

# Data Science Workshop-1 (CSE 2195)

## ASSIGNMENT-6: PANDAS

- Convert the list ['100', '200', 'python', '300.12', '400'] to a Series object than sort it. The output should be ['100', '200', '300.12', '400', 'python'].
- Write a Pandas program to convert the first column of a DataFrame as a Series.
- Create a data frame containing names of first five presidents of India, and years served by them. Once the data frame is created, add another column containing their home state for the first four presidents.
- Convert s1=[1,2,3,4,2] and s2=[3,4,5,6] to two series objects. Find elements in s1, which are not present in s2.
- Write a Pandas program to find the index of the first occurrence of the smallest and largest value of a given series. If the input is [1,1, 3, 7,88, 12, 88, 23, 3, 1, 9, 0], the output should be 0 and 4.
- Convert L=['Cry', 'Apple', 'Orange', 'Sky', 'Banana'] to a pandas series. Create a new series with the elements which has a vowel. Create another series which starts with a vowel.
- Write a Pandas program to calculate the number of characters in each word in a given series.
- Create a 4\*4 data frame where each entry is from the standard normal distribution. Create another matrix which is element wise square root of the previous matrix. Find the number of NaN values in each column. Find the mean and standard deviation of each column.
- Create a 4\*4 matrix, where each row contains Student's name, registration no, sec and cgpa.  
D='Name':['A','B','C','D'], 'Reg no':[1,2,3,4], 'Sec':['i','ii','iii','iv'], 'CGPA':[7,8,9,10] Find the name, registration no of the student with highest cgpa.(Use np.max on cgpa column) What is the row and column number of the cell with the highest CGPA? (Convert dataframe to numpy array and see which indices has maximum value)
- Create a dataframe object, df containing some Nan values. Count the number of missing values in each column of df. Which column has the maximum number of missing values?(create a function to count nan values in a column, use df.apply and that function)
- Create a 4\*4 matrix, entries from np.arange(16). Reverse the entries of the matrix. The matrix will look like 
$$\begin{bmatrix} 15 & 14 & 13 & 12 \\ 11 & 10 & 9 & 8 \\ 7 & 6 & 5 & 4 \\ 3 & 2 & 1 & 0 \end{bmatrix}.$$
- Create a 4\*4 matrix where each entries are from standard normal distribution. Add a column to this data frame, which contains row wise minimums.
- Create a 4\*4 matrix entries from standard normal distribution. Add another row where all entries are NaN. Replace missing values in 'col1' and 'col2' columns with their respective mean.