Lab – Formulas and Functions

Objectives

In this lab you will learn the basics of using formulas and functions in Microsoft Excel.

Part 1: Using Excel Text Functions: CONCAT, LEN, LEFT, RIGHT, and MID

Part 2: Using Excel Statistics Functions: COUNT, COUNTBLANK, COUNTA, COUNTIF, AVERAGEIF, MINIFS, and MAXIFS

Part 3: Experiment with Other Excel Functions

Background / Scenario

Excel has eleven different categories of built-in functions. Many of the functions used by data analysts belong to the Text and Statistical function categories. The Text function category includes CONCAT, LEN, LEFT, RIGHT, and MID. The Statistical function category includes COUNT, COUNTBLANK, COUNTA, COUNTIF, AVERAGEIF, MINIFS, and MAXIFS. This lab will work through the purpose and usage of each of these functions.

Required Resources

- Mobile device or PC/laptop with a browser, Microsoft Excel, and internet access.

Note: The precise steps to format and manipulate data in Excel can vary between platforms and versions. The instructions in this lab are based on the free version of Excel available from Office.com and may have to be modified to match the user's platform, software, or version to achieve the results shown in this lab.

Instructions

Part 1: Using Excel Text Functions

Excel has several text (also known as string) functions that help data analysts work with textual data. These functions allow analysts to change text, change case, find a string, count the length of a string, and more.

Step 1: Open Lab Workbook

Open the **Bike Sales_Functions_Lab** workbook.

Step 2: Use the CONCAT Function to Combine Data

The CONCAT function, short for "concatenate," joins strings of text together. The function can include actual text strings enclosed in quotation marks as well as cell and range references. It also allows you to combine different types of data like numbers, dates, and text strings into a single cell.

One use case for the CONCAT function is to combine different pieces of important information for a human to be able to read at a glance, instead of going to multiple columns to find this information. In this step we will use the CONCAT function with our bike sales data to

simulate this use case, and combine the sales order number, order quantity, product subcategory, and date into a new "Sales_Summary" column.

    a.   Select cell U1 and enter the header text "Sales Summary"

This will be the column for the combined sales data

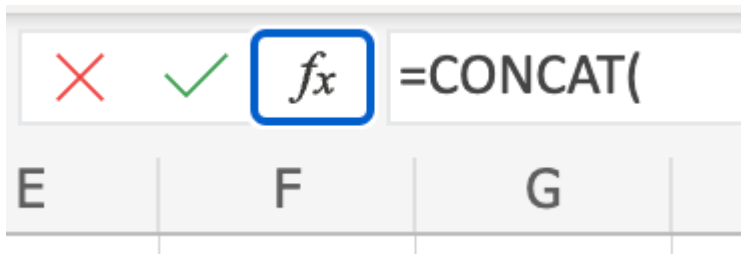   b.   Select cell U2 and type (don't copy and paste) **=CONCAT**

Excel may display a description of the function, "Concatenates a list or range of text strings," below the formula bar.
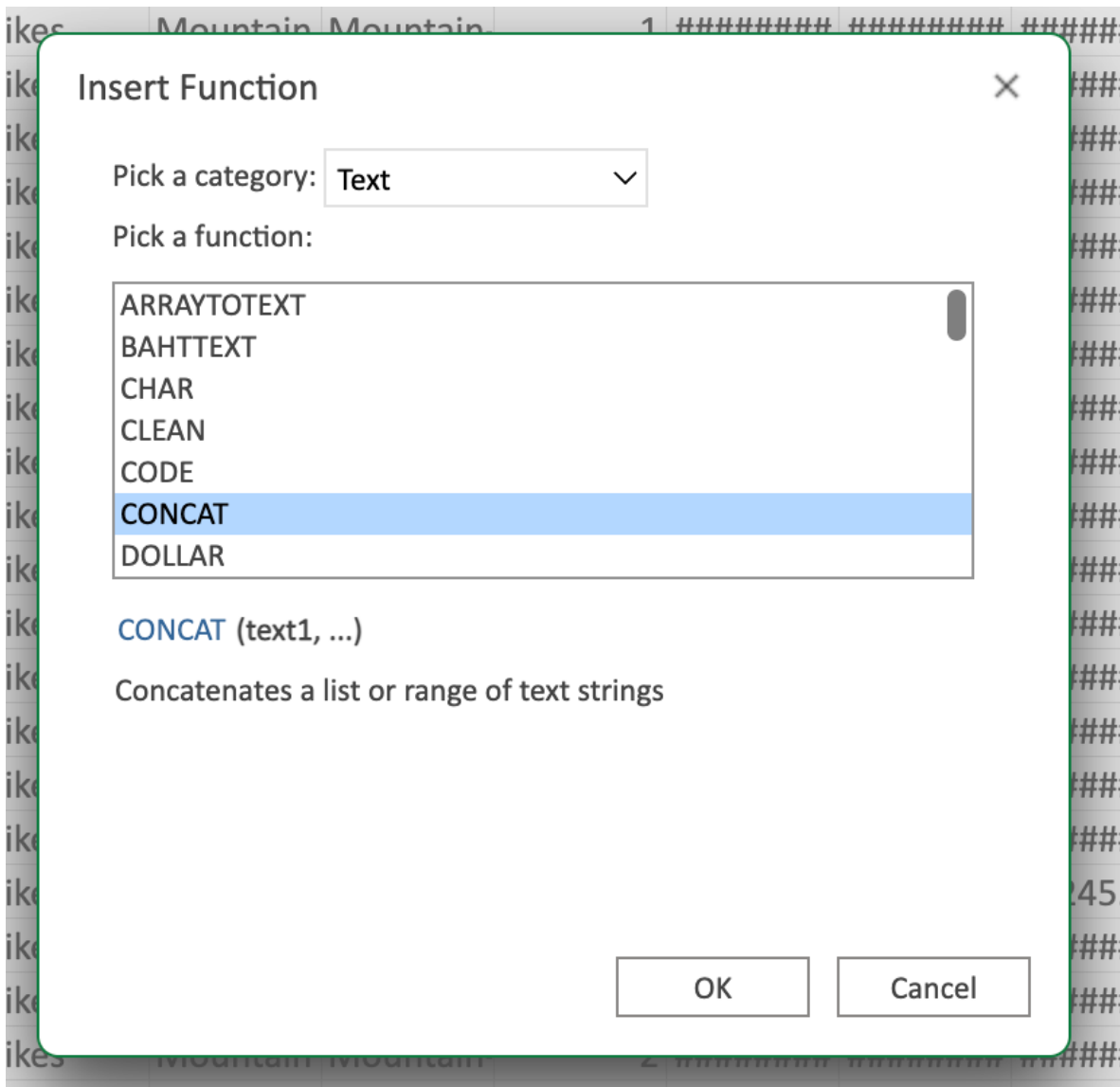
    c.   Add a parenthesis after the keyword so it reads =CONCAT(

Excel may display the syntax of the function, or how it must be written, below the formula bar.

=CONCAT(
CONCAT **(text1**, [text2], ...)

    d.   If it doesn't do this automatically, you can access the same information by clicking the Function button to the left of the formula bar, choosing "Text" in the Category drop-down, and selecting CONCAT from the list of functions.

✕	✓	*fx*	=CONCAT(
E		F	G

Insert Function   ×

Pick a category: | Text | ⌄

Pick a function:

> ARRAYTOTEXT
> BAHTTEXT
> CHAR
> CLEAN
> CODE
> **CONCAT**
> DOLLAR

CONCAT (text1, ...)

Concatenates a list or range of text strings

OK   Cancel

e. Notice that the text strings and cell references inside the parenthesis are separated by commas.

f. Note: if an actual text string (instead of a cell or range reference) is used, it must be enclosed by quotation marks.

g. Finish the writing the function in cell **U2** so that it will combine the order number (in cell **A2**) with the product subcategory (in cell **L2**):

1. In cell **U2**, finish typing = **CONCAT(A2, L2)**

Cell **U2** should now display the text "000261695Mountain Bikes" (it's also OK if the leading zeroes are missing).

2. To make the results more readable, let's edit the formula to add a colon and a blank space between the order number and "Mountain Bikes".

h. **=CONCAT(A2, ": ", L2).**

Note: Type the characters directly into the sheet, rather than copying and pasting from the lab, or you may get an error message.

1. For a challenge, edit the function to add additional information that includes the order quantity and the date. Be sure to enter a blank space between the order quantity and "Mountain Bikes," and add a dash and spaces between "Mountain Bikes" and the sales date.
Note: Excel stores dates as simple five-digit numbers; they have to be formatted using a TEXT function to be human-readable. **=CONCAT(A2,": ",N2, " ",L2," - ", TEXT(B2, "mm/dd/yyyy"))**

The results in cell U2 should be "000261695: 4 Mountain Bikes – 12/01/2021"

Note: When Excel thinks a function is text, it adds an apostrophe before the equals sign. If your function is showing up as text instead of returning data from other cells, check that there is no apostrophe before the equals sign in the formula bar.

1. Copy or drag the formula from cell **U2** to the remaining cells in **Column U**, from **U3** through **U89**.

i. Remove formulas from **column U**

1. The cells in **Column U** contain the **CONCAT** formulas. To remove the formulas, highlight and copy **Column U**. Then use **Paste** > **Paste Values** to paste just the values into **Column V**.

2. Delete **Column U** as the formulas are no longer needed.

What is the result if the formula in step 3 is written using just a cell reference to B2 to include the date, as =CONCAT(A2,": ",N2, " ",L2," - ", B2), rather than using the TEXT function?

Answer//The result for the B2 cell will display the **raw numeric value** of the date instead of the formatted date

Step 3: Use the LEFT, RIGHT and MID functions to Extract Information

Column M in the data set contains the product description of the bikes sold. It gives the bike model, the color, and the size. What if you needed this information separated into three separate columns for analysis? You can use the LEFT, RIGHT and MID functions to do this:

- The LEFT function will return a specified number of characters from the start (left) of a text string.

- The RIGHT function will return a specified number of characters from the end (right) of a text string.

- The MID function will return a specified number of characters from the middle of a text string.

a. First, create three new columns to the right of Column M, Product Description, that will contain the separated information. These will become columns N, O, and P.

b. Name the columns **Model**, **Color**, and **Size** respectively.

c. In cell **N2** enter the function **=LEFT(M2, 12)** to separate out the bike model. This function tells Excel to return the first 12 text characters from the left in cell M2.

The result should be "Mountain-200" in cell **N2**.

d. In cell **O2** enter the function **=MID(M2, 14,5)** to separate out the bike color. This function tells Excel to return the 5 characters after the 14th character (counting from the left) in cell **M2**.

The result should be "Black" in cell **O2**.

e. In cell **P2** enter the function **=RIGHT(M2,2)** to separate out the bike size. This function tells Excel to return the last 2 characters from the right (or end of the text string) in cell **M2**.

The result should be "46" in cell **P2**.

f. Copy and paste the functions in cells **N2**, **O2**, and **P2** down to the remaining cells in the columns.

g. While the functions returned the expected results for cells in row 2, notice that in other rows, some of the text is cut off as shown below. In row 3, the color is missing the "r" at the end of Silver. In rows 4 and 5, the –W is cut off the model name, and the color is shown as "W Sil." This is because the model names and colors of the product descriptions in Column M are not all the same number of characters. This is a common issue when splitting text between columns.

	M	N	O	P
	Product_Description	Model	Color	Size
1				
2	Mountain-200 Black, 46	Mountain-200	Black	46
3	Mountain-200 Silver, 42	Mountain-200	Silve	42
4	Mountain-400-W Silver, 46	Mountain-400	W Sil	46
5	Mountain-400-W Silver, 42	Mountain-400	W Sil	42
6	Mountain-200 Black, 46	Mountain-200	Black	46

There are multiple ways to correct this issue; one is to modify the formulas in these cells based on the length of the model names and colors. First, we will apply a filter to **Column M** to display only specific bike models, and then adjust the functions for those cells to the matching number of characters.

h. Highlight **Column M**, and then click **Sort & Filter** in the Editing menu group and select **Filter** in the dropdown menu.

i. Click the filter down arrow in the **Column M** header and click "Select All" to uncheck all the boxes.

j. Then check the boxes next to entries that are silver and that don't have the "–W" in the model name. This will be a total of 6 boxes. Then, click Apply.

Notice these are all missing the "r" in the color silver. To fix this, the MID function for these rows needs to return 6 characters after character 14, rather than 5.

	M	N	O	P
1	**Product_Description** ⏷	**Model**	**Color**	**Size**
3	Mountain-200 Silver, 42	Mountain-200	Silve	42
12	Mountain-200 Silver, 38	Mountain-200	Silve	38
17	Mountain-200 Silver, 38	Mountain-200	Silve	38
19	Mountain-200 Silver, 42	Mountain-200	Silve	42

k. Select the first cell, which should be **O3**, in the color column, **Column O** and adjust the formula

To read **=MID(M3, 14,6)**

l. Copy or drag this formula to the remaining cells in the column.

m. Next, click the filter down arrow in the **Column M** header and click "Select All" to uncheck all the boxes.

n. Then check the boxes next to entries that do have a –W in the model name. This will be a total of 3 boxes. Then, click Apply.

	M	N	O	P
1	**Product_Description** ⏷	**Model**	**Color**	**Size**
4	Mountain-400-W Silver, 46	Mountain-400	W Sil	46
5	Mountain-400-W Silver, 42	Mountain-400	W Sil	42
9	Mountain-400-W Silver, 42	Mountain-400	W Sil	42
14	Mountain-400-W Silver, 46	Mountain-400	W Sil	46

These rows will need an adjustment to the **LEFT** function in **Column N** to capture the whole model name and to the **MID** function in **Column O** to capture the whole color name.

o. In **Column N**, change the LEFT formula to return the first 14 characters.

=LEFT(M4, 14)

p. In Column O we need to change the MID formula to return the 6 characters (rather than 5) after the 16th character (rather than the 14th).

=MID(M4, 16,6)

q. Copy these formulas to the remaining cells in the columns.

r. Once the functions are correct, remove the filter from **Column M** by selecting **Filter** under **Sort & Filter**.

The functions in columns N, O, and P should now display the correct number of characters.

	M	N	O	P
1	**Product_Description**	**Model**	**Color**	**Size**
2	Mountain-200 Black, 46	Mountain-200	Black	46
3	Mountain-200 Silver, 42	Mountain-200	Silver	42
4	Mountain-400-W Silver, 46	Mountain-400-W	Silver	46
5	Mountain-400-W Silver, 42	Mountain-400-W	Silver	42
6	Mountain-200 Black, 46	Mountain-200	Black	46

    s.   Remove formulas from columns N, O, and P.

        1.   The cells in columns N, O, and P contain the LEFT, MID, and RIGHT formulas. To remove the formulas, add three new columns to the right of column P. Highlight and copy columns N, O, and P, then use Paste > Paste Values to paste the values into the three new columns,

        2.   Delete columns N, O, and P that contain the formulas.

    t.   Delete all columns added in Part 1 (columns N, O, P and U) to restore the worksheet to its original state before proceeding to Part 2 of the lab.

Part 2: Using Excel Statistical Functions

Excel has many built-in statistical functions that can perform all kinds of mathematical and statistical calculations. Some of the more common statistical functions used by data analysts include COUNT, COUNTBLANK, COUNTA, COUNTIF, AVERAGEIF, MINIF, and MAXIF. Let's learn how to use these functions to work with data.

Step 1: Use the COUNT Function to Find Cells with Non-numeric Values

    a.   The COUNT function in Excel returns the count, or number, of numeric values in a set or range of cells or values. This means it counts cells with numbers and dates, but not blank or text cells.

For this example, we will check Column N, Order_Quantity, for any blank or non-numeric (and therefore invalid) entries.

    b.   Add a new blank column to the right of **Column N**, Order_Quantity, this becomes **Column O**.

    c.   In cell O2, type the function **=COUNT(N2:N89)**

The result should be 88. Scroll to the end of the data to see that there are 89 rows filled. The top row is the headers, meaning there are 88 rows of data in the worksheet. The results of our COUNT function match the number of data rows we have, which means all the cells in column N contain numeric data and can be presumed valid.

f_x	=COUNT(N2:N89)	
N	**O**	
Order_Quantity		
4	88	

Now, let's introduce some non-numeric data and see how that changes the result of the COUNT function.

    d. Delete the contents of cell **N5** to make it a blank cell.

    e. Note that this changes the count of numeric cells in cell **O2** to 87 because cell **N5 now** contains no numeric value.

    f. Right-click on the cell and select Number Format from the drop-down menu. A pop-up will appear, and in the field labeled Category, choose Text. This will change the data type of cell **N5** to **Text**. **Now** type **1** into cell **N5**. Note that Excel left-justifies the number, because it is reading it as text.

    g. Note that the count remains 87, because the value of "1" in cell **N5** is now considered text and not numeric. The COUNT function only counts cells containing numeric values.

    h. Delete the contents of cell **N5**, change the data type back to General, and enter 1 to return the cell to its original value. The function in cell **O2** should again return a result of 88.

    i. Delete the COUNT function from cell **O2**.

Step 2: Use the COUNTA Function to Find Non-Blank Cells

The COUNTA function returns the count of non-blank cells in a range, whether the cells contain numeric values or text values.

Again, we will check **Column N**, this time to make sure there are no blank cells.

    a. Select cell O2 and type the function **=COUNTA(N2:N89)**

The result should be 88, indicating that all 88 cells from N2 to N89 contain data.

    b. As before in Step 2, delete the contents of cell **N5** to make it a blank cell.

Note, as before, that the count of numeric cells changes to 87 because cell **N5** contains no value.

    c. Change the data type of cell **N5** to **Text** again, and then type **1** in the cell again.

    d. This time the function returns 88, because the COUNT function counts cells containing both numeric and text values.

    e. Return cell **M5** back to the **General** data type and enter **1** to return the cell to its original value.

f. Delete the COUNTA function from cell **O2**.

Step 3: Use the COUNTBLANK Function to Find Blank Cells

Before analyzing data, it is a good idea to check for missing data. When importing data or with manual data entry, some cells that are supposed to have values in them may end up blank. The COUNTBLANK function is useful for identifying the presence of blank cells in a range.

The COUNTBLANK function returns the count of blank cells in a range of cells.

For this example, we will check Column N, Order_Quantity, for any blank cell entries.

a. Select cell O2 to enter the COUNTBLANK function

Type the function **=COUNTBLANK(N2:N89)**

The result should be 0 which is expected since there are no blank cells in cells N2 through N89.

b. Delete the values in a few cells in column M and observe the output of the function.

For each cell in which the values were deleted, the function should increase the count of blank cells.

c. Use the undo button at the left of the toolbar (or **Command+Z**) to return the cells to their original values. Delete column O.

Step 4: Use the COUNTIF Function to Find Non-Blank Cells

Another important task for data analysts to perform prior to analyzing data is to check for duplicate entries in a column. This, and many other tasks, can be accomplished with the COUNTIF function.

The COUNTIF function returns the count of cells that meet a specific criterion.

For this example, we will check for duplicated order numbers in column A, Sales_Order #.

a. Add a new column to the right of Column A. This creates a new column B.

b. Select cell B2 and type the function **=COUNTIF(A$2:A2,A2);** then copy the function to cells B3 to B89. This is telling Excel to count the total number of values in column A that match the value to the left of the formula. If you double click on the cells below B2, you can see the last cell reference is referring to the cell for that row in column A.

The result should be 1 for all the cells in column B, which means there is only one of each order number. In other words, there are no duplicate sales order numbers in column A.

c. Create some duplicate order numbers in **Column A**:

1. Copy order number 000261695 from cell **A2** into cell **A3** to make a duplicate entry.

2. Copy order number 000261700 from cell **A7** to cell **A8** to make another duplicate entry.

The function in cells B3 and B8 should now give a result of 2, indicating a second instance of the sales order numbers 000261695 and 000261700.

    d. Change the sales order number in cell A4 also to 000261695.

The function in cell B4 should now give a result of 3, indicating a third instance of the order number 000261695.

	A	B
1	Sales_Order #	
2	000261695	1
3	000261695	2
4	000261695	3
5	000261698	1
6	000261699	1
7	000261700	1
8	000261700	2

    e. Return the sale order numbers in cells A3, A4, and A8 to their original values using the Undo button or keyboard shortcut.

    f. Delete column B.

Step 5: Use the AVERAGEIF Function

The AVERAGE IF function finds the average of all the cells in a range that meet a given criterion. For this you will use the AVERAGEIF function to find the bike shop's average revenue by age group.

    a. In cell U2, enter the text "Youth"

    b. Select cell V2 to enter the AVERAGEIF function

The result should be 3533, which is the average revenue from mountain bikes purchased by youths

    c. In cell U3, enter the text "Young Adults" and in cell U4, enter the text "Adults"

Type the function **=AVERAGEIF(G2:G89, "Youth (<25)", S2:S89)**

    d. The result should be 3533, which is the average revenue from mountain bikes purchased by youths.

    e. In cell **U3**, enter the text "Young Adults" and in cell **U4**, enter the text "Adults"

    f. Write (do not copy and paste) the appropriate formula for calculating the average revenue from mountain bikes sold to the Young Adults and Adults age groups in cells **V3** and **V4** respectively:

1. For cell **V3**, use function **=AVERAGEIF(G2:G89, "Young Adult (25-34)",S2:S89)**

2. For cell **V4** use function **=AVERAGEIF(G2:G89, "Adult (35-64)",S2:S89)**

3. Change the number type in cells **V2, V3**, and **V4** to USD currency.

g. Add the header "Average Revenue by Age Group" in cell **V1**. You should see these results:

h.	Average Revenue by Age Group
Youth	$3,533.00
Young Adults	$3,859.55
Adults	$4388.43

Step 6: Use the MINIFS and MAXIFS Functions

The MINIFS function in Excel returns the minimum value from a set of values specified by one or more criteria. The MAXIFS function returns the maximum value from a set of values specified by one or more criteria.

For this example, you will use the MINIFS function to find the order from Australia that generated the smallest amount of revenue.

a. In cell **W1**, type the header "Minimum Revenue from Australia" in bold and expand the column as needed to see the text.

b. Select cell W2 to enter the **MINIFS** function to find the smallest amount revenue from Australia.

c. Type the function **=MINIFS(S2:S89, I2:I89, "Australia")**

The result should be 565. Next, we'll use the MAXIFS function to find the order from Australia that generated the largest amount of revenue.

d. In cell **X1**, type the header "Maximum Revenue from Australia" in bold and expand the column as needed to see the text.

e. Select cell **X2** to enter the **MAXIFS** function to find the largest amount revenue from Australia.

f. Type the function **=MAXIFS(S2:S89, I2:I89, "Australia")**

The result should be 13500.

g. Change the data type in cells **W2** and **X2** to USD currency.

You should see these results:

Minimum Revenue from Australia	Maximum Revenue from Australia

$565.00	$13,500.00

What function could you use if you wanted to add up the revenue from bikes sales limited to just Australia? (Hint, this function adds cells values in a range based on a given criteria.) Use the "Insert Function" button under the Formulas tab on the toolbar to look through available functions and see if you can find it.

Answer // SUMIF function adds the values in a range that meet a specific condition.

    a.   Save the workbook as **Bike Sales_Functions_Lab 4.3.3_Part2**

Part 3: Experiment with Other Excel Functions

There are many more functions that are useful to data analysts for preparing, cleaning, and searching datasets. Explore the various function cateASgories in Excel and experiment by applying them to a dataset. Some specific functions to try are those that convert text case, such as LOWER, PROPER, and UPPER, and those that return information, such as LEN, FIND, and SEARCH.