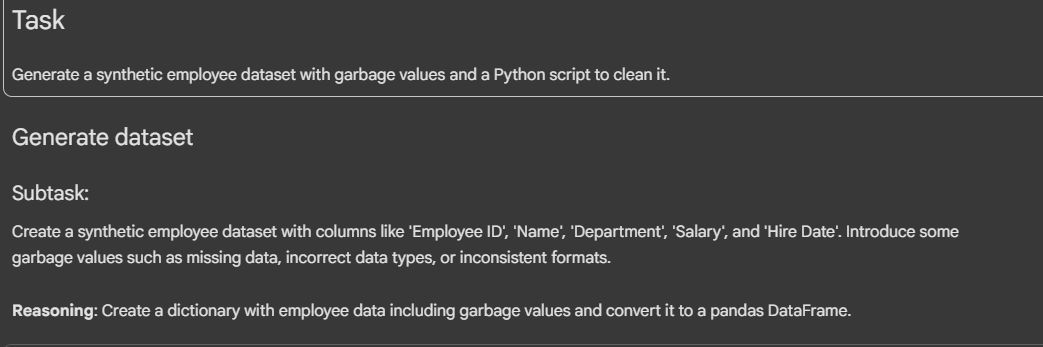
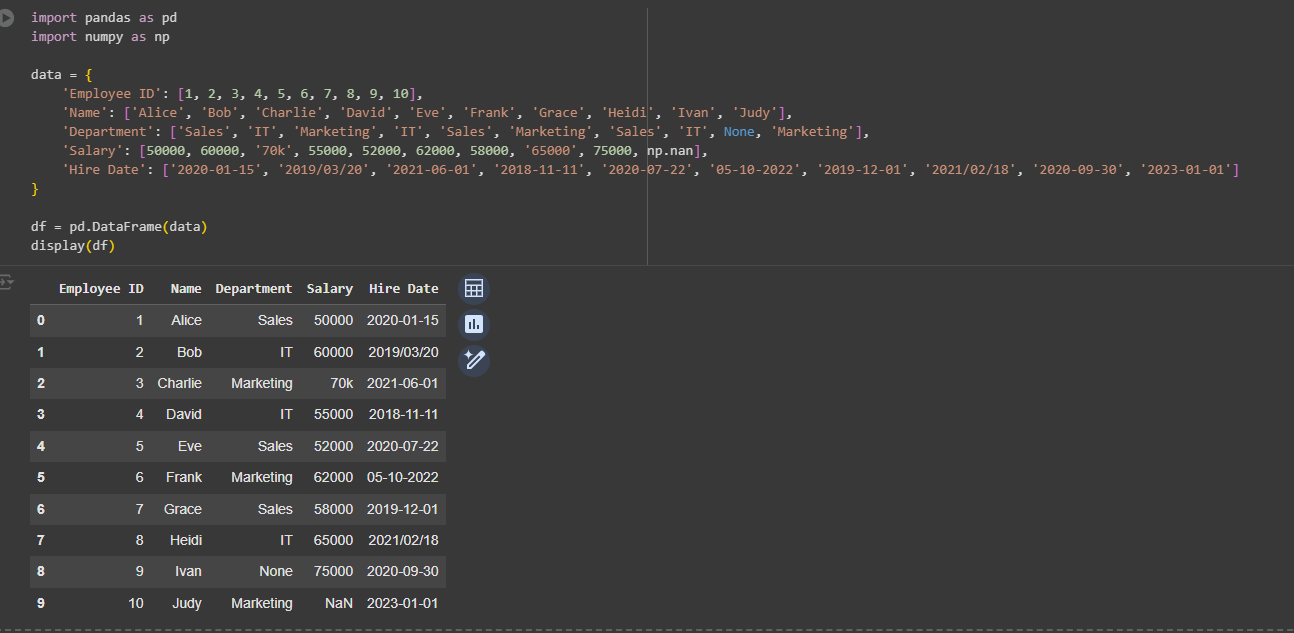
AI ASSIGNMENT-17.4

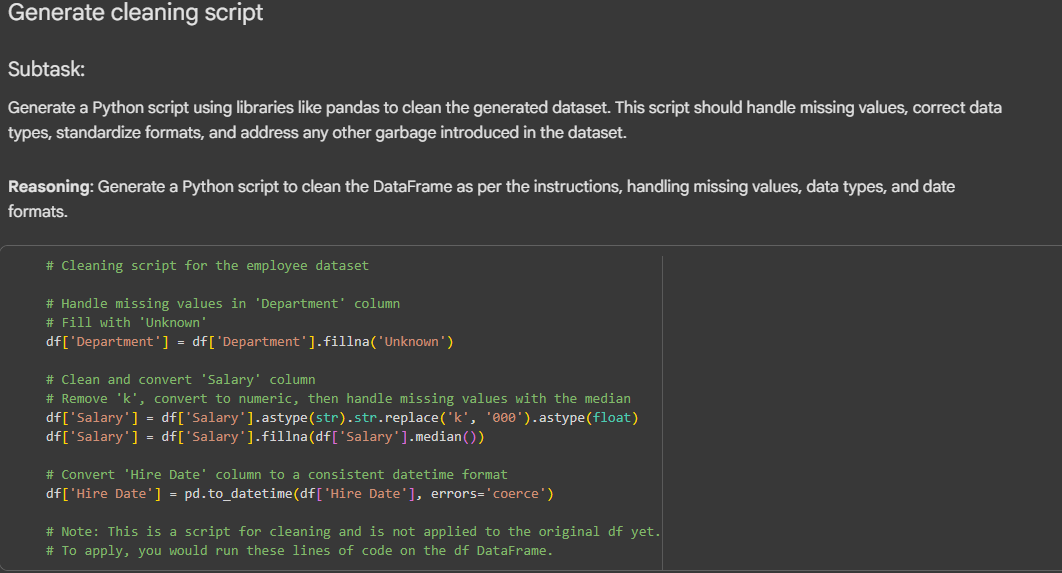
Task 1 – Employee Data Preprocessing  
Task:  
Use AI to generate a Python script for cleaning an employee dataset.  
Instructions:  
• Handle missing values in columns (salary, department,  
joining\_date).  
• Convert the "joining\_date" column into proper datetime format.  
• Standardize department names (e.g., "HR", "hr", "Human  
Resources" → "HR").  
• Encode categorical variables (department, job\_role).  
Expected Output:  
• A cleaned Pandas DataFrame with consistent departments, proper  
dates, and encoded features.

PROMPT: 

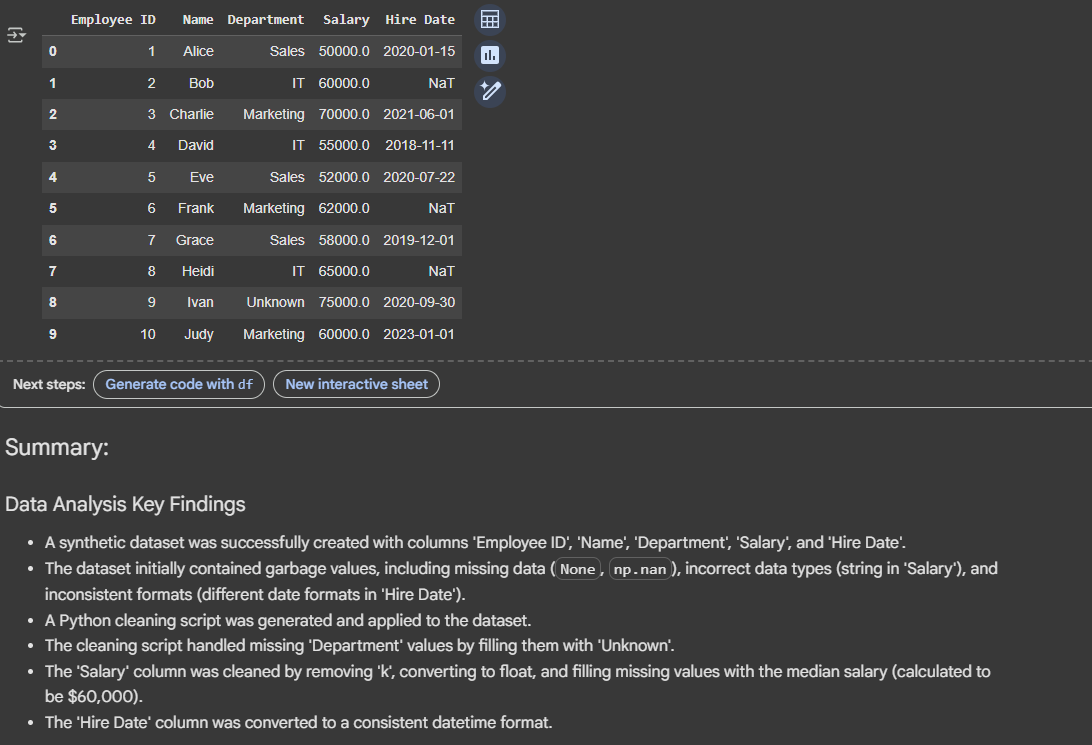
CODE:

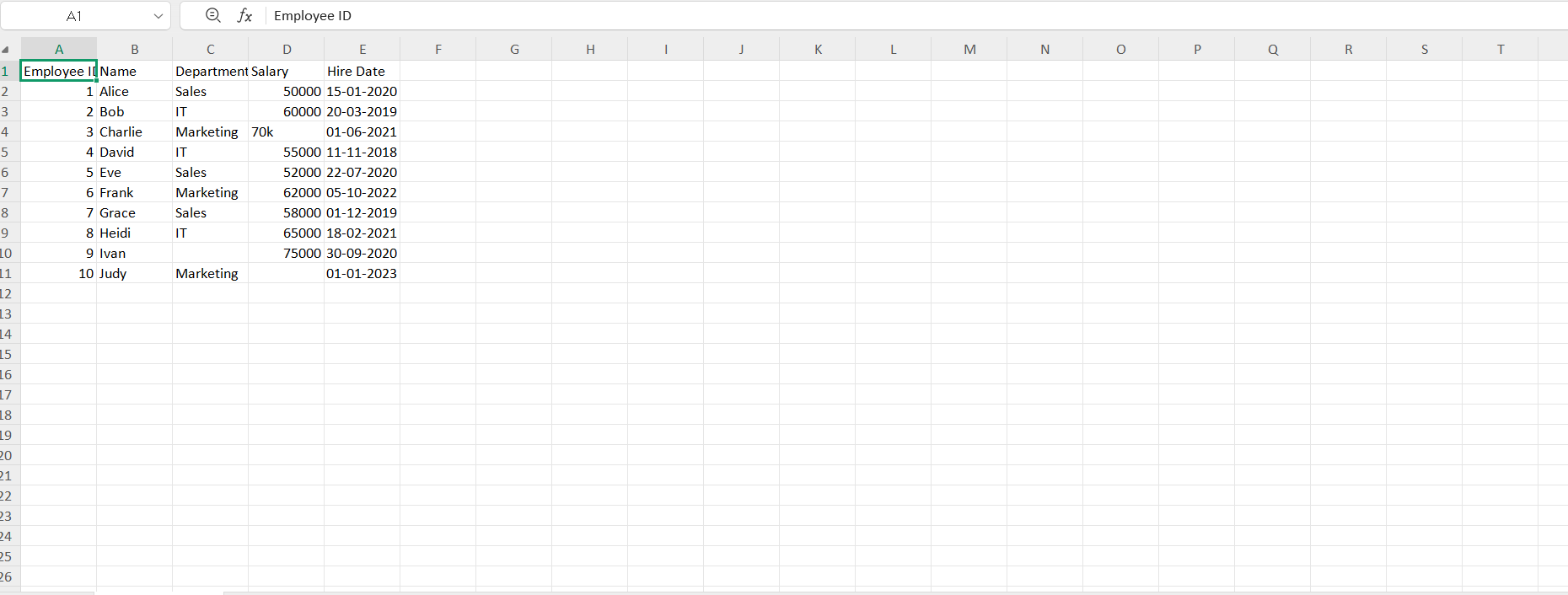


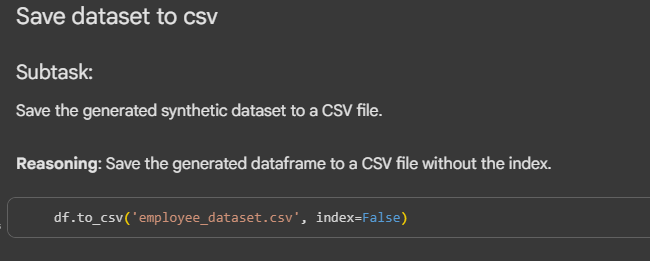


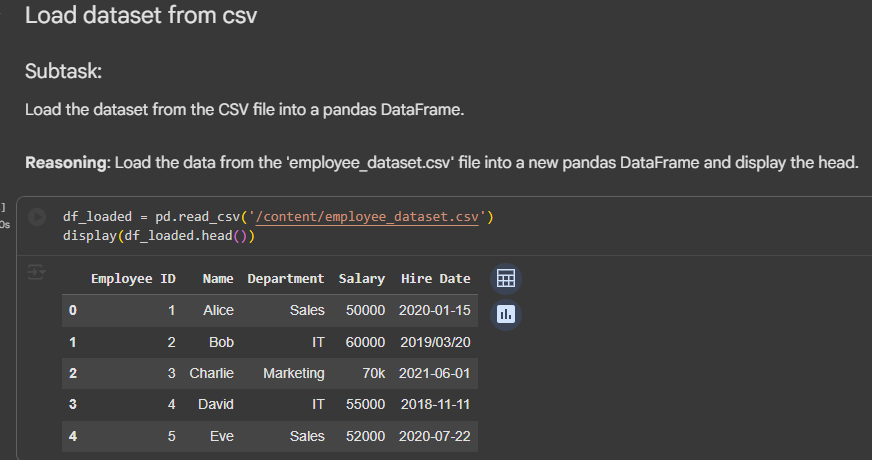


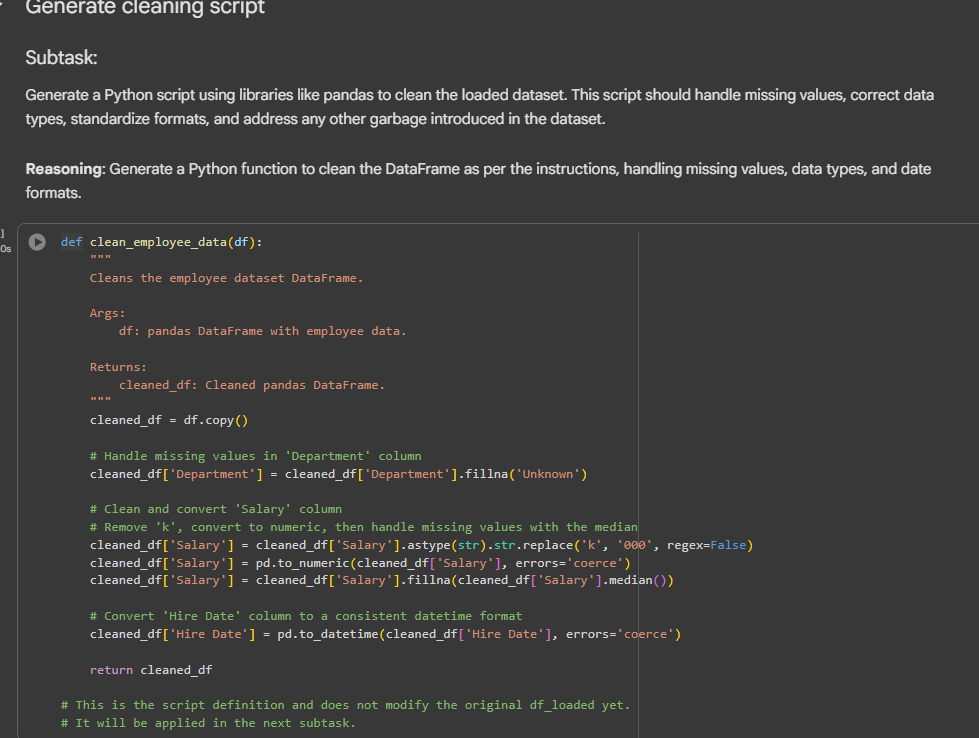


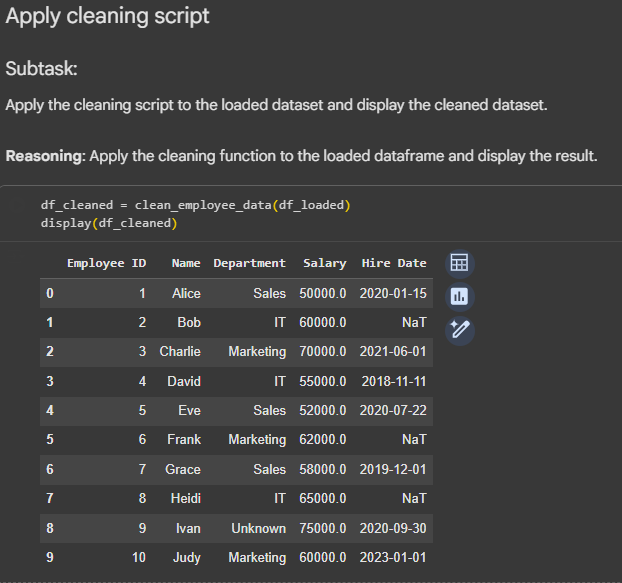




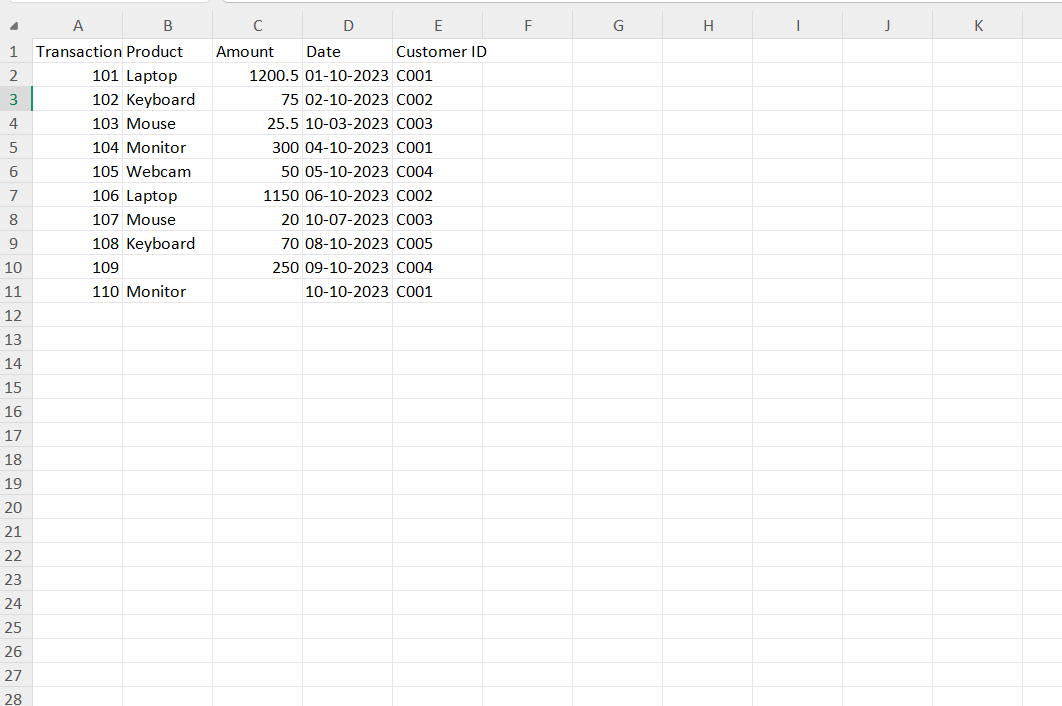


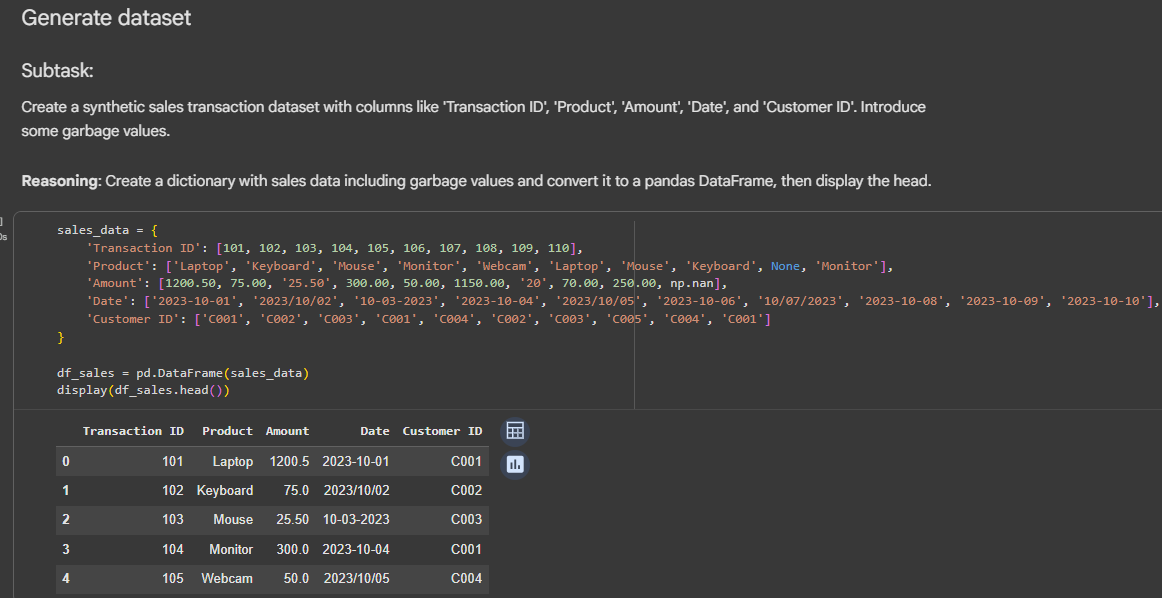


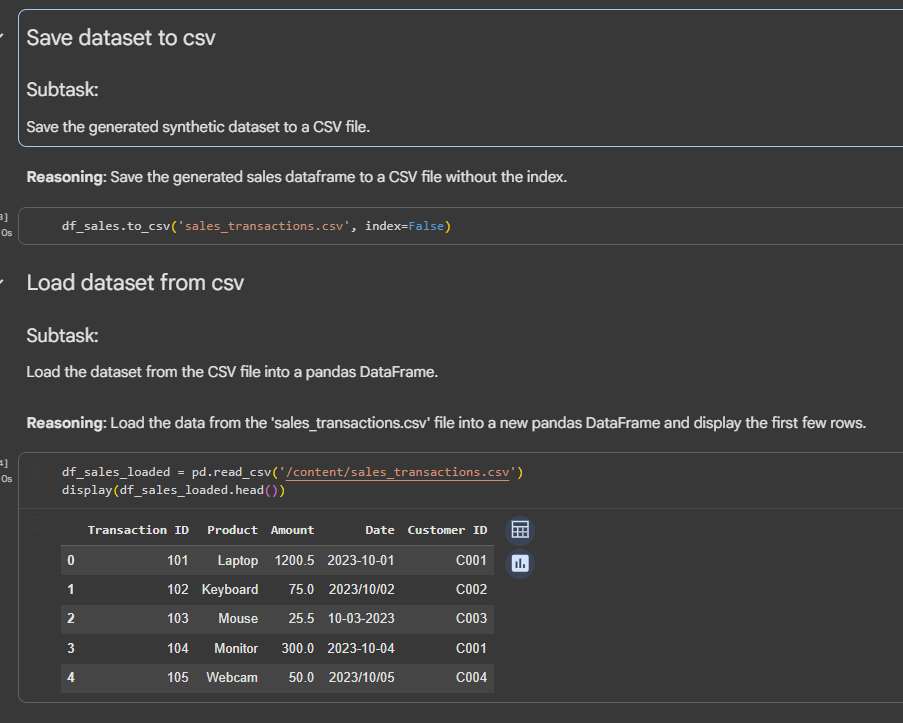


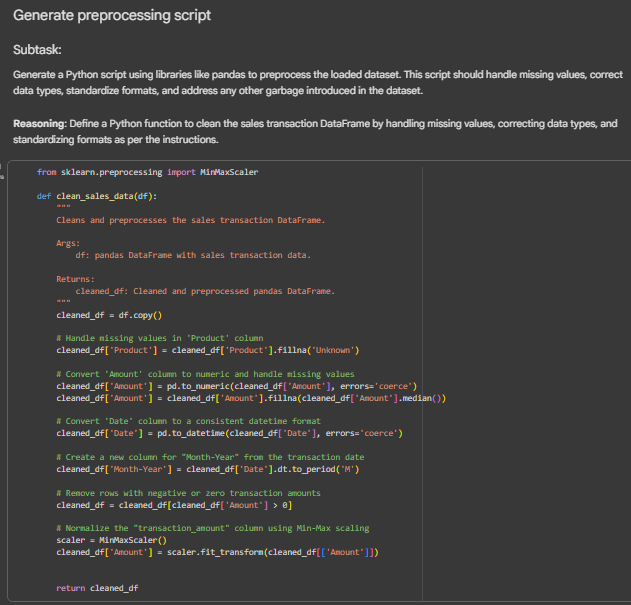


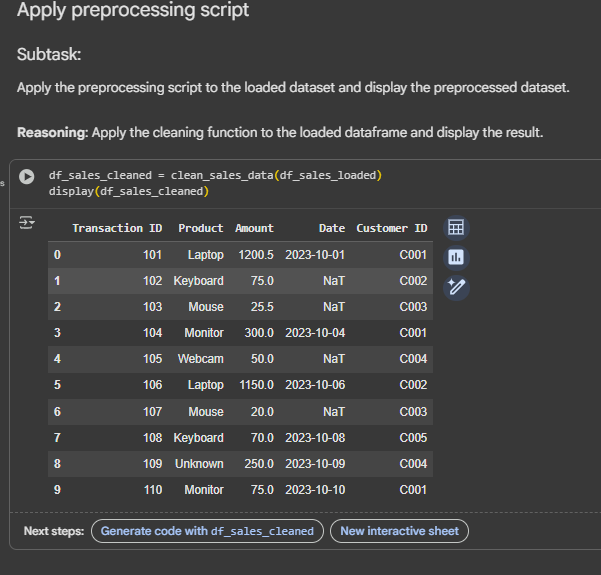
Task 2 – Sales Transaction Data Preprocessing  
Task:  
Use AI to generate a script for preprocessing a sales transaction dataset.  
Instructions:  
• Convert transaction dates to proper datetime format.  
• Create a new column for “Month-Year” from the transaction date.  
• Remove rows with negative or zero transaction amounts.  
• Normalize the "transaction\_amount" column using Min-Max  
scaling.  
Expected Output:  
• A preprocessed DataFrame with valid dates, normalized amounts,  
and no invalid records.





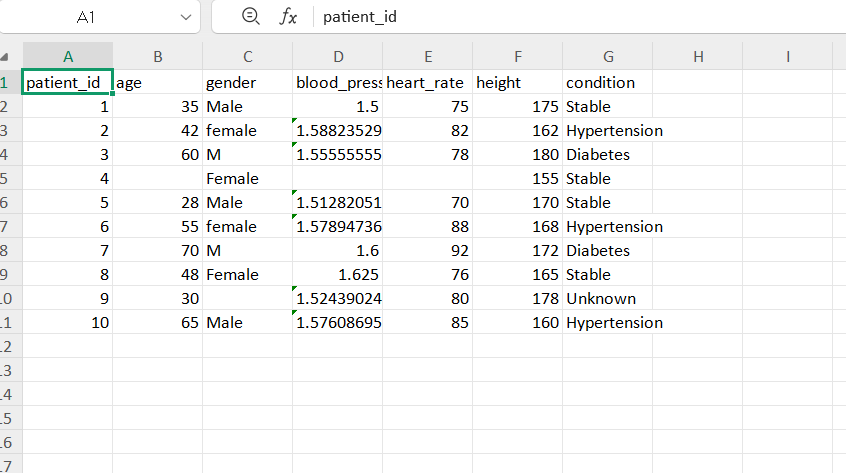


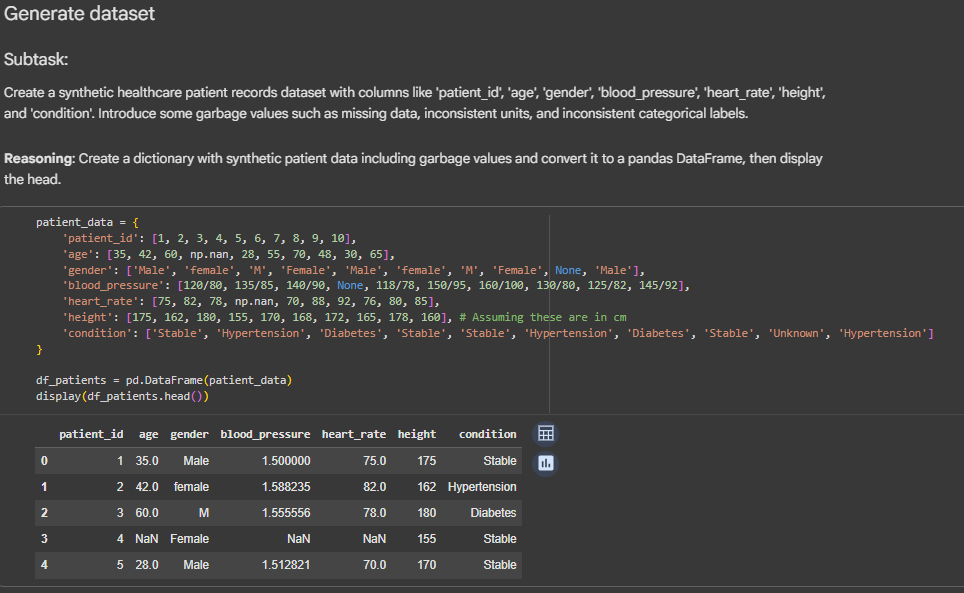


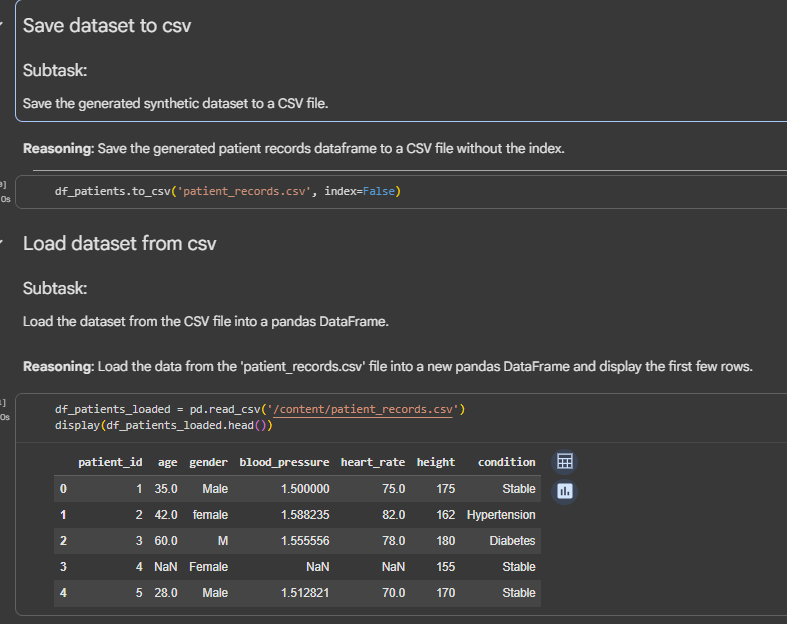


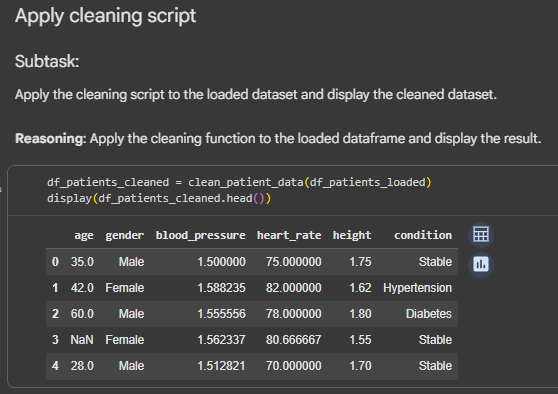
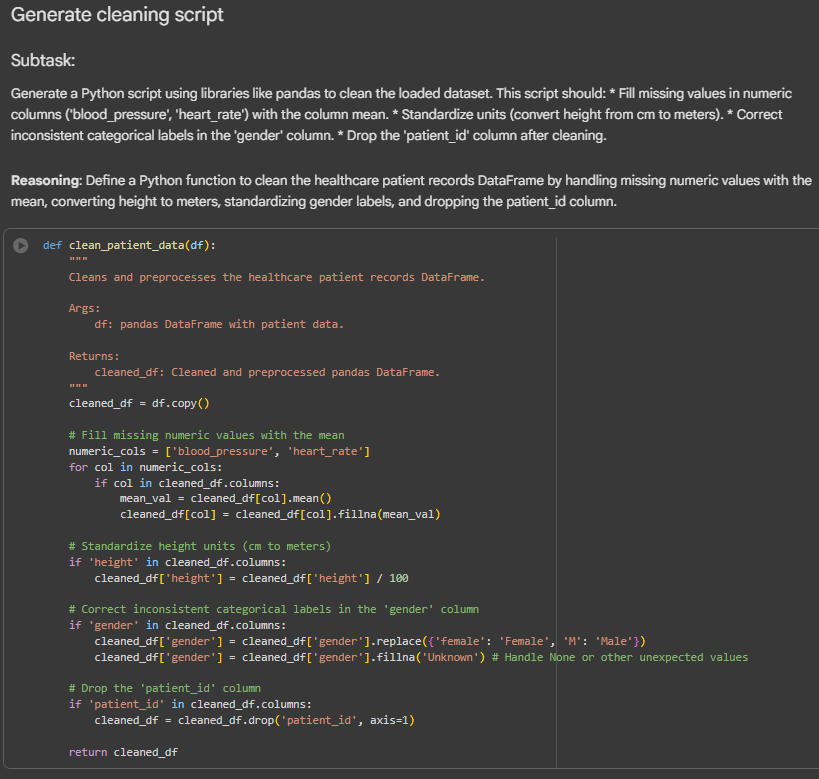
Task 3 – Healthcare Patient Records Cleaning  
Task:  
Use AI to generate a script for cleaning healthcare patient records.  
Instructions:  
• Fill missing values in numeric columns (e.g., blood\_pressure,  
heart\_rate) with column mean.  
• Standardize units (convert height from cm to meters).  
• Correct inconsistent categorical labels (e.g., "M", "Male", "male"

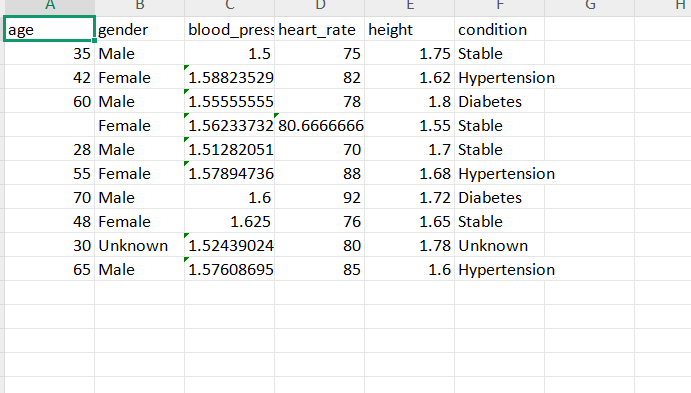
→ "Male").  
• Drop irrelevant columns such as patient\_id after cleaning.  
Expected Output:  
• A cleaned healthcare dataset suitable for ML model training



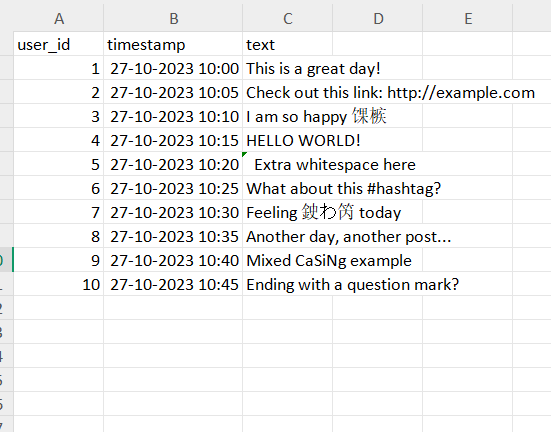


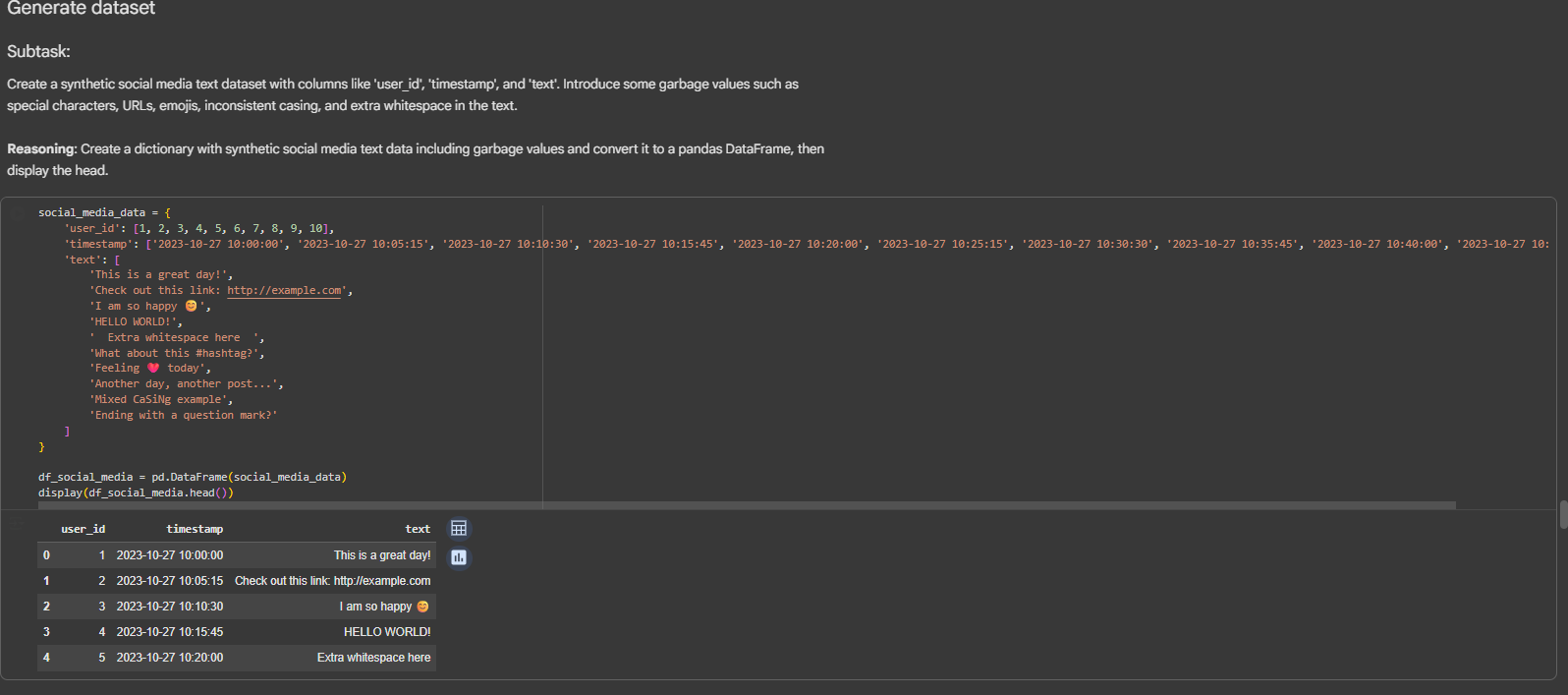


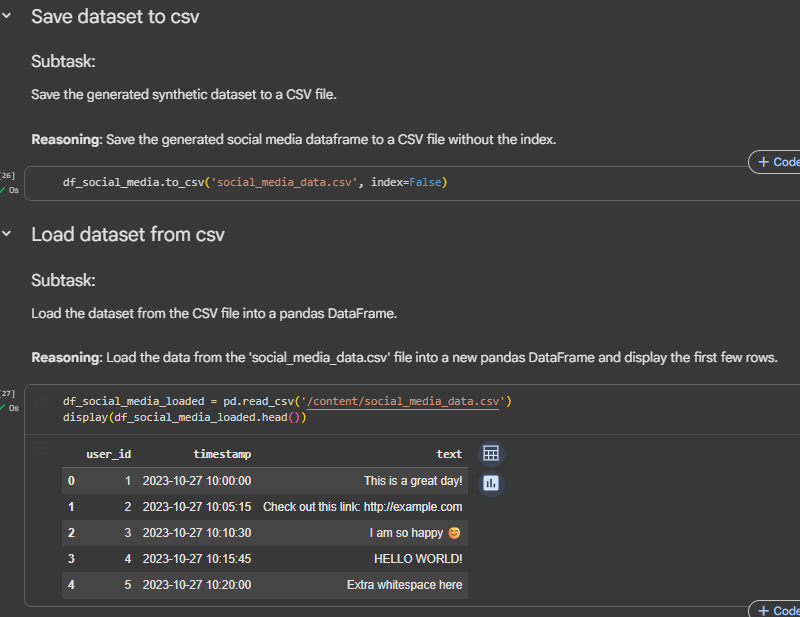




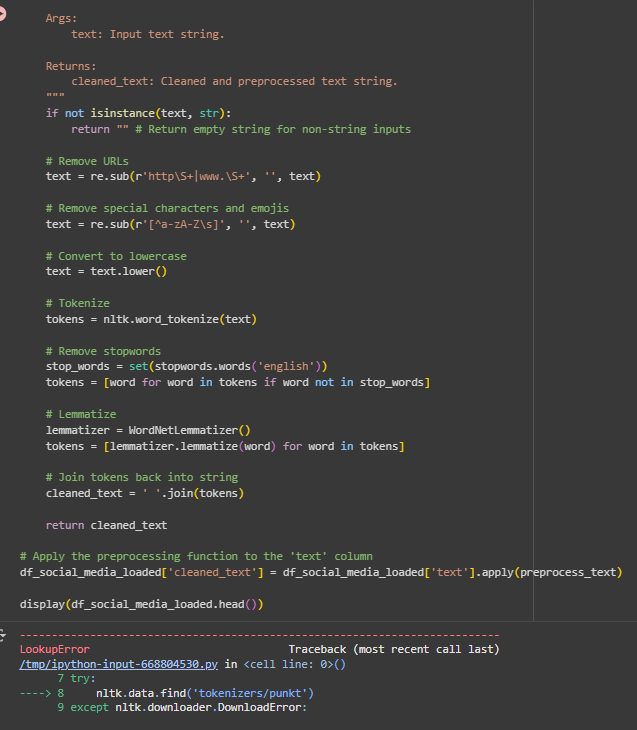
Task 4 – Social Media Sentiment Dataset Preparation  
Task:  
Use AI to write a script to preprocess a social media text dataset.  
Instructions:  
• Remove special characters, URLs, and emojis from text.  
• Convert all text to lowercase.  
• Tokenize and remove stopwords.  
• Apply lemmatization for standardizing words.  
Expected Output:  
• A processed dataset with clean text, ready for NLP sentiment  
analysis



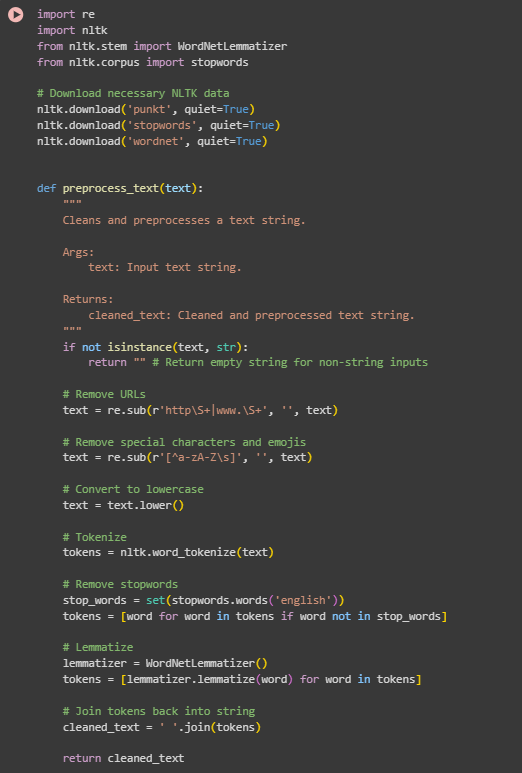


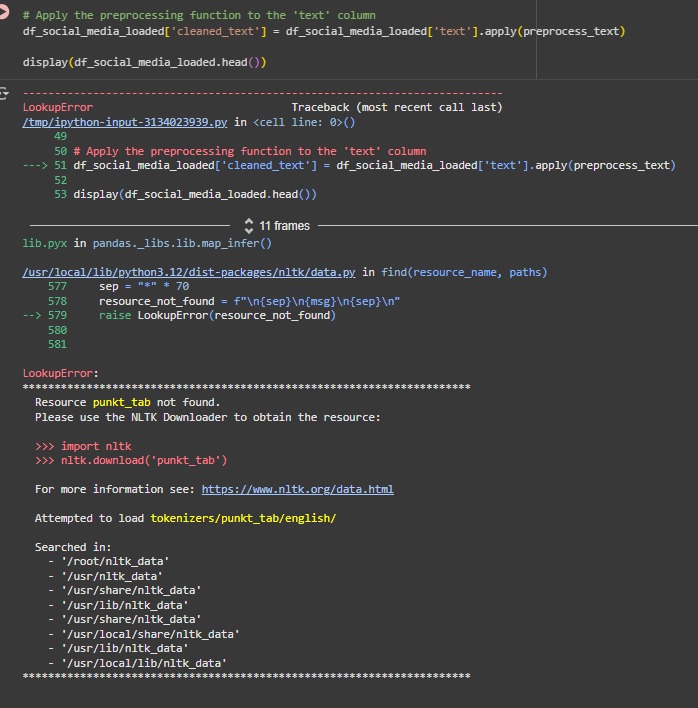


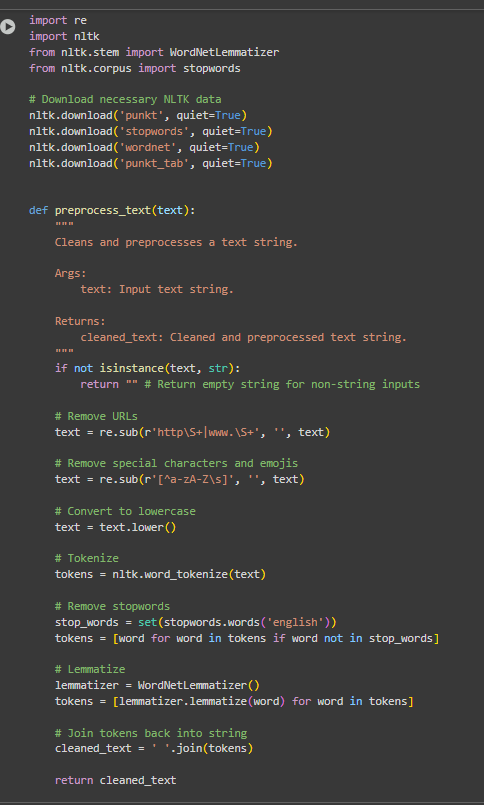


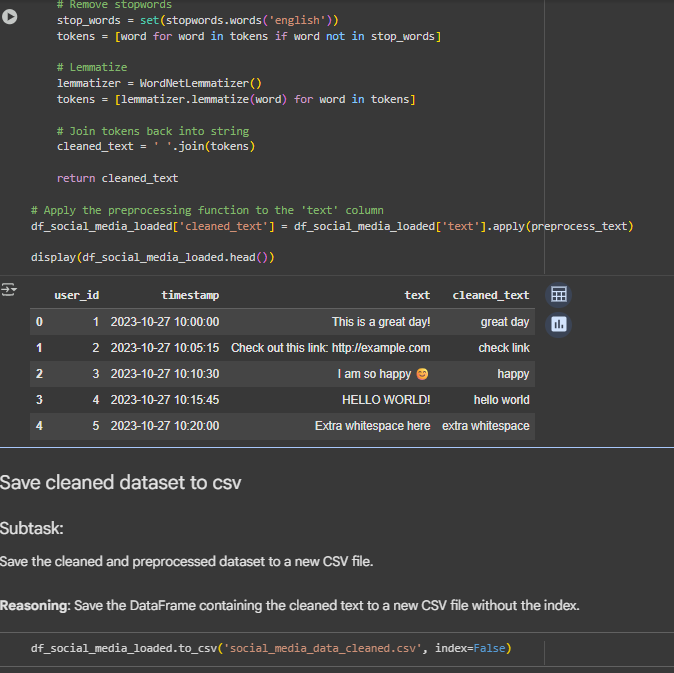


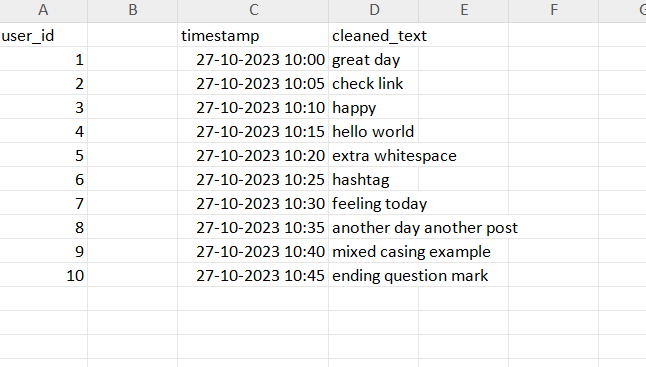












Task 5 – Financial Dataset Feature Engineering  
Task:  
Use AI to create a preprocessing script for a financial dataset.  
Instructions:  
• Handle missing values in stock price and volume.  
• Create new features such as moving average (7-day, 30-day).  
• Normalize continuous variables using StandardScaler.  
• Encode categorical columns (sector, company\_name).  
Expected Output:  
• A feature-engineered DataFrame with new indicators and  
normalized values for ML tasks.

