**Python Assignment-6 Document**

**Numpy & Pandas**

* **YASH BHAKTA**

**1. Create a 1D NumPy array with 10 integers. Print the array and its data type.**

**INPUT:-**

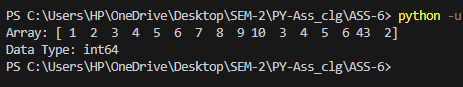
import numpy as np

array = np.array([1,2,3,4,5,6,7,8,9,10,3,4,5,6,43,2])

print("Array:", array)

print("Data Type:", array.dtype)

**OUTPUT:-**

****

**2.** **Generate an array of 10 random integers between 1 and 100 using NumPy.**

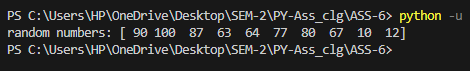
**INPUT:-**

**import numpy as np**

**arr = np.random.randint(1,101, size = 10)**

**print("random numbers:",arr)**

**OUTPUT:-**

****

**3.** **Calculate the mean, median, mode, standard deviation, and variance for the array: [12,**

**15, 12, 18, 21, 24, 24, 24, 27, 30]**

**INPUT:-**

**import numpy as np**

**from scipy import stats as st**

**arr=np.array([12,15, 12, 18, 21, 24, 24, 24, 27, 30])**

**print("Original array:",arr)**

**print("Mean: ",np.mean(arr))**

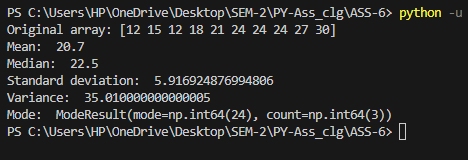
**print("Median: ",np.median(arr))**

**print("Standard deviation: ",np.std(arr))**

**print("Variance: ",np.var(arr))**

**print("Mode: ",st.mode(arr))**

**OUTPUT:-**

****

**4.** **Create a list of integers from 1 to 10. Convert it into a NumPy array and apply all five statistical functions.**

**INPUT:-**

**import numpy as np**

**from scipy import stats as st**

**l = [1,2,3,4,5,6,7,8,9,10]**

**print("List: ",l)**

**arr=np.array(l)**

**print("Original array:",arr)**

**print("Mean: ",np.mean(arr))**

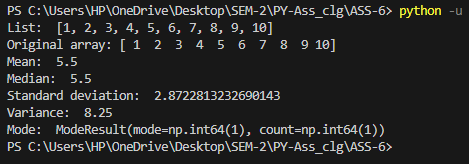
**print("Median: ",np.median(arr))**

**print("Standard deviation: ",np.std(arr))**

**print("Variance: ",np.var(arr))**

**print("Mode: ",st.mode(arr))**

**OUTPUT:-**

****

**5.** **Take a list of marks scored by 5 students in 3 subjects. Convert it into a 2D NumPy array and:**

**- Calculate the average marks of each student.**

**- Find the subject with the highest average.**

**INPUT:-**

**import numpy as np**

**l = [**

**[75,85,90],**

**[85,98,95],**

**[90,84,72],**

**[76,90,94],**

**[90,70,75]**

**]**

**arr=np.array(l)**

**print("Original array:")**

**print(arr)**

**student\_avg=np.mean(arr, axis=1)**

**print("Student average:",student\_avg)**

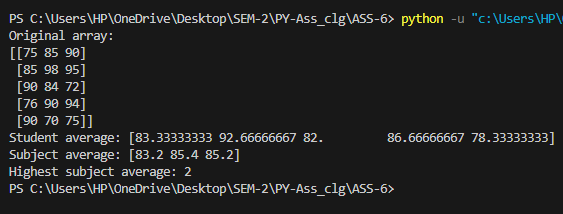
**subject\_avg=np.mean(arr, axis=0)**

**print("Subject average:",subject\_avg)**

**h\_subavg=np.max(subject\_avg)**

**print("Highest subject average:",h\_subavg+1)**

**OUTPUT:-**

****

**6.** **Print the first and last rows of the DataFrame using .head() and .tail().**

**INPUT:-**

**import pandas as pa**

**data={**

**'Name':['Tom','Jerry','Mickey','Donald'],**

**'Age':[20,21,22,23],**

**'City':['New York','Los Angeles','Chicago','Houston'],**

**}**

**print("Data Dictionary: ")**

**print(pa.DataFrame(data))**

**print("---------------------------------------")**

**print("First Row: ")**

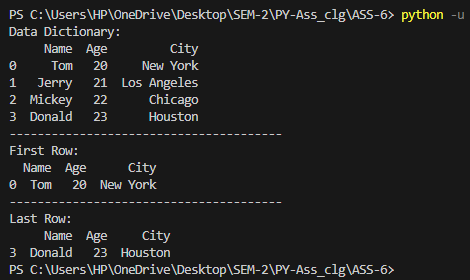
**print(pa.DataFrame(data).head(1))**

**print("---------------------------------------")**

**print("Last Row: ")**

**print(pa.DataFrame(data).tail(1))**

**OUTPUT:-**

****

**Q7:** **Create a DataFrame using the dictionary: {‘Product’: [Pen,Notebook, Eraser], Price:[10, 25, 5],Stock: [100, 50, 200]}. Print the entire DataFrame.**

**INPUT:-**

import pandas as pa

data={

    'Product':['Pen','Nootbook','Eraser'],

    'Price':[20,21,22],

    'Stock':[100,200,300]

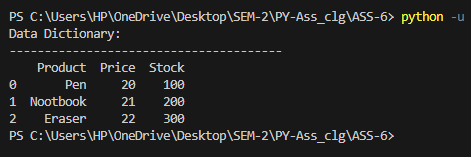
}

print("Data Dictionary: ")

print("---------------------------------------")

print(pa.DataFrame(data))

**OUTPUT:-**

****

**Q8: Add a new column Discounted\_Price = Price - 10%.**

**INPUT:-**

import pandas as pa

data={

    'Product':['Pen','Nootbook','Eraser'],

    'Price':[20,21,22],

    'Stock':[100,200,300]

}

df=pa.DataFrame(data)

print("Data Dictionary: ")

print("---------------------------------------")

print(df)

print("---------------------------------------")

df['Discounted']=df['Price']\*0.9

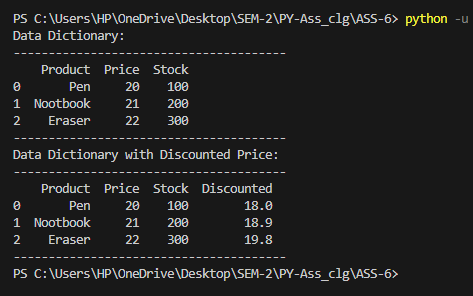
print("Data Dictionary with Discounted Price: ")

print("---------------------------------------")

print(df)

print("---------------------------------------")

**OUTPUT:-**

****

**Q9:** **Create a sample data.csv file with student details (Name, RollNo, Marks) and:**

**- Read it using pd.read\_csv().**

**- Print the total number of students.**

**INPUT:-**

import pandas as pa

data=pa.read\_csv('data.csv')

print("Data Dictionary: ")

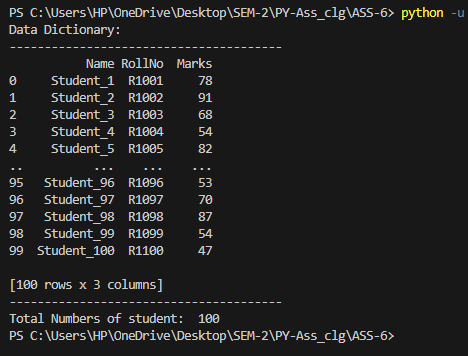
print("---------------------------------------")

print(data)

print("---------------------------------------")

print("Total Numbers of student: ", len(data))

**OUTPUT:-**

****

**Q10:** **Load a CSV file and display only the names of students who scored more than 75 marks.**

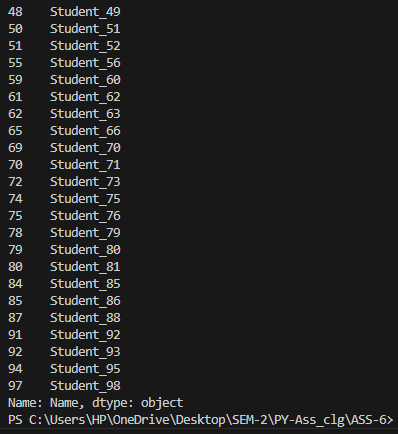
**INPUT:-**

import pandas as pd

data = pd.read\_csv('data.csv')

print(data[data['Marks'] > 70]["Name"])

**OUTPUT:-**

****

**Q11:** **Given a DataFrame of employees with Name, Dept, and Salary:**

**- Retrieve the first 3 rows.**

**- Retrieve only the Name and Salary columns.**

**- Display the salaries of employees in the HR department.**

**INPUT:-**

import pandas as pa

data={

    'Name':['Tom','Jerry','Mickey','Donald','John'],

    'department':['CS','IT','HR','HR','Finance'],

    'salary':[20000,21000,22000,23000,25000]

}

df=pa.DataFrame(data)

print("Data Dictionary: ")

print("---------------------------------------")

print(df.head(3))

print("---------------------------------------")

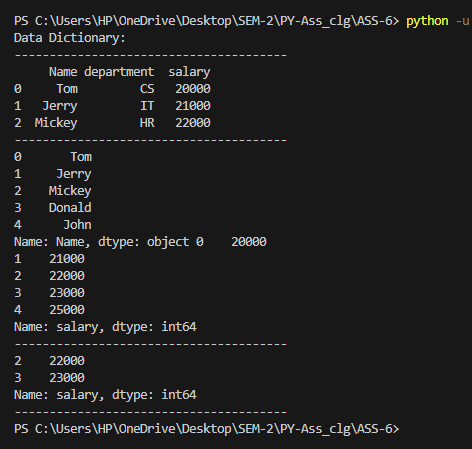
print(df['Name'],df['salary'])

print("---------------------------------------")

print(df[df['department']=='HR']['salary'])

print("---------------------------------------")

**OUTPUT:-**

****

**Q12:** **Access the value at row index 2 and column index 1 using iloc.**

**INPUT:-**

import pandas as pd

data = {

'name':['Anash', 'Daksh', 'Deep'],

'age' :[19, 18, 20]

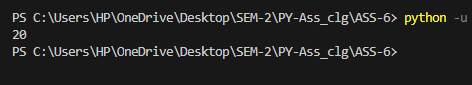
}

df = pd.DataFrame(data)

result = df.iloc[2,1]

print(result)

**OUTPUT:-**

****

**Q13:** **Use loc to get the rows where Salary > 50000.**

**INPUT:-**

import pandas as pd

data = {

'Name':['rushabh','ved','daksh','anash'],

'Dept':['Hr','finances','IT','Marketing'],

'salary':[60000, 50000, 55000, 45000]

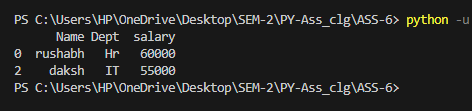
}

df = pd.DataFrame(data)

highest\_salary = df.loc[df ['salary'] > 50000]

print(highest\_salary)

**OUTPUT:-**

****

**Q14:** **Create a DataFrame with at least 5 rows and 3 columns. - Retrieve the third row using iloc. - Retrieve the values from the second and fourth rows, and the first two columns using iloc.**

**INPUT:-**

import pandas as pd

data = {

'Name':['yash','ved','daksh','anash'],

'Dept':['Hr','finances','IT','Marketing'],

'salary':[50000, 60000, 45000, 75000]

}

df = pd.DataFrame(data)

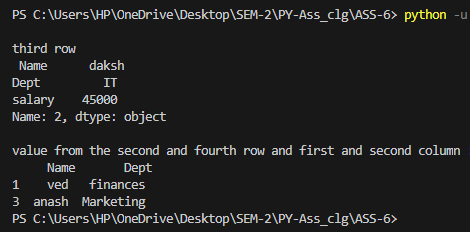
third\_row = df.iloc[2]

print("\nthird row\n",third\_row)

subset\_row = df.iloc[[1,3],[0,1]]

print("\nvalue from the second and fourth row and first and second column :\n",subset\_row)

**OUTPUT:-**

****

**Q15:** **Using loc, display:**

**- All columns for the row with index 0.**

**- A slice of rows and selected columns using labels.**

**INPUT:-**

import pandas as pd

data = {

'Name':['yash','deep','krishna','daksh'],

'Dept':['Hr','finances','IT','Marketing'],

'salary':[50000, 60000, 45000, 75000]

}

df = pd.DataFrame(data)

row\_index\_0 = df.loc[0, :]

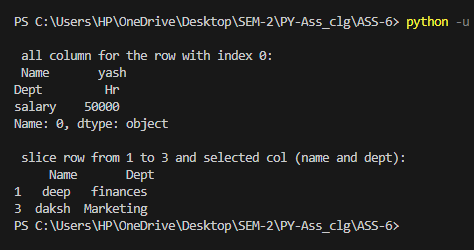
print("\n all column for the row with index 0:\n", row\_index\_0)

slice\_column\_index = df.loc[[1, 3] , ['Name', 'Dept']]

print("\n slice row from 1 to 3 and selected col (name and dept):\n",

slice\_column\_index)

**OUTPUT:-**

****