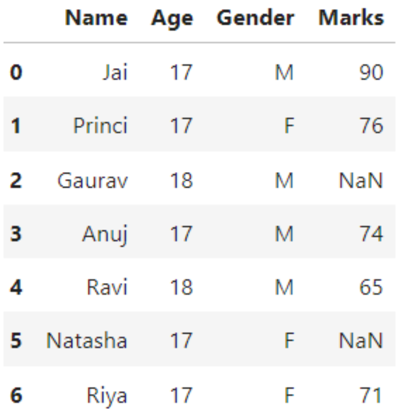
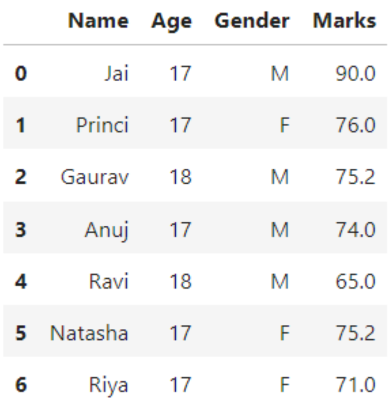
```
import pandas as pd

# Assign data
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
```



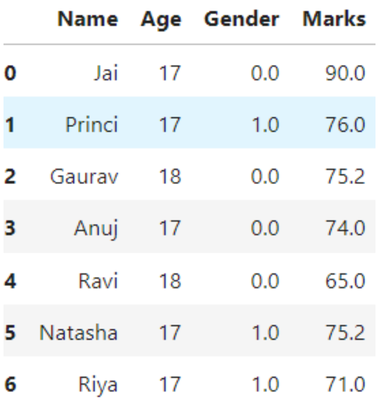
	Name	Age	Gender	Marks
0	Jai	17	M	90
1	Princi	17	F	76
2	Gaurav	18	M	NaN
3	Anuj	17	M	74
4	Ravi	18	M	65
5	Natasha	17	F	NaN
6	Riya	17	F	71

```
# Dealing with missing values
# Compute average

 c = avg = 0

    for ele in df['Marks']:

              if str(ele).isnumeric():

                  c += 1

                  avg += ele

 avg /= c

# Replace missing values
 df = df.replace(to_replace="NaN", value=avg)
# Display data
df
```



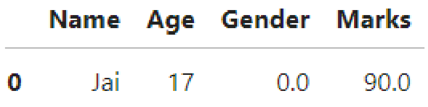
	Name	Age	Gender	Marks
0	Jai	17	M	90.0
1	Princi	17	F	76.0
2	Gaurav	18	M	75.2
3	Anuj	17	M	74.0
4	Ravi	18	M	65.0
5	Natasha	17	F	75.2
6	Riya	17	F	71.0

```
# Data Replacing in Data Wrangling
# in the GENDER column, we can replace the Gender column data by categorizing them into different numbers.

# Categorize gender
df['Gender'] = df['Gender'].map({'M': 0, 'F': 1, }).astype(float)

# Display data

df
```



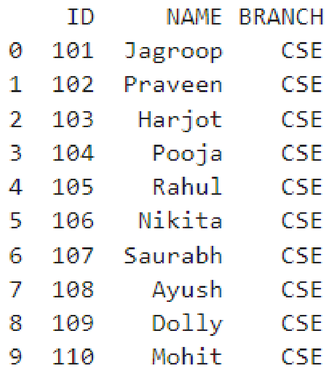
	Name	Age	Gender	Marks
0	Jai	17	0.0	90.0
1	Princi	17	1.0	76.0
2	Gaurav	18	0.0	75.2
3	Anuj	17	0.0	74.0
4	Ravi	18	0.0	65.0
5	Natasha	17	1.0	75.2
6	Riya	17	1.0	71.0

```
# Filter top scoring students

df = df[df['Marks'] >= 80].copy()

df
```



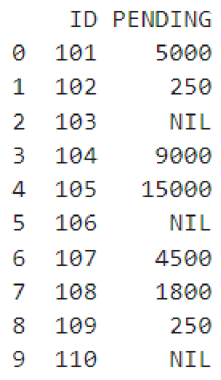
	Name	Age	Gender	Marks
0	Jai	17	0.0	90.0

```
# Data Wrangling  Using Merge Operation
#Merge operation is used to merge two raw data into the desired format.

import pandas as pd


# creating DataFrame for Student Details 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', 'CSE']})

# printing details

 print(details)
```



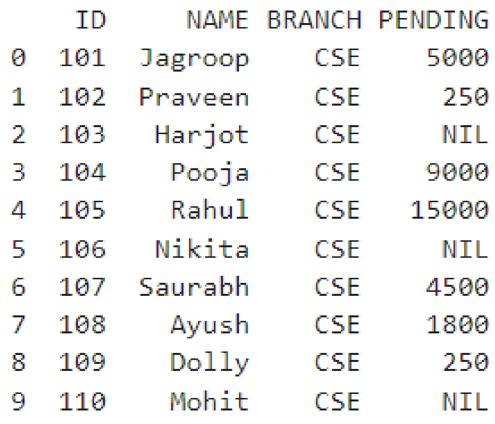
```
    ID     NAME BRANCH
0  101  Jagroop    CSE
1  102  Praveen    CSE
2  103   Harjot    CSE
3  104    Pooja    CSE
4  105    Rahul    CSE
5  106   Nikita    CSE
6  107  Saurabh    CSE
7  108    Ayush    CSE
8  109    Dolly    CSE
9  110    Mohit    CSE
```

```
import pandas as pd

# Creating Dataframe for Fees_Status

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']})

# Printing fees_status

 print(fees_status)
```



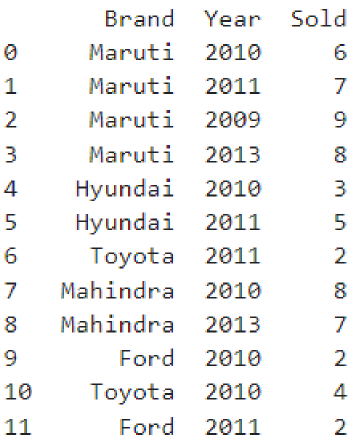
```
    ID PENDING
0  101    5000
1  102     250
2  103     NIL
3  104    9000
4  105   15000
5  106     NIL
6  107    4500
7  108    1800
8  109     250
9  110     NIL
```

```
import pandas as pd

# Creating Dataframe
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', 'CSE']})

# Creating Dataframe
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']})

# Merging Dataframe
print(pd.merge(details, fees_status, on='ID'))
```



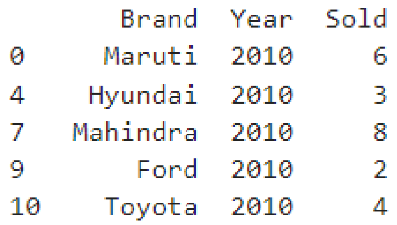
```
    ID     NAME BRANCH PENDING
0  101  Jagroop    CSE    5000
1  102  Praveen    CSE     250
2  103   Harjot    CSE     NIL
3  104    Pooja    CSE    9000
4  105    Rahul    CSE   15000
5  106   Nikita    CSE     NIL
6  107  Saurabh    CSE    4500
7  108    Ayush    CSE    1800
8  109    Dolly  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

```
# Data Wrangling Using Grouping Method #
Using groupby() method.
import pandas as pd

# Creating Data
        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]}

# Creating Dataframe of car_selling_data df =
pd.DataFrame(car_selling_data)

# printing Dataframe print(df)
```



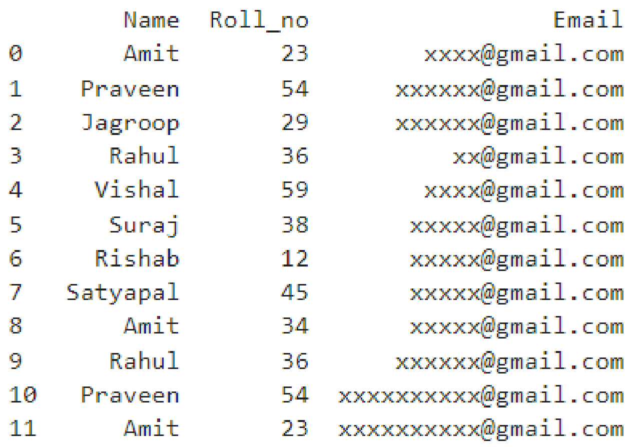
```
       Brand  Year  Sold
0     Maruti  2010     6
1     Maruti  2011     7
2     Maruti  2009     9
3     Maruti  2013     8
4    Hyundai  2010     3
5    Hyundai  2011     5
6     Toyota  2011     2
7   Mahindra  2010     8
8   Mahindra  2013     7
9       Ford  2010     2
10    Toyota  2010     4
11      Ford  2011     2
```

```
# Creating Dataframe to use Grouping methods[DATA OF THE YEAR 2010]:


import pandas as pd


# Creating Data
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]}

# Creating Dataframe for Provided Data

 df = pd.DataFrame(car_selling_data)

# Group the data when year = 2010

grouped = df.groupby('Year')
print(grouped.get_group(2010))
```



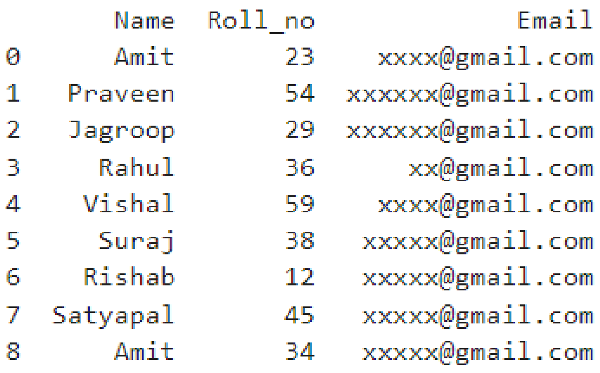
```
       Brand  Year  Sold
0     Maruti  2010     6
4    Hyundai  2010     3
7   Mahindra  2010     8
9       Ford  2010     2
10    Toyota  2010     4
```

```
# Data Wrangling  by Removing Duplication
# Pandas duplicates() method helps us to remove duplicate values from Large Data

# Syntax: DataFrame.duplicated(subset=None, keep='first')

#Here subset is the column value where we want to remove the Duplicate value.

#In keeping, we have 3 options :

#if keep ='first' then the first value is marked as the original rest of all values if occur will be removed as
it is considered duplicate. #if keep='last' then the last value is marked as the original rest the above
same values will be removed as it is considered duplicate values.
#if keep ='false' all the values which occur more than once will be removed as all are considered
duplicate values.


import pandas as pd

# Initializing Data
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',
                  'xxxxxxxxxx@gmail.com',
'xxxxxxxxxx@gmail.com']}

# Creating Dataframe of Data

 df = pd.DataFrame(student_data)

# Printing Dataframe

 print(df)
```



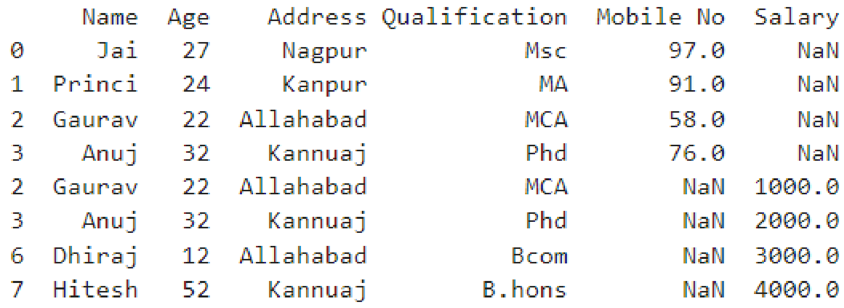
```
        Name  Roll_no                 Email
0       Amit       23        xxxx@gmail.com
1    Praveen       54      xxxxxx@gmail.com
2    Jagroop       29      xxxxxx@gmail.com
3      Rahul       36          xx@gmail.com
4     Vishal       59        xxxx@gmail.com
5      Suraj       38       xxxxx@gmail.com
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  xxxxxxxxxx@gmail.com
11      Amit       23  xxxxxxxxxx@gmail.com
```

```
# Removing Duplicate data from the Dataset using Data wrangling:


import pandas as pd
# initializing Data
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',
                    'xxxxxxxxxx@gmail.com',
'xxxxxxxxxx@gmail.com']}

# creating dataframe

df = pd.DataFrame(student_data)

# Here df.duplicated() list duplicate Entries in ROllno.

# So that ~(NOT) is placed in order to get non duplicate values. non_duplicate =
df[~df.duplicated('Roll_no')]

# printing non-duplicate values

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
```

```
# Creating New Datasets Using the Concatenation of Two Datasets In Data Wrangling.


import pandas as pd

# Define a dictionary containing employee 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]}

# Define a dictionary containing employee data
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]}

# Convert the dictionary into DataFrame
df = pd.DataFrame(data1,index=[0, 1, 2, 3])

# Convert the dictionary into DataFrame

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

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

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
```