Lab Assignment-17.1

**NAME: P. Umesh Reddy ROLLNO: 2403A510F9**

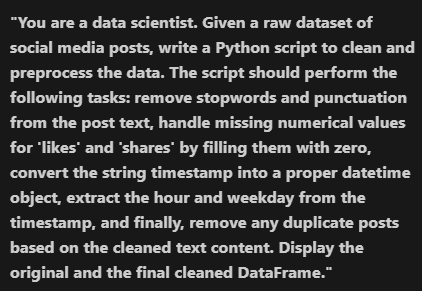
**BATCHNO: 06**

**TASK1:**

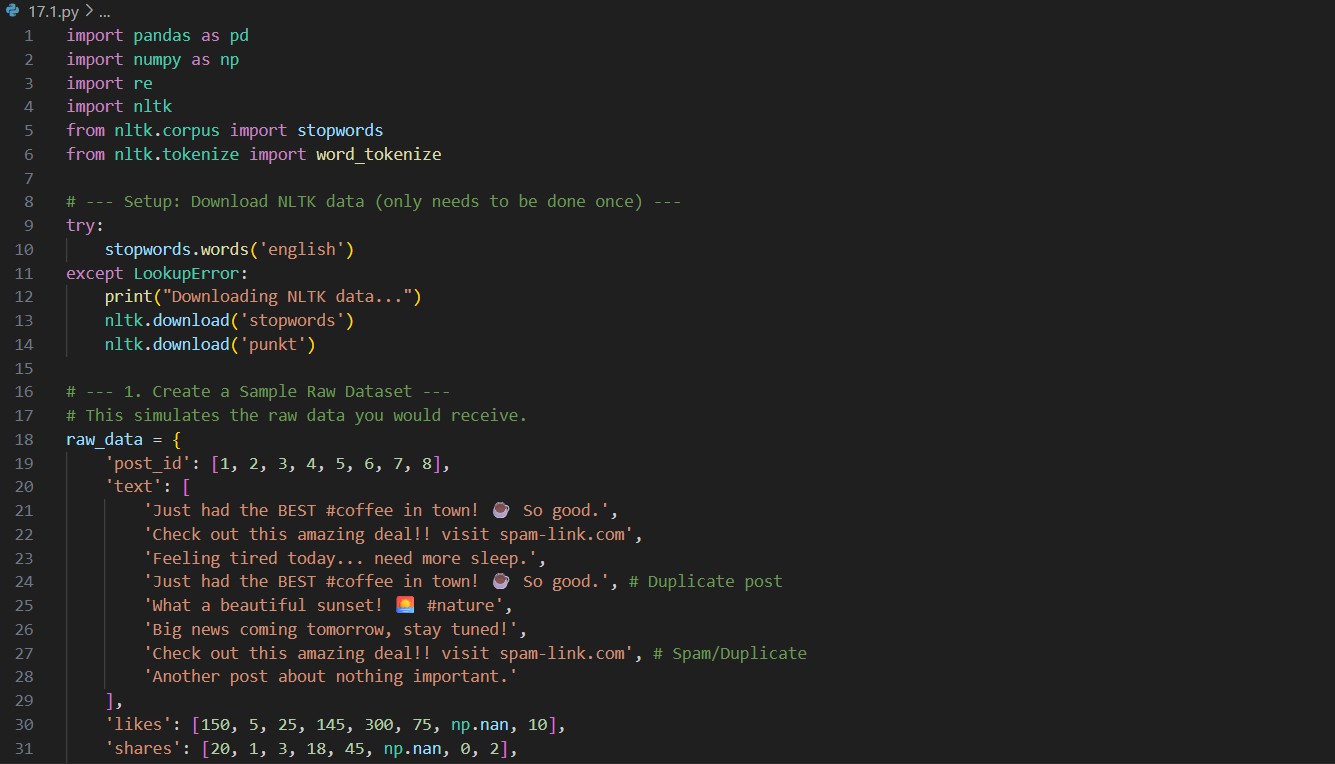
Task: Clean raw social media posts dataset. Instructions:

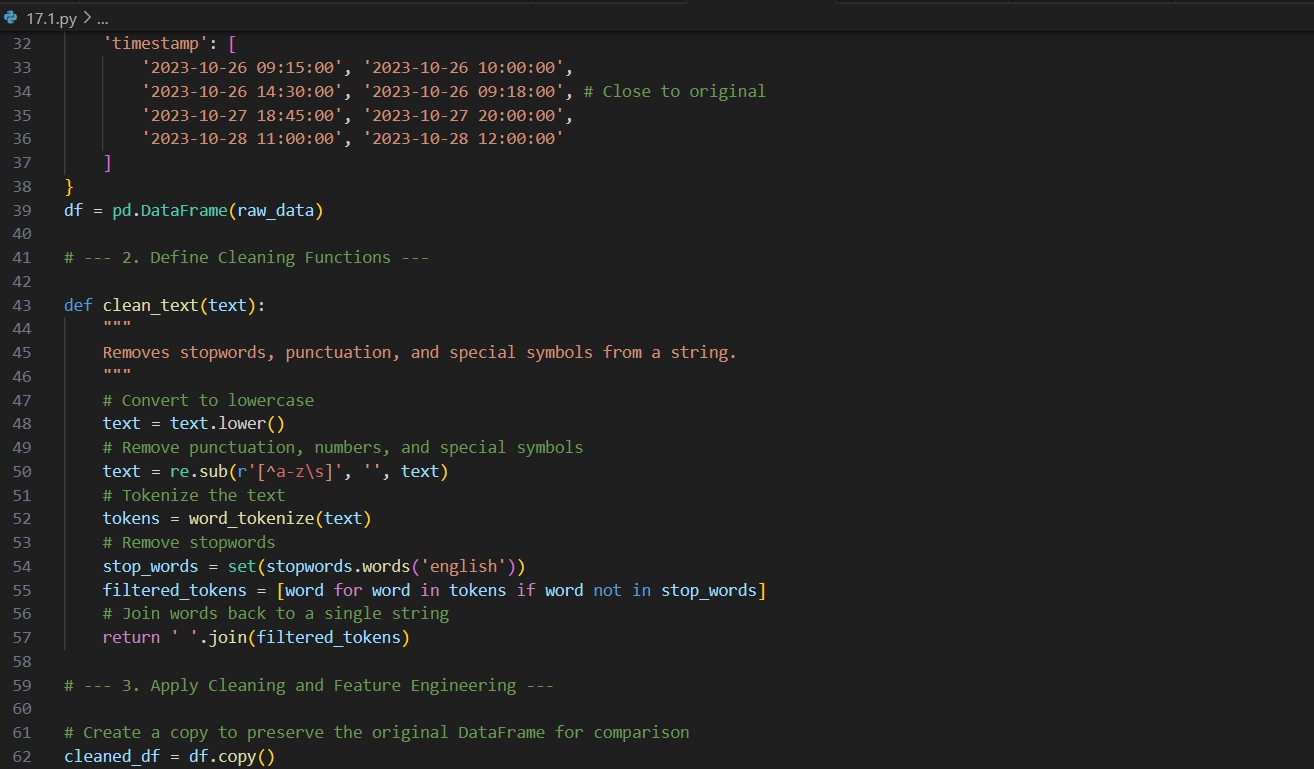
* Remove stopwords, punctuation, and special symbols from post text.
* Handle missing values in likes and shares columns.
* Convert timestamp to datetime and extract features (hour, weekday).
* Detect and remove spam/duplicate posts.

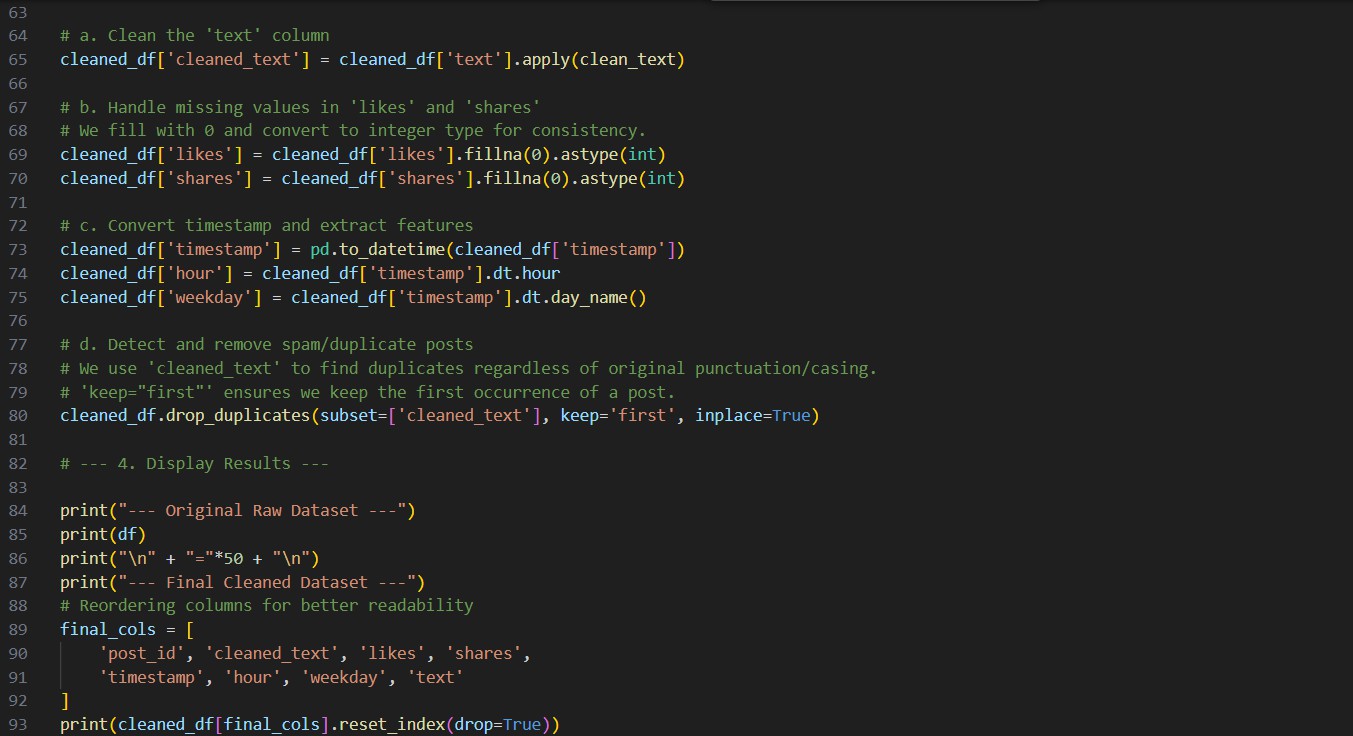
**PROMPT:**



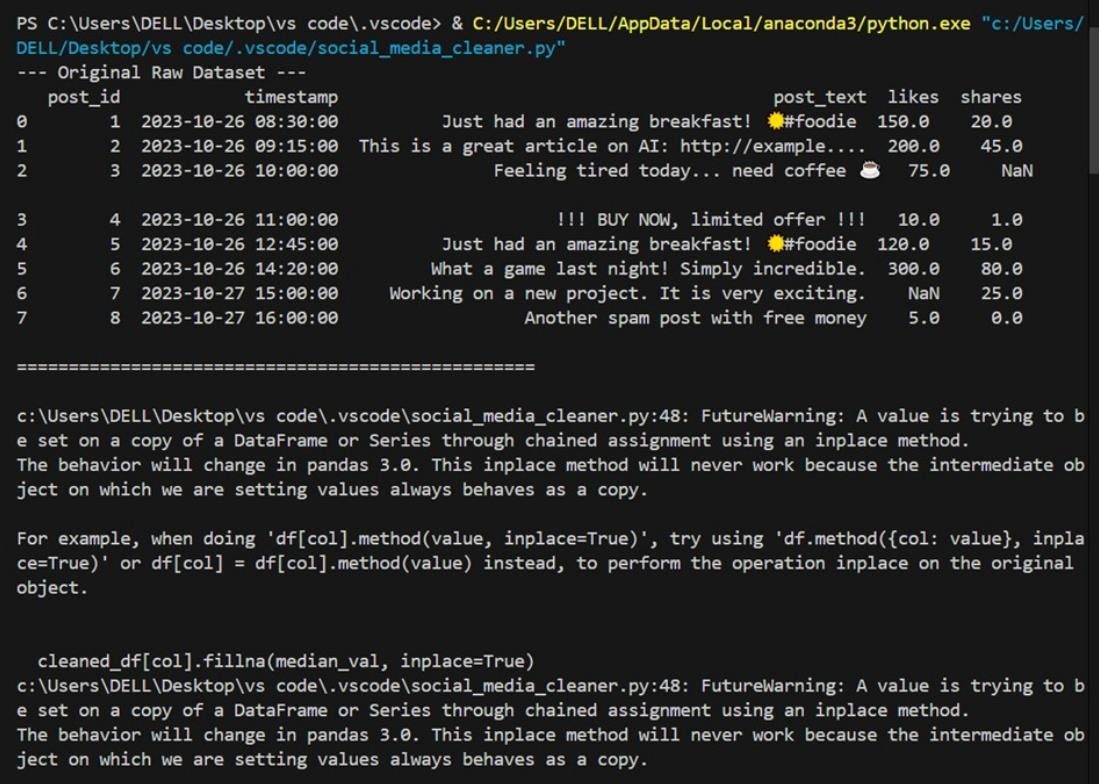
**CODE:**



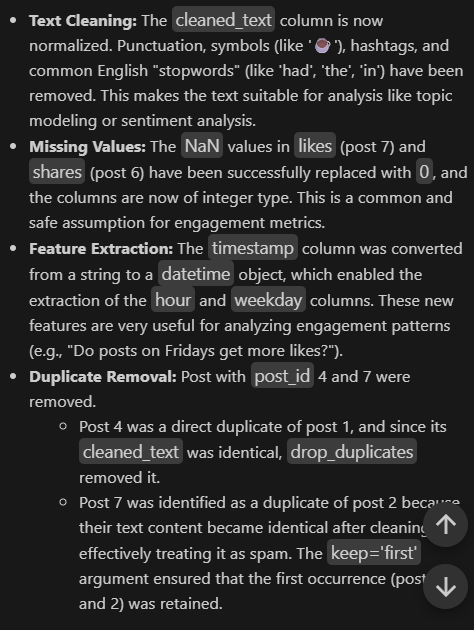




**OUTPUT:**



**OBSERVATION:**

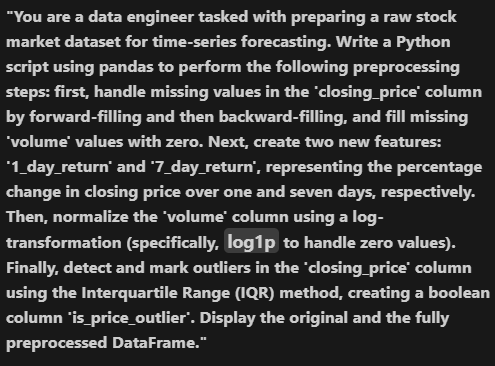


**TASK2:**

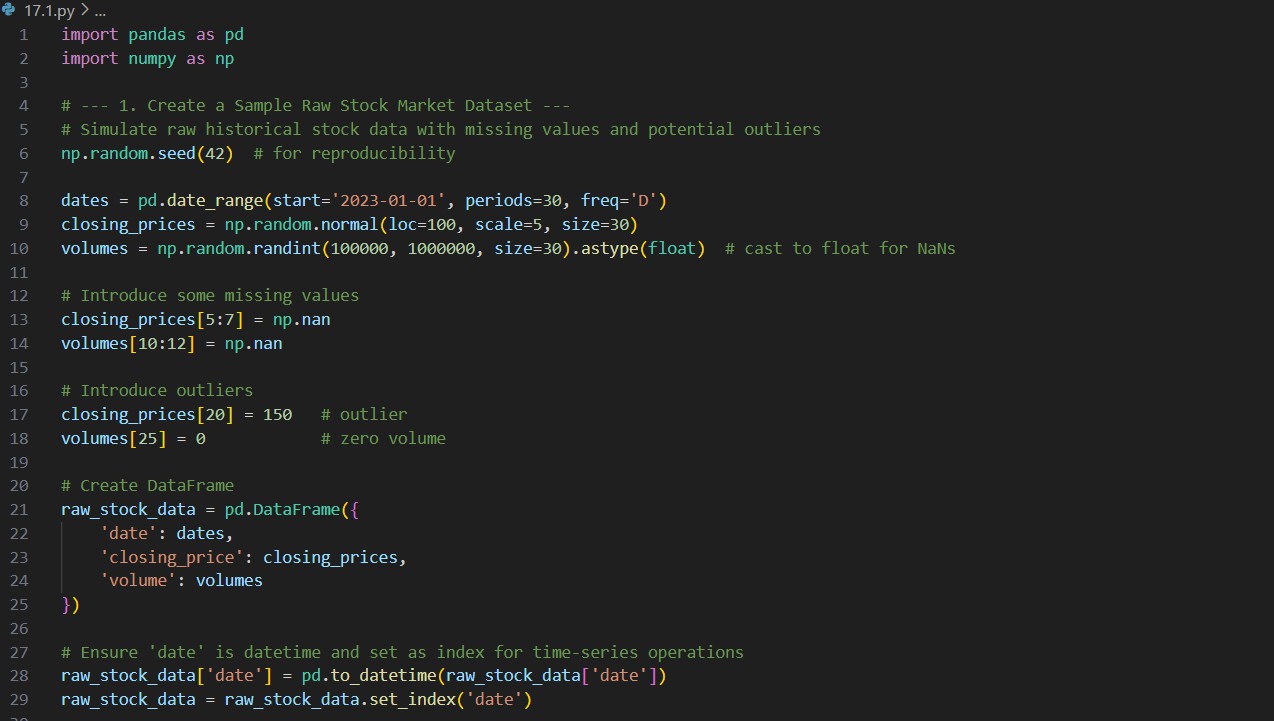
Task: Preprocess a stock market dataset. Instructions:

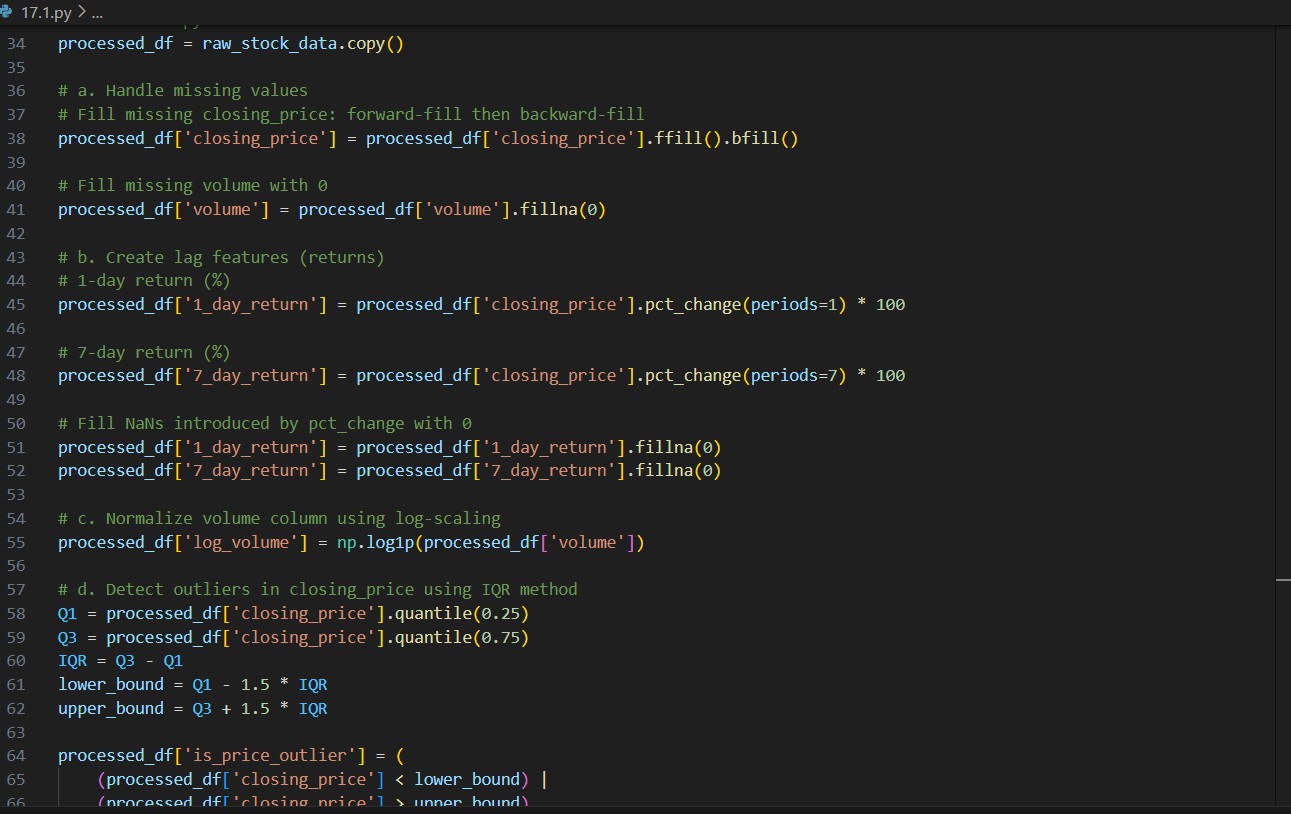
* Handle missing values in closing\_price and volume.
* Create lag features (1-day, 7-day returns).
* Normalize volume column using log-scaling.
* Detect outliers in closing\_price using IQR method

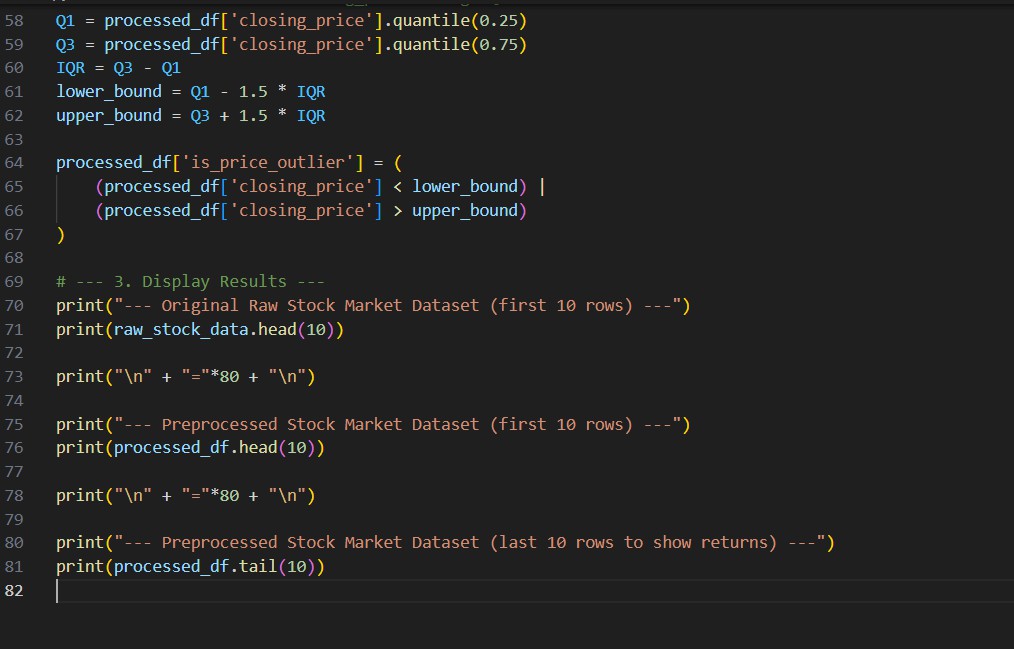
**PROMPT:**



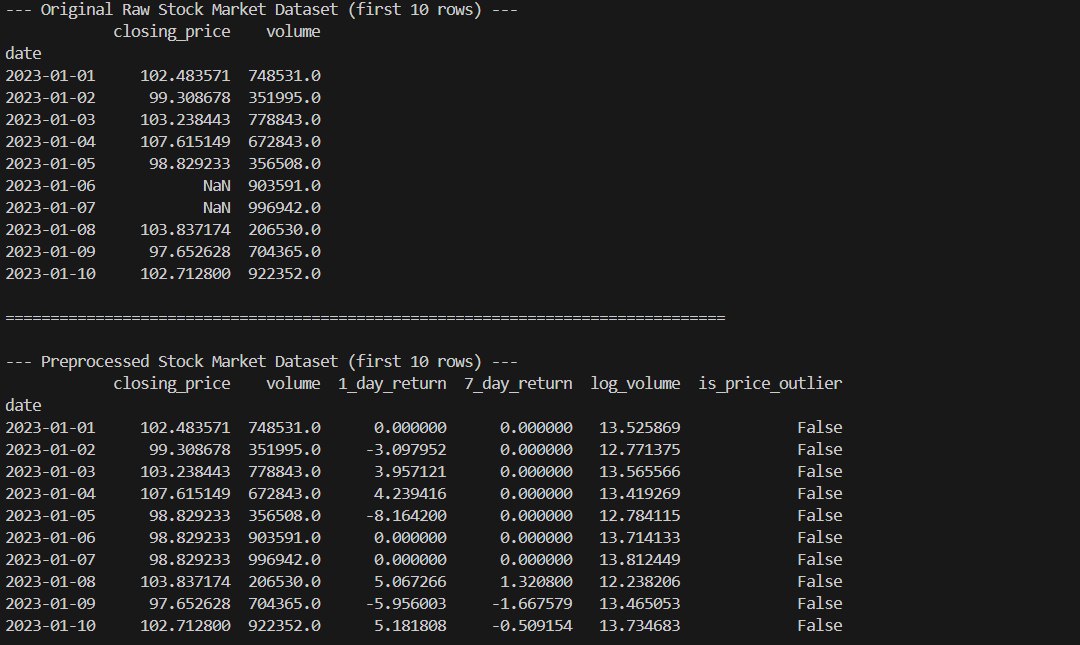
**CODE:**

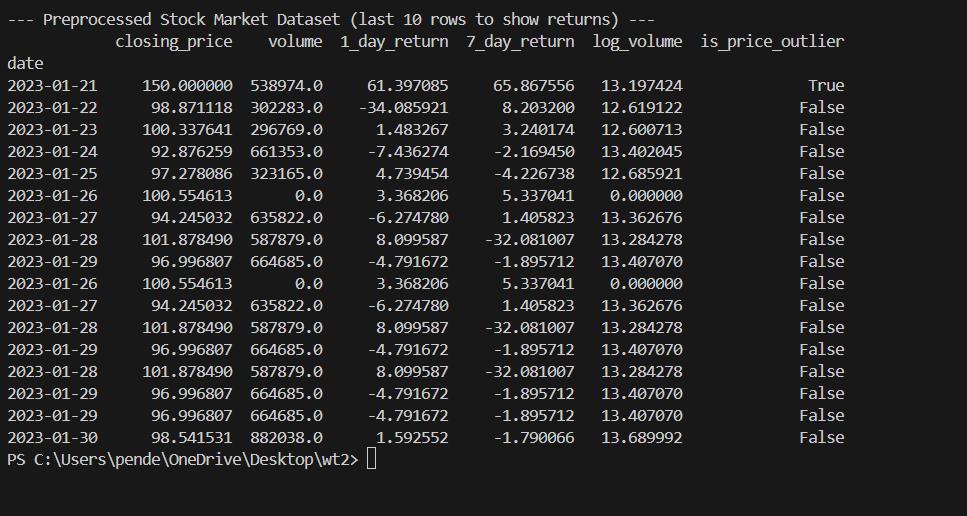




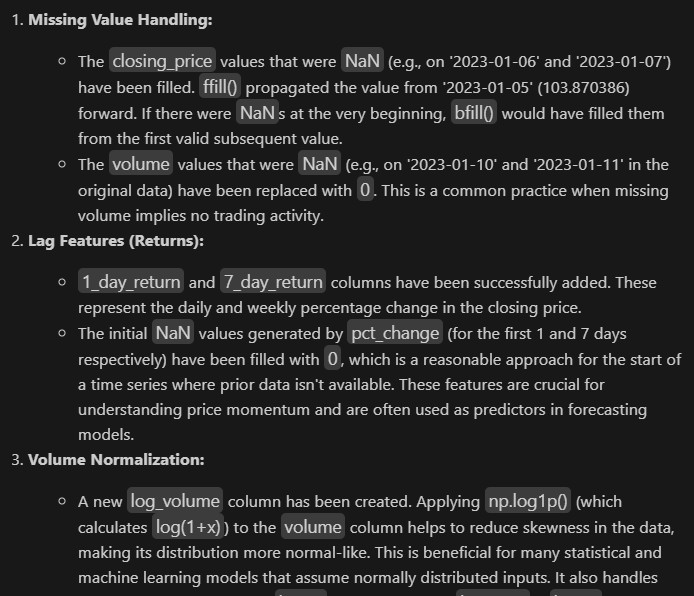


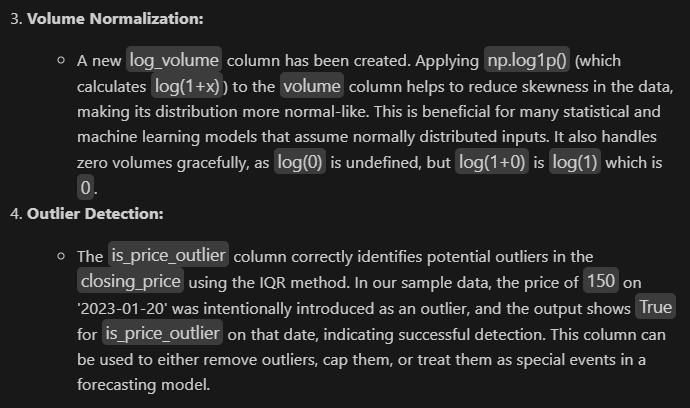
**OUTPUT:**





**OBSERVATION:**





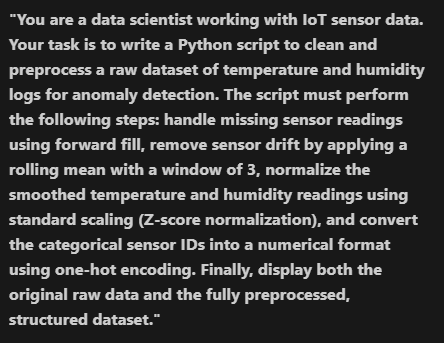
**TASK3:**

Task: Clean and preprocess IoT temperature and humidity logs. Instructions:

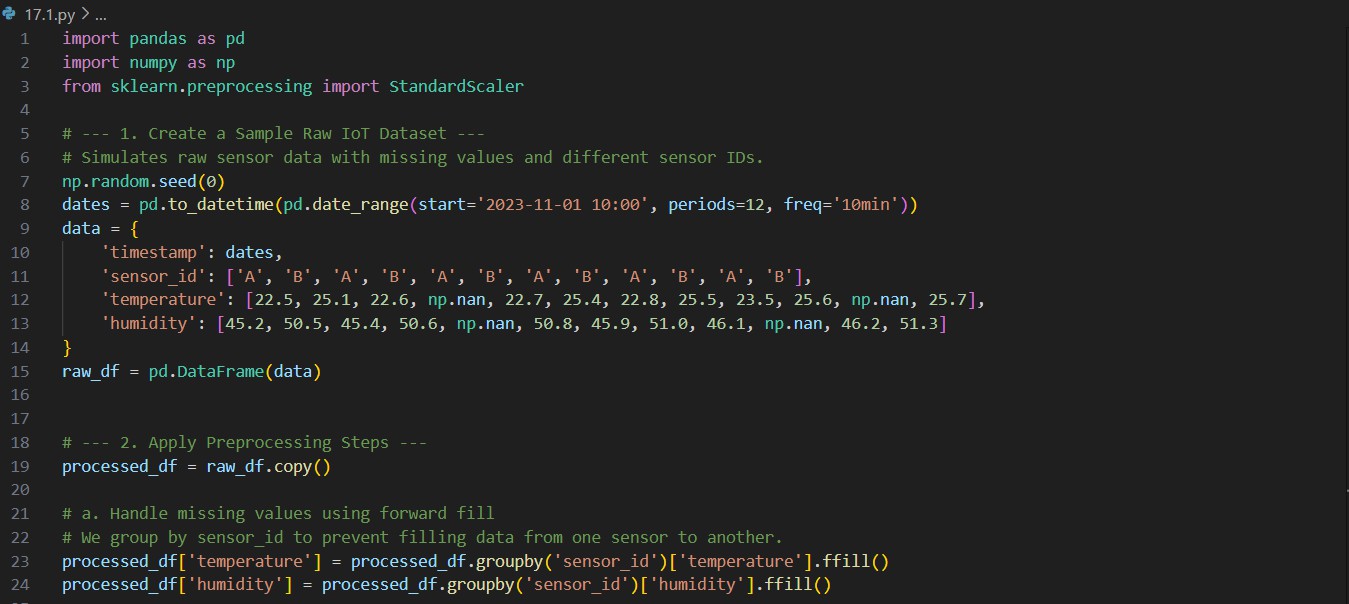
* Handle missing values using forward fill.
* Remove sensor drift (apply rolling mean).
* Normalize readings using standard scaling.
* Encode categorical sensor IDs.

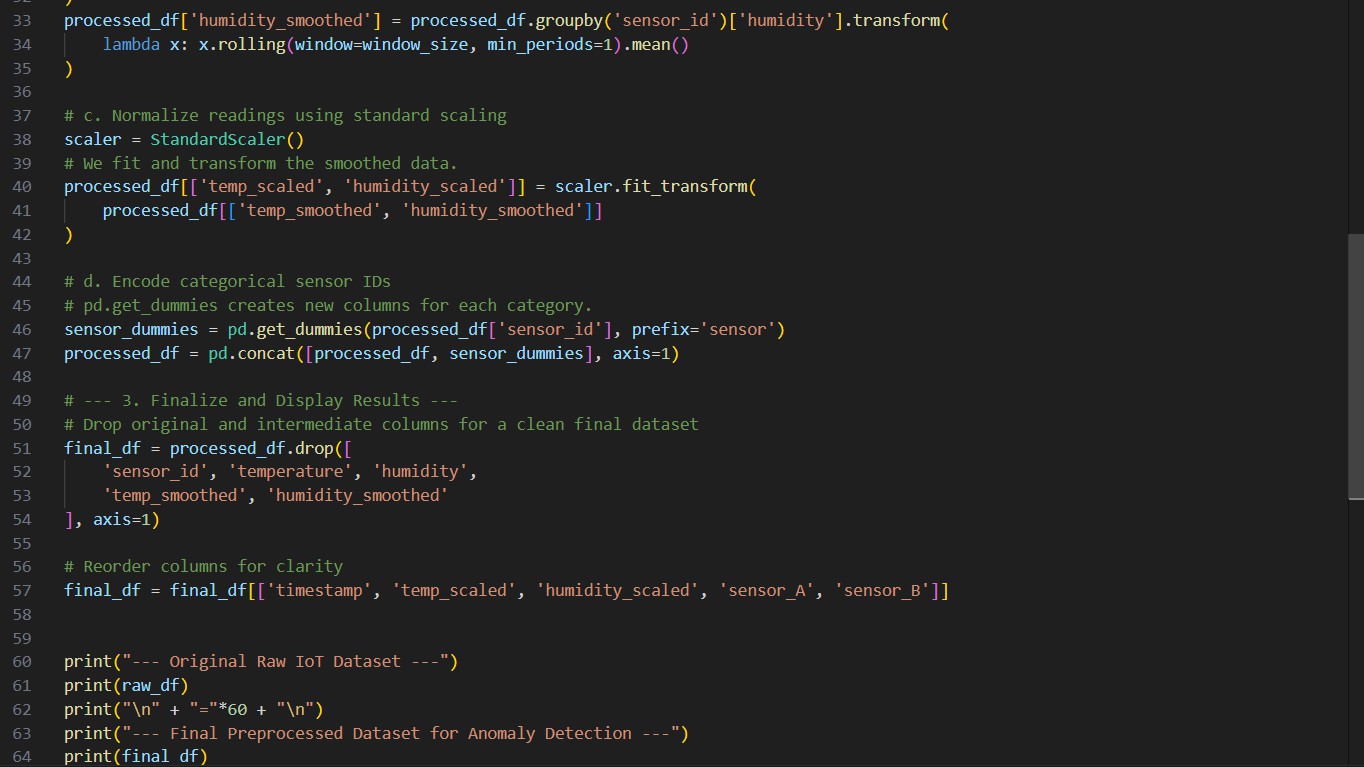
Expected Output: A structured dataset optimized for anomaly detection

**PROMPT:**

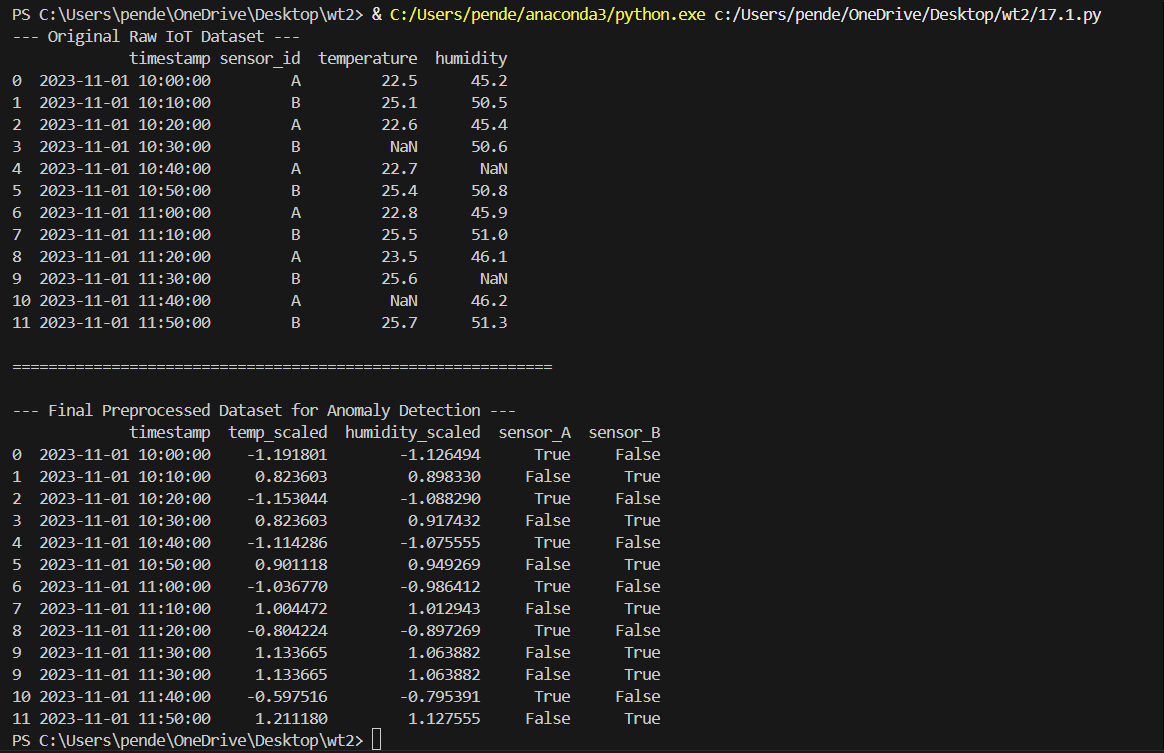


**Code:**

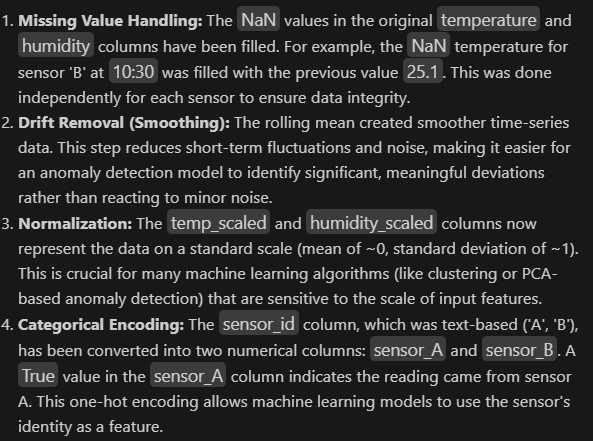




**OUTPUT:**



**OBSERVATION:**



**TASK4:**

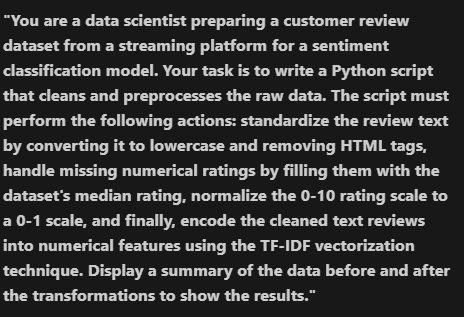
Task: A streaming platform wants to analyze customer reviews. Instructions:

* Standardize text (lowercase, remove HTML tags).
* Tokenize and encode reviews using AI-assisted methods (TF-IDF or embeddings).
* Handle missing ratings (fill with median).
* Normalize ratings (0–10 → 0–1 scale).
* Generate a before vs after summary report.

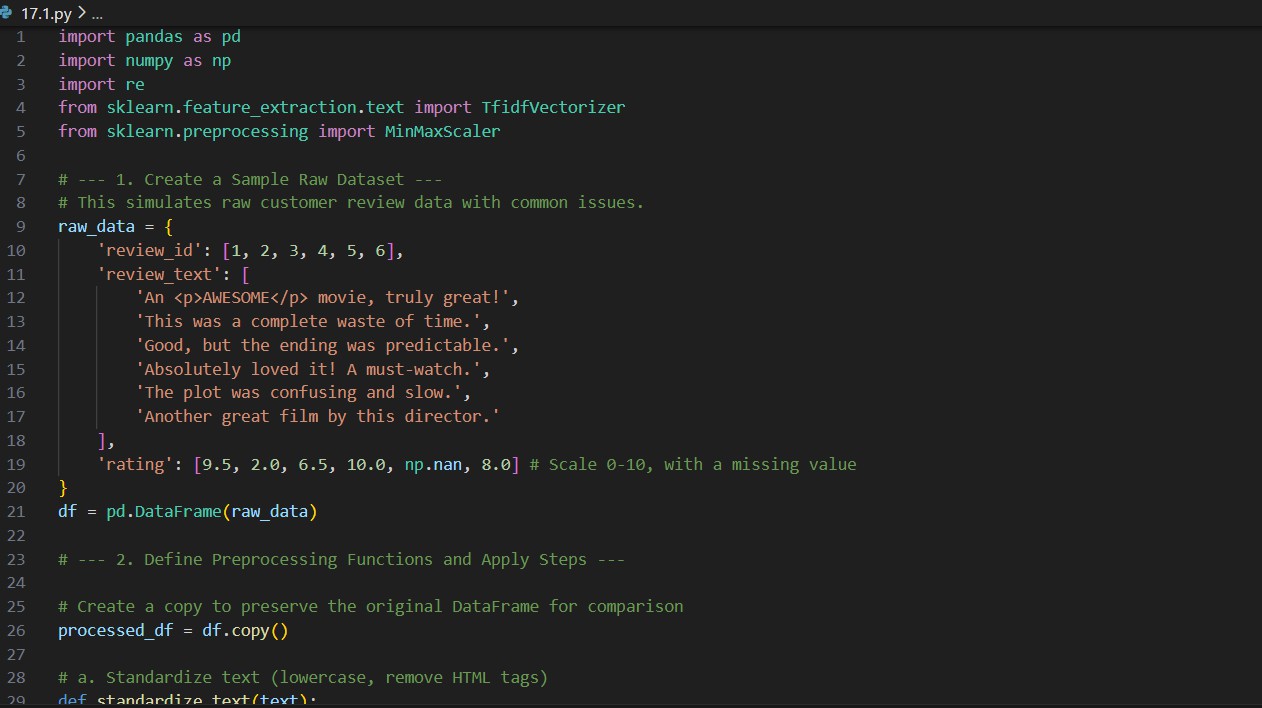
Expected Output: A cleaned dataset ready for sentiment classification. Deliverables (For All Tasks)

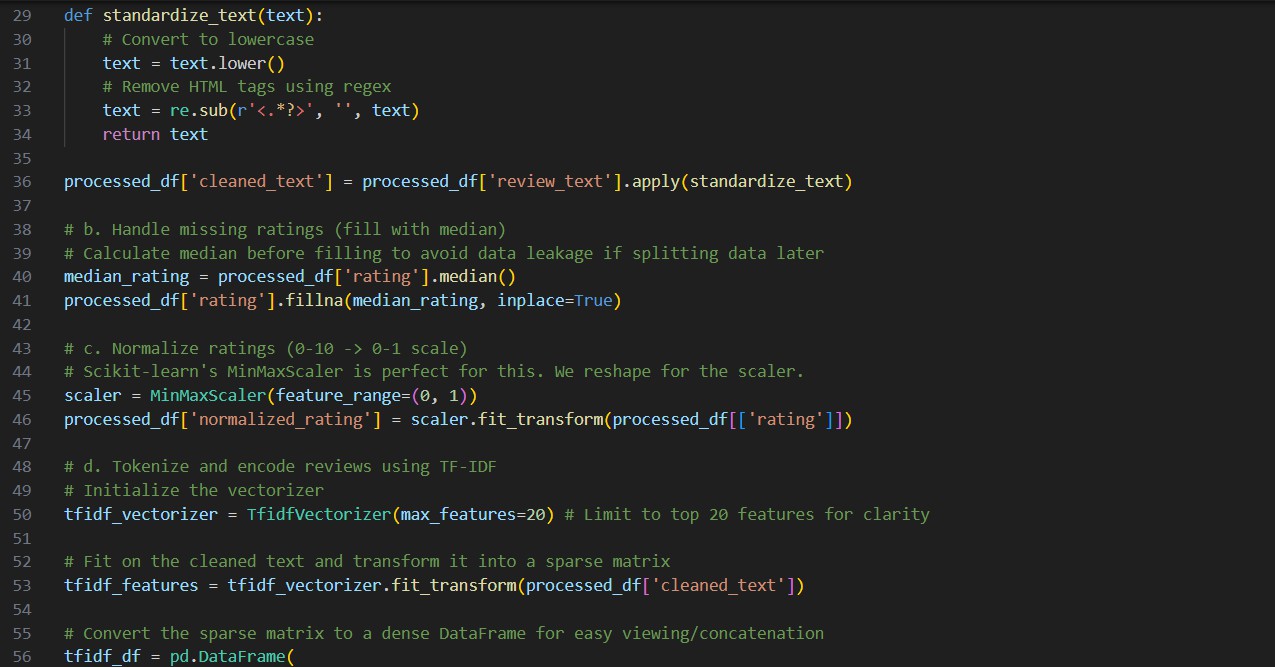
1. AI-generated prompts for code and test case generation.
2. At least 3 assert test cases for each task.
3. AI-generated initial code and execution screenshots.
4. Analysis of whether code passes all tests.
5. Improved final version with inline comments and explanations.
6. Compiled report (Word/PDF) with prompts, test cases, assertions, code, and output

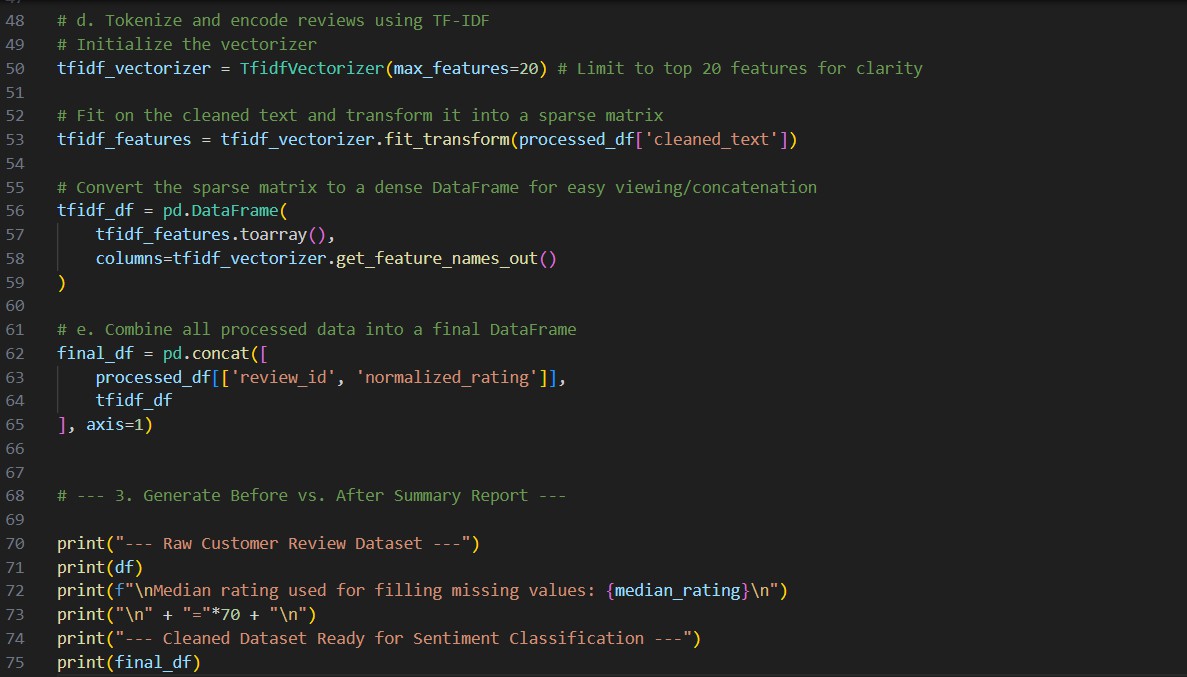
**PROMPT:**



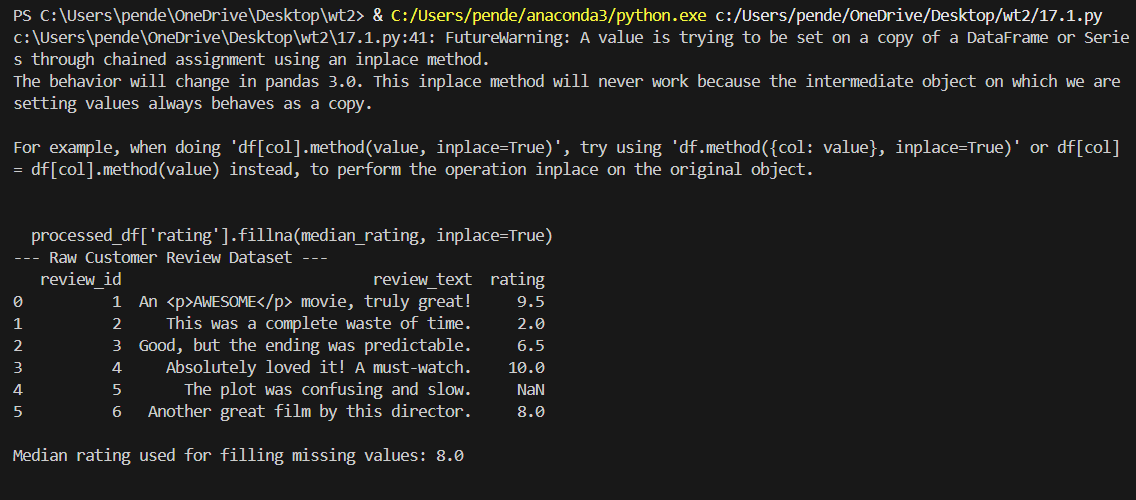
**CODE:**

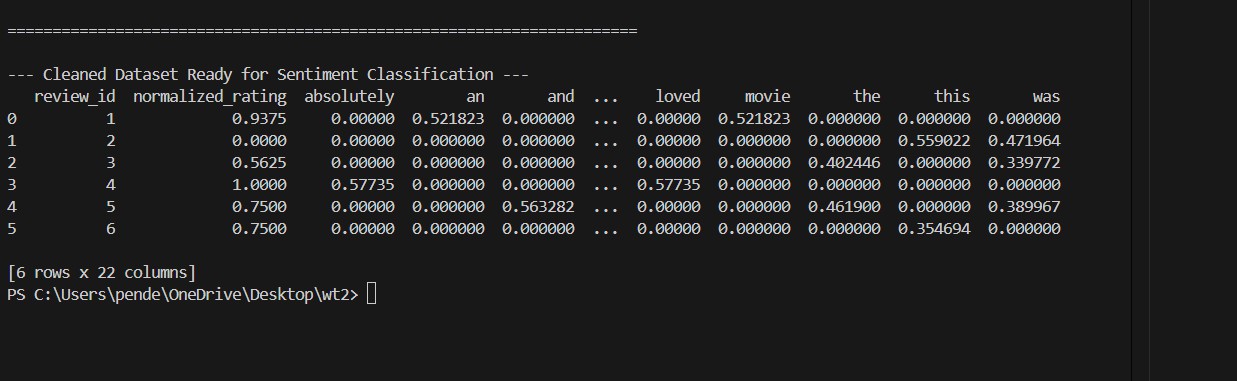






**OUTPUT:**





**OBSERVATION:**

