**Solution: Preliminary Data Exploration**

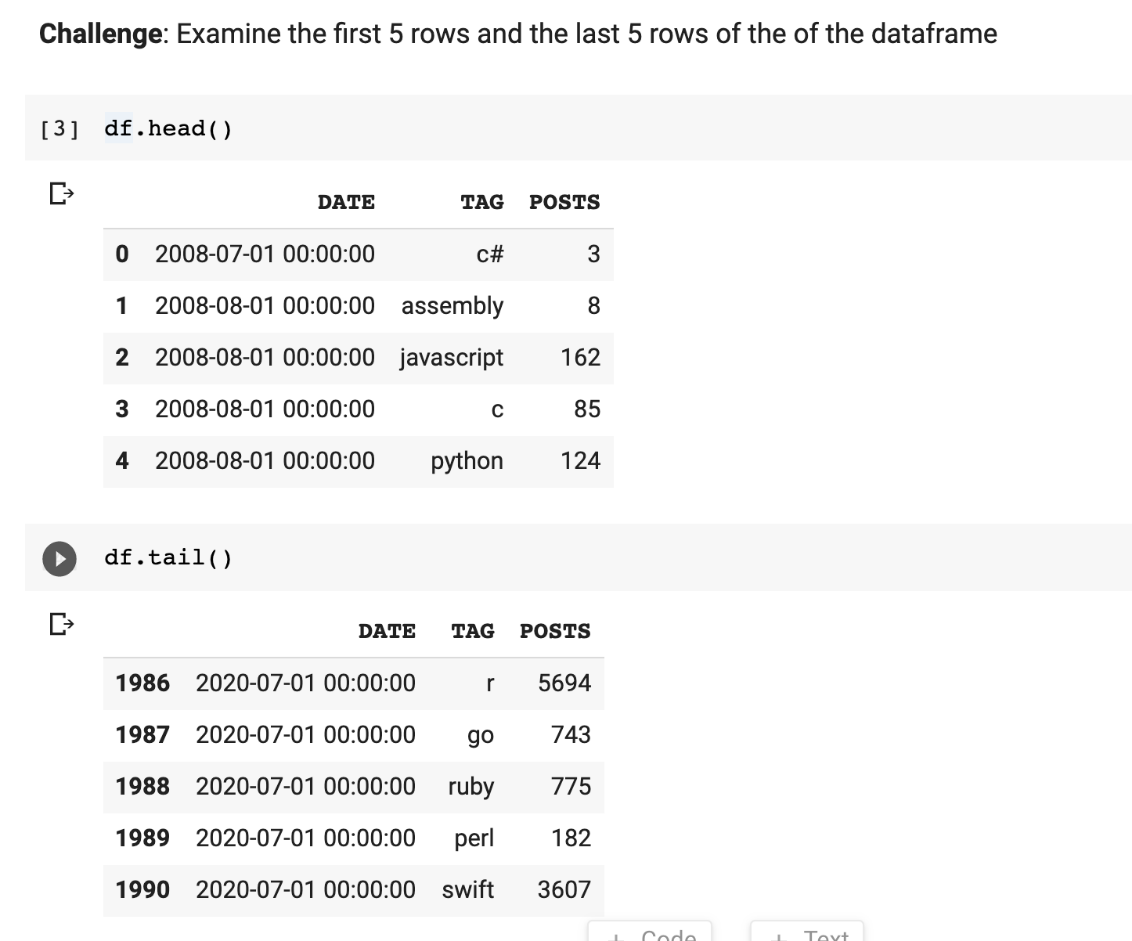
I hope the last steps were fairly straightforward. First, we import pandas and then we can call read\_csv(), where we can provide some additional arguments, like the names for our columns.

1. df = pd.read\_csv('QueryResults.csv', names=['DATE', 'TAG', 'POSTS'], header=0)

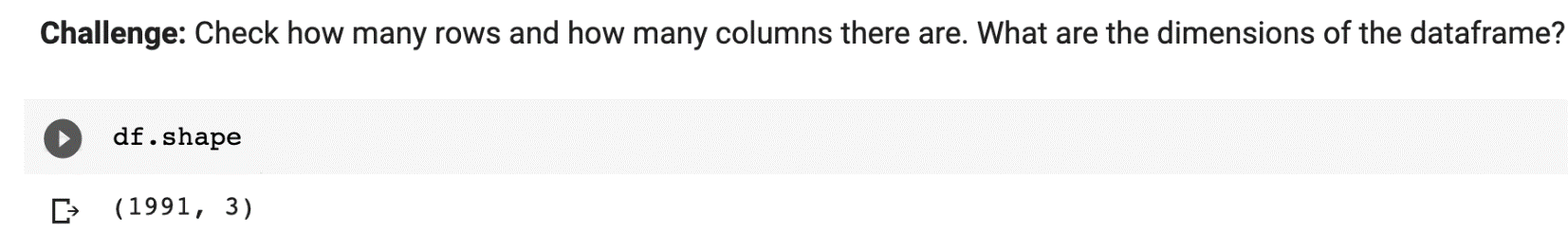
Setting the header row to 0 allows us to substitute our own column names.



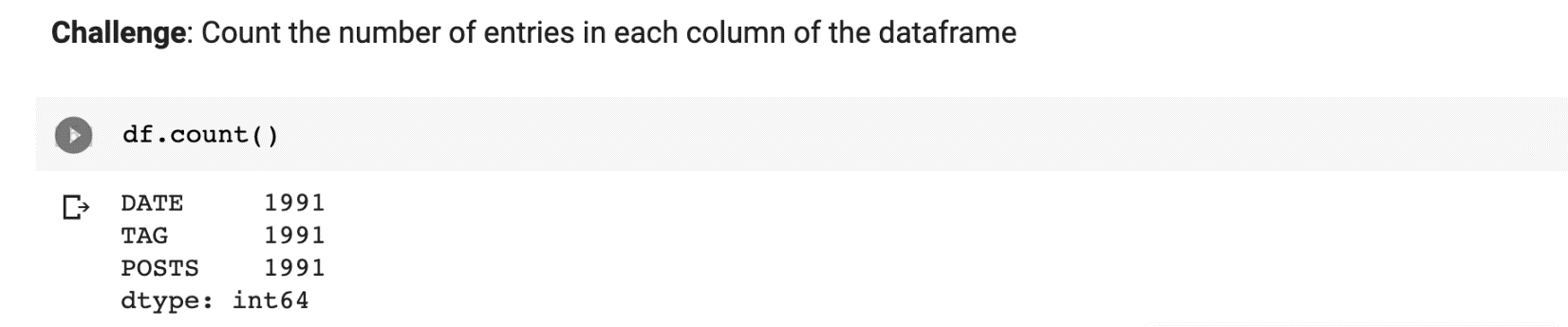
Next, we use .head() and .tail() to look at the first and last 5 rows. This allows us to verify that our column naming worked as intended.



To check the dimensions of the DataFrame, we use our old friend .shape. This tells us we have 1991 rows and 3 columns.



To count the number of entries in each column we can use .count(). Note that .count() will actually tell us the number of non-NaN values in each column.



**Next Challenge**

The TAG is the name of the programming language. So, for example in July 2008, there were 3 posts tagged with the language C#. Given that the TAG serves as our category column, can you figure out how to count the number of posts per language? Which programming language had the most high number of posts since the creation of Stack Overflow? (Hint: you may need to review one of yesterday's lessons).

Also, some languages are older like C and other languages are newer (like Swift). The dataset starts in July 2008, so some languages will not have any posts for every month. Can you count how many months of posts exist for each programming language?