Encoding: 1.Converting catagorical data into numerical data

One-Hot-Encoders

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import OneHotEncoder

df = pd.read_csv(r"C:\Mypythonfiles\Salary_EDA.csv")
df.head()
```

```
    Age  Gender Education Level          Job Title  Years of
Experience  \
0  32.0    Male      Bachelor's  Software Engineer
5.0
1  28.0  Female        Master's       Data Analyst
3.0
2  45.0    Male             PhD     Senior Manager
15.0
3  36.0  Female      Bachelor's    Sales Associate
7.0
4  36.0  Female      Bachelor's    Sales Associate
7.0

     Salary
0   90000.0
1   65000.0
2  150000.0
3   60000.0
4   60000.0
```

Filter Categorical features

```
categorical_cols = ['Education Level']
```

Define and Apply Encoder

```
encoder = OneHotEncoder(drop=None,sparse_output=False)

encoded_data = encoder.fit_transform(df[categorical_cols])

print(encoded_data)
```

```
[[1. 0. 0. 0.]
 [0. 1. 0. 0.]
```

```
 [0. 0. 1. 0.]
 ...
 [1. 0. 0. 0.]
 [1. 0. 0. 0.]
 [0. 0. 1. 0.]]
```

the encoded data is in the form of array.Now we need to convert the encoded features into data frame with categories as column names

```
encoded_df = pd.DataFrame(encoded_data,
columns=encoder.get_feature_names_out(categorical_cols))
encoded_df.head()
```

```
   Education Level_Bachelor's  Education Level_Master's  Education
Level_PhD  \
0                         1.0                       0.0
0.0
1                         0.0                       1.0
0.0
2                         0.0                       0.0
1.0
3                         1.0                       0.0
0.0
4                         1.0                       0.0
0.0

   Education Level_nan
0                  0.0
1                  0.0
2                  0.0
3                  0.0
4                  0.0
```

```
encoded_df.drop('Education Level_nan',axis=1,inplace = True)
encoded_df.head()
```

```
   Education Level_Bachelor's  Education Level_Master's  Education
Level_PhD
0                         1.0                       0.0
0.0
1                         0.0                       1.0
0.0
2                         0.0                       0.0
1.0
3                         1.0                       0.0
0.0
4                         1.0                       0.0
0.0
```

```
final_df = pd.concat([df,encoded_df],axis=1)
final_df.head()
```

```
    Age  Gender Education Level          Job Title  Years of
Experience  \
0  32.0    Male      Bachelor's  Software Engineer
5.0
1  28.0  Female        Master's       Data Analyst
3.0
2  45.0    Male             PhD     Senior Manager
15.0
3  36.0  Female      Bachelor's    Sales Associate
7.0
4  36.0  Female      Bachelor's    Sales Associate
7.0

     Salary  Education Level_Bachelor's  Education Level_Master's  \
0   90000.0                         1.0                       0.0
1   65000.0                         0.0                       1.0
2  150000.0                         0.0                       0.0
3   60000.0                         1.0                       0.0
4   60000.0                         1.0                       0.0

   Education Level_PhD
0                  0.0
1                  0.0
2                  1.0
3                  0.0
4                  0.0
```

Label Encoder

```
from sklearn.preprocessing import LabelEncoder

df1 = pd.read_csv(r"C:\Mypythonfiles\Salary_EDA.csv")
df1.head()
```

```
    Age  Gender Education Level          Job Title  Years of
Experience  \
0  32.0    Male      Bachelor's  Software Engineer
5.0
1  28.0  Female        Master's       Data Analyst
3.0
2  45.0    Male             PhD     Senior Manager
15.0
3  36.0  Female      Bachelor's    Sales Associate
7.0
4  36.0  Female      Bachelor's    Sales Associate
```

```
7.0

     Salary
0   90000.0
1   65000.0
2  150000.0
3   60000.0
4   60000.0

le = LabelEncoder()
df1['Gender_encoded'] = le.fit_transform(df['Gender'])
df1.head()

    Age  Gender Education Level          Job Title  Years of
Experience  \
0  32.0    Male      Bachelor's  Software Engineer
5.0
1  28.0  Female        Master's       Data Analyst
3.0
2  45.0    Male             PhD     Senior Manager
15.0
3  36.0  Female      Bachelor's    Sales Associate
7.0
4  36.0  Female      Bachelor's    Sales Associate
7.0

     Salary  Gender_encoded
0   90000.0               1
1   65000.0               0
2  150000.0               1
3   60000.0               0
4   60000.0               0

le1 = LabelEncoder()
df1['Education_Level_encoded'] = le1.fit_transform(df['Education
Level'])
df1.head()

    Age  Gender Education Level          Job Title  Years of
Experience  \
0  32.0    Male      Bachelor's  Software Engineer
5.0
1  28.0  Female        Master's       Data Analyst
3.0
2  45.0    Male             PhD     Senior Manager
15.0
3  36.0  Female      Bachelor's    Sales Associate
7.0
4  36.0  Female      Bachelor's    Sales Associate
7.0
```

```
     Salary  Gender_encoded  Education_Level_encoded
0   90000.0               1                        0
1   65000.0               0                        1
2  150000.0               1                        2
3   60000.0               0                        0
4   60000.0               0                        0
```

```
from sklearn.preprocessing import MinMaxScaler

df2 = pd.read_csv(r"C:\Mypythonfiles\Salary_EDA.csv")
df2.head()
```

```
    Age  Gender Education Level          Job Title  Years of
Experience  \
0  32.0    Male      Bachelor's  Software Engineer
5.0
1  28.0  Female        Master's       Data Analyst
3.0
2  45.0    Male             PhD     Senior Manager
15.0
3  36.0  Female      Bachelor's    Sales Associate
7.0
4  36.0  Female      Bachelor's    Sales Associate
7.0

     Salary
0   90000.0
1   65000.0
2  150000.0
3   60000.0
4   60000.0
```

```
ms = MinMaxScaler()
df2['Salary_scaler'] = ms.fit_transform(df[['Salary']])
df2.head()
```

```
    Age  Gender Education Level          Job Title  Years of
Experience  \
0  32.0    Male      Bachelor's  Software Engineer
5.0
1  28.0  Female        Master's       Data Analyst
3.0
2  45.0    Male             PhD     Senior Manager
15.0
3  36.0  Female      Bachelor's    Sales Associate
7.0
4  36.0  Female      Bachelor's    Sales Associate
7.0

     Salary  Salary_scaler
```

```
0   90000.0       0.359103
1   65000.0       0.258963
2  150000.0       0.599439
3   60000.0       0.238935
4   60000.0       0.238935
```

Z-Score Normalization

```
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
df2['Salary_StandardScaler']=sc.fit_transform(df[['Salary']])
df2.head()
```

```
    Age  Gender Education Level          Job Title  Years of
Experience  \
0  32.0    Male      Bachelor's  Software Engineer
5.0
1  28.0  Female        Master's       Data Analyst
3.0
2  45.0    Male             PhD     Senior Manager
15.0
3  36.0  Female      Bachelor's    Sales Associate
7.0
4  36.0  Female      Bachelor's    Sales Associate
7.0

     Salary  Salary_scaler  Salary_StandardScaler
0   90000.0       0.359103              -0.211488
1   65000.0       0.258963              -0.733148
2  150000.0       0.599439               1.040496
3   60000.0       0.238935              -0.837480
4   60000.0       0.238935              -0.837480
```