



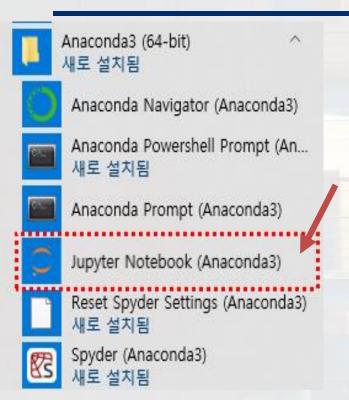
### 오늘의 학습

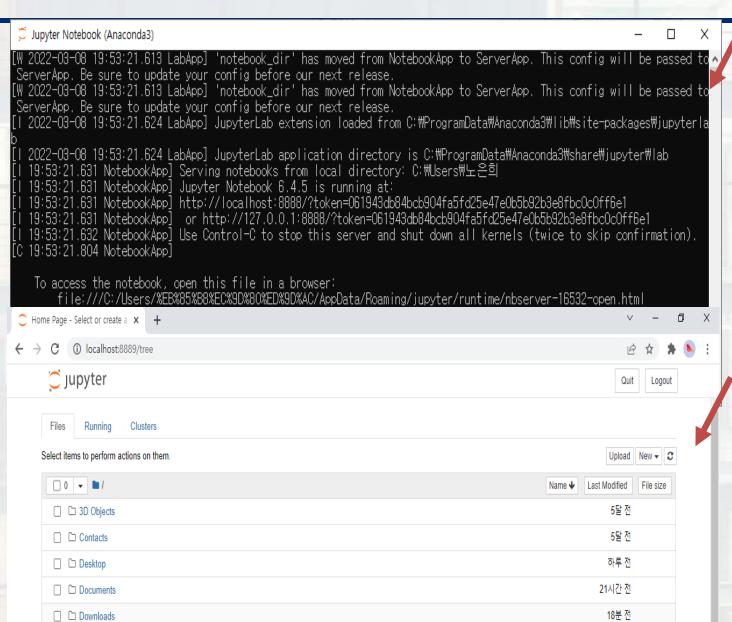
### 학습내용

- 리스트, 딕셔너리 자료형 이해 데이터 분석을 위한 라이브러리 Pandas



### Jupyter Notebook(주피터 노트북) 실행하기

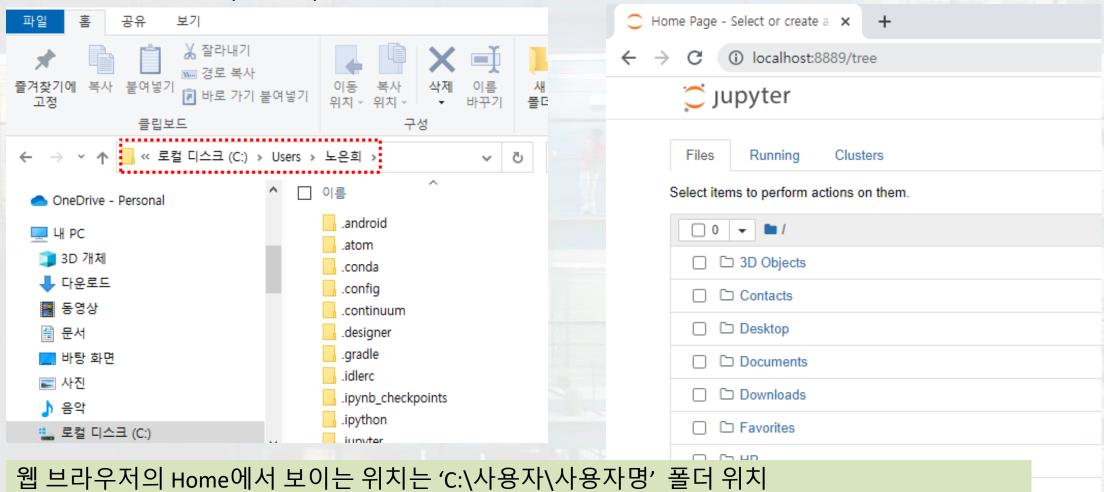






### Jupyter Notebook(주피터 노트북)에서 코딩하기

#### 시작폴더는 탐색기 c:\사용자\사용자명



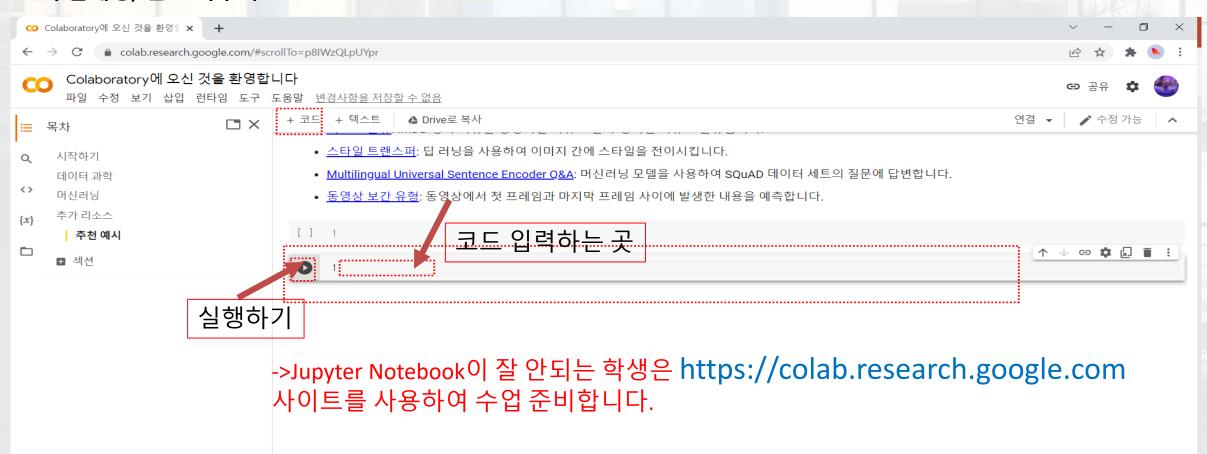
웹 브라우저의 Home에서 보이는 위치는 'C:\사용자\사용자명' 폴더 위치 웹브라우저의 home에서 새로운 노트북을 생성하면 'C:\사용자\사용자명' 폴더 위치에 저장됨



### 코랩에서 시작하기 https://colab.research.google.com

구글 코랩(Google Colaboratory)은 구글에서 제공하는 Jupyter Notebook 구글 클라우드를 기반으로 하기 때문에 PC보다 좋은 성능 및 접근성을 제공하나 보안등의 이유로 접속이 차단된 곳에서는 사용 못할 수도 있음 [준비]

- 구글계정, 웹 브라우저





# 리스트와 리스트 인덱싱

리스트를 만들 때는 대괄호([])로 감싸 주고 각 요솟값은 쉼표(,)로 구분

- 리스트 인덱싱과 슬라이싱
  - 리스트 인덱싱
    - 인덱스(index) : 리스트 항목에 붙어있는 번호, 0부터 번호 시작
  - "red","green","blue" 문자열 데이터를 갖는 colors\_list 생성하기

```
colors_list = ["red","green","blue"]

리스트의 각 항목 접근시
인덱스 번호 사용

red"

green"

"green"

"blue"
```

>>> colors\_list[0] red

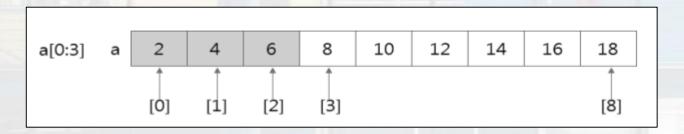
## 리스트 슬라이싱

- 리스트 슬라이싱
  - []안에 인덱스로 범위를 지정하여 리스트의 일부를 잘라줌

### 형식

리스트[시작 인덱스:끝인덱스]

끝인덱스는 가져오려는 내용보다 1 크게 지정



[2, 4, 6, 8, 10, 12, 14, 16, 18]

[2, 4, 6]

a = [2,4,6,8,10,12,14,16,18]
print(a)
print(a[0:3])

가져오려는 끝 인덱스보다 1 크게 지정

>>> a[:3] ← 0부터 2까지
[3, 6, 9]
>>> a[3:] ← 3부터 끝까지
[12, 15, 18, 27]
>>> a[:] ← 0부터 끝까지

[3, 6, 9, 12, 15, 18, 27]

>>> a = [3,6,9,12,15,18,21,27]

### 딕셔너리

- 딕셔너리(dictionary)
  - 중괄호 {}로 묶어서 구성
  - 키(key)와 값(value)의 순서쌍을 갖는 자료형
  - 각각의 요소는 Key : Value 형태로 이루어져 있고 쉼표(,)로 구분
  - Key에는 변하지 않는 값을 사용하고, Value에는 변하는 값과 변하지 않는 값 모두 사용

형식 {Key1:Value1, Key2:Value2, Key3:Value3, ...}

예

```
a = \{ \text{'name':'Hong', 'phone':'01012345678', 'birth': '020918'} \}

a = \{ \text{'a': } [1,2,3,4,5] \}
```



### ■ [실습] Jupyter Notebook 환경->리스트와 딕셔너리

### 리스트 실습

```
In [1]: 1 a=[1,3,5,7,9]
2 a
```

Out[1]: [1, 3, 5, 7, 9]

In [2]: 1 a[0]

Out [2]: 1

In [3]: 1 a[1:3]

Out [3]: [3, 5]

In [4]: 1 a[-1]

Out[4]: 9

### 딕셔너리 실습

```
In [5]: | 1 | b = ('name' 'Hong', 'phone' '01012345678', 'birth' '020918')
Out[5]: {'name': 'Hong', 'phone': '01012345678', 'birth': '020918'}
In [6]: 1 b['hobby'] = "독서"
Out[6]: {'name': 'Hong', 'phone': '01012345678', 'birth': '020918', 'hobby': '독서'}
In [7]: 1 b['name']
Out [7]: 'Hong'
```



### ┗ [실습] 코랩(Colab)환경->리스트와 딕셔너리

#### 리스트 실습

```
[4] 1 a=[1,3,5,7,9]
2 a
[1, 3, 5, 7, 9]
```

- 1 a[0]
- 8
- [6] 1 a[1:3]
  - [3, 5]
- [7] 1 a[-1]

#### 딕셔너리 실습

'Hong'



## Python과 Pandas 자료형 비교

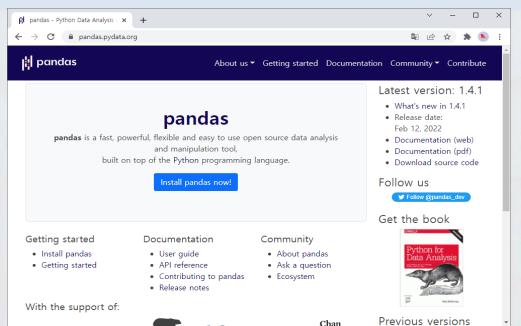
Python 자료형	Pandas 자료형	설명
string	object	문자열
int	int64	정수
float	float64	실수형



### 데이터 분석을 위한 주요 라이브러리

#### Pandas

- 데이터 분석에서 자주 사용하는 테이블 형태를 다룰 수 있는 라이브러리
- 데이터 분석과 처리를 쉽게 할 수 있게 도와주는 라이브러리
- TensorFlow 및 Scikit-Learn, Matplotlib 등 데이터 분석 및 시각화 기능을 수행하기 위해 필수적으로 사용되는 라이브러리
- 아나콘다 배포판에 포함되어 있음
- pandas 홈페이지: <a href="http://pandas.pydata.org">http://pandas.pydata.org</a>
- Pandas에서는 기본적으로 정의되는 자료구조인 Series와 Data Frame을 사용

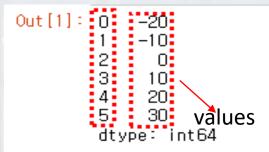




### Pandas 자료구조

1차원 자료구조 : Series 인덱스와 값을 함께 저장하는 1차원 데이터 구조





2차원 자료구조 : DataFrame 인덱스와 컬럼으로 구성된 2차원 데이터 구조



columns

thon 데이터문석
100 80
88 90
90 100
95 80
85 88
1



### 데이터 분석을 위한 주요 라이브러리\_Pandas

#### Pandas

Pandas 임포트 import pandas as pd

Series 데이터 생성하기

s = pd.Series(seq\_data)

DataFrame 데이터 생성하기

df = pd.DataFrame(data [, index = index\_data, columns = columns\_data])

[참고] seq\_data: 시퀀스란 데이터를 순서대로 하나씩 나열하여 나타낸 데이터 구조로 시퀀스의 각 요소에는 특정 위치(~번째)의 데이터를 가리키는 인덱스가 지정됩니다. 첫 번째 인덱스는 0이고, 두 번째 인덱스는 1 (자료구조 예>리스트, 튜플, 문자열)



#### **Pandas**

### 1차원 자료구조 : Series

1차원 자료구조 : Series

- 리스트를 원소를 생성하는 1차원 자료구조
- Series의 원소 인덱스는 0부터 시작하는 정수로 인덱스가 자동부여
- 세로축 라벨을 index, 입력한 시퀀스 데이터를 values라고 함

1차원 자료구조 : Series

s = pd.Series(seq\_data)

#### 리스트

- 여러 개의 데이터를 하나로 묶어서 처리할 수 있도록 해주는 자료형
- 리스트 만들기
  - 비어있는 리스트 만들 경우 []만 지정
  - 항목(item)들을 쉼표(,)로 분리하여 대괄호([])안에 넣기



### Series로 데이터 생성하기[실습]

### 1차원 자료구조 : Series 자료형

Series 생성은 pd.Series함수로 하며 ([ ]) 괄호 형태로 원소들을 넣어준다. Series를 불러오면 원소들과 함께 0,1,2,3이라는 번호가 앞에 매겨진다.

#### Series 객체 생성

예> 1월부터 6월까지 평균 데이터(-20, -10, 0, 10, 20, 30)

In [2]: 1 s = pd.Series([-20,-10,0,10,20,30])
2 s

Out [2]: 0 -20 1 -10 2 0 3 10 4 20 5 30

dtype: int64

#### Series 객체 접근하기

In [3]: 1 s[0] Out[3]: -20

In [4]: 1 s[2]

Out [4]: 0



### Series로 데이터 생성하기[실습]

Series 함수는 인덱스와 자료형 등을 같이 지정할 수 있는데 각각 []괄호를 쓴다.

### Series 객체 생성(Index 지정)

```
In [5]: 1 s = pd.Series([-20,-10,0,10,20,30],index=['1월','2월','3월','4월','5월','6월'])

1월 -20
2월 -10
3월 0
4월 10
5월 20
6월 30
dtype: int64

In [6]: 1 s['3월'] # index 3월에 해당하는 데이터 출력

Out[6]: 0
```



### Series로 데이터 생성하기[실습]

Series 데이터 생성

- 딕셔너리 데이터를 사용하면 index와 value 값을 한꺼번에 입력 가능

```
In [7]: 1 s = pd.Series({'1월':-20,'2월':-10,'3월':0,'4월':10,'5월':20,'6월':30})

1월 -20
2월 -10
3월 0
4월 10
5월 20
6월 30
dtype: int64
```



### **DataFrame**

### 2차원 자료구조 : DataFrame

- 가장 널리 사용되는 자료구조로 행과 열이 있는 테이블 형태의 2차원 구조
- 자동 생성된 index, columns를 갖는 DataFrame 데이터
- 엑셀의 스프레드시트와 유사

df = pd.DataFrame(data [, index = index\_data, columns = columns\_data])

data: 시퀀스 타입의 데이터(리스트, 딕셔너리, NumPy 배열, Series, DataFrame) 입력

	이름	학과	Al	Python	데이터분석	→ columns
0	홍길동	웹공학	90	100	80	
1	이사랑	전자	85	88	90	
2	강백호	컴퓨터	100	90	100	
3	김지수	법	90	95	80	- 37
4	박희수	역사문화	90	85	88	- 8

	Α	В	С	D	Е
1	사용년월	노선번호	노선명	표준버스정류장![	버스정류장ARS번
2	202201	741	741번(진관차고지~헌인릉	100000001	1001
3	202201	N37	N37번(진관공영차고지~쇰	100000001	1001
4	202201	470	470번(상암차고지~안골미	100000001	1001
5	202201	N37	N37번(송파공영차고지~진	100000001	1001
6	202201	100	100번(하계동~용산구청)	100000002	1002
7	202201	107	107번(민락동차고지~동대	100000002	1002
8	202201	104	104번(강북청소년수련관년	100000002	1002
9	202201	171	171번(용마문화복지센터~	100000002	1002
10	202201	172	172번(하계동~월드컵2.3단	100000002	1002

index



### [실습하기] 딕셔너리 데이터 만들기

- 딕셔너리(dictionary)
  - · 중괄호 {}로 묶어서 구성
  - 키(key)와 값(value)의 순서쌍을 갖는 자료형

```
data = {
                  '이름': ['홍길동','이사랑','강백호','김지수','박희수'],
'학과': ['웹공학','전자','컴퓨터','법','역사문화'],
'Al': [90, 85, 100, 90, 90],
                   'Python' : [100, 88, 90, 95, 85],
                   '데이터분석': [80, 90, 100, 80, 88]
           8 data
Out[1]: {'이름': ['홍길동', '이사랑', '강백호', '김지수', '박희수'], '학과': ['웹공학', '전자', '컴퓨터', '법', '역사문화'],
          'Al': [90, 85, 100, 90, 90],
          'Python': [100, 88, 90, 95, 85],
          '데이터분석': [80, 90, 100, 80, 88]}
In [2]: 1 type(data)
Out [2]: dict
          1 data['0|름']
Out[3]: ['홍길동', '이사랑', '강백호', '김지수', '박희수']
In [4]: 1 data['Al']
Out[4]: [90, 85, 100, 90, 90]
```

```
덕셔너리

data = {
  '이름': ['홍길동','이사랑','강백호','김지수','박희수'],
  '학과': ['웹공학','전자','컴퓨터','법','역사문화'],
  'Al': [90, 85, 100, 90, 90],
  'Python': [100, 88, 90, 95, 85],
  '데이터분석': [80, 90, 100, 80, 88]
}
data
```

딕셔너리 변수 = {Key1:Value1, Key2:Value2, Key3:Value3, …}



### DataFrame 객체 생성

### 2차원 자료구조 : DataFrame

import pandas as pd
df = pd.DataFrame(data)

df = pd.DataFrame(data [, index = index\_data, columns = columns\_data])

data : 시퀀스 타입의 데이터(리스트, 딕셔너리, NumPy 배열, Series, DataFrame) 입력

#### DataFrame 객체 생성

In [5]: 1 import pandas as pd 2 df = pd.DataFrame(data)

In [6]: 1 dt

Out [6]:

	이름	학과	ΑI	Python	데이터분석
0	홍길동	웹공학	90	100	80
1	이사랑	전자	85	88	90
2	강백호	컴퓨터	100	90	100
3	김지수	뜝	90	95	80
4	박희수	역사문화	90	85	88



### DataFrame 데이터 접근

### 2차원 자료구조 : DataFrame

### 데이터 접근

```
In [7]:
             df['이름']
        0 홍길동
1 이사랑
2 강백호
3 김지수
4 박희수
Name: 이름, dtype: object
Out [7]: 0
In [8]:
             df[['이름','학과']]
Out[8]:
              이름
                      학과
                    웹공학
          1 이사랑
                      전자
         2 강백호
                     컴퓨터
          3 김지수
```



### DataFrame 객체 생성(index지정)

### DataFrame 객체 생성(Index 지정)

```
In [9]: 1 df = pd.DataFrame(data, index=['1번','2번','3번','4번','5번'])
2 df
```

#### Out [9]:

	이름	학과	AI	Python	데이터분석
1번	홍길동	웹공학	90	100	80
2번	이사랑	전자	85	88	90
3번	강백호	컴퓨터	100	90	100
4번	김지수	H	90	95	80
5번	박희수	역사문화	90	85	88

```
In [10]: 1 df.index
```

Out[10]: Index(['1번', '2번', '3번', '4번', '5번'], dtype='object')



### DataFrame 객체 생성(columns지정)

### DataFrame 객체 생성(Columns 지정)

data중에서 원하는 column만 선택하거나, 순서 변경 지정

```
In [11]: 1 df = pd.DataFrame(data, columns = ['이름','학과','데이터분석']) 2 df
```

#### Out [11]:

	이름	학과	데이터분석
0	홍길동	웹공학	80
1	이사랑	전자	90
2	강백호	컴퓨터	100
3	김지수	11	80
4	박희수	역사문화	88

In [12]: 1 df = pd.DataFrame(data, columns = ['이름','데이터분석','학과']) 2 df

#### Out [12]:

학과	데이터분석	이름	
웹공학	80	홍길동	0
전자	90	이사랑	1
컴퓨터	100	강백호	2
법	80	김지수	3
역사문화	88	박희수	4

# csv파일살펴보기

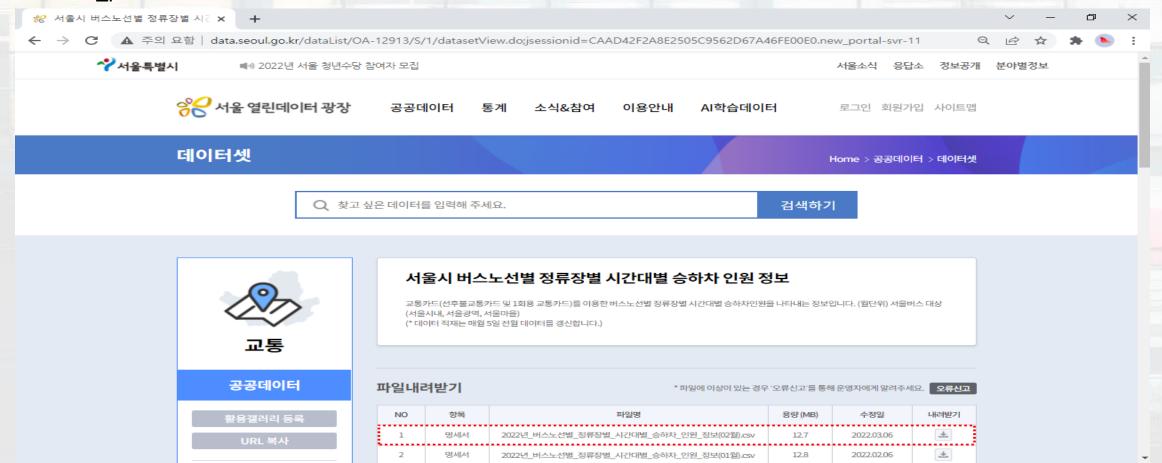


- CSV파일은 각각의 데이터값을 콤마(,)로 구분하는 파일 형식
- 정부에서 운영하는 공공데이터포털(www.data.go.kr)이 제공하는 일반적인 파일 형식
- 데이터 분석 전문가들이 자주 사용하는 형식
- CSV파일은 엑셀파일처럼 사용이 가능하고 엑셀과 메모장에서 파일을 열어 수정가능



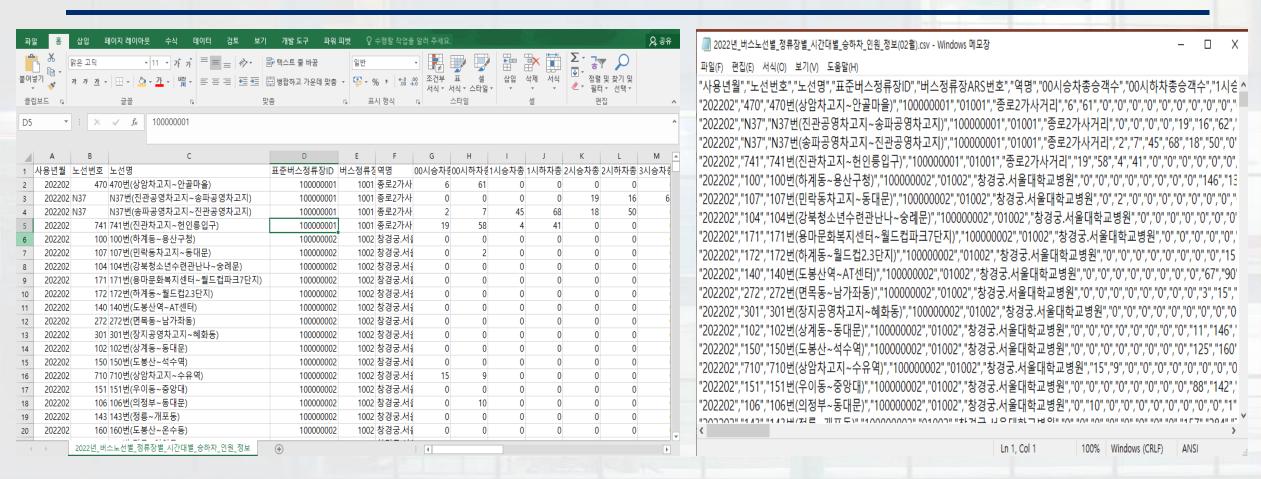
### [자료 다운받기] 🖾 2022년\_버스노선별\_정류장별\_시간대별\_승하차\_인원\_정보(02월).csv

http://data.seoul.go.kr/dataList/OA-12913/S/1/datasetView.do;jsessionid=CAAD42F2A8E2505C9562D67A46FE00 E0.new\_portal-svr-11





### 다운받은 자료 확인 🛍 2022년\_버스노선별\_정류장별\_시간대별\_승하차\_인원\_정보(02월).csv

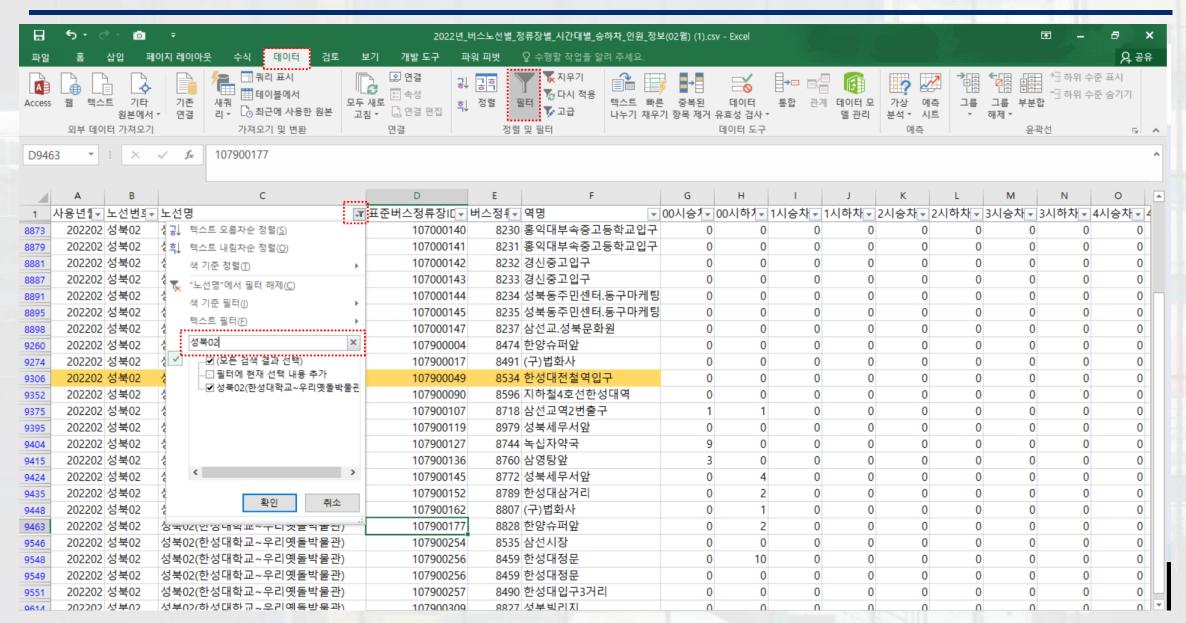


CSV파일을 엑셀에서 열어 수정 가능

CSV파일을 메모장에서 파일을 열어 수정가능



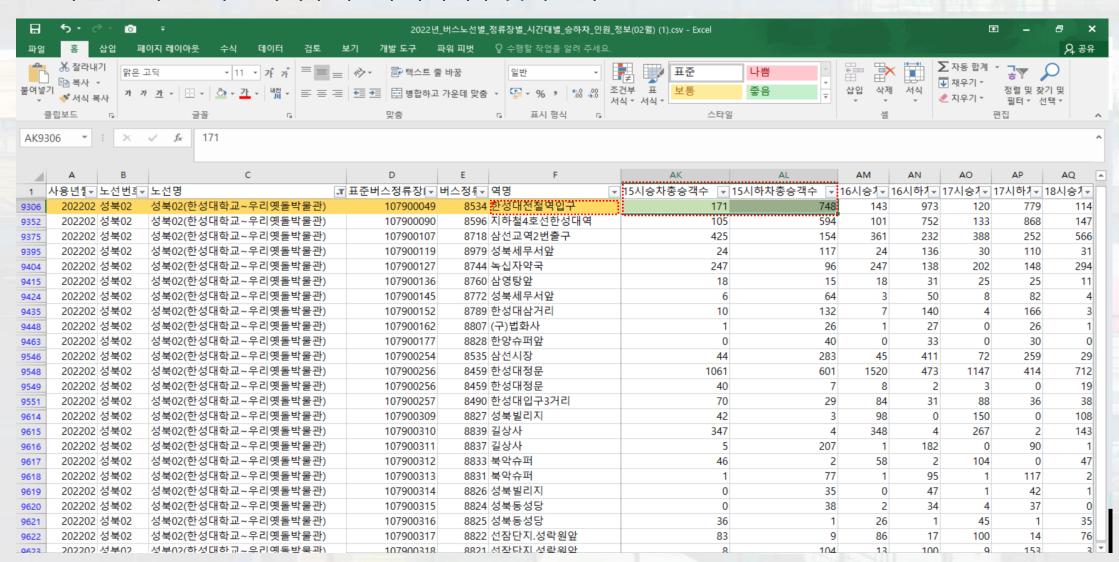
### 엑셀에서 '성북02'버스 한성대전철역입구 시간대별 승하차인원 조회하기





### 엑셀에서 '성북02'버스 한성대전철역입구 시간대별 승하차인원 조회하기

'성북02'번 버스 3시 승차자수와 3시 하차자수 찾아보기





### 서울 성북02번 마을버스 노선

※ 서울 성북02번 마을 버스노선 시작 한성대정문 → 한양슈퍼앞 → 법화사 → 한성대입구3거리 → 성북세무서앞 → 삼선시장 → 한성대전철 역입구 → 삼선교.성북문화원 → 성북동주민센터.동구마케팅고 → 경신중고입구 → 홍익대부속중고등 학교입구 → 선잠단지.성락원앞 → 성북동성당 → 선잠로3길입구 → 성북빌리지 → 북악슈퍼 → 길상 사 → 한국가구박물관입구 → 우리옛돌박물관.정법사 → 한국가구박물관입구 → 길상사 → 북악슈퍼 → 성북빌리지 → 선잠로3길입구 → 성북동성당 → 선잠단지.성락원앞 → 홍익대부속중고등학교입구 → 경신중고입구 → 성북동주민센터.동구마케팅고 → 지하철4호선한성대역 → 삼선교역2번출구 → 녹 십자약국 → 삼영탕앞 → 성북세무서앞 → 한성대삼거리 → 법화사 → 한양슈퍼앞 → 한성대정문 → ※ 서울 성북02번 마을 버스노선 끝