Telegraph Platform Engineer Exercise

When a user visit a page in Telegraph website, a record is created in a hitlog file which shows the detail of the visit. This file is available for Data Team on daily basis.

We want to find out top 3 Telegraph articles which lead users to the registration.

As an example if we have these three different journeys for three different users

user1: article#1 -> article#2 -> article#3 -> registration

user2: article#1 -> registration

user3: article#2 -> article#1 -> registration

article#1 is the most influential article with total number of 3 as it was part of all journeys.

Build a pipeline in python 2.7 which uses hitlog.csv as a source file and output a csv file which has top 3 influential topics and their total numbers.

* Hitlog is a comma delimited csv file with the following fields: page\_name, page\_url, user\_id, timestamp
* Registration url is '/register'
* Article url prefix is '/articles/'
* Output file should be comma delimited csv file with the following fields: page\_name,page\_url,total
* Try to write reusable, maintainable code
* Include unit testing as part of your code

Please do not hesitate to contact us if you have any question.