Lecture Note #7: Some More Functions

BUSI 201: Business Data Analysis

Fall 2023

Topic 1. The `HLOOKUP` Function

Similar to the `VLOOKUP` function we covered in a previous class, the `HLOOKUP` function returns a certain value. While the `VLOOKUP` function looks *vertically* down the first column of the table array to find the unique identifier, the `HLOOKUP` function looks *horizontally* across the first row of the table array to find the identifier.

Figure 1: VLOOKUP

Figure 2: HLOOKUP

`HLOOKUP` is used when the unique ID is aligned in a single row, as shown in Figure 2, and `VLOOKUP` is appropriate when the data is arranged like Figure 1, where the unique ID is aligned in a single column. The syntax of `HLOOKUP` is similar to that of `VLOOKUP`.

```
= HLOOKUP(LOOKUP VALUE,
          RANGE WHERE LOOKUP VALUE IF LOCATED,
          ROW NUMBER IN THE RANGE CONTAINING VALUE TO RETURN,
          [EXACT/APPROXIMATE MATCH])
```

Navigate to the worksheet HLOOKUP_DATA in the workbook BUSI201-LEC07-Workbook.xlsx. You should find a table with information on some real estate properties on the market in various regions across the U.S. Suppose you want to extract information on the county the property is located in, the number of bedrooms and bathrooms, and the asking price of four properties: RE008, RE002, RE010, and RE001. Fill out the table located in cells B20 : F24.

Figure 3: The HLOOKUP_DATA Sheet

For instance, to find the number of bedrooms for property code RE002 in cell D22, we can use the HLOOKUP function as follows:

$$= \text{HLOOKUP}\,(\$J2, \$C\$2:\$L\$17, 5, 0)$$

Topic 2. The `XLOOKUP` Function

We have covered two distinct types of lookup functions: the `VLOOKUP` for data that are stacked *vertically*, and the `HLOOKUP` for data that are stacked *horizontally*. It is challenging to think of any other method of organizing data that is comprehensible, while neither being stacked vertically nor horizontally. So, you may wonder, what is the purpose of this `XLOOKUP` function?

The primary limitation of the `VLOOKUP` and `HLOOKUP` functions is that the lookup value (unique identifier) must be located in the first column/row, respectively. This restriction exists because `VLOOKUP` returns values to the *right* of the lookup value, and `HLOOKUP` returns values *below* the lookup value. For instance, refer to Figure 4, which represents a hypothetical data sheet from a hospital system, where individuals are uniquely identified by their patient ID number.

	Date	Full Name	Age	Patient ID	Procedure	Cost	Return	Network	Pharma		Patient ID	Full Name	Date	Cost
3	Aug/07/23	Emily Johnson	55	PID-001	Blood Pressure Measurement	$ 27.88	No	No	Yes		PID-006			
4	Aug/12/23	Benjamin Smith	55	PID-002	Blood Tests	$ 267.61	Yes	Yes	No		PID-002			
5	Aug/20/23	Ava Williams	17	PID-003	X-ray	$ 103.57	No	No	No		PID-021			
6	Aug/24/23	Liam Anderson	67	PID-004	Ultrasound	$ 377.08	No	Yes	No		PID-018			
7	Aug/28/23	Olivia Brown	52	PID-005	Electrocardiogram	$ 273.07	No	No	Yes					
8	Sep/02/23	Jackson Martinez	33	PID-006	Colonoscopy	$ 29.80	No	No	No					
9	Sep/06/23	Sophia Davis	42	PID-007	Endoscopy	$ 370.67	Yes	Yes	Yes					
10	Sep/10/23	Noah Wilson	68	PID-008	MRI	$ 13.96	No	No	Yes					
11	Sep/13/23	Emma Garcia	52	PID-009	CT Scan	$ 303.60	No	Yes	No					
12	Sep/15/23	William Jones	15	PID-010	Mammogram	$ 352.11	Yes	Yes	Yes					
13	Aug/04/23	Isabella Taylor	30	PID-011	Biopsy	$ 354.63	No	Yes	Yes					
14	Aug/17/23	James Clark	62	PID-012	Surgery	$ 276.26	Yes	No	No					
15	Aug/22/23	Mia Hernandez	25	PID-013	Laparoscopy	$ 98.18	No	Yes	Yes					
16	Aug/31/23	Jackson Martinez	33	PID-006	Catheterization	$ 207.10	Yes	No	Yes					
17	Sep/01/23	Charlotte White	50	PID-015	Dialysis	$ 51.72	No	Yes	No					
18	Sep/05/23	Ethan Lewis	45	PID-016	Ventilation	$ 224.30	No	Yes	No					
19	Sep/08/23	Harper Hall	44	PID-017	Chemotherapy	$ 397.83	Yes	No	Yes					
20	Sep/11/23	Elijah Turner	32	PID-018	Radiation Therapy	$ 147.96	Yes	Yes	Yes					
21	Sep/14/23	Amelia Scott	70	PID-019	Physical Therapy	$ 164.92	Yes	Yes	Yes					
22	Sep/16/23	Alexander Rodriguez	57	PID-020	Cardiac Bypass Surgery	$ 114.86	Yes	Yes	No					

Figure 4: The `XLOOKUP_DATA` Sheet

While the data is organized *vertically*, we may be tempted to use the `VLOOKUP` function. However, we can observe that the unique patient ID is located in column E, which is in the middle of the table. Therefore, if we want to use the unique patient ID to look up information about this patient, we can use the `XLOOKUP` function. The syntax of the `XLOOKUP` function is as follows:

```
= XLOOKUP(LOOKUP VALUE,
         RANGE WHERE THE LOOKUP VALUE IS LOCATED,
         RANGE WHERE THE RETURN VALUE IS LOCATED,
         [WHAT TO RETURN WHEN LOOKUP VALUE NOT FOUND],
         [EXACT/APPROXIMATE MATCH],
         [SEARCH DIRECTION])
```

3

The arguments are similar, but not identical to those of VLOOKUP. Let's go over the correct solution for cell M3, where we attempt to find the full name of PID-006:

$$= \texttt{XLOOKUP}(\texttt{\$L3},\texttt{\$E\$3:\$E\$22},\texttt{\$C\$3:\$C\$22},\texttt{"Not Found"},0,1)$$

Please see Figure 5 for the color-coded zones chosen as arguments in the XLOOKUP function.

Figure 5: An Example of XLOOKUP

The Three "Optional" Arguments

There are three optional arguments in the XLOOKUP function, and they function as follows:

- 4th Argument:
  [WHAT TO RETURN WHEN LOOKUP VALUE IS NOT FOUND]
  Value to return when XLOOKUP cannot find the lookup value in the data.

- 5th Argument:
  [EXACT/APPROXIMATE MATCH]
  Identical to VLOOKUP. Typically, we use only the exact match, so we usually stick to "0."

- 6th Argument:
  [SEARCH DIRECTION]
  The default (1) is a "Top-to-Bottom" search. However, we can also use "-1" if we want Excel to search the data "Bottom-to-Top."

Topic 3. The `AND` / `OR` Functions

Up to this point, we have covered functions that can either take a single condition, such as the `IF` or `VLOOKUP` functions, and functions that can evaluate multiple conditions, like the `SUMIFS` function. For functions that only accept a single condition argument, we can use the `AND` and/or `OR` functions to assess multiple conditions in a single line.

Figure 6: The `AND_OR` Sheet

If we want to populate cell `M3` with an indicator to check if the order ID `ORD202309100001` generated revenue greater than \$500 and if the profit margin for filling that order was greater than 5%, we can use a simple `IF` function:

$$= \text{IF(I3>500,IF(K3>0.05,1,0),0)}$$

Meanwhile, if we aim to fill cell `N3`:

$$= \text{IF(I3>500,1,IF(K3>0.05,1,0))}$$

When checking between two conditions, this is relatively straightforward. However, we may encounter situations where we need to verify if an entry satisfies dozens of conditions simultaneously or if it satisfies at least one of a dozen conditions. In such scenarios, we can rely on the `AND` or `OR` functions to streamline this process.

The AND **Function**

Consider the worksheet AND_OR in the workbook BUSI201-LEC07-Workbook.xlsx. You will find some sales data, but for this exercise, we will focus on the revenue (column I) and the profit margin (column K). Suppose we want to create a column that acts as an indicator. For example, column M should display:

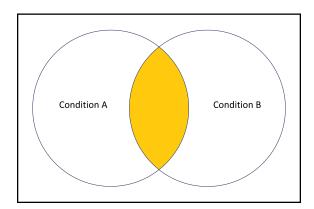$$M3 = \begin{cases} 1, & \text{if } \texttt{Revenue} > 500 \text{ and } \texttt{Profit Margin} > 5\% \\ 0, & \text{otherwise} \end{cases}$$

Figure 7: AND Function Diagram

To achieve this goal, we can use the AND function, which returns TRUE if all the conditions provided as arguments are met simultaneously. In other words, the AND function will return TRUE only if all logic tests pass, and it will return FALSE if any of the conditions are not met.

Visualized in a Venn diagram as shown in Figure 7, the AND function will yield TRUE for objects in the shaded zone, representing the intersection of two sets.

The basic syntax of the AND function is as follows:

$$= \texttt{AND} \left(\texttt{LOGIC\_TEST\_1}, \left[\texttt{LOGIC\_TEST\_2} \right], \cdots \right)$$

For cell M3, we can use the AND function as follows:

$$= \texttt{AND} \left(\texttt{I3} > 500, \texttt{K3} > 0.05 \right)$$

Please note that there is no need to use absolute/mixed references in this case, and percentages are represented in the numerical form of 0.05, not as 5%. Try filling the remaining cells in column M using the AND function and column O with the corresponding conditions.
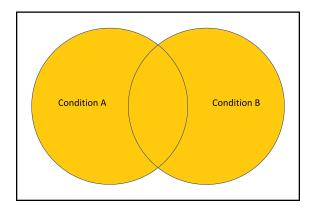
The OR Function

Figure 8: OR Function Diagram

Another useful function to employ when checking multiple conditions simultaneously is the OR function. In terms of set theory, the OR function is the *union* equivalent of the AND function, as shown in Figure 8.

In other words, it returns TRUE if *any* of the conditions are met and will return FALSE only if *none* of the conditions are met. We can use the OR function to fill out the cells in columns N and P.

Since the syntax of the OR function is identical to that of the AND function, cell N3 should read:

$$= \mathtt{OR}\left(\mathtt{I3} > 500, \mathtt{K3} > 0.05\right)$$

The AND and OR Functions with the IF Function

We can also combine the AND and OR functions in conjunction with the IF function to output 0 for FALSE and 1 for TRUE. The syntax will be:

$$= \mathtt{IF}\left(\mathtt{AND}\left(\mathtt{LOGIC\_TEST\_1}, \left[\mathtt{LOGIC\_TEST\_2}\right], \cdots\right), 1, 0\right)$$

Topic 4. Helper Columns

In many scenarios, especially when dealing with multiple conditions, it's useful to employ what is known as a *helper column*. Please navigate to the HELPER sheet, which is essentially a copy of the AND_OR sheet with two new columns. We will populate columns M and N with information that can assist us.

Figure 9: The HELPER Sheet

In cells M3 and N3 of the HELPER sheet, we can utilize the IF function to create "indicator" variables indicating whether the revenue generated from the order exceeds \$500 and whether the profit margin was greater than 5%. You can use the following formulas to achieve the result shown in Figure 10:

$$\texttt{M3:} = \texttt{IF(I3>500,1,0)}$$

$$\texttt{N3:} = \texttt{IF(J3>0.05,1,0)}$$

Figure 10: Helper Columns Filled Out

Now that the helper columns are ready, filling out the original table becomes more straightforward. For example, consider cell P3, where we want to check if the order generates more than $500 in revenue and also has a profit margin greater than 5%:

$$= \texttt{IF(M3+N3=2,1,0)}$$

The condition M3+N3=2 is met only when both the revenue and profit conditions are satisfied. Therefore, it's equivalent to using the AND function or writing out the slightly longer nested IF function.