# ORIE 4820: Spreadsheet-Based Modeling and Data Analysis Acme Order Analysis Lab Exercise (Part I) Spring 2013

The focus of this two-part exercise is *fundamental data analysis*. We will cover a wide range of Excel shortcuts, functions, and tools for summarizing, extracting, filtering, partitioning, and viewing data in ways that are both flexible and reusable.

The template file for this exercise is *Acme-order-analysis-Part-I.xlsm*. Download a copy of the file from the course Blackboard site and *save it on your computer*.

#### <u>Topics</u> and tools we will cover in Part I:

- Setting *Excel Options*
- Performing "quick view" summarization using the spreadsheet status bar
- Using *shortcuts* to select cell ranges, filter data, populate columns, view formulas, etc.
- *Naming* cell ranges and using names within functions
- Using AutoFilter tool to extract records that meet certain criteria
- Using *date and text* functions to cast dates into various forms: **text**, **month**, **weekday**, **day**
- Using *text concatenation* to create dynamic headers and titles: &
- Creating and manipulating *PivotTables* and *PivotCharts*
- Retrieving data using lookup and reference functions: vlookup, match

If you have problems or questions at any point during the exercise, please raise your hand.

### **Background:**

It is now January 2013. You are a business analyst at Acme Merchants, a small merchandising company based in the Midwestern United States that currently sells ten items in the *CallSign* product line: **Alpha, Bravo, Charlie, Delta, Echo, Foxtrot, Golf, Hotel, India,** and **Juliet.** 

The **Customer Orders** dataset on the worksheet of the same name contains a complete record of daily orders for *CallSign* products placed by Acme's top five customers from January 1, 2012 through December 31, 2012. The **Product Price Table** and **Product Cost Table** listed at the top of the worksheet give, respectively, the per-unit prices and costs that have been applied for each product type during 2012. The selling prices that Acme charges its customers for each product type are updated monthly (on the first day of each month). The purchasing costs that Acme pays to its *CallSign* suppliers are also updated on the first of each month.

#### **Section 1: Setting Excel Options**

- (1) The parameters for controlling the development environment that resided in many different menu locations in Excel 2003 (including options for Calculation, AutoCorrect, Saving, Error Checking, and Add-Ins, among others) have been *consolidated* in later versions of Excel under *Excel Options*. In general, you should check this area if you are looking for a parameter setting that you cannot find elsewhere. *To modify Excel Options*:
  - a. From the top-left corner of your screen, select File, and click Options.
  - b. Under Formulas, Calculation options, make sure that "Automatic" is selected. (In subsequent exercises when we work with data tables, we will discuss circumstances where you may want to modify this option during development.)
  - c. Under Customize Ribbon, make sure that "Developer" is *checked* on the Main Tabs ribbon. You can use this form to set and organize the options that you want to have available under any of the Main Tabs.
  - d. Under Quick Access Toolbar, you can set and organize the options that you want to have available on the Quick Access Toolbar (directly above the Main Tabs ribbon).
  - e. Under Add-Ins, make sure that Analysis Toolpak, Analysis Toolpak VBA, and Solver Add-In are *active*. If they are not, then use the Manage dropdown menu at the bottom of the form to activate these Excel Add-Ins.
- (2) Click OK.

### **Section 2: Naming Cell Ranges**

- (1) Let's begin by *verifying that there are 18,300 rows* (366 days x 5 customers x 10 products) in the Customer Orders dataset:
  - a. Select all of the data in the Order Quantity column, *not including* the header. A quick way to do this is to click on the top cell of the column containing data (cell D22), then hold down the Ctrl and Shift keys *together*, and press the down arrow (↓).
  - b. Any time two or more cells are selected, the status bar at the bottom of the screen (showing "Ready" on the left) will display summary statistics of the currently highlighted cells. If you do not see "Count", right click on the status bar and check the "Count" option. The statistics will only show as long as the cells are selected. This status bar summarization feature is an excellent tool for "sanity checking" numbers quickly.
  - c. To keep a permanent record count on the spreadsheet, select the cell K20, then type "=COUNT(D22:D18321)", and press Enter. Note that the **count** function *only* counts cell with numeric data, while the **counta** function counts cells with numeric or alphanumeric data.
- (2) In analyzing data and building decision models, it is frequently useful to *name* cell ranges. Naming cell ranges both increases spreadsheet clarity and reduces the likelihood of formula errors. Let's *name the cell containing the order count* that you just populated using the built-in Name Box:
  - a. Select cell K20.

- b. Click on the Name Box dropdown arrow in the upper-left hand corner of the worksheet.
- c. Type "Order\_Count" (or any name without spaces or special characters) and press Enter.
- d. Select any blank cell and type "=Order\_Count", and press Enter. It should display the same value that is in cell K20.

<u>Note</u>: When you select the Name Box dropdown arrow, *all* of the names in the workbook are displayed and can be selected to highlight a named range. In this workbook, Price\_Table and Cost\_Table are named ranges that we will use later.

- (3) Now *name each column of data* in the Customers Orders dataset after its header text:
  - a. Place the cursor anywhere within the dataset and press Ctrl-Shift-space. This will highlight the entire contiguous cell block.
  - b. Since we do not want row 20 included in the selection, hold down the Shift key and press the down arrow  $(\downarrow)$  once.
  - c. From the ribbon, select Formulas->Defined Names->Create from Selection.
  - d. Check the "Top row" box only, and click OK. The data in the columns will now be named after their respective headers, with the spaces in the headers replaced with underscores "\_". (See Appendix B for more information on naming restrictions.) To check this, click the Name Box dropdown arrow.
  - e. Select any blank cell and type "=SUM(Order\_Quantity)", and press Enter. This will give you exactly the same result as if you entered "=SUM(D22:D18321)". Note that you can use the F3-key to paste a name instead of typing it.

## **Section 3: Using Excel's AutoFilter**

- (1) Excel's *AutoFilter* is an extremely useful tool that allows you to identify records in datasets that meet specific criteria:
  - a. Place your cursor anywhere in the Customer Orders dataset.
  - b. From the ribbon, select Data->Sort & Filter->Filter (or Ctrl-Shift-L). Each field name will now have a drop down list that allows you to sort the records according to the field, extract records that have specific field values, or to set a customized filter. If you filter a dataset on multiple fields, only the records that satisfy *all* criteria will remain in the list. The records not meeting the criteria are *not deleted*, they are just hidden in a special way (see Appendix B for more information on filtered vs. hidden data). Note that the drop down arrow for a field turns into a *filter cone* when a filter or sort order has been set.
  - c. Example: To quickly determine the total amount of product Hotel sold to Customer 4 in 2012, select the appropriate field restrictions on the Customer and Product fields, then use the status bar summarization tool on the Order\_Quantity field it works on filtered lists! You can also make mass changes to the entries in filtered lists (including deleting rows) without affecting the filtered data.
  - d. To clear a filter from a field, click the filter cone and select "Clear Filter from <field>". To clear the filter from all fields, press Ctrl-Shift-L or select Data->Sort & Filter.>Filter.

(2) *Use the AutoFilter to answer the indicated questions <u>as quickly as you can</u>. You will need to customize the Text/Number Filters for some of the questions.* 

**Note:** Other mechanisms in Excel that allow you to summarize filtered data include:

- The **subtotal** function accepts two parameters: a *number* which corresponds to the type of summarization you want, and the *column of data* to summarize. When the *column of data* is actively filtered, any rows excluded from the filter result are also excluded from the subtotal.
- The **Subtotals tool** (Data->Outline->Subtotal) allows you to incorporate subtotals into a list or table according to user-specified changes (e.g., at every change in Customer). When the table is actively filtered, any rows excluded from the filter result are also excluded from the subtotals.
- Excel **Dfunctions** (e.g. dsum, daverage, dcount, etc.) summarize data stored in a list or table according to user-specified criteria. With Dfunctions, the criteria specify the records to be included in the summary, so you do not have to explicitly filter the data.

Although these tools and functions provide some additional capabilities, they can be awkward to use and exist primarily for backward compatibility with previous versions of Excel. In this class, we will focus on more powerful tools such as PivotTables and array functions.

### **Section 4: Using Date and Text Functions**

In order to analyze Acme's 2012 *CallSign* orders thoroughly, you will need to summarize, partition, and view the data over different periods of time. In order to facilitate this, we will augment the dataset with fields that specify the Order\_Month (i.e., January, February, etc.) and Order\_Day (i.e., Sunday, Monday, etc.) of each order.

- (1) **Populate the Order Month field** for each record using the **text** function:
  - a. In the top cell of the Order Month column (E22), type "=TEXT(A22,"mmmm")" and press Enter. This function accepts any numeric value as input and converts it to a text string in the format specified. The second parameter "mmmm" indicates that we want the full month name. If we had used "mmm" instead, abbreviated versions would be returned (i.e., Jan, Feb, etc.).
  - b. To see of list of sample text format options, from the Ribbon, select Home->Number. Click on the dropdown box and select "More Number Formats...". Under Category, click Custom, and the right-hand pane will show a sample of formats you can use.
  - c. Copy the formula down the column by *double-clicking on the small square fill handle in the lower-right corner* of cell E22.

<u>Note</u>: For a good overview of the **text** function and the many formats options available, see: <a href="http://office.microsoft.com/en-us/excel-help/text-function-HP010062580.aspx">http://office.microsoft.com/en-us/excel-help/text-function-HP010062580.aspx</a>. If we had wanted month *numbers* (i.e., 1, 2, etc.) instead of month names, we could have used the **month** function, which accepts a date as input and returns the number of the corresponding month.

- (2) **Populate the Order Day field** for each record using the **text** function:
  - a. In the top cell of the Order Day column (F22), type "=TEXT(A22,"dddd")" and press Enter. The second parameter "dddd" indicates that we want the full day name.

b. Copy the formula in F22 down the column.

**Note**: If we had wanted the *number* of the day of the week (i.e., 1, 2, ..., 7) instead of the name, we could have used the **weekday** function, which accepts a date as input and returns the number of the corresponding day of the week. Similarly, the **day** function returns the number of the day of the corresponding month (i.e., 1, 2, ..., 31).

#### **Section 5: Creating PivotTables and PivotCharts**

Excel's versatile PivotTable/PivotChart tool allows you to partition a dataset based on any field or combinations of fields you have defined. Once you choose the field(s) on which to partition (e.g., by Product and Order Month), you can select any data field you want to analyze (e.g. Order Quantity), and it will be summarized in a table according to the partition. You can also view the results graphically using a PivotChart.

#### (1) To create a PivotTable:

- a. From the ribbon, select Insert->Tables->PivotTable.
- b. In the "Create PivotTable" dialog box, specify the entire Customer Orders dataset, including the headers in row 21, select New Worksheet, and click OK.
- c. On the new worksheet, click inside the specified PivotTable area, and an additional set of PivotTable Tools options will appear on the ribbon. Under Options->Show/Hide, make sure that you show the "Field List".
- d. Drag the fields that you want to use to partition the data into the *Row Labels* and/or *Column Labels* areas below the field list. For example, drag Order Month into Row Labels and Product into Column Labels.
- e. Drag the Order Quantity field into the *Values* area. It will default to "Sum of Order Quantity", but by clicking on it and selecting "Value Field Settings," you can select another summarization statistic and/or display option. If you keep the default setting, you will see a breakdown of the total number of units ordered of each product by month.
- f. To see a more detailed breakdown of the data, you can use *multiple* Row Label and/or Column Label fields. For instance, drag Customer into the Row Label area and drop it *above* Order Month to see a Product by Order Month breakdown for each Customer. If you drop Customer *below* Order Month, you will get a Product by Customer breakdown for each Order Month. You can also select multiple Value fields, although typically you will select only one Value field at a time.
- g. You can control the segments of the partition that will appear in the PivotTable using the PivotTable Field List entries. For instance, if you want to see the data for Alpha, Bravo, and Charlie only, then click on the Product entry of the PivotTable Field List and *uncheck* all of the items except those three products.
- h. You can also use *slicers* to filter the data. Click anywhere in your PivotTable, and then select Options->Sort&Filter->Insert Slicer. Each slicer provides buttons that you can click to *include* data in a PivotTable (as opposed to having to *uncheck* boxes in a PivotTable Field List). Slicers also indicate the current filtering state, which makes it easy to understand what exactly is shown in a filtered PivotTable report.

- i. To edit the PivotTable, simply drag and drop items from the Field List into or out of the desired locations. When dragging a field out of the table, do not release the mouse until you see an "x".
- (2) PivotCharts are the graphical counterparts to PivotTables and can be very useful visualization tools, but it takes practice to get used to manipulating them. *To create a PivotChart*:
  - a. Click anywhere in your PivotTable, and then select Options->Tools->PivotChart. An "Insert Chart" dialog box will appear. Select the type of chart you want to insert (for instance, Line Chart with Markers).
  - b. Click inside the PivotChart area, and an additional set of PivotChart Options will appear on the ribbon. Under Analyze->Show/Hide, make sure that you show the "Field List" and the "Field Buttons". <u>Caution</u>: *Unless you show the "Field Buttons", the chart will not indicate whether you are viewing complete or filtered data.*
  - c. Fields shown in the *right legend* of the chart page correspond to the Column Label partition fields you selected for the PivotTable. Fields shown across the *bottom axis* of the chart correspond to the Row Label partition fields. (Once a PivotChart is created, Column Labels and Row Labels become *Legend Fields* and *Axis Fields*, respectively.) You can now control the segments of the partition to display using either the Field Buttons or the PivotTable Field List entries. *Since the PivotTable and PivotChart are linked together*, *changing an item on one will automatically change the other*.
  - d. Report Filter fields allow you to restrict the records included in the analysis and essentially act like an AutoFilter on the data before the partitioning begins. If no Report Filter fields are specified, then all of the data will be used in the summary. For example, click anywhere on the PivotTable to activate the Field List, and drag the Customer field into the Report Filter area. Above the PivotTable, you now have a Customer dropdown list from which you can select a specific customer or (by checking the Select Multiple Items box) any combination of customers. If more than one item is selected, you will be viewing the aggregated data across those items. A field can be used at most once as a Report Filter, Row Label, or Column Label. (It can be used as a Value field in addition to one of these, however.) Slicers (introduced in Excel 2010) can be used to produce the same results as Report Filters, with the added benefit that they clearly indicate the current filtering state. To select multiple items in a slicer list, hold down the Ctrl-key.
- (3) *To change the chart type*, right-click anywhere in the PivotChart and select Change Chart Type..., then choose the new type of chart you want to use. (Note: Line charts lend themselves well to time series data, while bar charts are frequently used for categorical data.)
- (4) *PivotTables and PivotCharts have a built-in grouping tool* that allows you to group Row Label and/or Column Label fields. To see this:
  - a. Drag Product into Column Labels in the PivotTable Field List.
  - b. Within the PivotTable itself, select multiple product labels (e.g., Bravo, Charlie, and Delta). Right-click, and select Group. The items will now be grouped within the PivotTable (and PivotChart), so that the group can be expanded or collapsed.
  - c. To edit a group name, highlight it, press the F2-key, and enter a new name.
  - d. To undo the grouping, right-click on the group label and select Ungroup.

# Section 6: Using the VLOOKUP and MATCH functions to Retrieve Data

Reference and lookup functions are among the most useful and time-saving tools in Excel. The **vlookup** function allows you to look up and retrieve information from a column-oriented table, based on either an *exact* lookup or a *range* lookup on the first column of data. (There is a corresponding function, **hlookup**, that allows you to look up and retrieve information from a row-oriented table, but **vlookup** is more commonly used.)

For instance, to get a complete picture of Acme's *CallSign* performance in 2012, you will need to determine what drove its *profits*, not just its order volumes. To do this, you need to compute the *sales revenue* and *cost of goods sold* associated with each order in the Customer Orders dataset. These figures depend, respectively, on the *unit selling price* of the product on the date the order was placed and the *unit cost* of the product on the date that Acme purchased the product from its supplier. (Since we do not have any information about Acme's purchasing policies at this point, *for simplicity we will assume that the unit cost associated with a customer order is the unit cost of the product on the order date.)* 

For each order, you need to use the order transaction date to *look up* the corresponding unit price in the Price\_Table and the corresponding unit cost in the Cost\_Table for the product sold. The **vlookup** function does precisely this. It takes four arguments as input. The syntax is:

 ${\bf vlookup}(lookup\_value,\ lookup\_table\_range,\ column\_to\_return,\ range\_value)$ 

where:

- *lookup\_value* is the value you want to use to isolate the correct row in a table. It is usually a reference to a cell that contains a numeric value or text. In this case, *lookup\_value* is the Date of the order transaction.
- *lookup\_table\_range* specifies the cell range of the data in the table (in this case, Price\_Table or Cost\_Table), *not including* the table headers. The first column of the range is the one in which *lookup\_value* will be sought.
- *column\_to\_return* is the relative column number of the *lookup\_table\_range* that contains the information you want to retrieve. In our case, there is *no one fixed value* for *column\_to\_return*, since the value depends upon the product that has been ordered: the prices for an Alpha order are in column 2 of the Price\_Table; the prices for Bravo are in column 3, and so on. Therefore, we need to make this entry dynamic.
- range\_value specifies whether you want an exact match on lookup\_value, or you want to find the row with the largest value that does not exceed lookup\_value. If you want an exact match, then range\_value should be set to 0 (i.e., FALSE); otherwise, it should be set to 1 (i.e., TRUE). In this case, range\_value should be set to 1, since the price we want to retrieve is dictated by the date range the order transaction date falls within (i.e., we are not looking for an exact match on the order transaction date itself).

<u>Note</u>: TRUE should be used for *range\_value* **only** when the first column of the table contains *lower range limits in sorted order*, and you want to look up the correct *range* for a particular value, *not* an exact match.

- (1) The sales revenue associated with each order transaction is the order quantity times the unit price. *Pre-populate cell G22* by typing "=\$D22\*VLOOKUP(\$A22,Price\_Table,2,1)" and pressing Enter. Note that the hard-coded 2 as the *column\_to\_return* parameter will return the unit prices for Alpha only. We will come back and fix this problem shortly.
- (2) Alternatively, press Shift-F3 to get function help and type "vlookup", or from the ribbon: select Formulas->Function Library->Lookup & Reference->VLOOKUP. Once the function is selected, you will be prompted and given descriptions for each function argument. These tools can be very helpful when you are unsure of a function's syntax.

To get the correct unit price (or unit cost) for each order transaction, we need the third **vlookup** parameter, *column\_to\_return*, to correspond to the column for the product that has been ordered. To do this, we will use the **match** function.

The **match** function takes as input a specified value, looks for a match to this value in a specified array, and returns the *relative position* of the matched item within the array. The difference between **match** and **vlookup** is that **match** returns the *position* of the matched item in the array (i.e., 1 if the matched item is in the first position, 2 if the second, and so on), *not* the item itself.

The syntax for the **match** function is:

match(lookup\_value, lookup\_array, match\_type)

where

- *lookup\_value* is the value you want to match in *lookup\_array*. It can be a value (number, text, or logical value) or a cell reference to a number, text, or logical value.
- *lookup\_array* is a contiguous range of cells or a reference to an array.
- *match\_type* is the number -1, 0, or 1:
  - o If *match\_type* is 1, the function finds the largest value that is less than or equal to *lookup\_value* and returns its relative position. If 1 is specified, then *lookup\_array* must be arranged in ascending order. (This is the default if *match\_type* is omitted.)
  - o If *match\_type* is 0, the function finds the first value that is exactly equal to *lookup\_value* and returns its relative position. The *lookup\_array* can be in any order.
  - o If match\_type is -1, the function finds the smallest value that is greater than or equal to *lookup\_value* and returns its relative position. If -1 is specified, then *lookup\_array* must be arranged in descending order.

For instance, **match**("Charlie", {"Date", "Alpha", "Bravo", "Charlie", "Delta", "Echo"}, 0) returns the number **4**, the relative position of "Charlie" within the specified array.

- (3) We are now ready to *fix the Sales\_Revenue field and populate the column*:
  - a. In cell G22, replace the **2** that is the third **vlookup** argument with "MATCH(\$C22,\$A\$6:\$K\$6,0)" and press Enter.
  - b. Copy this formula down the column.

Observe how we are using the **match** function here: we are matching the *product name* from the order record (in column C) against the *headers* in the Price\_Table in order to get the correct *column number* of the Price\_Table. The number returned from the **match** function serves as the *column\_to\_return* parameter within the **vlookup** function, so that the unit price returned is taken from the correct column of the table. Note that the last parameter of the **match** function is 0 since we want an *exact match on the product name*, but the last parameter of **vlookup** is a 1 (i.e., TRUE), since we are looking up the *range the order transaction date falls in*, and *not* an exact match on the order transaction date.

#### (4) To populate the cost of goods sold (COGS) field in a similar manner:

- a. The COGS associated with each order is the order quantity times the unit cost. In cell H22, type "=\$D22\*VLOOKUP(\$A22,Cost\_Table,MATCH(\$C22,\$M\$6:\$W\$6,0),1)" and press Enter.
- b. Copy this formula down the column.

#### (5) To populate the Gross\_Margin field:

- a. The gross margin associated with each order is simply the sales revenue minus the COGS. In cell I22, type "=G22-H22", and press Enter.
- b. Copy this formula down the column.

Assignment #1 is due Friday, February 8<sup>th</sup>. As part of this assignment, you will need to prepare a Powerpoint slide deck that:

- Communicates the relative order volume and profit contribution in 2012 of all *CallSign* products and customers; and
- Highlights specific *CallSign* product trends and customer ordering patterns that have had (or promise to have) a significant impact on Acme's bottom line.

You can spend the remainder of today's class getting started on this using the augmented dataset and the tools you have learned.