Unit 3.1 Graded Assignment:
Muhammad Khan (2303.KHI.DEG.027)
Qadeer Hussain (2303.KHI.DEG.006)

Daily Assignment :

Implement a label encoder for categorical data using pure Python, Pandas and NumPy.

Answer:

Fist of all we import the numpy and pandas libraries, then loads the Titanic dataset using the pd.read_csv function and stores it in the variable titanic, then selects the categorical features to encode and stores them in a list named categorical_features. The 'Sex' and 'Embarked' columns are selected as categorical features.

```python
import pandas as pd
import numpy as np

# Load dataset
titanic = pd.read_csv('titanic.csv')

# Select the categorical features
categorical_features = ['Sex', 'Embarked']
printed_features =['Name','Sex', 'Embarked']


def encoder_function(df, columns):
    encoded_df = df.copy()
    for column in columns:
        unique_values = df[column].unique()
        mapping = {}
        for i, value in enumerate(unique_values):
            mapping[value] = i
        encoded_df[column] = df[column].map(mapping)
    return encoded_df

encoded_titanic = encoder_function(titanic, categorical_features)


print("Original data:")
print(titanic[printed_features].head())
print("\nEncoded data:")
print(encoded_titanic[printed_features].head())
```

A encoder_function() function that takes a DataFrame and a list of columns to encode as input, and returns a new DataFrame with the selected columns encoded with integer labels. This function is defined using a for loop that iterates over the selected columns, creates a mapping between unique categorical values and integer labels using a dictionary comprehension, and applies the mapping to the column using the map method. The encoded DataFrame is returned as a copy of the input DataFrame.

The encoder_function() function is then applied to the titanic DataFrame using the categorical_features list as the columns to encode. The resulting encoded DataFrame is stored in the variable encoded_titanic.

So finally, prints both the original and encoded data for a subset of the features. The printed_features list is defined to select the 'Name', 'Sex', and 'Embarked' columns to print. The head method is used to print the first five rows of each DataFrame.

Output:

```
Original data:
                                                Name     Sex Embarked
0                                Kelly, Mr. James    male        Q
1                Wilkes, Mrs. James (Ellen Needs)  female        S
2                       Myles, Mr. Thomas Francis    male        Q
3                                Wirz, Mr. Albert    male        S
4  Hirvonen, Mrs. Alexander (Helga E Lindqvist)  female        S

Encoded data:
                                                Name  Sex  Embarked
0                                Kelly, Mr. James    0         0
1                Wilkes, Mrs. James (Ellen Needs)    1         1
2                       Myles, Mr. Thomas Francis    0         0
3                                Wirz, Mr. Albert    0         1
4  Hirvonen, Mrs. Alexander (Helga E Lindqvist)    1         1
```