

```
In [2]: """
Que-1What is the difference between Ordinal Encoding and Label Encoding?
Provide an example of when you might choose one over the other.
Ans-1
Ordinal Encoding-
should be used for ordinal variables (where order matters, like cold, warm, hot)

Label Encoding-
Label encoding should be used for non-ordinal (aka nominal) variables
(where order doesn't matter, like blonde, brunette)

Que-2Explain how Target Guided Ordinal Encoding works and provide an example of
when you might use it in a machine learning project.
Ans-2
In this method, we calculate the mean of each categorical variable based on the
output and then rank them

we need to use this technique when our feature strongly correlate with the Target
variable

Que-3 Define covariance and explain why it is important in statistical analysis
How is covariance calculated?
Ans-3Covariance indicates the relationship of two variables whenever one variable
changes. If an increase in one variable results in an increase in the other
variable, both variables are said to have a positive covariance. Decreases in
one variable also cause a decrease in the other.

it helps to understand how the one feature variable is correlated with other feature
variables

Que-4 For a dataset with the following categorical variables: Color (red, green,
blue), Size (small, medium, large), and Material (wood, metal, plastic), perform
label encoding using Python's scikit-learn library. Show your code and explain
the output.
"""

import pandas as pd
df=pd.DataFrame({"Color":["red","green","blue"],
                  "Size":["small","medium","large"],
                  "Material":["wood","metal","plastic"]})
```

```
In [3]: df.head(5)
```

```
Out[3]:
```

	Color	Size	Material
0	red	small	wood
1	green	medium	metal
2	blue	large	plastic

```
In [4]: # for Label Encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

```
In [29]: df.apply(le.fit_transform)
```

Out[29]:

	Color	Size	Material
0	2	2	2
1	1	1	0
2	0	0	1

```
"""
From the above output we see that in color-[red:2,green:1,blue:0]
Size-[small:2,medium:1,large:0]
Material-[wood:2,metal:1,plastic:0]"""
```

```
In [33]: """
Que-5 You are working on a machine learning project with a dataset containing
several categorical variables, including "Gender" (Male/Female), "Education
Level" (High School/Bachelor's/Master's/PhD),and "Employment Status"
(Unemployed/Part-Time/Full-Time). Which encoding method would you use for
each variable, and why?
Ans-5 i am using the label encoder are
for Gender variable i will use the Label Encoding because there is no level
for Education level i will used the Ordinal Encoder because there are level
for Employment status i will use the label encoder because there are no level

Que 6-You are analyzing a dataset with two continuous variables, "Temperature"
and "Humidity", and two categorical variables, "Weather Condition"
(Sunny/Cloudy/Rainy) and "Wind Direction" (North/South/East/West).
Calculate the covariance between each pair of variables and interpret the
results.
Ans 6
"""
df1=pd.DataFrame({"Temperature":[1,2,3,4],
                  "Humidity":[5.4,9.0,3.1,4.5],
                  "Weather":["Sunny", 'Cloudy', 'Rainy', 'Mid'],
                  "Wind Direction":["North", "South", "East", "West"]})
```

```
In [34]: df1.head()
```

```
Out[34]:
```

	Temperature	Humidity	Weather	Wind Direction
0	1	5.4	Sunny	North
1	2	9.0	Cloudy	South
2	3	3.1	Rainy	East
3	4	4.5	Mid	West

```
In [37]: import numpy as np
```

```
In [39]: from sklearn.preprocessing import LabelEncoder
```

```
In [40]: lr=LabelEncoder()
```

```
In [45]: df1=df1.apply(lr.fit_transform)
```

```
In [46]: # covariance of above variable are  
np.cov(df1)
```

```
Out[46]: array([[ 1.66666667, -0.33333333,  0.          , -1.33333333],  
                [-0.33333333,  1.66666667, -1.33333333,  0.          ],  
                [ 0.          , -1.33333333,  1.33333333,  0.          ],  
                [-1.33333333,  0.          ,  0.          ,  1.33333333]])
```

```
In [ ]:
```