

# 11

# FACEBOOK

## *How Facebook Use Big Data To Understand Customers*

### **Background**

Facebook, by some considerable margin, is still the world's biggest social network. It's used by everyone and their granny to keep in touch with friends, share special occasions and organize social events. Millions of people every day also use it to read news, interact with brands and make buying decisions.

Like all of the big social networks and search engines, it's essentially free to the end user. The company make the money they use to pay their 10,000-plus staff and keep their services online from businesses that pay to access the data Facebook collect on us as we use their services.

This year, the company announced they had attracted two million active advertisers, mostly small and medium-sized businesses, which pay for ads to appear in the feeds of people who may be interested in them.

### **What Problem Is Big Data Helping To Solve?**

Businesses need to sell products and services to survive. In order to do this, they need to find customers to sell to. Traditionally, this has

been done by advertising in a “broadcast” manner: newspaper, TV, radio and display advertising work on the principle that if you put your ad in the most prominent place you can afford a large number of people will see it and some of them are likely to be interested in what you’re offering.

However, this is obviously a hit-and-miss approach. For a large multinational company, it may be clear that a TV spot during the Super Bowl will increase their exposure and put their brand in front of potential customers. But a small business just starting out has to think far more carefully about the most efficient way of spending its limited marketing budget. These companies can’t afford to cover all the bases, so tools that can help them work out who their customers are, and where to find them, can be hugely beneficial.

### **How Is Big Data Used In Practice?**

The rapid expansion of the online world over the last two decades has provided advertisers with a simple way to do just that. Because websites are hosted on computers, not newspapers or billboards, each visitor can be independently identified by the software running the website. And Facebook, with 1.5 billion active monthly users, has access to far more user data than just about anyone else.<sup>1</sup>

Its data is also more personal – whereas services like Google can track our Web page visits (which incidentally Facebook can now also do) and infer much about us from our browsing habits, Facebook often have full access to straight-up demographic data about us such as where we live, work, play, how many friends we have, what we do in our spare time and the particular movies, books and musicians we like.

## FACEBOOK

A book publisher, for example, can then pay Facebook to put their adverts in front of a million people who like similar books, and match the demographic profiles of their customers.

Data collected by users as they browse Facebook is used to match them with companies which offer products and services that, statistically, they are likely to be interested in. Facebook undoubtedly hold one of the biggest and most comprehensive databases of personal information ever collated, and it is expanding every second of every day.

As well as a platform for sharing messages, Facebook is also a platform for running software. Over half a million apps have been created for Facebook so far, most of which take the advantage of access they have, via the extensive APIs (application program interfaces), to Facebook user data. These apps in turn gather data about how they are used that their developers use to target ads at their own customers.

Facebook also expands by buying out other companies and services and adding their data to its own. In recent years, the company have acquired the Instagram and WhatsApp services, putting more data about how we share pictures and instant messages at their disposal. More intriguingly, they also acquired virtual reality headset manufacturers Oculus. Some commentators have said this shows Facebook are interested in developing services to let us interact with each other in virtual reality, rather than simply on flat screens. Monitoring our behaviour in these new, immersive virtual worlds will undoubtedly be a very valuable source of new data in the near future.

## **What Were The Results?**

Facebook's tactic of leveraging their huge wealth of consumer data to sell advertising space led to their taking a 24% share of the US online display ads market in 2014, and generating \$5.3 billion in revenue

from ad sales. By 2017 this has been forecasted to be a 27% share, worth over \$10 billion.<sup>2</sup>

### **What Data Was Used?**

Facebook, together with its users, generates its own data. Users upload 2.5 million pieces of content every minute. This content is analysed for clues about us that can be used to segment us for advertisers. Additionally, they interact with other people's content as well as data stored in Facebook's own databases, which include business listings and databases of films, music, books and TV shows. Whenever we "Like" and share this content, it learns a little bit more about us.

In order to provide privacy, all of this data is anonymized when it is fed into the systems that match businesses with potential customers. All this really means is that your name is removed and replaced with a unique identifying code which can't be traced back to you.

### **What Are The Technical Details?**

Facebook is the most visited Web page in the world after Google's search engine – and the most frequent thing Google is used to search for is Facebook. It is said to account for around 10% of all online traffic. Of course, a Web service of this size requires a huge amount of infrastructure.

Its data centres are filled with its custom-designed servers, built using Intel and AMD chips, and power-saving technology to help cut down the huge cost of keeping so many machines running 24/7. The designs for the server systems have been made available as open-source documentation. Facebook also relies on open-source technology for its software, which is written in PHP and runs MySQL databases. Its programmers created the HipHop for MySQL compiler, which translates PHP code into C++ at runtime, allowing code to be executed far more quickly and reducing CPU load. It uses its own distributed storage system based on Hadoop's HBase platform to manage storage. It is

also known that Facebook makes use of Apache Hive for real-time analytics of user data.

### **Any Challenges That Had To Be Overcome?**

In line with most of the big online service providers, Facebook's biggest challenge has been gaining our trust. At the start, it wasn't unusual to find people who were highly sceptical of entering personal details into any online system, as it was impossible to know with any certainty what would be done with them. Even if every company in the world rigidly abided by the terms of their privacy and data-sharing policies, the most watertight policies in the world are powerless in the face of data loss or theft, such as hacking attacks.

From the start, Facebook attempted to win our trust by showing us they took privacy seriously. As full of holes and references to mysterious and unspecified "third parties" as they may have been, their privacy features were light years ahead of those offered by contemporaries, such as Myspace.

The fact there was at least an illusion of privacy was enough to get a lot of people on board the social media revolution. By default, anything a user shared was shared only with a trusted group of friends, in contrast to Myspace where initially posts were, by default, shared with the world. It also offered switches allowing individual aspects of a person's data to be made public or private. However, there have always been complaints that these options are confusing or difficult to find.

### **What Are The Key Learning Points And Takeaways?**

Facebook has revolutionized the way we communicate with each other online by allowing us to build our own network and choose who we share information about our lives with.

## **BIG DATA IN PRACTICE**

This data holds tremendous value to advertisers, who can use it to precisely target their products and services at people who are, according to statistics, likely to want or need them.

Targeted advertising is particularly useful to small businesses, who can't afford to waste their limited marketing budget paying for exposure to the wrong audience segment.

Gaining the trust of users is essential. Aside from data thefts and such illegal activity, users can become annoyed simply by being subjected to adverts they aren't interested in, too frequently. So it's in Facebook's interests, as well as the advertisers, to match them up effectively.

### **REFERENCES AND FURTHER READING**

1. Statista (2016) Number of monthly active Facebook users worldwide as of 3rd quarter 2015 (in millions), <http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>, accessed 5 January 2016.
2. eMarketer (2015) Facebook and Twitter will take 33% share of US digital display market by 2017, <http://www.emarketer.com/Article/Facebook-Twitter-Will-Take-33-Share-of-US-Digital-Display-Market-by-2017/1012274>, accessed 5 January 2016.

For more information about Facebook and Big Data, visit:

<http://www.wired.com/insights/2014/03/facebook-decade-big-data/>