Teaching Case

Analyzing Security Breaches in the U.S.: A Business Analytics Case-Study

Rachida F. Parks
rfparks@ualr.edu
University of Arkansas at Little Rock
Little Rock, AR

Lascelles Adams
adamsl@cookman.edu
Bethune-Cookman University
Daytona Beach, FL, USA

Abstract

This is a real-world applicable case-study and includes background information, functional organization requirements, and real data. Business analytics has been defined as the technologies, skills, and practices needed to iteratively investigate historical performance to gain insight or spot trends. You are asked to utilize/apply critical thinking skills that will produce measurable insights from historical performance data that can be transformed into actionable insights. By critical analysis, reporting and visualization, you will engage with the three major analytic activities: (1) extract, transform and load (ETL) the data; (2) create reports and visualization graphs using business intelligence tools (e.g., IBM Cognos Insight, Tableau, Excel, SQL Server Reporting Services (SSRS)); and (3) engage in critical thinking to identify actionable items that will assist with decision making or recommendations. The Privacy Rights Clearinghouse (PRC) Chronology of Data Breaches reported more than 800 million records breached from over 4500 data breaches since 2005. Security breaches continue to increase: Therefore, there is an urgent need to analyze their patterns and provide meaningful insights to support better decision making.

Keywords: Information security, Business intelligence, Business analytics, Teaching case, Critical thinking

1. CASE SUMMARY

Data breaches have gone mainstream. Whether it is Sony PlayStation Network, Target, or Anthem Health attack, more and more consumers are receiving notifications from companies stating that sensitive, personally identifiable information has been exposed and possibly compromised. Across all industries, as the number of data breaches has increased, eliminating data breaches and protecting business-critical data remains a top priority as well as government's interest in protecting its citizens.

The Bureau of Consumer Protection, Information Security Commission (ISC) Office recently organized a task force seeking to understand the nature of data breaches across the government, and selected industries including: retail, financial and insurance, and healthcare. The task force is seeking your help to collect and analyze the data obtained from the Privacy Rights Clearinghouse Chronology of Data Breaches.

Table 1. Organization Type	
Code	***Description***
BSO	Businesses Other
BSF	Businesses Financial and Insurance Services
BSR	Businesses Retail/Merchant
EDU	Educational Institutions
GOV	Government and Military
MED	Healthcare Medical Providers
NGO	Nonprofit Organizations

Privacy Rights Clearinghouse (PRC), accessible via http://www.privacyrights.org/data-breach, is a nonprofit corporation in California that was established in 1992. PRC keeps up-to-date information of data breaches across all industries and the government within the US. PRC aims to provide timely and historical information on data breaches and to educate stakeholders (e.g., consumers, businesses, and policymakers) of currents trends in data breaches. PRC reported more than 800 million records breached from over 4500 data breaches since 2005. These breaches span across financial institutions, retail, educational institutions, government and military, healthcare, non-profit and other businesses (see table 1). PRC also provides data about different types of breaches as outlined in table 2.

Table 2. Type of Breaches		
Code	***Description***	***How it occurred***
DISC	Unintended Disclosure	Information made public via web
HACK	Hacking or Malware	Electronic entry by outside party
CARD	Payment Card Fraud	Fraud via debit or credit cards accomplished by means other than hacking
INSD	Insider	Someone with legitimate access intentionally breaches
PHYS	Physical Loss	Non electronic records lost, discarded, or stolen
PORT	Portable device	Data lost, stolen or discarded through electronic devices
STAT	Stationary device	Lost, stolen or discarded stationary electronic device
UNKN	Unknown or other	Unknown

Data breaches are endangering the privacy and confidentiality of consumers and resulting in dire organizational consequences, such as reputation damage, monetary penalties, and civil and criminal liabilities. Organization leaders recognize the importance of keeping track of breaches trends and their impacts.

Therefore, there is an urgent need to have tools that communicate patterns difficult to see. As part of the task force, your team will produce a report to help the bureau understand the patterns of data breaches in the government and the selected industries identified above.

The case study includes a set of functional requirements (technical, business, and critical thinking), the data set, and the business intelligence tools to be used. Below are the detailed technical and business functional requirements.

2. TECHNICAL REQUIREMENTS

Extract – Transform – Load Process

Clean, meaningful and useful data play a very important role in getting better analysis and more insightful results. The ETL process consists of three steps; extract, transform, and load. First, the extract step consists of pulling data from a source. Next, the transform step involves converting the raw data into a form that can be loaded into a target system for further processing. Lastly, the load step entails putting the transformed data into the target system. In the end the ETL process take source data, cleans it, and integrates it into a target system.

Before beginning the analysis, you will perform an ETL process set using PRC data.

- Extraction: extract data from the PRC website. The dataset can be downloaded as a CSV file from http://www.privacyrights.org/data.
- Transformation: transform data into a suitable format as required by the target system. The transformation process takes place in Excel by saving the CSV file into an Excel file. You can delete the columns for street address, the name of the business, postal code, city, description, and a few unnamed columns. You can remove either the "total records" or "records breached" column to eliminate redundancy. In addition, a few columns need to have their headers changed. For example

"location" to "city" and the "state" column to be changed from ",". Additional columns can be derived from the "Date" column such as the year and the quarter. An example of a finalized and transformed dataset is provided in Appendix A.

- Loading: load the data into the business intelligence (BI) application. This starts by loading or importing the transformed file into your BI tool. During the loading, you have to identify dimensions and measures among the available attributes. You can choose several dimensions such as entity, type, state, year, and quarter. An example of measures is the number of breached records and the count of breaches

Business Intelligence Tools

The ISC Office and organization leaders want a tool that allows them to know the state of information breaches across different industries, impacted geographical areas (States with higher breaches) and type of breaches by which they are threatened.

BI tools are used to retrieve, analyze, and report data to support decision making by providing meaningful insights. Several BI tools are available and some providers offer trial versions or free student versions. Some of the BI tools currently available are EXCEL, SSRS, Tableau, and IBM Cognos Insights.

You will need to select the appropriate BI software to use and install it; keeping in mind that some BI tools are only windows compatible.

3. BUSINESS REQUIREMENTS

Reporting or Descriptive analytics

Descriptive analysis answers the "What" questions (See Appendix B) and provides a view of both current and historical results. Descriptive analytics tells the business how they are performing and help identify key issues in their current performances. Using the dataset provided by PRC, you need to address the requirements outlined below:

1. **Total Number of Breaches:** This report should outline the number of breaches by year; broken down by government and the selected industries (e.g., retail, financial and healthcare). The visual representation should allow the changes and trends throughout the years among the above-mentioned industries and government. Advanced report options include exploring any seasonal trends (e.g., quarterly analysis).

2. **Type of Data Breaches Report:** This report compares the different types of data breaches reported. This report can be broken into two sub-reports. The first sub-report will show different breach trends across all different years provided, while the second sub-report will represent different type of breaches broken down by industry to be able to see the most prevalent type of breaches within different industries. Advanced report options should combine the findings from the previous report (total number of breaches) to the type of data breaches. This will allow comparison of the highest occurrences of breaches to the type of breaches reported.

3. **Geographic Location of Breaches Report:** This report describes the geographic locations (States) associated with the breaches in the United States. Advanced report options can integrate this geographic representation to be interactively displayed on a U.S. map.

4. **Citizens Impacted by Breaches Report**: This report provides a comparison of the total individuals impacted by the breaches in comparison to the total number of breach occurrences. Advanced report options should consider adding the type of breaches and their location to this report for more meaningful insights.

5. **Cost Analysis of Breaches Report:** This is an advanced report options where you can include a new measure called "Estimated Cost of a Breach". This measure is calculated based on the average cost per breach published by the Ponemon Institute. By multiplying the records breached by the average cost per record you will be able to determine an estimated cost for an entire breach for an entity. In your reports, you can analyze (1) the total cost of breaches per industry, (2) the cost per type of breach and, (3) cost of breaches over the years.

Visual Communication

Organization leaders need to be able to make quick and accurate decisions. Therefore, a need for simple graphs that stand out and aid in their decision making is very important. These graphs provide insights into the data and address the

key issues of the problem identified in descriptive analytics reports.

You should create visual representation of all the above reports along with a short analysis of the reports.

4. ACTIONABLE INSIGHTS: CRITICAL AND ANALYTICAL THINKING

While having descriptive and historical reports is very important, business leaders are looking for the needle in the haystack. Meaning they want you to provide them with recommendations that are actionable. Your recommendations should be based on the insights gathered through the reports and supplemented by recommendations to safeguards against data breaches. These preventative recommendations can be:

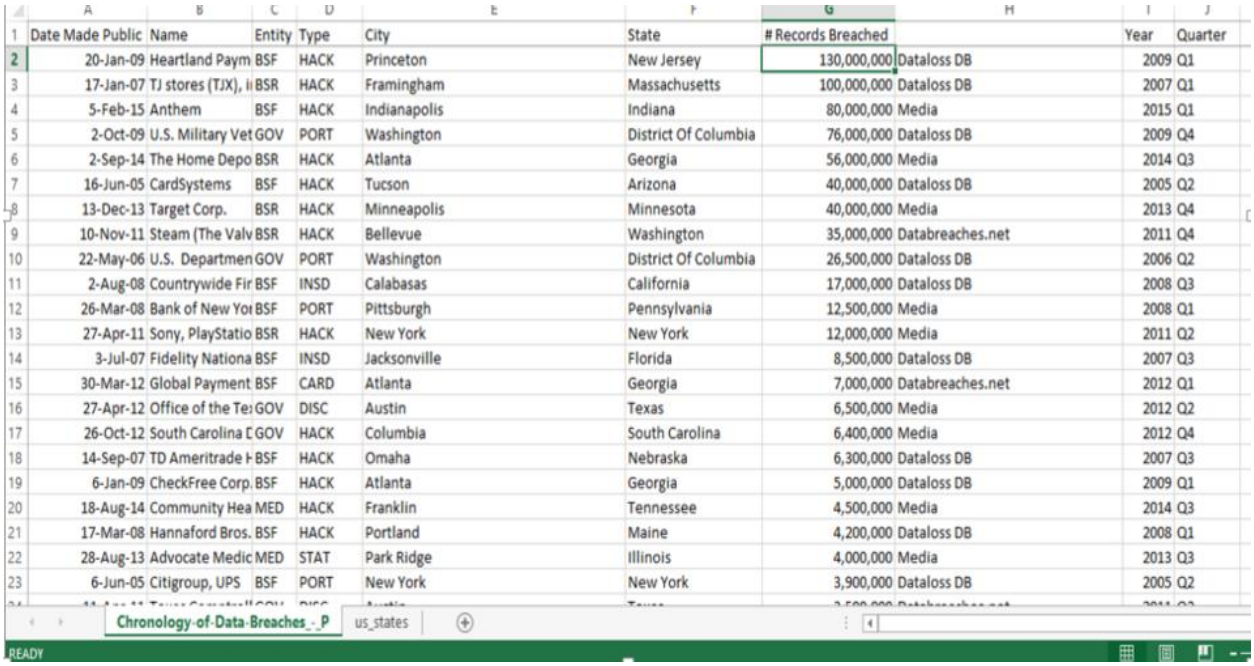
- Pertinent to specific types of breaches.
- Specific to a particular industry or government or span across all the industries and government.
- Either technical recommendations (e.g., encryption), human recommendations (e.g., education/training, awareness, social engineering), or policies.

5. CONCLUSION

As data breaches continue to increase, countermeasures can only be effective if they align with what is causing the breaches. Therefore, a thorough analysis of breaches and their trends is crucial. A real and up-to-date data set of security breaches is provided by the privacy rights clearinghouse. The Information Security Commission Office is seeking your help to clean the data, build and analyze reports, and ultimately provide insights into the state of data breaches in the U.S. for the purpose of better preventative safeguards.

Note: Teaching Notes and Case Supplements are available by contacting the authors

Appendix A – Sample of a Transformed Dataset



	A	B	C	D	E	F	G	H	I	J
1	Date Made Public	Name	Entity	Type	City	State	# Records Breached		Year	Quarter
2	20-Jan-09	Heartland Paym	BSF	HACK	Princeton	New Jersey	130,000,000	Dataloss DB	2009	Q1
3	17-Jan-07	TJ stores (TJX), i	BSR	HACK	Framingham	Massachusetts	100,000,000	Dataloss DB	2007	Q1
4	5-Feb-15	Anthem	BSF	HACK	Indianapolis	Indiana	80,000,000	Media	2015	Q1
5	2-Oct-09	U.S. Military Vet	GOV	PORT	Washington	District Of Columbia	76,000,000	Dataloss DB	2009	Q4
6	2-Sep-14	The Home Depo	BSR	HACK	Atlanta	Georgia	56,000,000	Media	2014	Q3
7	16-Jun-05	CardSystems	BSF	HACK	Tucson	Arizona	40,000,000	Dataloss DB	2005	Q2
8	13-Dec-13	Target Corp.	BSR	HACK	Minneapolis	Minnesota	40,000,000	Media	2013	Q4
9	10-Nov-11	Steam (The Valv	BSR	HACK	Bellevue	Washington	35,000,000	Databreaches.net	2011	Q4
10	22-May-06	U.S. Departmen	GOV	PORT	Washington	District Of Columbia	26,500,000	Dataloss DB	2006	Q2
11	2-Aug-08	Countrywide Fir	BSF	INSD	Calabasas	California	17,000,000	Dataloss DB	2008	Q3
12	26-Mar-08	Bank of New Yo	BSF	PORT	Pittsburgh	Pennsylvania	12,500,000	Media	2008	Q1
13	27-Apr-11	Sony, PlayStatio	BSR	HACK	New York	New York	12,000,000	Media	2011	Q2
14	3-Jul-07	Fidelity Nationa	BSF	INSD	Jacksonville	Florida	8,500,000	Dataloss DB	2007	Q3
15	30-Mar-12	Global Payment	BSF	CARD	Atlanta	Georgia	7,000,000	Databreaches.net	2012	Q1
16	27-Apr-12	Office of the Te	GOV	DISC	Austin	Texas	6,500,000	Media	2012	Q2
17	26-Oct-12	South Carolina D	GOV	HACK	Columbia	South Carolina	6,400,000	Media	2012	Q4
18	14-Sep-07	TD Ameritrade H	BSF	HACK	Omaha	Nebraska	6,300,000	Dataloss DB	2007	Q3
19	6-Jan-09	CheckFree Corp.	BSF	HACK	Atlanta	Georgia	5,000,000	Dataloss DB	2009	Q1
20	18-Aug-14	Community Hea	MED	HACK	Franklin	Tennessee	4,500,000	Media	2014	Q3
21	17-Mar-08	Hannaford Bros.	BSF	HACK	Portland	Maine	4,200,000	Dataloss DB	2008	Q1
22	28-Aug-13	Advocate Medic	MED	STAT	Park Ridge	Illinois	4,000,000	Media	2013	Q3
23	6-Jun-05	Citigroup, UPS	BSF	PORT	New York	New York	3,900,000	Dataloss DB	2005	Q2

Appendix B -- Questions to consider in your analysis

Type of question	Sample question pertaining to the data set provided
WHAT questions Descriptive Analytics	• What type of breaches? • What is the breakdown of breaches per year? • What the breakdown of breaches per geographic location? • What entities/type of entities are impacted the most? • What are the seasonal trends if any (quarterly analysis)? • What is the total number of breaches in comparison to the total of individuals impacted by the breaches?
WHY Questions Getting to the meat	These types of questions allow you to further analyze your findings. It involves more depth in your analysis/discussion getting and may require the use of other data sets. • Why is a particular breach type higher than others? • Why are breaches the highest in certain geographical areas (States)? • Why is it important to focus on the number of breaches rather than the number of individuals impacted or vice versa? • Why are breaches lower in certain States? (may have to check the stringency of their privacy/security policies)

HOW Questions
Prescriptive Analytics
(check figure below)
	• How do the above findings and discussion provide a competitive advantage or improve decision making?
	• How can one achieve the best outcome using the following?
	  ○ Technical countermeasures
	  ○ Physical countermeasures
	  ○ Policies and regulations
	  ○ Training and education