

## CHAPTER 9: EXERCISE 1

# ADVANCED DATA VISUALIZATION USING SAP LUMIRA DESKTOP

NITIN KALÉ, UNIVERSITY OF SOUTHERN CALIFORNIA  
NANCY JONES, SAN DIEGO STATE UNIVERSITY

Updated by Hossam Ali-Hassan for the SAP UA Analytics Summer Workshop June 2016 @ UWM.

### OBJECTIVE

Use advanced data visualization to discover trends in data sets.

### ACTIVITIES

- Import and prepare data
- Join data from several files into one dataset
- Perform data cleansing
- Perform data harmonization
- Create data visualizations

### SOFTWARE PREREQUISITES

- Microsoft Excel 2010 or newer
- SAP Predictive Analytics 2.4 (PA)

## UCC PRODUCTS REQUIRED

- None

## DATA REQUIRED

- GBI data files are available in the folder *GBI\_Data\_Files\_E9\_1*

## SCENARIO

We have noted some interesting and potentially informative trends in the 2007 – 2011 sales data at *Global Bike Inc.* We would like to explore these trends using data visualizations. Data trends can be used to make informed business decisions and plan our strategy for marketing, product development, customer service etc.

The multidimensional data for Global Bike Inc. can be modeled as a star schema as shown in Figure 1.

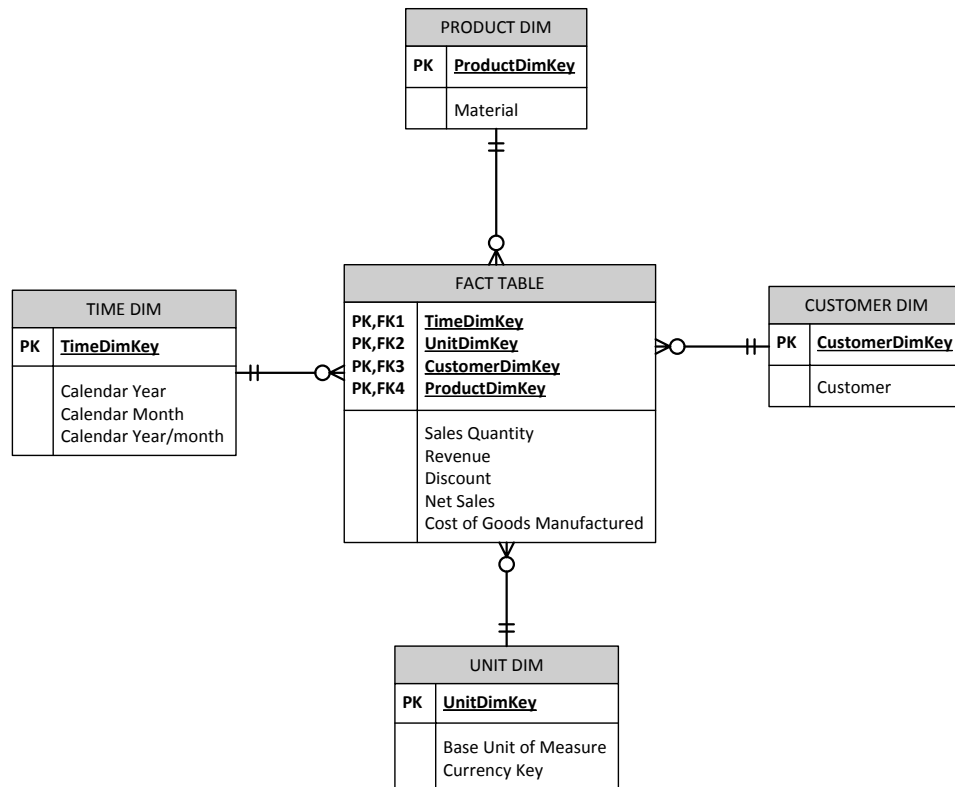


Figure 1: GBI Star Schema

The fact table consists of five key figures – *Sales Quantity*, *Revenue*, *Discount*, *Net Sales* and *Cost of Goods Manufactured*. These facts can be analyzed against four dimensions – *Time*, *Product*, *Customer* and *Unit*. These dimensions have attributes as shown above.

## TECHNIQUES FOR ADVANCED DATA VISUALIZATION

The human visual system has evolved to be particularly good at recognizing patterns. Data visualization has become a standard analytical tool which capitalizes on the ability of humans to recognize patterns within massive quantities of multi-dimensional data generated by business information systems. Many scientific studies have led to the creation of visualization models that utilize human perception and cognition.

When the number of dimensions is small, we can use standard graphing techniques for visualization e.g. *bar charts*, *line charts*, *histograms*, *pie charts* and *scatter plots*.

When the number of dimensions is large, there are several novel techniques for visualizing such data. They are categorized into the following major areas <sup>1</sup> – For more information on these visualization techniques, please refer to the journal reference in the footnote.

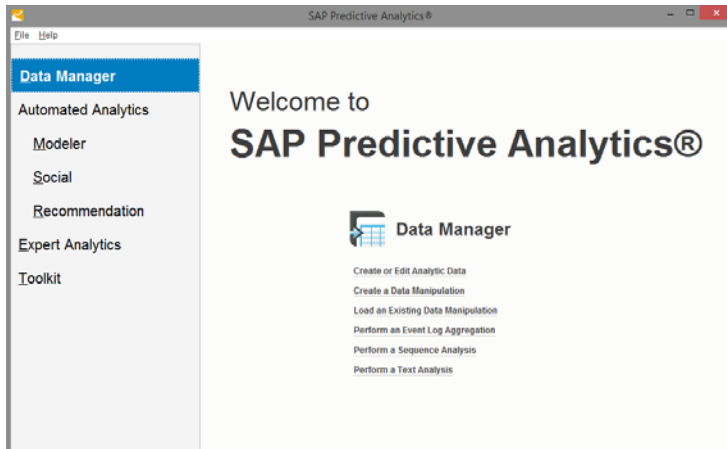
- A. Pixel-oriented Techniques
  - a. Space filling curves
  - b. Recursive pattern
  - c. Snake-Spiral
  - d. Circle segments
- B. Geometric Projection Techniques
  - a. Parallel coordinates
  - b. Scatter plot matrix
  - c. Hyperbox
  - d. Trellis display
  - e. Self-organizing maps
- C. Icon-based Techniques
  - a. Star glyphs
  - b. Color icons
  - c. Stick figures
  - d. Chernoff faces
- D. Hierarchical and Graph-based Techniques
  - a. Dimensional stacking
  - b. Cone trees
  - c. Mosaic plots
  - d. Fractal foam

## ADVANCED DATA VISUALIZATION FOR GLOBAL BIKE INC.

1. Launch SAP Predictive Analytics using *Start → All Programs → SAP Business Intelligence → SAP Predictive Analytics* or by clicking on the shortcut icon on your desktop. Choose Expert Analytics on the main menu. Click on Expert Analytics again when you get to the Welcome screen.

---

<sup>1</sup> Keim D. A., Kriegel H.-P. *Visualization Techniques for Mining Large Databases: A Comparison*, *Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, Dec. 1996, pp. 923-938.



## 2. Acquiring GBI data

- a. Click *File* → *New*
- b. In the *New Dataset* window, choose *Microsoft Excel*. Next
- c. Browse for the *Sales Transactions* file (in the GBI Data Files folder provided by your instructor)
- d. You should see a preview of the data being acquired. This is where you can select/deselect fields for importing. Import all fields.
- e. Click *Create*

## 3. Interface

- a. Click on the *Prepare* view
- b. See Figure 2 for the SAP Predictive Analysis interface. **Note: The interface changes occasionally with revisions of the software.** Explore the various panels and tools.

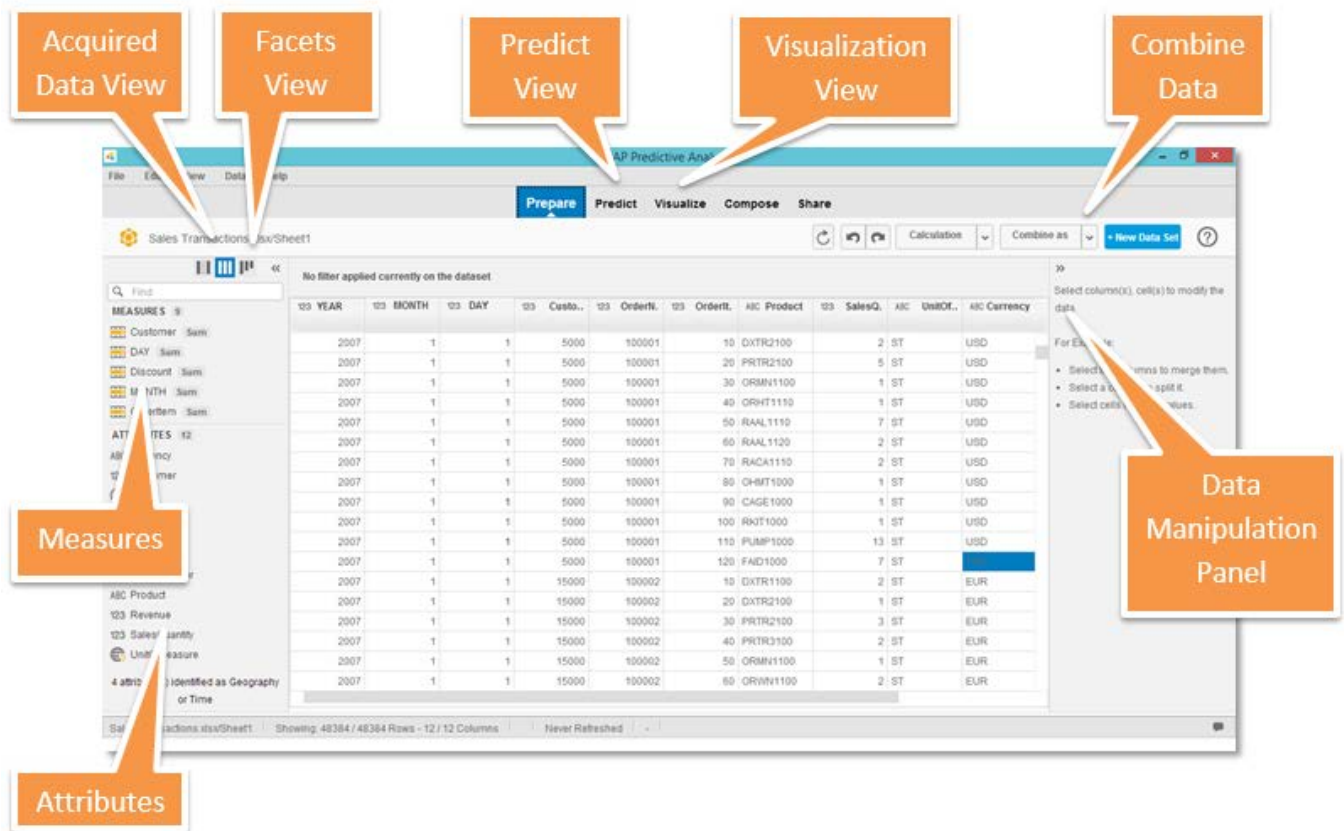
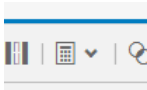


Figure 2

4. Look at the status bar at the bottom of the window, to see that there are 48,384 sales entries.
5. So far we have acquired *Transactional data* for GBI. We would like to acquire *master data* as well. Master data will help describe fields such as *Product* in more detail. For example, in addition to Product number, we might like to know its color, price, description etc. These attributes are usually available in the material master data in SAP ERP.
  - a. Select *Data – Combine – Merge* menu option
  - b. Click *Add New Dataset*
  - c. Choose MS Excel
  - d. Browse for the file *Material.xlsx* file
  - e. *Create*
  - f. You will now see the master data; e.g., *Product Category, Components, color* etc... for 28 materials.
6. Merging data

- a. In the Current Dataset list, choose (highlight) *Product*
  - b. In the Lookup Dataset list, choose (highlight) *Material*
  - c. Click *Merge*
  - d. The two data sets are merged (*joined* on the common attribute)
  - e. Now Add the *Customer* data and Merge following the steps above. Be sure to start from the Sales Transaction sheet before doing the merge, (using *customer* field in both the datasets). *If you have trouble merging the data, this is probably due to a data type mismatch. Edit the Excel file for Customer. Change the data type for which there is mismatch. Reacquire it. Then try to merge.*
  - f. Now Add the *Product Category* data and Merge (*product category* in both). *Note: There is only 61% match because some products in the master dataset do not have a category assigned to them. Ignore this and merge.*
  - g. Now Add the *Sales Organization* data and Merge (*Sales organization* in both)
  - h. Now Add the *Month* data and Merge
  - i. You notice that there is more than one column called *Language Key* in your new spreadsheet. Delete (click on cog wheel and select Remove) all but one of them so that you are left with only one Language Key column.
7. We have acquired and merged sales transaction data with master data.
8. Save as a file name of your choice.
9. We are now ready to manipulate and visualize this data
10. Since the values in *Customer*, *OrderNumber*, *OrderItem* etc. are numeric, PA has incorrectly identified them as **measures** (in the left panel).
- a. Click on the cog wheel for each errant measure. Remove.
  - b. Delete all measures except – *Discount*, *Revenue* , *Sales Price*, *SalesQuantity*, and *Transfer Price*
11. We will now add *Net Sales* as a new calculated measure
- a. Click on the Grid icon in the toolbar
  - b. Click on *Calculation*  → *New Calculated Dimension*.

- c. Attribute name: *Net Sales*.
  - d. In the Formula area enter the formula  $\{Revenue\} - \{Discount\}$ . You can do this by double clicking on each attribute and entering the operand in between them rather than typing them out.
  - e. Click *OK*
  - f. A new *attribute* (and a column) is created.
  - g. Click on the cog wheel for *Net Sales* and *Create a Measure*.
  - h. You should now see *Net Sales* as a measure (rename it to *Net Sales*, if different). *If you see it only as a Dimension, create it a Measure from it.*
12. In the *dimensions* list on the left, you see that several dimension have a globe icon with a question mark. This implies that there is a geographic hierarchy possible for that attribute. We will use these hierarchies and we will also do some data cleansing.
- a. Click on the cog wheel for *Country*. Select *Create a Geographic Hierarchy by Names*
  - b. In the *Geographic Data* window, choose *Country* as *Country*, *City* as *Location*.
  - c. Click *Confirm*
  - d. You see *US/Palo Alto was Not Found*. Click *Done*
  - e. The data set now contains four new Geographic Hierarchy aware columns – *SubRegion*, *Country*, *City*, *Region*
  - f. Delete the *SubRegion* column
  - g. Click on *City* column in the Grid, then *unresolved*. In the Replace tab (in the Manipulation Panel, to the right of the grid), Replace: *Palo Alto*. Click *Apply*.
  - h. Click on *Region* column, then *unresolved*. In the Replace tab, Replace: *California*. Apply.
  - i. Click on *Country (Geography)* column, then *unresolved*. In the Replace tab, Replace: *United States*. Apply.
13. The data is now cleansed and ready for Visualization, click on Visualize
- a. Several charting options are available for visualization – *bars, lines, pies, geographic, scatter/bubble, maps, radar, tag cloud etc.*



- b. Using the appropriate charting technique, answer the following questions. Hints are provided for each question.

Hint: Use a column chart. From *Measures*, drag *Revenue* into Y Axis, from *Dimensions*, drag *Year* into X Axis. In Y axis, use the cog wheel for Revenue to *Sort* it in descending.

**Question 1: What Year had the highest revenue? What was the revenue?**

Hint: Click Create new Visualization (click + at the bottom). Use a column chart. Y Axis – *Revenue*, X Axis – *Product*, *Product Description*. Sort Revenue in descending.

**Question 2: What material (with name) had the highest revenue? What was the revenue?**

Hint: Use a *line* chart. Y Axis – *Revenue*, X Axis – *Year*, Legend Color - *Country* (use *Country* from the attribute list)

**Question 3: Are the historical (year by year) revenue trends for the US and DE (Germany) similar or dissimilar?**

Hint: Use a *Column* chart. Y Axis – *Revenue*, X Axis – *Customer*, *Customer Description*, Trellis: Rows - *Year*

**Question 4: Did GBI ever gain or lose a customer? Explain**

Hint: Use a *heat* map. Area Color – *Revenue*, Area Name – *Month and Month Name*, Area Name 2 – *Year*

**Question 5: Is there seasonality in revenue during the year? If so, what month has the highest revenue? Is the seasonality similar from year to year?**

Hint: Use a *Column* chart. Y Axis –*Revenue*, X Axis – *Customer and Product*. Filter the Year (in Dimensions to 2011). Now add descriptions for the *customer* and *product* to the X-Axis.

**Question 6:** In 2011, for what Material did the highest Revenue from a single customer occur?

Hint: Use a *heat* map. Measure (Area Color) – *Revenue*, Attributes: Area Name – *Year*, Area Name 2 – *Product*.

**Question 7:** Are there any products that show dramatic change in revenue over time (years)? Does it have similar change by country?

Hint: Use a *line* chart. Y Axis – *Revenue*, X Axis – *Month*, Legend Color – *Product and Product Description*. In the resulting line chart, select the tallest peak by selecting its color in the Product legend, then *exclude*. Be sure you are not deleting any months (check the filters above the chart). Repeat to eliminate more peaks. As you zero into the products that have low seasonality, the one with the lowest seasonality (relatively flat line) will appear.

**Question 8:** Is there any material that does not display significant seasonality?

Hint: Use a *Column* Chart. Measures: Y-axis: *Revenue*. Dimensions: X-axis: *Customer*. For the Measure *Revenue* add a calculation (use the cog wheel) to your Revenue measure – *Percentage*. Delete the *Revenue* measure, leaving the Percentage Revenue in place.

**Question 9:** What customer has the highest percentage contribution to revenues? What has been the trend of that customer's percentage contribution over the years?

Hint: Use a *Geo Choropleth* chart to display revenue. Measures – *Revenue*, Geography –*Region*.

**Question 10:** Which region has the highest revenue?

14. Save and close.

