ALVAREZ College of Business

DA 6223 Data Analytics Tools and Techniques

Week 7 Meeting Agenda

I. Weeks 5&6 Recap with Questions

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

II. Transforming Tables

a) Append Table Task

When appending tables, all tables should be similar in structure.

- A column that is in some but not all input tables will have a missing value in each row of the result that comes from an input table that does not contain the column.
- Columns with the same name should also have the same data type. Otherwise, the Append Table task will fail.
- If the columns are of different lengths, SAS will still append the tables. However, truncation may occur. There will be a warning in the log that indicates that multiple lengths were specified.
- The formats from the first table will be used. There will be a note in the log that indicates that different formats were found.

Demonstration 1: Append EMPS_AU, EMPS_DE and EMPS_US. This demonstration illustrates appending tables with different structures.
Demonstration 2: Append orders09, orders08, and orders07 in this order. Specify **SummerOrders** as the name of the output table. Why do the Order_Date values appear in the **MM/DD/YYYY** format?

b) Split Columns Task

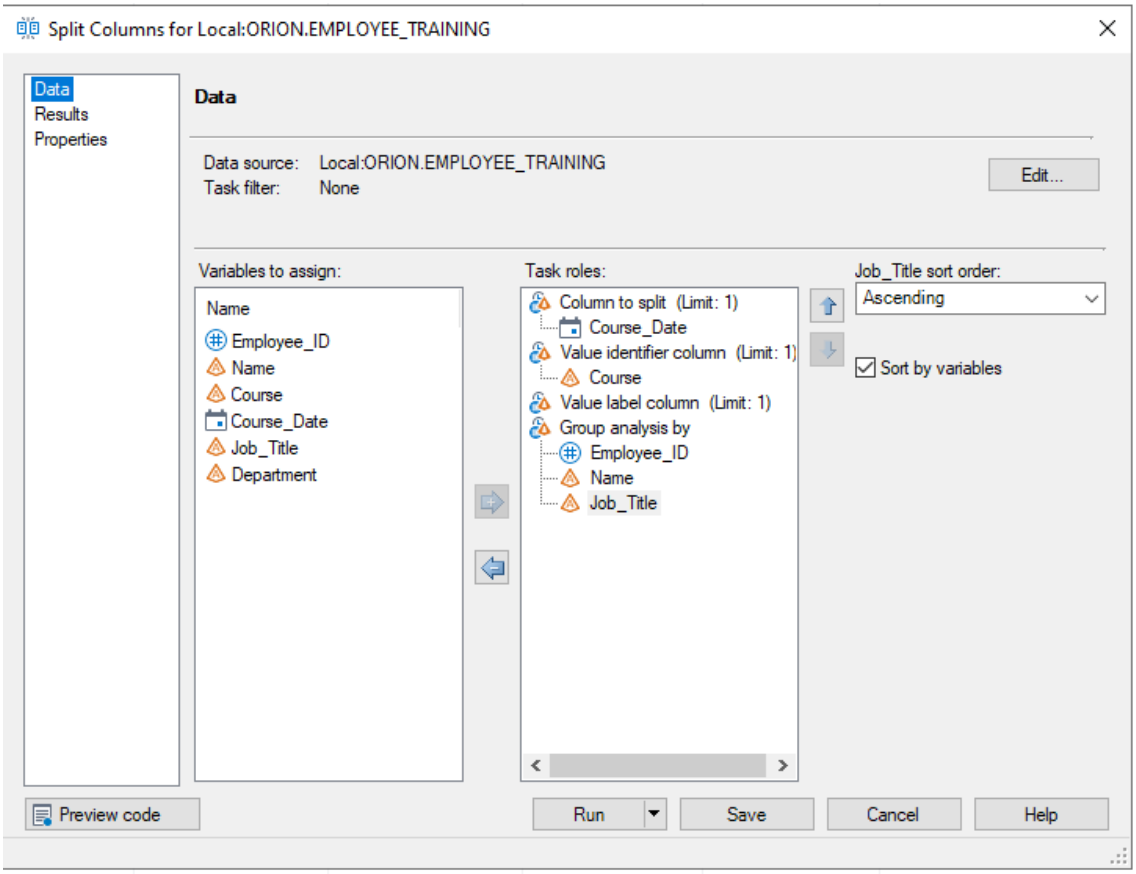
When assigning columns to task roles, there are three questions to ask:

- Which column is being split?
- Which column identifies the values being split?
- Which column or columns group the data?

https://documentation.sas.com/doc/en/egcdc/8.1/egdoccdc/egamotasks/p0wzr3yy5lbiq5n1dekt940lw38n.htm

Demonstration: A Human Resources manager at Orion Star needs a report that shows the training classes that employees have attended. Data for classes is stored in multiple rows for each employee,

but it would be easier if all class dates for each employee were in one row for the report. Use the Split Columns task to restructure the **employee_training** table from which a list report can be generated.



c) **Stack Columns Task**

When assigning data, there are two questions to ask.

- Which columns will be stacked?
- Which column contains the grouping information?

https://documentation.sas.com/doc/en/egcdc/8.1/egdoccdc/egamotasks/n012bqyr5tl5o6n1wawr9p0xbu6u.htm

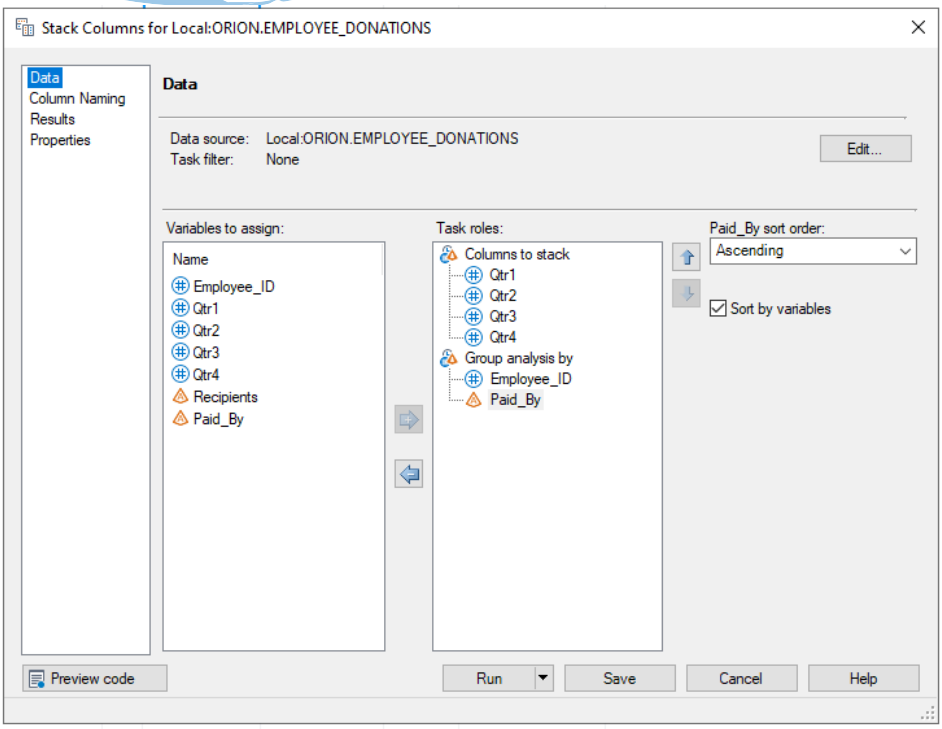
Demonstration: The Orion Star payroll manager would like a graphical report showing the charitable donations made each quarter by Orion Star employees. Use the Stack Columns task to restructure the **employee_donations** table from which a graph can be easily generated.

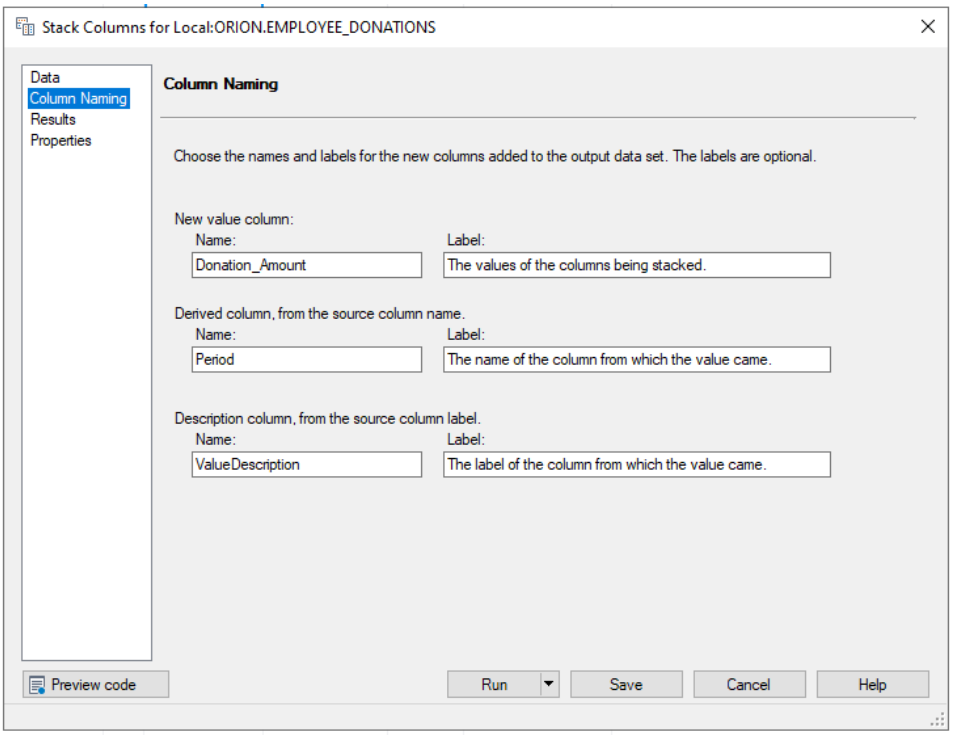
ALVAREZ
College of Business

DA 6223 Data Analytics Tools

and Techniques

Week 7 Meeting Agenda





Week 7 Meeting Agenda

III. Manipulating Data Values

a) Manipulating Numeric Values

Demonstration: Use the **Quarterly_Purchases** table to give customers promotions based on quarterly spending averages.

Query Builder

QuarterlyAvg: MEAN(t1.Qtr1Purchases,t1.Qtr2Purchases,t1.Qtr3Purchases,t1.Qtr4Purchases)

PromotionAmount:
(MEAN(t1.Qtr1Purchases,t1.Qtr2Purchases,t1.Qtr3Purchases,t1.Qtr4Purchases))*0.05

RoundedPromAmount:
ROUND(((MEAN(t1.Qtr1Purchases,t1.Qtr2Purchases,t1.Qtr3Purchases,t1.Qtr4Purchases))*0.05) , 5)

Demonstration: Continue with the previous exercise. This time give promotions on their birthdays.

PromotionDay: MDY(MONTH(t1.Customer_BirthDate), DAY(t1.Customer_BirthDate), YEAR(TODAY()))

The default delimiters for the PROPCASE function are a blank, forward slash, hyphen, open parenthesis, period, and tab. To use a different list of delimiters, specify a list of characters in a single set of quotation marks as the second argument in the function.

b) Manipulating Character Values

Demonstration: Continue the previous exercise, and create a new column for the Customer's First Name to be used in the promotional email.

Customer_FirstName: SCAN(t1.Customer_Name, 1)

Demonstration: Use the **Product_Orders** dataset to show the Product Name and Customer Names (as First Name and Last Name) who purchased the product group =007. The group starts at 5th character of the Product_ID.

Product_Name: PROPCASE(t1.Product_Name)

Customer_FirstName: SCAN(t1.Customer_Name, 2)

Week 7 Meeting Agenda

Customer_LastName: SCAN(t1.Customer_Name, 1)

Customer_Name: CATX(" ", (SCAN(t1.Customer_Name, 2)),(SCAN(t1.Customer_Name, 1)))

Filter: **SUBSTR(t1.Product_ID, 5, 3) = '007'**

c) **Converting Data Type**

Demonstration: Use **CharacterData** and convert the columns to the appropriate type.

Salary: INPUT(t1.Salary,10.)

Bonus: INPUT(t1.Bonus, dollar10.)

New_Salary: SUM((INPUT(t1.Salary,10.)),(INPUT(t1.Bonus, dollar10.)))

Hire_Date: INPUT(t1.HireDate, mmddyy10.)

Demonstration: Use **NumericData** and convert the columns to the appropriate type.

Phone_Number: PUT(t1.Phone, 10.)

Account_Num: PUT(t1.AccountNum, Z10.)

SSN PUT(t1.SocSecNum, SSN.)

IV. Creating Simple Queries with Tasks – JOINING TABLES

a) Join Types

i) Inner Join

ii) Full Outer Join

iii) Left Join

iv) Right Join

V. Q&A