Name: - Muhammad Faizan Rafique (2303.005.KHI.DEG)
Pair Partner 1: - Muhammad Huzaifa Waseem (2303-KHI-DEG-021)
Pair Partner 2: - Syed Muhammad Hammad Irshad (2303.KHI.DEG.032)
Note: Second pair partner was permitted by Trainer Yamna Tahir.

UNIT 3.1:

Assignment

Method 1: -

Before Label Encode

```
df = pd.read_csv("Iris.csv")
df
```

	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

After Label Encode

Explanation: In this code we first import pandas then create a function having two parameters first is taking dataset and second is taking column of that dataset. After that we selected data type includes object type. After that we converted that column type into category and assigned that column into main data frame.

```
import pandas as pd
def label_encoder(df, Column_name):
    obj_df = df.select_dtypes(include=['object']).copy()
    obj_df[Column_name] = obj_df[Column_name].astype('category')
    obj_df[Column_name] = obj_df[Column_name].cat.codes
    df[Column_name] = obj_df[Column_name]
    return df

df = pd.read_csv("Iris.csv")
label_encoded_df = label_encoder(df, "Species")
label_encoded_df
```

	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

Method 2: -

Before Label Encode

```
df = pd.read_csv("test.csv")
df
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...
413	1305	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN	S
414	1306	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105	C
415	1307	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN	S
416	1308	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN	S
417	1309	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN	C

418 rows × 11 columns

After Label Encode

In this code we first import pandas then create a function having one parameter taking dataset. After that we selected data type includes object type. We iterate that object data so we can get all object data. After that we converted all object column types into categories and assigned that column into the main data frame.

```
import pandas as pd
def label_encoder(df):
    obj_df = df.select_dtypes(include=['object']).copy()
    for data in obj_df :
        obj_df[data] = obj_df[data].astype('category')
        obj_df[data] = obj_df[data].cat.codes
        df[data] = obj_df[data]
    return df


df = pd.read_csv("test.csv")
label_encoded_df = label_encoder(df)
label_encoded_df
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	206	1	34.5	0	0	152	7.8292	-1	1
1	893	3	403	0	47.0	1	0	221	7.0000	-1	2
2	894	2	269	1	62.0	0	0	73	9.6875	-1	1
3	895	3	408	1	27.0	0	0	147	8.6625	-1	2
4	896	3	178	0	22.0	1	1	138	12.2875	-1	2
...	...	...	...	...	...	...	...	...	...	...	...
413	1305	3	353	1	NaN	0	0	267	8.0500	-1	2
414	1306	1	283	0	39.0	0	0	324	108.9000	22	0
415	1307	3	332	1	38.5	0	0	346	7.2500	-1	2
416	1308	3	384	1	NaN	0	0	220	8.0500	-1	2
417	1309	3	302	1	NaN	1	1	105	22.3583	-1	0