SERVEY.csv EDA

In [ ]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy import stats
```

In [ ]:
```
survey = pd.read_csv("C:/Users/ASUS/바탕 화면/wadadada/data/SURVEY.csv", encodin
```

In [ ]:
```
survey.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5720 entries, 0 to 5719
Columns: 110 entries, 조사차수 to 기준연도
dtypes: float64(46), int64(64)
memory usage: 4.8 MB
```

In [ ]:
```
survey_cols=list(survey.columns)
survey_cols
```

Out[ ]:
```
['조사차수',
 '일반_대표자_성별코드',
 '대표자연령대코드',
 '일반_합계종사자수',
 '일반_합계_남자종사자수',
 '일반_합계_여자종사자수',
 '일반_대표자종사자수',
 '일반_대표자_남자종사자수',
 '일반_대표자_여자종사자수',
 '일반_근로계약기간_1년이상_종사자수',
 '일반_근로계약기간_1년이상_남자종사자수',
 '일반_근로계약기간_1년이상_여자종사자수',
 '일반_근로계약기간_3개월이상1년미만_종사자수',
 '일반_근로계약기간_3개월이상1년미만_남자종사자수',
 '일반_근로계약기간_3개월이상1년미만_여자종사자수',
 '일반_근로계약기간_3개월미만_종사자수',
 '일반_근로계약기간_3개월미만_남자종사자수',
 '일반_근로계약기간_3개월미만_여자종사자수',
 '일반_무급가족종사자수',
 '일반_무급가족_남자종사자수',
 '일반_무급가족_여자종사자수',
 '일반_기타근로자종사자수',
 '일반_기타근로자_남자종사자수',
 '일반_기타근로자_여자종사자수',
 '일반_사업자형태코드',
 '산업중분류코드',
 '일반_프랜차이즈가맹점여부',
 '일반_창업형태코드',
 '일반_창업인수승계_연도',
 '일반_창업인수승계_월',
 '일반_사업장이전_경험여부',
 '일반_사업장이전_사유코드',
 '일반_현사업체_직전종사상지위항목코드',
 '일반_창업횟수',
 '창업_동기코드',
 '창업_준비기간_창업_준비기간_년수',
 '창업_준비기간_창업_준비기간_월수',
 '창업_준비활동중요성_사업계획서작성코드',
 '창업_준비활동중요성_시장조사코드',
 '창업_준비활동중요성_동종업종종사경험코드',
 '창업_준비활동중요성_창업교육코드',
 '창업_준비활동_사업계획서작성여부',
 '창업_준비활동_시장조사여부',
 '창업_준비활동_동종업종종사경험여부',
 '창업_준비활동_창업교육여부',
 '창업_어려움정도_입지선정코드',
 '창업_어려움정도_업종선택코드',
 '창업_어려움정도_자금조달코드',
 '창업_어려움정도_기술확보코드',
 '창업_어려움정도_인력확보코드',
 '창업_어려움정도_행정절차코드',
 '창업_어려움정도_경영방법코드',
 '창업_비용_총창업비용',
 '창업_비용_본인부담금액',
 '경영_영업기간_하루평균시간수',
 '경영_영업기간_월평균일수',
 '경영_영업기간_년간월수',
 '경영_고객결재방법_현금비율',
 '경영_고객결재방법_카드비율',
 '경영_고객결재방법_간편결제비율',
```

```
 '경영_고객결재방법_어음비율',
 '경영_고객결재방법_기타비율',
 '경영_고객결재방법_합계비율',
 '경영_점유형태코드',
 '경영_점유형태_임차형태코드',
 '경영_점유형태_보증금액',
 '경영_점유형태_월세금액',
 '경영_점유형태_일정비율지급금액',
 '경영_운영활동코드1',
 '경영_운영활동코드2',
 '경영_운영활동코드3',
 '경영_운영활동코드4',
 '경영_운영활동코드5',
 '경영_운영활동_디지털대응1코드',
 '경영_운영활동_디지털대응2코드',
 '경영_운영활동_디지털대응3코드',
 '경영_운영활동_디지털대응4코드',
 '경영_운영활동_디지털대응5코드',
 '경영_운영활동_디지털대응6코드',
 '경영_단체활동여부',
 '경영_매출금액',
 '경영_영업비용',
 '경영_영업비용_매출원가',
 '경영_영업비용_급여총액',
 '경영_영업비용_임차료',
 '경영_영업비용_기타금액',
 '경영_영업이익',
 '경영_전자상거래_매출실적여부',
 '경영_전자상거래_매출비율',
 '경영_부채여부',
 '경영_부채금액',
 '경영_애로사항1코드',
 '경영_애로사항2코드',
 '정부지원정책_지원경험코드',
 '정부지원정책_추진정책1코드',
 '정부지원정책_추진정책2코드',
 '코로나19_지원정책1코드',
 '코로나19_지원정책2코드',
 '사업전환_운영계획코드',
 '사업전환_운영계획_소유권승계계획여부',
 '사업전환_노후준비1해당여부',
 '사업전환_노후준비2해당여부',
 '사업전환_노후준비3해당여부',
 '사업전환_노후준비4해당여부',
 '사업전환_노후준비5해당여부',
 '사업체수가중값',
 '종사자수가중값',
 '매출금액가중값',
 '행정구역시도코드',
 '기준연도']
```

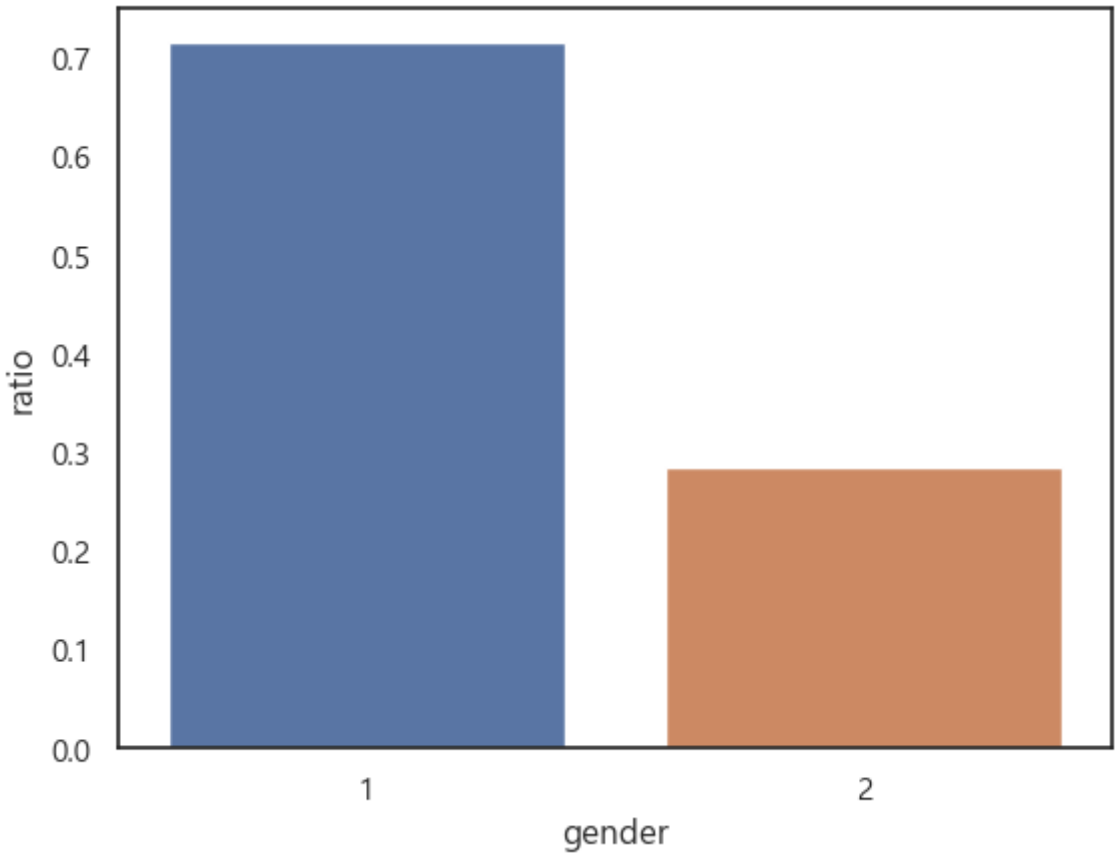
In [ ]:
```
# 대표자 정보
survey_1=survey[survey_cols[1:3]]
survey_1
```

Out[ ]:

	일반_대표자_성별코드	대표자연령대코드
0	1	20
1	1	20
2	1	20
3	1	20
4	1	20
...	...	...
5715	2	50
5716	2	60
5717	2	60
5718	2	70
5719	1	40

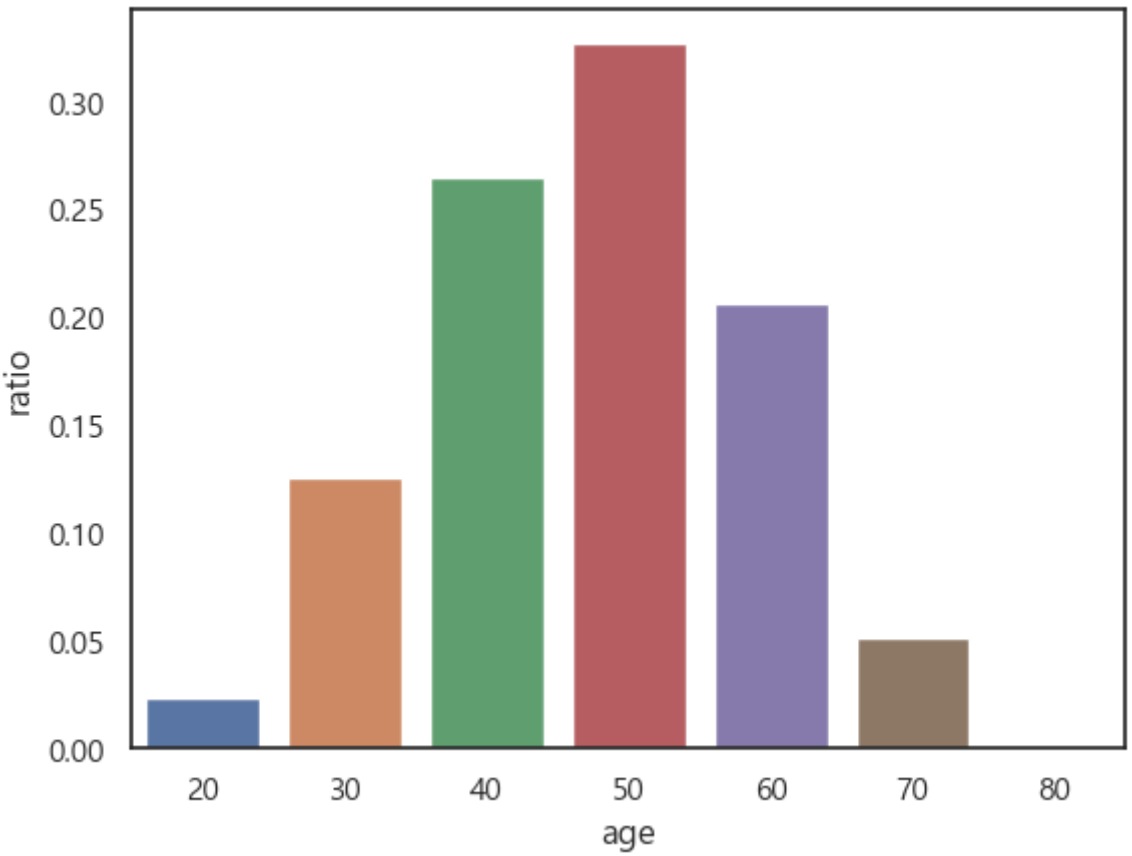
5720 rows × 2 columns

In [ ]:
```
sns.barplot(x='일반_대표자_성별코드', y='일반_대표자_성별코드', data=survey_1, es
plt.xlabel('gender')
plt.ylabel('ratio')
plt.show()
```

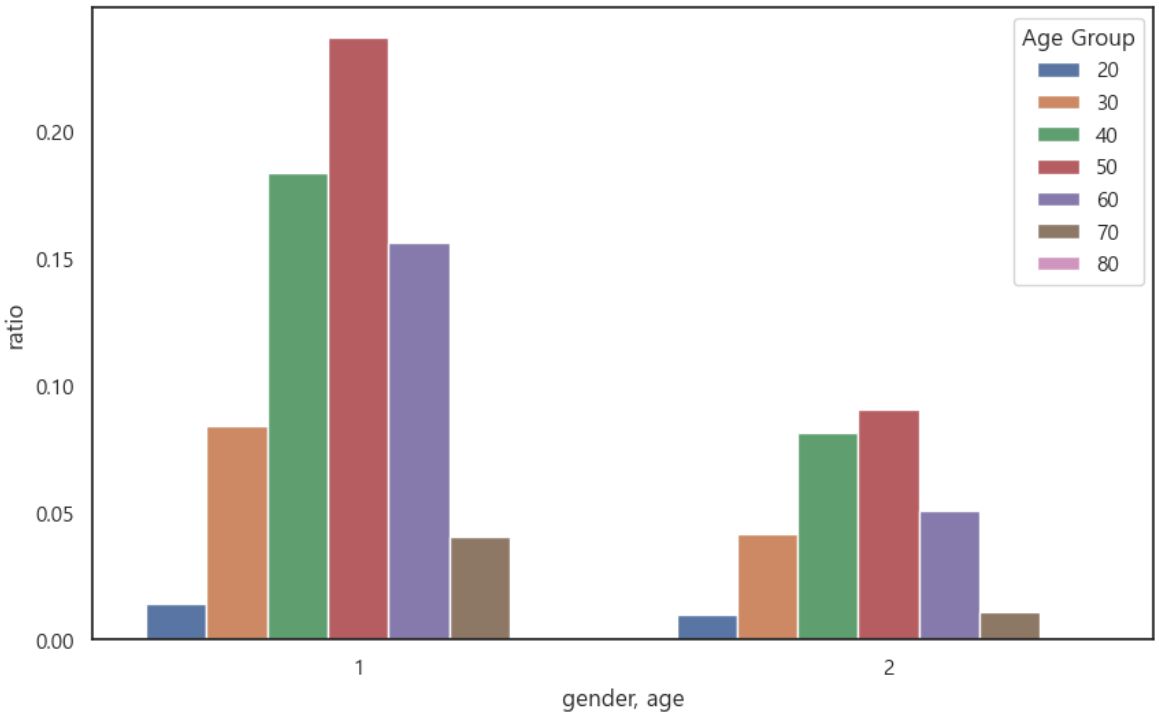


In [ ]:
```
sns.barplot(x='대표자연령대코드', y='대표자연령대코드', data=survey_1, estimator=
plt.xlabel('age')
```

```
plt.ylabel('ratio')
plt.show()
```



In [ ]:
```
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))
sns.barplot(x='일반_대표자_성별코드', y='대표자연령대코드', data=survey_1, estima

plt.xlabel('gender, age')
plt.ylabel('ratio')
plt.legend(title='Age Group', loc='upper right')

plt.show()
```



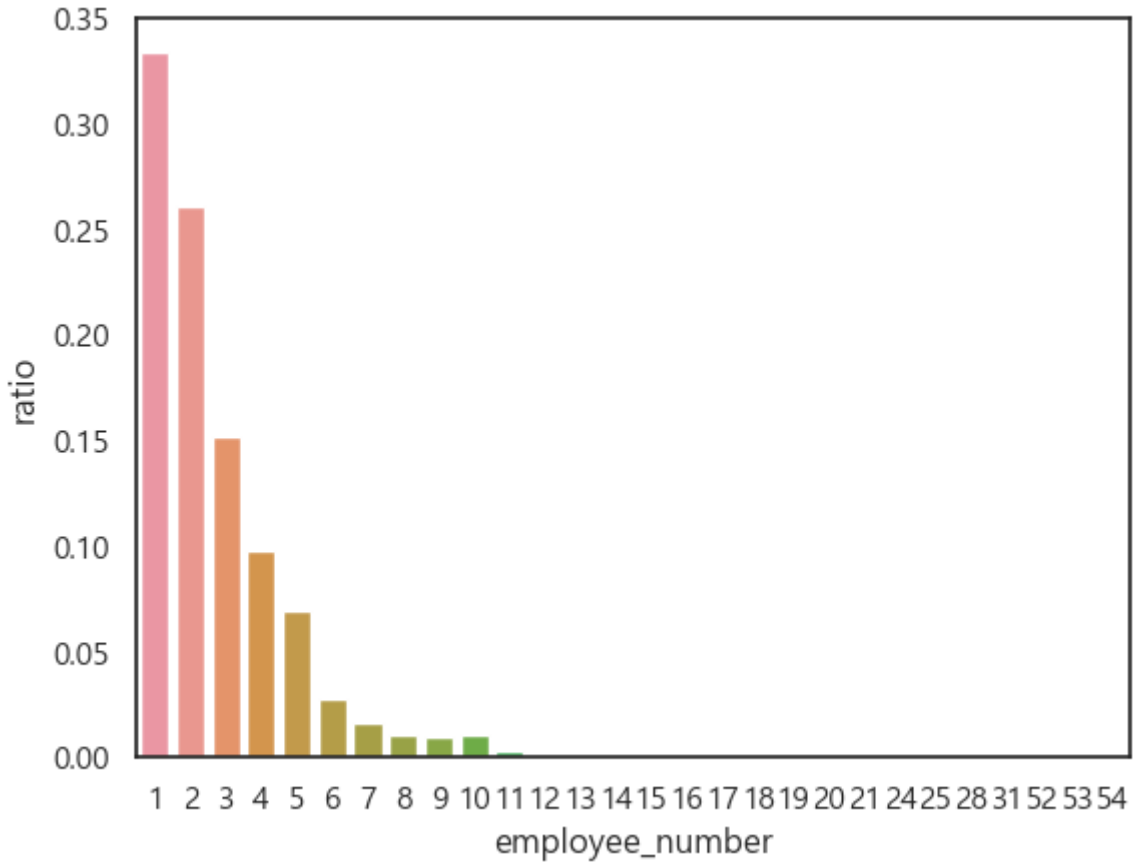
In [ ]:
```
# 종사자 수
survey_2=survey[survey_cols[3:24]]
survey_2
```

Out[ ]:

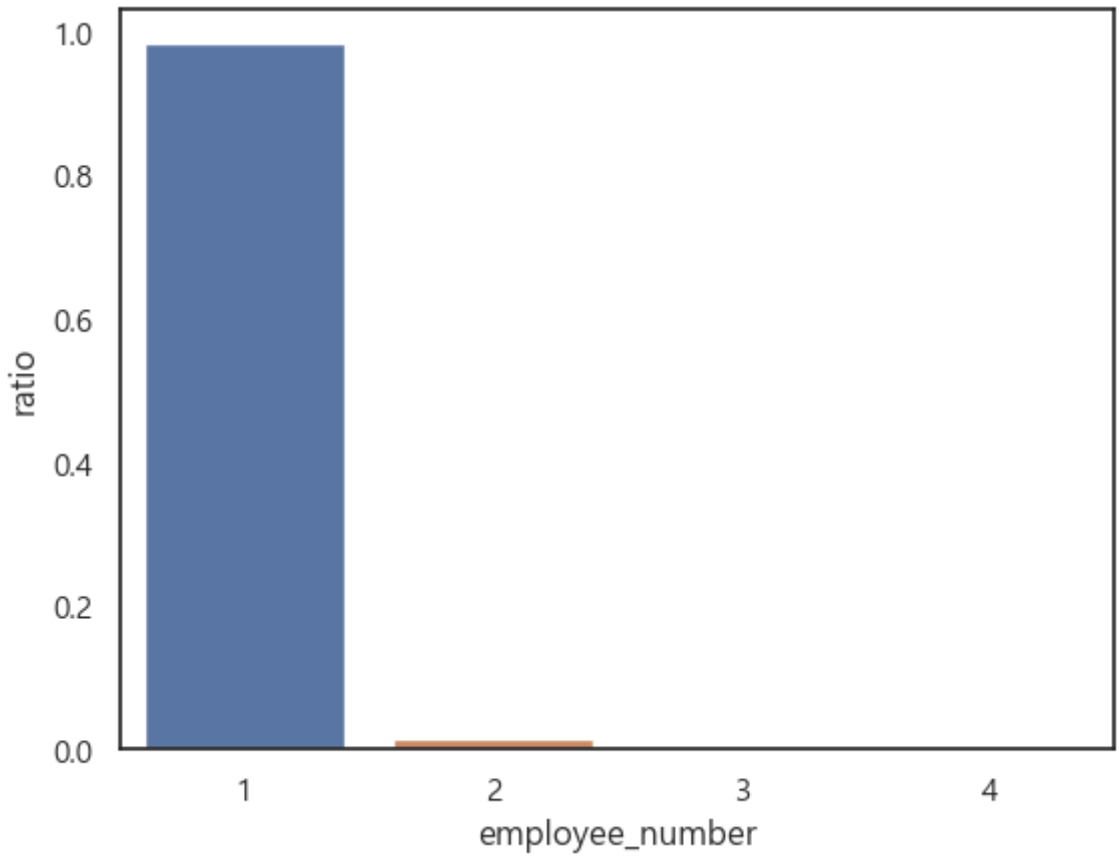
	일반_합계종사자수	일반_합계_남자종사자수	일반_합계_여자종사자수	일반_대표자종사자수	일반_대표자_남자종사자수	일반_대표자_여자종사자수	일반_근로계약기간_1년이상_종사자수	일반_근로계약기간_1년이상_남자종사자수	일반_근로계약기간_1년이상_여자종사자수	일반_근로계약기간_3개월이상1년미만_종사자수	...	일반_근로계약기간_3개월이상1년미만_여자종사자수	일반_근로계약기간_3개월미만_종사자수	일반_근로계약기간_3개월미만_남자종사자수	일반_근로계약기간_3개월미만_여자종사자수	일반_무급가족종사자수	일반_무급가족_남자종사자수	일반_무급가족_여자종사자수	일반_기타근로자종사자수	일반_기타근로자_남자종사자수	일반_기타근로자_여자종사자수
0	1	1	0	1	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	1	1	0	1	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	1	1	0	1	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	1	1	0	1	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	1	1	0	1	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5715	6	2	4	1	0	1	2	2	0	0	...	0	3	0	3	0	0	0	0	0	0
5716	1	0	1	1	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
5717	4	3	1	2	1	1	0	0	0	0	...	0	2	2	0	0	0	0	0	0	0
5718	4	1	3	1	0	1	1	0	1	1	...	1	0	0	0	1	1	0	0	0	0
5719	1	1	0	1	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5720 rows × 21 columns

In [ ]:
```
sns.barplot(x='일반_합계종사자수', y='일반_합계종사자수', data=survey_2, estimato
plt.xlabel('employee_number')
plt.ylabel('ratio')
plt.show()
```



In [ ]:
```
sns.barplot(x='일반_대표자종사자수', y='일반_대표자종사자수', data=survey_2, esti
plt.xlabel('employee_number')
plt.ylabel('ratio')
plt.show()
```

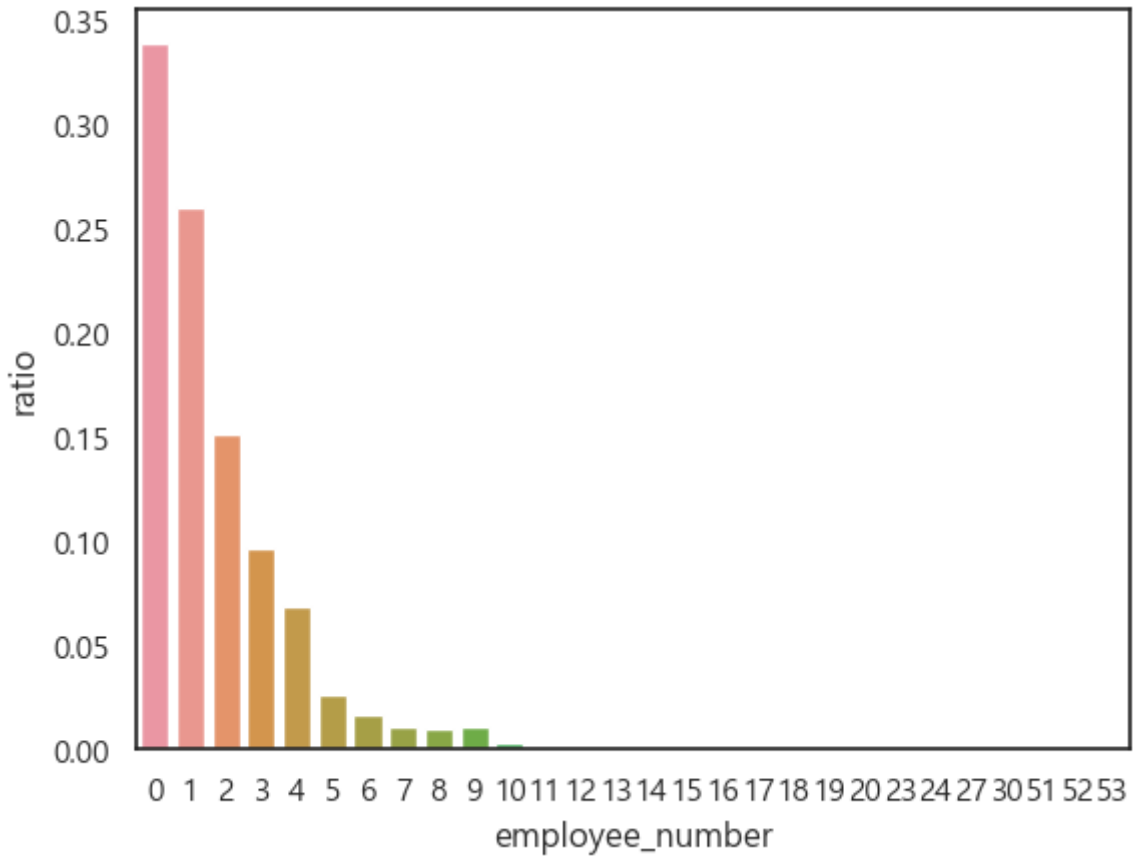


In [ ]:
```
survey_2["고용근로자수"]=survey_2["일반_합계종사자수"]-survey_2["일반_대표자종사자
survey_2["고용근로자수"]
```

```
C:\Users\ASUS\AppData\Local\Temp\ipykernel_3236\3525959332.py:1: SettingWithCopyW
arning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stabl
e/user_guide/indexing.html#returning-a-view-versus-a-copy
  survey_2["고용근로자수"]=survey_2["일반_합계종사자수"]-survey_2["일반_대표자종
사자수"]
```

Out[ ]:
```
0       0
1       0
2       0
3       0
4       0
       ..
5715    5
5716    0
5717    2
5718    3
5719    0
Name: 고용근로자수, Length: 5720, dtype: int64
```

In [ ]:
```
sns.barplot(x='고용근로자수', y='고용근로자수', data=survey_2, estimator=lambda
plt.xlabel('employee_number')
plt.ylabel('ratio')
plt.show()
```



In [ ]:
```
#사업 분류, 사업장 정보
survey_3=survey[survey_cols[24:33]]
```

```
survey_3
```

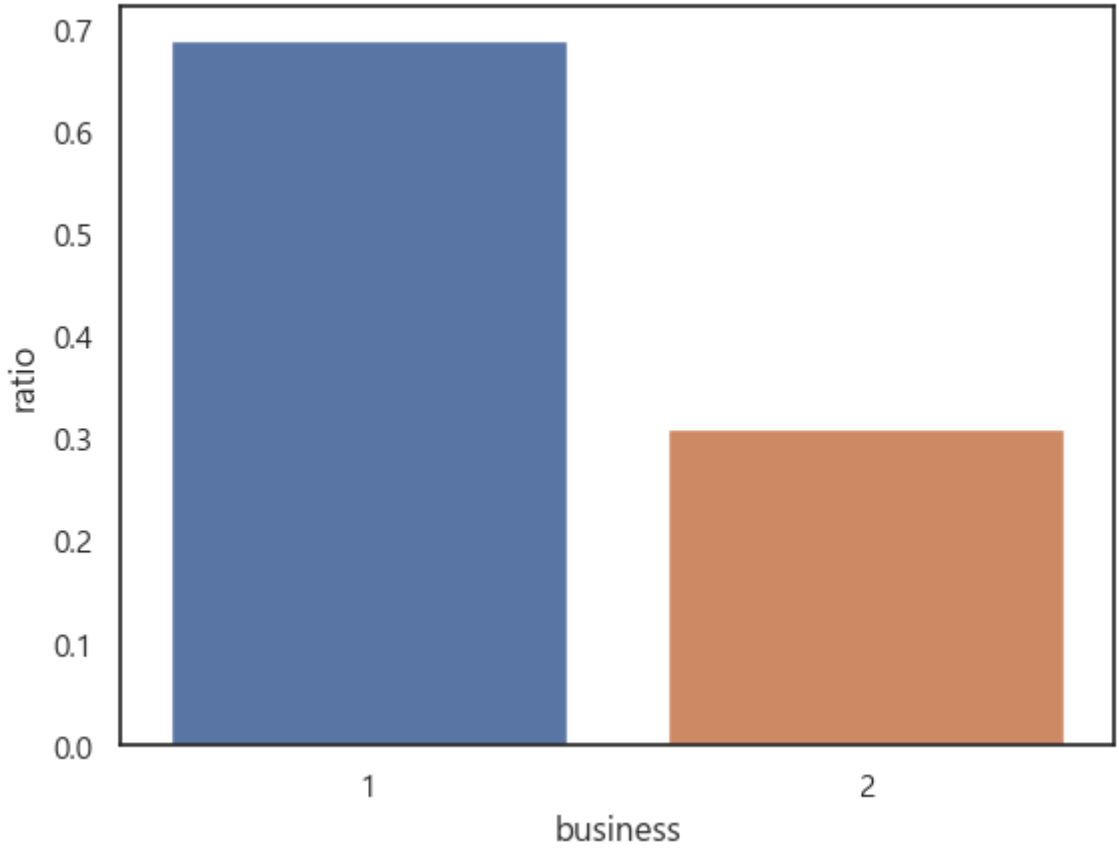
Out[ ]:

	일반_사업자형태코드	산업중분류코드	일반_프랜차이즈가맹점여부	일반_창업형태코드	일반_창업인수승계_연도	일반_창업인수승계_월	일반_사업장이전_경험여부	일반_사업장이전_사유코드	일반_현사업체_직전종사상지위항목코드
0	1	10	2	1	2020	3	2	0	3
1	1	10	2	3	2017	1	1	5	4
2	1	25	2	1	2017	7	2	0	3
3	1	28	2	1	2019	8	2	0	4
4	1	29	2	1	2019	7	2	0	4
...	...	...	...	...	...	...	...	...	...
5715	1	10	2	2	2018	9	1	4	2
5716	1	68	2	2	2015	2	1	3	2
5717	1	96	2	2	2017	4	2	0	1
5718	1	56	2	1	2020	10	1	4	2
5719	2	63	2	1	2021	6	2	0	2

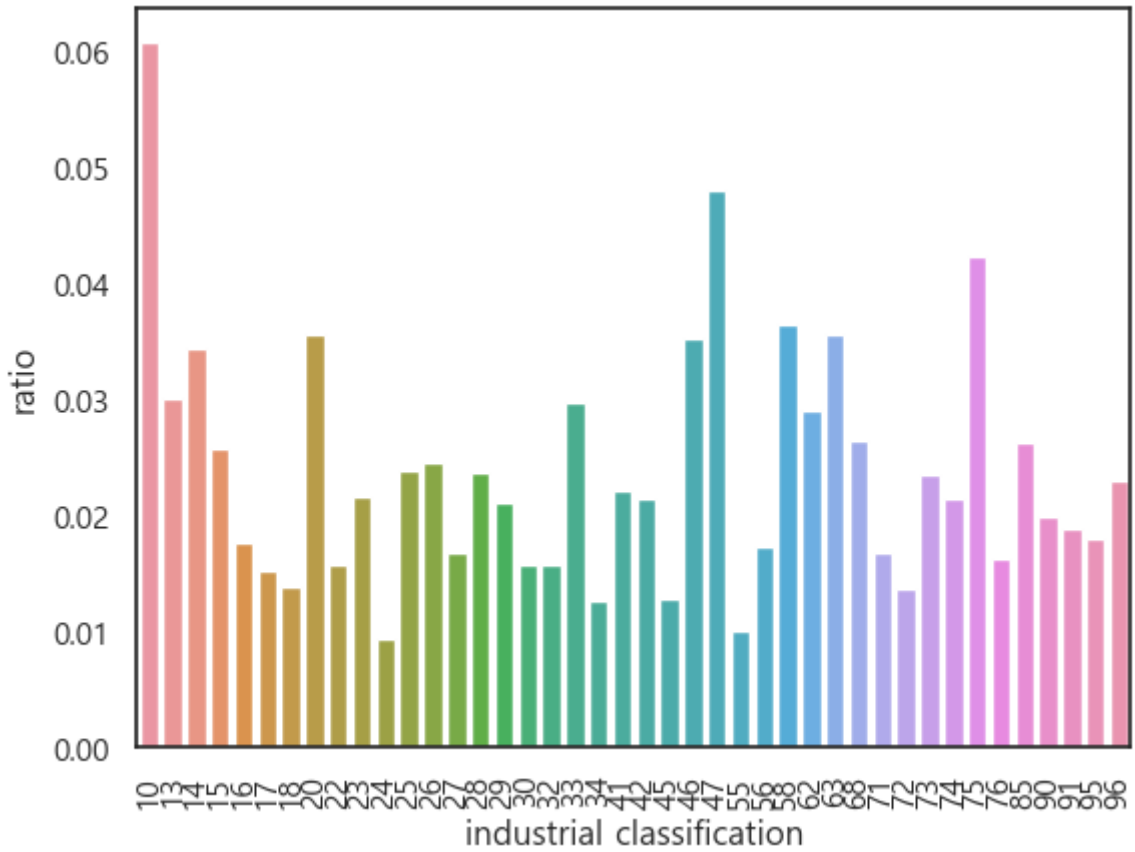
5720 rows × 9 columns

In [ ]:

```
sns.barplot(x='일반_사업자형태코드', y='일반_사업자형태코드', data=survey_3, esti
plt.xlabel('business')
plt.ylabel('ratio')
plt.show()
```



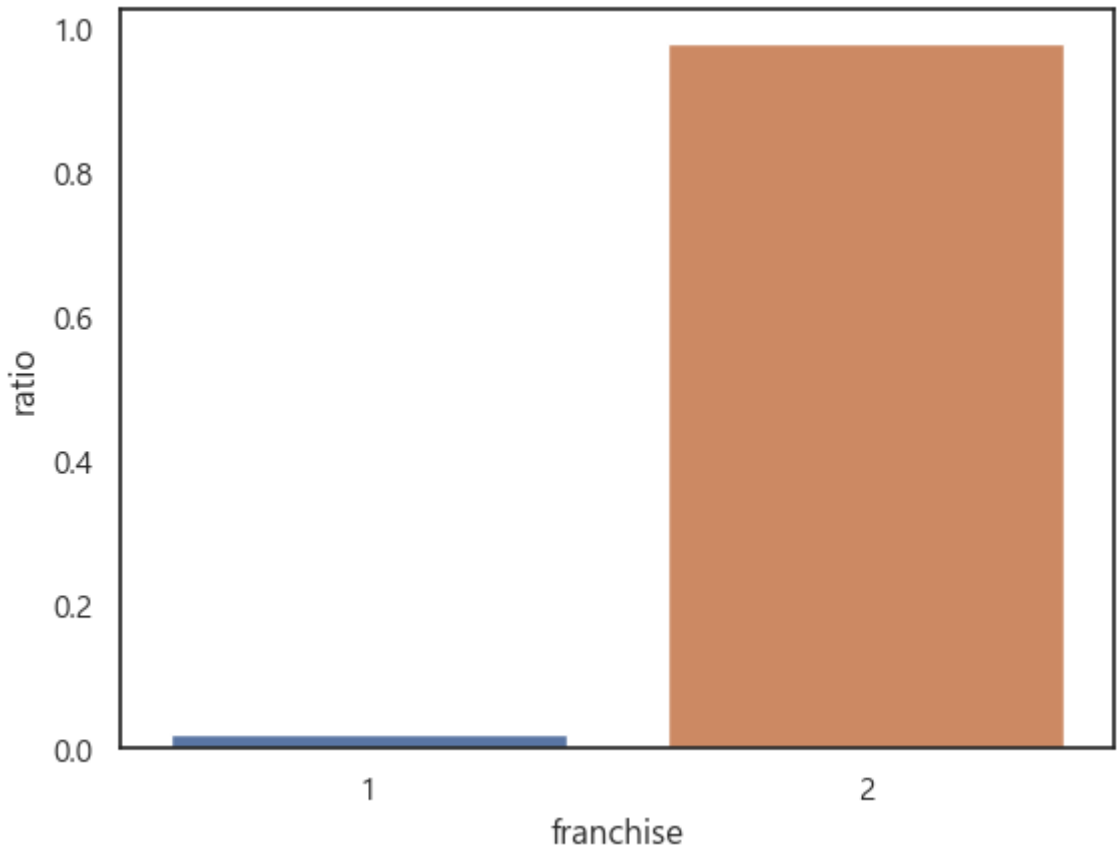
In [ ]:
```
sns.barplot(x='산업중분류코드', y='산업중분류코드', data=survey_3, estimator=lam
plt.xlabel('industrial classification')
plt.ylabel('ratio')
plt.xticks(rotation=90)
plt.show()
```



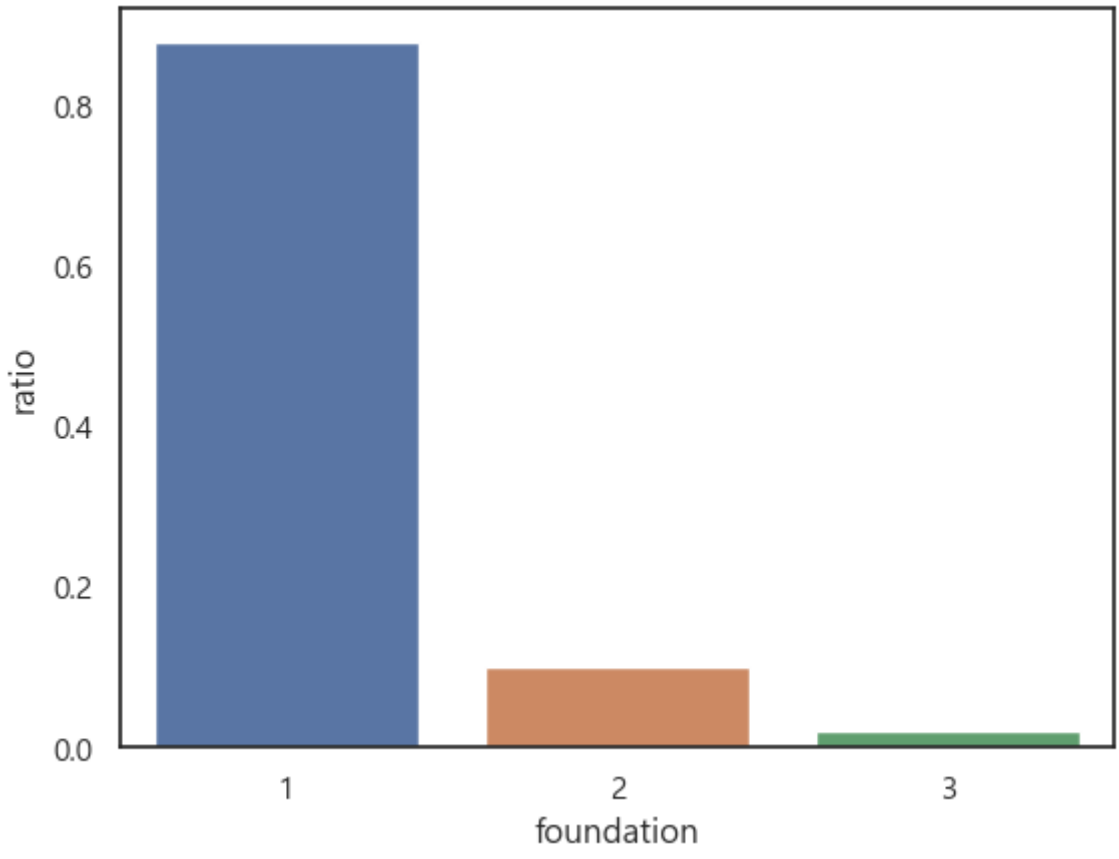
In [ ]:
```
survey_3.groupby("산업중분류코드").size().sort_values(ascending=False)
# 10: 식료품 제조업, 47: 소매업; 자동차 제외, 75: 사업지원 서비스업
```

Out[ ]:
```
산업중분류코드
10    348
47    275
75    242
58    209
20    204
63    204
46    202
14    197
13    172
33    170
62    166
68    152
85    151
15    148
26    141
25    137
28    136
73    135
96    132
41    127
23    124
42    123
74    123
29    121
90    114
91    108
95    103
16    101
56     99
27     96
71     96
76     93
22     90
30     90
32     90
17     87
18     79
72     78
45     73
34     72
55     58
24     54
dtype: int64
```

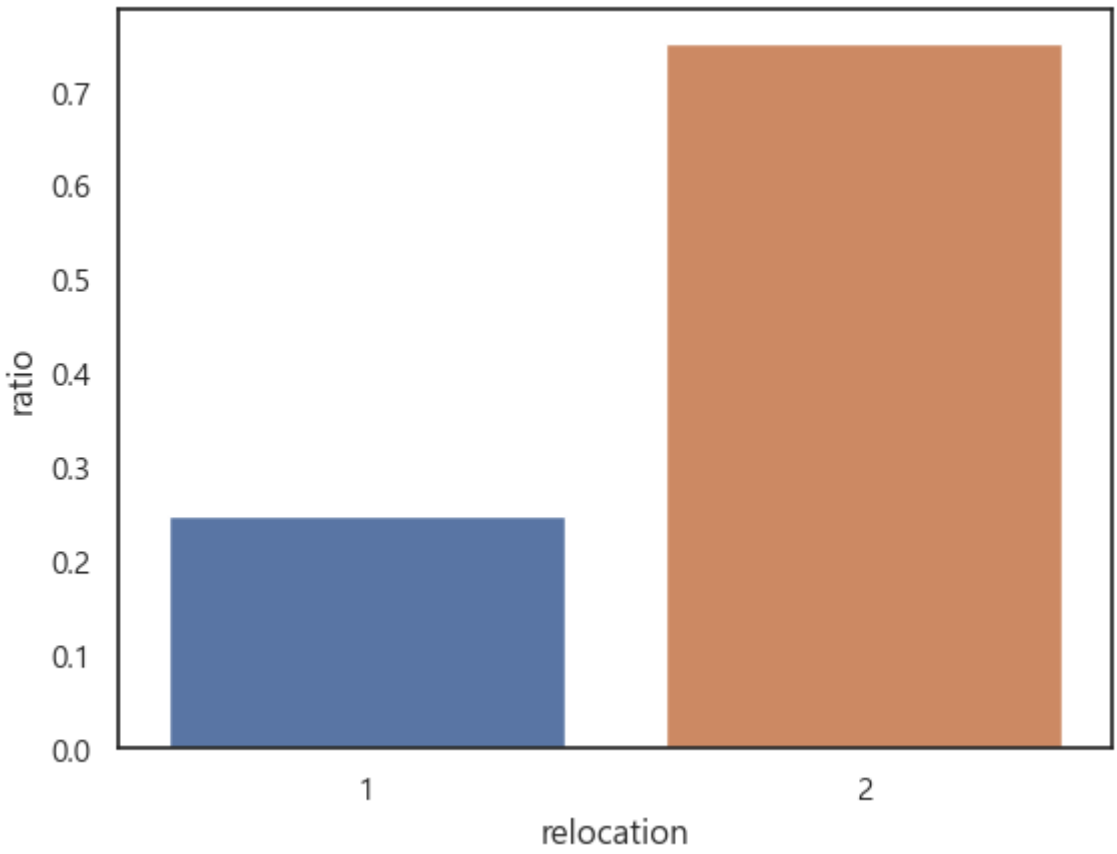
In [ ]:
```
sns.barplot(x='일반_프랜차이즈가맹점여부', y='일반_프랜차이즈가맹점여부', data=su
plt.xlabel('franchise')
plt.ylabel('ratio')
plt.show()
# 1: o, 2: X
```



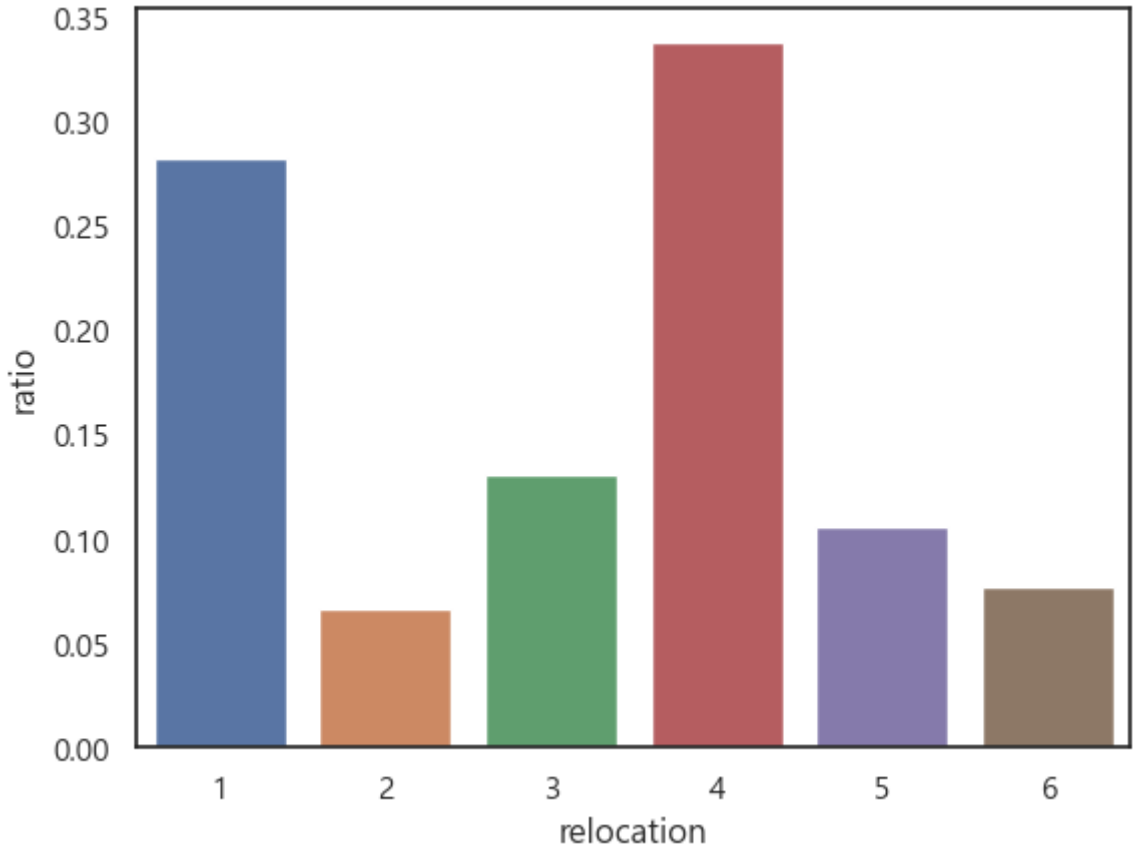
In [ ]:
```
sns.barplot(x='일반_창업형태코드', y='일반_창업형태코드', data=survey_3, estimato
plt.xlabel('foundation')
plt.ylabel('ratio')
plt.show()
# 1: 신규창업, 2: 인수창업, 3: 가업승계
```



In [ ]:
```
sns.barplot(x='일반_사업장이전_경험여부', y='일반_사업장이전_경험여부', data=surv
plt.xlabel('relocation')
plt.ylabel('ratio')
plt.show()
# 1: O, 2: X
```



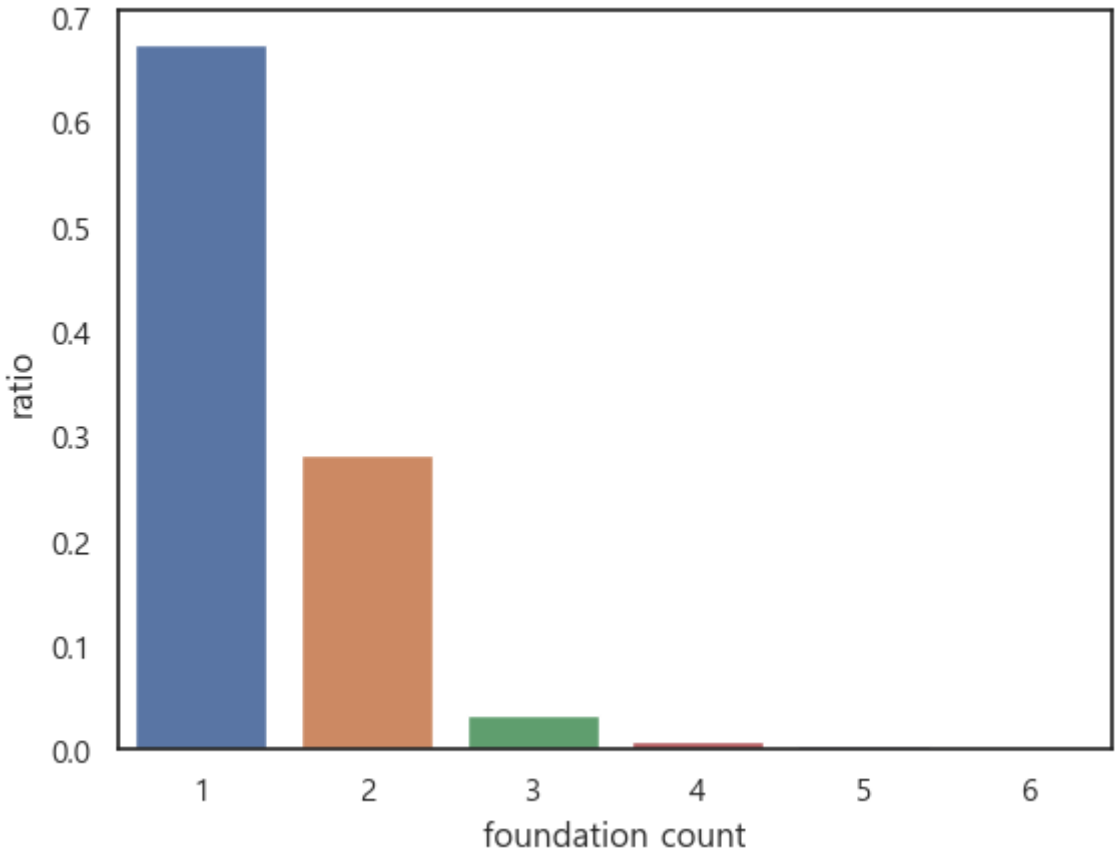
In [ ]:
```
survey_3_remove=survey_3[survey_3['일반_사업장이전_사유코드']!=0]
sns.barplot(x='일반_사업장이전_사유코드', y='일반_사업장이전_사유코드', data=surv
plt.xlabel('relocation')
plt.ylabel('ratio')
plt.show()
# 1: 사업 확장, 2: 상권쇠퇴, 3: 임대인의 계약 연장 거절, 4: 임대료 부담, 5: 철거,
```



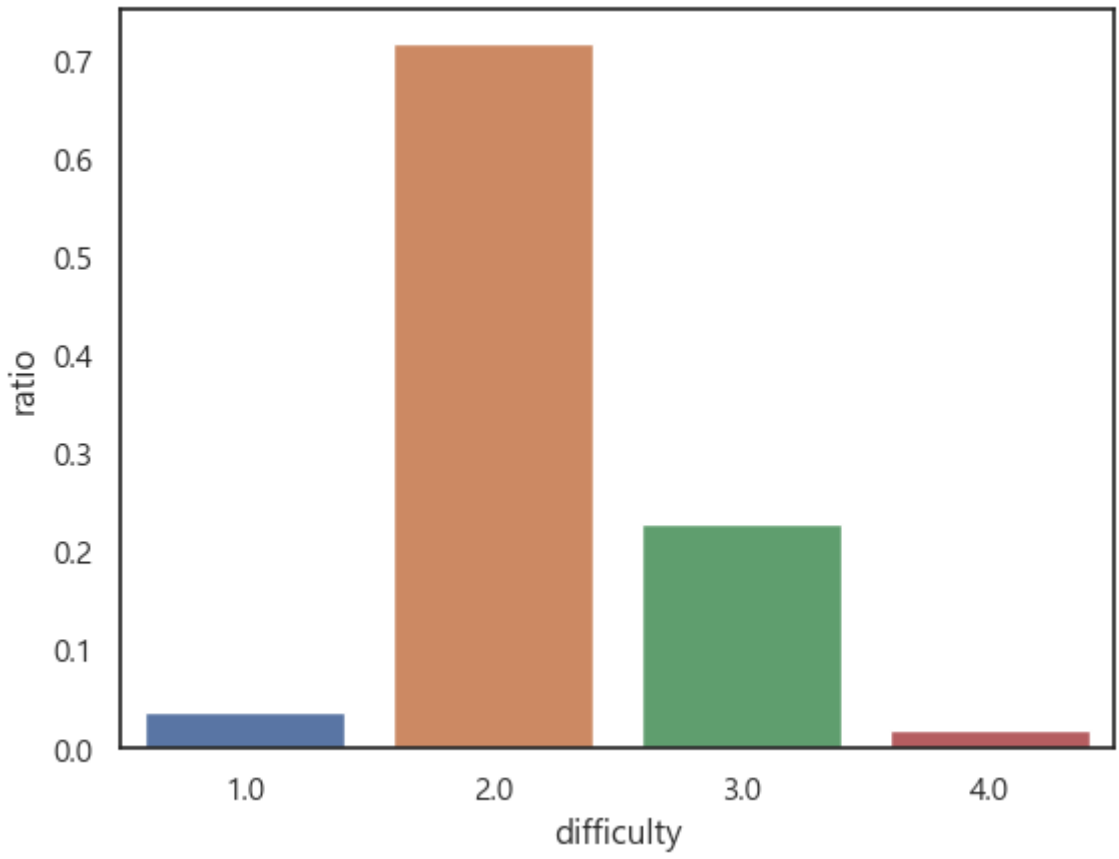
In [ ]:
```
# 창업 정보
survey_4=survey[survey_cols[33:54]]
survey_4.columns
```

Out[ ]:
```
Index(['일반_창업횟수', '창업_동기코드', '창업_준비기간_창업_준비기간_년수', '창
업_준비기간_창업_준비기간_월수',
       '창업_준비활동중요성_사업계획서작성코드', '창업_준비활동중요성_시장조사코
드', '창업_준비활동중요성_동종업종종사경험코드',
       '창업_준비활동중요성_창업교육코드', '창업_준비활동_사업계획서작성여부',
'창업_준비활동_시장조사여부',
       '창업_준비활동_동종업종종사경험여부', '창업_준비활동_창업교육여부', '창업
_어려움정도_입지선정코드',
       '창업_어려움정도_업종선택코드', '창업_어려움정도_자금조달코드', '창업_어
려움정도_기술확보코드',
       '창업_어려움정도_인력확보코드', '창업_어려움정도_행정절차코드', '창업_어
려움정도_경영방법코드', '창업_비용_총창업비용',
       '창업_비용_본인부담금액'],
      dtype='object')
```

In [ ]:
```
sns.barplot(x='일반_창업횟수', y='일반_창업횟수', data=survey_4, estimator=lambd
plt.xlabel('foundation count')
plt.ylabel('ratio')
plt.show()
```

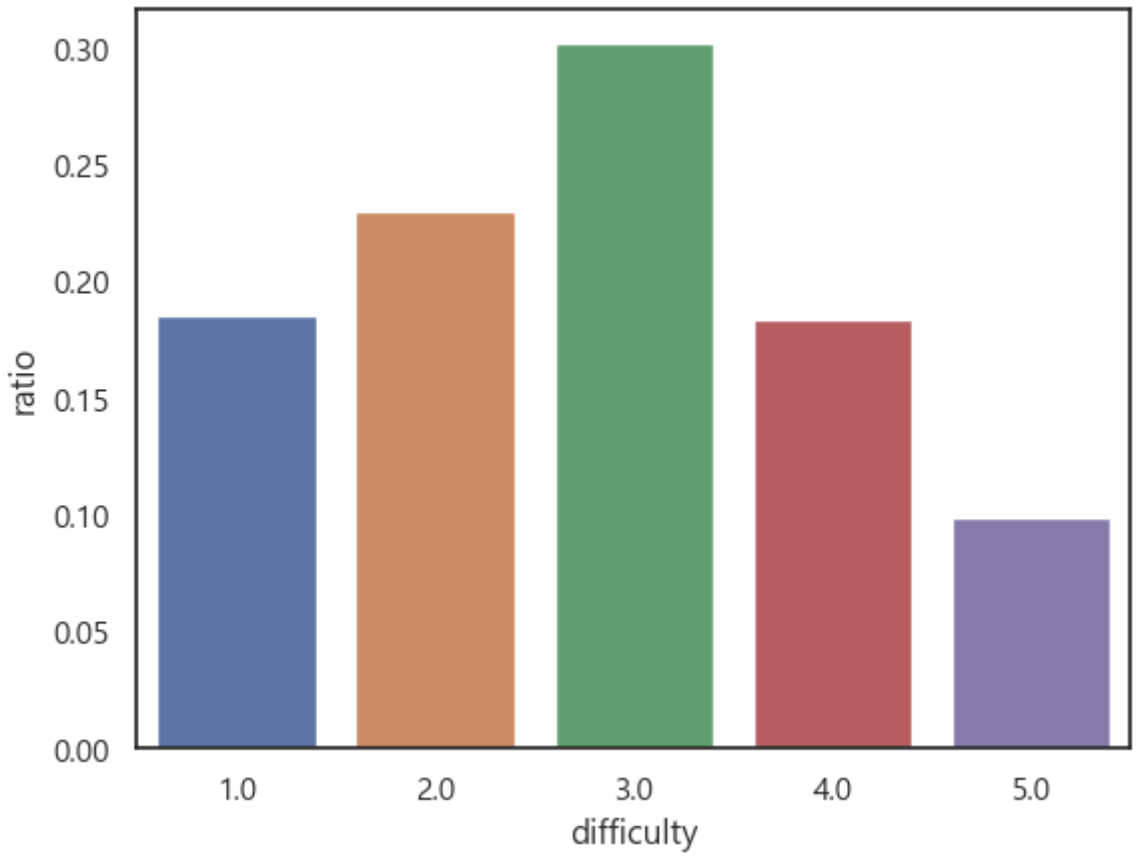


In [ ]:
```
survey_4_remove = survey_4[
    (survey_4['창업_동기코드'] == 1) |
    (survey_4['창업_동기코드'] == 2) |
    (survey_4['창업_동기코드'] == 3) |
    (survey_4['창업_동기코드'] == 4) |
    (survey_4['창업_동기코드'] == 5)]

sns.barplot(x='창업_동기코드', y='창업_동기코드', data=survey_4_remove, estimato
plt.xlabel('difficulty')
plt.ylabel('ratio')
plt.show()
```

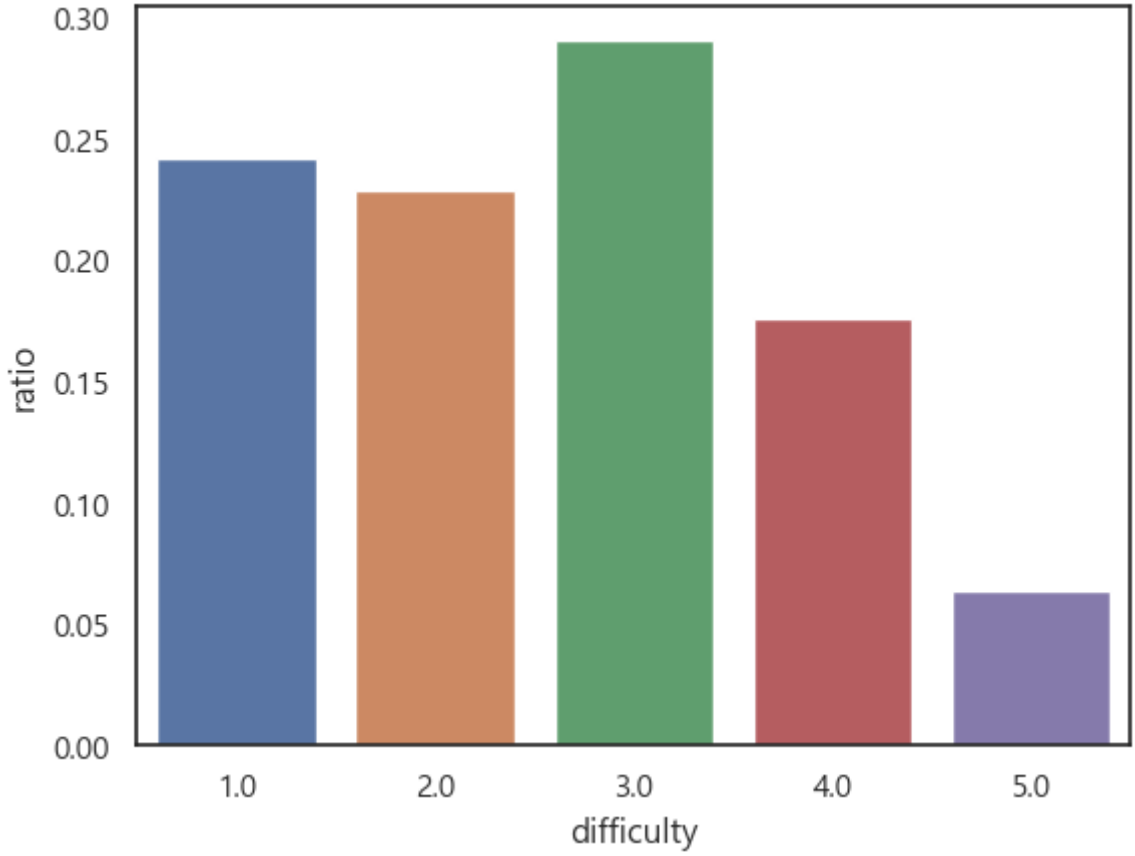


In [ ]:

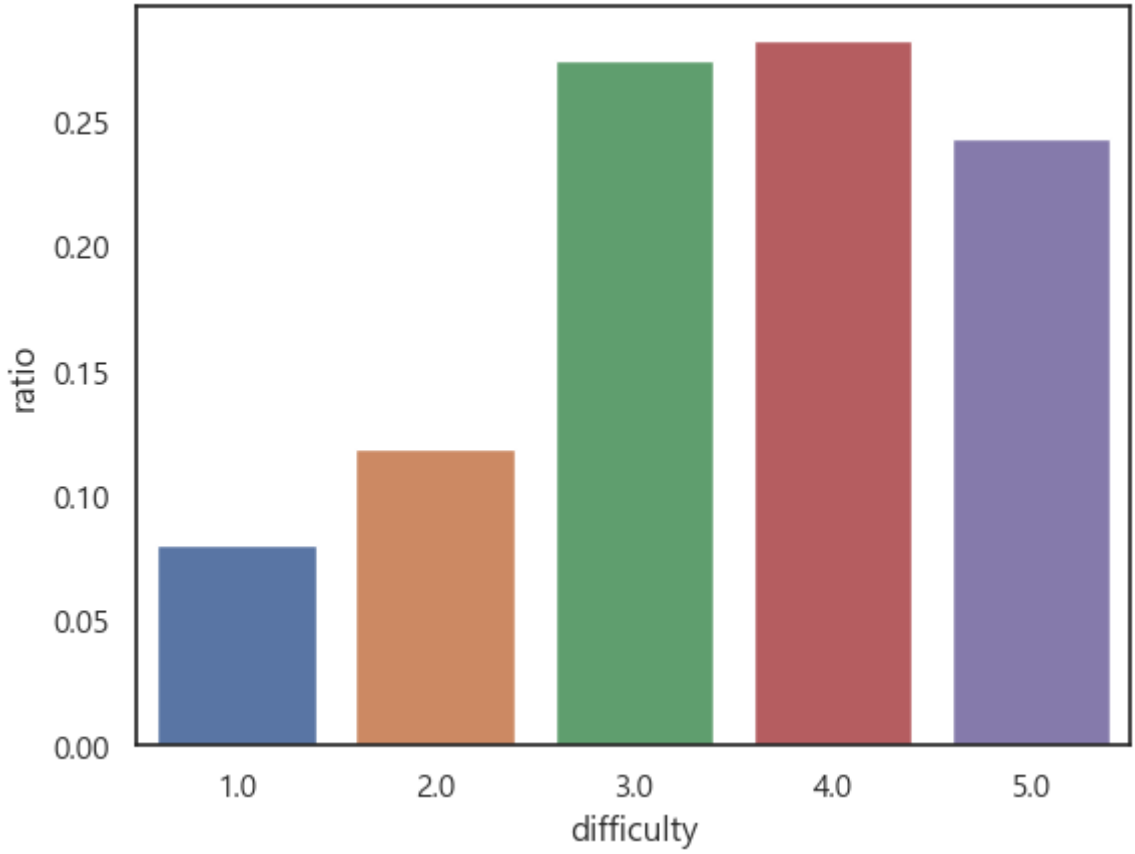
In [ ]:
```
survey_4_remove = survey_4[
    (survey_4['창업_어려움정도_입지선정코드'] == 1) |
    (survey_4['창업_어려움정도_입지선정코드'] == 2) |
    (survey_4['창업_어려움정도_입지선정코드'] == 3) |
    (survey_4['창업_어려움정도_입지선정코드'] == 4) |
    (survey_4['창업_어려움정도_입지선정코드'] == 5)]

sns.barplot(x='창업_어려움정도_입지선정코드', y='창업_어려움정도_입지선정코드', d
plt.xlabel('difficulty')
plt.ylabel('ratio')
plt.show()
```



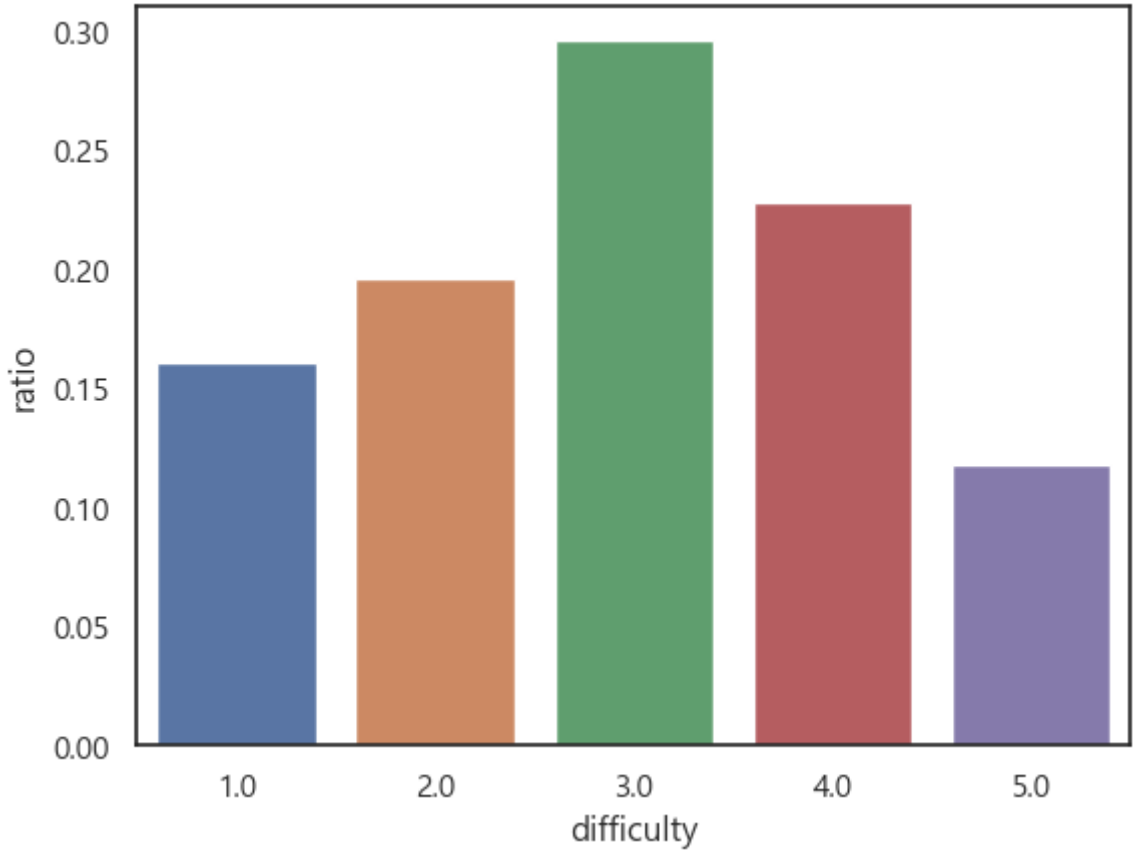
In [ ]:
```
survey_4_remove = survey_4[
    (survey_4['창업_어려움정도_업종선택코드'] == 1) |
    (survey_4['창업_어려움정도_업종선택코드'] == 2) |
    (survey_4['창업_어려움정도_업종선택코드'] == 3) |
    (survey_4['창업_어려움정도_업종선택코드'] == 4) |
    (survey_4['창업_어려움정도_업종선택코드'] == 5)]

sns.barplot(x='창업_어려움정도_업종선택코드', y='창업_어려움정도_업종선택코드', d
plt.xlabel('difficulty')
plt.ylabel('ratio')
plt.show()
```



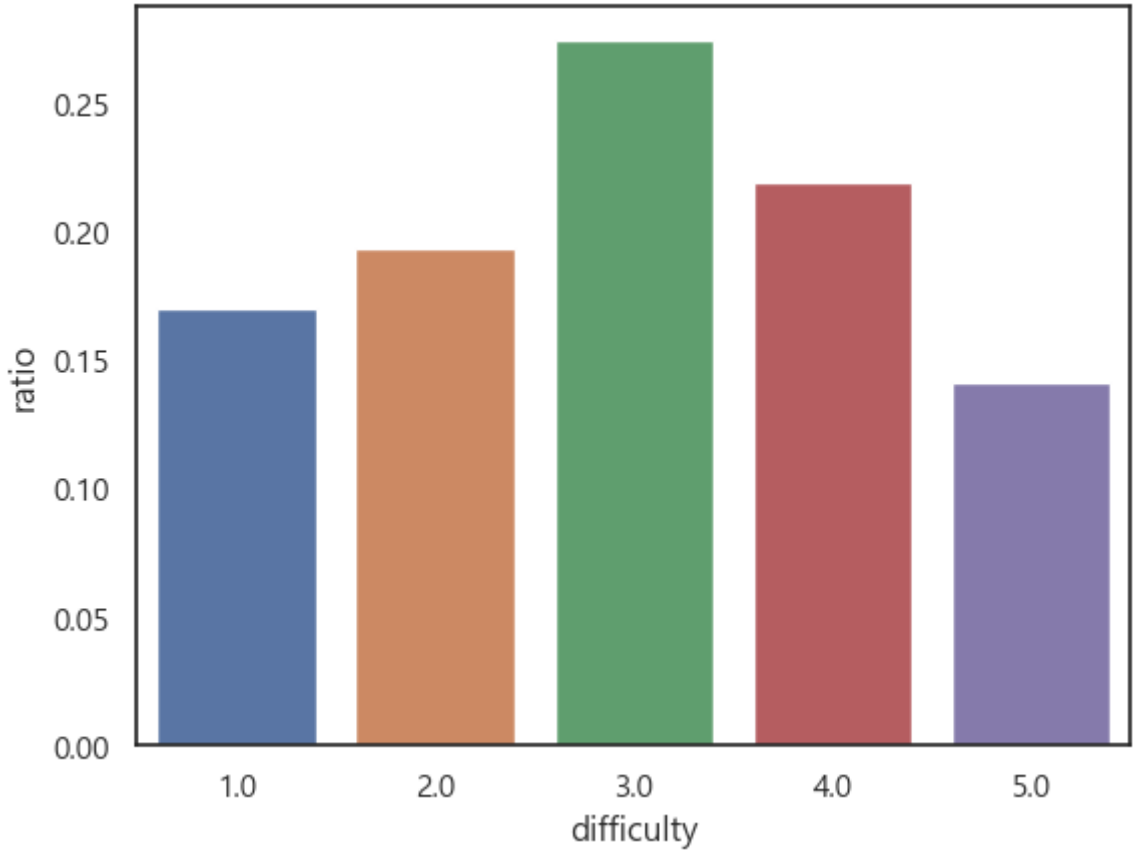
In [ ]:
```
survey_4_remove = survey_4[
    (survey_4['창업_어려움정도_자금조달코드'] == 1) |
    (survey_4['창업_어려움정도_자금조달코드'] == 2) |
    (survey_4['창업_어려움정도_자금조달코드'] == 3) |
    (survey_4['창업_어려움정도_자금조달코드'] == 4) |
    (survey_4['창업_어려움정도_자금조달코드'] == 5)]

sns.barplot(x='창업_어려움정도_자금조달코드', y='창업_어려움정도_자금조달코드', d
plt.xlabel('difficulty')
plt.ylabel('ratio')
plt.show()
```



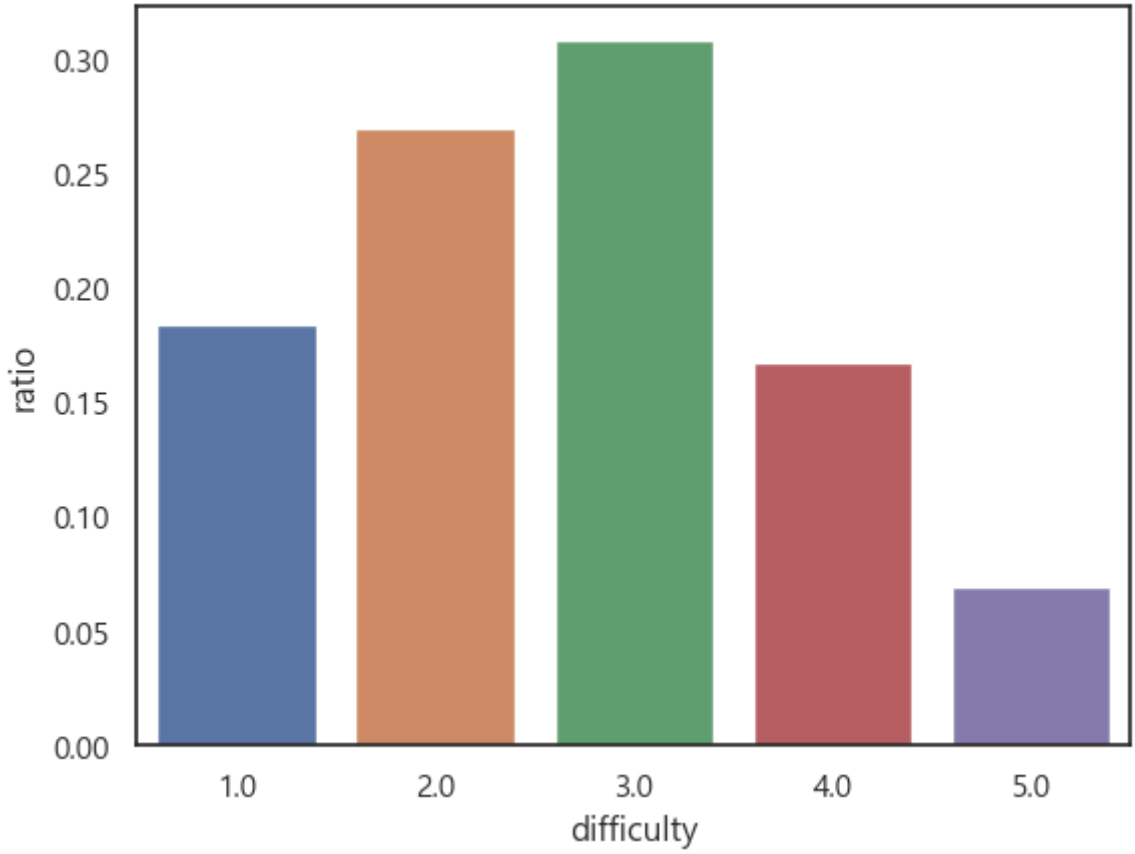
In [ ]:
```
survey_4_remove = survey_4[
    (survey_4['창업_어려움정도_기술확보코드'] == 1) |
    (survey_4['창업_어려움정도_기술확보코드'] == 2) |
    (survey_4['창업_어려움정도_기술확보코드'] == 3) |
    (survey_4['창업_어려움정도_기술확보코드'] == 4) |
    (survey_4['창업_어려움정도_기술확보코드'] == 5)]

sns.barplot(x='창업_어려움정도_기술확보코드', y='창업_어려움정도_기술확보코드', d
plt.xlabel('difficulty')
plt.ylabel('ratio')
plt.show()
```



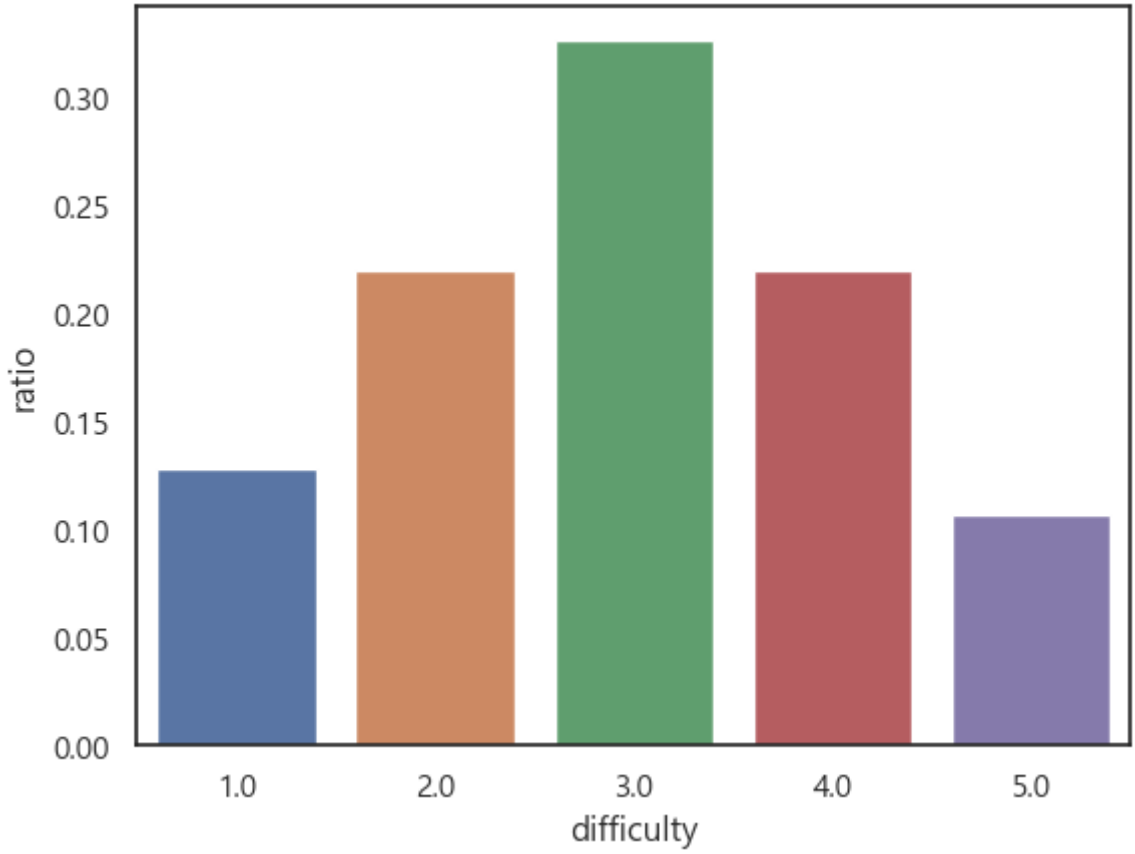
In [ ]:
```
survey_4_remove = survey_4[
    (survey_4['창업_어려움정도_인력확보코드'] == 1) |
    (survey_4['창업_어려움정도_인력확보코드'] == 2) |
    (survey_4['창업_어려움정도_인력확보코드'] == 3) |
    (survey_4['창업_어려움정도_인력확보코드'] == 4) |
    (survey_4['창업_어려움정도_인력확보코드'] == 5)]

sns.barplot(x='창업_어려움정도_인력확보코드', y='창업_어려움정도_인력확보코드', d
plt.xlabel('difficulty')
plt.ylabel('ratio')
plt.show()
```



In [ ]:
```
survey_4_remove = survey_4[
    (survey_4['창업_어려움정도_행정절차코드'] == 1) |
    (survey_4['창업_어려움정도_행정절차코드'] == 2) |
    (survey_4['창업_어려움정도_행정절차코드'] == 3) |
    (survey_4['창업_어려움정도_행정절차코드'] == 4) |
    (survey_4['창업_어려움정도_행정절차코드'] == 5)]

sns.barplot(x='창업_어려움정도_행정절차코드', y='창업_어려움정도_행정절차코드', d
plt.xlabel('difficulty')
plt.ylabel('ratio')
plt.show()
```



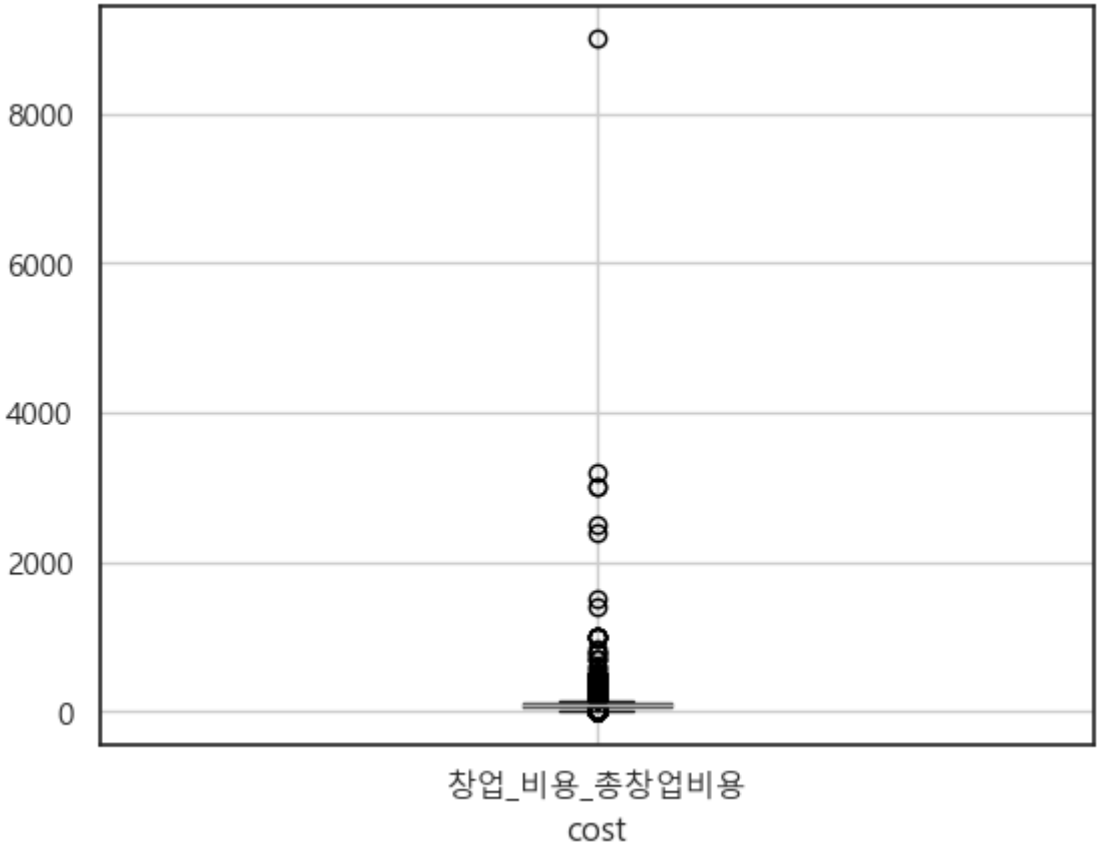
In [ ]:
```
survey_4_remove = survey_4[
    (survey_4['창업_어려움정도_경영방법코드'] == 1) |
    (survey_4['창업_어려움정도_경영방법코드'] == 2) |
    (survey_4['창업_어려움정도_경영방법코드'] == 3) |
    (survey_4['창업_어려움정도_경영방법코드'] == 4) |
    (survey_4['창업_어려움정도_경영방법코드'] == 5)]

sns.barplot(x='창업_어려움정도_경영방법코드', y='창업_어려움정도_경영방법코드', d
plt.xlabel('difficulty')
plt.ylabel('ratio')
plt.show()
```



In [ ]:
```
zscore_threshold = 1.96
df=survey_4['창업_비용_총창업비용']
df[(np.abs(stats.zscore(df)) > zscore_threshold)].values.ravel()
# outlier 없음
```

Out[ ]:
```
array([], dtype=float64)
```

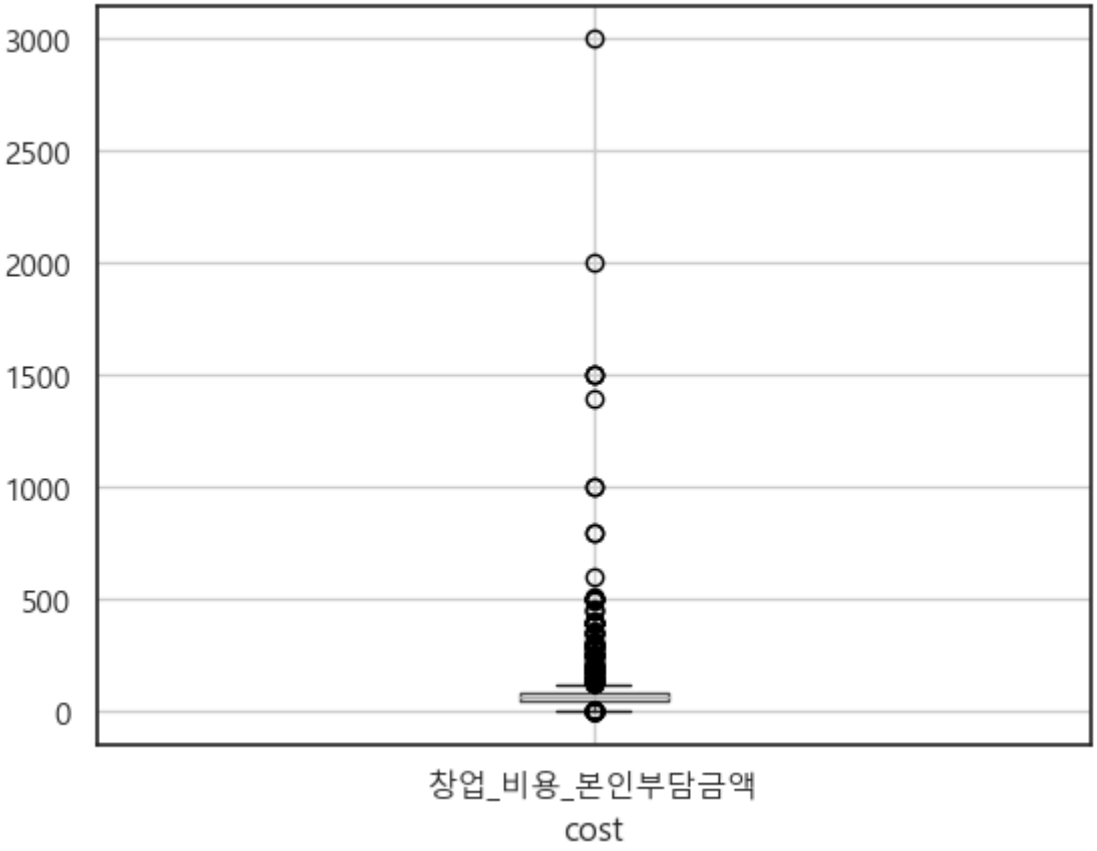
In [ ]:
```
plt.figure()
boxplot = survey_4.boxplot(column=["창업_비용_총창업비용"])
plt.xlabel('cost')
plt.show()
```



In [ ]:
```
zscore_threshold = 1.96
df=survey_4['창업_비용_본인부담금액']
df[(np.abs(stats.zscore(df)) > zscore_threshold)].values.ravel()
# outlier 없음
```

Out[ ]:
```
array([], dtype=float64)
```

In [ ]:
```
plt.figure()
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
boxplot = survey_4.boxplot(column=["창업_비용_본인부담금액"])
plt.xlabel('cost')
plt.show()
```



In [ ]:
```
df=survey_4[["창업_비용_총창업비용","창업_비용_본인부담금액"]]
df.describe()
```

Out[ ]:

	창업_비용_총창업비용	창업_비용_본인부담금액
count	5719.000000	5719.000000
mean	92.753989	71.468257
std	160.999964	78.767538
min	1.000000	0.000000
25%	70.000000	50.000000
50%	80.859386	66.394718
75%	107.713190	81.239264
max	9000.000000	3000.000000

In [ ]:
```
# 경영 영업 기간
survey_5=survey[survey_cols[54:57]]
survey_5.columns
```

Out[ ]:
```
Index(['경영_영업기간_하루평균시간수', '경영_영업기간_월평균일수', '경영_영업기
간_년간월수'], dtype='object')
```

In [ ]:
```
survey_5.describe()
```

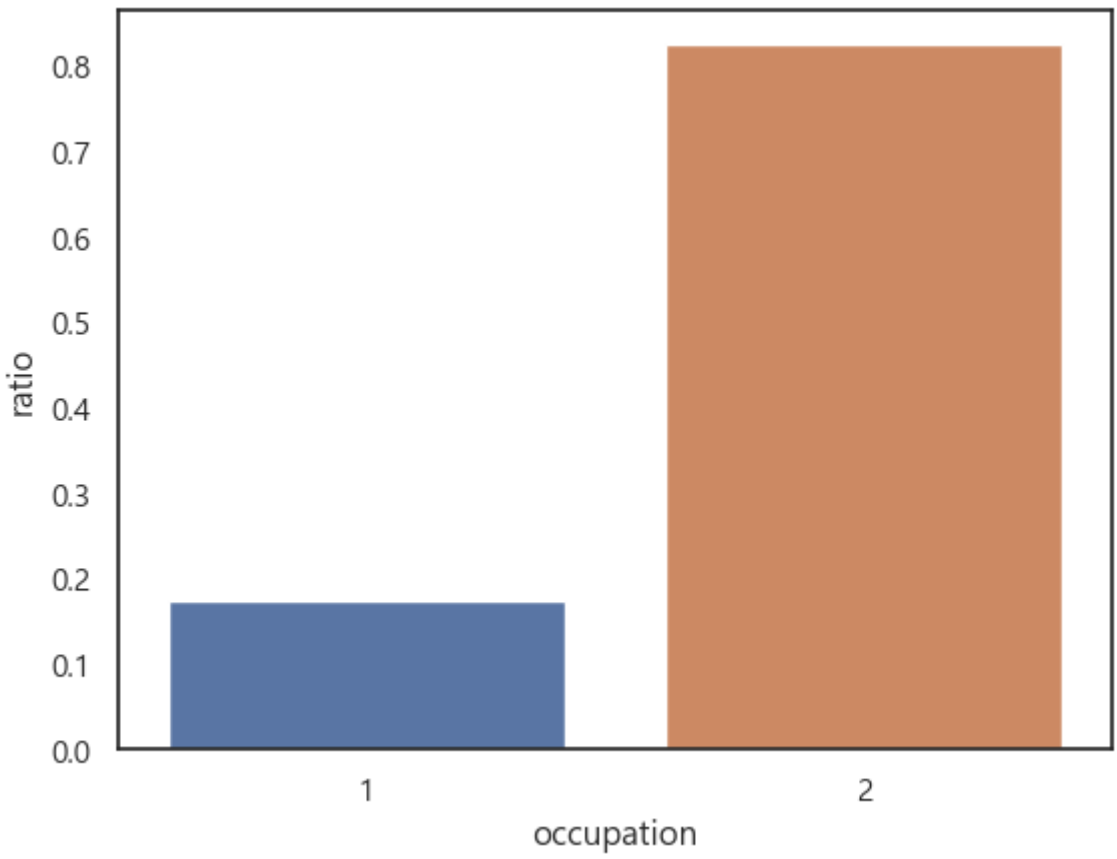
Out[ ]:

	경영_영업기간_하루평균시간수	경영_영업기간_월평균일수	경영_영업기간_년간월수
count	5720.000000	5720.000000	5720.000000
mean	9.183217	23.269755	11.771678
std	2.844024	3.394779	1.104269
min	1.000000	2.000000	1.000000
25%	8.000000	20.000000	12.000000
50%	8.000000	22.000000	12.000000
75%	10.000000	25.000000	12.000000
max	24.000000	30.000000	12.000000

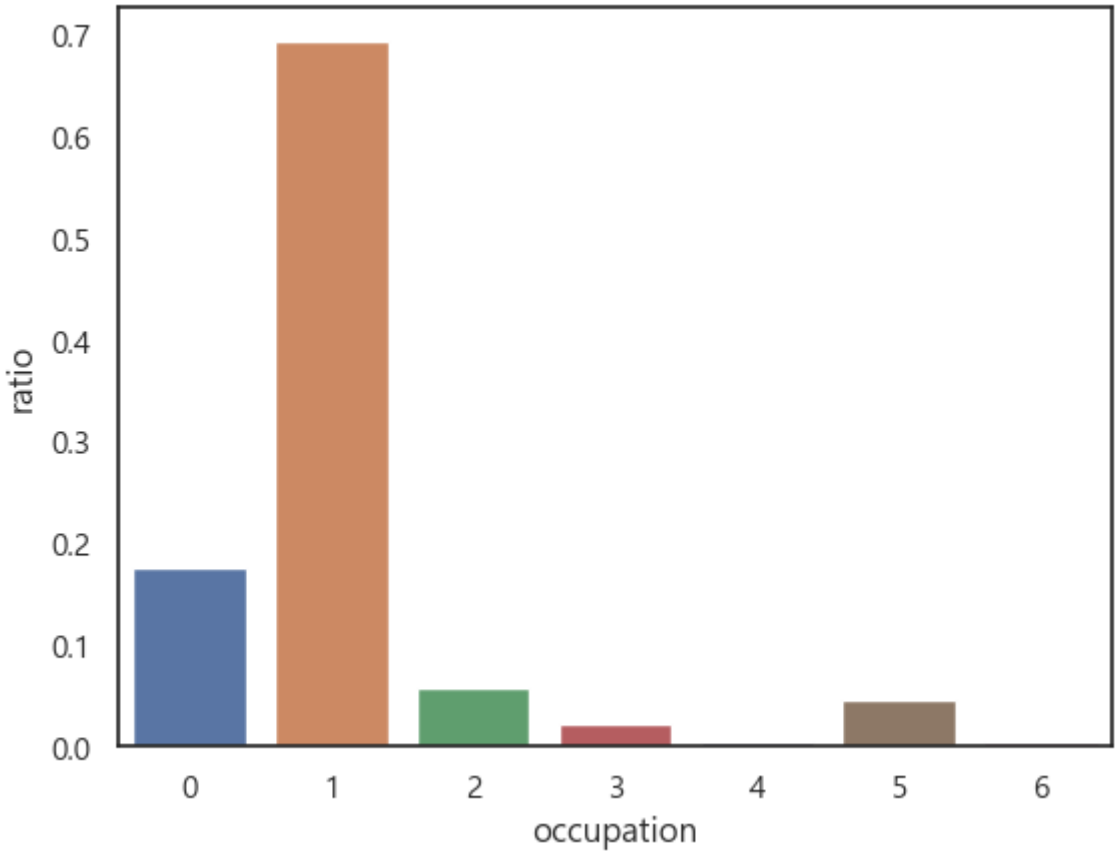
In [ ]:
```
# 경영 점유형태
survey_6=survey[survey_cols[63:68]]
survey_6.columns
```

Out[ ]:
```
Index(['경영_점유형태코드', '경영_점유형태_임차형태코드', '경영_점유형태_보증금액', '경영_점유형태_월세금액',
       '경영_점유형태_일정비율지급금액'],
      dtype='object')
```

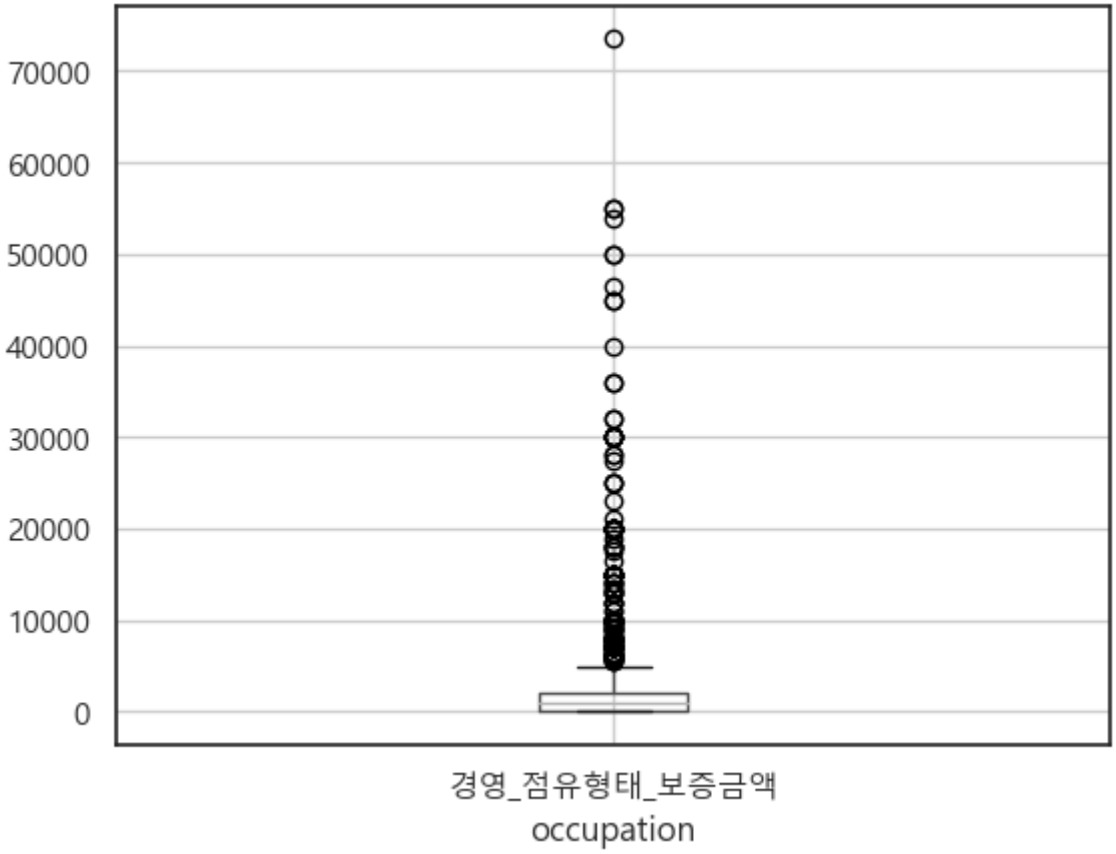
In [ ]:
```
sns.barplot(x='경영_점유형태코드', y='경영_점유형태코드', data=survey_6, estimato
plt.xlabel('occupation')
plt.ylabel('ratio')
plt.show()
# 1: 보증금 있는 월세, 2: 보증금 없는 월세
```



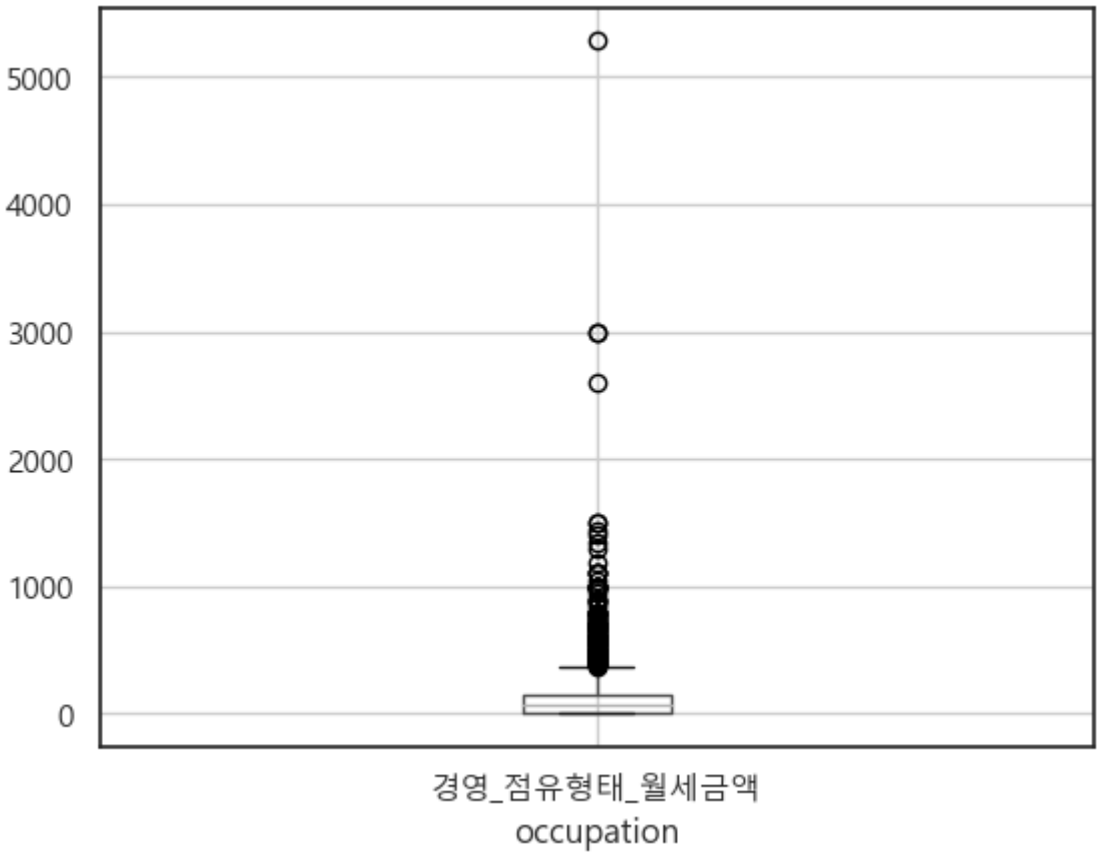
In [ ]:
```
sns.barplot(x='경영_점유형태_임차형태코드', y='경영_점유형태_임차형태코드', data=
plt.xlabel('occupation')
plt.ylabel('ratio')
plt.show()
# 0.소유 1.보증금 있는 월세 2.보증금 없는 월세 3.전세 4.매출대비 일정 비율 지급 5
```



In [ ]:
```
plt.figure()
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
boxplot = survey_6.boxplot(column=["경영_점유형태_보증금액"])
plt.xlabel('occupation')
plt.show()
```



In [ ]:
```
plt.figure()
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
boxplot = survey_6.boxplot(column=["경영_점유형태_월세금액"])
plt.xlabel('occupation')
plt.show()
```



In [ ]:
```
survey_6[["경영_점유형태_보증금액","경영_점유형태_월세금액"]].describe()
```

Out[ ]:

	경영_점유형태_보증금액	경영_점유형태_월세금액
count	5720.000000	5720.000000
mean	1791.913462	108.887238
std	3559.661244	169.449790
min	0.000000	0.000000
25%	0.000000	3.000000
50%	1000.000000	70.000000
75%	2000.000000	150.000000
max	73500.000000	5280.000000

In [ ]:
```
# 영업 비용
survey_7=survey[survey_cols[80:91]]
survey_7.columns
```

Out[ ]:
```
Index(['경영_매출금액', '경영_영업비용', '경영_영업비용_매출원가', '경영_영업비
용_급여총액', '경영_영업비용_임차료',
       '경영_영업비용_기타금액', '경영_영업이익', '경영_전자상거래_매출실적여
부', '경영_전자상거래_매출비율',
       '경영_부채여부', '경영_부채금액'],
      dtype='object')
```

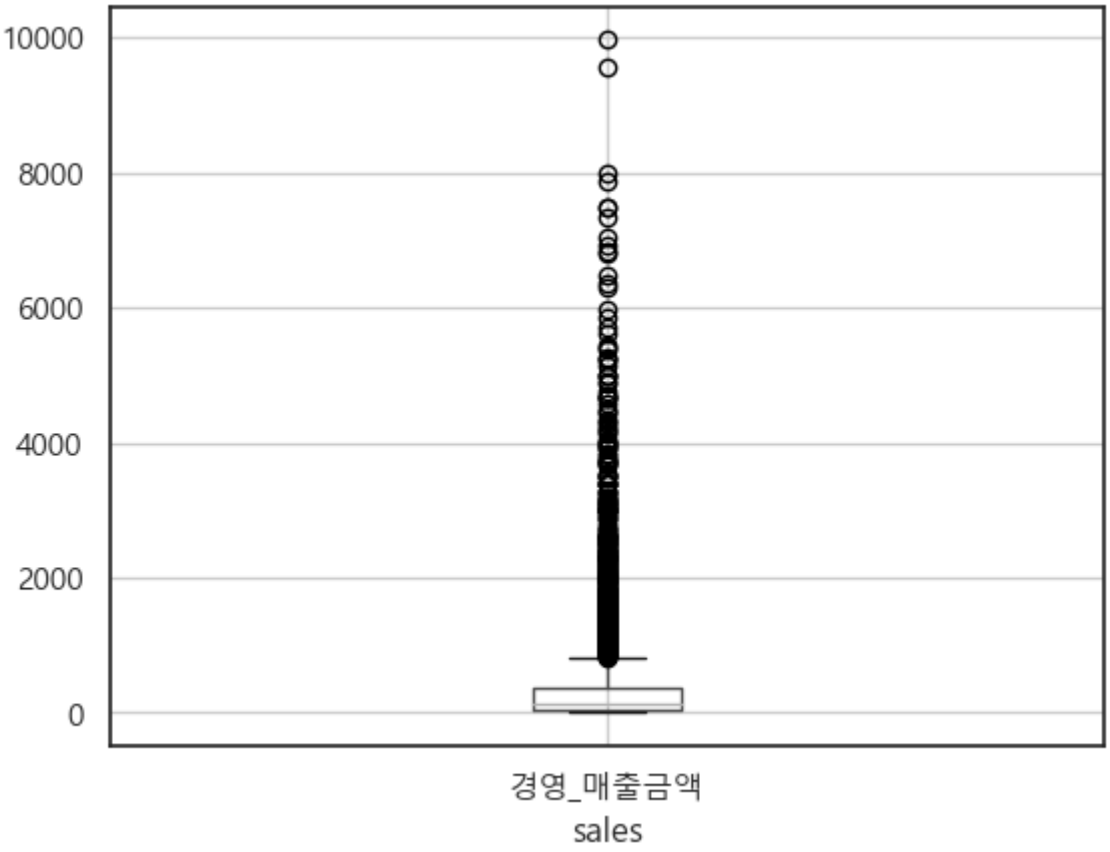
In [ ]:
```
survey_7[['경영_매출금액', '경영_영업비용', '경영_영업비용_매출원가', '경영_영업비
       '경영_영업비용_기타금액', '경영_영업이익', '경영_전자상거래_매출비율', '경
```

Out[ ]:

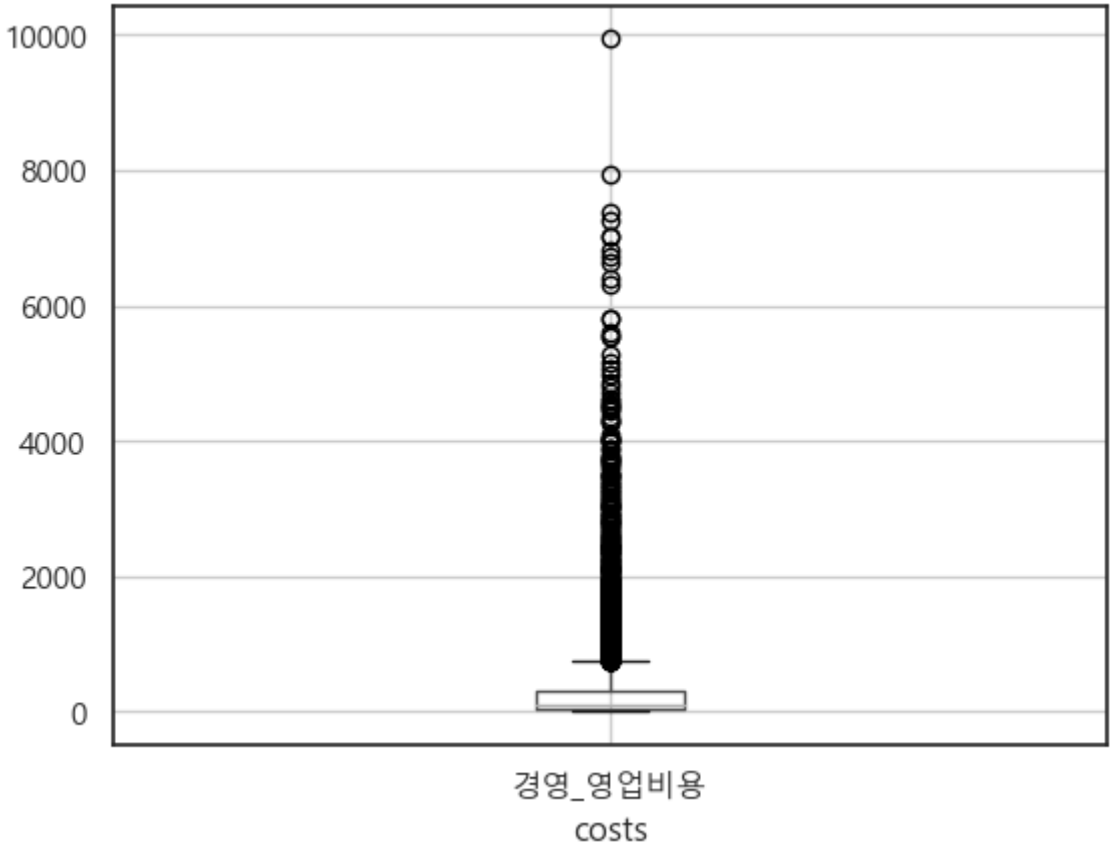
	경영_매출금액	경영_영업비용	경영_영업비용_매출원가	경영_영업비용_급여총액	경영_영업비용_임차료	경영_영업비용_기타금액
count	5720.000000	5720.000000	5720.000000	5720.000000	5720.000000	5720.000000
mean	365.378671	325.747727	165.386189	57.497203	17.230245	85.634091
std	716.020257	659.537908	449.070757	107.388822	44.198094	260.843311
min	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	50.000000	38.000000	2.000000	0.000000	3.000000	3.000000
50%	130.000000	111.000000	24.500000	24.000000	10.000000	12.000000
75%	360.000000	319.000000	120.000000	72.000000	20.000000	64.250000
max	9961.000000	9936.000000	6696.000000	2080.000000	1276.000000	4811.000000

In [ ]:
```
plt.figure()
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
boxplot = survey_7.boxplot(column=["경영_매출금액"])
```

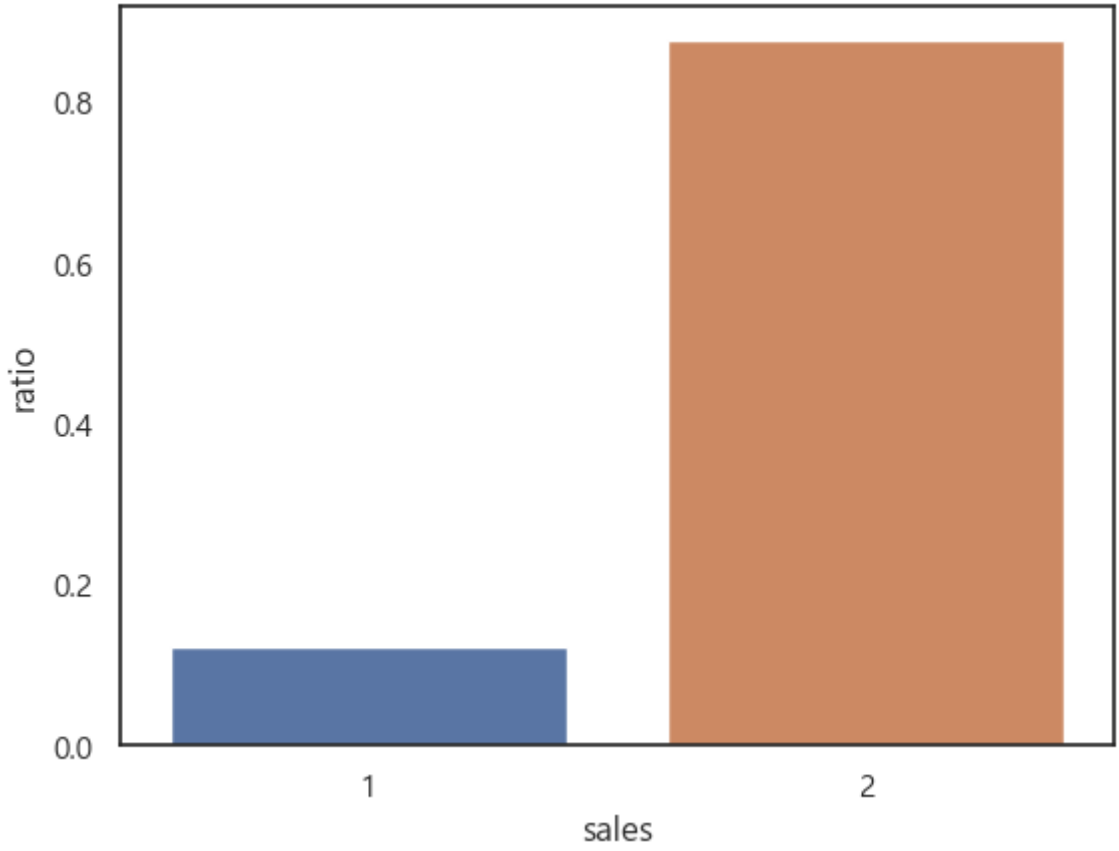
```
plt.xlabel('sales')
plt.show()
```



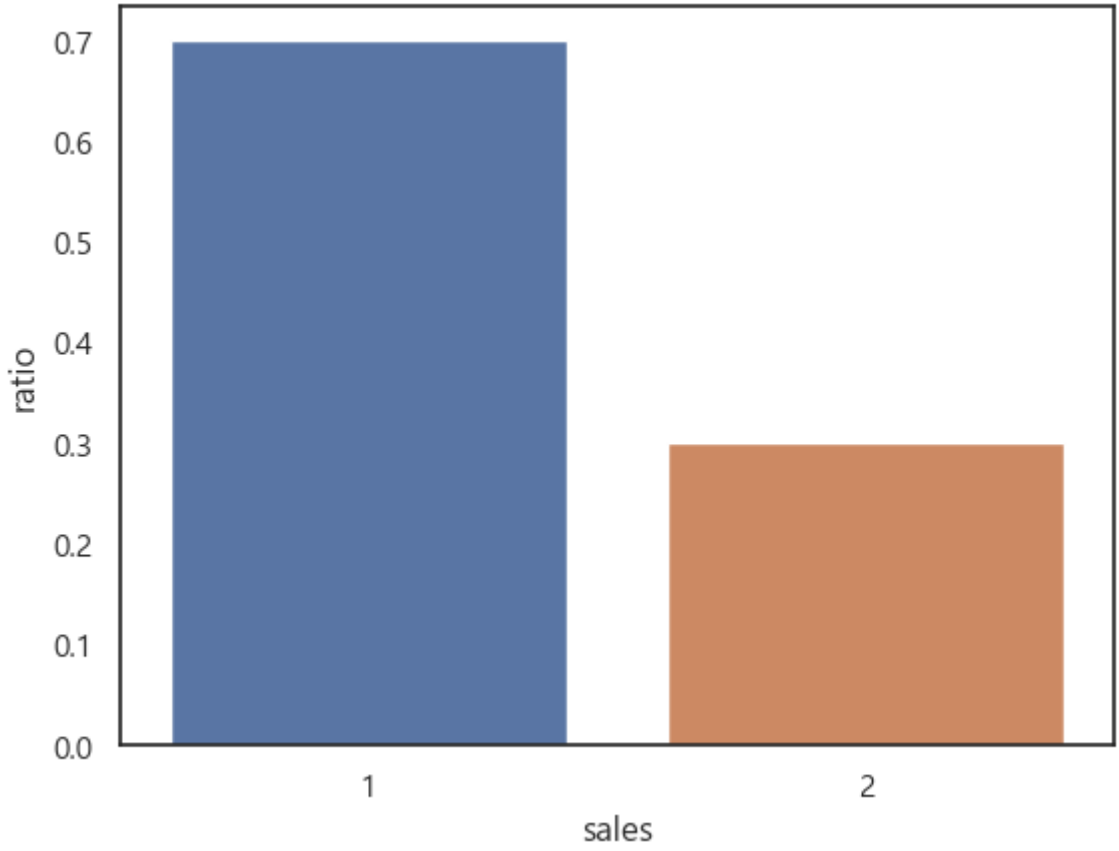
In [ ]:
```
plt.figure()
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
boxplot = survey_7.boxplot(column=["경영_영업비용"])
plt.xlabel('costs')
plt.show()
```



In [ ]:
```
sns.barplot(x='경영_전자상거래_매출실적여부', y='경영_전자상거래_매출실적여부', d
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
plt.xlabel('sales')
plt.ylabel('ratio')
plt.show()
# 1. Y 2. N
```



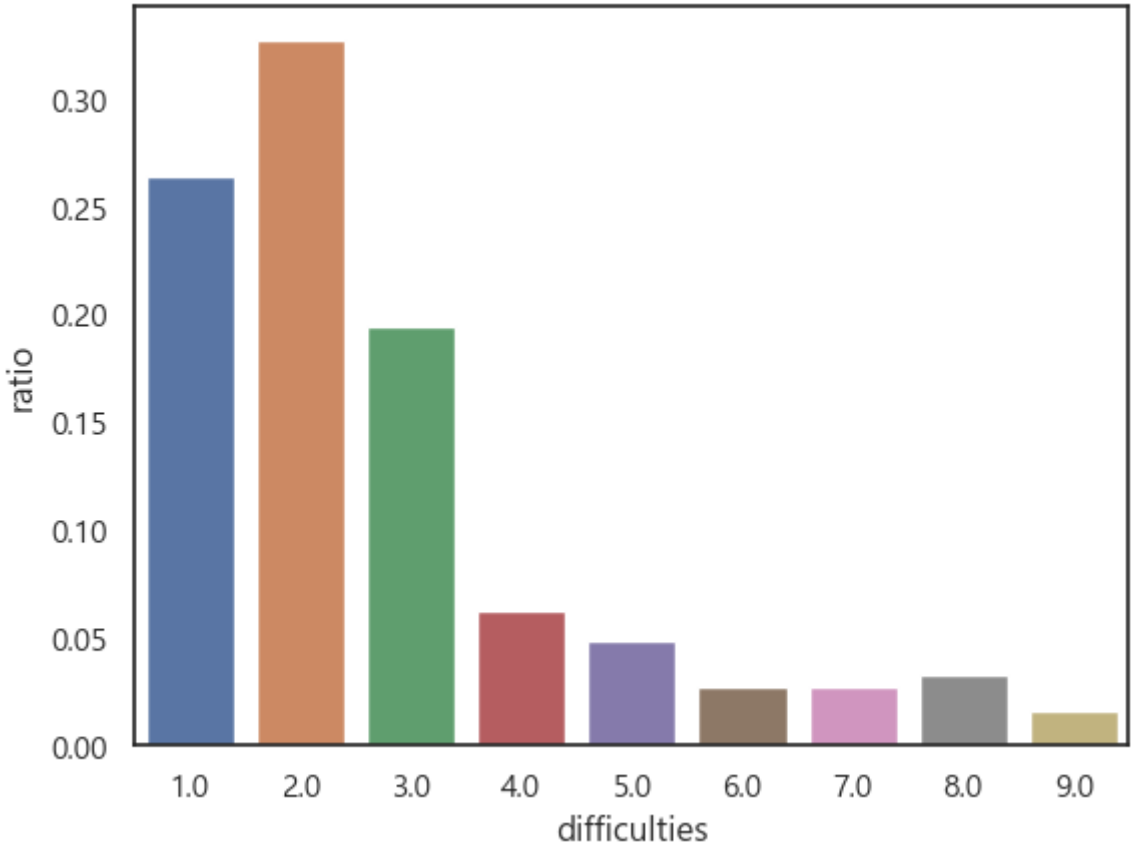
In [ ]:
```
sns.barplot(x='경영_부채여부', y='경영_부채여부', data=survey_7, estimator=lambd
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
plt.xlabel('sales')
plt.ylabel('ratio')
plt.show()
# 1. Y 2. N
```



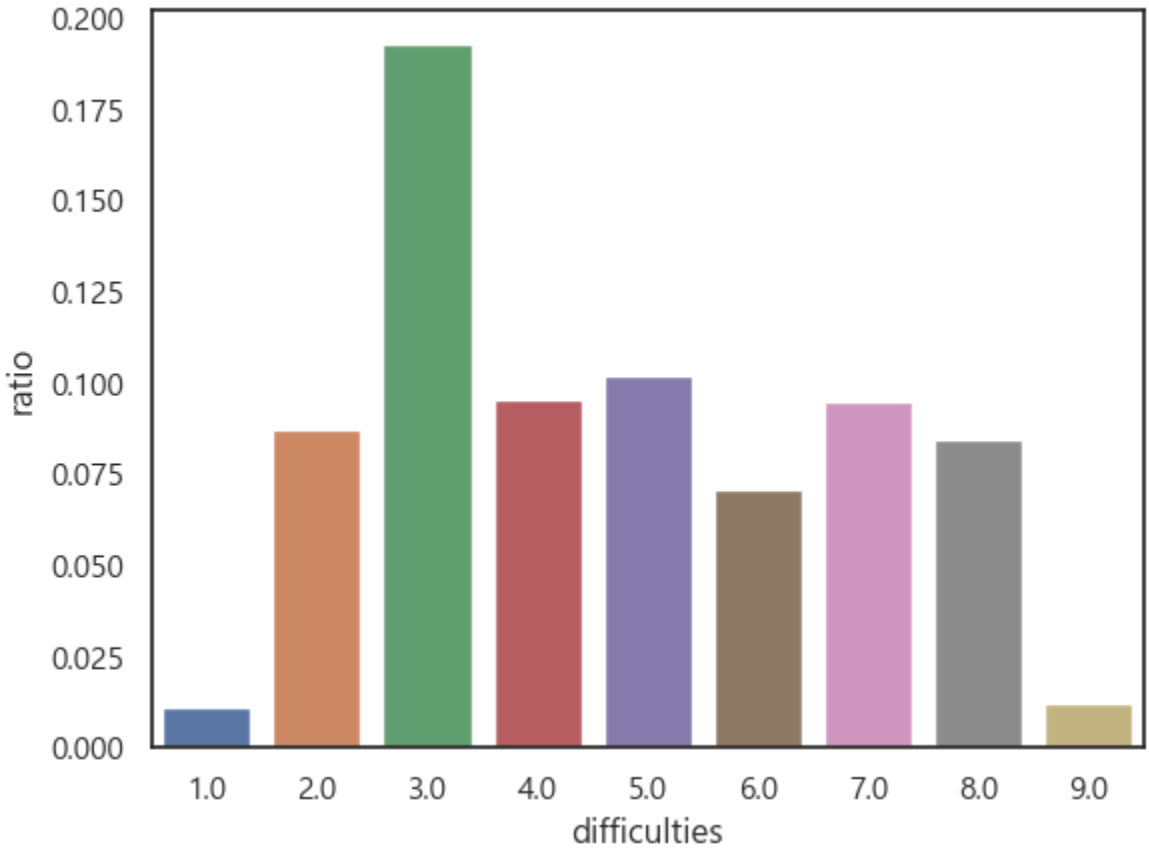
In [ ]:
```
# 경영 설문 조사
survey_8=survey[survey_cols[91:100]]
survey_8.columns
```

Out[ ]:
```
Index(['경영_애로사항1코드', '경영_애로사항2코드', '정부지원정책_지원경험코드',
'정부지원정책_추진정책1코드',
       '정부지원정책_추진정책2코드', '코로나19_지원정책1코드', '코로나19_지원정
책2코드', '사업전환_운영계획코드',
       '사업전환_운영계획_소유권승계계획여부'],
      dtype='object')
```

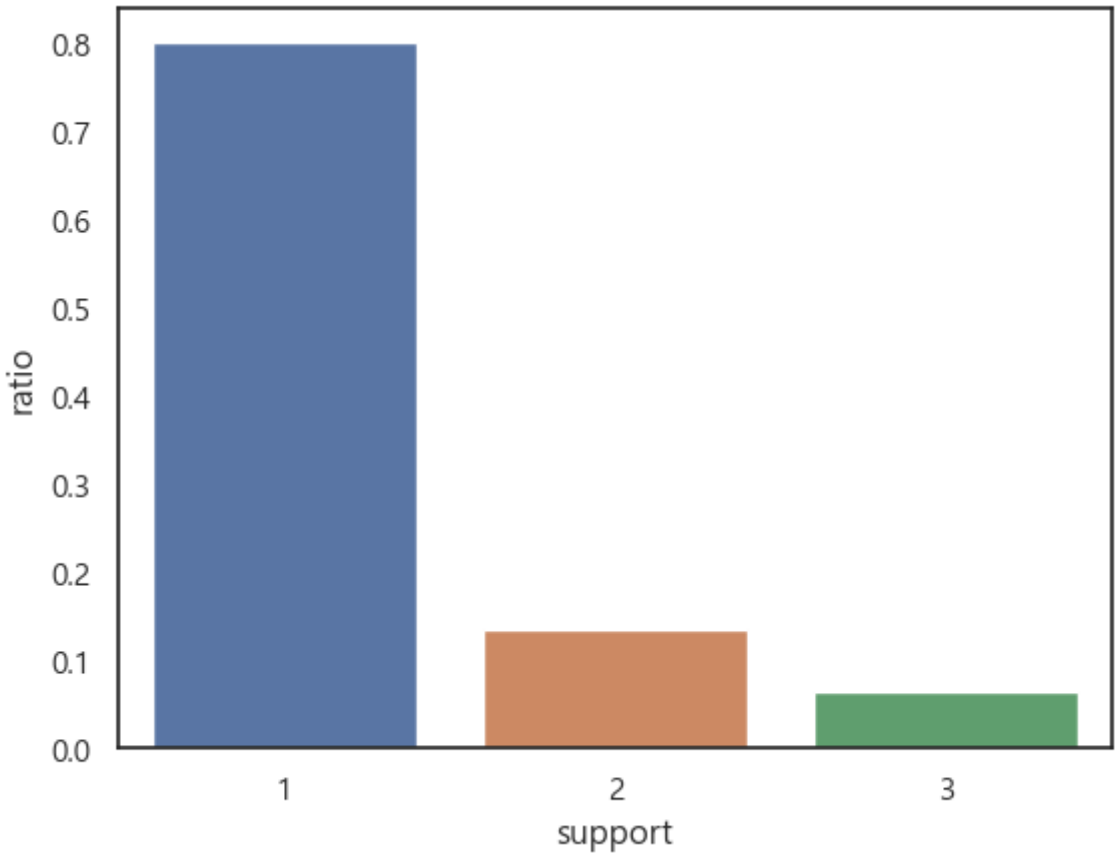
In [ ]:
```
sns.barplot(x='경영_애로사항1코드', y='경영_애로사항1코드', data=survey_8, estima
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
plt.xlabel('difficulties')
plt.ylabel('ratio')
plt.show()
```



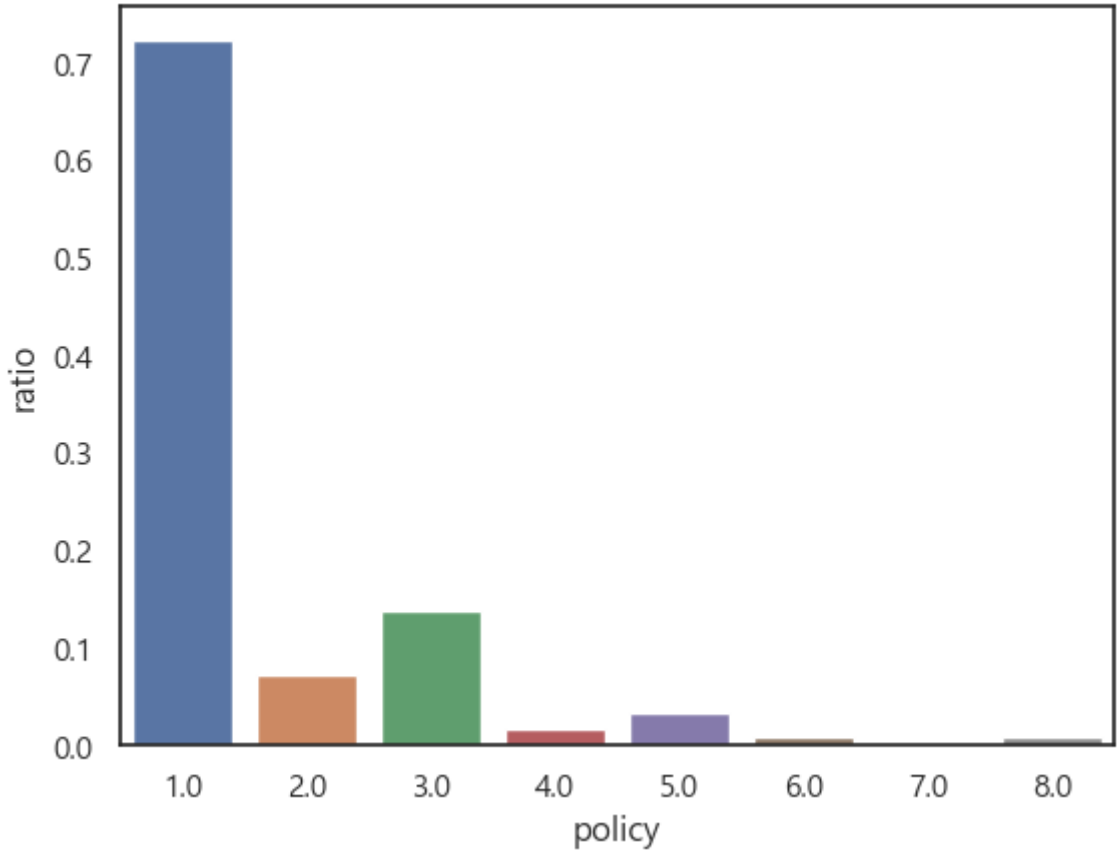
In [ ]:
```
sns.barplot(x='경영_애로사항2코드', y='경영_애로사항2코드', data=survey_8, estima
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
plt.xlabel('difficulties')
plt.ylabel('ratio')
plt.show()
```



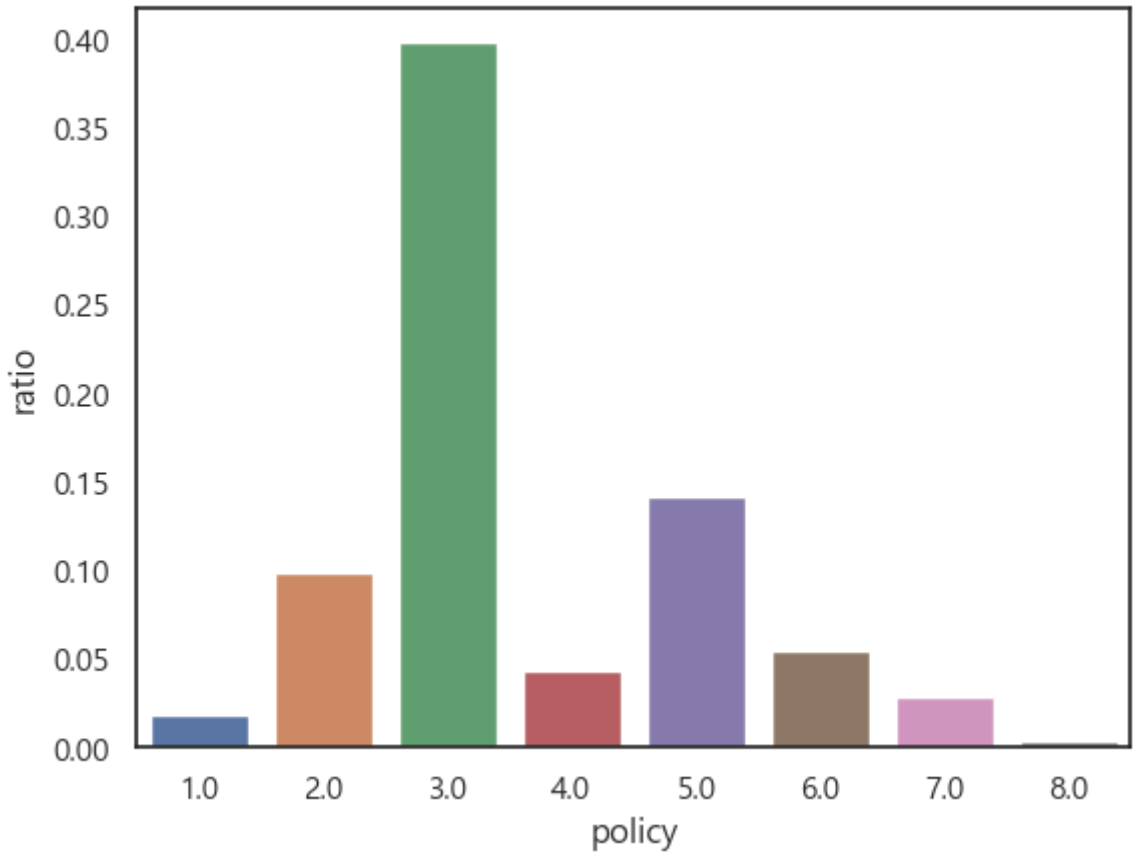
In [ ]:
```
sns.barplot(x='정부지원정책_지원경험코드', y='정부지원정책_지원경험코드', data=su
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
plt.xlabel('support')
plt.ylabel('ratio')
plt.show()
# 1: 지원 받음 2: 신청 X 3: 신청했으나 지원 X
```



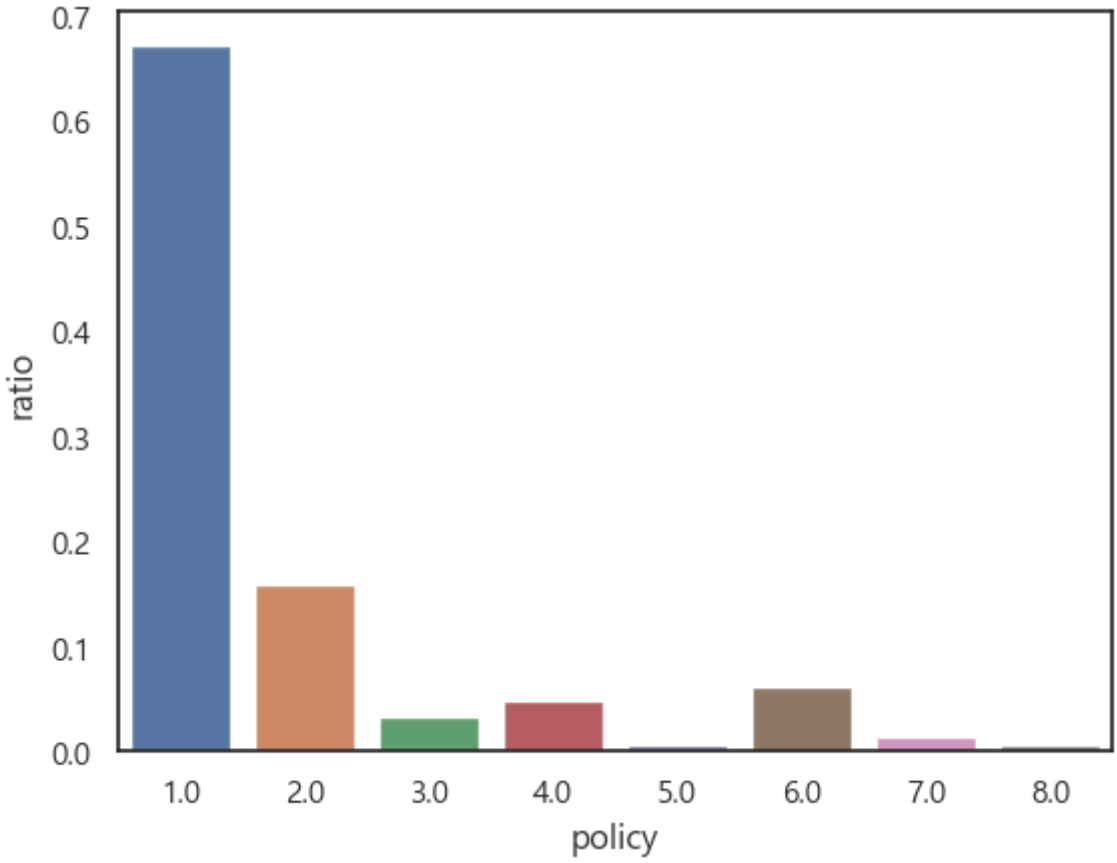
In [ ]:
```
sns.barplot(x='정부지원정책_추진정책1코드', y='정부지원정책_추진정책1코드', data=
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
plt.xlabel('policy')
plt.ylabel('ratio')
plt.show()
```



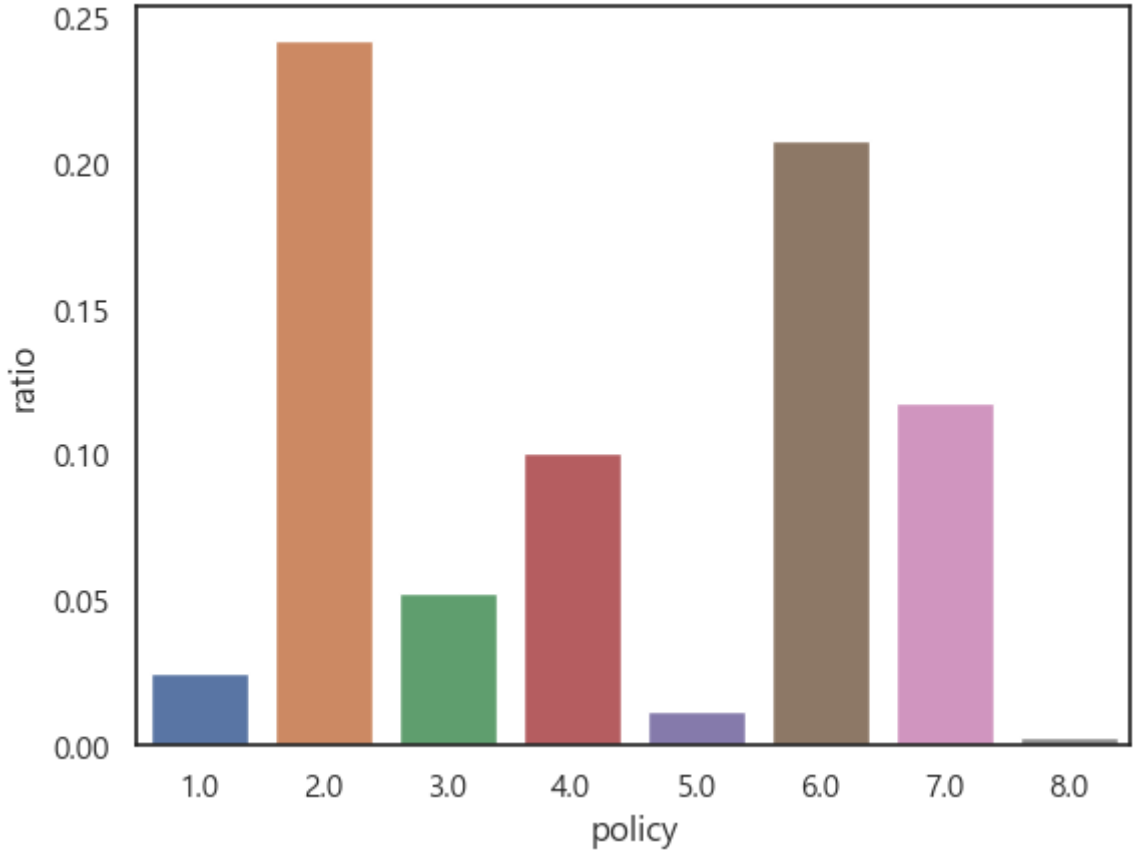
In [ ]:
```
sns.barplot(x='정부지원정책_추진정책2코드', y='정부지원정책_추진정책2코드', data=
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
plt.xlabel('policy')
plt.ylabel('ratio')
plt.show()
```



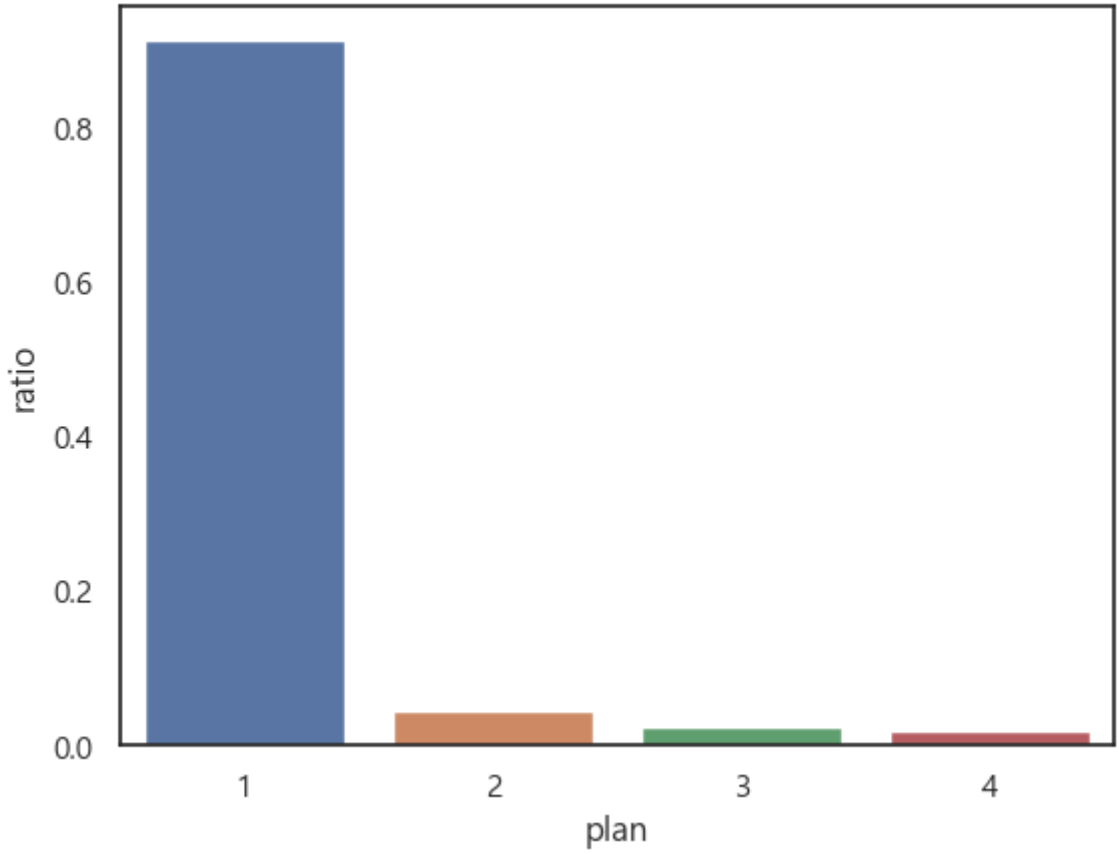
In [ ]:
```
sns.barplot(x='코로나19_지원정책1코드', y='코로나19_지원정책1코드', data=survey_
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
plt.xlabel('policy')
plt.ylabel('ratio')
plt.show()
```



In [ ]:
```
sns.barplot(x='코로나19_지원정책2코드', y='코로나19_지원정책2코드', data=survey_
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
plt.xlabel('policy')
plt.ylabel('ratio')
plt.show()
```



In [ ]:
```
sns.barplot(x='사업전환_운영계획코드', y='사업전환_운영계획코드', data=survey_8,
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
plt.xlabel('plan')
plt.ylabel('ratio')
plt.show()
# 1: 계속 운영, 2: 사업 전환, 3: 폐업 후 임금 근로자 전향, 4: 은퇴
```



In [ ]:
```
survey_statistic=survey[['일반_합계종사자수','창업_비용_총창업비용', "창업_비용_
       '경영_영업비용_기타금액', '경영_영업이익', '경영_전자상거래_매출비율', '경
```

In [ ]:
```
survey_statistic.corr(method='pearson', min_periods=1)
```

Out[ ]:

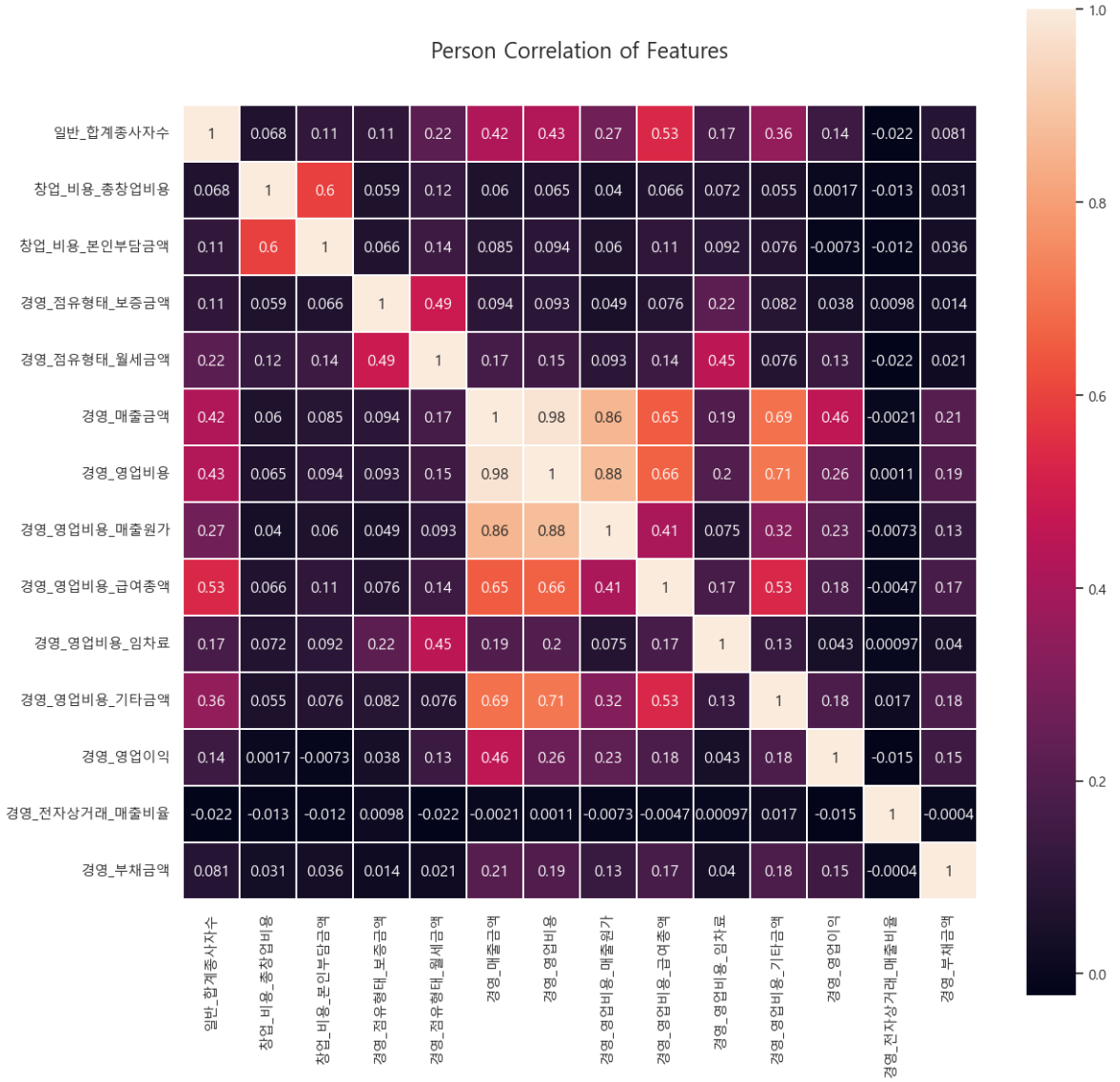
	일반_합계종사자수	창업_비용_총창업비용	창업_비용_본인부담금액	경영_점유형태_보증금액	경영_점유형태_월세금액	경영_매출금액	경영_영업비용	경영_영업비용_매출
일반_합계종사자수	1.000000	0.068304	0.109768	0.107674	0.217380	0.422272	0.425598	0.27368
창업_비용_총창업비용	0.068304	1.000000	0.595331	0.058988	0.123846	0.059914	0.064650	0.04009
창업_비용_본인부담금액	0.109768	0.595331	1.000000	0.066186	0.141858	0.085255	0.094288	0.05971
경영_점유형태_보증금액	0.107674	0.058988	0.066186	1.000000	0.485002	0.093628	0.092667	0.04865
경영_점유형태_월	0.217380	0.123846	0.141858	0.485002	1.000000	0.165127	0.147493	0.09346

	일반_합계종사자수	창업_비용_총창업비용	창업_비용_본인부담금액	경영_점유형태_보증금액	경영_점유형태_월세금액	경영_매출금액	경영_영업비용	경영_영업비용_매출원가
세금액								
경영_매출금액	0.422272	0.059914	0.085255	0.093628	0.165127	1.000000	0.977672	0.859532
경영_영업비용	0.425598	0.064650	0.094288	0.092667	0.147493	0.977672	1.000000	0.879321
경영_영업비용_매출원가	0.273688	0.040092	0.059713	0.048654	0.093464	0.859532	0.879321	1.000000
경영_영업비용_급여총액	0.534603	0.066091	0.106392	0.075843	0.143699	0.649029	0.662450	0.408654
경영_영업비용_임차료	0.169877	0.071917	0.091837	0.222496	0.454759	0.190116	0.196154	0.075368

	일반_합계_종사자수	창업_비용_총창업비용	창업_비용_본인부담금액	경영_점유형태_보증금액	경영_점유형태_월세금액	경영_매출금액	경영_영업비용	경영_영업비용_매출원기
경영_영업비용_기타금액	0.356054	0.055049	0.076242	0.081619	0.075807	0.692829	0.708668	0.32072
경영_영업이익	0.138991	0.001669	-0.007333	0.038008	0.134499	0.457005	0.259893	0.22781
경영_전자상거래_매출비율	-0.022457	-0.013040	-0.011524	0.009797	-0.022042	-0.002138	0.001129	-0.00727
경영_부채금액	0.081191	0.031018	0.036182	0.013663	0.021082	0.208643	0.192139	0.13146

In [ ]:
```
plt.figure(figsize=(14, 14))
plt.title("Person Correlation of Features", y = 1.05, size = 17)
sns.set(font="Malgun Gothic",
rc={"axes.unicode_minus":False}, style='white')
sns.heatmap(survey_statistic.corr(), linewidths = 0.1, vmax = 1.0,
            square = True,  linecolor = "white", annot = True, annot_kws = {"size
# 영업 이익 -> 12번째 col
```

Out[ ]:
```
<Axes: title={'center': 'Person Correlation of Features'}>
```


Person Correlation of Features
일반_합계종사자수 1 0.068 0.11 0.11 0.22 0.42 0.43 0.27 0.53 0.17 0.36 0.14 -0.022 0.081
창업_비용_총창업비용 0.068 1 0.6 0.059 0.12 0.06 0.065 0.04 0.066 0.072 0.055 0.0017 -0.013 0.031
창업_비용_본인부담금액 0.11 0.6 1 0.066 0.14 0.085 0.094 0.06 0.11 0.092 0.076 -0.0073 -0.012 0.036
경영_점유형태_보증금액 0.11 0.059 0.066 1 0.49 0.094 0.093 0.049 0.076 0.22 0.082 0.038 0.0098 0.014
경영_점유형태_월세금액 0.22 0.12 0.14 0.49 1 0.17 0.15 0.093 0.14 0.45 0.076 0.13 -0.022 0.021
경영_매출금액 0.42 0.06 0.085 0.094 0.17 1 0.98 0.86 0.65 0.19 0.69 0.46 -0.0021 0.21
경영_영업비용 0.43 0.065 0.094 0.093 0.15 0.98 1 0.88 0.66 0.2 0.71 0.26 0.0011 0.19
경영_영업비용_매출원가 0.27 0.04 0.06 0.049 0.093 0.86 0.88 1 0.41 0.075 0.32 0.23 -0.0073 0.13
경영_영업비용_급여총액 0.53 0.066 0.11 0.076 0.14 0.65 0.66 0.41 1 0.17 0.53 0.18 -0.0047 0.17
경영_영업비용_임차료 0.17 0.072 0.092 0.22 0.45 0.19 0.2 0.075 0.17 1 0.13 0.043 0.00097 0.04
경영_영업비용_기타금액 0.36 0.055 0.076 0.082 0.076 0.69 0.71 0.32 0.53 0.13 1 0.18 0.017 0.18
경영_영업이익 0.14 0.0017 -0.0073 0.038 0.13 0.46 0.26 0.23 0.18 0.043 0.18 1 -0.015 0.15
경영_전자상거래_매출비율 -0.022 -0.013 -0.012 0.0098 -0.022 -0.0021 0.0011 -0.0073 -0.0047 0.00097 0.017 -0.015 1 -0.0004
경영_부채금액 0.081 0.031 0.036 0.014 0.021 0.21 0.19 0.13 0.17 0.04 0.18 0.15 -0.0004 1
일반_합계종사자수
창업_비용_총창업비용
창업_비용_본인부담금액
경영_점유형태_보증금액
경영_점유형태_월세금액
경영_매출금액
경영_영업비용
경영_영업비용_매출원가
경영_영업비용_급여총액
경영_영업비용_임차료
경영_영업비용_기타금액
경영_영업이익
경영_전자상거래_매출비율
경영_부채금액
1.0
0.8
0.6
0.4
0.2
0.0