

- (i) import pandas as pd.
df = pd.read_csv("housing.csv")
print("Data loaded into Data Frame")
- (ii) print("\n Statistical information :")
print(df.describe())
- (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()
em = missing-value[missing-values > 0]
print(em)

1. ~~Diabetes Dataset~~ : Columns like Glucose Blood pressure and BMI had null handled by imputing mean and median

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

- (d) 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 scale the data to a fixed range (0 to 1) and is 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 assumed a normal distribution.

~~Signature~~
10/3/25