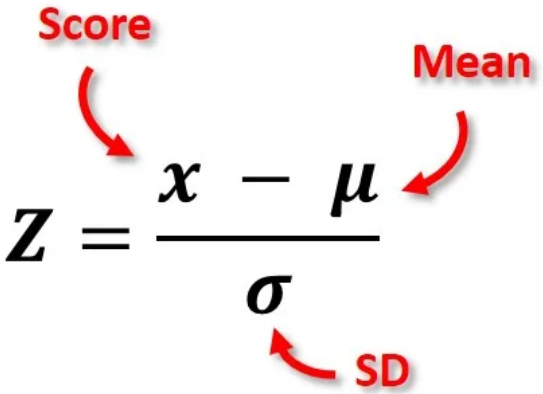
In [2]:
```
# import the packages
# read the data
# divide into numerical and categorical
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

file_path=r'C:\Users\omkar\OneDrive\Documents\Gen_AI\Data_files\Visadataset.csv'
visa_df=pd.read_csv(file_path)

cat=visa_df.select_dtypes(include='object').columns
num=visa_df.select_dtypes(exclude='object').columns
```

- Scaling is the One of Most important step before Model development
- Scaling means makes all the columns under one scale
- Scaling used to make all the columns or features comparable
- Some ML models works on Distance methods
    - Example age min: 0 max:100
    - Income might be lakhs crores so much bigger values
    - If we dont make age and income under one scale ML model treats Income is the Important varaible
    - When values are huge maths makes more complex so it is better to do lower down the values
    - All the features under one scale so easy to compare
        - Dollars and Ruppes we can not compare becuase two are different scales

Standard Scalar

- Z scale makes mean=0 and standard deviation always =1



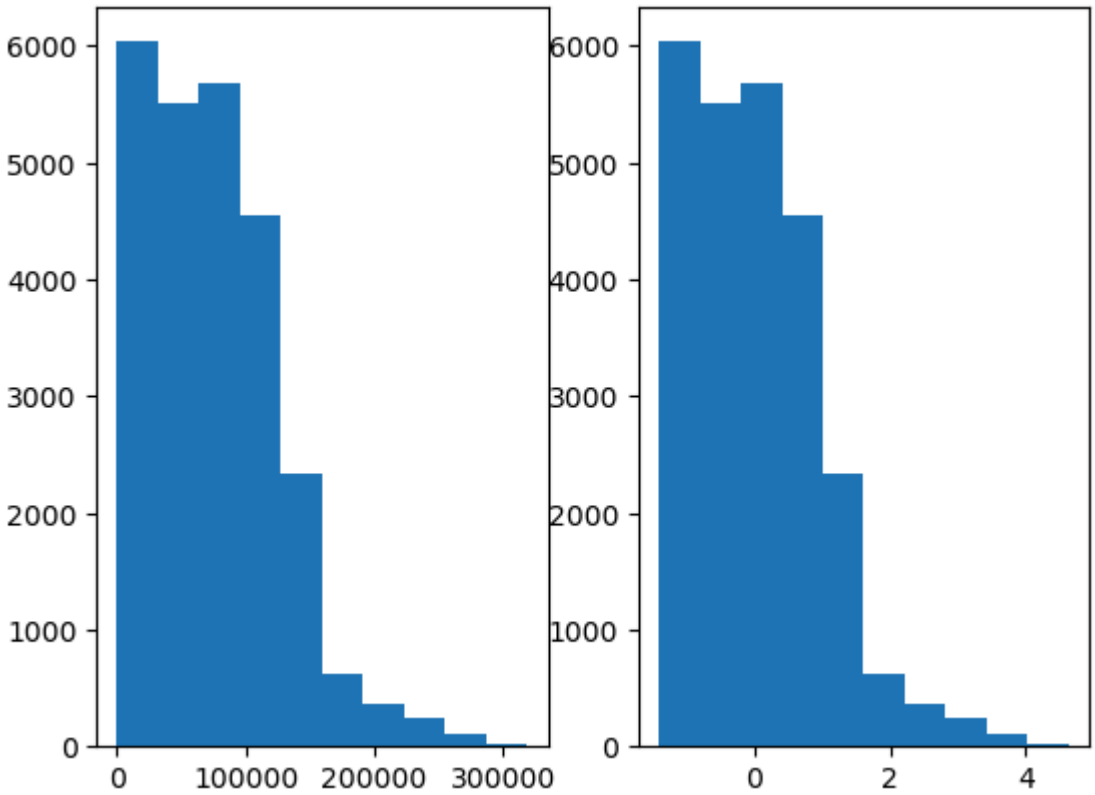
In [ ]:
```
# step-1: x= prevailawage data
# step-2: mean= mean of prevailange data
# step-3: sd = sd of prevailange data
# step-4: step1-step2 = x-mean
# step5: step4/step3  = (x-mean)/sd
```

In [6]:
```
x=visa_df['prevailing_wage']
x_mean=visa_df['prevailing_wage'].mean() # fit
x_sd=visa_df['prevailing_wage'].std()   # fit
wage_z=(x-x_mean)/x_sd # transform
```

In [12]:
```
wage_z.mean(),wage_z.std()
```

Out[12]:
```
(8.421660368899147e-17, 1.000000000000007)
```

In [16]:
```
plt.subplot(1,2,1).hist(x)
plt.subplot(1,2,2).hist(wage_z)
plt.show()
```

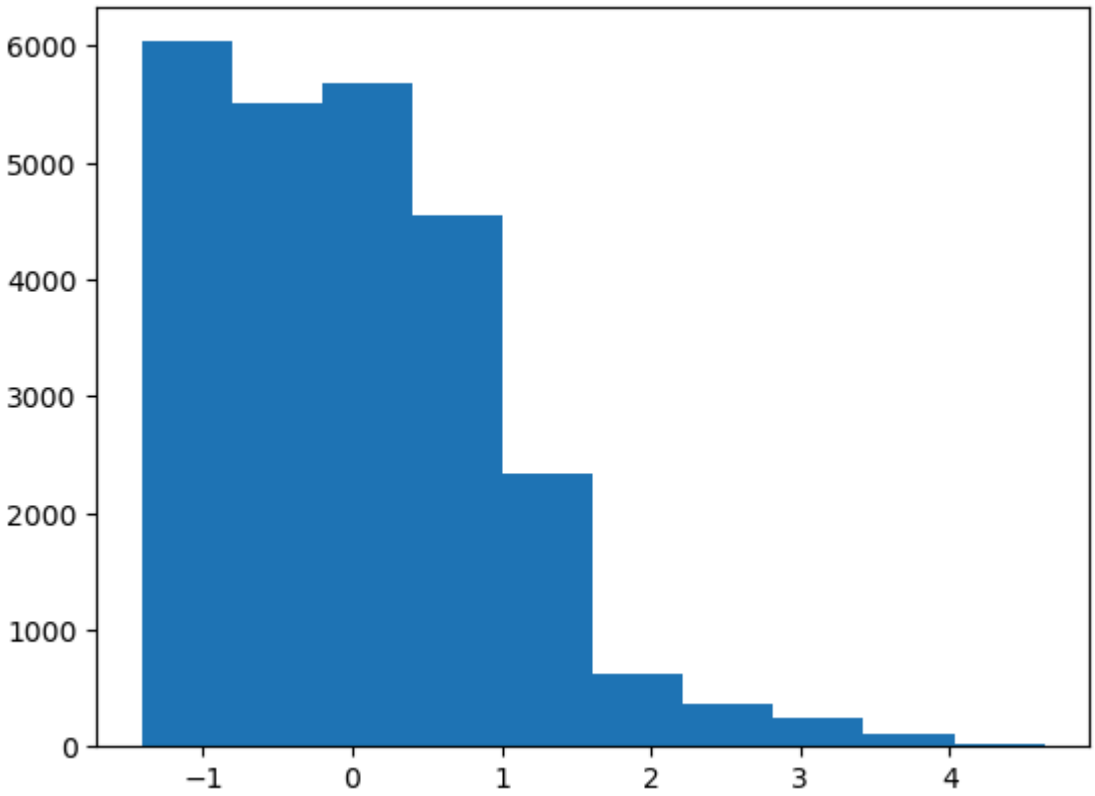


- sklearn
  - preprocessing
    - StandardScalar

In [33]:
```
wage_data=visa_df[['prevailing_wage']]
from sklearn.preprocessing import StandardScaler
ss=StandardScaler()
wage_ss=ss.fit_transform(wage_data)
```

In [35]:
```
plt.hist(wage_ss)
```

Out[35]:
```
(array([6038., 5504., 5681., 4551., 2334.,  624.,  373.,  240.,  114.,
         21.]),
 array([-1.40970956, -0.80531933, -0.20092909,  0.40346114,  1.00785137,
         1.61224161,  2.21663184,  2.82102207,  3.42541231,  4.02980254,
         4.63419278]),
 <BarContainer object of 10 artists>)
```



In [ ]:
```
# Assignemnet
apply MinMax Scale or Normalization on wage data
do based on formulae
do based on package
draw the histogram
min=0 max=1
```

In [ ]:
```
wage_data=visa_df[['prevailing_wage']]

from sklearn.preprocessing import StandardScaler
ss=StandardScaler()
wage_ss=ss.fit_transform[wage_data]

method function ()
[]  access the values
```