
▣ Ch17 데이터 분석을 위한 데이터 구조 다루기

Ch16과 Ch17 내용을 함께 다룬다. (이미 Python을 다룰 줄 아는 학생들을 대상으로 진행하기 때문에)

- > Python의 데이터 구조 형식 중에 데이터 분석에 관련된 쓰임새 위주로 살펴본다.
- > 하나의 값 만을 저장하는 스칼라(Scalar) 형 int, float, str, bool 자료형
- > 여러 개의 값을 사용할 수 있는 비스칼라 형 list, tuple, dictionary 자료형
- > 외장 Pandas 자료형인 series, data frame 자료형

■ Python 자료 구조 종류

[내장 자료형]

- >> 스칼라(Scalar) 형
- >> 리스트(list) 형
- >> 튜플(Tuple) 형
- >> 딕셔너리(Dictionary) 형

[외장 자료형] Pandas 자료형

- << 시리즈(Series) 형
- << 데이터 프레임(Data Frame) 형

In []:

▣ 스칼라(Scalar) 형

- > 하나의 값 만으로 구성된 자료 구조
 - > Data Type : int, float, str, bool 등
 - > 관련 함수 : int(), float(), str(), bool()
- <> 기본 자료형

```
In [10]: ## 스칼라(Scalar) 형 ##
## 생성 및 초기화
a = 3                      #int형 자료 구조 생성 및 초기화
print(type(a))              #자료구조 확인
b = 4.15                     #float형 자료 구조 생성 및 초기화
print(type(b))
c = 'hello, world!'         #str형 자료 구조 생성 및 초기화
print(type(c))

<class 'int'>
<class 'float'>
<class 'str'>
```

○ Python에서는 자료 구조가 모두 **class** 형태로 만들어진다.

○ 따라서 자료형에 대해 메서드들이 제공된다.

<> 데이터 타입 확인하기

```
In [19]: ## 타입 확인하기
a = 3.14
type(a)
```

```
Out[19]: float
```

```
In [20]: ## 타입 확인하기
a = 3.14
if type(a) is float:
    print('>> Float type')

>> Float type
```

<> 데이터 타입 바꾸기

```
In [13]: ## 타입 바꾸기: int(), float(), str()
a = 3.14
print(type(a), a)
b = int(a)                  # float > int
print(type(b), b)
c = str(b)                  # int > str
print(type(c), c)
d = float(c)                # str > float
print(type(d), d)
e = str(d)                  # float > str
print(type(a), e)

<class 'float'> 3.14
<class 'int'> 3
<class 'str'> 3
<class 'float'> 3.0
<class 'float'> 3.0
```

<> 자료형의 메서드 사용하기

```
In [127...]: c.upper()           # 'HELLO, WORLD!'
Out[127]:   'HELLO, WORLD!'
```

```
In [128...]: c.capitalize()      # 'Hello, world!'
```

```
Out[128]: 'Hello, world!'
```

```
In [129...]: c.title() # 'Hello, World!'
```

```
Out[129]: 'Hello, World!'
```

```
In [130...]: c.replace('!', '!!!') # 'Hello, World!!!'
```

```
Out[130]: 'hello, world!!!'
```

```
In [131...]: c.find(',') # 5
```

```
Out[131]: 5
```

```
In [132...]: c.split(',') # ['Hello', 'World!']
```

```
Out[132]: ['Hello', 'World!']
```

```
In [ ]:
```

▣ 비 스칼라 형

> 여러 개의 값을 사용할 수 있도록 구성된 자료 구조

> Data Type : list, tuple, dictionary

> 관련 함수 : list(), tuple(), dict()

```
In [ ]:
```

○ 리스트(List) 형 : []

> 여러 개의 값들을 나열된 자료 구조

> 기본적으로 순서번호(Index)로 값에 접근

> 값의 삽입, 삭제가 가능하다.

> 각 값들은 [와] 안에 ,로 구분하여 초기화 한다.

<> 추출하기

```
In [15]: ## 리스트(List) 형 : [ ] ##
```

```
## 생성 및 검색
```

```
a = [1, 2, 'a', 'b'] #리스트 생성 및 초기화
print(a[0]) #리스트는 0번째부터 시작 > a[0]는 한 개 값
print(a[-1]) #거꾸로 첫 번째
print(a[2:3]) #2번째부터 3번째 전까지 > a[2:3]는 한 개 값이지만 범위이므로 결과는 룩
print(a[2:]) #2번째부터 끝까지 > 결과는 룩은 값 list형
print(a[:2]) #처음부터 2번째 전까지 > 결과는 룩은 값 list형
```

```
1  
b  
['a']  
['a', 'b']  
[1, 2]
```

<> 수정 조작하기

```
In [16]: ## 값 변경  
a[2] = 'A'  
a
```

```
Out[16]: [1, 2, 'A', 'b']
```

관련 메서드

> **append()** 항목 값 추가하기

> **remove()** 항목 값으로 제거하기

> **insert(index, value)** index로 항목 추가하기

> **index()** 항목 값으로 index 값 얻기

> **sort()** 항목을 정렬하여 바꾸기 (매개변수 **reverse=False**가 기본값)

```
In [2]: ## 생성 및 조작  
a = [6, 1, 0, 3, 4, 5, 6]  
a.append(2)          # 맨 마지막에 2를 추가  
a.insert(2, 9)      # 2번째에 값 9를 추가  
a.remove(9)         # 3을 찾아 제거  
a.index(0)          # 'a'의 위치 값 반환  
a.sort(reverse=False) # 오름차순으로 정렬  
a
```

```
Out[2]: [0, 1, 2, 3, 4, 5, 6, 6]
```

관련 함수

> **len()** 항목 개수 구하기

> **sorted()** 항목을 정렬하여 리스트를 반환(원 리스트는 변화 없음)

> **del()** 특정 항목 제거하기

> **clear()** 모든 항목 제거하기

```
In [4]: ## 관련 함수: 정렬  
a = [5, 3, 7, 8, 2, 1]  
asort = sorted(a)    # 값들을 정렬  
asort
```

```
Out[4]: [1, 2, 3, 5, 7, 8]
```

```
In [3]: a = [5, 3, 7, 8, 2, 1]  
a.sort() # 리스트 자체를 정렬  
a
```

```
Out[3]: [1, 2, 3, 5, 7, 8]
```

```
In [16]: ## 관련 함수: 제거, 갯 수  
del(a[1])           #1번째 값을 제거  
print(a)  
len(a)  
  
[5, 8, 2, 1]  
Out[16]: 4
```

<> 1, 2차원 리스트 초기화 하기

```
In [4]: ## 1차원 리스트 초기화 방법  
arr = [0, 0, 0, 0, 0]  
arr
```

```
Out[4]: [0, 0, 0, 0, 0]
```

```
In [5]: ## 1차원 리스트 초기화 방법  
arr = [0]*5;  
arr
```

```
Out[5]: [0, 0, 0, 0, 0]
```

```
In [68]: ## 1차원 리스트 초기화 방법  
arr = [0 for i in range(5)];      #[ ]는 list()와 같은 역할  
arr = list(0 for i in range(5));  
arr
```

```
Out[68]: [0, 0, 0, 0, 0]
```

```
In [7]: ## 1차원 리스트 초기화 방법  
arr=[];          #만들고  
for i in range(5):  
    arr.append(0)    #추가하고  
arr
```

```
Out[7]: [0, 0, 0, 0, 0]
```

```
In [8]: ## 2차원 리스트 초기화 방법  
arr = [[0, 0, 0, 0], [0, 0, 0, 0], [0, 0, 0, 0]]  
arr
```

```
Out[8]: [[0, 0, 0, 0], [0, 0, 0, 0], [0, 0, 0, 0]]
```

```
In [9]: ## 2차원 리스트 초기화 방법  
arr = [[0 for j in range(4)] for i in range(3)];  
arr
```

```
Out[9]: [[0, 0, 0, 0], [0, 0, 0, 0], [0, 0, 0, 0]]
```

```
In [11]: ## 2차원 리스트 초기화 방법  
arr=[]; #1차원 리스트 생성  
for i in range(3):  
    temp = []    #열 생성  
    for j in range(4):  
        temp.append(0) #열 값 추가  
    arr.append(temp) #1차원 리스트에 열 추가  
arr
```

```
Out[11]: [[0, 0, 0, 0], [0, 0, 0, 0], [0, 0, 0, 0]]
```

<> 반복문으로 리스트 읽어내기

```
In [14]: ## 1차원 리스트 읽어내기
a = [1, 2, 'a', 'b']      #리스트 생성 및 초기화
for x in a:                #range() 대신에 리스트 자료 Unpacking
    print(x, end=' ')
```

```
1 2 a b
```

```
In [15]: ## 2차원 리스트 읽어내기
a = [1, 2, ['a', 'b'], 4]      #리스트 생성 및 초기화
for x in a:                  #range() 대신에 리스트 자료 Unpacking
    if type(x) is list:       #x가 list type인지 확인
        for y in x:
            print(y, end=' ')
    else:
        print(x, end=' ')
```

```
1 2 a b 4
```

[실습] List로 막대 그래프 그리기

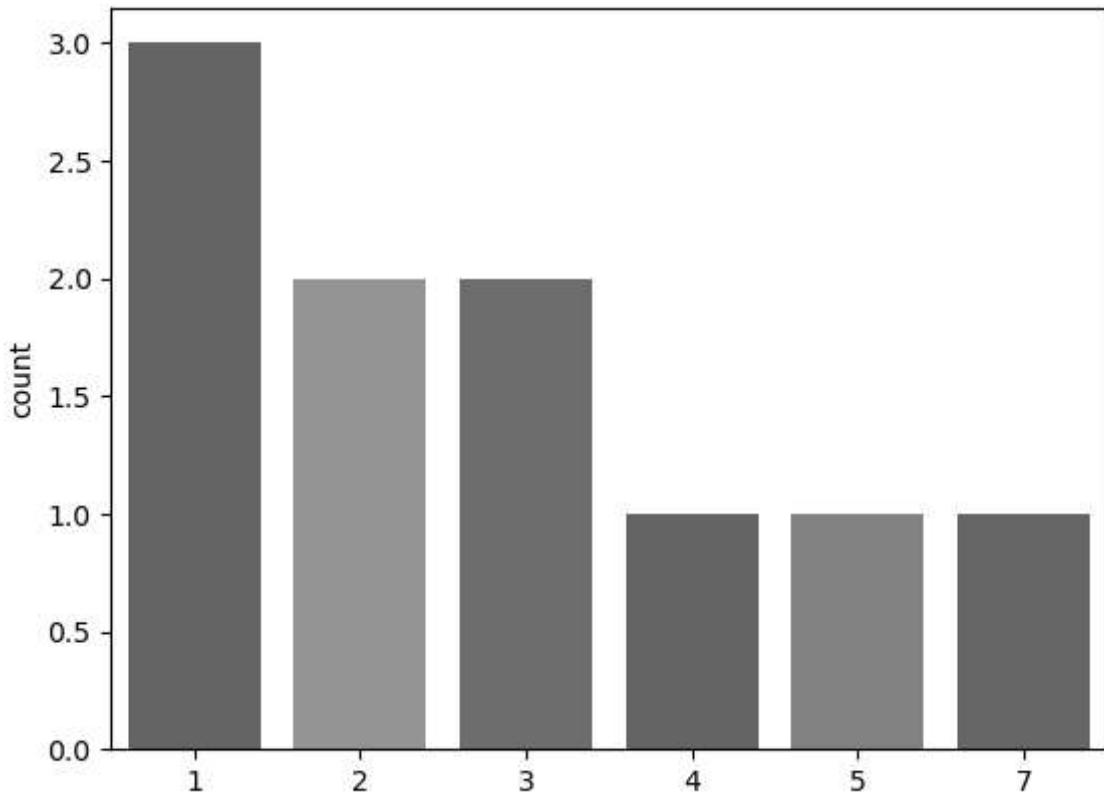
> import seaborn 모듈 사용

```
In [6]: ## 리스트(List) 형 : []
## List로 막대 그래프 그리기
alist = [1, 4, 5, 3, 2, 1, 7, 3, 1, 2]
print(alist)

import seaborn as sns
sns.countplot(x = alist)
```

```
[1, 4, 5, 3, 2, 1, 7, 3, 1, 2]
```

```
Out[6]: <AxesSubplot:ylabel='count'>
```



[실습-1] List로 막대 그래프 그리기 (응용)

- > List를 랜덤 수로 초기화 함. (range는 0~9까지)
- > List의 길이는 30으로 함
- > import seaborn 모듈 사용 빈도수 막대 그래프 그리기

```
In [ ]: ## 리스트(List) 형 : [ ]
## [실습] List를 랜덤 수로 초기화 함
## 랜덤 리스트 구성
import random
```

```
## List로 막대 그래프 그리기
import seaborn as sns
```

```
In [1]: ## seaborn 패키지 인터페이스 매개변수 확인
import seaborn as sb
sb.countplot?
```

```
Signature:
```

```
sb.countplot(  
    ...  
    *,  
    ... x=None,  
    ... y=None,  
    ... hue=None,  
    ... data=None,  
    ... order=None,  
    ... hue_order=None,  
    ... orient=None,  
    ... color=None,  
    ... palette=None,  
    ... saturation=0.75,  
    ... dodge=True,  
    ... ax=None,  
    ... **kwargs,  
)
```

```
Docstring:
```

```
Show the counts of observations in each categorical bin using bars.
```

A count plot can be thought of as a histogram across a categorical, instead of quantitative, variable. The basic API and options are identical to those for :func:`barplot`, so you can compare counts across nested variables.

Input data can be passed in a variety of formats, including:

- Vectors of data represented as lists, numpy arrays, or pandas Series objects passed directly to the ``x``, ``y``, and/or ``hue`` parameters.
- A "long-form" DataFrame, in which case the ``x``, ``y``, and ``hue`` variables will determine how the data are plotted.
- A "wide-form" DataFrame, such that each numeric column will be plotted.
- An array or list of vectors.

In most cases, it is possible to use numpy or Python objects, but pandas objects are preferable because the associated names will be used to annotate the axes. Additionally, you can use Categorical types for the grouping variables to control the order of plot elements.

This function always treats one of the variables as categorical and draws data at ordinal positions (0, 1, ... n) on the relevant axis, even when the data has a numeric or date type.

See the :ref:`tutorial <categorical_tutorial>` for more information.

Parameters

x, y, hue : names of variables in ``data`` or vector data, optional
Inputs for plotting long-form data. See examples for interpretation.

data : DataFrame, array, or list of arrays, optional
Dataset for plotting. If ``x`` and ``y`` are absent, this is interpreted as wide-form. Otherwise it is expected to be long-form.

order, hue_order : lists of strings, optional
Order to plot the categorical levels in, otherwise the levels are inferred from the data objects.

orient : "v" | "h", optional
Orientation of the plot (vertical or horizontal). This is usually inferred based on the type of the input variables, but it can be used to resolve ambiguity when both `x` and `y` are numeric or when plotting wide-form data.

color : matplotlib color, optional
Color for all of the elements, or seed for a gradient palette.

palette : palette name, list, or dict
Colors to use for the different levels of the ``hue`` variable. Should

```
.... be something that can be interpreted by :func:`color_palette`, or a
.... dictionary mapping hue levels to matplotlib colors. ....
saturation : float, optional
.... Proportion of the original saturation to draw colors at. Large patches
.... often look better with slightly desaturated colors, but set this to
.... ``1`` if you want the plot colors to perfectly match the input color
.... spec. ....
dodge : bool, optional
.... When hue nesting is used, whether elements should be shifted along the
.... categorical axis. ....
ax : matplotlib Axes, optional
.... Axes object to draw the plot onto, otherwise uses the current Axes. ....
kwargs : key, value mappings
.... Other keyword arguments are passed through to
.... :meth:`matplotlib.axes.Axes.bar`.
```

Returns

```
ax : matplotlib Axes
.... Returns the Axes object with the plot drawn onto it. ....
```

See Also

```
barplot : Show point estimates and confidence intervals using bars. ....
catplot : Combine a categorical plot with a :class:`FacetGrid`. ....
```

Examples

Show value counts for a single categorical variable:

```
.. plot::
.... :context: close-figs

.... >>> import seaborn as sns
.... >>> sns.set_theme(style="darkgrid")
.... >>> titanic = sns.load_dataset("titanic")
.... >>> ax = sns.countplot(x="class", data=titanic)
```

Show value counts for two categorical variables:

```
.. plot::
.... :context: close-figs

.... >>> ax = sns.countplot(x="class", hue="who", data=titanic)
```

Plot the bars horizontally:

```
.. plot::
.... :context: close-figs

.... >>> ax = sns.countplot(y="class", hue="who", data=titanic)
```

Use a different color palette:

```
.. plot::
.... :context: close-figs

.... >>> ax = sns.countplot(x="who", data=titanic, palette="Set3")
```

Use :meth:`matplotlib.axes.Axes.bar` parameters to control the style.

```
.. plot::
.... :context: close-figs
```

```

...     >>> ax = sns.countplot(x="who", data=titanic,
...                         facecolor=(0, 0, 0, 0),
...                         linewidth=5,
...                         edgecolor=sns.color_palette("dark", 3))

Use :func:`catplot` to combine a :func:`countplot` and a
:class:`FacetGrid`. This allows grouping within additional categorical
variables. Using :func:`catplot` is safer than using :class:`FacetGrid`
directly, as it ensures synchronization of variable order across facets:

.. plot::
   :context: close-figs

...     >>> g = sns.catplot(x="class", hue="who", col="survived",
...                         data=titanic, kind="count",
...                         height=4, aspect=.7);
File:    c:\users\admin\anaconda3\lib\site-packages\seaborn\categorical.py
Type:    function

```

In []:

○ 튜플(Tuple) 형 : ()

- > 리스트와 같이 여러 개의 값들을 나열된 자료 구조
- > 리스트와 다르게 한 번 초기화된 자료는 **수정할 수 없다.**
- > 고정된 값을 미리 정렬시켜서 튜플로 만들어 놓고 사용하면 검색 속도 빨라짐.

<> 생성하기

```

In [16]: ## Tuple 생성하기
          a = (1, 2, 3, 4, 5)    #튜플의 생성 및 초기화
          a

```

Out[16]: (1, 2, 3, 4, 5)

```

In [20]: # () 생략하고 튜플 만들기
          a = 1, 2, 3, 4, 5    #튜플의 패킹(Packing)
          a

```

Out[20]: (1, 2, 3, 4, 5)

<> 추출하기

```

In [19]: ## 튜플(Tuple) 형 : ( )
          ## 생성 및 검색
          a = (1, 2, 3, 4, 5)    #튜플의 생성 및 초기화
          print(a[0])
          print(a[1:3])      #범위로 추출하므로 결과는 tuple 형

```

1
(2, 3)

<> 튜플의 패킹(Packing)과 언패킹(Unpacking)

```
In [133]: ## 튜플(Tuple) 형 : ( )
## 패킹(Packing), 언패킹(Unpacking)
b = 10, 20, 30      #튜플의 패킹(Packing)
x, y, z = b        #튜플의 언패킹(Unpacking)
print(x, y, z)
```

10 20 30

<> 활용하기

```
In [13]: #지역 번호 만들기
zoneno = ([[['010'], '휴대전화'],
           ['02', ['서울특별시', '광명시', '과천시']],
           ['031', '경기도'],
           ['032', ['인천광역시', '부천시']],
           ['033', '강원도'],
           ['041', '충청남도'],
           ['042', ['대전광역시', '계룡시']],
           ['043', '충청북도'],
           ['044', '세종특별자치시'],
           ['051', '부산광역시'],
           ['052', '울산광역시'],
           ['053', ['대구광역시', '경산시']],
           ['054', '경상북도'],
           ['055', '경상남도'],
           ['061', '전라남도'],
           ['062', '광주광역시'],
           ['063', '전라북도'],
           ['064', '제주특별자치도']]))

#지역 번호표 출력
for no, zone in zoneno: #unpacking
    print(no, " : ", zone)
```

'010' : 휴대전화
02 : ['서울특별시', '광명시', '과천시']
031 : 경기도
032 : ['인천광역시', '부천시']
033 : 강원도
041 : 충청남도
042 : ['대전광역시', '계룡시']
043 : 충청북도
044 : 세종특별자치시
051 : 부산광역시
052 : 울산광역시
053 : ['대구광역시', '경산시']
054 : 경상북도
055 : 경상남도
061 : 전라남도
062 : 광주광역시
063 : 전라북도
064 : 제주특별자치도

[실습-2] Tuple 항목 검색하기

> 키보드로 지역명을 입력받아 zoneno의 번호를 출력해줌

```
In [17]: #지역 번호 검색(지역명으로 찾기)
inzone = input(">전화번호 지역 입력: ")
```

032 : ['인천광역시', '부천시']

In []:

○ 딕셔너리(Dictionary) 형 : { }

- > 키(Key)와 대응되는 값(Value)이 짹을 이루어 나열된 자료 구조
- > 값의 삽입, 삭제가 가능하다.
- > 순서번호(Index) 대신에 키 값으로 값에 접근

<> 생성하기

```
In [42]: ## 딕셔너리(Dictionary) 형 : { } ##
## Key : Value로 생성
a = {'A':90, 'B':80, 3:70, 4:60}          #딕셔너리 생성 및 초기화
a
```

```
Out[42]: {'A': 90, 'B': 80, 3: 70, 4: 60}
```

<> 조작 하기

추가하기

```
In [45]: ## 항목 추가하기
a['C'] = [79, 70] #맨 뒤에 추가됨
a
```

```
Out[45]: {'A': 90, 'B': 80, 3: 70, 4: 60, 'C': [79, 70]}
```

삭제하기

```
In [ ]: ## 항목 추가하기
del(a['C']) # id 키:값 쌍 삭제
a
```

```
Out[ ]: {'A': 90, 'B': 80, 3: 70, 4: 60}
```

값 변경하기

```
In [51]: ## 항목 값 변경하기
a['B'] = [89, 80] #Key가 없으면 추가됨
a
```

```
Out[51]: {'A': 90, 'B': [89, 80], 3: 70, 4: 60}
```

<> 추출하기

```
In [36]: ## 딕셔너리(Dictionary) 형 : { } ##
## 추출 : key 사용
a = {'A':90, 'B':80, 3:70, 4:60}          #딕셔너리 생성 및 초기화
a['A']
```

```
Out[36]: 90
```

```
In [42]: ## 추출 : key 사용
a['a']
```

```
-----
-- 
KeyError..... Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_20124\1838990921.py in <module>
      1 ## 추출 : key 사용
----> 2 a['a']

KeyError: 'a'
```

```
In [43]: ## 추출 : .get() 사용  
a.get('A')
```

```
Out[43]: 90
```

```
In [41]: ## 추출 : .get() 사용  
a.get('a')
```

Key와 Value 읽어내기

```
In [46]: ## Key 추출  
a.keys() #Key 추출
```

```
Out[46]: dict_keys(['A', 'B', 3, 4])
```

```
In [47]: ## Value 추출  
a.values() #Value 추출
```

```
Out[47]: dict_values([90, 80, 70, 60])
```

```
In [48]: ## Key와 Value 추출  
a.items() #Value 추출
```

```
Out[48]: dict_items([('A', 90), ('B', 80), (3, 70), (4, 60)])
```

<> 반복문으로 Dictionary 읽어내기

```
In [62]: ## Dictionary 키 읽어내기  
a = {'A':90, 'B':80, 3:70, 4:60}  
for x in a.keys(): #range() 대신에 Dictionary 자료 Unpacking  
    print(x)
```

```
A  
B  
3  
4
```

```
In [60]: ## [Dictionary] 반복문으로 값 읽어내기  
for x in a.values(): #range() 대신에 Dictionary 자료 Unpacking  
    print(x)
```

```
90  
80  
70  
60
```

```
In [28]: ## [Dictionary] 반복문으로 키, 값 읽어내기  
for x in a.items(): #range() 대신에 적용  
    print(x)  
for x, y in a.items(): #range() 대신에 적용  
    print(x, y)
```

```
('A', 90)  
( 'B', 80)  
(3, 70)  
(4, 60)  
A 90  
B 80  
3 70  
4 60
```

<> 활용하기

```
In [30]: #(소분류) 메뉴표 만들기
menu = {'COFFEE': [['에스프레소', 3.0], ['아메리카노', 3.0], ['카페라떼', 4.0], ['카프치노', 4.0], ['LATTE': [['말차라떼', 4.0], ['초코라떼', 4.0], ['카페라떼', 4.0]], 'TEA': [['청귤차', 4.0], ['자몽차', 4.0], ['레몬차', 4.0], ['카모마일', 4.5], 'ADE': [['자몽에이드', 4.5], ['레몬에이드', 4.5], ['청포도에이드', 4.5]], 'JUICE': [['망고', 4.5], ['바나나', 4.5], ['딸기', 4.5], ['키위', 4.5]], 'SMOOTHIE': [['청귤스무디', 4.5], ['요거트스무디', 4.5]], 'MILK TEA': [['흑당밀크티', 4.5], ['달고나밀크티', 4.5]]}

#[출력] 소분류 메뉴
print("[[ 메뉴 ]]")
for title, mlist in menu.items():
    print("## %s ##" %title)
    for mmenu, price in mlist: #리스트 언팩킹
        print("%-10s₩%5.1f" %(mmenu, price))
    print()

[[ 메뉴 ]]
## COFFEE ##
에스프레소 ..... 3.0
아메리카노 ..... 3.0
카페라떼 ..... 4.0
카프치노 ..... 4.0

## LATTE ##
말차라떼 ..... 4.0
초코라떼 ..... 4.0
카페라떼 ..... 4.0

## TEA ##
청귤차 ..... 4.0
자몽차 ..... 4.0
레몬차 ..... 4.0
카모마일 ..... 4.5

## ADE ##
자몽에이드 ..... 4.5
레몬에이드 ..... 4.5
청포도에이드 ..... 4.5

## JUICE ##
망고 ..... 4.5
바나나 ..... 4.5
딸기 ..... 4.5
키위 ..... 4.5

## SMOOTHIE ##
청귤스무디 ..... 4.5
요거트스무디 ..... 4.5

## MILK TEA ##
흑당밀크티 ..... 4.5
달고나밀크티 ..... 4.5
```

[실습-3] Dictionary 항목 검색하기

> 키보드로 대부분류 메뉴명을 입력받아 메뉴명 단가를 출력해줌

> 키보드 입력 값을 받아 모두 대문자 처리 후 작업

```
In [56]: #[주문] 대분류 메뉴 선택  
intitle = input(">대분류 메뉴명: ")
```

```
에스프레소..... 3.0  
아메리카노..... 3.0  
카페라떼..... 4.0  
카프치노..... 4.0
```

```
In [ ]:
```

▣ 외장 Pandas 자료형

> pandas라는 패키지를 import 해서 사용 가능한 추가 자료 구조

> Data Type : series, dataframe

> 관련 함수 : Series(), DataFrame()

```
In [ ]:
```

○ 시리즈(Series) 형

> 1차원 배열로 여러 값을 나열한 자료 구조

> 각 데이터 항목은 index로 식별된다.

> 데이터 프레임의 데이터를 구성하는 열 값을 시리즈로 추출하여 조작할 수 있다.

> pandas 패키지를 통해 사용할 수 있다.

<> 생성하기

```
In [58]: ## 시리즈(Series) 형 ##  
## index 미지정 생성하기  
import pandas as pd  
a = pd.Series([11, 22, 33, 44]) #index 자동 부여, 항목을 리스트 형태로 나열  
a
```

```
Out[58]: 0 ... 11  
1 ... 22  
2 ... 33  
3 ... 44  
dtype: int64
```

```
In [59]: ## 시리즈(Series) 형 ##  
## index 지정 생성하기  
b = pd.Series([11, 22, 33, 44], index = [1, 2, 3, 4]) #index 강제 부여  
b
```

```
Out[59]: 1 ... 11  
2 ... 22  
3 ... 33  
4 ... 44  
dtype: int64
```

<> 추가/수정 하기

> index가 없으면 추가, 있으면 수정

```
In [76]: ## index=3에 9 할당  
a = pd.Series([11, 22, 33, 44]) #index 자동 부여  
a[4] = 9 #index가 없으면 추가, 있으면 수정  
a
```

```
Out[76]: 0    11  
1    22  
2    33  
3    44  
4    9  
dtype: int64
```

```
In [79]: ## index=3에 9 할당  
b = pd.Series([11, 22, 33, 44], index = [1, 2, 3, 4]) #index 강제 부여  
b[4] = 9 #index가 없으면 추가, 있으면 수정  
b
```

```
Out[79]: 1    11  
2    22  
3    33  
4    9  
dtype: int64
```

<> 추출 하기

```
In [88]: ## 시리즈(Series) 형 ##  
## 생성 및 추출  
a = pd.Series([11, 22, 33, 44])  
print(a[1])      #값으로 반환  
print(a[1:2])    #Series로 반환
```

```
22  
1    22  
dtype: int64
```

```
In [102...]: ## 생성 및 추출  
b = pd.Series([11, 22, 33, 44], index = ['a', 'b', 'c', 'd'])  
print(b['1'])  
print(b['b'])  
print(b[['a', 'b']]) #List로 검색, Series로 반환
```

```
22  
22  
a    11  
b    22  
dtype: int64
```

<> 삭제하기

```
In [93]: b = pd.Series([11, 22, 33, 44], index = ['a', 'b', 'c', 'd'])  
del(b['b'])  
b
```

```
Out[93]: a    11  
c    33  
d    44  
dtype: int64
```

```
In [90]: ## 시리즈 조작
a = pd.Series([11, 22, 33, 44])

a['e'] = 50;           #새로운 값 추가
print(a)
a['e'] = 55;           #새로운 값으로 변경
print(a)
del(a['e']);           #값 제거
print(a)

0    11
1    22
2    33
3    44
e    50
dtype: int64
0    11
1    22
2    33
3    44
e    55
dtype: int64
0    11
1    22
2    33
3    44
dtype: int64
```

<> 반복문으로 읽어내기

```
In [95]: ## 시리즈를 for 반복문으로 읽어내기
a = pd.Series([11, 22, 33, 44]) #index 자동 부여
for x in a:                   #range() 대신에 적용
    print(x)
```

```
11
22
33
44
```

pandas 함수의 출력 결과 활용하기

```
In [60]: ## pandas 함수 활용
a = pd.Series([11, 22, 33, 44]) #index 자동 부여
a.value_counts() #항목별 분포 수
a.sum() #항목의 합
```

```
Out[60]: 110
```

```
In [61]: a.mean() #항목 평균
```

```
Out[61]: 27.5
```

```
In [62]: a.max() #최고값 항목
```

```
Out[62]: 44
```

```
In [63]: a.min() #최소값 항목
```

Out[63]: 11

In []:

[실습-4] List로 Series 만들기

- > List를 랜덤 수로 초기화 함. (range는 0~9까지)
- > List의 길이는 30으로 함
- > 각 숫자의 빈도 수를 index로 하는 Series 구성

In [69]:

```
## 리스트(List) 형 : []
## [실습] List를 랜덤 수로 초기화 함
## 랜덤 리스트 구성
import random
alist = []
for _ in range(30):
    x = random.randrange(0, 10)
    alist.append(x)
print(alist)

# 빈도 리스트 data 만들기

# 인덱스 리스트 idx 만들기

## pandas 패키지 활용 : 데이터 시리즈(2차원 자료 구조) 구성
import pandas as pd
ds = pd.Series(data, index = idx)
print(ds)

[1, 5, 2, 4, 1, 7, 1, 8, 0, 2, 4, 8, 6, 5, 9, 0, 9, 5, 7, 4, 7, 0, 8, 0, 1, 1, 8, 3, 8, 2]
0    4
1    5
2    3
3    1
4    3
5    3
6    1
7    3
8    5
9    2
dtype: int64
```

In []:

○ 데이터 프레임(Data Frame) 형

- > 행과 열로 나열되는 2차원적 자료 구조
- > 행과 열 값들은 딕셔너리 구조를 활용하여 초기화 한다.

> 하나의 열 값들은 시리즈형으로 추출하여 사용 가능하다.

> pandas 패키지를 통해 사용할 수 있다.

<> 생성하기

```
In [71]: ## 데이터 프레임(Data Frame) 형 ##
## 생성: index 자동 부여
import pandas as pd
df = pd.DataFrame({'Eng' : [87, 79, 80],
                   'Mat' : [80, 90, 80]}) #index 자동 부여
df
```

```
Out[71]:    Eng  Mat
0     87   80
1     79   90
2     80   80
```

```
In [70]: ## [Data Frame] 생성: index 강제 부여
import pandas as pd
df = pd.DataFrame({'Eng' : [87, 79, 80],
                   'Mat' : [80, 90, 80]},
                   index = ['Kims', 'Lees', 'Parks']) #index 강제 부여
df
```

```
Out[70]:    Eng  Mat
Kims    87   80
Lees    79   90
Parks   80   80
```

<> 추출하기

> Data Frame에서 변수를 추출하면 Series형으로 반환

```
In [72]: ## [Data Frame] 조작 : 열 추출
x = df['Eng']; # 'Eng' 열 값들을 시리즈(Series)로 추출, 결과는 Series
print(type(x))
x
```

```
<class 'pandas.core.series.Series'>
0    87
1    79
2    80
Name: Eng, dtype: int64
Out[72]:
```

```
In [74]: ## [Data Frame] 조작 : 열 추출
y = df[['Mat', 'Eng']] # ['Mat', 'Eng']는 List, 결과는 Data Frame
print(type(y))
y
```

```
<class 'pandas.core.frame.DataFrame'>
```

Out[74]:

	Mat	Eng
0	80	87
1	90	79
2	80	80

<> 추가/수정 하기

> index가 없으면 추가, 있으면 수정

```
In [151...]: ## [Data Frame] 조작 : 열 추가  
df['avg'] = (df['Eng'] + df['Mat']) / 2 # 'avg' 열을 만들어서 평균 값 추가  
df
```

Out[151]:

	Eng	Mat	avg
Kims	87	80	83.5
Lees	79	90	84.5
Parks	80	80	80.0

<> 행, 열 제거하기

> .drop()

```
In [152...]: ## [Data Frame] 관련 함수 : 행, 열 제거  
ddf = df.drop( 'Lees', axis = 'rows') # 행 이름으로 열 제거  
ddf = ddf.drop( 'avg', axis = 'columns') # 컬럼 이름으로 컬럼 제거  
ddf
```

Out[152]:

	Eng	Mat
Kims	87	80
Parks	80	80

In []:

pandas 함수의 출력 결과 활용하기

```
In [156...]: ## [Data Frame] 관련 함수 사용  
df.value_counts() # 'Eng' 열 값들의 개수 구하기
```

```
Out[156]: Eng Mat avg  
79 ... 90 ... 84.5 ... 1  
80 ... 80 ... 80.0 ... 1  
87 ... 80 ... 83.5 ... 1  
dtype: int64
```

```
In [157...]: df.mean() # 'Eng' 열 값들의 평균 구하기
```

```
Out[157]: Eng ... 82.000000  
Mat ... 83.333333  
avg ... 82.666667  
dtype: float64
```

```
In [154]: ## [Data Frame] 관련 함수 사용  
df['Mat'].mean() # 'Mat' 열 값들의 평균 구하기
```

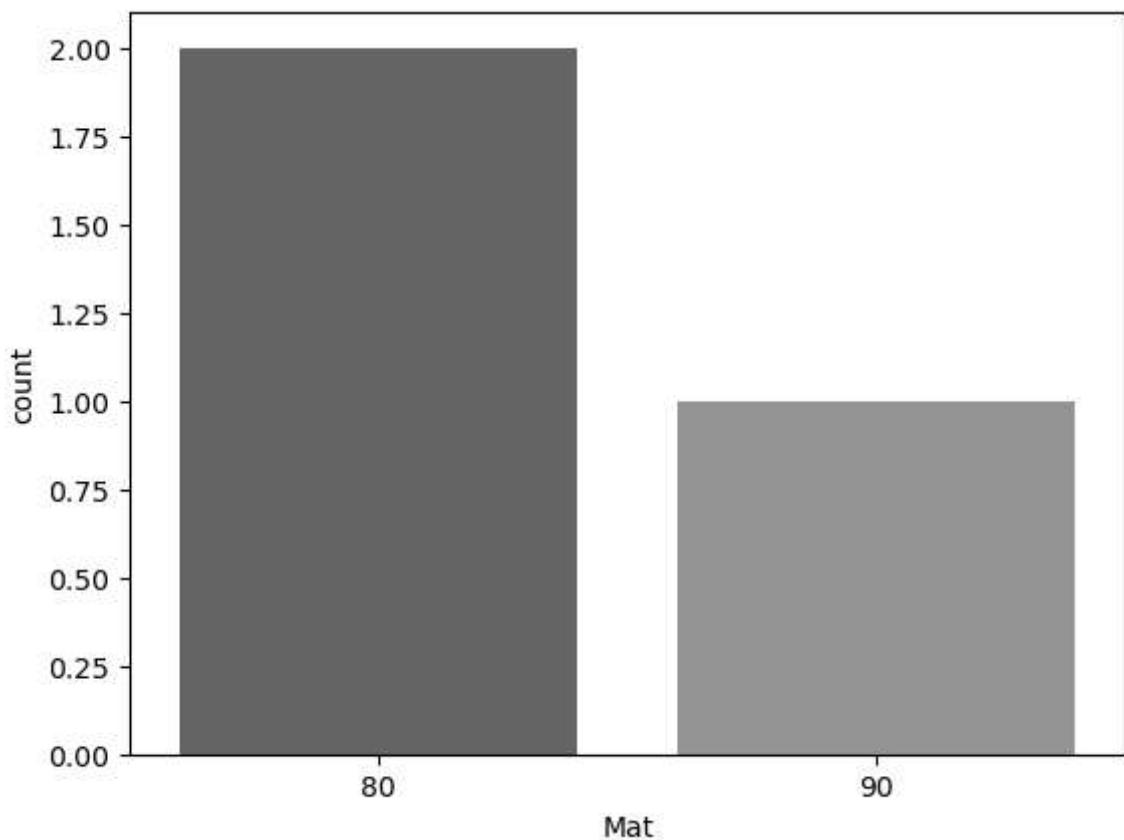
```
Out[154]: 83.33333333333333
```

```
In [ ]:
```

Data Frame으로 그래프 그리기

```
In [73]: ## [Data Frame] 그래프 그리기  
import seaborn as sns  
sns.countplot(data=df, x='Mat')
```

```
Out[73]: <AxesSubplot:xlabel='Mat', ylabel='count'>
```



```
In [75]: ## 함수 사용법 보기  
import pandas as pd  
pd.DataFrame?
```

```
Init signature:  
pd.DataFrame(  
    ... data=None,  
    ... index: 'Axes | None' = None,  
    ... columns: 'Axes | None' = None,  
    ... dtype: 'Dtype | None' = None,  
    ... copy: 'bool | None' = None,  
)  
Docstring:  
Two-dimensional, size-mutable, potentially heterogeneous tabular data.
```

Data structure also contains labeled axes (rows and columns). Arithmetic operations align on both row and column labels. Can be thought of as a dict-like container for Series objects. The primary pandas data structure.

Parameters

```
data : ndarray (structured or homogeneous), Iterable, dict, or DataFrame  
    ... Dict can contain Series, arrays, constants, dataclass or list-like objects. If  
    ... data is a dict, column order follows insertion-order. If a dict contains Series  
    ... which have an index defined, it is aligned by its index.  
  
    .... .. versionchanged:: 0.25.0  
    .... If data is a list of dicts, column order follows insertion-order.  
  
index : Index or array-like  
    ... Index to use for resulting frame. Will default to RangeIndex if  
    ... no indexing information part of input data and no index provided.  
columns : Index or array-like  
    ... Column labels to use for resulting frame when data does not have them,  
    ... defaulting to RangeIndex(0, 1, 2, ..., n). If data contains column labels,  
    ... will perform column selection instead.  
dtype : dtype, default None  
    ... Data type to force. Only a single dtype is allowed. If None, infer.  
copy : bool or None, default None  
    ... Copy data from inputs.  
    ... For dict data, the default of None behaves like ``copy=True``. For DataFrame  
    ... or 2d ndarray input, the default of None behaves like ``copy=False``.  
  
    .... .. versionchanged:: 1.3.0
```

See Also

```
DataFrame.from_records : Constructor from tuples, also record arrays.  
DataFrame.from_dict : From dicts of Series, arrays, or dicts.  
read_csv : Read a comma-separated values (csv) file into DataFrame.  
read_table : Read general delimited file into DataFrame.  
read_clipboard : Read text from clipboard into DataFrame.
```

Examples

Constructing DataFrame from a dictionary.

```
>>> d = {'col1': [1, 2], 'col2': [3, 4]}  
>>> df = pd.DataFrame(data=d)  
>>> df  
   col1  col2  
0     1     3  
1     2     4
```

Notice that the inferred dtype is int64.

```
>>> df.dtypes
```

```
col1 ... int64  
col2 ... int64  
dtype: object
```

To enforce a single dtype:

```
>>> df = pd.DataFrame(data=d, dtype=np.int8)  
>>> df.dtypes  
col1 ... int8  
col2 ... int8  
dtype: object
```

Constructing DataFrame from a dictionary including Series:

```
>>> d = {'col1': [0, 1, 2, 3], 'col2': pd.Series([2, 3], index=[2, 3])}  
>>> pd.DataFrame(data=d, index=[0, 1, 2, 3])  
   col1  col2  
0     0    NaN  
1     1    NaN  
2     2    2.0  
3     3    3.0
```

Constructing DataFrame from numpy ndarray:

```
>>> df2 = pd.DataFrame(np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]]),  
...                      columns=['a', 'b', 'c'])  
>>> df2  
   a  b  c  
0  1  2  3  
1  4  5  6  
2  7  8  9
```

Constructing DataFrame from a numpy ndarray that has labeled columns:

```
>>> data = np.array([(1, 2, 3), (4, 5, 6), (7, 8, 9)],  
...                  dtype=[("a", "i4"), ("b", "i4"), ("c", "i4")])  
>>> df3 = pd.DataFrame(data, columns=['c', 'a'])  
...  
>>> df3  
   c  a  
0  3  1  
1  6  4  
2  9  7
```

Constructing DataFrame from dataclass:

```
>>> from dataclasses import make_dataclass  
>>> Point = make_dataclass("Point", [("x", int), ("y", int)])  
>>> pd.DataFrame([Point(0, 0), Point(0, 3), Point(2, 3)])  
   x  y  
0  0  0  
1  0  3  
2  2  3  
File: c:\users\wadmin\anaconda3\lib\site-packages\pandas\core\frame.py  
Type: type  
Subclasses: SubclassedDataFrame
```

In []: