

AIM:

To load the Iris dataset and convert it into a DataFrame and express it.

PROGRAM CODE:

importing libraries.

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
from sklearn.preprocessing import StandardScaler,  
LabelEncoder.
```

Loading datasets.

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

```
Print(df.head(1))
```

cleaning the dataset.

Missing Values.

```
Print("Missing Values:\n", df.isnull().sum())
```

Removing duplicates.

~~```
df.drop_duplicates(inplace=True)
```~~~~```
Print(f"Number of duplicates: {df.duplicated().sum()}")
```~~

Data Transformation.

⇒ Encoding categorical variables.

```
label_encoder = LabelEncoder()
```

```
df['Species'] = label_encoder.fit_transform(df['Species'])
```

```
Print("Encoded Species:\n", df['Species'].head())
```

Do
⇒ Sc
Sc
df

OUTPUT:

| ID | SepalLengthcm | SepalWidthcm | PetalLengthcm | PetalWidthcm | Species |
|----|---------------|--------------|---------------|--------------|-------------|
| 0 | 1 | 5.1 | 3.5 | 1.4 | Iris-setosa |
| 1 | 2 | 4.9 | 3.0 | 1.4 | Iris-setosa |
| 2 | 3 | 4.7 | 3.2 | 1.4 | Iris-setosa |
| 3 | 4 | 4.6 | 3.1 | 1.5 | Iris-setosa |
| 4 | 5 | 5.0 | 3.6 | 1.4 | Iris-setosa |

Missing values:

ID 0 Encoded species:
SepalLengthcm 0 0 0
SepalWidthcm 0 1 0
PetalLengthcm 0 2 0
PetalWidthcm 0 3 0
Species 0 Name: Species, dtype: int64
dtype: int64

Number of duplicates: 0

| ID | SepalLengthcm | SepalWidthcm | PetalLengthcm | PetalWidthcm | Species |
|----|---------------|--------------|---------------|--------------|---------|
| 1 | -0.900681 | 1.032057 | -1.341272 | -1.312977 | 0 |
| 2 | -1.143017 | -0.124958 | -1.341272 | -1.312977 | 0 |
| 3 | -1.385353 | 0.337848 | -1.398138 | -1.312977 | 0 |
| 4 | -1.506521 | 0.106445 | -1.284407 | -1.312977 | 0 |
| 5 | -1.021849 | 1.263460 | -1.841272 | -1.312977 | 0 |

CORRELATION HEATMAP

| ID | SepalLengthcm | SepalWidthcm | PetalLengthcm | PetalWidthcm | Species |
|---------------|---------------|--------------|---------------|--------------|---------|
| SepalLengthcm | 1.00 | 0.72 | -0.40 | 0.88 | 0.90 |
| SepalWidthcm | 0.72 | 1.00 | -0.11 | 0.87 | 0.82 |
| PetalLengthcm | -0.40 | -0.11 | 1.00 | -0.42 | -0.36 |
| PetalWidthcm | 0.88 | 0.87 | -0.42 | 1.00 | 0.96 |
| Species | 0.90 | 0.82 | -0.36 | 0.96 | 1.00 |

RES

Dh

Data Transformation # Data Normalization
⇒ Scaling Numerical Variables.

Scaler = StandardScaler()

df[['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm',
'PetalWidthCm']] = scaler.fit_transform(df
[['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm',
'PetalWidthCm']])

Print(df.head())

Visualization. (Pairplot (visualizing relationships
sns.pairplot(df, hue = 'Species') between features))

plt.subtitle("Pairplot of Iris Dataset", size = 16)

plt.show()

Visualization (Correlation Heatmap (visualizing
correlations between numerical features))

plt.figure(figsize = (8, 6))

sns.heatmap(df.corr(), annot = True, cmap = 'coolwarm',
fmt = '.2f', linewidths = 0.5)

plt.title("Correlation Heatmap", size = 16)

plt.show()

RESULT:

Thus the data Preprocessing and cleaning
has been implemented successfully.