**Lab # 9**

**Dataset Preprocessing and Scaling techniques**

**OBJECTIVE**

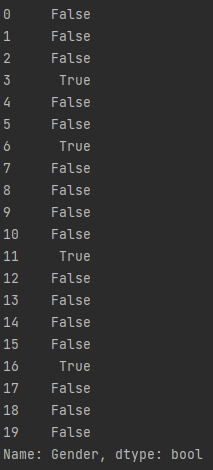
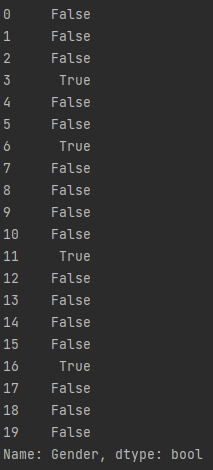
Checking the data set for missing values and outliers. Implementing Normalization and Standardization techniques to scale the values.

**Lab Tasks:**

1. Write a python code to fill all the null values in Gender column of employees.csv with “No Gender”. Print the first 10 to 30 rows of the data frame for visualization.

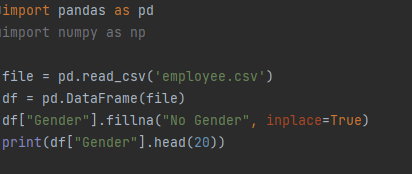
Employee.csv:

Gender:

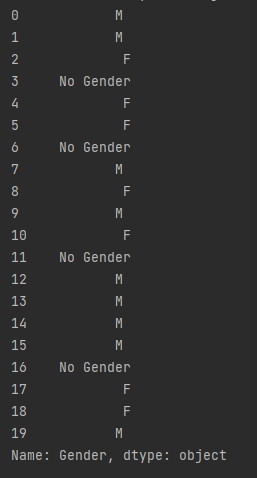
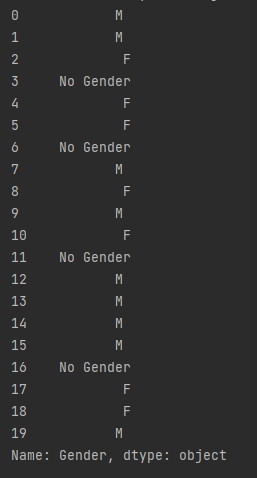
 

“True” shows that specific rows does not have any value.

Code:

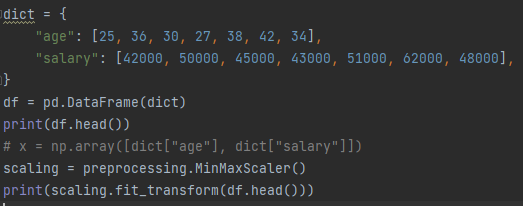


Output:

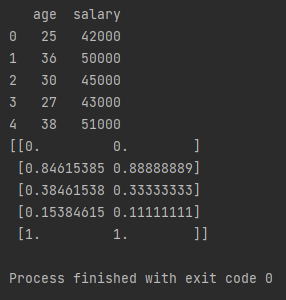
 

1. Write a python code to scale the values of features (Age and Salary) using Min-Max Normalization technique. Verify your answers by applying the formula mentioned above.

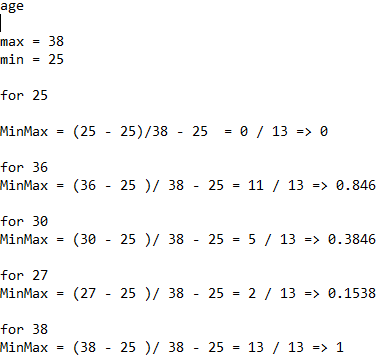
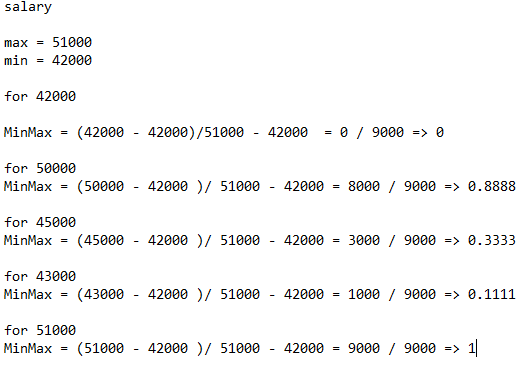
Code:



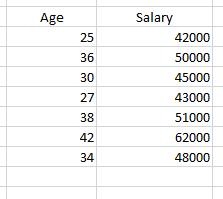
Output:



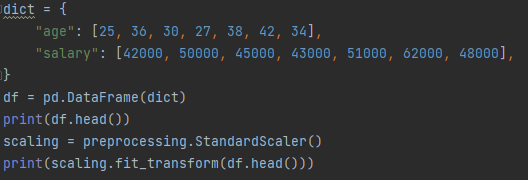
Verification:



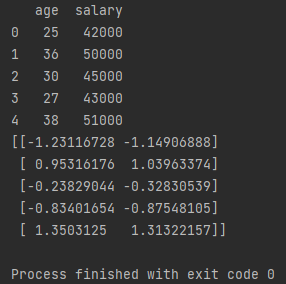
1. Write a python code to scale the values of features (Age and Salary) using Standardization technique. Verify your answers by applying the formula mentioned above.



Code:



Output:

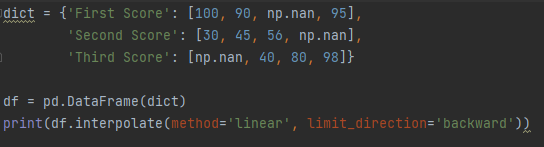


Verification:

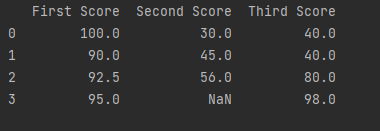
1. Given this dictionary, create a dataframe from dictionary and interpolate the missing values using backward interpolation. Hint: use interpolate().

dict = {**'First Score'**: [100, 90, np.nan, 95],  
 **'Second Score'**: [30, 45, 56, np.nan],  
 **'Third Score'**: [np.nan, 40, 80, 98]}

Code:



Output:



You will notice NaN in 4th row of Second Score column. It is because we interpolate using backward interpolation and the last row could not get filled as no row exists after it.