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
```

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
----> 2 import panda as pd
      3 df = pd.read_csv('D:\\gender.csv')
      4
      5 df.dropna(inplace = True)

~\anaconda3\lib\site-packages\panda\__init__.py in <module>
----> 1 from request import PandaRequest
      2 from models import Video, Cloud, Encoding, Profile, Notifications, Pa
ndaDict
      3 from models import GroupRetriever, SingleRetriever
      4 from models import PandaError
      5 from upload_session import UploadSession

ModuleNotFoundError: No module named 'request'
```

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 [ ]:

In [17]:
```
##cleaning the data
import pandas as pd
a=pd.read_csv('D:\\koushithachowdary.csv')
print(a)
print("after deleting empty cell")
new_pd= a.dropna()
print(new_pd)
file_name = 'koushitha.xlsx'

# saving the excel
new_pd.to_excel(file_name)
```

```
   s.no  tiffins        veg   nonveg
0     1     idly  mushroom   chicken
1     2     puri        dal   mutton
2     3  samosa       methi      NaN
3     4    maggi   chmanthi   prawns
4     5     vada    panneer     fish
after deleting empty cell
   s.no tiffins        veg   nonveg
0     1    idly  mushroom   chicken
1     2    puri        dal   mutton
3     4   maggi   chmanthi   prawns
4     5    vada    panneer     fish
```

In [7]:
```
import pandas as pd
df=pd.read_csv('D:\\names exercise 1.csv')
print(df.to_string())
```

```
     sno     name  age gender        ph.no     address
0      1      ani   18      f   9441282225  srikakulam
1      2     anki   21      f   9875546756       vizag
2      3     jhon   13      m   8756787657       nepal
3      4     mami   23      m   6453745465      guntur
4      5    mouni   17      f   5456576767     ponduru
5      6      sai   25      m   7678566578   hyderadad
6      7   suresh   24      m  78645689878    srilanka
7      8   rakesh   29      m  78657587689      london
8      9    bhanu   20      m  98793678664       delhi
9     10  dhanush   23      m  84665954879      gunpur
10    11     siri   40      f  38745864567     kolkata
11    12  bhavana   19      f  87675767859      bombai
12    13    likki   18      f  97896756878       dubai
13    14  manasas   19      f  98577654877        agra
14    15   sharmi   21      f  98789856789        puri
15    16    jessy   26      m  74567457679        pune
16    17     abhi   48      m  85748678078       vizag
17    18      anu   36      f  87856509098       nepal
18    19      kia   25      m  77980577890    mani[pur
19    20   koushi   29      f  76786547900     ongoole
20    21  deepika   30      f  76465587968     nellore
21    22   harshi   31      f  56634698709       paris
22    23     mega   27      f  78765696809     chennai
23    24    shami   22      f  76987685670     kurnool
24    25  keerthi   70      f  78565987909       aruku
```

In [9]:
```
import pandas as pd
df=pd.read_csv('D:\\names exercise 1.csv')
de=df.drop_duplicates(subset="gender")
print(de)
```

```
    sno  name  age gender       ph.no     address
0     1   ani   18      f  9441282225  srikakulam
2     3  jhon   13      m  8756787657       nepal
```

In [12]:
```
import pandas as pd
df=pd.read_csv('D:\\names exercise 1.csv')
de=df.drop_duplicates(subset="address",keep="first")
print(de)
```

	sno	name	age	gender	ph.no	address
0	1	ani	18	f	9441282225	srikakulam
1	2	anki	21	f	9875546756	vizag
2	3	jhon	13	m	8756787657	nepal
3	4	mami	23	m	6453745465	guntur
4	5	mouni	17	f	5456576767	ponduru
5	6	sai	25	m	7678566578	hyderadad
6	7	suresh	24	m	78645689878	srilanka
7	8	rakesh	29	m	78657587689	london
8	9	bhanu	20	m	98793678664	delhi
9	10	dhanush	23	m	84665954879	gunpur
10	11	siri	40	f	38745864567	kolkata
11	12	bhavana	19	f	87675767859	bombai
12	13	likki	18	f	97896756878	dubai
13	14	manasas	19	f	98577654877	agra
14	15	sharmi	21	f	98789856789	puri
15	16	jessy	26	m	74567457679	pune
18	19	kia	25	m	77980577890	mani[pur
19	20	koushi	29	f	76786547900	ongoole
20	21	deepika	30	f	76465587968	nellore
21	22	harshi	31	f	56634698709	paris
22	23	mega	27	f	78765696809	chennai
23	24	shami	22	f	76987685670	kurnool
24	25	keerthi	70	f	78565987909	aruku

In [13]:
```
import pandas as pd
df=pd.read_csv('D:\\names exercise 1.csv')
de=df.drop_duplicates(inplace=True)
print(de)
```

```
None
```

In [14]:
```
import pandas as pd
df=pd.read_csv('D:\\names exercise 1.csv')
de=df.drop_duplicates(subset="name")
print(de)
```

	sno	name	age	gender	ph.no	address
0	1	ani	18	f	9441282225	srikakulam
1	2	anki	21	f	9875546756	vizag
2	3	jhon	13	m	8756787657	nepal
3	4	mami	23	m	6453745465	guntur
4	5	mouni	17	f	5456576767	ponduru
5	6	sai	25	m	7678566578	hyderadad
6	7	suresh	24	m	78645689878	srilanka
7	8	rakesh	29	m	78657587689	london
8	9	bhanu	20	m	98793678664	delhi
9	10	dhanush	23	m	84665954879	gunpur
10	11	siri	40	f	38745864567	kolkata
11	12	bhavana	19	f	87675767859	bombai
12	13	likki	18	f	97896756878	dubai
13	14	manasas	19	f	98577654877	agra
14	15	sharmi	21	f	98789856789	puri
15	16	jessy	26	m	74567457679	pune
16	17	abhi	48	m	85748678078	vizag
17	18	anu	36	f	87856509098	nepal
18	19	kia	25	m	77980577890	mani[pur
19	20	koushi	29	f	76786547900	ongoole
20	21	deepika	30	f	76465587968	nellore
21	22	harshi	31	f	56634698709	paris
22	23	mega	27	f	78765696809	chennai
23	24	shami	22	f	76987685670	kurnool
24	25	keerthi	70	f	78565987909	aruku

In [12]:
```
import pandas as pd
df=pd.read_csv('D:\\passport1.csv')
df.aggregate({"Age":['min','max']})

print(df.to_string())
```

	S.no	Name	Age	Gender	Number	Address	passport
0	1	allu arjun	38	male	8500761113	hyderabad	Y
1	2	prabhas	35	male	1234567890	hyderabad	N
2	3	mahesh	42	male	8547452895	hyderabad	Y
3	4	ram charan	35	male	84785620445	hyderabad	N
4	5	yash	38	male	8745947567	karnataka	Y
5	6	allu arjun	38	male	8500761113	hyderabad	N
6	7	prabhas	35	male	1234567890	hyderabad	Y
7	8	mahesh	42	male	8547452895	hyderabad	Y
8	9	ram charan	35	male	84785620445	hyderabad	N
9	10	yash	38	male	8745947567	karnataka	Y

In [13]:
```
import pandas as pd
df=pd.read_csv('D:\\passport1.csv')
df.aggregate({"Name":['min','max']})

print(df.to_string())
```

```
   S.no        Name  Age Gender       Number    Address passport
0     1  allu arjun   38   male   8500761113  hyderabad        Y
1     2     prabhas   35   male   1234567890  hyderabad        N
2     3      mahesh   42   male   8547452895  hyderabad        Y
3     4  ram charan   35   male  84785620445  hyderabad        N
4     5        yash   38   male   8745947567  karnataka        Y
5     6  allu arjun   38   male   8500761113  hyderabad        N
6     7     prabhas   35   male   1234567890  hyderabad        Y
7     8      mahesh   42   male   8547452895  hyderabad        Y
8     9  ram charan   35   male  84785620445  hyderabad        N
9    10        yash   38   male   8745947567  karnataka        Y
```

In [14]:
```
import pandas as pd
df=pd.read_csv('D:\\passport1.csv')
print(df.aggregate({"Age":['min','max']}))
```

```
     Age
min   35
max   42
```

In [15]:
```
import pandas as pd
df=pd.read_csv('D:\\passport1.csv')
print(df.aggregate({"Name":['min','max']}))
```

```
          Name
min  allu arjun
max        yash
```

In [16]:
```
import pandas as pd
df=pd.read_csv('D:\\passport1.csv')
print(df.aggregate({"Age":['sum']}))
```

```
     Age
sum  376
```

In [17]:
```
import pandas as pd
df=pd.read_csv('D:\\passport1.csv')
print(df.aggregate({"Age":['mean']}))
```

```
       Age
mean  37.6
```

In [19]:
```
import pandas as pd
df=pd.read_csv('D:\\passport1.csv')
print(df.aggregate({"Age":['mean']}))
print(df.aggregate({"S.no":['max']}))
```

```
       Age
mean  37.6
     S.no
max    10
```

In [36]:
```
import pandas as pd
a=df.groupby(by="Address")
a.first()
```

Out[36]:

Address	S.no	Name	Age	Gender	Number	passport
hyderabad	1	allu arjun	38	male	8500761113	Y
karnataka	5	yash	38	male	8745947567	Y

In [39]:
```
import pandas as pd
a=df.groupby(by=['Address','Name'])
a.first()
```

Out[39]:

Address	Name	S.no	Age	Gender	Number	passport
hyderabad	**allu arjun**	1	38	male	8500761113	Y
	mahesh	3	42	male	8547452895	Y
	prabhas	2	35	male	1234567890	N
	ram charan	4	35	male	84785620445	N
karnataka	**yash**	5	38	male	8745947567	Y

In [41]:
```
import pandas as pd
a=df.groupby(by=['Address','Name'])
print(type(a))
print(pd.DataFrame(a))
```

```
<class 'pandas.core.groupby.generic.DataFrameGroupBy'>
                          0                                                  1
0  (hyderabad, allu arjun)     S.no        Name  Age Gender      Number   ...
1      (hyderabad, mahesh)     S.no    Name  Age Gender      Number    Add...
2     (hyderabad, prabhas)     S.no     Name  Age Gender      Number    Ad...
3  (hyderabad, ram charan)     S.no        Name  Age Gender       Number  ...
4        (karnataka, yash)     S.no  Name  Age Gender      Number    Addre...
```

In [ ]: