ALVAREZ College of Business

DA 6223 Data Analytics Tools and Techniques

Week 4 Meeting Agenda

I. Week 3 Recap with Questions

a) We'll go over several multiple-choice questions related to last week's material. This activity is not graded.

II. Introduction to PROC SQL

a) Introduction to the SQL Procedure

i) SELECT statement and clauses in the SELECT statement

ii) DESCRIBE TABLE statement

b) Data Set Options and PROC SQL Options

i) Data Set Option (OBS=)

ii) Controlling processing with INOBS= OUTOBS=

iii) Controlling display with NUMBER

III. PROC SQL Fundamentals

Generating Simple Reports

a) Filtering Rows Using the WHERE Clause, combining expressions using OR and AND.

Activity 1 – Orion Star wants to identify the customers who reside in Turkey, display the customer ID, date of birth, and country. In another query, identify customers who reside in Turkey or Germany. Finally, identify the customers reside in this set of countries (Australia, United States, or Canada).

b) Handling missing values with IS NOT NULL etc.

Activity 2 – A nonprofit organization has data about past donations. Use the PVA97NK dataset to identify donors who are younger than 30 years old. Is there a problem with this query? How can we fix it? Try filtering for donors who are younger than 30 yo, AND the age value is not missing. Use the BETWEEN operator to filter donors 20 – 30 years old.

c) Pattern matching with LIKE

Week 4 Meeting Agenda

Activity 3 – Freestyle query writing 😊 Filter customers either based on First name, last name or full name. Provide a pattern using wildcards such as % or _

d) Sorting the output with ORDER BY

Activity 4 – Filter employees reside in San Diego, and the postal code starts with 920. Use LIKE operator and wildcard % to write this query. Add an Order by clause to sort the table in decreasing order of postal code and ascending order of street number and street name.

e) Creating a new column

f) Subsetting Calculated Values

Activity 5 – Using the yrdif(date1, date2) function, calculate the fractional age of the customers based on their birthdates and today's date. Rename the newly created column: AgeFractional. Use a WHERE clause to filter the dataset based on this newly created column AgeFractional GE 70. What seems to be the error? How can we fix it? There are two ways: Either use the same function you create AgeFractional column in the WHERE clause or tell SAS this AgeFractional is a calculated column by adding CALCULATED keyword in front of the variable name in the WHERE clause.

g) Assigning Values to a New Column Conditionally

Activity 6 – Suppose we want to decode the order type column in the Orders table. Instead of 1, 2, 3, we want 'Retail', 'Catalog' and 'Online', respectively. Within the SELECT statement, create a new column called Order_Category. Use CASE-WHEN operator with the syntax when Order_Type = 1 then "Retail" and so on. Alternatively, you can use CASE-OPERAND syntax (more concise) such as: case Order_Type when 1 then "Retail" and so on. Do not forget to finish it with end and rename the column as Order_Category.

IV. Q&A

V. Exercise #3

VI. Quiz #1