Dissertation Report on

Evaluating the power of Big Data Analytics by assessing the effectiveness of Google Adsense

Submitted by:
Shivam Tiwari
DTU/2K12/MBA/60

Under the Guidance of:
Dr. Shikha N. Khera
Assistant Professor

DELHI SCHOOL OF MANAGEMENT
Delhi Technological University
Bawana Road, New Delhi 110042
May 2014

CERTIFICATE FROM THE INSTITUTE

This is to certify that the Project Report titled *"Evaluating the power of Big Data Analytics by assessing the effectiveness of Google Adsense"* is a bonafide work carried out by Mr. Shivam Tiwari of MBA 2012-14 and submitted to Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42 in partial fulfillment of the requirement for the award of the Degree of Masters of Business Administration.

Signature of Guide

Signature of Head (DSM)

Place:

Date:

Seal of Head

DECLARATION

I, Shivam Tiwari, student of MBA 2012-14 of Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42 declare that Dissertation Report on *"Evaluating the power of Big Data Analytics by assessing the effectiveness of Google Adsense"* submitted in partial fulfillment of Degree of Masters of Business Administration is the original work conducted by me.

The information and data given in the report is authentic to the best of my knowledge.

This Report is not being submitted to any other University for award of any other Degree, Diploma and Fellowship.

Shivam Tiwari

Place:

Date:

ACKNOWLEDGEMENT

The satiation and euphoria that accompany the successful completion of the project would be incomplete without the mention of the people who made it possible.

I would like to take the opportunity to thank and express my deep sense of gratitude to my faculty mentor Dr. Shikha N. Khera. I am greatly indebted to her for providing valuable guidance at all stages of the study, her advice, constructive suggestions, positive and supportive attitude and continuous encouragement, without which it would have not been possible to complete the project.

I would also like to thank Mr.Anuj Srivastava, a technology enthusiast and a passionate software engineer, who in spite of his busy schedule has co-operated with me continuously and indeed, his valuable contribution in the form of detailed description of the technical aspects of Google Ads, and the nitty gritties of associated tools have been certainly indispensable for my project work.

Last but certainly not the least, I am thankful to Ms. Tanya Saraswat and Ms. Sakshi Gupta, my friends and fellow students, for helping me with the statistical tools and their analysis.

I hope that I can build upon the experience and knowledge that I have gained and make a valuable contribution towards marketing in any industry in coming future.

Shivam Tiwari,
DSM-DTU 2012-14
New Delhi Signature

EXECUTIVE SUMMARY

The objective of the project was to understand the strengths of Big Data analytics, a term very widely used for obtaining vital information for the business out of large amount of data related to the users, customers or consumers. In the early phase of research, the project was planned with a focus on Hadoop (the biggest tool for data driven marketing used by organizations), success or failure of Data Driven marketing can largely be assessed through its results. But being an algorithm largely put in use by organizations, it was not possible to utilize it for the project considering the lack of time, resources and low scale nature of the study. However, a ray of light came in the form of the world's largest search engine and a frontrunner in exploiting new technologies, Google. Google analytics is an in-house developed program, which can assertingly be called the 'Google's version of Big Data Analytics' model that has revolutionized Data Driven Marketing when it comes to Online Advertising. It currently finds its implementation in offering the "Google Adword" and "Google Adsense" products. So this project is a humble attempt of evaluating the viability and effectiveness of Google Analytics and DoubleClick (in turn Big Data) in the attempt to grab the attention of online buyers through Adsense.

The project was started on 25th March, 2014 after researching a bit on all the relevant information regarding the topic, i.e., about Big Data analytics, its relation to Google Analytics, about adword and adsense and the technology on which they are based. After thorough research on the secondary data available on the internet, the problem statement was formed.

Since, the next part of my project was to develop the questionnaire directed towards understanding of real time effectiveness of Adsense, the detailed study of the data and information available in various websites provided me with a lead in developing questionnaire and my faculty mentor guided me in finalizing it. For this the questionnaire was imperatively directed towards understanding the response of customers towards random google ads popping up in the websites, which were actually based on their browsing history and RSS feeds. The respondents were avid internet users, not necessarily enthusiastic online shoppers. The sample size of the marketing

research was taken to be 85. The questionnaire contained various aspects related to customer psyche and behavior of Google Adsense like the relation between their search and browsing history and ads, online shoppers 'latest viewed products' in e-commerce portals and popping up of the same in google ads, relevance of the ads and their online shopping patterns based on the ads, etc. It is to be noted that the research tries to understand the effectiveness from the perspective of retail customers and its impact in their buying behavior, and in no sense it tries to examine the advantages that adsense may give to e-commerce portals or any other offline advertiser.

Most important part is analyzing the information gathered through the respondents. The real challenge lied in measuring the effectiveness of Google AdSense in the various verticals of Online Advertising and then linking the result to Big Data Analytics to understand the power, capabilities and opportunities in Data Science.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Big Data

Big data is a popular term used to describe the exponential growth and availability of data, both structured and unstructured. And big data may be as important to business – and society – as the Internet has become because more data may lead to more accurate analyses; more accurate analyses may lead to more confident decision making. And better decisions can mean greater operational efficiencies, cost reductions and reduced risk.

As far back as 2001, industry analyst Doug Laney (currently with Gartner) articulated the now mainstream definition of big data as the three Vs: volume, velocity and variety [*Source*: *http://www.sas.com/en_us/insights/big-data/what-is-big-data.html*]

- **Volume.** Many factors contribute to the increase in data volume. Transaction-based data stored through the years. Unstructured data streaming in from social media. Increasing amounts of sensor and machine-to-machine data being collected. In the past, excessive data volume was a storage issue. But with decreasing storage costs, other issues emerge, including how to determine relevance within large data volumes and how to use analytics to create value from relevant data.

- **Velocity.** Data is streaming in at unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time. Reacting quickly enough to deal with data velocity is a challenge for most organizations.

- **Variety.** Data today comes in all types of formats. Structured, numeric data in traditional databases. Information created from line-of-business applications. Unstructured text documents, email, video, audio, stock ticker data and financial transactions. Managing, merging and governing different varieties of data is something many organizations still grapple with.

The real issue is not that you are acquiring large amounts of data. It's what you do with the data that counts. The hopeful vision is that organizations will be able to take data from any source, harness relevant data and analyze it to find answers that enable 1) cost reductions, 2) time reductions, 3) new product development and optimized offerings, and 4) smarter business decision making. For instance, by combining big data and high-powered analytics, it is possible to:

- Determine root causes of failures, issues and defects in near-real time, potentially saving billions of dollars annually.
- Optimize routes for many thousands of package delivery vehicles while they are on the road.
- Analyze millions of SKUs to determine prices that maximize profit and clear inventory.
- Generate retail coupons at the point of sale based on the customer's current and past purchases.
- Send tailored recommendations to mobile devices while customers are in the right area to take advantage of offers.
- Recalculate entire risk portfolios in minutes.
- Quickly identify customers who matter the most.
- Use clickstream analysis and data mining to detect fraudulent behavior.

1.2 Data Driven Marketing

'Big' is good. 'Smart' is better. The first wave of the big data revolution was about scale, processing massive data inputs to render new insights about customers and markets. The next wave is more targeted – knitting together diverse sources, scaling the data against specific goals, crafting internal processes and policies that make the most of

the data, and figuring out how to quantify and attribute the diverse marketing that the marketers across the industry are already doing.

In Data Driven Marketing, or DDM, marketers explore some of the real-world challenges they face in finding the ROI in data investment. They start with the consumer and that complex "journey" that is itself a data gathering exercise across screens. But they also look at the technology "stack" marketers and agencies need to erect in order to track and serve that customer, often in real time. Then they engage two of the biggest challenges ahead: tracking consumers beyond the Web onto devices and offline, and using data to create more human and personal interactions along the way. And because data ultimately is coming from the consumer, marketers engage how companies become stewards of personal information. And finally they follow the data trail where it loops back to the starting point – attribution.

1.3 Big Data Analytics

Big data analytics refers to the process of collecting, organizing and analyzing large sets of data ("big data") to discover patterns and other useful information. Not only will big data analytics help you to understand the information contained within the data, but it will also help identify the data that is most important to the business and future business decisions. Big Data analysts basically want the *knowledge* that comes from analyzing the data.

Enterprises are increasingly looking to find actionable insights into their data. Many big data projects originate from the need to answer specific business questions such as what customers really think about our brand, and how can we increase our sales intelligence and close more deals. With the right big data analytics platforms in place, an enterprise can boost sales, increase efficiency, and improve operations, customer service and risk management.

To analyze such a large volume of data, big data analytics is typically performed using specialized software tools and applications for predictive analytics, data mining, text mining, forecasting and data optimization. Collectively these processes are separate but

highly integrated functions of high-performance analytics. Using big data tools and software enables an organization to process extremely large volumes of data that a business has collected to determine which data is relevant and can be analyzed to drive better business decisions in the future.

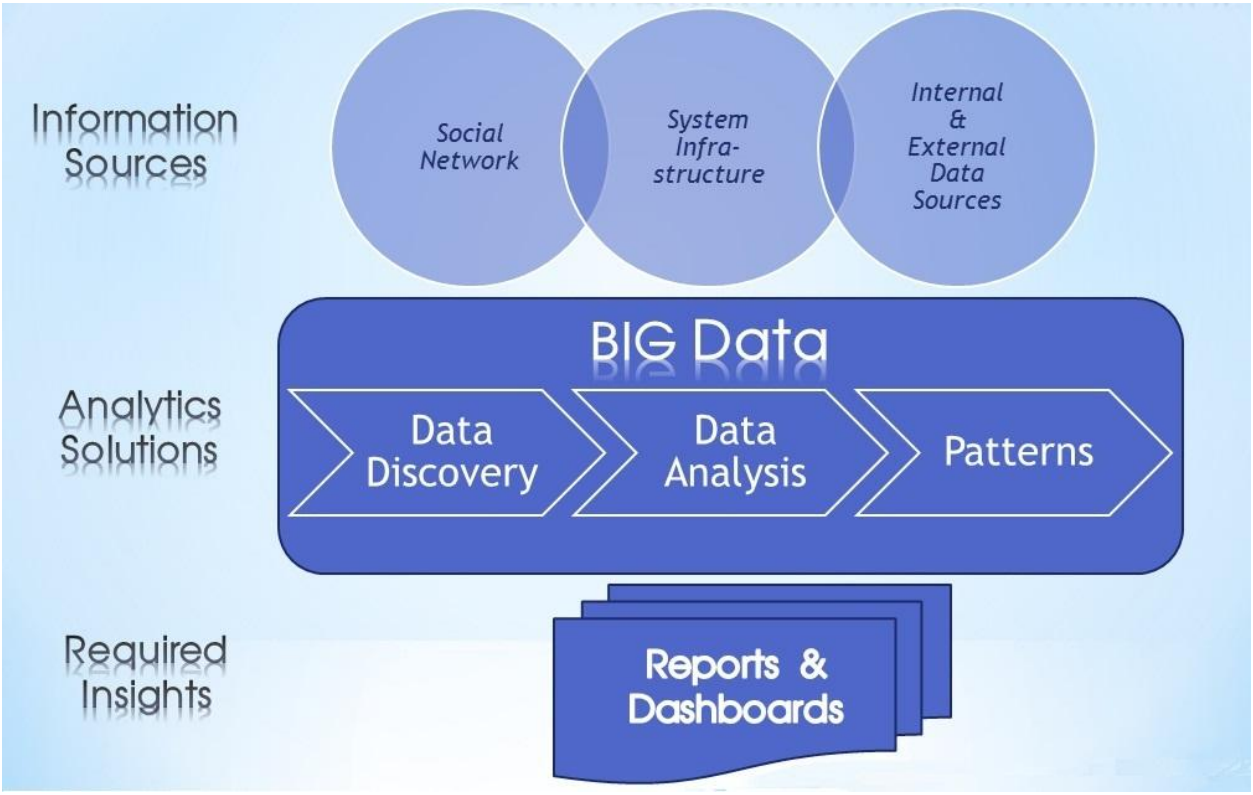

Figure 1.1 Concept behind Big Data Analytics

[*Source: http://blog.maia-intelligence.com/2013/01/16/high-performance-analytics-%E2%80%93-2/*]

1.4 Google Analytics as a tool for Big Data Analytics

Google Analytics is a service offered by Google that generates detailed statistics about a website's traffic and traffic sources and measures conversions and sales. The product is aimed at marketers as opposed to webmasters and technologists from which the industry of web analytics originally grew. It's the most widely used website statistics service.

The basic service is free of charge and a premium version is available for a fee.

Google Analytics can track visitors from all referrers, including search engines and social networks, direct visits and referring sites. It also displays advertising, pay-per-click networks, email marketing and digital collateral such as links within PDF documents.

Integrated with DoubleClick, AdWords and AdSense, users can now review online campaigns by tracking landing page quality and conversions (goals). Goals might include sales, lead generation, viewing a specific page, or downloading a particular file.

Google Analytics' approach is to show high-level, dashboard-type data for the casual user, and more in-depth data further into the report set. Google Analytics analysis can identify poorly performing pages with techniques such as funnel visualization, where visitors came from referrers, how long they stayed and their geographical position. It also provides more advanced features, including custom visitor segmentation.

Google Analytics e-commerce reporting can track sales activity and performance. The e-commerce reports show a site's transactions, revenue, and many other commerce-related metrics.

Google Analytics launched Real Time analytics as of September 29, 2011.

A user can have 50 site profiles. Each profile generally corresponds to one website. It is limited to sites which have a traffic of fewer than 5 million page views per month (roughly 2 page views per second), unless the site is linked to an AdWords campaign. [Source: http://en.wikipedia.org/wiki/Google_Analytics]

Internet and Social media is a key model of the velocity and variety which are associated with Big Data. With Internet, data is coming at you at an inconceivable speed and in a number of formats including videos and pictures, search keywords, browsing history and RSS feeds. Only Big Data Applications, i.e., Google Analytics in this case, can enable e-commerce portals to manage these online conversations in real-time. That is why internet is indeed mountains of big data waiting to be explored. Companies cannot miss the ability to make sense of what all the available data means to their business and

marketing strategies. Big Data is usually associated to having volume, velocity, variety, variability and complexity which are also features of Google Analytics. So they are willing subscribers of Google AdWords and AdSense technologies which enable them to reach out to more customers through various mediums of online advertising offered by Google.

1.5 Double Click

DoubleClick Digital Marketing (DDM) is an integrated ad-technology platform that enables agencies and advertisers to more effectively create, manage and grow high-impact digital marketing campaigns. DDM integrates world-class solutions to help buyers run holistic campaigns across multiple channels.

DoubleClick targets along various criteria. Targeting can be accomplished using IP addresses, business rules set by the client or by reference to information about users stored within cookies on their machines. Some of the types of information collected are:

- Web browser
- Operating System
- ISP
- Bandwidth
- Time of day

In addition, the cookie information may be used to target ads based on the number of times the user has been exposed to any given message. This is known as "frequency capping".

So it is evident that such a level of detail about internet users is being maintained by Google, which implies that it’s a lot of data per user; which in turn is multiplied by billions of Internet Users world wide. Hence, Google is utilizing the power of Big Data through AdSense, AdWords, Analytics and DoubleClick.

2. LITERATURE REVIEW

"Google is a colossus that sits astride access to information on the World Wide Web. Ubiquitous, useful, and often imitated-but seldom equaled-Google has lent its name to a verb: to *google* something (or someone) is to search for the thing or person on the Web. Google is also a forward looking corporation filled with brilliant thinkers and one of the largest companies in the world in terms of market capitalization.

The primary focus of this book is making money with the Google advertising applications: the AdSense and AdWords programs. These programs are closely related to Google's searching technology. AdSense ads are placed on your web site depending on the context of your site (in other words, Google's analysis of how your site is likely to be found). And AdWords ads are targeted using keywords and phrases-the same keywords and phrase used when searching for something with Google." (Davis, H., 2010)

Due to the fact that many Americans today are having problems with work and income, but still have time to spend on Facebook or other social media, it may be the right moment in internet marketing history to utilize a social medium they are already wedded to as a platform to make some revenue. The aim is to see if Facebook friends, on a daily basis, can actually be a good way to increase Google AdSense Revenue rather than relying on search engines to bring customers or clickers to your website without asking your social capital (Facebook friends) to intentionally click on the AdSense advertisement. It was discovered that AdSense clicks, web content searches, and website visits were significant in driving AdSense revenue through Facebook to any website. AdSense clicks and click-through-ratios (CTR) where significant for low content search value websites without revenue increase, while AdSense clicks and AdSense page views were significant for high content search value websites when Facebook was utilized to increase AdSense revenue. (Aimiuwu, E. E.; Bapna, S. & Ahmed, A., 2012)

The analysis of qualitative and quantitative data from any website and the competition, to drive a continual improvement of the online experience that customers, and potential

customers have, which translates into marketers desired outcomes (online and offline). (Kaushik A., 2012)

Google Analytics provides a core set of tools that supports some of the primary tasks that web analysts perform.

First and foremost, Google Analytics tracks many standard website metrics, like visits, unique visitors, pageviews, bounce rate, and abandonment rate. But more importantly, it can track business outcomes, called goals.

In addition to tracking goals, Google Analytics does a great job at tracking all different kinds of marketing initiatives. Many people believe that Google Analytics can only track AdWords, but it can track other types of paid searches, email marketing, display advertising, social media, and any other type of ad you can think of. (Cutroni, Justin, 2010)

Google Analytics does not identify requests originatingfrom different physical locations as belonging to the same user,and as all other cookie-based mechanism is prone to the cookiedeletion vulnerability noted in the comScore's and our studies.To prove the latter point participants of the study presented inthis paper Section 5 were asked to clear their cookies in order toascertain the impact on Google Analytics. As a result GoogleAnalytics recorded a predictable drop in the returning visitortraffic on the day the cookies were cleared.

Google Analytics -the industry leader in the user tracking is fooled by sporadiccookie clearing and the multitude of Internet accesslocations/devices and in our case overestimated the number of unique visitors by a factor of six during the month-long study. (Fomitchev, M. I.,2010)

The processes in figure 1 provide data, which is at the core of successful Internet marketing and suitable for data mining. The data may be described along the following dimensions:

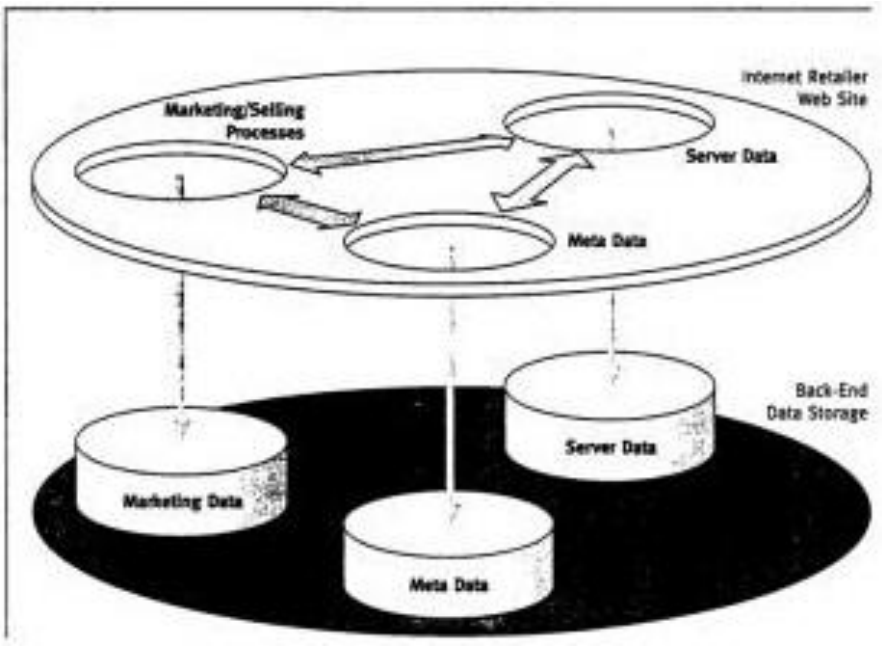

Figure 2. 1 The types of data available on the internet related to a user

[*Source: Data-Driven Marketing.', Electronic Markets, 1998 (pg. 3)*]

1. Server Data: data generated by the interactions between the persons browsing an individual site, and the web server. For example, the ip address of a computer browsing a retailing site.
2. Marketing Data: the data stored by the Internet retailer on products, customers, suppliers, etc. for example, the consumer responses to discounting.
3. Site 'meta' data: the data about the site, usually generated by dynamically and automatically after a site update. For example, location of product pages on a site, which are "leaf" pages, navigational pages, etc.

There are now hundreds of thousands of on-line Internet shopping providers worldwide, with thousands more appearing each month. Data generated by clients of the on-line retailers by their interaction with the site in browsing and buying. Clearly such large data sets, which may be distributed and heterogeneous, will contain useful information helpful to business marketing strategies, both for retrospective analyses as well as data-driven forecasting.

The potential for this new form of data-driven marketing is enormous-browsers and buyers of products can be identified and targeted with attractive offers and sales promotions. (Mulvenna, M. D.; Norwood, M. & Büchner, A. G., 1998)

As a result of their(social media) massive popularity, these(social networking) sites have been exploited as a platform for the viral marketing of content, products, and political campaigns. For instance, major movie studios place trailers for their movies on MySpace; US presidential candidate ran online political campaigns on You Tube; and individuals and amateur artists promote their songs, artwork, and blogs through these sites, all hoping to reach millions of online users. One of the distinguishing features of online social networks is the potential for information dissemination along the social links, i.e., information propagation among friends in social networks, one hop at a time. It is widely believed that such user-to-user exchanges, also known as "word-of-mouth" exchanges, can spread content, ideas, or information widely and quickly throughout the network. (Cha, M.; Mislove, A. & Gummadi, K. P., 2009)

Global businesses have entered a new era of decision making. The ability to gather, store, access, and analyze data has grown exponentially over the past decade, and companies now spend tens of millions of dollars to manage the information streaming in from suppliers and customers.

For all the breathless promises about the return on investment in Big Data, however, companies face a challenge. Investments in analytics can be useless, even harmful, unless employees can incorporate that data into complex decision making.(In Google AdSense case, the advertising). (Shah S., Horne A., and Capella´ J.,2012)

The problem is not with the data. We all know that data has to be structured into actionable information. Creation of KPI's and dashboard for each stratum of management is key to leveraging data within an organization framework. At each level, you have to understand what data is going to be most effective to measure, track and control. For example if you are in Customer Service and tracking "returned goods" as the only metric you got to control, so be it.. The lesser the volume of data, the more people can effectively act upon it. (Maiti S., 2011)

Plenty of retailers collect data about their stores and their shoppers, and many use the information to try to improve sales. Target Stores, for example, introduced a branded Visa card in 2001 and has used it, along with an arsenal of gadgetry, to gather data ever since. But Wal-Mart amasses more data about the products it sells and its shoppers' buying habits than anyone else, so much so that some privacy advocates worry about potential for abuse.(Hays C. L., 2004)

The "data scientist" is a high-ranking professional with the training and curiosity to make discoveries in the world of big data. The title has been around for only a few years. (It was coined in 2008 by D.J. Patil, and Jeff Hammerbacher, then the respective leads of data and analytics efforts at LinkedIn and Facebook.) But thousands of data scientists are already working at both start-ups and well-established companies. Their sudden appearance on the business scene reflects the fact that companies are now wrestling with information that comes in varieties and volumes never encountered before. If your organization stores multiple petabytes of data, if the information most critical to your business resides in forms other than rows and columns of numbers, or if answering your biggest question would involve a "mashup" of several analytical efforts, you've got a big data opportunity.

Much of the current enthusiasm for big data focuses on technologies that make taming it possible, including Hadoop (the most widely used framework for distributed file system processing) and related open-source tools, cloud computing, and data visualization. While those are important breakthroughs, at least as important are the people with the skill set (and the mind-set) to put them to good use. On this front, demand has raced ahead of supply. Indeed, the shortage of data scientists is becoming a serious constraint in some sectors. Greylock Partners, an early-stage venture firm that has backed companies such as Facebook, LinkedIn, Palo Alto Networks, and Workday, is worried enough about the tight labor pool that it has built its own specialized recruiting team to channel talent to businesses in its portfolio. "Once they have data," says Dan Portillo, who leads that team, "they really need people who can manage it and find insights in it."

If capitalizing on big data depends on hiring scarce data scientists, then the challenge for managers is to learn how to identify that talent, attract it to an enterprise, and make it

productive. None of those tasks is as straightforward as it is with other, established organizational roles. Start with the fact that there are no university programs offering degrees in data science. There is also little consensus on where the role fits in an organization, how data scientists can add the most value, and how their performance should be measured.(Davenport T.H., and Patil D.J.,2012)

One of the main concerns for any firm is when, how and to whom they should market their products. Firms make marketing decisions based on how much they know about their customers and potential customers. They may choose to mass market when they do not know much. With more information, they may market directly based on some observed characteristics. We provide strong evidence that whether and how well a consumer is linked to existing customers is a powerful characteristic on which to base direct marketing decisions. Our results indicate that a firm can benefit from the use of social networks to predict the likelihood of purchasing. Taking the network data into account improves significantly and substantially on both the firm's own marketing "best practices" and our best efforts to collect and model with traditional data.(Hill, S., Provost, F., and Volinsky, C, 2006)

3. PROJECT RATIONALE

Recently it has been noticed by many Internet users that the Ads displayed on the various websites contains the links to the Online shopping websites and more precisely to the products they have recently viewed on them.

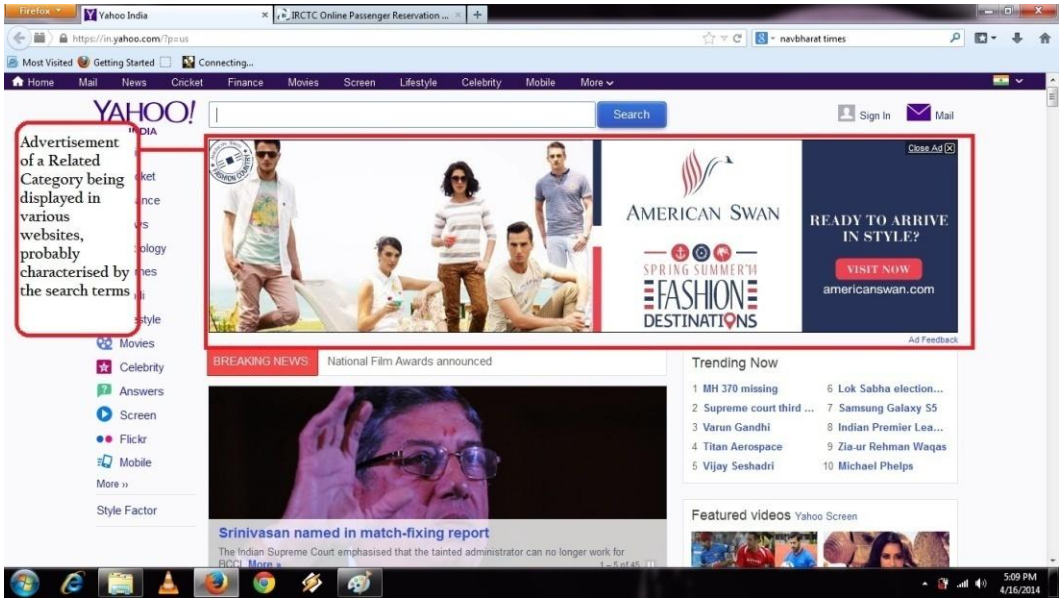

Figure 3.1 Advertisements of Related Merchandise

[*Source: Own Analysis*]

For example, If a user viewed a Wayfarer sunglass on Jabong.com and a blue shirt on myntra.com and made no purchases; the next time the user open a blog (S)he get to see the links to either the sunglass or the shirt popping up in the Ad space containing ads by Google.

Figure 3.2 Advertisement of Previously viewed products

[*Source: Own Analysis*]

Another user visited a site called Zoffio.com for the first time. The next day the user logged on to the Irctc (Indian Railways) website to book a ticket and the Ad for zoffio.com was displayed on the screen. This meant that Google was not only monitoring the user's viewed webpages and products displayed on them, but also keeping a record of user's browsing history.

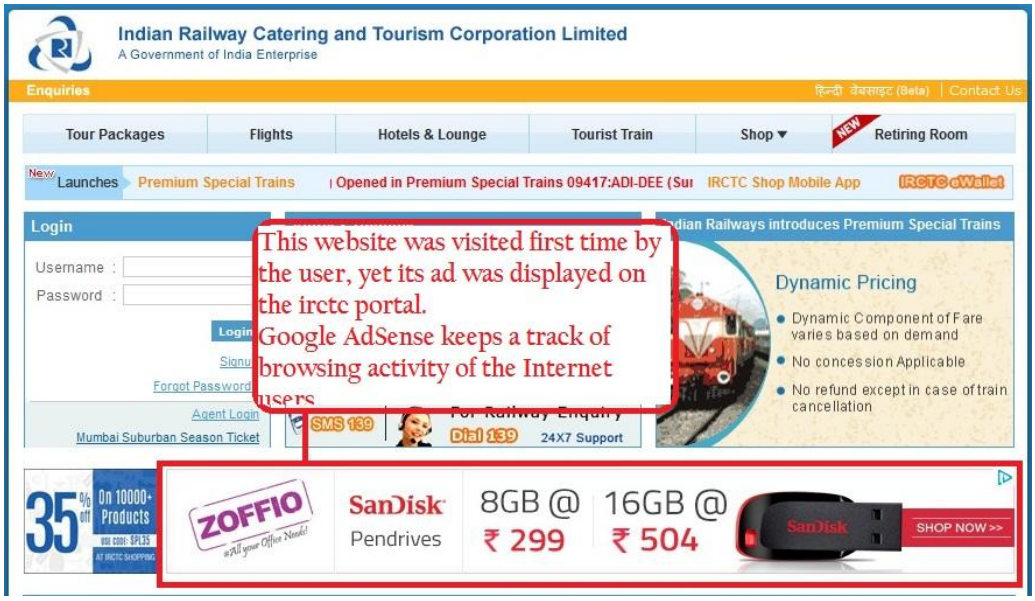

Figure 3.3 Advertisement of a website visited for the first time

[*Source: Own Analysis*]

Apart from these two incidents, another commonly witnessed phenomenon by Google search engine users is that they find Ads related to their search query popping up many times. Users that visit ecommerce portals for getting Top-ups, 3G packs and rate cutters recharged on their mobile connections get to see Ads from Vodafone displaying value offers surfacing frequently in the Google Ad space on Facebook (a major social media portal).

Figure 3.4 Advertisement related to browsing history

[*Source: Own Analysis*]

All such incidences lay open a research opportunity of understanding how Google Ads were so much in relevance to an internet user's online activities which formed the basis for this research work. A prominent research in this regard would definitely imply the strength of Big Data analytics in understanding the customer psyche and advertising according to it.

4. PROBLEM STATEMENT

Discussions about the dynamic between advertising and utilization of IT have often centered around the technologies being implemented in this age of digital engagement. However, as organizations realize the importance of delivering on a more customer-centric overall business strategy, big data has emerged as the real glue that has permanently cemented the relationship between the two departments.

Both marketing and IT believe that customer centricity must start with a corporate culture that focuses all strategies and programs around the needs of the customer. It is clear that both Marketers and Knowledge Workers see data—or more specifically, big data's role in delivering deep insights, understanding and intelligence about customers, markets and operational efficiencies—as critical to success.

Marketers agree that there is no shortage of data; in fact, data is seemingly everywhere and for many, the lure of collecting everything has led to a scenario where data repositories are overly bloated. Rather than fueling the enterprise with fact-based decision-making tools, the organization has slipped into a state of "analysis paralysis"; which means that companies are not able to draw the correct insights out of the available data, which in turn has led to raising a doubt about the efficiency of Big Data Analysis in gaining productivity. In this world where data is everywhere, it is easy to see why both IT and marketing executives feel the strain.

This project aims at studying the efficiency of Google AdSense, one of the products of Google which is operating on the amount of available data related to an Internet user and checking that whether the organization has been successful in deploying their analytical tool by reaching out to the consumer with more relevant Ads or not.

4.1 Research Scope does NOT cover

- Study of E-commerce websites
- Understanding the working model of AdSense or AdWords, and how they earn profits

- Any form of insights about SEO, PPC, Buzz Marketing
- Effectiveness of Online Advertising mediums other than Google Ads
- Online Shopper's buying pattern through other mediums than Google Ads
- Use of Big Data for Data Driven Decision making at a corporate scale
- Any other form of Advertising efficiency at B2B level

4.2 Objectives of the Research

- Measuring the relevance of the Ads displayed to internet users
- Understanding the alignment between nature of Ads displayed to users and their browsing history
- Gathering their response to such Ads, and judging the success of AdSense in luring the customers
- Correlating the success of Google AdSense to the success of Data Driven Marketing as a whole

4.3 Managerial Usefulness of the Project

This project help us in gaining an insight on the current level of success that Big Data has gained in analyzing customer needs and desires. Most of the customer-centric organizations are striving hard to garner insights about their consumers so as to increase the customer satisfaction level. Big Data has been a revolutionary move in this direction, but even after almost a decade since the introduction of this science ("Data Science") there have been no fixed algorithms to obtain accurate information out of the available data.

Most of the formulae devised for extracting knowledge out of data are generally adopted on an experimental basis, and so results might not be always positive when aligned to the desired outcomes.

This project is a modest approach towards understanding the levels of alignment of the efforts put in by a global organization and the results that are obtained in the form of

reactions from the end users. So, even though the research is based on the findings from end users of a service, it is directed towards understanding a bigger picture not restrained to a tool, an organization or an industry, but to the "data science" as a whole.

4.4 Data Collection

For the research purpose, both primary and the secondary sources have been used.

4.4.1 Primary Research

The primary source includes getting the questionnaire filled by various respondents using the internet. The questionnaire collected information about the demographics of the respondents, and several questions related to the project directed towards obtaining the views of the respondents on a five-point likert scale.

Sample size has been 85 respondents across the various demographics.

4.4.2 Secondary Research

Secondary data have been collected through reports of various agencies that are working towards development of this science through both business and academic viewpoints. The published findings, from agencies like SAS, CMO council, DOMO with the aid of academic institutions like Ivey league universities, INSEAD, etc., provided a useful source of information. Various articles published in Howard Business Review, and the periodicals by Gartner Inc. were also useful sources of information.

These are good especially for up-to-date information. They are frequently used in literature reviews because they offer a relatively concise, up-to-date format for research.

The fastest-growing source of information is on the Internet. But it too has its own advantages and disadvantages since anyone can post information on the Internet so the quality may not be reliable. Also, the information found may be intended for a general audience and so may not be suitable for inclusion in the literature review. However, many electronic journals (e-journals) are appearing on the Internet - if there is an editorial board that evaluates the work before publishing it in their e-journal, the quality can be more reliable.

By far the most widely used research method is collecting data through secondary research. This process involves collecting data from either the originator or a distributor of primary research. In other words, accessing information that others have already gathered.

Compared to primary research there are several advantages to the secondary approach, including ease of access and generally lower cost for acquiring the information.

Sources for secondary research are quite extensive. However, the Internet has changed how secondary research offering convenience, accessibility to a large array of information sources accesses, and generally low cost.

The project will be a result from contributions from various secondary sources which includes books, reference materials and websites.

4.5 Sample Design

A Sample Design is a definite plan for obtaining a sample from the population of Internet Users. It refers to the technique to the procedure adopted in selecting items for the sampling designs are as below:

4.5.1 Sample Size:

The sample consists of N Internet Users, majority of the respondents are from India, while few respondents were based in Developed nations like Australia, USA and France.

4.5.2 Sampling Method:

The sampling method used is judgmental non-probabilistic sampling and to some extent the snowball sampling is also utilized. An avid internet user may be able to reason out the questions in a better manner.

5. DATA ANALYSIS

5.1 Objective I: Awareness about the Google Ads services among respondents

The first part was to find out the level of awareness respondents have about the technology and tools that Google employs to make the Ads relevant.

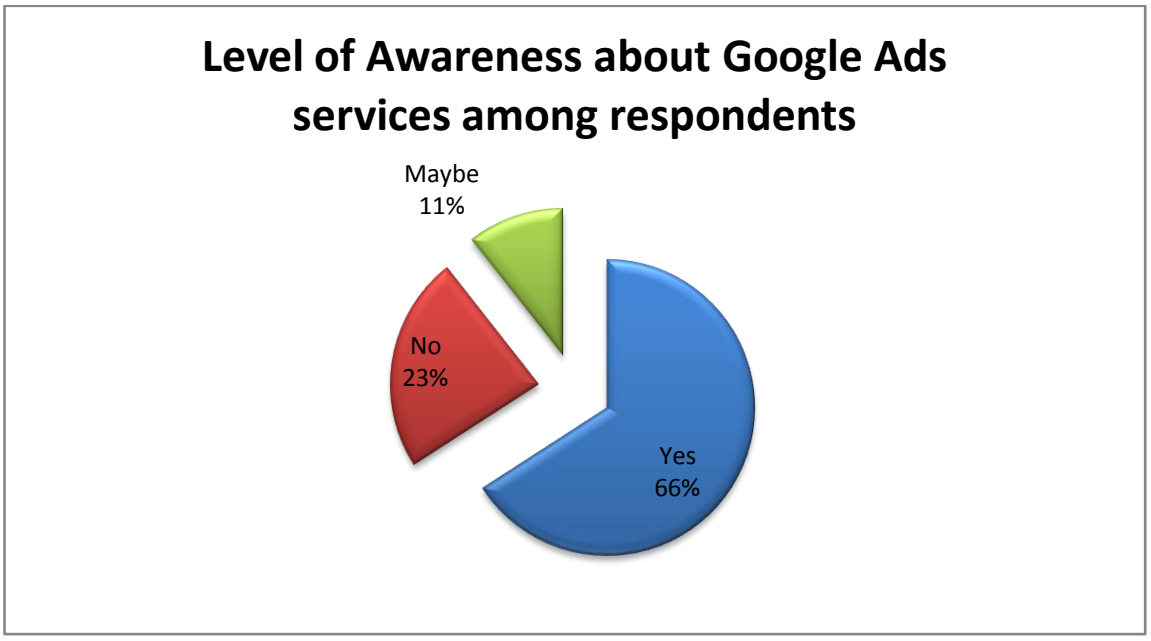

Figure 5. 1: Level of Awareness

[*Source: Own Analysis*]

INFERENCE

- 66% of respondents are aware of the services like Analytics, Double click and AdSense, Hence we can assume that Google has promoted its product well and has created a decent level of awareness among Internet Users.
- 23% of respondents are unaware of the existence of such services and how they help in displaying relevant Ads to them.
- 11% of respondents are unsure if they know about the services.

5.2 Objective II: Are Google Ads successful in grabbing attention of the Internet Users?

A) Likeness of noticing a Google Ad **B) Relation of Ads and user's search history**	**Result** **[1= highest impact** **5= lowest impact]**	**Interpretation**
Correlation (r) between A & B	0.62508221	The factors A and B are moderately correlated
Mean A	2.47058824	The likeability of noticing a Google Ad is good
SD A	1.37657473	There is a significant level of variation in the responses
Mean B	2.44705882	The users find the Ads somewhat related to their past search queries
SD B	1.20025674	There is low level of variation in the responses

Table 5.1: Ads and Search History

INFERENCE

- Google Ads are effective in gaining attention of the Internet users to a good extent and they are even related to the user's past search history
- There is a moderate level of correlation between the two phenomenon, which imply that the relation to user's search query does account for the attractiveness of the Google Ads but not to the full extent, so there might be other factors contributing towards the same

5.3 OBJECTIVE III: Relation between browsing history and Google Ads and associated effectiveness

A) Relation between user's recently viewed websites and their Ads **B) Likeability of visiting the sites again due to Ads**	**Result** **[1= highest impact 5= lowest impact]**	**Interpretation**
Correlation Between A & B	0.78095591	There is a high level of correlation
Mean A	2.30588235	Users often come across Ads of websites they have viewed recently
SD A	1.2912114	There is a significant level of deviation in the responses
Mean B	2.35294118	Users are likely towards visiting the websites again
SD B	1.24121563	The variability in responses is moderate

Table 5.2: Ads and Browsing History

INFERENCE

- Respondents feel that they often get to view the Ads of websites they have viewed recently, so there is indeed a relation between their browsing history and AdSense0
- Most of the respondents are either likely towards visiting the same websites again using the link provided by the Google Ads
- There is significantly high level of correlation between the recently viewed websites and likeability of visiting them again through Google Ads, which imply that Google AdSense is successful in enticing target customers

5.4 Objective IV: Effectiveness of Google Ads in Sales conversion with respect to Ecommerce portals

A) Frequency of online shopping **B) Frequency of Ads of recently viewed products being displayed** **C) Likeability of buying those products**	**Result** **[1= highest impact** **5= lowest impact]**	**Interpretation**
Correlation between B & C	0.62849453	There is a moderate level of correlation
Mean A	2.62352941	The respondents are almost regular online shoppers
SD A	1.02325892	There is a low level of variation in the buying trend of respondents
Mean B	2.50588235	Users frequently notice the Ads of recently viewed products
SD B	1.26889087	Low level of variation
Mean C	3.07058824	Respondents are neutral towards buying the product displayed
SD C	1.18108339	Low level of variation

Table 5.3: Ads and Online Shopping

INFERENCE

- The respondents are frequent online buyers, low level of variation in the responses imply that most of the respondents follow similar buying pattern
- The Ads of the products or related categories that respondents viewed in various Ecommerce portals get displayed very frequently

- However, the respondents are not very enthusiastic about buying those products even after frequent promotions of the same
- Hence, frequency of displaying of Ads of recently viewed products does not ensure that online shoppers will buy them. This is also evident statistically through the correlation coefficient which shows a moderate correlation between Ads and buying pattern of specific products

5.5 Objective V: Effectiveness of Ads containing Discount offers

A) Frequency of online shopping **B) Relevance of Discount offers being advertised**	**Result** **[1= highest impact** **5= lowest impact]**	**Interpretation**
Correlation AB	0.52731771	Moderate level of correlation
Mean A	2.62352941	The respondents are almost regular online shoppers
SD A	1.02325892	There is a low level of variation in the buying trend of respondents
Mean B	2.74117647	The respondents find the Ads of offers useful
SD B	1.07074675	There is a low level of variation in the perception of respondents towards discount offer Ads

Table 5.4: Ads and Online Offers

INFERENCE

- The respondents are routine Online Shoppers and they even find the discount offers and promotions being advertised to be somewhat useful
- There is a moderate level of correlation between Online shopping behavior and users interest in Ads containing discount offers, this implies that Online shoppers may find such Ads useful

5.6 Objective VI: Relevance of Google Ads as perceived by Internet Users

A) Likeness of noticing a Google Ad **B) Level of relevance of Google Ads perceived by respondents**	**Result [1= highest impact 5= lowest impact]**	**Interpretation**
Correlate A & B	0.69443362	The factors A and B are moderately correlated
Mean A	2.47058824	The likeability of noticing a Google Ad is good
SD A	1.37657473	There is a significant level of variation in the responses
Mean B	1.89411765	Respondents find the Google Ads to be highly relevant
SD B	0.97618706	There is a considerably small level of variation in the responses

Table 5.5: Relevance of Google Ads to users

INFERENCE

- Google is successful in gaining the attention of Internet users as there is a good level to which the Ads are being noticed by the respondents
- Respondents find the Ads to be highly relevant; they perceive the Ads as a useful tool through which they can get what they are searching for
- There is a moderately high level of correlation between the level to which respondents notice the Google Ads and their relevance; hence it is a win win situation for both the users and company

6. FINDINGS

An important facet of research is the assessment of effectiveness of Google AdSense and in turn understanding how Big Data is changing the way the various business processes, including marketing and advertising, is being done nowadays. Big Data Analytics has been employed in decision making by the various organizations in International market for years, but now with improved algorithms Big Data is also being used for demand forecasting (WalMart), creation of relevant services and features (Facebook) and Online Advertising (Google).

Through the new changes in the Analytical tools, and with the acquisition of DoubleClick, Google tries to understand the consumer online buying behavior in a detailed manner and displaying more relevant Ads to the Internet users. But, it is essential to study the effectiveness of the major performance indicators related to AdSense:

- From the research, it is figured out that a large percentage of Internet users are either fully or partially available of the Google's analytical tools, this means that Google has promoted its services in an effective manner.

- Most of the Internet users often notice the Ads, so Google is also successful in reaching out to the audience, also the users can relate the Ads to their search history and hence in the whole process they notice the Ads of products or services they were searching for.

- The respondents unanimously state that they have witnessed Ads of websites they have visited recently being displayed very often, and they even visit the websites again by using the link in the advertisement.

- Most of the respondents were regular online shoppers and they have noticed the Ads of products they have recently viewed in several E-commerce portals being displayed to them, and there is a chance that they may purchase the product on being reminded.

- But there are also many online shoppers who are not interested in buying the product, so Google Ads find moderate level of success in this aspect of displaying Ads of relevant products.
- On a similar note, shoppers find the Ads containing discount offers to be moderately interesting. One probable reason for this may be that shoppers seek out for discount coupons only from certain premium websites, as there are many ecommerce portals in India that offer discounts on their own merchandise on a year round basis, which does not hold much value to a regular shopper as they start considering such websites to be a hoax selling substandard product.
- However, the respondents do find the Ads to be relevant, informative and interesting.

7. RECOMMENDATIONS AND CONCLUSIONS

7.1 Conclusion

Google's strength is in its innovation. Its strength in innovation comes from the immense efforts company puts in for the R&D of new technologies and features which are unique when compared to products and services offered by competitors. Google provides services like AdSense and AdWords, which are backed by tools like Google Analytics and DoubleClick. They are the leading innovators in digital marketing tools and data analytics.

Google's greatest differentiator is in the diverse mix of their business portfolio and a lineup of mass clientele as well as the company's ability to produce what people need throughout the world, emotionally, aesthetically, and professionally. And it can be confidently said that as the Ecommerce segment in India will mature, Google Ads would become more relevant and useful for both online and offline shoppers.

The industry is changing fast. This means that the company has to think ahead, something that sounds easy but involves great risks. Huge sums must be invested in uncertain in-house research and development and/or must go toward open source developers and development tools. As companies develop their new marketing tools, they must be mindful of changes caused by the changing social norms, culture and demographics in usage of social media forums, websites, search engines and Ecommerce portals in countries all over the globe. The global competitive environment creates challenges and opportunities for the companies, and to understand such a diverse user population, Google has successfully implemented Big Data Analytics tools in understanding the needs and desires of its users. Hence, the victory of Google can be directly related to the power of Big Data and the immense opportunity that it brings along.

7.2 Recommendation

The main challenges for Google come from four areas. First, they must deal with changing demographics, evolving ecommerce sector and maturing online advertising mechanism. Second, they must create user acceptance for their product, Internet users should not treat the Ads as spam or be suspicious of them being malware. Third, Google must be constantly developing new algorithms of data assessment to encash any new and growing market segment. Finally, they must periodically measure the effectiveness of the AdSense.

Lastly, in terms of Big Data Analytics, every data analytics tool should regularly be updated with more accurate functions to improve its efficiency in extracting meaningful information from the huge pile of unrelated data. The Data Science still has realms of undiscovered opportunities and Data scientists are still working towards finding newer improved ways to use the data in getting accurate information relating to almost every domain. So when it comes to advertising by using data science, still a lot of other factors are required to be taken into consideration like study of 'social media' behavior of the internet users and the effective 'timing' when they can perform impulse buying. These two, along with many other such factors, can help in making the digital advertising more relevant and useful to the internet users.

8. LIMITATIONS AND FUTURE SCOPE OF THE STUDY

8.1 Limitations

- Research was conducted mostly from the respondents belonging to similar demographic background, so analysis and finding may not be representative of the actual facts.

- Non-Probability Sampling Technique was used for data collection so estimation of sampling errors was not possible. This leads to exclusion bias on the results based on the fact that how much information a sample can provide about population.

- Since Google AdSense has been updated with new features discussed in the report, the technical in depth information about how it is actually happening is still unknown.

- The current updates in the Google AdSense program made many of the previously conducted researches, and available secondary data absurd.

8.2 Future scope of the Study

- There are millions of Internet users in the world, so a more descriptive research covering a diverse sample from every strata of the population will make this research more meaningful and accurate.

- Apart from Google's widely used tools, other Data Analytic tools being used in the industries and their outcomes should also be studied to assess the Impact of Big Data on businesses in a more insightful manner.

- The process under consideration in this project is just Digital Advertising and measurement of its effectiveness, however, there are many other processes that

Big Data enables like Demand forecasting, decision making, requirements gathering, so a detailed study in other verticals can also be conducted.

Also, the E-commerce industry in India is still in its initial phase so the direction in which the business is moving is not predictable, so a lot of scope of study will arise as the business or the industry as a whole will mature because it will generate more scope for digital advertising means.

7. BIBLIOGRAPHY

7.1 Reference Books

- Fomitchev, M. I. (2010), How google analytics and conventional cookie tracking techniques overestimate unique visitors, *in* 'Proceedings of the 19th international conference on World wide web' , ACM, New York, NY, USA , pp. 1093--1094 .
- Hill, S., Provost, F., and Volinsky, C. Network-based marketing: Identifying likely adopters via consumer networks. Statistical Science, 21(2), 2006, pp. 256–276.

7.2 Scholarly Articles

- Laney, D. (2001), '3D Data Management: Controlling Data Volume, Velocity, and Variety' , Technical report, META Group .
- Davis, H. (2010), *Google Advertising Tools - Cashing in with AdSense, AdWords, and the Google APIs (2. ed.).* , O'Reilly .
- Aimiuwu, E. E.; Bapna, S. & Ahmed, A. (2012), How Web Search and Social Media Affect Google AdSense Performance., *in* 'AMCIS' , Association for Information Systems, .
- Cutroni, Justin. *Google Analytics - Understanding Visitor Behavior.*. : O'Reilly, 2010.
- Web Analytics: An hour a day (Wiley), Avinash Kaushik Conor Mc Guckin and Niall Crowley. Cyberpsychology, Behavior, and Social Networking. November 2012, 15(11): 625-629. doi:10.1089/cyber.2011.0460.
- Mulvenna, M. D.; Norwood, M. & Büchner, A. G. (1998), 'Data-Driven Marketing.', *Electronic Markets* **8** (3).
- Chen, Y.; Fay, S. & Wang, Q. (2011), 'The Role of Marketing in Social Media: How Online Consumer Reviews Evolve', *Journal of Interactive Marketing* **25** (2) , 85 - 94 .
- Cha, M.; Mislove, A. & Gummadi, K. P. (2009), A measurement-driven analysis of information propagation in the flickr social network, *in* 'WWW '09:

Proceedings of the 18th international conference on World wide web' , ACM, New York, NY, USA , pp. 721--730 .

- Martens, David and Provost, Foster, Pseudo-Social Network Targeting from Consumer Transaction Data (September 2011). Faculty of Applied Economics, University of Antwerp, Belgium; Departmentof Information, Operations a, Vol. , pp. -, 2011. Available at SSRN: http://ssrn.com/abstract=1934670
- Hill, Shawndra; Provost, Foster; Volinsky, Chris. Network-Based Marketing: Identifying Likely Adopters via Consumer Networks. Statistical Science 21 (2006), no. 2, 256-276.doi:10.1214/088342306000000222. http://projecteuclid.org/euclid.ss/1154979826.
- Shah S., Horne A., and Capella´ J. Good data won't guarantee good decisions. Harv Bus Rev, Apr 2012.
- Hays C. L. What they know about you. N Y Times, Nov.14, 2004. Davenport T.H., and Patil D.J. Data scientist: the sexiest job of the 21st century. Harv Bus Rev, Oct 2012
- Hill, S., Provost, F., and Volinsky, C. Network-based marketing: Identifying likely adopters via consumer networks. Statistical Science, 21(2), 2006, pp. 256–276.

7.3 Website References

- SAS. (2014, 04 29). Retrieved from: http://www.sas.com/en_us/insights/big-data/what-is-big-data.html
- SAS. (2014.04.29). Retrieved from: http://www.sas.com/reg/gen/corp/2255911
- Mediapost. (2014.04.27). Retrieved from: http://www.mediapost.com/ommaddm/
- "Web Analytics & Reporting – Google Analytics.". (2014.04.27). Retrieved from: 2005.http://www.google.com/analytics/.
- Social Media Today. (2014.04.26). Retrieved from: http://socialmediatoday.com/jayson-bowden/2066591/reasons-explore-big-data-social-media-analytics