

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.

Ticker	Asset	Asset Type	Quantity	Purchase Price	Current Price
AAPL	Apple Inc.	Stocks	100	\$15,000.00	\$17,500.00
MSFT	Microsoft	Stocks	75	\$20,000.00	\$22,000.00
AMZN	Amazon.com	Stocks	50	\$320,000.00	\$350,000.00
GOOGL	Google	Stocks	60	\$250,000.00	\$270,000.00
IBM	IBM	Stocks	40	\$13,000.00	\$12,000.00
TSLA	Tesla Inc.	Stocks	30	\$65,000.00	\$70,000.00
JNJ	Johnson & Johnson	Stocks	50	\$15,000.00	\$16,000.00
PG	Procter & Gamble	Stocks	60	\$13,000.00	\$14,000.00
XOM	Exxon Mobil	Stocks	70	\$6,000.00	\$6,500.00
NFLX	Netflix Inc.	Stocks	40	\$45,000.00	\$48,000.00
SPY	SPDR S&P 500 ETF	ETFs	25	\$40,000.00	\$42,000.00
XLK	Shares Tech ETF	ETFs	30	\$15,000.00	\$16,000.00
BND	Vanguard Bond ETF	ETFs	100	\$8,000.00	\$8,500.00
EEM	Shares Emerging Markets ETF	ETFs	35	\$4,500.00	\$4,800.00
USGOV	US Treasury	Bonds	500	\$100,000.00	\$102,000.00
CORP	Corporate Bond	Bonds	300	\$12,000.00	\$12,200.00
MUNI	Municipal Bond	Bonds	200	\$9,000.00	\$9,200.00
HYLD	High-Yield Bond	Bonds	150	\$7,500.00	\$7,800.00
TIPS	Treasury Inflation-Protected Securities	Bonds	75	\$11,000.00	\$11,200.00
VNQ	Real Estate ETF	Real Estate	15	\$10,000.00	\$10,500.00

Figure 1: VLOOKUP

Property Code	RE001	RE002	RE003	RE004
State	CA	TX	NY	FL
County	Los Angeles	Harris	Queens	Miami-Dac
City	Los Angeles	Houston	New York	Miami
Square Footage	2000	1800	2200	1600
Bedrooms	3	4	5	2
Bathrooms	2.5	3	4	2
Garage Spaces	2	2	3	1
Year of Construction	1990	1985	2005	1980
Renovation Year	2015	2020	2018	2019
Days on the Market	30	45	60	15
Asking Price	\$550,000	\$400,000	\$750,000	\$300,000
Flooring	Hardwood	Carpet	Hardwood	Tile
Heating	Central	Radiant	Forced Air	Central
Cooling	Central	Central	Central	Window
HOA	\$300	\$250	\$400	\$150

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 that you want to pull out 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.

	B	C	D	E	F	G	H	I	J	K	L	M
2	Property Code	RE001	RE002	RE003	RE004	RE005	RE006	RE007	RE008	RE009	RE010	
3	State	CA	TX	NY	FL	IL	AZ	WA	CO	NC	GA	
4	County	Los Angeles	Harris	Queens	Miami-Dade	Cook	Micropia	King	Denver	Wake	Fulton	
5	City	Los Angeles	Houston	New York	Miami	Chicago	Phoenix	Seattle	Denver	Raleigh	Atlanta	
6	Square Footage	2000	1800	2200	1600	2500	1900	2800	2100	2300	1950	
7	Bedrooms	3	4	5	2	4	3	4	3	4	3	
8	Bathrooms	2.5	3	4	2	3.5	2.5	3.5	2.5	3	2.5	
9	Garage Spaces	2	2	3	1	2	2	3	2	2	1	
10	Year of Construction	1990	1985	2005	1980	1995	2002	2010	1988	1998	2008	
11	Renovation Year	2015	2020	2018	2019	2016	2021	2017	2020	2015	2019	
12	Days on the Market	30	45	60	15	75	40	90	55	70	50	
13	Asking Price	\$550,000	\$400,000	\$750,000	\$300,000	\$600,000	\$350,000	\$800,000	\$450,000	\$475,000	\$380,000	
14	Flooring	Hardwood	Carpet	Hardwood	Tile	Hardwood	Carpet	Hardwood	Tile	Carpet	Hardwood	
15	Heating	Central	Radiant	Forced Air	Central	Radiant	Forced Air	Radiant	Forced Air	Radiant	Forced Air	
16	Cooling	Central	Central	Central	Window	Central	Central	Central	Central	Central	Central	
17	HOA	\$300	\$250	\$400	\$150	\$300	\$200	\$500	\$250	\$300	\$200	
18												
19												
20	Property Code	County	Bedrooms	Bathrooms	Asking Price							
21	RE008											
22	RE002											
23	RE010											
24	RE001											

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 covered two distinct types of lookup functions, the VLOOKUP for data that are stacked *vertically*, and the HLOOKUP for data that is stacked *horizontally*. It is difficult to think of any other method of organizing data that is both comprehensible, while neither being stacked vertically nor horizontally. Then you may ask what is the purpose of this XLOOKUP function?

The one weakness of the VLOOKUP and HLOOKUP functions are that the lookup value (unique indicator) must be located on the first column / row, respectively. This is because VLOOKUP will return values to the *right* of the lookup value, and HLOOKUP will return values *below* the lookup value. For instance, see Figure 4, which is a hypothetical data sheet from some hospital system, where individuals are uniquely identified by their patient ID number.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Date	Full Name	Age	Patient ID	Procedure	Cost	Return	Network	Pharma	Patient ID	Full Name	Date	Cost		
Aug/07/23	Emily Johnson	55	PID-001	Blood Pressure Measurement	\$ 27.88	No	No	Yes	PID-006					
Aug/12/23	Benjamin Smith	55	PID-002	Blood Tests	\$ 267.61	Yes	Yes	No	PID-002					
Aug/20/23	Ava Williams	17	PID-003	X-ray	\$ 103.57	No	No	No	PID-021					
Aug/24/23	Liam Anderson	67	PID-004	Ultrasound	\$ 377.08	No	Yes	No	PID-018					
Aug/28/23	Olivia Brown	52	PID-005	Electrocardiogram	\$ 273.07	No	No	Yes						
Sep/02/23	Jackson Martinez	33	PID-006	Colonoscopy	\$ 29.80	No	No	No						
Sep/06/23	Sophia Davis	42	PID-007	Endoscopy	\$ 370.67	Yes	Yes	Yes						
Sep/10/23	Noah Wilson	68	PID-008	MRI	\$ 13.96	No	No	Yes						
Sep/13/23	Emma Garcia	52	PID-009	CT Scan	\$ 303.60	No	Yes	No						
Sep/15/23	William Jones	15	PID-010	Mammogram	\$ 352.11	Yes	Yes	Yes						
Aug/04/23	Isabella Taylor	30	PID-011	Biopsy	\$ 354.63	No	Yes	Yes						
Aug/17/23	James Clark	62	PID-012	Surgery	\$ 276.26	Yes	No	No						
Aug/22/23	Mia Hernandez	25	PID-013	Laparoscopy	\$ 98.18	No	Yes	Yes						
Aug/31/23	Jackson Martinez	33	PID-006	Catheterization	\$ 207.10	Yes	No	Yes						
Sep/01/23	Charlotte White	50	PID-015	Dialysis	\$ 51.72	No	Yes	No						
Sep/05/23	Ethan Lewis	45	PID-016	Ventilation	\$ 224.30	No	Yes	No						
Sep/08/23	Harper Hall	44	PID-017	Chemotherapy	\$ 397.83	Yes	No	Yes						
Sep/11/23	Elijah Turner	32	PID-018	Radiation Therapy	\$ 147.96	Yes	Yes	Yes						
Sep/14/23	Amelia Scott	70	PID-019	Physical Therapy	\$ 164.92	Yes	Yes	Yes						
Sep/16/23	Alexander Rodriguez	57	PID-020	Cardiac Bypass Surgery	\$ 114.86	Yes	Yes	No						

Figure 4: The XLOOKUP_DATA Sheet

As the data is organized *vertically*, we may be tempted to use the VLOOKUP function. However, we can see 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] )
```

The arguments are similar, but not identical to that of VLOOKUP. We will go over a correct solution for cell M3, where we attempt to find the full name of PID-006:

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

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

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	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	Sep/04/23	Isabella Taylor	30	PID-011	Biopsy	\$ 354.63	No	Yes	Yes					
14	Sep/17/23	James Clark	62	PID-012	Surgery	\$ 276.26	Yes	No	No					
15	Sep/22/23	Mia Hernandez	25	PID-013	Laparoscopy	\$ 98.18	No	Yes	Yes					
16	Sep/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 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 NOT FOUND]
Value to return when XLOOKUP cannot find the lookup value in the data.
- 5th Argument:
[EXACT/APPROXIMATE MATCH]
Identical to VLOOKUP. We typically only use the exact match, so we usually stick to “0.”
- 6th Argument:
[SEARCH DIRECTION]
The default (or 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 covered functions that could either take a single condition such as the IF or VLOOKUP functions, and functions that can look at multiple conditions such as the SUMIFS function. For those functions that can only take a single line as the condition arguments, we can use the AND and/or OR functions to examine multiple conditions in one line.

The screenshot shows a Microsoft Excel spreadsheet with the following structure:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Order ID	Date	Customer	Sales Rep	Category	Item ID	Quantity	Revenue	Profit	Margin							
2	ORD202309100001	9/10/2023	John Smith	Alice Johnson	Electronics	E123	14	\$ 742.16	\$ 36.37	4.90%	Revenue > 500 AND Profit Margin > 5%	Revenue > 500 OR Profit Margin > 5%	Revenue < 500 AND Profit Margin < 5%	Revenue < 500 OR Profit Margin < 5%			
3	ORD202309100002	9/10/2023	Emily Davis	Michael Wilson	Clothing	C456	10	\$ 317.25	\$ 26.65	8.40%							
4	ORD202309100003	9/10/2023	Robert Brown	Jessica Miller	Home Appliances	HA789	1	\$ 111.46	\$ 3.90	3.50%							
5	ORD202309100004	9/9/2023	Sarah Johnson	David Smith	Electronics	E123	8	\$ 232.96	\$ 19.57	8.40%							
6	ORD202309100005	9/8/2023	James Wilson	Jennifer White	Clothing	C456	7	\$ 50.00	\$ 24.00	4.80%							
7	ORD202309100006	9/8/2023	Linda Johnson	Michael Wilson	Home Appliances	HA789	13	\$ 94.57	\$ 3.01	3.20%							
8	ORD202309100007	9/8/2023	William Brown	Alice Johnson	Electronics	E123	9	\$ 773.49	\$ 57.24	7.40%							
9	ORD202309100008	9/7/2023	Susan Johnson	Jessica Miller	Clothing	C456	8	\$ 83.67	\$ 6.86	8.20%							
10	ORD202309100009	9/7/2023	Michael Smith	David Smith	Home Appliances	HA789	10	\$ 351.30	\$ 18.62	5.30%							
11	ORD202309100010	9/6/2023	Olivia White	Jennifer White	Electronics	E123	5	\$ 685.80	\$ 25.37	3.70%							
12	ORD202309100011	9/6/2023	Henry Davis	Alice Johnson	Clothing	C456	5	\$ 561.59	\$ 55.60	9.90%							
13	ORD202309100012	9/5/2023	Emily Wilson	Michael Wilson	Home Appliances	HA789	11	\$ 714.00	\$ 26.42	3.70%							
14	ORD202309100013	9/5/2023	William Wilson	Jessica Miller	Electronics	E123	8	\$ 654.00	\$ 27.47	4.20%							
15	ORD202309100014	9/4/2023	Susan Davis	David Smith	Clothing	C456	10	\$ 971.59	\$ 43.72	4.50%							
16	ORD202309100015	9/4/2023	James Brown	Jennifer White	Home Appliances	HA789	8	\$ 724.91	\$ 7.97	1.10%							
17	ORD202309100016	9/3/2023	Linda Johnson	Alice Johnson	Electronics	E123	6	\$ 189.02	\$ 2.84	1.50%							
18	ORD202309100017	9/3/2023	Robert Smith	Michael Wilson	Clothing	C456	4	\$ 957.49	\$ 77.56	8.10%							
19	ORD202309100018	9/2/2023	Sarah Davis	Jessica Miller	Home Appliances	HA789	12	\$ 130.52	\$ 10.31	7.90%							
20	ORD202309100019	9/2/2023	Olivia Brown	David Smith	Electronics	E123	12	\$ 877.39	\$ 21.06	2.40%							
21	ORD202309100020	9/1/2023	Michael Wilson	Jennifer White	Clothing	C456	7	\$ 475.29	\$ 15.21	3.20%							
22	ORD202309100021																
23																	
24																	
25																	
26																	
27																	
28																	
29																	
30																	
31																	
32																	
33																	
34																	

Below the table, there is a navigation bar with tabs: VLOOKUP_DATA, HLOOKUP_DATA, XLOOKUP_DATA, AND_OR (which is selected), and HELPER.

Figure 6: The AND_OR Sheet

If we are to fill out the cell M3 with an indicator if the order ID ORD202309100001 generated a revenue greater than \$500, and if the profit margin for filling that order was greater than 5% with a simple IF function:

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

Meanwhile, if we are to fill cell N3:

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

When we are checking between two conditions, this is rather simple. However, we may run into situations where we are asked to check whether some entry satisfies dozens of conditions at the same time, or if it satisfies one of a dozen conditions. In these situations, we can rely on the AND or OR functions to make this process easier.

The AND Function

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

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

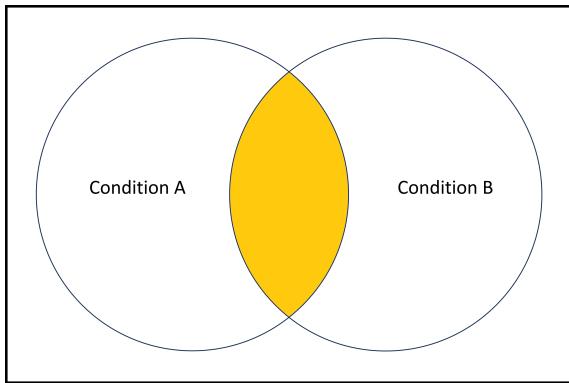


Figure 7: AND Function Diagram

To achieve this goal, we can use the AND function, which displays TRUE if the conditions we input as arguments are *jointly met*. That is, the AND function will return TRUE if *all* logic tests are passed, and will return FALSE if *any* of the conditions are not met.

Plotted in a Venn diagram shown in Figure 7, the AND function will display TRUE for objects in the shaded zone, equivalent to the intersection of two sets.

The basic syntax of the AND function is as follows:

$$= \text{AND} (\text{LOGIC_TEST_1}, [\text{LOGIC_TEST_2}], \dots)$$

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

$$= \text{AND} (I3 > 500, K3 > 0.05)$$

Note that we do not need to use absolute / mixed references in this case, and that percentages are represented in the numerical form of 0.05, not 5. Please try out filling the remainder of column M using the AND function, and column O with the corresponding conditions.

The OR Function

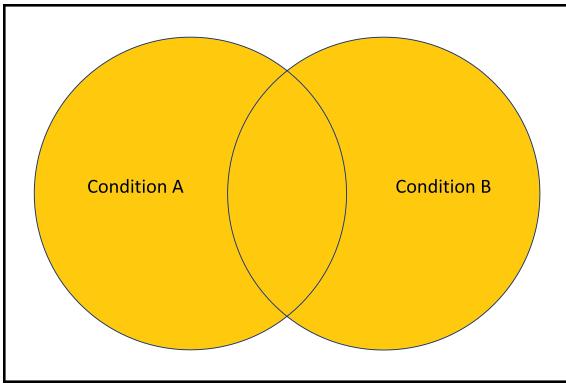


Figure 8: OR Function Diagram

Another useful function to use when trying to check multiple conditions at the same time is the OR function. In terms of talking about sets, the OR function is the *union* equivalent of the AND function as shown in Figure 8.

That is, 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:

$$= \text{OR} (\text{I3} > 500, \text{K3} > 0.05)$$

The AND and OR Functions with the IF Function

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

$$= \text{IF} (\text{AND} (\text{LOGIC_TEST_1}, [\text{LOGIC_TEST_2}], \dots), 1, 0)$$

Topic 4. Helper Columns

In many scenarios, especially when dealing with multiple conditions, is to use what is known as a *helper column*. Please navigate to the **HELPER** sheet, which is basically the **AND_OR** sheet copied with two new columns. We will fill out the columns M and N with information that can *help* us out.

The screenshot shows the 'HELPER' sheet in Excel. It contains the following columns:

- Original Data Columns (A-S):** Order ID, Date, Customer, Sales Rep, Category, Item ID, Quantity, Revenue, Profit, Margin.
- Newly Added Columns (M-S):**
 - M: Revenue > \$500
 - N: Profit Margin > 5%
 - Q: Revenue > 500 AND Profit Margin > 5%
 - R: Revenue > 500 OR Profit Margin > 5%
 - S: Revenue < 500 AND Profit Margin < 5%
 - T: Revenue < 500 OR Profit Margin < 5%

The data in columns M and N is filled with values based on the conditions defined in the formulas above them. The formulas are:

```

M3 := IF(I3>500,1,0)
N3 := IF(J3>0.05,1,0)

```

Figure 9: The **HELPER** Sheet

In cells M3 and N3 of the **HELPER** sheet, we can use the IF function to create “indicator” variables of whether the revenue generated from the order exceeds 500, and whether the profit margin was greater than 5%. We can use the following to get to Figure 10:

$$M3 := IF(I3>500,1,0)$$

$$N3 := IF(J3>0.05,1,0)$$

The screenshot shows the 'HELPER' sheet in Excel with the data filled out. The columns M and N now contain binary values (0 or 1) indicating whether each row meets the specified conditions. The formulas used are:

```

M3 := IF(I3>500,1,0)
N3 := IF(J3>0.05,1,0)

```

The data in columns M and N is filled with values based on the conditions defined in the formulas above them. The formulas are:

```

M3 := IF(I3>500,1,0)
N3 := IF(J3>0.05,1,0)

```

Figure 10: Helper Columns Filled Out

Now with the helper columns ready, we can fill out the original table a bit more easily. For instance, take 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%:

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

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