Name: Rahima Siddiqui(2303.KHI.DEG.030)
Peer Name: Osama Abdul Razzak(2303.KHI.DEG.029)

Assignment 3.1

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

In the beginning, I import the required libraries

```
import pandas as pd
```

```
import numpy as np
```

For implementing label encoder in categorical data, we used the ‘Iris.csv’ file
and store it into data variable

```
data = pd.read_csv('Iris.csv')
data
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

Name: Rahima Siddiqui(2303.KHI.DEG.030)
Peer Name: Osama Abdul Razzak(2303.KHI.DEG.029)

After that, we analyzed the target class ‘Species’ and store this class into ‘target’ variable

```
target = data['Species']
target
```

```
0         Iris-setosa
1         Iris-setosa
2         Iris-setosa
3         Iris-setosa
4         Iris-setosa
            ...
145    Iris-virginica
146    Iris-virginica
147    Iris-virginica
148    Iris-virginica
149    Iris-virginica
Name: Species, Length: 150, dtype: object
```

then we isolate the unique “Species” by using numpy.unique(‘target’) and introduce ‘label_Species’ as dict

```
unique_Species = np.unique(target)
unique_Species
```

```
array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

then we labeled to each unique species

```
for i in range(len(unique_Species)):
    label_Species[unique_Species[i]] = i
label_Species
```

and map it into ‘target’ variable and store it into ‘encode label’

```
: encode_label = target.map(label_Species)
  encode_label
```

```
: 0      0
  1      0
  2      0
  3      0
  4      0
        ..
  145    2
  146    2
  147    2
  148    2
  149    2
  Name: Species, Length: 150, dtype: int64
```

In the last, we introduce new column “Species_id” and store our encoded data into it

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Species_id
0	1	5.1	3.5	1.4	0.2	Iris-setosa	0
1	2	4.9	3.0	1.4	0.2	Iris-setosa	0
2	3	4.7	3.2	1.3	0.2	Iris-setosa	0
3	4	4.6	3.1	1.5	0.2	Iris-setosa	0
4	5	5.0	3.6	1.4	0.2	Iris-setosa	0
...	...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica	2
146	147	6.3	2.5	5.0	1.9	Iris-virginica	2
147	148	6.5	3.0	5.2	2.0	Iris-virginica	2
148	149	6.2	3.4	5.4	2.3	Iris-virginica	2
149	150	5.9	3.0	5.1	1.8	Iris-virginica	2

150 rows × 7 columns

Name: Rahima Siddiqui(2303.KHI.DEG.030)
Peer Name: Osama Abdul Razzak(2303.KHI.DEG.029)

we also did some variation like over write encoded data to ‘Species’ and drop the previous “Species_id” column

```
data['Species'] = encode_label
```

```
data = data.drop(['Species_id'],axis =1)
data
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	0
1	2	4.9	3.0	1.4	0.2	0
2	3	4.7	3.2	1.3	0.2	0
3	4	4.6	3.1	1.5	0.2	0
4	5	5.0	3.6	1.4	0.2	0
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	2
146	147	6.3	2.5	5.0	1.9	2
147	148	6.5	3.0	5.2	2.0	2
148	149	6.2	3.4	5.4	2.3	2
149	150	5.9	3.0	5.1	1.8	2

150 rows × 6 columns