Assignment: Movie Review Data Analysis

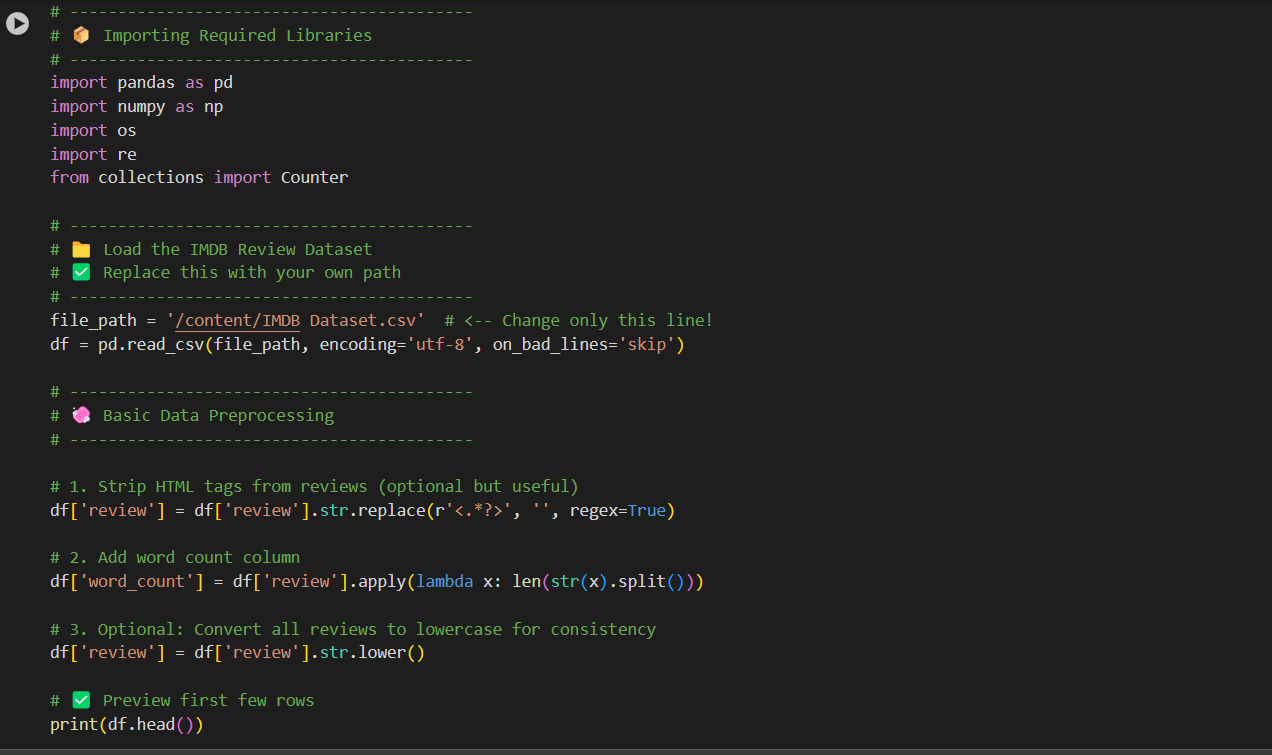
**Student Details**

* Name: UMESH KUYATE
* PRN: 202401080042
* Roll No: CS6-55
* Class: CS6

**Introduction**

In this assignment, I have worked with a real-world movie review dataset, performing text data analysis using Python libraries such as Pandas and Numpy. I formulated 20 problem statements based on the dataset and applied appropriate methods to find the solutions. Screenshots of the code execution are attached below each problem statement.

**NOTE : This is common code before running all 20 code**

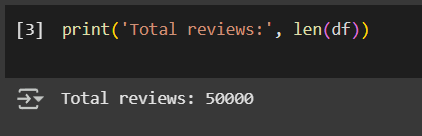
****

**Problem Statements and Solutions**

**Problem 1**

**Problem Statement:** Display the total number of reviews in the dataset.

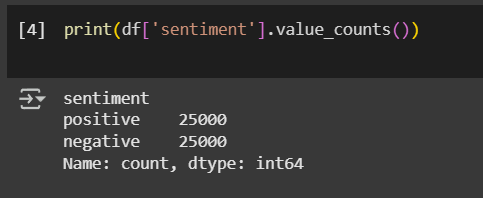
**Solution:**



**Problem 2**

**Problem Statement:** Count how many reviews are labeled as 'positive' and 'negative'.

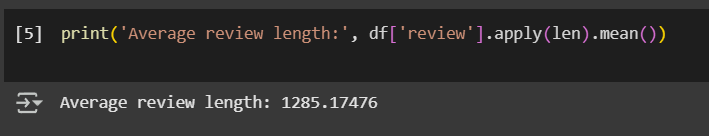
**Solution:**



**Problem 3**

**Problem Statement:** Find the average length of the reviews (in characters).

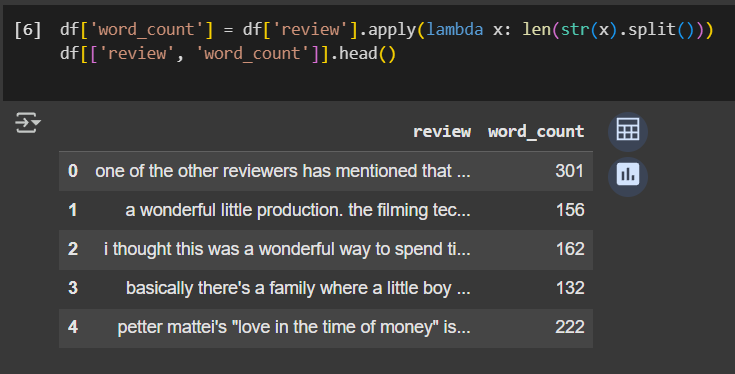
**Solution:**



**Problem 4**

**Problem Statement:** Add a new column showing the word count of each review..

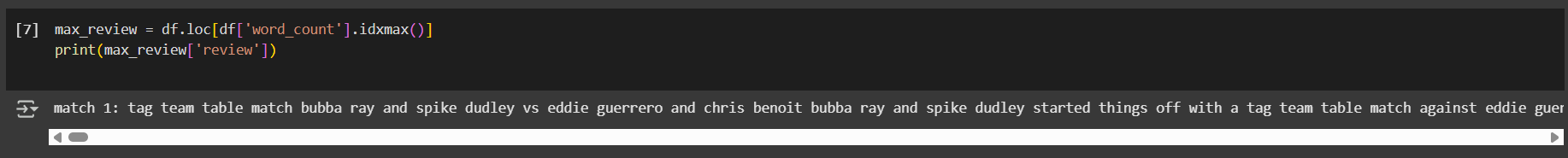
**Solution:**



**Problem 5**

**Problem Statement:** Find the review with the maximum word count.

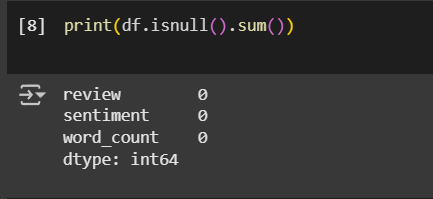
**Solution:**



**Problem 6**

**Problem Statement:Check if there are any missing values in the dataset.**.

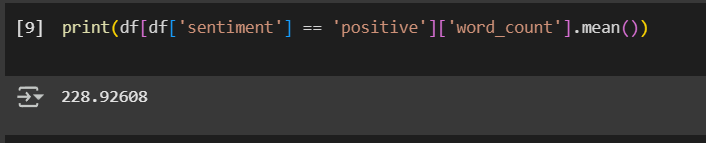
**Solution:**



**Problem 7**

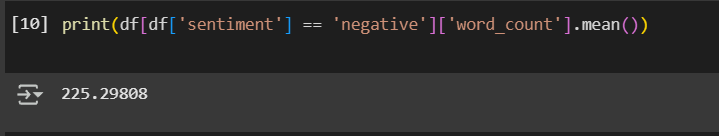
**Problem Statement:** Calculate the average word count for positive reviews.

**Solution:**



**Problem 8**

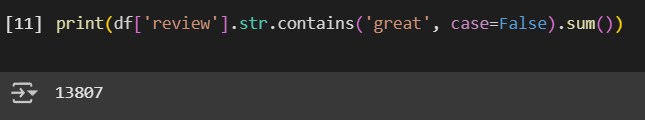
**Problem Statement:** Calculate the average word count for negative reviews.**Solution:**



**Problem 9**

**Problem Statement:** Count how many reviews contain the word 'great'.

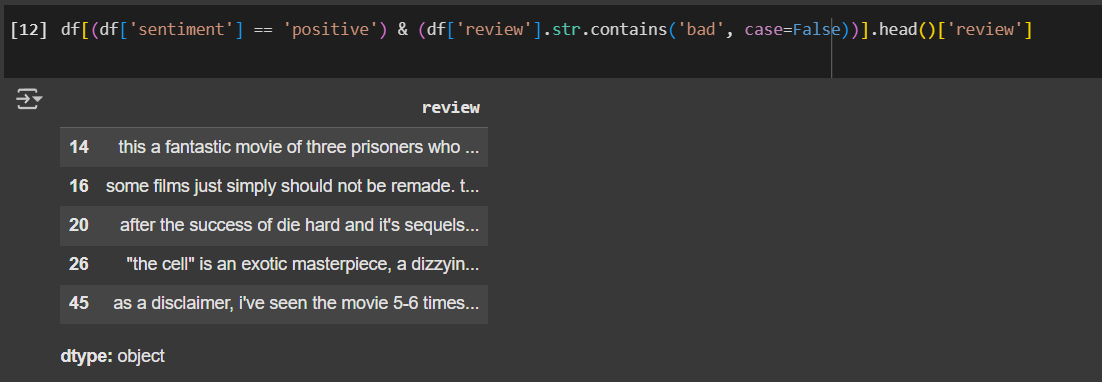
**Solution:**



**Problem 10**

**Problem StatementShow the first 5 positive reviews that mention 'bad'.**

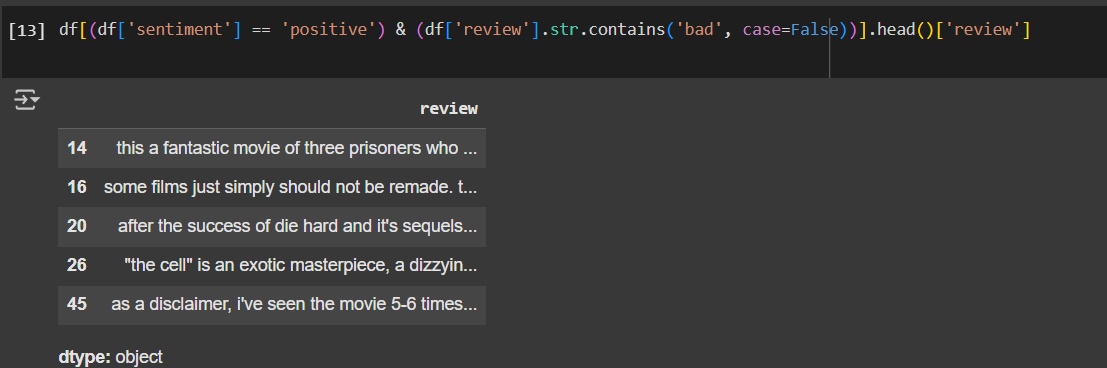
**Solution:**



**Problem 11**

**Problem Statement:** Remove HTML tags from reviews using regex..

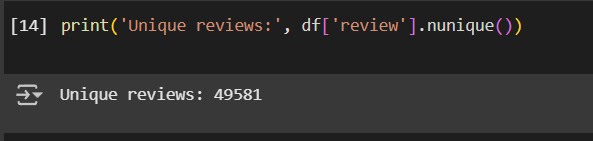
**Solution:**



**Problem 12**

**Problem Statement: Count the number of unique reviews in the dataset.**

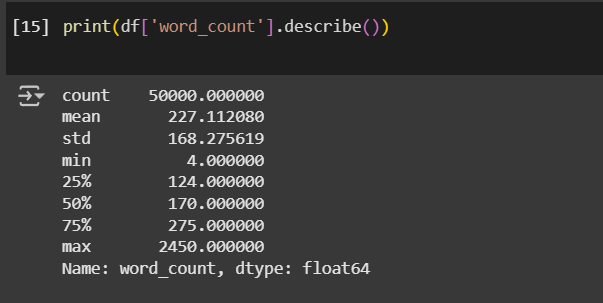
**Solution:**



**Problem 13**

**Problem StatementDisplay the distribution (min, max, mean) of word counts.**

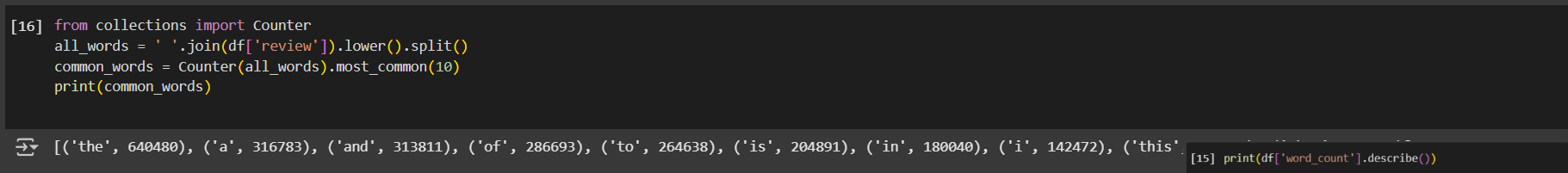
**Solution:**



**Problem 14**

**Problem Statement:** Show the 10 most frequent words in all reviews.

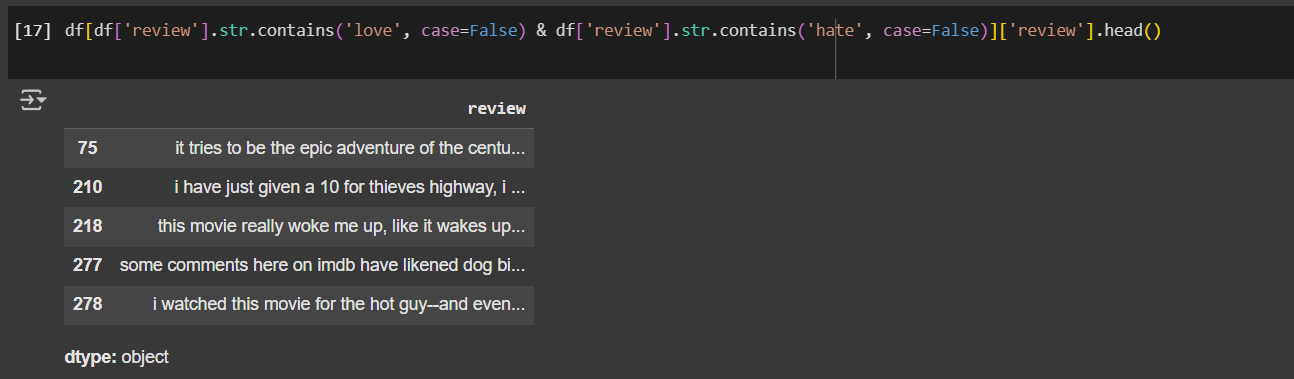
**Solution:**



**Problem 15**

**Problem Statement: Find reviews that mention both 'love' and 'hate'.**

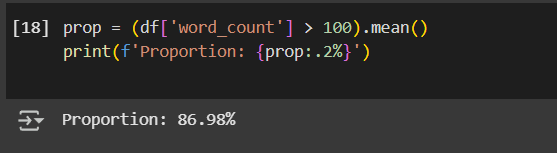
**Solution:**



**Problem 16**

**Problem StatementCalculate the proportion of reviews with more than 100 words.**

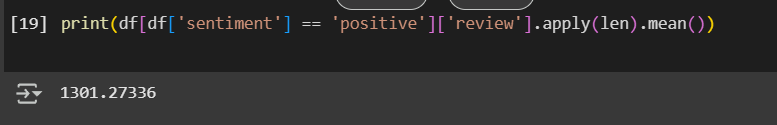
**Solution:**



**Problem 17**

**Problem Statement:** Get the average number of characters in positive reviews.

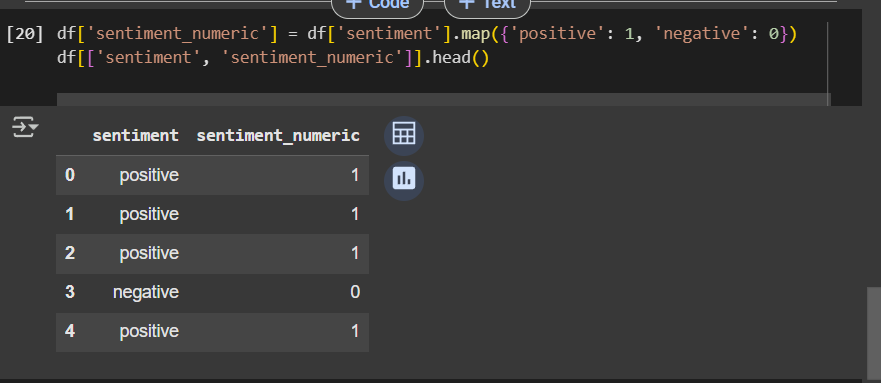
**Solution:**



**Problem 18**

**Problem Statement:** Convert sentiment labels to numeric (positive = 1, negative = 0).

**Solution:**



**Problem 19**

**Problem Statement:** Create a boolean column has\_excellent if 'excellent' appears in the review.

**Solution:**



**Problem 20**

**Problem Statement:** Find how many reviews have both more than 150 words and are labeled positive.

**Solution:**

