In [125]:
```
import pandas as pd
```

In [126]:
```
dict1 = {
    "Name" :
["Abhinav","Shyam","lakshman","Anuradha","Samikshya","Shruti","Bishal"],
    "Marks" : [98,45,92,67,78,88,90],
    "Gender" : ["Male","Male","Male","Female","Female","Female","Male"]
}
df = pd.DataFrame(dict1)
```

First top 3 data

In [127]:
```
df.head(3)
```

Out[127]:

	Name	Marks	Gender
0	Abhinav	98	Male
1	Shyam	45	Male
2	lakshman	92	Male

last 3 items

In [128]:
```
df.tail(3)
```

Out[128]:

	Name	Marks	Gender
4	Samikshya	78	Female
5	Shruti	88	Female
6	Bishal	90	Male

#Shape of datasets

In [129]:
```
df.shape
```

Out[129]: (7, 3)

In [130]:
```
print("Numbers of rows", df.shape[0])
```

```
Numbers of rows 7
```

In [131]:
```
print("Number of columns", df.shape[1])
```

```
Number of columns 3
```

#information about dataset

In [132]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7 entries, 0 to 6
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   Name    7 non-null      object
 1   Marks   7 non-null      int64 
 2   Gender  7 non-null      object
dtypes: int64(1), object(2)
memory usage: 300.0+ bytes
```

#Check Null value

In [133]:
```
df.isnull().sum(axis = 0)
```

Out[133]:

	0
Name	0
Marks	0
Gender	0

dtype: int64

In [134]:
```
df.isnull().sum(axis = 1)
```

Out[134]:

	0
0	0
1	0
2	0
3	0
4	0
5	0
6	0

dtype: int64

#Overall Statistics of data

In [135]:
```
df.describe()
```

Out[135]:

	Marks
count	7.000000
mean	79.714286
std	18.391251
min	45.000000
25%	72.500000
50%	88.000000
75%	91.000000
max	98.000000

In [136]:
```
df.describe(include = "all")
```

Out[136]:

	Name	Marks	Gender
count	7	7.000000	7
unique	7	NaN	2
top	Abhinav	NaN	Male
freq	1	NaN	4
mean	NaN	79.714286	NaN
std	NaN	18.391251	NaN
min	NaN	45.000000	NaN
25%	NaN	72.500000	NaN
50%	NaN	88.000000	NaN
75%	NaN	91.000000	NaN
max	NaN	98.000000	NaN

#find unique value from gender column

In [137]:
```
df["Gender"].unique()
```

Out[137]:
```
array(['Male', 'Female'], dtype=object)
```

In [138]:
```
len(df["Gender"].unique())
```

Out[138]: 2

In [139]:
```
df["Gender"].nunique()
```

Out[139]: 2

In [140]:
```
df["Gender"].value_counts()
```

Out[140]:

	count
Gender	
Male	4
Female	3

dtype: int64

#total numbers of students having marks between 90 and 100 (inclusive)

In [141]:
```
df[df["Marks"] >= 90]["Marks"]
```

Out[141]:

	Marks
0	98
2	92
6	90

dtype: int64

In [142]:
```
len(df[(df["Marks"] >= 90) & (df["Marks"] <= 100)])
```

Out[142]: 3

In [143]:
```
len(df[df["Marks"].between(90,100)])
```

Out[143]: 3

In [144]:
```
sum(df["Marks"].between(90,100))
```

Out[144]: 3

#finding average marks

In [145]:
```
print(df["Marks"].mean())
```

```
79.71428571428571
```

In [146]:
```
print(df["Marks"].median())
```

```
88.0
```

In [147]:
```
print(df["Marks"].max())
```

```
98
```

In [148]:
```
print(df["Marks"].min())
```

```
45
```

#apply method

In [149]:
```
def marks(x):
  return x/2
df["Half"] = df["Marks"].apply(marks)
df.head()
```

Out[149]:

	Name	Marks	Gender	Half
0	Abhinav	98	Male	49.0
1	Shyam	45	Male	22.5
2	lakshman	92	Male	46.0
3	Anuradha	67	Female	33.5
4	Samikshya	78	Female	39.0

#lambda function

In [150]:
```
df["Marks"].apply(lambda x: x/2.5)
```

Out[150]:

	Marks
0	39.2
1	18.0
2	36.8
3	26.8
4	31.2
5	35.2
6	36.0

dtype: float64

#builtin functions

In [151]:
```
df["Name"].apply(len)
```

Out[151]:

	Name
0	7
1	5
2	8
3	8
4	9
5	6
6	6

dtype: int64

#Map functions

In [152]:
```
df["Gender"].map({"Male": 1, "Female": 0})
```

Out[152]:

	Gender
0	1
1	1
2	1
3	0
4	0
5	0
6	1

dtype: int64

In [153]:
```
df["M/F"] = df["Gender"].map({"Male":1, "Female": 0})
df
```

Out[153]:

	Name	Marks	Gender	Half	M/F
0	Abhinav	98	Male	49.0	1
1	Shyam	45	Male	22.5	1
2	lakshman	92	Male	46.0	1
3	Anuradha	67	Female	33.5	0
4	Samikshya	78	Female	39.0	0
5	Shruti	88	Female	44.0	0
6	Bishal	90	Male	45.0	1

In [154]:
```
df.drop("M/F", axis =1)
```

Out[154]:

	Name	Marks	Gender	Half
0	Abhinav	98	Male	49.0
1	Shyam	45	Male	22.5
2	lakshman	92	Male	46.0
3	Anuradha	67	Female	33.5
4	Samikshya	78	Female	39.0
5	Shruti	88	Female	44.0
6	Bishal	90	Male	45.0

In [156]:
```
df.drop(["M/F","Half"], axis =1, inplace=True)
df
```

Out[156]:

	Name	Marks	Gender
0	Abhinav	98	Male
1	Shyam	45	Male
2	lakshman	92	Male
3	Anuradha	67	Female
4	Samikshya	78	Female
5	Shruti	88	Female
6	Bishal	90	Male

#print name of the column

In [159]:
```
df.columns
```

Out[159]: Index(['Name', 'Marks', 'Gender'], dtype='object')

In [160]:
```
df.index
```

Out[160]: RangeIndex(start=0, stop=7, step=1)

In [ ]:
```

```

#sort the dataframe asper the marks

In [165]:
```
df.sort_values(by="Marks", ascending = False)
```

Out[165]:

	Name	Marks	Gender
0	Abhinav	98	Male
2	lakshman	92	Male
6	Bishal	90	Male
5	Shruti	88	Female
4	Samikshya	78	Female
3	Anuradha	67	Female
1	Shyam	45	Male

#display name and marks of female students

In [167]:
```
df[df["Gender"] == "Female"][["Name","Marks"]]
```

Out[167]:

	Name	Marks
3	Anuradha	67
4	Samikshya	78
5	Shruti	88

In [170]:
```
df[df["Gender"].isin(["Female"])][["Name","Marks"]]
```

Out[170]:

	Name	Marks
3	Anuradha	67
4	Samikshya	78
5	Shruti	88

In [ ]:
```

```