

Experiment–3

Aim:

Understanding data preprocessing using Pandas.

Code:

```
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4
5 df = pd.read_csv("./iris.csv")
6 print("Shape (rows, columns) :", df.shape)
7 print("Total samples:", df.shape[0])
8 print("Total features:", df.shape[1])
9 print("\nColumns:")
10 print(df.columns.tolist())
11
12 print("\nFirst 5 rows:")
13 print(df.head())
14
15 print("\nData types & missing values:")
16 print(df.info())
17
18 print("\nMissing values per column:")
19 print(df.isnull().sum())
20
21 target_col = df.columns[-1] # assumes last column is target
22 print(f"\nTarget column: '{target_col}'")
23 print("Number of classes :", df[target_col].nunique())
24 print("\nClass distribution:")
25 print(df[target_col].value_counts())
26
27 # 1. Countplot
28 plt.figure(figsize=(10,5))
29 sns.countplot(data=df, x=target_col, palette="viridis")
30 plt.title(f"Countplot of {target_col}")
31 plt.xticks(rotation=45)
32 plt.show()
33
34 # 2. Pairplot (only if dataset is small)
35 if df.shape[1] <= 12:
36     sns.pairplot(df, hue=target_col, diag_kind='kde', corner=True)
37     plt.show()
38
39 # 3. Correlation Heatmap (numeric columns only)
40 numeric_df = df.select_dtypes(include=['float64', 'int64'])
41 plt.figure(figsize=(10,8))
```