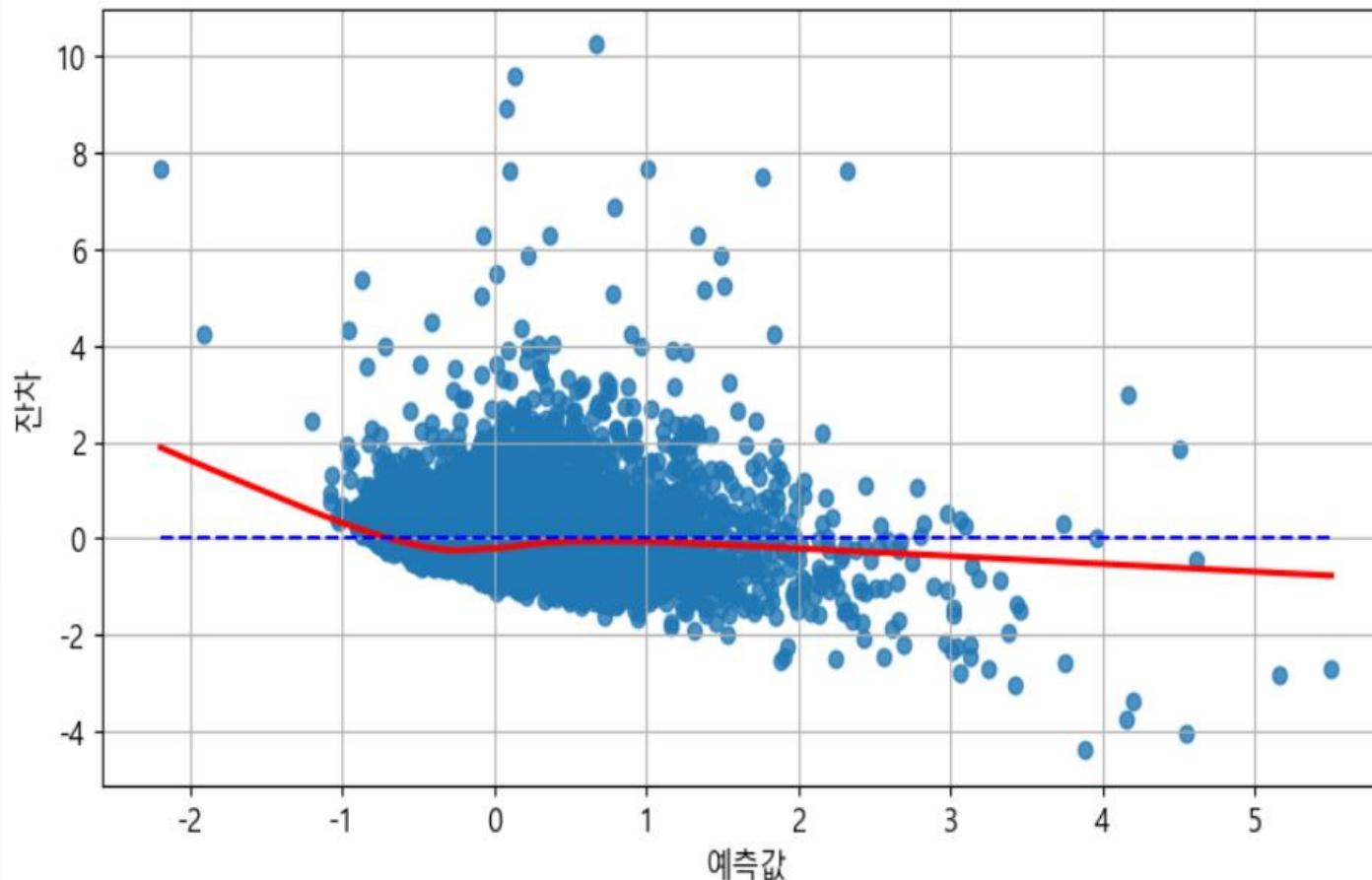
MLR 결과 시각화



7. 결과 보고

Result
결과 해석

잔차분석(선형성 확인)

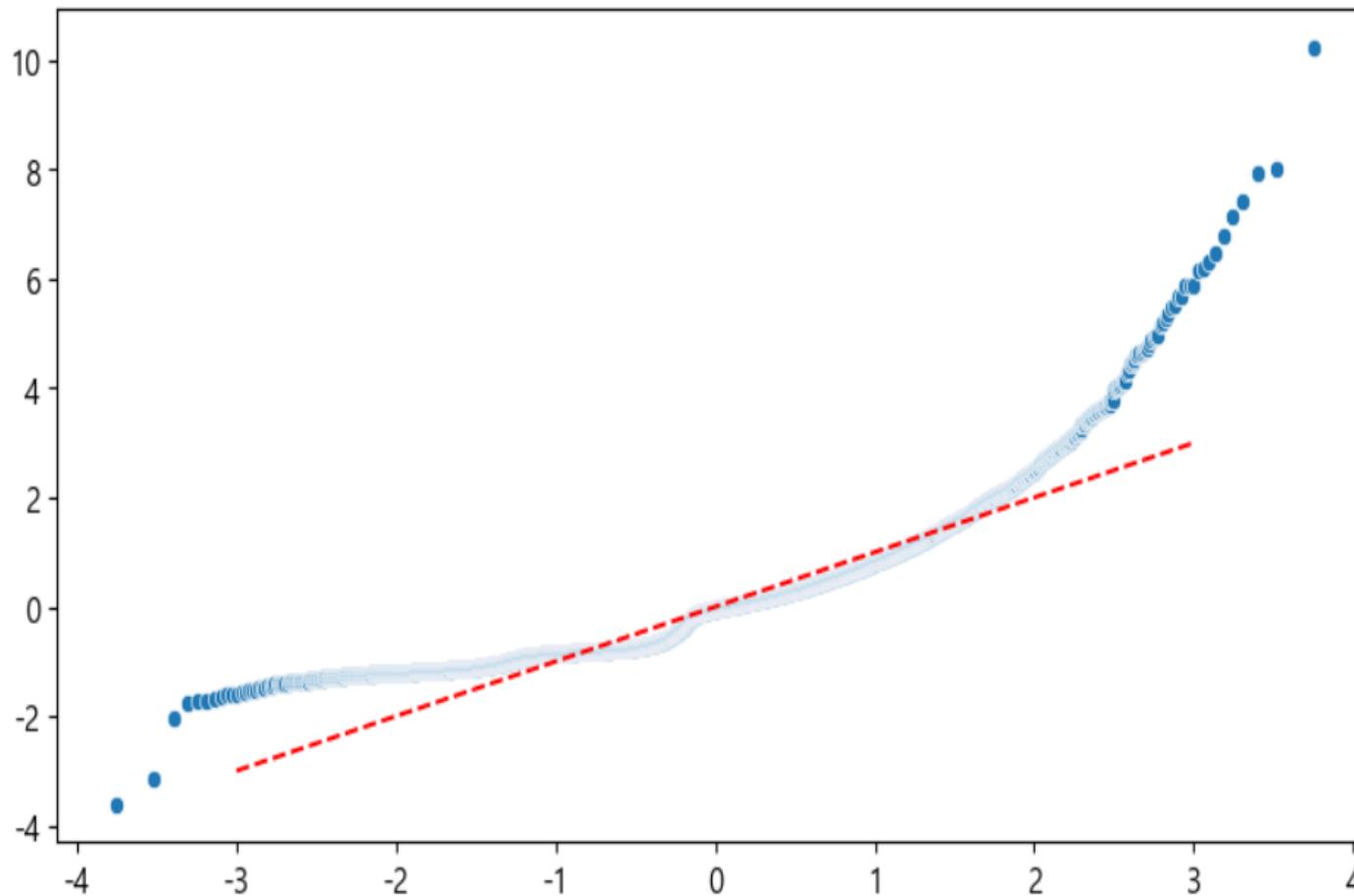


잔차의 추세선이 점선에서 크게 벗어나지 않는 것으로 보아 잔차의 선형성을 확인 할 수 있음

7. 결과 보고

Result
결과 해석

정규성 확인(Q-Q plot)



특정 구간에서 잔차값이 정규분포를 벗어난 것을 확인 할 수 있음

오른쪽 끝이 길게 뻗어져 있는 것으로 보아 평균보다 큰 값이 많음을 예측할 수 있음



Thank You!