2738. Count Occurrences in Text

Premium

Solved

Medium | Topics

SQL Schema > Pandas Schema >

Table: `Files`

```
+-------------+---------+
| Column Name | Type    |
+-------------+---------+
| file_name   | varchar |
| content     | text    |
+-------------+---------+
file_name is the column with unique values of this table.
Each row contains file_name and the content of that file.
```

Write a solution to find the number of files that have at least one occurrence of the words **'bull'** and **'bear'** as a **standalone word**, respectively, disregarding any instances where it appears without space on either side (e.g. 'bullet', 'bears', 'bull.', or 'bear' at the beginning or end of a sentence will **not** be considered)

Return *the word 'bull' and 'bear' along with the corresponding number of occurrences in* ***any order.***

The result format is in the following example.

Example 1:

```
Input:
Files table:
+------------+----------------------------------------------------------------------------------+
| file_name  | content                                                                          |
+------------+----------------------------------------------------------------------------------+
| draft1.txt | The stock exchange predicts a bull market which would make many investors happy. |
| draft2.txt | The stock exchange predicts a bull market which would make many investors happy, |
|            | but analysts warn of possibility of too much optimism and that in fact we are    |
|            | awaiting a bear market.                                                          |
| draft3.txt | The stock exchange predicts a bull market which would make many investors happy, |
|            | but analysts warn of possibility of too much optimism and that in fact we are    |
|            | awaiting a bear market. As always predicting the future market is an uncertain   |
|            | game and all investors should follow their instincts and best practices.         |
+------------+----------------------------------------------------------------------------------+
Output:
+------+-------+
| word | count |
+------+-------+
| bull | 3     |
| bear | 2     |
+------+-------+
Explanation:
- The word "bull" appears 1 time in "draft1.txt", 1 time in "draft2.txt", and 1 time in "draft3.txt". Therefore, the total number of occurrences for the word "bull" is 3.
- The word "bear" appears 1 time in "draft2.txt", and 1 time in "draft3.txt". Therefore, the total number of occurrences for the word "bear" is 2.
```

Seen this question in a real interview before? 1/5

Yes No

Accepted 10,754/19.6K | Acceptance Rate 54.8%

Topics

Discussion (11)

Code

Pandas · Auto

```python
import pandas as pd

def count_occurrences(files: pd.DataFrame) -> pd.DataFrame:

    bull_cnt = files[files.content.str.contains(' bull ') ].shape[0]
    bear_cnt = files[files.content.str.contains(' bear ') ].shape[0]
    return pd.DataFrame( {'word': ['bull', 'bear'],
                          'count': [bull_cnt, bear_cnt]})
```

Saved | Ln 9, Col 5

Testcase | Test Result

Accepted Runtime: 226 ms

Case 1

Input

Files =

```
| file_name  | content
|
| ---------- | ------------------------------------------------------------------------------------------------
| draft1.txt | The stock exchange predicts a bull market which would make many investors happy.
```