ALVAREZ College of Business

DA 6223 Data Analytics Tools and Techniques

Week 11 Meeting Agenda

I. Transforming Data Part II

1. Recoding Columns using Query Builder

a. Replace Values

b. Replace a Range

Demonstration 1: To further analyze profit per order, management would like to recode the values of Order_Type to include a detailed description of the order type rather than a number. In addition, management would like to categorize each order by ranges of Profit. See slide #3 of Transforming Data Part II for the order type and profit categories. Use the Query Builder's Recode Columns option to complete this task. Then use another Query Builder to summarize the output table to get the total profit and total quantity per the order type and profit category.

	Order_Type_Detail	Profit_Category	SUM_of_Profit	SUM_of_Quantity
1	Retail Sales		$-2.82	2
2	Retail Sales	$0 to $100	$9,119.97	393
3	Catalog Sales	$0 to $100	$5,035.35	185
4	Internet Sales	$0 to $100	$3,182.60	130
5	Retail Sales	$100.01 to $500	$10,708.88	146
6	Catalog Sales	$100.01 to $500	$8,065.75	84
7	Internet Sales	$100.01 to $500	$7,009.75	91
8	Catalog Sales	$500.01 and Above	$5,500.10	24
9	Retail Sales	$500.01 and Above	$4,157.70	18
10	Internet Sales	$500.01 and Above	$1,390.70	5

c. Replace Condition

Demonstration 2: A manager at Orion Star is interested in the number of available men's and women's shoes versus all other categories. Use the Query Builder to create a new column based on the words found in the **Product_Name** column of the **shoe_vendors** table. Then use the One-Way Frequencies task to compare the availability of the different categories of shoes.

Week 11 Meeting Agenda

Men's and Women's Shoe Styles

The FREQ Procedure

Shoe_Categories	Frequency	Percent
Men's Shoes	174	48.20
Other	108	29.92
Women's Shoes	79	21.88

Writing CASE Expressions

Demonstration 3: Modify the Financial Quarters Query you created in Activity 1.

Using the purchased_products table, create a new column named Financial_Quarter that holds the correct quarter values for Orion Star's financial year. Because the Orion Star financial year starts December 1, the QUARTER function cannot be used, and quarters must be grouped as follows:

Quarter	Time Period
1	12-2 (December through February)
2	3-5 (March through May)
3	6-8 (June through August)
4	9-11 (September through November)

Then, create a table that summarizes the number of sales and the sum of total_retail_price by quarter.

Demonstration 4: Writing a CASE Expression with the WEEKDAY Function

Add the orders table to the project. Create a query named **Order Day Query** to group orders based on the day of the week on which the order was placed. Name the output table **OrderDay.** Include the columns Order_ID and Order_Type in the new table. Use the WEEKDAY function and

the Order_Date column to create a new column named OrderDOW, which is the day of the week on which each order was placed. Create a new column named WeekdayGroup based on the value of OrderDOW. WeekdayGroup has a value of *Monday-Thursday* when OrderDOW has a value of *2*, *3*, *4*, or *5*, and a value of *Friday-Sunday* when OrderDOW has a value of *1*, *6*, or *7*.

Do more purchases occur during Monday-Thursday or Friday-Sunday?

2. Creating and Applying Custom Formats

Format names can contain only letters, digits, and underscores. They cannot begin or end with a digit. The maximum length depends on the format type. The name of a character format cannot be longer than 31 characters. The name of a numeric format cannot be longer than 32 characters. For a complete set of naming rules, click Help through the Create Format task.

In the **Define formats window**, you build a table of correspondences to associate data values with formatted values. There are two parts to the format definition: the label and the range. You define a label to specify the text to be displayed in place of the original data values. Then you define a range to specify one or more ranges of stored values to be translated into that text.

There are two types of ranges: discrete or range. **Discrete** accepts single values such as CA or OK. **Range** accepts a low-end value (such as 100) and a high-end value (such as 200) to define a range of values. Endpoints can be designated as inclusive or exclusive.

Format definitions are **case-sensitive**. Therefore, the text supplied in the Values field of the Create Format task must be identical to the text in the column or columns to which you apply the format in the data table. If an equivalent value in the data table's column is stored in a variety of cases (for example, OK, ok, oK, or Ok), you must define each possible case in the format definition.

When defining ranges, you cannot enter the keywords **High** and **Low** in the Values boxes. You *must* select them from the drop-down list.

The following symbols indicate whether the range endpoints are included or excluded:

○—○	Excludes the values on the left and on the right

Week 11 Meeting Agenda

●—○	Includes the value on the left but excludes the value on the right.
○—●	Excludes the value on the left but includes the value on the right.
●—●	Includes the values on the left and on the right

Demonstration 5:

Regional managers at Orion Star would like to compare profits based on order type and region. The data includes coded values that must be displayed with descriptive labels to create the desired report. Use the Create Format task to provide descriptive labels for Order_Type, and use the Create Format from Data Set task to display region names instead of countries.

Creating a Format Task: Create a user-defined format that displays Order_Type values (1, 2, or 3) with more descriptive labels. Because you format numeric values (1, 2, and 3), change the Format type value to Numeric and then enter OrderTypeFmt in the Format name field.

Select Define Formats in the selection pane. The label should be Retail Sales if the value is 1. Similarly, add catalog and internet sales as well.

Creating a Format from a Data Set Task: In the ORION library, right-click COUNTRY_REGION_LOOKUP and select Properties. Click the Columns tab. Notice that the length of the region column is 30. That value is used later when you create the format. Click Close.

Name the format RegionFmt. Assign country_code as the discrete value and region as the label. Change the maximum label length to 30. Click Run to generate the RegionFmt format. View the log to confirm that the format was created successfully. Close the Create Format tab.

Applying User-Defined Formats: Add the orion_profit table to the project. Launch the Summary Tables task. Drag **Profit** to the Analysis variables role. Drag **Customer_Country** and **Order_Type** to the Classification variables role. Change the Heading format of both variables to the User Defined formats, e.g. **$REGIONFMT**. and **OrderTypeFmt.**

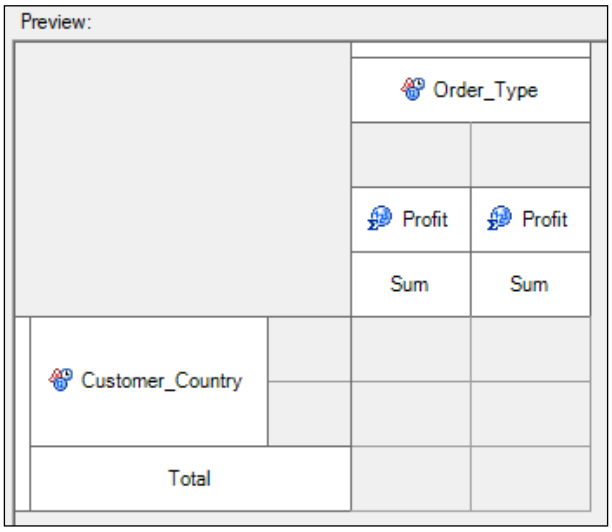
ALVAREZ
College of Business

DA 6223 Data Analytics Tools

and Techniques

Week 11 Meeting Agenda

Create the summary table in this form:



Summary Tables

	Order Type		
	Catalog Sales	Internet Sales	Retail Sales
Africa	$1,587.75	$1,008.55	.
Asia/Pacific	$4,024.60	$706.00	$8,040.69
Europe	$5,234.25	$2,814.00	.
North America	$7,754.60	$7,054.50	$15,943.04
Total (ALL)	$18,601.20	$11,583.05	$23,983.73