

⋮ 문제해결을 위한 데이터 분석 및 시각화

데이터 분석을 위한 Pandas(3)_기초통계

한성대학교 노은희 교수

“미래로 향하는 새로운 이정표”



오늘의 학습

학습내용

- 기초통계 이해

[파일 다운받기]

 서울글로벌센터 상담실적.csv

서울글로벌센터 상담실적.csv - Excel

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 개발 도구 파워 피벗 ? 수행할 작업을 알려 주세요

잘라내기 붙여넣기 복사 서식 복사 클립보드 글꼴 맞춤 표시 형식

맑은 고딕 11 가 가 텍스트 줄 바꿈 일반

가 가 간 가 나 가 나 병합하고 가운데 맞춤 % , .00 0.0

A1 X ✓ fx 년월

	A	B	C	D	E	F	G	H	I	J	K
1	년월	기관	전화	방문	이메일	게시판	화상상담	이동상담	합계		
2	202202	교육	128	11	1	0	0	0	140		
3	202202	교통	4	1	0	0	0	0	5		
4	202202	금융	17	3	0	0	0	0	20		
5	202202	기타	123	5	2	0	0	0	130		
6	202202	노무	348	16	1	0	0	0	365		
7	202202	문화/관광	6	1	0	0	0	0	7		
8	202202	보건/복지	462	43	9	0	0	0	514		
9	202202	비즈니스	14	2	0	0	0	0	16		
10	202202	안전	14	0	1	0	0	0	15		
11	202202	여성/가족	19	0	1	0	0	0	20		
12	202202	전년도 총	1807	195	33	0	0	0	2035		
13	202202	정보통신	11	2	0	0	0	0	13		
14	202202	주거	38	5	0	0	0	0	43		
15	202202	출입국	357	72	2	0	0	0	431		
16	202202	행정	266	34	16	0	0	0	316		
17	202201	교육	153	14	1	0	0	0	168		
18	202201	교통	4	2	0	0	0	0	6		
19	202201	금융	9	0	0	0	0	0	9		
20	202201	기타	136	12	4	0	0	0	152		

서울글로벌센터 상담실적

파일 불러오기 `read_csv()` 함수를 사용해서 csv 파일을 불러와서 DataFrame으로 저장

```
import pandas as pd
df = pd.read_csv('서울글로벌센터 상담실적.csv', encoding = 'cp949')
df
```

`encoding = 'cp949' => 한글 인코딩의 한 종류`

	년월	기관	전화	방문	이메일	게시판	화상상담	이동상담	합계
0	202202	교육	128	11	1	0	0	0	140
1	202202	교통	4	1	0	0	0	0	5
2	202202	금융	17	3	0	0	0	0	20
3	202202	기타	123	5	2	0	0	0	130
4	202202	노무	348	16	1	0	0	0	365
5	202202	문화/관광	6	1	0	0	0	0	7
6	202202	보건/복지	462	43	9	0	0	0	514
7	202202	비즈니스	14	2	0	0	0	0	16
8	202202	안전	14	0	1	0	0	0	15
9	202202	여성/가족	19	0	1	0	0	0	20
10	202202	전년도 총계	1807	195	33	0	0	0	2035

중간생략

	A	B	C	D	E	F	G	H	I	J	K
1	년월	기관	전화	방문	이메일	게시판	화상상담	이동상담	합계		
2	202202	교육	128	11	1	0	0	0	140		
3	202202	교통	4	1	0	0	0	0	5		
4	202202	금융	17	3	0	0	0	0	20		
5	202202	기타	123	5	2	0	0	0	130		
6	202202	노무	348	16	1	0	0	0	365		
7	202202	문화/관광	6	1	0	0	0	0	7		
8	202202	보건/복지	462	43	9	0	0	0	514		
9	202202	비즈니스	14	2	0	0	0	0	16		
10	202202	안전	14	0	1	0	0	0	15		
11	202202	여성/가족	19	0	1	0	0	0	20		
12	202202	전년도 총	1807	195	33	0	0	0	2035		
13	202202	정보통신	11	2	0	0	0	0	13		
14	202202	주거	38	5	0	0	0	0	43		
15	202202	출입국	357	72	2	0	0	0	431		
16	202202	행정	266	34	16	0	0	0	316		
17	202201	교육	153	14	1	0	0	0	168		
18	202201	교통	4	2	0	0	0	0	6		
19	202201	금융	9	0	0	0	0	0	9		
20	202201	기타	136	12	4	0	0	0	152		

'서울글로벌센터 상담실적.csv',

데이터프레임 자료형 보기

```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype  
---  --
 0   년월       30 non-null    int64  
 1   기관       30 non-null    object  
 2   전화       30 non-null    int64  
 3   방문       30 non-null    int64  
 4   이메일     30 non-null    int64  
 5   게시판     30 non-null    int64  
 6   화상상담   30 non-null    int64  
 7   이동상담   30 non-null    int64  
 8   합계       30 non-null    int64  
dtypes: int64(8), object(1)
memory usage: 2.2+ KB
```

null : 데이터가 비어 있는것(결측값)

non-null : 비어 있지 않는 데이터

object : 문자열

describe()로 통계 출력하기

```
1 df.describe() # 통계 모두 출력
```

	년월	전화	방문	이메일	게시판	화상상담	이동상담	합계
count	30.000000	30.000000	30.000000	30.000000	30.0	30.0	30.0	30.000000
mean	202201.500000	240.466667	26.666667	4.466667	0.0	0.0	0.0	271.600000
std	0.508548	451.645171	52.693868	8.950952	0.0	0.0	0.0	510.48203
min	202201.000000	1.000000	0.000000	0.000000	0.0	0.0	0.0	2.000000
25%	202201.000000	11.500000	1.250000	0.000000	0.0	0.0	0.0	13.500000
50%	202201.500000	30.500000	5.000000	1.000000	0.0	0.0	0.0	36.500000
75%	202202.000000	308.000000	15.750000	3.500000	0.0	0.0	0.0	335.500000
max	202202.000000	1807.000000	205.000000	34.000000	0.0	0.0	0.0	2039.000000

'전화'데이터만 보기

전화 데이터만 보기

```
1 first_row = df['전화']  
2 first_row
```

```
0    128  
1      4  
2     17  
3    123  
4    348  
5      6  
6    462  
7     14  
8     14  
9     19  
10   1807  
11     11  
12     38  
13    357  
14    266  
15    153  
16      4  
17      9  
18    136  
19    322  
20      5  
21    409  
22      1  
23     13  
24     16  
25   1800  
26      7  
27     23
```

```
28    445
```

```
29    257
```

```
Name: 전화, dtype: int64
```

'전화'데이터에 대한 min()/max()/mean()

```
1 first_row.min() # min() 최소값 구하기
```

```
1
```

```
1 first_row.max() # max() 최대값 구하기
```

```
1807
```

```
1 first_row.mean() # mean() 평균값 구하기
```

```
240.46666666666667
```


연도별로 그룹화하여 평균 구하기

Python pandas의 groupby() 연산자를 사용하여 집단, 그룹별로 데이터를 집계, 요약

groupby('기준이 될 컬럼명')

```
1 print(df.groupby('년월')['전화'].mean())
```

년월별로 전화상담 평균 구하기

```
년월
202201    240.000000
202202    240.933333
```

```
1 print(df.groupby('년월')['전화'].count())
```

년월별로 전화상담 일수구하기

```
년월
202201     15
202202     15
```

```
1 print(df.groupby('년월')['전화'].sum())
```

년월별로 전화상담 총횟수구하기

```
년월
202201    3600
202202    3614
Name: 전화, dtype: int64
```