

Paper 020-2012

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole

Stephen Overton, Zencos Consulting, Cary, NC, USA

ABSTRACT

Business information can be consumed many ways using the SAS Enterprise Business Intelligence Platform. SAS® BI Dashboard can be used to report high level key performance indicators at-a-glance through the SAS® Information Delivery Portal. Detailed web-based reports can also be surfaced from SAS Information Delivery Portal through SAS® Web Report Studio. This paper will present an information system that integrates the functionality of these tools to answer business questions faster and with a greater understanding of the key drivers of business critical data. This paper will also present the data infrastructure needed to support this type of information system through the use of OLAP technology, effective data architecture, report linking, and information maps.

CONCEPT

Rather than generate flat reports in the SAS Enterprise BI platform, this paper presents a system that uses multiple data storage concepts and reporting layers to present the same information but with much faster results. The basic concept is to break the analysis into a series of linked dashboards and web reports, subset the data through each link, summarize the subset if needed, and continually link to specific data until the final answer is reached. This is shown conceptually in Figure 1. A critical piece to this information system is the data infrastructure which supports it. Without an effective data model, performance can suffer and additional data complexities can arise. In the end, the user should be presented with a much more reasonable number of rows of data – enough to fit on a single page. This keeps the end user in a single, controlled environment and promotes faster decision making. This paper assumes the reader has an advanced understanding of the SAS Enterprise BI platform and knowledge of data modeling techniques.

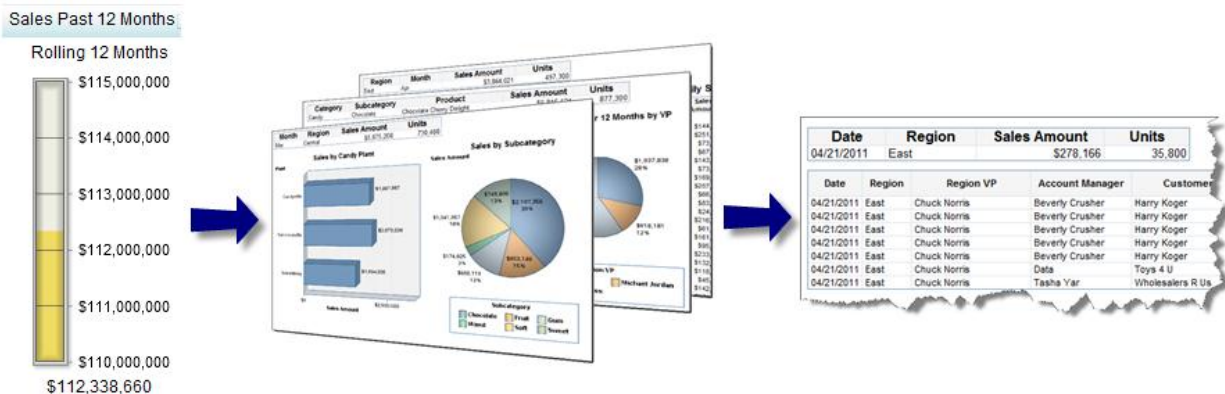


Figure 1. Conceptual Guided Analysis to Key Data

The general idea of the guided drill-through methodology is to think in terms of key measures and logical ways to breakdown these measures. Reporting requirements may not fully leverage the capabilities that come with the SAS Enterprise BI platform due to strict visual requirements. Sometimes business users request reports with a large number of columns and a very large number of rows. In this situation, the end users typically performs their own analysis using whatever means necessary, adding time to arrive at a data-driven conclusion.

EVERYTHING STARTS AND ENDS WITH THE DATA

The latest trend in Information Technology is “big data”. Data is growing at an exponential pace as it becomes more critical to competing effectively using analytical technology such as SAS. The concepts around data management described in this paper are very beneficial for large data repositories. This paper uses small sample datasets from the SAS® Enterprise Guide 4.3 sample library. The candy sales data is transformed into a star schema data structure to demonstrate an efficient structure to store data. This is shown in Figure 2 and used throughout this paper for example reports. The code used to build this sample star schema is referenced at the bottom of this paper.

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

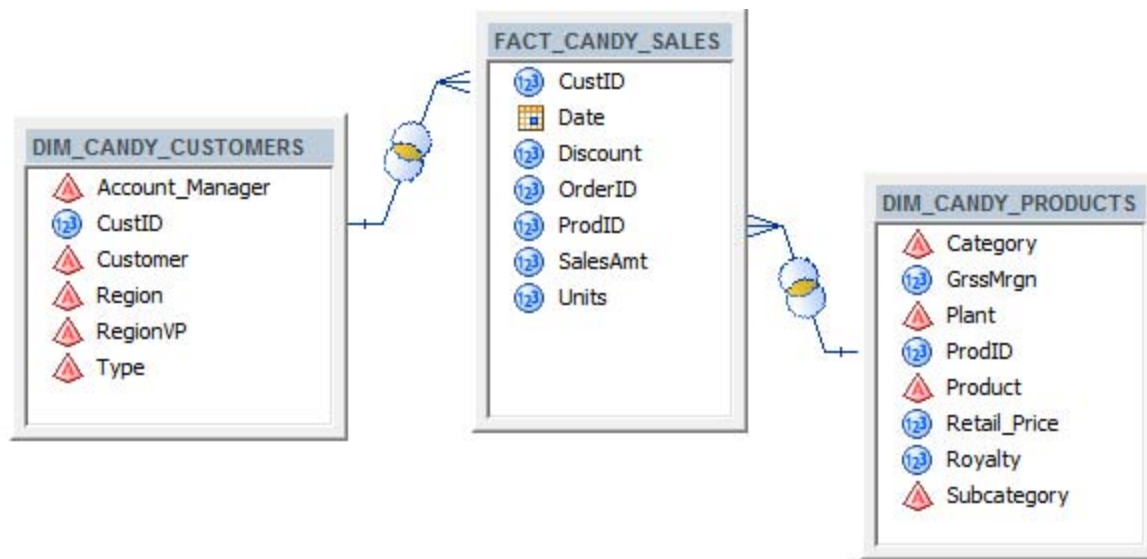


Figure 2. Sample Candy Sales Star Schema Data Model

The star schema data model is an efficient way to store data for reporting because retrieval is very fast. Dimensions provide an easy way to eliminate redundancy. Integrity constraints between fact and dimension tables subset data faster and enforce data quality. Storing data in an enterprise database environment, such as Oracle or SQL Server, can provide even more power because of the higher performance query optimizer and better indexing options available. More information on star schemas and indexing can be found in the References section at the bottom of this paper.

ADVANTAGES OF USING A STAR SCHEMA DATA STRUCTURE IN SAS

Using a star schema design in SAS is beneficial in the OLAP cube build process and accessing the data directly using a relational information map. This paper leverages both types of information maps to show the different uses. OLAP-based information maps are used for higher level summary data. Relational information maps are specifically used when data needs to be subset ahead of time using filter prompts.

BUILDING OLAP CUBES

SAS OLAP Cubes can be built from a star schema using either SAS OLAP Cube Studio® or writing the PROC OLAP code in SAS Enterprise Guide. If the source data is very large, SAS will build the cube much faster using a star schema because the OLAP server has less redundant data to process to build dimensions. Building a cube from a flat detail table will result in longer processing times because the entire dataset has to be scanned for values rather than a much smaller dimension table in a star schema. For example, if the source data contains 1 million records and a Region dimension is defined using a Region column, 1 million records are scanned. If the source data is structured into a star schema, the distinct regions are scanned from a much smaller table. Assuming the dimension table in the star schema only contains regions, this could be a dozen records at the most.

Tip: Keep cardinality low between dimensions and fact tables when designing a star schema. This will dramatically reduce dimension build times in an OLAP cube because of the reduced number of records to scan.

This paper used a sample OLAP cube that summarizes the candy sales data. In environments with large data warehouses or data marts, OLAP technology provides a fast way to report critical data through dashboards or web reports. The structure of this cube is shown in the sample information map defined in Figure 3.

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

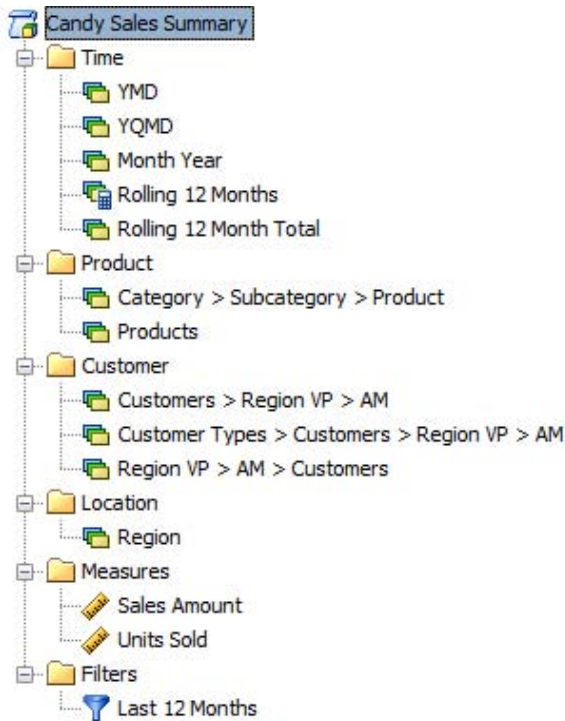


Figure 3. Candy Sales OLAP Cube Structure

DEFINING THE RELATIONAL INFORMATION MAP USED FOR DRILL THROUGH REPORTING

Data Items

The main structure of the relational information map used in this paper is shown in Figure 4. The Data Items shown are mapped to the columns in the candy sales star schema. The data items are logically the same as what is available in the OLAP information map defined in Figure 4.

Tip: Organize data items into key subject areas or focus groups for a more user friendly interface for Web Report Studio report writers. This is also demonstrated in Figure 4.

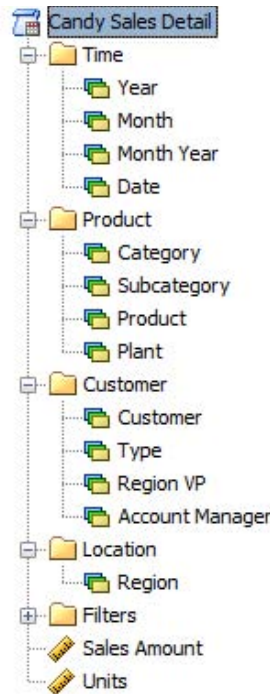


Figure 4. Relational Information Map of Candy Sales Star Schema

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

Filters

A key piece to drilling through data in web reports with report linking are the filter prompts. These provide the inputs for the incoming report to pass data that is selected or drilled upon. These are not possible in OLAP-based information maps.

Some fields may need custom data items created in order for the filter prompt to work properly. In the case of the Year and Month data items shown in Figure 5, the numeric values are converted to character values due to the way SAS Web Report Studio links reports. This is documented more in SAS Sample Note 40009 listed in the References section of this paper.

Tip: Organize special data items, which are specialized for filter prompts within a folder as shown in Figure 5.

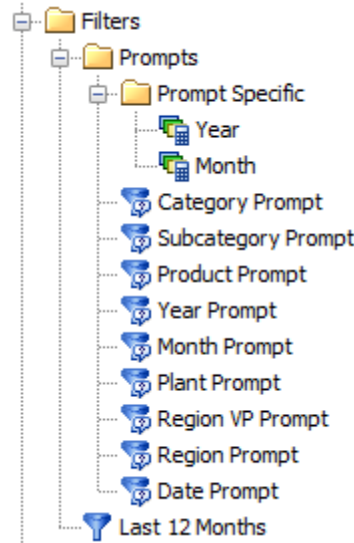


Figure 5. Filters Defined in Relational Information Map of Candy Sales Star Schema

EFFICIENT QUERIES USING SAS INFORMATION MAPS

Another major benefit of using a star schema to store statistical data is the fast retrieval time. When querying a star schema directly using a SAS information map, tables are selectively joined as respective data items are selected. If a piece of data is not needed, the table is not joined and the amount of data to process is less, thus increasing the overall speed. This is demonstrated in the query below and in Figure 6.

```
select
  dim.Region,
  fact.SalesAmt
from
  fact_candy_sales fact
  inner join dim_candy_customers dim
    on (fact.CustID = dim.CustID)
```

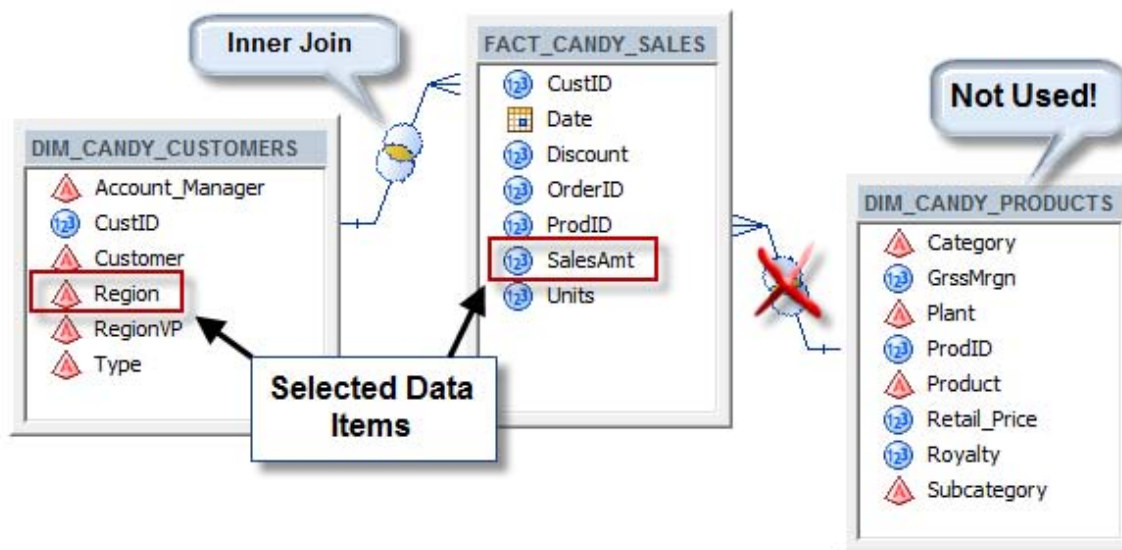


Figure 6. SAS Information Map Querying a Star Schema Efficiently

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

Another major benefit of a star schema design in an information map is the improved performance of dynamic filter prompts. Since there is less data to process for distinct values, the prompt is generated much faster at report run time, improving the user experience. For large datasets this can dramatically improve the user experience and provide a powerful way to report big data.

DEFINING THE INFORMATION PATH FOR THE GUIDED ANALYSIS

Now that the data has been introduced, let's define an information path to make some sense out of the data in an efficient manner. Figure 7 shows a flattened view of the candy sales data provided in a sample Web Report.

Date ▲	Region	Region VP	Account Manager	Customer	Type	Category	Subcategory	Product
01/01/2007	East	Chuck Norris	Data	Toys 4 U	Retailer	Candy	Soft	Nougatty Swirls
01/01/2007	East	Chuck Norris	Tasha Yar	Wholesalers R Us	Wholesale	Nuts	Sweet	CinnaPecans
01/01/2007	East	Chuck Norris	Beverly Crusher	Harry Koger	Grocery	Nuts	Fruit	Cherry Delight All Fruit
01/01/2007	Central	Smedley Snodgrass	Worf	Super Low Wholesaler	Wholesale	Nuts	Fruit	Cherry Delight All Fruit
01/01/2007	Central	Smedley Snodgrass	Geordi La Forge	Bulls Eye Emporium	Retailer	Nuts	Sweet	Carob N Almonds
01/01/2007	East	Chuck Norris	Data	Toys 4 U	Retailer	Candy	Gum	Sparkle Peppermint Gum
01/01/2007	West	Michael Jordan	Deanna Troi	Nile Online	Retailer	Nuts	Fruit	Cherry Delight All Fruit
01/01/2007	Central	Smedley Snodgrass	William Riker	Floor Mart	Retailer	Candy	Soft	Strawberry Chewers
01/01/2007	Central	Smedley Snodgrass	Worf	Super Low Wholesaler	Wholesale	Candy	Soft	Strawberry Chewers
01/01/2007	East	Chuck Norris	Tasha Yar	Wholesalers R Us	Wholesale	Nuts	Fruit	Cherry Delight All Fruit
01/01/2007	Central	Smedley Snodgrass	Jean-Luc Picard	Land of Fun	Retailer	Candy	Chocolate	Fruity Choco-Rolls
01/02/2007	Central	Smedley Snodgrass	Worf	Super Low Wholesaler	Wholesale	Candy	Soft	Strawberry Chewers
01/01/2007	East	Chuck Norris	Tasha Yar	Wholesalers R Us	Wholesale	Candy	Sweet	Cherry Delight All Fruit

Figure 7. Flat View of Candy Sales Data

Rather than report the flattened candy sales data as-is, this paper demonstrates a guided approach to analyze the data and drill down specific paths to arrive at answers faster. The following diagram shown in Figure 8 defines an example drill path for the candy sales data. The top level is the most generic view of the raw data which only provides a single number. Subsequent screens are linked using the out-of-the-box functionality of SAS BI Dashboard and SAS Web Report Studio. Each screen is intended to either drill to a lower level of detail or summarize the data further. Higher levels are best surfaced through the OLAP summarization of the candy sales data. As lower levels are drilled, the relational information map is used to query the star schema.

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

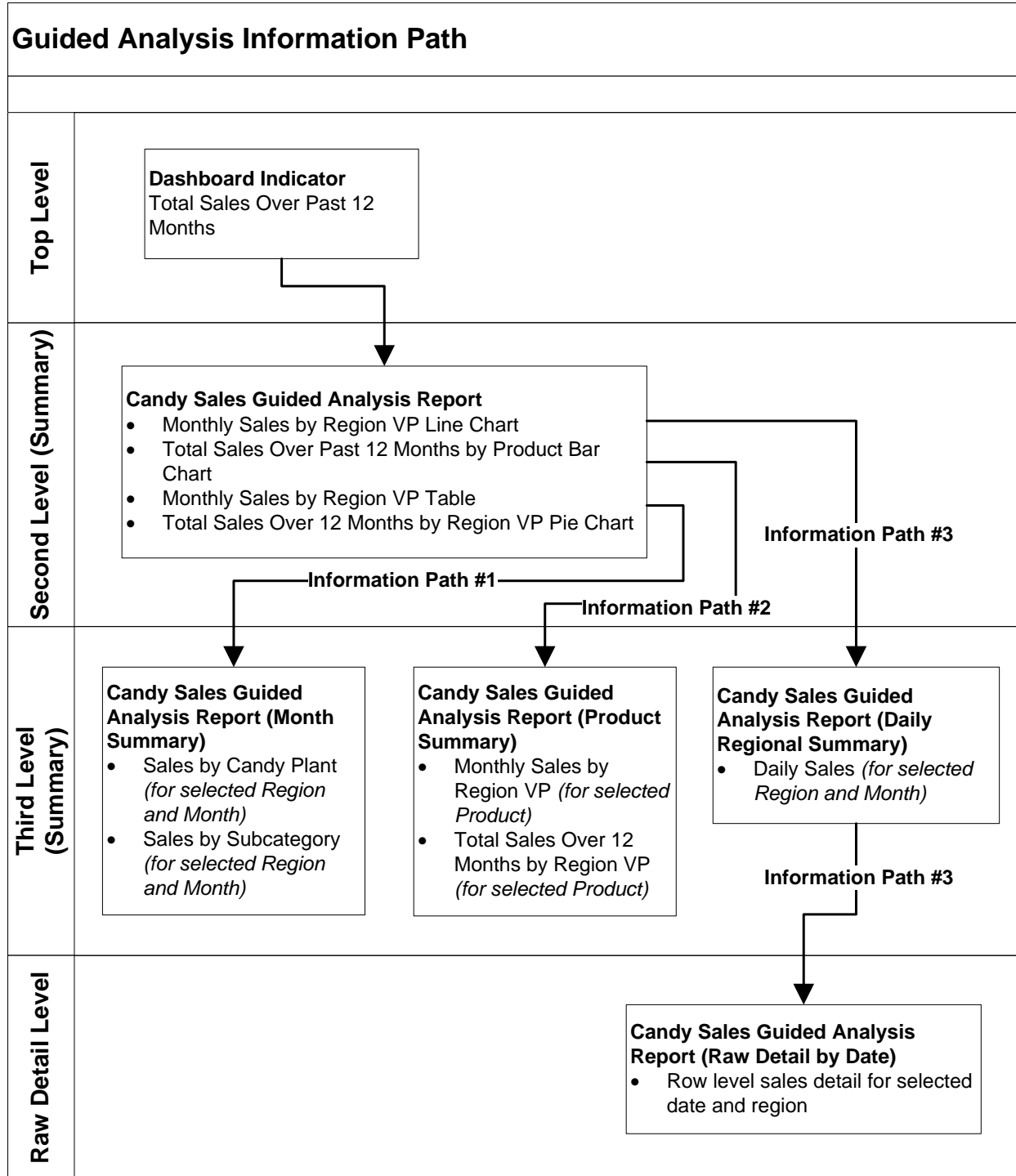


Figure 8. Guided Analysis Information Paths

Each layer of the information path reports the candy sales data a certain way. Each layer presents critical information at a certain aggregate level then provides a link or multiple links to drill down to the next level. Typical paths to drill down are time, location, product hierarchies, and organization structures.

EXAMPLE GUIDED ANALYSIS USING DASBOARDS AND WEB REPORTS

Suppose a Sales Manager wants to analyze the past 12 months of sales for the candy company. A dashboard would be an excellent way to provide quick access to key metrics. Subsequent web reports can be used to provide

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

information about which products are selling the most, what regions are performing the best, and how sales change over time.

TOP LEVEL BI DASHBOARD

The top level dashboard is shown in Figure 9. The first indicator is used to demonstrate the guided analysis. The data generated in this indicator is from the Candy Sales OLAP cube defined in the beginning of this paper. The past 12 months are defined using a custom member, which aggregates the last 12 months into a single member shown in the Sales Past 12 Months indicator.



Figure 9. Top Level BI Dashboard

DRILLING INTO THE DATA USING SAS WEB REPORT STUDIO

The Rolling 12 Months indicator is linked to a report in SAS Web Report Studio which begins the guided analysis. This is defined in the second level of the diagram in Figure 8. The second level of the guided analysis brings the user into SAS Web Report Studio as shown in Figure 10. The data for this report is surfaced through the same OLAP cube that the dashboard indicator uses. Any of the highlighted areas can be clicked and will link the user into the next level of analysis. Report Linking is used to send users to sections within the same report. More information on report linking within the same report can be found in the Reference section below.

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

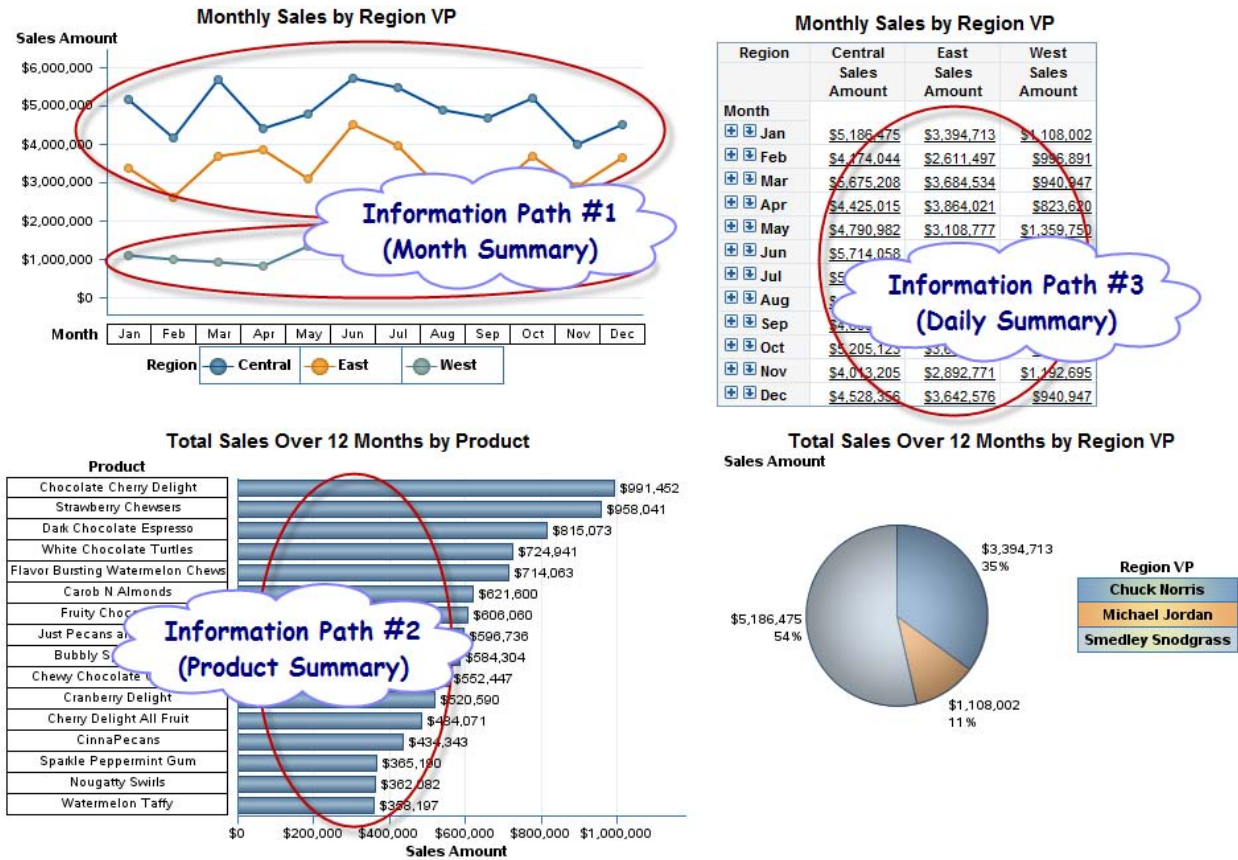


Figure 10. Second Level Web Report Studio Report

The three information paths highlighted above in Figure 10 will breakdown candy sales over the past 12 months into a few different areas.

1. Month Summary
2. Product Summary
3. Daily Summary → Raw Sales Detail

Information Path #1 (Month Summary)

The first information path summarizes the month and region. Clicking any of the points on the three lines will pass the respective month and region to the next report. This is setup using the reporting linking feature of SAS Web Report Studio. The Sales Manager viewing this report can now determine which candy plants are producing more sales and what subcategory of products are sold for the given month and region. The table at the top of the report confirms the correct member is drilled upon and provides contextual detail for the report. This is shown in Figure 11.

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

Month	Region	Sales Amount	Units
Mar	Central	\$5,675,208	730,400

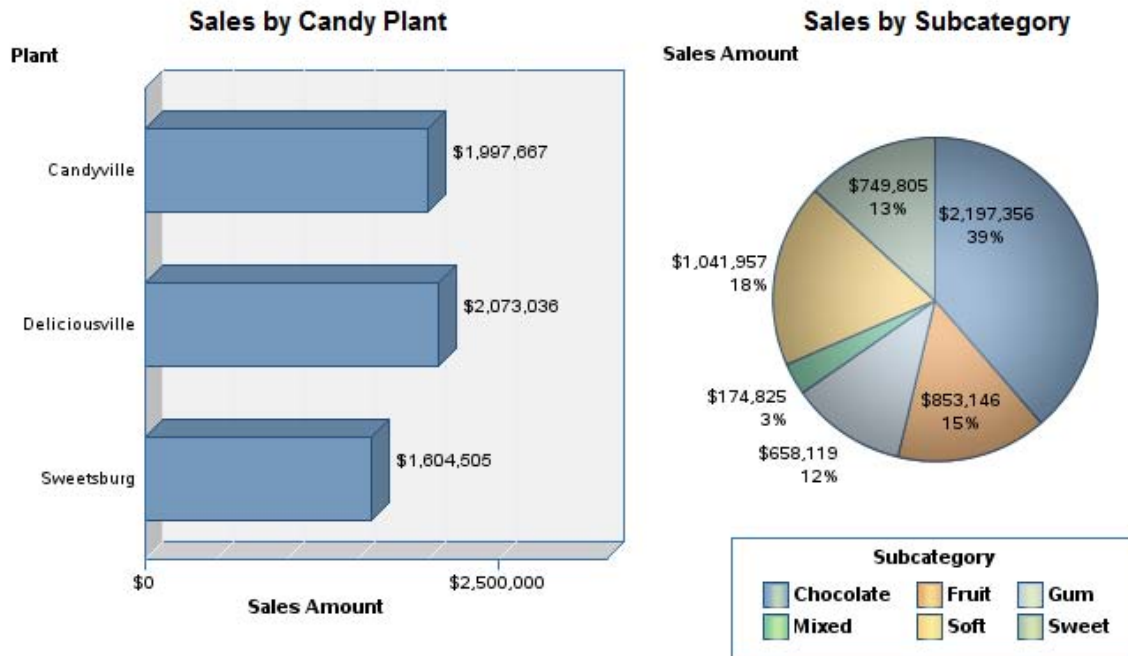


Figure 11. Information Path #1 Summarizing Month and Regional Sales Data

Information Path #2 (Product Summary)

The second information path summarizes product sales over the past 12 months. This section of the report only focuses on a single product type that was selected from the previous screen. The table at the top of the report confirms the correct member is drilled upon and provides contextual detail for the report. This is shown in Figure 12.

Category	Subcategory	Product	Sales Amount	Units
Candy	Chocolate	Chocolate Cherry Delight	\$6,816,621	877,300

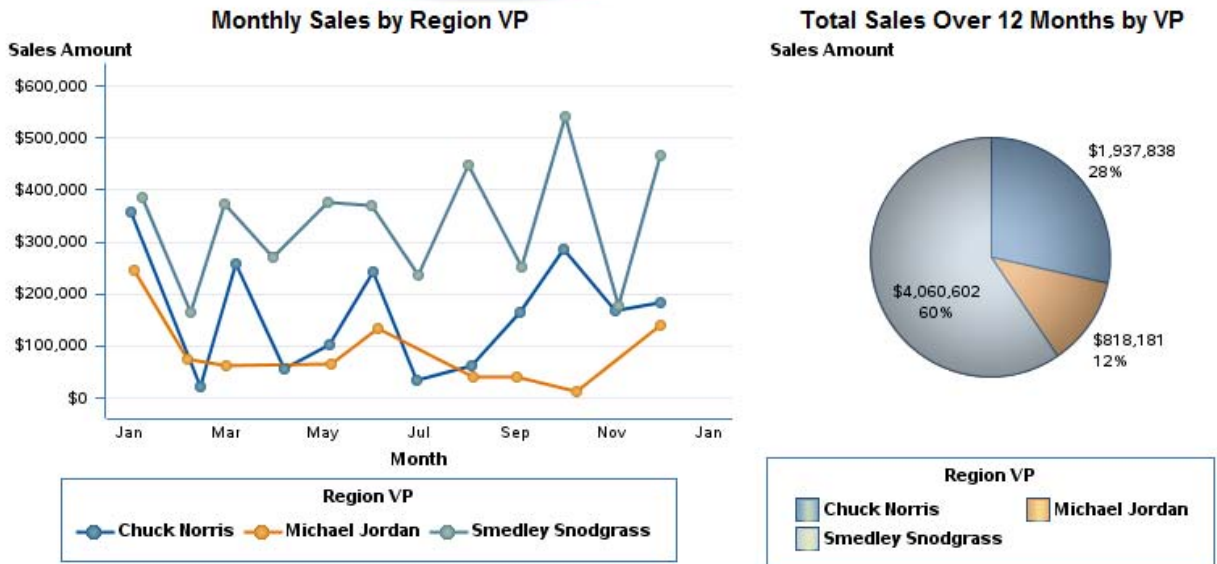


Figure 12. Information Path #2 Summarizing Product Sales Over the Past 12 Months

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

Information Path #3 (Daily Summary)

The third information path provides a day to day view of sales for the month which was drilled. From this screen, the Sales Manager can drill through the specific day of the month to get more insight. Figure 13 highlights the clickable region of the progressive bar chart which links to the final level of this information path.

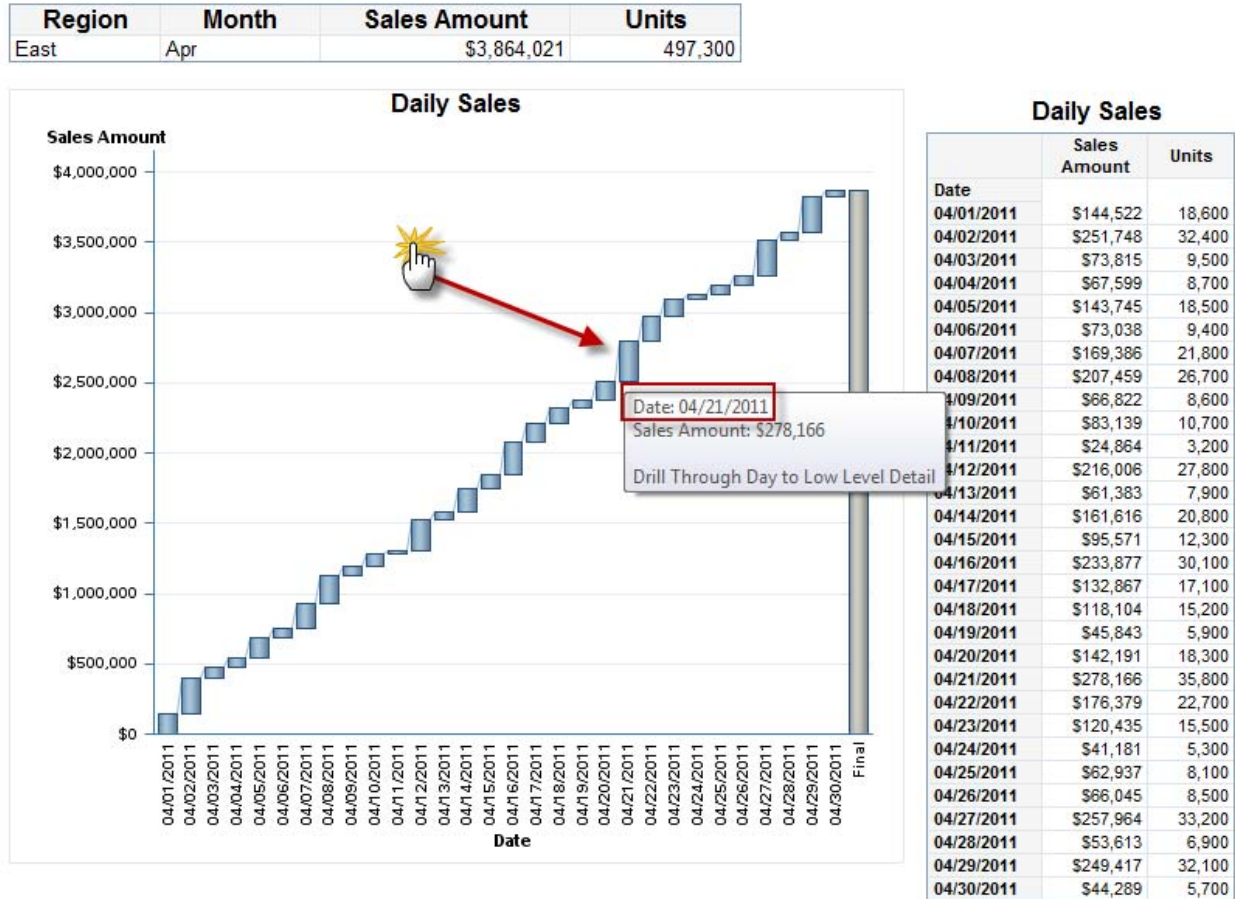


Figure 13. Information Path #3 Summarizing Daily Sales by Region

The final level of the third information path provides the low level detail which was clicked from the previous screen. This is shown in Figure 14. Notice the only date available to the report is the date clicked on the previous screen (April 21, 2011).

Date	Region	Sales Amount	Units
04/21/2011	East	\$278,166	35,800

Date	Region	Region VP	Account Manager	Customer	Type	Category	Subcategory	Product
04/21/2011	East	Chuck Norris	Beverly Crusher	Harry Koger	Grocery	Candy	Chocolate	Fruity Choco-Rolls
04/21/2011	East	Chuck Norris	Beverly Crusher	Harry Koger	Grocery	Candy	Gum	Bubbly Sparkle Gum
04/21/2011	East	Chuck Norris	Beverly Crusher	Harry Koger	Grocery	Candy	Soft	Flavor Bursting Water
04/21/2011	East	Chuck Norris	Beverly Crusher	Harry Koger	Grocery	Candy	Soft	Nougatty Swirls
04/21/2011	East	Chuck Norris	Beverly Crusher	Harry Koger	Grocery	Nuts	Fruit	Cranberry Delight
04/21/2011	East	Chuck Norris	Data	Toys 4 U	Retailer	Candy	Gum	Bubbly Sparkle Gum
04/21/2011	East	Chuck Norris	Tasha Yar	Wholesalers R Us	Wholesale	Candy	Soft	Watermelon Taffy

Figure 14. Final Layer of Guided Analysis

The final layer shown in Figure 14 is the end of the guided analysis demonstrated in this paper. For this sample sales data mart, the number of rows left to analyze is much more reasonable. The end user has 7 rows of data that

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

represent the lowest level of data available. Additional pie charts or bar charts could be added to summarize the 7 rows.

QUICK TIP FOR ORGANIZING MULTIPLE SECTIONS IN WEB REPORTS

All activity after the initial dashboard shown above in Figure 9 is contained within a single web report to minimize the number of reports in metadata. Each drill through level and drill path is organized into sections. The section naming scheme shown in Figure 15 organizes the sections further. The first number represents the level of drill through while the second number represents the drill path. The third drill path is the longest in this example. Since the first level is the dashboard indicator, the web report sections start with the second level.

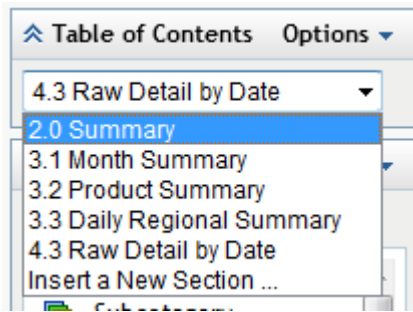


Figure 15. Naming Scheme for Web Report Studio Sections

CONCLUSION

Providing a guided approach to analyzing data can be more effective if the analysis paths can be defined. If the path to analyze data cannot be determined, simply observing how end users analyze the raw detail data can provide the logical steps required to determine the results desired. In summary, this paper describes a methodology automating the steps to analyze data using SAS. This reduces the time to come to data-driven conclusions and standardizes the method for which this is accomplished. The goal of the methodology described in this paper is to make better business decisions faster and with a greater level of understanding of contributing factors.

REFERENCES

- SAS code which creates the candy sales star schema referenced in this paper is available at <http://www.stephenoverton.net/SASCode/SGF2012/>
- SAS Support. "Sample 41526: Using SAS® Web Report Studio to pass prompt values between sections in the same report." March 4, 2012. Available at <http://support.sas.com/kb/41/526.html>.
- SAS Support. "Sample 40009: Using report linking to display detail data in an OLAP-based SAS® Web Report Studio report." March 4, 2012. Available at <http://support.sas.com/kb/40/009.html>.
- Wikipedia. "Star Schema." March 4, 2012. Available at http://en.wikipedia.org/wiki/Star_schema.
- Raithel, Michael A. May 2004. "Creating and Exploiting SAS® Indexes." SUGI 29 Proceedings. Rockville, MD. Available at <http://www2.sas.com/proceedings/sugi29/123-29.pdf>.
- Wikipedia. "Database Indexing." March 4, 2012. Available at http://en.wikipedia.org/wiki/Database_index.

RECOMMENDED READING

- Aanderud, T and Hall, A. 2012. *Building Business Intelligence with SAS®: Content Development Examples*. Cary, NC: SAS Press <http://www.bi-notes.com/sas-bi-book-all-about/>

Lost in Wonderland? Methodology for a Guided Drill-Through Analysis Out of the Rabbit Hole, continued

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Stephen Overton

Zencos Consulting (<http://www.zencos.com>)

Work Email: soverton@zencos.com

Personal Email: stephen.overton@gmail.com

Personal Website: <http://www.stephenoverton.net>

LinkedIn: <http://www.linkedin.com/in/overton>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.