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 [8]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel 2.csv')
print(df.to_string())
```

```
    S.no       Name  Age Gender      Ph no    Address
0      1    Mounika   18      F  701312790      vizag
1      2     Mourva   20      M  789477247    chennai
```

In [18]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel Sheet 1.csv')
print(df.to_string())
de = df.drop_duplicates(subset="Name",keep="last")
print(de)
```

```
    S.no       Name  Age Gender      Ph no    Address Having pp not havep
0      1    Mounika   18      F   78593777      vizag          y          N
1      2     Mourya   20      M   98872870    chennai          N          Y
2      3   Ramadevi   40      F   98237982    tekkali          Y          N
3      4   Nagaraju   53      M   75635453       sklm          N          Y
4      5    Preethi   33      F   54547454  hyderabad          Y          N
5      6       Raja   37      M   78565342    kolkata          N          Y
6      7    Shreyas    3      M   87654544     mumbai          Y          N
7      8   Ruchitha   21      F   87745636     palasa          N          Y
8      9       Ravi   32      M   86432109       pune          N          N
9     10     Mahesh   29      M   56544220       jamu          Y          N
10    11       Siva   31      M   89745643     nelore          N          Y
11    12       Abhi   20      F   87642212      rajam          Y          N
12    13        Ani   19      F   87565444     jaipur          N          Y
13    14  Choornika   19      F   87765564     ongole          Y          N
14    15     prajna   18      F   98765544     kurnol          N          Y
15    16     Jhansi   19      F   87476765   srilanka          Y          N
16    17    Deepthi   18      F   87567342   banglore          N          Y
17    18    Sandhya   19      F   54433456  jharkhand          Y          N
18    19     chathu    7      F   95688633     kadapa          N          Y
19    20    Satwick    8      M   96586533    srinagar          Y          N
20    21      vedha    7      F  559489i42    manipur          N          Y
21    22    charvik    6      M    4098873      dubai          Y          N
    S.no       Name  Age Gender      Ph no    Address Having pp not havep
0      1    Mounika   18      F   78593777      vizag          y          N
1      2     Mourya   20      M   98872870    chennai          N          Y
2      3   Ramadevi   40      F   98237982    tekkali          Y          N
3      4   Nagaraju   53      M   75635453       sklm          N          Y
4      5    Preethi   33      F   54547454  hyderabad          Y          N
5      6       Raja   37      M   78565342    kolkata          N          Y
6      7    Shreyas    3      M   87654544     mumbai          Y          N
7      8   Ruchitha   21      F   87745636     palasa          N          Y
8      9       Ravi   32      M   86432109       pune          N          N
9     10     Mahesh   29      M   56544220       jamu          Y          N
10    11       Siva   31      M   89745643     nelore          N          Y
11    12       Abhi   20      F   87642212      rajam          Y          N
12    13        Ani   19      F   87565444     jaipur          N          Y
13    14  Choornika   19      F   87765564     ongole          Y          N
14    15     prajna   18      F   98765544     kurnol          N          Y
15    16     Jhansi   19      F   87476765   srilanka          Y          N
16    17    Deepthi   18      F   87567342   banglore          N          Y
17    18    Sandhya   19      F   54433456  jharkhand          Y          N
18    19     chathu    7      F   95688633     kadapa          N          Y
19    20    Satwick    8      M   96586533    srinagar          Y          N
20    21      vedha    7      F  559489i42    manipur          N          Y
21    22    charvik    6      M    4098873      dubai          Y          N
```

In [16]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel 2.csv')
de = df.drop_duplicates(inplace=True)
print(de)
```

```
None
```

In [5]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel Sheet 1.csv')
df.aggregate({"Age":['max','min']})

print(df.to_string())
```

```
    S.no       Name  Age Gender      Ph no    Address Having pp not havep
0      1    Mounika   18      F   78593777      vizag          y          N
1      2     Mourya   20      M   98872870    chennai          N          Y
2      3   Ramadevi   40      F   98237982    tekkali          Y          N
3      4   Nagaraju   53      M   75635453       sklm          N          Y
4      5    Preethi   33      F   54547454  hyderabad          Y          N
5      6       Raja   37      M   78565342    kolkata          N          Y
6      7    Shreyas    3      M   87654544     mumbai          Y          N
7      8   Ruchitha   21      F   87745636     palasa          N          Y
8      9       Ravi   32      M   86432109       pune          N          N
9     10     Mahesh   29      M   56544220       jamu          Y          N
10    11       Siva   31      M   89745643     nelore          N          Y
11    12       Abhi   20      F   87642212      rajam          Y          N
12    13        Ani   19      F   87565444     jaipur          N          Y
13    14  Choornika   19      F   87765564     ongole          Y          N
14    15     prajna   18      F   98765544     kurnol          N          Y
15    16     Jhansi   19      F   87476765   srilanka          Y          N
16    17    Deepthi   18      F   87567342   banglore          N          Y
17    18    Sandhya   19      F   54433456  jharkhand          Y          N
18    19     chathu    7      F   95688633     kadapa          N          Y
19    20    Satwick    8      M   96586533   srinagar          Y          N
20    21      vedha    7      F  559489i42    manipur          N          Y
21    22    charvik    6      M    4098873      dubai          Y          N
```

In [9]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel Sheet 1.csv')
df.aggregate({"Ph no":['max','min']})

print(df.to_string())
```

	S.no	Name	Age	Gender	Ph no	Address	Having pp	not havep
0	1	Mounika	18	F	78593777	vizag	y	N
1	2	Mourya	20	M	98872870	chennai	N	Y
2	3	Ramadevi	40	F	98237982	tekkali	Y	N
3	4	Nagaraju	53	M	75635453	sklm	N	Y
4	5	Preethi	33	F	54547454	hyderabad	Y	N
5	6	Raja	37	M	78565342	kolkata	N	Y
6	7	Shreyas	3	M	87654544	mumbai	Y	N
7	8	Ruchitha	21	F	87745636	palasa	N	Y
8	9	Ravi	32	M	86432109	pune	N	N
9	10	Mahesh	29	M	56544220	jamu	Y	N

In [11]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel Sheet 1.csv')
print(df.aggregate({"Age":['max','min']}))
```

```
     Age
max   53
min    3
```

In [12]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel Sheet 1.csv')
df.aggregate({"Ph no":['max','min']})
```

Out[12]:

	Ph no
max	98872870
min	4098873

In [13]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel Sheet 1.csv')
print(df.aggregate({"Age":['sum']}))
```

```
     Age
sum  477
```

In [14]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel Sheet 1.csv')
print(df.aggregate({"Age":['mean']}))
```

```
            Age
mean  21.681818
```

In [16]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel Sheet 1.csv')
print(df.aggregate({"Age":['mean']}))
print(df.aggregate({"S.no":['max']}))
```

Age

In [20]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel Sheet 1.csv')
de = pd.DataFrame
print(de)
```

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

In [21]:
```
import pandas as pd
print(df.to_string())
de = df.drop_duplicates(subset="Name",keep="first")
print(de)
```

	S.no	Name	Age	Gender	Ph no	Address	Having pp	not havep
0	1	Mounika	18	F	78593777	vizag	y	N
1	2	Mourya	20	M	98872870	chennai	N	Y
2	3	Ramadevi	40	F	98237982	tekkali	Y	N

In [22]:
```
import pandas as pd
df = pd.read_csv('D:\\Excel Sheet 1.csv')
print(df.to_string())
de = df.drop_duplicates(inplace=False)
print(de)
```

	S.no	Name	Age	Gender	Ph no	Address	Having pp	not havep
0	1	Mounika	18	F	78593777	vizag	y	N
1	2	Mourya	20	M	98872870	chennai	N	Y
2	3	Ramadevi	40	F	98237982	tekkali	Y	N
3	4	Nagaraju	53	M	75635453	sklm	N	Y
4	5	Preethi	33	F	54547454	hyderabad	Y	N
5	6	Raja	37	M	78565342	kolkata	N	Y
6	7	Shreyas	3	M	87654544	mumbai	Y	N
7	8	Ruchitha	21	F	87745636	palasa	N	Y
8	9	Ravi	32	M	86432109	pune	N	N
9	10	Mahesh	29	M	56544220	jamu	Y	N
10	11	Siva	31	M	89745643	nelore	N	Y
11	12	Abhi	20	F	87642212	rajam	Y	N
12	13	Ani	19	F	87565444	jaipur	N	Y
13	14	Choornika	19	F	87765564	ongole	Y	N
14	15	prajna	18	F	98765544	kurnol	N	Y
15	16	Jhansi	19	F	87476765	srilanka	Y	N
16	17	Deepthi	18	F	87567342	banglore	N	Y
17	18	Sandhya	19	F	54433456	jharkhand	Y	N
18	19	chathu	7	F	95688633	kadapa	N	Y
19	20	Satwick	8	M	96586533	srinagar	Y	N
20	21	vedha	7	F	559489i42	manipur	N	Y
21	22	charvik	6	M	4098873	dubai	Y	N

	S.no	Name	Age	Gender	Ph no	Address	Having pp	not havep
0	1	Mounika	18	F	78593777	vizag	y	N
1	2	Mourya	20	M	98872870	chennai	N	Y
2	3	Ramadevi	40	F	98237982	tekkali	Y	N
3	4	Nagaraju	53	M	75635453	sklm	N	Y
4	5	Preethi	33	F	54547454	hyderabad	Y	N
5	6	Raja	37	M	78565342	kolkata	N	Y
6	7	Shreyas	3	M	87654544	mumbai	Y	N
7	8	Ruchitha	21	F	87745636	palasa	N	Y
8	9	Ravi	32	M	86432109	pune	N	N
9	10	Mahesh	29	M	56544220	jamu	Y	N
10	11	Siva	31	M	89745643	nelore	N	Y
11	12	Abhi	20	F	87642212	rajam	Y	N
12	13	Ani	19	F	87565444	jaipur	N	Y
13	14	Choornika	19	F	87765564	ongole	Y	N
14	15	prajna	18	F	98765544	kurnol	N	Y
15	16	Jhansi	19	F	87476765	srilanka	Y	N
16	17	Deepthi	18	F	87567342	banglore	N	Y
17	18	Sandhya	19	F	54433456	jharkhand	Y	N
18	19	chathu	7	F	95688633	kadapa	N	Y
19	20	Satwick	8	M	96586533	srinagar	Y	N
20	21	vedha	7	F	559489i42	manipur	N	Y
21	22	charvik	6	M	4098873	dubai	Y	N

In [23]:
```
import pandas as pd
a = de.groupby(by='Address')
a.first()
```

Out[23]:

Address	S.no	Name	Age	Gender	Ph no	Having pp	not havep
banglore	17	Deepthi	18	F	87567342	N	Y
chennai	2	Mourya	20	M	98872870	N	Y
dubai	22	charvik	6	M	4098873	Y	N
hyderabad	5	Preethi	33	F	54547454	Y	N
jaipur	13	Ani	19	F	87565444	N	Y
jamu	10	Mahesh	29	M	56544220	Y	N
jharkhand	18	Sandhya	19	F	54433456	Y	N
kadapa	19	chathu	7	F	95688633	N	Y
kolkata	6	Raja	37	M	78565342	N	Y
kurnol	15	prajna	18	F	98765544	N	Y
manipur	21	vedha	7	F	559489i42	N	Y
mumbai	7	Shreyas	3	M	87654544	Y	N
nelore	11	Siva	31	M	89745643	N	Y
ongole	14	Choornika	19	F	87765564	Y	N
palasa	8	Ruchitha	21	F	87745636	N	Y
pune	9	Ravi	32	M	86432109	N	N
rajam	12	Abhi	20	F	87642212	Y	N
sklm	4	Nagaraju	53	M	75635453	N	Y
srilanka	16	Jhansi	19	F	87476765	Y	N
srinagar	20	Satwick	8	M	96586533	Y	N
tekkali	3	Ramadevi	40	F	98237982	Y	N
vizag	1	Mounika	18	F	78593777	y	N

In [24]:
```
import pandas as pd
a = de.groupby(by=['Address','Name'])
a.first()
```

Out[24]:

Address	Name	S.no	Age	Gender	Ph no	Having pp	not havep
banglore	**Deepthi**	17	18	F	87567342	N	Y
chennai	**Mourya**	2	20	M	98872870	N	Y
dubai	**charvik**	22	6	M	4098873	Y	N
hyderabad	**Preethi**	5	33	F	54547454	Y	N

Address	Name	S.no	Age	Gender	Ph no	Having pp	not havep
jaipur	Ani	13	19	F	87565444	N	Y
jamu	Mahesh	10	29	M	56544220	Y	N
jharkhand	Sandhya	18	19	F	54433456	Y	N
kadapa	chathu	19	7	F	95688633	N	Y
kolkata	Raja	6	37	M	78565342	N	Y
kurnol	prajna	15	18	F	98765544	N	Y
manipur	vedha	21	7	F	559489i42	N	Y
mumbai	Shreyas	7	3	M	87654544	Y	N
nelore	Siva	11	31	M	89745643	N	Y
ongole	Choornika	14	19	F	87765564	Y	N
palasa	Ruchitha	8	21	F	87745636	N	Y
pune	Ravi	9	32	M	86432109	N	N
rajam	Abhi	12	20	F	87642212	Y	N
sklm	Nagaraju	4	53	M	75635453	N	Y
srilanka	Jhansi	16	19	F	87476765	Y	N
srinagar	Satwick	20	8	M	96586533	Y	N

In [25]:
```
import pandas as pd
a = de.groupby(by=['Address','Name'])
print(type(a))
print(pd.DataFrame(a))
```

```
<class 'pandas.core.groupby.generic.DataFrameGroupBy'>
```

In [ ]: