**Carnegie Mellon University, Heinz College**

**Business Intelligence & Data Mining with SAS Suite**

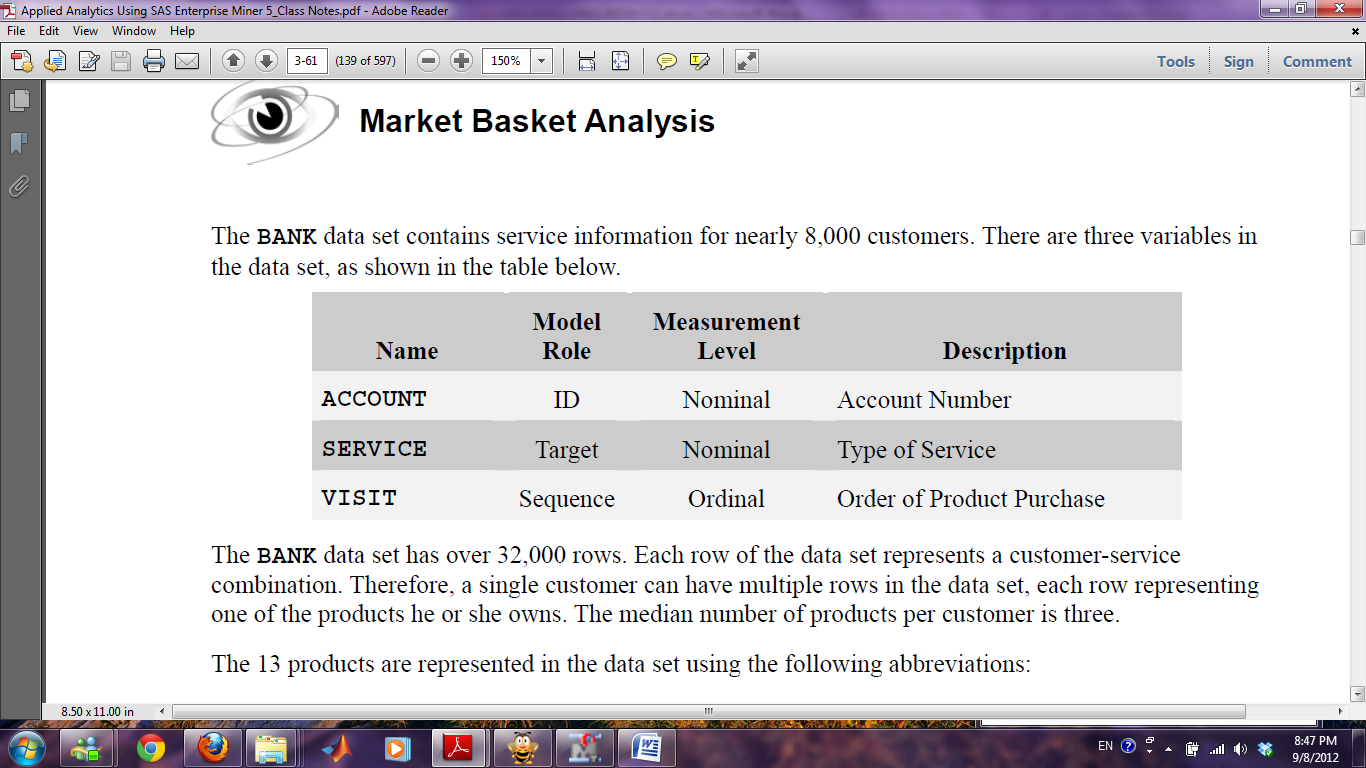
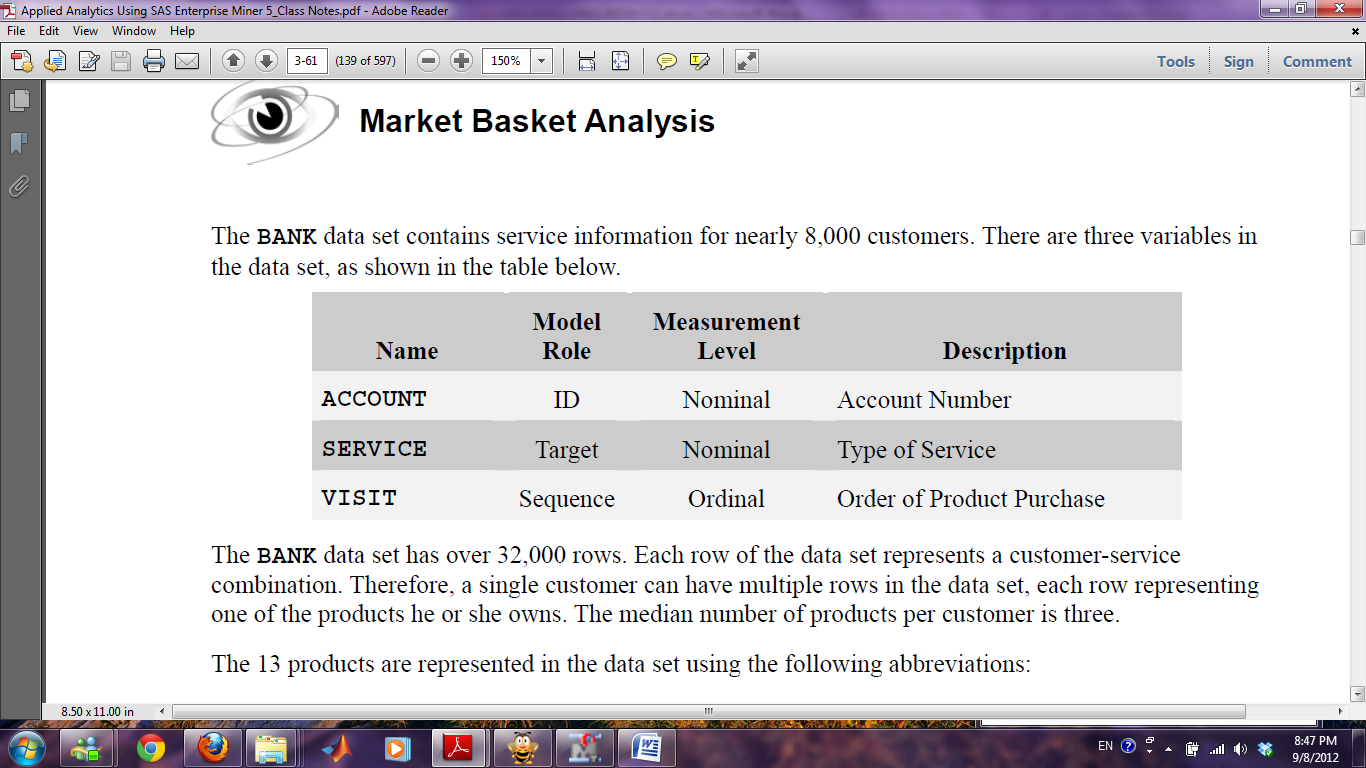
**Fall 2012 (94832 Mini 1)**

**Assignment 1 - Market Basket Analysis**

**(Due 9/18/2012)**

**Image you were the manager of the PNC bank (Pittsburgh branch). You would like to find out what financial services are more likely to be used together by your customers, so that you and your team can better design your financial services in the coming year. After consulting with a senior BI specialist, you decide to conduct a market basket analysis.**

You told the IT department to prepare a detailed customer account-level data set on individual usage of financial services. The IT department has just finished the data preparation -- a BANKdata set that contains service information for nearly 8,000 customers. There are three variables in the data set, as shown in the table below.



The BANKdata set has over 32,000 rows. Each row of the data set represents a customer- service combination. Therefore, a single customer can have multiple rows in the data set, each row representing one of the products he or she owns. The median number of products per customer is three.

The 13 products are represented in the data set using the following abbreviations:

ATM automated teller machine debit card

AUTO automobile installment loan

CCRD credit card

CD certificate of deposit

CKCRD check/debit card

CKING checking account

HMEQLC home equity line of credit

IRA individual retirement account

MMDA money market deposit account

MTG mortgage

PLOAN personal/consumer installment loan

SVG saving account

TRUST personal trust account

**Questions:**

**(Note: It is recommended that you provide screen shots of your analysis along with the discussion.)**

**a.** Open a new diagram in your Exercises project. Name the diagram **BANK**. Create a new data source for the data set **AAEM. BANK (SAS table--> AAEM-->BANK)**. Use the Explore function to plot the variable distribution. Provide your variable plots. What did you learn from the variable exploration?

**b.** Notice that at this point you are not interested in variable **VISIT** during the analysis. You are interested in the association rules in variable **SERVICE** for each **ACCOUNT.** Explain what model role should you assign to each variable and why.

**(**Add the node for the **BANK** data set and an Association node to the diagram. Change the setting for Export Rule by ID to Yes and leave the remaining default settings for the Association node and run the analysis. )

**c.** Examine the results of your association analysis. What is the highest lift value for the resulting rules? Which 10 rules have the highest lift values? How do you interpret your results?

**d.** Suppose you are particularly interested in the other financial services that customers use when they have automobile loans. Use the link graph to examine what are the association rules that involve automobile loans. Based on the rules, what other products and services would you recommend to customers when they take automobile loans?

**e.** Based on your analysis, what would you plan to do in the coming year?