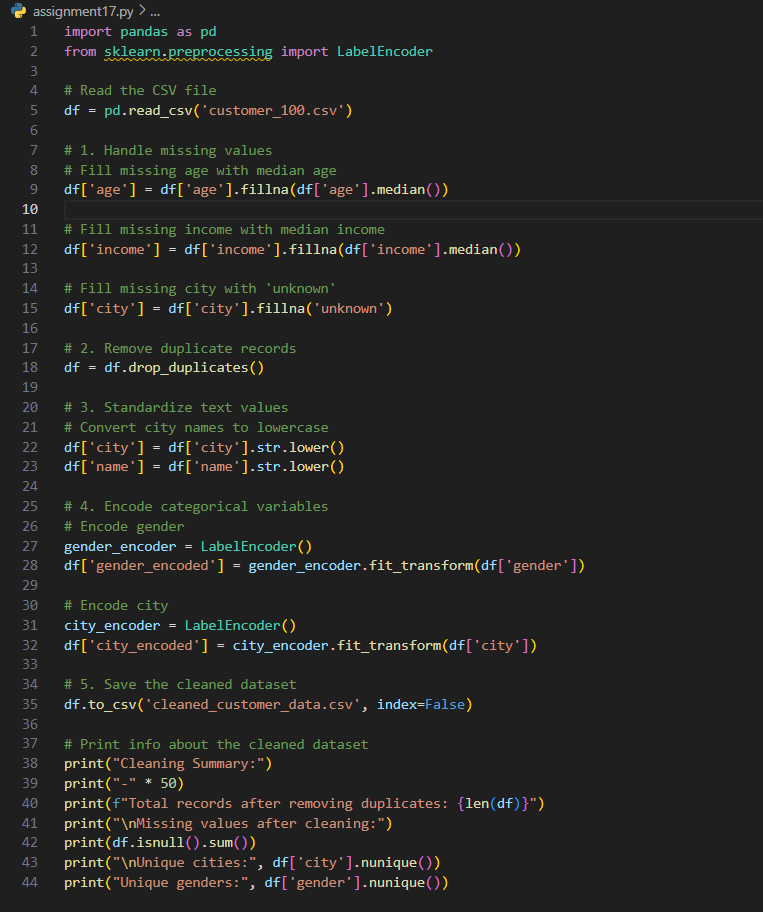
Name : Suraj Jadhav

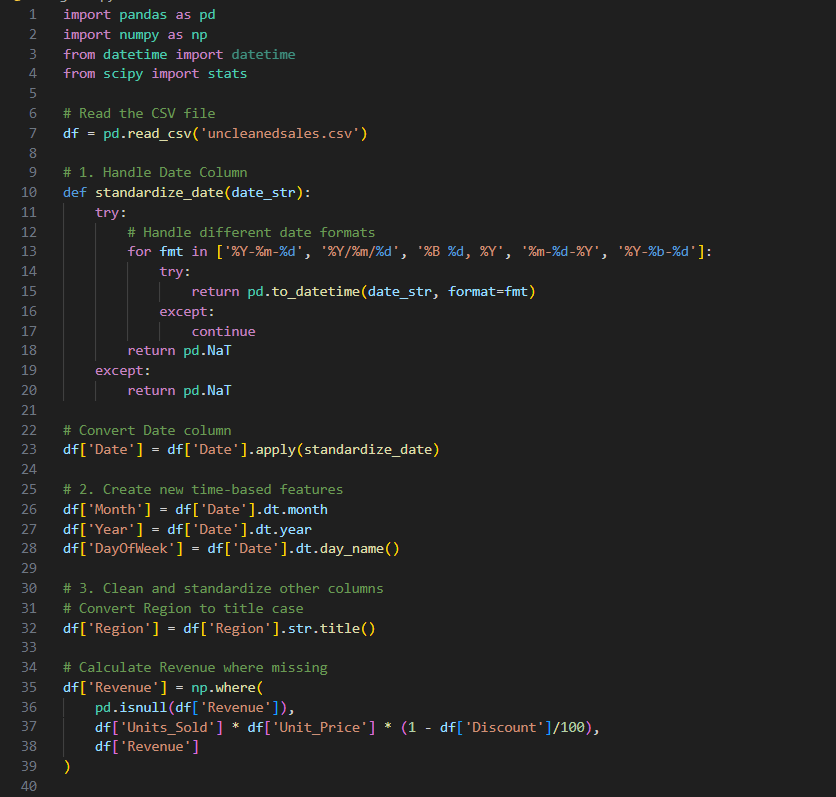
Roll no: 2403a51300

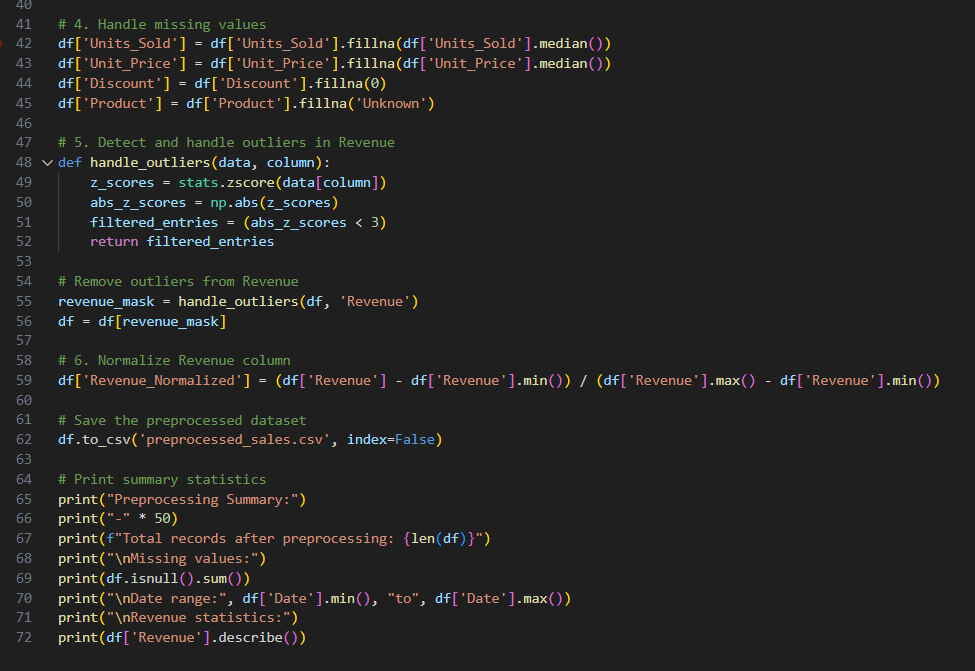
Batch: 13

Task 1 – Customer Data Cleaning  
Task:  
Use AI to generate a Python script for cleaning a customer dataset.  
Instructions:  
• Handle missing values in columns (age, income, city).  
• Remove duplicate records.  
• Standardize text values (e.g., city names in lowercase).  
• Encode categorical variables (gender, city).  
Expected Output:  
• A cleaned Pandas DataFrame ready for analysis

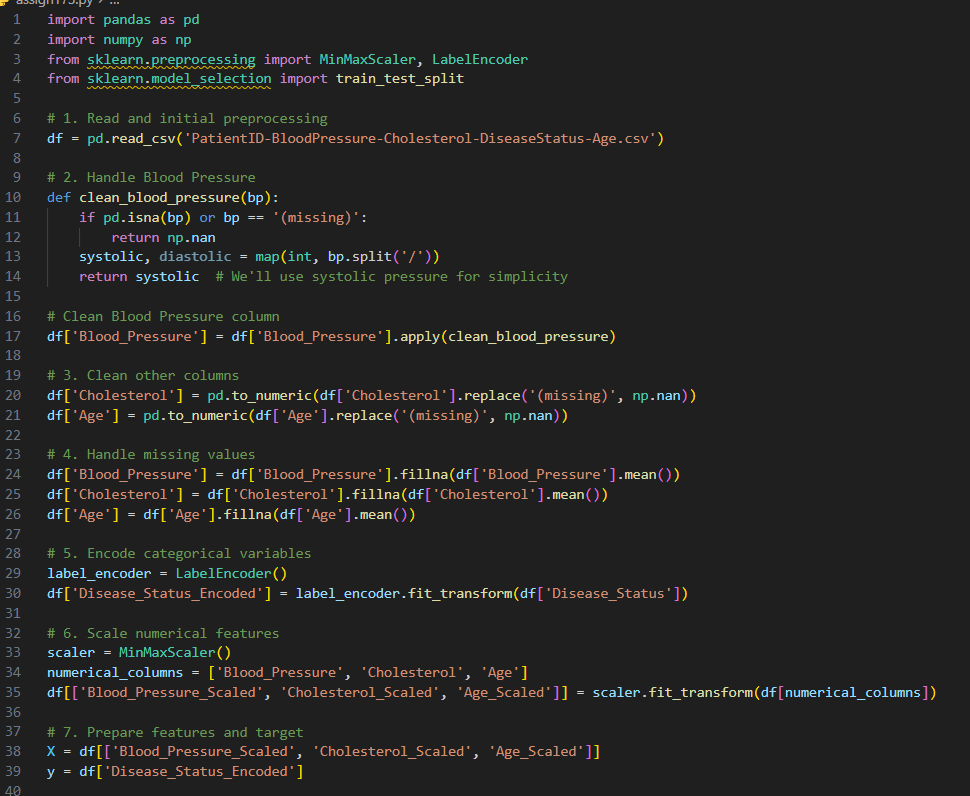


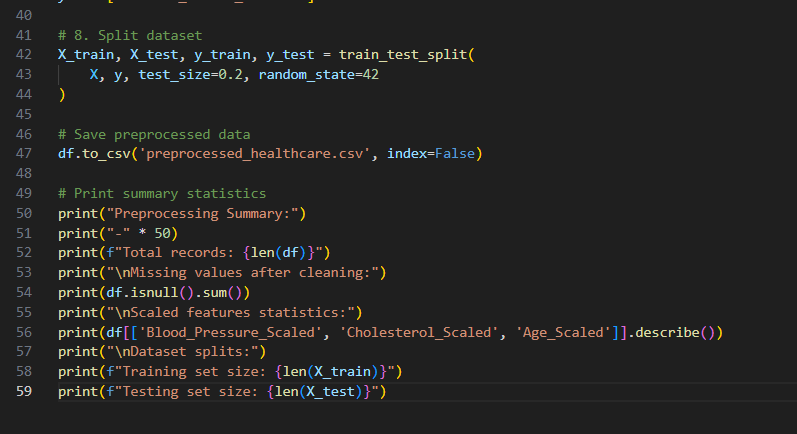
Task 2 – Sales Data Preprocessing  
Task:  
Preprocess sales transaction data using AI assistance.  
Instructions:  
• Convert date columns into datetime format.  
• Create new features (month, year, day of week).  
• Normalize sales amount column.  
• Detect and handle outliers in transaction amounts





Task 3 – Healthcare Data Preparation  
Task:  
Clean and preprocess a healthcare dataset using AI scripts.  
Instructions:  
• Handle missing values in blood\_pressure, cholesterol.  
• Encode categorical features like disease\_status (Yes/No).  
• Scale numerical features (min-max or standard scaling).  
• Split dataset into training and testing sets





Task 4 – Real-Time Application: E-commerce Product Data

Cleaning  
Scenario:  
An e-commerce company has product catalog data with  
inconsistencies.  
Requirements:  
• Standardize product names and categories.  
• Fill missing prices with median values.  
• Remove duplicate entries.  
• Normalize product ratings (0–5 scale).  
• Generate a summary report of data quality improvements (before  
vs after cleaning).

