الجمهورية اليمنية

جامعة إب

كلية العلوم

قسم علوم الحاسوب وتقنية المعلومات

تكليف مقرر

تنقيب بيانات - عملي

Data Mining

المحاضرة الثالثة

عمل الطالب :

أسامة سعيد محمد حمود سعيد

إشراف :

أ. مالك المصنف

2024 - 2025

```
import pandas as pd
import numpy as np


temp = [4,5,6,78,2]
ser = pd.Series(temp)
print (f'The Tabel is : \n{ser}')
print (f'The Data Type of the DF is {type(ser)}')
```

```
The Tabel is :
0     4
1     5
2     6
3    78
4     2
dtype: int64
The Data Type of the DF is <class 'pandas.core.series.Series'>
```

```
data = {
   "Name":["Malek","Ali","Ahmed"],
   "Age":[23,25,28]
}
df = pd.DataFrame(data)
print (f'The Table is {df}')
print (f'The Data Type of The Df is {type(df)}')
```

```
The Table is     Name  Age
0  Malek   23
1    Ali   25
2  Ahmed   28
The Data Type of The Df is <class 'pandas.core.frame.DataFrame'>
```

```
dataset = pd.read_csv('mushrooms.csv')
print (dataset)
```

```
      class cap-shape cap-surface cap-color bruises odor gill-attachment  \
0         p         x           s         n       t    p               f
1         e         x           s         y       t    a               f
2         e         b           s         w       t    l               f
3         p         x           y         w       t    p               f
4         e         x           s         g       f    n               f
...     ...       ...         ...       ...     ...  ...             ...
8119      e         k           s         n       f    n               a
8120      e         x           s         n       f    n               a
8121      e         f           s         n       f    n               a
8122      p         k           y         n       f    y               f
8123      e         x           s         n       f    n               a

      gill-spacing gill-size gill-color  ... stalk-surface-below-ring  \
0                c         n          k  ...                        s
1                c         b          k  ...                        s
2                c         b          n  ...                        s
3                c         n          n  ...                        s
4                w         b          k  ...                        s
...            ...       ...        ...  ...                      ...
8119             c         b          y  ...                        s
8120             c         b          y  ...                        s
8121             c         b          n  ...                        s
8122             c         n          b  ...                        k
8123             c         b          y  ...                        s

      stalk-color-above-ring stalk-color-below-ring veil-type veil-color  \
0                          w                      w         p          w
1                          w                      w         p          w
2                          w                      w         p          w
3                          w                      w         p          w
4                          w                      w         p          w
...                      ...                    ...       ...        ...
8119                       o                      o         p          o
8120                       o                      o         p          n
8121                       o                      o         p          o
8122                       w                      w         p          w
8123                       o                      o         p          o

      ring-number ring-type spore-print-color population habitat
0               o         p                 k          s       u
1               o         p                 n          n       g
2               o         p                 n          n       m
3               o         p                 k          s       u
4               o         e                 n          a       g
...           ...       ...               ...        ...     ...
8119            o         p                 b          c       l
8120            o         p                 b          v       l
8121            o         p                 b          c       l
```

```
8122          o         e                w         v       l
8123          o         p                o         c       l

[8124 rows x 23 columns]
```

```
dataset.shape
```

```
(8124, 23)
```

```
dataset.head()
```

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number
0	p	x	s	n	t	p	f	c	n	k	...	s	w	w	p	w	o
1	e	x	s	y	t	a	f	c	b	k	...	s	w	w	p	w	o
2	e	b	s	w	t	l	f	c	b	n	...	s	w	w	p	w	o
3	p	x	y	w	t	p	f	c	n	n	...	s	w	w	p	w	o
4	e	x	s	g	f	n	f	w	b	k	...	s	w	w	p	w	o

5 rows × 23 columns

```
dataset.head(20)
```

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number
0	p	x	s	n	t	p	f	c	n	k	...	s	w	w	p	w	
1	e	x	s	y	t	a	f	c	b	k	...	s	w	w	p	w	
2	e	b	s	w	t	l	f	c	b	n	...	s	w	w	p	w	
3	p	x	y	w	t	p	f	c	n	n	...	s	w	w	p	w	
4	e	x	s	g	f	n	f	w	b	k	...	s	w	w	p	w	
5	e	x	y	y	t	a	f	c	b	n	...	s	w	w	p	w	
6	e	b	s	w	t	a	f	c	b	g	...	s	w	w	p	w	
7	e	b	y	w	t	l	f	c	b	n	...	s	w	w	p	w	
8	p	x	y	w	t	p	f	c	n	p	...	s	w	w	p	w	
9	e	b	s	y	t	a	f	c	b	g	...	s	w	w	p	w	
10	e	x	y	y	t	l	f	c	b	g	...	s	w	w	p	w	
11	e	x	y	y	t	a	f	c	b	n	...	s	w	w	p	w	
12	e	b	s	y	t	a	f	c	b	w	...	s	w	w	p	w	
13	p	x	y	w	t	p	f	c	n	k	...	s	w	w	p	w	
14	e	x	f	n	f	n	f	w	b	n	...	f	w	w	p	w	
15	e	s	f	g	f	n	f	c	n	k	...	s	w	w	p	w	
16	e	f	f	w	f	n	f	w	b	k	...	s	w	w	p	w	
17	p	x	s	n	t	p	f	c	n	n	...	s	w	w	p	w	
18	p	x	y	w	t	p	f	c	n	n	...	s	w	w	p	w	
19	p	x	s	n	t	p	f	c	n	k	...	s	w	w	p	w	

20 rows × 23 columns

```
dataset.tail()
```

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ri nur
8119	e	k	s	n	f	n	a	c	b	y	...	s	o	o	p	o	
8120	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	n	
8121	e	f	s	n	f	n	a	c	b	n	...	s	o	o	p	o	
8122	p	k	y	n	f	y	f	c	n	b	...	k	w	w	p	w	
8123	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	o	

5 rows × 23 columns

```
dataset.tail(20)
```

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ri nur
8104	e	k	s	n	f	n	a	c	b	y	...	s	o	o	p	o	
8105	e	k	s	n	f	n	a	c	b	y	...	s	o	o	p	n	
8106	e	k	s	n	f	n	a	c	b	o	...	s	o	o	p	o	
8107	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	o	
8108	p	k	y	e	f	y	f	c	n	b	...	s	p	w	p	w	
8109	e	b	s	w	f	n	f	w	b	w	...	s	w	w	p	w	
8110	e	x	s	n	f	n	a	c	b	o	...	s	o	o	p	o	
8111	e	k	s	w	f	n	f	w	b	p	...	s	w	w	p	w	
8112	e	k	s	n	f	n	a	c	b	o	...	s	o	o	p	n	
8113	p	k	y	e	f	y	f	c	n	b	...	k	p	p	p	w	
8114	p	f	y	c	f	m	a	c	b	y	...	y	c	c	p	w	
8115	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	o	
8116	p	k	y	n	f	s	f	c	n	b	...	k	p	w	p	w	
8117	p	k	s	e	f	y	f	c	n	b	...	s	p	w	p	w	
8118	p	k	y	n	f	f	f	c	n	b	...	s	p	w	p	w	
8119	e	k	s	n	f	n	a	c	b	y	...	s	o	o	p	o	
8120	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	n	
8121	e	f	s	n	f	n	a	c	b	n	...	s	o	o	p	o	
8122	p	k	y	n	f	y	f	c	n	b	...	k	w	w	p	w	
8123	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	o	

20 rows × 23 columns

```
dataset.columns
```

```
Index(['class', 'cap-shape', 'cap-surface', 'cap-color', 'bruises', 'odor',
       'gill-attachment', 'gill-spacing', 'gill-size', 'gill-color',
       'stalk-shape', 'stalk-root', 'stalk-surface-above-ring',
       'stalk-surface-below-ring', 'stalk-color-above-ring',
       'stalk-color-below-ring', 'veil-type', 'veil-color', 'ring-number',
       'ring-type', 'spore-print-color', 'population', 'habitat'],
      dtype='object')
```

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   class                     8124 non-null   object
 1   cap-shape                 8124 non-null   object
 2   cap-surface               8124 non-null   object
 3   cap-color                 8124 non-null   object
 4   bruises                   8124 non-null   object
 5   odor                      8124 non-null   object
 6   gill-attachment           8124 non-null   object
```

```
 7   gill-spacing              8124 non-null   object
 8   gill-size                 8124 non-null   object
 9   gill-color                8124 non-null   object
 10  stalk-shape               8124 non-null   object
 11  stalk-root                8124 non-null   object
 12  stalk-surface-above-ring  8124 non-null   object
 13  stalk-surface-below-ring  8124 non-null   object
 14  stalk-color-above-ring    8124 non-null   object
 15  stalk-color-below-ring    8124 non-null   object
 16  veil-type                 8124 non-null   object
 17  veil-color                8124 non-null   object
 18  ring-number               8124 non-null   object
 19  ring-type                 8124 non-null   object
 20  spore-print-color         8124 non-null   object
 21  population                8124 non-null   object
 22  habitat                   8124 non-null   object
dtypes: object(23)
memory usage: 1.4+ MB
```

```
dataset.describe()
```

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	r
count	8124	8124	8124	8124	8124	8124	8124	8124	8124	8124	...	8124	8124	8124	8124	8124	
unique	2	6	4	10	2	9	2	2	2	12	...	4	9	9	1	4	
top	e	x	y	n	f	n	f	c	b	b	...	s	w	w	p	w	
freq	4208	3656	3244	2284	4748	3528	7914	6812	5612	1728	...	4936	4464	4384	8124	7924	

4 rows × 23 columns

```
dataset.select_dtypes('int64').columns
```

```
Index([], dtype='object')
```

```
dataset.select_dtypes('object').columns
```

```
Index(['class', 'cap-shape', 'cap-surface', 'cap-color', 'bruises', 'odor',
       'gill-attachment', 'gill-spacing', 'gill-size', 'gill-color',
       'stalk-shape', 'stalk-root', 'stalk-surface-above-ring',
       'stalk-surface-below-ring', 'stalk-color-above-ring',
       'stalk-color-below-ring', 'veil-type', 'veil-color', 'ring-number',
       'ring-type', 'spore-print-color', 'population', 'habitat'],
      dtype='object')
```

```
dataset.bruises.unique()
```

```
array(['t', 'f'], dtype=object)
```

```
dataset.odor.unique()
```

```
array(['p', 'a', 'l', 'n', 'f', 'c', 'y', 's', 'm'], dtype=object)
```

```
dataset.population.unique()
```

```
array(['s', 'n', 'a', 'v', 'y', 'c'], dtype=object)
```

```
dataset.habitat.unique()
```

```
array(['u', 'g', 'm', 'd', 'p', 'w', 'l'], dtype=object)
```

```
dataset.bruises.value_counts()
```

```
bruises
f    4748
t    3376
Name: count, dtype: int64
```

```
dataset.population.value_counts()
```

```
population
v    4040
y    1712
s    1248
n     400
a     384
```

```
c     340
Name: count, dtype: int64
```

Start coding or generate with AI.