Lecture Note #4: Functions

BUSI 201: Business Data Analysis

Fall 2023

Topic 1. The Basic Structure of Excel Functions

Excel allows its users to automate complex calculations and tasks using various functions. Each function serves a unique purpose and has its own syntax, which we must learn to fully utilize Excel as a tool. However, the basic structure remains consistent across different types of functions:

$$= \texttt{FUNCTION}\,(\texttt{ARGUMENT\ \#1}, \texttt{ARGUMENT\ \#2}, [\texttt{ARGUMENT\ \#3}], \cdots, \texttt{ARGUMENT\ \#N})$$

Functions always begin with an = sign, followed by the function name and the arguments in parentheses. Some arguments are strictly required, while others are optional. Optional arguments are typically enclosed in square brackets. As shown in Figure 1, this matches the syntax Excel uses to report which arguments are required, and which are optional.

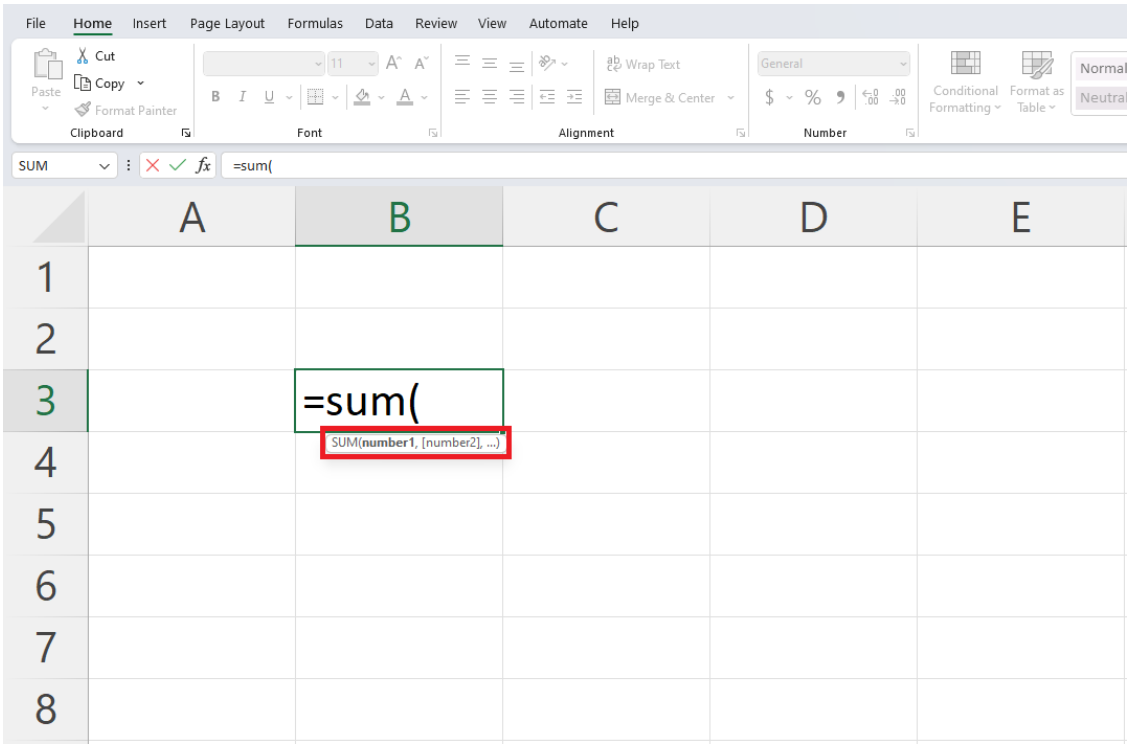


Figure 1: Function Inputs

Topic 2. How To

If you already have some working knowledge of the function you intend to use, you can navigate to the cell of choice and simply start typing in the function and arguments. If the cell you chose is empty, you may start by typing the = key. However, if the cell you chose is already populated with another function (so, if you are editing a pre-existing function), you may navigate to the cell, and press F2 to start editing the formula. To reference certain cells to use as arguments for the function, you may either type in the address of the cells (e.g. `A1:C5`), or use your mouse to select cells directly.

If you are not sure which function you wish to use, or you would like some more help from Excel, you may rely on the "Insert Function" feature located next to the formula bar. For windows users, you may also use the hotkey `shift`+F3 to access the insert function feature.

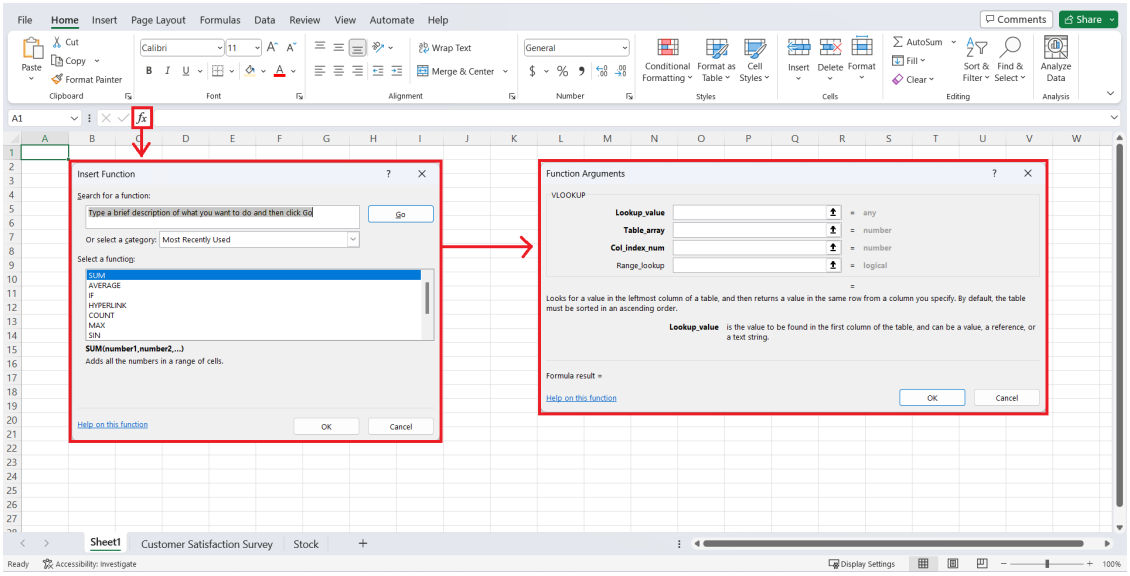


Figure 2: Insert Function Feature

As you can see in Figure 2, once you open the insert function window, you may search for a specific function, and call up a more detailed guide on the arguments of said function.

Topic 3. Functions: Calculating and "Locating"

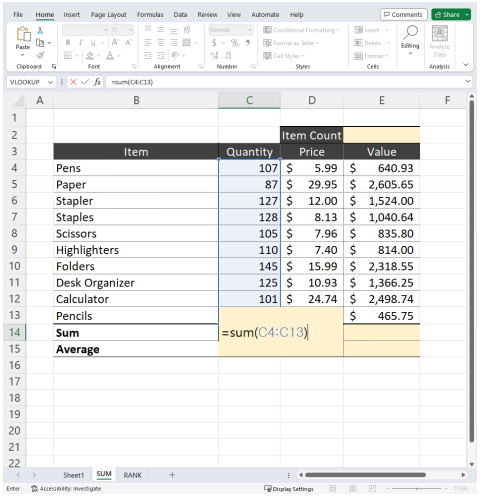
Some of the most basic, yet frequently used functions are the `SUM`, `AVERAGE`, and `COUNTA` functions. Please open the `SUM` sheet of the file `BUSI201-LEC04-Workbook.xlsx` which contains a table that keeps track of a hypothetical company's stockpile of office supplies.

The `SUM` Function

The sum function returns the sum of all cells included in the argument. So essentially, it is the `+` operator made simple. The syntax of the `SUM` function would be:

```
= SUM(ADDRESS OF CELLS CONTAINING THE VALUES YOU WOULD LIKE TO ADD)
```

Let us go through an example to make it clear. The cells `C14` and `E14` from this worksheet is supposed to track the sum of the quantity of goods, and the sum of the value of the goods in storage, so we may apply the `SUM` function.



	Item	Quantity	Price	Value
			Item Count	
4	Pens	107	$ 5.99	$ 640.93
5	Paper	87	$ 29.95	$ 2,605.65
6	Stapler	127	$ 12.00	$ 1,524.00
7	Staples	128	$ 8.13	$ 1,040.64
8	Scissors	105	$ 7.96	$ 835.80
9	Highlighters	110	$ 7.40	$ 814.00
10	Folders	145	$ 15.99	$ 2,318.55
11	Desk Organizer	125	$ 10.93	$ 1,366.25
12	Calculator	101	$ 24.74	$ 2,498.74
13	Pencils			$ 465.75
14	**Sum**	=sum(C4:C13)		
15	**Average**			

Figure 3: `SUM`

The correct formula to find the total number of individual items in storage, which is the intended content of cell `C14` is:

```
= SUM(C4:C13)
```

You could get the same result by manually adding all cells using the operator `+` as well:

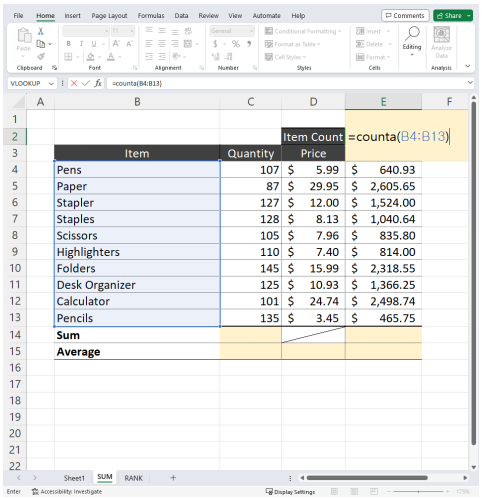
```
= C4 + C5 + C6 + C7 + C8 + C9 + C10 + C11 + C12 + C13
```

Hopefully we can all agree that using the `SUM` function, even when there are only 10 entries, is by far the superior way to perform addition.

As the function performs an addition, only numerical information is allowed. Any values that are not numerical (including empty cells) will be ignored in calculating the sum. While this is probably not a limiting factor for most, the maximum number of arguments that you can input in the `SUM` function is 255. Please use the `SUM` function to fill out cell `E14` to display the sum of the value of all office supplies in storage.

The `COUNTA` Function

The `COUNTA` function will count the number of non-empty cells in the given range dictated by the arguments of the function. The syntax of the `COUNTA` function is identical to the `SUM` function.



	A	B	C	D	E	F
1						
2				Item Count	=counta(B4:B13)	
3		Item	Quantity	Price		
4		Pens	107	$ 5.99	$ 640.93	
5		Paper	87	$ 29.95	$ 2,605.65	
6		Stapler	127	$ 12.00	$ 1,524.00	
7		Staples	128	$ 8.13	$ 1,040.64	
8		Scissors	105	$ 7.96	$ 835.80	
9		Highlighters	110	$ 7.40	$ 814.00	
10		Folders	145	$ 15.99	$ 2,318.55	
11		Desk Organizer	125	$ 10.93	$ 1,366.25	
12		Calculator	101	$ 24.74	$ 2,498.74	
13		Pencils	135	$ 3.45	$ 465.75	
14		**Sum**				
15		**Average**				
16						
17						
18						
19						
20						
21						
22						

Figure 4: `COUNTA`

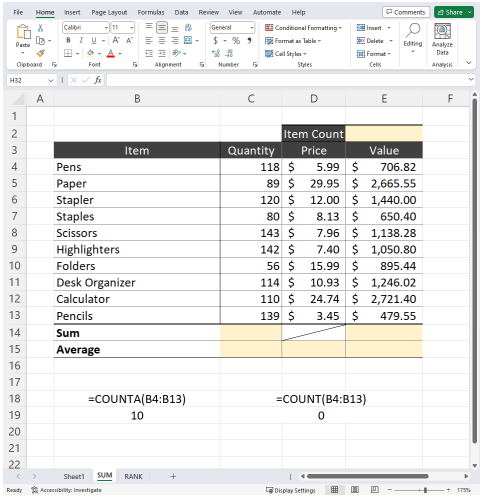
To report the number of unique items stored in storage for this firm, we can use the `COUNTA` function in cell `E2`. One possible forumla may be:

$$= \texttt{COUNTA(B4:B13)}$$

Note that it is stated that it is one possible formula. In fact, we can use any of the columns ranging from `B` to `E` to get the same answer in this specific case. This is possible as the other columns also have a value assigned to each row, matching that of the item description in column B. There is no one correct rule to use in real life, and the argument that you choose should depend on each specific case.

The `COUNT` Function

One commonly made mistake is using the `COUNT` function when you mean to use the `COUNTA` function. The `COUNT` function is a similar function, but only counts the number of cells in the range that contains numerical values.



	A	B	C	D	E	F
1						
2				Item Count		
3		Item	Quantity	Price	Value	
4		Pens	118	$ 5.99	$ 706.82	
5		Paper	89	$ 29.95	$ 2,665.55	
6		Stapler	120	$ 12.00	$ 1,440.00	
7		Staples	80	$ 8.13	$ 650.40	
8		Scissors	143	$ 7.96	$ 1,138.28	
9		Highlighters	142	$ 7.40	$ 1,050.80	
10		Folders	56	$ 15.99	$ 895.44	
11		Desk Organizer	114	$ 10.93	$ 1,246.02	
12		Calculator	110	$ 24.74	$ 2,721.40	
13		Pencils	139	$ 3.45	$ 479.55	
14		**Sum**				
15		**Average**				
16						
17						
18		=COUNTA(B4:B13)		=COUNT(B4:B13)		
19		10		0		
20						
21						
22						

Figure 5: `COUNT` vs `COUNTA`

Recall the status bar trick, where there were two separate types of "count" options. One was a "Numerical Count," and the other was simply "Count." In functions, for some reason, Excel decided to let `COUNT` correspond to the "Numerical Count," and `COUNTA` correspond to "Count."

The results depicted below the main table in figure 5 illustrates the difference in results when applying `COUNT` and `COUNTA` functions on cells with text data. Please try out applying the `COUNT` and `COUNTA` functions on the cell range `C4:C13` to verify that they will return the same result, as column `C` consists of numerical data.

The AVERAGE Function

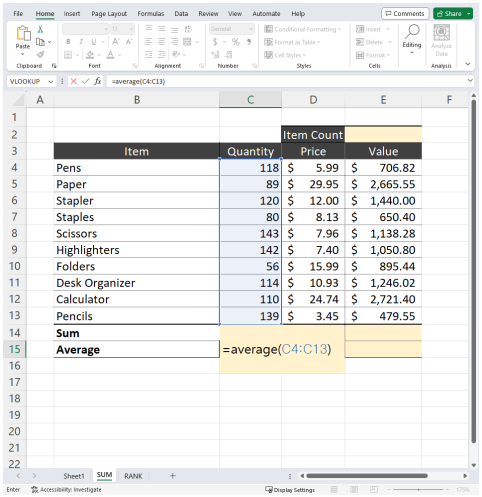


Figure 6: AVERAGE

The cells C15, D15, and E15 each aim to find the average value of the respective columns. The syntax is identical to the SUM, COUNTA, and COUNT functions. That is:

$$= \texttt{AVERAGE(C4:C13)}$$

Similar to the SUM function, all values related to this function must be numerical in nature. Any values that are not numerical (or empty) will be ignored in calculating the average value.

Please fill out the cells D15 and E15 and find the average unit price of each type of good in storage, and the average of each items' total worth, respectively.

Since the AVERAGE function returns the arithmetic mean of the selected values, the output from the AVERAGE function is identical to the result of taking the ratio of the SUM and COUNTA functions:

$$\texttt{AVERAGE(C4:C13)} = \frac{\texttt{SUM(C4:C13)}}{\texttt{COUNTA(C4:C13)}}$$

The AVERAGE, MEDIAN, and MODE Functions

The MEDIAN and MODE functions share the same syntax with the AVERAGE function, and return the sample median and mode respectively. The sample median is the middle value of the sample data, while the mode is the value that most often appears in the set of data.

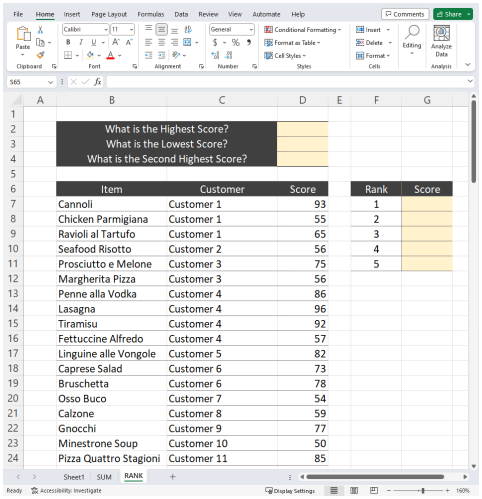
For instance, for a sequence of numbers $\{1, 2, 3, 5, 5\}$:

- AVERAGE = 3.2
- MEDIAN = 3
- MODE = 5

It is interesting to note that while the AVERAGE function "calculates" the value for us, the MEDIAN and MODE functions "locate" a certain value out of an array of numbers. Next we will be examining some more functions that "locate" certain values for the user.

The `MAX` and `MIN` Functions

Recall the other two variables that was readily available for the user in the status bar of Excel; the maximum and minimum values of an array of numbers. These are also available in the functions `MAX` and `MIN`, respectively. The syntax is identical to all previous examples, where the only argument(s) is the address of the cells that contain the numerical data you wish to evaluate. Please navigate to the `RANK` sheet of the workbook `BUSI201-LEC04-Workbook.xlsx`.



What is the Highest Score?		
What is the Lowest Score?		
What is the Second Highest Score?		

Item	Customer	Score	Rank	Score
Cannoli	Customer 1	93	1	
Chicken Parmigiana	Customer 1	55	2	
Ravioli al Tartufo	Customer 1	65	3	
Seafood Risotto	Customer 2	56	4	
Prosciutto e Melone	Customer 3	75	5	
Margherita Pizza	Customer 3	56		
Penne alla Vodka	Customer 4	86		
Lasagna	Customer 4	96		
Tiramisu	Customer 4	92		
Fettuccine Alfredo	Customer 4	57		
Linguine alle Vongole	Customer 5	82		
Caprese Salad	Customer 6	73		
Bruschetta	Customer 6	78		
Osso Buco	Customer 7	54		
Calzone	Customer 8	59		
Gnocchi	Customer 9	77		
Minestrone Soup	Customer 10	50		
Pizza Quattro Stagioni	Customer 11	85		

Figure 7: Restaurant Menu Ratings

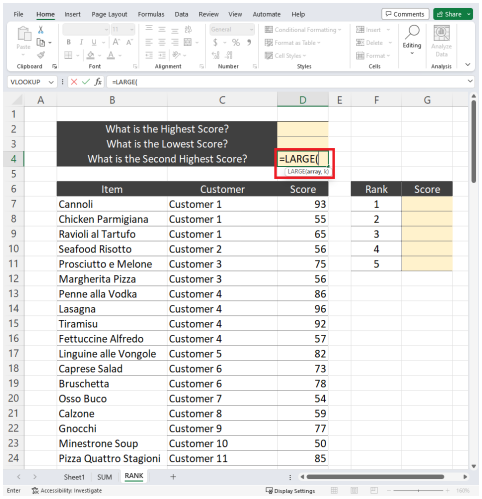
The worksheet consists of feedback from customers from an imaginary Italian restaurant. Suppose that we are primarily interested in what is the absolutely highest and lowest score any dish received. Those are the items to fill out in cells `D2` and `D3`, repsectively. Each value can be found by using the `MAX` and `MIN` functions:

$$= \texttt{MAX(D7:D106)}$$
$$= \texttt{MIN(D7:D106)}$$

Similar to the functions we covered up to this point, the `MAX` and `MIN` functions ignore any cell that include non-numerical inputs and empty inputs.

The `LARGE` and `SMALL` Functions

Even with the `MAX` and `MIN` functions, we run into a problem with the next empty cell of `D4`. How do we find the second highest score in the list? Here is where the `LARGE` function comes into play.



What is the Highest Score?		
What is the Lowest Score?		
What is the Second Highest Score?	=LARGE(	

Item	Customer	Score	Rank	Score
Cannoli	Customer 1	93	1	
Chicken Parmigiana	Customer 1	55	2	
Ravioli al Tartufo	Customer 1	65	3	
Seafood Risotto	Customer 2	56	4	
Prosciutto e Melone	Customer 3	75	5	
Margherita Pizza	Customer 3	56		
Penne alla Vodka	Customer 4	86		
Lasagna	Customer 4	96		
Tiramisu	Customer 4	92		
Fettuccine Alfredo	Customer 4	57		
Linguine alle Vongole	Customer 5	82		
Caprese Salad	Customer 6	73		
Bruschetta	Customer 6	78		
Osso Buco	Customer 7	54		
Calzone	Customer 8	59		
Gnocchi	Customer 9	77		
Minestrone Soup	Customer 10	50		
Pizza Quattro Stagioni	Customer 11	85		

Figure 8: `LARGE`

The `LARGE` function searches the array of numbers it is given, and looks for the `k`th largest value. Naturally, the `LARGE` function has two mandatory arguments; the address for the array of numbers, and the position of interest `k`. The function can be used in cell `D4` as follows:

$$= \texttt{LARGE(D7:D106,k)}$$

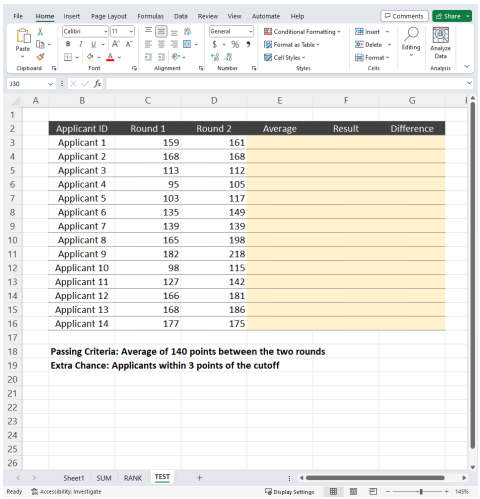
In words, this means "Find the `k`th largest numerical value in cells `D7:D106`." Meanwhile, the `SMALL` function does the exact opposite, and returns the `k`th smallest numerical value in the given array of numbers. Note that:

$$\texttt{LARGE(ARRAY,1)} = \texttt{MAX(ARRAY)}$$

Please fill out the top 5 review scores attained in this restaurant in the cells `G7:G11`.

Topic 4. Functions: Logical

One family of functions that sees regular use consists of functions that return values *conditionally* upon meeting certain thresholds or satisfying certain conditions.



Applicant ID	Round 1	Round 2	Average	Result	Difference
Applicant 1	159	161			
Applicant 2	168	168			
Applicant 3	113	112			
Applicant 4	95	105			
Applicant 5	103	117			
Applicant 6	135	149			
Applicant 7	139	139			
Applicant 8	165	198			
Applicant 9	182	218			
Applicant 10	98	115			
Applicant 11	127	142			
Applicant 12	166	181			
Applicant 13	168	186			
Applicant 14	177	175			

Passing Criteria: Average of 140 points between the two rounds
Extra Chance: Applicants within 3 points of the cutoff

Figure 9: `TEST`

Suppose that you are a manager in the human resource department of a firm, and you are involved in the hiring process. You have a list of applicants, and their standardized test results. A hypothetical example of this sheet can be found in the `TEST` sheet of the workbook `BUSI201-LEC04-Workbook.xlsx` as shown in Figure 9.

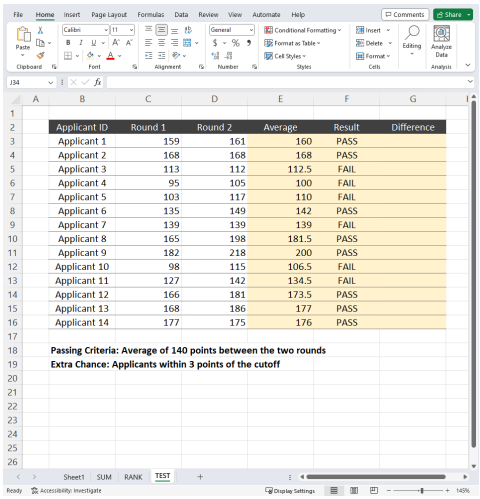
The first task is to find the applicants' average scores. We can use the `AVERAGE` function covered in the previous topic to calculate the average values in column `E`. But how do we make Excel tell us who passed and who failed in column `F`? We can use the `IF` function to sort applicants into the pass and fail categories.

The `IF` Function

The `IF` function has three arguments: two mandatory arguments and one optional argument.

$$= \texttt{IF}(\texttt{CONDITION}, \texttt{OUTPUT IF TRUE}, [\texttt{OUTPUT IF FALSE}])$$

If used correctly, the `IF` function will run a logic test to see if the `CONDITION` is met, then return `OUTPUT IF TRUE` if the condition is met, and return the `OUTPUT IF FALSE` when the condition is not met. If you choose to leave the optional `OUTPUT IF FALSE` blank, `FALSE` will be returned in its place.



Applicant ID	Round 1	Round 2	Average	Result	Difference
Applicant 1	159	161	160	PASS	
Applicant 2	168	168	168	PASS	
Applicant 3	113	112	112.5	FAIL	
Applicant 4	95	105	100	FAIL	
Applicant 5	103	117	110	FAIL	
Applicant 6	135	149	142	PASS	
Applicant 7	139	139	139	FAIL	
Applicant 8	165	198	181.5	PASS	
Applicant 9	182	218	200	PASS	
Applicant 10	98	115	106.5	FAIL	
Applicant 11	127	142	134.5	FAIL	
Applicant 12	166	181	173.5	PASS	
Applicant 13	168	186	177	PASS	
Applicant 14	177	175	176	PASS	

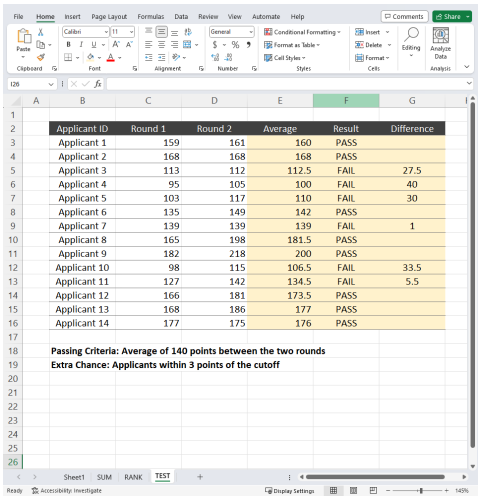
Passing Criteria: Average of 140 points between the two rounds
Extra Chance: Applicants within 3 points of the cutoff

Figure 10: `IF`

Let us return to the case at hand to see what we can do. We want to assign individuals a `PASS` if their average score is greater than or equal to the cutoff score of 140. Those who do not meet this threshold are assigned a `FAIL`. Translating this sentence into something that Excel will understand and filling cell `F3`:

$$= \texttt{IF(E3>=140,"PASS","FAIL")}$$

Copying and pasting cell `F3` to the other cells in the column, we can assign `PASS` and `FAIL` according to the same rules as shown in Figure 10. Now we should turn our attention to the second rule that applicants that fail, but are within 3 points of the cutoff will be given an extra chance.



Applicant ID	Round 1	Round 2	Average	Result	Difference
Applicant 1	159	161	160	PASS	
Applicant 2	168	168	168	PASS	
Applicant 3	113	112	112.5	FAIL	27.5
Applicant 4	95	105	100	FAIL	40
Applicant 5	103	117	110	FAIL	30
Applicant 6	135	149	142	PASS	
Applicant 7	139	139	139	FAIL	1
Applicant 8	165	198	181.5	PASS	
Applicant 9	182	218	200	PASS	
Applicant 10	98	115	106.5	FAIL	33.5
Applicant 11	127	142	134.5	FAIL	5.5
Applicant 12	166	181	173.5	PASS	
Applicant 13	168	186	177	PASS	
Applicant 14	177	175	176	PASS	

Passing Criteria: Average of 140 points between the two rounds
Extra Chance: Applicants within 3 points of the cutoff

Figure 11: `IF`

The function we want to write would be something like "For those who failed, calculate the difference between the cutoff and their average score, and for those who passed, don't show me anything." Translating this into Excel starting at cell `G3`:

```
=IF(F3="FAIL",E3-140,"")
```

Copying and pasting cell `G3` to the other cells in the column, we see that the differences for the failing candidates are calculated, while the passing applicants did not receive any value as shown in Figure 11. There are some other ways to handle this situation, such as using the `AND` function, or embedding `IF` functions in another `IF` function.