ISSN (Print): 0972-2467
ISSN (Online): 0976-2477
SRELS Journal of Information Management, Vol 55(5), October 2018, p. 247-253 DOI: 10.17821/srels/ 2018/v55i5/132073

Correlation between Daily Page Views and View Duration: An Exploration of National Library Websites

Sukanya Mandal[1] and Arijit Das[2*]

[1]Dr. B. R. Ambedkar Institute of Education, Baruipur - 743610, Kolkata, West Bengal, India; sukanya24.mandal@gmail.com

[2]Department of Library and Information Science, Jadavpur University, Jadavpur - 700032, Kolkata, West Bengal, India; arijitdas399@gmail.com

Abstract

Generally page view and view duration help to estimate the popularity of a site they are also determining factors for estimating the extent of use of that site. This paper seeks to establish a "correlation" between the two variables, i.e. daily page views per visitor and daily time on site (view duration) of twenty five national library websites around the world to see the degree of association. A scatter diagram indicating the nature of association between these two variables is drawn. The Pearson's Correlation Coefficient (r) is 0.744, which indicates a strong positive relation between these two variables. The pattern of scatter diagram also represents a positive correlation, which means the association between the variables is direct, indicating thereby that if the value of average page view is high then the value of daily time on site also will be high, and low values are associated with low values.

Keywords: Correlation Coefficient, National Libraries, Page View, Scatter Diagram, Website Analysis, View Duration

1. Introduction

With rapid advancements in information and communication, library and information professionals are facing new challenges continuously. The need for digital information is increasing. As a result libraries have started providing facilities and services to the users through websites. National libraries are no exception to this. "A National Library is a library specifically established by the government of a country to serve as the preeminent repository of information for that country. Unlike public libraries, these rarely allow citizens to borrow books. Often, they include numerous rare, valuable, or significant works (Verma & Brahma, 2017)". This study tries to examine the association, if any, between page view and daily time on site. This association notifies page view as a strong predictor for visit duration of a website.

"The World Wide Web, an information space, is a highly complex conglomerate of all types of information carriers produced by all kinds of people and searched by all kinds of users (Bjorneborn, 2001)".

"Web Analytics Association Standards (2006) committee defined the three most important metrics as Unique Visitors, Visits/Sessions, and Page Views; and, also categorized search engine marketing metrics through counts (visits…), ratios (page views per visits….), and Key Performance Indicators (KPIs) (Association, 2006)". "Pageview statistics are useful to describe and predict the behaviour of clients on Internet sites. Typical questions that are related to visitor behaviour are the frequency and length of visits during a certain time period, the entrance and exit locations of visitors, the percentage of visitors who reach key pages (such as a sign-up page, cash register, etc.), the paths

*Author for correspondence

they take, the traffic trend, the prediction of traffic spikes, the accommodation of server space for increased traffic, the adjustment for browser technology, the evaluation of behaviour variations among subsets of customers and the change during sales, etc., etc. However, these questions are difficult to answer because of the existence of several boundary conditions: human behaviour is very stochastic and data can be incomplete or noisy caused by the existence of proxy servers, fire walls, caching, browser settings, and cookies (Gelder & et.al., 2011)".

2. Objectives

- The purpose of this study is to identify relationships between two different sets of data, related to the use of website, i.e. page views per visitor and time spent on the site taken as variables, to measure the association and to understand the nature of relationship among them, and
- To investigate effectiveness and usability of the national library websites. Usability of the website is assumed to be strongly correlated with page views and daily time on site.

3. Limitations of Study

- "It is important to remember that the results can be misleading because in some cases the visitor may have been interacting with your pages and site content or they could have left the browser window open and were not actually viewing your page (support.google, 2018)", and
- The results taken from website evaluation tool may vary day to day.

4. Methodology

4.1 Selection of Variables

Two variables are selected to estimate the correlation which indicates the degree and nature of association among them. These two are average page view per visitor and daily time spent on site.

- Pageview: "A pageview is defined as a view of a page on the site that is being tracked by a tracking tool. If a user clicks reload after reaching the page, this is counted as an additional pageview. If a user navigates to a different page and then returns to the original page, a second pageview is recorded as well (support.google, 2018)". "Page view, as its name shows, is the number of times a particular page (URL) is viewed by a user. Alexa.com makes it clear though that, if a particular user visits the same URL multiple times on the same day, all those visits will be counted as one (Avangate, 1999)", and
- Daily Time on site: "Average time on site is a type of visitor report that provides data on the amount of time (in minutes or seconds) visitors have spent on your website. When viewing the time on site report in your Web analytics (webopedia, 2018)."

4.2 Selection of National libraries

Twenty five national libraries were selected around the world. At the time of selection of these libraries preference has been given to cover all the zones of the world. Five continents are covered, viz., 9 countries from Asia-Oceania, 7 from Europe, 3 from North America, 2 from South America and 4 from Africa.

4.3 Selection of Website Evaluation Tool

The page views and visit duration are generally tracked by website evaluation tools such as, Google Analytics, Alexa Siteinfo, etc. In this study Alexa was taken for collection of data on August 30, 2018.

Alexa is owned by Amazon since 1999. Its toolbar collects data on Internet browsing behavior and transmits them to the Alexa website, where they are stored and analyzed. This is the basis for the company's web traffic reporting, including its well-known Alexa Rank. According to its website, Alexa provides web traffic data, global rankings, and other information on 30 million websites (wikipedia, 2018)". It also gives global traffic rank, which measures the of a website *vis-à-vis* against other websites over the preceding 3 months.

4.4 Data collection and Statistical Methods

Alexa Siteinfo has been used to collect data on page views per visitor and time spent on site of selected national library websites. The degree of correlation is measured by correlation coefficient, which is a ratio indicating the extent to which the changes in one variable are accompanied by the changes in concerned variable. Here, Pearson's Correlation Coefficient method has been applied to measure the degree of association between two variables.

5. Data analysis and Interpretation

The first step in correlation analysis is to visualize the relationship between the independent variable and the dependable variable. Daily page views per visitor are taken as the independent variable and daily time on site is the dependent variable.

Table 1 represents daily page views per visitor and time spent on site of twenty five national library websites around the world. Nine national libraries have been chosen from Asia-Oceania, National Library of China and National Library of Bangladesh found high rates in terms of Avg. page views per visitor and time spent on site. The highest values for page view and view duration are found for the National Library of China

Table 1. Daily Page views per visitor (avg. no. of pages) and daily time on site (in mm:ss & only mm) of the enlisted libraries

Continent	Country	Name of National Library	Daily Pageviews per Visitor (avg. no. of pages)	Daily Time on Site (mm:ss)	Daily Time on Site (mm)
Asia- Oceania	*India*	National Library of India	3.50	2:44	2.73
	China	National Library of China	6.60	4:57	4.95
	Pakistan	National Library of Pakistan	3.30	3:29	3.48
	Singapore	National Library Board	4.00	3:26	3.43
	Japan	National Diet Library	3.50	2:36	2.60
	Bangladesh	National Library of Bangladesh	4.00	4:40	4.66
	Sri Lanka	National Library of Sri Lanka	2.10	1:56	1.93
	Australia	National Library of Australia	1.87	2:08	2.13
	New Zealand	National Library of New Zealand	3.40	3:14	3.23
Europe	*United Kingdom*	The British Library	3.50	3:12	3.20
	France	Bibliothèque Nationale de France	4.39	4:42	4.70
	Ireland	National Library of Ireland	3.60	3:46	3.76
	Germany	Deutsche National Bibliothek	1.90	2:26	2.43
	Finland	The National Library of Finland	5.20	8:59	8.98
	Russia	The National Library of Russia	4.60	4:30	4.50
	Netherlands	The National Library of Netherlands	2.60	2:40	2.66
North America	*United States*	Library of Congress	3.21	4:21	4.35
	Canada	Library and Archives Canada	4.40	4:02	4.03
	Mexico	Biblioteca Nacional de Mexico	3.69	4:37	4.61
South America	*Argentina*	Biblioteca Nacional Mariano Moreno	5.70	3:54	3.90
	Brazil	Biblioteca Nacional	1.90	2:29	2.48
Africa	*South Africa*	National Library of South Africa	2.60	3:35	3.58
	Egypt	Egyptian National Library and Archives	1.90	2:03	2.05
	Kenya	National Library Service of Kenya	3.20	5:05	5.08
	Tunisia	National Library of Tunisia	8.90	7:22	7.36

Accessed on August 30, 2018, Source: https://www.alexa.com/siteinfo

(6.60 and 4.95). National library websites of Finland, Russia (Europe), Argentina (South America) and Tunisia (Africa) found high rates of page views and view duration. In North America, the highest page view is in Library and Archives Canada and view duration is in Biblioteca Nacional de Mexico.

In Asia-Oceania the lowest number of page views is the National Library of Australia (1.87) and view duration is the National Library of Sri Lanka (1.93). Deutsche National Bibliothek (Europe) found low rates of page views and visit duration. In North America Library of Congress has a low page view (3.21) and Library and Archives Canada has a low view duration (4.03). In South America, Biblioteca Nacional (Brazil) has a low page view (1.90) view duration (2.48). Egyptian National Library and Archives has a low page view (1.90) and view duration (2.05).

Table 2 shows that the National Library of Tunisia has the highest Avg. page view, i.e. 8.90 and The National Library of Finland has most time spent on site, i.e. 8.98.

5.1 Correlation Coefficient

In statistics Correlation coefficients are used to measure how strong a relationship is between two variables. There are different correlation coefficients: Pearson's correlation (also called Pearson's R the product-moment correlation) is a correlation coefficient commonly used in linear regression (Figure 1).

"Correlation coefficient formulas are used to find how strong the relationship is between data. The formulas return a value between -1 and 1, where:

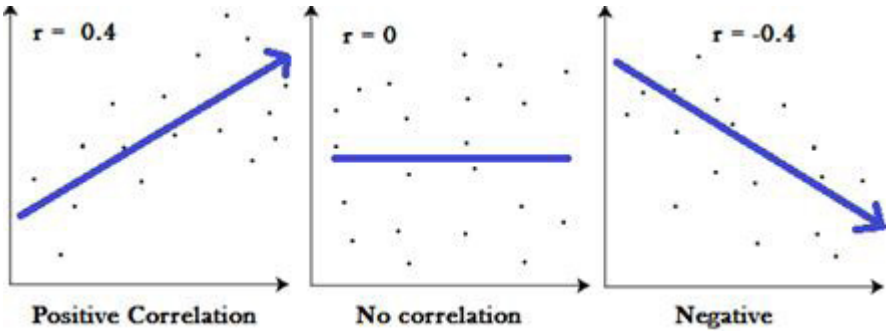
- 1 indicates a strong positive relationship,
- –1 indicates a strong negative relationship, and
- A result of zero indicates no relationship at all (Avangate Home Page, 1999)".

5.2 Scatter Diagram

Statistical data on two variables are geometrically represented by a point on the graph paper – the values of one variable being shown along the *X*-axis (independent variable) and those of the other variable along *Y*-axis (dependent variable) (Figure 2).

This scatter diagram represents the nature of association between the two variables. The pattern of points on the scatter diagram shows a linear path suggesting that the relationship between the two variables is strongly positive (Figure 2).

In other words, the association between the variables is direct; indicating thereby that if the number of page views is high then the value of daily time on site also will be high and if page view is low then the value of daily time on site also will be low.



Source: http://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/

Figure 1. Pattern of correlation.

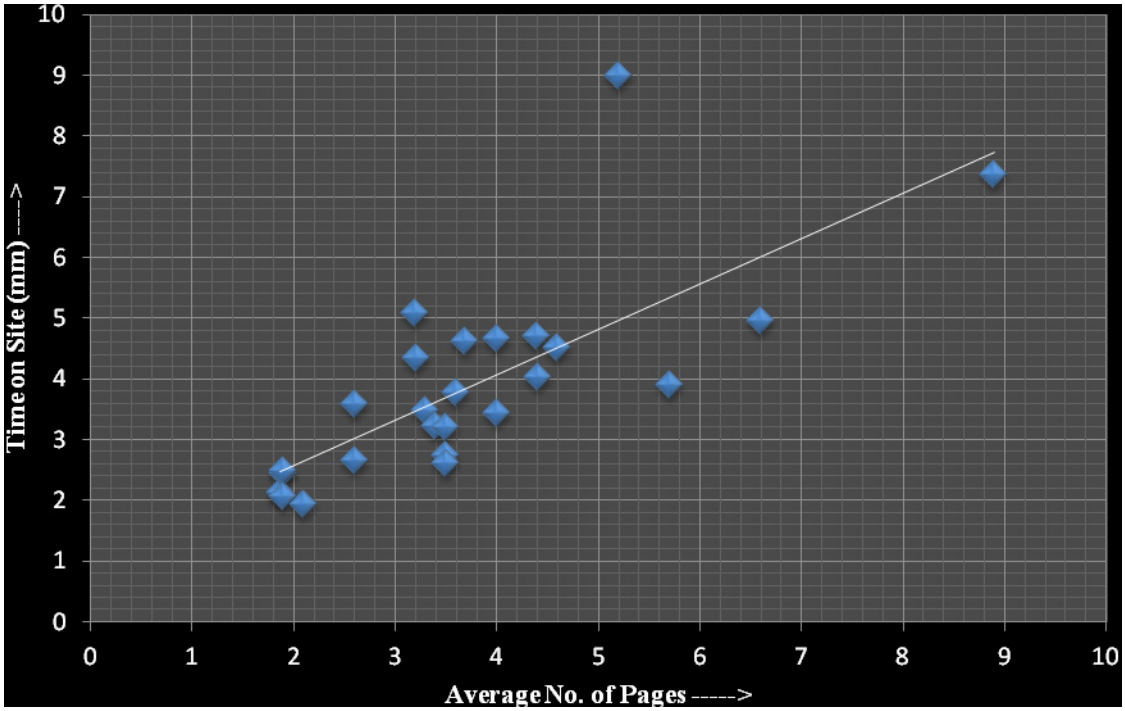


Figure 2. Scatter diagram showing correlation b/w daily page views and view duration of 25 National Library Websites.

5.3 Calculation of Correlation Coefficient

Table 2. The daily avg. page views per visitor (X variable) and daily time on site in minutes (Y variable) of twenty five libraries

Daily Avg. Pageviews per Visitor (X)	Daily Time on Site in minutes (Y)	XY	X^2	Y^2
3.50	2.73	9.55	12.25	7.45
6.60	4.95	32.67	43.56	24.50
3.30	3.48	11.48	10.89	12.11
4.00	3.43	13.72	16.00	11.76
3.50	2.60	9.1	12.25	6.76
4.00	4.66	18.64	16.00	21.71
2.10	1.93	4.05	4.41	3.72
1.87	2.13	3.98	3.49	4.54
3.40	3.23	10.98	11.56	10.43
3.50	3.20	11.20	12.25	10.24
4.39	4.70	20.63	19.27	22.09
3.60	3.76	13.54	12.96	14.14
1.90	2.43	4.62	3.61	5.90
5.20	8.98	46.69	27.04	80.64

4.60	4.50	20.7	21.16	20.25
2.60	2.66	6.92	6.76	7.07
3.21	4.35	13.96	10.30	18.92
4.40	4.03	17.73	19.36	16.24
3.69	4.61	17.01	13.61	21.25
5.70	3.90	22.23	32.49	15.20
1.90	2.48	4.71	3.61	6.15
2.60	3.58	9.31	6.76	12.816
1.90	2.05	3.89	3.61	4.20
3.20	5.08	16.26	10.24	25.81
8.90	7.36	65.50	79.21	54.17
$\Sigma X = 93.56$	**$\Sigma Y = 96.81$**	**$\Sigma XY = 409.07$**	**$\Sigma X^2 = 412.66$**	**$\Sigma Y^2 = 438.066$**

$\overline{X}$ $= \Sigma X / N,$ $\overline{Y}$ $= \Sigma Y / N$

$= 93.56/25$ $= 96.81/25$

$= 3.742$ $= 3.872$

$\overline{X}^2$ $= 3.742^2$ $\overline{Y}^2$ $= 3.872^2$

$= 14.003$ $= 14.992$

Calculation of Pearson's Correlation Coefficient (r) from Table 2,

$$r = \frac{\sum XY - n\overline{X}\overline{Y}}{\sqrt{\left(\sum X^2 - n\overline{X}^2\right)\left(\sum Y^2 - n\overline{Y}^2\right)}}$$

$$r = \frac{409.04 - 25 \times 3.72 \times 3.872}{\sqrt{\left(412.66 - 25 \times 14.003\right)\left(438.066 - 25 \times 14.992\right)}}$$

$$r = \frac{409.04 - 362.22}{\sqrt{\left(412.66 - 350.075\right)\left(438.066 - 374.8\right)}}$$

$$r = \frac{46.82}{\sqrt{\left(62.585\right)\left(63.266\right)}}$$

$$r = \frac{46.82}{\sqrt{\left(62.585\right)\left(63.266\right)}}$$

$$r = \frac{46.82}{\sqrt{-3959.50261}}$$

$$r = \frac{46.82}{62.92}$$

Answer, $r = 0.744$

(Note: The value of r is calculated from Table 2, i.e. 0.744, which indicates a strongly positive relation between the two variables taken for the study.)

6. Findings

- There is a strong positive relation between daily average page views per visitor and daily time on Site suggesting that if the average page view per visitor is high then the time spent on site also will be high, and low view will result in low time spent, and
- In terms of page views per visitor and view duration, websites of National Library of Tunisia, National Library of China, The National Library of Finland, Biblioteca Nacional Mariano Moreno, Argentina found higher rates.

7. References

Association, W. A. (2006). Web Analytics "Big Three" Definitions. Washington DC; 20037.

Avangate Home Page (1999) – How important is Alexa ranking? Available at http://www.avangate.com/articles/ alexa-ranking-99.htm

Bjorneborn, L. and Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics,50*,65–82.https://doi.org/10.1023/A:1005642218907

Das, N. G. (2009). Statistical methods. New Delhi: Tata McGraw Hill Education; 269–316p. PMCid: PMC2710084.

Gelder, P. V.; Beiger, G. and Berger, M. (2011). Statistical analysis of pageviews on web sites. Acceced on 2018. Available at: https://pdfs.semanticscholar.org/9b4b/a3ceef973ffa48e-5d71a93eded2c12f7add8.pdf

Goon, A. M.; Gupta, M. K. and Dasgupta, B. (1996). Basic statistics. Calcutta: World Press; 242-78p.

Rao, I. K. R. (2009). Quantitative methods for library and information science. New Delhi: New Age International; 98-113p. PMCid: PMC2797950.

support.google (2018). The difference between Google Ads Clicks, and Sessions, Users, Entrances, Pageviews, and Unique Pageviews in Analytics. Accessed on 24.8.2018. Available at: https://support.google.com/analytics/answer/1257084?hl=en a

Verma, M. K. and Brahma, K. (2017). A webometric analysis of National Libraries' websites in South Asia. *Annals of Library and Information Studies,* 64(2), 116–24.

Webopedia (2018). Average time on site – Web analytics. Accessed on 24.8.2018. Available at: https://www.webopedia.com/TERM/A/average_time_on_site.html.

Wikipedia (2018). Alexa Internet. Accessed on 24.8.2018. Available at: https://en.wikipedia.org/wiki/Alexa_Internet.