Vintage Analytics and Data Warehouse Design

Ido Millet, Department of Management Information Systems, Pennsylvania State University, Erie, PA, USA
Syed S. Andaleeb, Department of Marketing, Pennsylvania State University, Erie, PA, USA
John L. Fizel, Department of Economics, Pennsylvania State University, Erie, PA, USA

ABSTRACT

Vintage Analytics is a useful technique for a variety of domains where performance depends on experience or age. However, this technique might be underutilized due to lack of awareness and due to difficulties in data preparation. This paper provides examples of Vintage Analytics, including one detailed scenario demonstrating benefits and typical challenges in the context of a sales database. It then proposes data warehouse design guidelines to address these difficulties.

Keywords: Business Intelligence, Credit Risk Analysis, Customer Value Analysis, Data Warehousing, Mini-Dimensions, Slowly Changing Dimensions, Survival Analysis, Time Dimension, Vintage Analysis, Vintage Models

1. INTRODUCTION

Vintage charts have long been used to select wines, and to consume them at their optimal age. While empirical evidence (Weil, 2011) casts doubts about the reliability of wine charts, the same approach of tracking performance by age and origination date has proven valuable for other domains such as credit risk (Siarka, 2011) and healthcare (Collett, 2003). While each domain may use a different name, let us define Vintage Analytics as the analysis of how age and experience influence performance.

Vintage Analytics (VA) is a useful but challenging technique. When discussing “Time Issues” in data mart design, Kimball (1997) highlights it as an advanced application need that is “seen repeatedly in data warehouse environments.” As a case

DOI: 10.4018/ijbir.2014040104
in point, Kimball describes the need to analyze all customers who had their credit limit raised to $1,000:

*Once this triggering event occurs, you want to study the behavior of this cohort group. You want to measure their purchases and their payments as a function of the time after the granting of credit. You may want to ask what is the average time until these customers have a credit default (if they do).* (Kimball, 1997)

According to Kimball, even well-designed data marts are not suited to this type of analysis, and he knows of "no application development or decision-support environment that handles this kind of application automatically" (1997).

VA is difficult because experience and age change with time and require calculations. Therefore, to facilitate VA we should consider alternative approaches for preparing the data for analysis. For example, to support data mining of factors influencing online procurement auctions, Millet, Parente, Fizel, & Venkataraman (2004) suggest that bid times "could be stored in terms of elapsed time since the start of the auction or before its end, instead of clock time" (p. 178).

Perhaps due to lack of awareness, VA might be underutilized even among experienced business intelligence (BI) practitioners. A convenience sample of 8,992 reports gathered from seven organizations with advanced BI skills identified age-related calculations in only 2.4 percent of the reports. Furthermore, the percent of reports with age calculations varied widely across these seven organizations, ranging from a low of 0.2 percent to a high of 5.5 percent.

Another source of difficulty in using VA, is methodological complexity. Allison (1984, p. 9) discusses “censoring” (missing data cases that are common in survival analysis) and “time-varying explanatory variables” as two typical features that “create major problems for standard statistical methods.” Rahimi (2011) provides a good example of how age, period, and cohort effects are analyzed to show that the Nazi occupation of Norway during World War II was followed by reduced risks for some types of cancer. Rather than addressing advanced statistical and methodological issues, this paper aims to promote VA through improved awareness and data preparation methods.

The first section of this paper provides examples of VA. This is then followed by a case study of applying VA to typical sales data. The case can be used by BI instructors to demonstrate the implementation and value of VA. The case also demonstrates solutions for typical difficulties encountered when deriving age and experience measures. The sample database, queries, and reports used for this case are available for download upon request.

The last section of this paper proposes data warehouse design guidelines for facilitating VA. Some of these guidelines are related to the concepts of Slowly Changing Dimensions and Mini-Dimensions, as advanced by Kimball (1996).
How Business Intelligence Creates Value: An Empirical Investigation
[www.igi-global.com/article/how-business-intelligence-creates-value/83476?camid=4v1a](www.igi-global.com/article/how-business-intelligence-creates-value/83476?camid=4v1a)