

# Business Intelligence for the Micro-ISV

## Using Spreadsheets to Support Decision Making

Frank S. Rietta  
Software Developer  
www.rietta.com

### 1 Introduction

#### 1.1 Overview

To some, business intelligence (BI) is a buzzword used in management circles to garner support for an IT project. To others, it may sound like a dressed-up way of saying data mining. BI is neither a product nor a system, but it is a method of collecting data from various sources and turning it into actionable information [4]. BI projects typically implement a decision-support system (DSS) that helps managers make decisions based on facts instead of relying on intuition. Even though full scale BI projects are usually associated with larger organizations who need to use IT systems to effectively communicate cross-functional information, even small businesses can take advantage of some of these BI techniques.

Many organizations do not take advantage of the readily available data about their businesses. Unused data is an untapped source of competitive advantage for most companies [1]. Often web site and payment processor logs are automatically collected but ignored. In order to understand the big picture of a business, it is important to pull the available data from multiple sources into a flexible analytical tool. One such tool is the spreadsheet, such as Microsoft Excel, that is capable of pulling information from multiple sources and performing various analytical reporting functions [9]. This paper will discuss methods for aggregating data from web logs, payment reports, and other data into a spreadsheet.

With the modern prevalence of off-the-shelf software being mixed

with in-house developments, it is as important as ever for business managers to have access to operational and financial reports that integrate information from multiple sources [5]. Spreadsheets provide a convenient tool for this important task as they can quickly pull data from multiple sources and allow the business managers to experiment with information and even make up unique charts without the need to directly manipulate the business database server or servers. Moreover, a savvy spreadsheet user does not have to be an expert programmer or web developer in order to work with the data once it has been retrieved from the primary source.

#### 1.2 Why Business Intelligence

BI is an umbrella term that covers many activities, including data mining and business analysis. A typical BI project will, minimally, include the following steps:

- Identify what information is needed to answer questions in order to help make informed decisions.
- Identify sources of data to be converted to that information.
- Perform data analysis to help understand the raw data.
- Build an extract, transform, and load (ETL) process to gather and cleanse the raw data so that it can be plugged into a decision-support system (DSS).
- Build a DSS, which can be a spreadsheet.

While a formal BI process in a large organization would consist of many more steps, these are enough to get started on the process of converting data to information. Furthermore, even though there are a lot of BI tools, some that are rather expensive, most projects can be done with a little thinking and tools that you either already have or that can be obtained free-of-charge.

### 1.3 Why Spreadsheets

Spreadsheets, such as Microsoft Excel, Gnumeric, Lotus 1-2-3, and Open Office Calc, are good for analyzing data and have numerous computation, statistical, and graphing capabilities. A spreadsheet program lets you enter text, numbers, or formulas in a network of cells so that, by changing the content of certain cells, you can automatically perform an involved set of calculations and update various graphs [8]. Microsoft Excel calls an individual spreadsheet a worksheet and a collection of worksheets forms a workbook.

Spreadsheets are a versatile utility for data analysis and business modeling [9] and, more importantly, are able to pull information from multiple sources, validate data, use hundreds of statistical analysis functions, and use numerous graphing functions.

### 1.4 Sources of Data

Modern businesses use many different information technology systems, many of which produce large quantities of data in the form of logs. Some typical sources of raw data that businesses can easily find include the following:

- Raw web site logs
- Analyzed web site logs
- Registration processor logs
- SQL databases

More exotic sources of useful intelligence data can, and should, undoubtedly be found and used. Data analysis is the biggest challenge to all BI decision-support projects [4]. However, utilizing these typical sources is a good place to start.

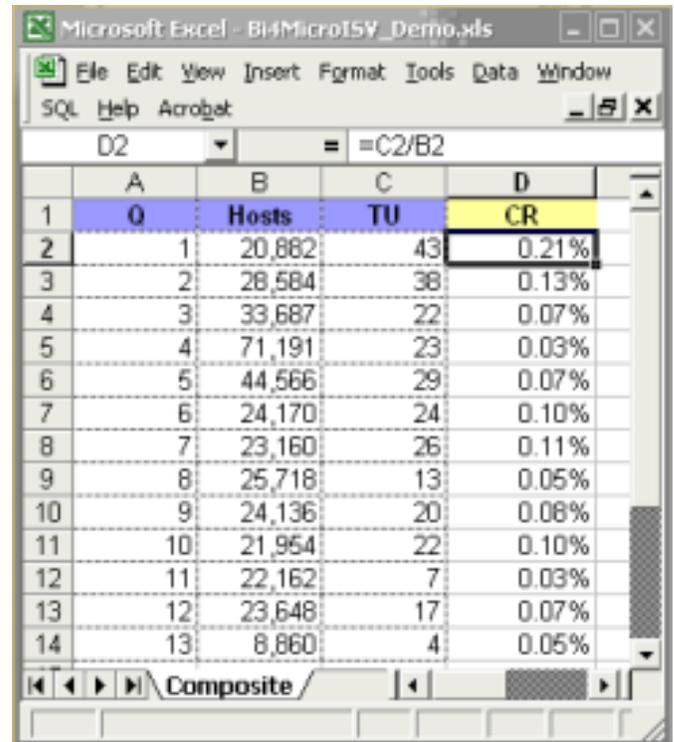
## 2 Metrics Useful for Shareware Businesses

Selling software on the internet is a great hobby for programmers, but to really make a business out of it one must be able to measure how well the business is doing. There are a number of e-metrics that will help show if a product is making money and achieving a proper return on investment [2]. While it is worthwhile to consider all of the various e-metrics, this paper will focus on the conversion ratios and cost per visit measures.

### 2.1 Overall Conversion Ratio (OCR)

The overall conversion ratio (OCR), **Purchases/Total Visits**, is the easiest ratio to calculate, but it is generally hard to make much use of since the visit count typically includes a lot of noise from search engine spiders and other non prospective-customer activity. Figure

Figure 1. The conversion ratio for a web site is the number of purchases over the total number of visitors. This ratio typically includes a lot of noise from search engine spiders and other non-prospective-customer activity.



	A	B	C	D
1	Q	Hosts	TU	CR
2	1	20,882	43	0.21%
3	2	28,584	38	0.13%
4	3	33,687	22	0.07%
5	4	71,191	23	0.03%
6	5	44,566	29	0.07%
7	6	24,170	24	0.10%
8	7	23,160	26	0.11%
9	8	25,718	13	0.05%
10	9	24,136	20	0.08%
11	10	21,954	22	0.10%
12	11	22,162	7	0.03%
13	12	23,648	17	0.07%
14	13	8,860	4	0.05%

1 shows one example of computing the OCR using a spreadsheet.

**NOTE:** *Visits* are not the same thing as *hits*! A single visitor can generate hundreds of hits on a web site. Hits are not prospective customers, so it is far better to count the unique visitors to the web site. However, search engine spiders are also unique visitors, so a more specific measure is needed.

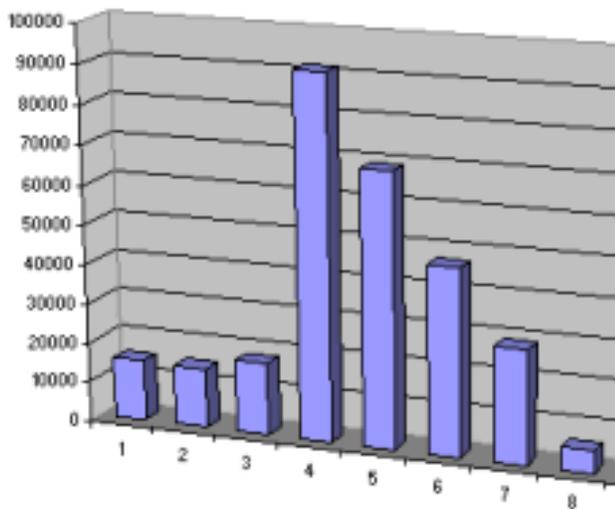
### 2.2 Download Conversion Ratio (DLCR)

The download conversion ratio (DLCR), **Total Downloads/Total Visits**. Assuming you are in the business of offering a try-before-you-buy product to prospective customers, it is fair to say that a unique visitor who downloads a copy of the trial software is a prospective customer. This ratio is more useful than the OCR as there is typically less noise from automated spiders.

### 2.3 Trial Conversion Ratio (TCR)

The trial conversion ration (TCR), **Purchases/Total Downloads**, is a great measure for how well a product sells itself to prospective customers. A low TCR indicates that there may be critical problems

**Figure 2. Plotting a statistic over time, such as quarters, shows how things are changing. The X-axis shows the eight quarters from 1 Jan 2003 to 31 Dec 2004. The Y-axis shows the total number of downloads of EXE and ZIP files. These measures can be plugged into calculations for DLCR and TCR.**



in the program that need to be addressed:

- It does not install properly or is otherwise dead-on-arrival.
- It crashes or behaves incorrectly.
- It has been cracked.
- It is falsely identified as malware by one or more anti-virus or anti-spyware programs.
- It is low quality work for which the prospective customer just does not want to open his wallet.

## 2.4 Cost Per Visit

The cost per visit (CPV), **Total Marketing Costs/Total Visitors**, is an important measure for those spending advertising money to generate traffic. It is a good idea to track this statistic on a per-page or per-product basis. Price Per Click (PPC) is a related measure generally associated with advertisement services such as Google's AdWords and similar programs.

## 3 Understanding Data over Time

The useful ratios introduced in the previous section are all examples of statistical measures. Each is applied to data from a particular time period and is therefore a static indicator. However, the spreadsheet makes it easier to plot data from multiple time periods. Figure 2 shows an example of how the total sum of downloads

varies among quarters. This type of data is already becoming information that tells a story of how the business is functioning.

## 4 Generic Steps for a Mini-BI Project

### 4.1 Collecting Data

The technical process of collecting data is often called "data mining" [3]. The idea is to construct an ETL process to extract and clean the data from one or more sources so that it can be used together to compute the information needed for business analysis.

When using spreadsheets, one typically wants to create a worksheet for each source of data. Sometimes the source data is already in a form that is directly usable in this form. At other times, a scripted program might need to be used to convert the raw data into a tabular form. Once the data is ready for the worksheets, the spreadsheet application has several methods for importing the data.

Microsoft Excel has three main external data import functions:

- Web Queries that pull data from tables on a web page
- Database Queries that pull data from relational databases
- Import Text Files which allows data to be gleaned from ASCII text files

#### 4.1.1 Formula to Convert Dates to Quarters

Raw data frequently contains fixed dates and times. The built-in spreadsheet functions can be used to convert the dates to a series of quarters. For instance, the following cell formula converts the fixed dates in column B2 to an integer representing the quarter:

$$=INT((MONTH(B2)+2)/3) + ((YEAR(B2) - 2003) * 4)$$

The "2003" constant indicates that the series starts in January, 2003. This formula can be easily tweaked as needed.

#### 4.1.2 Web Query for Purchase History

The web query is useful for pulling the payment processor's report into a worksheet. Once the data is in the worksheet, the dates are converted to quarters and a pivot table is used to summarize the sales by quarters.

#### 4.1.3 Database Query for Internal Information

When a relational database contains valuable information, one can use SQL queries to pull data into a worksheet [5]. A combination of SQL functions and spreadsheet functions can be used to manipulate

the cell data as needed.

#### 4.1.4 Text Import for Download Data

Suppose that we have monthly web site log analysis reports and need to find the number of unique visitors that downloaded particular file types, such as EXE and ZIP. A shell script can be used to extract and clean the data and write a comma separated values (CSV) file. The tabulated data in the CSV file can then be imported into a worksheet for analysis.

## 4.2 Analyzing the Data

In general, statistical analysis is concerned with the analysis, interpretation, and presentation of data. Numerous methods exist for describing data and also showing the nature of changes over time.

Descriptive statistics, such as the arithmetic mean, median, and mode are often used to summarize simple data sets. When understanding variability is important, common methods for measuring quantitative data are variance, standard deviation, range, interquartile range, and absolute deviation [7].

When trends are present in the data, descriptive statistics are not useful. In these cases, the moving average and linear regression analysis are useful for time-series forecasting [6].

Microsoft Excel and other spreadsheet programs provide the functionality for both basic and advanced statistical analysis and graphing. A DSS usually consists of a model built upon regression analysis and mathematical modeling on one or more data sets.

## 4.3 Scenarios and Goal-Seeking

A well implemented DSS will help show how current trends play out in the future given various situations. Excel provides functionality to test multiple scenarios and find ways to adjust the inputs so that a particular output is given. Scenarios, Goal Seek, Monte Carlo simulations, and other statistical functions are all options available to help understand what may happen to the business [9, 6].

## 5 Further Reading

This short paper has touched upon on some of the possibilities and benefits that a small software company can realize through business intelligence processes. Each business has a substantial data at the available. Those who convert that data into actionable information gain considerable competitive advantage [1, 4].

Moss and Atre [4] and Witten and Frank [3] are good places to get

started for a deeper understanding of business intelligence and data mining. Winston [9] is a very valuable reference for doing analysis and modeling with spreadsheets.

For advanced BI projects, there are a number of BI products available at various prices. However, one should not underestimate the value of the spreadsheet. Spreadsheets are a versatile utility for data analysis, business modeling, and decision support.

## 6 Acknowledgments

Special thanks to Dr. Alok Srivastava, professor of Managerial Sciences at Georgia State University, for introducing me to the broader implications of business intelligence and decision support systems.

## 7 References

- [1] A. Dragoon. Business intelligence gets smart(er). <http://www.cio.com/archive/091503/smart.html>, 2003. [Online; accessed 29-April-2006].
- [2] B. Drolias. Shareware marketing e-metrics. [http://www.shareware-marketing.net/shareware\\_marketing\\_e-metrics.html](http://www.shareware-marketing.net/shareware_marketing_e-metrics.html), 2003.
- [3] E. F. Ian H. Witten. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. Morgan Kaufmann, San Francisco, CA, 2005.
- [4] L. T. Moss and S. Atre. *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. Addison-Wesley Professional, Boston, MA, 2003.
- [5] F. S. Rietta. Business analysis of web application information. [http://www.rietta.com/whitepapers/mysql\\_webapps\\_excel.html](http://www.rietta.com/whitepapers/mysql_webapps_excel.html), 2005.
- [6] C. Z. S. Christian Albright, Wayne L. Winston. *Data Analysis and Decision Making: With Microsoft Excel*. Thomas South-Western, Mason, OH, 2006.
- [7] Wikipedia. Descriptive statistics — Wikipedia, the free encyclopedia. [http://en.wikipedia.org/w/index.php?title=Descriptive\\_statistics&oldid=49874313](http://en.wikipedia.org/w/index.php?title=Descriptive_statistics&oldid=49874313), 2006. [Online; accessed 29-April-2006].
- [8] G. Williams. Lotus development corporations 1-2-3. <http://www.aresluna.org/attached/computerhistory/articles/spreadsheets/lotus123review>, 1982.
- [9] W. L. Winston. *Microsoft Excel Data Analysis and Business Modeling*. Microsoft Press, Redmond, WA, 2004.