

# A Study of Web Mining Application on E-Commerce using Google Analytics Tool

Y. Thushara  
Research Scholar  
Department of CSA  
SCSVMV University  
Kanchipuram, India

V. Ramesh, PhD  
Assistant Professor,  
Department of CSA  
SCSVMV University  
Kanchipuram, India

## ABSTRACT

Web mining is the application of data mining techniques to ascertain knowledge from WWW. Web mining is applied in E-Commerce to know the browsing behavior of customers. The problem in E-Commerce is we don't have any idea about our visitors till they place an order, after that you will be able to gain access to significant personal information from the sale. The main aim of this work is to define a good alternative to stay successful in E-Commerce business by understand customers better. The objective of the research work is to analysis what appear interesting the visitor to buy user product. Web Usage Mining allows the seller to observe, examine and discover patterns from composed information to form a primary statistical basis for decision making. The method of research work is defined by the precondition to properly use Web Usage Mining is to gather qualitative visitor data, that is to become able to know, whether a visitor has followed a link to get to a seller's shop, whether a convinced product has placed a seller's site first in a search engine etc. The research work is best used for E-Commerce business categories and development purpose. To analyze the visitor click stream of event generation we using Google Analytics Tool. This study discusses about the existing approach of tracing visitor data and highlights its shortcoming and rectifying them to gains a huge success in the field of E-Commerce. This approach will guide E-Commerce business into a profitable point in economy. This ability can help a vendor to track visitors and viewed goods that can visualize the way visitors use their E-Commerce superstore and categorize them so the seller can respond to their particular needs.

## Keywords

E-Commerce, Google Analytics, Web Mining, Web Usage Mining

## 1. INTRODUCTION

Web mining is the term of applying data mining techniques to automatically discover and extract useful information from the World Wide Web documents and services. Although Web mining puts down the roots deeply in data mining, it is not equivalent to data mining. The unstructured feature of Web data triggers more complexity of Web mining. Web mining research is actually a converging area from several research communities, such as Database, Information Retrieval, Artificial Intelligence, and also psychology and statistics as well. Web mining involves the analysis of web server logs of a website.

Web Usage Mining is one of the application of data mining techniques to discover usage patterns from Web data, in order to understand and better serve the needs of Web-based applications. It tries to discover the useful information

from the secondary data derived from the interactions of the users while surfing on the Web. Usage data captures the identity or origin of Web users along with their browsing behavior at a web site. It deals with studying the data generated by web surfer's sessions or behaviors. On the contrary, web usage mining mines the secondary data derived from the interactions of the users with the web. The secondary data includes the data from the proxy server logs, browser logs, web server access logs, user profiles, user sessions, user queries, registration data, bookmark data, mouse clicks and scrolls, cookies and any other data which are the results of these interactions.

Electronic commerce (e-commerce or EC) described the buying, selling, and exchanging of products, services, and information via computer networks, primarily the Internet. Most of the people view the term commerce as describing transactions conducted between business partners. The term electronic commerce seems fairly narrow, so many use the term e-business (electronic business) instead. It refers to a broad definition of EC, not just buying and selling, but also servicing organization customers, collaborating with business partners, and conducting electronic transactions within an.

Google Analytics lists the different reports user will use to in the navigation bar on the left side of user screen. The user can see an Audience, Advertising, Traffic Sources, Content, and Conversions report. Each report features a drop down menu with subsections.

Google Analytics is a free web analytics tool which provides sophisticated reports about traffic on user website. It lets user measure user advertising ROI as well as track user video and social networking sites and applications. Google analytics tool also offers Geo Targeting (targeting visitors in a specific geographic area), executive dashboard, site overlay, traffic conversion funnels, e-commerce reports and visitor segmentation analysis. The reports allow user to drill down and learn how visitors reach user site, on which page they exit, what they buy, how long they stay on each page and much more. The analysis offered by Google Analytics means that user can tailor the pages to suit user visitors, monitor the results of any advertising user do in a way that is impossible with printed or broadcast media. User can also set-up a Sales Funnel and monitor which components are not working as they should.

In this work we framed some research questions and find the solutions for that. A Research Questions a statement that identifies the phenomenon to be studied. Here we are going to raise some questions that will guide the research process.

1. Identify Number of Users and Audience Overview of the E-Commerce website.

2. Identify Number of Users and Traffic Overview of the E-Commerce website.
3. Finding the Active Users and Transactions in the E-Commerce Site.
4. Analyzed and Identify number of users all over the world are highly using E-Commerce Website.
5. Analyzed and Identify which state of users are using highly in E-Commerce Website.

## **2. REVIEW OF LITERATURE**

**M.Karthik and S.Swathi (2013)** did a study on usage of web mining technology to provide security for E-Commerce websites by framing a system consists of web structure mining, web content mining decision analysis and security analysis.

**V.ShanmugaPriya & S.Sakthivel(2013)**proposed a new method for web data extraction by removing noise that is present in the web documents. They designed three phases of data extraction. In the first phase list of web documents are selected, second phase documents are pre-processed, in the final phase results are presented to users.

**Vivek Dogne, Anurag Jain, et.al (2015)** did a survey and described about the abundance of information available on the World Wide Web (WWW), the issue of how to extract useful knowledge from the Web has gained significant attention among researchers in data mining and knowledge discovery areas. Web mining is applied to reflect the importance of Webpages and to predict the web domain visits of various users. They tried to deliver a survey of the rapidly rising area of Web usage mining, which is the order of current technology. They produce a common overview of Web usage mining

**R.Shanthi and S.P.Rajagopalan(2013)** compared different webmining algorithms to identify the context associated with the web design of an E-Commerce portal and proposed webpage collection web mining algorithm to manage time and space complexity.

**Neha Sharma & PawanMakhija (2015)** described and provides the novel approach for session identification .The growth of World Wide Web is incredible as it can be seen in present days. Web usage mining plays an important role in the personalization of Web services, adaptation of Web sites, and the improvement of Web server performance. It applies data mining techniques to discover Web access patterns from Web log data. In order to discover access patterns, Web log data should be reconstructed into sessions.

**ChhaviRana (2012)** researched about the web usage mining and focuses on techniques that could predict user behaviour while the user interacts with the Web. It tries to make sense of the data generated by the Web surfer's sessions or behaviors. There is an attempt to provide an overview of the state of the art in the research of web usage mining, while discussing the most relevant tools available in the sphere as well as the niche requirements that the current variety of tools lack. It will give an outlook on the existing tools, their specialized focus with respect to an applicative objectives and the need for a more comprehensive new entrant in this sphere in the light of the current scenario .

**Mohammad Amin Omidvar, Vahid Reza Mirabi et.al (2011)** develops a flexible methodology to analyse the effectiveness of different variables on various dependent variables which all are times series and especially shows how to use a time series regression on one of the most important and primary index (page views per visit) on Google analytic

and in conjunction it shows how to use the most suitable data to gain a more accurate result. Search engine visitors have a variety of impact on page views which cannot be described by single regression. On one hand referral visitors are well- fitted on linear regression with low impact. On the other hand, direct visitors made a huge impact on page views. The higher connection speed does not simply imply higher impact on page views and the content of web page and the territory of visitors can help connection speed to describe user behavior. Returning visitors have some similarities with direct visitors.

**Ravinder Singh & Bhumika Garg (2014)** did a research and described about web usage mining aims to automatically discover and analysed the pattern in click stream and associated data is self-controlled or generated as a result of user interaction with web browsers, on one or more websites .Web Usage mining is the combination of the two approaches that is Web Caching and Web Pre-fetching. Their proposed work presents an approach that integrates web caching and web pre-fetching approach to improve the performance of proxy server's cache. The Hybrid approach Web caching and Web pre-fetching can complement each other since the Web caching technique exploits the temporal locality and compares to Web pre-fetching technique utilizes the spatial locality of Web objects. Finally they concluded how the response time of hit taken from the user cache is less as compare to the data taken directly from the log file

## **3. RESEARCH METHODOLOGY**

### **3.1 Introduction**

Follow for immediate access to additional data and probability than ever before, because they lengthen across the world and reach peoples in almost every country. With the huge networking and telecommunications capabilities of E-Commerce, most business can be very profitable ventures if they were able to track patterns. The ability to endorse and sell goods and services online allows even the smallest firms to revenue with very little capital or overhead. Electronic commerce explains the trade, selling, advertising, and exchanging of products as well as services, and information via computer networks, primarily the Internet. Some people view the term commerce as describing dealings conducted between business partners. The term electronic commerce seems reasonably narrow, so many use the term e-business instead. It refers to a broad definition of EC, not just buying and selling, but also servicing customers, work together with business partners, and carry out electronic transactions within an organization.

### **3.2 Innovation of GAT**

Google Analytics is a dedicated web analytic tool from Google Inc. that helps administrators to track visitors click stream, event patterns and collect all kinds of useful data regarding them and when it comes to exporting raw data, things get difficult; GAT seems to keep the collected data and has not built an appropriate function to export raw data. A sequence of workarounds and proficiency are required to retrieve the essential unprocessed visitor data from Google Analytics.

### **3.3 Proposed Model**

Our model has a vision to implement in the GAT .The main goal is to track user click stream via using Google Analytics tool for tracking the e-commerce websites. We mainly proposed the idea in the tool to find the traffic information about the E-Commerce websites and finally it generated the report and displayed in the Google Analytics dashboard.

### 3.4 Architecture of the GAT

When a visitor accesses any section of user website (e.g. www.abc.com) a request is made to the Web server to display the page. The Web server processes the request and sends back the response to the browser. Then the browser loads the page. While loading it executes the JavaScript GA tracking code present in that webpage. The main logic of the JavaScript tracking code is present in the page view function.

1. When visitor accesses any section of user website, a request is made to the web server to display the page. The web server processes the request and sends back the response to the browser is shown in the Figure 4.5.
2. Browser loads the page; while loading it automatically executes the JavaScript GA tracking code present in that webpage.
3. The main logic of the javascript GA tracking code is present in the track page view() function.
4. Google Analytics now extracts user cookie values. It updates/creates the cookies as essential and sends the request to GAT server where the data is processed. After the processing of all the data then generation of Google Analytics Report.
5. This data is finally available for administrators to view in user's Dashboard. This is indeed very straight forward process and GA is one of the leading free analytics tool in the market available today.

The data that Google Analytics uses to provide all the information in user reports comes from these sources:

- The HTTP request of the user
- Browser/system information
- First-party cookies

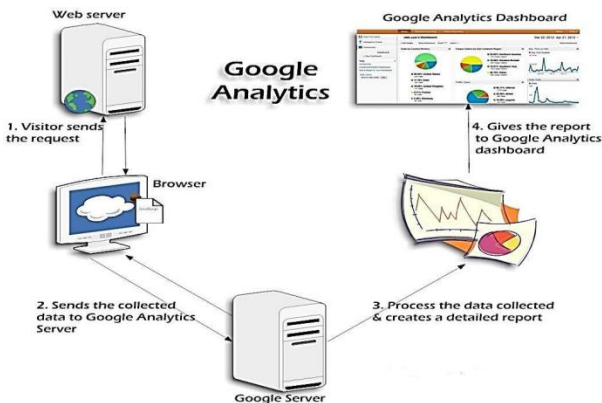


Fig 3.1 Google Analytics Architecture

The HTTP request for any web page to the web server contains details about the browser and the info about the computer which making the request, such as the Location, Device type info, hostname, the browser type, referrer, and language. In addition, the Document Object Model of most browsers provides access to more detailed browser and system information, such as Java and Flash support and screen resolution etc. Analytics uses this sort information in constructing reports like the Map location, type of browser, and referring Sites reports, system Language its calendar etc.. Analytics also sets and reads first-party cookies on user users' browsers in order to obtain user session and any ad campaign information from the page request where the JS GAT tracking

code is inserted. The GAT Code also reads the double-click cookie to get information about the Display Features.

When all this information is collected, it is sent to the Analytics servers in the form of a single-pixel GIF image request where it contains a long list of parameters attached in it. The data contained in the GIF request is the data sent to the Google servers, which then gets processed and ends up in user reports.

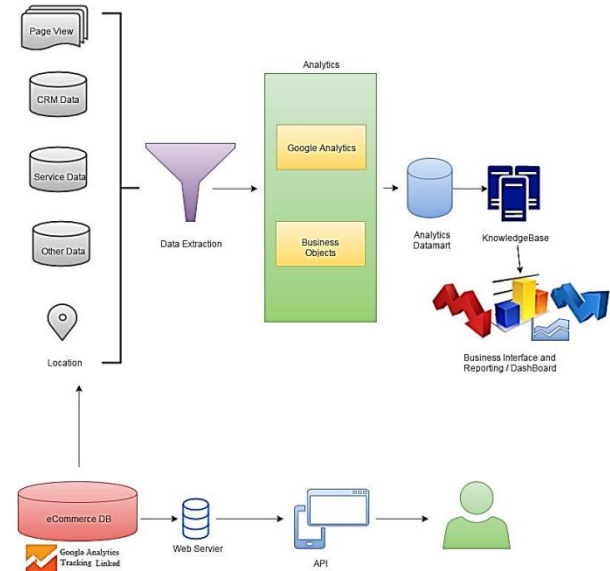


Figure 3.2 GAT with opencart integrated E-Commerce Architecture

#### Data Extraction

#### CRM Data

The data which are processed from user click stream are then extracted from the tracing code are added into a lead generation form. Google analytic serve will store those data as hidden form elements that the visitor cannot able to see. When the visitor submits the form the GA will automatically triggered and connected to the other info about the visitor entered into the form (such as name, contact information e-mail id, number, etc.).

#### Other Data

#### Data Collection & Management

Data collection and management with Google Analytics provides a single, accurate view of the customer that can be customized to user needs and shared across the organization.

#### Data Consolidation

Google delivers integrated solutions that preserve data integrity, reduce friction, and seamlessly connect disparate data sources.

#### Data Analytics & Reporting

Reports can be segmented and filtered to reflect the needs of user business. Real-time views let user know which new content is popular, how much traffic today's new promotion is driving to user site, and which tweets and blog posts draw the best results.

#### Ecommerce Reporting

Use Ecommerce reporting to understand customer behaviour, improve online merchandising, and drive revenue. Gain insight into the customer's path to purchase, visitor interest in

user products, and the performance of user shopping campaigns. User can also view Enhanced Ecommerce reports to identify segments of customers who added items to the cart, but never completed the purchase.

### Page View Data

The Page Analytics Extension allows user to see how customers interact with user web pages, including what they click and don't click. Use these insights to optimize user website layout, improve user experience, and increase conversions. When user view a web page for which user have Google Analytics access, user will see:

- Google Analytics metrics: Page views, Unique Page views, Average time on page, Bounce Rate, and %Exit
- Number of active visitors, in real time
- In-page click analysis: (where users click)

User can use the Google Analytics date comparison and segmentation tools directly in the extension. Pages user are tracking with the Google Analytics code for an account user Google account login has access to will appear like this in user Chrome browser as shown in the Figure.

## 4. RESULTS

A Research Questions (RQ) a statement that identifies the phenomenon to be studied. Here we are going to raise some questions that will guide the research process.

### RQ: 1 Identify Number of Users and Audience Overview of the E-Commerce Website.

- In this the author able to find the number of Sessions, Number of Users, Page views, and Average Session Duration and finally the total % of New Sessions.

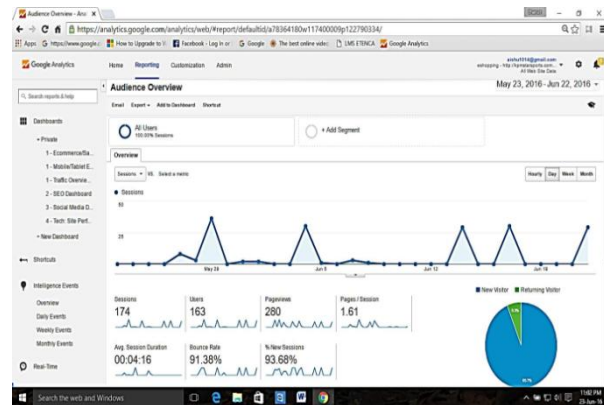


Figure 4.1 Identify Number of Users and Audience Overview of the E-Commerce Website.

Finally the author find the number of users are 163 and audience overview report of the E-Commerce website. The chart represents the new visitors and the returning visitors. The graph represents the sessions of the users day wise. It also report the pages views and the bounce rate.

### RQ: 2 Identify Number of Users and Traffic Overview of The E-Commerce Website.

- In the Reporting Page we can see the overall report of the Traffic Overview and it contains the Number of Visitors, Number of Pageviews ,New Visitors,Visits by Location, Average Time people stay on site, % of New Visitors, Overall Bounce

Rate%,Visitors Locality, Mobile Visitors, Top Pages (By View).

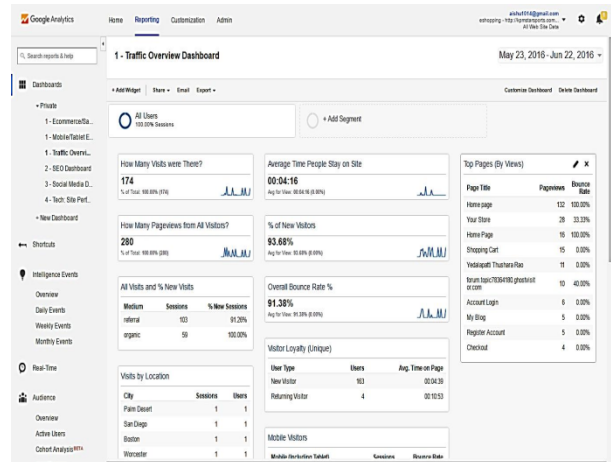


Figure 4.2 Identify Number of Users and Traffic Overview of the E-Commerce Website.

In the above figure the author find the users and traffic overview and the final result we gained. Such as visitors 174, page views 280,% of new visitors 93.68.

### RQ: 3 Finding the Active Users and Transactions in the E-Commerce Site

- In this pages it shows the active users daily and finally report for the month in the graph basics

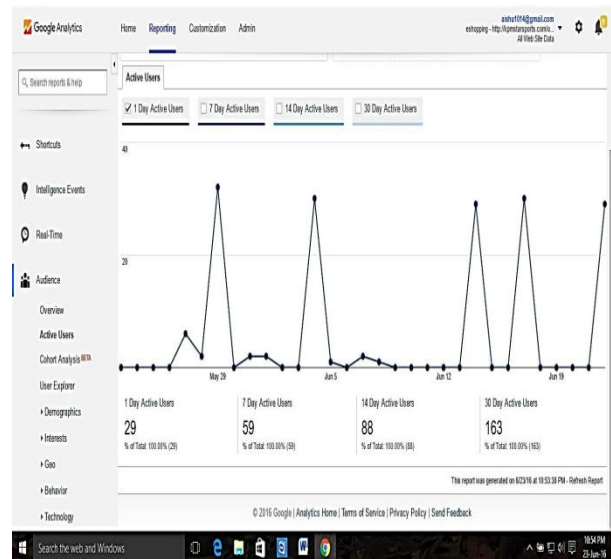


Figure 4.3 Finding the Active Users in the E-Commerce Site



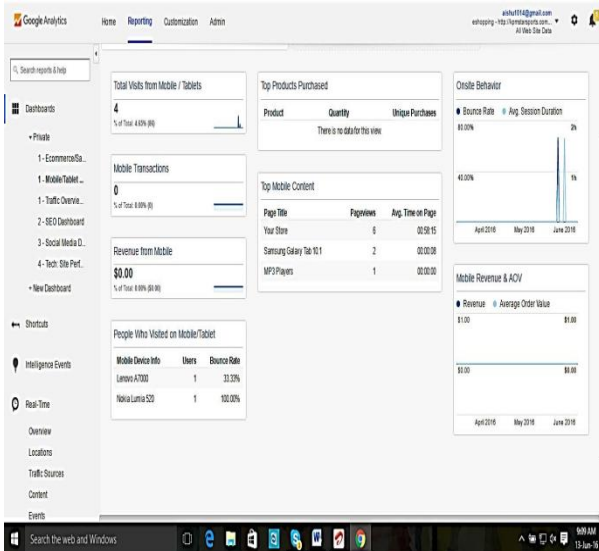


Figure 4.3.1 Finding Transactions in the E-Commerce Site

Finally the author find that active users by 1 day, 14 day and till end of the month. The transaction describes the purchase product and total visitors who visit the products.

**RQ: 4 Analysed and Identify No. of Users all over the world are highly using E-Commerce Website.**

- In this page we analysis the overall users around the world. It described the Acquisitions, Behaviour, Conversions of the users for different country users.

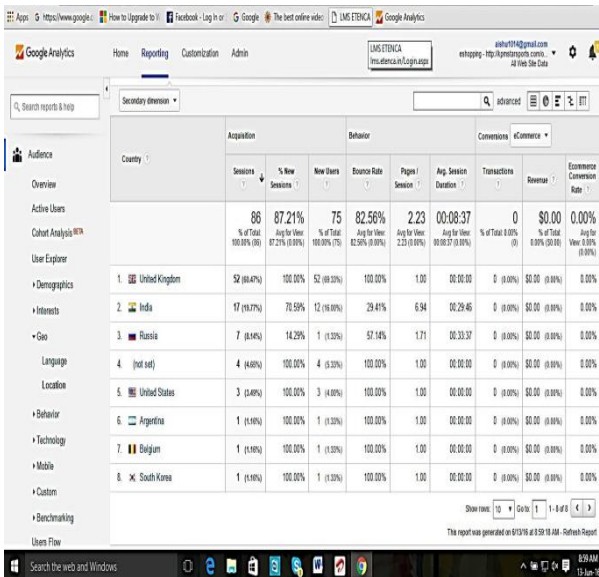


Figure 4.4 Analysed and Identify number of users all over the world are highly using E-Commerce Website.

Finally, the author identify that UK users are the highly using the E-Commerce Website with 52% and India is at second position with 17% percentage of users are using the E-Commerce websites.

**RQ: 5 Analysed And Identify which state of users are**

- using highly in E-Commerce Website.
- In this the author analysed and find which state has highest number of users are using this sites. So, we compare two different states. Such as, Tamil Nadu and Karnataka. Finally the author find the result as

Tamil Nadu has lot of users are using the E-Commerce site as shown.

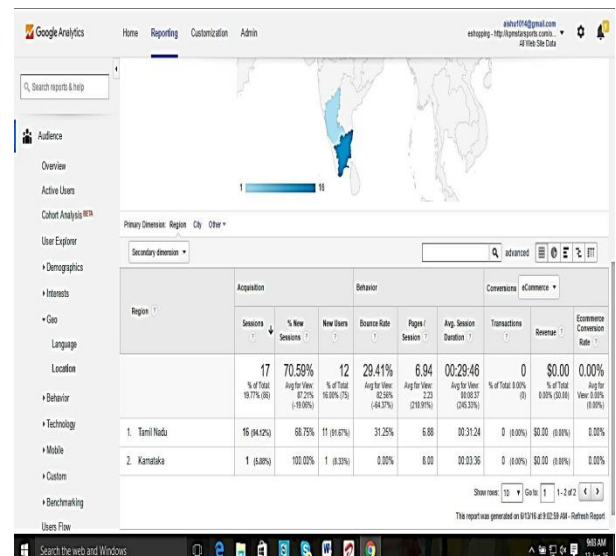


Figure 4.5 Analysed And Identify which state of users are using highly in E-Commerce Website.

**5. CONCLUSION**

In this research work the authors analyzed and also identified whether Google Analytics can be considered as a state-of-the-art alternative to collect data for web usage mining. The principle is to cluster customer segments by using automatic discovery and analysis of patterns in E-Commerce website which input data comes from web log of various e-commerce websites.

The authors analyzed one of the leading E-Commerce Software (OPENCART) to track the information of the web users. The authors implement into the Google Analytics Tool for the report of traffic information of the users. So far we success in traffic information and analyzed the overall users in different countries and authors also identify the traffic of the users in India at Tamil Nadu and as well as the Karnataka state. Hence, determine the relationship between Web data mining and E-Commerce and also to apply Web mining technology in E-Commerce.

**5.1 Suggestions And Recommendations**

This paper has highlighted a number of topics on which have been beneficial for the further research. The authors suggested to the Google Analytic Tool makers has to implement that how the sales information can able track and generate the report in the Google Analytics without any help of any E-Commerce software. Several areas where the information is lacking are highlighted in the literature review. In particular, there is lack of study of about web analytics. But there are a number of additional areas for further research that have been highlighted by the studies undertaken for this thesis. There is a problem while implementing the tracking code in the GAT.

Although this research was carefully analyzed and prepared, the author still aware of its limitations and shortcomings. Following are some limitations that we observed during our research.

- To implement the tracking code in the tool we have use one the E-Commerce software to insert the code then only it can able to generate the report. In GAT it is difficult to implement the sales details.

- The authors can't able to track the on moving E-Commerce Website because we unable to access the servers of any E-Commerce sites. Such as Flipkart, Amazon and Snapdeal websites.

## 5.2 Scope For Future Study

For the future research work, the researchers can able to calculate the overall sales information and total income of the E-Commerce websites in Google Analytics. It can also calculate the discount rates of the products and coupons also. Moreover the authors also find the discount sales for the special product in season wise. There is lot of scope to work on the Google Analytics tool. It is very useful for the future research workers.

## 6. REFERENCES

- [1] Vivek Dogne, Anurag Jain, Susheel Jain (2015), "Evolving Trends and its Application in Web Usage Mining: A Survey", *International Journal of Soft Computing and Engineering*, Vol-4 No-6, PP:98-102.
- [2] Shanmuga Priya.V, Sakhivel.S (2013), "An Implementation Of Web Personalization Using Web Mining Techniques", *International Journal of Computer Science and Mobile Computing*, Vol. 2, No. 6, PP:145 – 150.
- [3] Suhasini Parvatikar ,Bharti Joshi(2014), "Analysis of User Behaviour through Web Usage Mining", *International Conference on Advances in Science and Technology*, PP:27-32.
- [4] Ravinder Singh, BhumikaGarg (2014), "Hybrid Approach for Performance of Web Page Response through Web Usage Mining", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 4, No. 7, PP:1069-1073.
- [5] Shanthi.R ,Dr.Rajagopalan S.P (2013), "An Efficient Web Mining Algorithm To Mine Web Log Information", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 1, No 7, PP:1491-1500.
- [6] Munilatha .R, K.Venkataramana (2014), "A Study On Nos And Techniques Of Web Mining", *International Journal of Computer Science and Mobile Computing*, Vol. 3, No. 5, May 2014, PP.331 – 341.
- [7] PankajM.Meshram, GauriA.Chaudhary(2014), "Mining of Web Logs Using Preprocessing and Clustering", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 2, No 12, PP:7543-7546.
- [8] Menaka .P,Prathimadevi .A(2015), "A Survey on Web Mining and Its Techniques", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol 5, No 9,PP:276-280.
- [9] NehaSharma &PawanMakhija(2015), "Web usage Mining: A Novel Approach for Web user Session Construction", *Global Journal of Computer Science and Technology: E Network, Web & Security* Vol.15 , No 3,PP:15-20.
- [10] Mohammad Amin Omidvar, Vahid Reza Mirabi And NarjesShokry(2011), "Analyzing The Impact Of Visitors On Page Views With Google Analytics", *International Journal of Web & Semantic Technology*, Vol.2, No.1,PP:14-32.
- [11] Karthik.M, Swathi.S(2013), "Secure web mining framework for e- commerce websites", *International Journal of Computer Trends and Technology (IJCTT) - vol4 No.5*, PP:1042-1046.
- [12] GopalPandey, Swati Patel, VidhuSinghal, AkshayKansara(2013), "A Process Oriented Perception of Personalization Techniques in Web Mining", *International Journal of Science and Modern Engineering* Vol-1, No-2, PP:26-30
- [13] ChhaviRana (2012), "A Study of Web Usage Mining Research Tools", *Int. J. Advanced Networking and Applications*, Vol.3, No.6, PP:1422-1429.
- [14] BhavnaBeri, Parminder(2013), "Web Analytics: Increasing Website's Usability and Conversion Rate ", *International Journal of Computer Applications*, Vol.72– No.6, PP:35-39.
- [15] AnkitaKusmakar, Sadhna Mishra(2013), " Web Usage Mining: A Survey on Pattern Extraction from Web Logs", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol 3, No 9, PP:834-839.
- [16] Ananthi.J(2014), "A Survey Web Content Mining Methods and Applications for Information Extraction from Online Shopping Sites", *International Journal of Computer Science and Information Technologies*, Vol. 5 (3) ,PP: 4091-4094.
- [17] AmitPratap Singh, Dr. Jain . R.C (2014), " A Survey on Different Phases of Web Usage Mining for Anomaly User Behaviour Investigation", *International Journal of Emerging Trends & Technology in Computer Science*, Vol.3, No 3,PP:70-75.