

## 4.0-Target Guided Ordinal Encoding

January 3, 2024

### 0.1 Target Guided Ordinal Encoding

It is a technique used to encode categorical variables based on their relationship with the target variable. This encoding technique is useful when we have a categorical variable with a large number of unique categories, and we want to use this variable as a feature in our machine learning model.

In Target Guided Ordinal Encoding, we replace each category in the categorical variable with a numerical value based on the mean or median of the target variable for that category. This creates a monotonic relationship between the categorical variable and the target variable, which can improve the predictive power of our model.

```
[1]: import pandas as pd

# create a sample dataframe with a categorical variable and a target variable
df = pd.DataFrame({
    'city': ['New York', 'London', 'Paris', 'Tokyo', 'New York', 'Paris'],
    'price': [200, 150, 300, 250, 180, 320]
})
```

```
[5]: df
```

```
[5]:
```

	city	price
0	New York	200
1	London	150
2	Paris	300
3	Tokyo	250
4	New York	180
5	Paris	320

```
[8]: ## calculate the mean price for each city
mean_price=df.groupby('city')['price'].mean().to_dict()
mean_price
```

```
[8]: {'London': 150.0, 'New York': 190.0, 'Paris': 310.0, 'Tokyo': 250.0}
```

```
[9]: ## replace each city with its mean price
df['city_encoded']=df['city'].map(mean_price)
```

```
[10]: df
```

```
[10]:      city  price  city_encoded
0  New York    200         190.0
1   London    150         150.0
2   Paris     300         310.0
3   Tokyo     250         250.0
4  New York    180         190.0
5   Paris     320         310.0
```

```
[12]: import seaborn as sns

sns.load_dataset('tips')
```

```
[12]:      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]

```
[ ]: 
[ ]: 
[ ]: 
[ ]: 
[ ]: 
[ ]: 
[ ]: 
[ ]:
```