Binarization methods on given data set

In [1]:
```
# Python code explaining how
# to Binarize feature values

""" PART 1
    Importing Libraries """

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# Sklearn library
from sklearn import preprocessing
```

In [2]:
```
data_set =  pd.read_csv(r"C:\Users\prati\Desktop\Data_for_Missing_Values.csv")
data_set.head()
```

Out[2]:

	Country	Age	Salary	Purchased
0	France	44	72000	No
1	Spain	27	48000	Yes
2	Germany	30	54000	No
3	Spain	38	61000	No
4	Germany	40	54000	Yes

In [3]:
```
# here Features - Age and Salary columns
# are taken using slicing
# to binarize values
age = data_set.iloc[:, 1].values
salary = data_set.iloc[:, 2].values
print ("\nOriginal age data values : \n",  age)
print ("\nOriginal salary data values : \n",  salary)
```

```
Original age data values :
 [44 27 30 38 40 35 38 48 50 37]

Original salary data values :
 [72000 48000 54000 61000 54000 58000 52000 79000 83000 67000]
```

In [4]:
```
""" PART 4
    Binarizing values """

from sklearn.preprocessing import Binarizer

x = age
x = x.reshape(1, -1)
y = salary
y = y.reshape(1, -1)

# For age, let threshold be 35
# For salary, let threshold be 61000
binarizer_1 = Binarizer(threshold=35, copy=True)
binarizer_2 = Binarizer(threshold=6100, copy=True)
#binarizer_1.fit_transform(x)
```

```
#binarizer_1 = Binarizer(35)
#binarizer_2 = Binarizer(61000)
# Transformed feature
print ("\nBinarized age : \n", binarizer_1.fit_transform(x))
print ("\nBinarized salary : \n", binarizer_2.fit_transform(y))
```

```
Binarized age :
 [[1 0 0 1 1 0 1 1 1 1]]

Binarized salary :
 [[1 1 1 1 1 1 1 1 1 1]]
```

In [ ]: