

데이터 3 전처리

- 월별 그룹화를 위해 'base_year', 'base_month' 열 결합 -> 'y_m' 열 생성
- 'sex', 'resid_reg_pop', 'foreign_pop' 열을 이용하여 -> 외국인, 제주도민 성별 거주인구 생성.
- 행정부에서 제공하는 "주민등록인구통계" 외부데이터에서 나이 열을 활용 -> 연령별 '제주도민_60이상', '제주도민_60미만' 파생 변수 생성

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

import warnings
warnings.filterwarnings(action='ignore')
```

제주시 거주민 데이터 전처리 결합

```
In [2]: # 파일 불러오기
df_1 = pd.read_csv('03_거주인구_RESIDENT_POP.csv', encoding='cp949')
df_1.head()
```

```
Out[2]:
```

	base_year	base_month	city	emd_cd	emd_nm	sex	resid_reg_pop	foreign_pop	total_pop
0	2018	1	제주시	50110590	건입동	남성	5085	146	5231
1	2018	1	제주시	50110590	건입동	여성	4715	82	4797
2	2018	1	제주시	50110256	구좌읍	남성	7965	368	8333
3	2018	1	제주시	50110256	구좌읍	여성	7609	184	7793
4	2018	1	서귀포시	50130253	남원읍	남성	9806	428	10234

```
In [3]: df_1.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3612 entries, 0 to 3611
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype  
---  --
0   base_year       3612 non-null   int64   
1   base_month      3612 non-null   int64   
2   city            3612 non-null   object   
3   emd_cd          3612 non-null   int64   
4   emd_nm          3612 non-null   object   
5   sex             3612 non-null   object   
6   resid_reg_pop   3612 non-null   int64   
7   foreign_pop     3612 non-null   int64   
8   total_pop       3612 non-null   int64   
dtypes: int64(6), object(3)
memory usage: 254.1+ KB
```

```
In [4]: # 행정동명 하나당 emd_cd는 한개이다.
# 'emd_cd' 와 'emd_nm' 은 1:1 관계
for i in df_1.emd_nm.unique():
    display(i, df_1[df_1.emd_nm == i].emd_cd.unique())
```

```
'건입동'
array([50110590], dtype=int64)
'구좌읍'
array([50110256], dtype=int64)
'남원읍'
array([50130253], dtype=int64)
'노형동'
array([50110660], dtype=int64)
'대륜동'
array([50130590], dtype=int64)
'대정읍'
array([50130250], dtype=int64)
'대천동'
array([50130600], dtype=int64)
'도두동'
array([50110690], dtype=int64)
'동흥동'
array([50130570], dtype=int64)
'봉개동'
array([50110620], dtype=int64)
'삼도1동'
array([50110550], dtype=int64)
'삼도2동'
array([50110560], dtype=int64)
'삼양동'
array([50110610], dtype=int64)
'서흥동'
array([50130580], dtype=int64)
'성산읍'
array([50130259], dtype=int64)
'송산동'
array([50130510], dtype=int64)
'아라동'
array([50110630], dtype=int64)
'안덕면'
array([50130310], dtype=int64)
'애월읍'
array([50110253], dtype=int64)
'연동'
array([50110650], dtype=int64)
'영천동'
array([50130560], dtype=int64)
'예래동'
array([50130620], dtype=int64)
'오라동'
array([50110640], dtype=int64)
'외도동'
array([50110670], dtype=int64)
'용담1동'
array([50110570], dtype=int64)
'용담2동'
array([50110580], dtype=int64)
'우도면'
array([50110330], dtype=int64)
'이도1동'
array([50110530], dtype=int64)
'이도2동'
array([50110540], dtype=int64)
'이호동'
array([50110680], dtype=int64)
```

```

'일도1동'
array([50110510], dtype=int64)
'일도2동'
array([50110520], dtype=int64)
'정방동'
array([50130520], dtype=int64)
'조천읍'
array([50110259], dtype=int64)
'중문동'
array([50130610], dtype=int64)
'중앙동'
array([50130530], dtype=int64)
'천지동'
array([50130540], dtype=int64)
'추자면'
array([50110320], dtype=int64)
'표선면'
array([50130320], dtype=int64)
'한경면'
array([50110310], dtype=int64)
'한림읍'
array([50110250], dtype=int64)
'화북동'
array([50110600], dtype=int64)
'효돈동'
array([50130550], dtype=int64)

```

```

In [5]: ## column 이름 변경, 행정동 코드 제거
df_1 = df_1.drop('emd_cd', axis = 1)
df_1 = df_1.rename(columns={'emd_nm': 'location'})
df_1.head()

```

```

Out[5]:
   base_year  base_month  city location sex  resid_reg_pop  foreign_pop  total_pop
0      2018           1   제주시  건입동  남성           5085           146         5231
1      2018           1   제주시  건입동  여성           4715            82         4797
2      2018           1   제주시  구좌읍  남성           7965           368         8333
3      2018           1   제주시  구좌읍  여성           7609           184         7793
4      2018           1  서귀포시  남원읍  남성           9806           428        10234

```

```

In [6]: # month 변수 만들기 (ex. 01,02 ..)
df_1['base_month'] = df_1['base_month'].apply(lambda x: "{:0>2d}".format(x))

# 년-월 형태로 만들기
df_1['y_m'] = df_1[['base_year', 'base_month']].apply(lambda x: '-'.join(x.astype(str)),
df_1.head()

```

```

Out[6]:
   base_year  base_month  city location sex  resid_reg_pop  foreign_pop  total_pop  y_m
0      2018           01   제주시  건입동  남성           5085           146         5231  2018-01
1      2018           01   제주시  건입동  여성           4715            82         4797  2018-01
2      2018           01   제주시  구좌읍  남성           7965           368         8333  2018-01
3      2018           01   제주시  구좌읍  여성           7609           184         7793  2018-01

```

	base_year	base_month	city	location	sex	resid_reg_pop	foreign_pop	total_pop	y_m
4	2018	01	서귀포 시	남원읍	남성	9806	428	10234	2018-01

```
In [7]: #필요 없는 열 제거
df_1 = df_1.drop({'base_year', "base_month"},axis=1)
df_1.head()
```

```
Out[7]:
```

	city	location	sex	resid_reg_pop	foreign_pop	total_pop	y_m
0	제주시	건입동	남성	5085	146	5231	2018-01
1	제주시	건입동	여성	4715	82	4797	2018-01
2	제주시	구좌읍	남성	7965	368	8333	2018-01
3	제주시	구좌읍	여성	7609	184	7793	2018-01
4	서귀포시	남원읍	남성	9806	428	10234	2018-01

```
In [8]: # 제주도 데이터만 추출
df_1 = df_1.loc[df_1["city"]=="제주시"]
```

남성 거주자 데이터 변수 생성

```
In [9]: #남성 거주자 데이터 생성
temp = df_1[df_1.sex == "남성"]
temp
```

```
Out[9]:
```

	city	location	sex	resid_reg_pop	foreign_pop	total_pop	y_m
0	제주시	건입동	남성	5085	146	5231	2018-01
2	제주시	구좌읍	남성	7965	368	8333	2018-01
6	제주시	노형동	남성	26286	821	27107	2018-01
14	제주시	도두동	남성	1602	43	1645	2018-01
18	제주시	봉개동	남성	2089	37	2126	2018-01
...
3592	제주시	조천읍	남성	13017	279	13296	2021-06
3600	제주시	추자면	남성	874	240	1114	2021-06
3604	제주시	한경면	남성	4627	237	4864	2021-06
3606	제주시	한림읍	남성	10891	2090	12981	2021-06
3608	제주시	화북동	남성	12062	169	12231	2021-06

1092 rows × 7 columns

```
In [10]: # 열 이름 정의
temp.rename(columns={"resid_reg_pop" : "resid_reg_pop_남", "foreign_pop" : "foreign_pop_남"},axis = 1)
temp = temp.drop("sex",axis = 1)
temp
```

Out[10]:

	city	location	resid_reg_pop_남	foreign_pop_남	total_pop_남	y_m
0	제주시	건입동	5085	146	5231	2018-01
2	제주시	구좌읍	7965	368	8333	2018-01
6	제주시	노형동	26286	821	27107	2018-01
14	제주시	도두동	1602	43	1645	2018-01
18	제주시	봉개동	2089	37	2126	2018-01
...
3592	제주시	조천읍	13017	279	13296	2021-06
3600	제주시	추자면	874	240	1114	2021-06
3604	제주시	한경면	4627	237	4864	2021-06
3606	제주시	한림읍	10891	2090	12981	2021-06
3608	제주시	화북동	12062	169	12231	2021-06

1092 rows × 6 columns

여성 거주자 데이터 변수 생성

In [11]:

```
#여성 거주자 데이터 변수 생성
temp2 = df_1[df_1.sex == "여성"]
temp2
```

Out[11]:

	city	location	sex	resid_reg_pop	foreign_pop	total_pop	y_m
1	제주시	건입동	여성	4715	82	4797	2018-01
3	제주시	구좌읍	여성	7609	184	7793	2018-01
7	제주시	노형동	여성	27083	768	27851	2018-01
15	제주시	도두동	여성	1424	55	1479	2018-01
19	제주시	봉개동	여성	1838	39	1877	2018-01
...
3593	제주시	조천읍	여성	12422	242	12664	2021-06
3601	제주시	추자면	여성	752	11	763	2021-06
3605	제주시	한경면	여성	4531	100	4631	2021-06
3607	제주시	한림읍	여성	10341	1140	11481	2021-06
3609	제주시	화북동	여성	12238	161	12399	2021-06

1092 rows × 7 columns

In [12]:

```
# 열이름 정의
temp2.rename(columns={"resid_reg_pop" : "resid_reg_pop_여", "foreign_pop" : "foreign_pop_여"}, inplace=True)
temp2 = temp2.drop("sex",axis = 1)
temp2
```

Out [12]:

	city	location	resid_reg_pop_여	foreign_pop_여	total_pop_여	y_m
1	제주시	건입동	4715	82	4797	2018-01
3	제주시	구좌읍	7609	184	7793	2018-01
7	제주시	노형동	27083	768	27851	2018-01
15	제주시	도두동	1424	55	1479	2018-01
19	제주시	봉개동	1838	39	1877	2018-01
...
3593	제주시	조천읍	12422	242	12664	2021-06
3601	제주시	추자면	752	11	763	2021-06
3605	제주시	한경면	4531	100	4631	2021-06
3607	제주시	한림읍	10341	1140	11481	2021-06
3609	제주시	화북동	12238	161	12399	2021-06

1092 rows × 6 columns

```
In [13]: #데이터 병합
df2 = pd.merge(temp,temp2,how = "left", on = ["y_m","city","location"])
df2.head()
```

Out [13]:

	city	location	resid_reg_pop_남	foreign_pop_남	total_pop_남	y_m	resid_reg_pop_여	foreign_pop_여	total_pop_여
0	제주시	건입동	5085	146	5231	2018-01	4715	82	4797
1	제주시	구좌읍	7965	368	8333	2018-01	7609	184	7793
2	제주시	노형동	26286	821	27107	2018-01	27083	768	27851
3	제주시	도두동	1602	43	1645	2018-01	1424	55	1479
4	제주시	봉개동	2089	37	2126	2018-01	1838	39	1877

```
In [14]: # 총거주인구 열 생성
df2["total_pop"] = df2.total_pop_남 + df2.total_pop_여
df2 = df2.drop({'total_pop_남','total_pop_여'},axis=1)
df2.head()
```

Out [14]:

	city	location	resid_reg_pop_남	foreign_pop_남	y_m	resid_reg_pop_여	foreign_pop_여	total_pop
--	------	----------	-----------------	---------------	-----	-----------------	---------------	-----------

	city	location	resid_reg_pop_남	foreign_pop_남	y_m	resid_reg_pop_여	foreign_pop_여	total_pop
0	제주 시	건입동	5085	146	2018-01	4715	82	10028
1	제주 시	구좌읍	7965	368	2018-01	7609	184	16126
2	제주 시	노형동	26286	821	2018-01	27083	768	54958
3	제주 시	도두동	1602	43	2018-01	1424	55	3124
4	제주 시	봉개동	2089	37	2018-01	1838	39	4003

```
In [15]: # 그룹화
df2 = df2.groupby(["y_m", "city", "location"]).sum()
df2 = df2.reset_index()
df2.head()
```

```
Out[15]:
```

	y_m	city	location	resid_reg_pop_남	foreign_pop_남	resid_reg_pop_여	foreign_pop_여	total_pop
0	2018-01	제주 시	건입동	5085	146	4715	82	10028
1	2018-01	제주 시	구좌읍	7965	368	7609	184	16126
2	2018-01	제주 시	노형동	26286	821	27083	768	54958
3	2018-01	제주 시	도두동	1602	43	1424	55	3124
4	2018-01	제주 시	봉개동	2089	37	1838	39	4003

행정부에서 제공하는 "주민등록인구통계" 외부데이터에서 나이 열을 활용

```
In [16]: ## 주민등록인구통계 로드
resid_1 = pd.read_csv('주민등록인구통계/201801_201806_연령별인구현황_월간_제주시.csv', encoding='utf-8')
resid_2 = pd.read_csv('주민등록인구통계/201807_201812_연령별인구현황_월간_제주시.csv', encoding='utf-8')
resid_3 = pd.read_csv('주민등록인구통계/201901_201906_연령별인구현황_월간_제주시.csv', encoding='utf-8')
resid_4 = pd.read_csv('주민등록인구통계/201907_201912_연령별인구현황_월간_제주시.csv', encoding='utf-8')
resid_5 = pd.read_csv('주민등록인구통계/202001_202006_연령별인구현황_월간_제주시.csv', encoding='utf-8')
resid_6 = pd.read_csv('주민등록인구통계/202007_202012_연령별인구현황_월간_제주시.csv', encoding='utf-8')
resid_7 = pd.read_csv('주민등록인구통계/202101_202106_연령별인구현황_월간_제주시.csv', encoding='utf-8')
```

```
In [17]: ## 주민등록인구통계 데이터에서 필요한 열만 추출
def mm(resid):
    resid = resid.set_index("행정구역")
    col_na = resid.columns.tolist()
    col_na = col_na[1::21]
    resid = resid.loc[:, col_na]
```

```

resid = resid.reset_index()
return resid

```

```

In [18]: #데이터프레임 결합
resid_m = mm(resid_1)
for i in [resid_2,resid_3,resid_4,resid_5,resid_6,resid_7]:
    resid_m = pd.merge(resid_m,mm(i),how = "left", on = ["행정구역"])

resid_m.head()

```

```

Out[18]:

```

	행정구역	2018 년01 월_계_ 연령 구간 인구 수	2018 년02 월_계_ 연령 구간 인구 수	2018 년03 월_계_ 연령 구간 인구 수	2018 년04 월_계_ 연령 구간 인구 수	2018 년05 월_계_ 연령 구간 인구 수	2018 년06 월_계_ 연령 구간 인구 수	2018 년07 월_계_ 연령 구간 인구 수	2018 년08 월_계_ 연령 구간 인구 수	2018 년09 월_계_ 연령 구간 인구 수	...	2020년 09월_ 계_연령 구간인 구수	2020 년10 월_계_연 령구간 인구수
0	제주특별자치 도 제주시 (5011000000)	88,634	88,934	89,349	89,697	90,128	90,470	90,878	91,350	91,860	...	102,221	102,7
1	제주특별자치 도 제주시 한 림읍 (5011025000)	5,603	5,618	5,637	5,659	5,701	5,730	5,745	5,789	5,815	...	6,347	6,3
2	제주특별자치 도 제주시 애 월읍 (5011025300)	7,786	7,797	7,842	7,870	7,910	7,954	7,999	8,043	8,074	...	8,891	8,9
3	제주특별자치 도 제주시 구 좌읍 (5011025600)	4,849	4,844	4,857	4,861	4,882	4,902	4,920	4,929	4,944	...	5,297	5,3
4	제주특별자치 도 제주시 조 천읍 (5011025900)	5,751	5,783	5,814	5,836	5,849	5,863	5,900	5,928	5,973	...	6,613	6,6

5 rows × 43 columns

```

In [19]: # 행정구역열 전처리
resid_m.행정구역 = resid_m.apply(lambda x: x.행정구역[12:], axis = 1)
resid_m.행정구역 = resid_m.apply(lambda x: x.행정구역.split("(")[0], axis = 1)
resid_m.head()

```

```

Out[19]:

```

	행정구역	2018 년01 월_계_ 연령 구간 인구 수	2018 년02 월_계_ 연령 구간 인구 수	2018 년03 월_계_ 연령 구간 인구 수	2018 년04 월_계_ 연령 구간 인구 수	2018 년05 월_계_ 연령 구간 인구 수	2018 년06 월_계_ 연령 구간 인구 수	2018 년07 월_계_ 연령 구간 인구 수	2018 년08 월_계_ 연령 구간 인구 수	2018 년09 월_계_ 연령 구간 인구 수	...	2020년 09월_ 계_연령 구간인 구수	2020년 10월_ 계_연령 구간인 구수	2020 년11 월_계_연 령구간 인구수
0	88,634	88,934	89,349	89,697	90,128	90,470	90,878	91,350	91,860	...	102,221	102,749	103,15	

	행정구역	2018 년01 월_계_연령 구간 인구 수	2018 년02 월_계_연령 구간 인구 수	2018 년03 월_계_연령 구간 인구 수	2018 년04 월_계_연령 구간 인구 수	2018 년05 월_계_연령 구간 인구 수	2018 년06 월_계_연령 구간 인구 수	2018 년07 월_계_연령 구간 인구 수	2018 년08 월_계_연령 구간 인구 수	2018 년09 월_계_연령 구간 인구 수	...	2020년 09월_계_연령 구간인 구수	2020년 10월_계_연령 구간인 구수	2020년 11월_계_연령 구간인 구수
1	한림읍	5,603	5,618	5,637	5,659	5,701	5,730	5,745	5,789	5,815	...	6,347	6,383	6,388
2	애월읍	7,786	7,797	7,842	7,870	7,910	7,954	7,999	8,043	8,074	...	8,891	8,921	8,951
3	구좌읍	4,849	4,844	4,857	4,861	4,882	4,902	4,920	4,929	4,944	...	5,297	5,312	5,331
4	조천읍	5,751	5,783	5,814	5,836	5,849	5,863	5,900	5,928	5,973	...	6,613	6,644	6,661

5 rows × 43 columns

```
In [20]: # 전처리를 위한 데이터프레임 Transform
resid_m = resid_m.set_index("행정구역")
resid_t = resid_m.T
resid_t = resid_t.reset_index()
resid_t["index"] = resid_t.apply(lambda x: x["index"][:7],axis = 1)
resid_t["index"] = resid_t.apply(lambda x: x["index"].replace("년","-"),axis = 1)
resid_t.head()
```

	행정구역	index	한림읍	애월읍	구좌읍	조천읍	한경면	추자면	우도면	일도1동	...	화북동	삼양동	봉개동	아라동	오라동
0	2018-01	88,634	5,603	7,786	4,849	5,751	3,078	759	548	962	...	3,877	3,197	903	4,085	1,823
1	2018-02	88,934	5,618	7,797	4,844	5,783	3,083	761	554	959	...	3,900	3,230	904	4,100	1,827
2	2018-03	89,349	5,637	7,842	4,857	5,814	3,096	758	554	966	...	3,911	3,247	909	4,155	1,849
3	2018-04	89,697	5,659	7,870	4,861	5,836	3,120	757	550	965	...	3,932	3,271	916	4,197	1,868
4	2018-05	90,128	5,701	7,910	4,882	5,849	3,140	760	556	967	...	3,947	3,275	918	4,215	1,878

5 rows × 28 columns

```
In [21]: # 60세 이상 제주도민 데이터 삽입 함수 형성
def mm2(loca):
    temp = df2.loc[df2["location"]==loc, :]
    temp = temp.reset_index()
```

```
temp = temp.drop("index",axis = 1)
temp.loc[:, "60세이상"] = resid_t.loc[:, loca]
return temp
```

```
In [22]: # 기존 데이터 프레임 df2에 외부데이터 60세 이상 제주도민 데이터 삽입 함수 형성
df2["60세이상"] = 0
df_s = pd.DataFrame(columns=['y_m', 'city', 'location', 'resid_reg_pop_남', 'foreign_pop_남',
                             'resid_reg_pop_여', 'foreign_pop_여', 'total_pop', '60세이상'])

for i in df2.location.unique():
    df_s = pd.concat([df_s, mm2(i)])
```

```
In [23]: df_s.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1092 entries, 0 to 41
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   y_m                   1092 non-null   object
1   city                  1092 non-null   object
2   location              1092 non-null   object
3   resid_reg_pop_남     1092 non-null   object
4   foreign_pop_남       1092 non-null   object
5   resid_reg_pop_여     1092 non-null   object
6   foreign_pop_여       1092 non-null   object
7   total_pop            1092 non-null   object
8   60세이상             1092 non-null   object
dtypes: object(9)
memory usage: 85.3+ KB
```

```
In [24]: # 자료형 데이터 정수형으로 변경
for i in range(3, (len(df_s.columns)-1)):
    df_s.iloc[:, i] = df_s.iloc[:, i].astype(int)
```

```
In [25]: # 자료형 데이터 정수형으로 변경
df_s["60세이상"] = df_s.apply(lambda x : x["60세이상"].replace(',',''), axis = 1)
df_s.loc[:, "60세이상"] = df_s.loc[:, "60세이상"].astype(int)
df_s.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1092 entries, 0 to 41
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   y_m                   1092 non-null   object
1   city                  1092 non-null   object
2   location              1092 non-null   object
3   resid_reg_pop_남     1092 non-null   int32
4   foreign_pop_남       1092 non-null   int32
5   resid_reg_pop_여     1092 non-null   int32
6   foreign_pop_여       1092 non-null   int32
7   total_pop            1092 non-null   int32
8   60세이상             1092 non-null   int32
dtypes: int32(6), object(3)
memory usage: 59.7+ KB
```

```
In [26]: # 열 이름 변경
df_s.rename(columns={"resid_reg_pop_남" : "제주도민_남", "resid_reg_pop_여" : "제주도민_여",
                    "foreign_pop_남" : "외국인거주_남", "foreign_pop_여" : "외국인거주_여", "60세이상" : "제주도민_60이상"}, inplace=True)
df_s.head()
```

Out[26]:

	y_m	city	location	제주도민_남	외국인거주_남	제주도민_여	외국인거주_여	total_pop	제주도민_60이상
0	2018-01	제주시	건입동	5085	146	4715	82	10028	2478
1	2018-02	제주시	건입동	5090	142	4724	82	10038	2484
2	2018-03	제주시	건입동	5072	141	4711	81	10005	2494
3	2018-04	제주시	건입동	5070	143	4718	86	10017	2484
4	2018-05	제주시	건입동	5070	148	4730	90	10038	2504

In [27]:

```
# 60세 미만 파생변수 열 생성
df_s["제주도민_60미만"] = df_s.apply(lambda x: x.total_pop - x.제주도민_60이상, axis = 0)
df_s.head()
```

Out[27]:

	y_m	city	location	제주도민_남	외국인거주_남	제주도민_여	외국인거주_여	total_pop	제주도민_60이상	제주도민_60미만
0	2018-01	제주시	건입동	5085	146	4715	82	10028	2478	7550
1	2018-02	제주시	건입동	5090	142	4724	82	10038	2484	7554
2	2018-03	제주시	건입동	5072	141	4711	81	10005	2494	7511
3	2018-04	제주시	건입동	5070	143	4718	86	10017	2484	7533
4	2018-05	제주시	건입동	5070	148	4730	90	10038	2504	7534

In [28]:

```
# 열 순서 변경
df_s = df_s.loc[:,["y_m","city","location","제주도민_여","외국인거주_여","제주도민_남","외국인거주_남","total_pop","제주도민_60미만","제주도민_60이상"]]
```

In [29]:

```
# csv파일 변환
df_s.to_csv("3번데이터_성별_연령_제주시.csv",encoding = "cp949",index=False)
```

서귀포시 거주민 데이터 전처리 결합

In [30]:

```
# 파일 불러오기
df_1 = pd.read_csv('03_거주인구_RESIDENT_POP.csv',encoding='cp949')
df_1.head()
```

Out[30]:

	base_year	base_month	city	emd_cd	emd_nm	sex	resid_reg_pop	foreign_pop	total_pop
0	2018	1	제주시	50110590	건입동	남성	5085	146	5231
1	2018	1	제주시	50110590	건입동	여성	4715	82	4797

	base_year	base_month	city	emd_cd	emd_nm	sex	resid_reg_pop	foreign_pop	total_pop
2	2018	1	제주시	50110256	구좌읍	남성	7965	368	8333
3	2018	1	제주시	50110256	구좌읍	여성	7609	184	7793
4	2018	1	서귀포시	50130253	남원읍	남성	9806	428	10234

```
In [31]: # 서귀포시 데이터만 추출
df_1 = df_1.loc[df_1["city"] == "서귀포시"]
```

```
In [32]: df_1.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1428 entries, 4 to 3611
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   base_year       1428 non-null   int64
1   base_month      1428 non-null   int64
2   city            1428 non-null   object
3   emd_cd          1428 non-null   int64
4   emd_nm          1428 non-null   object
5   sex            1428 non-null   object
6   resid_reg_pop   1428 non-null   int64
7   foreign_pop     1428 non-null   int64
8   total_pop       1428 non-null   int64
dtypes: int64(6), object(3)
memory usage: 111.6+ KB
```

```
In [33]: # 행정동명 하나당 emd_cd는 한개이다.
# 'emd_cd' 와 'emd_nm' 은 1:1 관계
for i in df_1.emd_nm.unique():
    display(i, df_1[df_1.emd_nm == i].emd_cd.unique())
```

```
'남원읍'
array([50130253], dtype=int64)
'대륜동'
array([50130590], dtype=int64)
'대정읍'
array([50130250], dtype=int64)
'대천동'
array([50130600], dtype=int64)
'동홍동'
array([50130570], dtype=int64)
'서홍동'
array([50130580], dtype=int64)
'성산읍'
array([50130259], dtype=int64)
'송산동'
array([50130510], dtype=int64)
'안덕면'
array([50130310], dtype=int64)
'영천동'
array([50130560], dtype=int64)
'예래동'
array([50130620], dtype=int64)
'정방동'
array([50130520], dtype=int64)
'중문동'
```

```
array([50130610], dtype=int64)
'중앙동'
array([50130530], dtype=int64)
'천지동'
array([50130540], dtype=int64)
'표선면'
array([50130320], dtype=int64)
'효돈동'
array([50130550], dtype=int64)
```

```
In [34]: ## column 이름 변경, 행정동 코드 제거
df_1 = df_1.drop('emd_cd', axis = 1)
df_1 = df_1.rename(columns={'emd_nm': 'location'})
df_1.head()
```

```
Out[34]:
```

	base_year	base_month	city	location	sex	resid_reg_pop	foreign_pop	total_pop
4	2018	1	서귀포시	남원읍	남성	9806	428	10234
5	2018	1	서귀포시	남원읍	여성	9306	200	9506
8	2018	1	서귀포시	대륜동	남성	6836	83	6919
9	2018	1	서귀포시	대륜동	여성	6637	95	6732
10	2018	1	서귀포시	대정읍	남성	10360	841	11201

```
In [35]: # month 변수 만들기 (ex. 01,02 ...)
df_1['base_month'] = df_1['base_month'].apply(lambda x: "{:0>2d}".format(x))

# 년-월 형태로 만들기
df_1['y_m'] = df_1[['base_year', 'base_month']].apply(lambda x: '-'.join(x.astype(str)),
df_1.head()
```

```
Out[35]:
```

	base_year	base_month	city	location	sex	resid_reg_pop	foreign_pop	total_pop	y_m
4	2018	01	서귀포 시	남원읍	남 성	9806	428	10234	2018- 01
5	2018	01	서귀포 시	남원읍	여 성	9306	200	9506	2018- 01
8	2018	01	서귀포 시	대륜동	남 성	6836	83	6919	2018- 01
9	2018	01	서귀포 시	대륜동	여 성	6637	95	6732	2018- 01
10	2018	01	서귀포 시	대정읍	남 성	10360	841	11201	2018- 01

```
In [36]: #필요 없는 열 제거
df_1 = df_1.drop({'base_year', "base_month"}, axis=1)
df_1.head()
```

```
Out[36]:
```

	city	location	sex	resid_reg_pop	foreign_pop	total_pop	y_m
4	서귀포시	남원읍	남성	9806	428	10234	2018-01

	city	location	sex	resid_reg_pop	foreign_pop	total_pop	y_m
5	서귀포시	남원읍	여성	9306	200	9506	2018-01
8	서귀포시	대륜동	남성	6836	83	6919	2018-01
9	서귀포시	대륜동	여성	6637	95	6732	2018-01
10	서귀포시	대정읍	남성	10360	841	11201	2018-01

남성 거주자 데이터 변수 생성

```
In [37]: # 남성 거주자 데이터 추출
temp = df_1[df_1.sex == "남성"]
temp
```

```
Out[37]:
```

	city	location	sex	resid_reg_pop	foreign_pop	total_pop	y_m
4	서귀포시	남원읍	남성	9806	428	10234	2018-01
8	서귀포시	대륜동	남성	6836	83	6919	2018-01
10	서귀포시	대정읍	남성	10360	841	11201	2018-01
12	서귀포시	대천동	남성	6685	107	6792	2018-01
16	서귀포시	동홍동	남성	11124	624	11748	2018-01
...
3594	서귀포시	중문동	남성	5938	151	6089	2021-06
3596	서귀포시	중앙동	남성	1605	32	1637	2021-06
3598	서귀포시	천지동	남성	1754	60	1814	2021-06
3602	서귀포시	표선면	남성	6314	211	6525	2021-06
3610	서귀포시	효돈동	남성	2660	39	2699	2021-06

714 rows × 7 columns

```
In [38]: # 열이름 정의
temp.rename(columns={"resid_reg_pop" : "resid_reg_pop_남", "foreign_pop" : "foreign_pop_남"})
temp = temp.drop("sex",axis = 1)
temp
```

```
Out[38]:
```

	city	location	resid_reg_pop_남	foreign_pop_남	total_pop_남	y_m
4	서귀포시	남원읍	9806	428	10234	2018-01
8	서귀포시	대륜동	6836	83	6919	2018-01
10	서귀포시	대정읍	10360	841	11201	2018-01
12	서귀포시	대천동	6685	107	6792	2018-01
16	서귀포시	동홍동	11124	624	11748	2018-01
...
3594	서귀포시	중문동	5938	151	6089	2021-06

	city	location	resid_reg_pop_남	foreign_pop_남	total_pop_남	y_m
3596	서귀포시	중앙동	1605	32	1637	2021-06
3598	서귀포시	천지동	1754	60	1814	2021-06
3602	서귀포시	표선면	6314	211	6525	2021-06
3610	서귀포시	효돈동	2660	39	2699	2021-06

714 rows × 6 columns

여성 거주자 데이터 변수 생성

```
In [39]: # 여성 거주자 데이터 추출
temp2 = df_1[df_1.sex == "여성"]
temp2
```

```
Out[39]:
```

	city	location	sex	resid_reg_pop	foreign_pop	total_pop	y_m
5	서귀포시	남원읍	여성	9306	200	9506	2018-01
9	서귀포시	대륜동	여성	6637	95	6732	2018-01
11	서귀포시	대정읍	여성	10725	677	11402	2018-01
13	서귀포시	대천동	여성	6475	137	6612	2018-01
17	서귀포시	동홍동	여성	11569	642	12211	2018-01
...
3595	서귀포시	중문동	여성	5625	154	5779	2021-06
3597	서귀포시	중앙동	여성	1753	40	1793	2021-06
3599	서귀포시	천지동	여성	1712	55	1767	2021-06
3603	서귀포시	표선면	여성	6033	123	6156	2021-06
3611	서귀포시	효돈동	여성	2597	18	2615	2021-06

714 rows × 7 columns

```
In [40]: # 열이름 정의
temp2.rename(columns={"resid_reg_pop" : "resid_reg_pop_여", "foreign_pop" : "foreign_pop_여"}, inplace=True)
temp2 = temp2.drop("sex", axis = 1)
temp2
```

```
Out[40]:
```

	city	location	resid_reg_pop_여	foreign_pop_여	total_pop_여	y_m
5	서귀포시	남원읍	9306	200	9506	2018-01
9	서귀포시	대륜동	6637	95	6732	2018-01
11	서귀포시	대정읍	10725	677	11402	2018-01
13	서귀포시	대천동	6475	137	6612	2018-01
17	서귀포시	동홍동	11569	642	12211	2018-01
...

	city	location	resid_reg_pop_여	foreign_pop_여	total_pop_여	y_m
3595	서귀포시	중문동	5625	154	5779	2021-06
3597	서귀포시	중앙동	1753	40	1793	2021-06
3599	서귀포시	천지동	1712	55	1767	2021-06
3603	서귀포시	표선면	6033	123	6156	2021-06
3611	서귀포시	효돈동	2597	18	2615	2021-06

714 rows × 6 columns

```
In [41]: # 남성 여성 데이터 병합
df2 = pd.merge(temp,temp2,how = "left", on = ["y_m","city","location"])
df2.head()
```

```
Out[41]:
```

	city	location	resid_reg_pop_남	foreign_pop_남	total_pop_남	y_m	resid_reg_pop_여	foreign_pop_여	total_pop_여
0	서귀포시	남원읍	9806	428	10234	2018-01	9306	200	9506
1	서귀포시	대륜동	6836	83	6919	2018-01	6637	95	6732
2	서귀포시	대정읍	10360	841	11201	2018-01	10725	677	11402
3	서귀포시	대천동	6685	107	6792	2018-01	6475	137	6612
4	서귀포시	동홍동	11124	624	11748	2018-01	11569	642	12211

```
In [42]: # 총 거주인구 데이터 생성
df2["total_pop"] = df2.total_pop_남 + df2.total_pop_여
df2 = df2.drop({'total_pop_남','total_pop_여'},axis=1)
df2.head()
```

```
Out[42]:
```

	city	location	resid_reg_pop_남	foreign_pop_남	y_m	resid_reg_pop_여	foreign_pop_여	total_pop
0	서귀포시	남원읍	9806	428	2018-01	9306	200	19740
1	서귀포시	대륜동	6836	83	2018-01	6637	95	13651

	city	location	resid_reg_pop_남	foreign_pop_남	y_m	resid_reg_pop_여	foreign_pop_여	total_pop
2	서귀포시	대정읍	10360	841	2018-01	10725	677	22603
3	서귀포시	대천동	6685	107	2018-01	6475	137	13404
4	서귀포시	동홍동	11124	624	2018-01	11569	642	23959

```
In [43]: # 그룹화
df2 = df2.groupby(["y_m", "city", "location"]).sum()
df2 = df2.reset_index()
df2.head()
```

```
Out[43]:
```

	y_m	city	location	resid_reg_pop_남	foreign_pop_남	resid_reg_pop_여	foreign_pop_여	total_pop
0	2018-01	서귀포시	남원읍	9806	428	9306	200	19740
1	2018-01	서귀포시	대륜동	6836	83	6637	95	13651
2	2018-01	서귀포시	대정읍	10360	841	10725	677	22603
3	2018-01	서귀포시	대천동	6685	107	6475	137	13404
4	2018-01	서귀포시	동홍동	11124	624	11569	642	23959

행정부에서 제공하는 "주민등록인구통계" 외부데이터에서 나이 열을 활용

```
In [44]: ## 주민등록인구통계 로드
resid_1 = pd.read_csv('주민등록인구통계/201801_201806_연령별인구현황_월간_서귀포.csv', encoding='utf-8')
resid_2 = pd.read_csv('주민등록인구통계/201807_201812_연령별인구현황_월간_서귀포.csv', encoding='utf-8')
resid_3 = pd.read_csv('주민등록인구통계/201901_201906_연령별인구현황_월간_서귀포.csv', encoding='utf-8')
resid_4 = pd.read_csv('주민등록인구통계/201907_201912_연령별인구현황_월간_서귀포.csv', encoding='utf-8')
resid_5 = pd.read_csv('주민등록인구통계/202001_202006_연령별인구현황_월간_서귀포.csv', encoding='utf-8')
resid_6 = pd.read_csv('주민등록인구통계/202007_202012_연령별인구현황_월간_서귀포.csv', encoding='utf-8')
resid_7 = pd.read_csv('주민등록인구통계/202101_202106_연령별인구현황_월간_서귀포.csv', encoding='utf-8')
```

```
In [45]: ## 주민등록인구통계 데이터에서 필요한 열만 추출하는 함수
def mm(resid):
    resid = resid.set_index("행정구역")
    col_na = resid.columns.to_list()
    col_na = col_na[1::21]
    resid = resid.loc[:, col_na]
    resid = resid.reset_index()
    return resid
```

```
In [46]: ## 주민등록인구통계 데이터 병합
resid_m = mm(resid_1)
```

```
for i in [resid_2,resid_3,resid_4,resid_5,resid_6,resid_7]:
    resid_m = pd.merge(resid_m,mm(i),how = "left", on = ["행정구역"])

resid_m.head()
```

Out[46]:

	행정구역	2018 년01 월_계_ 연령 구간 인구 수	2018 년02 월_계_ 연령 구간 인구 수	2018 년03 월_계_ 연령 구간 인구 수	2018 년04 월_계_ 연령 구간 인구 수	2018 년05 월_계_ 연령 구간 인구 수	2018 년06 월_계_ 연령 구간 인구 수	2018 년07 월_계_ 연령 구간 인구 수	2018 년08 월_계_ 연령 구간 인구 수	2018 년09 월_계_ 연령 구간 인구 수	...	2020 년09 월_계_ 연령 구간 인구 수	2020 년10 월_계_ 연령 구간 인구 수
0	제주특별자치 도 서귀포시 (5013000000)	43,508	43,626	43,792	43,993	44,152	44,326	44,490	44,755	45,015	...	49,402	49,652
1	제주특별자치 도 서귀포시 대정읍 (5013025000)	5,330	5,330	5,347	5,371	5,398	5,398	5,416	5,437	5,484	...	5,904	5,939
2	제주특별자치 도 서귀포시 남원읍 (5013025300)	5,554	5,577	5,583	5,609	5,624	5,659	5,668	5,701	5,735	...	6,171	6,199
3	제주특별자치 도 서귀포시 성산읍 (5013025900)	4,642	4,653	4,668	4,682	4,683	4,703	4,718	4,744	4,770	...	5,248	5,275
4	제주특별자치 도 서귀포시 안덕면 (5013031000)	3,031	3,036	3,048	3,084	3,098	3,103	3,114	3,132	3,144	...	3,382	3,394

5 rows × 43 columns

```
In [47]: # 행정구역열 전처리
resid_m.행정구역 = resid_m.apply(lambda x: x.행정구역[13:], axis = 1)
resid_m.행정구역 = resid_m.apply(lambda x: x.행정구역.split("(")[0], axis = 1)
resid_m.head()
```

Out[47]:

	행정 구역	2018 년01 월_계_ 연령 구간 인구 수	2018 년02 월_계_ 연령 구간 인구 수	2018 년03 월_계_ 연령 구간 인구 수	2018 년04 월_계_ 연령 구간 인구 수	2018 년05 월_계_ 연령 구간 인구 수	2018 년06 월_계_ 연령 구간 인구 수	2018 년07 월_계_ 연령 구간 인구 수	2018 년08 월_계_ 연령 구간 인구 수	2018 년09 월_계_ 연령 구간 인구 수	...	2020 년09 월_계_ 연령 구간 인구 수	2020 년10 월_계_ 연령 구간 인구 수	2020 년11 월_계_ 연령 구간 인구 수
0		43,508	43,626	43,792	43,993	44,152	44,326	44,490	44,755	45,015	...	49,402	49,652	49,852
1	대 정 읍	5,330	5,330	5,347	5,371	5,398	5,398	5,416	5,437	5,484	...	5,904	5,939	5,967
2	남 원 읍	5,554	5,577	5,583	5,609	5,624	5,659	5,668	5,701	5,735	...	6,171	6,199	6,213

	행정구역	2018 년01 월_계_연령 구간_인구 수	2018 년02 월_계_연령 구간_인구 수	2018 년03 월_계_연령 구간_인구 수	2018 년04 월_계_연령 구간_인구 수	2018 년05 월_계_연령 구간_인구 수	2018 년06 월_계_연령 구간_인구 수	2018 년07 월_계_연령 구간_인구 수	2018 년08 월_계_연령 구간_인구 수	2018 년09 월_계_연령 구간_인구 수	...	2020 년09 월_계_연령 구간_인구 수	2020 년10 월_계_연령 구간_인구 수	2020 년11 월_계_연령 구간_인구 수
		성 산 읍	안 덕 면											
3	성 산 읍	4,642	4,653	4,668	4,682	4,683	4,703	4,718	4,744	4,770	...	5,248	5,275	5,285
4	안 덕 면	3,031	3,036	3,048	3,084	3,098	3,103	3,114	3,132	3,144	...	3,382	3,394	3,413

5 rows × 43 columns

```
In [48]: # 전처리를 위한 데이터프레임 Transform
resid_m = resid_m.set_index("행정구역")
resid_t = resid_m.T
resid_t = resid_t.reset_index()
resid_t["index"] = resid_t.apply(lambda x: x["index"][:7],axis = 1)
resid_t["index"] = resid_t.apply(lambda x: x["index"].replace("년","-"),axis = 1)
resid_t.head()
```

```
Out[48]: 행정구역      index      대정읍      남원읍      성산읍      안덕면      표선면      송산동      정방동      중앙동      천지동      효돈동      영천동      동홍동      서홍동

0  2018-01  43,508  5,330  5,554  4,642  3,031  3,333  1,298  696  1,140  1,032  1,503  1,514  3,898  1,857
1  2018-02  43,626  5,330  5,577  4,653  3,036  3,346  1,307  697  1,134  1,033  1,510  1,516  3,898  1,869
2  2018-03  43,792  5,347  5,583  4,668  3,048  3,364  1,309  697  1,144  1,039  1,514  1,516  3,915  1,877
3  2018-04  43,993  5,371  5,609  4,682  3,084  3,383  1,309  698  1,152  1,045  1,518  1,522  3,949  1,885
4  2018-05  44,152  5,398  5,624  4,683  3,098  3,391  1,309  706  1,158  1,049  1,531  1,532  3,961  1,898
```

```
In [49]: # 60세 이상 제주도민 데이터 삽입 함수 형성
def mm2(loca):
    temp = df2.loc[df2["location"]==loc,: ]
    temp = temp.reset_index()
    temp = temp.drop("index",axis = 1)
    temp.loc[:, "60세이상"] = resid_t.loc[:, loca]
    return temp
```

```
In [50]: # 기존 데이터 프레임 df2에 외부데이터 60세 이상 제주도민 데이터 삽입 함수 형성
df2["60세이상"] = 0
df_s = pd.DataFrame(columns=['y_m', 'city', 'location', 'resid_reg_pop_남', 'foreign_pop_남',
                             'resid_reg_pop_여', 'foreign_pop_여', 'total_pop', '60세이상'])
```

```
for i in df2.location.unique():
    df_s = pd.concat([df_s,mm2(i)])
```

In [51]: df_s.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 714 entries, 0 to 41
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   y_m                    714 non-null   object
1   city                   714 non-null   object
2   location               714 non-null   object
3   resid_reg_pop_남      714 non-null   object
4   foreign_pop_남        714 non-null   object
5   resid_reg_pop_여      714 non-null   object
6   foreign_pop_여        714 non-null   object
7   total_pop             714 non-null   object
8   60세이상              714 non-null   object
dtypes: object(9)
memory usage: 55.8+ KB
```

In [52]: # 자료형 데이터 정수형으로 변경
for i in range(3,(len(df_s.columns)-1)):
 df_s.iloc[:,i] = df_s.iloc[:,i].astype(int)

In [53]: # 자료형 데이터 정수형으로 변경
df_s["60세이상"] = df_s.apply(lambda x : x["60세이상"].replace(',',''), axis = 1)
df_s.loc[:, "60세이상"] = df_s.loc[:, "60세이상"].astype(int)
df_s.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 714 entries, 0 to 41
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   y_m                    714 non-null   object
1   city                   714 non-null   object
2   location               714 non-null   object
3   resid_reg_pop_남      714 non-null   int32
4   foreign_pop_남        714 non-null   int32
5   resid_reg_pop_여      714 non-null   int32
6   foreign_pop_여        714 non-null   int32
7   total_pop             714 non-null   int32
8   60세이상              714 non-null   int32
dtypes: int32(6), object(3)
memory usage: 39.0+ KB
```

In [54]: #열이름 정의
df_s.rename(columns={"resid_reg_pop_남" : "제주도민_남","resid_reg_pop_여" : "제주도민_여",
 ,"foreign_pop_여" : "외국인거주_여","60세이상" : "제주도민_60이상"}, i
df_s.head()

Out[54]:

	y_m	city	location	제주도민_남	외국인거주_남	제주도민_여	외국인거주_여	total_pop	제주도민_60이상
0	2018-01	서귀포시	남원읍	9806	428	9306	200	19740	5554
1	2018-02	서귀포시	남원읍	9821	430	9321	205	19777	5577

	y_m	city	location	제주도민_남	외국인거주_남	제주도민_여	외국인거주_여	total_pop	제주도민_60이상
2	2018-03	서귀포시	남원읍	9837	424	9341	204	19806	5583
3	2018-04	서귀포시	남원읍	9850	428	9353	205	19836	5609
4	2018-05	서귀포시	남원읍	9878	437	9374	201	19890	5624

```
In [55]: # 60세 미만 거주 인구수 파생변수 열 생성
df_s["제주도민_60미만"] = df_s.apply(lambda x: x.total_pop - x.제주도민_60이상, axis = 0)
df_s.head()
```

Out[55]:

	y_m	city	location	제주도민_남	외국인거주_남	제주도민_여	외국인거주_여	total_pop	제주도민_60이상	제주도민_60미만
0	2018-01	서귀포시	남원읍	9806	428	9306	200	19740	5554	14186
1	2018-02	서귀포시	남원읍	9821	430	9321	205	19777	5577	14200
2	2018-03	서귀포시	남원읍	9837	424	9341	204	19806	5583	14223
3	2018-04	서귀포시	남원읍	9850	428	9353	205	19836	5609	14227
4	2018-05	서귀포시	남원읍	9878	437	9374	201	19890	5624	14266

```
In [56]: # 열 순서 정의
df_s = df_s.loc[:,["y_m","city","location","제주도민_여","외국인거주_여","제주도민_남","제주도민_60미만","제주도민_60이상","total_pop"]]
```

```
In [57]: # csv파일 변환
df_s.to_csv("3번데이터_성별_연령_서귀포시.csv",encoding = "cp949",index=False)
```

제주시 서귀포시 결합

```
In [58]: #데이터로드
df_1 = pd.read_csv("3번데이터_성별_연령_제주시.csv", encoding = "cp949")
df_2 = pd.read_csv("3번데이터_성별_연령_서귀포시.csv", encoding = "cp949")
```

```
In [59]: # 데이터 결합
df_1 = pd.concat([df_1,df_2])
df_1
```

Out[59]:

	y_m	city	location	제주도민_여	외국인거주_여	제주도민_남	외국인거주_남	제주도민_60이상	제주도민_60미만	total_pop
0	2018-01	제주시	건입동	4715	82	5085	146	2478	7550	10028
1	2018-02	제주시	건입동	4724	82	5090	142	2484	7554	10038

	y_m	city	location	제주도 민_여	외국인거 주_여	제주도 민_남	외국인거 주_남	제주도민 _60이상	제주도민 _60미만	total_pop
2	2018-03	제주 시	건입동	4711	81	5072	141	2494	7511	10005
3	2018-04	제주 시	건입동	4718	86	5070	143	2484	7533	10017
4	2018-05	제주 시	건입동	4730	90	5070	148	2504	7534	10038
...
709	2021-02	서귀 포시	효돈동	2616	21	2673	44	1748	3606	5354
710	2021-03	서귀 포시	효돈동	2609	20	2653	41	1752	3571	5323
711	2021-04	서귀 포시	효돈동	2607	19	2660	40	1761	3565	5326
712	2021-05	서귀 포시	효돈동	2605	19	2661	41	1766	3560	5326
713	2021-06	서귀 포시	효돈동	2597	18	2660	39	1764	3550	5314

1806 rows × 10 columns

In [60]:

```
# csv파일 변환
df_1.to_csv("3번 데이터_전처리.csv", encoding = "cp949", index=False)
```