



## Transcript for Session 007

Listen to the podcast session, see resources & links:

<http://chandoo.org/session7/>

### Transcript:

Hello everyone. Welcome back to chandoo.org podcast. This show is dedicated to making you awesome in data analysis, charting, dashboards and VBA using Microsoft Excel.

Thank you so much for joining me in session 7 of the podcast. You can visit <http://www.chandoo.org/session7/> to download the show notes, transcript, links and resources mentioned in this podcast. Let us move in to our episode which is about AweSUM. That's the title of this episode - I'm not referring to the qualitative adjective 'awesomeness' but how certain SUM() functions can be effectively used in Excel.

The SUM() functions, which constitute the functions that are used to calculate the sum of a bunch of numbers, are very useful, easy to understand and simple to apply. However, not many people are familiar with all the capabilities of the SUM() functions. So today I am going to share these with you so that you can do better data analysis. Before we jump into the actual SUM() functions and learn them better, I want to share some exciting personal news with you. The end of April-beginning of May is really one of the best times of the year for me, when I am in India, because that's when my favourite fruit (which is mango) comes into season. I wait for this time every year! This year is extra special for me because my house, which I moved into three years ago, has a mango tree. However, when we moved in, it was a small tree and so we had to wait for a couple of years before it started to bear fruit. It's fully blossomed this year and has more than 150 mangoes! They'll probably be ripe by the last week of May and that's when we'll pick them. Ripe mangoes are available in the market around the last week of April. So I'm really excited as we've started purchasing our first batch of mangoes from the market. Yay! This is personal news but it's exciting for me as it's my favourite fruit and I look forward to this time every year.

Let's move into the podcast and talk about the SUM() functions. If I were to equate the SUM() functions with anything in my life (outside Excel), I would think of them as mangoes - the only difference being that I can use SUM() function any time of the year whereas mangoes are only available in the summer months!

When it comes to SUM() functions, there are a bunch of formulas available. Let's start with the very basic ones. You all know the '+' operator - we learnt this way back in school and much before we learnt Excel. The arithmetic operator '+' sign is used to add up a bunch of numbers. The same holds good in Excel too. If you have a couple of numbers in the cells A1 and A2 and you'd like to add them up, you don't have to write any functions or anything in Excel. All you need to do is type '=' in a cell (the '=' operator indicates that you are about to type a formula) and then select A1, press the '+' key and select A2. You've just written a small equation/expression in Excel in which you are asking Excel, "hey, go and get the value in cell A1 and A2 and add them up."



The '+' operator is that simple. We know it intuitively because the '+' operator is something that we see even before we start using Excel. There's nothing mysterious about the '+' operator; it's something we all know. - And it works in Excel. Just like there's a '+' operator, there's a '-' operator for subtraction, a '\*' operator for multiplication and a '/' operator for division. These are pretty straightforward and as a passionate Excel user spending time learning from chandoo.org's podcast, I assume you know these. That's not why you've tuned into my podcast. The '+' operator is one of the most basic and fundamental operators out there that we can use. But, the disadvantage with it is that it can be somewhat clumsy to use. Let us extend the example of adding a bunch of numbers that are in cells A1 and A2. Instead of just adding the numbers in A1 and A2, let us now add the numbers from cell A1 to A5. To write a formula for this, I would do something like:

```
=A1+A2+A3+A4+A5
```

Although there is nothing wrong with this operation and you can do it, this is the kind of thing that can get you into trouble. Out of boredom, mistake or oversight, we might accidentally include or exclude a cell. This can create a lot of trouble for you when someone starts auditing the workbook or sees a total that is not adding up properly. And then you realise that, by mistake, you've added the value in cell A7 as well, instead of just A1 to A5.

Although the '+' operator is good and it's technically the right way to do it, it's the kind of thing that can create a lot of clumsiness and some errors. For practical reasons, as a smart Analyst, you don't want to use the '+' operator for everything. And that brings us to the next set of SUM() functions that are available in Excel.

Let's start with the obvious ones - one of the most obvious and simplest ways to calculate the sum of a bunch of cells very quickly is to select them and look at the status bar of your Excel workbook. The Excel status bar is a place at the bottom of your screen that quickly tells you what's going on with Excel and gives you useful information about the data you've selected. This status bar, by default, shows the sum of values that you've selected. All you have to do is select a range of numbers and look at the status bar which will tell you that the sum is 350 for example. So you can look at the status bar to see the sum.

Likewise, the status bar can also tell you the average, minimum and maximum of the numbers etc. You can configure the kinds of summaries that are displayed in the status bar by right clicking on the status bar and choosing the kinds of statistics you want there.

That's a quick and simple way to understand the sum of numbers.

But the approach of looking at the status bar is not very repeatable. By that I mean that if I want to see the sum of another bunch of numbers, I must manually select that range and look at the status bar. This creates extra work every time I want to know the sum. It is not a very smart to do it, but it is a quick way to do it. You can quickly look at the status bar to see the sum, but you won't be able to do this if your job is really to calculate totals for numbers every now and then. In that case you would be looking at the status bar all day long! The same issue with the clumsiness of the '+' operator comes into play here too - you could make mistakes, accidentally select an extra cell or something like that.



That brings us to the next best option. - This is really an inbuilt feature of Excel called AutoSum. All you have to do is select a bunch of cells and press the AutoSum button which looks like the sigma sign ( $\Sigma$ ). It looks like the Greek letter for E, i.e. the letter E with a slightly slanted edge towards the right hand side. If you click on it, it will give you the automatic sum formula. There are two ways to use it:

1. The first option is to select the bunch of cells, for example cells A1 to A5, and then press 'Alt'+ '='. When you press that, Excel will insert a SUM formula right beneath the last cell - in this case it will insert a formula in cell A6 which is the total sum of the numbers in cells A1 to A5. Technically, this is equivalent to you writing the formula. However, since you are smart (and lazy!) you don't want to type the formula yourself and you ask Excel to automatically sum it for you. Pressing the keyboard shortcut 'Alt'+ '=' will make Excel calculate the automatic sum for the range of cells you've selected and place it right beneath it.
2. You can use the ribbon button for AutoSum which you can find in the 'Home' ribbon towards the right hand side. It looks the Greek symbol sigma ( $\Sigma$ ), and when you click on it after selecting a bunch of cells, Excel will automatically insert the SUM() formula for you. This is another way to calculate the sum.

So far you've learnt about the '+' operator, looking at the status bar and using AutoSum. AutoSum is a shortcut to writing the formula yourself and it can be handy if you just need to write the SUM() formula. Apart from adding up a bunch of numbers, AutoSum can also give you formulas for AVERAGE() and other simple statistics like calculating minimum and maximum values. But that's about it.

Our focus today is to talk about SUM() formulas, so we'll limit ourselves to that. Apart from these three obvious methods (the '+' operator, status bar and AutoSum), there is one more method which is not common but very useful. This method is called using tables.

If you have sales data for a company called Acme Corporation in cells A1 to A5. As an Analyst, you see this kind of data quite often. However, it's not very realistic data. In real life, the data never starts from cell A1 because you probably have a header name on the first row like sales, revenue, number of customers etc. Imagine that cell A1 holds the header 'sales' and then cells A2 onwards contain the numbers. In fact we could stretch this example a little further and imagine that column A contains the month in which the total sale happened, and column B contains the actual sales figures. This is more realistic. Column A contains months and column B contains the sales figures for that particular month. A1 would have the title 'month' and B1 would have the title 'sales'. The numbers and dates would flow in the downward direction.

When you have data like this, you could use an Excel feature called 'Tables' to treat all of this data together. I am going to leave a link to what tables are, how to use them, why they are important for a Data Analyst like you etc. on the show notes page. Please visit <http://www.chandoo.org/session7/> to access all the links and resources available in this podcast

If you have a table of this data, you could go the 'table design' options and insert a bottom row which would be a summary row where you could display quick statistics, one of which would be the total. Technically, this is like AutoSum, but the beauty of this functionality is that through using tables if you add some extra data, the bottom row would automatically move down and all the data would be



between the header row and the totals row. This is how tables can help you with calculating the sum automatically.

There are four obvious and simple ways to do this - the '+' operator, AutoSum, status bar and a table's total row. The ones we've discussed are the most obvious ones.

Now let's get into the serious matters of summing data - i.e., how do I sum for some practical real-world scenarios. And that's where a beautiful and very powerful functions called SUMIFS() comes into the picture.

This is a very powerful function that every aspiring Analyst should learn. SUMIFS() function is a very powerful and beautiful function without which you would spend a lot of time to answer questions that can be done in just a fraction of seconds by using SUMIFS(). I will link to an article in which I talked about SUMIFS() formula in-depth and showed a lot of examples highlighting its importance.

Let's talk about what the SUMIFS() function does and why you need to use it. Let's go back to the example of the sales of a company where you have the month name and the total sales for that month. Let's say you are the boss of the company and you're looking at this data which has been recorded for the last 24 months. It's 28th April 2014 as I record this podcast. So let's imagine the data starts from April 2012 and ends in March 2014. - We have these 24 months and their corresponding sales figures. If I'm looking at this sales data as a CEO, Sales Manager or a department head, I'll never have a question as simple as "what re the total sales in all the 24 months?" I might ask the question, but for practical and analytical reasons, this kind of sum is practically useless.

As a manager, I might be more interested to ask questions like, "What are our total sales in the latest quarter?" Or, I might ask, "What are the total sales in the latest holiday season (defining holiday season as November to January for our company)?"

Likewise, if we have data for individual customer names instead of by month - I might ask, "What are the total sales where the customer name is ABC.com?" These are the kind of questions that you would ask as an Analyst. Stretching this a little further, as a Manager or as an Analyst, you would ask, "What are the total sales for the customer ABC.com in the month March 2014?"

The question that you're asking has two parts - what is the total for 'customer = ABC.com' and 'month = March 2014'. The first part of the question is asking for the sum (or total) and the second part is specifying some conditions - i.e., if customer is equal to ABC.com and if month is equal to March 2014. So, it's a SUM with some IF conditions and that's how the SUMIFS() formula comes into the picture and hence the name.

The SUMIFS() formula can give you the sum of a bunch of numbers that meet any number of criteria that you specify. It has two components - one is the column that you want to sum up which is the sales figures in our example (say column B or C); and the second part is the conditions where condition1 could be 'customer name = ABC.com' and condition 2 could be 'date = March 2014'. Explaining actual syntax and examples is not really feasible in a podcast. But, I want to plant the idea of SUMIFS() into your mind so that you can appreciate what it does. You can visit the show notes pages which is at <http://www.chandoo.org/session7/> where I will provide you a link to a detailed description of the



SUMIFS() formula, some examples and tutorials because I've covered it in great detail on my website. So I will provide links to that.

SUMIFS() helps you answer real world common questions that people generally ask. Nobody will really ask you the total sales since the company was incorporated as that would be a really huge number for companies like Walmart. And nobody is interested to know that figure really. People are interested in knowing what happened in the last quarter, what happened to a particular customer or vendor, or the sales a particular employee made. So, it's usually a SUM with some conditions and that's where the SUMIFS() formula comes into the picture.

Those of you who are using an earlier version of Excel like Excel 2003 or prior to that, please keep in mind that the SUMIFS() function is not supported in those versions. SUMIFS() formula is a new function added to Excel in Excel 2007. But I assume that a majority of our users on chandoo.org and the people who are watching this podcast have probably moved onto better versions of Excel like Excel 2010 or Excel 2013 and so you don't have to worry. However, if you are using an earlier version of Excel, please keep in mind that there is a version of SUMIFS() that is available called SUMIF().

SUMIF() is the original formula that Microsoft had, but it can only take one condition at a time. It can give you the sum of sales where the customer name is equal to ABC.com. Or, it can give you the sum of sales where the date is equal to March 2014. But it can't give you the sum of sales where customer name is equal to ABC.com and the month is equal to March 2014. The SUMIF() formula doesn't work for multiple conditions as it can only take one condition. So Microsoft eventually added the plural version of this formula called SUMIFS() from Excel 2007 onwards.

If you must calculate this kind of multi-conditional sum in earlier versions of Excel then you could use the SUMPRODUCT() formula or DSUM() where D stands for database. You can also use a technique in Excel called conditional sum wizard. It is part of Excel 2003 and is basically a version of SUMIFS() constructed through the SUM() formula and some ARRAY formula techniques. It's way too complicated if you ask me. If you are using Excel 2007 and above, just use the SUMIFS() formula and be done with it! There's no need to be worried about ARRAY formulas or even the SUMPRODUCT() formula.

I will provide links to SUMPRODUCT(), conditional sum wizard and how to achieve these results in earlier versions of Excel in the show notes pages. Please visit <http://www.chandoo.org/session7/> for that.

Moving on, let's talk about some practical scenarios where the SUMIFS() formula would work really beautifully:

1. Let's say you want to sum up all positive numbers in a range of cells and for the sake of simplicity let's call this range 'list' (range name). Since this is a podcast, some of the things like syntax are irrelevant here. So, I'll explain the basic technique. I will read out the formula, but you can simply ignore that part. The formula would be:  
`=SUMIFS(list,list,">0")`

The part within the double quotes is the most important part here. If I put a condition with a greater than or less than sign within the double quotes (like we have ">0" in the above formula), it tells SUMIFS() to evaluate the condition within the double quotes against the condition



column. Here we're asking Excel to check values greater than zero within the list. It'll check each item in the list to see if it's greater than zero, and if it is, it will be added up. Otherwise, it will be ignored. This is how you will calculate the sum of all the positive numbers. I will paste the actual formula and also include an example workbook to practice this on the show notes page located at <http://www.chandoo.org/session7/>.

2. If I want to calculate the sum of sales that happened in March 2014 - Let's say you're looking at the sales data and the sales are captured for each and every day. So you have sales for 28 February, 1 March onwards till 1 April and beyond. Sales are recorded in one row for each day. If I am looking at that kind of data and I write a SUMIFS() formula with a condition that goes loosely like this:

=sum of sales(sales column,date column,"March 2014").

This is how we picture it in our mind because what we are saying is that we want the sum of sales for the month March 2014. But, Excel does not understand what March 2014 means and that's the real problem. You and I know that March 2014 is the month that starts on 1 March 2014 and ends on 31 March 2014. When I tell you that I am visiting Europe in July 2014, I don't have to spell it out that I will be in Europe from 1 July to 31st July. That part is irrelevant, because you and I know that I mean the month of July. However, Excel is a computer program and it isn't as smart as we really think it is. So we need to spell out to Excel what we mean by March 2014 because Excel can only understand a single date. If I say:

=sum of sales(sales column,date column,"1 March 2014"),

Excel can give us an answer because it understands that date very clearly.

It will go and look at your data and when it sees that there is only record with that data, it sums up the total for that date (1 March 2014) and it gives you the answer. But, if I want to ask Excel what is the total sales for March then I must ask it the right question. The question will be something like this:

=SUMIFS(sales column, date column,">=1 March 2014",date column,"<1 April 2014").

Essentially, we are giving the SUMIFS() formula two conditions - the first condition is that the date should be greater than or equal to 1 March 2014 and the second condition is that the date should be less than 1 April 2014. We must specify the entire boundary, and the boundary for the month of March is that it starts on 1 March 2014 and it ends right before 1 April 2014. When you specify this kind of boundary with greater than and less than symbols, then SUMIFS() formula will know the items to include in the sum and then it'll do its job properly. Again, understanding the syntax and very specific nuances of this formula can be tricky in a podcast. So visit <http://www.chandoo.org/session7/>, where I will provide a link on how to add up a bunch of numbers that meet a date criteria.

3. Let's say that you are looking at the sales data by individual customers - so you have the customer names in one column and the sales data in another column. For the sake of argument, let's assume that you are the CEO of WalMart and so you probably have billions of customers. Let's imagine, for a moment, that all this data is housed in Excel and you're looking at this data and wondering how much you are selling to Taco Bell. The challenge is that when sales are made to Taco Bell, the customer name is not always Taco Bell because they have some



accounting codes. So you have 17 different Taco Bell accounts in your system like Taco Bell 1, Taco Bell 2, Taco Bell 3 and so on. They have 17 different Taco Bell names which are used in some seemingly random fashion in your sales records. If we write a SUMIFS() formula to calculate the sales, it will give us a value of zero because there is no customer called Taco Bell (they are called Taco Bell 1, Taco Bell 2 etc.). Since they have 17 different versions of their name, you'll end up writing 17 different formulas! And that's not pretty at all!

This is where wildcards come into the picture. What do they do? They simplify the job of answering questions like "what is the sum of sales where the customer name starts with Taco Bell?" The formula will go something like:  
`=SUMIFS(sales column,customer column,"Taco Bell*").`

The star symbol used here is called an asterix (press Shift+8 to type it). Essentially, you're saying that you don't care what's written after Taco Bell or how many characters there are - as long as the names starts with Taco Bell, you want the sum of it. This kind of wildcard formula can sum up all the different names of Taco Bell that appear in the system, as long as they start with the words Taco Bell.

To quickly re-cap the three scenarios that we talked about are summing up all the positive numbers where you would use a greater than operator for the condition statement; summing all the sales within the month of March where we need to specify the boundary conditions ( $\geq$  1 March 2014 and  $<$  1 April 2014); and summing all the sales where the customer name starts with Taco Bell where we would use "Taco Bell\*". This is how you would use SUMIFS() for these three practical scenarios. SUMIFS() is a really great function that works with a lot of other scenarios too, but these three are the things that many people don't know and that's why I thought it's important to highlight them.

Before we jump into other functions that deal with summing a bunch of numbers, let us quickly understand some of the cousins of this SUMIFS() formula. Just like there is a SUMIFS() formula, there is also a formula called COUNTIFS. It counts the items that meet a particular criteria or a bunch of criteria. So instead of summing the sales where customer name is equal to ABC.com and month is equal to April 2014, you just want to know how many transactions that customer did. You would use the COUNTIFS() formula in this case. It will give you the count of items where the customer is equal to ABC.com and the month is April 2014.

Similarly, there is a function called AVERAGEIFS(). It calculates the average meeting various criteria specified.

These are the important functions that Excel has - SUMIFS(), COUNTIFS() and AVERAGEIFS(). They are the plural functions as they can take more than one condition.

There are also singular functions of all these functions - SUMIF(), COUNTIF() and AVERAGEIF(). They work with one condition at a time, so if you just have one condition, you can use these formulas.

For the sake of simplicity and uniformity in your workbooks or dashboards, I encourage you to purely rely on the plural versions of these formulas because that way you can add multiple conditions at any



time and you don't have to change the formula syntax or anything like that. So use the plural version everywhere even if you have just one condition.

We have spoken about the SUMIFS() formula which is really the most important summing function in Excel - in fact, I would say that it is one of the most important functions in Excel (apart from the lookup functions). In case you don't know what lookup functions are, I highly encourage you to listen to the second episode of our podcast called VTALKUP. You can visit <http://www.chandoo.org/session2/> to learn more about VLOOKUP functions and how they can help you.

Moving on, apart from these scenarios, there is also a scenario where we do some adhoc analysis on the data. Let us imagine that you are looking at the sales data and you've filtered it to a smaller set using the filter functionality of Excel because the data set was very large. So you're now looking at a smaller subset of the data. Anybody who has ever filtered data in Excel knows that when you filter, Excel hides the rows that are not required. Technically you are filtering, but in the background, Excel is just hiding the rows that are not required to be shown.

We would like any summing or counting that we are doing on this filtered data to only take into account the visible cells and not the ones that are hidden away by the filter. But the SUMIFS() formula, AutoSum, status bar etc. work for the entire data. So if you have selected a bunch of cells like A1:A10 and you are using a SUM() formula like:

```
=SUM(A1:A10),
```

then even if cells A3 or A7 are hidden, the SUM() formula will still take them into account.

In such cases you might want to use a function that only includes the visible items and ignores anything that is filtered. There are two functions in Excel that can help you with this kind of scenario, called SUBTOTAL() and AGGREGATE(). For summing or counting, we can safely imagine that both of them do the same thing, i.e. both SUBTOTAL() and AGGREGATE() will behave the same way.

The AGGREGATE() function has some new features and it is available from Excel 2010 onwards. The SUBTOTAL() function is available right from Excel 2003 onwards.

We will skip the AGGREGATE() function for now and talk about it in a future article of chandoo.org.

Let us talk a little bit about the SUBTOTAL() function. The SUBTOTAL() function can give you a subtotal of whatever is visible. There is an argument in the SUBTOTAL() function that you can specify, in order to include the hidden cells also.

To use the SUBTOTAL() function, you need to specify the range that is needed for the SUBTOTAL() function and the kind of sub-total statistic you want, whether you want the sum, count, average, minimum or maximum etc., and whether you want to include or exclude the hidden rows. There are different function codes for each of them.

My suggestion is that instead of trying to memorize all of them while listening to the podcast, you can simply go and select a bunch of data and start writing a SUBTOTAL() formula if you haven't already done it. That's the best and easiest way to learn it.





There is also one other way which is actually more awesome! Just visit <http://www.chandoo.org/> and search for SUBTOTAL(). If you're lazy and smart, just like I am, you could visit <http://www.chandoo.org/session7/> where I will provide a link to how to use the SUBTOTAL() function better.

That is a little bit about the SUBTOTAL() functionality. There is also the AGGREGATE() function, and I encourage you to play with it, by searching on Google or simply typing it in an Excel sheet and seeing what it can do.

In a nutshell, these are the various SUM() function in Excel. They're quite awesome and very powerful.

There are some things and their formulas that we have omitted from this discussion, like SUMPRODUCT() and DSUM(). These are the two functions that we didn't discuss in detail. Keep in mind that both of them have slightly complicated syntax and usage scenarios. I encourage you to first master the SUMIFS() formula and other types of summing options like SUBTOTAL() and AutoSum. Once you feel confident and familiar with them, and feel like you've mastered them, then go and spend some time learning about the other functions. SUMPRODUCT() is especially complicated and it would take at least a couple of weeks to master, if you're just starting with it. In fact, I would say that it took me more than a year to really appreciate what it really does.

Don't be taken aback by its complexity, but just keep in mind that it is a difficult function to understand and appreciate, when you are just starting out with Excel. Again, I will provide a link to the SUMPRODUCT() formula, how it can help you and some advanced example scenarios on the show notes pages at <http://www.chandoo.org/session7/>.

To summarize, the things that we have talked about in the podcast - we talked about the obvious ways of summing like the '+' operator, AutoSum, status bar and table total row. Then we talked about the SUMIFS() formula which is a really beautiful, powerful and all-encompassing formula. We talked about three specific scenarios - summing positive numbers, summing values in a month like March 2014, and summing values where the customer name starts with Taco Bell. We also talked about summing scenarios where we are filtering the data. And, we talked about the SUBTOTAL() formula. Other things that we also talked about are COUNTIFS() and AVERAGEIFS() formulas, which are cousins of the SUMIFS() formula, and how they behave with multiple conditions.

That's about it for this particular podcast. I hope you have enjoyed the episode. I really like talking about SUM() formulas. Any time that I do a live or online class, or any time that I talk about formulas with anyone, I talk about the SUMIFS() formula because it is a very powerful, flexible and versatile formula that I can apply in any type of analytical situation.

I highly encourage you to spend some time with these formulas so that you can also realize how awesome they are!

Thank you so much for spending some time with me in this episode. I hope you have enjoyed it and I will see you again in the next session of chandoo.org's podcast.

Until then, stay awesome. Bye.