Problem Statement

Given a file of web traffic data, write a solution to find the top N 3-step sequences that users follow. Unique users are identified by a session id, and actions are identified by a page id.

The file is in CSV format with 3 values on each line. It will include a header row. The file is sorted by timestamp.

Output should be formatted with each line being a unique 3-step sequence followed by the number of occurrences of that sequence, formatted as "<step1>,<step2>,<step3>:<count>" and ordered by count. Records for users that did not visit at least 3 pages should be ignored.

The solution can be submitted in the language of your choice. Code should include comments describing your algorithm.

Example

Given the file:

```
timestamp,session_id,page_id
2023-01-01T00:00:01Z,abc-123,search
2023-01-01T00:00:39Z,abc-456,detail-2
2023-01-01T00:02:03Z,abc-123,detail-2
2023-01-01T00:10:14Z,abc-123,detail-1
2023-01-01T00:11:14Z,abc-789,search
2023-01-01T00:14:01Z,abc-456,detail-1
2023-01-01T00:14:15Z,abc-456,cart
2023-01-01T00:21:01Z,abc-123,cart
```

Your code should produce:

```
detail-2,detail-1,cart:2
search,detail-2,detail-1:1
```

Bonus

Add support for supplying a sequence length, rather than specifically looking for 3 step sequences.