# One-Hot-Encoding

Encoding -convert catagerical data into numarical data

```
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:\Users\Bhoomika.G\OneDrive\Documents\
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 catogiracal feature

```
catogarical_cols=['Education Level']  # fit_transform-it convert the 
non-numirical data to numirical data
```

# Define and apply the encoder

```
encoder=OneHotEncoder(drop=None,sparse_output=False) # it not drop 
hole data in that it show the unique

encoded_data=encoder.fit_transform(df[catogarical_cols]) # modify the 
original row in matrix (convert numarical type)
```

#the encodede data is in the form of array dot now we need to convert the ecoded data into dataframe with catageriy as follows the name Encode dataframes

```
encoded_df=pd.DataFrame(encoded_data,columns=encoder.get_feature_names
_out(catogarical_cols)) # get_feature_names_out it take unique values
the help to encoder

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(columns=['Education Level_nan'], inplace=True)
# it used to drop the columns

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

Fdf=pd.concat([df,encoded_df], axis=1)  #axis=1 column in
pandas,axix=0 row in pandas revers in numpy
Fdf.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 encoding

```
from sklearn.preprocessing import LabelEncoder

df1=pd.read_csv(r"C:\Users\Bhoomika.G\OneDrive\Documents\
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
```

```
le1=LabelEncoder()
df1['Gender_encoder']=le1.fit_transform(df1['Gender']) #fit_transform
is used to convert the  0/1 value assened insted of male and female
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_encoder
0   90000.0               1
1   65000.0               0
2  150000.0               1
3   60000.0               0
4   60000.0               0
```

```
le2=LabelEncoder()
df1['Education level encoded']=le2.fit_transform(df1['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_encoder  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
```

# Scaling

```
it will convert the bigger value to smaller value by dividing bigger
value (up to 0 to 1)

from sklearn.preprocessing import MinMaxScaler

df2=pd.read_csv(r"C:\Users\Bhoomika.G\OneDrive\Documents\
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

sk1= MinMaxScaler()
df2['Salary_skla']=sk1.fit_transform(df2[['Salary']]) #fit_transform
is used to convert the  0/1 value assened insted of male and female
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_skla
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
```

```python
from sklearn.preprocessing import StandardScaler
```

```python
df3=pd.read_csv(r"C:\Users\Bhoomika.G\OneDrive\Documents\
Salary_EDA.csv")
df3.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
```

```python
ss=StandardScaler()
df3['Salary_Stander']=ss.fit_transform(df3[['Salary']]) #fit_transform
is used to convert the  0/1 value assened insted of male and female
df3.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_Stander
0   90000.0       -0.211488
1   65000.0       -0.733148
2  150000.0        1.040496
3   60000.0       -0.837480
4   60000.0       -0.837480
```