

ONE-HOT-ENCODING

```
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("C:\my files\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 1) encoder = non numerical data into numerical data

```
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 form of array.now we need to encoded data into a data frame with convert the encoded features into a dataframe with categories as column name

```
encoded_df =
pd.DataFrame(encoded_data,columns=encoder.get_feature_names_out(catego
rical_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(columns=["Education Level_nan"],inplace = True,axis =
1)
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("C:\my files\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
```

```
myle1 = LabelEncoder()
df1["Gender_encoded"] = myle1.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

```
myle2 = LabelEncoder()
df1["educationlevel_encoded"] = myle2.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	educationlevel_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

standardization

Min Max Scaler

```
from sklearn.preprocessing import MinMaxScaler
```

```
df2 = pd.read_csv("C:\my files\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

```
mys1 = MinMaxScaler()
df2["Salary_scaler"] = mys1.fit_transform(df2[["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
```

```
Normalizer = StandardScaler()
df2[["Salary_Standardized"]] =
Normalizer.fit_transform(df2[['Salary']])
df2[["Salary", "Salary_Standardized"]]
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_Standardized
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