Lab -1

python

On.1 Write a program code , filename as "housing.csv"
i) To load .csv file into dataframe
ii) To display information of all columns
iii) To display statistical information of all numerical
iv) To display the count of unique labels for "ocean proximity" column.
v) To display which attributes in dataset have missing values count greater than zero.

```
→ import pandas as pd
file_name = "housing.csv"
df = pd.read_csv (file_name)
print (df. info())
print (df. describe())
print (df ['Ocean proximity'].count())
missing_values = df. isnull. sum()
columns_with_missing = missing_values [missing_values >0]
print (columns_with_missing)
```

On.2 Write python code implement following data processing techniques for diabetes dataset.

1) Data Cleaning: Handling missing values, Handling categorical data, Handling outliers.

2) Data transformations:- Min-Max Scaler / Normatized, standard scaler.

```python
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler,
StandardScaler, LabelEncoder.
diabetes = pd.read_csv ("diabetes.csv")
diabetes.drop(columns = ['ID', 'No.Pation'], inplace=True)
diabetes.dropna (inplace =true)
diabetes [['Gender', 'CLASS']] = diabetes [['Gender', 'CLASS']].
apply (LabelEncoder().fit_transform)


diabetes = diabetes [(np.abs ((diabetes.select_dtypes (
include = [np.number]) -diabetes.mean())/diabetes.
std())< 3). all (axis =1)]


scaler = MinMaxScaler()
diabetes_scaled = pd.DataFrame (scaler.fit_transform
(diabetes.drop(columns = ['class'])), columns =
diabetes.columns[:-1]))


diabetes_scaled ['CLASS'] = diabetes ['CLASS'].values
print (diabetes_scaled.head())
```

① ✓ 6/3/25