

Date: 2026-02-09

Owner: Htet Khant Linn

Week: 2

```
In [44]: import pandas as pd
```

Series

```
In [45]: Age = pd.Series([10, 20, 30, 40], index = ['age1', 'age2', 'age3', 'age4'])
```

```
In [46]: print(Age.age3)
```

30

```
In [47]: print(Age.age1)
```

10

```
In [48]: # Filtering values of the Series
Filtered_Age = Age[Age>10]
print(Filtered_Age)
```

```
age2    20
age3    30
age4    40
dtype: int64
```

```
In [49]: Filtered_Age_2 = Age[(Age > 10) & (Age < 40)]
print(Filtered_Age_2)
```

```
age2    20
age3    30
dtype: int64
```

```
In [50]: Filtered_Age_3 = Age[Age.between(20, 30)]
print(Filtered_Age_3)
```

```
age2    20
age3    30
dtype: int64
```

```
In [51]: # Calling values of the Series
Age.values
```

```
Out[51]: array([10, 20, 30, 40])
```

```
In [52]: # Calling indexes of the Series
Age.index
```

```
Out[52]: Index(['age1', 'age2', 'age3', 'age4'], dtype='str')
```

DataFrame

```
In [53]: import numpy as np
```

```
In [54]: # Creating DataFrame
```

```
DF = np.array([[20, 10, 8], [25, 8, 10], [27, 5, 3], [30, 9, 7]])  
# Creating array first
```

```
print(DF)  
type(DF)
```

```
[[20 10  8]  
 [25  8 10]  
 [27  5  3]  
 [30  9  7]]
```

```
Out[54]: numpy.ndarray
```

```
In [55]: Data_set = pd.DataFrame(DF)  
print(Data_set)  
type(Data_set)
```

```
   0  1  2  
0 20 10  8  
1 25  8 10  
2 27  5  3  
3 30  9  7
```

```
Out[55]: pandas.DataFrame
```

```
In [56]: Data_set = pd.DataFrame(DF, index = ['S1', 'S2', 'S3', 'S4'])  
print(Data_set)
```

```
   0  1  2  
S1 20 10  8  
S2 25  8 10  
S3 27  5  3  
S4 30  9  7
```

```
In [57]: Data_set = pd.DataFrame(DF, index = ['S1', 'S2', 'S3', 'S4'], columns = ['Age',  
print(Data_set)
```

```
   Age  Grade1  Grade2  
S1  20      10      8  
S2  25       8     10  
S3  27       5      3  
S4  30       9      7
```

```
In [58]: # Adding another column
```

```
Data_set['Grade3'] = [9, 6, 7, 10]  
print(Data_set)
```

```
   Age  Grade1  Grade2  Grade3  
S1  20      10      8       9  
S2  25       8     10       6  
S3  27       5      3       7  
S4  30       9      7     10
```

Indexing the DataFrame

.loc and iloc

.loc is label-based indexing and include the end

.iloc is integer-based indexing and exclude the end

```
In [59]: Data_set.loc['S1']
```

(Single Brackets): Returns a pandas Series. This is a one-dimensional array re

```
Out[59]: Age      20
Grade1    10
Grade2     8
Grade3     9
Name: S1, dtype: int64
```

```
In [60]: Data_set.loc[['S1']]
```

(Double Brackets): Returns a pandas DataFrame. By passing a list ['S1'] inside

```
Out[60]:
```

	Age	Grade1	Grade2	Grade3
S1	20	10	8	9

```
In [61]: Data_set.loc['S1':'S2']
```

```
Out[61]:
```

	Age	Grade1	Grade2	Grade3
S1	20	10	8	9
S2	25	8	10	6

```
In [62]: Data_set.iloc[1, 2]
```

```
Out[62]: np.int64(10)
```

```
In [63]: Data_set.iloc[:, 0]
```

all row and first column

```
Out[63]: S1    20
S2    25
S3    27
S4    30
Name: Age, dtype: int64
```

```
In [64]: Data_set.iloc[:, 3]
```

```
Out[64]: S1     9
S2     6
S3     7
S4    10
Name: Grade3, dtype: int64
```

```
In [65]: Data_set.iloc[:, 1:3] # doesn't not include the end
```

```
Out[65]:
```

	Grade1	Grade2
S1	10	8
S2	8	10
S3	5	3
S4	9	7

```
In [66]: Data_set.iloc[:, :3]
```

```
Out[66]:
```

	Age	Grade1	Grade2
S1	20	10	8
S2	25	8	10
S3	27	5	3
S4	30	9	7

```
In [67]: Data_set.iloc[1:2, :]
```

```
Out[67]:
```

	Age	Grade1	Grade2	Grade3
S2	25	8	10	6

```
In [68]: Data_set.iloc[:3, :]
```

```
Out[68]:
```

	Age	Grade1	Grade2	Grade3
S1	20	10	8	9
S2	25	8	10	6
S3	27	5	3	7

```
In [69]: Data_set.iloc[:, :-1]
```

```
Out[69]:
```

	Age	Grade1	Grade2
S1	20	10	8
S2	25	8	10
S3	27	5	3
S4	30	9	7

Drop and Replace

```
In [72]: Data_set_drop = Data_set.drop('Grade1', axis = 1)
print(Data_set_drop)
```

	Age	Grade2	Grade3
S1	20	8	9
S2	25	10	6
S3	27	3	7
S4	30	7	10

```
In [ ]: Data_set_change1 = Data_set.replace(10, 12) # Change all value for 10
print(Data_set_change1)
```

	Age	Grade1	Grade2	Grade3
S1	20	12	8	9
S2	25	8	12	6
S3	27	5	3	7
S4	30	9	7	12

```
In [ ]: Data_set_change2 = Data_set.replace({30:10, 9:30}) # changing two values at once
print(Data_set_change2)
```

	Age	Grade1	Grade2	Grade3
S1	20	10	8	30
S2	25	8	10	6
S3	27	5	3	7
S4	10	30	7	10

```
In [75]: Data_set.head()
```

```
Out[75]:
```

	Age	Grade1	Grade2	Grade3
S1	20	10	8	9
S2	25	8	10	6
S3	27	5	3	7
S4	30	9	7	10

```
In [76]: Data_set.head(1)
```

```
Out[76]:
```

	Age	Grade1	Grade2	Grade3
S1	20	10	8	9

```
In [77]: Data_set.tail(1)
```

```
Out[77]:
```

	Age	Grade1	Grade2	Grade3
S4	30	9	7	10

```
In [82]: Data_set.sort_values(by = ['Grade1'], ascending = True)
```

Out[82]:

	Age	Grade1	Grade2	Grade3
S3	27	5	3	7
S2	25	8	10	6
S4	30	9	7	10
S1	20	10	8	9

In [83]: `Data_set.sort_values(by = ['Age'], ascending= False)`

Out[83]:

	Age	Grade1	Grade2	Grade3
S4	30	9	7	10
S3	27	5	3	7
S2	25	8	10	6
S1	20	10	8	9

In [88]: `Data_set.sort_index(axis = 0, ascending = False)`
rearranges data based on its index labels rather than values

Out[88]:

	Age	Grade1	Grade2	Grade3
S4	30	9	7	10
S3	27	5	3	7
S2	25	8	10	6
S1	20	10	8	9

Importing Dataset

In [89]: `DF2 = pd.read_csv("household_electricity_consumption.csv")`

In [90]: `DF2.info()`

```
<class 'pandas.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Time            25 non-null    int64
1   E_Plug          24 non-null    float64
2   E_Heat          25 non-null    int64
3   Price           25 non-null    str
4   Temperature     25 non-null    int64
5   No. Occupants   25 non-null    int64
dtypes: float64(1), int64(4), str(1)
memory usage: 1.3 KB
```

In [91]: `DF2.head()`

Out[91]:

	Time	E_Plug	E_Heat	Price	Temperature	No. Occupants
--	------	--------	--------	-------	-------------	---------------

0	1	24.0	28	10	-15	12
1	2	17.0	32	12	-17	12
2	3	16.0	34	11	-19	12
3	3	16.0	34	11	-19	12
4	4	16.0	33	12	-18	12

```
In [ ]: DF2.to_csv('output1.csv') # this will add extra index
```

```
In [97]: DF3 = pd.read_csv('output1.csv')
DF3.head()
```

Out[97]:

	Unnamed: 0	Time	E_Plug	E_Heat	Price	Temperature	No. Occupants
--	------------	------	--------	--------	-------	-------------	---------------

0	0	1	24.0	28	10	-15	12
1	1	2	17.0	32	12	-17	12
2	2	3	16.0	34	11	-19	12
3	3	3	16.0	34	11	-19	12
4	4	4	16.0	33	12	-18	12

```
In [98]: DF2.to_csv('output2.csv', index = False)
```

```
In [99]: DF4 = pd.read_csv('output2.csv')
DF4.head()
```

Out[99]:

	Time	E_Plug	E_Heat	Price	Temperature	No. Occupants
--	------	--------	--------	-------	-------------	---------------

0	1	24.0	28	10	-15	12
1	2	17.0	32	12	-17	12
2	3	16.0	34	11	-19	12
3	3	16.0	34	11	-19	12
4	4	16.0	33	12	-18	12

```
In [106... two_cols = DF2.iloc[:, 1:3]
print(two_cols.head())
```

	E_Plug	E_Heat
0	24.0	28
1	17.0	32
2	16.0	34
3	16.0	34
4	16.0	33

```
In [105... all_columns_not_last_columns = DF2.iloc[:, :-1]
print(all_columns_not_last_columns.head())
```

	Time	E_Plug	E_Heat	Price	Temperature
0	1	24.0	28	10	-15
1	2	17.0	32	12	-17
2	3	16.0	34	11	-19
3	3	16.0	34	11	-19
4	4	16.0	33	12	-18

```
In [108... first_six_hours = DF2.iloc[:7, :]
print(first_six_hours)
```

	Time	E_Plug	E_Heat	Price	Temperature	No. Occupants
0	1	24.0	28	10	-15	12
1	2	17.0	32	12	-17	12
2	3	16.0	34	11	-19	12
3	3	16.0	34	11	-19	12
4	4	16.0	33	12	-18	12
5	5	16.0	30	10	-14	12
6	6	16.0	31	10	-16	12