treating-missing-values-1

April 15, 2025

```python
[1]: import pandas as pd
     import numpy as np
```

```python
[2]: dataset = pd.read_excel('/content/Employee_Data.xls')
```

```python
[3]: dataset.isnull().sum()
```

```
[3]: Employee Id     0
     First Name      0
     Last Name       0
     Department      9
     Age             6
     Experience     13
     Salary          0
     dtype: int64
```

```python
[4]: x=dataset.iloc[:,3:6].values
```

```python
[5]: x
```

```
[5]: array([['Asset Management', 38.0, 23.0],
            ['Customer Service', 26.0, 10.0],
            ['Human Resources', nan, 14.0],
            ['Public Relations', nan, 12.0],
            ['Customer Service', nan, 13.0],
            ['Advertising', nan, 9.0],
            ['Research and Development', nan, 24.0],
            ['Asset Management', nan, 2.0],
            ['Legal Department', 40.0, 30.0],
            ['Advertising', 24.0, 23.0],
            ['Legal Department', 35.0, 3.0],
            [nan, 31.0, nan],
            [nan, 58.0, nan],
            [nan, 51.0, nan],
            [nan, 27.0, nan],
            [nan, 54.0, nan],
            [nan, 55.0, nan],
```

```
['Media Relations', 58.0, nan],
['Research and Development', 59.0, nan],
['Media Relations', 45.0, 4.0],
['Sales and Marketing', 50.0, 2.0],
['Advertising', 55.0, 19.0],
['Tech Support', 40.0, 20.0],
['Advertising', 51.0, 21.0],
['Media Relations', 24.0, 12.0],
['Asset Management', 57.0, 25.0],
['Legal Department', 24.0, 9.0],
['Public Relations', 46.0, 6.0],
['Asset Management', 29.0, 26.0],
['Customer Relations', 46.0, 13.0],
['Quality Assurance', 30.0, 29.0],
['Accounting', 57.0, nan],
[nan, 47.0, nan],
[nan, 41.0, nan],
[nan, 50.0, nan],
['Public Relations', 49.0, nan],
['Customer Relations', 28.0, 6.0],
['Public Relations', 44.0, 29.0],
['Finances', 40.0, 19.0],
['Advertising', 30.0, 25.0],
['Finances', 56.0, 21.0],
['Accounting', 26.0, 24.0],
['Finances', 44.0, 16.0],
['Payroll', 36.0, 28.0],
['Tech Support', 43.0, 24.0],
['Finances', 47.0, 15.0],
['Finances', 29.0, 11.0],
['Customer Service', 52.0, 7.0],
['Customer Relations', 42.0, 23.0],
['Tech Support', 44.0, 5.0],
['Finances', 23.0, 8.0],
['Payroll', 57.0, 16.0],
['Asset Management', 40.0, 9.0],
['Media Relations', 28.0, 24.0],
['Advertising', 53.0, 7.0],
['Accounting', 56.0, 3.0],
['Customer Relations', 48.0, 20.0],
['Customer Service', 54.0, 7.0],
['Media Relations', 37.0, 2.0],
['Advertising', 57.0, 8.0],
['Human Resources', 50.0, 17.0],
['Legal Department', 46.0, 9.0],
['Media Relations', 28.0, 3.0],
['Legal Department', 26.0, 14.0],
```

```
       ['Payroll', 54.0, 11.0],
       ['Public Relations', 31.0, 18.0],
       ['Legal Department', 33.0, 2.0],
       ['Media Relations', 48.0, 27.0],
       ['Quality Assurance', 41.0, 20.0],
       ['Human Resources', 44.0, 1.0],
       ['Tech Support', 38.0, 28.0],
       ['Accounting', 34.0, 11.0],
       ['Quality Assurance', 28.0, 27.0],
       ['Advertising', 23.0, 23.0],
       ['Customer Relations', 52.0, 12.0],
       ['Finances', 33.0, 22.0],
       ['Human Resources', 45.0, 28.0],
       ['Legal Department', 37.0, 30.0],
       ['Research and Development', 26.0, 27.0],
       ['Asset Management', 54.0, 19.0],
       ['Tech Support', 51.0, 29.0],
       ['Quality Assurance', 26.0, 20.0],
       ['Accounting', 26.0, 10.0],
       ['Customer Service', 35.0, 4.0],
       ['Sales and Marketing', 38.0, 23.0],
       ['Tech Support', 33.0, 14.0],
       ['Advertising', 38.0, 17.0],
       ['Accounting', 43.0, 21.0],
       ['Accounting', 32.0, 7.0],
       ['Media Relations', 32.0, 15.0],
       ['Human Resources', 51.0, 7.0],
       ['Public Relations', 25.0, 4.0],
       ['Asset Management', 58.0, 23.0],
       ['Accounting', 41.0, 25.0],
       ['Finances', 25.0, 24.0],
       ['Tech Support', 52.0, 16.0],
       ['Customer Relations', 52.0, 3.0],
       ['Quality Assurance', 30.0, 5.0],
       ['Research and Development', 32.0, 27.0],
       ['Quality Assurance', 46.0, 24.0]], dtype=object)
```

```python
[6]: from sklearn.impute import SimpleImputer
     imputer = SimpleImputer(missing_values=np.nan, strategy='mean')

     # Fit the imputer to the desired columns (here columns 1 to 3)
     imputer.fit(x[:, 1:3])

     # Transform the dataset by replacing missing values with the mean
     x[:, 1:3] = imputer.transform(x[:, 1:3])
```

```python
[7]: imputer = SimpleImputer(missing_values=np.nan, strategy='median')

     # Fit the imputer to the desired columns (here columns 1 to 3)
     imputer.fit(x[:, 1:3])

     # Transform the dataset by replacing missing values with the mean
     x[:, 1:3] = imputer.transform(x[:, 1:3])
```

```python
[8]: x
```

```
[8]: array([['Asset Management', 38.0, 23.0],
            ['Customer Service', 26.0, 10.0],
            ['Human Resources', 40.93617021276596, 14.0],
            ['Public Relations', 40.93617021276596, 12.0],
            ['Customer Service', 40.93617021276596, 13.0],
            ['Advertising', 40.93617021276596, 9.0],
            ['Research and Development', 40.93617021276596, 24.0],
            ['Asset Management', 40.93617021276596, 2.0],
            ['Legal Department', 40.0, 30.0],
            ['Advertising', 24.0, 23.0],
            ['Legal Department', 35.0, 3.0],
            [nan, 31.0, 15.781609195402298],
            [nan, 58.0, 15.781609195402298],
            [nan, 51.0, 15.781609195402298],
            [nan, 27.0, 15.781609195402298],
            [nan, 54.0, 15.781609195402298],
            [nan, 55.0, 15.781609195402298],
            ['Media Relations', 58.0, 15.781609195402298],
            ['Research and Development', 59.0, 15.781609195402298],
            ['Media Relations', 45.0, 4.0],
            ['Sales and Marketing', 50.0, 2.0],
            ['Advertising', 55.0, 19.0],
            ['Tech Support', 40.0, 20.0],
            ['Advertising', 51.0, 21.0],
            ['Media Relations', 24.0, 12.0],
            ['Asset Management', 57.0, 25.0],
            ['Legal Department', 24.0, 9.0],
            ['Public Relations', 46.0, 6.0],
            ['Asset Management', 29.0, 26.0],
            ['Customer Relations', 46.0, 13.0],
            ['Quality Assurance', 30.0, 29.0],
            ['Accounting', 57.0, 15.781609195402298],
            [nan, 47.0, 15.781609195402298],
            [nan, 41.0, 15.781609195402298],
            [nan, 50.0, 15.781609195402298],
            ['Public Relations', 49.0, 15.781609195402298],
            ['Customer Relations', 28.0, 6.0],
```

```
['Public Relations', 44.0, 29.0],
['Finances', 40.0, 19.0],
['Advertising', 30.0, 25.0],
['Finances', 56.0, 21.0],
['Accounting', 26.0, 24.0],
['Finances', 44.0, 16.0],
['Payroll', 36.0, 28.0],
['Tech Support', 43.0, 24.0],
['Finances', 47.0, 15.0],
['Finances', 29.0, 11.0],
['Customer Service', 52.0, 7.0],
['Customer Relations', 42.0, 23.0],
['Tech Support', 44.0, 5.0],
['Finances', 23.0, 8.0],
['Payroll', 57.0, 16.0],
['Asset Management', 40.0, 9.0],
['Media Relations', 28.0, 24.0],
['Advertising', 53.0, 7.0],
['Accounting', 56.0, 3.0],
['Customer Relations', 48.0, 20.0],
['Customer Service', 54.0, 7.0],
['Media Relations', 37.0, 2.0],
['Advertising', 57.0, 8.0],
['Human Resources', 50.0, 17.0],
['Legal Department', 46.0, 9.0],
['Media Relations', 28.0, 3.0],
['Legal Department', 26.0, 14.0],
['Payroll', 54.0, 11.0],
['Public Relations', 31.0, 18.0],
['Legal Department', 33.0, 2.0],
['Media Relations', 48.0, 27.0],
['Quality Assurance', 41.0, 20.0],
['Human Resources', 44.0, 1.0],
['Tech Support', 38.0, 28.0],
['Accounting', 34.0, 11.0],
['Quality Assurance', 28.0, 27.0],
['Advertising', 23.0, 23.0],
['Customer Relations', 52.0, 12.0],
['Finances', 33.0, 22.0],
['Human Resources', 45.0, 28.0],
['Legal Department', 37.0, 30.0],
['Research and Development', 26.0, 27.0],
['Asset Management', 54.0, 19.0],
['Tech Support', 51.0, 29.0],
['Quality Assurance', 26.0, 20.0],
['Accounting', 26.0, 10.0],
['Customer Service', 35.0, 4.0],
```

```
['Sales and Marketing', 38.0, 23.0],
['Tech Support', 33.0, 14.0],
['Advertising', 38.0, 17.0],
['Accounting', 43.0, 21.0],
['Accounting', 32.0, 7.0],
['Media Relations', 32.0, 15.0],
['Human Resources', 51.0, 7.0],
['Public Relations', 25.0, 4.0],
['Asset Management', 58.0, 23.0],
['Accounting', 41.0, 25.0],
['Finances', 25.0, 24.0],
['Tech Support', 52.0, 16.0],
['Customer Relations', 52.0, 3.0],
['Quality Assurance', 30.0, 5.0],
['Research and Development', 32.0, 27.0],
['Quality Assurance', 46.0, 24.0]], dtype=object)
```