Web Log Mining of Server Log Data of 'Counselling Website' Using Microsoft Excel and Implementation of Preprocessing Algorithm

Neeraj Kandpal[1], Dr.Devesh Kumar Bandil[2], Dr.Shyam Sunder Gupta[3], M. S. Shekhawat[4]

[1]Research Scholar, Suresh Gyan Vihar University, Jaipur, Rajasthan.
[2]Associate Professor, Suresh Gyan Vihar University, Jaipur, Rajasthan.
[3]Professor, Institute of technology & Management, Gwalior (M.P.)
[4]Department of Physics, Govt. Engineering College, Bikaner, Rajasthan.

Abstract:
Web log mining is essential for the generation of useful patterns from web log data. Web log mining is an indispensable tool to improve website services and servers performance. It gives valuable information to website administrators for the advancement of performance of a website. Web log data mining is possible from a wide variety of available tools. We have introduced a method for web log mining using Microsoft Excel. In this paper we have used different features of Microsoft Excel like macros, pivot tables, filtering tools, graphical tools for preprocessing of web log data. A preprocessing algorithm is proposed for data cleaning of web log data using these features. These preprocessed web log files used for pattern discovery and pattern analysis. After data preprocessing various usage patterns generated to reveal actual behaviour of visitors at the time of traversal at website 'rajeducon.com'. Various patterns are also generated for raw web log data using Microsoft Excel.
Keywords: Web Log Mining, Preprocessing, Data Cleaning, Microsoft Excel.

1.   Introduction:

Oren Etzioni coined the term web mining for analysis of data obtained from the web server of a website[1]. Web mining is the process of uncovering the information hidden in raw web log data. The web data embedded at the webpage like text, audio, video, tables, graphics are content data type. Interpage or intra page hyperlinks and webgraphs form structured data type. The usage data of visitor traversal to websites is web log data type. Based on web data type, web mining is broadly divided into following types.

1.1 Web content mining

Web content mining represents the analysis of contents of a web page of a website. Lots of structured and unstructured data like text, tables, videos, audios, images, graphs etc attached with webpage. This content is used to classify the web document and may be used for indexing purposes for personalisation.

1.2 Web structure mining

Web structure mining deals with the hyperlinks picture of a website. Hyperlinks may exist between the contents of the same webpage or may be in between the webpages of the different websites. Hyperlink structure and document structure are two possible classification of web structure mining[2]. HTML and XML organize the same webpage structure using various tags and come under document structure. Link structure is responsible for connecting different parts of the same webpage or web pages of others websites.

1.3 Web log mining

Web log mining or web usage mining is the generation of human readable usage patterns by analysis of web log data. Web logs are records of clickstreams of users at the time of their visit to a website. Web logs obtained from the server are text files containing records of the users clickstreams. These web log data are highly unprocessed data and may not be used to generate patterns. So there is a need to preprocess or clean the raw web log files. The cleaned data then used for pattern discovery and pattern analysis. Visitors browsing and access patterns are the output of web log mining[3].

Web log mining is broadly divided into three parts.

(i) Preprocessing

Preprocessing is the filtering of the less useful data from raw web log data. Data cleaning is the important step in preprocessing. Unnecessary data in raw web log files are removed in the data

cleaning step of preprocessing. The quality of the results of web log mining extensively depends on depth of preprocessing[5]. There are lots of web mining tools available for preprocessing and analysis of web log data. These tools provide ready made reports for pattern analysis. In this paper we have used Microsoft Excel-2019 for preprocessing of web log data.

(ii) Pattern discovery

In pattern discovery, human understandable reports generated using sophisticated web mining algorithms. The classification, clustering analysis, association rules and sequential pattern analysis are used for the generation of interesting patterns from web log data.

(iii) Pattern analysis

The analysis of patterns obtained at the pattern discovery phase, comes under the pattern analysis part of web log mining. The intuitive knowledge is the base for evaluation of useful patterns.

Microsoft Excel-2019 provides several features for the generation of usage patterns and analysis of these patterns.

2.   Related Work:

Gordon S. Linoff described the process of data analysis using Microsoft Excel in his book 'Data Analysis Using SQL and Excel[4]. Andrea Dominique Cortez et al extracted the features of web log data using Microsoft Excel[7]. They cleaned and shorted the dataset and placed the data in separate columns according to their specific feature. They explore the filtering and decision tree operator features for the classification of data. Neeraj kandpal et al proposed three modified preprocessing algorithms for improved data cleaning and better pattern generation results[8]. Chintan H. Makwana proposed a preprocessing algorithm using Microsoft Excel for preprocessing of web log data[9]. Researchers give analysis of web log data using different pattern analysis algorithms[10]. Researchers explained the various features of spreadsheet software Microsoft Excel like pivot tables, statistical functions, graphical features, which may be used for effective data analysis[12]. According to them Data Analysis Excel Add InToolPak is very useful for the data analysis. Babandi Usman et al used the Microsoft Excel along with Rapid Minor for analysis of students data to evaluate performance of students[11].

Researchers have used WebLog Expert, Analog, Analog Stats, Webalizer and Google Analytics web mining tools for the analysis of web log data[13, 14]. Nanhay Singh et al briefly described the web usage mining process and gave analysis of web log data of NASA web server[15]. Jabed Al Faysal et al proposed an efficient algorithm for association rule mining for web log data[16]. Researchers discussed various web mining tools with their unique features[17]. According to researchers preprocessing, discovery of pattern, information collection, and pattern analysis are the steps of internet usage mining[302]. The next page to visit suggestions proposed by Shiva Asadianfam et al using case based reasoning method[18].

3.   Web Log Data Source:

We have used the web log data of June-2018 from the server of website 'rajeducon.com'. The 'rajeducon.com' is an official website of Directorate of Secondary Education, Government of Rajasthan, India and is used for counselling of selected candidates for posting in government schools. The web log text file of June-2018 contains a total of 1,040,943 rows. Each row represents one visitor's traversal at a website. It contains IP addresses, user identification details, date-time of visit, bytes transferred, resource requested, referrer sites, status of request and hardware-software details of the visitors device. It may also contain the cookies information. As the data is directly taken from the website so it contains each and every activity at the website in the month of June-2018. The original data file is a .gz file. We extracted the web log data from the .gz file in the form of a text file. We have used Microsoft Excel program for arranging and storing web log data text files. The text file is imported to MS Excel. Space is used as a separator for the separation of various fields of web log files. Similar fields in each record are stored column wise and used for analysis.

4.   Analysis of Web Log Data using Microsoft Excel:

There are lots of web mining tools available for the analysis of web log data. These tools are based on different complex algorithms based on data mining. These tools are specially designed to meet the specific requirements of web log mining.

Microsoft Excel is a very powerful spreadsheet program, which is widely used for commercial as well as academic complex calculations. The features of Microsoft Excel like pivot tables, macro

programming and graphical features can be effectively used for the analysis of web log data. In this paper we have used the features of Microsoft Excel for analysis of web log data. We have applied Microsoft Excel for data cleaning of web log files. The cleaned web log file is used for pattern discovery and pattern identification part of web log mining.

5.   Result And Discussion

5.1 Important finding from analysis of raw web log data.

The server log data of website 'rajeducon.com' analyzed using Microsoft Excel-2019. Some important results and their analysis are given below.

(i)   The Table(1) shows hits and bandwidth statistics of the website 'rajeducon.com'. It is obvious that the total 1,040,943 hits encountered in the month of June-2018 and a total of 52.92 GB of bandwidth used. The unique IP address recorded by the server is 35,688. Unique IP addresses represent a single user but due to proxy servers sometimes it may represent more than one user. There are 26,247 hits due to automated programsi.e. bots or crawlers. They consumed a bandwidth of 0.73 GB.

Hits Summary of June-2018	
Total Hits	1,040,943
Total Visitors Hits	1,014,696
Bots/Spiders Hits	26,247
Average Hits Per Day	34,698
Average Hits Per Visitor	33,823
Unique IP Hits	35,688
Bandwidth Summary of June-2018	
Total Bandwidth	52.92 GB
Total Visitors Bandwidth	52.19 GB
Bots/Spiders Bandwidth (robot+crawler+spider)	0.73 GB
Average Bandwidth used Per Day	1.76 GB

Table(1): Hits and Bandwidth Summary for Month June-2018.

(ii)   According to Table(2) 'Success' code, 200 and 'File not found' code, 404 are the most occurring status code in the web log file of June-2018. We have examined the reason for the large amount of '404errors'.It is found that the lots of '404 errors' is due to the use of saved links of earlier counselling in visitor's browsers. These links we have already removed. Some exhausted links present at the home page are also responsible for this. We have modified the error page for '404 errors' to inform the visitors that they are using the exhausted page links and diverted them to the home page of 'rajeducon.com'.

For the research and analysis of behavior of visitors the web log data related to the status code other than 200 are not useful. So we have eliminated the data related to the records of status code entries of error 206, 301, 302, 304, 404 and 508 in the data cleaning phase of preprocessing.

| S.No.	Status Code	Count of Hits	Percentage
1.	200 (Success)	698,059	67.07
2.	404(File not found Error)	227,154	21.83
3.	304(Not modified)	63,714	06.12
4.	301(Moved permanently)	27,118	02.61
5.	206(partial content)	20,737	01.99
6.	302(Moved temporarily)	3,797	00.36
7.	508(Resource limit reached)	141	00.01

Table(2): Hits according to Status code count

(iii)   The Figure(1) shows daily hits for the month of June-2018 on the website 'rajeducon.com'. Report shows the count of daily visitors at the website. By using a macro filter, the daily hits count is accessed as given in the following figure. The days of the greater visitors traversal are the scheduled days of counselling.
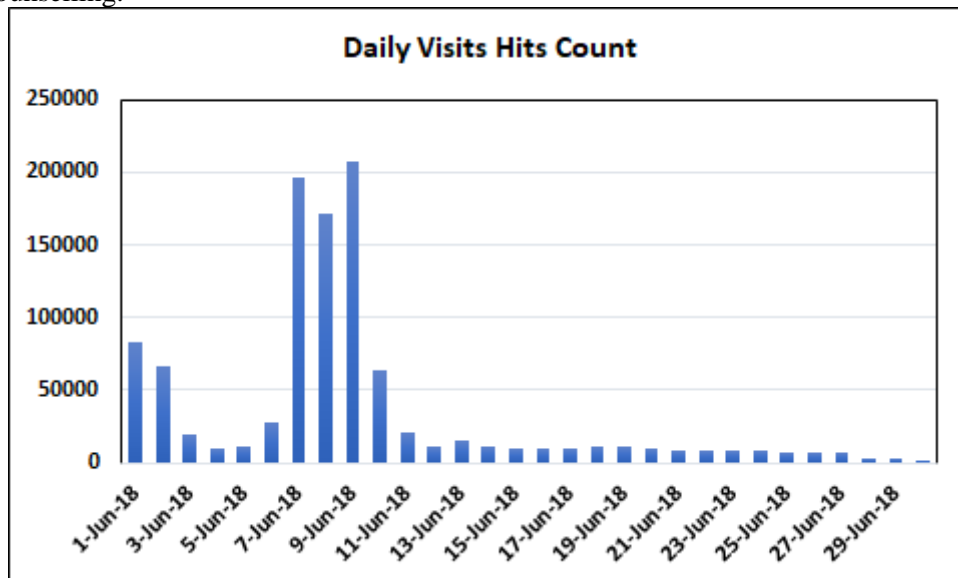
Figure(1): Date Wise visitors.

(iv) The Figure(2) shows that nearly 928,212 hits occur by device mobile, which is nearly 89% of total hits. Desktop users are 112,492 whereas 239 hits are encountered using Tablet PC. This report clearly shows that a website should contain features useful for mobile users.
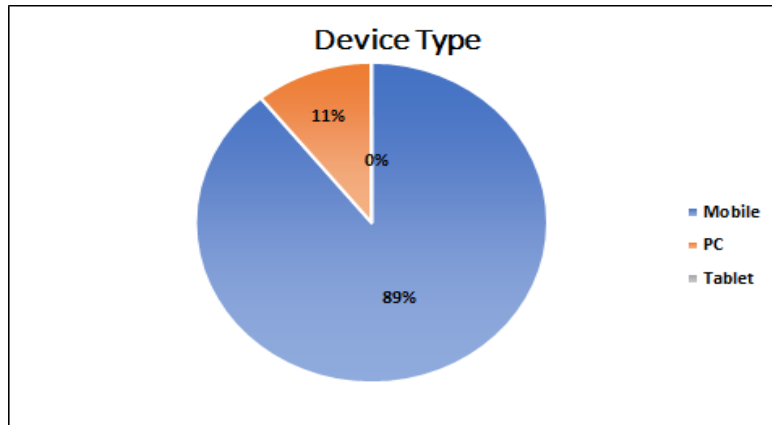
Figure(2): Device type used to access website 'rajeducon.com'

(v) Top referrer websites

Referrer websites are the websites which diverted the traffic towards our website. This report is very useful to review the security of a website. The Table(3) shows the top 10 referrer websites for the website 'rajeducon.com.'

Referrer Websites	Hits
https://www.google.co.in/	6528
android-app://com.google.android.googlequicksearchbox	3681
https://m.facebook.com/	1771
http://www.google.com/	1437
http://rajsevak.com/rajshiksha	1198
http://m.facebook.com/	903
http://www.google.co.in/	689
http://googleweblight.com/	595
android-app://com.google.android.googlequicksearchbox/https/www.google.com	432
https://www.bing.com/	201

Table(3): Top Referrer Websites

(vi) Method Requested

The 'GET' method is the most popular method as it requests data from the server. Our website is used for the real time updation of the counselling data. These data are delivered to the visitor as on request. So GET is the most popular method requested[Figure(3)].
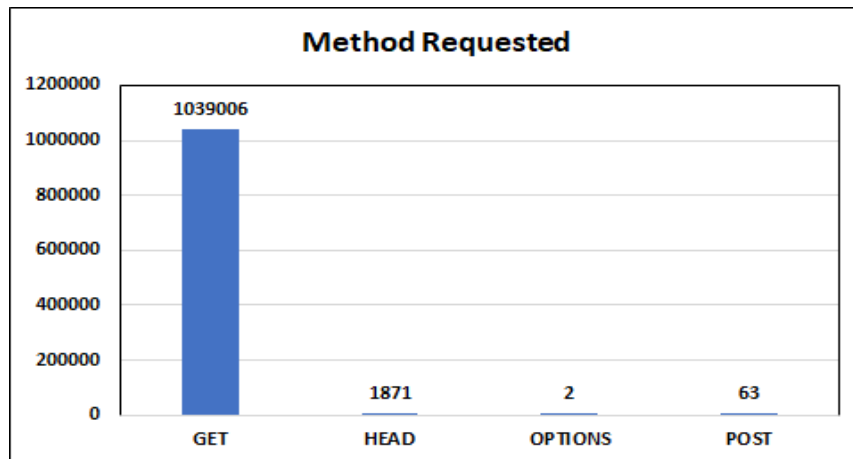
Figure(3): Method Requested count.

(vii) Most Popular 10 pages

The most popular pages are the most visited pages. These pages are definitely the source of most valuable information. These pages in the website can be used for the important announcements and may be used for the promotional purpose as advertisement.

Page URL	Total Hits
http://www.rajeducon.com/	47625
http://rajeducon.com/hos/pri/index.php	23897
http://rajeducon.com/lect/hin/	21240
http://rajeducon.com/lect/geo/	16589
http://rajeducon.com/lect/his/	15030
http://rajeducon.com/lect/pol/	10171
http://rajeducon.com/a/wait/wait.html	8948
http://rajeducon.com/hos/pri/	6697
http://rajeducon.com/pri/	4246
http://rajeducon.com/feedback.html	4083

Table(4): Most Popular 10 pages
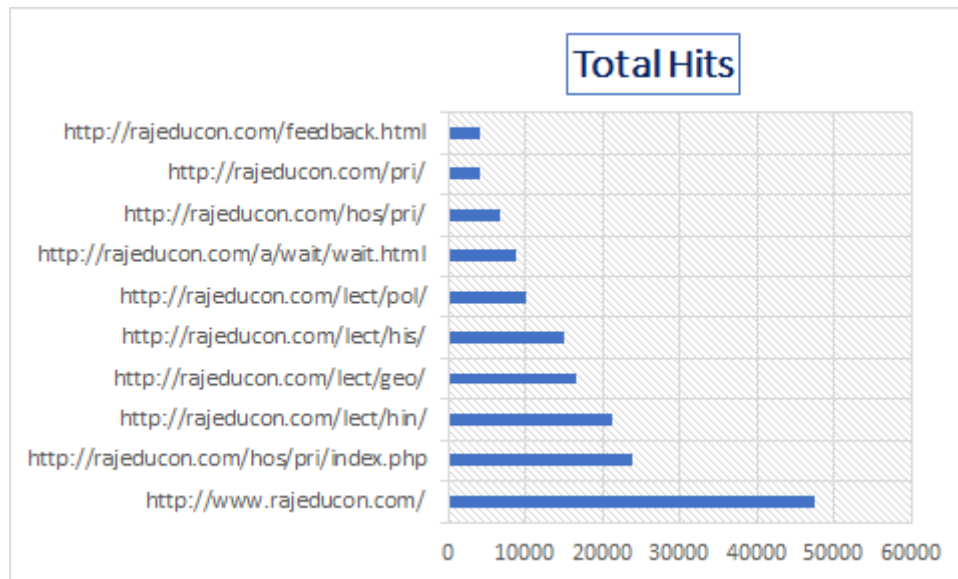
Figure(4): Most Popular 10 pages.

(viii) Most downloaded files

The following table shows the most downloaded files by visitors. The count shows the usefulness of the content on the webpage.

File Name	Download Count
/introduction.pdf	258
/contact.pdf	189
/Directortate.pdf	57
/hanumangarh/pdf/OrderEnglish.pdf	2
/hanumangarh/pdf/OrderL1_2.pdf	2

Table(5): Most downloaded files.

(xi) Most used Operating systems

Nearly 90% of visitors used the Android operating system, which is normally used by smartphones. This maintains the result of device count which suggests that mostly mobile devices are used to access the website 'rajeducon.com'. Windows 7, iPhone, Windows XP and Linux are other most used operating systems in the list of top 5 OS.
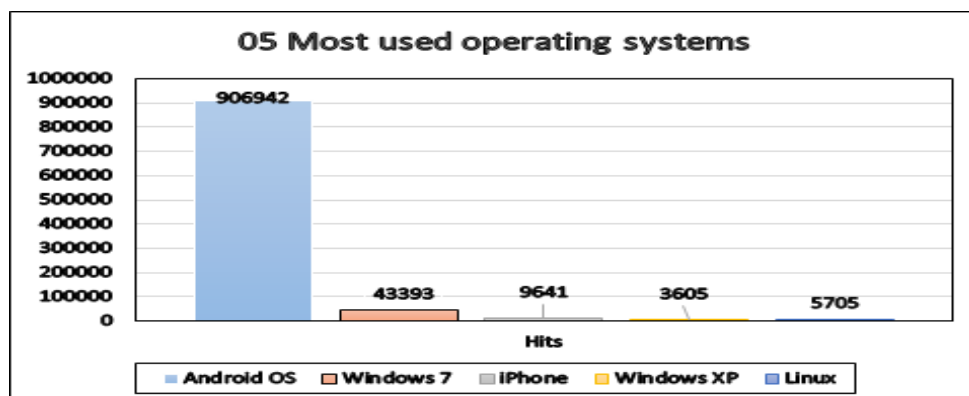
Figure(5): Most used Operating systems.

(x) Most used browsers

The Figure(6) shows that Google Chrome is the most used browser and Android Browser, Firefox, Internet Explorer and Mobile Safari are some other important ones. Report is useful to modify the code according to specific features and requirements of the browser. The display of a web page modified to enhance the experience in the most used browsers.
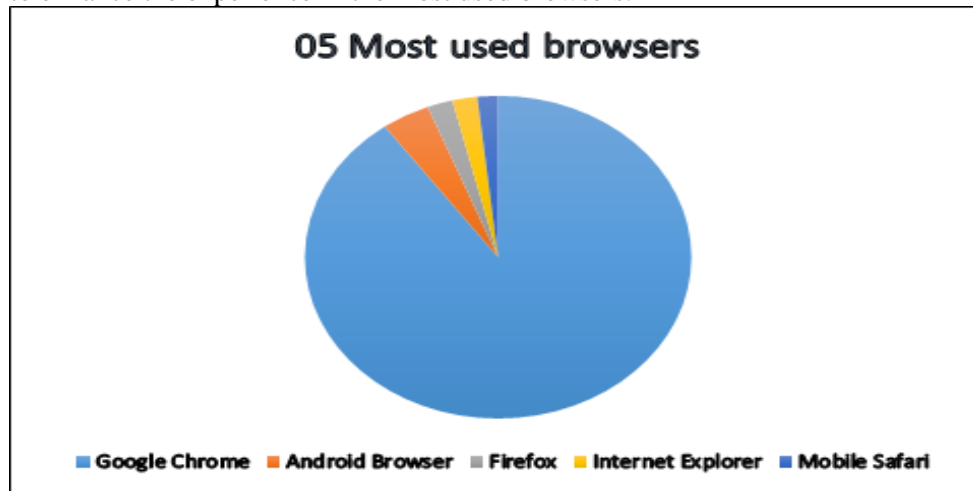
Figure(6): Most used browsers.

5.2 Preprocessing (Data cleaning) of a web log file using Microsoft Excel-2019.

The original web log file contains all visitor's hits at the website. This raw file contains records related to formatting information of a webpage, image files, automated programm visits and other files that do not represent the behaviour of visitors. So these are not important for the pattern discovery and pattern generation phase. We have removed these unnecessary data from the raw file which we have obtained from the server of website 'rajeducon.com'. The cleaned web log file is used for the pattern discovery and pattern generation phase of web log mining. The following steps represent data cleaning with Microsoft Excel-2019 using macros and filter properties of Excel.

5.2.1 Algorithm for Data Cleaning:

Input File: Raw web log file Original_db.

Output File: Cleaned web log file Cleaned_db.

Step-1: Open Original_db, which is a .gz file. Extract raw web log data from Original_db into the form of a text file.

Step-2: Import web log data text file at Microsoft Excel directly in one column.

Step-3: Arrange log data column wise taken space as a seperator. Use Microsoft Excel 'Text to Columns' option for this purpose. Give column headings to each field of web log data.

Step-4: Apply macro and other text filter options of Microsoft Excel for web log data filtering.

Step-5: Filter records having status code fields other than '200' and filter records having resource method fields other than 'GET'. Save remaining records in a separate file Temp_db.

Step-6: Open Temp_db. Filter records having resource requested field extensions as '.jpeg'. '.jpg', '.css', '.js', '.pdf', '.png', '.woff'(Web open font files), '.mp4', .ttf, .ico. Overwrite remaining records in a file Temp_db.

Step-7: Open Temp_db. Filter web robot programs by eliminating the records having words 'bot', 'crawler' or 'spider' in resource requested fields. Overwrite remaining records in a file Temp_db.

Step-8: Save records from file Temp_db in a file Cleaned_db after Step-7.

Step-9: Cleaned web log file, Cleaned_db is used to generate various statistics and graphical representation of web usage patterns. Use pivot tables for this purpose.

The favicon.ico files were also removed. It is short for favorite icon. Also Called website icon, URL icon, Bookmark icon etc. and this file contains one or more small icons associated with a website and used by the graphical web browsers. These are the small icons that appear in front of the url. The .ttf file type removed from the raw data as these does not represent the visitorsbehaviour (True type font .ttf files are common font type and it gives a fair control in the hands of developers to control the display of fonts at a website.)

The following example macro in Visual Basic is used for one of the data cleaning processes of a web log file.

```
Sub Filterdate()
' Macro1 Macro
Columns("A:P").Select
Selection.AutoFilter
ActiveSheet.Range("$A$1:$P$1040944").AutoFilter Field:=10, Criteria1:="GET"
ActiveSheet.Range("$A$1:$P$1040944").AutoFilter Field:=13, Criteria1:="200"
ActiveSheet.Range("$A$1:$P$1040944").AutoFilter Field:=4, _
            Criteria1:=">=6/1/2018", _
Operator:=xlAnd, _
            Criteria2:="<=6/30/2018"
End Sub
```

Table(6): Example Macro program to filter data.

Discription of example of Table(6): The above macro filters the raw web log file and gives the visitors records having status code '200' and request method 'GET', visited in between June 1, 2018 to June 30, 2018.

5.2.2 Results of Data Cleaning Process.

The following table shows the results of the data cleaning process.

S.No.	Description	Count
1.	Total records in original raw database of web log file.	1,040,943
2.	Records removed in the data cleaning Phase	870,900
3.	Records after data cleaning phase	170,043
4.	Percentage of records removed	83.66%

Table(7): Data cleaning of web log file.

It is obvious from Table(7) that nearly 84% unnecessary data has been removed from the web log file. This result proves that Excel's filtering properties gives excellent results which can be compared to the results of any other web mining tool.

5.3 Patterns discovery and Patterns analysis after web log data cleaning.

The raw web log file contains 1,040,943 records of visitors traversal. After preprocessing the raw file the total of 870,900 records were eliminated. The cleaned web log file contains 170,043 records. These records represent the clickstream behaviour of visitors of the website 'rajeducon.com'. The following points summarizes some patterns of visitorsbehaviour after preprocessing.

 (i) The operating system used by visitors.

As in the raw data file, the preprocessed file also has visitors mostly using the Android OS to visit the website 'rajeducon.com'. Android OS is used by nearly 92% visitors. As in Table(8), some other operating systems gave their little contribution.

S.No.	Operating System	Count
1.	Android OS	155870
2.	Windows	9343
3.	iPhone	2035
4.	Mac OS	2463
5.	iPad	56

| 6. | Others | 276 |
| | Total | 170043 |

Table(8): Operating System used: In Cleaned web log data file.

(ii) Most popular 10 webpages.

The most popular pages in preprocessed files truly represent the visitorsbehaviour as these incorporate only deliberate attempts. The /hos/pri/index.php is the most popular file among users of the website. The least visited pages should be checked for possible improvements. The report of most and least visited pages are a valuable asset for website administrators for improving the contents and structure of a website.

S.No.	Page Requested	Count
1.	/hos/pri/index.php	19858
2.	/lect/geo/	10447
3.	/lect/hin/	9904
4.	/lect/his/	7924
5.	/lect/pol/	5628
6.	/hos/pri/	4662
7.	/pri/	2412
8.	/lect/bio/	1957
9.	/a/wait/wait.html	1927
10.	/lect/che/	1259

Table(9): Most popular 10 pages: In Cleaned web log data file.

(iii) Bytes transferred by IP address.

The Table(10) gives the IP addresses to which the maximum bytes have been transferred. A lot of data transferred in a very short interval of time to a single IP address may be an attempt to break the password of sensitive information. These IP addresses should be checked for the authenticity of the user.

S.No.	IP Address	Sum of Bytes Transferred
1.	64.233.172.236	2977126
2.	157.37.184.213	2742427
3.	168.235.200.33	2191084
4.	157.37.165.14	2062728
5.	64.233.172.234	2028754
6.	171.79.232.228	1961457

7.	180.179.10.196	1899191
8.	168.235.195.162	1659426
9.	157.37.176.47	1653442
10.	106.207.190.94	1653388

Table(10): Bytes transferred by IP address: In Cleaned web log data file.

(iv) Top Referrer Websites

Top 10 referrer websites given in Table(11). These results are similar to results obtained from raw web log files. These are the websites which diverted the traffic towards our website. For security reasons, it is better to take a close look at referrer website reports, for any unusual traffic.

Referrer Websites	Numbers
https://www.google.co.in/	5860
android-app://com.google.android.googlequicksearchbox	3223
http://www.google.com/	1243
http://rajsevak.com/rajshiksha	786
http://m.facebook.com/	626
http://www.google.co.in/	497
http://googleweblight.com/	419
https://m.facebook.com/	391
android-app://com.google.android.googlequicksearchbox/https/www.google.com	344
https://www.bing.com/	168

Table(11): Top Referrer Websites: In Cleaned web log data file.

(v) In a preprocessed web log file, out of 170,043 visitors the total of 156,664 visitors used mobile devices. It is nearly 92% of total visitors. The Remaining of 8% visitors used desktop and tablet devices. In the raw data file the mobile user count is 89%. This result suggests that increase in mobile user count is mainly due to elimination of automated programs in the data cleaning phase.

(vi) Maximum bytes transferred to a webpage.

The pages with the maximum bytes transferred are the most rich and useful pages for the visitor of the website. These can be used for the announcement purpose. These pages are most popular among visitors and should be modified regularly so that their value remains intact.
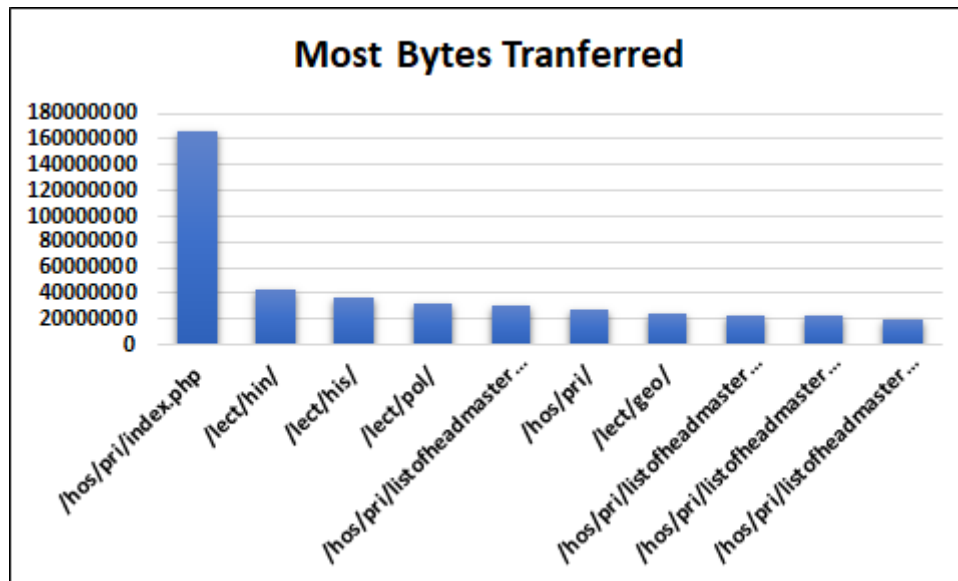
Figure(7): Most bytes transferred to a webpage: In Cleaned web log data file.

(vii) **Number of Times Web Page accessed By Visitors**

Websites may contain many pages and the number of times the page accessed by the visitor gives its popularity. The Figure(8) gives the number of times visitors accessed the webpage. These pages should be improved for optimal use of bandwidth.
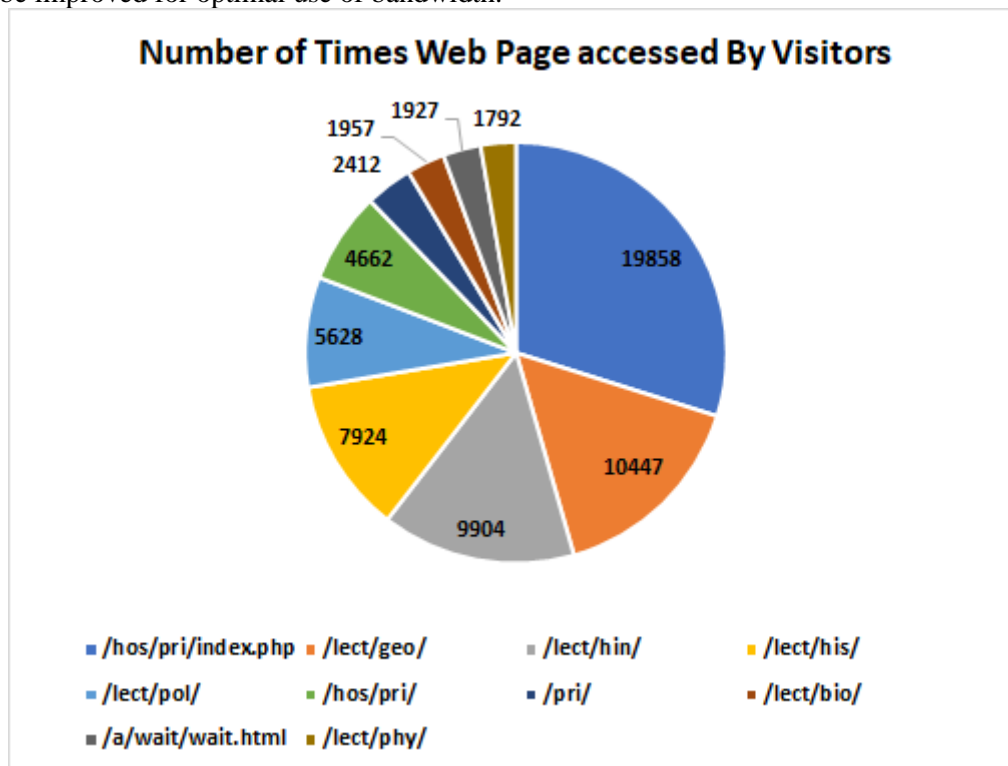
Figure(8): Number of Times Web Page accessed By Visitors: In Cleaned web log data file.

 6. Conclusion:

We have used various features of Microsoft Excel like macros, filtering properties for data cleaning. Then we have used various graphical tools and pivot tables of Microsoft Excel for pattern generation. The raw web log file used for generation of lots of statistical information. Analysis of raw data also provides useful information for improvement of websites.

All the records in the web log data file does not represent the visitorsbehaviour. The image files at homepage like .jpeg, .jpg and files used for formatting of web pages like .css, .js are downloaded automatically at the time of visit. So they are not useful for the pattern generation. Similarly visits of

bots or crawlers are also recorded by the server and are deleted from the web log file in the data cleaning phase. We have proposed an algorithm for data cleaning using properties of Microsoft Excel. In the data cleaning process a total of 870,900 records were omitted out of 1,040,943 records from raw web log data file, by using macros and other filtering properties of Microsoft Excel. After deletion of unnecessary records, the remaining records in the web log file are 170,043. Thus we have achieved nearly 84% reduction of web log data.

We have used pivot tables to record the general statistics from raw web log data files. We have found that a total 1,040,943 hits occurred in the month of June-2018 at website 'rajeducon.com' and total bandwidth 52.92 GB used by them. The total of 26,247 robot programs visited the website. It is also found that 89% visitors used mobile to access the website. The error report suggests that nearly 22% visitors don't access the resource they requested. We have removed the cause of this error by removing links of exhausted pages and also program the code for the error page for redirection. The list of most and least visited pages gives valuable insights about the popularity and usefulness of web pages. The results of the analysis are used for improving the website 'rajeducon.com'. We have achieved an increase in visitors traffic by applying results of web log data analysis to our website[19].

7. References:

[1] Etzioni O (1996), "The World-Wide Web: quagmire or gold mine?" ACM, Vol. 39.

[2] Santosh Kumar, Ravi Kumar, A Study on Different Aspects of Web Mining and Research Issues, IOP Conf. Series: Materials Science and Engineering 1022 (2021) 012018. doi:10.1088/1757-899X/1022/1/012018.

[3] Cooley R, Mobasher B, Srivatsava J. Web mining: Information and pattern discovery on the World Wide Web. 9th IEEE International Conference on Tools with Artificial Intelligence. Newport Beach, CA. 1997. P.558-67.

[4] Srivastava J, Cooley R, Deshpande M, Tan P-N. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. ACM SIGKDD Explorations Newsletter, 2000;1(2):12-23.

[5] Cooley R, Mobasher B, Srivastava J. Data preparation for mining world wide web browsing patterns. knowledge and information systems. 1999; 1(1):5-32.

[6] Data Analysis Using SQL and Excel, Gordon S. Linoff, Wiley Publishing Inc. January 2016.

[7] Andrea Dominique Cortez, Paolo Gabriel Gamab, Joseph Emmanuelle Julian, Benjamin Gabriel Tan, and Jocelynn Cu, Extracting Features from Web Logs for Web Usage Mining, Presented at the DLSU Research Congress 2017 De La Salle University, Manila, Philippines June 20 to 22, 2017.

[8] Neeraj Kandpal, Dr.DeveshkumarBandil, M. S. Shekhawat, International Journal of Advanced Science and Technology, Improved Preprocessing Algorithms to Filter Web Log Data based on Specific Requirements of Pattern Analysis. Vol. 29, No. 3, (2020), pp. 547-553.

Paper presented in Conference

[9] Chintan H. Makwana, Kirit R. Rathod, An Efficient Technique for Web Log Preprocessing using Microsoft Excel, International Journal of Computer Applications , Volume 90 – No 12, March 2014, Page-0975 – 8887.

[10] Prof. Amit Narote, Sana Afsheen Ansari, Sagar Singh Bangari, Rakshanda Khan , Jay Patel, Analysis of Web Server Log Data by Web Usage Mining, International Journal of Scientific & Engineering Research, Volume 9, Issue 3, March-2018.

[11] Babandi Usman, Rabi'uAdamu and Sani Salisu, Data Mining: Predicting of Student Performance Using Classification Technique, International Journal of Information Processing and Communication (IJIPC) Vol. 8 No. 1 [May, 2020], pp. 92-101.

[12] Boggu Anil Kumar, V Rahamathulla, A Novel Approach of Quantitative Data Analysis using Microsoft Excel, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Volume 3, Issue 4, 2018.

[13] Neeraj Kandpal, Dr. Devesh Kumar Bandil, M. S. Shekhawat, Analysis of Web Server Logs of Websites Using Different Web Mining Tools, Solid State Technology, Vol. 63 No. 2 (2020), P.1492 - 1506.

[14] Kandpal N., Bandil D.K., Shekhawat M.S. (2021) Improving Website by Analysis of Web Server Logs Using Web Mining Tools. In: Goar V., Kuri M., Kumar R., Senjyu T. (eds) Advances in Information Communication Technology and Computing. Lecture Notes in Networks and Systems, vol 135. Springer, Singapore. https://doi.org/10.1007/978-981-15-5421-6_50

[15] NanhaySingh ,Achin Jain , Ram Shringar Raw, COMPARISON ANALYSIS OF WEB USAGE MINING USING PATTERN RECOGNITION TECHNIQUES, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.4, July 2013.

[16] Jabed Al Faysal, Md. Anisur Rahman, RokebulAnam, An Efficient Approach for Mining Association Rules from Web Log Data, International Journal of Scientific and Research Publications, Volume 10, Issue 12, December 2020 Page:644.

[17] Arrshad Ali, Mohd. Faizan Farooqui, AN EXHAUSTIVE REVIEW ON WEB MINING TOOLS AND APPLICATIONS, JOURNAL OF CRITICAL REVIEWS,   VOL 7, ISSUE 19, 2020.

[18] Shiva Asadianfam, HoshangKolivand, SimaAsadianfam, A new approach for web usage mining using case based reasoning, SN Applied Sciences (2020) 2:1251.

[19] Neeraj Kandpal , Prof. H P Singh , M. S. Shekhawat, Application of Web Usage Mining for Administration and Improvement of Online Counseling Website, International Journal of Applied Engineering Research, Volume 14, Number 7 (2019) pp. 1431-1437.