

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import OrdinalEncoder, OneHotEncoder, LabelEncoder
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from scipy import stats

dfd = pd.read_csv("/content/Dataset of Diabetes .csv")
dfa = pd.read_csv("/content/adult.csv", delimiter=";", on_bad_lines='skip', engine='python') # Modified Line

def add_missing_values(df, missing_fraction=0.05):
    np.random.seed(42)
    num_missing = int(missing_fraction * df.size)

    missing_rows = np.random.randint(0, df.shape[0], num_missing)
    missing_cols = np.random.randint(0, df.shape[1], num_missing)

    for row, col in zip(missing_rows, missing_cols):
        df.iat[row, col] = np.nan

    return df

dfd = add_missing_values(dfd)
dfa = add_missing_values(dfa)

print("Missing values in Diabetes dataset before removal:\n", dfd.isnull().sum())
print("\nMissing values in Adult dataset before removal:\n", dfa.isnull().sum())

dfd_cleaned = dfd.dropna()
dfa_cleaned = dfa.dropna()

print("\nMissing values in Diabetes dataset after removal:\n", dfd_cleaned.isnull().sum())
print("\nMissing values in Adult dataset after removal:\n", dfa_cleaned.isnull().sum())

➡ Missing values in Diabetes dataset before removal:
  ID      60
No_Pation 44
Gender    56
AGE       39
Urea      58
Cr        48
HbA1c     42
Chol      47
TG        49
HDL       54
LDL       40
VLDL      46
BMI       51
CLASS     48
dtype: int64

Missing values in Adult dataset before removal:
  age      2801
workclass 2813
fnlwgt    2827
education 2902
educational-num 2912
marital-status 2982
occupation 3000

```

```

relationship    2905
race            2918
gender          2964
capital-gain    2866
capital-loss    2866
hours-per-week  2892
native-country  2879
income          2823
dtype: int64

```

Missing values in Diabetes dataset after removal:

```

ID            0
No_Pation    0
Gender        0
AGE           0
Urea          0
Cr            0
HbA1c         0
Chol          0
TG            0
HDL           0
LDL           0
VLDL          0
BMI           0
CLASS         0
dtype: int64

```

Missing values in Adult dataset after removal:

```

age           0
workclass     0
fnlwgt        0
education     0
educational num 0

```

```
categorical_cols_dfd = dfd.select_dtypes(include=['object']).columns
```

```
categorical_cols_dfa = dfa.select_dtypes(include=['object']).columns
```

```
print("Categorical Columns in Diabetes Dataset:", categorical_cols_dfd)
```

```
print("Categorical Columns in Adult Dataset:", categorical_cols_dfa)
```

```
label_encoder = LabelEncoder()
```

```
for col in categorical_cols_dfd:
```

```
    dfd[col] = label_encoder.fit_transform(dfd[col].astype(str))
```

```
for col in categorical_cols_dfa:
```

```
    dfa[col] = label_encoder.fit_transform(dfa[col].astype(str))
```

```
print("\nEncoded Diabetes Dataset:\n", dfd.head())
```

```
print("\nEncoded Adult Dataset:\n", dfa.head())
```

```

→ Categorical Columns in Diabetes Dataset: Index(['Gender', 'CLASS'], dtype='object')
   Categorical Columns in Adult Dataset: Index(['workclass', 'education', 'marital-status', 'occupation',
        'relationship', 'race', 'gender', 'native-country', 'income'],
        dtype='object')

```

Encoded Diabetes Dataset:

	ID	No_Pation	Gender	AGE	Urea	Cr	HbA1c	Chol	TG	HDL	LDL	\
0	502.0	17975.0	0	50.0	4.7	46.0	4.9	4.2	0.9	2.4	1.4	
1	735.0	NaN	1	26.0	4.5	62.0	4.9	3.7	NaN	NaN	2.1	
2	420.0	47975.0	0	50.0	4.7	46.0	4.9	4.2	0.9	2.4	1.4	
3	680.0	87656.0	0	50.0	4.7	46.0	4.9	4.2	0.9	2.4	1.4	
4	504.0	34223.0	1	33.0	NaN	46.0	4.9	4.9	1.0	0.8	2.0	

	VLDL	BMI	CLASS
0	0.5	24.0	0
1	0.6	23.0	0
2	0.5	24.0	0
3	0.5	24.0	0
4	NaN	21.0	0

Encoded Adult Dataset:

	age	workclass	fnlwgt	education	educational-num	marital-status	\
0	25.0	4	226802.0	1	7.0	4	
1	38.0	4	89814.0	11	9.0	2	
2	28.0	2	336951.0	16	12.0	2	
3	44.0	4	160323.0	15	10.0	2	
4	18.0	0	103497.0	15	10.0	4	

	occupation	relationship	race	gender	capital-gain	capital-loss	\
0		7	3	2	1	0.0	0.0
1		5	0	4	1	0.0	0.0
2		11	6	4	1	0.0	0.0
3		7	0	2	1	7688.0	0.0
4		0	3	4	0	0.0	0.0

	hours-per-week	native-country	income
0	40.0	39	0
1	50.0	39	0
2	40.0	39	1
3	40.0	39	1
4	30.0	39	0

Start coding or [generate](#) with AI.