NAME - ADYA KUMARI ROLL - 22352002

```
import pandas as pd;
data = {'Name': ['Jai', 'Princi', 'Gaurav',
    'Anuj', 'Ravi', 'Natasha', 'Riya'],
    'Age': [17, 17, 18, 17, 18, 17, 17],
    'Gender': ['M', 'F', 'M', 'M', 'M', 'F', 'F'],
    'Marks': [90, 76, 'NaN', 74, 65, 'NaN', 71]}
# Convert into DataFrame
df = pd.DataFrame(data)
# Display data
df
```

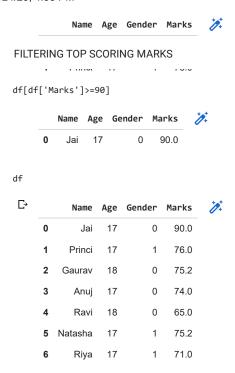
	Name	Age	Gender	Marks
0	Jai	17	М	90
1	Princi	17	F	76
2	Gaurav	18	M	NaN
3	Anuj	17	М	74
4	Ravi	18	М	65
5	Natasha	17	F	NaN
6	Riya	17	F	71

```
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)
# Display data
df
```

	Name	Age	Gender	Marks
0	Jai	17	М	90.0
1	Princi	17	F	76.0
2	Gaurav	18	М	75.2
3	Anuj	17	М	74.0
4	Ravi	18	М	65.0
5	Natasha	17	F	75.2
6	Riya	17	F	71.0

Categorizing data. (In this we categorized gender for male its $\bf 0$ and for female its one .

```
\label{eq:df_def} $$ df['Gender']=df['Gender'].map({'M':0,'F':1,}).astype(int) $$ df $$
```



- MERGING TWO DATASETS

```
Syntax for it is: pd.merge( data_frame1,data_frame2, on="field")
```

```
#CREATING TWO NEW DATASET DETAILS and FEE STATUS
details=pd.DataFrame({
    'ID':[101,102,103,104,105,106,107,108,109,110],
    'Name':['Jagroop', 'Praveen', 'Harjot','Pooja', 'Rahul', 'Nikita','Saurabh', 'Ayush', 'Dolly', 'Mohit'],
    'Branch':['CSE', 'CSE', 'CSE', 'CSE', 'CSE', 'CSE', 'CSE', 'CSE', 'CSE']
})
details
         ID
                Name Branch
     0 101
             Jagroop
                        CSE
                        CSE
        102 Praveen
     2 103
               Harjot
                        CSE
                        CSE
     3
        104
               Pooja
     4 105
               Rahul
                        CSE
     5 106
               Nikita
                        CSE
                        CSE
     6 107 Saurabh
                        CSE
     7 108
               Ayush
     8 109
                Dolly
                        CSE
     9 110
               Mohit
                        CSE
fee_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']
})
fee_status
```

	ID	PENDING	1
0	101	5000	
1	102	250	
2	103	NIL	
3	104	9000	
4	105	15000	
5	106	NIL	
6	107	4500	

new=pd.merge(details,fee_status,on="ID")
new

	ID	Name	Branch	PENDING
0	101	Jagroop	CSE	5000
1	102	Praveen	CSE	250
2	103	Harjot	CSE	NIL
3	104	Pooja	CSE	9000
4	105	Rahul	CSE	15000
5	106	Nikita	CSE	NIL
6	107	Saurabh	CSE	4500
7	108	Ayush	CSE	1800
8	109	Dolly	CSE	250
9	110	Mohit	CSE	NIL

GROUP BY

groupby is used for grouping the data according to the categories and applying a function to the categories.

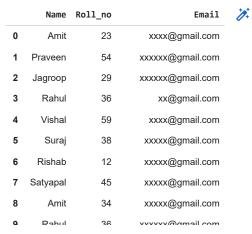
Syntax for it is: df.groupby('parameter name onto which you want to create group')

- Removing Duplicate

Pandas duplicated() method is used to remove duplicate from the dataset

Syntax:dataframe_name.duplicated('column_name')

for getting non duplicate data add \sim (not) sign before the dataframe in the syntax



#for getting the duplicated value
dff[dff.duplicated('Roll_no')]

1	Email	Roll_no	Name	
	xxxxxx@gmail.com	36	Rahul	9
	xxxxxxxxx@gmail.com	54	Praveen	10
	xxxxxxxxxx@gmail.com	23	Amit	11

#for getting the non duplicated value \sim (NOT) is used dff[\sim dff.duplicated('Roll_no')]



Concatenation of Two Datasets

Syntax for it is : pd.concat(data_frame1,data_frame2)

	Name	Age	Address	Qualification	Mobile No
0	Jai	27	Nagpur	Msc	97
1	Princi	24	Kanpur	MA	91
2	Gaurav	22	Allahabad	MCA	58
3	Anuj	32	Kannuaj	Phd	76

	Name	Age	Address	Qualification	Salary	7
2	Gaurav	22	Allahabad	MCA	1000	
3	Anuj	32	Kannuaj	Phd	2000	
6	Dhiraj	12	Allahabad	Bcom	3000	
7	Hitesh	52	Kannuaj	B.hons	4000	

res=pd.concat([d1,d2])
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

√ 0s completed at 4:38 PM