-: LAB: 01 :-

Date 03/3/25
Page 3

Data Pre-Processing Techniques :-

(i) import pandas as pd.
df = pd.read_csv("housing.csv")
print("Data loaded into DataFrame")

(ii) print("\n Information of all columns:")
print(df.info())

(iii) print("\n Stastifical information:")
print(df.describe())

(iv) print("\n Count of unique labels for ocean proximity column:")
print(df['Ocean Proximity'].value_counts())

(v) print("\n Columns with missing values:")
missing_values = df.isnull().sum()
m = missing_values[missing_values > 0]
print(m)

1] Diabetes Dataset: Columns like Glucose, Blood Pressure and BMI had missue value Handled by,

imp string mean as media

Adult income: Columns like occupation and nature country had missue values handled by mode or dropna()

Date __/__/__
Page _____

2) Diabetes Dataset: The outcome column is categorical, encoded using label encoding.

Adult Income Dataset: Columns like work class, education were categorical encoded using one-Hot encoding.

3) Min-Max Scaling: Scales the data to a fixed range (0 to 1) and is used when data is bounded the model is sensstive to the scale.

Standardization: Scales the data to have and mean of 0 and a standard deviation of 1 and is used when the data is normally distributed or when the model assumes a normal distribution.

10/3/25