Transcript for Session 028

Listen to the podcast session, see resources & links:

http://chandoo.org/session28/

Transcript:

Hi there, welcome to http://chandoo.org/ podcast. This podcast is aimed to make you awesome in data analysis, charting, dashboards and VBA using Microsoft Excel. Thank you so much for joining me in another episode of http://chandoo.org/ podcast.

Today, we are going to talk about a very interesting and relevant topic for all of us. **How do you take real world business rules and business logic and transform it to an Excel understandable way?** This is a challenge that many of us face. We can understand how VLOOKUP works and how you can sum all the sales against a particular product in a particular region. These things are very straightforward. But, when it comes to taking a complex business rule or business logic statement and transforming it into an Excel understandable way, most of us would struggle.

When I started using Excel way back in 2003, I wasn't really doing a lot of analysis. Most of the time, I had an Excel workbook on my computer and I was using it to keep track of my progress of exam preparation. I was preparing for MBA entrance examinations back then and so I had a workbook where I would keep track of my progress with respect to mock examinations (dummy tests) before attempting the final examinations and where I was making mistakes etc. So, even though I started using Excel in 2003, I wasn't really doing anything related to business logic or business modeling until late 2005. Sometime during my MBA, I started using Excel and one of the very first things that we do when we start using Excel seriously is that we try to mimic real world problems in Excel and solve them. I was trying to implement a lot of business modeling problems in Excel and every time that I had to express business logic or a business rule, I would invariably fail. I wouldn't know how to express that in Excel. I could understand the arithmetic/math part but when it came to the logic part, I would be invariably tongue-tied. It was good that I had some programming background and so I could, more or less, intuitively understand the process required to take the business rule and transform it into a computer understandable way but the initial few steps were hard. Those leaps took me several weeks.

One thing that I often tell my students when I run a live class is about when I started using the IF formula for the very first time. It took me one full afternoon to solve a simple problem using the IF formula. I don't remember the exact problem but it went something like this - we had to calculate the grade point of students based on the marks they had scored. This was a homework problem given to us by our

professor. We had a bunch of marks scored by some students and we had to calculate the grade point for each of them. I knew that this would require using some sort of IF logic because if the marks are less than 40 then the grade point would be F, if the marks are between 40 to 60 then the grade point would be D and so on. If you had asked me to write a program for it, I would have happily written it because I know programming. But, I didn't know Excel very much at that time. So, I struggled a lot and even though I could finally write the IF formula, for some reason, Excel wouldn't accept the formula. It kept throwing an error at me and I spent one entire afternoon really wrestling with it and pulling my hair out to figure out where I was going wrong with it. Only after I realised the silly mistakes that I was making, I could make amends. This is what I mean about making a quantum leap from one place where I had no idea about the IF formula to the other area where I was comfortable expressing logic and business rules using IF formulas

Today, I want to share that kind of knowledge with you and explain the **key ingredients** or key characters when you want to take business rules and business logic and transform them into Excel understandable format. My motivation for this is not just the IF formula wrestling that I did almost a decade okay. Recently, yesterday I think, somebody commented on http://chandoo.org/ asking a simple question and that got me thinking about how many people struggle with these kinds of logic situations. The problem goes something like this - they wanted to find out whether a person is male or female based on the name of the person. Now, this is not possible in all cultures and all regions but, in certain cultures, just by looking at the name, you can very quickly tell if a person is male or female. This is because of certain cultural nuances that they follow. For example, the person from Myanmar who posted this question clearly said, "If the name starts with Muong, So, Min or Kun then the person is male. Likewise, if the name starts with Ma, Nan or Nee then the person is female." How would you write a formula to figure out the gender based on the name in a cell. You know the plain English rule that if the name is beginning with these certain versions then you can be sure that the person is male and if the name begins with the other three versions then the person is female. If neither is the case then we will be undecided and we won't know the gender and we'll need to either look at a photo of the person or maybe use another heuristic to figure out their gender. By the end of this podcast, you will be able to understand how to write a formula to figure out the gender of a person. This is what prompted me to record this podcast.

When it comes to expressing business logic in Excel, there are, in my opinion, **four main ingredients** that you need. The first one is **logical formulas**. The second ingredient is **support formu**las. The third ingredient is **informational formulas**. And, the fourth ingredient is **operators**. The good thing is that most of these, even though they sound like technical mumbo-jumbos, are really simple to understand. Let's delve into each of those four types of ingredients that we require and understand what they do.

The first ingredient is **logical formulas**. These are the logical formulas or operators in Excel that'll help us formulate a complex business condition or business rule in Excel. Whenever we talk about a business

condition or business rule, we tend to use the words **AND**, **OR**, and **NOT** quite often. For example, in the earlier case, where we were talking about figuring out the gender of the person from the name, we said that if the name begins with Muong, So, Min OR Kun then the person is male. So, we are using the word OR there. Likewise, sometimes, we might say that if the day of the week is Monday OR Tuesday then do this. We are again using OR there. Likewise, if we say that if the number of marks that the student scored are greater than 40 AND less than 60 then, here, we are using the word AND. Likewise, we can say that if the person's designation is NOT Manager then do this. So, we tend to use these words AND, OR and NOT many times when we are expressing business logic. These very words are also available in Excel as formulas. You could use the AND formula and parse on multiple conditions; each parameter of that formula has to be a condition or a logical value. It could be 'age greater than 40'. That's a condition and it will return true or false. Likewise, it could be 'designation=Manager'. This is another condition where you are checking the equality.

Likewise, you could do any of these things and when you pass on a bunch of parameters to the AND formula, it will return a value of true if all of the conditions are true. That means that every one of the conditions has to be met otherwise it will return false. This is how the AND function works. The OR function takes similar parameters like the AND function - you can parse on a bunch of conditions or logical or boolean values and it will return a value of true if any one of them is true. For example, you could say:
=OR(Day of week=Monday, Day of week=Tuesday)

So, if the day of the week happens to be either Monday or Tuesday then it will return true and, for every other value, it will return false. The OR function takes a bunch of input values and the output will be true if at least one of them is true. The NOT function is where you parse on a single value, condition, boolean or logical value and it will return the negation of it. So, if you pass on TRUE to the NOT function then it will return FALSE. Likewise, if you parse on FALSE to the NOT function then it will return TRUE.

Using a combination of these three, you could pretty much conjure any real world condition or business rule in Excel. To give you an example, we want to check the gender of a person from the name, so we would say something like assuming the first part of the name is X:
=OR(X="Muong",X="So", X="Min",X="Kun")

We are writing an OR function where we are checking if the first part of the name (designated by X) is equal to any one of those four. So, you would simply pass on those four as four different conditions and the OR would return TRUE if any one of them happens to be true. This is pretty much how you could express any business condition. Sometimes, you may have to mix and match OR and NOT in a single function. We will talk about that a little later in the podcast.

There is also one extra function that is available especially in Excel 2013 and beyond called **XOR**. XOR is basically exclusive OR. This is where you want to check for a business condition which would be TRUE only if one of the values is TRUE or one of the values is FALSE. A good way to check this would be

something like either this OR that. I want to return TRUE only in such a case. If both of them are true then we want it to be FALSE and if both of them are FALSE then we want it to be FALSE.

Here is one example of the XOR condition. Let's say that you are operating in a big company where you have branches all over the U.S. and in one year you had really stellar success and so you wanted to offer 10% bonus to everyone in the marketing department or everybody working in the New York office - you have offices all over the place but New York is the place where you had really high success and so you want to reward people working in the New York office irrespective of what their department is. In addition, you want to pay bonus to anyone in the marketing department anywhere in the country. But, you don't want to give 10% to the marketing department of New York; you want to give them something higher. Here is your exclusive OR condition; you are saying:
=XOR(Department="Marketing",Region="New York")

The condition will be true only if you are in the marketing department or if you are in New York. If you happen to be a person working in the marketing department in New York then it will be FALSE for you because we want to give you something higher than 10%; maybe we want to give 25% bonus to you. You can use XOR to check these conditions. It is newly introduced in Excel 2013. Of course, if you want to check similar conditions in prior versions of Excel, you don't have to worry whether it can be done or not. There are ways in which you can create an XOR function by creatively mixing AND, OR and NOT functions. I will leave a link to an article that explains how to check XOR conditions in all versions of Excel in the show notes of this podcast. This is the first bucket where we have understood the logical conditions required to express business logic. They are AND, OR, NOT and, in Excel 2013, XOR as well.

The second set of functions that will help us express business logic are **support functions**. These are formulas or function where we usually put the logical functions. A good example is the **IF** formula. Whenever we are checking for a logical condition, it is obvious that the output of that condition would determine whether to give the output for this part of the IF formula or that part of the IF formula. So, the IF formula is really like a branching path in our business conditions. For example, the gender name determination logic is an example where we would obviously use an IF formula like:
=IF(OR(X=".....",X=".....",.....),"Male","Female")

So, the **IF formula will help us drive the output part of the business rule**. The way to visualize this is that the business rule would contain some condition and what should be done if the condition is true or false. The IF formula takes care of the 'what to do' or the output part and the conditions are expressed by your AND, OR and NOT functions.

Likewise, there are also formulas such as **CHOOSE**, **IFERROR** and **IFNA**. We can use any of these to drive similar IF logic that is available through the IF formula. Since our scope for this podcast is limited to the discussion on logical functions, I am not going to talk about the IF, CHOOSE and other formulas but, in the show notes, I will provide a link to all those functions so that you can learn more about them.

The third group of functions that are helpful when it comes to these kinds of business rules are the **information functions**. Information functions or formulas help us understand certain information about a particular cell or data that is available. A good example is **ISBLANK**. For example, if we want to determine business logic based on whether a cell is blank, how would you check it? Of course, you would check if the cell is equal to 0 or "" but a better option would be to use the ISBLANK function. There are many such IS functions that check a logical condition for us automatically. There are functions **ISEVEN**, **ISOR**, **ISLOGICAL**, **ISNUMBER**, **ISTEXT** etc. All of these are self-explanatory. They check a cell or a range or a reference and tell us the critical information about them in a logical true or false way.

The last part of our business logic puzzle is **operators**. There are some operators that are helpful for us when it comes to expressing business logic. The very first set of operators is **brackets**. The brackets will be helpful when we are expressing complex business logic that requires nesting of a lot of AND, OR or NOT functions. **When you have to mix a couple of AND, OR or NOT functions, it is advisable to use these brackets so that Excel will calculate the conditions in the correct order and so that we get the correct output.** They are just like the brackets that we use in expressing arithmetic calculations or expressions. For example, if you want to take the sum of 2+3 and multiply it by 4 then if you simply write:

=2+3*4

then you will get 2 plus 3 times 4, i.e. you will get a result of 14. The result we were expecting was 5 times 4 which is 20. So, to prevent the incorrect result, we need to put 2+3 in brackets and 4 outside the bracket like this:

=(2+3)*4

Now, the addition of 2 and 3 will happen first and then the multiplication will happen. Likewise, when you are writing business logic conditions using AND, OR and NOT formulas, if there is a scenario where you must first evaluate one function and then go and evaluate another function then you should use these brackets in a careful way so that the priority or the order of calculations is imposed on Excel.

Apart from brackets, there are two other operators that are very critical and they are **+** and *****. If you really look at it very closely, the **AND operator is multiplying and the OR operator is adding**. Let me give you one example. We all know the boolean values of true and false correspond to 1 and 0 if you look at them purely from a number point of view. So, true is 1 and false is 0. If you have a bunch of true and false values and you run an AND function on it, the output of it would be true only if everything is true otherwise it will be false. The way to visualize this is as follows - let's say that you are checking the AND condition of 3 values one of which is true and 2 of which are false - so we have true, false and false - the output would obviously be false because there are two false values and hence the output will be false. If you look at the numeric representation, it is 1, 0 and 0. 1 multiplied by 0 and then multiplied by 0 again is 0 because anytime that you multiply by 0 the output is automatically 0. So, you can visualize * or multiplication as similar to AND. This is particularly useful when you are trying to do logical calculations on a bunch of arrays or ranges and you feel that the AND and OR functions are not working the way you

want. You could really do the arithmetic part of it, i.e. replace the AND with * and it will give a similar result.

Likewise, let us say you want to do an OR calculation. So, you have the same scenario of true, false and false which is 1, 0 and 0 and you want to calculate the OR of it. You could simply add them up like 1+0+0 and check whether the sum is greater than or equal to 1. Since it is 1, it is obviously true. One way to visualize the OR calculation is that it is simply adding and the AND calculation is multiplying. Those of who have had some sort of prior computer engineering or electronics background, you could really go back to the days when you were learning those things and you could see that - we have AND and OR calculations in computers provided by small computer transistors - so everything in a computer is either adding or multiplying. That's the origin of how computers became what they are today; they were able to calculate massive amounts of data because all they were really doing was adding or multiplying and also working with the NOT part. When you combine these three creatively, you can really do anything. Anyway, that's enough about the jargon and technical part of these things. **You could use plus for calculating OR and the * for calculating AND values if there is a situation where you can't use the logical functions or if you feel more comfortable with their arithmetic counterparts.**

Now that we understand, at an abstract level, what the ingredients for framing business logic are, i.e. logical functions, support functions, information functions and the operators - let's go over a couple of examples. The first example would obviously be the gender from the name formula. Assuming the first part of the name is already extracted into a variable or a name called X then the formula would go like this:
=IF(OR(X="Moung",X="So",X="Min",X="Kun"),"Gender is Male",IF(X="Ma",X="Nan",X="Nee"),"Gender is Female","Gender cannot be determined")

If the gender is not male then we can assume it is female but it would be prudent to check for female also because there could be some names that begin with neither male-specific nor female-specific values. If both conditions are not true, i.e. the name is not beginning with the four male-specific names nor the three female-specific names then we would simply print that the gender cannot be determined because the logic is only for the 7 types of first names but there could be other names that are out there. For example, if I am on the list then since my name starts with Purna and it is not part of any of the seven options then if we use default logic that if the name does not begin with the four male-specific names then it is female then I will be considered as a female!

Let’s move on to the second example. I don't know if this is true or not anymore but a couple of years ago, we had a law in India where the marital age in India for males was 21 years or older and 18 years or older for females. So, if you are a guy and you want to get married then you can't get married if you are younger than 21. Likewise, if you are woman, then you cannot get married if you are younger than 18 years. So, how do you check, based on the age and gender available in two cells, i.e. let's say that age is

in A1 and gender is in B1 and I want to determine if the person is eligible for marriage or not. So, to check this, the conditions would go like this:
=IF(OR(AND(A1>21,B1="Male"),AND(A1>18,B1="Female")),"Eligible for marriage","Not eligible")

It is better if everything can be combined into one big condition. If either of the two AND functions are true, i.e. for males the age is greater than 21 and for females the age is greater than 18 then they are eligible for marriage. This is another kind of logic that can be expressed in Excel using AND and OR functions.

The third example is a classic requirement. When we are writing SUMIF or COUNTIF formulas, we are often saying that we want to count or sum up the sales for all the products where the product name is equal to A in region 6. However, what if you have a situation where you want to sum up product sales for product A in regions 4 or 5 or 6, i.e. you want to sum the sales against a bunch of regions and the regions could be one of those three? In such cases, we can't really set up an OR condition inside the SUMIFS formula because the SUMIFS function is set up to only check one condition at a time. So, you could say:
=SUMIFS(summing columns,product name,"A", region,4)**+**SUMIFS(summing columns,product name,"A", region,5)**+**SUMIFS(summing columns,product name,"A", region,6)

Essentially because we couldn't write an OR formula inside SUMIFS since the SUMIFS function won't even understand what to do with the OR, **we are using the numerical representation of the OR function**. We are nowhere saying OR in the expression but we are using + and any time that you use +, it really means OR, i.e. we want to add this or that or that. This is where the **+ operator is a replacement for the OR function**.

Likewise, if you wanted to check the AND condition, you could do a multiplication. SUMIFS is not a very good example but there are many situation through SUMPROUDUCT or other situations where you can do a * multiplication to check whether both conditions are true or not through the AND function.

In a nutshell, this is how you can take real world business logic or rules and express them in Excel through the help of the IF formula and logical functions like AND, OR, NOT in a very simple way. The **biggest challenge** when you want to take real world business rules and transform them into Excel **is having clarity**. I am pretty sure that if it is your first time when you want to go and write an IF function with this mix of AND, OR and NOT functions, you will definitely struggle a little bit. This happened to me despite having a lot of programming background and having used Excel for a couple of years before writing the IF function. So, don't beat yourself and don't feel bad if you are trying to attempt to create a business rule in Excel and not getting success. Instead, try to spend some time creating simpler conditions and, if all else fails, you could create a very long NESTED IF function and then slowly improve it until you can come up with the smallest possible IF formula. Again, there is nothing wrong with having a longer formula; the only problem that I find with long NESTED IF formulas is that they are sometimes difficult to understand and tricky to change because business rules can change many times. We can have

a rule that males can marry after 21 years but sometime later, the government can say that irrespective of whether you are male or female, you need to wait till you are 21 years before you get married. So, the rules can always change and, when the change happens, you want to incorporate the change into your formulas as quickly as possible and if you have a very long winded NESTED IF function then it can be tricky.

This is how I would go and express business logic. I hope you enjoyed this podcast. Here is a little something for you - IF you enjoyed this podcast AND you want to support http://chandoo.org/, please go to our iTunes page and leave an honest review there. I hope you enjoyed this podcast. See you in the next one. Bye.