Lab – Preparing Data

1. Objectives

Part 1: Import the Data into Microsoft Excel

Part 2: Get the Month out of a Date

Part 3: Split Full Name Column into Different Columns

Part 4: Concatenate Columns

1. Background / Scenario

Preparing the data is a critical step in data analysis. You may receive data from different sources and you will need to manipulate the data and extract the necessary information.

In this lab, you will explore how to use the look up information function. You will also explore how to split and concatenate columns of data.

1. Required Resources

* 1 PC with Microsoft Excel installed (Other spreadsheet programs, such as LibreOffice or Google Sheets, may be used, but they may not have the same formula syntax)

1. Import the Data into Microsoft Excel
   1. Create a .csv file.
      1. Open text editor and paste the following text.

Full Name, Join Date

Olivia Jones, 02/28/2010

Lucas Smith, 9/14/2011

Ava Williams, 4/12/2012

Liam Johnson, 7/31/1999

* + 1. Save the file as a .csv.
  1. Import the .csv file into Excel.
     1. Open Excel.
     2. Select **Data** from the menu. Select **Get External Data** > **From Text**.
     3. Select the saved .csv file and click **Import**.
     4. The Text Import Wizard opens. In the Step 1 of 3 dialog box, select the **My Data has headers** checkbox. Click **Next**.
     5. In the Step 2 of 3 dialog box, unselect **Tab** checkbox and select the **Comma** checkbox. Click **Next**.
     6. Select the **Join Date** column, and select **Date**. Verify that the dropdown displays **MDY**. Click **Finish**.
     7. Click **OK** to open the data in the existing worksheet.

1. Get the Month out of a Date

Excel has other functions that allow you to parse out the information from a date. Furthermore, it allows you to lookup values in a different table. In this part, you will extract the month out of a date and use vlookup to convert it into a string that spells out the name of the month.

The MONTH function returns the month as a number from 1 (January) to 12 (December). VLOOKUP function looks for a value in the leftmost column of a table and returns a value in the same row from the column you chose.

* 1. Add a lookup table.
     1. To lookup the month, a table with the name of each month is created. In a new sheet, enter 1 – 12 in the first column and the name of month in the second column.
     2. Rename Sheet2 to Months.
  2. Parse the month information from the date.

In an empty cell in column C row 2, enter **=month(B2)**. The data in cell B2 is 2/28/2010. What did this formula return? Verify the number format is set to General (Right-click **Format Cells** > **General)**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* 1. Look up the name of the month.
     1. In same cell (C2), enter **=VLOOKUP(MONTH(B2),Months!A2:B13,2,FALSE)**. What did this formula return?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In the vlookup function, the first argument is the name of the item. In this case, it was MONTH(B2). The second argument specifies the cell range that contains the data that VLOOKUP will search and return the lookup value (Months!A2:B13). The column index number is the third argument (2).

You provided the month as a value using the Month function (MONTH(B2)). Next, the location of the lookup table was specified (Months!A2:B13). The third parameter indicates the column index number. Because the information in the Month column is desired, the column index number is 2. The last argument indicates whether to look for approximate matches. In this case, the desired result is an exact, so the last argument is FALSE.

In this example, VLOOKUP searches through the table in the Months sheet until it finds 2 in the first column. Then it will move to the second column to find the Month and returns February.

* + 1. Copy and paste the formula into other empty cells in the same column.
    2. Name the column **Join Month**.

1. Split Full Name Column into Different Columns

You can split text in one or more cells and parse it to put it across multiple cells. In this part, you will split the Full Name column into two columns, Given Name and Family Name. You will split the data using some of the Excel functions: RIGHT(), LEFT(), and LEN().

* + 1. Select the **Formula** menu. Click **Text**. In the list of functions, read the description of these functions: RIGHT, LEFT, LEN, and FIND.

What do the RIGHTor LEFT functions provide?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What does the LEN function return?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What does the FIND function return?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* + 1. Using the name in A2, you can find the length of the string in the cell by entering **=len(A2)** into a blank cell in column D, row 2.

How many characters are in the name Olivia Jones?

* + 1. If you only want to display the first 4 characters in a string, you can enter **=LEFT(A2, 4)** into the same cell.

What was the result?

* + 1. Add the column **Given Name** in column D, row 1. Under the Given Name heading, enter **=LEFT(A2, LEN(A2)-(LEN(A2)-FIND(" ",A2)+1))**.

What was the output?

* + 1. Copy and paste the formula to the empty cells in the same column. Observe the results.
    2. Explain how the formula is applied to the data in the Given Name column.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* + 1. Enter **Family Name** as the header in column E row 1. Based on the Given Name, write a formula for Family Name and copy and paste the formula into other empty cells under the Given Name column.
    2. What formula did you use to parse out the Family name?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Concatenate Columns

Now that you have split the full names into family name and given name, you will concatenate them into one column, with the full name in this format: family name, given name. You will use the ampersand (&) symbol to join the data from different columns together.

* + 1. Enter **Full Name (Family Name, Given Name)** in column F row 1.
    2. In column F row 2, enter **=E2&", "&D2**. What was the results? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ J

In the formula, E2 refers to Jones and D2 refers to Olivia. The &", "& joins Jones and Olivia with comma and space (, ).

* + 1. Copy and paste the formula to rest of the column.
    2. If you would like to copy the manipulated data into another spreadsheet, you may want to copy and paste the new data as values instead of formulas.
    3. Copy the new data and navigate to a new sheet. Right-click to **Paste Value**. Click into the cells that had formulas. What are in the cells?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* + 1. In the Join Date column, what is number format for the date? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

How do you change the number format back to the date? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Reflection

When would you want to paste the formulas versus values into a new spreadsheet?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_