Ruthvek 22070521031

```
[57]: dataset.isnull().sum()
```

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

```
[58]: dataset
```

[58]:

	Employee Id	First Name	Last Name	Department	Age	Experience	Salary
0	1	Camden	Savage	Asset Management	38.0	23.0	46975
1	2	Justina	Mcconnell	Customer Service	26.0	10.0	30711
2	3	Stewart	Moon	Human Resources	NaN	14.0	25049
3	4	Tobias	Fletcher	Public Relations	NaN	12.0	33309
4	5	Raja	Sims	Customer Service	NaN	13.0	25274
...	...	...	...	...	...	...	...
95	96	Nerea	Haney	Tech Support	52.0	16.0	17756
96	97	Nadine	Franks	Customer Relations	52.0	3.0	31886
97	98	Aquila	Jacobson	Quality Assurance	30.0	5.0	39812
98	99	Wang	Doyle	Research and Development	32.0	27.0	30742
99	100	Beau	Richardson	Quality Assurance	46.0	24.0	30128

100 rows × 7 columns

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

```
[60]: x
```

```
[60]: 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],
```

Ruthvek 22070521031

```
[77]: from sklearn.impute import SimpleImputer
      imputer=SimpleImputer(missing_values='NaN',strategy='median')
      imputer =SimpleImputer().fit(x[:,1:3])
      x[:,1:3]=imputer.transform(x[:,1:3])
      print(x)
```

```
[['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]
```

```
[81]: from sklearn.impute import SimpleImputer
      imputer=SimpleImputer(missing_values='NaN',strategy='most_frequent')
      imputer =SimpleImputer().fit(x[:,1:3])
      x[:,1:3]=imputer.transform(x[:,1:3])
      print(x)
```

```
[['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]
```