

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

연습문제(2) 폴이

작업준비

```
from pca import pca
from pandas import read_excel, DataFrame
from matplotlib import pyplot as plt
import seaborn as sb
from sklearn.preprocessing import StandardScaler

import sys
import os
sys.path.append(os.path.dirname(os.path.abspath(__file__)))
from helper import my_ols
```

문제 1

데이터 가져오기

```
df = read_excel("https://data.hossam.l...")
df
```

	age	sex	bmi	
0	0.038076	0.050680	0.061696	0.0
1	-0.001882	-0.044642	-0.051474	-0.0
2	0.085299	0.050680	0.044451	-0.0
3	-0.089063	-0.044642	-0.011595	-0.0
4	0.005383	-0.044642	-0.036385	0.0
...
437	0.041708	0.050680	0.019662	0.0
438	-0.005515	0.050680	-0.015906	-0.0

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선택된 요인을 사용하여 회귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

	age	sex	bmi	
439	0.041708	0.050680	-0.015906	0.0
440	-0.045472	-0.044642	0.039062	0.0
441	-0.045472	-0.044642	-0.073030	-0

442 rows × 11 columns

독립변수만 추출

```
x_train = df.drop(['target'], axis=1)
x_train.head()
```

	age	sex	bmi	
0	0.038076	0.050680	0.061696	0.021
1	-0.001882	-0.044642	-0.051474	-0.02
2	0.085299	0.050680	0.044451	-0.00
3	-0.089063	-0.044642	-0.011595	-0.03
4	0.005383	-0.044642	-0.036385	0.021

전체 요인에 대한 회귀분석

```
xnames = list(df.columns)
xnames.remove("target")

model1 = my_ols(df, y='target', x=xnames)
model1
```

<helper.OlsResult at 0x16b2454d2d0>

요약 결과

연습문제(2) 폴이

작업준비

문제 1

- 데이터 가져오기
- 독립변수만 추출
- 전체 요인에 대한 회귀분석
 - 요약 결과
 - 결과표
 - 예측값 생성
- 차원 축소 적용
 - PCA 분석
 - 시각화를 통한 기여도 확인
 - 선택된 요인을 사용하여 회귀분석 수행
 - 요약 결과
 - 결과표
- 두 가지 모델의 결과 비교

문제 2

- 데이터 가져오기
- 데이터 전처리
 - 명목형 변수 확인
 - 불필요한 변수를 제거
 - 데이터 표준화
 - 종속변수 분리
 - 독립변수 분리
- PCA
 - 주성분 분석 시행
 - 기여도 확인
- 실제 임대료와 예측된 임대료

model1.summary

OLS Regression Results			
Dep. Variable:	target	R-squared:	0.511
Model:	OLS	Adj. R-squared:	0.507
Method:	Least Squares	F-statistic:	46.27
Date:	Wed, 26 Jul 2023	Prob (F-statistic):	3.83e-62
Time:	16:30:08	Log-Likelihood:	-238.4
No. Observations:	442	AIC:	4794.4
Df Residuals:	431	BIC:	4839.9
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>
Intercept	152.1335	2.576	59.061	0.000
age	-10.0099	59.749	-0.168	0.871
sex	-239.8156	61.222	-3.917	0.000
bmi	519.8459	66.533	7.813	0.000
bp	324.3846	65.422	4.958	0.000
s1	-792.1756	416.680	-1.901	0.061
s2	476.7390	339.030	1.406	0.160
s3	101.0433	212.531	0.475	0.634
s4	177.0632	161.476	1.097	0.275
s5	751.2737	171.900	4.370	0.000

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선택된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

s6	67.6267	65.984	1.025	0.3
----	---------	--------	-------	-----

Omnibus:	1.506	Durbin-Watson:	2.029
Prob(Omnibus):	0.471	Jarque-Bera (JB):	1.404
Skew:	0.017	Prob(JB):	0.496
Kurtosis:	2.726	Cond. No.	227.

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

결과표

model1.table

		B	표준오차	β	
종속 변수	독립 변수				
target	age	-10.0099	59.749	0	-0.
	sex	-239.8156	61.222	0	-3.
	bmi	519.8459	66.533	0	7.8
	bp	324.3846	65.422	0	4.9
	s1	-792.1756	416.680	0	-1.
	s2	476.7390	339.030	0	1.4
	s3	101.0433	212.531	0	0.4
	s4	177.0632	161.476	0	1.0
	s5	751.2737	171.900	0	4.3

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선택된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

		B	표준오차	β	
종속 변수	독립 변수				
	s6	67.6267	65.984	0	1.0

예측값 생성

```
y1 = model1.fit.predict(x_train)
y1
```

```
0      206.116677
1      68.071033
2     176.882790
3     166.914458
4     128.462258
...
437    193.017565
438    101.055814
439    121.226681
440    211.858945
441     53.447275
Length: 442, dtype: float64
```

차원 축소 적용

PCA 분석

```
pca_model = pca(n_components=len(x_train))
fit = pca_model.fit_transform(x_train)
fit['topfeat']
```

```
[pca] >Extracting column labels from data
[pca] >Extracting row labels from data
[pca] >The PCA reduction is performed
```

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

[pca] >Fit using PCA.

[pca] >Compute loadings and PCs.

[pca] >Compute explained variance.

[pca] >Outlier detection using Hotell

[pca] >Multiple test correction applic

[pca] >Outlier detection using SPE/Dmc

	PC	feature	loading	type
0	PC1	s4	0.428834	best
1	PC2	s1	0.573027	best
2	PC3	bp	0.513571	best
3	PC4	sex	-0.679860	best
4	PC5	age	-0.686876	best
5	PC6	s6	0.805066	best
6	PC7	s5	0.647302	best
7	PC8	bp	-0.477360	best
8	PC9	s4	0.773817	best
9	PC10	s1	0.709775	best
10	PC7	bmi	-0.519874	weak
11	PC10	s2	-0.563195	weak
12	PC2	s3	0.506239	weak

시각화를 통한 기여도 확인

```
pca_df = fit['topfeat']
pca_df.sort_values('loading', ascending=True)
pca_df.set_index('feature', inplace=True)
pca_df['loading']
```

```
xlabel = pca_df['loading'].index
y = pca_df['loading'].values
x = list(range(0, len(y)))
sb.lineplot(x=x, y=y)
plt.xticks(x, xlabel, rotation=90)
```

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

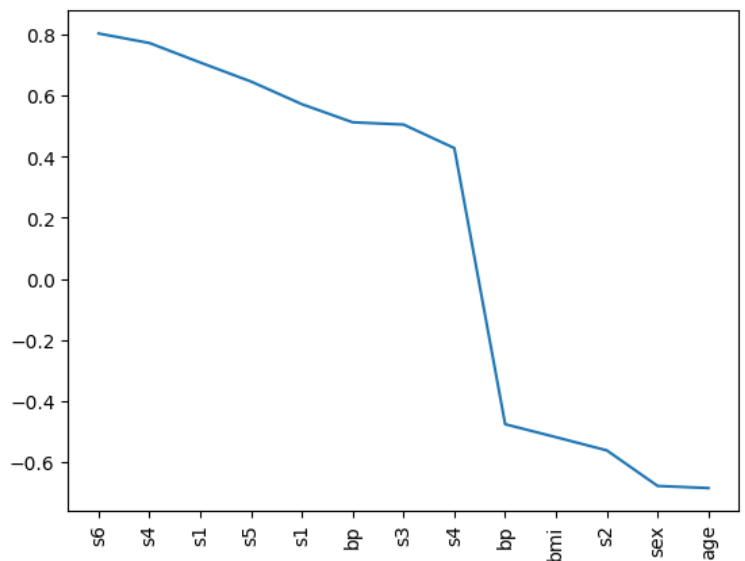
PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

```
plt.show()
plt.close()
```



선정된 요인을 사용하여 회귀분석 수행

```
model2 = my_ols(df, y='target', x=['s6', 's4', 's1', 's5', 's1', 'bp', 's3', 's4', 'bp', 'bmi', 's2', 'sex', 'age'])
```

요약 결과

```
model2.summary
```

OLS Regression Results

Dep. Variable:	target	R-squared:	0.410
Model:	OLS	Adj. R-squared:	0.400
Method:	Least Squares	F-statistic:	51.5
Date:	Wed, 26 Jul 2023	Prob (F-statistic):	7.70e-48
Time:	16:30:08	Log-Likelihood:	-242
No. Observations:	442	AIC:	4871

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

Df Residuals:	435	BIC:	4900
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>
Intercept	152.1335	2.823	53.899	0.000
s6	126.4813	70.640	1.790	0.083
s4	35.9208	174.587	0.206	0.843
s1	-138.8137	123.913	-1.120	0.265
s5	658.9295	84.467	7.801	0.000
bp	387.7776	67.065	5.782	0.000
s3	-238.8947	149.211	-1.601	0.117

Omnibus:	5.904	Durbin-Watson:	2.075
Prob(Omnibus):	0.052	Jarque-Bera (JB):	5.522
Skew:	0.222	Prob(JB):	0.0632
Kurtosis:	2.680	Cond. No.	87.7

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

결과표

```
y2 = model2.fit.predict(x_train)
y2
```


연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

```

0      187.914775
1       67.214133
2     162.506693
3     159.833966
4     131.053881
...
437    204.313806
438    120.827759
439    140.662926
440    184.205295
441     63.602616
Length: 442, dtype: float64

```

두 가지 모델의 결과 비교

```

rdf = DataFrame({
    's6': df['s6'],
    'y': df['target'],
    'y1': y1,
    'y2': y2
})

rdf

```

	s6	y	y1	
0	-0.017646	151	206.116677	187.914
1	-0.092204	75	68.071033	67.214
2	-0.025930	141	176.882790	162.506
3	-0.009362	206	166.914458	159.833
4	-0.046641	135	128.462258	131.053
...
437	0.007207	178	193.017565	204.313
438	0.044485	104	101.055814	120.827
439	0.015491	132	121.226681	140.662
440	-0.025930	220	211.858945	184.205

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선택된 요인을 사용하여 회귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

	s6	y	y1	
441	0.003064	57	53.447275	63.6020

442 rows × 4 columns

```
mdf1 = melt(rdf, id_vars=['s6'], value_vars=['y', 'y1'])
mdf1
```

	s6	variable	value
0	-0.017646	y	151.000000
1	-0.092204	y	75.000000
2	-0.025930	y	141.000000
3	-0.009362	y	206.000000
4	-0.046641	y	135.000000
...
879	0.007207	y1	193.017565
880	0.044485	y1	101.055814
881	0.015491	y1	121.226681
882	-0.025930	y1	211.858945
883	0.003064	y1	53.447275

884 rows × 3 columns

```
mdf2 = melt(rdf, id_vars=['s6'], value_vars=['y', 'y1'])
mdf2
```

	s6	variable	value
0	-0.017646	y	151.000000
1	-0.092204	y	75.000000
2	-0.025930	y	141.000000

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

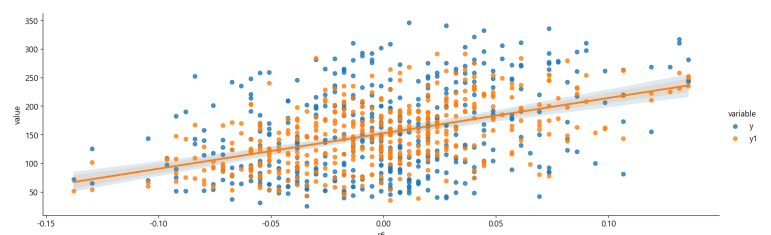
실제 임대료와 예측된 임대료

	s6	variable	value
3	-0.009362	y	206.000000
4	-0.046641	y	135.000000
...
879	0.007207	y2	204.313806
880	0.044485	y2	120.827759
881	0.015491	y2	140.662926
882	-0.025930	y2	184.205295
883	0.003064	y2	63.602616

884 rows × 3 columns

```
plt.rcParams["font.family"] = 'AppleGothic'
plt.rcParams["font.size"] = 12
plt.rcParams["figure.figsize"] = (10, 10)
plt.rcParams["axes.unicode_minus"] = False
```

```
sb.lmplot(data=mdf1, x="s6", y="value")
plt.show()
plt.close()
```



```
sb.lmplot(data=mdf2, x="s6", y="value")
plt.show()
plt.close()
```



연습문제(2) 풀이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

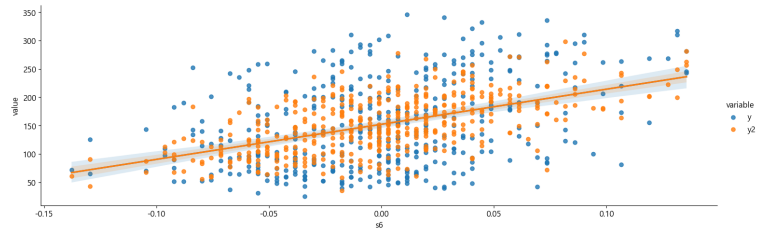
PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

연습문제2_풀이.ipynb



문제 2

데이터 가져오기

```
df = read_excel("https://data.hossam.l")
df
```

	rent	bedrooms	bathrooms	size
0	2550	0.0	1	480
1	11500	2.0	2	200
2	4500	1.0	1	916
3	4795	1.0	1	975
4	17500	2.0	2	480
...
3534	4210	1.0	1	532
3535	6675	2.0	2	988
3536	1699	0.0	1	250
3537	3475	1.0	1	651
3538	4500	1.0	1	816

3539 rows × 17 columns

데이터 전처리

명목형 변수 확인

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

값의 종류가 지나치게 많으므로 이 변수는 제외하
기로 함

```
df['neighborhood'].value_counts()
```

```
neighborhood
Upper West Side      579
Upper East Side      500
Midtown East         460
Midtown West         314
Financial District   268
Chelsea              182
Flatiron             132
Midtown              119
Tribeca              119
East Village         108
Battery Park City    104
Midtown South        85
Central Harlem       82
West Village         67
Greenwich Village    66
Gramercy Park        61
Soho                  58
Washington Heights   54
East Harlem          41
Lower East Side      41
Central Park South   23
Hamilton Heights     16
Morningside Heights  13
Inwood                12
Nolita                9
Chinatown             8
Roosevelt Island     5
Long Island City      4
Stuyvesant Town/PCV   3
Little Italy          3
West Harlem           2
Manhattanville        1
Name: count, dtype: int64
```

값의 종류가 없으므로 이 변수도 제외하기로 함

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선택된 요인을 사용하여 회귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

```
df['borough'].value_counts()
```

```
borough
Manhattan    3539
Name: count, dtype: int64
```

불필요한 변수를 제거

```
df2 = df.drop(['neighborhood', 'borough'])
df2.head()
```

	rent	bedrooms	bathrooms	size_sq
0	2550	0.0	1	480
1	11500	2.0	2	2000
2	4500	1.0	1	916
3	4795	1.0	1	975
4	17500	2.0	2	4800

데이터 표준화

```
scaler = StandardScaler()
std_values = scaler.fit_transform(df2)
std_df = DataFrame(std_values, columns=df2.columns)
std_df
```

	rent	bedrooms	bathrooms
0	-0.818669	-1.397410	-0.611790
1	2.011480	0.669863	1.056257
2	-0.202044	-0.363774	-0.611790
3	-0.108760	-0.363774	-0.611790
4	3.908786	0.669863	1.056257

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

	rent	bedrooms	bathrooms
...
3534	-0.293747	-0.363774	-0.611790
3535	0.485729	0.669863	1.056257
3536	-1.087770	-1.397410	-0.611790
3537	-0.526167	-0.363774	-0.611790
3538	-0.202044	-0.363774	-0.611790

3539 rows × 15 columns

종속변수 분리

```
y_train = std_df.filter(['rent'])
y_train.head()
```

	rent
0	-0.818669
1	2.011480
2	-0.202044
3	-0.108760
4	3.908786

독립변수 분리

```
x_train = std_df.drop(['rent'], axis=1)
x_train.head()
```

	bedrooms	bathrooms	size_sqft	mi
0	-1.397410	-0.611790	-0.962011	0.7
1	0.669863	1.056257	2.218694	-0.1
2	-0.363774	-0.611790	-0.049651	-0.5

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선택된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

	bedrooms	bathrooms	size_sqft	mi
3	-0.363774	-0.611790	0.073811	-0.3
4	0.669863	1.056257	8.077886	-0.3

PCA

주성분 분석 시행

```
pca_model = pca(n_components=len(x_train))
fit = pca_model.fit_transform(x_train)
fit['topfeat']
```

```
[pca] >Extracting column labels from data
[pca] >Extracting row labels from data
[pca] >The PCA reduction is performed
[pca] >Fit using PCA.
[pca] >Compute loadings and PCs.
[pca] >Compute explained variance.
[pca] >Outlier detection using Hotelling's
[pca] >Multiple test correction applied
[pca] >Outlier detection using SPE/Dmod
```

	PC	feature	loading	t
0	PC1	has_elevator	0.453207	b
1	PC2	size_sqft	0.573482	b
2	PC3	building_age_yrs	-0.638744	b
3	PC4	has_patio	0.743626	b
4	PC5	min_to_subway	0.624977	b
5	PC6	min_to_subway	0.672089	b
6	PC7	no_fee	0.666304	b
7	PC8	building_age_yrs	0.673353	b
8	PC9	has_roofdeck	0.801673	b

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

	PC	feature	loading	t
9	PC10	has_dishwasher	-0.650467	b
10	PC11	has_gym	0.834446	b
11	PC12	has_elevator	0.675102	b
12	PC13	bedrooms	-0.726891	b
13	PC14	size_sqft	0.784195	b
14	PC2	bathrooms	0.564578	w
15	PC8	floor	0.546436	w
16	PC10	has_washer_dryer	0.617406	w
17	PC12	has_doorman	-0.619009	w

기여도 확인

```
pca_df = fit['topfeat']
pca_df.sort_values('loading', ascending=True)
pca_df.set_index('feature', inplace=True)
pca_df['loading']
```

```
xlabel = pca_df['loading'].index
y = pca_df['loading'].values
x = list(range(0, len(y)))
sb.lineplot(x=x, y=y)
plt.xticks(x, xlabel, rotation=90)
plt.show()
plt.close()
```

연습문제(2) 풀이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

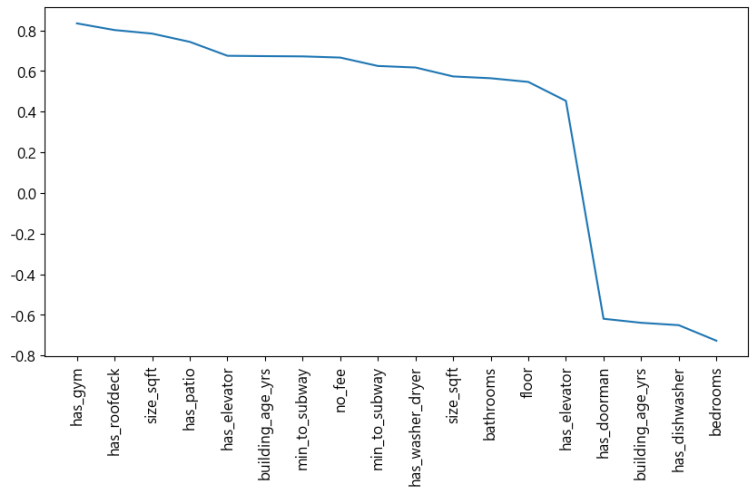
PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

연습문제2_풀이.ipynb



```
ols_result = my_ols(std_df, y='rent',
ols_result.summary
```

OLS Regression Results

Dep. Variable:	rent	R-squared:	0.771
Model:	OLS	Adj. R-squared:	0.77
Method:	Least Squares	F-statistic:	1109
Date:	Wed, 26 Jul 2023	Prob (F-statistic):	0.00
Time:	17:19:13	Log-Likelihood:	-237
No. Observations:	3539	AIC:	4778
Df Residuals:	3527	BIC:	4852
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t
Intercept	-9.346e-	0.008	-1.17e-

연습문제(2) 폴이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

	17		14
has_gym	-0.0065	0.011	-0.581
has_roofdeck	0.0006	0.010	0.061
size_sqft	0.6919	0.014	50.417
has_patio	-0.0069	0.008	-0.852
has_elevator	0.0032	0.011	0.288
building_age_yrs	-0.0981	0.009	-10.815
min_to_subway	-0.0326	0.008	-3.996
no_fee	-0.0185	0.008	-2.199
has_washer_dryer	0.0167	0.009	1.896
bathrooms	0.1951	0.014	14.260
floor	0.0831	0.009	9.497

Omnibus:	1011.570	Durbin-Watson:	2.077
Prob(Omnibus):	0.000	Jarque-Bera (JB):	10053
Skew:	1.068	Prob(JB):	0.00
Kurtosis:	10.976	Cond. No.	3.69

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

실제 임대료와 예측된 임대료 비교

```
rdf = DataFrame({
    'x': std_df['has_gym'],
    'std_rent': std_df['rent'],
    'std_y': ols_result.fit.predict(x,
```

연습문제(2) 풀이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회
귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

```

'real_rent': df['rent'],
'real_y': ols_result.fit.predict(c
})

```

rdf

	x	std_rent	std_y	r
0	2.171931	-0.818669	-0.832595	2
1	-0.460420	2.011480	1.564857	1
2	-0.460420	-0.202044	0.236502	4
3	2.171931	-0.108760	-0.033789	4
4	2.171931	3.908786	5.537699	1
...
3534	2.171931	-0.293747	-0.628387	4
3535	2.171931	0.485729	0.378069	6
3536	-0.460420	-1.087770	-1.252172	1
3537	2.171931	-0.526167	-0.493272	3
3538	2.171931	-0.202044	-0.175468	4

3539 rows × 5 columns

```

mdf1 = melt(rdf, id_vars=['x'], value_
mdf1

```

	x	variable	value
0	2.171931	std_rent	-0.818669
1	-0.460420	std_rent	2.011480
2	-0.460420	std_rent	-0.202044
3	2.171931	std_rent	-0.108760
4	2.171931	std_rent	3.908786
...

연습문제(2) 풀이

작업준비

문제 1

데이터 가져오기

독립변수만 추출

전체 요인에 대한 회귀분석

요약 결과

결과표

예측값 생성

차원 축소 적용

PCA 분석

시각화를 통한 기여도 확인

선정된 요인을 사용하여 회귀분석 수행

요약 결과

결과표

두 가지 모델의 결과 비교

문제 2

데이터 가져오기

데이터 전처리

명목형 변수 확인

불필요한 변수를 제거

데이터 표준화

종속변수 분리

독립변수 분리

PCA

주성분 분석 시행

기여도 확인

실제 임대료와 예측된 임대료

	x	variable	value
7073	2.171931	std_y	-0.628387
7074	2.171931	std_y	0.378069
7075	-0.460420	std_y	-1.252172
7076	2.171931	std_y	-0.493272
7077	2.171931	std_y	-0.175468

7078 rows × 3 columns

```
sb.lmplot(data=mdf1, x="x", y="value")
plt.show()
plt.close()
```

