

PANDAS

```
In [1]: import pandas as pd

In [17]: #dictionary i.e. dict for assign all the data
dict={'Name':['Komal','Saurav','Gaurav','Nisha','Gautam','Anjali','Shivam'],
      'Age':[17,17,18,17,18,17,17],
      'Gender':['F','M','M','F','M','F','M'],
      'Marks':[90,76,'NaN',74,65,'NaN',71]}
```

```
In [3]: #df is a variable for dataframe
df=pd.DataFrame(dict)
```

```
In [4]: #show the dataframe in a tabular format
df
```

```
Out[4]:
```

	Name	Age	Gender	Marks
0	Komal	17	F	90
1	Saurav	17	M	76
2	Gaurav	18	M	NaN
3	Nisha	17	F	74
4	Gautam	18	M	65
5	Anjali	17	F	NaN
6	Shivam	17	M	71

```
In [6]: c=avg=0
for ele in df['Marks']:
    if str(ele).isnumeric():
        c+=1
        avg+=ele
    avg/=c
    df=df.replace(to_replace="NaN",value=avg)
df
```

```
Out[6]:
```

	Name	Age	Gender	Marks
0	Komal	17	F	90.0
1	Saurav	17	M	76.0
2	Gaurav	18	M	90.0
3	Nisha	17	F	74.0
4	Gautam	18	M	65.0
5	Anjali	17	F	90.0
6	Shivam	17	M	71.0

```
In [7]: #male map to 0 and female map to 1 as a float type
df['Gender']=df['Gender'].map({'M':0,'F':1}).astype(float)
df
```

```
Out[7]:
```

	Name	Age	Gender	Marks
0	Komal	17	1.0	90.0
1	Saurav	17	0.0	76.0
2	Gaurav	18	0.0	90.0
3	Nisha	17	1.0	74.0
4	Gautam	18	0.0	65.0
5	Anjali	17	1.0	90.0
6	Shivam	17	0.0	71.0

```
In [8]: df=df[df['Marks']>=70].copy()
df
```

```
Out[8]:
```

	Name	Age	Gender	Marks
0	Komal	17	1.0	90.0
1	Saurav	17	0.0	76.0
2	Gaurav	18	0.0	90.0
3	Nisha	17	1.0	74.0
5	Anjali	17	1.0	90.0
6	Shivam	17	0.0	71.0

```
In [9]: import pandas as pd
```

```
In [23]: #dataframe for student details
details=pd.DataFrame({
    'ID':[101,102,103,104,105,106,107,108,109,110],
    'Name':['Anju','Sanju','Ram','Moni','Rahul','Shyam','Ramu','Riya','Shivu','Rani'],
    'Branch':['CSE','CSE','CSE','CSE','CSE','CSE','CSE','CSE','CSE','CSE']})
```

```
In [24]: print(details)
```

	ID	Name	Branch
0	101	Anju	CSE
1	102	Sanju	CSE
2	103	Ram	CSE
3	104	Moni	CSE
4	105	Rahul	CSE
5	106	Shyam	CSE
6	107	Ramu	CSE
7	108	Riya	CSE
8	109	Shivu	CSE
9	110	Rani	CSE

```
In [28]: import pandas as pd
```

```
In [29]: fees_status=pd.DataFrame({
    'ID':[101,102,103,104,105,106,107,108,109,110],
    'Pending':[5000,250,'NIL',9000,15000,'NIL',4500,1800,250,'NIL']})
```

```

In [30]: print(feas_status)

   ID Pending
0  101   5000
1  102    250
2  103    NIL
3  104   9000
4  105  15000
5  106    NIL
6  107   4500
7  108   1800
8  109    250
9  110    NIL

In [31]: print(pd.merge(details,feas_status,ons='ID'))

   ID  Name Branch Pending
0  101  Anju   CSE   5000
1  102  Sanju   CSE    250
2  103   Ram   CSE    NIL
3  104  Moni   CSE   9000
4  105  Rahul   CSE  15000
5  106  Shyam   CSE    NIL
6  107   Ramu   CSE   4500
7  108   Riya   CSE   1800
8  109  Shivu   CSE    250
9  110   Rani   CSE    NIL

In [32]: #Data wrangling
import pandas as pd

In [33]: car_selling_data={'Brand':['Maruti','Maruti','Maruti','Maruti','Hyundai','Hyundai','Toyota','Mahindra','Mahindra','Ford','Toyota','Ford'],
   'Year':[2010,2011,2009,2013,2010,2011,2011,2010,2013,2010,2010,2011],
   'Sold':[6,7,9,8,3,5,2,8,7,2,4,2]}

In [34]: df=pd.DataFrame(car_selling_data)

In [35]: grouped=df.groupby('Year')
print(grouped.get_group(2010))

   Brand  Year  Sold
0  Maruti  2010    6
4  Hyundai  2010    3
7  Mahindra  2010    8
9    Ford  2010    2
10  Toyota  2010    4

In [54]: import pandas as pd

In [66]: student_data={'Name':['Amit','Praveen','Jagroop','Rahul','Vishal','Suraj','Rishab','Satyapal','Amit','Rahul','Praveen','Amit'],
   'Roll_no':[23,54,29,36,59,38,12,45,34,36,54,23],
   'Email':['xxxx@gmail.com','xxxxxx@gmail.com','xxxxxx@gmail.com','xx@gmail.com','xxxx@gmail.com','xxxxx@gmail.com','xxxxx@gmail.com',
   'xxxxx@gmail.com','xxxxx@gmail.com','xxxxxx@gmail.com','xxxxxxxx@gmail.com','xxxxxxxx@gmail.com']]}

In [67]: df=pd.DataFrame(student_data)

In [69]: print(df)

   Name  Roll_no  Email
0    Amit      23  xxxx@gmail.com
1  Praveen     54  xxxxxx@gmail.com
2  Jagroop     29  xxxxxx@gmail.com
3   Rahul     36   xx@gmail.com
4  Vishal     59  xxxx@gmail.com
5   Suraj     38  xxxxx@gmail.com
6  Rishab     12  xxxxx@gmail.com
7  Satyapal    45  xxxxx@gmail.com
8    Amit     34  xxxxx@gmail.com
9   Rahul     36  xxxxxx@gmail.com
10  Praveen    54  xxxxxxxx@gmail.com
11  Amit      23  xxxxxxxx@gmail.com

In [70]: non_duplicate=df[~df.duplicated('Roll_no')]

In [71]: print(non_duplicate)

   Name  Roll_no  Email
0    Amit      23  xxxx@gmail.com
1  Praveen     54  xxxxxx@gmail.com
2  Jagroop     29  xxxxxx@gmail.com
3   Rahul     36   xx@gmail.com
4  Vishal     59  xxxx@gmail.com
5   Suraj     38  xxxxx@gmail.com
6  Rishab     12  xxxxx@gmail.com
7  Satyapal    45  xxxxx@gmail.com
8    Amit     34  xxxxx@gmail.com

In [72]: import pandas as pd

```

```

In [73]: #employees data
data1={'Name':['Jai','Princi','Gaurav','Anuj'],
       'Age':[27,24,22,32],
       'Address':['Nagpur','Kanpur','Allahabad','Kannuaj'],
       'Qualification':['Msc','MA','MCA','Phd'],
       'Mobile No.':[97,91,58,76]}

In [74]: data2={'Name':['Gaurav','Anuj','Dhiraj','Hitesh'],
               'Age':[22,32,12,52],
               'Address':['Allahabad','Kannuaj','Allahabad','Kannuaj'],
               'Qualification':['MCA','Phd','Bcom','B.hons'],
               'Salary':[1000,2000,3000,4000]}

In [77]: df=pd.DataFrame(data1,index=[0,1,2,3])

In [78]: df1=pd.DataFrame(data2,index=[2,3,6,7])

In [79]: res=pd.concat([df,df1])

In [80]: print(res)

```

	Name	Age	Address	Qualification	Mobile No	Salary
0	Jai	27	Nagpur	Msc	97.0	NaN
1	Princi	24	Kanpur	MA	91.0	NaN
2	Gaurav	22	Allahabad	MCA	58.0	NaN
3	Anuj	32	Kannuaj	Phd	76.0	NaN
2	Gaurav	22	Allahabad	MCA	NaN	1000.0
3	Anuj	32	Kannuaj	Phd	NaN	2000.0
6	Dhiraj	12	Allahabad	Bcom	NaN	3000.0
7	Hitesh	52	Kannuaj	B.hons	NaN	4000.0