

2주차-2

1. 오늘은 무엇을 배우는가?

2. Pandas

2.1. 패키지 import

2.2. 데이터 읽기, 인덱스 설정

2.3. 데이터 요약 : describe()

2.4. 데이터 정렬

2.5. 열 이름 바꾸기 rename(columns={"기존이름": "바꿀이름"})

2.6. 조건 인덱싱

2.7. 행, 열 인덱싱 나아가기 : 교통량.loc[1:3, ["A", "B"]]

1. 오늘은 무엇을 배우는가?

1. 패키지 import

2. 데이터 읽기, 인덱스 설정

3. 데이터 요약 : describe()

4. 데이터 정렬

a. 특정 열 : sort_values(by="열 이름")

b. 인덱스 기준 : sort_index()

5. 열 이름 바꾸기 rename(columns={"기존이름": "바꿀이름"})

6. 조건 인덱싱

a. 숫자 범위로 인덱싱 : 교통량[교통량["2021년 상행"] > 2000]

b. 문자로 인덱싱 : str.contains("찾으려는 문자")

7. 행, 열 인덱싱 나아가기 : 교통량.loc[1:3, ["A", "B"]]

2. Pandas

2.1. 패키지 import

- import를 통해 외부 라이브러리의 함수를 활용할 수 있습니다.

```
import pandas as pd
```

2.2. 데이터 읽기, 인덱스 설정

- 인덱스는 파일을 빠르게 검색하는 데 필수적입니다.

```
# 데이터를 읽고, 인덱스를 설정합니다.  
# 인덱스는 파일을 빠르게 검색하는 데 필수적입니다.  
  
교통량 = pd.read_excel("교통량.xlsx")  
교통량 = 교통량.set_index("번호", drop=True)
```

2.3. 데이터 요약 : describe()

- describe() 함수는 데이터에 대한 간단한 요약을 보여줍니다.

```
교통량.describe()
```

	2020년 상행	2020년 하행	2021년 상행	2021년 하행
count	25.00	25.00	25.00	25.00
mean	4546.24	4505.36	4444.36	4134.12
std	9322.65	9130.24	8680.81	7446.61
min	32.00	28.00	49.00	36.00
25%	802.00	783.00	840.00	812.00
50%	1408.00	1557.00	1267.00	1722.00
75%	3729.00	4020.00	3686.00	3927.00
max	46477.00	45154.00	42871.00	36557.00

2.4. 데이터 정렬

- sort_values(by="열 이름")을 통해 특정 열을 기준으로 정렬할 수 있습니다.
출력을 보면 2021년 하행 기준 가장 작은 값이 제일 먼저 나오는 결과를 확인할 수 있습니다.

```
교통량_정렬 = 교통량.sort_values(by="2021년 하행") # 2021년 하행선 교통량 기준으로 정렬합니다.
```

	구분	관측지점	2020년 상행	2020년 하행	2021년 상행	2021년 하행
번호						
18	군도4호	구산면 유산리 92-5	32	28	49	36
17	군도3호	구산면 신촌삼거리	314	268	451	444
19	군도5호	구산면 수정리 359-6	1250	1181	492	533
24	군도16호	진전면 고사리 753-8	581	549	691	721
7	지방도	진전면 시락	780	765	885	774

- `sort_index()`를 활용하여 이전에 설정해둔 값을 기준으로 정렬할 수 있습니다.
`sort_index()`를 통해 인덱스(이 데이터에서는 '번호')를 기준으로 정렬한 결과를 돌려줍니다.

```
교통량_정렬.sort_index()
```

	구분	관측지점	2020년 상행	2020년 하행	2021년 상행	2021년 하행
번호						
1	국지도	내서읍 용담리 162-2	13227	15259	13761	12874
2	국지도	의창구 등을 신방리	6452	7346	7355	6987
3	국지도	북면 내곡 탐앤탐스	3729	4463	3686	3927
4	국지도	북면 신촌리 농기계센터 앞	5745	5672	6856	5785
5	국지도	북면 신촌리 북면운동장 앞	4733	4020	5492	5783

2.5. 열 이름 바꾸기 `rename(columns={"기존이름":"바꿀이름"})`

- 인터넷에서 받은 데이터의 경우 데이터의 열 이름을 이해하기 편하도록 바꿀 일이 생깁니다.

예) 영어 열 이름을 한글로 바꾸기, 괄호가 들어간 열 이름을 간단한 이름으로 바꾸기

```
교통량 = 교통량.rename(columns={"관측지점":"장소"}) # 관측지점이라는 열 이름을 장소라는 이름으로 바꿉니다.
```

구분

장소

국지도

내서읍 용담리 162-2

국지도

의창구 동읍 신방리

2.6. 조건 인덱싱

- 데이터[데이터[열 이름] > 원하는 값]의 형태로 범위에 따른 인덱싱이 가능합니다.

```
교통량[교통량["2021년 상행"] > 2000] # 2021년 상행선 교통량 2000 이상만 출력합니다.
```

	구분	장소	2020년 상행	2020년 하행	2021년 상행	2021년 하행
번호						
1	국지도	내서읍 용담리 162-2	13227	15259	13761	12874
2	국지도	의창구 동읍 신방리	6452	7346	7355	6987
3	국지도	북면 내곡 탐앤탐스	3729	4463	3686	3927
4	국지도	북면 신촌리 농기계센터 앞	5745	5672	6856	5785
5	국지도	북면 신촌리 북면운동장 앞	4733	4020	5492	5783
6	지방도	진동면 요장리	3013	2792	1940	2474
8	지방도	성산구 불모산동	46477	45154	42871	36557
13	군도6호	동읍 다호리 65-22	3042	3007	2817	2562
16	군도1호	내서읍 평성리 661-4	10833	8727	9835	9480

- str.contains("찾으려는 문자")의 형태로 문자열을 포함하는 데이터만 볼 수 있습니다.
예) 특정 지역만 찾기

```
교통량[교통량["장소"].str.contains("북면")]
```

구분	번호	장소	2020년 상행	2020년 하행	2021년 상행	2021년 하행
3	국지도	북면 내곡 탐앤탐스	3729	4463	3686	3927
4	국지도	북면 신촌리 농기계센터 앞	5745	5672	6856	5785
5	국지도	북면 신촌리 북면운동장 앞	4733	4020	5492	5783
9	지방도	진북면 지산리	1807	1557	2039	1765
14	군도8호	북면 감계리 189-2	1602	2168	1811	1861
15	군도10호	북면 신촌리 402-13	1927	1690	2068	1722
21	군도10호	진북면 신촌리 신촌마을길 2	661	647	804	784
22	군도11호	진북면 예곡리 의림로 99	705	696	840	812

2.7. 행, 열 인덱싱 나아가기 : 교통량.loc[1:3, ["A", "B"]]

- 저번 시간에 열 데이터는 데이터["열 이름"]의 형태로 접근하고, 행은 데이터.loc["인덱스"]의 형태로 접근할 수 있다고 배웠습니다.
- .loc 속성을 활용하면 특정 행 뿐만 아니라 행, 열을 함께 고를 수 있습니다.

```
교통량.loc[1, "장소"] # 단 하나만 고릅니다.
```

```
# 7. 행, 열 인덱싱 나아가기 : 교통량.loc[1:3, ["A", "B"]]
```

```
교통량.loc[1, "장소"] # "장소"에 대괄호 포함하지 않음!
```

```
'내서읍 용담리 162-2'
```

```
교통량.loc[1] # 인덱스 1에 대해 가로 줄(행 한 줄 전체)을 선택합니다.
```

```
교통량.loc[1]
```

```
구분          국지도  
장소          내서읍 용담리 162-2  
2020년 상행    13227  
2020년 하행    15259  
2021년 상행    13761  
2021년 하행    12874  
Name: 1, dtype: object
```

교통량.loc[1:2] # 인덱스 1에서 인덱스 2까지에 대해 가로 줄(행 전체)을 선택합니다.

```
교통량.loc[1:2]
```

	구분	장소	2020년 상행	2020년 하행	2021년 상행	2021년 하행
번호						
1	국지도	내서읍 용담리 162-2	13227	15259	13761	12874
2	국지도	의창구 동읍 신방리	6452	7346	7355	6987

교통량.loc[1:2, ["장소"]] # 인덱스 1에서 인덱스 2까지에 대해 장소 열만 선택합니다.

```
교통량.loc[1:2, ["장소"]]
```

장소

번호

```
1  내서읍 용담리 162-2  
2  의창구 동읍 신방리
```

교통량.loc[1:2, ["장소", "2020년 상행"]] # 인덱스 1에서 인덱스 2까지에 대해 장소, 2020년 상행선 데이터를 선택합니다.

```
교통량.loc[1:2, ["장소", "2020년 상행"]]
```

장소 2020년 상행

번호

1	내서읍 용담리 162-2	13227
2	의창구 동읍 신방리	6452