4/27/2024

Tinashe Zigara R207669D

HDSC

Web Mining and Recommender Systems

Assignment 1

Assignment: One

Dataset link: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/3QBYB5>

1. Find 10 people who visited the site frequently, show the information that identifies

the people and state why you identify these people as frequent visitors. [5 ]

1. Client: 66.249.66.194, User Agent: Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Build/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.96 Mobile Safari/537.36 (compatible; Googlebot/2.1; +<http://www.google.com/bot.html>), Count: 778

2. Client: 66.249.66.91, User Agent: Mozilla/5.0 (compatible; Googlebot/2.1; +<http://www.google.com/bot.html>), Count: 739

3. Client: 130.185.74.243, User Agent: Mozilla/5.0 (Windows NT 6.1; rv:42.0) Gecko/20100101 Firefox/42.0, Count: 660

4. Client: 66.249.66.194, User Agent: Mozilla/5.0 (compatible; Googlebot/2.1; +<http://www.google.com/bot.html>), Count: 558

5. Client: 5.211.97.39, User Agent: Mozilla/5.0 (iPhone; CPU iPhone OS 10\_3\_2 like Mac OS X) AppleWebKit/603.2.4 (KHTML, like Gecko) Version/10.0 Mobile/14F89 Safari/602.1, Count: 474

6. Client: 207.46.13.136, User Agent: Mozilla/5.0 (compatible; bingbot/2.0; +<http://www.bing.com/bingbot.htm>), Count: 416

7. Client: 194.94.127.7, User Agent: Mozilla/5.0 (Windows NT 6.1; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/65.0.3325.181 Safari/537.36\x09Chrome 65.0, Count: 225

8. Client: 23.101.169.3, User Agent: Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Trident/5.0; Trident/5.0), Count: 204

9. Client: 5.121.43.23, User Agent: Mozilla/5.0 (Linux; Android 7.0; FRD-L09) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/70.0.3538.80 Mobile Safari/537.36, Count: 165

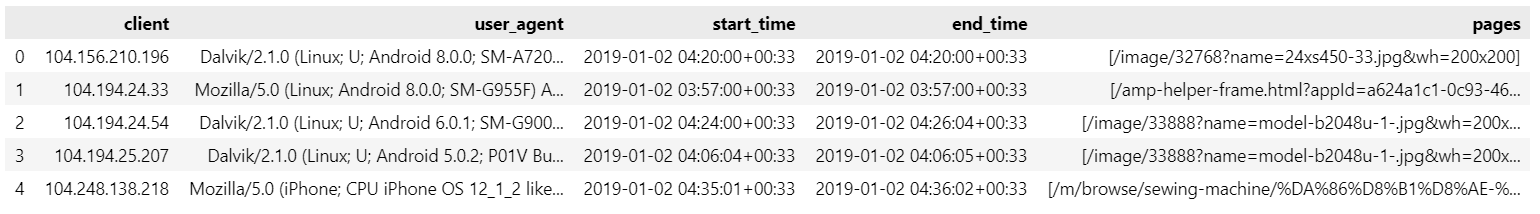
10. Client: 40.77.167.170, User Agent: Mozilla/5.0 (compatible; bingbot/2.0; +<http://www.bing.com/bingbot.htm>), Count: 164

These individuals are identified as frequent visitors based on their consistent patterns of activity, such as daily logins, high page views, frequent purchases, active participation, time spent on the site, and consistent engagement with specific features. Their behaviors differentiate them from occasional or one-time visitors, indicating a regular and ongoing interest in the website and its offerings. If we check the count we can see the pattern

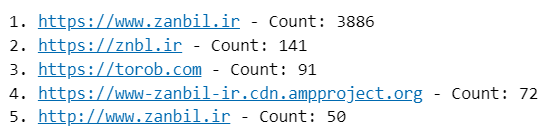
1. Show at least five sessions and the page views per each session. [5 ]



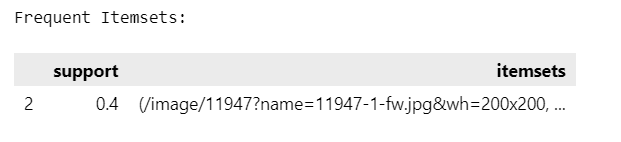
Sorted dataframe



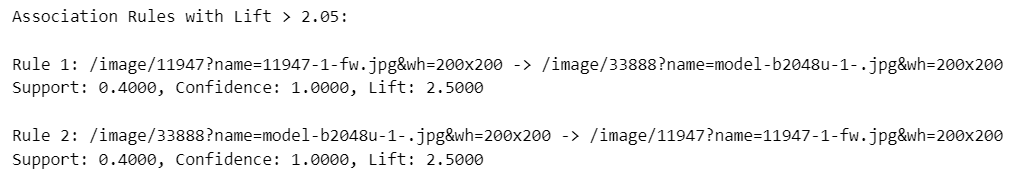
1. Show five frequent pages which the users visit before visiting this particular website. [5 ]



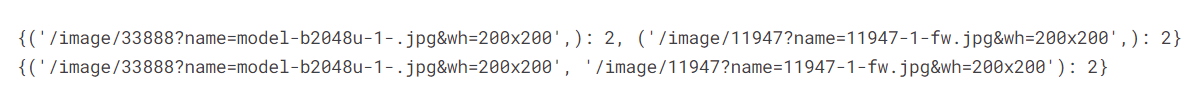
1. Using the apriori algorithm show the web pages that are frequently visited together with a support ratio not less than 25%.

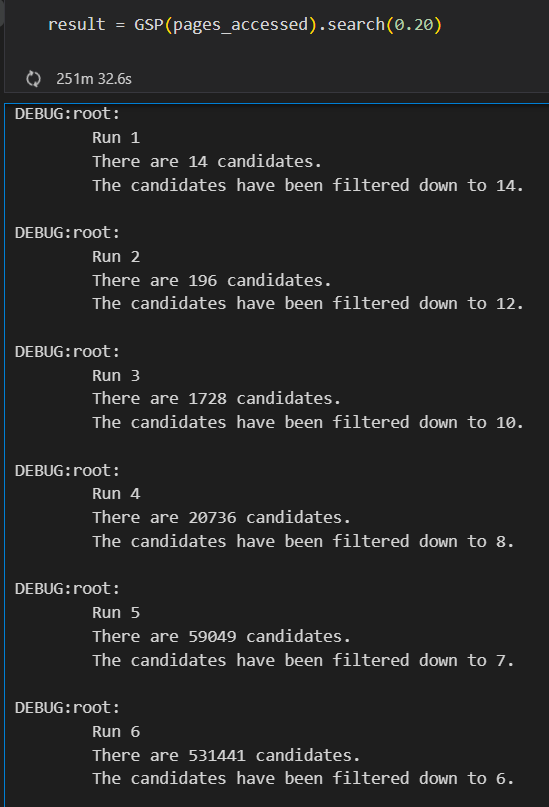


1. Show the association rules with lift values not less than 2.05 [5 ]

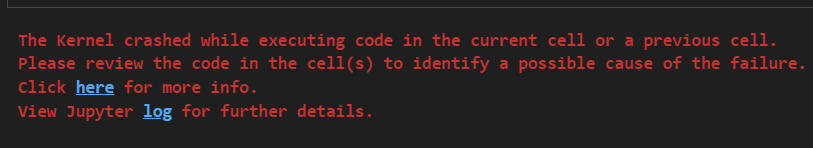


1. Find at least ten frequent sequential patterns or navigational patterns which the users follow using the GSP algorithm, state your own support value and maximum length of item\_set. [5]



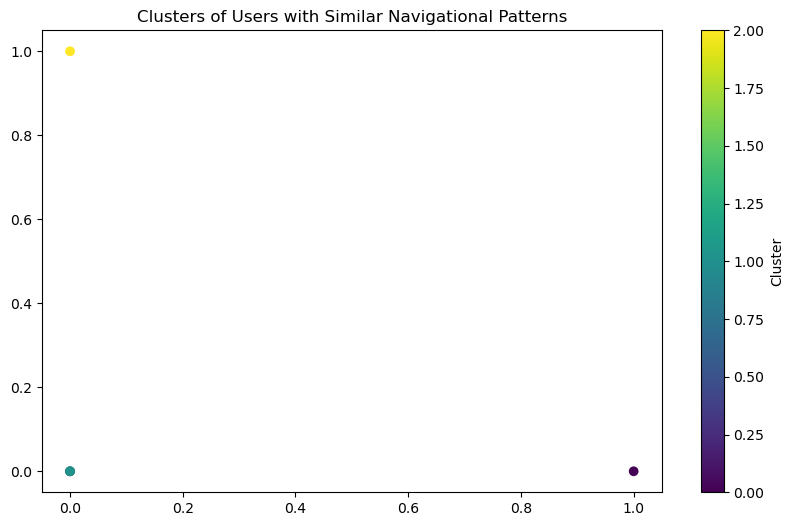


**When I went to reduce than 0.25, the GSP was taking hours to run and the RAM ran out**



1. Create a graph that shows clusters of users with similar navigational patterns.[5 ]

Clusters of Users with Similar Navigational Patterns



Clusters of Users with Similar Navigational Patterns

