## Transcript for Session 021

Listen to the podcast session, see resources & links:

http://chandoo.org/session21/

**Transcript:**

Hello everyone. Welcome to session 21 of chandoo.org podcast. If you are thinking why my voice is sounding a little odd it's because I'm recording this podcast from London Heathrow airport on my way back to India from Dallas. I had a 4 hour wait in London and I was just checking my emails and having a healthy breakfast and roaming around the airport to see if I can buy something nice for my wife. Suddenly, I thought that I'm sitting here doing pretty much nothing; why don't I record a short format podcast. So that's what I'm doing! Please excuse me if you hear any flight announcements, crying babies or running passengers. I've tried to come to a quiet corner but you know how it is in airports. You don't have much freedom!

Today let's talk about how you can **quickly compare 2 lists of values**. They could be numbers, names, alpha-numeric values like invoice numbers, vendor IDs, customer codes or product names. **How do we compare two lists of values and quickly identify if there are any duplicates, overlaps or any values that are only present in one list or another.**

This is something that many people often ask and it is also something that I find we end up wasting a lot of time on to figure out. Usually one of the biggest challenges with Excel is that it doesn't contain real data. It is coming from some other place - often from a database, .csv, or as a list of values in an email which you paste into Excel. There is always the scope of combining data from two different systems or two different times. For example, you get a list of all the products sold this month and then next month, you get another list of products sold in that month and you just want to figure out the overlaps so that you can do a comparative study. When that's the requirement, I see that most often we end up wasting a lot of time because we don't know how to do a quick comparison.

Basically, when it comes to comparison, there are **5 techniques** in my mind. Of course, there could be more because Excel has tons of features that could be used to achieve the same objective or output. But, I think 5 of them come to mind the most.

The first and most obvious one is **manual comparison**. Now, this is not really awesome but this is what some people resort to do because they don't know better. I used to do manual comparisons too when I started working. Earlier on I really had no clue what a formula was, how to use it or any feature to find out if there is a duplicate or not. So, I would manually compare. With manual comparisons you always have chances of errors. However, you don't need to know any techniques for manual comparison. You can just use your brute force and maybe a couple of cups of coffee or tea to keep you going and you're done in an afternoon! But let's not talk about manual comparison because you're not listening to this podcast to waste your time. You're listening to this so that you can be awesome.

The next technique that I have for you is called **conditional formatting**. Starting with the Excel 2007 version, Microsoft has introduced a feature in conditional formatting to quickly highlight duplicate values. If you have one list and you select it and go to your home ribbon and select conditional formatting and highlight duplicate values, it will highlight the duplicate values in that list. What if you have two lists like this month versus next month or system A versus system B? In such cases, you would select the first list, hold down the Ctrl key and select the second list. By holding down the Ctrl key, you are able to select multiple ranges in your workbook. Once you select them, go to conditional formatting and select highlight duplicate values. That'll highlight all the duplicate values in both lists. It will consider both lists as the total list and any item that appears in both will get highlighted. This is a very fast and quick way to get immediate results - a visual indication of the overlaps - whatever is not highlighted are the unique values. You can also go the other way. Instead of highlighting duplicates, you can highlight unique values. That way only the values that are present in one of the lists will be highlighted. One of the assumptions here is that there are no duplicates within each list. That means that if you have a list of products that are sold in this month then no product name is repeated twice within that list because in that case conditional formatting will highlight those values also. Keep that in mind. The conditional formatting technique is pretty awesome according to me. Anytime that I demonstrate it in my live classes, online training programs or even if I just tell people this while talking to them, they are super excited! It shows them how a powerful thing like that can be done in Excel in just a fraction of a minute or maybe even in a couple of seconds. That's conditional formatting for you. It's the top most method that I recommend.

The second method is that sometimes you have data where the number of rows in both data sets is the same but there are certain differences between one data set and the other data set. A good example is that let's say you are selling to 500 vendors and between the two months the vendors have not changed. It's supposed to be the same 500 names. However, when you import the data into Excel, you realise that they are not matching and you don't know where the mistake is happening. This is where the **row differences feature** in Excel helps. As the name suggests, it can take two ranges or two sets of values and compare them row by row. So, row 1 in list 1 will be compared with row 1 in list 2, row 2 in the first list will be compared with row 2 of the other list and so on. When you select row differences and press OK, it will highlight all the cells that do no match in both rows. This is pretty useful because when you have something like that you can quickly see where the differences are and figure out the reasons for the differences. For example, maybe there is an extra space at the end of the vendor name or the vendor name was abbreviated in one list and fully written out in the other list. This is a very powerful feature but it is only useful if you have the same number of rows in both lists and you just want to highlight where the differences are. To access the row differences feature, you just need to **select both ranges by using Ctrl, press F5 which will open up the Go To box, click on special and choose row differences**. That's your second technique.

The third technique for finding duplicates is to use a formula like **VLOOKUP**. The purpose of VLOOKUP is to lookup a value in a list and tell you whether it is present or not. If it is present, you get a corresponding value. So, this is useful for us to compare one list with another and quickly see if a value in the first list is present in the second list or vice versa. While VLOOKUP is powerful, it has one drawback. If a value is not found, it returns a #N/A error. So you need to be aware of that because when you compare with VLOOKUP you also need to figure out what to do with the errors. Obviously having an error means that a match was not found. But maybe by using an IFERROR formula you can change the #N/A to 'value not found' or something similar. That's using VLOOKUP.

The last technique that I have for you is using the **COUNTIF** formula. If you just want to count how many times a value is present in the other list, you can use the COUNTIF formula instead of the VLOOKUP formula. This way you won't have to deal with error messages. If the value is not found, you'll just get 0 indicating that the value is found 0 times in the other list. If it is found multiple times, you'll get 1, 2 or 3 depending on how many times the value is found in the other list.

These are the four techniques that I often use to compare two lists. If you are wondering why I am not telling you the syntax for the VLOOKUP and COUNTIF formulas, it's because while I can explain the syntax in an audio podcast, visualizing it, trying to remember it and implementing it in Excel could be difficult for you. So, I will add an example workbook and also link to a couple of articles on the show notes page of this podcast. Please head to http://chandoo.org/session21/ for details on the VLOOKUP and COUNTIF formula syntaxes and an example workbook.

There you go; these are four techniques for comparing two lists. I also have a small **bonus tip** for you. If you have just one list but you're struggling with a lot of duplicate values and you're trying to get rid of them, you could use the **remove duplicates button**. This is in the Data ribbon. Select the range of values and select the remove duplicates button from the Data ribbon and your list will be instantly cleaned.

That's about it. I know my flight is about to start boarding so I'm going to sign off now. Stay awesome and I'll see you again in another episode. Thanks. Bye.