Transcript for Session 018

Listen to the podcast session, see resources & links:

http://chandoo.org/session18/

Transcript:

Hi. Welcome to chandoo.org podcast session 18. This podcast is dedicated to making you awesome in data analysis, charting, dashboards and VBA using Microsoft Excel.

Let's get into session 18 of the podcast. Please remember that you can go to http://chandoo.org/session18/ to access the show notes, transcript, links and resources mentioned in this episode.

Today we are going to talk about pivot tables. The title of this episode is **'Don't be a pivot table virgin'**. Although it sounds like something trying to provoke a lot of attention, the reality is that I find many people are pivot table virgins. I mean that they don't use the pivot table feature in Excel at all. They go to great lengths to avoid them. This always surprises me because I used to be like that a while ago. But, ever since I started using pivot table 5 years ago, I have never looked back! Pivot tables have saved me a great deal of time. They provided me elegant solutions in situations that I thought were impossible to solve through formulas or through some other way. Today we are going to talk about pivot tables. Keep in mind that this is a short format podcast. That means that we will just talk about the basics and a few introductory topics. And, I will leave you with a ton of resources in the end.

Before we talk about pivot tables, let me share a couple of happy announcements with you. The first one is that **our podcast completed 100,000 downloads**! That's right. We are 100k downloads strong now. Yay! This is actually a very good achievement to reach in the six months since we launched the podcast. I think I launched it around 18 March 2014. And, in the last six months, we went from 0 downloads to 100,000 downloads. This is really good progress. And, it just took 17 episodes. I am glad that so many of you are tuning into the podcast and enjoying it. Thank you so much for loving it. You are really awesome.

The second happy announcement that I have for you is that **my upcoming live class in Houston, Texas has been sold out**! In the last two episodes I have been telling you that there are a few spots left that you can go and join. But it turns out that all the spots for the class have been filled up. I am really looking forward to meeting those 25 individuals who have signed up for the class and help them become awesome in Excel. I will share more updates with you in a future podcast, once the class has been completed, so that you can also learn about the things that happened in the class and what the reactions were.

Now, let's talk about pivot tables. What are pivot tables? That's the very first question that you might be asking especially if you have never used pivot tables before or if you have never cared to understand what they are doing. You may be using pivot tables already because they are part of a business reporting

template or some other workbook that your colleague has prepared and that you have inherited. But, you have not really spent a lot of time trying to understand what a pivot table is.

To appreciate what pivot tables are, I think it is important to understand how business data looks and what business reporting and Business Managers are asking for. To keep this a little more interesting and imaginative, let's imagine that you are the Chief Analytic Officer of Walmart, Texas. Walmart is a big company and has branches all over the world. We pick Walmart quite often because it is a company that most of us can easily visualize. All of us have a good idea about this company and quite a few of us have been inside a Walmart at least once in our lifetimes. It is a retail outlet company. They sell all sorts of things right from needles all the way up to jumbo jets in one of their stores. You are the Analytics Chief of their Texas operations. You're not the Analytic Chief of the entire US or the entire world; you're the person who's responsible for a single state. As part of your monthly data collection (or whatever you might want to call it), you receive a big file that also contains all the product level transactions within your state in a .csv file. Let's just imagine that the data fits into a .csv file because, in reality, a company like Walmart would probably have billions of transactions in any given month and such huge data wouldn't be in a .csv file. It would probably be in a database.

But, humor me here and just imagine that you're getting the .csv file which contains a bunch of columns. The columns are the date of the transaction, the store in which the transaction happened (for the sake of simplicity just imagine that within Texas they only have 10 stores), product category (candy for example), product name (one of the candy names), how many units are sold and at what price. This is the structure of the table. There are 6 columns - date, store, product category, product name, units sold and price. This 6-columnar data is what you get at the end of each month. You're looking at one of the recent reports. One thing is striking about the data that you have received. It is tall and thin. In other words, it is tall and narrow. By that we mean that the file has millions of rows and so it is really tall. If you look at it on a computer screen, it is a really tall rectangle. It is quite tall, but it has only 6 columns. So, it is tall and narrow. No Business Manager or Analytic Chief in their right mind would ask for a print out of all the million transactions in the past month. That kind of question is never asked in any Board room nor will it ever be asked in any sensible scenario. People may ask such a question if you are auditing something because you require the specifics in that scenario. But, purely from an analytic or insight point of view, you don't really require all that data. What you really want is a transformation of the data. The word transformation is the key here. Bookmark it in your mind. We'll visit it in a few minutes.

What kind of transformation? For example, given this data, as an Analytics Chief you may be interested to know the answers to specific questions like candy sales on Fridays. What are our candy sales on Fridays? This could be an interesting question to explore purely from an analytics point of view. For example, you have a hunch or hypothesis that you tend to sell more candy on Fridays than the rest of the week. Maybe because on the impending weekend and kids being home, moms or dads visiting the store on Friday might pick up an extra bag of candy so that the kids have something to chew when they are at home on the weekend and spending all their energy playing or ruining things. This is a hypothesis that you have. To test this hypothesis, you really want to know what the candy sales look like on Fridays. This could be one of the reporting or information needs from an analytic point of view. Likewise, you might be asking questions about the shoes sold in the Dallas stores. Although you are responsible for the entire Texas area, you just want to narrow it down and look at the total number of shoes (as a category) that you are selling within the stores of the Dallas metropolitan region. Likewise, you might be asking a

question about the top 10 product categories in the first week of this month. Again, it's a very specific question that helps you understand where things stand and which are the outperforming product categories. These are the kinds of questions that Managers, Reporting Chiefs, Analytic Chiefs, CFO's or CEO's will ask. Nobody ever asks for a print out of all the million transactions.

When this is the criteria, we are trying to transform the data into a format that we want. Given the million rows of data, we want to transform it so that we can only look at candy sales happening on Fridays. Or, we can only look at the shoe sales happening in the Dallas stores. Or, the top ten product categories out of all this data. This is the kind of transformation we are talking about. We are taking the data and we would like to transform it so that we can answer specific questions or so that we can provide specific types of information.

This kind of transformation, in computer terminology, can also be called as pivoting. That is, we take the data and we pivot it. The dictionary definition of pivoting would be to turn one on another [or something like that]. What we're trying to do is that we're trying to pivot the data so that we can get it into the shape that we want. Here the shape could be product categories arranged in descending order of the sales in week 1 with only the top 10 displayed. Or, candy sales happening on weekdays but only shown on Friday. That's the kind of transformation that we want. That's the kind of pivoting that we want. That's exactly what pivot tables do. At this point you might be thinking, "Hey Chandoo, wouldn't a SUMIFS formula or VLOOKUP formula or some other formula like SUMPRODUCT help us answer such a question?" Obviously, yes. Those formulas are also essentially built so that you can transform data to information. These formulas are also good and we use them often. For example, we could write a SUMIFS formula to get the answer to a question about how many shoes were sold in the Dallas stores. Let's say that Dallas has 3 stores and we want to answer the question for the 1st store. So, we could write a SUMIFS formula as follows:
=SUMIFS(quantity sold column, store ID column = "Dallas store # 1", product category column = "shoes")

That will give us the answer for shoes sold in the Dallas stores. You can use the SUMIFS formula for that. But, the challenge with the SUMIFS formula is that it involves a bit of typing and writing. And, this formula only answers the question for the product category of shoes and for stores in Dallas.

What if the reporting requirement immediately changes? After looking at this, you have a hunch that maybe you shouldn't be exploring the shoes category but that you would like to explore the swimsuits category instead.

To change this, you have to immediately go back and change the formula. In many business situations, the person who creates the formulas is not the person who consumes them. There is a difference between the creator and the consumer. As an Analyst listening to this podcast, you are likely to be the creator. You are the one who is creating various dashboards or reports or analytical outputs. But, most likely, you will not be the person who consumes them. It could be your boss, the CEO, a regulatory authority or it could be somebody else in your company who consumes this. When the requirement of the consumer changes - i.e. when the CEO wants to look at the swimsuit sales instead of the shoe sales - then the consumer must came back to the creator (you) and ask you to change it. This creates an unnecessary back and forth even to answer a simple question like, "Hey, The shoes are fine. I want to find out about the swimsuits now." This is where pivot tables can quickly help the consumers get their answers without your intervention. Once you set it up, your boss could quickly change some simple

screen setting (like change the filter or click on a button) and see the same report for a different store or see the same report for a different category etc.

That's how pivot tables help. Pivot tables are designed so that people who don't know SUMIF formulas or people who cannot use SUMIF formulas or in situations where you may not want to use SUMIF formulas, you can still get the answer. Some of the situations where we may not be able to use formulas etc. will be elaborated in the resources section of this podcast.

Let's move on. Pivot tables can help us quickly answer questions like this. You might be asking, "All of this is fine. I want to create my first pivot table. How do I go about it?" So, let's do that.

Before we create our very first pivot table, **let us understand the terminology** for when you are setting up pivot tables. Although pivot tables have a lot of different terminology and technical words that you come across once you start playing with them, there are really four different items that you should understand. The first one is called **labels**. The second one is called **values**. Understanding these separately might not help, but what they do together is what we will try to understand. Go back to the example of Walmart where we had 6 columns of data. Those 6 columns were date, store, product category, product name, units sold and price. All of these 6 columns are data columns, but if you observe closely, the last two columns (units sold and price) are the 2 real data points. These are the two based on which we would do some sort of calculation. Those two columns can be termed as values in pivot table terminology. You might also hear the words **value fields**. This is because each column in our table is considered as a field, so naturally the number columns are considered as value fields. In most business data, the last couple of columns are value fields. This is because the general structure of business data is like this. Depending on your internal company database structure or the format in which you get the data dump (.csv file or flat file or some other way), this could change but in most scenarios it will usually be the last couple of columns that form the value fields. Everything else can be considered as labels. In a nutshell, this is what label and value fields are. These are the two things that matter most when you are setting up a pivot table. There are two more things that we'll talk about but these two are important. Now that we've understood this is loose terms, let's get a little more technical. The value field is the one which offers the data part. It is really what matters when we are doing calculations. In our case, quantity and price are the value fields.

Now comes the ‘everything else’. What are they? They are the ones that are telling us information about the data. They aren't really data but they are giving us data about data. These kinds of things are called meta-data in the computer world. It's really data about data. That's what meta-data means. Again, you don't have to complicate things in your head. The simple fact is that there is data in your dump. In your entire data set there are a couple of columns that contain numbers and they are called value fields. Everything else is a label. Here, everything else really means columns that quantify the data. Again, all of this makes a lot of sense once you start playing with your very first pivot table. Trying to imagine all of this in an audio podcast might be tricky. I definitely understand that. In fact, it took me several years to grasp the whole concept of pivot tables and create my first few pivot tables. I created a couple of pivot tables very early on in my Excel career, probably in 2005 or so, but I did not understand them much. I really did that purely out of an automatic requirement to set up pivot tables for a report.

Later on in my career, in the year 2008 or 2009, I started thinking hard about them and only then understood what pivot tables are. Don't feel bad if you do not get the concept right away. It could take

some time.

The two things that matter most are labels and values and we've understood them. The third concept that's equally important is called **groups**. Sometimes the label fields can be combined to create groups. A classic example, as per our data, is product category and product name. Certain product names belong to a certain product category. For example, shoes are a category. Within that, we could have different types of shoes that we are selling based on the shoe size, brand name, gender and age group for which the shoes are manufactured, and based on the requirement etc. This is nothing but grouping.

Another classical grouping that you can obviously observe in a natural setting is dates. Dates logically fall into weeks, weeks into months, months into quarters and quarters fall into years. This is nothing but a naturally created hierarchy for you. Grouping or hierarchies is the third concept of pivot tables.

The fourth concept is **filters**. Once the pivot table is set up, you might want to filter it in such a way that only the information that you want is displayed. For example, we might want to see only the top 10 product categories in week 1. So, we would like to remove all the other weeks and all the other product categories that are not in the top 10. This is nothing but filtering. We don't want to see things that we don't need in the report or in the information that we are creating.

Filtering can be done in four different ways. The four concepts that you need to understand are labels, values, grouping and filters. Within filters, there are **four different types of filtering**:

1. report filters
2. slicers
3. label filters, and
4. value filters

Now, I am going to provide only the teasing information here. I am not even going to talk about the four types because that's where the next section of the podcast comes into the picture where we will be creating our very first pivot table. This is all that is there to understand about pivot tables before you create your very first pivot table and lose your pivot table virginity.

Let's go ahead and build the pivot table. Obviously you may be thinking, "Hey Chandoo, I am driving, going to my job, lying down on my couch or commuting to work and listening to this podcast. How do you suppose we will create a pivot table?" Well, you are right. You are not going to create any pivot tables while listening to this podcast and probably driving across the country or running on a trail.

Instead, what I have for you is a set of resources so that you could create your very first pivot table. I will be providing a link to a detailed **pivot table tutorial** that is really set up so that anybody can lose their virginity as far as pivot tables are concerned. That link can be accessed from http://chandoo.org/session18/. Go to that link for a detailed step-by-step tutorial that'll take you through various steps of constructing a pivot table from raw data, how to customize it, how to change the way it looks, how to format it and how to modify the layout of a pivot table. It also explains all these things that we just understood - what are labels, what are value fields, what is a row label, what is a column label etc. That particular page also contains a sample data set. You can download that data set. It is similar to the Walmart data set that we just spoke about and it contains more than 1000 rows. It gives you something to play with. You don't have to ruin your business data; you can take this sample

data set, play with it and understand how pivot tables behave.

If you are curious to know how to learn more about pivot tables, there are a couple of ways. The very first way is obviously on http://chandoo.org/ itself. We have a bunch of tutorials, articles, examples and case studies about pivot tables. Again, the link will be on http://chandoo.org/session18/. Go there and explore the pivot tables page that is linked from there. You can learn a lot more about pivot tables by simply browsing or spending 30 minutes to 1 hour every couple of days until you are done with all the articles. That's the most easy and simple way to do it.

The second way [these ways are not in any hierarchy; all of them are equally good] that I recommend is to get a book. There are many books on pivot tables. One book that I certainly recommend is a book **written by Bill Jenin and mike Alexander**, both of whom are my good friends. They are excellent authors. They have a great sense of humor, they really bring the subject alive and help you understand how to use them, where to use them, how to set them up, how to customize them etc. The entire plethora on everything about pivot tables is covered in this book. The book title is "**Excel 2013 Pivot Table Data Crunching**". I believe they have similar books for earlier versions of Excel also. Even if you are running an older version of Excel, I highly recommend that you get the latest version of the book because that way you can also use the book as soon as your company upgrades to a newer version of Excel. The book is named "Excel 2013 Pivot Table Data Crunching".

The third way to learn pivot tables is to join a course. Again, I don't want to sound like I am selling something to you in this podcast but my **Excel School online training program** offers excellent amount of information on pivot tables. Pivot tables are just one module in this 12-module course that I run. If you prefer, you can go in for the course. The course helps you get a holistic perspective of Excel and use it for analyzing data and making presentations, visualizations or dashboards very easily. As part of that, learning pivot tables is important. So, go in for the Excel School online training program if you want to learn this pivot table concept along with everything else in Excel through a course.

Those are my recommendations. I hope you enjoyed this short format podcast on pivot tables. If you have some suggestions, comments or anything else to share about pivot tables, please head to http://chandoo.org/session18/ and leave your reviews or comments there.

Thank you so much for listening and I wish you a fantastic day ahead. Bye.