**AI ASSISTED CODING**

Name: H.V.S.Vyshnavi

Htno: 2403A51182

Batch: 06

**Lab-Assignment: 17.1**

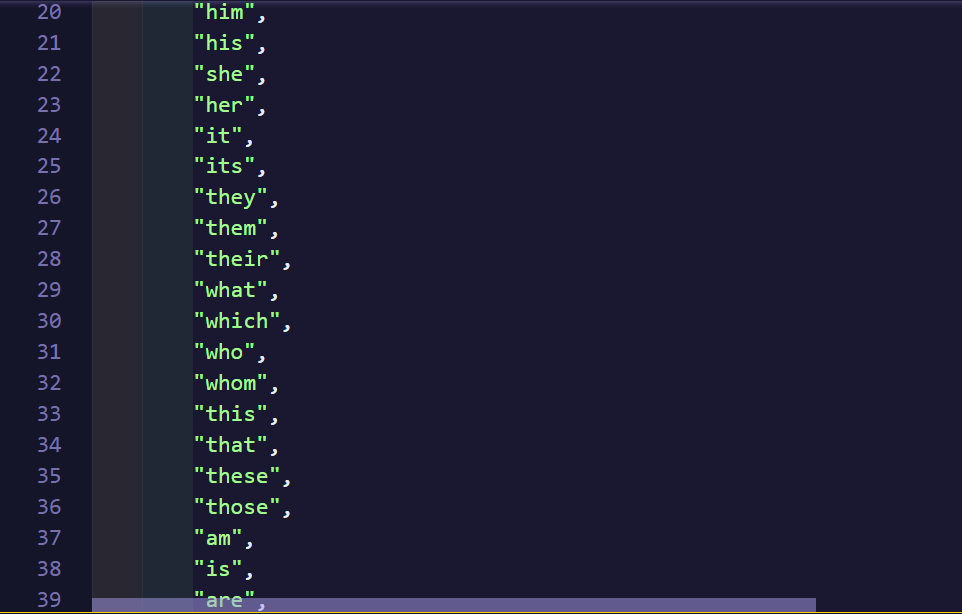
**Task-1:**

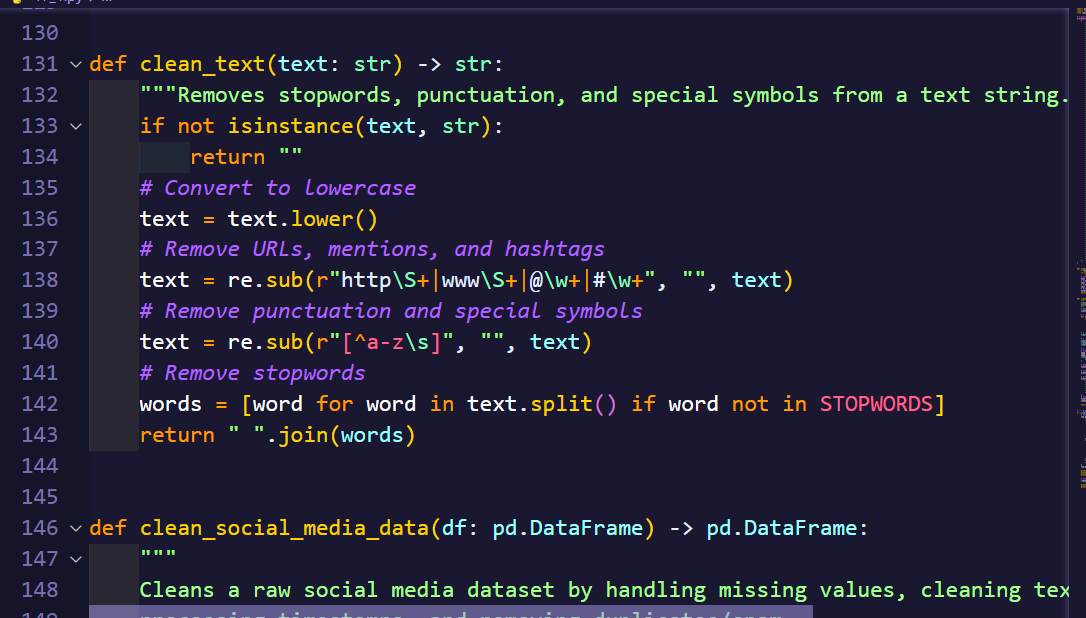
**Prompt:**

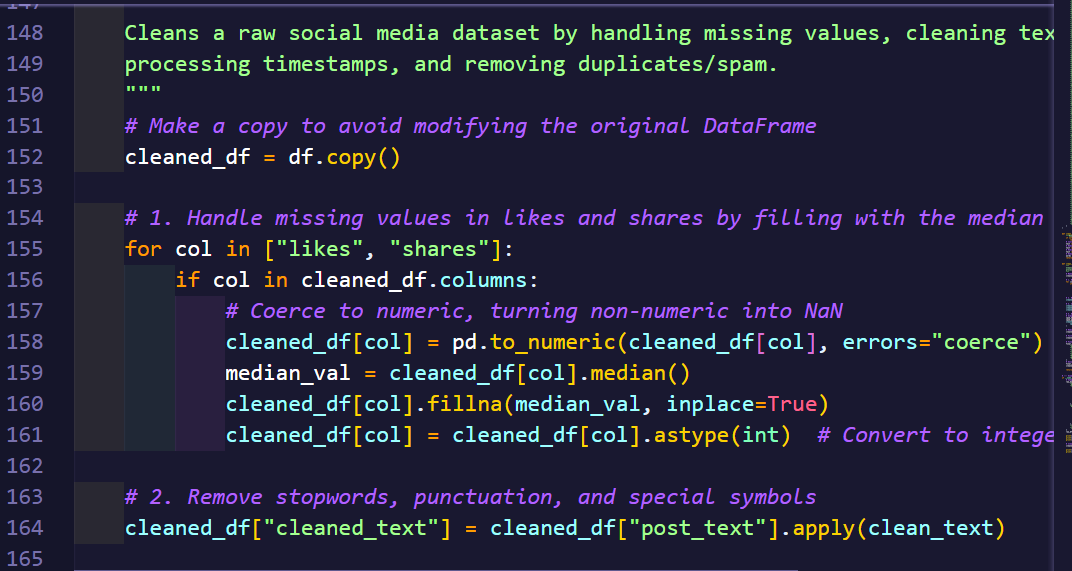
Write Python code to clean social media posts by removing stopwords, punctuation, and special symbols from text, handling missing values in likes and shares using the median, converting timestamps to datetime, extracting hour and weekday features, and removing duplicate or spam posts.

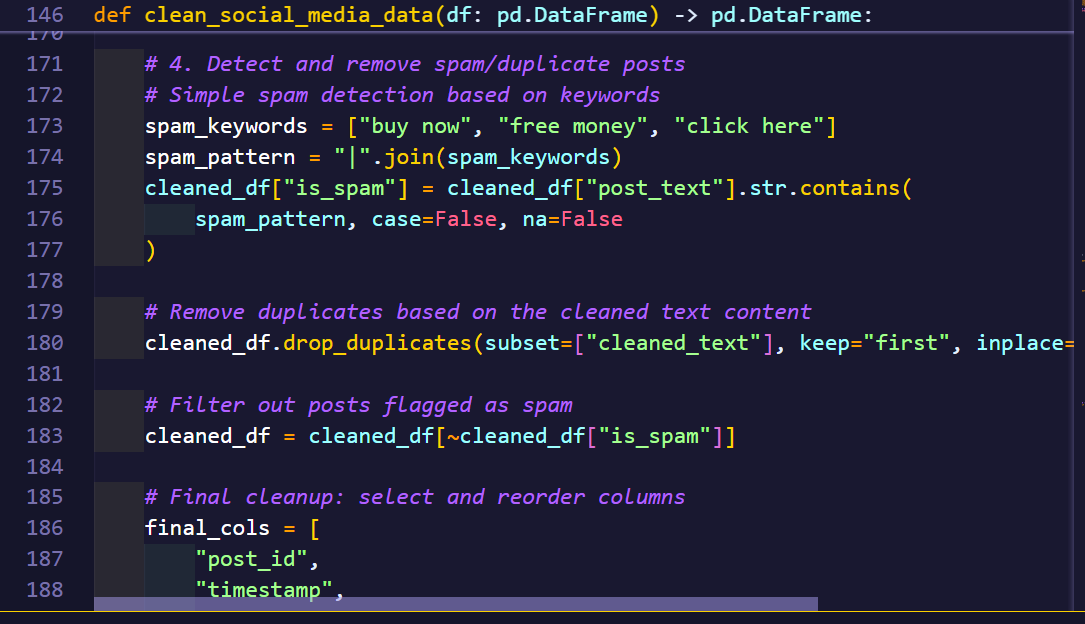
**Code:**

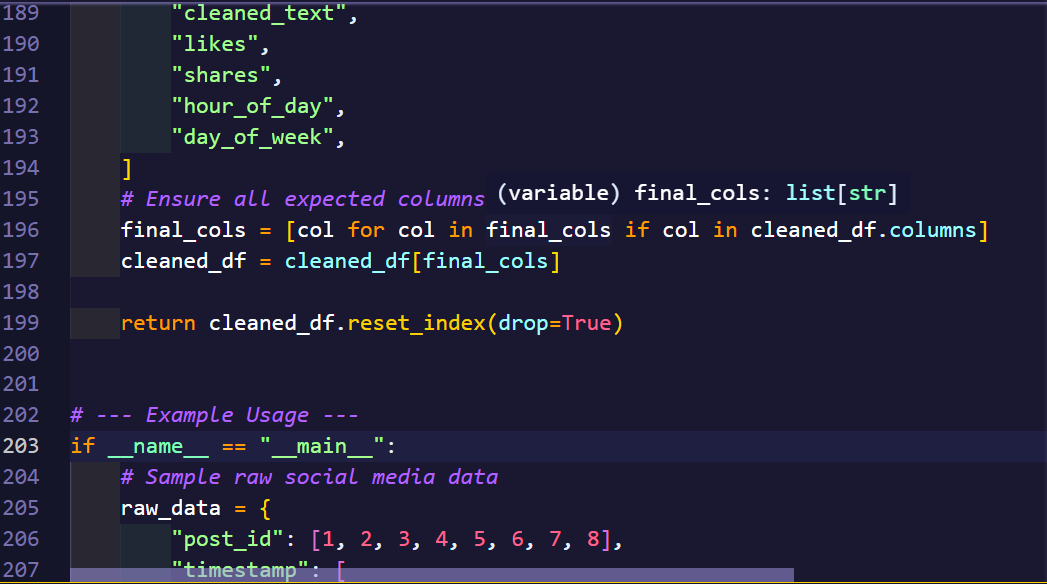


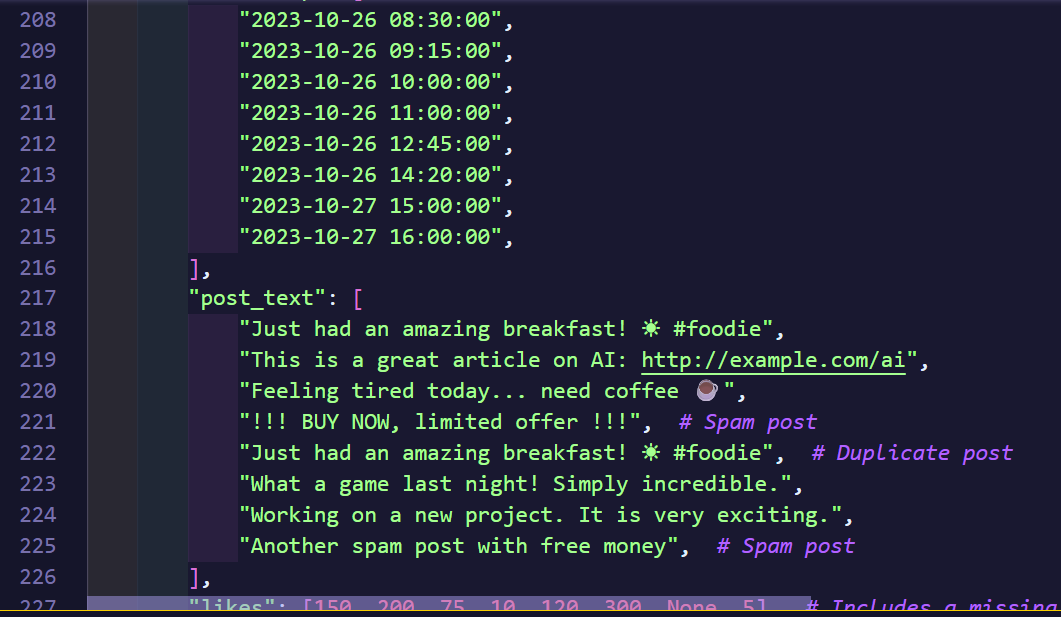


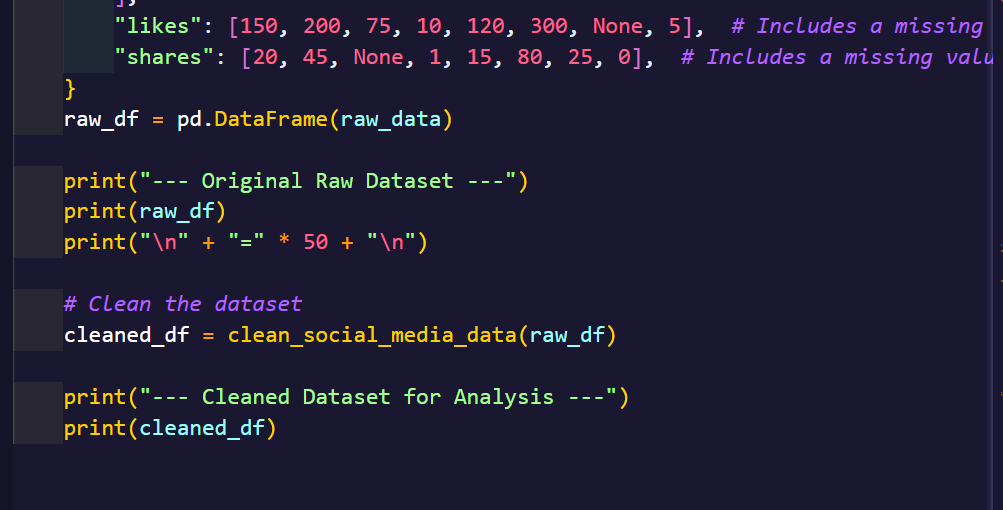




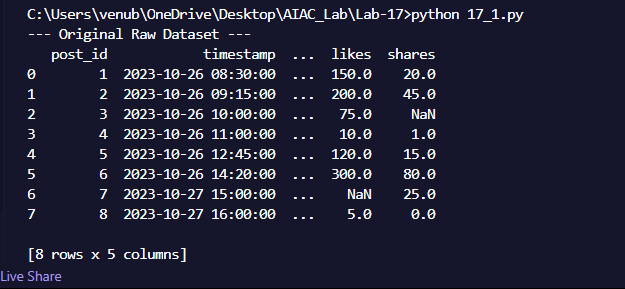
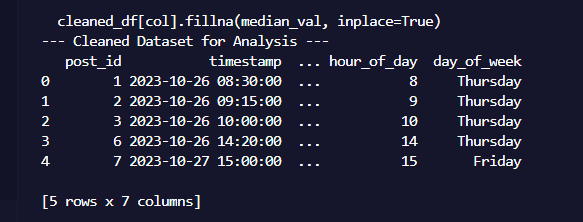








**Output:**

****

**Observation:**

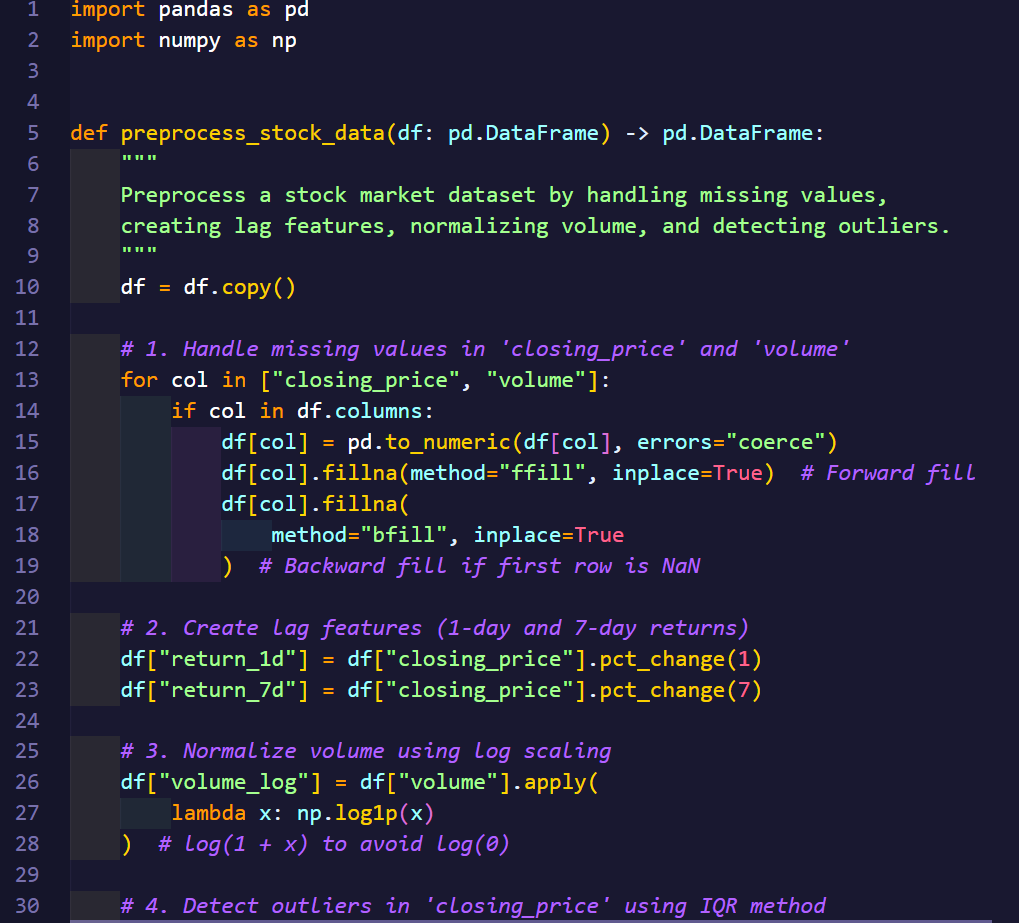
This task focuses on cleaning and preprocessing social media data to make it suitable for analysis. It involves removing noise from the text, such as stopwords, punctuation, URLs, mentions, and hashtags. Additionally, it handles missing values in numeric columns like likes and shares by replacing them with median values, and ensures these columns are of integer type. Time-related features, such as the hour of day and day of the week, are extracted from timestamps to allow for temporal analysis. The task also identifies and removes duplicate posts as well as spam posts containing keywords like “buy now”or “free money,” resulting in a cleaner and more reliable dataset for downstream tasks.

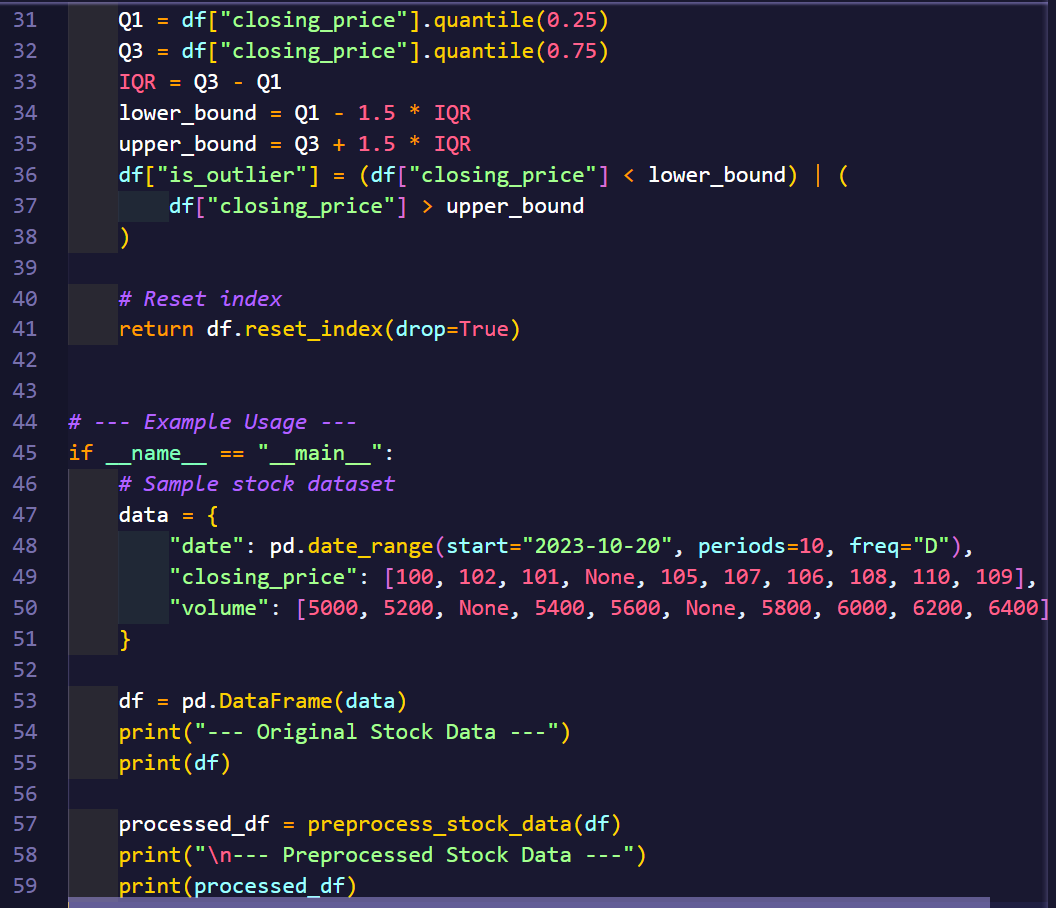
**Task-2:**

**Prompt:**

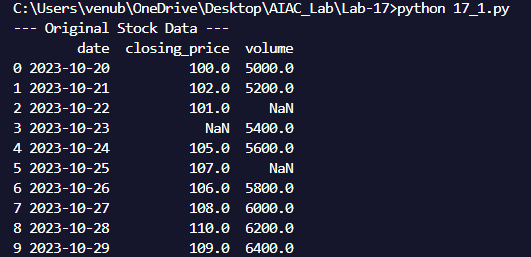
Write Python code to preprocess stock market data by handling missing values in closing\_price and volume, creating 1-day and 7-day lag return features, applying log-scaling to volume, and detecting outliers in closing\_price using the IQR method.

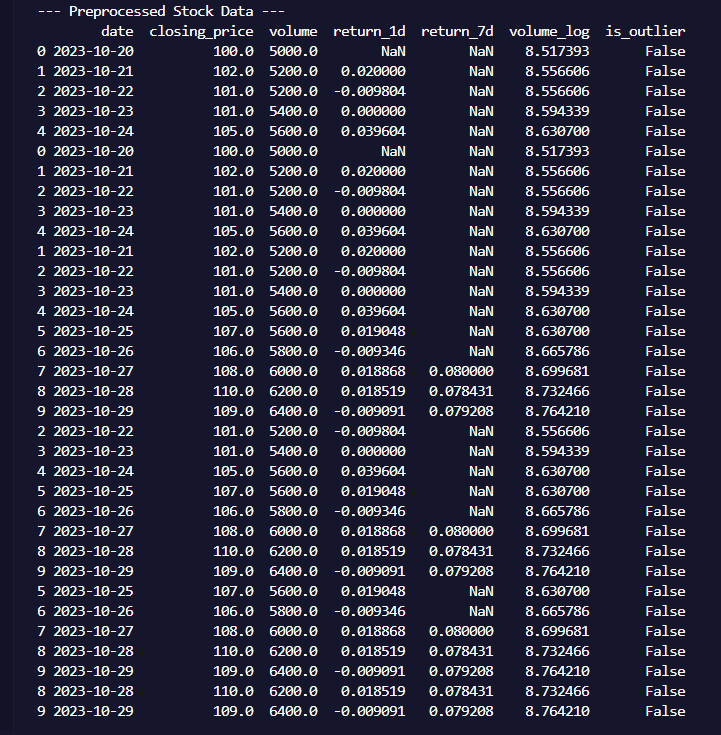
**Code:**

****

****

**Output:**

****

****

**Observation:**

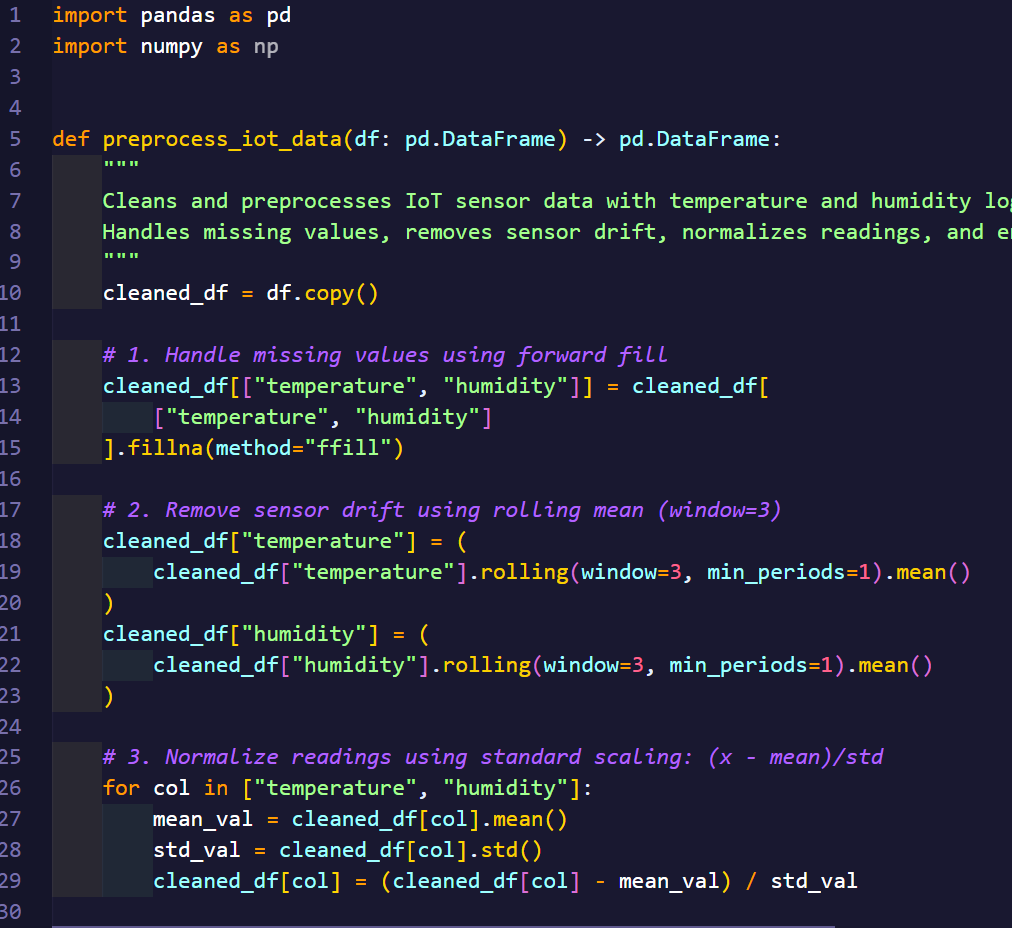
This deals with preprocessing financial or stock market data for predictive modeling or analysis. Missing values in key columns are filled with appropriate statistics, such as the mean or median. Lag features are created to capture temporal dependencies, and numeric columns like trading volume are normalized to bring them onto a comparable scale. Outliers are detected and handled to prevent them from skewing model predictions. This ensures the dataset is consistent, structured, and suitable for time-series forecasting or machine learning models.

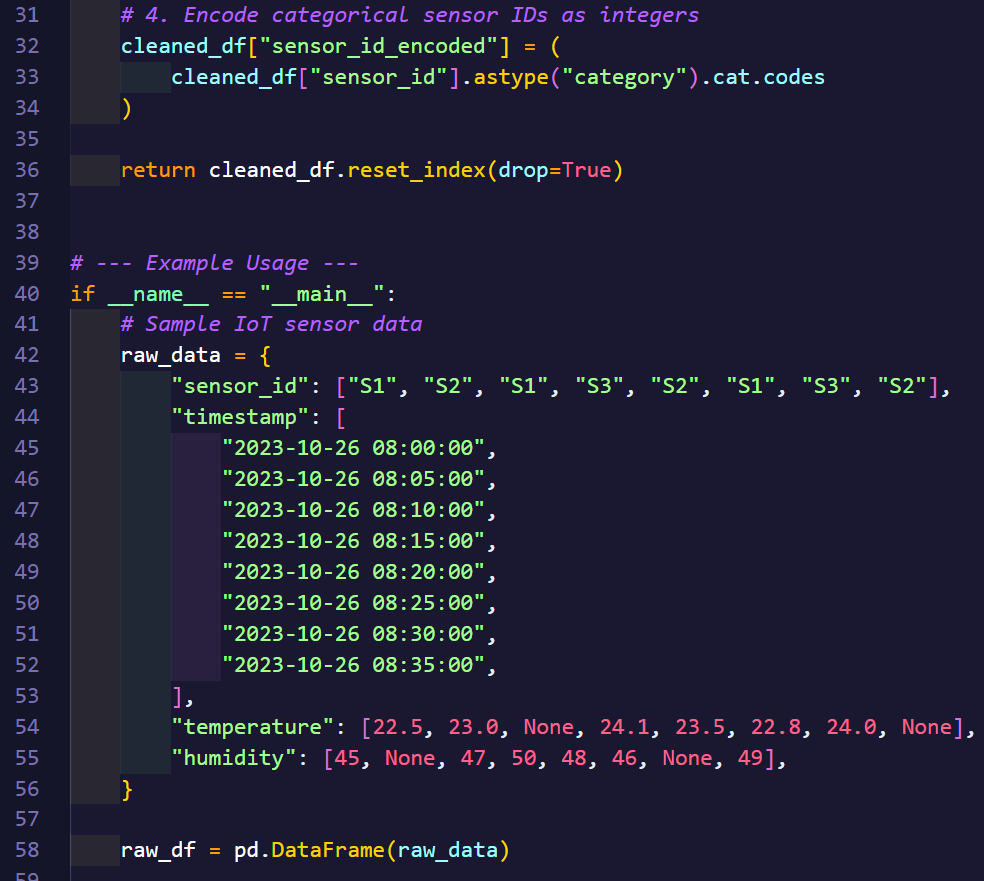
**Task-3:**

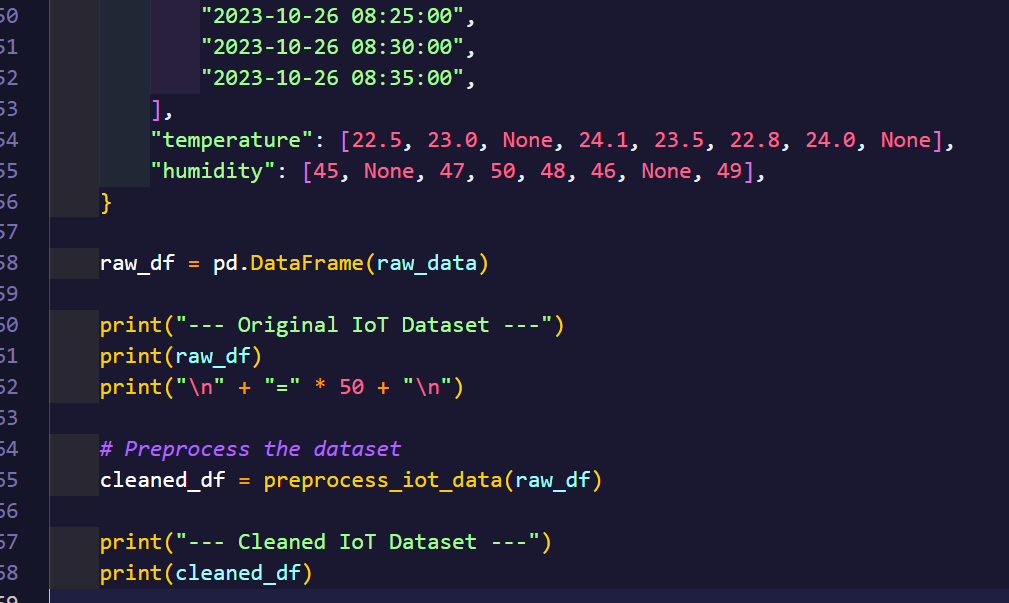
**Prompt:**

Write Python code to clean IoT sensor data by forward-filling missing values, applying a rolling mean to remove drift, normalizing temperature and humidity with standard scaling, and encoding categorical sensor IDs.

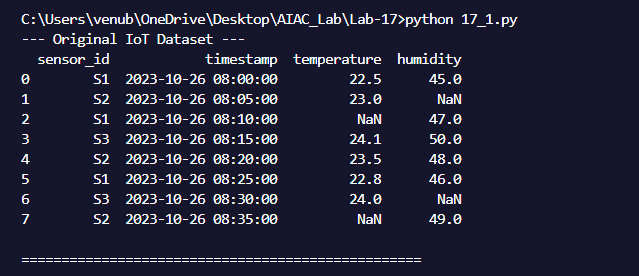
**Code:**

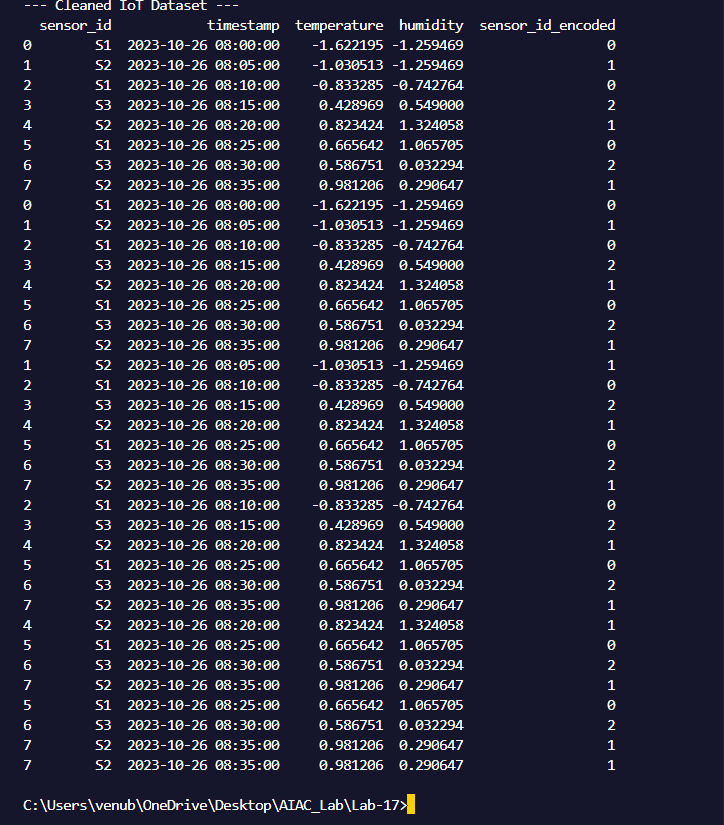
****

****

****

**Output:**

****

****

**Observation:**

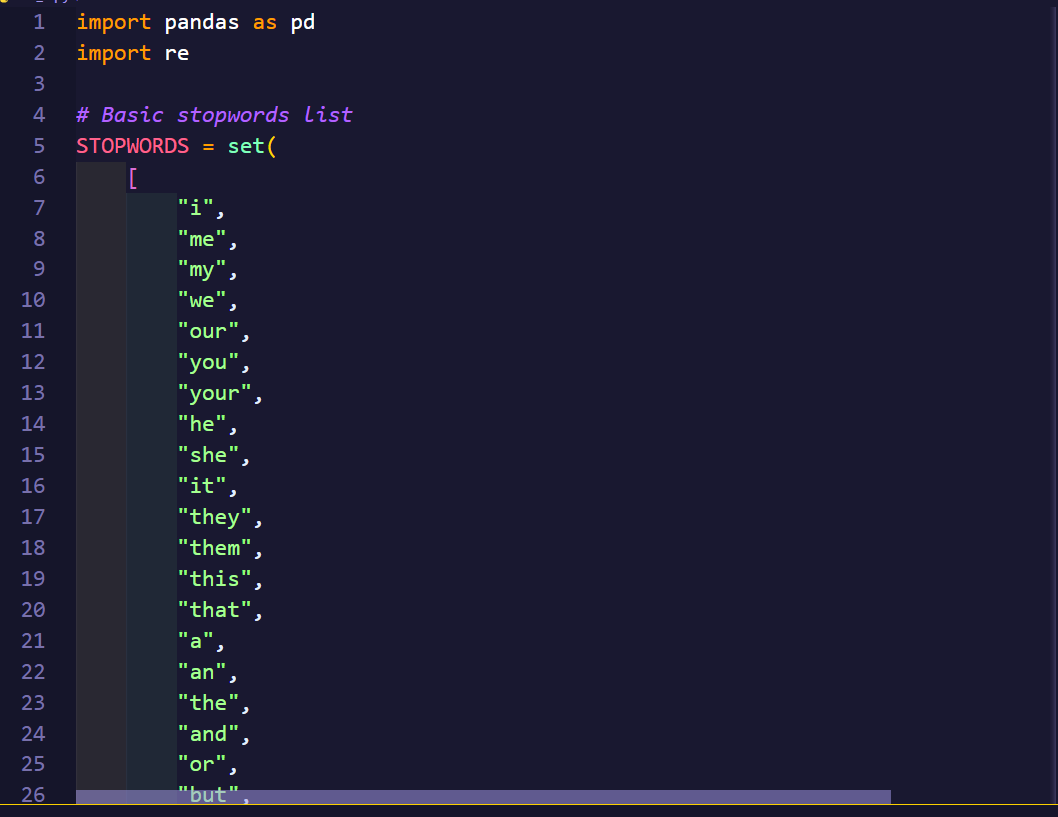
This task involves preprocessing IoT or sensor data, which often contains missing readings, noise, and inconsistencies due to sensor drift or device errors. Missing values are filled using appropriate imputation methods, and rolling averages or other smoothing techniques are applied to reduce noise. Sensor IDs or categorical features are encoded to numeric form so that models can process them effectively. The cleaned and normalized dataset is then ready for tasks like anomaly detection, predictive maintenance, or trend analysis.

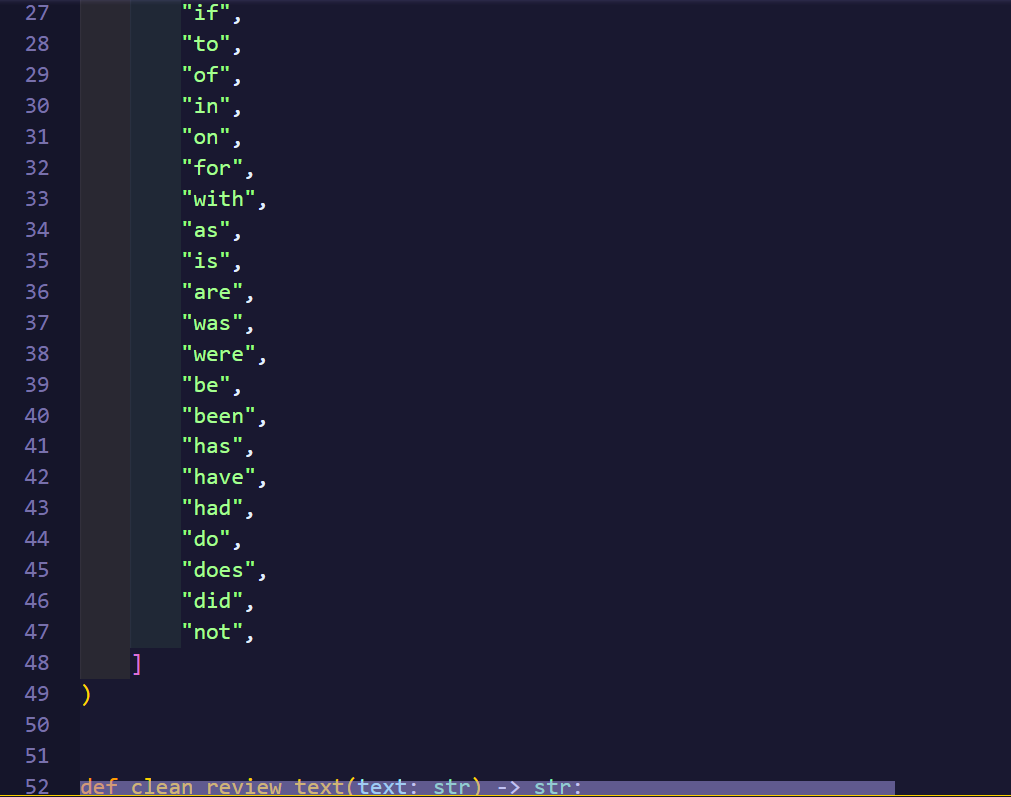
**Task-4:**

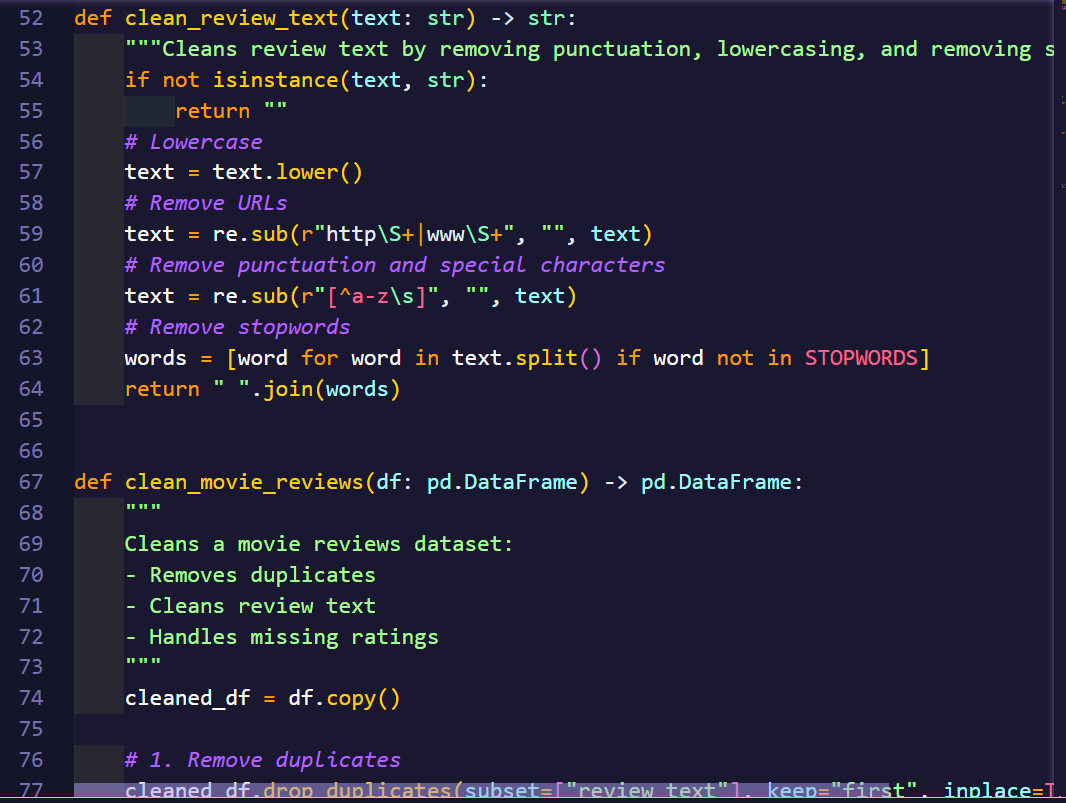
**Prompt:**

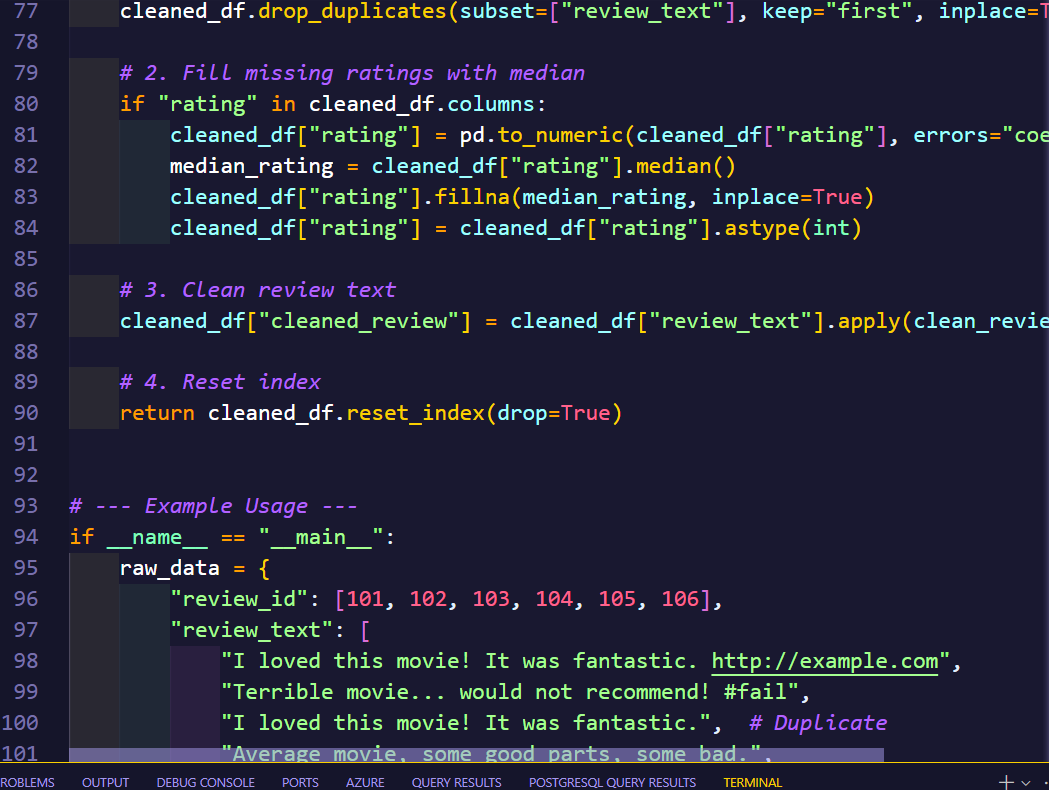
The prompt asked to clean reviews by lowercasing text, removing HTML tags, encoding using TF-IDF or embeddings, filling missing ratings with the median, normalizing ratings (0–10 → 0–1), and creating a before–after summary.

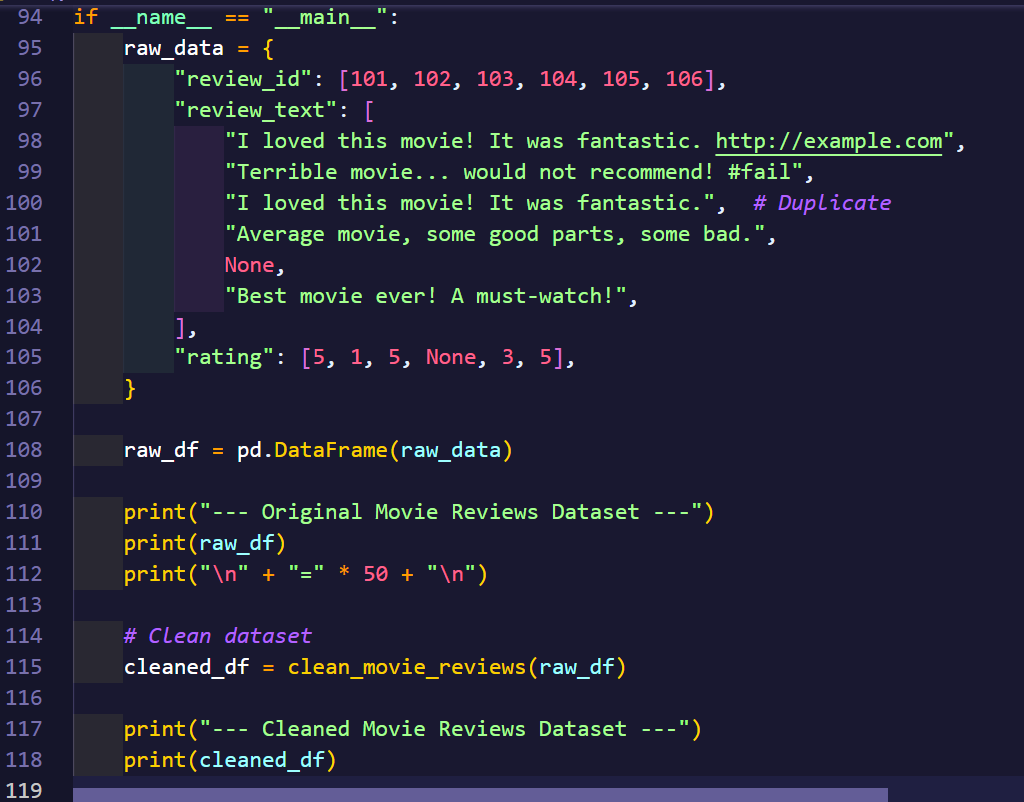
**Code:**

****

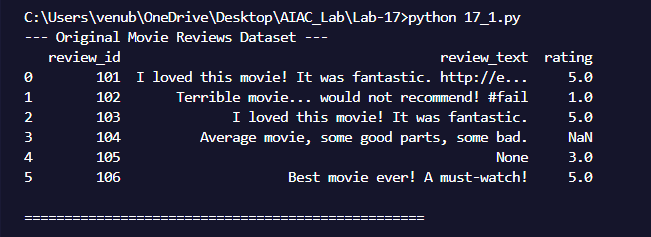
****

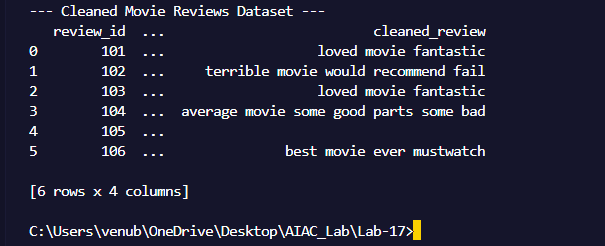
****

****

****

**Output:**

****

****

**Observation:**

This focuses on cleaning and preparing textual review data, such as movie or product reviews, for sentiment analysis or recommendation systems. Text preprocessing includes lowercasing, removing HTML tags, and optionally removing stopwords or special characters. The text is then encoded using methods like TF-IDF or embeddings to transform it into numeric features usable by machine learning models. Ratings are preprocessed by filling missing values with the median and normalizing them to a 0–1 scale. The task may also include generating before-and-after summaries to compare the raw and cleaned datasets, ensuring the text is ready for further analysis or modeling.