Applied Machine Learning Workshop (CSE 3193)

ASSIGNMENT-3: CRASH COURSE ON PANDAS

1. Write a program to create a dataframe from the given list and display it.
   Name: ['Ankit', 'Aishwarya', 'Shaurya', 'Shivangi','Ram', 'Alex','James','John','Supriya','Abhinash']
   Year: [1, 2, 3, 2, 1, 4, 3, 2, 1, 2]
   Branch: ['CSE', 'EE', 'CSIT', 'CSE (AIML)', 'CSE (AIML)', 'CSE', 'CSE (AIML)', 'CSIT', 'CSE', 'EE', 'EE']

2. Write a Python code to add the given row ('Alekh,' 3,'CSE') where the columns are ['Name," Year," Branch'] to the data frame created in Q.1.

3. Write Python code to check the size and dimensions of the new data frame created in Q.2.

4. Write a python program to add a column named 'CGPA' with values present in list1 to the dataframe created in Q.3 and print the result.
   list1=[8.9,8.8,9.0,7.8,7.5,9.10,9.2,8.4,9.4,8.1,7.4]

5. Write a Python code to create a data frame from the given list, which contains the marks of five different subjects of 5 first years of students.
   Ankit = [90, 92, 89, 81, 94]
   Aishwarya = [91, 81, 91, 71, 86]
   Ram = [85, 86, 83, 80, 75]
   James = [97, 96, 88, 67, 65]
   Alex = [93, 89, 78, 87, 65]
   Perform the following operations on the data frame:
   i) Change the column header as the student name.
   ii)Write a Python code to change the row label for the above data frame as
   {0:'Math',1:'ICP',2:'Physics',3:'English',4:'DSA'} and print the data frame.
   iii)Print the list of students with Boolean values to compare scores> 90 in Math.
   iv)Write a Python statement to print the marks of 'Ankit' and 'Aishwarya' in subjects Math, ICP, and Physics.
   v) Write python statement to insert a column at position 2 with values [92,89,86,93,95].

6. Write code to create and display a DataFrame from a specified dictionary data that has the index labels. Sample Python dictionary data and list labels:
   stock = {'Name': ['RAM','Register','Keyboard','Mouse','Flash Drive'], 'Price': [2500, 1000, 1500, 200, 600], 'Quantity': [10, 12, 2, 3, 12]}
   Write Python code to perform the following task on the data frame:
   i) Add a column Total, which is the product of Price and Quantity
   ii) Add a new item named 'Scanner' having a price of 3500 and Quantity 10 total as a product of price and quantity at location 5.
   iii) Drop the column quantity.

7. Write a Python code to create a data frame employee with the following details.
   Employee={'Name':['Atul', 'Shyam', 'Anmol', 'Sheetal','Dhruv'], 'Designation':['Manager', 'Analyst', 'Storekeeper', 'Manager','Analyst'], 'Salary':[56000,35000,20000,60000,380000], 'Bonus':[15000,10000,8000,18000,12000]}
   Perform the following operations on the data frame.

i) Change the row label by [E101,E102,E103,E104,E105].
ii) Write a statement to display Name, Designation and Salary columns from the above employee DataFrame by passing the list of columns into dataframe and using loc .
iii) Write a statement to display Name and Salary columns using iloc from the above DataFrame.
iv) Write a statement to display all information from Employee ids 'E102' to 'E104' using iloc.
v) List all the salary more than 35000.

8. Download the dataset for the liver patient provided in the given link and perform the following operations. https://archive.ics.uci.edu/dataset/225/ilpd+indian+liver+patient+dataset
i) Load the CVS file and display the shape, size and columns of the dataset.
ii) Print the first 5 rows and last 5 rows of the dataset.
iii) Add this heading to the column 'Age', 'Gender', 'TotalBilirubin', 'DirectBilirubin', 'TotalProteins', 'Albumin', 'A/G ratio', 'SGPT', 'SGOT','Alkphos','Class'
iv) Print the column which contains Age and Gender.
v) Print the list of datasets where the Age is greater than 40.
vi) Write a Python statement to count the number of records for Male and Female Patients.
vii) Write a Python statement to find the percentage of female patients.
viii) Write a Python statement to find the minimum of total protein given in the dataset.
ix) Write a Python statement to enlist the three smallest values of DirectBilirubin from the given data set.
x) Write a Python code to extract the first 5 records based on the descending order of the Total Proteins column.