Option #1: Retail BI Use Case: Analytics Application

Scott Miner

Colorado State University – Global Campus

**Option #1: Retail BI Use Case: Analytics Application**

**Part A**

The Critical Thinking task in Module 3 (n.d.) asks the student to provide detailed information on a business intelligence solution system that will answer the questions posed by one of the analytic applications in Table 1.1 of *Business intelligence, analytics, and data science: A managerial perspective.* I chose to provide a solution that answers the questions posed by Market Basket Analysis (MBA).

**MBA Basics**

MBA allows analysts to understand patterns in how products are purchased together. MBA answers questions such as, (a) “What products should I combine to create a bundle offer?,” (b) “Should I combine products based on slow-moving and fast-moving characteristics?,” and (c) “Should I create a bundle from the same category or different category line?” (Sharda, Delen, & Turban, n.d., p. 33).

MBA involves three key measures: (a) **frequency** measures how often a product is purchased, (b) **support** measures the probability of product purchases compared to overall purchases, and (c) **confidence** measures the conditional probability of purchases between individual products. Using these measures, analysts discover new insights into customers’ buying behaviors (“Market Basket Analysis,” 2008).

**Benefits of MBA**

MBA offers many benefits to the organization:

* **Profitable advertising and promotions.** MBA allows retailers to understand when and where discounts will make an effect on sales, allowing marketers to reduce unnecessary markdowns and increase gross margins.
* **Precise customer targeting improves ROI.** MBA allows retailers to increase incremental profits by promoting the right products to the right customers at the right times.
* **Better loyalty card promotions.** Time-variant data available in data warehouses (DWs) allow analysts to make use of longitudinal MBA, analyzing customers’ lifetime purchasing behavior, shopping frequencies, and shifts in shopping categories.
* **Attract more traffic into the store.** Once MBA is used to determine what brings a customer into a store, it can be used to bring that customer back into the store.
* **Increase the size and value of the market basket.** Retailers can identify, for instance, if any customer segments have stopped purchasing a product recently. Retailers can then market to these customers accordingly (Gordon, 2008).

**MBA Success Stories**

Instances of retailers employing MBA and DW architectures are abundant in the literature. Dasari and Kurhekar (2011) describe the study of Fresh Greens, a vegetable retailer in Bangalore, India. Fresh Greens used MBA to determine that customers frequently bought Tomato hybrid together with Onion economy. The retailer then applied a discounted bundle price to these products and placed them together on the shelf, increasing the sales of both products. Fresh Greens used MBA to improve its margins by bundling regularly purchased products with those having higher margins.

Similarly, Gordon (2008) describes the example of a toy retailer who used MBA to determine whether it was worth selling a product with a low-margin (e.g., video-game system) without also selling a product with a high-margin (e.g., game accessories and software). This retailer used data obtained through its loyalty card program to create a campaign that led to customer retention by targeting customers likely to purchase product accessories. Further, Breslin (2004) describes a retail chain that used MBA to reveal that new fathers bought diapers together with alcohol. By grouping these products on store shelves, retailers made sales skyrocket. The architecture of the system, which incorporated a DW, is credited with the discovery.

**Characteristics of a DW**

To create actionable insights using MBA it is imperative retailers incorporate a DW. Parzinger (2001) describes a DW as “a managed database in which the data is subject-oriented, integrated, time-variant, and nonvolatile” (p. 11). The following four characteristics distinguish a DW from a transactional database:

* **Subject oriented.** DWs contain data that is oriented to decision-making. Subject areas in a DW may span multiple business units, allowing for redundant data as this improves the run time of ad hoc analytical queries.
* **Data Integration.** Because DWs centralize data from multiple disparate sources, IT personnel must “cleanse” the data before loading it into the data warehouse, using extraction, transform, and loading (ETL) techniques to standardize the data for decision-making.
* **Time-Variant.** The key factor that distinguishes a DW from a transactional database may be time-variance. Data in a transactional database last 60 to 90 days, while data in a DW typically last five to ten years.
* **Nonvolatile.** DWs only allow for data loading and data access. Inserts, updates, and changes, as are customary to transactional systems, are not permitted in DWs (Parzinger, 2001).

**Benefits of a DW**

Now that we know some of the characteristics of DWs, let us describe several of the benefits:

* **Simplification of data access.** Managers and executives who may not have advanced technical skills can access the data needed for key decisions without having to ask IT personnel to intervene.
* **A single image of the business.** DWs allow for one centralized version of the truth, enabling decision-makers to trust their source of information and focus on key decisions.
* **Better and more timely information.** Enhanced data quality means enhanced decision making at a faster rate.
* **Enhanced system performance.** By offloading complex analytical queries from transactional systems to the DW, DWs can enhance the performance of transactional systems (Parzinger, 2001).

**DW Models**

**The Inmon Model.**

The Inmon model stresses a top-down development strategy and expands upon traditional tools for operational database development, including ERDs. Inmon’s model is also known as the hub-and-spoke architecture and was found to be the most prevalent architecture at 39% in a study conducted by Ariyachandra and Watson (2008) of 454 database warehouse personnel. It is the costliest of all the models to implement, as it is a complete overhaul of the entire enterprise, with an atomic data warehouse supporting one or more dependent data marts (DMs). Each DM contains consistent data summarized for various business units. End-users have less of a say in the development process, which relies more heavily on IT (Breslin, 2004).

**The Kimball Model.**

The Kimball model is also known as the bus architecture and uses dimensional modeling with tables rather than entity-attribute data models. Tables are either fact or dimension tables. Fact tables (e.g., sales) contain metrics (e.g., frequency) and dimensional tables (e.g., product, date, store, category, promotion) contain attributes (e.g., description, brand, department, weight, shelf life, etc.). The data in dimensional tables contain attributes of data in the fact tables in denormalized forms. Kimball’s development process incorporates a bottom-up approach, building one data mart per business unit. The model requires all DMs to be modeled within conformed dimensions, and integration is achieved through the data bus. The DW is the sum of all the organization’s DMs (Breslin, 2004).

**Conclusion**

Which architecture is the most effective? In the research study conducted by Ariyachandra and Watson (2008), Inmon’s model and Kimball’s model were found to perform similarly on measures of (a) Information Quality, (b) System Quality, (c) Individual Impacts, and (d) Organizational Impacts. The hub-and-spoke architecture is the costliest to implement and requires the most time and planning. The bus architecture generates a quicker ROI. Parzinger (2001) writes, “Sustained competitive advantage can come only from an organization’s knowledge workers interpreting, sharing, and acting on the information and knowledge that a data warehouse can provide” (p. 15). In effect, the DW is only as valuable as the insights it generates that lead to long term competitive advantages and increased market shares. MBA allows retailers to achieve these capabilities by analyzing product affinities. Retailers can then provide a more personalized shopping experience. This BI solution system answers the questions MBA poses.

**Part B: SAS code – Reading and Importing Files**

Part B of the assignment asks the student to write SAS code to import a .CSV file and read an SAS data file. I provided the code used to generate the program, as well as output from the program. I provided additional verification output to confirm the .CSV file was imported. Figure 2 displays an alphabetical list of variables in the .CSV file. Lastly, Figure 3 provides a screenshot of the program execution that shows the file explorer window, the code window, the results viewer, and the log output side-by-side. From this overview screenshot, we can observe no errors are present in the log output.

**SAS Code**

/\*\*\*\*\*\*\*\*\*\*\*

Step A: Read a file called somedata.sas7bdat

Note: MYSASLIB points to '/folders/myfolders/sasuser.v94/SASDATA/'

\*\*\*\*\*\*\*\*\*\*\*/

PROC MEANS DATA=MYSASLIB.SOMEDATA;

RUN;

/\*\*\*\*\*\*\*\*\*\*\*

Step B: Import an Excel data file (CSV) called CARSMPG.CSV

\*\*\*\*\*\*\*\*\*\*\*/

PROC IMPORT OUT=WORK.MPG\_FOR\_CARS

DATAFILE="/folders/myfolders/sasuser.v94/SASDATA/CARSMPG.csv" DBMS=CSV

REPLACE;

GETNAMES=YES;

DATAROW=2;

RUN;

/\*\*\*\*\*\*\*\*\*\*\*

Verify that input (CSV) was imported via visual output

\*\*\*\*\*\*\*\*\*\*\*/

%LET LIBNAME=WORK;

%LET DATASETNAME=MPG\_FOR\_CARS;

ODS SELECT VARIABLES;

PROC DATASETS;

CONTENTS DATA=&LIBNAME..&DATASETNAME;

RUN;

**SAS Output**

**Step A: Output.**

| **Variable** | **Label** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| ID AGE TIME1 TIME2 TIME3 TIME4 STATUS SEX | ID Number Age on Jan 1, 2000 Baseline 6 Months 12 Months 24 Months Socioeconomic Status | 50 50 50 50 50 50 50 50 | 374.2200000 10.4600000 21.2680000 27.4400000 30.4920000 30.8380000 3.9400000 0.4000000 | 167.4983143 2.4261332 1.7169551 2.6590623 3.0255942 3.5307333 1.3311036 0.4948717 | 101.0000000 4.0000000 17.0000000 21.3000000 22.7000000 21.2000000 1.0000000 0 | 604.0000000 15.0000000 24.2000000 32.3000000 35.9000000 36.1000000 5.0000000 1.0000000 |

Figure . PROC MEANS output for somedata.sas7bdata file

**Step B: Output (verification of CSV import).**

| **Alphabetic List of Variables and Attributes** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **#** | **Variable** | **Type** | **Len** | **Format** | **Informat** |
| **6** | AUTOMATIC | Num | 8 | BEST12. | BEST32. |
| **12** | AWD | Num | 8 | BEST12. | BEST32. |
| **1** | BRAND | Char | 10 | $10. | $10. |
| **9** | CITYMPG | Num | 8 | BEST12. | BEST32. |
| **8** | CYLINDERS | Num | 8 | BEST12. | BEST32. |
| **7** | ENGINESIZE | Num | 8 | BEST12. | BEST32. |
| **10** | HWYMPG | Num | 8 | BEST12. | BEST32. |
| **13** | HYBRID | Num | 8 | BEST12. | BEST32. |
| **3** | MINIVAN | Num | 8 | BEST12. | BEST32. |
| **2** | MODEL | Char | 14 | $14. | $14. |
| **5** | PICKUP | Num | 8 | BEST12. | BEST32. |
| **11** | SUV | Num | 8 | BEST12. | BEST32. |
| **4** | WAGON | Num | 8 | BEST12. | BEST32. |

Figure . Alphabetic List of Variables and Attributes for CARSMPG.CSV File

**SAS Overview Screenshot.**

A screenshot of a social media post

Description automatically generated

Figure 3. SAS File Explorer, Code Window, Results, and Log from Part B (importing data)

References

Ariyachandra, T., & Watson, H. J. (2008). Which Data Warehouse Architecture is Best? Communications of the ACM, 51(10), 146–147. <https://doi.org/10.1145/1400181.1400213>

Ariyachandra, T., & Watson, H. J. (2008). Which Data Warehouse Architecture Is Most Successful? Business Intelligence Journal, 11(1), 4–6.

Breslin, M. (2004). Data Warehousing Battle of the Giants: Business Intelligence Journal, 6–20.

Dasari, S., & Kurhekar, A. S. (2011). Retail Analytics Using SAS: Experience of Fresh Greens Pvt. Ltd., Bangalore. IUP Journal of Supply Chain Management, 8(1), 7–22.

Gordon, L. (2008). Leading Practices in Market Basket Analysis. Fact Point Group, 8.

Market Basket Analysis. (2008, April 9). MetaTheory. https://metatheory.wordpress.com/2008/04/09/market-basket-analysis/

Module 3: Critical Thinking. (n.d.). Retrieved December 31, 2019, from <https://csuglobal.instructure.com/courses/16501/assignments/332797>

Parzinger, M. J. (2001). Creating Competitive Advantage Through Data Warehousing. Information Strategy: The Executive’s Journal, 17(4), 10.

Sharda, R., Delen, D., & Turban, E. (n.d.). Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4/e.