Chapter 21
Social Media and Web Analytics

Vishnuprasad Nagadevara

1 Introduction

Social media has created new opportunities to both consumers and companies. It has become one of the major drivers of consumer revolution. Companies can analyze data available from the web and social media to get valuable insights into what consumers want. Social media and web analytics can help companies measure the impact of their advertising and the effect of mode of message delivery on the consumers. Companies can also turn to social media analytics to learn more about their consumers. This chapter looks into various aspects of social media and web analytics.

2 What Is Social Media and Web Analytics?

2.1 Why Is It Different? What All Does It Cover?

Social media analytics involves gathering information from social networking sites such as Facebook, LinkedIn and Twitter in order to provide businesses with better understanding of customers. It helps in understanding customer sentiment, creating

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-319-68837-4_21) contains supplementary material, which is available to authorized users.

V. Nagadevara (✉)
IIM-Bangalore, Bengaluru, Karnataka, India
e-mail: nagadevara_v@isb.edu

B. Pochiraju, S. Seshadri (eds.), *Essentials of Business Analytics*, International Series in Operations Research & Management Science 264,
https://doi.org/10.1007/978-3-319-68837-4_21

customer profiles and evolving appropriate strategies for reaching the right customer at the right time. It involves four basic activities, namely, listening (aggregating what is being said on social media), analyzing (such as identifying trends, shifts in customer preferences and customer sentiments), understanding (gathering insights into customers, their interests and preferences and sentiments) and strategizing (creating appropriate strategies and campaigns to connect with customers with a view to encourage sharing and commenting as well as improving referrals). One of the major advantages of social media analytics is that it enables businesses to identify and encourage different activities that drive revenues and profits and make real-time adjustments in the strategies and campaigns. Social media analytics can help businesses in targeting advertisements more effectively and thereby reduce the advertising cost while improving ROI.

On the other hand, web analytics encompasses the process of measuring, collecting and analyzing web traffic data to understand customer behaviour on a particular website. Web analytics can help in improving user experience leading to higher conversion rates. Companies can tweak the design and functionality of their websites by understanding how users interact with their websites. They can track user behaviour within the website and how users interact with individual elements on each page of the website. Web analytics can also help in identifying the most profitable source of traffic to the website and determining which referrals are important in terms of investing in marketing efforts. Google Analytics is probably the best tool available free of cost to any website owner for tracking and analyzing the traffic to their website. These analytics include sources of traffic, bounce rates, conversions, landing pages and paid search statistics. It is easy to integrate this with Google AdWords.

It is obvious that social media analytics is very different from web analytics. The data sources are different. These two complement each other and when used together in tandem can provide deep insights into the traffic patterns, users and their behaviour. For example, one can measure the volume of visitors to the website based on referrals by different social networks. By ranking these social networks based on the traffic generated, one can determine how to focus on the right networks. We can even determine the influencers in the networks and their behaviour on our website.

2.2 What Additional Information/Details Can It Provide?

Many companies have started leveraging the power of social media. A particular airline keeps the customers informed through social media about the delays and the causes for such delays. In the process, the airline is able to proactively communicate the causes for the delays and thereby minimize the negative sentiment arising out of the delays. In addition, the airline company is also able to save much of the time of its call centre employees, because many customers already knew about the delays as well as the reasons associated with the delays and hence do not make calls to the call centre.

Social media analytics can be effectively used to gauge customers' reaction to a new product or service introduced into the market. Social media gives an opportunity to listen to the person on the street. The simplest way to do this is to scan various social media sites such as Twitter, Facebook or LinkedIn or various discussion forums and blogs. By analyzing these messages and blogs, the company can understand the customer perceptions about the new product or service. It is also possible to analyze the trend of customer perceptions overtime as well as the response to various advertising and marketing campaigns. This response can be measured almost on a real-time basis.

Search engine optimization (SEO) is another technique to acquire customers when they are looking for a specific product or service or even an organization. For example, when a customer initiates a search for a product, say a smartphone, there is a possibility of getting overloaded and overwhelmed with the search results. These results contain both "paid listings" and "organic listings". Paid search listings are advertisements that businesses pay to have their ads displayed when users do a search containing specific keywords. On the other hand, "Organic search listings" are the listings of web pages returned by the search engine's algorithms that closely match the search string. Companies can make use of SEO in order to get their websites listed higher in the search results, especially higher among the organic listings. They can also initiate online advertisements and achieve higher visibility.

It is worthwhile for the companies to monitor the postings on various social media. One can use the posts as well as various blogs and reviews to constantly monitor customer opinions through opinion mining and sentiment analysis. Constant analysis of the opinions and sentiment analysis can help companies not only to understand the customer sentiments but also to develop strategies to further promote positive sentiment or carry out damage control in case of negative sentiment. Companies can also learn from negative comments to improve their products and services. In addition, they can monitor customers' response to the corrective actions taken by them.

Websites can use cookies as well as IP addresses to identify past history of customers and their preferences. Based on past history and browsing habits, customers can be served appropriate ads. For example, if we can predict that the customer likes outdoor sports, he or she can be targeted with an offer for a trekking package. The predictions can be made even with respect to a new arrival on the website. For example, if the customer has a Facebook account and arrives through viewing an ad on Facebook, we can identify his or her hobbies and interests and present appropriate ads.

It has become very common for customers to look at product (or service) reviews and recommendations that are posted on various websites, before making a decision. Many websites (such as TripAdvisor) present an analysis of the reviews as well as a summary of the reviews to help customers make the appropriate decision. While presenting the summary of the reviews, companies can make appropriate offerings to facilitate quick acceptance.

Analysis of frequently used keywords on search engines can help in identifying the current issues and concerns. It can reveal public sentiment towards emerging

issues as well as public attitude towards political parties. It can help in identifying pockets of political support. It can also help in creating appropriate campaigns and policy formulation to take full advantage of the prevailing sentiments.

It is very common to have leaders and followers in social media. By using network analytics (such as network and influence diagrams), organizations can identify the most influential persons in the network. Organizations can take the help of such influential persons to promote a new product or service. They can be requested to contribute an impartial review or article about the product or service. Such a review (or a recommendation) can help in promoting the new product or service by resolving uncertainty or hesitancy in the minds of the followers.

Network and influence diagrams are used very effectively in identifying key players in various financial frauds or drug laundering cartels. One such study identified the perpetrators of 9/11 attacks by using a network diagram based on the information available in the public domain. The network diagram could clearly identify the key players involved in 9/11 starting from the two Cole bombing suspects who took up their residences in California as early as 1999[1].

Social media analytics can also be used for public good. Data from mobile telephones are used to identify and predict traffic conditions. For example, Google Traffic analyzes data sourced from a large number of mobile users. Cellular telephone service providers constantly monitor the location of the users by a method called "trilateration" where the distance of each device to three or more cell phone towers is measured. They can also get the exact location of the device using GPS. By calculating the speed of users along a particular length of road, Google generates a traffic map highlighting the real-time traffic conditions. Using the existing traffic conditions across different routes, better alternative routes can be suggested.

With the advent of smartphones, mobile devices have become the most popular source for consuming information. This phenomenon has led to the development of new approaches to reach consumers. One such approach is geo-fencing. It involves creating a virtual perimeter for a geographical area and letting companies know exactly when a particular customer (or potential customer) is likely to pass by a store or location. These virtual perimeters can be dynamically generated or can be predefined boundaries. This approach enables companies to deliver relevant information or even pass on online coupons to the potential customer. The concept of geo-fencing can be combined with other information based on earlier search history, previous transactions, demographics, etc. in order to better target the customer with the right message.

In the rest of the chapter, we describe some applications in greater detail: display advertising in real time, A/B experiments for measuring value of digital media and handling e-retailing challenges, data-driven search engine advertising, analytics of digital attribution and strategies and analytics for social media and social enterprises and mobile analytics and open innovation.

[1]Valdis Krebs, "Connecting the Dots; Tracking Two Identified Terrorists" available at http://www.orgnet.com/tnet.html (last accessed on Jan 18, 2018).

3 Display Advertising in Real Time

The Internet provides new scope for creative approaches to advertising. Advertising on the Internet is also called online advertising, and it encompasses display advertisements found on various websites and results pages of search queries and those placed in emails as well as social networks. These display advertisements can be found wherever you access the web. As in the case of any other mode of advertising, the objective of display advertisement is to increase sales and enhance brand visibility. The main advantage of display advertisements is that all the actions of the user are trackable and quantifiable. We can track various metrics such as the number of times it was shown, how many clicks it received, post-click-and-view data and how many unique users were reached. These display advertisements can be all pervasive and be placed anywhere and on any type of web pages. They can be in the form of text, images, videos and interactive elements. Another advantage of display advertisements is that the conversion and sales are instantaneous and achieved with a single click.

There are different types of display advertisements. The most popular one is the banner advertisement. This is usually a graphic image, with or without animation, displayed on a web page. These advertisements are usually in the GIF or JPEG images if they are static, but use Flash, JavaScript or video if there are animations involved. These banner advertisements allow the user to interact and transact. These can be in different sizes and can be placed anywhere on the web page (usually on the side bar). You need to carry out appropriate tests or experimentation (refer to the section on Experimental Designs in this chapter) to know what works best for you. You can design your banner as a single- or multiple-destination URL(s).

There are banners that appear between pages. As you move from one page to the next through clicks, these advertisements are shown before the loading of the next page. These types of advertisements are referred to as interstitial banners.

The display advertisements can be opened in a new, small window over the web page being viewed. These are usually referred to as pop-ups. Once upon a time, these pop-up advertisements were very popular, but the advent of "pop-up blockers" in the web browsers had diminished the effectiveness of the pop-up advertisements. Users can be very selective in allowing the pop-ups from preselected websites. There are also similar ones called pop-under where an ad opens a new browser window under the original window.

Some of the local businesses can display an online advertisement over a map (say Google Maps). The placement of the advertisement can be based on the search string used to retrieve the map. These are generally referred to as map advertisements.

Occasionally, an advertisement appears as a translucent film over the web page. These are called floating advertisements. Generally, these advertisements have a close button, and the user can close the advertisement by clicking on the close button. Sometimes, these advertisements float over the web page for a few seconds before disappearing or dissolving. In such a case, there is no click-through involved, and hence it is difficult to measure the effectiveness of such ads.

Websites are generally designed to display at a fixed width and in the centre of the browser window. Normally, this leaves considerable amount of space around the page. Some display advertisements take advantage of these empty spaces. Such advertisements are called wallpaper advertisements.

3.1 How to Get the Advertisements Displayed?

There are many options for getting the advertisements displayed online. Some of these are discussed below.

One of the most popular options is placing the advertisements on social media. You can get your ads displayed on social media such as Facebook, Twitter and LinkedIn. In general, Facebook offers standard advertisement space on the right-hand side bar. These advertisements can be placed based on demographic information as well as hobbies and interests which can make it easy to target the right audience. In addition to these ads, Facebook also offers engagement ads which will facilitate an additional action point such as a Like button or Share button or a button to participate in a poll or even to add a video. Sponsored stories or posts can also be used to promote a specific aspect of a brand. These stories or posts can appear as news feeds. You can even publicize an existing post on Facebook.

Twitter also allows advertisements. Some promotional tweets appear at the top of the user's time line. The section "Who to Follow" can also be used to have your account recommended at a price. Usually, the payment is made when a user follows a promoted account. The Trends section of Twitter is also available for advertisements. While this section is meant for the most popular topics at any particular time, the space is also available for full-service Twitter ads customers.

LinkedIn allows targeted advertisements with respect to job title, job function, industry, geography, age, gender, company name and company size, etc. These advertisements can be placed on the user's home page, search results page or any other prominent page.

Online advertisements can be booked through a premium media provider, just like one would book from a traditional advertising agency. A premium media provider usually has access to high-profile online space and also can advise on various options available.

Another option is to work with an advertising network. Here, a single sales entity can provide access to a number of websites. This option works better if the collection of websites are owned or managed by a common entity, such as HBO or Times Inc. The Google Display Network is another such entity. Usually, the advertising network can offer targeting, tracking and preparing analytic reports in addition to providing a centralized server which is capable of serving ads to a number of websites. This advertising network can also advise you based on various factors that influence the response, such as demographics or various topics of interest.

If you are looking for advertising inventory (unsold advertising space), advertising exchanges can help. The publishers place their unsold space on the exchange, and it is sold to the highest bidder. The exchange tries to match the supply and demand through bidding.

One of the fastest-growing forms of display advertising is mobile advertising. Mobile advertising includes SMS and MMS ads. Considering that mobile is an intensely interactive mass media, advertisers can use this media for viral marketing. It is easy for a recipient of an advertisement to forward the same to a friend. Through this process, users also become part of the advertising experience.

There are blind networks such as www.buzzcity.com which will help you to target a large number of mobile publishers. A special category of blind networks (can be called as premium blind networks) can be used to target high-traffic sites. The sites www.millennialmedia.com and www.widespace.com are good examples of premium blind networks.

Ad servers play a major role in display advertisement. These servers can be owned by publishers themselves or by a third party and are used to store and serve advertisements. The advantage of the ad servers is that a simple line of code can call up the advertisement from the server and display it on the designated web page. Since the ad is stored at one single place, any modifications are to be carried out at one place only. In addition, the ad servers can supply all the data with respect to the number of impressions, clicks, downloads, leads, etc. These statistics can be obtained from multiple websites. One example of a third-party ad server is *Google DoubleClick*.

3.2 Programmatic Display Advertising

Programmatic advertising is "the automation of the buying and selling of desktop display, video, FBX, and mobile ads using real-time bidding. Programmatic describes how online campaigns are booked, flighted, analyzed and optimized via demand-side software (DSP) interfaces and algorithms"[2].

Traditionally, it is assumed that online visitors are exposed to a display ad when they somehow arrive on a website. The targeting of the ad is based on limited knowledge with respect to the relevance of the content, geography, type of device, etc. The advertiser enters into an agreement with the publisher with respect to the number of insertions at a certain price. In other words, the display ad is bought in the old-fashioned way, negotiating the number of exposures/insertions and the price, and served to a large number of potential consumers with a fond hope that somehow it will reach the right people.

Today, programmatic display delivers display ads in real time with specific messages based on each individual consumer's profile and behaviour. With the

[2]Gurbaksh Chahal, Chairman & CEO of RadiumOne.

evolution of recent technologies, the focus is shifting to understanding each individual customer and exposing him or her to the right display. This process is driven by real-time technologies using specially designed algorithms. The result is programmatic display advertising.

There are opportunities to display billions of ads every day, and advertisers can use them effectively to improve conversion rates. The technology enables the advertiser to obtain immediate information about where the ad was displayed, to whom it was displayed and for how long. This helps the advertiser to analyze and review the display ad very quickly and take any action that is required, including stopping of the ad, if it is not performing as expected.

The main components of programmatic display advertising are as follows:

(a) Supply-Side Platform (SSP)

The SSP helps the publishers to better manage and optimize their online advertising space and advertising inventory. SSP constantly interacts with the ad exchange and the demand-side platform (DSP). Admeld (www.admeld.com) and Rubicon (https://rubiconproject.com/) are two examples of SSPs.

(b) Demand-Side Platform (DSP)

The DSP enables the advertisers to set and apply various parameters and automate the buying of the displays. It also enables them to monitor the performance of their campaigns. Turn (now Amobee, https://www.amobee.com/), AppNexus (https://www.appnexus.com/) and Rocket Fuel (https://rocketfuel.com/) are some of the DSPs.

(c) Ad Exchange

Ad exchanges such as Facebook Ad Exchange or DoubleClick Ad Exchange facilitate purchase of available display inventory through auctions. These auctions are automated and take place within milliseconds, before a web page loads on the consumer's screen. These enable the publishers to optimize the price of their available inventory.

(d) Publisher

Publishers are those who provide the display ad inventory.

(e) Advertiser

The advertiser bids for the inventory in real time depending on the relevance of the inventory.

Real-time bidding (RTB) is an integral process of programmatic display advertising and is described in Fig. 21.1.

Whenever a consumer makes a request for a web page and if there is available space for a display ad, the information regarding the consumer/visitor, context of the web page requested and earlier web behaviour is sent to an ad exchange through the publisher. The ad exchange auctions the available space to various DSPs, and the winning bid/display ad is passed on to the publisher. The ad is displayed on the consumer's screen while the requested page is being loaded. This process is completed within a few milliseconds so that there is no delay in the page loading adversely affecting the user experience.

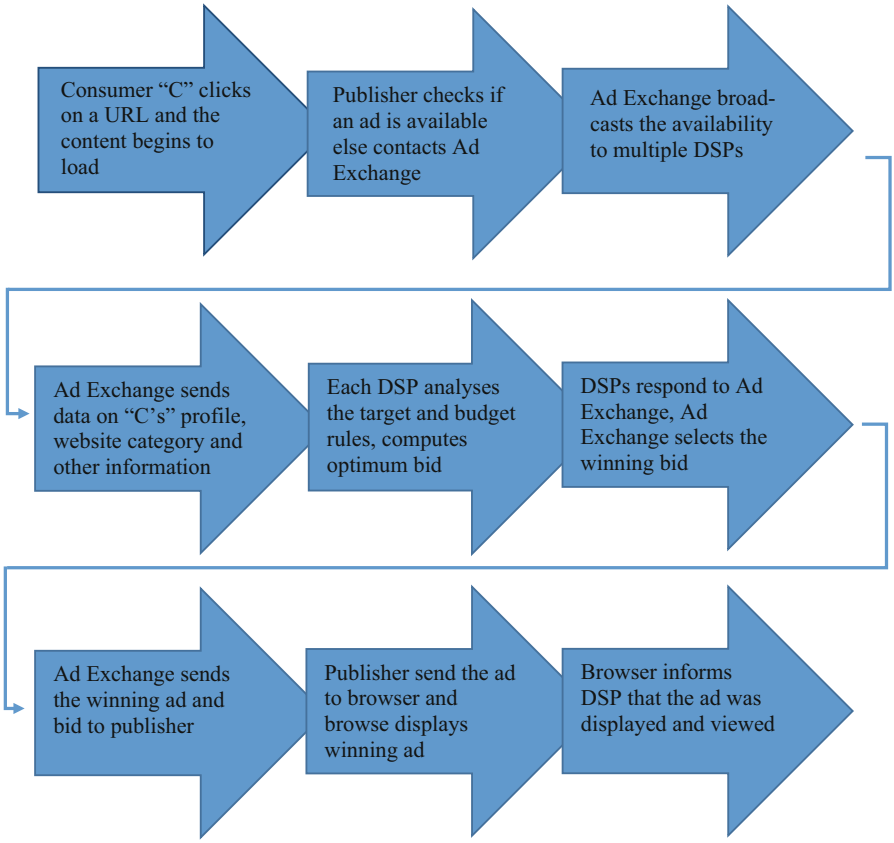


Fig. 21.1 Flowchart of real-time bidding

The entire process described above takes less than half a second. In other words, the entire process is completed and the display ad is shown while the browser is loading the requested page on the consumer's screen.

This process of matching the right display to the right consumer is completely data driven. Data with respect to the context, who is to see the display, the profile of the consumer, who is a good target, etc. is part of the process. In order to complete the process within a very short time span, it is necessary to build all the rules in advance into the system. The advertiser would have analyzed all the data and identified the appropriate profile, the possible number of exposures of the ad and the rules for bidding beforehand, based on his own analytics. The rules try to match the information about the context, profile of the consumer and space available received from the publisher with the requirement and trigger automatic bidding.

The programmatic display advertisement benefits both the publishers and advertisers. The advertisers benefit by effectively targeting only those who match the existing profiles. These profiles can be obtained by analyzing their own data or

from third parties. With the context built into the process, the advertiser can reach consumers who are browsing content that is most relevant to the product or service that the advertiser is offering. They can be very selective in terms of the website and the right context. They do not have to tie themselves to a pre-negotiated price and quantity. They will be paying only for the relevant exposure that was made to the most relevant target audience. The advertiser can analyze data with respect to visits to the website, bounce rates, earlier marketing efforts, etc. in order to improve the conversions. It also enables the advertiser to quickly review and revise the advertisement strategies instead of waiting for the entire campaign to be completed.

The publishers gain by maximizing the revenue through auctioning the available space. Each ad is auctioned to the highest bidder based on the context and consumer profile. It also allows them to optimize the available advertising space.

Programmatic display advertising opens up yet another opportunity in display advertising. The advertisers can use dynamic creative optimization (DCO) to improve the conversion rates. DCO involves breaking an ad into a number of components and creating different versions of each component. These components can be with respect to content, visuals, colours, etc. These components are then dynamically assembled to suit a particular consumer based on the context, profile, demographics as well as earlier browsing history. These details along with any other information available (such as time of the day or weather at that particular time) are fed into the DCO platform. The display ad is assembled based on this information, using predetermined rules before it is sent to the publisher's server. Thus, DCO can take advantage of the targeting parameters received from the ad exchange to optimally create (assemble) the appropriate ad.

It is expected that programmatic display advertising in the USA alone will reach more than $45 billion by 2019 (Fig. 21.2).

4 A/B Experiments for Measuring Value of Digital Media and Handling E-Retailing Challenges

Web technology companies such as Amazon, Facebook, eBay, Google and LinkedIn are known to use A/B testing in developing new product strategies and approaches. A/B testing (also called A/B splits or controlled experimentation) is one of the widely used data-driven techniques to identify and quantify customer preferences and priorities. This dates back to Sir Ronald A. Fisher's experiments at the *Rothamsted Agricultural Experimental Station* in England in the 1920s. It was called A/B splits because the approach was to change only one variable at a time. The publication of Ronald Fisher's book *The Design of Experiments* changed the approach where the values of many variables are changed simultaneously and the impact of each of the variables is estimated. These experimental designs use the concept of ANOVA (discussed in the earlier chapter) extensively, with appropriate modifications.

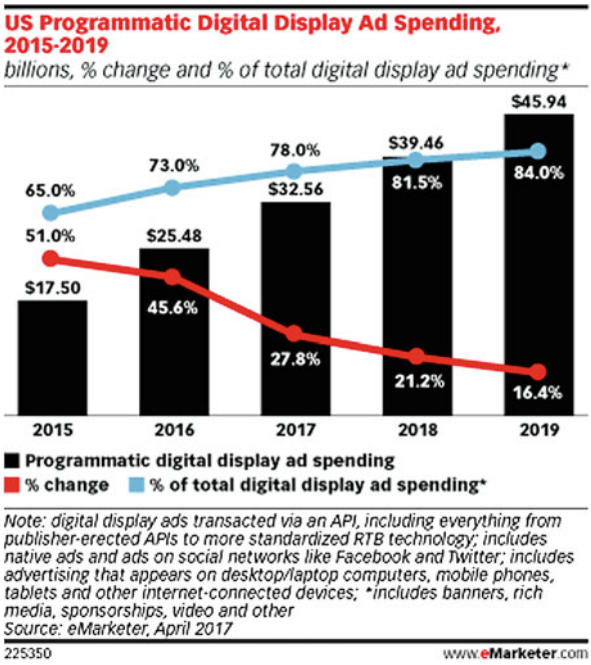


Fig. 21.2 US programmatic digital display ad spending

4.1 Completely Randomized Design

The simplest of the experimental designs is the completely randomized design (CRD). Internet experiments that present one of several advertising messages to users of search engines are typical examples of CRD. While there are a predetermined set of k advertisement messages offered to the users, each user is exposed to one and only one message. The response is measured by the hit ratio (proportion of those who click on the message to access the particular website in response to the message). Consider a scenario where an experiment is conducted to evaluate the effect of a display advertisement. Three different types of advertisements (namely, A, B and C) were designed and displayed randomly on five different search engines, and the experiment is carried for 3 weeks. The users are taken to a specific web page when the users click on the display advertisement. The number of visitors to the landing page is counted and used as a metric for the effectiveness of the display. Here, the three types of displays are called the treatments. A particular type of display is assigned randomly for a particular search engine in a given week for displaying to the users. The summary of the visitors based on the three different displays is presented in Table 21.1.

The above observations come from three different treatment groups, each group having five observations (replications). In general, each observation can be denoted by Y_{ij} where i represents the treatments and j represents the replication number. Even though the above example has an equal number of replications for each

Table 21.1 Number of visitors to the web page based on type of display

	Type of display		
	A	B	C
Number of visitors	2565	2295	2079
	864	2430	3051
	1269	2133	2619
	2025	1350	3105
	2241	864	3348
n	5	5	5
Average	1793	1814	2840
Variance	497,032	457,593	250,193
Number of clicks	574	592	501
	175	575	630
	240	449	507
	444	302	587
	405	160	547

treatment, it is not necessary to have an equal number of replications. The above experiment is called balanced because each treatment has an equal number of replications.

The above data can be analyzed for any significant difference between population means using ANOVA.

H_0: $\mu 1 = \mu 2 = \mu 3$.
H1: at least one μ is different.

The F test shown in Table 21.3 indicates that there is significant difference in the response to different displays.

The CRD can be represented as a regression model. Since everyone is familiar with regression analysis and many statistical packages have regression modules, representing and analyzing CRD as a regression model makes it easy to analyze and interpret the results.

The regression model for CRD is

$$Y_{ij} = \mu + \beta_i + \varepsilon_{ij}$$

where Y_{ij} is the response corresponding to ith treatment and jth replication

μ is the overall mean

β_i is the treatment effect and

ε_{ij} is the random error

Since the treatments are nominal variables, these are represented as dummy variables. As there are three treatments, these will be represented by two dummy variables. Type of Display "A" is represented by D1 = 1 and D2 = 0, "B" is represented by D1 = 0 and D2 = 1, and "C" is represented by D1 = 0 and D2 = 0. The data reformatted with dummy variables as required for the regression and the results of regression analysis are presented in Table 21.2 (a) and (b).

Table 21.2 (a) Reformatted data with dummy variables for the CRD experiment. (b) Regression results of the CRD experiment

(a) Reformatted data with dummy variables for the CRD experiment				
Yij	D1	D2		
2565	1	0		
864	1	0		
1269	1	0		
2025	1	0		
2241	1	0		
2295	0	1		
2430	0	1		
2133	0	1		
1350	0	1		
864	0	1		
2079	0	0		
3051	0	0		
2619	0	0		
3105	0	0		
3348	0	0		
(b) Regression results of the CRD experiment				
Regression statistics				
Multiple R	0.6531			
R square	0.4265			
Adjusted R square	0.3309			
Standard error	633.7240			
Observations	15			
	df	*SS*	*MS*	*F*
Regression	2	3,584,347	1,792,174	4.4625
Residual	12	4,819,273	401,606.1	
Total	14	8,403,620		
	Coefficients	*Standard error*	*t stat*	*P-value*
Intercept	2840.4	283.41	10.0222	0.0000
D1	−1047.6	400.8022	−2.6138	0.0226
D2	−1026.0	400.8022	−2.5599	0.0250

It can be seen that the ANOVA table calculated in Table 21.3 is identical to the ANOVA table obtained from the regression analysis. In addition, the regression coefficients corresponding to the dummy variables D1 and D2 are negative and statistically significant. This implies that Display C (which was left out in creating the dummy variables) has resulted in significantly higher visitors than Displays A and B. The intercept and other regression coefficients can be interpreted as the differences between mean responses of the three displays. The pairwise differences in the treatment effects can be obtained by post hoc tests.

Table 21.3 ANOVA table

Source of variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	3,584,347	2	1,792,174	4.4625	0.0356	3.8853
Within groups	4,819,273	12	401,606			
Total	8,403,620	14				

Table 21.4 Click-through rates

	Block (search engine)					
Treatment	SE1	SE2	SE3	SE4	SE5	Average
A	0.2236	0.2021	0.1892	0.2193	0.1806	0.2030
B	0.2580	0.2365	0.2107	0.2236	0.1849	0.2227
C	0.2408	0.2064	0.1935	0.1892	0.1634	0.1987
Average	0.2408	0.2150	0.1978	0.2107	0.1763	0.2081

4.2 Randomized Complete Block Design

In the above experiment, it is possible that there is an effect of the search engine, in addition to the effect of the type of display. In other words, there are two sources of variation, the type of display and the search engine. Since the displays are randomly assigned to the search engines and weeks, it is possible that all the 3 weeks selected for the first search engine could have been assigned Display A, while no week is assigned for Display A for the second search engine. It is necessary to run the experiment in "blocks" if we need to isolate the effect of display as well as the effect of the search engine. Such a design is called the randomized complete block design. In other words, the experiment should be run in blocks such that the three types of display are tested on each search engine on each of the weeks. When the visitors come to the landing page, they can "click" on the page to obtain additional information. The number of clicks is recorded, and the click-through rate is calculated as Click-through rate (CTR) = Number of clicks/Number of visitors.

The design and the results are shown in Table 21.4.

In a general randomized block design, there are k treatments and b blocks. The observations are represented by Y_{ij} where i represents the treatments and j represents the blocks.

The above data can be analyzed for differences between means by ANOVA, except that the sums of squares corresponding to treatments and blocks have to be estimated separately. The formulae for calculating these sums of squares are presented as follows:

$$SS(Treatment) = b\sum_{i=1}^{k}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2$$

$$SS(Block) = k \sum_{i=1}^{b} \left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2$$

$$SS(Total) = \sum_{i=1}^{k} \sum_{j=1}^{b} \left(Y_{ij} - \overline{Y}_{..}\right)^2$$

$$SS(Error) = SS(Total) - SS(Treatment) - SS(Block)$$

$$SS(Error)\ can\ also\ be\ calculated\ as\ \sum_{i=1}^{k} \sum_{j=1}^{b} \left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}\right)^2$$

where

$$Y_{ij}\ is\ the\ measurement\ corresponding\ to\ ith\ treatment\ and\ jth\ block,$$

$$\overline{Y}_{i.}\ is\ the\ average\ of\ ith\ treatment,$$

$$\overline{Y}_{.j}\ is\ the\ average\ of\ jth\ block\ and$$

$$\overline{Y}_{..}\ is\ the\ overall\ mean.$$

The degree of freedom for treatment is $k - 1$ and that of block is $b - 1$. While the total degree of freedom is $kb - 1$, the error degree of freedom is $(k - 1)(b - 1)$.

The ANOVA table for the above data is presented in Table 21.5.

It can be concluded from the ANOVA table that there are significant differences in responses between types of display as well as the search engines. The statistical significance is much higher (p-value = 0.0008) in the case of block (search engine) than that of treatment (type of display).

Table 21.5 ANOVA table for the experiment data

ANOVA						
Source of variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Treatment	0.0016	2	0.0008	7.6239	0.0140	4.4590
Block	0.0067	4	0.0017	15.5385	0.0008	3.8379
Error	0.0009	8	0.0001			
Total	0.0092	14				

The regression model for the randomized complete block design can be represented by

$$Y_{ij} = \mu + \beta_i + \pi_j + \varepsilon_{ij}$$

where Y_{ij} is the response for *i*th treatment and *j*th block
μ is the mean
β_i is the effect of Treatment *i*
π_j effect of Block *j* and
ε_{ij} is the random error

The coding of the above data for regression analysis is shown in Table 21.6. The dummy variables corresponding to the treatments (types of display) are represented by T_i, and those corresponding to blocks (search engines) are represented by SE_j.

As usual, the dummy variables corresponding to T3 and SE5 are omitted. The regression output is presented in Table 21.7.

It can be seen that the regression sum of squares is equal to the sum of the "treatment sum of squares" and the "block sum of squares". Consequently, the "F" value is different. It can also be seen that all the treatment effects and block effects are significant indicating that the effects of T1 and T2 are significantly better than that of T3. Similarly, the effects of the first four search engines (SE1 to SE4) are significantly better than that of SE5. The pairwise comparisons can be obtained by running post hoc tests.

Table 21.6 Coding for regression analysis

		Y	T1	T2	SE1	SE2	SE3	SE4
A	SE1	0.2236	1	0	1	0	0	0
B	SE1	0.2580	0	1	1	0	0	0
C	SE1	0.2408	0	0	1	0	0	0
A	SE2	0.2021	1	0	0	1	0	0
B	SE2	0.2365	0	1	0	1	0	0
C	SE2	0.2064	0	0	0	1	0	0
A	SE3	0.1892	1	0	0	0	1	0
B	SE3	0.2107	0	1	0	0	1	0
C	SE3	0.1935	0	0	0	0	1	0
A	SE4	0.2193	1	0	0	0	0	1
B	SE4	0.2236	0	1	0	0	0	1
C	SE4	0.1892	0	0	0	0	0	1
A	SE5	0.1806	1	0	0	0	0	0
B	SE5	0.1849	0	1	0	0	0	0
C	SE5	0.1634	0	0	0	0	0	0

Table 21.7 Results of regression analysis

Regression statistics					
Multiple R	0.9520				
R square	0.9063				
Adjusted R square	0.8361				
Standard error	0.0104				
Observations	15				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	0.0084	0.0014	12.9003	0.0010
Residual	8	0.0009	0.0001		
Total	14	0.0092			
	Coefficients	*Standard error*	*t stat*	*P-value*	
Intercept	0.1668	0.0071	23.4828	0.0000	
T1	0.0043	0.0066	0.6537	0.5316	
T2	0.0241	0.0066	3.6608	0.0064	
SE1	0.0645	0.0085	7.5955	0.0001	
SE2	0.0387	0.0085	4.5573	0.0019	
SE3	0.0215	0.0085	2.5318	0.0352	
SE4	0.0344	0.0085	4.0510	0.0037	

4.3 Analytics of Multivariate Experiments

The real power of experimental designs is felt when we have to estimate the effects of a number of variables simultaneously. For example, consider a scenario where a company is contemplating a particular display advertisement. They have identified three different variables (factors), each having two levels, to test. These are font (traditional font vs. modern), background colour (white vs. blue) and click button design (simple "Okay" vs. large "Click Now to Join"). The ad copies are randomly displayed to each viewer, and the conversion rate (defined as those who click to reach the website and join as members (free of cost)) is calculated. In the traditional experimental design, such as CRD, we will first decide which one is likely to be most important. Let us say the click button is the most important. Then, we would combine each of the two levels of click button with one of the other two factors (say, traditional font and white background) and run four replications. This actually involves eight runs (four each of simple button + traditional font + white background and large button + traditional font + white background). The resulting conversion rate can be used to decide which type of click button is more effective (say, large button). Now, we will select the background colour for experimentation. Since we already have the combination with white background, we will now select blue background and combine it with "large button" (since it was more effective) and traditional font and run four replications. Suppose the results show that blue background is more effective. Now, we select the combination of large button and

Table 21.8 Coding factorial design

Factor	Level 1 (code)	Level 2 (code)
Font	Traditional (−1)	Modern (+1)
Background colour	White (−1)	Blue (+1)
Click button design	Simple "Okay" (−1)	Large "Click Now to Join" (+1)

blue background and combine it with traditional font and modern font. We already have four runs of traditional font, blue colour and large button. Now we have to carry out four runs of the combination of modern font, blue colour and large button. Thus, we have a total of 16 runs. These 16 runs will help us to estimate the effects of font, background colour and type of click button. But it is also possible that there can be interaction effects between these factors. For example, a combination of small button with blue background and modern font could be much more effective than any other combination. Notice that we did not experiment with this particular combination at all, and hence, we have no way of estimating this effect. Same is true with many other interactions.

The factorial design developed by Ronald Fisher is a much better approach. With three factors and two levels for each factor, there are eight possible combinations. These eight combinations can be displayed randomly to the viewers and the conversion rate calculated. It is important that each combination is displayed with equal probability. This process involves only eight runs instead of 16 runs required in the earlier approach. Table 21.8 shows the factorial design of the above experiment with two replications. The levels of each factor are represented by +1 and −1. The coding is as follows:

There are only two levels for each of the factors in our experiment. Hence, these designs are called two-level factorial designs (since there are three factors, this design is referred to as 2^3 factorial design). In general, there can be many more levels for each factor. The change in the response (conversion rate) when the level of the factor is changed from −1 to +1 is called the "main effect". For example, the main effect for the factor "font" is the change in the conversion rate when the font is changed from "traditional (−1)" to "modern (+1)". When the effect of one factor is influenced by another factor (a typical example is water and fertilizer in agricultural experiments), it implies that there is a synergy between these two factors. Such effects are called interaction effects. In Table 21.9, the coding of interaction variables is the multiplication of the corresponding columns.

In order to isolate the effects of each factor and the interactions, we need to calculate the sum of squares corresponding to each main effect and interaction effect. The ANOVA table for the data of the above experiment is presented in Table 21.10. These results are obtained by running the model in R using the following code:

```
> Twoway_anova <- aov(Conversion_Rate ~ Font + Background
   + Click + FB + FC + BC + FBC, data=factorial_experiment)
> summary.aov(Twoway_anova)
```

Table 21.9 Factorial design with two replications

	Main effects			Interaction effects				
Conversion rate (%)	Font (F)	Background colour (B)	Click button (C)	FB	FC	BC	FBC	
1.36	−1	−1	−1	+1	+1	+1	−1	REPLICATION 1
2.04	+1	−1	−1	−1	−1	+1	+1	
1.53	−1	+1	−1	−1	+1	−1	+1	
1.87	+1	+1	−1	+1	−1	−1	−1	
3.23	−1	−1	+1	+1	−1	−1	+1	
5.78	+1	−1	+1	−1	+1	−1	−1	
2.04	−1	+1	+1	−1	−1	+1	−1	
5.61	+1	+1	+1	+1	+1	+1	+1	
0.68	−1	−1	−1	+1	+1	+1	−1	REPLICATION 2
2.38	+1	−1	−1	−1	−1	+1	+1	
0.51	−1	+1	−1	−1	+1	−1	+1	
2.38	+1	+1	−1	+1	−1	−1	−1	
2.04	−1	−1	+1	+1	−1	−1	+1	
4.59	+1	−1	+1	−1	+1	−1	−1	
2.55	−1	+1	+1	−1	−1	+1	−1	
6.12	+1	+1	+1	+1	+1	+1	+1	

Table 21.10 ANOVA table for 2^3 factorial experiment

Dependent variable: conversion rate						
	Df	Sum sq.	Mean sq.	F value	Pr(>F)	
Font (F)	1	17.703	17.703	54.149	7.93E−05	***
Background colour (B)	1	0.016	0.016	0.05	0.8291	
Click button (C)	1	23.064	23.064	70.547	3.07E−05	***
FB	1	0.219	0.219	0.669	0.4372	
FC	1	3.658	3.658	11.188	0.0102	*
BC	1	0.045	0.045	0.138	0.7198	
FBC	1	0.305	0.305	0.934	0.3622	
Residuals	8	2.615	0.327			

Signif. codes: ***, 0.001; **, 0.01; *, 0.05
R squared = 0.945 (adjusted R squared = 0.897)

It can be seen from the above table that the main effects of "font" and "click button" are significant and the effect of "background colour" is not significant. In addition, only the interaction between the "font" and "click button" is significant. All other interactions are not significant.

The mean conversion rates for each level of the factors and the interactions are presented in Table 21.11.

The mean effects of different factors and the interactions are also presented in Table 21.11 (overall column). The way these effects need to be interpreted is that the conversion rate will go up by 2.1038 when we change the font from traditional

Table 21.11 Mean conversion rates

	Mean conversion rates		
	1	−1	Overall (level 2 − level 1)
Font (F)	3.8463	1.7425	2.1038
Background colour (B)	2.8263	2.7625	0.0637
Click button (C)	3.9950	1.5938	2.4013
FB	2.9113	2.6775	0.2338
FC	3.2725	2.3163	0.9563
BC	2.8475	2.7413	0.1063
FBC	2.9325	2.6563	0.2763

Table 21.12 Results of regression analysis

Regression statistics				
Multiple R	0.9722			
R square	0.9451			
Adjusted R square	0.8970			
Standard error	0.5718			
Observations	16.0000			
ANOVA				
	df	*SS*	*MS*	*F*
Regression	7	45.0099	6.4300	19.6677
Residual	8	2.6155	0.3269	
Total	15	47.6254		
	Coefficients	*Standard error*	*t stat*	*P-value*
Intercept	2.7944	0.1429	19.5486	0.0000
Font (F)	1.0519	0.1429	7.3586	0.0001
Background colour (B)	0.0319	0.1429	0.2230	0.8291
Click button (C)	1.2006	0.1429	8.3992	0.0000
FB	0.1169	0.1429	0.8176	0.4372
FC	0.4781	0.1429	3.3448	0.0102
BC	0.0531	0.1429	0.3716	0.7198
FBC	0.1381	0.1429	0.9663	0.3622

to modern. The increase in the conversion rate is only 0.0637 when we change the background colour from white to blue.

The conclusion is that the company should use modern font with large click button with "Click Now to Join". The background colour does not matter.

More or less similar information could be obtained by carrying out a regression on the conversion rate with the columns of main effects and interaction effects in Table 21.9. The results of the regression analysis are presented in Table 21.12.

You can notice that the p-values of each effect in the regression analysis and the ANOVA in Table 21.10 match exactly. The intercept is nothing but the overall mean, and the regression coefficients corresponding to each factor or interaction are the shifts (positive or negative) from the overall mean.

The above example deals with two-factorial experimental design. The same model can be expanded to scenarios where there are more than two levels for the factors. The real problem will be the number of possible runs needed for the experiment. If there are six factors with two levels each, then the experiment will require 64 runs, not counting the replications. In such situations, one can use "fractional factorial designs". The discussion on fractional factorial designs is beyond the scope of this book. Interested students can read any textbook on experimental designs.

The interaction effects can be gauged better by drawing the interaction graphs. Figure 21.3 shows the interaction graphs for the three two-factor interactions (FB, FC and BC). When the two lines in the graph run parallel to each other, it indicates that there is no interaction between the two factors. A comparison between the two graphs, FB and FC, indicates that the conversion rate increases significantly when both font and click button are set at +1 level.

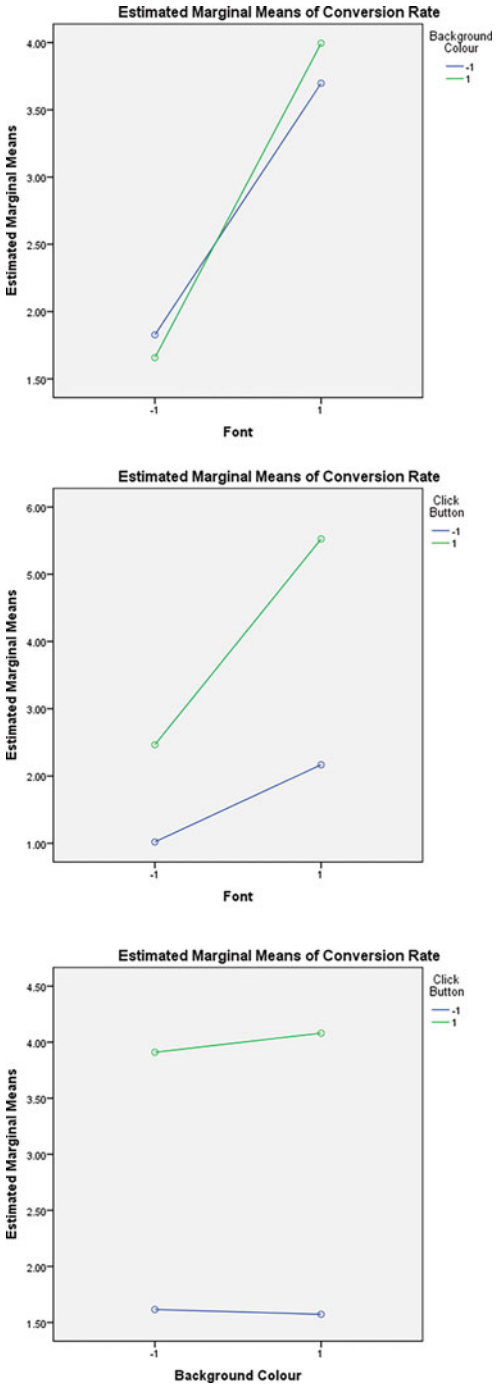
4.4 Orthogonality

An experimental design is said to be orthogonal if for any two design factors, each factor level combination has the same number of runs. The design specified above is an orthogonal design. Consider any two factors in the experiment and the effect on the response variable is studied for four possible combinations. There are exactly two runs (not counting the replications) for each combination. In addition, if you take any two columns (other than the response column) in Table 21.9 and multiply the corresponding elements and total them, the total is always zero. This also implies that the correlation between any two columns in Table 21.9 (other than the response column) is zero. This is a characteristic of the two-level factorial design. Because of this orthogonal nature of the design, all the effects can be estimated independently. Thus, the main effect of "font" does not depend on the main effect of "click button".

Experimental designs are extensively used in many social networks such as Facebook, Twitter and LinkedIn to make data-driven decisions. LinkedIn actually created a separate platform called XLNT (pronounced as Excellent) to carry out experiments on a routine basis. The platform can support more than 400 experiments per day with more than 1000 metrics. They have been using this platform for deploying experiments and analyzing them to facilitate product innovation. Their experiments range from visual changes in the home pages to personalizing the subject lines in emails[3].

[3]Ya Xu et. al., "From Infrastructure to Culture: A/B Testing Challenges in Large Scale Social Networks", KDD'15, 11–14 August 2015, Sydney, NSW, Australia.

Fig. 21.3 Interaction graphs



4.5 Data-Driven Search Engine Advertising

Today, data-driven marketing has become ubiquitous. Data-driven marketing strategies are based on insights obtained from data collected through customer touch points, customer interactions and customer demographics. This data is analyzed to understand customer behaviour and make predictions. These become inputs to better marketing and advertising efforts. The advertising efforts are aimed at enhancing customer experience through personalization. The approach helps companies to convey the right message at the right time to the right target. Companies can create highly customized campaigns delivering a personalized and focussed message to the consumer. With the amount of data at their disposal, companies can identify the groups of consumers that can be easily targeted, based on the consumer profiles and by identifying actionable strategies.

Consumers keep moving from one device to another (mobile to laptop to digital TV to tablet). They move from e-commerce sites to social websites. Data-driven advertising enables companies to follow the customer through various devices as well as different platforms. Companies can target the consumer with consistent messages across different channels. They can make relevant offers, specially designed to a particular consumer. For example, if a customer has purchased a mobile, he or she can be offered various accessories that can go with the specific model of mobile purchased.

Using techniques such as big data analytics, companies can stay up to date with the changing trends and preferences of consumers. One can even predict future changes in tastes and preferences. This data, coupled with appropriate marketing and advertising strategies, can actually help make the best changes.

Through data-driven advertising, marketers are able to reach consumers online irrespective of their physical locations. Individual consumers can be identified and selected to receive specific, highly focussed messages based on their behavioural patterns that will ultimately facilitate conversion. Implementing dynamic advertisement online which enables the consumer to interact can be a great source of obtaining first-hand information from the customer.

Today's consumers extensively rely on search engines to obtain information that they need before purchasing any product or service. By understanding and identifying what these consumers are searching for, marketers can create focussed messages. Once they identify what consumers are searching for, they can optimize the paid searches as well as create better content on their own websites that are in line with the frequent consumer searches and ultimately drive the traffic to their own websites.

Example: How Is Paid Search Done?

Let us take an example with Google AdWords to understand how paid search works. When a customer starts a search, Google starts the process by identifying the keywords that are relevant for the search string entered by the customer. These keywords need not necessarily be part of the search string entered by the customer. For example, if the search string entered is "social media marketing", keywords such

as "digital marketing" and "search engine marketing" may be found to be relevant. Then, it starts looking at the available AdWords from the pool of advertisers and determines if an auction for the available advertisement is possible.

On the other side, advertisers identify various keywords that they want to bid on, along with the amount that they are willing bid. They also pair each of the keywords with specific advertisements. These advertisements can either be predesigned or put together by combining different objects based on various parameters received from the search engine (Google, in this case).

Then, Google enters the keywords it deems relevant into the auction. Google determines the ranks of different possible keywords and the associated advertisements based on two factors, namely, the maximum bid and the quality score. The maximum bid is what is specified by the advertiser for a particular keyword. The quality score is usually determined by the click-through rate (CTR), relevance of the ad and the landing page. The rank of the advertisement is determined by multiplying the quality score with the maximum bid amount. This rank is used to decide on the position of the advertisement in the search results page.

One of the interesting aspects of the above model is that the bid amount of an advertiser need not be the highest among the competing advertisers in order to get the highest rank. If a particular advertisement (paired with a specific keyword) has a high-quality score, even a much lower bid amount can result in a top rank. Google understands this and correspondingly calculates the cost per click (CPC).

Let us consider an example where a customer entered a search string "detergent for baby bottles". Based on this search string, Google determines that the keywords "dish washing", "dishwasher soap" and "baby soap" are relevant and identifies four possible advertisers. The details of the bid amounts and quality scores of each of the four advertisers are given in Table 21.13.

Advertiser 1 is ranked as number 1 even though bid amount is the lowest. This is because of the high-quality score. Similarly, the highest bidder, Advertiser 4, is ranked lowest because of the low-quality score. Based on the above, the advertisement of Advertiser 1 is displayed.

When the winning ad is displayed and the customer clicks on it, the cost per click of the winning ad is calculated using the rank of the advertiser who is ranked just next to the winning ad and the quality score of the winning ad. In the above example, the winning score of the next best rank is 18 (Advertiser 2), and the quality score of the winner is 10. The CPC is calculated as $18/10 + 0.01 = 1.81$. The presumption here is that even if Advertiser 1 bids an amount of $1.81, its rank will be 18.1 (given

Table 21.13 Bid amounts and quality scores of advertisers

Advertiser	Maximum bid ($)	Quality score	Rank = bid*quality score	Actual CPC
Advertiser 1	2	10	20	1.81
Advertiser 2	3	6	18	2.01
Advertiser 3	4	3	12	3.51
Advertiser 4	5	2.1	10.5	5

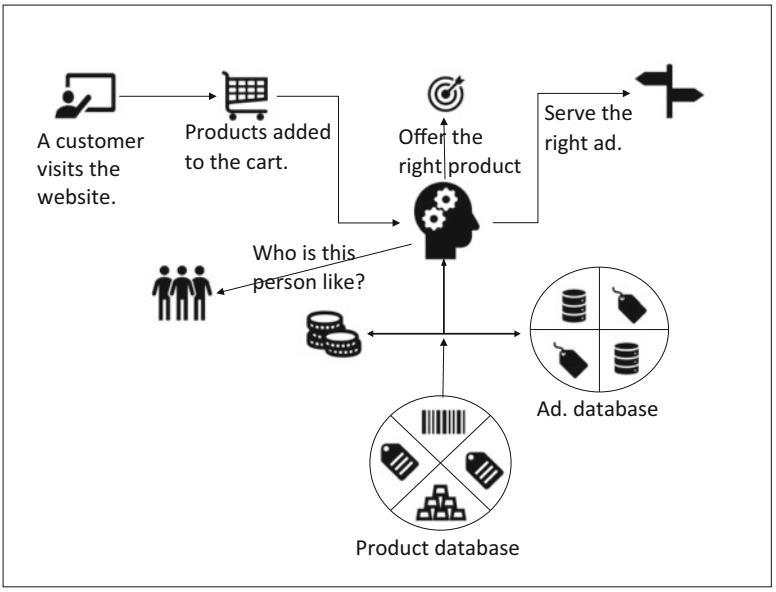


Fig. 21.4 Flowchart of recommendations

that the quality score is 10) which is marginally higher than the next best rank. The addition of one cent ($0.01) in the calculation formula is to ensure that the winning ad is marginally higher and not tied with the next best ranked ad.

Prediction Model for Ad and Product

Similar methods can be employed to display appropriate ads and/or make appropriate recommendations when customers access e-commerce sites for purchasing a product or service. The process is briefly presented in Fig. 21.4. Let us consider a customer who visits a particular e-commerce website looking for a stroller. She logs in and goes through different models and selects an infant baby stroller and adds the same to her shopping cart. Her past search history indicates that she had also been searching for various products for toddlers. In addition, her demographic profile is obtained from her log-in data as well as her previous purchasing decisions. This information can be fed into a product database which identifies a "duo stroller (for an infant and a toddler)" which is marginally higher in cost as compared to the one she had already selected. An advertisement corresponding to the duo stroller is picked up from an ad database and displayed to the customer along with a discount offer. The customer clicks on the ad, gets into the company website of the duo stroller, checks the details and goes through some reviews. She goes back to the e-commerce website and buys the duo stroller with the discount offered.

A similar process can be applied to make recommendations such as "those who purchased this item also bought these" to various customers based on simple market basket analysis. Other product recommendations can be made based on customer profiles, past browsing history, or similar items purchased by other customers.

4.6 Analytics of Digital Attribution

Traditionally, TV and print media advertising had been considered as the most effective marketing medium. But, in the recent years, digital ads have been outperforming all other media. With technology enabling advertisers to track consumers' digital footprints and purchasing activities in the digital world, advertisers are able to gain more insights into the behaviour of the consumers. Nevertheless, the consumers are simultaneously exposed to many other types of advertising in the process of making online purchases. Digital media constantly interacts with other media through multichannel exposures, and they complement each other in making the final sale. In such a scenario, attribution of the share of various components of digital media as well as other media is becoming more and more important.

The consumer today is exposed to multiple channels, each providing a different touch point. Imagine a consumer, while watching a TV show, comes across an ad for Samsung S8 smartphone and searches for "Samsung smartphones" on Google and comes across a pop-up ad for Samsung C9. He clicks on the ad and browses the resulting website for various comments and reviews on C9. He watches a couple of YouTube videos on C9 by clicking on a link given in one of the reviews. Then he goes to a couple of e-commerce sites (say, Amazon and Flipkart) and does a price comparison. A couple of days later, he receives email promotions on C9 from both the e-commerce sites. He comes across a large POS advertisement for C9 at a neighbourhood mobile store and stops and visits the store to physically see a C9 that is on display. A couple of days later, he receives a promotional mailer from Axis Bank which offers a 10% cashback on C9 at a particular e-commerce site (say Amazon) and finally buys it from the site.

The question here is: How much did each of these ads influence the consumer's decision and how should the impact of each of these channels be valued? These questions are important because the answers guide us in optimizing our advertisement spend on different channels.

Today, the explosion of data from various sources provides unmatched access to consumer behaviour. Every online action on each and every website is recorded, including the amount of time spent on a particular web page. Data from transactions from retail stores, credit card transactions, call centre logs, set-top boxes, etc. are available. All this information can be analyzed to not only understand consumer behaviour but also evaluate the contribution of each marketing channel.

An attribution model is the rule, or set of rules, that determines how credit for sales and conversions is assigned to touch points in conversion paths. Attribution is giving credit to different channels that the company employs to promote and broadcast the advertising message. Digital attribution pertains to attributing credit to various components for providing the marketing message in the online world using various forms of media and online platforms. It is true that the digital world offers phenomenal opportunities for measurability and accountability. Nevertheless, it is more challenging in the digital world to disentangle the impacts of various forms of advertising and those of different platforms employed. Some of the actions are not

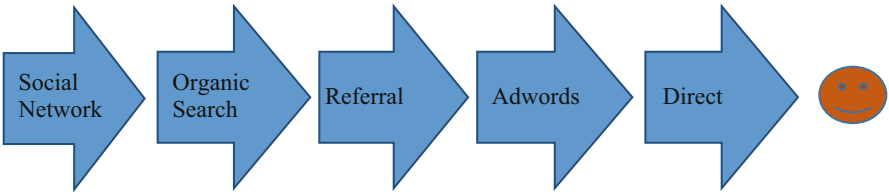


Fig. 21.5 Funnel showing different steps (channels) till conversion

controllable by the advertiser. For example, a search-based ad can be displayed only when the consumer initiates a search for a product or service on a particular search engine. It is always challenging to understand what triggers the consumer to initiate such search, leave alone influencing the consumer to initiate such an action!

There are different models for digital attribution, each having their own advantages and disadvantages. Before discussing the merits and demerits of each of these models, let us consider the following funnel which resulted in a conversion (Fig. 21.5).

Attribution Models

(i) Last Interaction/Last Click Attribution Model

This is one of the very early models of attribution and is used as a standard model in many web analytics because it is the simplest among all models. Here, the entire credit is given to the last interaction or click just before the conversion. In the above funnel, "Direct" is given the entire 100% credit. This model is used mainly because there is no ambiguity regarding the final action that resulted in the conversion. Since only the last touch point is considered, the cookie expiration period becomes irrelevant. This model is easy to implement and becomes appropriate when the funnels take a long time from initiation to conversion. The disadvantage is that there are many other interactions before the last interaction, and it does not recognize any of those. On the other hand, it is more appropriate when no information is available for visitor-centric analysis.

(ii) Last Non-direct Click Attribution Model

This model underplays the direct traffic. It gives credit to the campaign that was initiated just before conversion. These campaigns can be referring sites, emails, displays on search engines, social websites or even organic searches. "Direct channel" is often defined as when the visitor manually enters the URL in his browser. Most often, any visitor who does not have any referral tag is treated as "Direct". In other words, direct channel is where there is no referral tag and hence, at the same time, attributing the conversion to "Direct" can be misleading. This model avoids these issues involved with direct channel. There are many instances where a consumer actually types the URL because of the brand name and image. Since this model ignores the direct traffic, it is tantamount to undervaluing the brand recognition and brand value. All the efforts that generally go into creating such brand recognition are ignored by this model.

In the above funnel, "AdWords" is given the entire 100% credit.

(iii) Last AdWords Click Attribution Model

This model gives 100% credit to the most recent AdWords ad before converting. This model ignores the contribution of all other channels in the funnel. It looks at only the paid search. Since it ignores all other AdWords, it becomes difficult to optimize the AdWords campaign by identifying and valuing other AdWords that have contributed to the conversion. Needless to say, "AdWords" is given the entire 100% credit in the above funnel.

(iv) First Interaction/First Click Attribution Model

This is reverse of the last interaction/last click attribution model. The entire credit is given to the so-called first click. While there is no ambiguity about the last click, there is no certainty about the first click. Nevertheless, this model gives 100% credit to that particular market channel that brought the customer to the website and, consequently, results in overemphasizing one single part of the funnel, especially the top of the funnel channel. In our example, "Social Network" is given the complete credit. It may take a long time between the first touch point and the conversion. It is possible that the tracking cookie expires during this period, and consequently, the credit is given to a touch point which happens to be first in the cookie expiration window.

(v) Linear Attribution Model

In this model, each step involved, starting from the first interaction to the last interaction or click, is given equal weightage. Thus, this is the simplest of the multi-touch attribution models. The positive aspect of this model is that it considers all the channels involved in the funnel. While this recognizes the importance of each of the intermediate steps, it is very unlikely that the contribution of each step is the same. Each of the channels in the above funnel gets 20% credit. On the other hand, the major role played in the entire process of conversion could be that of social network influence, which actually started the funnel. Nevertheless, this model is better than attributing the entire credit to one single channel.

(vi) Time Decay Attribution Model

The model assigns different credits to different steps involved. It assigns the maximum credit to the last touch point before the conversion, and the touch point just before that gets less credit and so on. It assumes that the touch points that are closest to the conversion are the most influential. There are instances where the first touch point that initiated the entire process is not likely to get appropriate credit, especially if it is far away from conversion. One possible distribution of the credit for conversion in the above funnel could be Social Network 10%, Organic Search 15%, Referral 20%, AdWords 25% and Direct 30%.

(vii) Position-Based Attribution Model

This model, by default, attributes 40% of the credit to the first and last touch points each, and the remaining 20% is distributed among the remaining touch points. It is obviously not necessary to use the default values, and these weightages can be adjusted based on the objective (data-centric) or subjective analysis. By combining this with the time decay model, one can create a customized model. This model is best suited when the importance is on lead generation which got the visitor in and

the touch point which clinched the conversion. At the same time, it does not ignore other touch points.

The "Social Network" and the "Direct" touch points get 40% each, whereas the remaining three will get 6.67% each, in the above funnel.

(viii) Custom or Algorithmic Attribution Model

Based on the merits and demerits of the models described earlier, the analyst can build a custom attribution model. These custom models are built based on data on customer behaviour obtained at different stages. It is being acknowledged that these "evidence-based" models are better and more realistic. Unfortunately, these models are not easy to build. The main principle is to estimate which of the touch points contribute to what extent based on the available customer data so that it represents a more accurate picture of the customer's journey from the initiation to conversion. This is called a custom or algorithmic attribution model. These models not only make use of customer data but also use statistical techniques which can lead to continuous optimization of the advertisement budget.

A simple approach for building an algorithmic attribution model starts with identifying the key metrics that are to be used to measure the relative effectiveness of each channel. It is important to associate an appropriate time period with this. This is to be followed with the cost per acquisition for each channel.

With the present technology, the data with respect to various touch points is obtained without much difficulty. This data can be used to build models to predict conversion or otherwise. Generally, the target variable is binary, whether the conversion is successful or not. The predictor variables are the various touch points. To start with, the metric for effectiveness of the model can be the prediction accuracy. Based on the contribution of each of the predictor variables, the attribution to various touch points can be calculated. Consider building a logistic regression model for predicting the success of conversion. Once we achieve the prediction accuracy levels for both the training and testing datasets, the coefficients of the logistic regression can be used to calculate the percentage attribution to various touch points or channels[4].

While it is true that any of the predictive models can be employed for this purpose, models such as logistic regression or discriminant analysis are more amicable since the coefficients corresponding to different predictor variables are available with these models. We can also use black box methods such as artificial neural networks or support vector machines. In such a case, we can assign the attribution values to different predictor variables based on the "sensitivity index" obtained after building the model.

[4]For more details, see Shao, Xuhui and Lexin Li, "Data-driven Multi-touch Attribution Models", KDD'11, 21–24 August 2011, San Diego, California, USA.

Other algorithmic models that can be used for calculating the attribution are as follows:

- Survival analysis
- Shapley value
- Markov models
- Bayesian models

We can optimize the budget allocation for each of the channels by using the attribution values (obtained from the predictive models) and the costs estimated earlier (CPA, CPM, etc.). Even a simple optimization technique such as budget-constrained maximization can yield significant results.

Let us consider an example of calculating cost of acquisition (CoA) under different attribution models. Let us consider a customer who purchased an "Amazon Home" through the following process:

1. Yajvin, the customer, first clicks on AdWords and visits the website.
2. Then he visits his Facebook page and clicks on the ad displayed on Facebook, visits the website again and checks out the functionality of the device.
3. Afterwards, he visits the website again through his Twitter account and looks at the technical details.
4. Then, he directly comes to the website and checks on various reviews.
5. Finally, he clicks on an offer with a discount that he received on email and purchases the device.

Let us assume that the advertisement expenditure is as follows: AdWords, $12; Facebook, $15; Twitter, $20; direct, $0; email, $8.

The cost of acquisition under each of the attribution models can be calculated based on the above information. Table 21.14 provides the details about the ad spend on each channel, the weightages for each channel under each model and the calculated CoA.

Similarly, data-based attribution models can be used to estimate the contribution of each channel. This is important to understand which channels are actually driving the sales. Based on this, advertisers can spend money more effectively and maximize ROI. The example below demonstrates the attribution across three different channels using the Shapley value approach.

Table 21.14 Ad spend, weightages and CoA for various channels

Channel	Ad spend	First touch (%)	Last touch (%)	Linear (%)	Position based (%)	Time decay (%)
AdWords	$12.00	100	0	20	40	10
Facebook	$15.00	0	0	20	7	15
Twitter	$20.00	0	0	20	7	20
Direct	$0.00	0	0	20	7	25
Email	$8.00	0	100	20	40	30
CoA		$12.00	$8.00	$11.00	$10.33	$9.85

Table 21.15 Watches sold based on channels

Channel	Number sold
AdWords + Facebook + email	256
AdWords + Facebook	192
AdWords + email	190
Facebook + email	128
AdWords	180
Facebook	120
Email	64

Table 21.16 Permutations of channels and their contributions

	Contribution of channel by order in the permutation			Contribution of specific channel		
Permutation	First channel	Second channel	Third channel	A (AdWords)	B (Facebook)	C (email)
ABC	180	12	64	180	12	64
ACB	180	10	66	180	66	10
BAC	120	72	64	72	120	64
BCA	120	8	128	128	120	8
CAB	64	126	66	126	66	64
CBA	64	64	128	128	64	64
Total contribution across all permutations				814	448	274
Average contribution across all permutations				135.67	74.67	45.67
Percentage contribution				52.99%	29.17%	17.84%

Let us consider an example where the company uses three channels for promoting its product, a smartwatch. These channels are AdWords (Channel A), Facebook (Channel B) and email (Channel C). Based on the data, the number of watches sold through each channel (and each possible combination of channels) is obtained and summarized in Table 21.15.

The company managed to sell 256 smartwatches when the customers exposed (used) all the three channels, while it could sell only 64 smartwatches when the customers used only email and nothing else. These numbers are obtained based on the analysis of purchase data through different channels. Considering that there are three channels (A, B and C) in this example, there are six possible permutations for combining these three channels. These permutations are A→B→C, A→C→B, B→A→C, B→C→A, C→A→B and C→B→A.

In the first permutation, Channel A contributes 180 (contribution of AdWords alone), Channel B contributes 12 (channels A and B together contribute 192, and hence the contribution of B is $192 - 180 = 12$), and Channel C contributes 64 (all the three channels together contribute 256, while A and B together contribute 192, and hence the contribution of C is $256 - 192 = 64$). Similarly, the contribution of each of the channels corresponding to each permutation is calculated and presented in Table 21.16.

Across all the six permutations, the contribution of AdWords is 180, 180, 72, 128, 126 and 128. The total of these values is 814, and the average is 135.67. Similarly, the averages for Facebook and email are 74.67 and 45.67, respectively. These values are converted into percentages which are presented in the table. These values are referred to as "Shapley values" (named after the Nobel Laureate Lloyd Shapley).

Based on the above analysis, the company should invest more in AdWords and least in email. As a matter of fact, the advertisement budget can be distributed across the three channels in the same ratio as the percentage contributions.

The above approach requires large amounts of data. The company needs to obtain data with respect to each and every combination (all possible subsets as well as individual channels) of the channels employed. If there are *n* channels, the data has to be obtained for $2^n - 1$ subsets. Implementing experimental designs could be a possible approach to obtain the required data. Once an optimization strategy for budget allocation across different channels is evolved and implemented, constant monitoring of the channels is necessary for further fine-tuning.

5 Strategies for Mobile Devices

With the popularity of smart mobiles in the recent days, more than half of search traffic started to emanate from mobiles. These devices are also the popular medium for interacting within social networks. In addition, Google's mobile ranking algorithm includes mobile-friendly and mobile usability factors as well as availability of mobile apps in its indexing. Consequently, those with mobile-friendly websites and/or mobile apps get much higher ranks and appear at the top of the search results. Consequently, it is becoming more and more important for businesses to evolve a mobile-oriented strategy in order to improve effectiveness of their marketing campaigns. It is becoming necessary to create mobile marketing strategies which improve customer experience while using mobiles at every stage of the customer purchase funnel.

Two important strategies that businesses need to adopt are to create mobile-friendly websites and mobile apps. Users in the initial stages of the purchase funnel are most likely to be using the website rather than downloading and installing the app. On the other hand, mobile apps allow for better interaction and facilitate more creativity in engaging the customer. In other words, it is necessary for businesses to create their own mobile-friendly websites as well as create specific apps.

A mobile website is a website which is designed and created for specifically viewing on a smartphone or a tablet. It needs to be optimized so that it responds or resizes itself to suit the display based on the type of device. We need to understand that customers use these devices at different stages in the purchase funnel. Businesses can accelerate the purchase process through sales alerts, display advertisements, providing QR codes, extending special discounts and issuing discount coupons. It is easy to integrate the mobile-based campaigns with different social media sites so that the customers can interact with others regarding the

products and services within their social networks. The websites need to be optimized so that they load faster and they are easy to navigate, and click buttons need to be large enough and have short menu structures. The websites and apps should also ensure that there is minimum amount of typing required. It is also a good idea to allow for maps showing the location since many customers tend to use mobiles when they are on the go.

The website or app should allow users to connect to various social media platforms. This should also include a feature which will make it easy for customers to share the information with others in the network. The apps have an additional advantage. The app stays on the mobile screen, whether the customer is using the app or not. Every time the customer looks at the screen, they see the name of the app or name of the brand which acts as a constant reminder.

Geolocation is an important aspect of the mobile strategy. It is easy to integrate this into mobile apps. Businesses will be able to identify the location of the customer at any particular moment. Data can be collected on places that the customer visits on a regular basis (such as where the customer generally takes a walk). With this kind of information, the app can automatically provide various promotions or exclusive offers that are currently available at a store that is located nearest to the customer. Many of the mobile devices today come equipped with "near-field communication (NFC)". NFC can be useful in locating the customer within a particular store or facility, and the app can draw the user's attention to any items nearby or special discounts based on the past browsing/search/purchase behaviour of the customer through SMS or notifications. This is especially useful when the customer is physically close to the product and at a stage where he or she is ready to make a decision.

It is also important for the app to be able to operate offline. For example, the user could download the catalogue and browse through the offerings without relying on Wi-Fi or the mobile signal.

Ultimately, full benefit of a mobile strategy can be extracted only when the mobile channel is integrated with other channels. The customer should be able to seamlessly move from his or her mobile to any other channel and reach the brand or product or service.

Thus, the mobile strategy should be such that it provides enough customization to leverage the advantages of a mobile or tablet device while integrating with other channels so that the customer has a coherent experience across all channels.

6 The Future of Social Media Analytics

The past decade has seen a phenomenal growth of social media which has changed personal and professional lives of people all over the world. As networking through social media grew, businesses started leveraging social media platforms to reach out to customers directly to attract and retain them. Business organizations found innovative ways to listen to customers' voices through social media and better understand their needs. At the same time, development of technologies provided

opportunities to analyze large amounts of unstructured data generated by social media so that the businesses can become more responsive. Today, social media analytics is being used to obtain deeper insights into customer preferences. It also opened new avenues to innovate and experiment. The same social media analytics can be used to communicate with customers at exactly the right moment with the right message to influence the decision-making process. In addition, it opened a new approach to reach the customer through multiple channels. Customer engagement through multiple channels has not only become the need of the hour but imperative to drive home the message and influence decision-making. The available technology is making it easier to understand the behaviour of the customer through his/her use of social media and also to understand the effectiveness of each channel in contributing to the final decision of the customer. Today, social media analytics is still evolving and yet to mature. New applications are emerging on a daily basis while throwing up additional challenges in customer engagement. At the same time, privacy concerns are becoming serious issues, both from ethical and legal positions. The next few years will be challenging to find a proper balance between privacy concerns and the needs of businesses to engage customers more effectively.

Electronic Supplementary Material

All the datasets, code, and other material referred in this section are available in www.allaboutanalytics.net.

- Data 21.1: adspends.csv
- Data 21.2: antiques_devices.csv
- Data 21.3: bid_qs_advt.csv
- Data 21.4: bid_qs_gmat_orgs.csv
- Data 21.5: factorial_experiment.csv
- Data 21.6: furnimart.csv
- Data 21.7: global_time.csv
- Data 21.8: indo_american.csv
- Data 21.9: membership_drive_isha.csv
- Data 21.10: modern_arts.csv
- Data 21.11: watches_sales.csv
- Data 21.12: webpage_visitors.csv
- Data 21.13: SMWA_Solutions.csv
- Code 21.1: factorial_experiment.R

Exercises

Ex. 21.1 Modern Arts (India) has initiated a special email campaign with three different subject lines. These are as follows:

(a) Subject line 1: "Welcome to Modern Arts"
(b) Subject line 2: "Special Invitation to Our Modern Art Exhibition"
(c) Subject line 3: "Are You a Fan of Modern Art?"

A total of 30,000 unique email IDs were selected and divided randomly into four groups of 7500 each. Each of these groups was further randomly divided into three groups of 2500 each. Emails with exactly the same font and content (body) were sent to each of these groups with one of the three subject lines, and the response was recorded. In other words, the experiment consisted of three treatments and four replications. The responses are presented below:

Subject line	Sends	Opens	Clicks
	Replication 1		
Subject line 1	2500	278	96
Subject line 2	2500	405	136
Subject line 3	2500	222	62
	Replication 2		
Subject line 1	2500	314	87
Subject line 2	2500	461	155
Subject line 3	2500	187	59
	Replication 3		
Subject line 1	2500	261	82
Subject line 2	2500	421	147
Subject line 3	2500	216	58
	Replication 4		
Subject line 1	2500	289	79
Subject line 2	2500	436	128
Subject line 3	2500	192	41

There are two response rates, namely, "open rate" and "click-through rate". Test to find out which subject line is the best with respect to each of the response rates.

Ex. 21.2 It was revealed that each replication was sent to a different mail account. All the 7500 emails of Replication 1 were actually addressed to Gmail accounts. Similarly, all mails of Replication 2 were sent to Outlook mail accounts. All mails of Replication 3 were sent to Yahoo mail accounts. All mails of Replication 4 were sent to AOL mail accounts. Given this information, Modern Arts (India) decided to consider this as a completely randomized block design in order to look at the effect of treatments and blocks.

Test if there is a significant block effect. Which is the best subject line? Carry out the analysis for both the response rates.

Ex. 21.3 FurniMart is a hub for furniture enthusiasts to both sell and buy specially designed furniture. FurniMart operates through its own website with an online catalogue. The traffic to the website comes mainly from three sources, those who type the URL and reach the website (direct), those who come through AdWords

and those who respond to display advertisements from social networks. It had been their experience that many customers added products to their carts directly from the catalogue pages. The existing design of the website displays a graphic  to facilitate adding products to the cart. Customers select a particular product and click on the graphic button to add the item to their shopping carts. It was felt that a bigger "call-to-action" (CtA) button is likely to lead to better conversion rates.

FurniMart decided to experiment with three different types of buttons. These are displayed below:

	Treatment 1	Treatment 2	Treatment 3
Design of the call-to-action button	Add to Cart	+ CART	

FurniMart created three different websites, namely, HTTP://FurniMart.COM/T1, HTTP://FurniMart.COM/T2 and HTTP://FurniMart.COM/T3, each displaying a different CtA button. The traffic coming from each of the three sources is randomly diverted to each of the three sites such that the total traffic from each source is equally distributed to the three websites. The conversion rates are summarized in the table below:

Source of traffic	Treatment 1	Treatment 2	Treatment 3
Direct	12.20	14.14	11.65
AdWords	13.59	14.42	11.93
Social networks	12.48	12.20	10.54

Analyze the above data to identify which CtA button is the best for conversion.

Ex. 21.4 Consider the above data (Question 3) as a completely randomized block design in order to look at the effect of treatments and blocks. Test if there is a significant block effect. Which is the best treatment?

Ex. 21.5 Global Time is a local dealer for the "Ultimate" brand of smartwatches. Whenever any potential customer searches for smartwatches, Ultimate bids along with other smartwatch sellers. When any customer who is located within the geographical area of Global Time clicks on Ultimate's ad, the visitor is taken to Global Time's website using the geolocation feature. Global Time is trying to revamp its website in order to improve its conversion rates. They have identified three different aspects (treatments) of the website that they want to tweak. These aspects are as follows:

(a) Currently, there is no video on the home page. The proposal is to add a 90 s video showing the features of "Ultimate" smartwatch.

(b) At present, the "Buy" button is at the right side of the web page, vertically centred. The proposal is to shift it to the bottom right so that there is more space for more visuals on the page.
(c) At present, the page displays testimonial in text form. The proposal is to include a small photo of the customer who had given the testimonial.

Global Time decided to carry out A/B testing on these three aspects. The details of the treatments and the corresponding conversion rates are given in the table below:

Treatments			Number of visitors	
Video	Location of "Buy" button	Testimonials	Reaching Global's website	Placing an order for Ultimate
Replication 1				
No video	Right centre	Text only	2590	115
Add video	Right centre	Text only	2458	205
No video	Bottom right	Text only	2406	165
Add video	Bottom right	Text only	2557	144
No video	Right centre	Text with photo	2409	101
Add video	Right centre	Text with photo	2458	103
No video	Bottom right	Text with photo	2561	161
Add video	Bottom right	Text with photo	2587	181
Replication 2				
No video	Right centre	Text only	2519	170
Add video	Right centre	Text only	2574	193
No video	Bottom right	Text only	2476	112
Add video	Bottom right	Text only	2546	125
No video	Right centre	Text with photo	2595	100
Add video	Right centre	Text with photo	2459	110
No video	Bottom right	Text with photo	2562	99
Add video	Bottom right	Text with photo	2445	190

What should "Global Time" do with respect to the three aspects? Are there any interaction effects between these three aspects?

Ex. 21.6 Akshita started a search on Google for organizations which provide GMAT training. A quick analysis of the relevant AdWords by Google found that there are five advertisements that are available for display on the search results page. The bid amounts as well as the quality scores are presented in the table below.

Calculate the ranks and identify the first two advertisers whose ads will be displayed to Akshita. Also, calculate the CPC for each of the advertisers.

Ex. 21.7 Indo-American Consultancy Services (IACS) specializes in placing Indian graduates with considerable work experience with clients in the USA. They advertise their services with display ads in Twitter, LinkedIn and AdWords. When the potential customers click on the display ad, they are taken to the company's website,

Advertiser	Maximum bid ($)	Quality score
Elite Management Training Institute	2	10
Gem Management Training	7	2
International Coaching Institute	4	3
Management Trainings Unlimited	5	2.1
National Management Center	3	6

and the customers are encouraged to register on the website and upload their CVs. Once the potential customer uploads the CV, it is considered as conversion.

Channel	Number of conversions
Twitter + LinkedIn + AdWords	2944
Twitter + LinkedIn	2185
Twitter + AdWords	2208
LinkedIn + AdWords	1472
Twitter	736
LinkedIn	1380
AdWords	2070

Based on the above data, carry out appropriate attribution to each of the three channels.

Caselet 21.1: Membership Drive at ISHA[5]

Vishnuprasad Nagadevara

Initiative for Special Healthcare Access (ISHA) is an organization started by Karthik to provide easy access to private corporate health services at reasonable cost to its members. Karthik, after completing his MBA with health services management specialization, joined a major corporate hospital chain headquartered in Bangalore. He has worked with the chain for 15 years and finally reached the level of chief administrative officer for the entire chain. During his tenure at the hospital chain, he had toyed with an idea of making healthcare accessible to as many persons as possible at affordable rates. In the process, he has set up an organization called "Initiative for Special Healthcare Access (ISHA)" which invites membership of individuals. ISHA made contractual arrangements with a number of private healthcare providers to offer various services to its members. He had put together contractual agreements with different corporate hospitals, diagnostic

[5]For classroom discussion only.

service centres and pharmacists. The members of ISHA benefit from steeply discounted services from these private healthcare providers. The service providers benefit from economies of scale through increased demand for their services from the members of ISHA.

As a part of the agreement, members of ISHA receive a certain number of free consultations from any doctor at major hospitals including dental consultations. In addition, they also get two free dental cleaning and scaling at certain dental clinics. The members are also eligible for two free "complete health check-ups" per year. The participating pharmacists give a minimum discount of 20% on medicines subject to minimum billed amount. The members also get discounts on various diagnostic tests including radiology tests.

These benefits are available to the members of ISHA. The membership is available on an individual as well as family basis. The family membership covers a maximum of four members. Additional members can be added into the family membership by paying an additional amount per person.

The economics of the entire model depends on acquiring a critical mass of members. ISHA decided to take advantage of the increasing web access to push its membership drive. They have initiated an email campaign to enrol members with very limited success. They have also realized that campaigns in print media cannot be targeted in a focussed manner leading to high campaign costs and low conversion rates. Karthik finally decided to resort to web-based campaign using display advertising with organic search as well as with AdWords. ISHA hired "Web Analytics Services India Limited (WASIL)" that has expertise in creating, testing and running web-based advertising campaigns.

WASIL is willing to work with ISHA on a result-based payment model. The fees that are payable to WASIL will depend on the success of the campaign in terms of acquiring the members. WASIL has put together a team of three members to create and run the campaign. Rajeev is heading the team with Subbu and Krithika as the other two members. The team decided to first design a set of objects that can be put together into a display advertisement based on keywords in search strings.

"Blue is always associated with health, like it is with BlueCross", said Subbu. "We should have blue colour in the border. That will make it noticeable and will definitely lead to better click-through rates". Subbu and other members of his team are discussing the changes that need to be made in the display advertisement for ISHA. Currently, ISHA does not use any colour in its advertisement. It is very simple with plain white background and bold font in black. It does not have a click button either. The potential visitor can click anywhere on the ad, and it will take the visitor to ISHA's home page.

Rajeev agreed with Subbu that there should be a border with a colour. Rajeev is passionate about green and feels that it gives a soothing feeling which could be easily associated with health services. Krithika suggested that they can use two different versions, one with blue and the other with green. Since all of them agreed that a colour in the border is absolute necessary, the A/B testing can be done with the two colours.

"We need to have a CTA (Call-to-Action) button. Since we are already experimenting with blue and green on the border, the CTA has to be red", Krithika said. "Here again, we have ample scope for experimentation. Should we try two different colours?" asked Rajeev. The rest of the team members did not agree with different colours. They felt that there are already two colours, one in the border and the other on the CTA. The text will have a different colour, at least black. They felt that having more than three colours in a display ad can make it look gaudy and can also be jarring to the visitor. Rajeev said, "We can't leave it as a button. We need to put some 'call-to-action text' on the button, like 'Click Here to Save'. It will draw more attention. There are enough studies to show that an associated text will always reinforce and yield better results". Subbu felt that they should add some more information into the click-to-action text. "Putting a number such as 'Save up to 70%' will get better response". Rajeev was not convinced. He felt that highlighting such a large percentage saving might possibly make people suspicious of the benefits. Many people think that such large discounts are not possible in healthcare, even though the labs do offer 70% discount on specified lab tests. After a prolonged discussion, the team decided to try out both the versions as another treatment in A/B testing.

The present design of ISHA's ad is that the visitors, when they click on the ad, are taken to the home page. It is expected that the visitor will first look at the home page and navigate from there to other relevant pages. The team felt that it is not enough to make the visitor click on the ad and reach the website. The real conversion is when the visitor becomes a paid member, at least on trial basis. Any member can withdraw his or her membership within 15 days from the registration and get full refund. The team felt that the landing page will have a major impact on conversion. The team members also believed that the landing page should be dynamic. It should be based on the search string or the AdWords leading to the display ad. If the visitors are looking for lab tests, the landing page should accordingly be the one with lab tests, with a listing of laboratories located in a particular town and the corresponding discounts. On the other hand, if the visitor was searching for physician consultation, the landing page should correspond to physicians or respective clinics. Finally, the team agreed that the landing page will also be treated as a treatment with home page being the control.

The team decided to use one more treatment in their experimentation. It was understood that ISHA is a new player with a relatively new concept. The organization is not yet an established one. The team members as well as Karthik felt that putting ISHA's logo on the display ad will increase the visibility. The top line of the ad will show either ISHA without logo or ISHA with logo on the right side. The general layout of the ad is shown in Fig. 21.6.

The team concluded that they should go ahead with the four treatments. The team summarized the final proposal as shown in the table below.

WASIL and Karthik approved the team's proposal, and WASIL ran the campaign with the above four treatments over a period of 1 month. Data was collected in terms of number of potential visitors exposed to each type of advertisement (sample

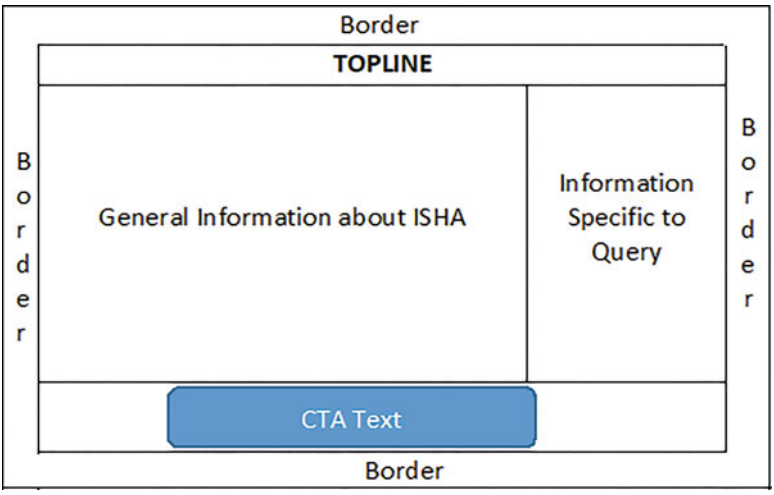


Fig. 21.6 General layout of the ad

Treatment details		
Treatment	Level 1 (−1)	Level 2 (+1)
Border colour	Blue	Green
Top line	Without logo	With logo
CTA	"Click Here to Save"	"Save up to 70%"
Landing page	Home page	Relevant page

size), number of visitors who clicked on the ad (clicks) and, finally, the number of conversions (members). The dataset "membership_drive_isha.csv" is available on the book's website.

Questions

(a) Which of the four treatments is more effective? What is the right level for each of the treatments?
(b) Are there any interaction effects between the treatments?
(c) What would be the final design for the ad in order to maximize the conversions?

Caselet 21.2: Antiques in Modernity[6]

Vishnuprasad Nagadevara

The Meeting

"I think it is Wanamaker who said '*Half the money I spend on advertising is wasted; the trouble is I don't know which half*'. I can't accept that today. Our advertising cost is almost 12% of our revenue, and I need to know where it is going", said Yajvin. "I mean we know exactly where we are spending the money, but I need to know what we are getting back. What is the effect of each of the channels that we are investing in?"

Yajvin is briefing Swetha on her new assignment. Yajvin is the CEO of the online company "Antiques in Modernity (AiM)". Antiques in Modernity specializes in modern electronic devices which have the look of antiques. Swetha heads the analytics consulting company "NS Analytics Unlimited", which provides consultancy services to various companies. In addition to providing consultancy services, NS Analytics also provides training to the client companies so that they can become self-sufficient as much as possible in terms of analytics. Their motto is "We make ourselves redundant by building expertise". NS Analytics has been hired by Antiques in Modernity to analyze their advertising spend and also advise them on how best to improve the ROI on their advertising investment.

"Our products are very special. Look at this turntable. It looks like an antique, and you can actually play a Beatles' gramophone record on this. But it can be used as a Bluetooth speaker; it can connect to you home assistant such as Amazon Echo or Google Home; it can even connect to your mobile. You can stream music from this turntable to any other device even in your backyard!" said Yajvin (Fig. 21.7).

"Similar is the case with our antique wall telephone. It can be used as a wired telephone or as a cordless phone. We are in the process of redesigning it so that you

Fig. 21.7 AiM turntable

[6]For classroom discussion only.

can carry the receiver outside your house and use it as a telephone. But let us come back to our problem. As I said, we invest a lot of money in advertising in different channels. We need to find out the effect of each of these channels. I do understand that many of these channels do complement each other in today's markets. Can we somehow isolate the effects of each, so that our ad spend can be optimized?" asked Yajvin.

Swetha responded saying that there are many models that can be used in order to address the problem, but such models require large amounts of reliable data. She also said that each of these models can give different results, and one needs to understand the assumptions involved in each of these models so that the one which is most applicable to a particular scenario can be picked. Yajvin put her in touch with his chief information officer, Skanda, so that Swetha can get a feel for the type of data that is available with the company and also explain her data requirements.

Skanda explained the data available with them. "We mainly depend on online advertisement. We do invest a small amount in the print media, but most of our advertising is on the social networking websites, AdWords and the like. We also get good amount of direct traffic into our website. Since all our sales are through online only, it makes sense for us to work this way", said Skanda. "We do have a system of tracking our potential customers through different channels. We try to collect as much data, reliably, as possible."

Antiques in Modernity

Antiques in Modernity was a start-up, started in Sunnyvale, California, by Yajvin and Skanda 3 years ago. Yajvin graduated from one of the prestigious technical institutes in India with a major in electronics and went on to do his MBA from a well-known management school in Europe. After his MBA, he worked as a product development manager in a European company. After 5 years with the company, he moved to San Jose as a marketing manager for a software product company. While he moved around from one company to another in the next 12 years, he remained in the marketing function.

Yajvin's hobby was to collect antiques, especially very old electronic products which were in full working condition. He used to collect old electronic products from whatever source possible. If they were not in a working condition, he would work on them and get them into working condition. He also used to sell some of them and also gift some of them away. In the process, he got bitten by the entrepreneur bug and decided to float a company with an aim to manufacture specially designed electronic products with latest technology, but look like antiques. The company designs the products, sources the components from China and Far East and assembles them in the USA. Skanda was his classmate from India and joined him in the venture. They decided that since these are niche products, it will be safer to market them through their own website. They felt that such a strategy will give them complete control on the way the products can be advertised and marketed.

The fact that both of them are comfortable with web-based marketing had a major role to play in making the decision. They had also decided to use as much of online advertising as possible.

AiM uses display ads on social networks, especially LinkedIn, Twitter and Facebook. They also use Google AdWords in order to display their ads based on the search strings used by potential customers. They also keep track of customers who reach their site through organic search. Since their products are sold only through online from their own website, the final conversion is when the customer places the order. They use various methods to trace the channels from which customers reach their website.

The Data

During their meeting, Skanda promised Swetha that he can provide details of each and every funnel starting from the first visit to their website, as well as the referrer by a customer (or potential customer). Swetha felt that there is no reason to look at the funnels that are incomplete which may or may not lead to conversion at some later date. She requested for data only on funnels which resulted in final conversion. There are also many repeat customers who type the URL directly and reach the website and make purchases. Similarly, there are some who purchase items on their very first visit to the website. Skanda told her that he can provide details of each funnel corresponding to each and every conversion. He felt that such detailed data could be useful because AiM sells different products and the profit margins are different for different products. On the other hand, Swetha felt that such detail was not necessary because the advertisements are not specific to any particular product. Even the display ads which are put together on the fly, by AiM based on AdWords or search strings, are not product specific. The main theme of these ads is that their products are latest in technology, but packaged as antiques. They are not really antiques either. Hence, she suggested that Skanda summarize the data "funnel-wise". She also suggested that all the social network channels can be clubbed into one channel for the purpose of initial analysis. "We can drill down into different social networks separately at a later date. As a matter of fact, you will be able to do it yourself after we train your people", she said.

Finally, they have agreed to concentrate on four channels: social networks (Channel A), AdWords (Channel B), organic search (Channel C) and direct (Channel D). It was also decided to maintain the actual order of channels within each funnel. Each funnel is to be read as the sequence of the channels. For example, ABCD implies A→B→C→D. Swetha explained that the order becomes important for estimating the contribution of each channel under different models. Then there was a question of the final metric. Should the final metric for conversion be revenue, profit margin or just the number of items sold? AiM is currently going through a major costing exercise, especially in terms of assigning the fixed/non-variable costs to different products. It was felt that the option of profit margin is not appropriate until the

costing exercise is completed. Skanda and Yajvin felt that the initial exercise can be made based on the sales quantity (number of items sold) and the method can easily be extended to revenue at a later date. Swetha assured them that they will just have to change the values in a simple spreadsheet and everything else will get recalculated automatically!

Swetha received the summarized data as required by her 2 days after her meeting with Yajvin and Skanda. The data "antiques_devices.csv" is available on the book's website.

Further Readings

Abhishek, V., Despotakis, S., & Ravi, R. (2017). *Multi-channel attribution: The blind spot of online advertising*. Retrieved March 16, 2018, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2959778.

Fisher, T. (2018). ROI in social media: A look at the arguments. *Database Marketing & Customer Strategy Management, 16*(3), 189–195. Tracy L. Tuten, Michael R. Solomon, Social Media Marketing, Sage Publishing.

Ganis, M., & Kohirkar, A. (2016). *Social media analytics*. New York, NY: IBM Press.

Gardner, J., & Lehnert, K. (2016). What's new about new media? How multi-channel networks work with content creators. *Business Horizons, 59*, 293–302.

Hawn, C. (2017). Take two aspirin and tweet me in the morning: How twitter, facebook, and other social media are reshaping health care. *Health Affairs, 28*(2), 361.

Kannan, P. K., Reinartz, W., & Verhoef, P. C. (2016). The path to purchase and attribution modeling: Introduction to special section. *International Journal of Research in Marketing, 33*, 449–456.

Ledolter, J., & Swersey, A. J. (2007). *Testing 1 - 2 - 3: Experimental design with applications in marketing and service operations*. Palo Alto, CA: Stanford University Press.

Oh, C., Roumani, Y., Nwankpa, J. K., & Hu, H.-F. (2017). Beyond likes and tweets: Consumer engagement behavior and movie box office in social media. *Information & Management, 54*(1), 25–37.

WilliamRibarsky, D. X. W., & Dou, W. (February 2014). Social media analytics for competitive advantage. *Computers & Graphics, 38*, 328–331.

Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social media mining*. Cambridge: Cambridge University Press.