Assignment - Data Cleaning & Wrangling

✔ Create following data:

data = pd.DataFrame([[2., 4, 3.], [1., 5, NA],[NA, 2, NA], [NA, 6.5, 3.]])

Use above data and do the following:

1. identify missing values and give Boolean output

2. Replace the missing values

3. Replace missing values with 0

4. Replace missing values with the mean of each column

✔ Create DataFrame with two columns as “Column1’ and ‘Column2”. Use numpy to create random 10 numbers and fill in the two columns as mentioned above. After the data creation, check if there are any duplicated values. If any remove the same.

✔ Create data = pd.Series([1., -888., 2., -999., -1000., 3.]). Replace all the values in sequence with values to be 8,2,4 respectively.

✔ df1 = pd.DataFrame({'key': ['g', 'b', 'a', 'c', 'a', 'b', 'b'],'data1': range(7)}).

df2 = pd.DataFrame({'key': ['a', 'b', 'd'],'data2': range(3)}).

Merge above two dataset and do Inner Join, Outer Join, Left Join and Left Join separately.

✔ Use data = pd.DataFrame(np.arange(6).reshape((2, 3)),

index=pd.Index(['Kolkata', 'Chennai'], name='state'), columns=pd.Index(['one', 'two', 'three'], name='number'))

Perform stacking and unstacking based on state name and then based on columns.