# SASEG – Variable Transformation

**For Regression analysis**

(Fall 2016)

**Sources** (adapted with permission) **-**

Ron Freeze Course and Classroom Notes

Enterprise Systems, Sam M. Walton College of Business, University of Arkansas, Fayetteville

Microsoft Enterprise Consortium

IBM Academic Initiative

SAS® Multivariate Statistics Course Notes & Workshop, 2010

SAS® Advanced Business Analytics Course Notes & Workshop, 2010

Microsoft® Notes

Teradata® University Network

*For educational uses only - adapted from sources with permission. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission from the author/presenter.*

Example:*Pelican Stores, a division of National Clothing, is a chain of women’s apparel stores operating throughout the country. The chain recently ran a promotion in which discount coupons were sent to customers of other National Clothing stores. Data collected for a sample of 100 in-store credit card transactions at Pelican Stores during one day while the promotion was running are contained in the file named PelicanStores. The Proprietary Card method of payment refers to charges made using a National Clothing charge card. Customers who made a purchase using a discount coupon are referred to as promotional customers and customers who made a purchase but did not use a discount coupon are referred to as regular customers. Because the promotional coupons were not sent to regular Pelican Stores customers, management considers the sales made to people presenting the promotional coupons as sales it would not otherwise make. Of course, Pelican also hopes that the promotional customers will continue to shop at its stores. Pelican would like to know the significance of the variables on Net Sales.*

SASEG provides several tools to analyze given variables in a problem. Prior to transforming variables, it is expected that descriptive statistics must be ran in order to understand the variables to be transformed. In the above case, **Type of Customer, Method of Payment, Gender** and **Marital Status** must be modified to be categorical variables in order to use these in a regression.

# C:\Program Files\PowerServ\CourseGraphics\demo_eye.jpgExercise – Variable Transformation

Transform the variables of **Type of Customer,** **Method of Payment, Gender** and **Marital Status** for the PELICANSTORES dataset. Remember that in order to accomplish this there must k-1 (k being the number of categories for the variables) dummy variables created. Remember that the categorical variables have the following categories for each variable.

**Type of Customer Method of Payment Gender Marital Status**

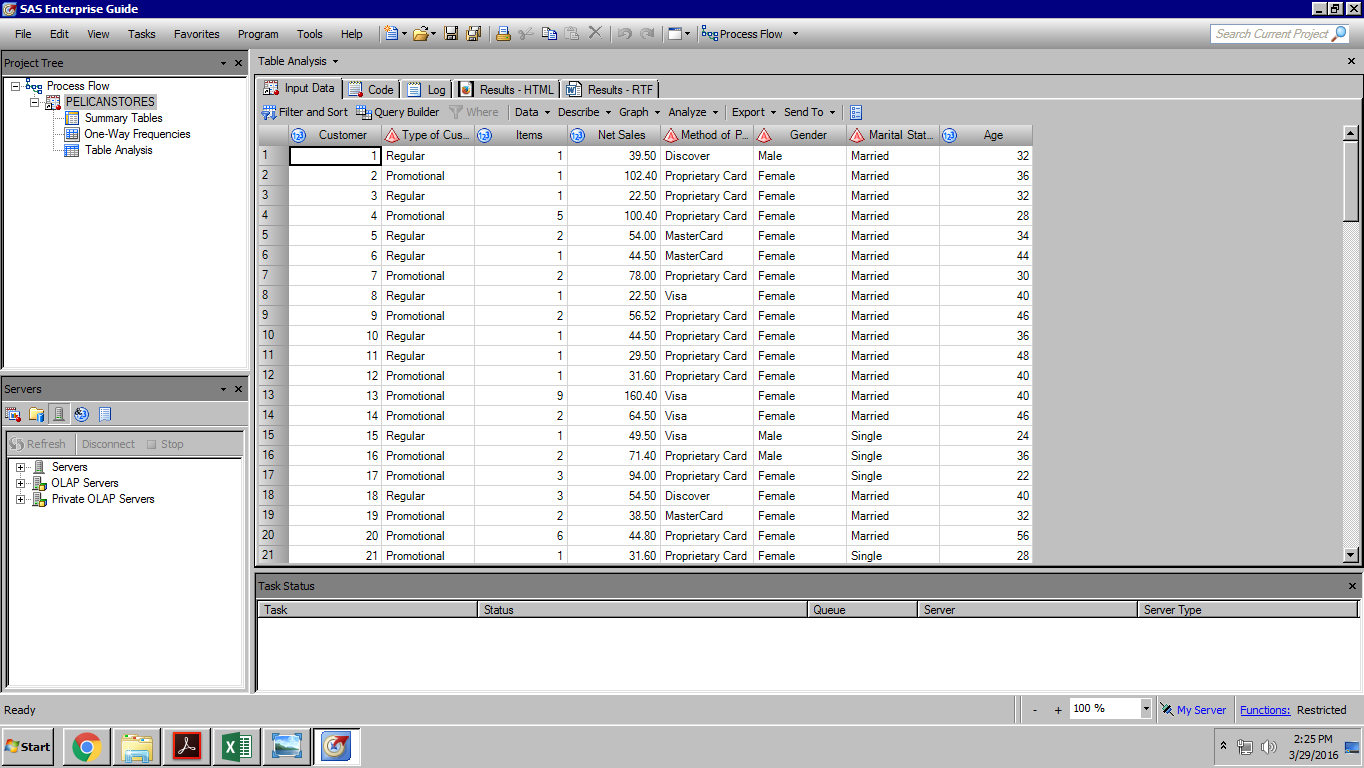
Regular Discover Male Married

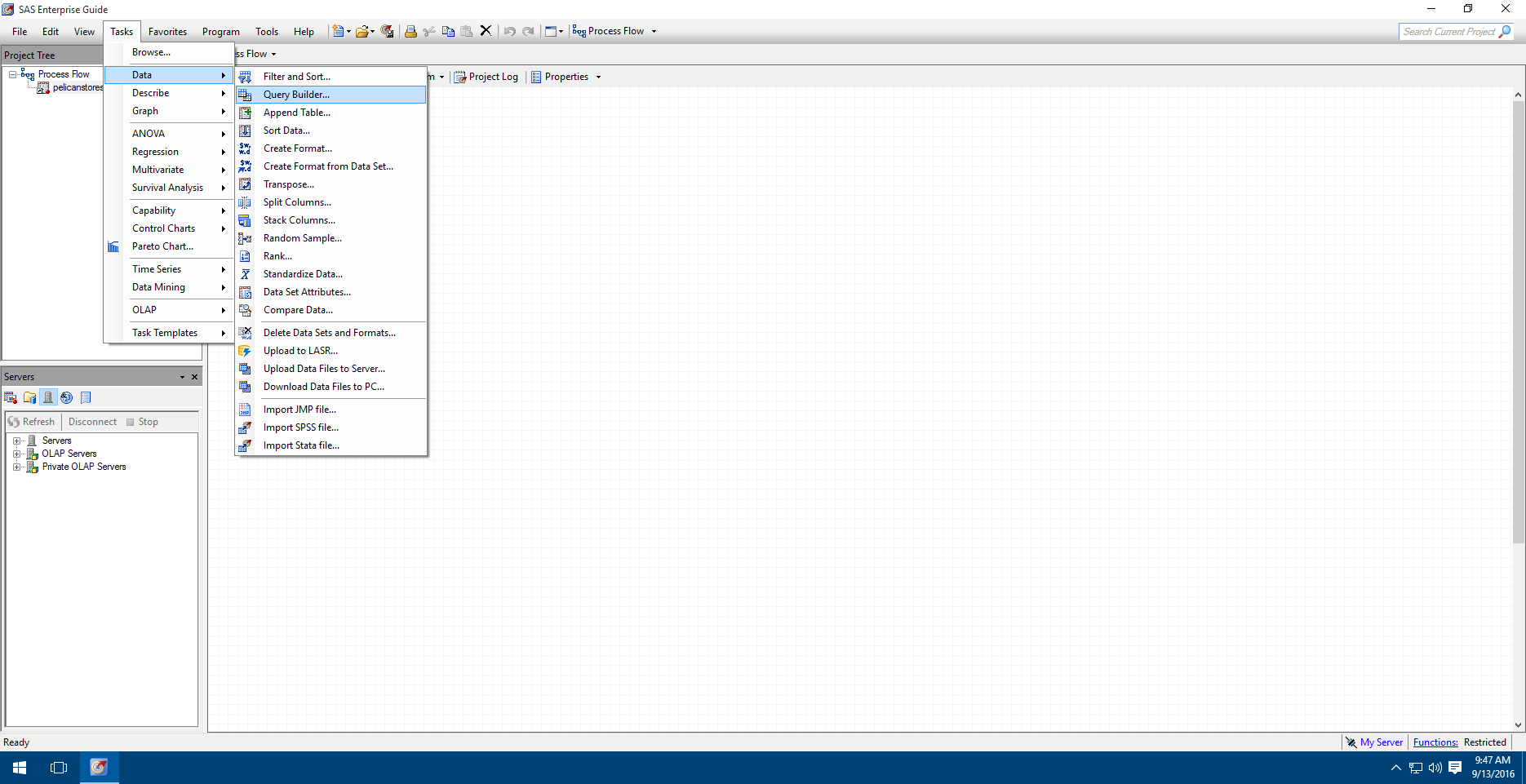
Promotional Proprietary Female Single

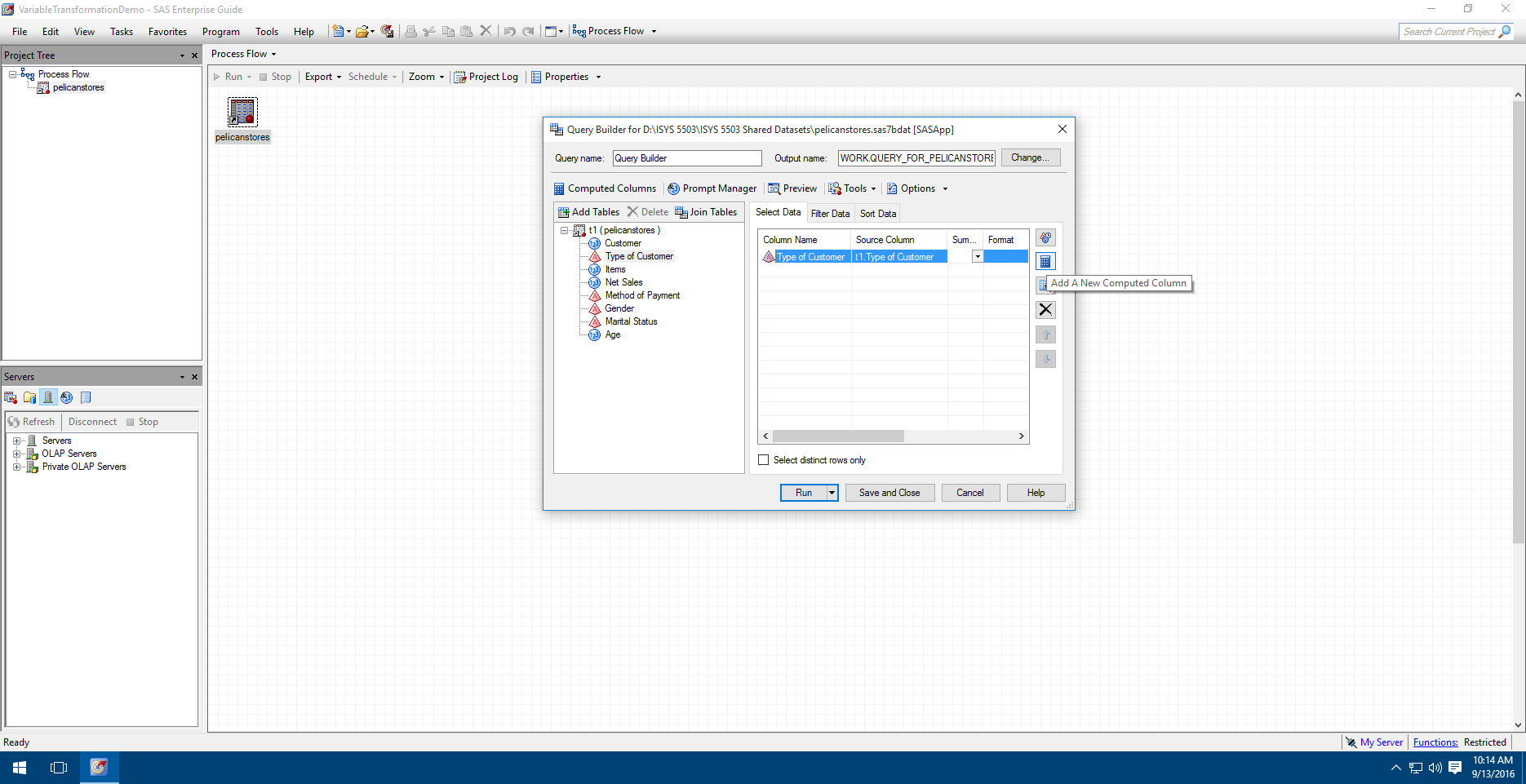
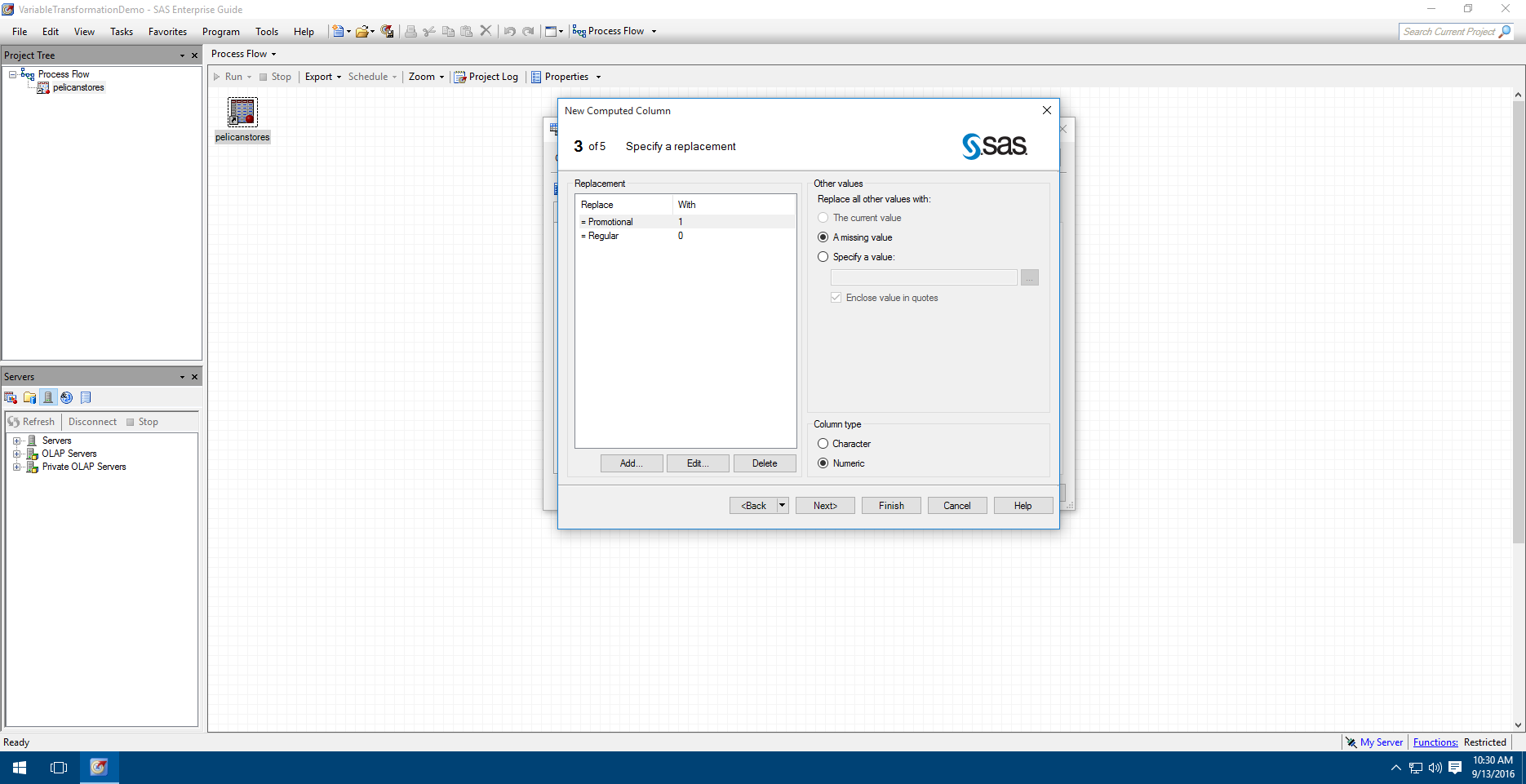
MasterCard

Visa

American Express

1. Open the **PelicanStores** SAS Dataset using the following path: **File > Open >Data--> Servers > SASApp-->Files > D: > ISYS 5503--> ISYS 5503 Shared Datasets--> Pelican Stores**



1. Select **Task** -> Data -> **Query Builder**
2. Make sure the **Select** **Data** tab is selected. (Note that you have the ability in the query to **filter** **Data** and **Sort** **Data**
3. Drag and drop **Type of Customer** to the “**Drop a column here to add it to the query**”.
4. From the Query Builder pop-up, click on the “**Add A New Computed Column**”.
5. The computed column will be where we insert a “0” for Regular customers and “1” for Promotional customers. The dummy variable will be named TypCust.
6. A series of 4 pop-ups for New Computed Column will appear. Make the following selections for each pop-up and select **Next** after the completion of each step.
   1. (1 of 5) Select **Recoded column**
   2. (2 of 5) Highlight **Type of Customer**
   3. (3 of 5) Select **Numeric** under Column type.
   4. (3 of 5) Select **Add**, select **Add**, type in “**Regular**” under the **Replace** **Values**, enter **0** under **With this value:**, select **OK**.
   5. (3 of 5) Make sure the box “enclose this value in quotes” is unchecked.
   6. (3 of 5) Repeat for “**Promotional**” using a **1**. Your result should look like the screen shot.
   7. (4 of 5) Enter TypCust as the Column Name:. Enter **Type of Customer Dummy Variable** as the Label:.
   8. (4 of 5) Change the Format to **BESTD11**.
   9. **Finish**
7. **Run** the query.

# C:\Program Files\PowerServ\CourseGraphics\demo_eye.jpgExercise – On Your Own Transformations

Using the same technique used above, transform the variables of **Method of Payment, Gender** and **Marital Status** into the required number of dummy variables.