

Full Length Research Paper

Validating Google analytics tips for micro-firms

Beatriz Plaza^{1*}, Pilar Gonzalez Casimiro¹, M. Paz Moral Zuazo¹ and Idurre Ostolaza²

¹Faculty of Economics, University of the Basque Country, Avda. Lehendakari Agirre 8348015 Bilbao, Spain.

²Basquetour, Basque Tourism Agency. Agencia Vasca de Turismo. Alameda Urquijo, 36 – 5. Edificio Plaza Bizkaia. 48011 Bilbao, Basque Country, Spain.

Accepted 11 February, 2011

In today's hyper-competitive and increasingly cost-conscious business environment, making a website findable is critical to its success. Effective website traffic monitoring and benchmarking may lead to better web site strategies, and more competitive micro-firms. In this context, the aim of this article is to validate Plaza's Google Analytics e-Metrics for micro-firms. The focus of this article is experimental and features the analysis of a case study.

Key words: Website, online advertising, bounce rate.

INTRODUCTION

Performance measurement of websites is becoming a critical issue for effective online marketing, at big and small firms alike. Website data are easy to collect, but analysis and interpretation are time and money-consuming. Google Analytics is a free service offered by Google that returns elaborated statistics about a website's traffic. Google Analytics can track visitors from search engines, including all referrers, display advertising, pay-per-click networks, email marketing and digital collateral such as links within PDF documents (Wikipedia, 2010). Using cross-sectional data, Plaza (2010) presents an experiment done with the information that Google Analytics offers about the number of visits on a website and their precedence: organic results in search engines, links from referral web pages or direct access. The importance of Plaza's articles is not the particular case study, but the methodology employed to arrive at those results.

The author's case study must be presented only as a way to explain the methodology, because it is a particular case and it should be validated for different websites. More experiments are needed with different data sets, so that the method has more general value. The aim of this article is precisely to test the methodology developed by Plaza (2010) through a different case study. The article is

structured as follows. Firstly, a literature review on Google Analytics is supplied. Secondly, the author validates the e-Metrics initiated by Plaza (2010), in order to evaluate the usefulness of each traffic source (which includes direct visits, referral entries and search engine arrivals). This is then followed by the conclusions.

LITERATURE REVIEW

Several scientific articles have analysed the use of Google Analytics and evaluated its usefulness as a web analytics tool. Fang (2007), Bhatnagar (2009) used Google Analytics to evaluate and develop a library website, utilizing the ordinary reports from Google Analytics; although, without developing specific metrics. Hasan et al. (2009) suggests specific web metrics that are useful for quickly indentifying potential usability problems of e-commerce websites. Betty (2009) explores the use of Google Analytics to track usage statistics for interactive Shockwave Flash (.swf) files, the common file output for screen cast and Flash projects. Plaza (2009, 2011) explores some statistical matters with regards to the use of Google Analytics data, in combination with time series methodology. Finally, Plaza (2010) sets up Google Analytics e-Metrics for micro-firms using basic cross-sectional data. The aim of this article is to re-test this methodology for a specific Adventure Sports Tourism website: <http://www.troka.com>.

*Corresponding author: E-mail: beatriz.plaza@ehu.es.

HYPOTHESES TESTING

Plaza (2010) performed cross-sectional analysis with Google analytics showing that:

Rule #1: Return visits navigate deeper into the website and stay longer (that is, there is more time spent at the site and/or a greater number of pages viewed per visit).

Rule #2: The less the bounce rate (that is, the number of error visits), the longer the visit length (with regard to the time spent at the site and/or the number of pages viewed per visit).

Rule #3: The greater the return visit rate, the less the bounce rate (that is, the number of error visits).

These rules are now tested for an active sports tourism website: <http://www.troka.com>. Business professionals take action based on measurable results using analytics to improve the effectiveness of the web content management and leverage information in their website. The aim of this work is to validate Plaza's Key Performance Indicators (KPI) using Google Analytics. Our tested website is <http://www.troka.com>, an active sports tourism firm. Troka Abentura, S.L. was born like 'Active Tourism and Adventure Sports' company in 1999. The firm's website pursues the dissemination of information in the field of the company's products. In June 2007 the Webmaster started to analyse web traffic using Google Analytics (<https://www.google.com/analytics/>). From 15 June 2007 to 14 June 2010, Google Analytics registers 57,095 entries. Of those visits, 48,675 came directly to this site, referring sites sent 2,210 visits, and search engines sent a total of 6,210 visits, mainly through Google. Direct visits are, by far, the main source of entries for <http://www.troka.com>: almost 85% of the total incoming visits (Figure 1). But how deep into the website do direct visits navigate, in comparison with other traffic sources? Are direct visits more effective than search engine entries? Which is the most effective traffic source? How deep do Google entries navigate? Which are the most effective keywords?

For the purpose of providing a preliminary answer to these questions, the already stated 3 rules (and key metrics) are applied in order to measure simple cross-sectional data as follows:

1. The first step is to collect all the data (Table 1): The number of visits for each traffic source, session length (that is, the time spent on the site and/or the number of pages viewed per visit), the bounce rate and the return visits rate. These indicators correspond to average values for the period 15 June 2007 to 14 June 2010.

2. Next, the traffic sources have to be sorted according to traffic volume. A ranking from the highest to the lowest traffic volume source is established. Next, the top ten are selected.

3. Then, the top traffic sources with the highest return rate should be selected. The key questions are the following: Which are the traffic sources that generate traffic and also produce a high return rate? Which are the traffic sources that produce entries and achieve a return rate above average? The answers can be seen in Figures 2, 3 and 4. In general, return visits travel deeper into the website and bounce less. It is for this reason that 'maximizing return rate' criteria has been chosen. However, the website owner can either choose to minimize 'bounce rate' or to maximize session length - with regards to the time spent at the site and/or the number of pages viewed per visit. For this particular website, it can be seen that the most effective traffic sources are the keyword 'troka', direct traffic and Google.

4. Next, a scatter plot is created for the return rate against the number of pages viewed per visit for all the main traffic sources (Figure 2). From Figure 2, it can be seen that there is a positive relationship between return rates and the number of pages viewed per visit for the traffic sources (Rule No. 1).

5. The next stage is to scatter plot bounce rates against the number of pages viewed per visit (Rule No. 2 in Figure 3). The aim here is to identify the qualified low bounce traffic sources. The referrals <http://www.mybilbao.bizkaia.com> and <http://www.turismo.gorliz.net> and the keyword 'aventura' in search engines qualify with the lowest bounce rate (Figure 3).

6. The next stage is to scatter plot bounce rate against return rate for all the traffic sources (Figure 4). From Figure 4, it can be seen here that, there is a negative relationship between bounce rate and return rate (Rule No. 3). The keyword 'troka' qualifies both a high return rate and the low bounce rate. Direct traffic and the referral www.troka.com perform also relatively well, in terms of higher return rates and qualified low bounce rate traffic.

The webmaster can quantify the relationships that underlie these graphs through very simple regression analysis, as can be seen in Tables 2, 3 and 4. Several regressions are undertaken. The Breusch-Godfrey Serial Correlation LM Test is used to check autocorrelation. The White Test is used to test heteroskedasticity, and the Jarque-Bera statistic to test normality of residuals. The presence of outliers is corrected through the use of dummies. The regressions are well-adjusted. The fitted estimations are in Tables 2, 3 and 4.

According to the reading of the results in Table 2, a 1% increase in the return rate leads to a 0.11 increase in the number of pages viewed per visit. Furthermore, according to the Intercept Dummy Variables, the referrals www.adclick.es and <http://es-es.facebook.com> under perform. In other words, it is clear for this particular website that return behaviour increases visit duration.

A 1% increase in the Bounce rate leads to a 0.07

Table 1. Traffic sources for www.troka.com (average values from 15 June 2007 to 14 June 2010).

	Variable	Visits	Pages per visit	Average time on site	Bounce rate (%)	Return visits rate (%)
Traffic sources	Total	57.095	6.0	0:02:19	46.7	35.6
	Direct traffic	48.675	6.2	0:02:09	47.5	36.7
	Referring sites	2.210	6.1	0:02:51	38.6	25.5
	Search engines	6.210	4.8	0:03:25	42.9	30.4
Top 7 referring sites by traffic	gorlizaterpetxea.com	1.243	7.7	0:03:02	32.5	23.1
	ruraliberica.com	100	7.0	0:06:16	27.0	26.0
	zankyou.com	53	1.6	0:00:57	83.0	3.8
	turismo.gorliz.net	36	6.8	0:05:48	25.0	30.6
	mybilbaobizkaia.com	35	3.3	0:02:08	22.9	11.4
	adclick.es	34	5.3	0:05:47	44.1	64.7
	facebook.com	22	2.7	0:01:04	54.6	59.1
Search engines	Google	5.555	5.0	0:03:37	40.5	32.5
	Bing	246	2.5	0:01:23	70.3	9.4
	Search	241	2.7	0:01:39	61.0	16.2
	Yahoo	66	3.4	0:02:16	62.1	7.6
Top 15 keywords by traffic	All keywords	6.210	48	0:03:25	42.9	30.4
	troka	3.339	6.3	0:04:41	27.8	45.8
	abentura	1.246	4.7	0:03:15	44.7	26.4
	aventura	650	5.9	0:03:28	25.5	24.0
	Gorliz	575	5.7	0:03:49	37.2	27.7
	albergue	342	4.9	0:03:13	34.2	17.5
	troka.com	269	9.7	0:08:02	28.6	50.6
	bizkaia	241	5.1	0:05:25	35.7	22.8
	surf	209	3.6	0:02:11	32.5	27.8
	vasco	206	3.4	0:02:07	43.2	16.5
	vizcaya	155	4.0	0:02:44	43.2	8.4
	deporte	120	3.2	0:01:33	51.7	10.0
	curso	111	3.1	0:01:53	42.3	26.1
	euskadi	102	3.5	0:02:09	44.1	6.9
	actividad	88	3.8	0:02:05	44.3	7.9
tirolina	86	4.9	0:04:43	66.3	24.4	

Source: Google Analytics for www.troka.com

decrease in the number of pages viewed per visit (Table 3). The negative relationship between Visit Duration and Bounce Rate (error visits) makes sense. Finally, a 1% increase in the Bounce Rate leads to a 0.40% decrease in the return rate (Table 4). Furthermore, according to the Intercept Dummy Variables, the referrals www.adclick.es and http://es-es.facebook.com perform above average, showing a higher return rate. Summarising, the lesser the bounce rate, the better the website's performance. On average, traffic sources with a high bounce rate shows that, the webmaster failed to meet his/her expectations.

Conclusions

Google Analytics tips for micro-firms have been re-tested with a different data set. This new test confirms that, traffic sources that work most effectively are those: 1) which drive a higher traffic volume; 2) which have the highest return rate; 3) which have the largest visit length and 4) which have the lowest bounce rate.

The results obtained here with cross-sectional data from the website http://www.troka.com are consistent with those results obtained by Plaza (2010) with data from the

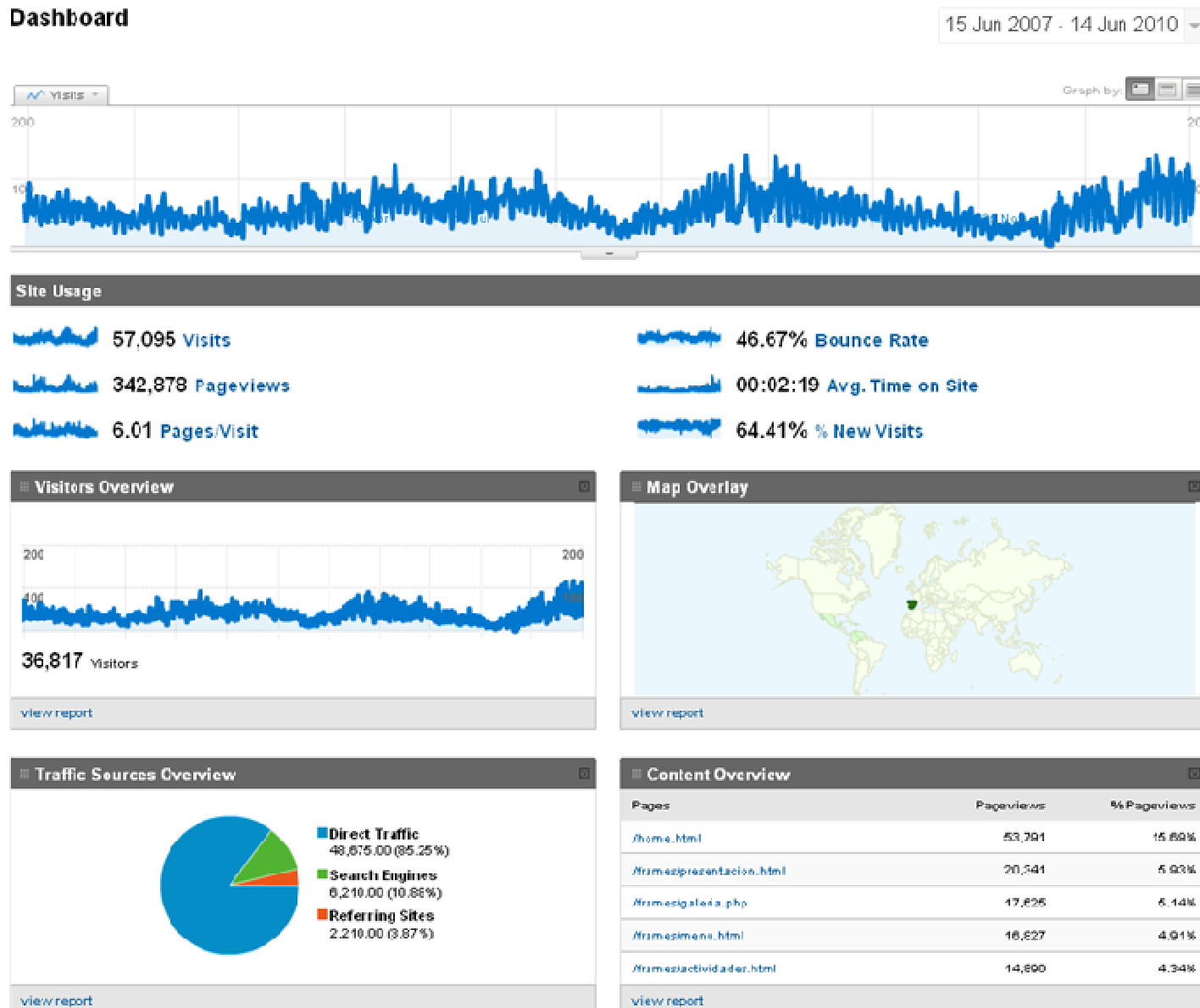


Figure 1. Google Analytics overview for www.troka.com (daily data, 15 June 2007 to 14 June 2010). Source: Google Analytics for www.troka.com.

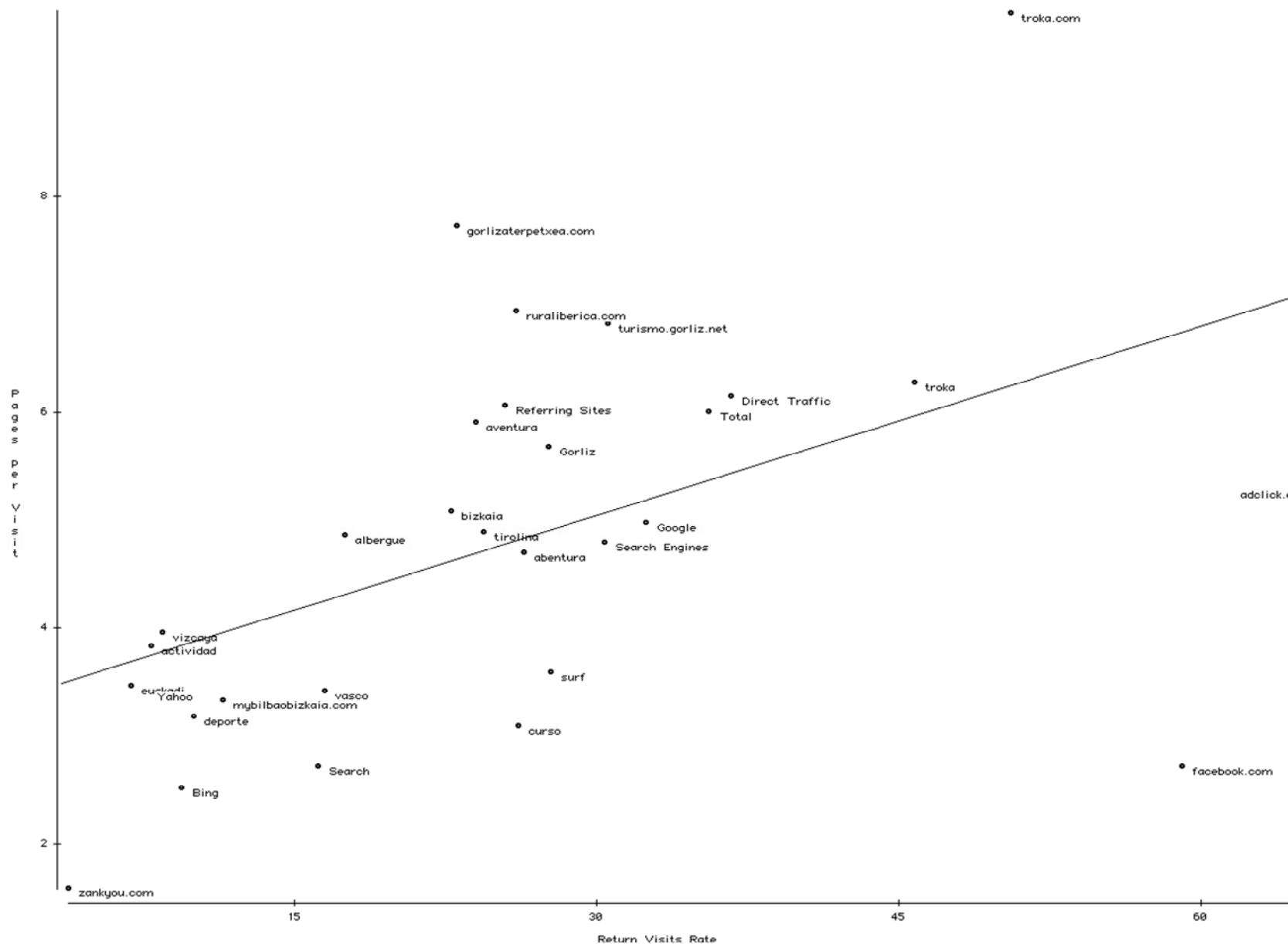


Figure 2. Traffic sources for www.troka.com. Rule #1, return visits navigate deeper into the website and stay longer (average values from 15 June 2007 to 14 June 2010).

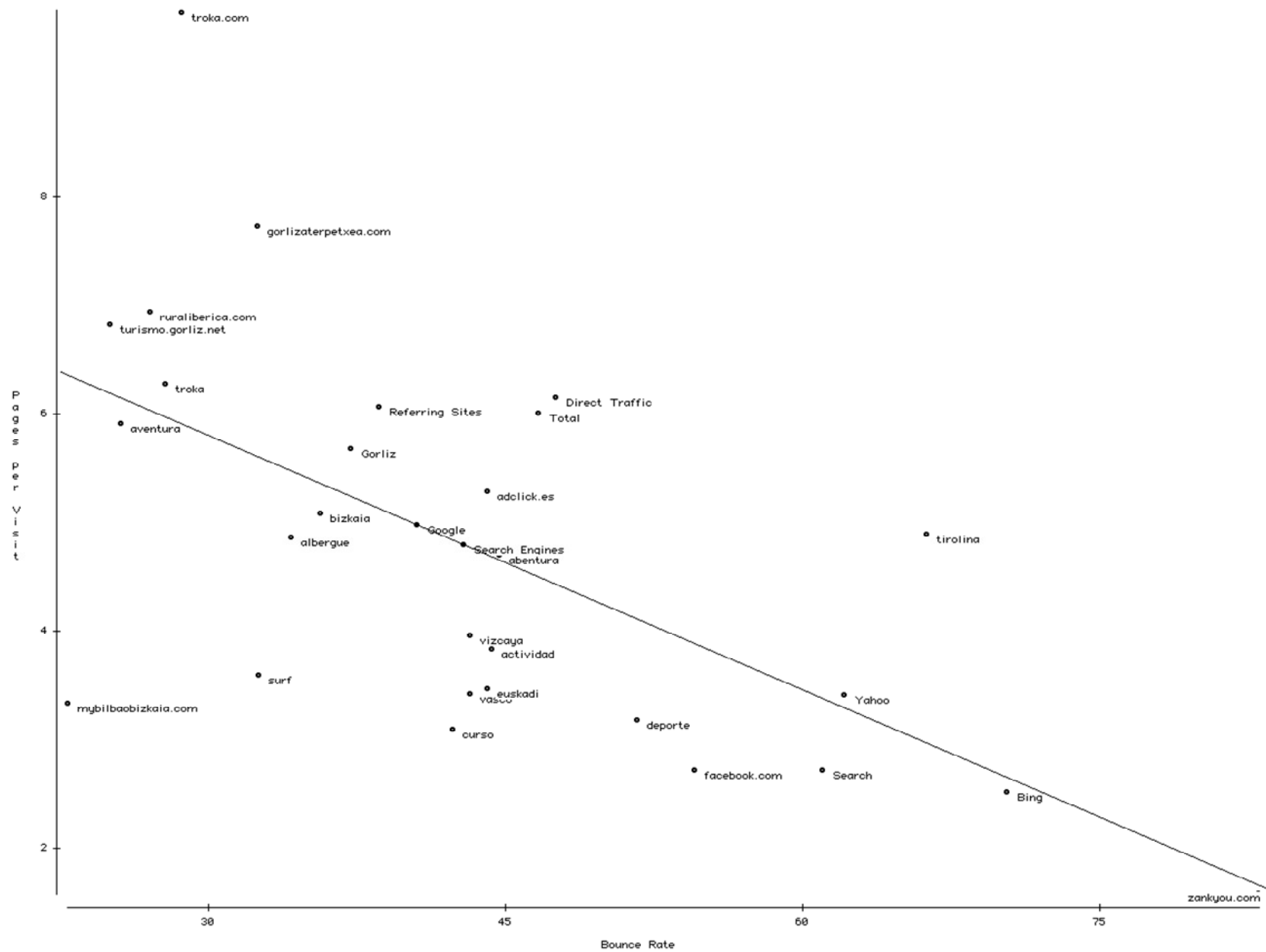


Figure 3. Traffic sources for www.troka.com. Rule #2, the smaller the bounce rate, the greater the visit duration (average values from 15 June 2007 to 14 June 2010).

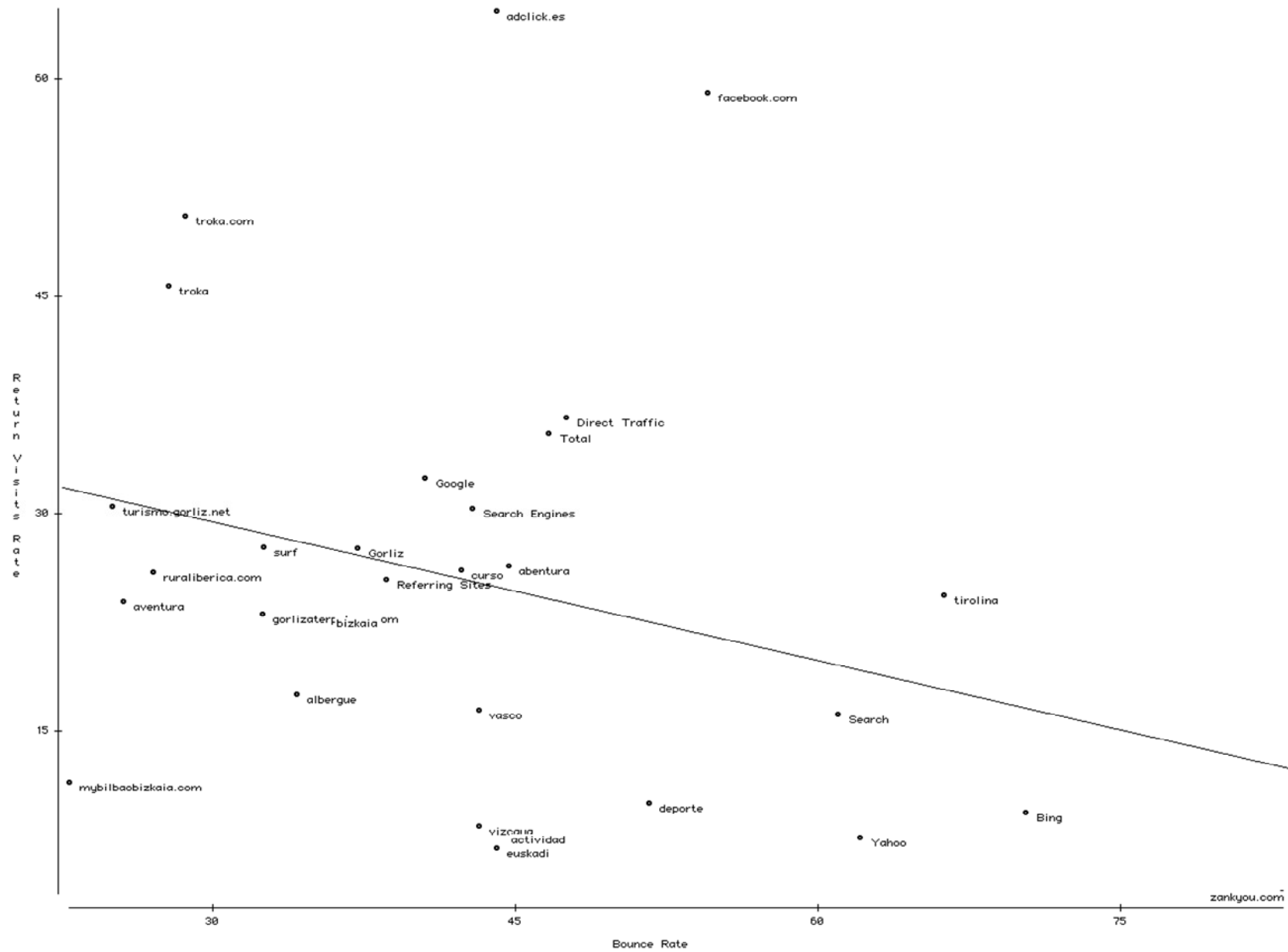


Figure 4. Traffic sources for www.troka.com. Rule #3, the greater the return visit rate, the smaller the bounce rate, (average values from 15 June 2007 to 14 June 2010).

Table 2. Regression for pages per visit (average values from 15 June 2007 to 14 June 2010).

Variable	Coefficient	Standard error	t-Statistic	Probability
Constant	2.18	0.47	4.60	0.000
Return rate	0,11	0.01	6.26	0.000
Dummy adclick.es (in-link)	-4.45	1.41	-3.15	0.000
Dummy facebook (in-link)	-6.36	1.35	-4.68	0.000
N = 30				
R ² = 0.62	F-statistic = 14.21		Prob (F-statistic) = 0.00	
Breusch-Godfrey Serial Correlation LM Test:	F-statistic 0.14		Probability 0.70	
White Heteroskedasticity Test:	F-statistic 0.62		Probability 0.64	
Jaque-Bera 3.19			Probability 0.63	

Table 3. Regression for pages per visit (average values from 15 June 2007 to 14 June 2010).

Variable	Coefficient	Standard error	t-Statistic	Probability
Constant	8.16	0.82	9.89	0.000
Bounce rate	-0.07	0.01	-4.33	0.000
N = 30				
R ² = 0.40	F-statistic = 18		Prob(F-statistic) = 0.00	
Breusch-Godfrey Serial Correlation LM Test:	F-statistic 1.06		Probability 0.35	
White Heteroskedasticity Test	F-statistic 2.03		Probability 0.14	
Jaque-Bera 1.71			Probability 0.42	

Table 4. Regression for return rate (average values from 15 June 2007 to 14 June 2010).

Variable	Coefficient	Standard error	t-Statistic	Probability
Constant	39.81	6.23	6.38	0.000
Bounce rate	-0.40	0.13	-2.92	0.007
Dummy adclick.es (in-link)	42.65	10.7	3.98	0.000
Dummy facebook (in-link)	41.22	10.8	3.80	0.000
N = 30				
R ² = 0.57	F-statistic = 11.59		Prob(F-statistic) = 0.00	
Breusch-Godfrey Serial Correlation LM Test	F-statistic 2.15		Probability 0.21	
White Heteroskedasticity Test:	F-statistic 0.71		Probability 0.59	
Jaque-Bera 0.46			Probability 0.79	

web 'Scholars on Bilbao' <http://www.scholars-on-bilbao.info>.

Future work calls for running the same experiment in other websites with different web architecture and dissimilar aims.

ACKNOWLEDGEMENT

The Basque Government provided financial support for

this Project (SAIOTEK SA-2010/00072). We are also grateful to Basquetour, which kindly gave technical support. They are not responsible for our interpretations.

REFERENCES

- Art4pax Foundation (2010). Scholars on Bilbao. URL: <http://www.scholars-on-bilbao.info>.
- Betty P (2009). Assessing homegrown library collections: Using google analytics to track use of screencasts and flash-based learning objects. *J. Electron. Resour. Librariansh.*, 21(1): 75-92.

- Bhatnagar A (2009). Web Analytics for Business Intelligence. Online. 33(6): 32-35.
- Fang W (2007). Using Google Analytics for Improving Library Website Content and Design: A Case Study. Library Philosophy and Practice, LPP Special Issue on Libraries and Google.
- Hasan LM, Proberts SA (2009). Using Google Analytics to Evaluate the Usability of E-Commerce Sites. Lecture Notes Comput. Sci., 5619: 697-706.
- Plaza B (2009). Monitoring web traffic source effectiveness with Google Analytics. An experiment with time series. Aslib Proceedings, 61(5): 474-482.
- Plaza B (2010). Google Analytics Intelligence for Information Professionals. Online, 34(5): 33-37.
- Plaza B (2011). Google Analytics for measuring website performance. Tour. Manage., 32(3): 477-481.
- Wikipedia (2010). Google Analytics. Retrieved 2010-10-14.