ASSIGNMENT 1

ALFIDO TECH

TASK 1====

DATA PREPROCESSING

CODE==

Import necessary libraries

```
import pandas as pd

from sklearn.model_selection

 import train_test_split from sklearn.pALreprocessing

 import LabelEncoder, StandardScaler
```

Step 1: Create sample data

```
data = { 'Name': ['Alice', 'Bob', 'Carol', 'David', 'Eva'],

'Age': [25, 30, None, 35, 28],

'Salary': [50000, None, 60000, 70000, 55000],

'Gender': ['Female', 'Male', 'Female', 'Male', 'Female'],

 'Purchased': ['Yes', 'No', 'Yes', 'No', 'Yes'] }

df = pd.DataFrame(data)

 print("Initial Data:") print(df)
```

=== Step 2: Handle missing values ===

df['Age'].fillna(df['Age'].mean(), inplace=True) df['Salary'].fillna(df['Salary'].mean(), inplace=True)

=== Step 3: Encode categorical columns ===

label_enc = LabelEncoder()

for col in ['Name', 'Gender', 'Purchased']:

df[col] = label_enc.fit_transform(df[col])

print("\nEncoded Data:")

print(df)

=== Step 4: Feature scaling ===

scaler = StandardScaler()

scaled_df = pd.DataFrame(scaler.fit_transform(df), columns=df.columns)

print("\nNormalized Data:")

print(scaled_df)

=== Step 5: Split into features (X) and target (y) ===

X = scaled_df.drop(columns=['Purchased'])

y = scaled_df['Purchased']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

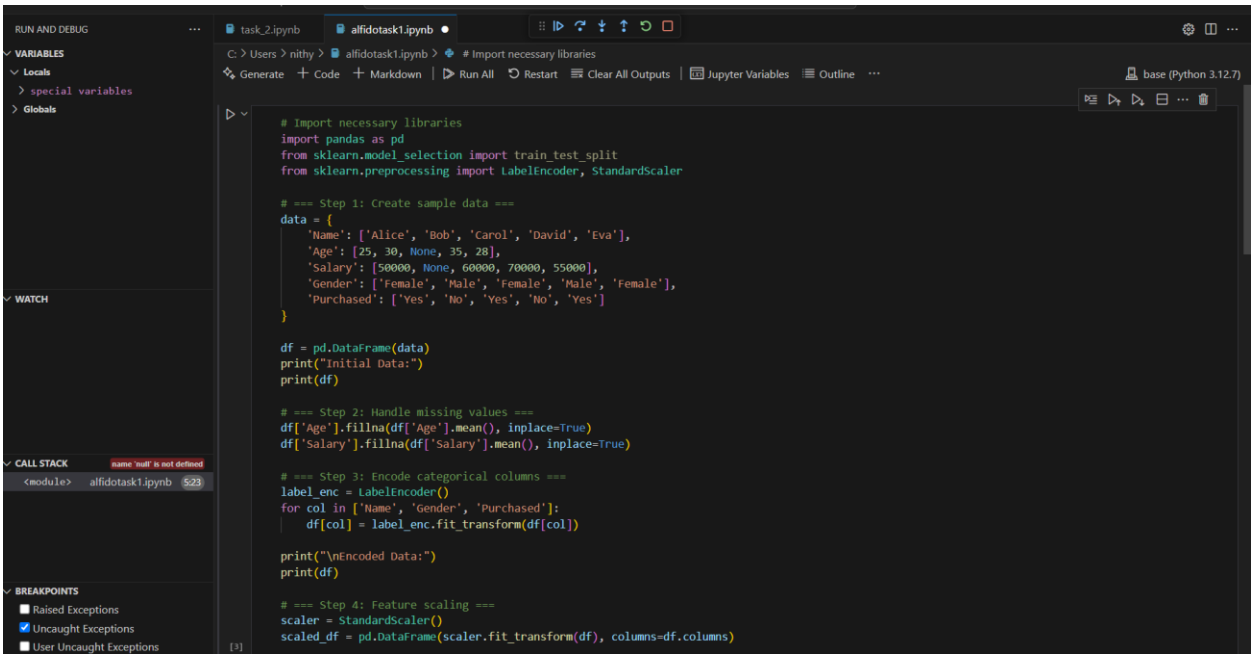
print("\nX_train:")

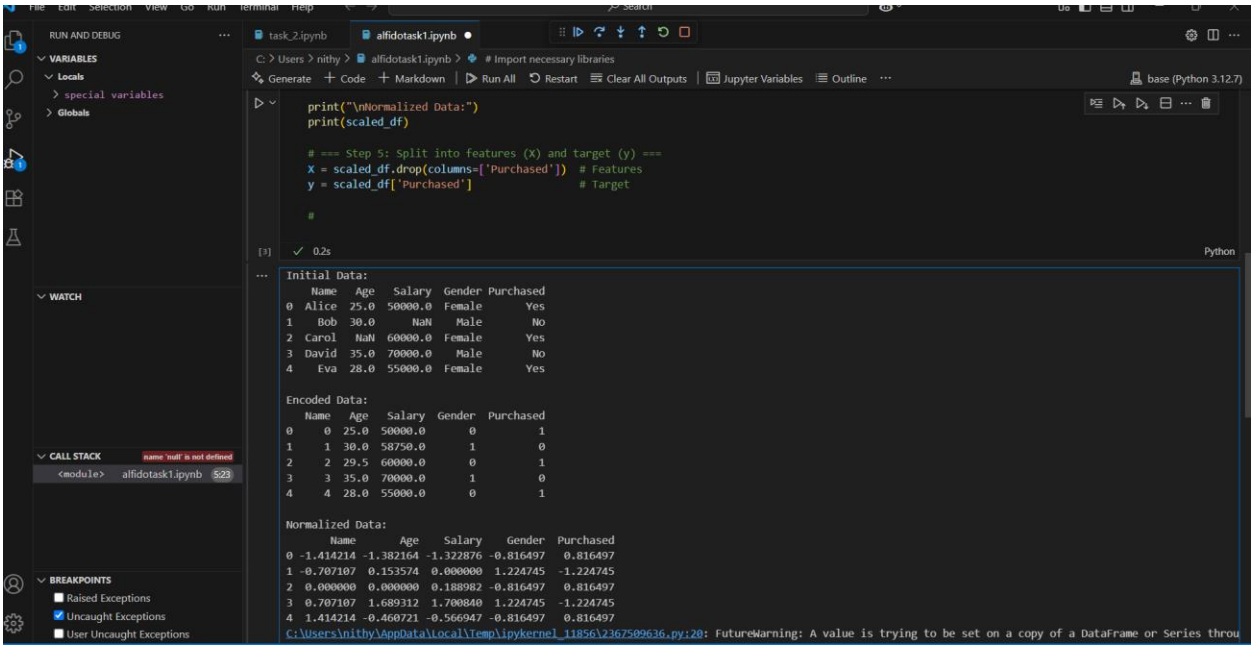
print(X_train)

print("\ny_train:")

print(y_train)

OUTPUT==




RUN AND DEBUG
VARIABLES
Locals
special variables
Globals
WATCH
CALL STACK
name 'null' is not defined
<module> alfidotask1.ipynb 523
BREAKPOINTS
Raised Exceptions
Uncaught Exceptions
User Uncaught Exceptions
task_2.ipynb
alfidotask1.ipynb
C: > Users > nithy > alfidotask1.ipynb > # Import necessary libraries
Generate + Code + Markdown | Run All Restart Clear All Outputs | Jupyter Variables Outline
base (Python 3.12.7)
print("\nNormalized Data:")
print(scaled_df)
# === Step 5: Split into features (X) and target (y) ===
X = scaled_df.drop(columns=['Purchased']) # Features
y = scaled_df['Purchased'] # Target
#
[3] ✓ 0.2s
Python
Initial Data:
Name Age Salary Gender Purchased
0 Alice 25.0 50000.0 Female Yes
1 Bob 30.0 NaN Male No
2 Carol NaN 60000.0 Female Yes
3 David 35.0 70000.0 Male No
4 Eva 28.0 55000.0 Female Yes
Encoded Data:
Name Age Salary Gender Purchased
0 0 25.0 50000.0 0 1
1 1 30.0 58750.0 1 0
2 2 29.5 60000.0 0 1
3 3 35.0 70000.0 1 0
4 4 28.0 55000.0 0 1
Normalized Data:
Name Age Salary Gender Purchased
0 -1.414214 -1.382164 -1.322876 -0.816497 0.816497
1 -0.707107 0.153574 0.000000 1.224745 -1.224745
2 0.000000 0.000000 0.188982 -0.816497 0.816497
3 0.707107 1.689312 1.700840 1.224745 -1.224745
4 1.414214 -0.460721 -0.566947 -0.816497 0.816497
C:\Users\nithy\AppData\Local\Temp\ipykernel_11856\2367509636.py:20: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series throu