

```
In [1]: import pandas

mydataset = {
    'cars': ["BMW", "Volvo", "Ford"],
    'passings': [3, 7, 2]
}

myvar = pandas.DataFrame(mydataset)

print(myvar)
```

	cars	passings
0	BMW	3
1	Volvo	7
2	Ford	2

```
In [3]: import pandas as pd

mydataset = {
    'cars': ["Jagvuer", "Audi", "Swift"],
    'passings': [6, 5, 4]
}

myvar = pd.DataFrame(mydataset)

print(myvar)
```

	cars	passings
0	Jagvuer	6
1	Audi	5
2	Swift	4

```
In [4]: import pandas as pd

a = [6, 7, 4]

myvar = pd.Series(a)

print(myvar)
```

0	6
1	7
2	4

dtype: int64

```
In [5]: import pandas as pd

print(pd.__version__)

1.3.4
```

```
In [6]: print(myvar[0])
```

6

```
In [2]: ## create labels
import pandas as pd

a = [7, 9, 7]

myvar = pd.Series(a, index = ["M", "C", "A"])

print(myvar)
```

```
M    7
C    9
A    7
dtype: int64
```

```
In [3]: ## Data frames
import pandas as pd

data = {
    "calories": [420, 380, 390],
    "duration": [50, 40, 45]
}

#Load data into a DataFrame object:
df = pd.DataFrame(data)

print(df)
```

```
   calories  duration
0        420         50
1        380         40
2        390         45
```

```
In [6]: ## Locate Row
print(df.loc[2])
```

```
calories    390
duration     45
Name: 2, dtype: int64
```

```
In [7]: ## use a List of indexes
print(df.loc[[0, 1]])
```

```
   calories  duration
0        420         50
1        380         40
```

```
In [9]: ## Name indexes
import pandas as pd

data = {
    "calories": [420, 380, 390],
```

```
"duration": [50, 40, 45]
}

df = pd.DataFrame(data, index = ["day1", "day2", "day3"])

print(df)
```

	calories	duration
day1	420	50
day2	380	40
day3	390	45

```
In [10]: ## Locate name indexes
## Refer to the name indexing:
print(df.loc["day2"])
```

calories	380
duration	40

Name: day2, dtype: int64

```
In [11]: ## Load a file in a data frame
import pandas as pd

df = pd.read_csv('data.csv')

print(df)
```

	name	price
0	Book	25
1	Coke	50
2	Cake	74
3	Pizza	150
4	Burger	95
5	Sandwich	80
6	Watch	5000
7	Mobile	25000

```
In [12]: ## Read csv files
import pandas as pd

df = pd.read_csv('data.csv')

print(df.to_string())
```

	name	price
0	Book	25
1	Coke	50
2	Cake	74
3	Pizza	150
4	Burger	95
5	Sandwich	80
6	Watch	5000
7	Mobile	25000

```
In [24]: ## Data frame (exporting from excel)
import pandas as pd
```

```
df = pd.read_csv('C:\\Users\\CSE22004\\Documents\\VU21CSEN0101010\\Excel 1.csv')
print(df)
```

	S.NO	veg	price
0	1	panner	120
1	2	Mushrrom	150
2	3	cabbage	60
3	4	potato	50

```
In [20]: ## Max rows

import pandas as pd

print(pd.options.display.max_rows)

9999
```

```
In [15]: ## max number of rows to display the entire data frame:

import pandas as pd

pd.options.display.max_rows = 9999

df = pd.read_csv('data.csv')

print(df)
```

	name	price
0	Book	25
1	Coke	50
2	Cake	74
3	Pizza	150
4	Burger	95
5	Sandwich	80
6	Watch	5000
7	Mobile	25000

```
In [33]: # series in pandas as float value
import pandas as nsk
c=[1,7.5,8.6,4]
z=nsk.Series(c)
print(z)
```

0	1.0
1	7.5
2	8.6
3	4.0

dtype: float64

```
In [35]: # series in pandas as int value
import pandas as nsk
c=[1,7,6]
z=nsk.Series(c)
print(z)
```

```
0    1
1    7
2    6
dtype: int64
```

```
In [6]: ##cleaning the data
import pandas as pd

df = pd.read_csv('D:\\gender,age.csv')

df = df.dropna()

print(new_df.to_string())
```

```
   s.no  name  age gender
0     1  jhon  17.0     m
1     2   ani  18.0     f
2     3  anki  19.0     f
```

```
In [16]: ##replace null
import pandas as pd

df = pd.read_csv('D:\\gender,age.csv')

df = df.dropna()

print(df.to_string())
```

```
   s.no  name  age gender
0     1  jhon  17.0     m
1     2   ani  18.0     f
2     3  anki  19.0     f
```

```
In [15]: ##remove all rows with null
import panda as pd
df = pd.read_csv('D:\\gender.csv')

df.dropna(inplace = True)

print(df.to_string())
```

```
-----  
ModuleNotFoundError                                Traceback (most recent call last)  
~\AppData\Local\Temp\ipykernel_10756\1641218250.py in <module>  
    1 ##remove all rows with null
```

In [19]: *##Replace NULL values with the number e:*

```
import pandas as pd  
  
df = pd.read_csv('D:\\gender.csv')  
  
df.fillna("e", inplace = True)  
print(df)
```

	s.no	name	age	gender
0	1	jhon	17	m
1	2	ani	18	f
2	3	anki	19	f
3	4	kul	20	e

In [20]: *##Calculate the MEAN, and replace any empty values with it:*

```
import pandas as pd  
  
df = pd.read_csv('D:\\age.csv')  
  
x = df["age"].mean()  
  
df["age"].fillna(x, inplace = True)  
  
print(df.to_string())
```

	s.no	name	age	gender
0	1	jhon	19.0	m
1	2	ani	18.0	f
2	3	anki	19.0	f
3	4	kul	20.0	m

In [21]: *##Calculate the mode, and replace any empty values with it:*

```
import pandas as pd  
  
df = pd.read_csv('D:\\age.csv')  
  
x = df["age"].mode()  
  
df["age"].fillna(x, inplace = True)  
  
print(df.to_string())
```

	s.no	name	age	gender
0	1	jhon	18.0	m
1	2	ani	18.0	f
2	3	anki	19.0	f
3	4	kul	20.0	m

In [22]: *##Calculate the median, and replace any empty values with it:*

```
import pandas as pd
```

```
df = pd.read_csv('D:\\age.csv')

x = df["age"].median()

df["age"].fillna(x, inplace = True)

print(df.to_string())
```

	s.no	name	age	gender
0	1	jhon	19.0	m
1	2	ani	18.0	f
2	3	anki	19.0	f
3	4	kul	20.0	m

```
In [24]: ##cleaning wrong data
import pandas as pd

df = pd.read_csv('D:\\names.csv')

for x in df.index:
    if df.loc[x, "age"] > 20:
        df.loc[x, "age"] = 40

print(df.to_string())
```

	s,no	names	age
0	1	ani	18
1	2	anki	20
2	3	sri	40
3	4	sai	40
4	5	ram	40

```
In [ ]:
```