Members

Huzefa Anver (2303.KHI.DEG.002)

Syed Mohammad Anjil Hussain Rizvi (2303.KHI.DEG.031)

```python
import pandas as pd

def Label_encoding(daata) -> dataframe:
    listt = []
    for j,i in enumerate(daata.columns):

        if daata.iloc[:,j].dtype == 'O':
            d =daata[i].drop_duplicates().reset_index(drop=True)
            g =dict(zip([i for i in d],[i for i in d.index]))

            for j in daata[i]:
                listt.append(g[j])

            daata[f"Label_encoded {i}"] = listt
            listt = []
```

Defining a function that takes data frame as an input and label encoding all the columns values which are of type object (non-numerical or categorical). Columns having numerical values remain the same

Inside the function:

1st step: looping through columns of dataframe

2nd step: Checking if columns are of type object (non-numerical or categorical)

3rd step: Dropping all the duplicates of data frame column and resetting int and resetting index

4th step: creating a dictionary that takes zip function as an argument and zip function is taking two lists , 1st list is the string values of data frame columns and 2nd list is the index (note: that data frame columns here is the data frame with non duplicate values)

Zip function returns 1st value of 1st list, 1st value of second list and both to tuples

So dictionary would get tuples from zip function containing two values from each list, dictionary would appoint first value as key and second as its value

5th step: we are looping through original data frame columns having duplicates passing values as dict keys to get their values which are numerical value and then appending ppending them into list to record

```
dataframe = pd.DataFrame({"Names":["Taha","Ubaid","Ali","Huzefa","Taha","Talha","Talha","Huzefa","Talha"],
             "Department":["CS","CS","SE","SE","CS","EE","EE","SE","EE"],
             "col":[1,2,3,4,5,6,7,8,9]})

Label_encoding(dataframe)
dataframe
```

	Names	Department	col	Label_encoded Names	Label_encoded Department
0	Taha	CS	1	0	0
1	Ubaid	CS	2	1	0
2	Ali	SE	3	2	1
3	Huzefa	SE	4	3	1
4	Taha	CS	5	0	0
5	Talha	EE	6	4	2
6	Talha	EE	7	4	2
7	Huzefa	SE	8	3	1
8	Talha	EE	9	4	2

Creating a dataframe and then passing dataframe to label_encoding function as we can see it works as it encoding only categorical and not numerical columns