# DAY 10 MORNING ASSESSMENT

**#create a table with columns and rows  
# column name : students,technology,marks  
# add the column teacher  
# and delete the technology later  
# and try to to practice on all the example methods**  
  
import pandas as pd  
var={'students':['a','b','c','d','e','f','g','h','i','j','k','l','m','n','o'],  
 'technology':['AI','AI','ML','ML','AI','ML','AI','ML','AI','ML','AI','ML','AI','AI','ML'],  
 'marks':[90,91,95,87,99,100,66,43,24,90,76,43,22,14,56]}  
df=pd.DataFrame(var)  
#print(df)  
df['teacher']=[ "Aanya","Rohan","Meera","Arjun","Kavya","Dev","Ishita","Nikhil","Tara","Vihaan","Saanvi","Aarav","Anaya","Reyansh","Diya"]  
print(df)  
del df['technology']  
print(df)  
print(df['marks']>50)  
print(df.describe())  
sorted\_df = df.sort\_values(by=['marks'],ascending=True)  
print(sorted\_df)

o/p:

C:\Users\padamati.vidya\PycharmProjects\Firstproject\.venv\Scripts\python.exe C:\Users\padamati.vidya\PycharmProjects\Firstproject\test\_pandas.py

students technology marks teacher

0 a AI 90 Aanya

1 b AI 91 Rohan

2 c ML 95 Meera

3 d ML 87 Arjun

4 e AI 99 Kavya

5 f ML 100 Dev

6 g AI 66 Ishita

7 h ML 43 Nikhil

8 i AI 24 Tara

9 j ML 90 Vihaan

10 k AI 76 Saanvi

11 l ML 43 Aarav

12 m AI 22 Anaya

13 n AI 14 Reyansh

14 o ML 56 Diya

students marks teacher

0 a 90 Aanya

1 b 91 Rohan

2 c 95 Meera

3 d 87 Arjun

4 e 99 Kavya

5 f 100 Dev

6 g 66 Ishita

7 h 43 Nikhil

8 i 24 Tara

9 j 90 Vihaan

10 k 76 Saanvi

11 l 43 Aarav

12 m 22 Anaya

13 n 14 Reyansh

14 o 56 Diya

0 True

1 True

2 True

3 True

4 True

5 True

6 True

7 False

8 False

9 True

10 True

11 False

12 False

13 False

14 True

Name: marks, dtype: bool

marks

count 15.000000

mean 66.400000

std 30.429779

min 14.000000

25% 43.000000

50% 76.000000

75% 90.500000

max 100.000000

students marks teacher

13 n 14 Reyansh

12 m 22 Anaya

8 i 24 Tara

7 h 43 Nikhil

11 l 43 Aarav

14 o 56 Diya

6 g 66 Ishita

10 k 76 Saanvi

3 d 87 Arjun

0 a 90 Aanya

9 j 90 Vihaan

1 b 91 Rohan

2 c 95 Meera

4 e 99 Kavya

5 f 100 Dev

# PANDAS

1. import pandas as pd  
from io import StringIO  
csv\_data="""Name,Age,Department,Salary  
Alice,28,HR,52000  
Bob,35,Engineering,75000  
Charlie,30,HR,48000  
David,40,Marketing,61000  
Eve,25,Engineering,50000"""  
# Use StringIO to simulate reading from a file  
df = pd.read\_csv(StringIO(csv\_data))  
  
print(df.head()) #first 5 rows

o/p:

Name Age Department Salary

0 Alice 28 HR 52000

1 Bob 35 Engineering 75000

2 Charlie 30 HR 48000

3 David 40 Marketing 61000

4 Eve 25 Engineering 50000

2. print(df.shape) #gives num of rows and columns

o/p: (5, 4)

3. print(df.describe())

o/p:

Age Salary

count 5.00000 5.000000

mean 31.60000 57200.000000

std 5.94138 11122.050171

min 25.00000 48000.000000

25% 28.00000 50000.000000

50% 30.00000 52000.000000

75% 35.00000 61000.000000

max 40.00000 75000.000000

# Data Selection and Filtering

4. print(df['Age'])

o/p:

0 28

1 35

2 30

3 40

4 25

Name: Age, dtype: int64

5. print(df['Salary']>5000)

o/p:

0 True

1 True

2 True

3 True

4 True

Name: Salary, dtype: bool

6. print(df[(df['Department'] == 'HR') & (df['Age'] > 30)])

o/p:

Name Age Department Salary

0 Alice 31 HR 52000

# Data cleaning

7. import pandas as pd  
from io import StringIO  
csv\_data="""Name,Age,Department,Salary  
Alice,,HR,52000  
Bob,35,Engineering,  
Charlie,30,HR,48000  
David,40,Marketing,61000  
Eve,25,Engineering,50000  
Hash,50,Finance,"""  
# Use StringIO to simulate reading from a file  
df = pd.read\_csv(StringIO(csv\_data))

missing\_values=df.isnull().sum()  
print(missing\_values[missing\_values>0])

o/p:

Age 1

Salary 2

dtype: int64

Process finished with exit code 0

8. df['Salary']=df['Salary'].fillna(0)  
print(df)

o/p:

Name Age Department Salary

0 Alice NaN HR 52000.0

1 Bob 35.0 Engineering 0.0

2 Charlie 30.0 HR 48000.0

3 David 40.0 Marketing 61000.0

4 Eve 25.0 Engineering 50000.0

5 Hash 50.0 Finance 0.0

9. df=df.drop\_duplicates()  
df=df.reset\_index(drop=True)  
print(df)

o/p:

Name Age Department Salary

0 Alice NaN HR 52000.0

1 Bob 35.0 Engineering NaN

2 Charlie 30.0 HR 48000.0

3 David 40.0 Marketing 61000.0

4 Eve 25.0 Engineering 50000.0

5 Hash 50.0 Finance NaN

10. df\_sorted=df.sort\_values(by='Age',ascending=False)  
print(df\_sorted)

o/p:

Name Age Department Salary

5 Hash 50.0 Finance NaN

3 David 40.0 Marketing 61000.0

1 Bob 35.0 Engineering NaN

2 Charlie 30.0 HR 48000.0

4 Eve 25.0 Engineering 50000.0

0 Alice NaN HR 52000.0

11. avg\_salary=df.groupby('Department')['Salary'].mean()  
print(avg\_salary)

o/p:

Department

Engineering 50000.0

Finance NaN

HR 50000.0

Marketing 61000.0

Name: Salary, dtype: float64

12. unique\_deparment=df.drop\_duplicates(subset=['Department'])  
print(unique\_deparment)

o/p:

Name Age Department Salary

0 Alice NaN HR 52000.0

1 Bob 35.0 Engineering NaN

3 David 40.0 Marketing 61000.0

5 Hash 50.0 Finance NaN