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 Statistical 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()
cm = missing_values[missing_values>0]
print(cm)
```

(1) Diabetes Dataset: Columns like Glucose, Blood Pressure and BMI had missing values handled by imputing mean or median

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

PAGE NO :
DATE :

(2) Diabetes Dataset: The Outcome column is categorical, encoded using Label Encoding.

Adult Income Dataset: Columns like workclass, education were categorical encoded using one-hot Encoding.

(3) Min-Max Scaling scales the data to a fixed range (0 to 1) and it used when data is bounded the model is sensitive to the scale.

Standardization scales the data to have a 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.

02.03.25