

Untitled

January 3, 2024

#NOMINAL OR OHE

```
[1]: import pandas as pd
      from sklearn.preprocessing import OneHotEncoder
```

```
[2]: # creating a dataframe with a categorical feature
      df=pd.DataFrame({'color':['red','blue','green','green','red','blue']
                       })
```

```
[3]: df
```

```
[3]:   color
0    red
1   blue
2  green
3  green
4    red
5   blue
```

```
[4]: #Creating the instance of the One hot encoder
      encoder =OneHotEncoder()
```

```
[11]: encoded=encoder.fit_transform(df[['color']])
```

```
[13]: import pandas as pd
      encoded_df=pd.DataFrame(encoded.toarray(),columns=encoder.
      ↪get_feature_names_out())
```

```
[15]: pd.concat([df,encoded_df],axis=1)
```

```
[15]:   color  color_blue  color_green  color_red
0    red           0.0           0.0          1.0
1   blue           1.0           0.0          0.0
2  green           0.0           1.0          0.0
3  green           0.0           1.0          0.0
4    red           0.0           0.0          1.0
5   blue           1.0           0.0          0.0
```

```
[16]: import seaborn as sns
```

```
[20]: df1=sns.load_dataset('tips')
```

```
[21]: df1
```

```
[21]:
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
..
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

[244 rows x 7 columns]

```
[24]: encoded1=encoder.fit_transform(df1[['sex']])
```

```
[26]: encoded_df1=pd.DataFrame(encoded1.toarray(),columns=encoder.  
    ↪get_feature_names_out())
```

```
[27]: encoded_df1
```

```
[27]:
```

	sex_Female	sex_Male
0	1.0	0.0
1	0.0	1.0
2	0.0	1.0
3	0.0	1.0
4	1.0	0.0
..
239	0.0	1.0
240	1.0	0.0
241	0.0	1.0
242	0.0	1.0
243	1.0	0.0

[244 rows x 2 columns]

```
[28]: pd.concat([df1,encoded_df1],axis=1)
```

```
[28]:
```

	total_bill	tip	sex	smoker	day	time	size	sex_Female	sex_Male
0	16.99	1.01	Female	No	Sun	Dinner	2	1.0	0.0

1	10.34	1.66	Male	No	Sun	Dinner	3	0.0	1.0
2	21.01	3.50	Male	No	Sun	Dinner	3	0.0	1.0
3	23.68	3.31	Male	No	Sun	Dinner	2	0.0	1.0
4	24.59	3.61	Female	No	Sun	Dinner	4	1.0	0.0
..
239	29.03	5.92	Male	No	Sat	Dinner	3	0.0	1.0
240	27.18	2.00	Female	Yes	Sat	Dinner	2	1.0	0.0
241	22.67	2.00	Male	Yes	Sat	Dinner	2	0.0	1.0
242	17.82	1.75	Male	No	Sat	Dinner	2	0.0	1.0
243	18.78	3.00	Female	No	Thur	Dinner	2	1.0	0.0

[244 rows x 9 columns]

[]: