


Feature Scaling ~ Is a practice where you preprocess the dataset in order to work with the ML model.
Categorical Variable ~ These are type of values, that are non-numerical.

Label Encoding ~ It is a procedure to handle the categorical variable with respect to Ordinal Mechanism.
Ordinal Mechanism ~ It is a ranking system.

```
import pandas as pd

df = pd.read_csv('/content/titanic.csv')

df.head()
```



| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked |
|---|-------------|----------|--------|--|--------|------|-------|-------|---------------------|---------|-------|----------|
| 0 | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN | S |
| 1 | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| 2 | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S |
| 3 | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily Mav Peel) | female | 35.0 | 1 | 0 | 113803 | 53.1000 | C123 | S |

```
from sklearn.preprocessing import LabelEncoder

label_encoder = LabelEncoder()

label_encoding_for_sex = label_encoder.fit_transform(df.Sex)

# label_encoding_for_sex

label_encoder.classes_

array(['female', 'male'], dtype=object)

df['Sex_Encoded'] = label_encoding_for_sex

df.head()
```

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked | Sex_Encoded |
|---|-------------|----------|--------|---|--------|------|-------|-------|---------------------|---------|-------|----------|-------------|
| 0 | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN | S | 1 |
| 1 | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C | 0 |
| 2 | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S | 0 |

```
df.Embarked.isna().sum()

2

df.Embarked.unique()
```

```
array(['S', 'C', 'Q', nan], dtype=object)

df.dropna(subset = ['Embarked'], inplace=True)

df.Embarked.unique()

array(['S', 'C', 'Q'], dtype=object)

embarked_encoded = label_encoder.fit_transform(df.Embarked)

# embarked_encoded

label_encoder.classes_

array(['C', 'Q', 'S'], dtype=object)

df['Embarked_encoded'] = embarked_encoded

df.head()
```

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked | Sex_Encoded | Embarked_encoded |
|---|-------------|----------|--------|---|--------|------|-------|-------|-----------|---------|-------|----------|-------------|------------------|
| 0 | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN | S | 1 | |
| 1 | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C | 0 | |
| | | | | Heikkinen | | | | | | | | | | |